M.Sc. Final Year
Physics, MP 08(A)

COMMUNICATION ELECTRONICS

मध्यप्रदेश भोज (मुक्त) विश्वविद्यालय – भोपाल

MADHYA PRADESH BHOJ (OPEN) UNIVERSITY - BHOPAL

COURSE WRITERS

Dr Samiksha Sirakwar, Assistant Professor, Integrated Basic Science, School for Physical and Decision Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow.

Dr Bhupendra Pratap Singh, Former Faculty Deptt. of Physics University of Lucknow.

Units (1.0-1.6.1, 1.10, 1.12, 2.2.5, 2.3.1, 2.3.2, 3.2, 3.3.5-3.3.7, 5.2.5, 5.3.3 - 5.3.6, 5.5-5.7)

Vijai Singh, Lecturer, IMS Engineering College, Ghaziabad.

Varsha Singh, Lecturer, JSSATE, Noida

Units (1.9)

Rajneesh Agrawal, Senior Scientist, Deptt. Of Information Technology, Govt of India, New Delhi

Units (2.2, 3.3-3.3.2, 5.3, 5.4)

Dr. Archana Mahamuni, Associate Professor, Marketing and Operations, Area Chair- Marketing, IILM University Gurugram, Haryana

Units (1.7, 1.8, 1.11, 1.13-1.17, 2.0-2.1, 2.2.1-2.2.4, 2.2.6-2.2.8, 2.3, 2.3.3-2.3.9, 2.4-2.8, 3.0-3.1, 3.3.3, 3.3.4, 3.8-3.12, 5.0-5.2, 5.2.1,, 5.2.2, 5.2.3, 5.2.4, 5.2.6, 5.3.1, 5.8-5.12)

Neha Gupta, Assistant Professor, Department of Electronics and Communication Engineering, Dev Bhoomi Group of Institutions, Dehradun, Uttarakhand

Unit (4)

Published by Registrar, MP Bhoj (Open) University, Bhopal in 2020

VIKAS®

Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

SYLLABI-BOOK MAPPING TABLE

Communication Electronics

Syllabi	Mapping in Book
UNIT: 1 **Signal and Noise:** Fourier Series, Sampling Function, Response of Linear System, Normalized Power, Normalized Power in Fourier Expansion, Power Spectral Density, Relationship between Fourier and Laplace Transforms, Transformation Theorems, Dirac Delta Function, Gate Function, Energy Density Function. Convolution, Impulse Response, Convolution Integral, Physical Interpretation on Convolution, Parseval's Theorem. **Noise:** Physical Sources of Noise, Shot Noise, 1 /f Noise, Thermal Noise, External Noise, Internal Noise. Noise Calculations, Noise Temperature, Noisy Two Port Network, Interference and Noise in AM & Pulse Modulation - Pre-Emphasis and De-Emphasis, Demodulation in the Presence of Noise.	**Unit-1:** Signal and Noise: Fourier Transforms and Convolution **(Pages 3-78)**
UNIT 2 **Amplitude Modulation:** Necessity of Modulation, Principle of Amplitude Modulation, Modulation Index, Power Relation, Multitone Modulation, AM Wave Generation, AM Square Law Modulator, Switching Modulator. **Demodulation of AM:** Synchronous Detection - Nonlinear Demodulation, Suppress Carrier AM Demodulator, Envelope Detector, Square Law Demodulator, DSB-SC and SSB Modulation Systems, Sideband and Carrier Power, Method of Generation and Detection of DSB-SC and SSB, Independent SideBand (ISB) System, Vestigial SideBand (VSB) Modulation.	**Unit-2:** Amplitude Modulation and Demodulation **(Pages 79-141)**
UNIT 3 **Sampling and Analogue Pulse Modulation:** Sampling Theory, Sampling Analysis, Ideal Sampling, Idea Reconstruction, Natural Sampling, Major Problems in Practical Sampling. Types of Analogue Pulse Modulation, Pulse Modulation Characteristics, Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PDM): Analysis, Generation and Recovery of PDM. Pulse Position Modulation (PPM): Analysis, Generation and Recovery. Comparison of PDM and PPM. Signal to Noise Ratio in Pulsed System (PAM, PDM & PPM).	**Unit-3:** Sampling and Analogue Pulse Modulation **(Pages 143-181)**
UNIT 4 **Microwave Devices and Communication:** Klystron, Magnetron and Travelling Wave Vibes, Velocity Modulation: Basic Principles of Two Cavity Klystron and Reflex Klystron Principles of Operation of Magnetrons. **Transferred Electron Devices:** Gunn Effect, Principles of Operation, Modes of Operation, IMPATT Diode, TRAPATT Diode. Advantages and Disadvantages of Microwave Transmission, Loss in Free Space, Propagation of Microwaves, Atmospheric Effects of Propagation. Freshed Zone Problem, Ground Reflection, Fading Sources, Detectors, Components, Antennas used in Microwave Communication Systems.	**Unit-4:** Microwave Transmission, Communication and Transferred Electron Devices **(Pages 183-214)**
UNIT 5 **Digital Communication:** Advantages and Disadvantages of Digital Communication, Bit Transmission, Signaling Rate, Error Probability, Digital Filtering, Delta Modulation, Pulse Code Modulation (PCM), PCM Generation, Binary Coding, PCM Bandwidth, PCM Reception Noise: Quantization Noise Analysis, SIN Ratio and Channel Capacity of PCM. **Code:** Error Detection and Correction Codes, Digital Carrier Systems, Teleprinters and Telegraphs Circuits, Radio Telegraphs Transmitters.	**Unit-5:** Digital Communications: PCM, Codes and Telegraphs **(Pages 215-283)**

CONTENTS

INTRODUCTION

NOTES

In telecommunication, Communications Electronics (CE) is the specialized field concerned with the use of electronic devices and systems for the acquisition or acceptance, processing, storage, display, analysis, protection, disposition, and transfer of information.

Communications Electronics (CE) includes the wide range of responsibilities and actions relating to, electronic devices and systems used in the transfer of ideas and perceptions; electronic sensors and sensory systems used in the acquisition of information devoid of semantic influence; electronic devices and systems intended to allow friendly forces to operate in hostile environments and to deny to hostile forces the effective use of electromagnetic resources.

In signal processing, noise is a general term for unwanted (and, in general, unknown) modifications that a signal may suffer during capture, storage, transmission, processing, or conversion. Sometimes the word is also used to mean signals that are random (unpredictable) and carry no useful information; even if they are not interfering with other signals or may have been introduced intentionally, as in comfort noise. A Fourier series is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic.

Amplitude Modulation (AM) is a modulation technique used in electronic communication, most commonly for transmitting messages with a radio wave. 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. AM was the earliest modulation method used for transmitting audio in radio broadcasting. It was developed during the first quarter of the 20th century beginning with Roberto Landell de Moura and Reginald Fessenden's radiotelephone experiments in 1900.

Pulse Amplitude Modulation (PAM) is an analog modulating scheme in which the amplitude of the pulse carrier varies proportional to the instantaneous amplitude of the message signal. In natural PAM, a signal sampled at the Nyquist rate is reconstructed, by passing it through an efficient Low Pass Frequency (LPF) with exact cutoff frequency. Fundamentally, PAM is a form of signal modulation where the message information is encoded in the amplitude of a series of signal pulses. It is an analog pulse modulation scheme in which the amplitudes of a train of carrier pulses are varied according to the sample value of the message signal. Demodulation is performed by detecting the amplitude level of the carrier at every single period.

Microwave is a form of electromagnetic radiation with wavelengths ranging from about one meter to one millimeter corresponding to frequencies between 300 MHz and 300 GHz, respectively. The prefix micro- in microwave is not meant to suggest a wavelength in the micrometer range. Rather, it indicates that microwaves are 'Small' (having shorter wavelengths), compared to the radio waves used prior to microwave technology. Data transmission and data reception or, more broadly, data communication or digital communications is the transfer and reception of data in the form of a digital bitstream or a digitized analog signal over a point-to-point or point-to-multipoint communication channel. The data are represented as an electromagnetic signal, such as an electrical voltage, radio wave, microwave, or infrared signal.

NOTES

This book, *Communications Electronics*, is divided into five units, which will help to understand the basic concepts of signal and noise, Fourier series, sampling function, response of linear system, normalized power, normalized power in Fourier expansion, transformation theorems, Dirac delta function, convolution, Parseval's theorem, shot noise, 1/f noise, thermal noise, external noise, internal noise. noise calculations, noise temperature, noisy two port network, Amplitude Modulation (AM), necessity of modulation, principle of amplitude modulation, modulation index, multitone modulation, AM wave generation, demodulation of AM, suppress carrier AM demodulator, envelope detector, square law demodulator, DSB-SC and SSB modulation systems, sampling and analogue pulse modulation, sampling theory, sampling analysis, types of analogue pulse modulation, microwave devices and communication, basic principles of two cavity Klystron and reflex Klystron principles of operation of Magnetrons, transferred electron devices, digital communication, advantages and disadvantages of digital communication, bit transmission, signaling rate, delta modulation, Pulse Code Modulation (PCM), error detection and correction codes, teleprinters and telegraphs circuits. The book follows the Self-Instruction Mode or the SIM format wherein each unit begins with an 'Introduction' to the topic followed by an outline of the 'Objectives'. The content is presented in a simple and structured form interspersed with Answers to 'Check Your Progress' for better understanding. A list of 'Summary' along with a 'Key Terms' and a set of 'Self-Assessment Questions and Exercises' is provided at the end of each unit for effective recapitulation.

UNIT 1 SIGNAL AND NOISE: FOURIER TRANSFORMS AND CONVOLUTION

Structure

1.0 INTRODUCTION

A signal is a function that conveys information about a phenomenon. In electronics and telecommunications, it refers to any time varying voltage, current, or electromagnetic wave that carries information. A signal may also be defined as an observable change in a quality such as quantity. As per the 'Signals and Systems' classification, the signals can be classified based on different standard criteria, namely according to the different feature of values, classified into analog signals and digital signals; according to the determinacy of signals, classified into deterministic signals and random signals; according to the strength of signals, classified into energy signals and power signals.

Fundamentally, the two main types of signals are 'Analog Signal' and 'Digital Signal'. Digital signals are quantized, while analog signals are continuous. An analog signal is any continuous signal for which the time varying feature of the signal is a representation of some other time varying quantity, i.e., analogous to another time varying signal. It differs from a digital signal, in which the continuous quantity is a representation of a sequence of discrete values which can only take on one of a finite number of values. A digital signal is a signal that is constructed from a discrete set of waveforms of a physical quantity to represent a sequence of discrete values. A logic signal is a digital signal with only two possible values and describes an arbitrary bit stream.

The Fourier series is named in honour of Jean-Baptiste Joseph Fourier (1768–1830), who made significant contributions to the study of trigonometric series. Fundamentally, the 'Fourier Series' is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic. As such, the summation is a synthesis of another function. The discrete-time Fourier transform is an example of Fourier series. The process of deriving weights that describe a given function is a form of Fourier analysis. For functions on unbounded intervals, the analysis and synthesis analogies are Fourier transform and inverse transform. The Fourier series has many such applications in electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing, quantum mechanics, econometrics, shell theory, etc.

Convolution is, basically, a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other. The term convolution refers to both the result function and to the process of computing it. Convolution is, therefore, a mathematical way of combining two signals to form a third signal. Characteristically, the single is referred as the most significant technique in digital signal processing. Using the strategy of impulse decomposition, systems are described by means of a signal called the impulse response. Convolution is important because it relates the three signals of concern, namely the input signal, the output signal, and the impulse response.

In signal processing, 'Noise' is a general term for unwanted and, in general, unknown modifications that a signal may suffer during capture, storage, transmission, processing, or conversion. Sometimes the word is also used to mean signals that are random (unpredictable) and carry no useful information; even if they are not interfering with other signals or may have been introduced intentionally, as in comfort noise. The noise measures have been typically defined to measure noise in signal processing, specifically in absolute terms, relative to some standard noise level, or relative to the desired signal level.

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. Modulation, therefore, means altering some aspect of a continuous wave carrier signal with an information-bearing modulation waveform, such as an audio signal which represents sound, or a video signal which represents images. The carrier wave, which has a much higher frequency than the message signal, carries the information. At the receiving station, the message signal is extracted from the modulated carrier by demodulation. Amplitude Modulation (AM) is a modulation technique used in electronic communication, most commonly for transmitting messages with a radio wave. Pulse Amplitude Modulation (PAM) is a form of signal modulation where the message information is encoded in the amplitude of a series of signal pulses.

In this unit, you will study about the signals, Fourier series, sampling function, response of linear system, normalized power, normalized power in Fourier expansion, power spectral density, relationship between Fourier and Laplace transforms, transformation theorems, Dirac Delta function, gate function, energy density function. convolution, impulse response, convolution integral, physical interpretation on convolution, Parseval's theorem, noise, physical source of noise, external noise, internal noise, shot noise, 1/f noise, thermal noise, noise calculations, noise temperature, noisy two port network, interference and noise in AM and Pulse modulation, pre-emphasis and de-emphasis, and demodulation in the presence of noise.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of signals and Fourier series
- Explain sampling function and response of a linear system
- Discuss about the normalized power and the normalized power in Fourier expansion
- Define what power spectral density is
- Analyse the relationship between Fourier and Laplace transforms
- Describe transformation theorems, Dirac Delta function and gate functions
- Elaborate on energy density function
- Explain convolution impulse response, convolution integral and physical interpretation on convolution
- Comprehend on the Parseval's theorem
- Understand the basic concept of noise and physical sources of noise
- Define the various types of noise, such as external noise and internal noise, shot noise, 1/f noise or flicker noise and thermal noise
- Do the noise calculations and measure the noise temperature
- Discuss the noisy two port network
- Understand noise interference in Amplitude Modulation (AM) and Pulse Modulation (PM)
- Explain how demodulation happens in the presence of noise
- Know what pre-emphasis and de-emphasis are

NOTES

1.2 FOURIER SERIES

A Fourier series is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic. As such, the summation is a synthesis of another function. The discrete time Fourier transform is an example of Fourier series. The process of deriving weights that describe a given function is a form of Fourier analysis. For functions on unbounded intervals, the analysis and synthesis analogies are Fourier transform and inverse transform.

Fourier originally defined the Fourier series for real valued functions of real arguments and using the sine and cosine functions as the basis set for the decomposition. A Fourier series, however, can be used only for periodic functions, or for functions on a bounded (compact) interval.

A non-sinusoidal periodic function can be expressed by a sum of a set of sinusoidal oscillating fruitions or periodic functions, i.e., mainly sine and cosine functions or complex exponentials. The expansion of the non-sinusoidal periodic function is known as the 'Fourier Series.'

Or

The Fourier series expansion of every piecewise smooth function expressed on a finite interval There are many different ways to express a function in Fourier space, the most common being by expanding it into a series of sines and cosines.

i.e.,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx$$

The coefficients a_o, a_n and bn are the Fourier coefficients of function $f(x)$.

Fourier analysis was founded on the idea of using Fourier series to represent and analyze periodic phenomena, and then using the Fourier transform to extend those insights to non-periodic phenomena. The Fourier series can be transformed into a Fourier transform by considering nonperiodic events (and consequently nearly any generic function) as a limit case of periodic occurrences as the period approaches infinity.

Formulating Fourier series, in sines and cosines, came to be seen as being reformulated in terms of orthogonality, linear operators, and Eigen functions as a more comprehensive framework. Eigenfunction expansions of differential equation solutions were born from this, which has since become a common approach in a wide range of fields and applications. The Fourier transform has become the basis for defining the objects of study in modern partial differential equations, while yet remaining a tool for solving individual problems. Much of this progress is based on the amazing relationship between Fourier transforms and convolution, which was also seen in the early days of the Fourier series. Mathematicians were forced (by engineers and physicists) to reassess how general the notion of 'Function' may be, and what kinds of functions can be — and should be — accepted into the operating theatre of calculus, to enable the techniques to be applied with increasing generality. Both differentiation and integration were generalized in the service of Fourier analysis.

Several techniques employ Fourier analysis with the symmetries of the objects under investigation. This may bring to mind, crystals and crystallography, and it would be correct, but mathematicians are more likely to think of number theory and Fourier analysis on groups.

Periodic Functions

If the values of a function repeat after an equal interval of x, the function is said to be periodic.

i.e., $f(x) = f(x + \mathrm{T}) = f(x + 2\mathrm{T}) = f(x + 3\mathrm{T})$……...

$f(x)$is therefore said to be the periodic fraction of x with period T.

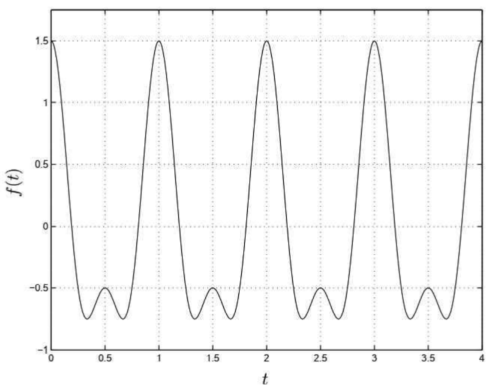
Such as:

(i) sin x, cos x, sec x, cosec x: T = 2π

(ii) tan x, cot x: T = À

(iii) sin (5x+2À) = sin 5x = sin 5(x + 2À/5): time period T = 2À/5.

Example 1: Consider the function $f(t)$ = cos 2$\grave{A}t$ + 1/2cos 4$\grave{A}t$ and the corresponding graph as shown below.



Individual terms have periods of 1 and 1/2, respectively, but the sum is periodic and has a period of 1:

$$f(t+1) = \cos 2\pi(t+1) + \tfrac{1}{2}\cos 4\pi(t+1)$$
$$= \cos(2\pi t + 2\pi) + \tfrac{1}{2}\cos(4\pi t + 4\pi) = \cos 2\pi t + \tfrac{1}{2}\cos 4\pi t = f(t)\,.$$

There is no smaller value of T for which $f(t + \mathrm{T}) = f(t)$. The overall pattern repeats every 1 second, but if this function represents some kind of wave would it have frequency 1 Hz? Somehow, it is not so. It has one *period* but one would probably say that it has, or contains, two frequencies, one cosine of frequency 1 Hz and one of frequency 2 Hz.

A periodic function $f(x)$ can be expanded into a Fourier series consisting of the terms:

(i) $\boldsymbol{a_o}$**:** A constant term or DC component.

(ii) $\boldsymbol{a_1 \cos x}$ **and** $\boldsymbol{b_1 \sin x}$**:** Components at fundamental frequency.

(iii) ***a_2 cos 2x, a_3 cos 3x,….. and b_2 sin 2x, b_3 sin 3x….:*** Harmonics, having frequencies multiple of fundamental frequency.

Note: When a function and its derivatives are continuous, then the function can be expanded by power of x by Taylor series. However, under specific conditions, one may expand both continuous and discontinuous forms of functions employing Fourier series.

Dirichlet's Condition for a Fourier Series

Suppose a function $f(x)$ is defined in the interval (a, b) and if,

(i) It is single valued in the interval (a, b)

(ii) It is bounded in the interval (a, b)

(iii) In the interval, there are only a finite number of maxima and minima (a, b)

(iv) In the interval, there are only a finite number of discontinuities (a, b) (v) $f(x + T) = f(x)$ where $[T = b \; a]$, for all values of x outside $[a, b]$ then

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx$$

Useful Integrals

Following are some significant integrals.

(i) $\int_0^{2\pi} \sin nx dx = 0$ (ii) $\int_0^{2\pi} \cos nx dx = 0$ (iii) $\int_0^{2\pi} \sin^2 nx dx = \pi$

(iv) $\int_0^{2\pi} \cos^2 nx dx = \pi$ (v) $\int_0^{2\pi} \sin nx \sin mx \, dx = 0 \quad [m \neq n]$

(vi) $\int_0^{2\pi} \cos nx \cos mx \, dx = 0 \; [m \neq n]$ (vii) $\int_0^{2\pi} \sin nx \cos mx \, dx = \int_0^{2\pi} \sin nx \cos nx \, dx = 0$

(viii) $\sin n\pi = 0, \cos n\pi = (-1)^n$.

Evaluation of Fourier Coefficient

The Fourier coefficient can be evaluated as follows.

Case 1: Suppose a function $f(x)$ defined in the interval $(0, 2\pi)$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx$$

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)dx; \quad a_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos nx\,dx; \quad b_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin nx\,dx$$

Case 2: Suppose a function $f(x)$ defined in the interval $(-\pi, \pi)$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx; \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\, dx; \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\, dx$$

Case 3: Suppose a function $f(x)$ defined in the interval $(0, 2l)$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{1} + \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{1}$$

$$a_0 = \frac{1}{1}\int_{0}^{21} f(x)dx; \quad a_n = \frac{1}{l}\int_{0}^{2l} f(x)\cos\frac{n\pi x}{l}dx; \quad b_n = \frac{1}{l}\int_{0}^{2l} f(x)\sin\frac{n\pi x}{l}dx$$

Case 4: Suppose a function $f(x)$ defined in the interval $(-l, l)$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi x}{1} + \sum_{n=1}^{\infty} b_n \sin\frac{n\pi x}{1}$$

$$a_0 = \frac{1}{l}\int_{-l}^{l} f(x)dx; \quad a_n = \frac{1}{1}\int_{-1}^{1} f(x)\cos\frac{n\pi x}{1}dx; \quad b_n = \frac{1}{1}\int_{-1}^{1} f(x)\sin\frac{n\pi x}{1}dx$$

The Fourier Series' Nature for Even and Odd Functions

Case I: Even Function: A function is said to be even if $f(-x) = f(x)$.

For an even function:

$$\int_{-\pi}^{\pi} f(x)dx = 2\int_{0}^{\pi} f(x)dx$$

$$a_0 = \frac{2}{\pi}\int_{-\pi}^{\pi} f(x)\,dx = \frac{1}{\pi}\int_{0}^{\pi} f(x)\,dx$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = \frac{2}{\pi}\int_{0}^{\pi} f(x)\cos nx\,dx$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = 0$$

This series will contain only **cosine** terms, and the expansion of the Fourier series will be stated as,

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos nx$$

As a result, it is often referred to as the **Fourier Cosine Series.**

Case II: Even Function: A function is said to be even if $f(-x) = -f(x)$.

NOTES

For an odd function:

$$\int_{-\pi}^{\pi} f(x)dx = 0$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx = 0 \quad \cdots\cdots$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = 0$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{2}{\pi}\int_{0}^{\pi} f(x)\sin nx\,dx$$

There will be just **sine** terms in this series, and the Fourier series expansion can be expressed as ,

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx$$

It is also known as the **Fourier Sine Series** because of this**.**

Half Range Series

Let us consider that the function is defined in the interval $(0, \pi)$ and it is significant regardless of the function's form outside of $(0, \pi)$.

To get cosine half range series, we take $f(x)$ as an even function in the interval $(-\pi, \pi)$ then,

$$a_0 = \frac{2}{\pi}\int_{0}^{\pi} f(x)dx\,;\; a_n = \frac{2}{\pi}\int_{0}^{\pi} f(x)\cos nx\,dx\,;\; b_n = 0\,;$$

To obtain the sine half range series, we first assume that $f(x)$ is an odd function in the interval $(-\pi, \pi)$ then,

$$a_0 = 0;\; a_n = 0;\; b_n = \frac{2}{\pi}\int_{0}^{\pi} f(x)\sin nx\,dx\,;$$

Following is the Half-Range Cosine Series and Half-Range Sine Series:

$$\int_{-c}^{+c}[f(x)]^2\,dx = c\left\{\frac{1}{2}a_0^2 + \sum_{n=1}^{\infty}\left(a_n^2 + b_n^2\right)\right\}$$

Half-range cosine series : $\int_{0}^{c}[f(x)]^2\,dx = \frac{c}{2}\left[\frac{a_0^2}{2} + \sum_{n=1}^{\infty} a_n^2\right]$

Half-range sine series : $\int_{0}^{c}[f(x)]^2\,dx = \frac{c}{2}\left[\sum_{n=1}^{\infty} b_n^2\right]$

Full Range Fourier Series

The Fourier series is an infinite series expansion involving trigonometric functions. A periodic waveform $f(t)$ of Period $p = 2L$ has a Fourier series given by:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi t}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi t}{L}\right)$$

$$= \frac{a_0}{2} + a_1 \cos\left(\frac{\pi t}{L}\right) + a_2 \cos\left(\frac{2\pi t}{L}\right) + a_3 \cos\left(\frac{3\pi t}{L}\right) + \ldots + b_1 \sin\left(\frac{\pi t}{L}\right) + b_2 \sin\left(\frac{2\pi t}{L}\right)$$

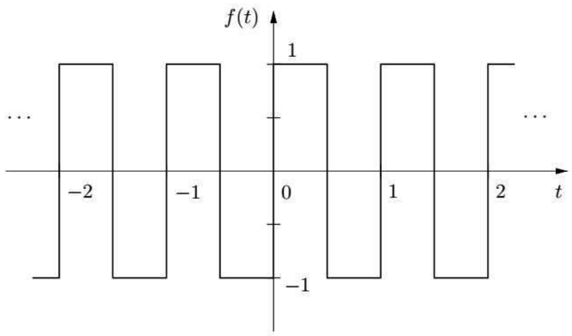
$$+ b_3 \sin\left(\frac{3\pi t}{L}\right) + \cdots$$

Where, a_n and b_n are the **Fourier coefficients**.

And, $a_0/2$ is the **mean value**, sometimes referred to as the **DC level**.

Example 3: **Fourier Series to calculate the Fourier Coefficients**

Consider a square wave of Period 1 as illustrated below in the figure**.**



Let us calculate the Fourier coefficients. The function is,

$$f(t) = \begin{cases} +1 & 0 \le t < \frac{1}{2} \\ -1 & \frac{1}{2} \le t < 1 \end{cases}$$

When the function is applied to the interval $0 \le t \le 1$, the zeroth coefficient represents the average value of the function.

and then extended to be periodic of Period 1. When the function is applied to the interval 0 *d" t d"* 1, the zeroth coefficient represents the average value of the function. Obviously, this is zero. For the other coefficients we have,

$$\begin{aligned} \hat{f}(n) &= \int_0^1 e^{-2\pi i n t} f(t)\, dt \\ &= \int_0^{1/2} e^{-2\pi i n t}\, dt - \int_{1/2}^{1} e^{-2\pi i n t}\, dt \\ &= \left[-\frac{1}{2\pi i n} e^{-2\pi i n t}\right]_0^{1/2} - \left[-\frac{1}{2\pi i n} e^{-2\pi i n t}\right]_{1/2}^{1} = \frac{1}{\pi i n}\left(1 - e^{-\pi i n}\right) \end{aligned}$$

We should thus consider the infinite Fourier series,

$$\sum_{n \neq 0} \frac{1}{\pi i n}\left(1 - e^{-\pi i n}\right) e^{2\pi i n t}$$

NOTES

We can phrase this in a more straightforward manner by first recognizing that

$$1 - e^{-\pi i n} = \begin{cases} 0 & n \text{ even} \\ 2 & n \text{ odd} \end{cases}$$

So, the series becomes,

$$\sum_{n \text{ odd}} \frac{2}{\pi i n} e^{2\pi i n t}.$$

Now combine the positive and negative terms and put them to work for you

$$e^{2\pi i n t} - e^{-2\pi i n t} = 2i \sin 2\pi n t.$$

Substituting this into the series and writing $n = 2k + 1$, our final answer is,

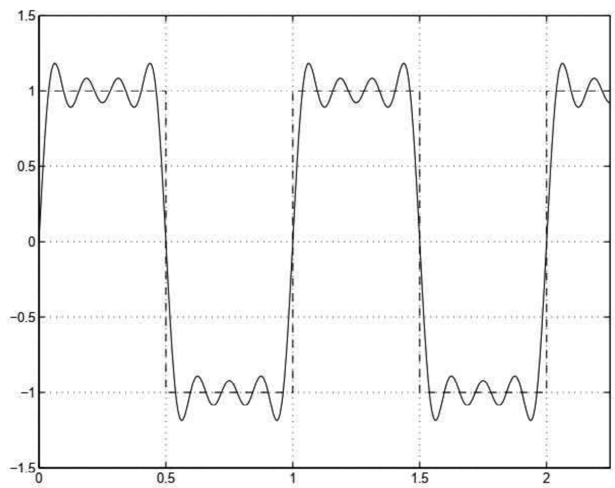
$$\frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \sin 2\pi(2k+1)t.$$

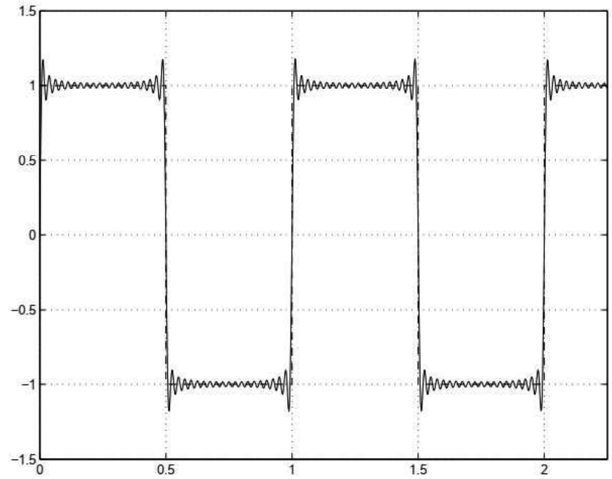
It is important to note that the function $f(t)$ is odd, which corresponds to the Fourier series having only sine terms.

What kind of series is this? In what sense does it converge, if at all, and to what does it converge, i.e., can we represent $f(t)$ as a Fourier series through the equation,

$$f(t) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \sin 2\pi(2k+1)t\,?$$

The sums of terms up to frequencies 9 and 39, respectively, are depicted in the graphs below.





A strange phenomenon can be seen. It can be certainly seen that the general shape is a square wave, but there is trouble at the corners. In hindsight, it is clear that it should not have expected a function like the square wave to be represented by a ***finite*** sum of complex exponentials. Because a finite sum of continuous functions is continuous, and the square wave has jump discontinuities, a finite sum of continuous functions is continuous. Thus, for maybe the first time, one of those theorems from calculus that seemed so pointless at the time makes an appearance.

The sum of two continuous functions or a finite number of them is continuous. Whatever may be concluded about a Fourier series representation for a square wave, it must contain arbitrarily high frequencies.

Why do Instruments Sound Different?

More precisely, why do two instruments sound different even when they are playing the same note? It is because the note they produce is not a single sinusoid of a single frequency, not the A at 440 Hz, for instance, but a sum (literally) of many sinusoids, each of which contributes a different amount. The complex wave that reaches your ear is the combination of many ingredients. Two instruments sound different because of the harmonics they produce and because of the strength of the harmonics.

Orthogonality: $\int_0^{\pi} \sin nx \, \sin kx \, dx = 0 \text{ if } n \neq k$

Explanation: Since $\sin nx \sin kx = \frac{1}{2} \cos (n-k)x - \frac{1}{2} \cos (n+k)x$

Orthogonality of the sines is demonstrated by integrating $\cos mx$ with $m = n - k$ and $m = n + k$, respectively.

Complex Fourier Series

There is a single formula for all the **complex coefficients** $\boldsymbol{c_k}$, rather than separate formulas for a_0, a_k, and b_k. $F(x)$ may also be a complex function (as in quantum mechanics). The complex infinite series for a 2π-periodic function:

$$F(x) = c_0 + c_1e^{ix} + c_{-1}e^{-ix} + \cdots = \sum_{n=-\infty}^{\infty} c_n e^{inx}$$

NOTES

Here e^{inx} can be combined with e^{-inx} into 2cos nx if every $c_n = -c_{-n}$. The cosine series is then applied to an even function. We use $e^{inx} - e^{-inx} = 2i \sin nx$ if every $c_n = -c_{-n}$. The sine series for an odd function follows, with the c's being pure imaginary.

To find c_k, multiply the previous equation by e^{-ikx} (not e^{ikx}), then integrate from $-\pi$ to π:

$$\int_{-\pi}^{\pi} F(x)e^{-ikx}dx = \int_{-\pi}^{\pi} c_0e^{-ikx}dx + \int_{-\pi}^{\pi} c_1e^{ix}e^{-ikx}dx + \cdots + \int_{-\pi}^{\pi} \boldsymbol{c_k e^{ikx} e^{-ikx} dx} + \cdots$$

The complex exponentials are orthogonal. Except for the period where $n = k$ and $e^{ikx} - e^{-ikx} = 1$, every integral on the right side is zero. The integral of 1 is 2π. That surviving term gives the formula for c_k:

The Fourier coefficients are,

$$\int_{-\pi}^{\pi} F(x)e^{-ikx}\,dx = 2\pi c_k \qquad \text{for} \quad k = 0, \pm 1, \ldots$$

1.3 SAMPLING FUNCTION

The process of converting a continuous time signal, typically not quantized, to a discrete time signal is known as sampling, usually quantized. It is also known as the discretization of the process of measuring the instantaneous values of a continuous time signal.

A sample is a portion of data that is continuous in the time domain and is taken from the entire data set. When a source creates an analogue signal that must be digitized, the signal must be discretized in time, using 1s and 0s, i.e., High or Low. Sampling is the process of discretizing an analogue signal.

A continuous time signal x-t and a sampled signal x_s-t are shown in Figure 1.1. The sampled signal x_s-t is obtained by multiplying x-t by a periodic impulse train.

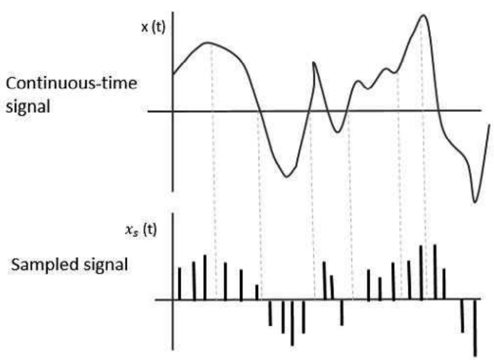


Fig. 1.1 *Sampling of a Signal*

Note: Sampling if done properly (Nyquist theorem is satisfied), does not introduce distortion.

Sampling Rate

The interval between the samples should be fixed in order to discretize the signals. A sampling period T_s can be used to describe this gap.

$$\text{Sampling Frequency} = \frac{1}{Ts} = f_s$$

Where, T_s denotes the sample time, while f_s denotes the sampling frequency or rate.

The reciprocal of the sample duration is the sampling frequency. The sample frequency is also known as the sampling rate. The sampling rate refers to the number of samples obtained per second or for a specific set of values.

The sampling rate should be taken into account when reconstructing an analogue signal from a digitized source. The sampling rate should be such that neither the data in the message signal is lost nor is it overlapped. As a result, a rate was established for this, known as the Nyquist rate.

Nyquist Rate

Assume a signal is band-limited, with no frequency components exceeding W Hertz. W stands for the highest frequency. For such a signal, the sampling rate should be twice the highest frequency for effective replication of the original signal, which means $f_S = 2W$, where f_S is the sampling rate and W is the highest frequency.

This sampling rate is referred to as the Nyquist rate.

Sampling Theorem

The sampling theorem, often known as the Nyquist theorem, establishes the theory of a sufficient sample rate in terms of bandwidth for the class of band-limited functions. According to the sampling theorem, "A signal can be accurately reproduced if it is sampled at a rate fs greater than twice its highest frequency W."

To illustrate this sampling theorem, take a band-limited signal, that is, one with a value greater than zero between some $-W$ and W Hertz.

Such a signal is represented as $x(f) = 0$ for $|f| > W$

Figure 1.2 depicts the frequency domain representation of the continuous time signal X-t.

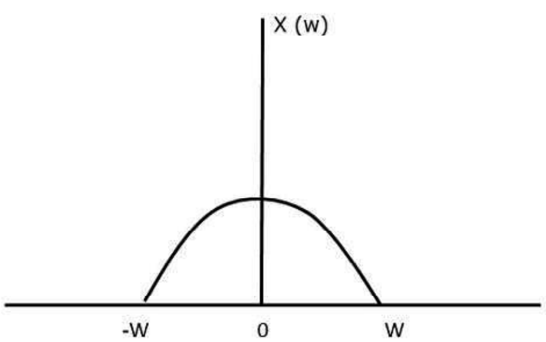


Fig. 1.2. *Band Limited Signal*

There is also a need for sample frequency, a frequency at which information is not lost even after sampling is necessary. As a result, a sampling rate of Nyquist rate, which is two times the maximum frequency, must be used. It is the sample rate that matters most.

It is possible to recover the original signal from *X-t* if it is sampled above the Nyquist rate, but not if it is sampled below it. Figure 1.3 depicts a signal if it is sampled at a higher rate than *2W* in the frequency domain.

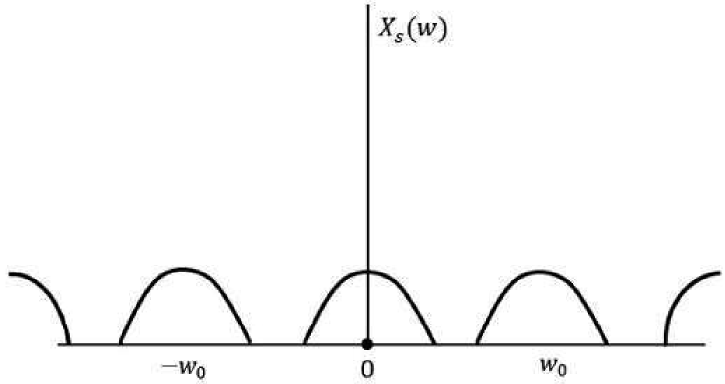


Fig. 1.3. *Fourier Transform of a Signal* X_s*-(t)*

Shown in Figure 1.3 is a signal's Fourier transform $X_s(t)$. As you can see, there is no loss of information here. Recovery is achievable because there is no muddled data.

In the case of $X_s(t)$, the Fourier Transform is,

$$X_s(w) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(w - nw_0)$$

Where T_s = Sampling Period and $w_0 = 2\pi/T_s$

When the sample rate is twice as high as the highest frequency, we will observe what happens (2*W*).

That means, $f_s = 2W$

In which f_s is the sampling rate, and *W* the greatest possible rate. Figure 1.4 depicts the end result. There is no loss of data as a result of this procedure. Because of this, this is a good sample rate.

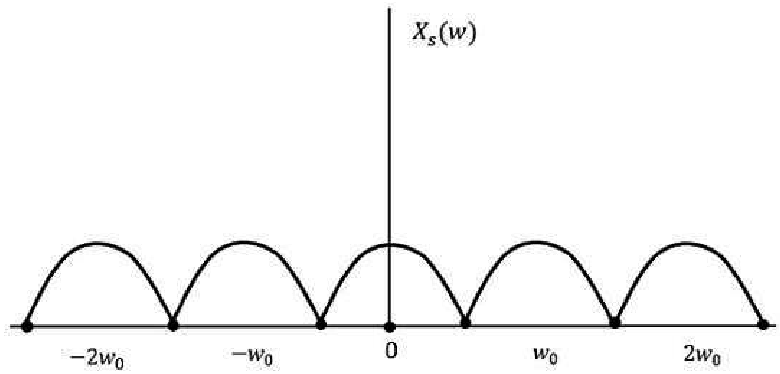


Fig. 1.4. *If fs* = 2*W, There is no Loss of Information*

NOTES

Now, let us look at the condition,

$f_s < 2W$

The resultant pattern will look like Figure 1.5.

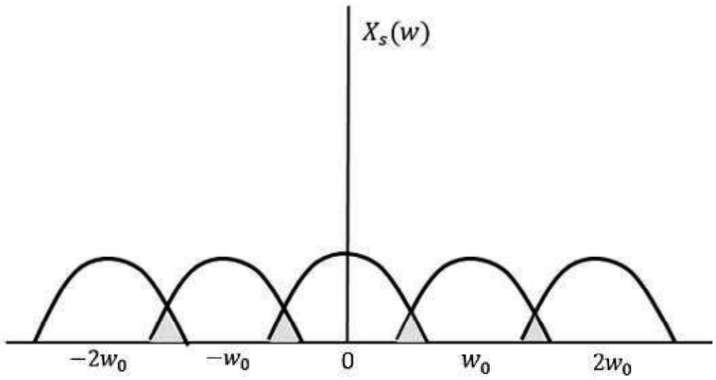


Fig. 1.5. *If fs < 2W, There is Loss of Information*

As illustrated in Figure 1.5, information is overlaid, resulting in data mixing and loss. Aliasing is the term used to describe this undesirable phenomenon of overlapping.

Aliasing

Aliasing is defined as, "The phenomenon in which a signal's high-frequency component assumes the identity of a low-frequency component in the spectrum of its sampled form".

To mitigate the effect of aliasing, the following precautions are taken:

(i) In the PCM transmitter section, a low pass anti-aliasing filter is used prior to the sampler to eliminate undesirable high frequency components.

(ii) After filtering, the signal is sampled at a rate slightly greater than the Nyquist rate.

(iii) Choosing a sampling rate greater than the Nyquist rate simplifies the design of the receiver's reconstruction filter.

1.4 RESPONSE OF LINEAR SYSTEM

Let us call the responses of a system to the inputs $x_1(t)$ and $x_2(t)$, $y_1(t)$ and $y_2(t)$, respectively. The system is said to have the additivity property if the response to the input $x_1(t) + x_2(t)$ is $y_1(t) + y_2(t)$ for any choice of $x_1(t)$ and $x_2(t)$. Let $y(t)$ represent a system's response to the input $x(t)$, and a represent a complex constant. If, for any choice of $x(t)$ and a, the response to the input $ax(t)$ is $ay(t)$, the system is said to possess the homogeneity property.

If a system possesses both the additivity and homogeneity properties, it is said to be linear. Otherwise, it is said to be nonlinear.

Consider a continuous Linear Time Invariant (LTI) system with impulse response $h(t)$. Assume that the system is causal and stable at all times. When this system is subjected to a continuous time random process $X(t)$, the output response is also a continuous time random process $Y(t)$. The linear system is called a discrete time system if the random processes X and Y are discrete time signals. The statistical and spectral features of the output random process $Y(t)$ are the focus of this unit.

System Response

As demonstrated in Figure 1.6, apply a random process $X(t)$ to a continuous linear time invariant system with an impulse response of $h(t)$. The output response $Y(t)$ is a random process as well. The convolution integral, $Y(t) = h(t) * X(t)$, can be used to express it. That is, the output response is $Y(t) = \int_{-\infty}^{\infty} h(\tau) X(t-\tau) d\tau$.

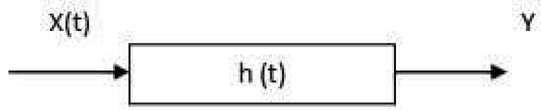


Fig. 1.6 *Schematic for a Signal Processing*

Mean Value of Output Response

Assume that the random process $X(t)$ is a stationary process in the broadest sense.

Mean value of output response = E [$Y(t)$], Then,

$$E[Y(t)] = E[h(t) * X(t)]$$

$$= E\left[\int_{-\infty}^{\infty} h(\tau)X(t-\tau)d\tau\right]$$

$$= \int_{-\infty}^{\infty} h(\tau)E[X(t-\tau)]d\tau$$

But $E[X(t-\tau)] = \bar{X}$ = Constant, since $x(t)$ is WSS.

Then $E[Y(t)] = \bar{Y} = \bar{X}\int_{-\infty}^{\infty} h(\tau)\, d\tau$. Also if $H(\omega)$ is the Fourier transform of $h(t)$ then $H(\omega) = \int_{-\infty}^{\infty} h(\tau)e^{-j\omega t}\, dt$. At $\omega = 0$, $H(0) = \int_{-\infty}^{\infty} h(t)\, dt$ is called the zero frequency response of the system. Substituting this we get $E[Y(t)] = \bar{Y} = \bar{X}\, H(0)$ is constant. Thus, the mean value of the output response $Y(t)$ of a WSS random process is equal to the product of the mean value of the input process and the zero frequency response of the system.

Mean square value of output response is,

$$E[Y^2(t)] = E[(h(t) * X(t))^2]$$

$$= E[(h(t) * X(t))\,(h(t) * X(t))]$$

$$= E\left[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)X(t-\tau_2)d\tau_2\right]$$

$$= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(t-\tau_1)X(t-\tau_2)h(\tau_1)h(\tau_2)d\tau_1 d\tau_2\right]$$

$$E[Y^2(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[X(t-\tau_1)X(t-\tau_2)]h(\tau_1)h(\tau_2)d\tau_1 d\tau_2$$

Where τ_1 and τ_2 are shifts in time intervals, If input $X(t)$ is a WSS random process then,

$$E\,[X(t-\tau_1)X(t-\tau_2)] = R_{XX}(\tau_1-\tau_2)$$

Therefore $E\,[Y^2(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau_1-\tau_2)\,h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2$

This expression is independent of time t. And it represents the Output power.

Here, WSS stands for Wavelength Selective Switch.

Autocorrelation Function of Output Response

The autocorrelation of Y (t) is,

$$R_{YY}(\tau_1,\tau_2) = E\,[Y\,(t_1)\;Y\,(t_2)$$

$$= E\,[(h\,(t_1) * X\,(t_1))\;(h\,(t_2) * X\,(t_2))]$$

$$= E\,[\int_{-\infty}^{\infty} h(\tau_1)X(t_1-\tau_1)d\,\tau_1 \int_{-\infty}^{\infty} h(\tau_2)X(t_2-\tau_2)d\,\tau_2]$$

$$= E\,[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(t_1-\tau_1)X(t_2-\tau_2)h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[\,X(t_1-\tau_1)X(t_2-\tau_2)]h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2$$

We know that,

$$E\,[X(t_1-\tau_1)X(t_2-\tau_2)] = R_{XX}(t_2-t_1+\tau_1-\tau_2).$$

If input $X(t)$ is a WSS random process, Let the time difference $\tau = t_1 - t_2$ and $t = t_1$. Then,

$E\,[X(t-\tau_1)X(t+\tau-\tau_2)] = R_{XX}(\tau+\tau_1-\tau_2)$. Then

$$R_{YY}(t,t+\tau) = R_{YY}(t,\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau+\tau_1-\tau_2)\,h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2$$

If $R_{XX}(\tau)$ is the autocorrelation function of $X(t)$, then $R_{YY}(\tau) = R_{XX}(\tau) * h(\tau)\,(-\tau)$

The output autocorrelation function is found to be a function of just τ. As a result, the output random process, $Y(t)$, is a WSS (Wavelength Selective Switch) random process as well.

If $X(t)$ is a WSS random process, then the cross-correlation function between $X(t)$ and $Y(t)$ is as follows:

$$R_{XY}(t,t+\tau) = E\,[X\,(t)\;Y\,(t+\tau)]$$

$$R_{XY}(\tau) = E\,[X\,(t)\int_{-\infty}^{\infty} h(\tau_1)\;X\,(t+\tau-\tau_1)\;d\tau_1]$$

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} E\,[X\,(t)\;X\,(t+\tau-\tau_1)]\;h(\tau_1)d\tau_1]$$

$R_{XY}(\tau) = \int_{-\infty}^{\infty} R_{XX}(\tau-\tau_1)]\;h(\tau_1)d\tau_1$ which is the convolution of $R_{XX}(\tau)$ and $h(\tau)$.

Therefore $R_{XY}(\tau) = R_{XX}(\tau) * h(\tau)$ similarly we can show that

$$R_{YX}(\tau) = R_{XX}(\tau) * h(-\tau)$$

This shows that $X(t)$ and $Y(t)$ are jointly WSS or Wavelength Selective Switch. Additionally, we may link the autocorrelation and cross-correlation functions as,

$$R_{YY}(\tau) = R_{XY}(\tau) * h(-\tau)$$

$$R_{YY}(\tau) = R_{YX}(\tau) * h(\tau)$$

Spectral Characteristics of a System Response

Consider the random process $X(t)$ to be a WSS (Wavelength Selective Switch) random process with the autocorrelation function (τ) applied via an LTI system. Notably, the output response $Y(t)$ is likewise a WSS, as are the processes $X(t)$ and $Y(t)$. The Fourier transform of the correlation functions can be used to derive the power spectrum characteristics of the output process $Y(t)$.

Power Density Spectrum of Response

Consider the application of a random process $X(t)$ to an LTI (Linear Time Invariant) system with a transfer function $H(\omega)$. $Y(t)$ is the output response. If the input process has a power spectrum of $S_{XX}(\omega)$, then the output response has a power spectrum of:

$$S_{YY}(\omega) = \left|H(\omega)\right|^2 S_{XX}(\omega)$$

Proof: Let $R_{yy}(\tau)$ be the output response's autocorrelation $Y(t)$. Then the response's power spectrum is the Fourier transform of $R_{yy}(\tau)$.

Therefore $S_{yy}(\omega) = F[S_{YY}(\omega)] = \int_{-\infty}^{\infty} R_{yy}(\tau) e^{-j\omega t} d\tau$

We know that,

$$R_{YY}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau + \tau_1 - \tau_2)\, h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2$$

Then,

$$S_{YY}(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_{XX}(\tau + \tau_1 - \tau_2)\, h(\tau_1)h(\tau_2)d\,\tau_1 d\tau_2\, e^{-j\omega\tau} d\tau$$

$$= \int_{-\infty}^{\infty} h(\tau_1) \int_{-\infty}^{\infty} h(\tau_2) \int_{-\infty}^{\infty} R_{XX}(\tau + \tau_1 - \tau_2)\, e^{-j\omega\tau} d\tau\, d\tau_2\, d\tau_1$$

$$= \int_{-\infty}^{\infty} h(\tau_1)e^{j\omega\tau_1} \int_{-\infty}^{\infty} h(\tau_2)e^{j\omega\tau_2} \int_{-\infty}^{\infty} R_{XX}(\tau + \tau_1 - \tau_2)\, e^{-j\omega\tau}e^{j\omega\tau_1}e^{j\omega\tau_2} d\tau\, d\tau_2\, d\tau_1$$

Let $\tau + \tau_1 - \tau_2 = t$, $d\tau = dt$

Therefor $S_{YY}(\omega) = \int_{-\infty}^{\infty} h(\tau_1)e^{j\omega\tau_1} d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)e^{j\omega\tau_2} d\tau_2 \int_{-\infty}^{\infty} R_{XX}(t)\, e^{-j\omega t} dt$

We know that $H(\omega) = \int_{-\infty}^{\infty} h(\tau)e^{-j\omega t}\, dt.$

Therefore $S_{YY}(\omega) = H^*(\omega)\, H(\omega)\, S_{XX}(\omega) = H(-\omega)H(\omega)\, S_{XX}(\omega)$

Therefore $S_{YY}(\omega) = S_{XX}(\omega)$. Hence proved.

In a similar fashion, we can demonstrate that the cross power spectral density function is,

$$S_{XY}(\omega) = S_{XX}(\omega)\, H(\omega) \text{ and } S_{YX}(\omega) = S_{XX}(\omega)\, H(-\omega)$$

NOTES

1.5 NORMALIZED POWER AND NORMALIZED POWER IN FOURIER EXPANSION

The concept of normalized power is defined on the basis of Parseval's theorem and normalized power in a periodic waveform which is often employed in 'Communication Engineering' and allied areas as a measure of signal strength. Typically, it is defined as the average power that can be transported specifically to 1 Ω resistance when the periodic waveform is considered as a voltage waveform applied to that specific resistor.

Parseval's Theorem and Normalized Power in a Periodic Waveform

The idea of normalized power in a periodic waveform is frequently used as a measure of signal intensity. It is defined as the average power given to a 1 Ω resistance when the periodic waveform is viewed as a voltage waveform applied to the resistor. The averaging is performed over any interval equal to the waveform's period is given by,

$$P_{\text{n}} = \frac{1}{T}\int_{-0.5T}^{0.5T} [v(t)]^2\, \mathrm{d}t. \qquad \ldots(1.1)$$

Characteristically, the normalized power in a periodic signal can be defined.

Normalized power of a waveform is equal to zeroth coefficient in the exponential Fourier series of its squared version.

If we think of $v_3(t) = [v(t)]^2$ as a new time-function, the term on the right side of Equation 1.1 can be identified as the DC component of $v_3(t)$. Hence, P_n must be equal to the exponential Fourier series coefficient of $v_3(t)$, $\tilde{\upsilon}_{3k}$ for $k = 0$.

The multiplication-in-time property of Fourier series states that if $v_1(t)$ and $v_2(t)$ are two periodic waveforms with same period and $v_3(t) = v_1(t) \times v_2(t)$, then, the exponential Fourier series coefficients of $v_3(t)$ is given by:

$$\tilde{v}_{3k} = \sum_{n=-\infty}^{\infty} \tilde{v}_{1n}\tilde{v}_{2(k-n)} \text{ for } -\infty < k < \infty,$$

where $\tilde{\upsilon}_{1\text{n}}$ and $\tilde{\upsilon}_{2\text{n}}$ are the exponential Fourier series coefficients of $v_1(t)$ and $v_2(t)$, respectively. We use this property with $v_1(t) = v_2(t) = v(t)$ and evaluate the exponential Fourier series coefficient of,

$$[v(t)]^2 \text{ for } k = 0 \text{ as } \tilde{v}_{30} = \sum_{n=-\infty}^{\infty} \tilde{v}_n\tilde{v}_{-n}.$$

NOTES

Therefore,

$$P_{\mathrm{n}} = \frac{1}{T}\int_{-0.5T}^{0.5T}[v(t)]^2 \mathrm{d}t = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_{-n} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_n^* = \sum_{n=-\infty}^{\infty} |\tilde{v}_n|^2 = |\tilde{v}_0|^2 + 2\sum_{n=1}^{\infty} |\tilde{v}_n|^2.$$

This is Parseval's Theorem on normalized power of periodic waveforms.

The trigonometric Fourier series for $v(t)$ is,

$$v(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t,$$

where $a_0 = \tilde{v}_0 = \frac{1}{T}\int_{-T/2}^{T/2} v(t)\mathrm{d}t,$

$$a_n = \tilde{v}_n + \tilde{v}_{-n} = \tilde{v}_n + \tilde{v}_n^* = 2\,\mathrm{Re}(\tilde{v}_n) = \frac{2}{T}\int_{-T/2}^{T/2} v(t)\cos n\omega_0 t\ \mathrm{d}t \text{ for } n = 1, 2, 3 ...$$

$$a_n = -\tilde{v}_n + \tilde{v}_{-n} = -\tilde{v}_n + \tilde{v}_n^* = 2\,\mathrm{Im}(\tilde{v}_n) = \frac{2}{T}\int_{-T/2}^{T/2} v(t)\sin n\omega_0 t\ \mathrm{d}t \text{ for } n = 1, 2, 3 ...$$

Therefore, Parseval's Theorem can be expressed in terms of trigonometric Fourier series coefficients as,

$$P_{\mathrm{n}} = a_0^2 + \sum_{n=1}^{\infty}\left(\frac{a_n^2 + b_n^2}{2}\right).$$

The second form of trigonometric Fourier series is shown below.

$$\therefore v(t) = c_0 + \sum_{n=1}^{\infty} c_n \cos(n\omega_0 t - \phi_n)$$

$$\text{where } c_0 = \tilde{v}_0, c_n = \sqrt{a_n{}^2 + b_n{}^2} = 2\tilde{v}_n\tilde{v}_n^* = 2|\tilde{v}_n|$$

$$\text{and } \phi_n = \tan^{-1}\frac{b_n}{a_n} = -\angle \text{ of } \tilde{v}_n \text{ for } n = 1, 2, 3...$$

$$\therefore P_{\mathrm{n}} = c_0^2 + \sum_{n=1}^{\infty}\frac{c_n^2}{2}$$

The normalized power of a periodic waveform $v(t)$, P_n, is given by,

$$P_{\mathrm{n}} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_{-n} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_n^* = \sum_{n=-\infty}^{\infty} |\tilde{v}_n|^2 = |\tilde{v}_0|^2 + 2\sum_{n=1}^{\infty} |\tilde{v}_n|^2$$
$$= a_0^2 + \sum_{n=1}^{\infty}\frac{(a_n^2 + b_n^2)}{2}$$
$$= c_0^2 + \sum_{n=1}^{\infty}\frac{c_n^2}{2}.$$

Parseval's Power Relation for a Periodic Waveform

Though the multiplication-in-time property easily led us to Parseval's theorem, it does not help us to see the significance of this theorem. Neither does it tell us how this total normalized power is distributed among various frequency components.

Hence, we use the trigonometric Fourier series

$v(t) = c_0 + \sum_{n=1}^{\infty} c_n \cos(n\omega_0 t - \phi_n)$ for further appreciation of P_n.

Consider a simpler situation in which $v(t)$ contains just three components.

$v(t) = c_0 + c_m \cos(m\omega_0 t - \phi_m) + c_k \cos(k\omega_0 t - \phi_k)$, k and m are integers.

$$\begin{aligned}\therefore [v(t)]^2 &= c_0^2 + c_m^2 \cos^2(m\omega_0 t - \phi_m) + c_k^2 \cos^2(k\omega_0 t - \phi_k)\\ &+2c_0 c_m \cos(m\omega_0 t - \phi_m) + 2c_0 c_k \cos(k\omega_0 t - \phi_k)\\ &+2c_m c_k \cos(m\omega_0 t - \phi_m)\cos(k\omega_0 t - \phi_k).\end{aligned}$$

$$\begin{aligned}\therefore [v(t)]^2 &= c_0^2 + \frac{1}{2}c_m^2 + \frac{1}{2}c_k^2 + \frac{1}{2}c_m^2 \cos 2(m\omega_0 t - \phi_m) + \frac{1}{2}c_k^2 \cos 2(k\omega_0 t - \phi_k)\\ &+2c_0 c_m \cos(m\omega_0 t - \phi_m) + 2c_0 c_k \cos(k\omega_0 t - \phi_k)\\ &+c_m c_k \cos[(m+k)\omega_0 t - (\phi_m + \phi_k)] + c_m c_k \cos[(m-k)\omega_0 t - (\phi_m - \phi_k)].\end{aligned}$$

k and m are integers. Thus, if k ‘“ m, the frequencies $m\omega_o$, $k\omega_o$, $2m\omega_o$, $2k\omega_o$, $(m-k)\omega_o$ and $(m+k)\omega_o$ are integer multiples of ω_o. Hence, all the cosine waves in $[v(t)]^2$ will have integer number of cycles in T_s, where T is the period of $v(t)$. Therefore, their average over one T will be zero.

$$\begin{aligned}\therefore P_n &= \frac{1}{T}\int_{-0.5T}^{0.5T}[v(t)]^2 dt = \frac{1}{T}[c_0^2 + \frac{1}{2}c_m^2 + \frac{1}{2}c_k^2]\times T\\ &= c_0^2 + \frac{1}{2}c_m^2 + \frac{1}{2}c_k^2 = c_0^2 + \left(\frac{c_m}{\sqrt{2}}\right)^2 + \left(\frac{c_k}{\sqrt{2}}\right)^2\end{aligned}$$

Parseval’s relation interpreted in terms of DC component and RMS (Root Mean Square) values of harmonic components.

Generalizing the result for infinite term Fourier series,

$$\therefore P_n = \frac{1}{T}\int_{-0.5T}^{0.5T}[v(t)]^2 dt = c_0^2 + \sum_{n=1}^{\infty}\left(\frac{c_n}{\sqrt{2}}\right)^2 \qquad ....(1.2)$$

i.e., P_n = (DC Component)^2 $P_n = \sum_{n=0}^{\infty}$ + (RMS Value of nth Harmonic Component)^2

Since the RMS value of a DC component is same as its value, we can express this as,

$$P_n = \sum_{n=0}^{\infty}(\text{r.m.s value of } n\text{th harmonic component})^2. \qquad ....(1.3)$$

Square root of this quantity will give the RMS value of $v(t)$ itself.

$$\text{R.M.S value of } v(t) = \sqrt{\sum_{n=1}^{\infty}(\text{r.m.s value of } n^{th} \text{ harmonic component})^2} \qquad ...(1.4)$$

The normalized power of a particular harmonic component with amplitude c_n when acting alone will be $0.5\, c_n^2$. Equation 1.2 shows that it contributes the same amount to the total power even when it is acting along with other harmonics.

R.M.S. Value of a Non-Sinusoidal Periodic Waveform

Each harmonic component in the trigonometric Fourier series of a waveform contributes to normalized power. We can assign the power contributed by a particular component to its frequency and plot the information against frequency or harmonic order as a line spectrum. This spectral plot is called the discrete power spectrum. However, it will be a single-sided spectrum since we derived it from trigonometric Fourier series. Spectral lines will be located at 0, ω_0, $2\omega_0$, $3\omega_0$, etc., and the length of the spectral line will be proportional to $0.5\, c_n^2$. By Parseval's theorem,

$$P_{\mathrm{n}} = \frac{1}{T}\int_{-0.5T}^{0.5T}\left[v(t)\right]^2 \mathrm{d}t = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_{-n} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_n^* = \sum_{n=-\infty}^{\infty} \left|\tilde{v}_n\right|^2 .$$

Does Power Obey Superposition Principle?

Consider two arbitrary waveforms $v_1(t)$ and $v_2(t)$. Let average of $[v_1(t)]^2$ and $[v_2(t)]^2$ over some interval be a_1 and a_2, respectively. Will the average of $[v_1(t) + v_2(t)]^2$ over the same interval be $a_1 + a_2$? The answer depends on whether the average of $2v_1(t)v_2(t)$ in that interval is zero or not. In general, it is not zero, and average of $[v_1(t) + v_2(t)]^2$ is not the same as the sum of averages of $[v_1(t)]^2$ and $[v_2(t)]^2$. Thus, the power does not obey superposition principle in general.

But, if $v_1(t)$ and $v_2(t)$ are two sinusoids with different frequencies, and, if their frequencies are integer multiples of some basic frequency, the average of $2v_1(t)v_2(t)$ in an interval that is equal to the period corresponding to the basic frequency, is zero.

Hence, if $v(t)$ is a mixture of harmonically related sinusoids and DC, the normalized.

Therefore, we can draw the two-sided discrete power spectrum by plotting two lines of $2\left|\tilde{v}_n\right|^2$ height proportional to at $n\omega_0$ and $-n\omega_0$. We had noted earlier that, two spectral components located at $\pm n\omega_0$ in the two-sided magnitude and phase spectra based on exponential Fourier series, have to be treated as an integral unit rather as $2\left|\tilde{v}_n\right|^2$ individual components. Those two components always go together and form a real sinusoid. Similarly, it is understood that the power spectral components located at $\pm n\omega_0$ in the two-sided power spectrum always go together to make a total contribution of to P_n.

Check Your Progress

1. Define the term Fourier series.
2. State the Fourier series expansion.
3. What is full range Fourier series?
4. What is sampling?
5. Define Nyquist rate.
6. When is the response of a system said to be linear?
7. State about the normalized power in a periodic signal.
8. What does multiplication-in-time property of Fourier series states? Why is this property used?

1.6 POWER SPECTRAL DENSITY

A Power Spectral Density (PSD) is a metric that compares the power content of a signal to its frequency. Broadband random signals are often described using a PSD. The spectral resolution used to digitize the signal is used to normalize the PSD's amplitude. PSD depicts the variation's **strength (energy)** as a function of frequency. In other words, it displays which frequencies have high variations and which frequencies have minor variations. PSD is measured in energy (variance) per frequency (width), and energy within a certain frequency range can be obtained by integrating PSD within that frequency range.

PSD is a very useful in determining the frequencies and amplitudes of oscillatory signals in time series data. For example, let assume you are operating a factory with many machines and some of them have motors inside. You detect unwanted vibrations from somewhere. You may be able to glean information about the location of offending machines by examining PSD, which contains information about vibration frequencies. PSD is still relevant in the absence of exclusively oscillatory signals in the data. For instance, if you have sales data from an ice cream parlor, you can obtain an approximation of the summer peak in sales by inspecting the PSD of your data. PSD is frequently computed and plotted to obtain a "feel" for data during the early stages of time series analysis. PSD is similar to a simple time series plot, except that we are looking at time series as a function of frequency rather than time. Frequency is a transformation of time, thus examining changes in the frequency domain is simply another approach to examine time series data variances. PSD indicates which frequency bands have significant variance, which can be highly valuable for subsequent study.

1.6.1 Energy Density Function

The energy spectral density quantifies the frequency dependence of the energy in a signal or time series. The term 'Energy' is used here in a broad sense of signal processing, i.e., the energy E of a signal $x(t)$ is as follows:

NOTES

$$E \triangleq \int_{-\infty}^{\infty} |x(t)|^2 \, dt.$$

The energy spectral density is best appropriate for transients—that is, signals with a finite total energy. Whether the signal is finite or not, Parseval's theorem (or Plancherel theorem) provides an alternative equation for the signal's energy:

$$\int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |\hat{x}(f)|^2 \, df,$$

where:

$$\hat{x}(f) \triangleq \int_{-\infty}^{\infty} e^{-i2\pi ft} x(t) \, dt$$

The Plancherel theorem (sometimes called the Parseval–Plancherel identity) is a result in harmonic analysis, proven by Michel Plancherel in 1910. It states that the integral of a function's squared modulus is equal to the integral of the squared modulus of its frequency spectrum.

This is the Fourier transform value of $x(t)$ at frequency f (in Hz). Additionally, the theorem holds true in discrete-time instances. Because the right-hand integral is the signal's energy, the integrand ^{2} can be read as a density function characterizing the energy contained in the signal at frequency f. As a result, the energy spectral density of $x(t)$ can be defined as follows:

$$\bar{S}_{xx}(f) \triangleq |\hat{x}(f)|^2$$

Wiener–Khinchin theorem states that the function $S_{xx}(f)$ and the autocorrelation of $x(t)$ generate a Fourier transform pair.

1.7 TRANSFORMATION THEOREMS

Following are the significant transformation theorems in mathematical analysis.

1.7.1 Fourier Transform

A Fourier Transform (FT) is a mathematical transform that decomposes functions depending on space or time into functions depending on spatial or temporal frequency, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes. The term Fourier transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of space or time.

The Fourier transform of a function of time is a complex valued function of frequency, whose magnitude (absolute value) represents the amount of that frequency present in the original function, and whose argument is the phase offset of the basic sinusoid in that frequency. The Fourier transform is not limited to functions of time, but the domain of the original function is commonly referred to as the time domain. There is also an inverse Fourier transform that mathematically synthesizes the original function from its frequency domain representation, as proven by the Fourier inversion theorem.

Fourier transform of function $f(t)$ is given by,

$$F\,f(t) = F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt$$

Properties

1. **Linearity Property**

   If $f(t) \leftrightarrow F(\omega)$ and $y(t) \leftrightarrow Y(\omega)$

   Then linearity property states that

   $af(t) + by(t) \leftrightarrow aF(\omega) + bY(\omega)$

2. **Time Shifting Property**

   If $f(t) \leftrightarrow F(\omega)$

   Then Time shifting property states that

   $f(t - t_0) \leftrightarrow e^{-j\omega t_0}F(\omega)$

   $f(t+t_0) \leftrightarrow e^{j\omega t_0}F(\omega)$

3. **Frequency Shifting Property**

   If $f(t) \leftrightarrow F(\omega)$

   Then frequency shifting property states that

   $e^{j\omega_0 t}.f(t) \leftrightarrow F(\omega - \omega_0)$

   $e^{-j\omega_0 t}.f(t) \leftrightarrow F(\omega + \omega_0)$

4. **Time Reversal Property**

   If $f(t) \leftrightarrow F(\omega)$

   Then Time reversal property states that

   $f(-t) \leftrightarrow F(-\omega)$

5. **Time Scaling Property**

   If $f(t) \leftrightarrow F(\omega)$

   Then Time scaling property states that

   $f(at) \leftrightarrow (1/a)F(\omega/a)$

6. **Differentiation and Integration Properties**

   If $f(t) \leftrightarrow F(\omega)$

   Then Differentiation property states that

   $df(t)/dt \leftrightarrow j\omega.F(\omega)$

   $d^n f(t)/dt^n \leftrightarrow (j\omega)^n.F(\omega)$

   Time Integration property states that

   $\int f(t)dt \leftrightarrow (1/j\omega)F(\omega)$

7. **Multiplication and Convolution Properties**

   If $f(t) \leftrightarrow F(\omega)$ & $y(t) \leftrightarrow Y(\omega)$

Then multiplication property states that

$f(t).y(t) \leftrightarrow (1/2\pi)F(\omega)*Y(\omega)$

Convolution property states that

$f(t)*y(t) \leftrightarrow F(\omega).Y(\omega)$

Fourier Transform of some elementary functions:

1	$e^{-at}u(t)$	$\frac{1}{a+j\omega}$	$a > 0$
2	$e^{at}u(-t)$	$\frac{1}{a-j\omega}$	$a > 0$
3	$e^{-a\vert t\vert}$	$\frac{2a}{a^2+\omega^2}$	$a > 0$
4	$te^{-at}u(t)$	$\frac{1}{(a+j\omega)^2}$	$a > 0$
5	$t^n e^{-at}u(t)$	$\frac{n!}{(a+j\omega)^{n+1}}$	$a > 0$
6	$\delta(t)$	1	
7	1	$2\pi\delta(\omega)$	
8	$e^{j\omega_0 t}$	$2\pi\delta(\omega-\omega_0)$	
9	$\cos\omega_0 t$	$\pi[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]$	
10	$\sin\omega_0 t$	$j\pi[\delta(\omega+\omega_0)-\delta(\omega-\omega_0)]$	
11	$u(t)$	$\pi\delta(\omega)+\frac{1}{j\omega}$	
12	$\operatorname{sgn} t$	$\frac{2}{j\omega}$	
13	$\cos\omega_0 t\, u(t)$	$\frac{\pi}{2}[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]+\frac{j\omega}{\omega_0^2-\omega^2}$	
14	$\sin\omega_0 t\, u(t)$	$\frac{\pi}{2j}[\delta(\omega-\omega_0)-\delta(\omega+\omega_0)]+\frac{\omega_0}{\omega_0^2-\omega^2}$	
15	$e^{-at}\sin\omega_0 t\, u(t)$	$\frac{\omega_0}{(a+j\omega)^2+\omega_0^2}$	$a > 0$
16	$e^{-at}\cos\omega_0 t\, u(t)$	$\frac{a+j\omega}{(a+j\omega)^2+\omega_0^2}$	$a > 0$
17	$\operatorname{rect}\left(\frac{t}{\tau}\right)$	$\tau\operatorname{sinc}\left(\frac{\omega\tau}{2}\right)$	
18	$\frac{W}{\pi}\operatorname{sinc}(Wt)$	$\operatorname{rect}\left(\frac{\omega}{2W}\right)$	
19	$\Delta\left(\frac{t}{\tau}\right)$	$\frac{\tau}{2}\operatorname{sinc}^2\left(\frac{\omega\tau}{4}\right)$	
20	$\frac{W}{2\pi}\operatorname{sinc}^2\left(\frac{Wt}{2}\right)$	$\Delta\left(\frac{\omega}{2W}\right)$	
21	$\sum_{n=-\infty}^{\infty}\delta(t-nT)$	$\omega_0\sum_{n=-\infty}^{\infty}\delta(\omega-n\omega_0)$	$\omega_0=\frac{2\pi}{T}$
22	$e^{-t^2/2\sigma^2}$	$\sigma\sqrt{2\pi}e^{-\sigma^2\omega^2/2}$	

[A] Fourier Sine Transforms.

They can be subdivided in two, namely, the infinite Fourier sine transform and the Finite Fourier sine transforms.

[a_1] *The Infinite Fourier sine Transform* of a function $F(x)$ of x such that $0 < x < \infty$ is denoted by $f_s\,(n)$, n being a positive integer and is defined as

$$f_s\,(n) = \int_0^\infty F(x) \sin nx\, dx$$

Here the *Inverse Fourier sine transform of* $f_s\,(n)$ and defined as

$$F(x) = \frac{2}{\pi}\int_0^\infty f_s\,(n) \sin nx\, dx$$

Thus if $f_s\,(n) = f_s\,[F(x)]$, then $F(x) = f_s^{-1}\,[f_s\,(n)]$

where f is the symbol for Fourier transform and f^{-1} for its inverse.

[B] Fourier Cosine Transforms.

They can also be subdivided into two, namely, Infinite and Finite cosine transforms.

[b_1] *The Infinite Fourier Cosine Transform* of $F\,(x)$ for $0 < x < \infty$, is defined as

$$f_c\,(n) = \int_0^\infty F(x) \cos nx\, dx, n \text{ being a positive integer.}$$

Here the function $F\,(x)$ is called as the *Inverse cosine transform* of $f_c\,(n)$ is defined as

$$F\,(x) = \frac{2}{\pi}\int_0^\infty f_c\,(n) \cos nx\, dx$$

Thus if $f_c\,(n) = f_c\,[F(x)]$, then $F(x) = f_c^{-1}\,[f_c\,(n)]$

[C] The Complex Fourier Transforms.

The Complex Fourier Transform of a function $F(x)$ for $-\infty < x < \infty$, is defined as

$$f\,(n) = \int_{-\infty}^\infty F\,(x)\, e^{inx}\, dx$$

where e^{inx} is said to be the *Kernel* of the transform.

The inversion formula is $F(x) = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^\infty f\,(n)\, e^{-inx}\, dn$

[D] Parseval's Identity for Fourier Integrals.

It is stated as $\displaystyle\int_{-\infty}^\infty |F(x)|^2\, dx = \frac{1}{2\pi}\int_{-\infty}^\infty |f(n)|^2\, dn$

where $f\,(n)$ is the Fourier transform of $F(x)$.

[E] Multiple Fourier Transforms.

If $F\,(x, y)$ be a function of two variables x and y, then assuming it to be the function of x only, its fourier transform $\phi\,(n, y)$ is given by

$$\phi\,(n,\, y) = \int_{-\infty}^\infty F\,(x, y)\, e^{inx}\, dx$$

NOTES

Now if $f(n, l)$ be the Fourier complex transform of $f(n, y)$ which is regarded as function of y only then

$$f(n, l) = \int_{-\infty}^{\infty} \phi(n, y)\, e^{ily}\, dy$$

These two results when combined, give

$$f(n, l) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, e^{i(nx + ly)}\, dx\, dy$$

and the inversion formula is

$$f(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(n, l)\, e^{-i(nx + ly)}\, dn\, dl$$

Similarly in case of three variables x, y, z, we have

$$f(n, l, m) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y, z)\, e^{i(nx + ly + mz)}\, dx\, dy\, dz$$

and $$f(x, y, z) = \frac{1}{8\pi^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(n, l, m)\, e^{-i(nx + ly + mz)}\, dn\, dl\, dm$$

Note 1. The result may be generalized for any number of variables.

Note 2. In case the Fourier transforms are finite such that $F(x, y)$ is a function of two independent variables x, y where $0 \le x \le \pi$ and $0 \le y \le \pi$, then the sine transform of $F(x, y)$ is given by

$$f_s(n, l) = \int_0^{\pi}\int_0^{\pi} F(x, y) \sin nx \sin ly\, dx\, dy$$

and the inversion formula is

$$F(x, y) = \frac{4}{\pi^2}\sum_{n=1}^{\infty}\sum_{l=1}^{\infty} f_s(n, l) \sin nx \sin ly$$

[F] Convolution or Faltung Theorem for Fourier Transforms.

If $F(x)$ and $G(x)$ are two functions such that $-\infty < x < \infty$ then their Faltung or Convolution F^*G is defined as

$$H(x) = F^*G = \int_{-\infty}^{\infty} F(n)\, G(x - n)\, dn \qquad \text{... (21)}$$

It is worth noting that the Fourier Transform of the Convolution of $F(x)$ and $G(x)$ is the product of their Fourier transforms, *i.e.*,

$$f[F^*G] = f[F]\, f[G] \qquad \text{... (22)}$$

Since $f[F^*G] = \int_{-\infty}^{\infty} H(x)\, e^{-inx}\, dx$ by definition

$$= \int_{-\infty}^{\infty} F(x)\, e^{-inx}\, dx \int_{-\infty}^{\infty} G(x)\, e^{-inx}\, dx$$

$$= f[F].\, f[G].$$

1.7.2 Laplace Transforms

If $F(t)$ be a function of t defined for all positive values of t (*i.e.,* $t \ge 0$), then the Laplace transform of $F(t)$ denoted by $L\{F(t)\}$ or $\tilde{F}(s)$ or $f(s)$ is defined by the expression

$$L\{F(t)\} = \tilde{F}(t) = f(s) = \int_0^\infty e^{-st} F(t)\, dt$$

where s is a parameter (real or complex).

If the integral $\int_0^\infty e^{-st} F(t)\, dt$ converges for some value of s, then the Laplace transform of $F(t)$ is said to exist, otherwise it does not exist.

Laplace Transform of Derivatives

If $F(t)$ is continuous for $t \geq 0$ and of exponential order as $t \to \infty$ while $F'(t)$ is sectionally continuous, i.e., $F'(t)$ is of class A for $t \geq 0$, and if $L\{F(t)\} = f(s)$, then $L\{F'(t)\} = s\,f(s) - F(0)$.

Properties

1. **Linearity Property**

   If $f(t) \leftrightarrow F(S)$ and $y(t) \leftrightarrow Y(S)$

   Then linearity property states that

   $af(t) + by(t) \leftrightarrow aF(S) + bY(S)$

2. **Time Shifting Property**

   If $f(t) \leftrightarrow F(S)$

   Then Time shifting property states that

   $f(t - t_0) \leftrightarrow e^{-St_0} F(S)$

   $f(t + t_0) \leftrightarrow e^{St_0} F(S)$

3. **Frequency Shifting Property**

   If $f(t) \leftrightarrow F(S)$

   Then frequency shifting property states that

   $e^{S_0 t}.f(t) \leftrightarrow F(S - S_0)$

   $e^{-S_0 t}.f(t) \leftrightarrow F(S + S_0)$

4. **Time Reversal Property**

   If $f(t) \leftrightarrow F(S)$

   Then Time reversal property states that

   $f(-t) \leftrightarrow F(-S)$

5. **Time Scaling Property**

   If $f(t) \leftrightarrow F(S)$

   Then Time scaling property states that

   $f(at) \leftrightarrow (1/a)F(S/a)$

6. **Differentiation and Integration Properties**

   If $f(t) \leftrightarrow F(S)$

   Then Differentiation property states that

   $df(t)/dt \leftrightarrow S.F(S)$

   $d^n f(t)/dt^n \leftrightarrow (S)^n.F(S)$

NOTES

Time Integration property states that

$\int f(t)dt \leftrightarrow (1/S)F(S)$

7. Multiplication and Convolution Properties

If $f(t) \leftrightarrow F(S)$ and $y(t) \leftrightarrow Y(S)$

Then multiplication property states that

$f(t).y(t) \leftrightarrow F(S) * Y(S)$

Convolution property states that

$f(t) * y(t) \leftrightarrow F(S).Y(S)$

Laplace Transform of some elementary functions:

$f(t)$	$F(s) = \mathcal{L}[f(t)]$	
$f(t) = 1$	$F(s) = \frac{1}{s}$	$s > 0$
$f(t) = e^{at}$	$F(s) = \frac{1}{(s-a)}$	$s > a$
$f(t) = t^n$	$F(s) = \frac{n!}{s^{(n+1)}}$	$s > 0$
$f(t) = \sin(at)$	$F(s) = \frac{a}{s^2 + a^2}$	$s > 0$
$f(t) = \cos(at)$	$F(s) = \frac{s}{s^2 + a^2}$	$s > 0$
$f(t) = \sinh(at)$	$F(s) = \frac{a}{s^2 - a^2}$	$s > \lvert a \rvert$
$f(t) = \cosh(at)$	$F(s) = \frac{s}{s^2 - a^2}$	$s > \lvert a \rvert$
$f(t) = t^n e^{at}$	$F(s) = \frac{n!}{(s-a)^{(n+1)}}$	$s > a$
$f(t) = e^{at}\sin(bt)$	$F(s) = \frac{b}{(s-a)^2 + b^2}$	$s > a$
$f(t) = e^{at}\cos(bt)$	$F(s) = \frac{(s-a)}{(s-a)^2 + b^2}$	$s > a$
$f(t) = e^{at}\sinh(bt)$	$F(s) = \frac{b}{(s-a)^2 - b^2}$	$s - a > \lvert b \rvert$
$f(t) = e^{at}\cosh(bt)$	$F(s) = \frac{(s-a)}{(s-a)^2 - b^2}$	$s - a > \lvert b \rvert$

Other Properties

[A] Linearity Property

A Laplace transform $L\{F(t)\}$ is said to be linear if for every pair of function $F_1(t)$ and $F_2(t)$ and for every pair of constants C_1 and C_2, we have

$$L\{C_1F_1(t) + C_2F_2(t)\} = C_1\, L\{F_1(t)\} + C_2\, L\{F_2(t)\}$$
$$= C_1 f_1(s) + C_2 f_2(s)$$

where $f_1(s)$ and $f_2(s)$ are linear transforms of $F_1(t)$ and $F_2(t)$ respectively.

We have $L\{F_1(t)\} = f_1\,(s) = \int_0^\infty e^{-st} F_1(t)\, dt$

and $L\{F_2(t)\} = f_2(s) = \int_0^\infty e^{-st} F_2(t)\, dt$

so that $L\{C_1\, F_1(t)\} = C_1\, f_1\,(s) = \int_0^\infty e^{-st} C_1\, F_1(t)\, dt = C_1 L\{F_1(t)\}$

and $L\{C_2\, F_2(t)\} = C_2\, f_2(s) = \int_0^\infty e^{-st} C_2 F_2(t)\, dt = C_2 L\,\{F_2(t)\}$

$$\therefore L\{C_1F_1(t) + C_2F_2(t)\} = \int_0^\infty e^{-st}\{C_1\, F_1(t)\, dt + C_2\, F_2(t)\}\, dt \text{ by definition}$$

$$= \int_0^\infty e^{-st} C_1 F_1(t)\, dt + \int_0^\infty e^{-st} C_2 F_2(t)\, dt$$
$$= C_1\, L\{F_1(t)\} + C_2\, L\{F_2(t)\}$$
$$= C_1 f_1(s) + C_2 f_2(s)$$

The result may be generalized for any number of functions and for the same number of arbitrary constants, *i.e.,*

$$L\left\{\sum_{r=1}^{n} C_r\, F_r\,(t)\right\} = \sum_{r=1}^{n} C_r\, L\,\{F_r(t)\}.$$

[B] First Translation (or Shifting) Property

If $f\,(s)$ be the Laplace transform of $F(t)$, then the Laplace transform of e^{at} $F(t)$ is $f\,(s - a)$, where a is any real or complex number, i.e., if

$L\{F(t)\} = f(s)$, then $L\,\{e^{at}\, F(t)\} = f(s - a)$.

Given, $L\{F(t)\} = \int_0^\infty e^{-st} F(t)\, dt = f\,(s)$

$$\therefore \quad L\,\{e^{at}\, F(t)\} = \int_0^\infty e^{-st} e^{at} F(t)\, dt$$

$$= \int_0^\infty e^{-(s-a)t} F(t)\, dt$$

$$= \int_0^\infty e^{-ut} F(t)\, dt \text{ by putting } u = s - a$$

$$= f(u)$$
$$= f(s - a)$$

NOTES

[C] Second Tanslation (or Shifting) Property

If $L\{F(t)\} = f(s)$ *and* $G(t) = \begin{cases} F(t-a), & t > a \\ 0, & t < a \end{cases}$

Then $L\{G(t)\} = e^{-as} f(s)$.

We have $$L\{G(t)\} = \int_0^\infty e^{-st} G(t)\, dt$$

$$= \int_0^a e^{-st} G(t)\, dt + \int_a^\infty e^{-st} G(t)\, dt$$

$$= \int_0^a e^{-st} \cdot 0\, dt + \int_a^\infty e^{-st} F(t-a)\, dt$$

$$= \int_a^\infty e^{-st} \cdot F(t-a)\, dt$$

$$= \int_0^\infty e^{-s(u+a)} F(u)\, du$$, by taking $u = t - a$, *i.e.,* $du = dt$.

when $t = a$, $u = 0$ and when $t = \infty$, $u = \infty$.

$$= e^{-sa} \int_0^\infty e^{-su} F(u)\, du$$

$$= e^{-sa} f(s)$$

[D] The Change of Scalar Property

If $L\{F(t)\} = f(s)$, *then* $L\{F(at)\} = \frac{1}{a} f\left(\frac{s}{a}\right)$

We have $$L\{F(t)\} = \int_0^\infty e^{-st} F(t)\, dt = f(s)$$

$$\therefore \quad L\{F(at)\} = \int_0^\infty e^{-st} F(at)\, dt \qquad \text{(on replacing } t \text{ by } at)$$

$$= \int_0^\infty e^{-su/a} F(u) \frac{du}{a} \qquad \text{(by taking } at = u)$$

$$= \frac{1}{a} \int_0^\infty e^{-pu} F(u)\, du \qquad \left(\text{where } p = \frac{s}{a}.\right)$$

$$= \frac{1}{a} \int_0^\infty e^{-pu} F(u)\, dt \qquad \text{(replacing } u \text{ by } t)$$

$$= \frac{1}{a} f(p)$$

$$= \frac{1}{a} f\left(\frac{s}{a}\right) \qquad \because\ p = \frac{s}{a}.$$

1.7.3 Relationship Between Fourier and Laplace Transforms

Fourier transform is the special case of Laplace transform which is evaluated keeping the real part zero. Generally, the Fourier transform is used for analysis in frequency domain whereas Laplace transform is typically used for analysis in s-domain, it is not frequency domain. The Fourier transform helps to study and analyse anything in the frequency domain whereas Laplace transform is specifically done for complex analysis, when anything is not easy and simple to analyse in time domain, then we convert it into s domain and then take the inverse Laplace transform to complete the analysis.

Laplace Transform

$$f(t) \rightarrow \int_{-\infty}^{\infty} e^{-st}\, f(t)$$

Put $\quad s = \sigma + j\omega$

$$f(t) \rightarrow \int_{-\infty}^{\infty} e^{-(\sigma + j\omega)t}\, f(t)$$

if $\quad \sigma = 0$, then

$$= \int_{-\infty}^{\infty} e^{-j\omega t}\, f(t)$$

$$= F(\omega)$$

1.8 DIRAC DELTA FUNCTION

The Dirac delta distribution (δ distribution), also known as the unit impulse symbol, is a generalized function or distribution over the real numbers, whose value is zero everywhere except at zero, and the integral over the entire real line is equal to one.

The current understanding of the impulse is as a linear functional that maps every continuous function to its value at zero, or as the weak limit of a sequence of bump functions, which are zero over most of the real line, with a tall spike at the origin. Bump functions are thus sometimes called 'Approximate' or 'Nascent' delta distributions.

The delta function was introduced by physicist Paul Dirac as a tool for the normalization of state vectors. It also has uses in probability theory and signal processing. The Kronecker delta function, which is usually defined on a discrete domain and takes values 0 and 1, is the discrete analog of the Dirac delta function.

The graph of the Dirac delta is usually thought of as following the whole x-axis and the positive y-axis. The Dirac delta is used to model a tall narrow spike function (an impulse), and other similar abstractions, such as a point charge, point mass or electron point. For example, to calculate the dynamics of a billiard ball being struck, one can approximate the force of the impact by a Dirac delta. In doing so, one not only simplifies the equations, but one also is able to calculate the motion of the ball by only considering the total impulse of the collision without a detailed model of all of the elastic energy transfer at subatomic levels (for instance).

To be specific, suppose that a billiard ball is at rest. At time $t = 0$ it is struck by another ball, imparting it with a momentum P, in kg m/s. The exchange of momentum is not actually instantaneous, being mediated by elastic processes at the molecular and subatomic level, but for practical purposes it is convenient to consider that energy transfer as effectively instantaneous. The force therefore is $P\delta(t)$. The units of $\delta(t)$ are s^{-1}.

To model this situation more rigorously, suppose that the force instead is uniformly distributed over a small time interval $\Delta t = [0, T]$. That is,

$$F_{\Delta t}(t) = \begin{cases} P/\Delta t & 0 < t \leq T, \\ 0 & \text{otherwise.} \end{cases}$$

Then the momentum at any time t is found by integration:

$$p(t) = \int_0^t F_{\Delta t}(\tau)\, \mathrm{d}\tau = \begin{cases} P & t > T \\ P\,t/\Delta t & 0 < t \leq T \\ 0 & \text{otherwise.} \end{cases}$$

Now, the model situation of an instantaneous transfer of momentum requires taking the limit as $\Delta t \to 0$, giving,

$$p(t) = \begin{cases} P & t > 0 \\ 0 & t \leq 0 \end{cases}$$

Here the functions $F_{\Delta t}$ are thought of as useful approximations to the idea of instantaneous transfer of momentum.

The delta function allows us to construct an idealized limit of these approximations. Unfortunately, the actual limit of the functions (in the sense of pointwise convergence) $\lim_{\Delta t \to 0} F_{\Delta t}$ is zero everywhere but a single point, where it is infinite. To make proper sense of the Dirac delta, we should instead insist that the property,

$$\int_{-\infty}^{\infty} F_{\Delta t}(t)\, \mathrm{d}t = P,$$

which holds for all $\Delta t > 0$, should continue to hold in the limit. So, in the equation $F(t) = P\,\delta(t) = \lim_{\Delta t \to 0} F_{\Delta t}(t)$, it is understood that the limit is always taken outside the integral.

Definitions

The Dirac delta can be roughly thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

And which is also constrained to satisfy the identity,

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1.$$

This is merely a heuristic characterization. The Dirac delta is not a function in the traditional sense as no function defined on the real numbers has these properties. The Dirac delta function can be rigorously defined either as a distribution or as a measure.

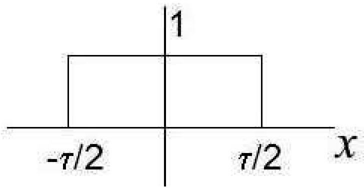
The Dirac delta is not truly a function, at least not a usual one with domain and range in real numbers. For example, the objects $f(x) = \delta(x)$ and $g(x) = 0$ are equal everywhere except at $x = 0$ yet have integrals that are different. According to Lebesgue integration theory, if f and g are functions such that $f = g$ almost everywhere, then f is integrable if and only if g is integrable and the integrals of f and g are identical. A rigorous approach to regarding the Dirac delta function as a mathematical object in its own right requires measure theory or the theory of distributions.

1.9 GATE FUNCTION (RECTANGULAR PULSES)

We define a gate function as rectangular pulses of unit height and unit width, centred at the origin.

$$rect\left(\frac{x}{\tau}\right) = \begin{cases} 0 & |x| > \tau/2 \\ 0.5 & |x| = \tau/2 \\ 1 & |x| < \tau/2 \end{cases}$$

Graphical representation:



Check Your Progress

9. Define the Power Spectral Density (PSD).
10. What does energy spectral density quantify?
11. State about the Fourier Transform (FT).
12. Differentiate between the Fourier transform and the Laplace transform.
13. Define Dirac delta distribution.
14. Give definition of Dirac delta function.
15. What do you mean by a gate function?

1.10 CONVOLUTION AND IMPULSE RESPONSE

Convolution can be used to determine a system's zero-state response (i.e., the reaction to an arbitrary input when the system's initial conditions are zero) to an arbitrary input by using the system's impulse response. Given a system's impulse response, $h(t)$, and an input signal, $f(t)$, the output signal, $y(t)$, is the product of $h(t)$ and $f(t)$:

$$y(t) = h(t) * f(t) = f(t) * h(t)$$
$$= \int_{-\infty}^{+\infty} h(t-\lambda)f(\lambda)d\lambda = \int_{-\infty}^{+\infty} h(\lambda)f(t-\lambda)d\lambda$$

NOTES

If $h(t)$=0 for t<0 and $f(t)$=0 for t<0 we can change the limits of integration such that:

$$y(t) = \int_0^t h(t-\lambda)f(\lambda)d\lambda = \int_0^t h(\lambda)f(t-\lambda)d\lambda$$

Impulse Response Function

Frequently, the systems are described in terms of their transfer function in the frequency domain. This process was so frequent that it was totally neglected that the system's response can be described in the time domain. This can be done using convolution.

In the temporal domain, a system's impulse response function $h(t)$ is used to describe it. This function expresses the system's response at time t to a unit impulse or -function input given at time $t = 0$.

Suppose current time is t and an impulse was administered to the system in the past at time τ. The response now is $y(t) = h(t-\tau)$.

Assume that a series of impulses of varying intensity x is provided. The portion of the response $dy(t)$ that is due to an impulse from a previous time τ is

$$\mathrm{d}y(t) = x(\tau)h(t-\tau)\mathrm{d}\tau$$

So that the total response now is

$$y(t) = \int_{-\infty}^{t} x(\tau)h(t-\tau)\mathrm{d}\tau\ ,$$

Where all $\tau \leq t.$

Now suppose that h is causal. If $\tau > t$ then $h(t-\tau) = 0$, and therefore, provided $h(t)$ is causal, the time response $y(t)$ to an input $x(t)$ is:

$$y(t) \quad = \quad \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\mathrm{d}\tau \quad = \quad x * h$$

The output is the CONVOLVED input after the Impulse Response Function has been applied.

1.10.1 Convolution Integral

In the physical sciences, the convolution integral is often used. $f_1(t) * f_2(t)$ is the symbol for the convolution integral of two functions $f_1(t)$ and $f_2(t)$.

$$f_1(t) * f_2(t) \equiv \int_{-\infty}^{\infty} f_1(\tau)f_2(t-\tau)d\tau$$

1.10.2 Physical Interpretation on Convolution

In this case, the second function is inverted vertically, denoted by $f_2(-\tau)$. The function is then moved to the right by t seconds to obtain $f_2(t-\tau)$. This is multiplied by the first function to obtain a third function. By computing the integral, we obtain the area beneath the graph of this third function. That's convolution.

NOTES

CONTINUOUS-TIME SYSTEMS The **Zero-state Response** can be written as the **convolution integral** of the Input and the Unit Impulse Response. If $f(t)$ and $h(t)$ are causal, the limits of integration are 0 to t.	$\overbrace{y(t)}^{\text{Zero-state Response}} = \underbrace{f(t)}_{\text{Input}} * \overbrace{h(t)}^{\text{Unit Impulse Response}} = \overbrace{\int_0^t f(\tau)\, h(t-\tau)\, d\tau}^{\text{Convolution Integral}}, \quad t \geq 0$

DISCRETE-TIME SYSTEMS The **Zero-state Response** can be written as the **convolution sum** of the Input and the Unit Impulse Response:	$\overbrace{y[k]}^{\text{Zero-state Response}} = \underbrace{f[k]}_{\text{Input}} * \overbrace{h[k]}^{\text{Unit Impulse Response}} = \overbrace{\sum_{m=-\infty}^{\infty} f[m]\, h[k-m]}^{\text{Convolution Sum}}$

Properties of Convolution

Commutative: $f_1(t) * f_2(t) = f_2(t) * f_1(t)$

Associative: $f_1(t) * [f_2(t) * f_3(t)] = [f_1(t) * f_2(t)] * f_3(t)$

Distributive: $f_1(t) * [f_2(t) * f_3(t)] = f_1(t) * f_2(t) + f_1(t) * f_3(t)$

Shift Property: If $f_1(t) * f_2(t) = c_1(t)$ then $f_1(t) * f_2(t-T) = c_1(t-T)$
$f_1(t-T) * f_2(t) = c_1(t-T)$

Convolution with an impulse: $f_1(t) * \delta(t) = f(t)$

Width Property: If $f_1(t)$ and $f_2(t)$ have durations of T_1 and T_2 respectively, then the duration of $f_1(t) * f_2(t)$ is $T_1 + T_2$.

1.11 PARSEVAL'S THEOREM

In mathematical analysis, Parseval's identity, named after Marc-Antoine Parseval, is a fundamental result on the summability of the Fourier series of a function. Geometrically, it is generalised Pythagorean theorem for inner-product spaces (which can have an uncountable infinity of basis vectors).

Informally, the identity asserts that the sum of the squares of the Fourier coefficients of a function is equal to the integral of the square of the function,

$$\| f \|^2_{L^2(-\pi,\pi)} = \int_{-\pi}^{\pi} | f(x) |^2 \, dx = 2\pi \sum_{n=-\infty}^{\infty} | c_n |^2$$

Where the fourier coefficients c_n of f are given by

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{inx} dx.$$

More formally, the result holds as stated provided f is square-integrable or, more generally, in $L^2[-\pi, \pi]$. A similar result is the plancherel theorem, which asserts that the integral of the square of the fourier transform of a function is equal to the integral of the square of the function itself. In one-dimension, for $f \in L^2$ (R),

$$\int_{-\infty}^{\infty} | \hat{f}(\xi) |^2 \, d\xi = \int_{-\infty}^{\infty} | f(x) |^2 \, dx$$

NOTES

1. The Generalization of Parserval's Theorem

The result is

$$\int_{-\infty}^{\infty} f(t)g(t)dt = \frac{1}{2\pi}\int_{-\infty}^{-\infty} \bar{f}(\omega)\bar{g}(\omega)^* d\omega \qquad ...(1.5)$$

This has many names but is often called Plancherel's formula.

The key step in the proof of this is the use of the integral representation of the δ-function

$$\delta(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\pm i\tau\omega} d\omega \text{ or } \delta(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{\pm i\tau\omega} d\tau \qquad ...(1.6)$$

We firstly invoke the inverse Fourier transform

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{f}(\omega)e^{\pm i\omega t} d\omega \qquad ...(1.7)$$

And then use this to re-write the LHS of Equation (1.5) as

$$\int_{-\infty}^{\infty} f(t)g(t)^* dt \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{f}(\omega)e^{i\omega t} d\omega\right)$$

$$\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} g(\omega')\ldots d\omega'\right) dt \qquad ...(1.8)$$

Re-arranging the order of integration we obtain

$$\int_{-\infty}^{\infty} f(t)f(t)^* dt = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{f}(\omega)\bar{g}(\omega')\underbrace{\left(\int_{-\infty}^{\infty} e^{i(\omega-\omega')t} dt\right)}_{\text{Use delta } -\int_n} d\omega' d\omega$$

$$...(1.9)$$

The version of the integral representation of the δ-function we use in Equation (1.6) above is

$$\delta(\omega-\omega') = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it(\omega-\omega')} dt \qquad ...(1.10)$$

Using this in Equation (1.9), we obtain

$$\int_{-\infty}^{\infty} f(t)g(t)^* dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{f}(\omega)\left(\int_{-\infty}^{\infty} \bar{g}(\omega')^* \delta(\omega-\omega')d\omega'\right)d\omega \qquad ...(1.11)$$

Equation (1.11) comes about because of the $f(\omega)$ general δ function property

$$\int_{-\infty}^{\infty} F(\omega')\delta(\omega-\omega')d\omega'$$

Taking $g = f$ in we immediately obtain

$$\int_{-\infty}^{\infty} [f(t)]dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\bar{f}(\omega)|^2 d\omega \qquad ...(1.12)$$

2. The Convolution Theorem and the Auto-Correlation Function

The statement of the Convolution theorem is this: for two function $f(t)$ and $g(t)$ with Fourier transforms $F|f(t)| = \bar{f}(\omega)$ and $F|g(t)| = \bar{g}(\omega)$, with convolution integral defined by.

$$f * g = \int_{-\infty}^{\infty} f(u)g(t-u)du, \qquad ...(1.13)$$

Then the Fourier transform of this convolution is given by

$$F(f * g) = \bar{f}(\omega)\bar{g}(\omega). \quad \text{...(1.14)}$$

To prove Equation (1.14) we write it as

$$F(f * g) = \int_{-\infty}^{\infty} e^{-i\omega t}\left(\int_{-\infty}^{\infty} f(u)g(t-u)du\right)dt \quad \text{...(1.15)}$$

Now define $\tau = t - u$ and divide the order of integration to find

$$F(f * g) = \int_{-\infty}^{\infty} e^{-i\omega t} f(u)du \int_{-\infty}^{\infty} e^{-i\omega t} g(\tau)d\tau = \bar{f}(\omega) - \bar{g}(\omega) \quad \text{...(1.16)}$$

This step is allowable because the region of integration in the $\tau - u$ plane is infinite. As we shall later, with Laplace transforms this is not the case and requires more case.

The normalised auto-correlation function is related to this and is given by

$$\gamma(t) = \frac{\int_{-\infty}^{\infty} f(u) f^*(t-u)du}{\int_{-\infty}^{\infty} | f(u)|^2 \, du}$$

Practical Harmonic Analysis: If function is not given by a formula, but by a graph or by a table of corresponding values, then process of finding the Fourier series for the function is known as Harmonic analysis.

$$\text{As Mean} = \frac{1}{b-a}\int_a^b f(x)dx$$

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)dx$$

$$= \frac{2}{2\pi - 0}\int_0^{2\pi} f(x)dx$$

= 2 |Mean of $f(x)$ in $(0, 2\pi)$]

Similarly a_n = 2[Mean of $f(x)$ cos nx in $(0, 2\pi)$]

b_n = 2[Mean of $f(x)$ sin nx in $(0, 2\pi)$]

Example 2: The turning moment T units of the crank shaft of a steam engine for a series of values of the crank-angle θ in degrees:

θ:	0°	30°	60°	90°	120°	150°	180°
T:	0	5224	8097	7850	5499	2626	0

Find the first 4 terms in a series of sines to represent T also calculate T when $\theta = 75°$

Soution: $T = b_1 \sin\theta + b_2 \sin 2\theta + b_3 \sin 3\theta + b_4 \sin 4\theta$

b_n = 2[Mean of $f(x)$ sin nx in $(0, 2\pi)$)]

NOTES

0°	T	sin θ	sin 2θ	sin 3θ	sin 4θ	T sin θ	T sin 2θ	T sin 3θ	T sin 4θ
0°	0	0	0	0	0	–0	0	0	0
30°	5224	0.5	0.866	1	0.866	–2612	4523.984	5224	4523.984
60°	8097	0.866	0.866	0	–0.866	–7012.002	7012.002	0	–7012.002
90°	7850	1	0	–1	0	7850	0	–7850	0
120°	5499	0.866	–0.8660	0	0.866	4767.0831	4767.0831	0	4767.0831
150°	2626	0.5	–0.866	1	–0.866	1313	–2274.116	–2626	–2274.116
						23554.08591	14028.9531	0	0

$$b_1 = 2\,[\text{Mean of } T \sin\theta \text{ in } (0, 2\pi)]$$

$$= 2\left[\frac{23554.0851}{6}\right] = 7851.36$$

$$b_2 = 2\,[\text{Mean of } T \sin 2\theta \text{ in } (0, 2\pi)]$$

$$= 2\left[\frac{14028.9531}{6}\right] = 1358.7725$$

$$b_3 = 2\,[\text{Mean of } T \sin 3\theta \text{ in } (0, 2\pi)] = 0$$

$$b_4 = 2\,[\text{Mean of } T \sin 4\theta \text{ in } (0, 2\pi)] = 0$$

$$T = (7851.36)\sin\theta + (1558.7725)\sin 2\theta$$

At $\theta = 75°$

$$T = 7851.56 \sin 75° + (1558.7725) \sin 150°$$

$$= (7851.56)\,(.9659) + (1558.7725)\,(0.5)$$

$$= (7583.8218) + (799.38625)$$

Example 3: Find the Fourier series as far as the second harmonic to represent the function given by table below:

x :	0°	30°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°
$f(x)$:	2.34	3.01	3.69	4.15	3.69	2.20	0.83	0.51	0.88	1.09	1.19	1.64

x°	sin x	sin $2x$	cos x	cos $2x$	$f(x)$	$f(x)$ sin x	$f(x)$ sin $2x$	$f(x)$ cos x	$f(x)$ cos$x$
0°	0	0	1	1	2.34	0	0	2.340	2.340
30°	0.50	0.87	0.87	0.50	3.01	1.505	2.619	2.619	1.505
60°	0.87	0.87	0.30	–0.50	3.69	3.210	3.210	1.845	1.845
90°	1.00	0	0	–1.00	4.15	4.150	0	0	–4.150
120°	0.87	–0.87	–0.50	–0.50	3.69	3.210	–3.210	–1.845	–1.845
150°	0.50	–0.87	–0.87	0.50	2.20	1.100	–1.914	–1.914	1.100
180°	0	0	–1	1.00	0.83	0	0	–0.830	0.830
210°	–0.50	0.87	–0.87	0.50	0.51	–0.255	0.444	–0.444	0.255
240°	–0.87	0.87	–0.50	–0.50	0.88	–0.766	0.766	–0.440	–0.440
270°	–1.00	0	0	–1.00	1.09	–1.090	0	0	–1.090
300°	–0.87	–0.87	0.50	–0.50	1.19	–1.035	–1.035	0.595	–0.595
330°	–0.50	–0.87	0.87	0.50	1.64	–0.820	–1.427	1.427	0.820
					25.22	9.209	–0.547	3.353	–3.115

Solution: $b_1 = 2[\text{Mean of } f(x) \sin x]$

$$= 2\left(\frac{9.209}{12}\right) = 1.535$$

$b_2 = 2[\text{Mean of } f(x) \sin 2x]$

$$= 2\left(\frac{-0.547}{12}\right) = -0.091$$

$$f(x) = \frac{a_0}{2} + a_1 \cos x + a_2 \cos 2x + b_1 \sin x + b_2 \sin 2x$$

$$= 2.1018 + 0.557 \cos x - 0.519 \cos 2x + 11.535 \sin x - 0.091 \sin 2x$$

Example 4: The following values of y give the displacement of a certain machine part for the rotation x of the flywheel.

x :	0°	60°	120°	180°	240°	300°	360°
y :	1.98	2.15	2.77	–0.22	–0.31	1.43	1.93

Express y is Fourier series upto the IIIrd harmonic.

Solution: Let $y = \frac{a_0}{2} + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x)$

$$+(a_3 \cos 3x + b_3 \sin 3x)$$

x	y	$\cos x$	$\sin x$	$\cos 2x$	$\sin 2x$	$\cos 3x$	$\sin 3x$	$y \cos x$	$y \sin x$	$y \cos 2x$	$y \sin 2x$	$y \cos 3x$	$y \sin 2x$
0°	1.98	1.0	0	1.0	0	1.0	0	1.98	0	1.98	0	1.98	0
60°	2.15	0.5	0.866	–0.5	0.866	–1.0	0	1.075	1.8619	–1.075	1.8619	–2.15	0
120°	2.77	–0.5	0.866	–0.5	–0.866	1.0	0	–1.385	2.3988	–1.385	–2.3988	2.77	0
180°	–0.22	7.0	0	1.0	0	–1.0	0	0.22	0	0	0	.22	0
240°	–0.31	–0.5	–0.866	–0.5	0.866	1.0	0	0.155	0.2685	0.2685	–0.2685	–.31	0
360°	–1.43	0.5	–0.866	–0.5	–0.866	–1.0	0	0.715	–1.2383	–.715	–1.2383	–1.43	0
	7.8							2.76	3.2909	–1.4635	–2.0437	1.08	0

$b_3 = 0$

$$\therefore \quad y = 1.3 + (0.92 \cos x + 1.0969 \sin x)$$
$$+ (-0.4878 \cos 2x - 0.6812 \sin 2x) + (0.36 \cos 3x)$$

Example 5: The following table gives the variations of a periodic current over a period

t (sec) :	0	$T/6$	$T/3$	$T/2$	$2T/3$	$5T/6$	T
A (amp) :	1.98	1.30	1.05	1.30	–0.88	–0.25	1.98

Show that there is a direct current part of 0.75 amp. in the variable current, and obtain the amplitude of the first harmonic.

Solution: Let $A = \frac{a_0}{2} + a_1 \cos \frac{2\pi t}{T} + b_1 \sin \frac{2\pi t}{T} \cdots$

NOTES

t	A	$\cos\left(\frac{2\pi t}{T}\right)$	$\sin\left(\frac{2\pi t}{T}\right)$	$A\cos\frac{2\pi t}{T}$	$A\sin\frac{2\pi t}{T}$
.0	1.98	1	0	1.98	1.1258
T/6	1.3	0.5	0.866	0.65	0
T/3	1.05	–0.5	0.866	–0.525	0.9093
T/2	1.3	–1	0	–1.3	0
2T/3	–0.88	–0.5	–0.866	.44	0.76208
5T/6	–0.25	0.5	–0.866	–0.125	0.2165
	4.5			1.12	3.01348

a_0, a_1 and a_2

$$\therefore\ A = 0.75 + 0.373\cos\left(\frac{2\pi t}{T}\right) + 1.005\sin\left(\frac{2\pi t}{T}\right)$$

$\therefore$ A has a direct current part of 0.75 amp.

The amplitude of first harmonic is given by

$$= \sqrt{(.373)^2 + (1.005)^2}$$

$$= \sqrt{1.1491}$$

$$= 1.072$$

1.12 NOISE: SOURCE, TYPES AND MODULATION TECHNIQUES

All electronic systems contain noise, which is defined as voltage and current variations induced by the random motion of charged particles. In RF (Radio Frequency) and microwave receivers, which must extract information from extremely small signals, knowing noise and how it propagates through a system is especially important. Noise introduced by circuit parts can mask or obfuscate low-level signals, impairing speech or video reception, causing bit recognition in digital systems to be uncertain, and causing radar mistakes. For RF and microwave engineers, measuring the noise contributions of circuit elements in the form of noise factor or noise figure is a critical task.

1.12.1 Physical Sources of Noise

There are two types of noise sources: **internal** and **external** to the receiver.

1. The antenna picks up external sources of noise. Atmospheric noise, galactic/cosmic noise, man-made noise, interference 'Noise' generated by other users in a nearby channel (Adjacent Channel Interference, ACI) or the same channel (Co-Channel Interference, CCI), and so on are examples.
2. Components inside the receiver emit internal noise. Random processes like as charge flow in a device, or, at a more fundamental level, thermal vibrations in any component at a temperature above absolute zero, cause this noise.

Radio receivers are made up of noise-generating components. Noise is generated by all components, whether passive (such as, resistors) or active (transistor-based circuits). The device's useful operating range is limited by noise in active components.

The noise can be generated by:

- Thermal noise, which results from the thermal vibration of bound charges. It's also known as Johnson or Nyquist noise, and it's the most common type.
- Shot noise occurs when charge carriers in an active device, such as a solid-state device or an electron tube, fluctuate randomly.
- Recombination of charge carriers causes flicker noise, also known as 1/*f* noise, which can be found in active components.
- When compared to the above noise sources, quantum noise, which is created by the discrete or quantized nature of charge carriers and photons, is generally small.

1.12.2 Noise Types

Following are the types of noise.

Thermal Noise

The universe as a whole is in motion. Even seemingly inert things have random vibrations in their molecules. The human senses detect this random vibration as heat. The average kinetic energy of these random vibrations is measured by temperature. The same may be said of the molecules in insulators, conductors, and semiconductors, as well as charge carriers (electrons and holes) and all physical structures used to construct electronic devices. The power of thermal noise is calculated as follows:

$$\langle P_n \rangle = kTB.$$

This is a statistical measure expressed as an average, as shown by the brackets.

P_n = Noise power expressed in watts,

T = Absolute temperature in Kelvin,

B = Bandwidth in Hertz, and

$k = 1.3806488 \times 10^{-23} \, \frac{Joules}{Degree\ K}$ = Boltzmann's constant.

The thermal noise produced by electronic components presents itself as voltage and current oscillations. When monitored within a narrow bandwidth, the statistical distributions of voltage and current are roughly Gaussian.

Most of the frequencies utilized by RF and microwave engineers are approximately white in thermal noise in a perfect resistor. At very high frequencies, when the quantum nature of electromagnetic waves takes over, a departure from this flat frequency distribution occurs.

NOTES

$$P_n(f) = \frac{hfB}{e^{\frac{hf}{kT}}-1} + \frac{hfB}{2}$$

The photon energy, *hf*, is included in the above equation, where *h* is Planck's constant. The operating frequencies employed by RF and microwave engineers in the vast majority of applications are such that the j" 1 and above equation simplifies to the well-known equation showing a flat power spectral density for thermal noise $P_n(f) = kTB$. Extremely low noise temperatures and extremely high frequencies are the exceptions.

Shot Noise

Electrical current is commonly thought of as a continuous quantity by engineers. It is made up of individual electrons, each of which has a set charge. The present has been quantized. Current is made up of the impacts of individual electrons travelling from the source to the load in a circuit, rather than a continuous flow. Over time, the electrons arrive in a uniform distribution. Shot noise or corpuscular noise is the result of this fluctuation in arrival time. The power of the noise is proportional to the current. The effective noise current approximates a Gaussian distribution with the RMS value or standard deviation of

$$i_n = \sqrt{2BIq_e}$$

1/*f* Noise

Many processes, ranging from the Nile River's flooding patterns to the firing of human brain neurons, have been discovered to have random fluctuations with a power spectra density that changes approximately as 1/*f*. At extremely low frequencies, some noise phenomena deviate from white noise. The power spectrum density of this so-called 1/*f* noise, often known as pink noise or Flicker noise, approximates a curve that is inversely proportional to frequency. It can be dominating at low frequencies, but at frequencies ranging from a few Hz to a few kHz, depending on the devices in issue, it decreases below the flat thermal noise.

The 1/*f* effects are prevalent at low frequencies until their power spectral density falls below that of thermal noise. Quantum effects are present at very high frequencies. For most of the frequencies in which electronic devices operate, the power spectral density can be considered flat.

1.12.3 Noise Calculations and Noise Temperature

The PSD (Power Spectral Density) is solely dependent on the resistor's temperature. Because the PSD is constant with frequency (at least within the approximation's limitations), it is referred to as 'White' noise. Indeed, every type of noise process, even non-thermal ones, can be mimicked using an equivalent noise source by correctly altering the source's temperature. For example, we previously demonstrated that an amplifier generates noise due to the presence of components that generate thermal noise as well as active devices that generate their own noise (shot noise, 1/*f* noise or flicker noise, etc.). In total, if the noise power created by an arbitrary white noise source is known, the noise power can be described using an equivalent situation. One may say that the noise was generated by a resistor operating at an equivalent temperature T_e.

NOTES

$$T_e = \frac{N}{kB}$$

where N is the noise power created by the noise process in its entirety. It is critical to understand that the equivalent temperature is in no way representational of the physical temperature in this circumstance.

Effect of Noise on Radio Systems

The Signal-to-Noise Ratio, or SNR, is defined as the ratio of signal power to noise power in a radio link. It is used to measure the link's quality.

$$SNR = \frac{\text{average signal power}}{\text{average noise power}}.$$

SNR is frequently stated in decibels (dB) because it is a ratio (SNR_{dB} = 10 log SNR).

Noise Figure

It is critical to monitor Signal-to-Noise Ratio or SNR as it passes through a radio receiver, as components dramatically alter this signal ratio. For instance, an amplifier is theoretically supposed to increase the strength of a signal, but in practice produces some noise in the process. As a result, the signal-to-noise ratio at the amplifier's output is actually lower than it is at the input: the signal is amplified, but so is the noise that was present at the input. Thus, the amplifier boosts both signal and noise power; then, additional noise is introduced into the output signal as a result of the amplifier's noisiness. The ratio of the input SNR to the output SNR is referred to as a component's noise figure:

$$F = \frac{S_i/N_i}{S_o/N_o}$$

where S_i and N_i denote the signal and noise powers at the input, and So and No denote the signal and noise powers at the output. This ratio is always greater than one because the output signal-to-noise ratio is never greater than the input signal-to-noise ratio:

$$F \geq 1$$

Additionally, the noise figure can be given in decibels (dB). For a network that is completely noiseless, F = 1 or F = 0 dB.

Noise in Electronic Components

Thermal energy causes molecular vibrations in all physical matter. In fact, temperature is a measure of the average kinetic energy of the molecules in motion. Electronic devices are not exempt from this rule. At some point, all electronic equipment generate noise as a result of the vibration of molecules.

Resistors

Thermal noise is generated by resistors due to the random fluctuations of their internal molecules. It is frequently beneficial to consider a resistor's noise contribution in terms of its equivalent noise voltage or equivalent noise current.

These values can be calculated using knowledge of kits, or the thermal energy of particles, which is also referred to as the thermal noise floor. The equivalent noise voltage generated by thermal noise in a resistor can be calculated as follows:

NOTES

$$\langle e_n \rangle = \sqrt{4KTBR_0}$$

The comparable thermal noise current is calculated as follows:

$$\langle i_n \rangle = \frac{e_n}{R_0} = \sqrt{\frac{4KTB}{R_0}}$$

Resistor noise power has a flat power spectral density and is temperature and resistor value dependent.

Capacitors

As with other reactive elements, ideal capacitors exhibit no thermal noise. Capacitors, on the other hand, are constructed using impure conductors and dielectrics and hence have an associated resistance. The noise voltage of a capacitor can be calculated by comparing it to the noise voltage of a parallel combination of a resistor and a capacitor. Allowing the resistor to reach infinity provides us with the thermal noise power generated solely by the capacitor.

$$e_{nC}{}^2 = \frac{4KTB}{R_pC^2} \frac{1}{\omega^2 + \left(\frac{1}{R_pC}\right)^2}$$

When we allow the effective parallel resistance to reach infinity, an interesting event occurs. The overall noise power integrated across all frequencies is solely temperature and capacitance dependent.

$$\int_0^\infty e_{nC}{}^2 \, df = \frac{kT}{C}$$

Inductors

As with other reactive elements, ideal inductors display no thermal noise. However, true inductive components have losses in their windings and magnetic cores. These losses can be represented using an equivalent series resistor, which contributes to noise. It is simply reliant on the inductor's effective series resistance and is given by:

$$e_{nL} = \sqrt{4KTB\,\boldsymbol{Re(Z)}} = \sqrt{4KTBR_s}$$

Active Devices

Numerous noise sources can be found in active gadgets. Thermal noise is generated by each of its resistive elements. Currents with a bias contribute to shot noise. All circuit elements' intrinsic reactance contributes to the power spectral distribution of noise. It is frequently advantageous to model all noise sources within a circuit element in terms of an equivalent noise voltage and current, as seen in Figure 1.7 with the Op-Amp.

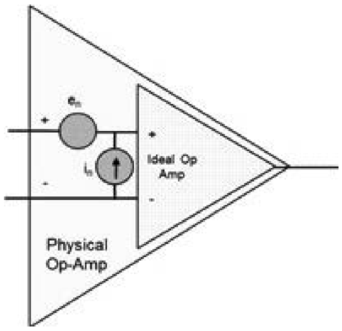


Fig. 1.7 *Op-Amp with Equivalent Noise Voltage and Current*

The real op-amp is modelled as an ideal noiseless op-amp with noise voltages and currents added at the input. Additionally, equivalent noise sources at the amplifier's output can be used to mimic the noise contributions. The noise voltage and noise current indicate the cumulative effect of all circuit parts inside the op-amp. Thermal energy is provided by resistive components. Currents with a bias contribute to shot noise. Reactive elements can alter the spectral distribution of noise power. This method can be used to represent transistors, RF amplifiers, and all other active components. The actual values of the equivalent noise sources can be determined either by measurements or through detailed modelling of the amplifier's internal circuitry.

All electronic systems contain noise. Understanding noise and the proper measurement, modelling, and accounting for its effects in a system is critical in RF and microwave receivers that must extract information from extremely small signals. Circuit elements can introduce noise that conceals or obscures low-level signals, impairing the signals being received. Measuring the noise contributions of circuit

$$F = \frac{S_i}{kT_0B} \cdot \frac{kGB(T_0 + T_e)}{GS_i} = 1 + \frac{T_e}{T_0} \geq 1.$$

This enables us to express the network's noise temperature as

NOTES

$$T_e = (F - 1)T_0.$$

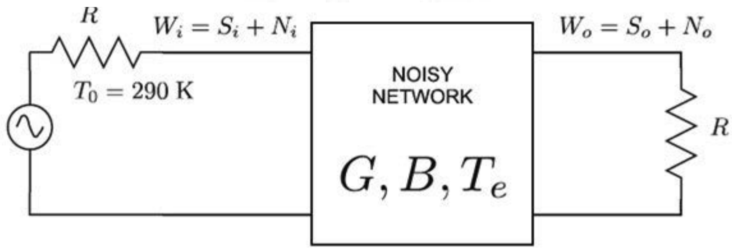


Fig. 1.8 *Noisy Two Port Network*

1.12.5 Interference and Noise in Amplitude Modulation (AM) and Pulse Modulation (PM)

Noise and Interference

Interference is a term that refers to artificial signals. Telephone lines are susceptible to interference from electricity lines (in the United States a distorted 60 Hz sinusoid). Cellular phone channels are susceptible to adjacent-cell phone talks that use the same signal frequency. The issue with this type of interference is that it operates in the same frequency band and has a similar structure to the desired communication signal.

Question: Suppose interference occupied a different frequency band; how would the receiver remove it?

Answer: If the interferer's spectrum does not overlap that of our communications channel—the interferer is out-of-band—we need only use a bandpass filter that selects our transmission band and removes other portions of the spectrum.

We use the notation $i(t)$ to represent interference. Because interference has man-made structure, we can write an explicit expression for it that may contain some unknown aspects (how large it is, for example).

Noise signals have little structure and arise from both human and natural sources. Satellite channels are prone to deep space noise caused by the galaxy's pervasive electromagnetic radiation. Thermal noise is a problem in all electrical circuits with resistors. Thus, in receiving small amplitude signals, receiver amplifiers will most certainly add noise as they boost the signal's amplitude. All channels are subject to noise, and we need a way of describing such signals despite the fact. We cannot create a formula for the noise signal in the same way that we can for the interference signal. White noise is the most often used noise model. It is totally defined by its frequency-domain properties.

- White noise has constant power at all frequencies.
- The phase of the noise spectrum is completely unpredictable at each frequency, it can be any value between 0 and 2π, and its value at any frequency has no relationship with the phase at any other frequency.

- When two distinct sources of noise combine, the resulting noise signal has a power equal to the sum of the component powers.

Due to frequency domain power, we define the power spectrum. Because of Parseval's Theorem, we define the power spectrum $\mathbf{P_s(f)}$ of a non-noise signal **s(t)** to be the magnitude squared of its Fourier transform as,

$$P_s(f)=(|S(f)|)^2$$

Integrating the power spectrum across any range of frequencies yields the amount of power contained in the signal in that band. Because signals must contain both positive and negative frequency components, we typically define the power in a spectral band as the integral over positive frequencies multiplied by two.

$$Power\ in\ [f_1, f_2] = 2\int_{f_1}^{f_2} P_s(f)df$$

Using the notation $n(t)$ to represent a noise signal's waveform, we define noise in terms of its power spectrum. For white noise, the power spectrum equals the constant.

$$\frac{N_0}{2}$$

With this definition, the power in a frequency band equals,

$$N_0(f_2 - f_1)$$

When we pass a signal through a linear, time-invariant system, the output's spectrum equals the product of the system's frequency response and the input's spectrum. Thus, the power spectrum of the system's output is given by,

$$P_y(f) = (|H(f)|)^2 P_x(f)$$

This result applies to noise signals as well. When we pass white noise through a filter, the output is also a noise signal but with power spectrum,

$$(|H(f)|)^2 \frac{N_0}{2}$$

Noise in Analog Communication Systems

The noise in analog communication systems can be evaluated through the performance of several analogue modulation methods in the presence of noise. The receiver's performance will be quantified in terms of the Signal-to-Noise Ratio (SNR) at the receiver's output, defined as,

$$\text{SNR}_o = \frac{\textit{average power of message signal at reciver output}}{\textit{average power of noise atthe reciver output}}$$

In some conditions, this is may not be the case, and the approximation methods can be used to figure out what the message and noise were doing at the receiver output.

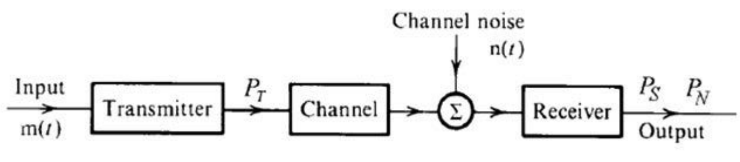


Fig. 1.9 *Model of an Analog Communication System*

NOTES

As depicted in the Figure 1.9, a modulated signal with power P_T is transmitted over an additive noise channel in the conventional communication system. The output SNR is $\text{SNR}_O = P_S = P_N$ because the signal and noise powers are P_S and P_N, respectively, at the receiver's output. By raising the transmitted power, this ratio can be increased as much as required. Considerations like as transmitter cost, channel capabilities, interference with other channels, and so on limit P_T's maximum value in practice. We will compare systems with the same transmitted power in order to fairly compare various modulation techniques. To compare the different modulation techniques, we also require a common measuring criterion. Baseband Signal-to-Noise Ratio (SNR) will be used in this case. Remember that all modulation schemes (i.e., the modulated signal is centered around a carrier frequency) are bandpass (i.e., An unmodulated communication system is known as **baseband**. Such a system is excellent for transmission across wires.

Baseband Communication System

Figure 1.10 shows a baseband communication system, where $m(t)$ is the band-limited message signal and W is its bandwidth.

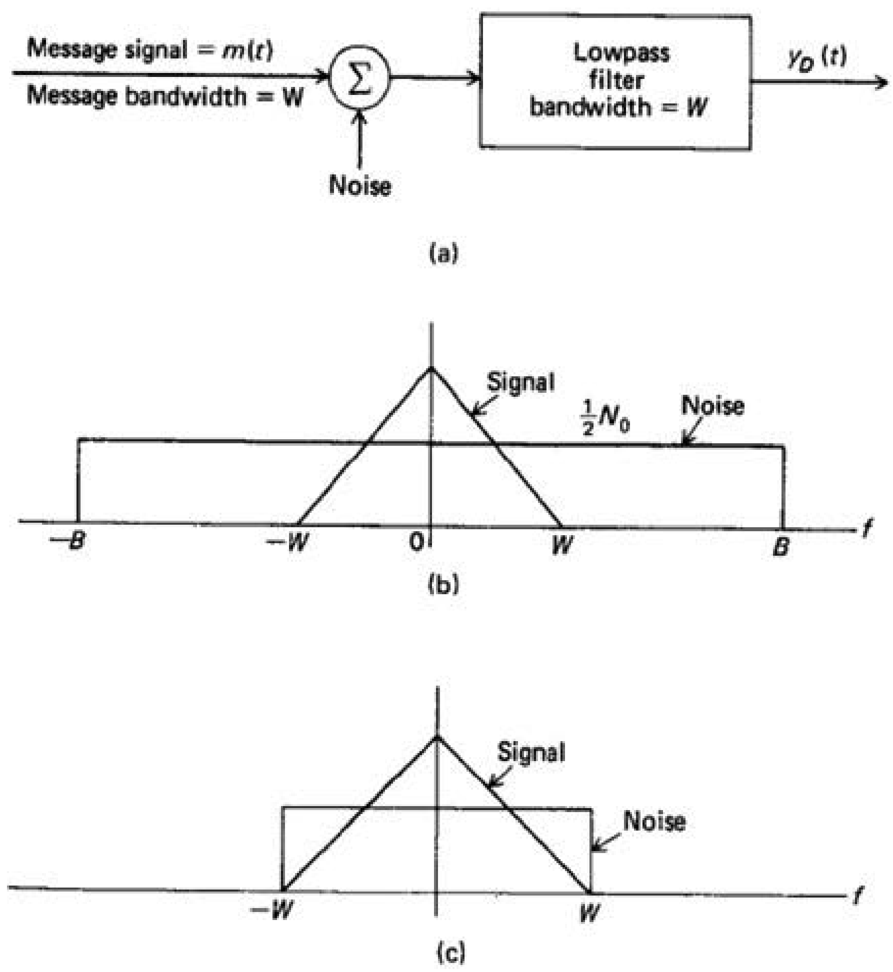


Fig. 1.10 *Signal Spectra at (a) Filter Input and Output (b) Model (c) Signal Spectra at Filter Output*

Figure 1.10 shows an example of a signal Power Spectral Density (PSD). In order to denote the average signal power, we will use the area under the triangle labelled 'Signal'. As depicted in Figure 1.10, we assume that the additive noise has a white, double-sided PSD of $N_O/2$ over a bandwidth $B > W$. Baseband systems broadcast with the same power as messages, therefore there is no difference between the two.

$$\text{i.e., } P_T = P.$$

SNR is improved by removing as much noise as possible with a low-pass filter with a bandwidth of W in the receiver. Figure 1.10 shows the PSD of the noise at the output of the LPF, and the average noise power is presented in the figure.

$$\int_{-w}^{w} \frac{N_O}{2} df = N_O W$$

Thus, the SNR at the receiver output is,

$$\text{SNR}_{\text{baseband}} = \frac{P_T}{N_O W}$$

For a baseband system, we have the option of boosting the transmission power, narrowing the message bandwidth, or reducing receiver noise to increase the SNR.

Amplitude Modulation

A sinusoidal carrier wave's *amplitude* is modulated linearly with the message signal in Amplitude Modulation (AM). In its most basic form, an AM signal appears as,

$$s(t)_{\text{AM}} = [A + m(t)] \cos(2\pi f_c t)$$

Where A is the amplitude of the carrier, f_c is the carrier frequency, and $m(t)$ is the message signal.

The ***modulation index***, μ is defined as,

$$\mu = \frac{m_p}{A}$$

Where m_p is the peak amplitude of $m_P = \max|m(t)|$$m(t)$, i.e.,

Recall that if μ = 1, (i.e., $A \geq m_p$), the envelope of $s(t)$ will have the same shape as the message $m(t)$, and so a simple envelope detector can be used to demodulate the AM signal. The availability of a very simple receiver is the primary advantage of AM, as its noise performance is not particularly good.

If an envelope detector is not possible to utilize, a technique called *synchronous detection* can be used. Figure 1.11 depicts the block diagram of a synchronous detector.

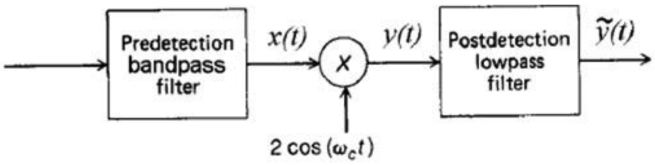


Fig. 1.11 *Synchronous Demodulator*

NOTES

The receiver's waveform is multiplied by a local carrier with the same frequency (and phase) as the *transmitter's carrier*. This effectively substitutes a cos2(φ) phrase for the cos(φ) term. From the individual's identity,

$$2\cos^2(x) = 1 + \cos 2(x)$$

As a result, the message signal is frequency translated down to baseband (i.e., $f = 0$) and up to twice the carrier frequency. The baseband message signal is then recovered using the low-pass filter. As one might assume, the primary disadvantage of this technique is that it necessitates the creation of a precisely synchronized local carrier signal with the transmitted carrier.

The AM signal is composed of two components: the carrier $\cos(2\pi f_c t)$ and the sidebands $m(t)\cos(2\pi f_c t)$. Due to the inefficiency of sending the carrier term, another form of AM worth considering is one in which the carrier is suppressed. This is referred to as a Double SideBand - Suppressed Carrier (DSB-SC) and is denoted by,

$$s(t)_{\text{DSB-SC}} = Am(t)\cos(2\pi f_c t)$$

In this situation, the signal's envelope does not resemble the original message signal, necessitating the employment of a synchronous detector for demodulation.

Noise in DSB-SC

The *predetection* signal (i.e., just before the multiplier) in figure is,

$$x(t) = s(t) + n(t)$$

The predetection filter's objective is to pass only frequencies in the vicinity of the carrier frequency, hence reducing the effect of out-of-band noise. After the predetection filter, the noise signal $n(t)$ is bandpass with a double-sided white PSD of N_o=1/2 and a bandwidth of $2W$ (centered on the carrier frequency). As a result, the predetection signal is represented using the bandpass representation.

$$x(t) = [Am(t) + n_c(t)]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

After multiplying by $2\cos(2\pi f_c t)$, this becomes,

$$\begin{aligned} y(t) &= 2\cos(2\pi f_c t)\, x(t) \\ &= Am(t)[1 + \cos(4\pi f_c t)] + n_c(t)[1 + \cos(4\pi f_c t))] - jn_s(t)\sin(4\pi f_c t) \end{aligned}$$

Where we have used $2\cos x \sin x = \sin(2x)$

Low pass filtering will remove all of the $2f_c$ frequency terms, leaving,

$$y(t) = Am(t) + n_c(t)$$

The signal power at the receiver output is,

$$P_s = E\{A^2m^2(t)\} = A^2E\{m^2(t)\} = \frac{A^2}{2}$$

Where, P is the power in the message signal $m(t)$. The power in the noise signal $n_c(t)$ is,

$$P_N = \int_{-W}^{W} \frac{N_O}{2} df = 2N_O W$$

Since from the PSD of $n_c(t)$ is No and the bandwidth of the LPF is W. Thus, for the DSB-SC synchronous demodulator, the SNR at the receiver output is,

$$\text{SNR}_O = \frac{A^2 P}{2N_O W}$$

To make a fair comparison with a baseband system, we need to calculate the transmitted power,

$$P_T = E\{Am(t)\cos(2\pi f_c t)\} = \frac{A^2 P}{2}$$

And substitution gives,

$$\text{SNR}_O = \frac{P_T}{N_O W}$$

Comparison gives,

$$\text{SNR}_{\text{DSB-SC}} = \text{SNR}_{\text{baseband}}$$

We conclude that a DSB-SC system provides no SNR performance gain over a baseband system.

It turns out that an SSB system also has the same SNR performance as a baseband system.

Noise in AM Synchronous Detection

For an AM waveform, the predetection signal is,

$$x(t) = [A + m(t) + n_C(t)]\cos(2\pi f_c t) - j\, n_s(t)\sin(2\pi f_c t)$$

After multiplication by 2 cos($2\pi f_c$t), this becomes,

$$y(t) = A[1 + \cos(4\pi f_c t] + m(t)[1 + \cos(4\pi f_c t]$$
$$+ n_c(t)[1 + \cos(4\pi f_c t] - n_s(t)\sin(2\pi f_c t)$$

After low pass filtering this becomes,

$$\tilde{y}(t) = A + m(t) + n_c(t)$$

It is worth noting that the DC term A can be readily removed using a DC block (e.g., a capacitor), and that most AM demodulators are not DC-coupled.

The signal power at the receiver output is,

$$P_s = E\{m^2(t)g = P$$

And the noise power is,

$$P_N = 2NoW$$

The SNR at the receiver output is, therefore,

$$\text{SNR}_o = \frac{P}{2N_O w}$$

The transmitted power for an AM waveform is,

$$P_T = \frac{A^2}{2} + \frac{P}{2}$$

By putting this into the baseband SNR, we can see that for a baseband system with the same transmitted power,

$$\text{SNR}_{\text{baseband}} = \frac{A^2+P}{2N_O W}$$

NOTES

As a result, we have the following equation for an AM waveform using a synchronous demodulator:

$$\text{SNR}_{\text{AM}} = \frac{P}{A^2+P} SNR_{baseband}$$

To put it another way, AM's performance is always inferior to that of a baseband system. This is due to the wasted power that occurs when the carrier is transmitted explicitly in the AM waveform.

Noise in AM Envelope Detection

Remember that an envelope detector can only be utilized if the following conditions are $\mu<1$. The **envelope** of the received signal is detected using an envelope detector. To see how this affects things, we will represent the received signal with phasors, as shown in Figure 1.12. The output of the receiver, designated by $E_i(t)$ in the diagram, will be:

$$y(t) = \text{envelope of } x(t)$$
$$= \sqrt{[A + m(t) + n_c(t)^2 + n_s^2(t)}$$

Figure 1.12 illustrates the signals present at an AM receiver are depicted as a phasor diagram.

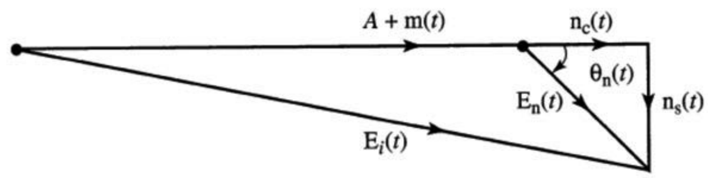


***Fig. 1.12** The Signals Present at an AM Receiver are Depicted as a Phasor Diagram*

This expression is a little more difficult to calculate the SNR at the receiver output. What must consider a $y(t)$ approximation in which the message and noise are additive.

(a) Small Noise Case

If we assume that the noise power is modest for practically all t, the receiver output can be simplified, i.e., $n(t) << [A + m(t)]$. Hence,

$$|A + m(t) + n_c(t)| >> |n_s(t)|$$

Then, most of the time,

$$y(t) \approx A + m(t) + n_c(t)$$

In the case of synchronous detection, this signal is equal to the post-detection signal. As a result, the output SNR is (again, ignoring the DC term A).

$$\text{SNR}_O = \frac{A^2 P}{2N_O W}$$

In terms of baseband SNR, this can be written as,

$$\text{SNR}_{\text{env}} = \frac{A^2}{2} + \frac{P}{2}$$
$$\text{SNR}_{\text{baseband}} = \frac{A^2+P}{2N_O W}$$

Note that whereas $\text{SNR}_{\text{AM}} = \frac{P}{A^2+P}\text{SNRbaseband}$ is valid always, the expression for SNR_{env} is only valid for small noise power.

Large Noise Case

Consider the case where the noise power is high, and we have practically all t; $n(t) >> [A + m(t)]$.

Rewrite as,

$$y^2(t) = [A + m(t) + n_c(t])^2 + n_s^2(t)$$
$$= [A + m^2(t) + n_c(t)^2 + 2Am(t) + 2An_c(t) + 2m(t)n_c(t) + n_s^2(t)$$

For $n_c(t) >> [A + m(t)]$, this reduces to,

$$y^2(t) \approx n_c^2(t) + n_s^2(t) + 2[A + m(t)]n_c(t)$$
$$= E_n^2(t)(1 + \frac{2[A+m(t)]n_c(t)}{E_n^2(t)})$$
$$E_n(t) = \sqrt{n_c^2(t) + n_s^2(t)}$$

Where,

$E_n(t)$ is the envelope of the noise (as described). But from the phasor diagram in Figure 1.12, we have, $n_c(t) = E_n(t)\cos(\theta_t)$, giving,

$$y(t) \approx E_n(t)\sqrt{1 + \frac{2[A+m(t)]\cos\theta_n(t)}{E_n(t)}}$$

Further, $\sqrt{1+x} \approx 1 + \frac{x}{2}$ for $x << 1$, so this reduces to

$$y(t) \approx E_n(t)(1 + \frac{[A+m(t)\cos\theta_n(t)]}{E_n(t)})$$
$$= E_n(t) + [A + m(t)\cos\theta_n(t)]$$

The most important point to remember is that the envelope detector's output has no component that is proportional to the message $m(t)$. The word $m(t)\cos\theta_n(t)$ refers to the message multiplied by a noise term $\cos\theta_n(t)$, and it has no bearing on $m(t)$ recovery. The message is corrupted to a significantly greater extent by this multiplicative effect than by the additive noise in our earlier analysis, resulting in a complete loss of information at the receiver. This causes a *threshold effect*, in which the detector's performance rapidly degrades below a certain carrier power level. Despite this threshold effect, we find that it has little impact in practice. This is because a signal with an output SNR of less than roughly 25 dB is of such poor quality that no one would want to listen to it. The threshold effect is rarely important for envelope detectors from a practical standpoint.

Frequency Modulation

We will now look at the SNR performance of frequency modulation systems after studying the influence of additive noise on amplitude modulation systems. There is a significant distinction between the two. The message information in AM is included within the signal's amplitude, and because noise is additive, it adds to the modulated signal directly. The message information in FM, on the other hand, is contained in the modulated signal's frequency. Because a signal's frequency may be characterised by its zero crossings, the influence of noise on an FM signal is determined by how much it alters the modulated signal's zero crossing. This implies that the influence of noise on an FM signal will be less than that on an AM signal, as we will demonstrate in this section.

Consider the general representation of a carrier waveform shown below:

$$s(t) = A\cos[\theta_i(t)]$$

The *instantaneous phase angle* is $\theta_i(t)$. We may define the *instantaneous frequency* by comparing this to the general waveform $A\cos(2\pi ft)$, where f is the frequency.

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}$$

The carrier's instantaneous frequency varies linearly with the message in an FM system, i.e.,

$$f_i(t) = f_c + k_f\, m(t)$$

Where *kf* is the *frequency sensitivity* of the modulator. Hence, the instantaneous phase is,

$$\theta_i(t) = 2\pi \int_{-\infty}^{t} f_i(\tau)d\tau = 2\pi f_c dt + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau$$

And the modulated signal is,

$$s(t) = A\cos[2\pi f_c dt + 2\pi k_f \textstyle\int_{-\infty}^{t} m(\tau)d\tau]$$

The FM signal has two characteristics to note:

(a) The Envelope is Constant.

(b) The Signal $s(t)$ is a Nonlinear Function of the Message Signal $m(t)$.

Bandwidth of FM

Let $m_p = \max|m(t)|$ be the peak message amplitude, so that the instantaneous frequency varies between $f_c - k_f m_p$ and $f_c + k_f m_p$. The *frequency deviation* is the difference between the instantaneous frequency and the carrier frequency.

$$\Delta f = k_f m_p$$

In the case of tone modulated FM, the *deviation ratio* (also known as the FM modulation index) is defined as,

$$\beta = \frac{\Delta f}{W}$$

Where, *W* is the bandwidth.

Unlike AM, FM's capacity is not only determined by the message bandwidth. The FM bandwidth for Small β is around twice that of the message bandwidth (referred to as narrowband FM). However, with large β (also known as wide-band FM), the bandwidth can be substantially larger. *Carson's rule is a useful guideline for estimating the transmission bandwidth of an FM signal:*

$B_T = 2W(\beta+1) = 2(\Delta f + W)$

Observe that for $\beta <<1$, $B_T \approx 2W$ (as is the case in AM). At the other extreme, for $\beta>>1$, $B_T \approx 2\Delta f$, which is independent of *W*.

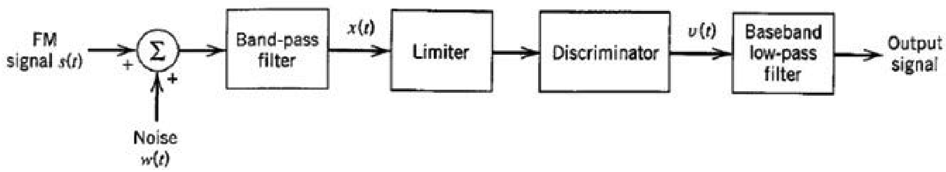


Fig. 1.13 *Model of an FM Receiver*

Noise in FM

Figure 1.13 shows the model of an FM receiver, where *s*(*t*) is the FM signal and *w*(*t*) is white Gaussian noise with power spectral density $N_O/2$. The bandpass filter is employed to remove any signals that fall outside of the bandwidth of $f_c \pm B_T/2$, and hence the receiver's predetection noise is bandpass with a bandwidth of BT. Because the envelope of an FM signal is continuous, the limiter is employed to reduce any amplitude changes. The discriminator is a device whose output is proportional to the instantaneous frequency deviation (i.e., it recovers the message signal), and the final baseband low-pass filter has a bandwidth of W, so it passes the message signal while filtering out out-of-band noise.

The predetection signal is,

$$x(t) = A\cos\left[2\pi f_c dt + 2\pi k_f \int_{-\infty}^{t} m(\tau)d\tau\right] + n_c(t)]\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

Let's start with the signal strength at the receiver output. When the predetection SNR is high, it can be demonstrated that noise has no effect on the output signal power. Thus, ignoring noise, the input signal's instantaneous frequency is,

$$f_i = f_c + k_f m(t)$$

And the discriminator's output is $k_f m(t)$ (which is supposed to merely report the instantaneous frequency's departure from the carrier frequency).

As a result, the output signal power is,

$$P_s = k_f^2 P$$

Where *P* is the average power of the message signal.

When calculating the noise power at the receiver output, it turns out that the noise output is roughly independent of the message signal at high predetection SNR. Only the carrier and noise signals are present in this situation. Thus,

$$\tilde{x}(t) = A\cos(2\pi f_c t) + n_c(t)\cos(2\pi f_c t) - n_s(t)\sin(2\pi f_c t)$$

This is depicted as a phasor diagram in Figure 1.14. We may observe from this diagram that the instantaneous phase is,

NOTES

$$\theta_i(t) = tan^{-1}\frac{n_s(t)}{A+n_c(t)},$$

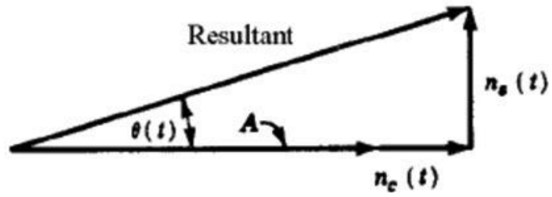


Fig. 1.14 *The FM Carrier and Noise Signals are Represented as a Phasor Diagram*

When it comes to large carrier power, most of the time,

$$\theta_i(t) = tan^{-1}\frac{n_s(t)}{A} \approx \frac{n_s(t)}{A}$$

The last sentence is derived from tan ∈≈∈ for tiny. The discriminator output, on the other hand, is the instantaneous frequency, which is given by,

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt} = \frac{1}{2\pi A}\frac{dn_s(t)}{dt}$$

We know the PSD of $n_s(t)$ shown in Figure 1.14, but what is the PSD of dn_s/dt? Fourier theory tells us that,

If, $x(t) \leftrightarrow X(f)$

Then, $$\frac{dx(t)}{dt} = j2\pi f x(f)$$

To put it another way, temporal differentiation is the same as transmitting a signal via a system with a transfer function of $H(f)=j2\pi f$. It can be demonstrated that if a signal with PSD $S_i(f)$ is fed into a linear system with transfer function $H(f)$, the PSD at the system's output is,

$$s_o(f) = |H(f)|^2 S_i(f)$$

If $n_s(t)$ has a PSD value of N_o inside the band $\pm B_T/2$, then $dn_s(t)/dt$ has a PSD $|j2\pi f|^2 N_O$. Figure 1.15 (a) and (b) illustrate the PSD of $dn_s(t)/dt$ before and after the baseband LPF.

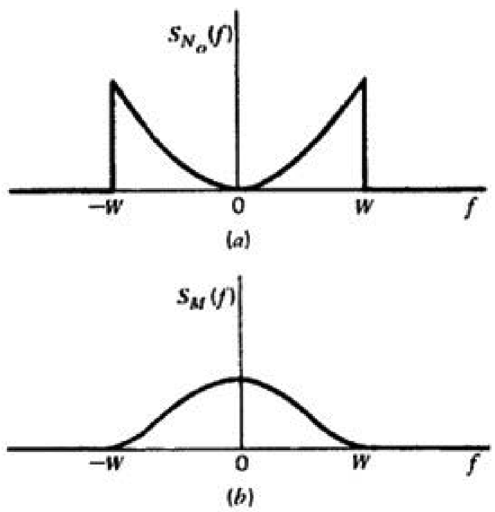


Fig. 1.15 *The Power Spectral Densities of (a) Noise at the FM Receiver's Output and (b) a Typical Message Signal*

NOTES

Returning to equation, we can determine the average noise power at the receiver output by knowing the PSD of $dn_s(t)/dt$. It is provided by,

$$P_N = \int_{-w}^{w} s_D(f)df$$

Where $S_D(f)$ is the PSD of the noise component at the discriminator output (i.e., the PSD of $f_i(t)$ in equation); the integration limits are set between $-W$ and W to reflect the output signal's low-pass filtering.

Thus,

$$P_N = \int_{-w}^{w}\left(\frac{1}{2\pi A}\right)^2 (2\pi f)^2 N_O df = \int_{-w}^{w}\frac{N_O}{A^2}(f)^2 df = \frac{2N_O w^3}{3A^2}$$

This calculation is critical because it demonstrates that the average noise power at an FM receiver's output is inversely proportional to the carrier power $A^2/2$. As a result, increasing the carrier power has a *noise quieting* effect. This is a significant benefit of FM systems.

Finally, we have that the SNR is at the output.

$$SNR_O = \frac{3A^2K_f^2 p}{2N_O W^3}$$

Due to the fact that the transmitted power of an FM waveform is,

$$P_T = \frac{A^2}{2}$$

Substitution into earlier equation gives,

$$\text{SNR}_{\text{FM}} = \frac{3K_f^2 p}{w^2} SNR_{baseband} = 3\beta^2 \frac{P}{m_p^2} SNR_{baseband}$$

The SNR expression assumes that the carrier power is greater than the noise power. The FM detector, like the AM envelope detector, exhibits a *threshold effect*. As carrier power diminishes, the FM receiver fails, as Haykin describes: "Initially, individual clicks are detected in the receiver output, but as the carrier-to-noise ratio decreases further, the clicks swiftly merge into a crackling or spitting sound". Experiments reveal that this noise mutilation is insignificant in the majority of cases where the predetection SNR (i.e., immediately after the receiver bandpass filter) is more than 10. In other terms, the threshold point occurs approximately,

$$\frac{A^2}{2N_O\beta_T} = 10$$

Where, recall $B_T = 2W(\beta+1)$. For predetection SNRs above this value, the output SNR is given by equation.

It is worth noting that, while equation implies that the output SNR of an FM system can be improved arbitrarily by increasing β while maintaining the signal power constant, examination of the equation reveals that this is not absolutely true. The reason for this is that if β increases too much, the condition in equation that we are above threshold may become false, implying that equation no longer yields an expression for the genuine SNR.

1.12.6 Pre-Emphasis and De-Emphasis

Pre-emphasis refers to increasing the modulating voltage's relative amplitudes at higher audio frequencies between 2 and approximately 15 KHz, and de-emphasis means decreasing those frequencies by the amount by which they are boosted. However, the transmitter does the pre-emphasis and the receiver performs the de-emphasis. The goal of this modification is to increase the signal-to-noise ratio of FM reception. The RC or L/Z network specifies a time constant of 75μs for pre- and de-emphasis.

Pre-Emphasis

As previously stated, noise has a greater effect on higher modulating frequencies in FM. For higher modulating frequencies (f_m), this impact can be mitigated by raising the value of the modulation index (f_m). This can be done by increasing the deviation Δf and Δf can be increased by increasing the amplitude of modulating signal at higher modulating frequencies. Thus, by artificially increasing the amplitude of higher frequency modulating signals, we can improve noise immunity at higher modulating frequencies. Pre-emphasis is a term that refers to the artificial enhancement of higher modulating frequencies. The pre-emphasis circuit depicted in Figure 1.16 is used to boost the higher frequency modulating signal.

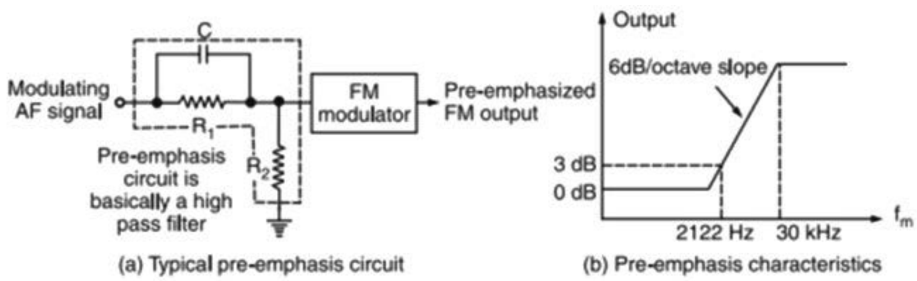


Fig. 1.16 *Circuit for Pre-Emphasis and Its Properties*

The modulating AF signal is processed via a high pass RC filter before being applied to the FM modulator, as shown in Figure 1.16(a). As *fm* increases, reactance of C decreases and modulating voltage applied to FM modulator goes on increasing. Figure 1.16 depicts the RC high pass network's frequency response characteristic (b). The boosting is carried out in accordance with this pre-determined curve. In the United States, the amount of pre-emphasis in FM transmission and sound transmission in television has been defined at 75 μ sec. A high pass filter serves as the basis for the pre-emphasis circuit. At the transmitter, the pre-emphasis is done. Figure 1.16(b) shows the frequency of the RC high pass network, which is 2122 Hz. As a result, as shown in Figure 1.16, the pre-emphasis circuit is employed at the transmitter.

The energy content of higher-frequency signals is increased by this pre-emphasis circuit, which causes them to become stronger than high-frequency noise components. This promotes intelligibility and fidelity while improving the signal-to-noise ratio.

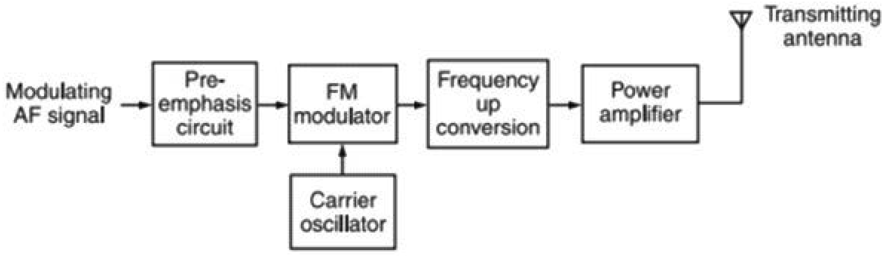


Fig. 1.17 *The Pre-Emphasis is Included in the FM Transmitter*

De-Emphasis

De-emphasis is a technique used at the receiver end to cancel out or compensate for the artificial boosting imparted to the higher modulating frequencies during the pre-emphasis process. That is, the de-emphasis circuit restores the artificially amplified high frequency signals to their original amplitude. The de-emphasis circuit of 75 μ sec is standard, as seen in Figure 1.18. It clearly indicates that it is a low pass filter. A frequency response curve that is 3 dB lower at a frequency with a 75 μ sec RC time constant equates to a 75 μ sec de-emphasis. i.e.,

$$f = \frac{1}{2\pi RC} = \frac{1}{2\pi \times 75 \times 10^{-6}} = 2{,}122 \text{ Hz.}$$

The FM signal is applied to the De-emphasis circuit after it has been demodulated. As f_m increases, the reactance of C decreases and the output of the de-emphasis circuit decreases as well, as illustrated in Figure 1.18.

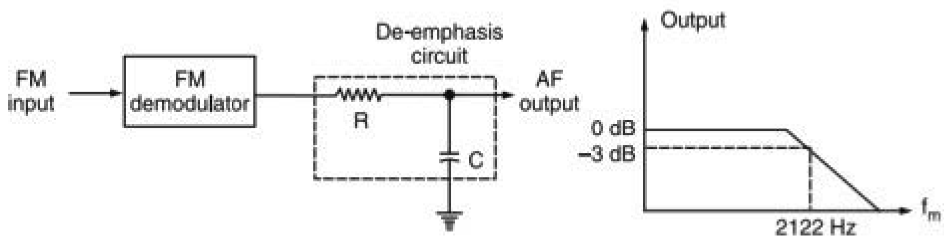


Fig. 1.18 *De-Emphasis Circuit and its Characteristics*

The combined effect of pre-emphasis and de-emphasis is to boost the strength of the high-frequency components during transmission, preventing them from being hidden by noise. Due to pre-emphasis and de-emphasis, the *S/N* ratio at the receiver's output is kept constant. Figure 1.19 illustrates the combined frequency response characteristic of the pre-emphasis and de-emphasis circuits.

NOTES

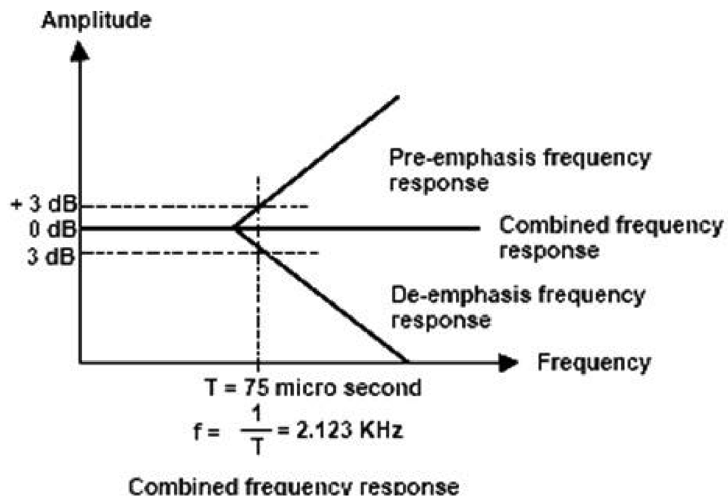


Fig. 1.19 Combined Frequency Response of Pre-Emphasis and De-Emphasis

Check Your Progress

16. Why is convolution used?
17. When is system's impulse response function h(t) used?
18. Why is convolution integral used?
19. What is Parseval's identity?
20. Define the effect of noise on radio systems.
21. How is noisy two port network adds noise to the transmission?
22. How is noise in analog communication systems evaluated?
23. Define the terms pre-emphasis and de-emphasis.
24. What is the combined effect of pre-emphasis and de-emphasis?

1.13 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A Fourier series is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic. As such, the summation is a synthesis of another function. A Fourier series, however, can be used only for periodic functions, or for functions on a bounded (compact) interval.

   A non-sinusoidal periodic function can be expressed by a sum of a set of sinusoidal oscillating fruitions or periodic functions, i.e., mainly sine and cosine functions or complex exponentials. The expansion of the non-sinusoidal periodic function is known as the 'Fourier Series'.

2. The Fourier series expansion of every piecewise smooth function is expressed on a finite interval. There are many different ways to express a function in Fourier space, the most common being by expanding it into a series of sines

and cosines.

i.e., $f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx$

The coefficients a_o, a_n and *bn* are the Fourier coefficients of function $f(x)$.

3. The Fourier series is an infinite series expansion involving trigonometric functions. A periodic waveform $f(t)$ of Period $p = 2L$ has a Fourier series given by:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi t}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi t}{L}\right)$$

$$= \frac{a_0}{2} + a_1 \cos\left(\frac{\pi t}{L}\right) + a_2 \cos\left(\frac{2\pi t}{L}\right) + a_3 \cos\left(\frac{3\pi t}{L}\right) + \ldots + b_1 \sin\left(\frac{\pi t}{L}\right) + b_2 \sin\left(\frac{2\pi t}{L}\right)$$

$$+ b_3 \sin\left(\frac{3\pi t}{L}\right) + \cdots$$

Where, a_n and b_n are the Fourier coefficients.

And, $a_0/2$ is the mean value, sometimes referred to as the DC level.

4. The process of converting a continuous time signal, typically not quantized, to a discrete time signal is known as sampling, usually quantized. It is also known as the discretization of the process of measuring the instantaneous values of a continuous time signal. A sample is a portion of data that is continuous in the time domain and is taken from the entire data set. When a source creates an analogue signal that must be digitized, the signal must be discretized in time, using 1s and 0s, i.e., High or Low. Sampling is the process of discretizing an analogue signal.

5. Assume a signal is band-limited, with no frequency components exceeding *W* Hertz. *W* stands for the highest frequency. For such a signal, the sampling rate should be twice the highest frequency for effective replication of the original signal, which means $f_S = 2W$, where f_S is the sampling rate and *W* is the highest frequency. This sampling rate is referred to as the Nyquist rate.

6. Consider the responses of a system to the inputs $x_1(t)$ and $x_2(t)$, $y_1(t)$ and $y_2(t)$, respectively. The system is said to have the additivity property if the response to the input $x_1(t) + x_2(t)$ is $y_1(t) + y_2(t)$ for any choice of $x_1(t)$ and $x_2(t)$. Let $y(t)$ represent a system's response to the input $x(t)$, and a represent a complex constant. If, for any choice of $x(t)$ and a, the response to the input $ax(t)$ is $ay(t)$, the system is said to possess the homogeneity property. If a system possesses both the additivity and homogeneity properties, it is said to be linear.

7. Characteristically, the normalized power in a periodic signal can be defined. Normalized power of a waveform is equal to zeroth coefficient in the exponential Fourier series of its squared version.

$$P_n = \frac{1}{T}\int_{-0.5T}^{0.5T} [v(t)]^2 \, dt.$$

NOTES

8. The multiplication-in-time property of Fourier series states that if $v_1(t)$ and $v_2(t)$ are two periodic waveforms with same period and $v_3(t) = v_1(t) \times v_2(t)$, then, the exponential Fourier series coefficients of $v_3(t)$ is given by:

$$\tilde{v}_{3k} = \sum_{n=-\infty}^{\infty} \tilde{v}_{1n}\tilde{v}_{2(k-n)} \text{ for } -\infty < k < \infty,$$

Where $\tilde{\upsilon}_{1n}$ and $\tilde{\upsilon}_{2n}$ are the exponential Fourier series coefficients of $v_1(t)$ and $v_2(t)$, respectively. We use this property with $v_1(t) = v_2(t) = v(t)$ and evaluate the exponential Fourier series coefficient of,

$$[v(t)]^2 \text{ for } k = 0 \text{ as } \tilde{v}_{30} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_{-n}.$$

Therefore,

$$P_n = \frac{1}{T}\int_{-0.5T}^{0.5T} [v(t)]^2 dt = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_{-n} = \sum_{n=-\infty}^{\infty} \tilde{v}_n \tilde{v}_n^* = \sum_{n=-\infty}^{\infty} |\tilde{v}_n|^2 = |\tilde{v}_0|^2 + 2\sum_{n=1}^{\infty} |\tilde{v}_n|^2.$$

This is Parseval's Theorem on normalized power of periodic waveforms.

9. A Power Spectral Density (PSD) is a metric that compares the power content of a signal to its frequency. Broadband random signals are often described using a PSD. The spectral resolution used to digitize the signal is used to normalize the PSD's amplitude. PSD depicts the variation's strength (energy) as a function of frequency. In other words, it displays which frequencies have high variations and which frequencies have minor variations. PSD is measured in energy (variance) per frequency (width), and energy within a certain frequency range can be obtained by integrating PSD within that frequency range.

10. The energy spectral density quantifies the frequency dependence of the energy in a signal or time series. The term 'Energy' is used here in a broad sense of signal processing, i.e., the energy E of a signal $x(t)$ is as follows:

$$E \triangleq \int_{-\infty}^{\infty} |x(t)|^2 \, dt.$$

11. A Fourier Transform (FT) is a mathematical transform that decomposes functions depending on space or time into functions depending on spatial or temporal frequency, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes. The term Fourier transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of space or time.

12. Fourier transform is the special case of Laplace transform which is evaluated keeping the real part zero. Generally, the Fourier transform is used for analysis in frequency domain whereas Laplace transform is typically used for analysis in s-domain, it is not frequency domain. The Fourier transform helps to study and analyse anything in the frequency domain whereas Laplace transform is specifically done for complex analysis, when anything is not

easy and simple to analyse in time domain, then we convert it into s domain and then take the inverse Laplace transform to complete the analysis.

13. The Dirac delta distribution (δ distribution), also known as the unit impulse symbol, is a generalized function or distribution over the real numbers, whose value is zero everywhere except at zero, and the integral over the entire real line is equal to one. The delta function was introduced by physicist Paul Dirac as a tool for the normalization of state vectors. It also has uses in probability theory and signal processing. The Kronecker delta function, which is usually defined on a discrete domain and takes values 0 and 1, is the discrete analog of the Dirac delta function.

14. The Dirac delta can be roughly thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

And which is also constrained to satisfy the identity,

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1.$$

This is merely a heuristic characterization. The Dirac delta is not a function in the traditional sense as no function defined on the real numbers has these properties. The Dirac delta function can be rigorously defined either as a distribution or as a measure.

15. A gate function is defined as rectangular pulses of unit height and unit width, centred at the origin.

16. Convolution can be used to determine a system's zero-state response (i.e., the reaction to an arbitrary input when the system's initial conditions are zero) to an arbitrary input by using the system's impulse response.

17. In the temporal domain, a system's Impulse Response Function $h(t)$ is used to describe it. This function expresses the system's response at time t to a unit impulse or -function input given at time $t = 0$.

18. In the physical sciences, the convolution integral is often used. $f_1(t) * f_2(t)$ is the symbol for the convolution integral of two functions $f_1(t)$ and $f_2(t)$.

$$f_1(t) * f_2(t) \equiv \int_{-\infty}^{\infty} f_1(\tau) f_2(t-\tau) d\tau$$

19. Parseval's identity, named after Marc-Antoine Parseval, is a fundamental result on the summability of the Fourier series of a function. Geometrically, it is generalised Pythagorean theorem for inner-product spaces (which can have an uncountable infinity of basis vectors).

20. The Signal-to-Noise Ratio, or SNR, is defined as the ratio of signal power to noise power in a radio link. It is used to measure the link's quality.

$$SNR = \frac{\text{average signal power}}{\text{average noise power}}.$$

SNR is frequently stated in decibels (dB) because it is a ratio ($SNR_{dB} = 10 \log SNR$).

21. The noisy two port network adds noise to the transmission. The network's noise is mimicked by an additional thermal noise generator operating at an equivalent temperature T_e and adding a power $kGBT_e$ to the network's output.

22. The noise in analog communication systems can be evaluated through the performance of several analogue modulation methods in the presence of noise. The receiver's performance will be quantified in terms of the Signal-to-Noise Ratio (SNR) at the receiver's output, defined as,

$$SNR_o = \frac{average\ power\ of\ message\ signal\ at\ reciver\ output}{average\ power\ of\ noise\ atthe\ reciver\ output}$$

In some conditions, this is may not be the case, and the approximation methods can be used to figure out what the message and noise were doing at the receiver output.

23. Pre-emphasis refers to increasing the modulating voltage's relative amplitudes at higher audio frequencies between 2 and approximately 15 KHz, and de-emphasis means decreasing those frequencies by the amount by which they are boosted. However, the transmitter does the pre-emphasis and the receiver performs the de-emphasis.

De-emphasis is a technique used at the receiver end to cancel out or compensate for the artificial boosting imparted to the higher modulating frequencies during the pre-emphasis process. That is, the de-emphasis circuit restores the artificially amplified high frequency signals to their original amplitude.

24. The combined effect of pre-emphasis and de-emphasis is to boost the strength of the high-frequency components during transmission, preventing them from being hidden by noise. Due to pre-emphasis and de-emphasis, the *S/N* ratio at the receiver's output is kept constant.

1.14 SUMMARY

- A Fourier series is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic. As such, the summation is a synthesis of another function.
- The discrete time Fourier transform is an example of Fourier series. The process of deriving weights that describe a given function is a form of Fourier analysis.
- For functions on unbounded intervals, the analysis and synthesis analogies are Fourier transform and inverse transform.

- Fourier originally defined the Fourier series for real valued functions of real arguments and using the sine and cosine functions as the basis set for the decomposition.
- A Fourier series, however, can be used only for periodic functions, or for functions on a bounded (compact) interval.
- A non-sinusoidal periodic function can be expressed by a sum of a set of sinusoidal oscillating fruitions or periodic functions, i.e., mainly sine and cosine functions or complex exponentials. The expansion of the non-sinusoidal periodic function is known as the 'Fourier Series'.
- The Fourier series expansion of every piecewise smooth function is expressed on a finite interval. There are many different ways to express a function in Fourier space, the most common being by expanding it into a series of sines and cosines.

  i.e., $$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos nx+\sum_{n=1}^{\infty}b_n\sin nx$$

  The coefficients a_o, a_n and bn are the Fourier coefficients of function $f(x)$.
- If the values of a function repeat after an equal interval of x, the function is said to be periodic.

  i.e., $$f(x) = f(x + \mathrm{T}) = f(x + 2\mathrm{T}) = f(x + 3\mathrm{T})\ldots\ldots..$$

  $f(x)$is therefore said to be the periodic fraction of x with period T.
- When a function and its derivatives are continuous, then the function can be expanded by power of x by Taylor series. However, under specific conditions, one may expand both continuous and discontinuous forms of functions employing Fourier series.
- The Fourier series is an infinite series expansion involving trigonometric functions. A periodic waveform $f(t)$ of Period $p = 2L$ has a Fourier series given by:

  $$f(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos\left(\frac{n\pi t}{L}\right)+\sum_{n=1}^{\infty}b_n\ \sin\left(\frac{n\pi t}{L}\right)$$

  $$=\frac{a_0}{2}+a_1\cos\left(\frac{\pi t}{L}\right)+a_2\cos\left(\frac{2\pi t}{L}\right)+a_3\cos\left(\frac{3\pi t}{L}\right)+\ldots+b_1\sin\left(\frac{\pi t}{L}\right)+b_2\sin\left(\frac{2\pi t}{L}\right)$$

  $$+b_3\sin\left(\frac{3\pi t}{L}\right)+\ldots$$

  Where, a_n and b_n are the Fourier coefficients.

  And, $a_0/2$ is the mean value, sometimes referred to as the DC level.
- The process of converting a continuous time signal, typically not quantized, to a discrete time signal is known as sampling, usually quantized. It is also known as the discretization of the process of measuring the instantaneous values of a continuous time signal.
- A sample is a portion of data that is continuous in the time domain and is taken from the entire data set. When a source creates an analogue signal

NOTES

that must be digitized, the signal must be discretized in time, using 1s and 0s, i.e., High or Low. Sampling is the process of discretizing an analogue signal.

- The sample frequency is also known as the sampling rate. The sampling rate refers to the number of samples obtained per second or for a specific set of values.
- Assume a signal is band limited, with no frequency components exceeding W Hertz. W stands for the highest frequency. For such a signal, the sampling rate should be twice the highest frequency for effective replication of the original signal, which means $f_S = 2W$, where f_S is the sampling rate and W is the highest frequency. This sampling rate is referred to as the Nyquist rate.
- The sampling theorem, often known as the Nyquist theorem, establishes the theory of a sufficient sample rate in terms of bandwidth for the class of band-limited functions. According to the sampling theorem, "A signal can be accurately reproduced if it is sampled at a rate fs greater than twice its highest frequency W".
- Aliasing is defined as, "The phenomenon in which a signal's high-frequency component assumes the identity of a low-frequency component in the spectrum of its sampled form".
- The system is said to have the additivity property if the response to the input $x_1(t) + x_2(t)$ is $y_1(t) + y_2(t)$ for any choice of $x_1(t)$ and $x_2(t)$.
- Let $y(t)$ represent a system's response to the input $x(t)$, and a represent a complex constant. If, for any choice of $x(t)$ and a, the response to the input $ax(t)$ is $ay(t)$, the system is said to possess the homogeneity property.
- If a system possesses both the additivity and homogeneity properties, it is said to be linear.
- The concept of normalized power is defined on the basis of Parseval's theorem and normalized power in a periodic waveform which is often employed in 'Communication Engineering' and allied areas as a measure of signal strength. Typically, it is defined as the average power that can be transported specifically to 1Ω resistance when the periodic waveform is considered as a voltage waveform applied to that specific resistor.
- Characteristically, the normalized power in a periodic signal can be defined. Normalized power of a waveform is equal to zeroth coefficient in the exponential Fourier series of its squared version.

$$P_n = \frac{1}{T}\int_{-0.5T}^{0.5T} [v(t)]^2\, dt.$$

- The multiplication-in-time property of Fourier series states that if $v_1(t)$ and $v_2(t)$ are two periodic waveforms with same period and $v_3(t) = v_1(t) \times v_2(t)$, then, the exponential Fourier series coefficients of $v_3(t)$ is given by:

$$\tilde{v}_{3k} = \sum_{n=-\infty}^{\infty} \tilde{v}_{1n}\tilde{v}_{2(k-n)} \text{ for } -\infty < k < \infty,$$

NOTES

Where $\tilde{\upsilon}_{1n}$ and $\tilde{\upsilon}_{2n}$ are the exponential Fourier series coefficients of $v_1(t)$ and $v_2(t)$, respectively.

- A Power Spectral Density (PSD) is a metric that compares the power content of a signal to its frequency. Broadband random signals are often described using a PSD. The spectral resolution used to digitize the signal is used to normalize the PSD's amplitude.
- PSD depicts the variation's strength (energy) as a function of frequency. In other words, it displays which frequencies have high variations and which frequencies have minor variations.
- PSD is measured in energy (variance) per frequency (width), and energy within a certain frequency range can be obtained by integrating PSD within that frequency range.
- The energy spectral density quantifies the frequency dependence of the energy in a signal or time series. The term 'Energy' is used here in a broad sense of signal processing, i.e., the energy E of a signal $x(t)$ is as follows:

$$E \triangleq \int_{-\infty}^{\infty} |x(t)|^2 \, dt.$$

- Fourier transform is the special case of Laplace transform which is evaluated keeping the real part zero. Generally, the Fourier transform is used for analysis in frequency domain whereas
- Laplace transform is typically used for analysis in s-domain, it is not frequency domain.
- The Fourier transform helps to study and analyse anything in the frequency domain whereas Laplace transform is specifically done for complex analysis, when anything is not easy and simple to analyse in time domain, then we convert it into s domain and then take the inverse Laplace transform to complete the analysis.
- A Fourier Transform (FT) is a mathematical transform that decomposes functions depending on space or time into functions depending on spatial or temporal frequency, such as the expression of a musical chord in terms of the volumes and frequencies of its constituent notes. The term Fourier transform refers to both the frequency domain representation and the mathematical operation that associates the frequency domain representation to a function of space or time.
- The Fourier transform of a function of time is a complex valued function of frequency, whose magnitude (absolute value) represents the amount of that frequency present in the original function, and whose argument is the phase offset of the basic sinusoid in that frequency.
- The Fourier transform is not limited to functions of time, but the domain of the original function is commonly referred to as the time domain. There is

NOTES

also an inverse Fourier transform that mathematically synthesizes the original function from its frequency domain representation, as proven by the Fourier inversion theorem.

- The Dirac delta distribution (δ distribution), also known as the unit impulse symbol, is a generalized function or distribution over the real numbers, whose value is zero everywhere except at zero, and the integral over the entire real line is equal to one.
- The delta function was introduced by physicist Paul Dirac as a tool for the normalization of state vectors. It also has uses in probability theory and signal processing. The Kronecker delta function, which is usually defined on a discrete domain and takes values 0 and 1, is the discrete analog of the Dirac delta function.
- The Dirac delta can be roughly thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

And which is also constrained to satisfy the identity,

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1.$$

This is merely a heuristic characterization.

- The Dirac delta is not a function in the traditional sense as no function defined on the real numbers has these properties. The Dirac delta function can be rigorously defined either as a distribution or as a measure.
- In the temporal domain, a system's impulse response function $h(t)$ is used to describe it. This function expresses the system's response at time t to a unit impulse or -function input given at time $t = 0$.
- In the physical sciences, the convolution integral is often used. $f_1(t) * f_2(t)$ is the symbol for the convolution integral of two functions $f_1(t)$ and $f_2(t)$.

$$f_1(t) * f_2(t) \equiv \int_{-\infty}^{\infty} f_1(\tau) f_2(t-\tau) d\tau$$

- Parseval's identity, named after Marc-Antoine Parseval, is a fundamental result on the summability of the Fourier series of a function. Geometrically, it is generalised Pythagorean theorem for inner-product spaces (which can have an uncountable infinity of basis vectors).
- All electronic systems contain noise, which is defined as voltage and current variations induced by the random motion of charged particles.
- In RF (Radio Frequency) and microwave receivers, which must extract information from extremely small signals, knowing noise and how it propagates through a system is especially important.

NOTES

- Noise introduced by circuit parts can mask or obfuscate low-level signals, impairing speech or video reception, causing bit recognition in digital systems to be uncertain, and causing radar mistakes. For RF and microwave engineers, measuring the noise contributions of circuit elements in the form of noise factor or noise figure is a critical task.
- The antenna picks up external sources of noise. Atmospheric noise, galactic/cosmic noise, man-made noise, interference 'Noise' generated by other users in a nearby channel (Adjacent Channel Interference, ACI) or the same channel (Co-Channel Interference, CCI), and so on are examples.
- Components inside the receiver emit internal noise. Random processes like as charge flow in a device, or, at a more fundamental level, thermal vibrations in any component at a temperature above absolute zero, cause this noise.
- Radio receivers are made up of noise-generating components. Noise is generated by all components, whether passive (such as, resistors) or active (transistor-based circuits). The device's useful operating range is limited by noise in active components.
- Thermal noise, which results from the thermal vibration of bound charges. It's also known as Johnson or Nyquist noise, and it's the most common type.
- Shot noise occurs when charge carriers in an active device, such as a solid-state device or an electron tube, fluctuate randomly.
- Recombination of charge carriers causes flicker noise, also known as $1/f$ noise, which can be found in active components.
- When compared to the above noise sources, quantum noise, which is created by the discrete or quantized nature of charge carriers and photons, is generally small.
- The Signal-to-Noise Ratio, or SNR, is defined as the ratio of signal power to noise power in a radio link. It is used to measure the link's quality.

$$SNR = \frac{Average\ Signal\ Power}{Average\ Noise\ Power}$$

- SNR is frequently stated in decibels (dB) because it is a ratio ($SNR_{dB} = 10 \log SNR$).
- Thermal noise is generated by resistors due to the random fluctuations of their internal molecules. It is frequently beneficial to consider a resistor's noise contribution in terms of its equivalent noise voltage or equivalent noise current.
- Interference is a term that refers to artificial signals. Telephone lines are susceptible to interference from electricity lines (in the United States a distorted 60 Hz sinusoid).
- The noisy two port network adds noise to the transmission. The network's noise is mimicked by an additional thermal noise generator operating at an equivalent temperature T_e and adding a power $kGBT_e$ to the network's output.

- At the network's output, the signal power is equal to $S_o = GS_i$. The total noise power, after the gain of the network has acted on it, is $kGB(T_0 + T_e)$, because the gain of the network has acted on the input noise power kBT_0. The network's noise figure is calculated as,

$$F = \frac{S_i}{kT_0B} \cdot \frac{kGB(T_0 + T_e)}{GS_i} = 1 + \frac{T_e}{T_0} \geq 1.$$

  This enables us to express the network's noise temperature as,

$$T_e = (F - 1)T_0.$$

- The noise in analog communication systems can be evaluated through the performance of several analogue modulation methods in the presence of noise. The receiver's performance will be quantified in terms of the Signal-to-Noise Ratio (SNR) at the receiver's output, defined as,

$$\text{SNR}_o = \frac{\textit{average power of message signal at reciver output}}{\textit{average power of noise atthe reciver output}}$$

- Carson's rule is a useful guideline for estimating the transmission bandwidth of an FM signal:

$$B_T = 2W(\beta+1) = 2(\Delta f + W)$$

- Pre-emphasis refers to increasing the modulating voltage's relative amplitudes at higher audio frequencies between 2 and approximately 15 KHz, and de-emphasis means decreasing those frequencies by the amount by which they are boosted.
- However, the transmitter does the pre-emphasis and the receiver performs the de-emphasis. The goal of this modification is to increase the signal-to-noise ratio of FM reception. The RC or L/Z network specifies a time constant of 75μs for pre-emphasis and de-emphasis.
- The energy content of higher-frequency signals is increased by this pre-emphasis circuit, which causes them to become stronger than high-frequency noise components. This promotes intelligibility and fidelity while improving the signal-to-noise ratio.
- De-emphasis is a technique used at the receiver end to cancel out or compensate for the artificial boosting imparted to the higher modulating frequencies during the pre-emphasis process. That is, the de-emphasis circuit restores the artificially amplified high frequency signals to their original amplitude.
- The combined effect of pre-emphasis and de-emphasis is to boost the strength of the high-frequency components during transmission, preventing them from being hidden by noise. Due to pre-emphasis and de-emphasis, the *S/N* ratio at the receiver's output is kept constant.

1.15 KEY TERMS

- **Fourier series:** A Fourier series is a periodic function composed of harmonically related sinusoids combined by a weighted summation. With appropriate weights, one cycle or period of the summation can be made to approximate an arbitrary function in that interval or the entire function if it too is periodic.
- **Sampling:** The process of converting a continuous time signal, typically not quantized, to a discrete time signal is known as sampling, usually quantized. It is also known as the discretization of the process of measuring the instantaneous values of a continuous time signal.
- **Aliasing:** Aliasing is defined as the phenomenon in which a signal's high-frequency component assumes the identity of a low-frequency component in the spectrum of its sampled form.
- **Power Spectral Density (PSD):** A Power Spectral Density (PSD) is a metric that compares the power content of a signal to its frequency. Broadband random signals are often described using a PSD. The spectral resolution used to digitize the signal is used to normalize the PSD's amplitude.
- **Fourier Transform:** The Fourier transform of a function of time is a complex valued function of frequency, whose magnitude (absolute value) represents the amount of that frequency present in the original function, and whose argument is the phase offset of the basic sinusoid in that frequency.
- **Dirac delta distribution:** The Dirac delta distribution (δ distribution), also known as the unit impulse symbol, is a generalized function or distribution over the real numbers, whose value is zero everywhere except at zero, and the integral over the entire real line is equal to one. The delta function was introduced by physicist Paul Dirac as a tool for the normalization of state vectors.
- **Convolution:** Convolution can be used to determine a system's zero-state response (i.e., the reaction to an arbitrary input when the system's initial conditions are zero) to an arbitrary input by using the system's impulse response.
- **Parseval's identity:** Parseval's identity, named after Marc-Antoine Parseval, is a fundamental result on the summability of the Fourier series of a function. Geometrically, it is generalised Pythagorean theorem for inner-product spaces (which can have an uncountable infinity of basis vectors).
- **Pre-emphasis:** Pre-emphasis refers to increasing the modulating voltage's relative amplitudes at higher audio frequencies between 2 and approximately 15 KHz.
- **De-emphasis:** De-emphasis is a technique used at the receiver end to cancel out or compensate for the artificial boosting imparted to the higher modulating frequencies during the pre-emphasis process, i.e., the de-emp hasis circuit restores the artificially amplified high frequency signals to their original amplitude.

1.16 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. What is signal?
2. Define Fourier series.
3. State the sampling function.
4. Define about the response of linear system.
5. What is normalized power in Fourier expansion?
6. What is power spectral density?
7. State the relationship between Fourier and Laplace transforms.
8. Name the transformation theorems.
9. State the Dirac Delta function.
10. What is convolution?
11. Define impulse response and convolution integral.
12. State the Parseval's theorem.
13. What are the physical sources of noise?
14. Differentiate between the external noise and the internal noise.
15. Name the types of noise.
16. How is noise calculation done?
17. What is noise temperature?
18. Define the noisy two port network.
19. What is interference in AM and PM?
20. State about pre-emphasis and de-emphasis.

Long-Answer Questions

1. Briefly discuss the signals with reference to Fourier series giving appropriate examples.
2. How the Fourier coefficient can be evaluated? Explain the Fourier series' nature for even and odd functions giving examples.
3. Explain sampling function, sampling theorem, sampling rate and Nyquist rate with the help of relevant examples.
4. Describe response of linear system by means of continuous Linear Time Invariant (LTI) system giving examples.
5. Discuss briefly the significance of normalized power giving examples.
6. Brief a detailed note on the Parseval's theorem and normalized power in a periodic waveform.

7. Discuss the power spectral density and energy density function giving relevant examples.
8. Explain briefly the transformation theorems for Fourier transform and Laplace transform.
9. Describe the Dirac Delta function giving definition.
10. Briefly discuss about the significance of convolution, impulse response, convolution integral and physical interpretation on convolution giving relevant examples.
11. Explain the Parseval's theorem.
12. Discuss in detail the significant properties of noise and physical sources of noise with the help of examples.
13. Brief a detailed note on external noise, internal noise, shot noise, 1 /f noise and thermal noise.
14. Explain the concept of noise calculations, noise temperature, and noisy two port network giving appropriate examples.
15. Elaborate on the interference and noise in Amplitude Modulation (AM) and Pulse Modulation (PM).
16. Analyse how demodulation happens in the presence of noise? Explain with the help of examples.
17. Discuss the characteristic features of pre-emphasis and de-emphasis giving examples.
18. Analyse the current i(amp) given by the table below ints onts constituent upto III harmonic.

$\theta°$:	0	30	60	90	120	150	180	210	240	270	300	330
i:	.0	24	33.5	27.5	18.2	13	0	–24	–33.5	–27.5	–18.2	–13

19. Find the Fourier series upto III harmonic.

x :	0	$\frac{\pi}{3}$	$\frac{2\pi}{3}$	π	$\frac{4\pi}{3}$	$\frac{5\pi}{3}$	2π
y :	0.8	0.6	0.4	0.7	0.9	1.1	0.8

(ii)

x :	0	30	60	90	120	150	180	210	240	270	300	330
y :	25	40	50.5	57.5	61.5	63.3	68.2	59.2	52.2	44.2	35.8	28.7

20. The turning moment T on the crank-shaft of a steam engine for the crank angle θ (degrees)

θ :	0°	15°	30°	45°	60°	75°	90°	105°	120°	135°	150°	165°	180°
T:	0	2.7	5.7	7	8.1	8.3	7.9	6.8	5.5	4.1	2.6	1.2	0

Expand T in a series of sines upto second harmonic.

1.17 FURTHER READING

Gurung, J.B. 2009. *Signals and Systems*. New Delhi: PHI Learning Pvt. Ltd.

Carlson, A Bruce and Paul B. Crilly. 2010. *Communication System*. Noida: Tata McGraw-Hill Education.

Lathi, B. P. and Zhi Ding. 2010. *Modern Digital and Analog Communication Systems*. Oxford: Oxford University Press.

Kennedy, George and Bernard Davis. 1993. *Electronic Communication Systems*. London: Glencoe/McGraw-Hill.

Roddy, Dennis and John G. Coolen. 1977. *Electronic Communications*. Noida: Pearson Education.

Greenwich, Edwin. 2012. *Analog Integrated Circuits*. Berlin: Springer Science & Business Media.

Millman, Jacob and Christos C. Halkias. 2001. *Integrated Electronics*. Noida: Tata McGraw-Hill Education.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey: Prentice-Hall.

Tanenbaum, Andrew. S. 2002. *Computer Networks*. New Jersey: Prentice-Hall PTR.

Basavaraj, B. and H.N. Shivashankar. 2004. *Basic Electronics*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 2 AMPLITUDE MODULATION AND DEMODULATION

NOTES

Structure

2.0 INTRODUCTION

Amplitude Modulation (AM) is a modulation technique used in electronic communication, most commonly for transmitting messages with a radio wave. 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. AM was the earliest modulation method used for transmitting audio in radio broadcasting. It was developed during the first quarter of the 20th century beginning with Roberto Landell de Moura and Reginald Fessenden's radiotelephone experiments in 1900. This original form of AM is sometimes called Double SideBand Amplitude Modulation (DSBAM), because the standard method produces sidebands on either side of the carrier frequency. Single SideBand (SSB) modulation uses bandpass filters to eliminate one of the sidebands and possibly the carrier signal, which improves the ratio of message power to total transmission power, reduces power handling requirements of line repeaters, and permits better bandwidth utilization of the transmission medium.

In electronics and telecommunications, modulation means varying some aspect of a continuous wave carrier signal with an information bearing modulation waveform, such as an audio signal which represents sound, or a video signal which represents images. In this sense, the carrier wave, which has a much higher frequency than the message signal, carries the information. At the receiving station, the message signal is extracted from the modulated carrier by demodulation. The simplest form of AM demodulator consists of a diode which is configured to act as envelope detector. Another type of demodulator, the product detector, can provide better quality demodulation with additional circuit complexity.

Demodulation is extracting the original information bearing signal from a carrier wave. A demodulator is an electronic circuit 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).

In this unit, you will study about the Amplitude Modulation (AM), necessity of modulation, principle of amplitude modulation, modulation index, power relation, multitone modulation, AM wave generation, AM square law modulator, switching modulator, demodulation of AM, synchronous detection - nonlinear demodulation, suppress carrier AM demodulator, envelope detector, square law demodulator, DSB-SC and SSB modulation systems, sideband and carrier power, method of generation and detection of DSB-SC and SSB, Independent SideBand (ISB) system, Vestigial SideBand (VSB) modulation.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance of Amplitude Modulation (AM)
- Understand the necessity of modulation
- Explain the principle of Amplitude Modulation (AM)
- Know what modulation index is
- Define power relation and multitone modulation
- Explain about the AM wave generation and AM square law modulator
- Analyse switching modulator
- Understand the concept demodulation of AM
- Define synchronous detection and nonlinear demodulation
- Discuss suppress carrier AM demodulator
- Explain envelope detector and square law demodulator
- Give specifications about DSB-SC and SSB modulation systems
- Elaborate on sideband and carrier power
- State the method of generation and detection of DSB-SC and SSB
- Explain Independent SideBand (ISB) system
- Know about the Vestigial SideBand (VSB) modulation

2.2 AMPLITUDE MODULATION

NOTES

Amplitude Modulation (AM) involves the modulation of the amplitude of the carrier as analog sine wave. It occurs when a signal to be modulated is applied to a carrier frequency. The carrier frequency may be a radio wave or light wave. The amplitude of carrier frequency changes in accordance with the modulated signal, while the frequency of carrier does not change and we get a complex wave as shown in Figure 2.1. It is the sum of three sinusoids of different frequencies. These are $f_c - f_m$, f_c, $f_c + f_m$. The sinusoid with frequency f_c has the same amplitude as the unmodulated carrier. The other two waves are called as lower and upper side band with frequency $f_c - f_m$, $f_c + f_m$ respectively and have equal amplitudes, which are proportional to the amplitude of the modulating signal. It is clear from the above that the *bandwidth* is equal to $2f_m$.

When combined, the resultant AM signal consists of the carrier frequency, plus upper and lower side bands. This is known as *Double SideBand-Amplitude Modulation (DSB-AM),* or more commonly referred to as AM. This is shown in Figure 2.1.

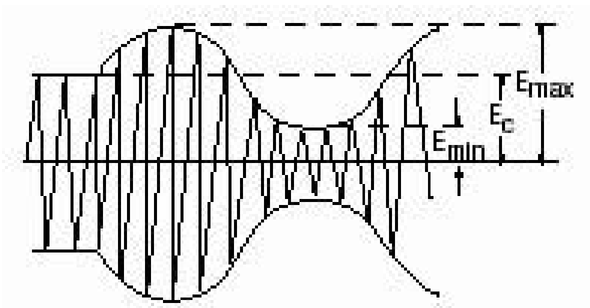


Fig. 2.1 *Amplitude Modulation of a Carrier Wave*

The carrier frequency may be suppressed or transmitted at a relatively low level. This requires that the carrier frequency be generated, or otherwise derived, at the receiving site for demodulation. This type of transmission is known as *Double SideBand - Suppressed Carrier (DSB-SC).*

It is also possible to transmit a single side band. The advantage is a reduction in analog bandwidth needed to transmit the signal. This type of modulation is known as *Single SideBand-Suppressed Carrier (SSB-SC)* and is ideal for Frequency Division Multiplexing (FDM).

Another type of analog modulation is known as *Vestigial Side Band modulation.* This is almost like Single Side band, except that the carrier frequency is preserved and one of the side bands is eliminated through filtering. Vestigial Side band transmission is usually found in television broadcasting. Amplitude modulation is rarely used individually as it is highly sensitive to the impacts of attenuation and line noise.

The modulating index is given as:

$$m = E_{max} - E_c / E_c$$

We may derive the following equation for modulating index m:

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

NOTES

Quadrature Amplitude Modulation

Quadrature Amplitude Modulation (QAM) is based on the basic amplitude modulation. It improves the performance of the basic amplitude modulation. In this technique two carrier signals are transmitted simultaneously. The two carrier signals are at the same frequency with a 90 degrees phase shift.

Data Rate

The number of signal changes transmitted per unit of time is called the data rate of the modem. That rate is usually expressed in terms of a unit known as a *baud*. The baud is the number of times per second the line condition can switch from "1" to "0". Data rate and transmission speed, which is expressed in terms of bits per second, usually are not the same, as several bits may be transmitted through the channel by the modem in each signal change (a few bits can be transmitted as one symbol).

C Shannon theorem states that the maximum capacity (bit rate) of a bandwidth limited transmission line with limited signal to noise ratio is given by $C=W*\log(1+S/N)$. Where C is the maximum capacity, W is the limited bandwidth and S/N is the power of the signal to noise ratio.

A telephone line, for example, has a bandwidth of 3000 Hz and maximum S/N of about 1000 (30 dB). Thus, theoretically, maximum data rate that can be achieved is about 30 Kbps (bits per second). Earliest modems that work through telephone lines had 1.2 Kbps. Today's modems reach data rates of 28.8 Kbps.

Generation of AM Waves

Wwe have already studied the basic concepts of amplitude modulation. In this section, we study the techniques to generate AM waves. These techniques can be classified as:

- Low Level Modulation
- High Level Modulation

Low Level Modulation: In this technique, AM waves are generated in the initial stage of amplification, i.e., at a low power level. The generated AM signal is then amplified with the help of a number of amplifier stages.

High Level Modulation: Under this technique, amplitude modulation takes place in the final stage of amplification and therefore modulation circuitry has to handle high power. For example, if the transmitter power is 1500 W and the modulation index is 1 then modulation power is 500 W. (33 per cent of transmitter power). The modulator circuitry must be able to deliver such a high power. Table 2.1 gives the comparison between low level AM and high level AM.

Table 2.1 *Comparison between Low Level and High Level AM*

S.No.	Parameter	Low Level Modulation	High Level Modulation
1.	Point at which modulation takes place	Modulation takes place in the initial stages of amplification.	Modulation takes place in the final stage of amplification.
2	Power level	Modulation circuitry has to handle low power.	Modulation circuitry has to handle high power.
3.	Complexity	Modulation circuitry is simple as it needs to handle low power.	Modulation circuitry is quite complex as it needs to handle high power.
4.	Prime factors in design	Prime requirement is simplicity.	Prime requirement is high efficiency and low distortion.
5.	Audio power	Low audio power is required to produce modulation.	High audio power is required to produce modulation.
6.	Design requirements of amplifier stages	Each amplifier stage following modulation, must handle sideband power as well as the carrier. All the subsequent amplifiers must have sufficient bandwidth for the sideband frequencies.	This is not the case with high level modulation because in this modulation takes place in the output stage.
7	Amplifier used	Linear amplifiers such as Class A amplifier are used because all stages must be capable of handling amplitude variations caused by the modulation.	High efficient Class C amplifiers are used.
8.	Amplifiers used	Transistors and Op-amps	Vacuum tubes and power transistors.
9.	Efficiency	Lower than high level modulators.	Very high.

Low Level Modulator

This simple modulator circuit consists of operational amplifier (op-amp) and a Field Effect Transistor (FET), as shown in Figure 2.2.

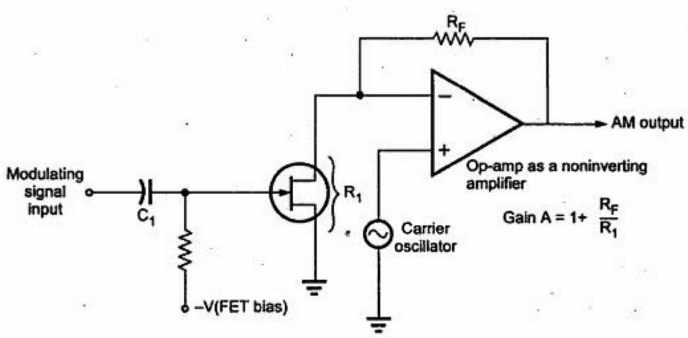


Fig. 2.2 *Low Level Modulation Using FET and Op-Amp*

Here, FET is used as a variable resistor and op-amp is used as a non-inverting amplifier for the carrier signal. For op-amp, R_F acts as a feedback resistor

NOTES

and the resistance of an N-channel junction FET acts as input resistance R_1. The gain of the circuit (non-inverting amplifier) is given by,

$$A = 1 + \frac{R_F}{R_1}$$

In the absence of the modulating signal FET provides a fixed resistance and therefore the gain of the noninverting amplifier is constant, giving a steady carrier output. On the applying modulating signal, the resistance of FET will vary. A positive going modulating input signal will cause the FET resistance to decrease, whereas a negative-going modulating signal will cause it to increase the resistance. Increasing the FET resistance causes the op-amp circuit gain to drop and vice versa. This results in an AM signal at the output of op-amp.

High Level Modulator

Basic Requirements

Generation of AM wave is done in two parts. The first part generates series of current pulses which are proportional to the modulating signal and the second part converts each pulse into a complete sine wave proportional in amplitude to the size of pulse. Figure 2.3 shows the block schematic of AM generator. As shown in Figure 2.3, there are two main blocks of AM generator: Class C amplifier and tuned circuit.

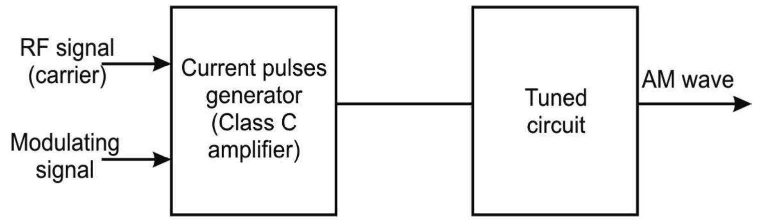


Fig. 2.3 *Block Schematic of AM Generator*

Class C Amplifier

It is possible to make the output current of a class C amplifier proportional to the modulating voltage by applying modulating voltage in series with the d.c. power supply voltage for the amplifier. Figure 2.4 (a) and (b) show the modulating signal and the output current waveform of the class C amplifier.

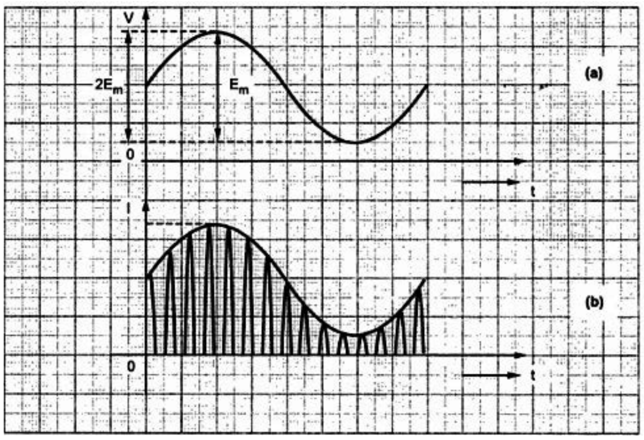


Fig. 2.4 *(a) Modulating Signal and (b) Current Pulses*

Tuned Circuit

Each current pulse applied to the tuned circuit initiates a damped oscillation in it. The amplitude of the oscillation is proportional to the size of the current pulse and decay rate is proportional to the time constant of the circuit. Since a series of pulses are fed to the tuned circuit, each pulse will generate a complete sine wave proportional in amplitude to the size of applied pulse. This will be followed by the next sine wave, proportional to the size of the next applied pulse and so on. Figure 2.5 (a) and (b) shows the input and output waveforms of the tuned circuit.

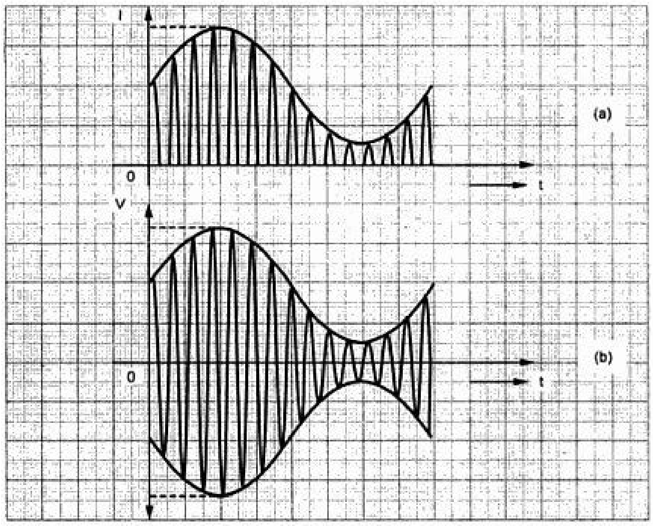


Fig. 2.5 *(a) Current Pulses and (b) Tuned Circuit Output (AM Wave)*

Suppression of Unwanted Sideband

Single-SideBand (SSB) modulation, which is a refinement of AM, helps to utilize electric power and bandwidth more efficiently. AM produces a modulated output signal which has twice the bandwidth of the original baseband signal. SSB controls this bandwidth doubling, and saves the power wasted on a carrier. Of course this is achieved at the price of somewhat increased device complexity. SSB is also used for long distance telephone transmissions lines, as part of Frequency Division Multiplexing (FDM) technique. SSB technique is used in many applications as it offers many advantages. One of its many advantages is that it allows good quality signals by using very narrow bandwidth with relatively low power over longer distances.

To get SSB one of the sidebands of the double sideband signal available at the output of balanced modulator have to be suppressed. There are three practical methods used to remove the unwanted sideband from the double sideband signal to get the single sideband signal. These are:

1. Filter Method
2. Phase Shift Method
3. The 'Third' Method

Each of the three methods removes either the upper or the lower sideband, depending on the specific circuit arrangements. We will discuss these methods in the following sections.

NOTES

Filter Method to Produce SSB

Figure 2.6 shows the block diagram of filter method to suppress one sideband. As shown in the diagram, the balanced modulator produces DSB (Double SideBand) Signal which contains both the sidebands. It is applied to sideband suppression filter to remove the unwanted sideband. The filter must have a flat passband and extremely high attenuation outside the passband. To get this type of response the Q of the tuned circuits must be very high. The required value of Q factor increases as the difference between modulating frequency and carrier frequency increases. Carrier frequency is usually same as the transmitter frequency. For higher transmitting frequencies, the required value of Q is so high that it cannot be achieved by practical methods. In such situations, initial modulation is done at a low frequency carrier, say 100 kHz by the balanced modulator, and then, the filter suppresses one of the sidebands. The frequency of the SSB signal generated at the output of filter is very low as compared to the transmitter frequency. The frequency is increased to the transmitter frequency by the balanced mixer and crystal oscillator. This process of boosting frequency is also called as up conversion. The SSB signal having frequency equal to the transmitter frequency is then amplified by the linear amplifiers.

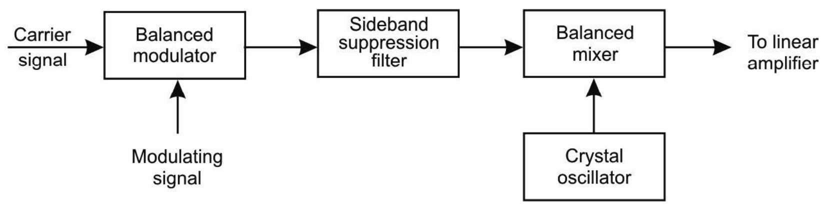


Fig. 2.6 *Filter Method to Suppress Sideband*

Phase Shift Method to Generate SSB

Figure 2.7 shows the block diagram of the phase shift method to generate SSB. The carrier signal is shifted by 90° and applied to the balanced modulator M_1. The modulating signal is also directly applied to this balanced modulator. The carrier signal is directly applied to the balanced modulator M_2. The modulating signal is phase shifted by 90° and applied to balanced modulator M_2. Both the modulators produce an output consisting of only sidebands. The upper balanced modulator (M_1) generates upper sideband and lower sideband, but each one is shifted by +90°. The lower balanced modulator (M_2) generates the upper and lower sidebands, but upper sideband is shifted by +90° whereas lower sideband is shifted by –90°. The summing amplifier adds the outputs of balanced modulators. Since upper sidebands of both the modulators are phase shifted by +90°, they are in phase and add up to produce double amplitude signal. But lower sidebands of the balanced modulators are 180° out of phase (+90°, –90°) and hence cancel each other out. Hence, the output of summing amplifier contains only upper sideband SSB signal. The carrier is already suppressed by the balanced modulators.

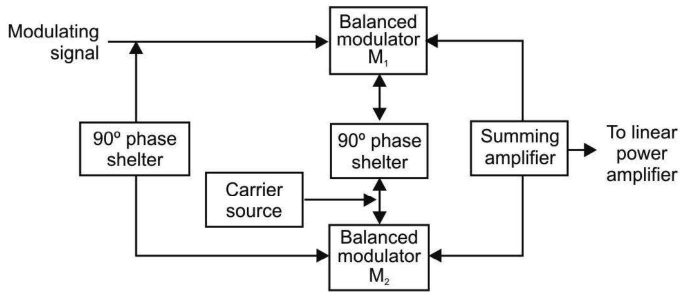


Fig. 2.7 *Phase Shift Method to Generate SSB Signal*

NOTES

Let us understand mathematically, how the sidebands add and cancel each other because of phase shifts. Input to the balanced modulator M_1 are, sin $\omega_m t$ and sin $(\omega_c t + 90°)$. Hence,

$$\text{Output of} \quad M_1 = \cos[(\omega_c t + 90°) - \cos[(\omega_c t + 90°) + \omega_m t]$$

$$= \cos[(\omega_c t - \omega_m t + 90°) - \cos[(\omega_c t + \omega_m t + 90°)\,] \qquad (2.1)$$

In Equation (2.1), observe that the first term represents LSB with +90° phase shift and the second term represents USB with +90° phase shift. Now the inputs to the balanced modulator M_2 are, sin $(\omega_m t + 90°)$ sin$\omega_c t$ +90°. Hence,

$$\text{Output of} \quad M_2 = \cos[(\omega_c t\,(\omega_m t + 90°) \;-\cos[(\omega_c t\,(\omega_m t + 90°)]$$

$$= \cos[(\omega_c t - \omega_m t - 90°) - \cos[(\omega_c t + \omega_m t + 90°)\,] \qquad (2.2)$$

In Equation (2.2), the first term represents LSB and has a phase shift of –90°. Similarly the second term represent USB with phase shift of +90°. When the signals of Equation (2.1) and (2.2) add in the summing amplifier, the lower sidebands (the first terms) cancel each other since they are out of phase. The second terms add since they have same phase shift of +90° (i.e., in phase). Thus SSB is generated at the output of summing amplifier.

Third Method to Generate SSB

In 1950, D.K. Weaver developed the *third method* of generating SSB. It is similar to the phase shift method, but in this method only the carrier signals are phase shifted by 90°. Figure 2.8 shows the block diagram of third method of SSB generation. It consists of four balance modulators, two carrier signal generators, two audio low pass filters and two 90° phase shift networks. In this system modulation occurs in two phases. In the first phase, modulating signal combines with the audio carrier. As shown in Figure 2.8, inputs for balance modulator (M_1) are unshifted modulating signal and 90° shifted audio carrier signal with frequency f_0. Therefore, output of balance modulator (M_1) contains the upper and lower sidebands, both shifted in phase by 90°. The inputs for balanced modulator (M_2) are unshifted modulating and audio carrier signals. Therefore, output of balanced modulator (M_2) contains the upper and lower sidebands without any phase shift. The outputs of both the balanced modulators are connected to the low pass filters. Low pass filters remove the upper sideband signals since the cut off frequency of low pass filters is f_0. The signals with suppressed uperside band ($F_0 - f_m$) are applied to the balance modulator (M_3) and balanced modulator (M_4), which are driven with the RF carrier frequency and its 90° shifted version, respectively. The

NOTES

output of balanced modulator (M_3) contains two sidebands, upper sideband $[f_e + f_0 - f_m)]$ shifted by +90° and lower sideband $[f_e - (f_0 - f_m)]$ shifted by –90º.

The output of balanced modulator (M_4) also contains two sidebands, upper sideband and lower sideband. The upper sideband, $[f_c + (f_0 - f_m)]$ shifted by +90º, which is in phase with the upper sideband of the balanced modulator (M_3) and adds directly to it, where as lower sideband $[f_e - (f_0 - f_m)]$ shifted by –90°, which is 180º out of phase with the lower sideband of the balanced modulator (M_3) and cancels it. As a result, the output of summing circuit gives frequency component ($f_c + f_0 - f_m$), which corresponds to the lower sideband of f_m on a carrier frequency of ($f_c + f_0$). The other sideband and carrier have been suppressed.

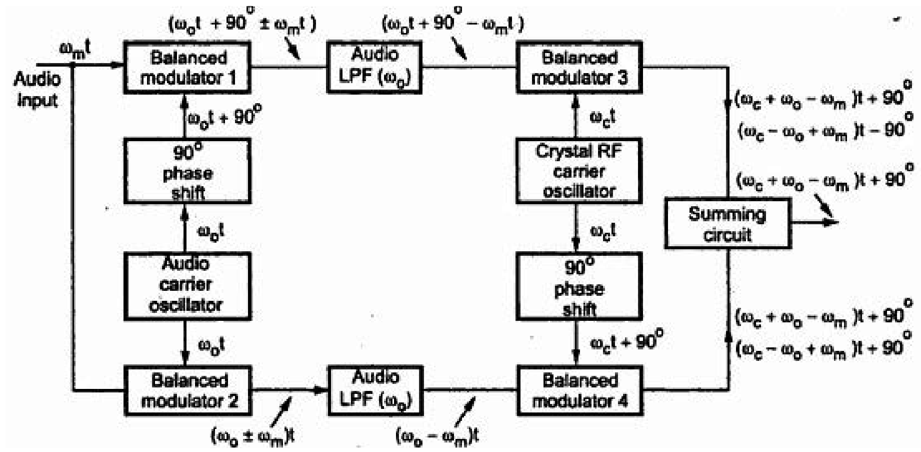


Fig. 2.8 *Block Diagram of Third Method SSB Generation*

Advantages of the 'Third' Method

1. It requires no sideband suppression filter and wideband audio phase shift network.
2. With carrier frequencies being constant, the phase shift network is rather a simple RC circuit.
3. Without critical parts or adjustments, correct output can be maintained.
4. Low frequencies can be transmitted.
5. Since the of the circuitry is at audio frequency layout and components tolerances are not critical.
6. Sidebands most also be switched quite easily.

Disadvantages of the Third Method

1. Amongest three methods it is the most complex
2. Extra crystal is required for sideband switching.
3. DC coupling is needed to avoid the loss of signal components close to the audio carrier frequency.

2.2.1 Necessity of Modulation

Why should modulation be employed when it is possible to directly transmit a baseband signal? Modulation is required because there are several limitations associated with baseband transmission that can be overcome using modulation. There is a translation of the baseband signal during the modulation process which converts the low frequency signal to a high frequency signal. The shift in frequency is in proportion to the carrier's frequency.

The following are the advantages that modulation provides:

• Reduction in the Height of Antenna

In the case of radio signal transmission, the height of the antenna has to be multiple of $\lambda/4$ (λ refers to the wavelength).

$$\lambda = c/f$$

Where,

c refers to the velocity of light.

f refers to the frequency of the signal to be transmitted.

So, the calculation of the minimum antenna height required to transmit a baseband signal of f = 10 kHz is:

$$\text{Minimum antenna height} = \frac{\lambda}{4} = \frac{c}{4f} = \frac{3\times10^8}{4\times10\times10^3} = 7500 \text{ meters, i.e., } 7.5 \text{ km}$$

It is practically not possible to install and antenna of this height.

With the modulated signal at f = 1 MHz, it is possible to give the minimum antenna height with:

$$\text{Minimum antenna height} = \frac{\lambda}{4} = \frac{c}{4f} = \frac{3\times10^8}{4\times10\times10^6} = 75 \text{ meters}$$

Practically, it is easy to install this antenna. So, with modulation it is possible to reduce the antenna's height to one which is feasible to install.

• Prevents Mixing of Signal

In case of transmission of baseband sound signals without modulation by more than one transmitter, the signals will all fall in the same frequency range - 0 to 20 kHz. This will result in all of the signals getting intermixed and leaving the receiver unable to separate one from the other.

On the other hand, if every baseband sound signal is used to modulate a different carrier, each one will fall in a separate frequency domain (different channels) slot and will prevent signal mixing.

• Increase in Range of Communication

A baseband signal is a low frequency signal. Low frequency signals when transmitted are unable to go long distances getting heavily attenuated.

Attenuation causes a reduction with increase in frequency of the transmitted signal, enabling the signal to travel longer distances.

The frequency, of the signal to be transmitted, is increased when there is modulation of the signal. This leads to the communication range being increased.

NOTES

• Enables Multiplexing

With multiplexing, it becomes possible to simultaneously transmit two or more signals over the same communication channel. This can only be achieved with modulation.

With multiplexing, several signals can make use of the same channel at the same time. To take an example, several different TV channels can use the same frequency range without one interfering or getting mixed with another. It even allows different frequency signals can be transmitted at the same time.

• Improved Reception Quality

Frequency modulation (FM) along with digital communication techniques like PCM, there is significant decrease in the effect of noise. Due to this, reception quality becomes much improved.

2.2.2 Principle of Amplitude Modulation

The modulation technique referred to as Amplitude Modulation (AM) is employed in the field of electronic communication, mainly for the purpose of information transmission by using a radio carrier wave. For implementation of amplitude modulation, the carrier wave's amplitude (signal strength) needs to be varied proportionate to the waveform which is to be transmitted. For example, the specific waveform may, for instance, correspond to the sounds to be reproduced by a loudspeaker, or the intensity of light in the pixels of a television. AM is a technique which is directly in contrast with both technique of Frequency Modulation (FM) where the carrier signal's frequency is varied, and phase modulation where there is the varying of the phase of the carrier signal.

The earliest method of modulation was the AM technique for the purpose of transmitting of voice by radio. Its development happed in the initial twenty years of the 20th century, starting with the 1900 when the radiotelephone experiments were conducted by Reginald Fessenden and Landell de Moura. Several communications make use of AM even today. Some examples being, portable two-way radios, VHF aircraft radio, Citizen's Band Radio, and in computer modems (in the form of QAM). "AM" is often used to refer to mediumwave AM radio broadcasting.

Amplitude Modulation: Forms

In the field of telecommunications and electronics, modulation implies the varying of a certain aspect of a higher frequency continuous wave carrier signal with an information-bearing modulation waveform, live a video signal that depicts an image or audio signal that depicts sound, so that the carrier will "Carry" the information. When the same gets to the defined destination, demodulation is used for the extraction of the information signal from the modulated carrier.

In the case of the amplitude modulation, that which is varied is the carrier oscillations strength or amplitude. Let us take an example. In the case AM radio communication, prior to its transmission, the amplitude of a continuous wave radio-frequency signal (sinusoidal carrier wave) is modulated by an audio waveform. The audio waveform modifies the amplitude of the carrier wave and determines the *envelope* of the waveform. When the frequency domain is taken into account, amplitude modulation results in a signal which has its power concentrated at the carrier frequency and two adjacent sidebands. Both of the sidebands are individually equal in bandwidth to that of the modulating signal, and are mirror image of the other. So, at times the standard AM is also referred to as 'Double-SideBand Amplitude Modulation' (DSB-AM) so that it can be told apart from sophisticated modulation methods which too are based on AM.

A disadvantage that plagues each and every technique of amplitude modulation, and not just standard AM, lies with the receiver as it amplifies and detects electromagnetic interference and noise in equal proportion to the signal. If there is to be an increase in the received signal-to-noise ratio, for example by a factor of 10 (which is 10 decibel improvement), it will need that there be an increase in the transmitter power by a factor of 10. This stands in contrast to digital audio and to Frequency Modulation (FM) in which the effect of such noise post demodulation gets reduced greatly till the reduced remains well above the reception threshold. This is the reason why AM broadcast is preferred for broadcasts and voice communication but is not sought for the purpose of high fidelity broadcasting or for music.

AM suffers for the disadvantage of power usage inefficiency. It is in the carrier signal that a minimum of two-thirds of the power gets concentrated. No part of the original information which it being transmitted is present in the carrier signal. Nevertheless, it being present creates a simple means of demodulation which employs envelope detection, providing a frequency and phase reference to extract the modulation from the sidebands. Certain AM based modulation systems require a lower transmitter power via a complete or partial removal of the carrier component, nevertheless these signals require costlier and more complex receivers. A copy might be regenerated of the carrier frequency (generally moved to the intermediate frequency) by the receiver, with a "Pilot" carrier (in reduced-carrier transmission or DSB-RC) which has been greatly reduced, for use with the process of demodulation. When the carrier is totally eliminated as it is in the case of double-sideband suppressed-carrier transmission, it is still possible to accomplish carrier regeneration through employing Costas phase-locked loop. Nevertheless, this will not work in the case of Single SideBand Suppressed-Carrier Transmission (SSB-SC), causing the "Donald Duck" sound from such receivers when slightly detuned. However, there is wide usage of single sideband in amateur radio and other voice communications since it is power efficient as well as bandwidth efficient (reducing RF bandwidth by half in comparison to standard AM). Furthermore, in case of broadcasting of the short wave and medium wave, standard AM with the full carrier enables reception by employing inexpensive receivers. The extra power cost are absorbed by the broadcaster to greatly increase potential audience.

The standard AM's carrier provides an amplitude reference which is lost in single or double-sideband suppressed-carrier transmission. In the receiver, the Automatic Gain Control (AGC) responds to the carrier so that the reproduced audio level stays in a fixed proportion to the original modulation. Furthermore, when there is suppressed-carrier transmission, there will not be any transmitted power when there is a pause in the modulation, so the AGC must respond to

NOTES

peaks of the transmitted power during peaks in the modulation. Such a situation will typically have what is referred to as *fast attack, slow decay* circuit that holds the AGC level for a second or more following peaks of this type, in between the program's short pauses or syllables. In the case of communications radios, such a thing is extremely acceptable, since the compression of the audio helps with intelligibility. On the other hand, it is not at all favored for music or normal broadcast programming, where a faithful reproduction of the original program, including its varying modulation levels, is expected.

AM's trivial form that is employed for the purpose of binary data transmission is called ON-OFF keying. ON-OFF keying is the most basic and simple form of *amplitude-shift keying*, and in this technology the zeros and ones are depicted with the absence or presence of a carrier. Use is made of ON-OFF keying by radio amateurs for the transmission of Morse code and here it is called Continuous Wave (CW) operation, despite the fact that this transmission is does not remain "Continuous", in the strictest sense. Quadrature amplitude modulation is a form of AM which is much more complex and it is used more commonly for digital data, while making more efficient use of the available bandwidth.

ITU Designations

International Telecommunication Union (ITU), in the year 1982, fixed the various types of amplitude modulation. The following Table 2.2 depicts these amplitude modulations.

Table 2.2 *Amplitude Modulation with their Designation*

Designation	Description
R3E	single-sideband reduced-carrier
Lincompex	linked compressor and expander
J3E	single-sideband suppressed-carrier
H3E	single-sideband full-carrier
C3F	vestigial-sideband
B8E	independent-sideband emission
A3E	double-sideband a full-carrier - basic AM scheme

Standard AM: Simplified Analysis

Let us look at a carrier wave (sine wave) whose amplitude will be depicted as A and frequency by f_c as given below:

$$c(t) = A\cdot\sin(2\pi f_c t)$$

Consider that the modulating waveform is represented by $m(t)$. in our example, the modulation to be just a sine wave whose frequency is f_m, which is a frequency that is way lower (for example an audio frequency) than f_c:

$$m(t) = M\cdot\cos(2\pi f_m t + \phi)$$

where m is amplitude of message/modulating signal.

$$y(t) = A \cdot \sin(2\pi f_c t) + \frac{M}{2}[\sin(2\pi(f_c + f_m)t + \phi) + \sin(2\pi(f_c - f_m)t - \phi)$$

Consider $\phi = 0°$

NOTES

Thus, there are three components of the modulated signal which are: the carrier wave $c(t)$ which is unchanged, and two pure sine waves (known as sidebands) with frequencies slightly above and below the carrier frequency f_c.

2.2.3 Modulation Index

The AM modulation index is a measure based on the ratio of the modulation excursions of the RF signal to the level of the unmodulated carrier. Following is how it is, therefore, defined.

$$H = \frac{\text{Peak Value of } m(t)}{A} = \frac{M}{A} = \frac{\text{Peak value of } M(t)}{\text{Peak value of } c(t)}$$

$$H = \frac{v_{max} - v_{min}}{v_{max} + v_{min}}$$

The modulation amplitude is the peak (be it negative or positive) change in the RF amplitude from its unmodulated value. Generally, we express the modulation index in the form of a percentage, and it may be displayed on a meter which is attached with the AM transmitter. So, if the carrier amplitude will be varying by 100% above (and below) its unmodulated level. In the case of, it varies by 50%. In case when there is 100% modulation, it is possible that at times the wave amplitude reach zero, and it will imply full modulation using standard AM and at times this is the target (so that highest possible signal-to-noise ratio can be attained) but this should not be crossed/exceeded. If the modulating signal is increased further than this point (called over modulation), it will result in a standard AM modulator failing, due to the fact that the negative excursions of the wave envelope are incapable of becoming lower than zero, leading to distortion/clipping of the modulation that is received. Typically, in transmitters there is a limiter circuit which prevents overmodulation, and/or a compressor circuit (more so with voice communications) to approach 100% modulation for maximum intelligibility above the noise. Circuits of this type are also called vogad.

Nevertheless, it is possible to have a modulation index exceeding 100%, with no distortion introduced, as with double-sideband reduced-carrier transmission. In that case, negative excursions beyond zero require the carrier phase's reversal. It is not possible to produce this thru employing the efficient high-level (output stage) modulation techniques that are used widely and specifically so in broadcast transmitters that are of high power. Furthermore, such a waveform is produced by a special modulator at a low level and this is followed with a linear amplifier. Also, it is not possible a standard AM receiver that uses an envelope detector to properly demodulating a signal of this kind. What is needed in this case is synchronous detection. Due to this, generally double-sideband transmission does not fall in the category of "AM" despite it forming an RF waveform which is identical to a standard AM till the modulation index stays lower than 100%. Majority of the times, systems of this type try to create a radical reduction of the carrier level compared to the sidebands reduced to zero. Each case of this type displays a loss of vale of the term "modulation index" as it refers to the ratio of the modulation amplitude to a rather small (even zero) remaining carrier amplitude.

NOTES

The modulation index is indicative of the amount by which the modulated variable varies around its unmodulated level. It relates to carrier frequency variations:

$$h = \frac{\Delta f}{f_m} = \frac{f_\Delta |x_m(t)|}{f_m}$$

Where,

- f_m is highest frequency component present in the modulating signal $x_m(t)$
- Δf is the peak frequency-deviation

In case of sine wave modulation, the modulation index is the ratio of the carrier wave's peak frequency deviation to the modulating sine wave's frequency. It is referred to as Narrowband FM when bandwidth is approximately $2f_m$. Sometimes, modulation index $h < 0.3$ rad is taken to be Narrowband FM while otherwise it is seen as being Wideband FM.

In the case of modulation systems that are digital (like Binary Frequency Shift Keying (BFSK)), in which the carrier is modulated by a binary signal, the modulation index is given by:

$$h = \frac{\Delta f}{f_m} = \frac{\Delta f}{\frac{1}{2T_s}} = 2\Delta f T_s$$

Where,

- T_s represents the symbol period
- $f_m = \frac{1}{2T_s}$ is the highest *fundamental* of modulating binary waveform.

With digital modulation, the transmission of the carrier f_c never happens. Instead, there is the transmission of one of two frequencies i.e. $f_c + \Delta f$ or $f_c - \Delta f$. The basis for the transmission is the modulating signal's binary state - 0 or 1.

If $h \gg 1$, the modulation will be referred to as *wideband FM*. This has a bandwidth of about $2f_\Delta$. The signal-to-noise ratio can be improved using the more bandwidth of wideband FM. For example, the signal-to-noise ratio will see an eight-fold improvement on keeping f_m constant and doubling the value of Δf.

In case of a tone-modulated FM wave, increasing the modulation index and keeping the modulation frequency constant, there is an increase in the (non-negligible) bandwidth of the FM signal and the spacing between spectra remains the same. While there is a decrease in the strength of some spectral components and there is an increase in some others. If there is an increasing of the modulation frequency while the frequency deviation is maintained as constant, there is an increase in the spacing between spectra.

There are two ways to classify frequency modulation:

- Narrowband – When the change in the carrier frequency is about the same as the signal frequency

- Wideband – When the change in the carrier frequency is much higher (modulation index > 1) than the signal frequency

AM

$$\begin{aligned} y(t) &= [1 + m(t)]\cdot c(\mathrm{t}) \\ &= [1 + M\cdot\cos(2\pi f_m t + \phi)]\cdot A\cdot\sin(2\pi f_c t) \end{aligned}$$

If use is made of prosthaphaeresis identities, it is possible to depict $y(t)$ as being the sum of three sine waves in the following manner.

In our example, which is explained below. With $M = 0.5$ the amplitude modulated signal $y(t)$ thus corresponds to the top graph (labelled "50% Modulation") in the following illustration.

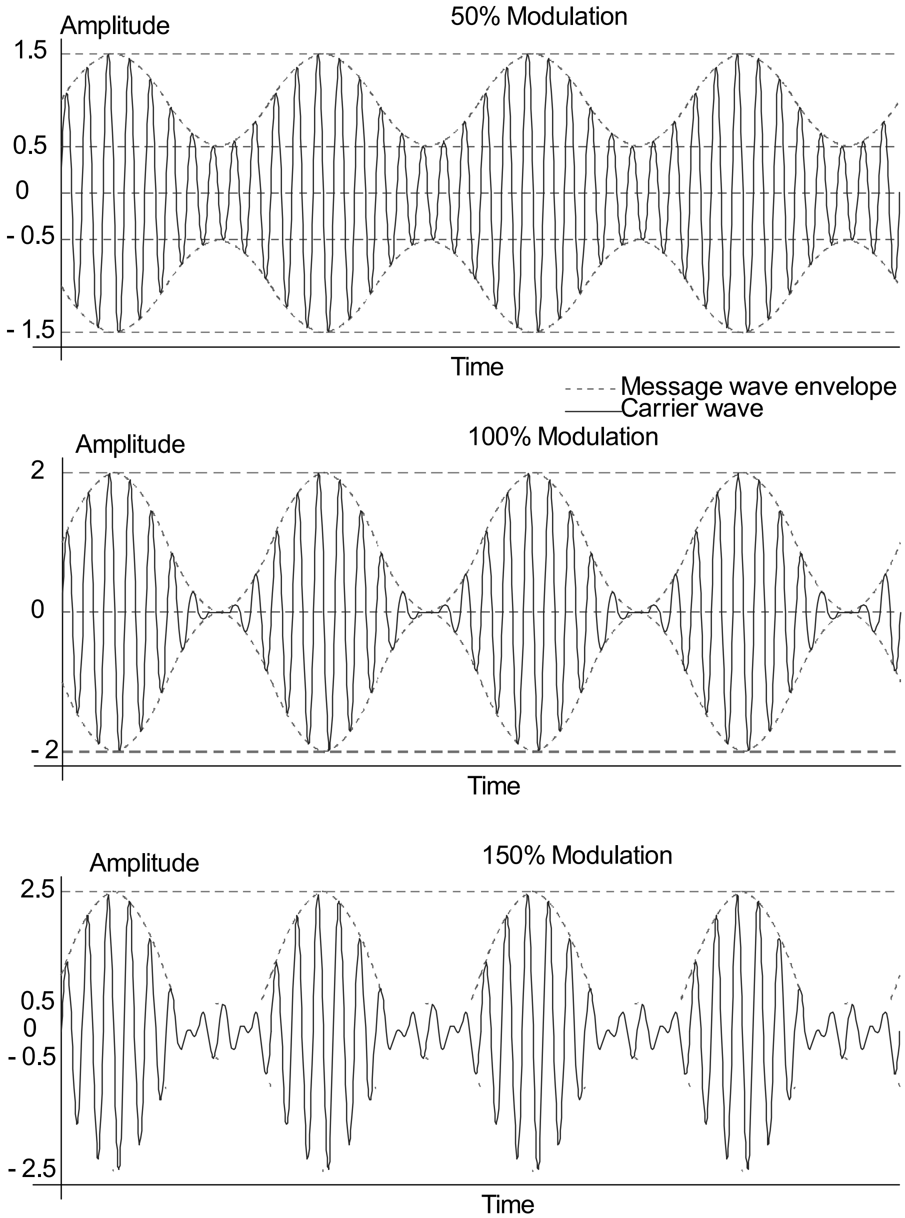


Fig. 2.9 *Modulation Depth, the Unmodulated Carrier has An amplitude of* 1

NOTES

2.2.4 Power Relation

The carrier component of the modulated wave has the similar amplitude as the unmodulated carrier, i.e., the amplitude of the carrier is unchanged; energy is either added or subtracted. The modulated wave contains extra energy in the two sideband components. Consequently, the modulated wave contains more power in comparison to the carrier had before modulation happened. Because the amplitude of the sidebands depends on the modulation index V_m/V_c, hence it is anticipated that the total power in the modulated wave depends on the modulation index also.

The **power relation** can be derived as follows.

The **total power** in the modulated wave will be,

$$P_t = \frac{V_{carr}^2}{R} + \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R} \text{ (rms)} \qquad \text{....(2.3)}$$

Where all three voltages are (rms) values (√2 converted to peak), and R is the resistance, (e.g., antenna resistance), in which the power is dissipated. The first term of Equation (2.3) is the unmodulated carrier power and is given by,

$$P_c = \frac{V_{carr}^2}{R} = \frac{(V_c/\sqrt{2})^2}{R}$$

$$= \frac{V_c^2}{2R} \qquad \text{....(2.4)}$$

Similarly,

$$P_{LSB} = P_{USB} = \frac{V_{SB}^2}{R} = \left(\frac{HV_c/2}{\sqrt{2}}\right)^2 \div R = \frac{H^2V_c^2}{8R} = \frac{H^2}{4}\frac{V_c^2}{2R} \qquad \text{....(2.5)}$$

Substituting Equations (2.4) and (2.5) into Equation (2.3), we have,

$$P_t = \frac{V_c^2}{2R} + \frac{H^2}{4}\frac{V_c^2}{2R} + \frac{H^2}{4}\frac{V_c^2}{2R} = P_c + \frac{H^2}{4}P_c \frac{H^2}{4}P_c$$

$$\frac{P_t}{P_c} = 1 + \frac{H^2}{2} \qquad \text{....(2.6)}$$

Equation (2.6) relates the total power in the amplitude-modulated wave to the unmodulated carrier power.

From Equation (2.6), the maximum power in the AM (Amplitude Modulation) wave is $P_t = 1.5P_c$ when $H = 1$. This is significant, because it is the **maximum power** that relevant amplifiers must be capable of handling without distortion.

Current Calculations

The situation which very often occurs in AM is that the modulated and unmodulated currents are easily measurable, and it is then essential to calculate the modulation index from them. This occurs when the antenna current of the transmitter is metered, and the problem may be resolved as follows.

NOTES

Let I_c be the unmodulated current and I_t the total, or modulated, current of an AM transmitter, both being rms values. If R is the resistance in which these currents flow, then,

$$\frac{P_t}{P_c} = \frac{I_t^2 R}{I_c^2 R} = \left(\frac{I_t}{I_c}\right)^2 = 1 + \frac{H^2}{2}$$

$$\frac{I_t}{I_c} = \sqrt{1+\frac{H^2}{2}} \text{ or } I_t = I_c\sqrt{1+\frac{H^2}{2}} \qquad \ldots(2.7)$$

2.2.5 Multitone Modulation

Modulation is the overlaying of information or the signal onto an electronic or optical carrier waveform. If the message signal contains single frequency component, then the resulting modulating signal is called as single tone modulated signal. But if the message signal contains more than one frequency component, then the resulting modulated signal is termed as multitone modulated signal, therefore, the multitone modulation contains message signals which has more than one frequency component.

In radio communications, a sideband is a band of frequencies higher than or lower than the carrier frequency, those are the result of the modulation process. The sidebands carry the information transmitted by the radio signal. The sidebands comprise all the spectral components of the modulated signal except the carrier. The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). When the message signal contains more than one frequency, then the corresponding modulation scheme is known as multitone modulation.

Expression of Multitone AM (Amplitude Modulation) Signal

Consider a message signal having following equation:

$$m(t) = A_{m1}\ cos\pi f_{m1}t + A_{m2}\ cos2\pi f_{m2}t$$

Equation for the amplitude modulated signal is given as:

$$S_{AM}(t) = A_c\{1 + K_a m(t)\}cos2\pi f_c t$$

$$S_{AM}(t) = A_c\{1 + K_a A_{m1} cos2\pi f_{m1}t + K_a A_{m2} cos2\pi f_{m2}t\}cos2\pi f_c t$$

Let $K_a Am_1 = u_1$ and $K_a Am_2 = u_2$

Therefore, $S_{AM}(t) = A_c\{1 + u_1 cos2\pi f_{m1}t + u_2 cos2\pi f_{m2}t\}cos2\pi f_c t$

$$S_{AM}(t) = A_c cos2\pi f_c t + A_c u_1 cos2\pi f_{m1} t cos2\pi f_c t + A_c u_2 cos2\pi f_{m2} t cos2\pi f_c t$$

$$\boldsymbol{S_{AM}(t) = A_c cos2\pi f_c t + \frac{A_c u_1}{2} cos2\pi(f_c - f_{m1})t + \frac{A_c u_1}{2} cos2\pi(f_c + f_{m1})t + \frac{A_c u_2}{2} cos2\pi(f_c - f_{m2})t + \frac{A_c u_2}{2} cos2\pi(f_c + f_{m2})t}$$

$$\ldots(2.8)$$

Here, $A_c cos2\pi f_c t$ = Carrier Frequency Component

$\frac{A_c u_1}{2} cos2\pi(f_c - f_{m1})t$ = First Lower SideBand (LSB_1)

NOTES

$\frac{A_c u_1}{2} cos2\pi(f_c + f_{m1})t$ = First Upper SideBand (USB_1)

$\frac{A_c u_2}{2} cos2\pi(f_c - f_{m2})t$ = Second Lower SideBand (LSB_2)

And,

$\frac{A_c u_2}{2} cos2\pi(f_c + f_{m2})t$ = Second Upper SideBand (USB_2)

Power of Multitone AM Signal/Sideband and Carrier Power

The total power of a multitone AM signal (P_t) is given by the sum of the powers of the carrier signal (P_c), lower sideband (P_{LSB}) and upper sideband (P_{USB}) as:

$$P_t = P_c + P_{LSB}(total) + P_{USB}(total)$$

Thus, $P_t = P_c + P_{LSB_1} + P_{LSB_2} + P_{USB_1} + P_{USB_2}$

To obtain the power factors, use $P = V^2/2R$ in Equation (2.8):

Thus, $$P_c = \frac{A_c^2}{2R}$$

$$P_{LSB_1} = P_{USB_1} = \frac{\left(\frac{A_c u_1}{2}\right)^2}{2R} = \frac{A_c^2 u_1^2}{8R}$$

$$P_{LSB_2} = P_{USB_2} = \frac{\left(\frac{A_c u_2}{2}\right)^2}{2R} = \frac{A_c^2 u_2^2}{8R}$$

Therefore, $$P_t = \frac{A_c^2}{2R} + \frac{A_c^2 u_1^2}{4R} + \frac{A_c^2 u_2^2}{4R}$$

$$P_t = \frac{A_c^2}{2R}\left[1 + \frac{u_1^2 + u_2^2}{2}\right]$$

$$P_t = P_c\left[1 + \frac{u_1^2 + u_2^2}{2}\right]$$

Let, $$u_1^2 + u_2^2 = u_t^2$$

$$P_t = P_c\left[1 + \frac{u_t^2}{2}\right]$$

Thus, $$P_t = P_c + P_{SB}$$

Hence, Sideband Power (P_{SB}) = $\frac{P_c u_t^2}{2}$

Therefore, Modulation Efficiency, $\eta = \frac{P_{SB}}{P_t}$

$$\eta = \frac{\frac{P_c u_t^2}{2}}{P_c\left[1 + \frac{u_t^2}{2}\right]} \times 100\% = \frac{u_t^2}{2 + u_t^2} \times 100\%$$

NOTES

2.2.6 AM Wave Generation

- Amplitude modulation refers to a process that is used to vary the amplitude of high frequency carrier signal in accordance with the amplitude of the low frequency modulating or information signal, while the carrier signal's phase and the frequency are kept constant.
- Consider the carrier voltage to be v_c and the modulating voltage to be v_m, both of which are being represented as follows:

  $$v_c = V_c \sin w_c t$$
  $$v_m = V_m \sin w_m t$$

- With amplitude modulation, the unmodulated carrier's (v_c) amplitude will be varied in proportion with the instantaneous modulating voltage $V_m \sin É_{mt}$.
- In case where modulation is not there, the carrier's amplitude will be equal to its unmodulated value and in the case where there is modulation, the carrier's amplitude will be varied by its instantaneous value (as shown in Figure 2.10).

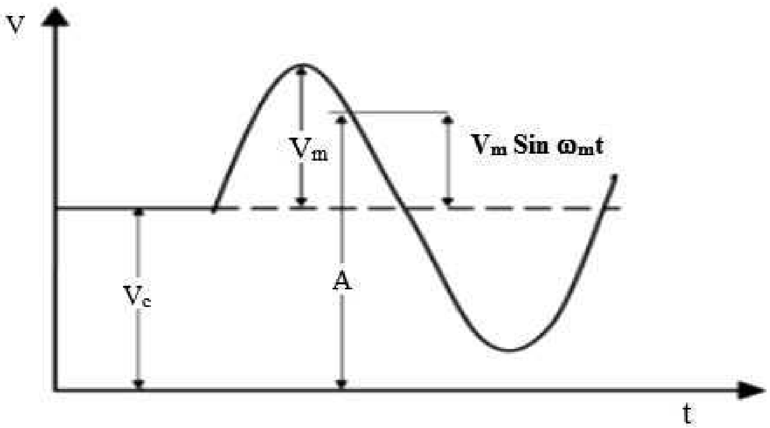


Fig. 2.10 *Amplitude of AM Wave*

- Illustration in Figure 2.11 given below provides an amplitude modulated wave's time domain representation depicting the carrier signal's modulation by the modulating signal.
- The modulating wave as well as the carrier wave in the case or AM wave is sinusoidal in nature yet the modulated wave is not. Following is the amplitude of the AM.

$$v_{AM} = V_c \sin w_c t + \frac{mV_c}{2}\cos(w_c - w_m)t - \frac{mV_c}{2}\cos(w_c + w_m)t$$

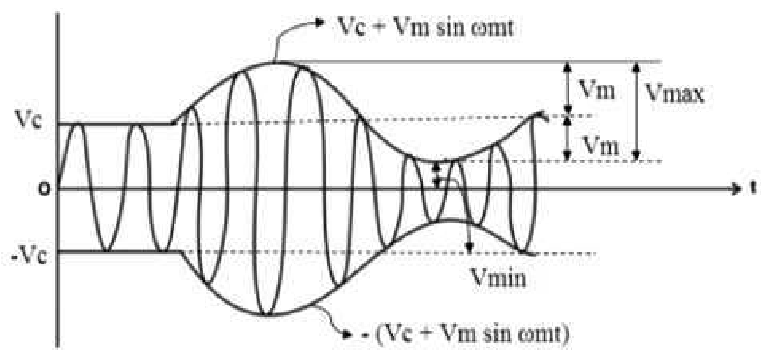


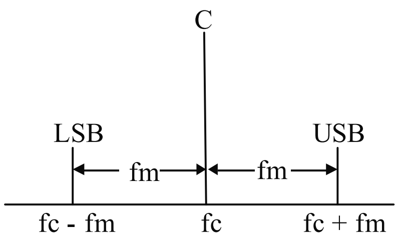
Fig. 2.11 *Time Domain Representation of AM Wave*

NOTES

- In light of the above, it is possible to say that besides the original signal two additional sine waves exist, one below and one above the carrier frequency. So, the complete AM signal comprises one carrier wave plus two additional frequencies one on rather side, referred to as side frequencies.
- The frequency below the carrier frequency is referred to as lower sideband and the one above the carrier frequency is referred to as upper sideband.
- The frequency of the Lower SideBand (LSB) is:

  $f_c - f_m$
- The frequency of the Upper SideBand (USB) is:

  $f_c + f_m$
- The illustration given below represents the spectrum of amplitude modulated wave.



- The amplitude of the central frequency (the carrier frequency f_c) is the highest. The sideband frequencies which lie on one on each side of the carrier frequency have lower amplitude. The following is used to provide the bandwidth of the amplitude modulated wave:

  $$BW = (f_c + f_m) - (f_c - f_m) = 2f_m$$

Generation of AM

While there are several forms of Amplitude Modulation (AM) generation, the normal AM or the Double SideBand (DSB) is generated with 'Switching-Modulator' or the 'Square-Law' modulator and the Double SideBand Suppressed Carrier (DSB-SC) with the balanced modulator or the 'Ring-Modulator'. When there is DSB generation by employing 'Switching Modulator', it becomes essential that the amplitude of the carrier remain much higher than the amplitude of the message due to the fact that the switching action has to remain dependent on just the carrier, though in the square-law modulator, the total amplitude of the message and carrier to be 'Low' enough to operate the active device in the 'Square-Law' region. So, it must be noted that there are strict conditions for operation associated with the two existing methods of discrete component realization of the AM. For DSB-SC, the 'Ring Modulator' which is employed is not of the kind with which has operating conditions that are difficult to maintain and strict. With the carrier, the DSB-SC becomes the DSB, and when DSB has the carrier removed it leads to DSB-SC modulation. Nevertheless, it is not possible to use the 'Square-Law' modulator or the 'Switching-Modulator' for generating DSB-SC. Similarly, the 'Ring-Modulator' cannot be employed for DSB generation.

2.2.7 AM Square Law Modulator

Amplitude Modulation (AM) is a process by which the wave signal is transmitted by modulating the amplitude of the signal. The AM and is commonly used in

NOTES

transmitting a piece of information through a radio carrier wave. The modulators generate amplitude modulated waves. The following two modulators generate AM wave.

- Square Law Modulator
- Switching Modulator

Following Figure 2.12 illustrates the block diagram of the square law modulator.

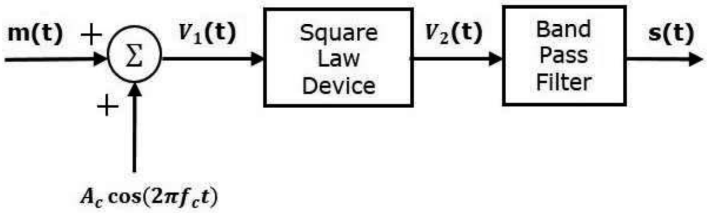


Fig. 2.12 *Block Diagram of the Square Law Modulator*

Let the modulating and carrier signals be denoted as, $m(t)$ and $A\cos(2\pi f_c t)$, respectively. These two signals are applied as inputs to the summer (adder) block. This summer block produces an output, which is the addition of the modulating and the carrier signal. Mathematically, we can write it as,

$$V_1 t = m(t) + A_c \cos(2\pi f_c t)$$

This signal $V_1 t$ is applied as an input to a nonlinear device like diode. The characteristics of the diode are closely related to square law.

$$V_2 t = k_1 V_1(t) + k_2 V_1^2(t) \qquad \text{.....(2.9)}$$

Where, k_1 and k_2 are constants.

Substitute $V_1 t$ in Equation 2.9.

$$V_2(t) = k_1[m(t) + A_c\cos(2\pi f_c t)] + k_2[m(t) + A_c\cos(2\pi f_c t)]^2$$

$$\Rightarrow V_2(t) = k_1 m(t) + k_1 A_c \cos(2\pi f_c t) + k_2 m^2(t) + k_2 A_c^2 \cos^2(2\pi f_c t) + 2k_2 m(t) A_c \cos(2\pi f_c t)$$

$$\Rightarrow V_2(t) = k_1 m(t) + k_2 m^2(t) + k_2 A_c^2 \cos^2(2\pi f_c t) + k_1 A_c\left[1 + \left(\frac{2k_2}{k_1}\right) m(t)\right]\cos(2\pi f_c t)$$

The last term of the above equation represents the desired AM wave, and the first three terms of the above equation are not essential. So, with the help of band pass filter, we can pass only AM wave and eliminate the first three terms.

Therefore, the output of square law modulator is,

$$s(t) = k_1 A_c\left[1 + \left(\frac{2k_2}{k_1}\right) m(t)\right]\cos(2\pi f_c t)$$

The standard equation of AM wave is,

$$s(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$

Where, K_a is the amplitude sensitivity.

NOTES

By comparing the output of the square law modulator with the standard equation of AM wave, we will get the scaling factor as k_1 and the amplitude sensitivity k_a as,

$$\frac{2k_2}{k1}$$

Square law modulators are used for Amplitude Modulation (AM). Square law modulators are highly nonlinear in low voltage region.

2.2.8 Switching Modulator

Following Figure 2.13 illustrates the block diagram of switching modulator.

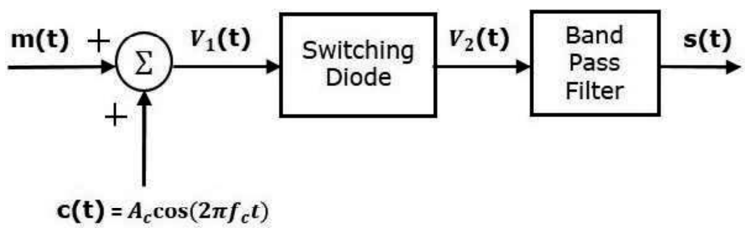


Fig. 2.13 *Block Diagram of Switching Modulator*

Switching modulator is similar to the square law modulator. The only difference is that in the square law modulator, the diode is operated in a non-linear mode, whereas, in the switching modulator, the diode must operate as an ideal switch.

Let the modulating and carrier signals be denoted as $m(t)$ and $c(t) = A_c \cos(2\pi f_c t)$, respectively. These two signals are applied as inputs to the summer (adder) block. Summer block produces an output, which is the addition of modulating and carrier signals. Mathematically, we can write it as,

$$V_1(t) = m(t) + c(t) = m(t) + A_c \cos(2\pi f_c t)$$

This signal $V_1(t)$ is applied as an input of diode. Assume, the magnitude of the modulating signal is very small when compared to the amplitude of carrier signal A_c. So, the diode's ON and OFF action is controlled by carrier signal $c(t)$. This means, the diode will be forward biased when $c(t) > 0$ and it will be reverse biased when $c(t) < 0$.

Therefore, the output of the diode is,

$$V_2(t) = \begin{cases} V_1(t) & if \quad c(t) > 0 \\ 0 & if \quad c(t) < 0 \end{cases}$$

We can approximate this as,

$$V_2(t) = V_1(t)\, x(t) \qquad \text{.....(2.10)}$$

Where, $x(t)$ is a periodic pulse train with time period, $T = \frac{1}{f_c}$ as shown below in Figure 2.14.

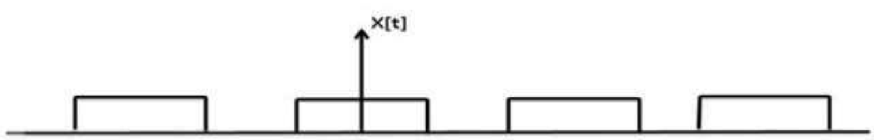


Fig. 2.14 *Periodic Pulse Train* x(t) *with Time Period* $T = \frac{1}{f_c}$

The Fourier series representation of this periodic pulse train is given as,

$$x(t) = \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n - 1}{2n-1}\cos(2\pi(2n-1)f_ct)$$

$$\Rightarrow x(t) = \frac{1}{2} + \frac{2}{\pi}\cos(2\pi f_ct) - \frac{2}{3\pi}\cos(6\pi f_ct) + \ldots$$

Substitute, $V_1(t)$ and $x(t)$ values in Equation 2.10.

$$V_2(t) = [m(t) + A_c\cos(2\pi f_ct)]\left[\frac{1}{2} + \frac{2}{\pi}\cos(2\pi f_ct) - \frac{2}{3\pi}\cos(6\pi f_ct) + \ldots\right]$$

$$V_2(t) = \frac{m(t)}{2} + \frac{A_c}{2}\cos(2\pi f_ct) + \frac{2m(t)}{\pi}\cos(2\pi f_ct) + \frac{2A_c}{\pi}\cos^2(2\pi f_ct) - \frac{2m(t)}{3\pi}\cos(6\pi f_ct) - \frac{2A_c}{3\pi}\cos(2\pi f_ct)\cos(6\pi f_ct) + \ldots$$

$$V_2(t) = \frac{A_c}{2}\left(1 + \left(\frac{4}{\pi A_c}\right)m(t)\right)\cos(2\pi f_ct) + \frac{m(t)}{2} + \frac{2A_c}{\pi}\cos^2(2\pi f_ct) - \frac{2m(t)}{3\pi}\cos(6\pi f_ct) - \frac{2A_c}{3\pi}\cos(2\pi f_ct)\cos(6\pi f_ct) + \ldots$$

The 1st term of the above equation represents the desired AM wave, and the remaining terms are unwanted terms. Thus, with the help of band pass filter, we can pass only AM wave and eliminate the remaining terms.

Therefore, the output of switching modulator is,

$$s(t) = \frac{A_c}{2}\left(1 + \left(\frac{4}{\pi A_c}\right)m(t)\right)\cos(2\pi f_ct)$$

The standard equation of AM wave is,

$$s(t) = A_c[1 + k_am(t)]\cos(2\pi f_ct)$$

Where, k_a is the amplitude sensitivity.

By comparing the output of the switching modulator with the standard equation of AM wave, we get the scaling factor as 0.5 and amplitude sensitivity k_a as,

$$\frac{4}{\pi A_c}$$

Check Your Progress

1. Define the term Amplitude Modulation (AM).
2. State about the Double SideBand - Suppressed Carrier (DSB-SC).
3. What is Single SideBand-Suppressed Carrier (SSB-SC)?
4. Define the term data rate. What does C Shannon theorem states?
5. State about the low level modulation and high level modulation.
6. Why should modulation be employed when it is possible to directly transmit a baseband signal?
7. What does standard AM's carrier provide?
8. Define modulation index.
9. State about multitone modulation.

2.3 DEMODULATION OF AM

NOTES

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, therefore, 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).

Demodulation was first used in radio receivers. In the wireless telegraphy radio systems used during the first 3 decades of radio (1884-1914) the transmitter did not communicate audio (sound) but transmitted information in the form of pulses of radio waves that represented text messages in 'Morse Code'. Therefore, the receiver merely had to detect the presence or absence of the radio signal and produce a click sound. The device that did this was called a detector. The first detectors were coherers, simple devices that acted as a switch. The term detector stuck, was used for other types of demodulators and continues to be used to the present day for a demodulator in a radio receiver.

The first type of modulation used to transmit sound over radio waves was Amplitude Modulation (AM), invented by Reginald Fessenden around 1900. An AM radio signal can be demodulated by rectifying it to remove one side of the carrier, and then filtering to remove the radio-frequency component, leaving only the modulating audio component. This is equivalent to peak detection with a suitably long time constant.

Techniques

There are several ways of demodulation depending on how parameters of the base-band signal, such as amplitude, frequency or phase are transmitted in the carrier signal. For example, for a signal modulated with a linear modulation like AM (Amplitude Modulation), we can use a synchronous detector. On the other hand, for a signal modulated with an angular modulation, we must use an FM (Frequency Modulation) demodulator or a PM (Phase Modulation) demodulator. Different kinds of circuits perform these functions.

Many techniques, such as carrier recovery, clock recovery, bit slip, frame synchronization, rake receiver, pulse compression, Received Signal Strength Indication (RSSI), error detection and correction, etc., are only performed by demodulators, although any specific demodulator may perform only some or none of these techniques.

2.3.1 Synchronous Detection - Nonlinear Demodulation

In electronics, a synchronous detector is a device that recovers information from a modulated signal by mixing the signal with a replica of the un-modulated carrier. This can be locally generated at the receiver using a phase-locked loop or other techniques. Synchronous detection preserves any phase information originally present in the modulating signal. Synchronous detection is a necessary component of any analog color television receiver, where it allows recovery of the phase

NOTES

information that conveys hue. Synchronous detectors are also found in some shortwave radio receivers used for audio signals, where they provide better performance on signals that may be affected by fading. To recover baseband signal, the synchronous detection technique is used.

For synchronous demodulation, a mixer is used. The incoming signal is fed into the signal input of the mixer, and a local oscillator signal on the same frequency as the carrier of the incoming signal is fed into the other. This mixing process converts the carrier to a 0Hz signal and the sidebands to their base band frequency band, i.e., it reconstitutes the audio.

As the carrier has a frequency of 0Hz, it appears as a DC voltage on the output - the DC level will depend on the phase between the carrier and the local oscillator. The sidebands of the AM signal will appear relative to zero frequency, i.e., as the original audio or other modulating signal. Figure 2.15 illustrates the synchronous demodulation.

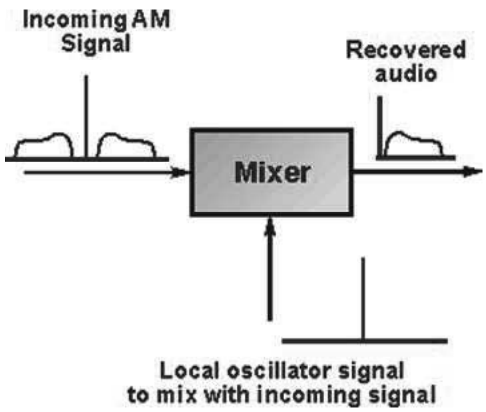


Fig. 2.15 *Synchronous Demodulation*

2.3.2 Square Law or Non-Linear Diode Detector/ Demodulation

Square law diode detection lies in the non-linear portion of dynamic current-voltage characteristic of a diode. It differs from the linear diode detection in the sense that in this case, the applied input carrier voltage is of small magnitude and hence is restricted to nonlinear portion of dynamic characteristic whereas in linear diode detectors, large amplitude modulated carrier voltage is applied to the diode and the operation takes place over linear region of the characteristic. The diode is biased for operation in nonlinear region of the characteristic of diode. The capacitor-resistor combination acts as load.

Since the operation takes place over the non-linear region of the characteristic, the current wave form has its lower half compressed. This result in the average value of the current. This average component consists of a steady or DC component and a time varying component of modulation frequency. The shunt capacitor C bypasses the entire radio frequency component leaving only the average component to flow through R, producing the desired detected output.

Square law demodulator is used to demodulate low level AM wave. Following Figure 2.16 illustrates the block diagram of the **square law demodulator**.

NOTES

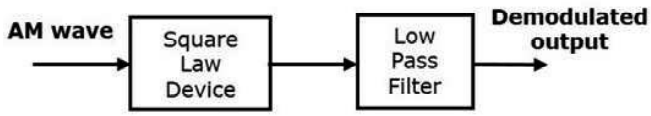


Fig. 2.16 *Block Diagram of the Square Law Demodulator*

This demodulator contains a square law device and low pass filter. The AM wave $V_1(t)$ is applied as an input to this demodulator.

The standard form of AM wave is,

$$V_1(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t)$$

We know that the mathematical relationship between the input and the output of square law device is,

$$V_2(t) = k_1 V_1(t) + k_2 V_1^2(t) \qquad \ldots(2.11)$$

Where,

$V_1(t)$ = Input of the Square Law Device, which is the AM Wave

$V_2(t)$ = Output of the Square Law Device

k_1 and k_2 = Constants

Substitute $V_1(t)$ in Equation 2.11,

$$V_2(t) = k_1(A_c[1 + k_a m(t)]\cos(2\pi f_c t)) + k_2(A_c[1 + k_a m(t)]\cos(2\pi f_c t))^2$$

$$\Rightarrow V_2(t) = k_1 A_c \cos(2\pi f_c t) + k_1 A_c k_a m(t)\cos(2\pi f_c t) +$$
$$k_2 A_c^2\left[1 + K_a^2 m^2(t) + 2k_a m(t)\right]\left(\frac{1+\cos(4\pi f_c t)}{2}\right)$$

$$\Rightarrow V_2(t) = k_1 A_c \cos(2\pi f_c t) + k_1 A_c k_a m(t)\cos(2\pi f_c t) + \frac{K_2 A_c^2}{2} +$$
$$\frac{K_2 A_c^2}{2}\cos(4\pi f_c t) + \frac{k_2 A_c^2 k_a^2 m^2(t)}{2} + \frac{k_2 A_c^2 k_a^2 m^2(t)}{2}\cos(4\pi f_c t) +$$
$$k_2 A_c^2 k_a m(t) + k_2 A_c^2 k_a m(t)\cos(4\pi f_c t)$$

In the above equation, the term $k_2 A_c^2 k_a m(t)$ is the scaled version of the message signal. It can be extracted by passing the above signal through a low pass filter and the DC component $\frac{k_2 A_c^2}{2}$ can be eliminated with the help of a coupling capacitor.

The circuit diagram of square law demodulator is shown in Figure 2.17.

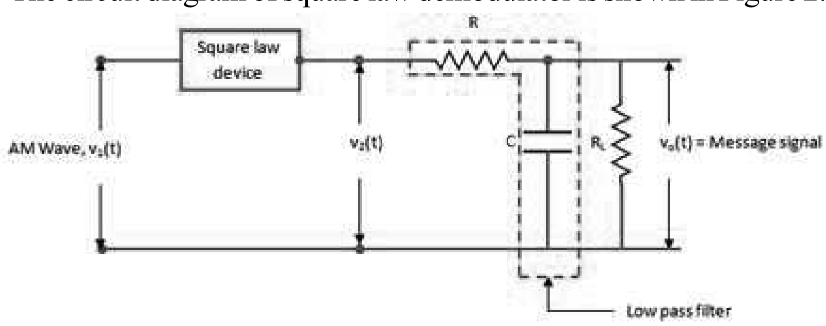


Fig. 2.17 *Circuit Diagram of Square Law Demodulator*

Figure 2.18 illustrates the generation of AM waves using the non-linear square law modulator circuit.

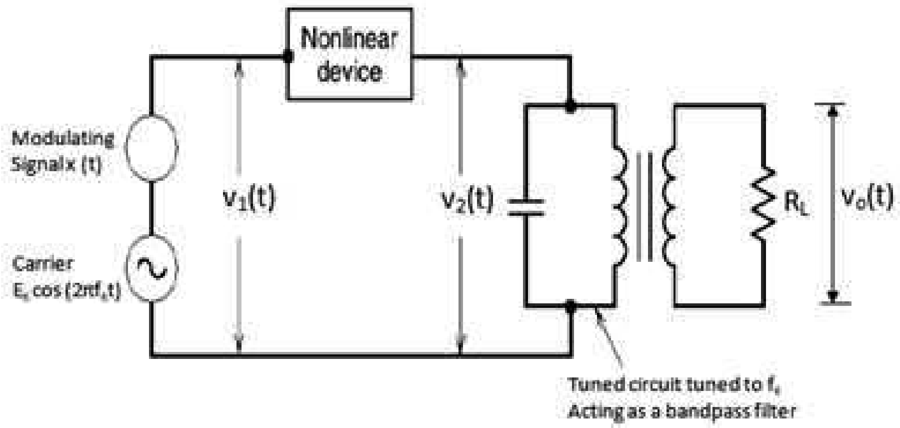


Fig. 2.18 *Non-Linear Square Law Modulator Circuit*

Due to non-linear region, the anode current flowing in the diode is expressed as:

$$i = a_1 e + a_2 e^2$$

where $e = E_c(1 + H\cos\omega_m t)\cos\omega_c t$, hence

$$i = a_1 E_c(1 + H_a\cos\omega_m t)\cos\omega_c t + a_2\left[E_c(1 + H_a\cos\omega_m t)\cos\omega_c t\right]^2$$

$$= a_1 E_c(1 + H_a\cos\omega_m t)\cos\omega_c t + a_2 E_c{}^2$$

$$(1 + H^2\cos^2\omega_m t + 2H_a\cos\omega_m t)\cos^2\omega_c t$$

$$= a_1 E_c\cos\omega_c t + a_1 E_c H\cos\omega_m t\cos\omega_c t + a_2 E_c{}^2\cos^2\omega_c t$$

$$+ a_2 E_c{}^2 H^2\cos^2\omega_m t\cos^2\omega_c t + 2H_a a_2 E_c{}^2\cos\omega_m t\cos^2\omega_c t$$

$$= a_1 E_c\cos\omega_c t + \frac{a_1 E_c H}{2}[\cos\frac{(\omega_c + \omega_m)t}{2} + \cos\frac{(\omega_c - \omega_m)t}{2}]$$

$$+ a_2 E_c{}^2\left(\frac{\cos 2\omega_c t + 1}{2}\right) + a_2 E_c{}^2 H^2\left(\frac{\cos 2\omega_m t + 1}{2}\right)\left(\frac{\cos 2\omega_c t + 1}{2}\right)$$

$$+ 2H\ a_2 E_c{}^2\cos\omega_m t\left(\frac{\cos 2\omega_c t + 1}{2}\right)$$

$$= a_1 E_c\cos\omega_c t + \frac{a_1 E_c H}{2}\cos\frac{(\omega_c + \omega_m)t}{2} + \frac{a_1 E_c H}{2}\ \cos\frac{(\omega_c - \omega_m)t}{2}$$

$$+ a_2 E_c{}^2\frac{\cos 2\omega_c t}{2} + \frac{a_2 E_c{}^2}{2} + \frac{a_2 E_c{}^2 H^2}{4} + \frac{a_2 E_c{}^2 H^2\cos 2\omega_m t}{4}$$

$$+ \frac{a_2 E_c{}^2 H\cos 2\omega_c t}{4} + \frac{a_2 E_c{}^2 H^2\, 2\cos 2\omega_m \text{t}\cos 2\omega_c t}{8}$$

$$+ \frac{a_2 E_c{}^2 H\ 2\cos 2\omega_m \text{t}\cos 2\omega_c t}{2} + H\ a_2 E_c{}^2\cos\omega_m t$$

NOTES

$$= a_1 E_c \cos \omega_c t + \frac{a_1 E_c H}{2} \cos \frac{(\omega_c + \omega_m)t}{2} + \frac{a_1 E_c H}{2} \cos \frac{(\omega_c - \omega_m)t}{2}$$

$$+ a_2 E_c{}^2 \frac{\cos 2\omega_c t}{2} + \frac{a_2 E_c{}^2}{2} + \frac{a_2 E_c{}^2 H^2}{4} + \frac{a_2 E_c{}^2 H^2 \cos 2\omega_m t}{4}$$

$$+ \frac{a_2 E_c{}^2 H^2 \cos 2\omega_c t}{4} + \frac{a_2 E_c{}^2 H^2}{8} cos 2(\omega_c + \omega_m)t$$

$$+ \frac{a_2 E_c{}^2 H^2}{8} cos 2(\omega_c - \omega_m)t + a_2 E_c{}^2 H^2 \cos \omega_m t$$

In the above expression, we get the terms in frequencies,

$$2\omega_c,\ 2(\omega_c + \omega_m),\ 2(\omega_c - \omega_m),\ \omega_m,\ 2\ \omega_m,\ (\omega_c + \omega_m),$$
$$(\omega_c - \omega_m),\ (2\omega_c + \omega_m),\ and\ (2\omega_c - \omega_m)$$

The RF (Radio Frequency) terms are passsed by the shunt capacitor C while the voltage corresponding to the frequencies ω_m and $2\omega_m$ are developed across the load resistor R. The term in frequency ?m is the desired output whereqas the term in frequency $2\omega_m$ gives the distrortion term. The second harmonic component should be kept as low as possible. If it is to kept low, say 10% of the fundamental component, then it is necessary that ($\frac{H^2}{4}$ (*Second Harmonic Amplitude*)< $\frac{H}{10}$ (*Fundamental Amplitude*) or $m_a < 0.4$, that is, the modulation index should be less that 0.4. Ordinary broadcast is generally performed at 40%.

2.3.3 Suppress Carrier AM Demodulator

Suppressed carrier transmission is a special case in which the carrier level is reduced below that required for demodulation by a normal receiver. We know that the carrier voltage $E_c \cos\omega_c t$ contains no information. The information is only contained in each of the two sidebands. Accordingly, **carrier** may be **suppressed** or eliminated. Further, the carrier takes up large portion of the total modulated power $\left(P_c = \frac{2}{3} P_t\right)$, therefore by suppressing the carrier, the power is saved. In ordinary AM radio broadcast, carrier is allowed along with the two side bands and the system is referred as the double side band system. Carrier is necessary for the reproduction of modulating signal in the detector of radio receiver. Accordingly, in ordinary AM broadcast, carrier is allowed to propagate. If the carrier is eliminated or suppressed, the system becomes Suppressed Carrier Double SideBand (DSB-SC) system. In this case, the carrier is reinserted in the radio receiver and the circuit becomes complicated as well as costly. Hence, DSB-SC may be used in point-to-point communication. The circuit of suppressed carrier balanced modulator using transistors.

The carrier voltage is applied in parallel to the input of the two transistors whereas the modulating voltage is applied in push pull to the same two transistors. In this case, the base voltage for the two transistors can be expressed as:

$$e_{b_1} = e_c + e_m$$

$$e_{b_1} = E_c cos\omega_c t + E_m cos\omega_m t \qquad \ldots\ldots. (2.12)$$

And $$e_{b_1} = e_c - e_m$$

$$e_{b_2} = E_c cos\omega_c t - E_m cos\omega_m t \qquad \ldots\ldots. (2.13)$$

The AC collector current for the two transistors can be written as:

$$i_{c_1} = a_1 e_{b_1} + a_2 e_{b_1}^2$$

$$= a_1\left(E_c cos\omega_c t + E_m cos\omega_m t\right) + a_2\left(E_c cos\omega_c t + E_m cos\omega_m t\right)^2$$

$$= a_1\left(E_c cos\omega_c t + E_m cos\omega_m t\right)$$

$$+ a_2\left(E_c{}^2 cos^2\omega_c t + E_m{}^2 cos^2\omega_m t + 2E_c E_m cos\omega_c t cos\omega_m t\right) \qquad \ldots (2.14)$$

And $$i_{c_2} = a_1 e_{b_2} + a_2 e_{b_2}^2$$

$$= a_1\left(E_c cos\omega_c t - E_m cos\omega_m t\right) + a_2\left(E_c cos\omega_c t - E_m cos\omega_m t\right)^2$$

$$= a_1\left(E_c cos\omega_c t - E_m cos\omega_m t\right)$$

$$+ a_2\left(E_c{}^2 cos^2\omega_c t + E_m{}^2 cos^2\omega_m t - 2E_c E_m cos\omega_c t cos\omega_m t\right) \qquad \ldots.. (2.15)$$

Hence, the output voltage is given by:

$$e_o = K\left(i_{c_1} - i_{c_2}\right)$$

$$= K[2a_1 E_m cos\omega_m t + 4a_2 E_c E_m cos\omega_c t cos\omega_m t]$$

$$= 2Ka_1 E_m cos\omega_m t + 2Ka_2 E_c E_m [\cos\left(\omega_c + \omega_m\right) t cos\left(\omega_c - \omega_m\right) t$$

Therefore, the output voltage consists of modulating frequency voltage and two sidebands. The carrier voltage has been suppressed. The tuned circuit in the output is tuned to resonate at frequency ω_c and so it responds to a band of frequencies centered about ω_c. Here in the output only two sidebands appear, i.e.,

$$e_o = 2Ka_2 E_c E_m [\cos\left(\omega_c + \omega_m\right) t cos\left(\omega_c - \omega_m\right) t$$

$$= 4Ka_2 E_c E_m cos\omega_c t cos\omega_m t$$

2.3.4 Envelope Detector

An **envelope detector**, sometimes called a peak detector, is an electronic circuit that takes a relatively high frequency amplitude modulated signal as input and provides an output, which is the demodulated envelope of the original signal. Figure 2.19 illustrates simple envelope demodulator circuit.

NOTES

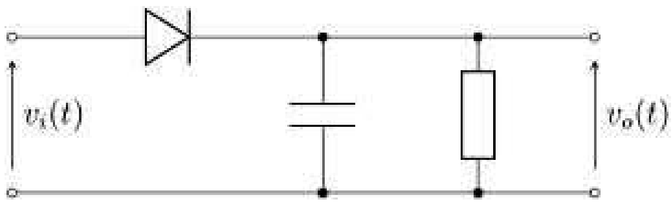


Figure 2.19 *Simple Envelope Demodulator Circuit*

Circuit Operation

The capacitor in the circuit stores charge on the rising edge and releases it slowly through the resistor when the input signal amplitude falls. The diode in series rectifies the incoming signal, allowing current flow only when the positive input terminal is at a higher potential than the negative input terminal.

Most practical envelope detectors use either half-wave or full-wave rectification of the signal to convert the AC audio input into a pulsed DC signal. Filtering is then used to smooth the result. This filtering is rarely perfect, and some 'Ripple' is likely to remain on the envelope follower output, particularly for low frequency inputs, such as notes from a bass instrument. Reducing the filter cutoff frequency gives a smoother output but decreases the high frequency response. Therefore, practical designs must reach a compromise.

Definition of the Envelope

Any AM or FM signal $x(t)$ can be written in the following form,

$$x(t) = R(t)\cos(\omega t + \varphi(t))$$

In the case of AM, $\varphi(t)$ the phase component of the signal is constant and can be ignored. In AM, the carrier frequency ω is also constant. Thus, all the information in the AM signal is in $R(t)$. $R(t)$ is called the **envelope** of the **signal**. Hence, an AM signal is given by the function,

$$x(t) = (C + m(t))\cos(\omega t)$$

With $m(t)$ representing the original audio frequency message, C the carrier amplitude and $R(t)$ equal to $C + m(t)$. Therefore, if the envelope of the AM signal can be extracted, then the original message can be recovered.

In the case of FM, the transmitted $x(t)$ has a constant envelope $R(t) = R$ and can be ignored. However, many FM receivers measure the envelope anyway for received signal strength indication.

Diode Detector

The simplest form of envelope detector is the diode detector. A diode detector is simply a diode between the input and output of a circuit, connected to a resistor and capacitor in parallel from the output of the circuit to the ground. If the resistor and capacitor are correctly chosen, the output of this circuit should approximate a voltage-shifted version of the original (baseband) signal. A simple filter can then be applied to filter out the DC component.

Precision Detector

An envelope detector can also be constructed using a precision rectifier feeding into a low pass filter.

Demodulation of Signals

An envelope detector can be used to demodulate a previously modulated signal by removing all high frequency components of the signal. The capacitor and resistor form a low pass filter to filter out the carrier frequency. Such a device is often used to demodulate AM radio signals because the envelope of the modulated signal is equivalent to the baseband signal.

2.3.5 Square Law Demodulator

The process of separating the original information or 'Signal' from the 'Modulated Carrier'. In the case of 'Amplitude or Frequency Modulation' it involves a device, called a **Demodulator** or **Detector**, which produces a signal corresponding to the instantaneous changes in amplitude or frequency, respectively. This signal corresponds to the original modulating signal.

In electronic signal processing, a square law detector is a device that produces an output proportional to the square of some input. For example, in demodulating radio signals, a semiconductor diode can be used as a square law detector, providing an output current proportional to the square of the amplitude of the input voltage over some range of input amplitudes. A square law detector provides an output directly proportional to the power of the input electrical signal. In radio transmission, this process is a major function of a receiver, in order to retrieve the desired signal.

2.3.6 DSB-SC and SSB Modulation Systems

Double-SideBand Suppressed Carrier (DSB-SC) transmission is transmission in which frequencies produced by Amplitude Modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level, ideally being completely suppressed. The name 'Suppressed Carrier' comes about because the carrier signal component is suppressed—it does not appear in the output signal. This is apparent when the spectrum of the output signal is viewed.

In the DSB-SC modulation, unlike in AM, the wave carrier is not transmitted; thus, much of the power is distributed between the side bands, which implies an increase of the cover in DSB-SC, compared to AM, for the same power use.

DSB-SC transmission is a special case of double-sideband reduced carrier transmission. It is used for radio data systems. This mode is frequently used in Amateur radio voice communications, especially on High Frequency (HF) bands.

Spectrum

DSB-SC is basically an amplitude modulation wave without the carrier, therefore reducing power waste, giving it a 50% efficiency. This is an increase compared to normal AM transmission (DSB) that has a maximum efficiency of 33.333%, since

NOTES

2/3 of the power is in the carrier which conveys no useful information and both sidebands containing identical copies of the same information. Single SideBand Suppressed Carrier (SSB-SC) is 100% efficient.

Following Figure 2.20 illustrates the 'Spectrum' plot of a DSB-SC signal:

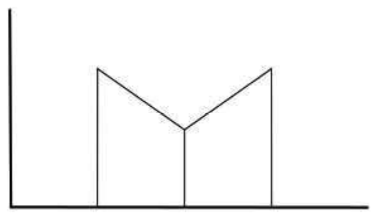


Fig. 2.20 *The 'Spectrum' Plot of a DSB-SC Signal*

Demodulation

For DSBSC, Coherent Demodulation is done by multiplying the DSB-SC signal with the carrier signal with the same phase as in the modulation process just like the modulation process. This resultant signal is then passed through a low pass filter to produce a scaled version of the original message signal.

$$\overbrace{\frac{V_m V_c}{2}\left[\cos((\omega_m+\omega_c)\,t)+\cos((\omega_m-\omega_c)\,t)\right]}^{\text{Modulated Signal}} \times \overbrace{V_c' \cos(\omega_c t)}^{\text{Carrier}}$$
$$= \left(\frac{1}{2}V_c V_c'\right)\underbrace{V_m \cos(\omega_m t)}_{\text{original message}} + \frac{1}{4}V_c V_c' V_m \left[\cos((\omega_m+2\omega_c)t)+\cos((\omega_m-2\omega_c)t)\right]$$

The equation above shows that by multiplying the modulated signal by the carrier signal, the result is a scaled version of the original message signal plus a second term. Since, $\omega_c \gg \omega_m$ this second term is much higher in frequency than the original message. Once this signal passes through a low pass filter, the higher frequency component is removed, leaving just the original message.

Distortion and Attenuation

For demodulation, the demodulation oscillator's frequency and phase must be the same as the modulation oscillator's, otherwise, distortion and/or attenuation will occur.

To observe this effect, take the following conditions:

- Message signal to be transmitted: $f(t)$
- Modulation (carrier) signal: $V_c \cos(\omega_c t)$
- Demodulation signal (with small frequency and phase deviations from the modulation signal): $V_c' \cos[(\omega_c + \Delta\omega)t + \theta]$

The resultant signal can then be given by,

$$f(t) \times V_c \cos(\omega_c t) \times V_c' \cos[(\omega_c + \Delta\omega)t + \theta]$$
$$= \frac{1}{2}V_c V_c' f(t) \cos(\Delta\omega \cdot t + \theta) + \frac{1}{2}V_c V_c' f(t) \cos[(2\omega_c + \Delta\omega)t + \theta]$$
$$\xrightarrow{\text{After low pass filter}} \frac{1}{2}V_c V_c' f(t) \cos(\Delta\omega \cdot t + \theta)$$

NOTES

The $\cos(\Delta\omega{\cdot}t + \theta)$ terms results in distortion and attenuation of the original message signal. In particular, if the frequencies are correct, but the phase is wrong, contribution from θ is a constant attenuation factor, also $\cos \Delta\omega{\cdot}t$ represents a cyclic inversion of the recovered signal, which is a serious form of distortion.

SSB (Single SideBand) Modulation Systems

In radio communications, **Single SideBand (SSB) modulation** 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.

Basic Concept

Radio transmitters work by mixing a Radio Frequency (RF) signal of a specific frequency, the carrier wave, with the audio signal to be broadcast. In AM transmitters this mixing usually takes place in the final RF amplifier (high level modulation). It is less common and much less efficient to do the mixing at low power and then amplify it in a linear amplifier. Either method produces a set of frequencies with a strong signal at the carrier frequency and with weaker signals at frequencies extending above and below the carrier frequency by the maximum frequency of the input signal. Therefore, the resulting signal has a spectrum whose bandwidth is twice the maximum frequency of the original input audio signal.

SSB takes advantage of the fact that the entire original signal is encoded in each of these 'Sidebands'. It is not necessary to transmit both sidebands plus the carrier, as a suitable receiver can extract the entire original signal from either the upper or lower sideband. There are several methods for eliminating the carrier and one sideband from the transmitted signal. Producing this single sideband signal is too complicated to be done in the final amplifier stage as with AM. SSB Modulation must be done at a low level and amplified in a linear amplifier where lower efficiency partially offsets the power advantage gained by eliminating the carrier and one sideband. Nevertheless, SSB transmissions use the available amplifier energy considerably more efficiently, providing longer-range transmission for the same power output. In addition, the occupied spectrum is less than half that of a full carrier AM signal.

SSB reception requires frequency stability and selectivity well beyond that of inexpensive AM receivers which is why broadcasters have seldom used it. In point-to-point communications, where expensive receivers are in common use already, they can successfully be adjusted to receive whichever sideband is being transmitted.

Mathematical Formulation

NOTES

Single SideBand (SSB) has the mathematical form of Quadrature Amplitude Modulation (QAM) in the special case where one of the baseband waveforms is derived from the other, instead of being independent messages:

$$s_{\text{ssb}}(t) = s(t) \cdot \cos(2\pi f_0 t) - \hat{s}(t) \cdot \sin(2\pi f_0 t).$$

Where $s(t)$ is the message (real-valued), $\hat{s}(t)$ is its Hilbert transform, and f_0 is the radio carrier frequency.

To understand this formula, we may express $s(t)$ as the real part of a complex-valued function, with no loss of information:

$$s(t) = \text{Re}\{s_a(t)\} = \text{Re}\{s(t) + j \cdot \hat{s}(t)\}$$

Where $s_a(t)$ represents the imaginary unit. $s(t)$ is the analytic representation of which means that it comprises only the positive-frequency components of $s(t)$:

$$\frac{1}{2}S_a(f) = \begin{cases} S(f), & \text{for } f > 0, \\ 0, & \text{for } f < 0, \end{cases}$$

Where $S_a(f)$and $S(f)$ are the respective Fourier transforms of and $s(t)$. Therefore, the frequency-translated function contains only one side of $S(f)$.

2.3.7 Sideband and Carrier Power

In radio communications, a **sideband** is a band of frequencies higher than or lower than the carrier frequency, that are the result of the modulation process. The sidebands carry the information transmitted by the radio signal. The sidebands comprise all the spectral components of the modulated signal except the carrier. The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). All forms of modulation produce sidebands.

Figure 2.21 illustrates the power of an AM radio signal plotted against frequency, where f_c is the carrier frequency, f_m is the maximum modulation frequency.

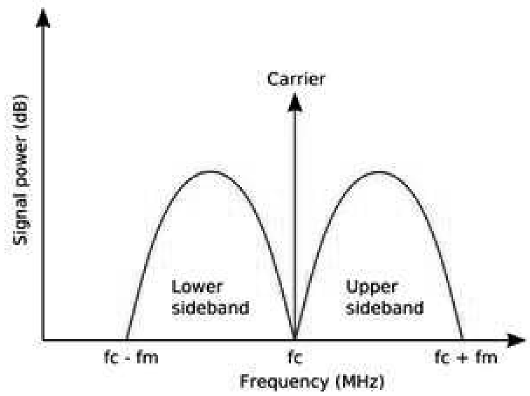


***Fig. 2.21** Power of an AM Radio Signal Plotted against Frequency, where f_c is the Carrier Frequency, f_m is the Maximum Modulation Frequency*

Sideband Creation

We can illustrate the creation of sidebands with one trigonometric identity as shown below:

$$\cos(A) \cdot \cos(B) \equiv \tfrac{1}{2}\cos(A+B) + \tfrac{1}{2}\cos(A-B)$$

Adding cos(A) to both sides, we have:

$$\cos(A) \cdot [1+\cos(B)] = \tfrac{1}{2}\cos(A+B) + \cos(A) + \tfrac{1}{2}\cos(A-B)$$

Substituting (for instance)andwhere t represents time:

$$\underbrace{\cos(1000\,t)}_{\text{carrier wave}} \cdot \underbrace{[1+\cos(100\,t)]}_{\text{amplitude modulation}} = \underbrace{\tfrac{1}{2}\cos(1100\,t)}_{\text{upper sideband}} + \underbrace{\cos(1000\,t)}_{\text{carrier wave}} + \underbrace{\tfrac{1}{2}\cos(900\,t)}_{\text{lower sideband}}$$

Adding more complexity and time variation to the amplitude modulation also adds it to the sidebands, causing them to widen in bandwidth and change with time. In effect, the sidebands 'Carry' the information content of the signal.

Amplitude Modulation

Amplitude Modulation (AM) of a carrier signal normally results in two mirror image sidebands. The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). For example, if a 900 kHz carrier is amplitude modulated by means of a 1 kHz audio signal, there will be components at 899 kHz and 901 kHz as well as 900 kHz in the generated radio frequency spectrum, therefore an audio bandwidth of 7 kHz will require a radio spectrum bandwidth of 14 kHz. In conventional AM transmission, as used by broadcast band AM stations, the original audio signal can be recovered or 'Detected' by either synchronous detector circuits or by simple envelope detectors because the carrier and both sidebands are present. This is sometimes called Double SideBand - Amplitude Modulation (DSB-AM), but not all variants of DSB are compatible with envelope detectors.

In some forms of AM, the carrier may be reduced, to save power. The term DSB reduced-carrier normally implies enough carrier remains in the transmission to enable a receiver circuit to regenerate a strong carrier or at least synchronise a phase-locked loop but there are forms where the carrier is removed completely, producing Double SideBand with Suppressed Carrier (DSB-SC). Suppressed carrier systems require more sophisticated circuits in the receiver and some other method of deducing the original carrier frequency. An example is the stereophonic difference (L-R) information transmitted in stereo FM broadcasting on a 38 kHz subcarrier where a low power signal at half the 38-kHz carrier frequency is inserted between the monaural signal frequencies (up to 15 kHz) and the bottom of the stereo information sub-carrier (down to 38–15 kHz, i.e., 23 kHz). The receiver locally regenerates the subcarrier by doubling a special 19 kHz pilot tone.

If part of one sideband and all of the other remain, it is called **vestigial sideband**, used mostly with television broadcasting, which would otherwise take up an unacceptable amount of bandwidth. Transmission in which only one sideband is transmitted is called Single SideBand or SSB modulation. SSB is the predominant voice mode on shortwave radio other than shortwave broadcasting. Since the sidebands are mirror images, which sideband is used is a matter of convention.

NOTES

In SSB, the **carrier** is suppressed, significantly reducing the electrical power (by up to 12 dB) without affecting the information in the sideband. This makes for more efficient use of transmitter power and RF bandwidth, but a beat frequency oscillator must be used at the receiver to reconstitute the carrier. If the reconstituted carrier frequency is wrong then the output of the receiver will have the wrong frequencies, but for speech small frequency errors are no problem for intelligibility. Another way to look at an SSB receiver is as an RF-to-audio frequency transposer: in USB mode, the dial frequency is subtracted from each radio frequency component to produce a corresponding audio component, while in LSB mode each incoming radio frequency component is subtracted from the dial frequency.

Total power in AM signal is the sum of carrier power and power of both sidebands (upper side band and lower side band).

$$P_t = P_c + P_{USB} + P_{LSB}$$
$$P_t = P_c + (P_c\mu^2/4) + (P_c\mu^2/4)$$
$$P_t = P_c\,(1+\mu^2/2)$$

P_t = Total power of AM single

P_c = Power carrier single

P_{USB} = Power of upper side band $(P_c\mu^2/4)$

P_{LSB} = Power of upper side band $(P_c\mu^2/4)$

μ = Modulation Index

In DSB-SC AM signal, carrier signal is suppressed from modulated single hence the transmitted total power will be only the power of sidebands. Hence for 100% modulation about 67% of total power is required for transmitted.

$$P_t = P_{USB} + P_{LSB}$$
$$P_t = (P_c\mu^2/4) + (P_c\mu^2/4)$$
$$P_t = P_c\,(\mu^2/2)$$

In SSB-SC AM signal system, carrier or either of one sideband is suppressed from modulated signal, hence the transmitted total power will be only the power of sideband.

$$P_t = P_{SB}$$
$$P_t = P_c\,(\mu^2/4)$$

Frequency Modulation

Frequency Modulation (FM) also generates sidebands, the bandwidth consumed depending on the modulation index - often requiring significantly more bandwidth than DSB. Bessel functions can be used to calculate the bandwidth requirements of FM transmissions.

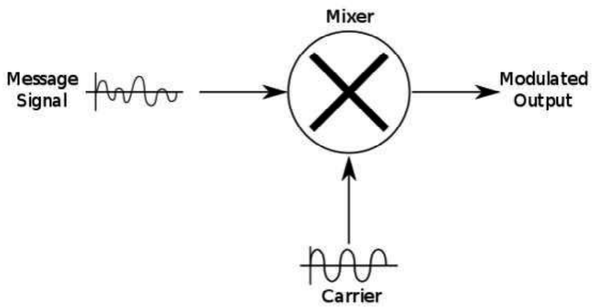
2.3.8 Method of Generation and Detection of DSB-SC and SSB

Generation of DSB-SC

DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal. The mathematical representation of this process is shown below, where the product-to-sum trigonometric identity is used.

NOTES

$$\underbrace{V_m \cos(\omega_m t)}_{\text{Message}} \times \underbrace{V_c \cos(\omega_c t)}_{\text{Carrier}} = \underbrace{\frac{V_m V_c}{2} \left[\cos((\omega_m + \omega_c)\, t) + \cos((\omega_m - \omega_c)\, t)\right]}_{\text{Modulated Signal}}$$



Balanced Modulators

A balanced modulator is a electronic circuit which generates a DSB-SC signals. In DSB-SC carrier is suppressed and only the sum and difference frequencies (sidebands) are left at the output. For generating SSB signal the resulting output of balance modulator is further processed by filter/phase-shifting circuits to eliminate one of the sidebands.

Lattice Modulators

Lattice modulator is one of the most popular and widely used balanced modulators in DSB-SC systems. The circuit of lattice modulator *or* the diode ring is shown in Figure 2.22, consisting of two centre tapped transformers *T1* and T2. Transformer *T*1 stands for an input transformer and T2 stands for an output transformer. The four diodes are connected in a bridge arrangement in the circuit. The modulating single (information which will be modulate) is applied to the primary coil of T1 and the resulting modulated signal will received from secondary coil of T2. The carrier single is applied to the centre tapped terminal of both transformers as arrangement shown in figure.

The working principle of the lattice modulator is relatively simple and depends on the switching condition of the diodes. Diode will be ON/OFF according to their forward and reverse baising conditions. The frequency and amplitude carrier single is usually considerably higher in than the modulating signal. The carrier signal turns the diodes off and on at a high rate of speed, used as a source of forward and reverse bias for the diodes.

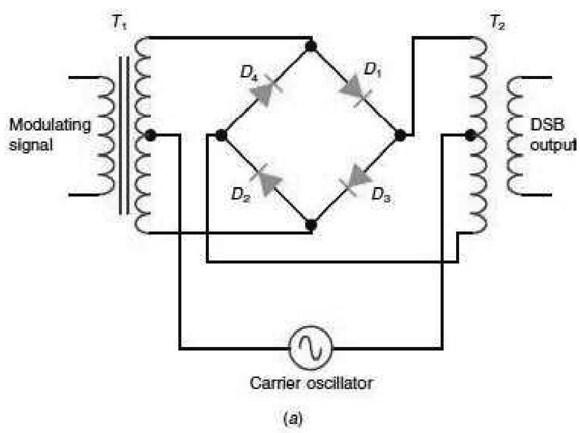


Fig 2.22 *Balanced Modulator (Lattice Modulator)*

When the modulating input is zero and the polarity of the carrier is positive, diodes *D*1 and *D*2 (mentioned in figure) are forward-biased and act as close switch. At the same instant *D*3 and *D*4 are reverse-biased and act as open switch. The current flows in the circuit divide equally in the upper and lower portions of the primary winding of *T*2. The magnetic field produced by current flows in the upper part of the secondary winding equal but opposite to the magnetic field produced by the current in the lower half of the secondary. Thus magnetic field cancels each other and no output is induced in the secondary, which means the carrier is effectively suppressed.

If the carrier is negative (reverses polarity), diodes *D*3 and *D*4 are in forward bias and act as close switch. At the same instant *D*1 and *D*3 are reverse-biased and act as open switch. Again, the current flows in the circuit divide equally in the upper and lower portions of the primary winding of *T*2. The magnetic field produced by current flows in the upper part of the secondary winding equal but opposite to the magnetic field produced by the current in the lower half of the secondary. Thus magnetic field cancels each other and no output is induced in the secondary. The degree of carrier suppression depends on the degree of precision with which the transformers are made and the placements of the center tap.

Now suppose a low-frequency modulating signal is applied to the primary of transformer *T*1. Then that single appears across the secondary of *T*1. As we discussed earlier the diodes will connect or disconnected between the secondary of *T*1 to the primary of *T*2 at different times depending upon the polarity of carrier single. When the carrier polarity is positive, diodes *D*1 & *D*2 acts as closed switches and *D*3 & *D*4 are act as open switches. As a result, the modulating signal at the secondary of *T*1 is applied to the primary of *T*2 through *D*1 and *D*2. When the carrier polarity negative, all the condition discussed above will opposite and the result is a 180° phase reversal. If the modulating signal is positive, the output will be negative, and vice versa.

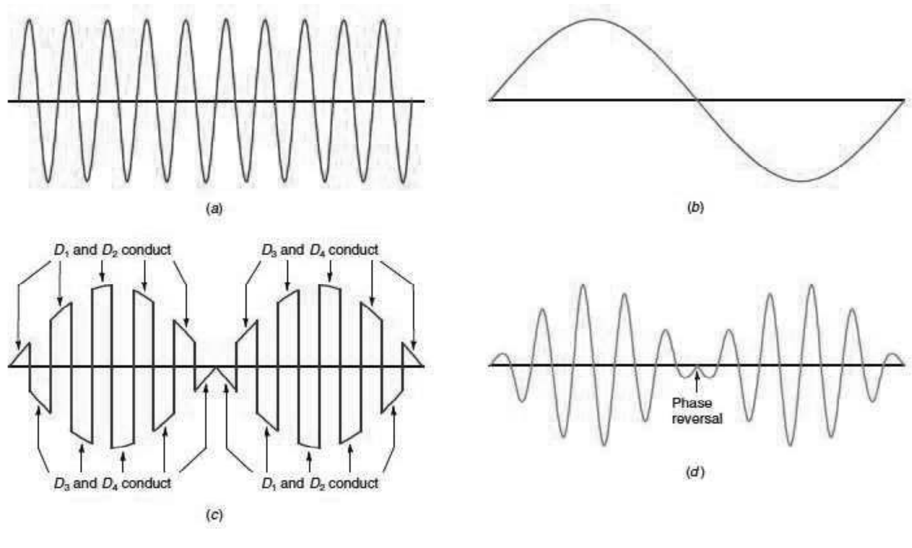


Fig 2.23 *Waveforms in lattice-type Balanced Modulator (a) Carrier (b) Modulating signal (c) DSB-SC at primary T2 (d) DSB-SC waveform*

NOTES

Generation of SSB

A most important characteristic of any SSB generator is the amount of out-of-band energy it produces, relative to the wanted output. In most cases this is determined by the degree to which the unwanted sideband is suppressed 3. A ratio of wanted to-unwanted output power of 40 dB was once considered acceptable commercial performance; but current practice is likely to call for a suppression of 60 dB or more, which is not a trivial result to achieve.

The Filter Method

For generating SSB signal the resulting output of balance modulator is processed by filter/phase-shifting circuits to eliminate one of the sidebands. Filter method is a simplest and most widely used method for generating SSB signals. Figure shows a general block diagram of an SSB transmitter of the filter method. The modulating signal, which may be a voice single sensed from a microphone, is applied to the audio amplifier and the amplified output is fed to one of the input of a balanced modulator. A high frequency carrier signal which is generated by crystal oscillator connected to the other one input of balanced modulator. The balanced modulator generate *double-sideband single without carrier*, hence the output of the balanced modulator is a *double-sideband (DSB)* signal, which consist both upper and lower side bands. The output of balanced modulator is passes through highly selective and sensitive band pass filter for selected either one of the sideband from double side band single for generate single side band modulated signal

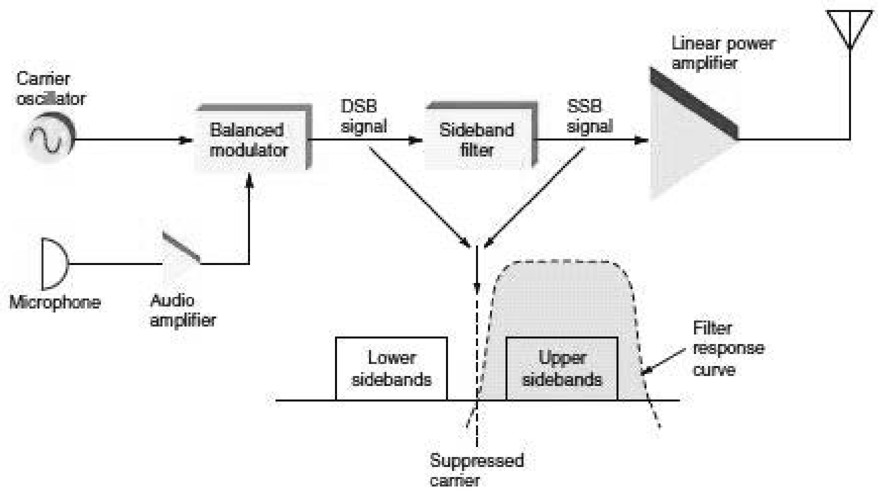


Fig 2.24 *An SSB transmitter using the filter method*

In filter method primary requirement of the circuit is a filter which can pass only the desired sideband. For passing standard voice frequencies from filters, there are of many filters designed for bandwidth of approximately 2.5 to 3 kHz. The frequency of carrier oscillator must be chosen appropriate for the fixed tuned filter so that the sidebands fall within the frequency range of band pass filter. However, the selection of the upper or lower sideband as a standard varies from manufacturer and it is necessary to know which has been used to properly receive an SSB signal.

NOTES

There are two methods of sideband selection.

1. **Two filters: -** Some modulators simply contain two filters for selecting either of one sideband. One of that filters will pass the upper sideband and another that will pass the lower sideband, and for the selection of desirable filter is switch can be employed as shown in figure.
2. **Two carrier frequencies: -** This is an alternative method is for providing two carrier frequencies to the balance modulator. For this purpose two crystal oscillator are used. The selection of the frequency from carrier oscillator is force to either the upper sideband or the lower sideband to appear in the band pass filter.

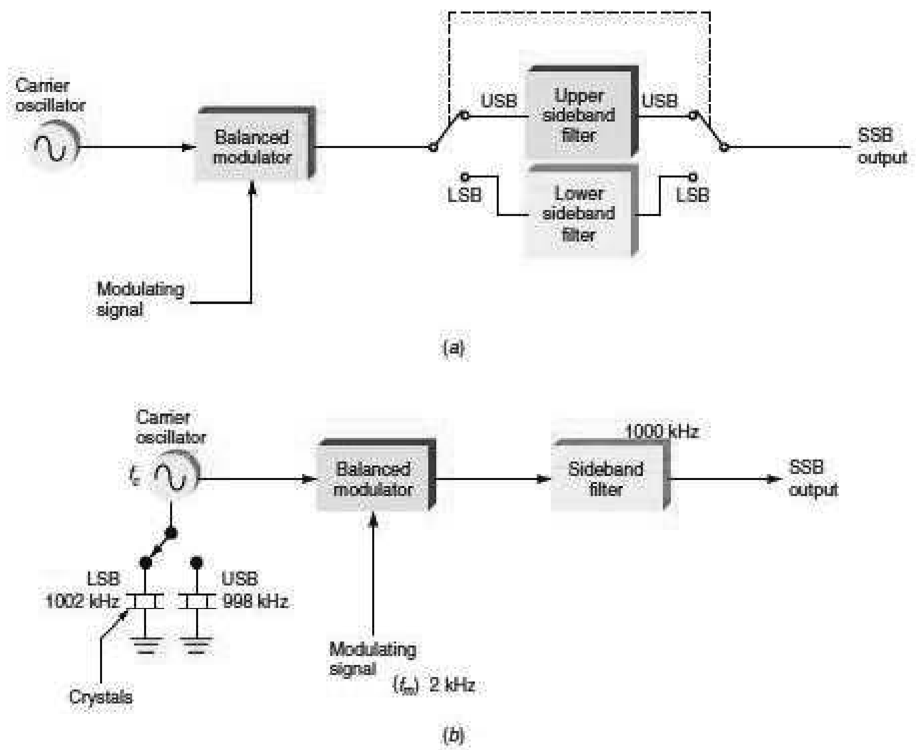


***Fig** 2.25 Methods for selecting the upper or lower sideband. (a) Two filters. (b) Two carrier frequencies.*

Phasing method

Another method for the SSB generation is phasing method. This method uses a phase-shift technique, which eliminate one of the sidebands from the double side band single and generate a single side band (SSB) single. A systematic block diagram of a phasing-type SSB generator is shown in Figure. There are two balanced modulator are used in the method, one is called upper balanced modulator (balanced modulator 1) and another one is called lower balanced modulator (balanced modulator2). The high frequency carrier signal and modulating single is applied directly to the upper balanced modulator in a phase and at the lower balanced modulator, the carrier and modulating signal is in 90° phase of actual phase. This phase difference between both balanced modulator actions causes the elimination of one sideband of the product of both balanced modulator outputs.

Suppose a carrier signal is $V_c \sin 2\pi f_c t$, and the modulating signal is $V_m sin 2\pi f_m t$, then the result of balanced modulator 1 is the product of these two signals:

$$(V_m sin 2\pi f_m t).\ (V_c \sin 2\pi f_c t)$$

Applying a common trigonometric

sin A sin B= 0.5[cos (A - B) - cos (A + B)]

$(V_m sin 2\pi f_m t)\quad (V_c \sin 2\pi f_c t) = 0.5 V_m V_c\ [\cos(2\pi f_c - 2\pi f_m)t - \cos(2\pi f_c + 2\pi f_m)t]$

$(2\pi f_c + 2\pi f_m)$ and $(2\pi f_c - 2\pi f_m)$ are the sum and difference frequencies or the upper and lower sidebands.

The 90° phase shifters create cosine waves of the carrier and modulating signals that are multiplied in balanced modulator 2 to produce

$(V_m cos 2\pi f_m t).\ (V_c \cos 2\pi f_c t).$

Applying a common trigonometric identity

sin A sin B= 5 0.5[cos (A - B) - cos (A + B)]

$(V_m sin 2\pi f_m t)\ (V_c \sin 2\pi f_c t) = 0.5\ V_m V_c [\cos(2\pi f_c - 2\pi f_m)t + \cos(2\pi f_c + 2\pi f_m)t]$

Now, When we add the sine expression given previously to the cosine expression just above, then the sum frequencies from both expressions are cancel each other and the difference frequencies from both expressions are added each other and producing only the lower sideband $[\cos(2\pi f_c + 2\pi f_m)t]$.

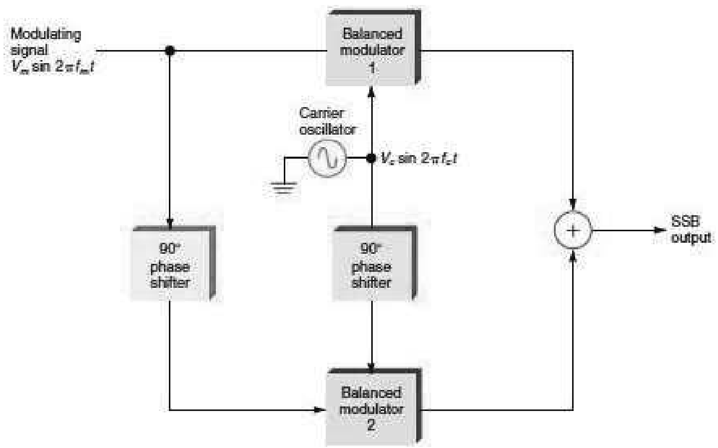


Fig 2.26 *An SSB generator using the phasing method.*

Detection of DSB-SC

Detector is the process of extracting an original message signal from DSB-SC signal. There are following demodulators (detectors) devices are used for demodulating DSB-SC signals at the received end.

- Coherent Detector/ synchronous detection
- Costas Loop

NOTES

Coherent Detector

When the same carrier signal (which is used for generating DSB-SC signal) is employed to detect the message signal (Original signal) is called **coherent** or **synchronous detection**.

At the received, same carrier frequency is added with DSB-SC for converting it into AM single and then and then demodulate it. Figure show the block diagram of the coherent detector. In this diagram, the local carrier frequency which having the same frequency and the phase of the carrier used in DSB-SC modulation, is multiplied with received DSB-SC signal for extracted the original message signal. The resulting signal is passes through a Low Pass Filter for getting desired message signal.

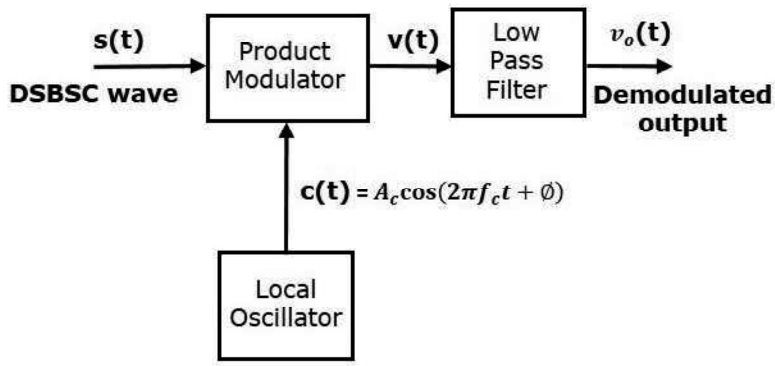


Fig 2.27 *Coherent Detector*

Let we have the DSB-SC single

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

The output of the local oscillator is

$$c(t)=A_c\cos(2\pi f_c t+\phi)$$

ϕ = phase difference between the local oscillator signal and the carrier signal.

Now, the output of product modulator as

$$v(t)=s(t)c(t)$$

Substitute, s(t) and c(t) values in the above equation.

$$v(t) = A_c \cos(2\pi f_c t)\, m(t)\, A_c\cos(2\pi f_c t+\phi)$$

$$v(t) = A_c^2 \cos(2\pi f_c t) \cos(2\pi f_c t+\phi)\, m(t)$$

$$v(t) = (A_c^2/2)\, [\cos(4\pi f_c t+\phi)+\cos\phi]m(t)$$

$$v(t)= (A_c^2/2)\cos \phi m(t) + (A_c^2/2) \cos(4\pi f_c t+\phi)m(t)$$

from the above equation, the first term is the scaled version of the message signal. It can be filtered by passing through a low pass filter. Therefore, the output of low pass filter is

$$v(t)= (A_c^2/2)\cos \phi m(t)$$

Costas Loop

Another method of DSB-SC detection is Costas loop. In the method both the carrier signal (used for DSBSC modulation) and the locally generated signal are in same phase. The block diagram of the Costas look method is shows in figure.

This method consists two product modulators, which have DSBSC signal as common input s(t). The carrier for both balanced modulator are in different in phase for each other and generated by Voltage Controlled Oscillator (VCO).

NOTES

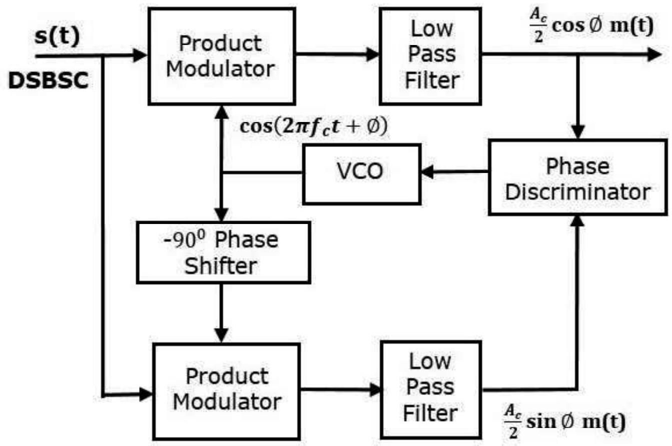


Fig 2.28 *Costas Loop*

The equation of DSB-SC signal-

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

Suppose the output of VCO be-

$$c_1(t) = \cos(2\pi f_c t+\phi)$$

The output of the upper product modulator-

$$v_1(t)=s(t)c_1(t)$$

Substitute the values of s(t) and $c_1(t)$ in the above equation.

$$v_1(t)= A_c \cos(2\pi f_c t)\ m(t)\ \cos(2\pi f_c t+\phi)$$

After simplification, the output of upper balanced modulator is -

$$v_1(t)= (A_c^2/2)\cos\phi m(t) + (A_c^2/2)\cos(4\pi f_c t+\phi)m(t)$$

After the elimination of higher frequencies, the output of the low pass filter is

$$v_1(t)= (A_c^2/2)\cos\phi m(t)$$

This signal is applied as the carrier input of the lower product modulator. The output of the lower product modulator is

$$v_2(t)=s(t)c_2(t)$$

Substituting the value of s(t) and $c_2(t)$ in the above equation.

$$v_2(t)=A_c \cos(2\pi f_c t)\ m(t).\sin(2\pi f_c t+\phi)$$

After simplification, the output of lower balanced modulator is-

$$v_2(t)= (A_c^2/2)\sin\phi m(t) + (A_c^2/2)\sin(4\pi f_c t+\phi)m(t)$$

After the elimination of higher frequencies, the output of the low pass filter is

$$v_2(t)= (A_c^2/2)\sin\phi m(t)$$

The value of $v_1(t)$ and $v_2(t)$ has -90^O phase difference, it means the result of both low pass filters has phase shift. These both single are further applied as inputs of the phase discriminator. The phase discriminator produces a DC control signal

NOTES

for correcting the phase of carrier frequency generated by VCO according to the phase difference between the signals of both filters ,so that the carrier signal (used for DSBSC modulation) and the locally generated signal (VCO output) can be in same phase.

Detection of SSB Signal

Recovering the original message signal from received SSBSC single is known as detection or demodulation of SSB-SC. Coherent detector is one of the most used processes for demodulating the SSB-SC single.

Coherent Detector

When the same carrier signal (which is used for generating SSB-SC signal) is employed to recover the message signal (Original signal) is called **coherent** or **synchronous detection**. At the received, same carrier frequency is added with SSB-SC for converting it into an original signal.

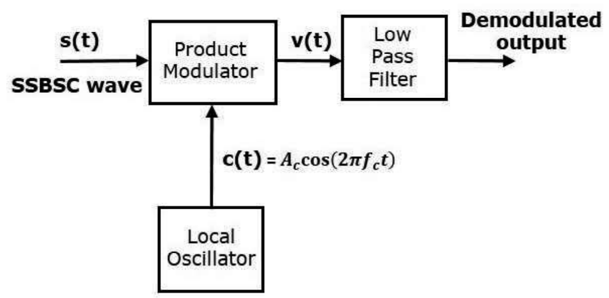


Fig 2.29 *Coherent Detector*

In this process, we use a local carrier, which having the same frequency and the phase of the carrier used in SSBSC modulation, is multiply with received SSB-SC modulated signal using product modulator and the result of this passes by a Low Pass Filter for recovering out low frequency message signal.

Suppose we have SSBSC signal which having a lower sideband.

$$s(t) = (A_m A_c/2)\cos[2\pi(f_c - f_m)t]$$

The single from the local oscillator is

$$c(t) = A_c\cos(2\pi f_c t)$$

The output of product modulator -

$$v(t) = s(t)c(t)$$

Substitute the value of s(t) and c(t) in the above equation.

$$v(t) = (A_m A_c/2)\cos[2\pi(f_c - f_m)t].\ A_c\cos(2\pi f_c t)$$

$$v(t) = (A_m A_c^2/2)\cos[2\pi(f_c - f_m)t].\cos(2\pi f_c t)$$

$$v(t) = (A_m A_c^2/4)\{\cos[2\pi(2f_c - f_m)t] + \cos(2\pi f_m t)\}$$

$$v(t) = (A_m A_c^2/4)\cos(2\pi f_m t) + (A_m A_c^2/2)\{\cos[2\pi(2f_c - f_m)t]$$

From the above equation, the first term is the message signal and It can be recovered by passing through a low pass filter. Therefore, the result of low pass filter is

$$v(t) = (A_m A_c^2/4)\cos(2\pi f_m t)$$

Bandpass Filtering

One method of producing an SSB signal is to remove one of the sidebands via filtering, leaving only either the Upper SideBand (USB), the sideband with the higher frequency, or less commonly the Lower SideBand (LSB), the sideband with the lower frequency. Most often, the carrier is reduced or removed entirely (suppressed), being referred to in full as Single SideBand - Suppressed Carrier (SSB-SC). Assuming both sidebands are symmetric, which is the case for a normal AM signal, no information is lost in the process. Since the final RF amplification is now concentrated in a single sideband, the effective power output is greater than in normal AM, the carrier and redundant sideband account for well over half of the power output of an AM transmitter. Though SSB uses substantially less bandwidth and power, it cannot be demodulated by a simple envelope detector like standard AM.

Hartley Modulator

An alternate method of generation known as a Hartley modulator, named after R. V. L. Hartley, uses phasing to suppress the unwanted sideband. To generate an SSB signal with this method, two versions of the original signal are generated, mutually 90° out of phase for any single frequency within the operating bandwidth. Each one of these signals then modulates carrier waves (of one frequency) that are also 90° out of phase with each other. By either adding or subtracting the resulting signals, a lower or upper sideband signal results. A benefit of this approach is to allow an analytical expression for SSB signals, which can be used to understand effects such as synchronous detection of SSB.

Shifting the baseband signal 90° out of phase cannot be done simply by delaying it, as it contains a large range of frequencies. In analog circuits, a wideband 90-degree phase-difference network is used. The method was popular in the days of vacuum tube radios, but later gained a bad reputation due to poorly adjusted commercial implementations. Modulation using this method is again gaining popularity in the homebrew and DSP fields. This method, utilizing the Hilbert transform to phase shift the baseband audio, can be done at low cost with digital circuitry.

Phasing Generator

The phasing method of SSB generation is based on the addition of two DSBSC signals, so phased that their upper sidebands are identical in phase and amplitude, whilst their lower sidebands are of similar amplitude but opposite phase. The two out-of-phase sidebands will cancel if added; alternatively, the in-phase sidebands will cancel if subtracted. The principle of the SSB phasing generator in illustrated in Figure 2.30. Notice that there are two 90^o phase changers. One operates at carrier frequency, the other at message frequencies. The carrier phase changer operates at a single, fixed frequency, ω rad/s. The message is shown as a single tone at frequency μ rad/s. But this can lie anywhere within the frequency range of speech, which covers several octaves. A network providing a constant 90^o phase shift over this frequency range is very difficult to design. This would be a wideband phase shifter, or Hilbert transformer.

NOTES

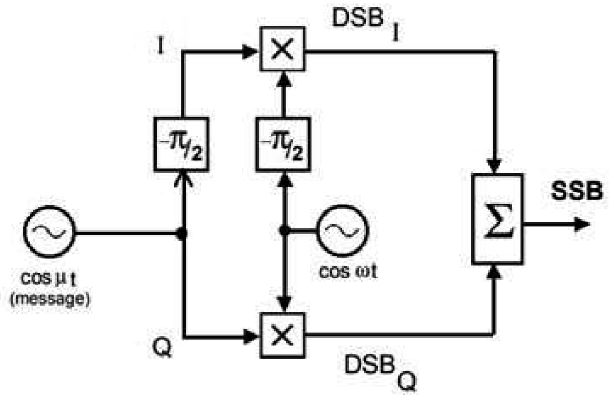


Fig. 2.30 *Principle of the SSB Phasing Generator*

Practically, a wideband phase splitter is used. This is shown in the arrangement of Figure 2.31.

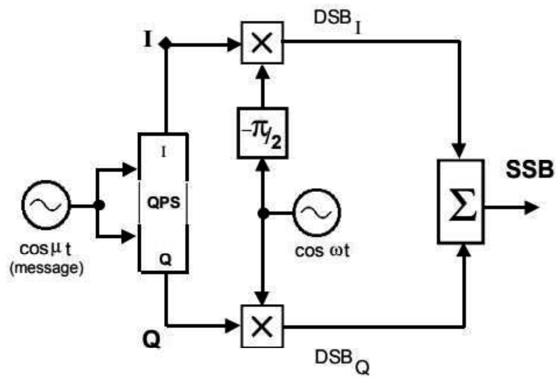


Fig. 2.31 *Practical Realization of the SSB Phasing Generator*

The wideband phase splitter consists of two complementary networks, namely **I** (Inphase) and **Q** (Quadrature). When each network is fed from the same input signal the phase difference between the two outputs is maintained at 90°. Note that the phase difference between the common input and either of the outputs is not specified; it is not independent of frequency.

At the single frequency μ rad/s the arrangements of Figure 2.30 and Figure 2.31 will generate two DSBSC. These are of such relative phases as to achieve the cancellation of one sideband, and the reinforcement of the other, at the summing output.

2.3.9 Independent SideBand (ISB) System

Independent SideBand (ISB) is an AM single sideband mode which is used with some AM radio transmissions. Normally each sideband carries identical information, but ISB modulates two different input signals — one on the Upper SideBand (USB), the other on the Lower SideBand (LSB). This is used in some kinds of AM stereo, sometimes known as the Kahn system.

ISB is a compromise between Double SideBand (DSB) and Single SideBand (SSB) — the other is Vestigial SideBand (VSB). If the sidebands are out of phase with each other, then Phase Modulation (PM) of the carrier occurs. AM and PM together then create Quadrature Amplitude Modulation (QAM). ISB may or may not have the carrier suppressed.

Suppressed Carrier ISB was employed in point-to-point (usually overseas) radiotelephony and radioteletype by shortwave (HF). In military use, ISB usually referred to a close pair of FSK (Frequency Shift Keying) radioteletype channels which could be demodulated by a single receiver, and employed in fleet broadcast, point-to-point, and between larger vessels and shore stations on HF (High Frequency) and UHF (Ultra High Frequency).

Frequency Shift Keying (FSK) is a frequency modulation scheme in which digital information is transmitted through discrete frequency changes of a carrier signal.

2.3.10 Vestigial SideBand (VSB) Modulation

In case of SSB modulation, when a sideband is passed through the filters, the band pass filter may not work perfectly in practice. As a result of which, some of the information may get lost. Hence to avoid this loss, a technique is chosen, which is a compromise between DSB-SC and SSB, called as **Vestigial SideBand (VSB)** technique. The word vestige which means 'A Part' from which the name is derived.

Both sidebands are not required for the transmission, as it is a waste. But a single band if transmitted, leads to loss of information. Hence, this technique has evolved.

Vestigial SideBand Modulation or VSB Modulation is the process where a part of the signal called as vestige is modulated, along with one sideband. A VSB signal can be plotted as shown in the following Figure 2.32.

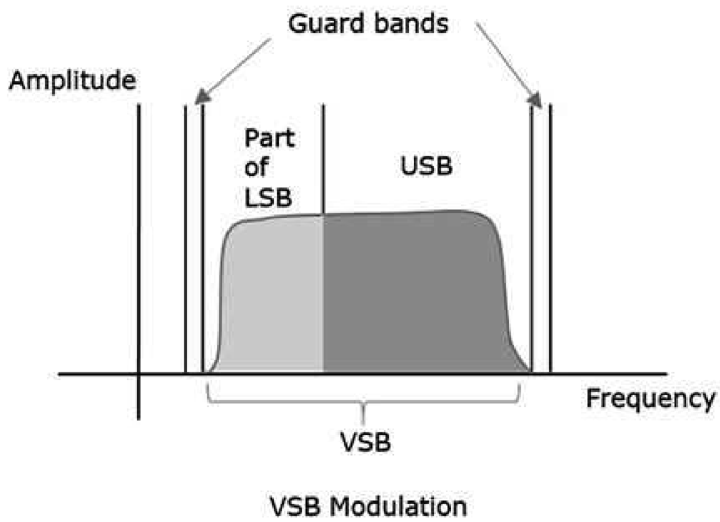


Fig. 2.32 *Vestigial SideBand or VSB Modulation*

Along with the Upper SideBand (USB), a part of the Lower SideBand (LSB) is also being transmitted in this technique. A guard band of very small width is laid on either side of VSB in order to avoid the interferences. VSB modulation is mostly used in television transmissions.

NOTES

Transmission Bandwidth

The transmission bandwidth of VSB modulated wave is represented as,

$$B=(f_{m}+ f_{v})\ Hz$$

Where,

f_m = Message Bandwidth

f_v = Width of the Vestigial SideBand (VSB)

The most prominent and standard application of VSB is for the transmission of television signals. Also, this is the most convenient and efficient technique when bandwidth usage is considered.

Limitation of Single SideBand (SSB) modulation being used for voice signals and not available for video/TV signals leads to the usage of Vestigial SideBand (VSB). In radio communication, a Vestigial SideBand (VSB) is a sideband that has been only partly cut off or suppressed. Television broadcasts (in analog video formats) use this method if the video is transmitted in AM, due to the large bandwidth used. It may also be used in digital transmission, such as the ATSC standardized 8VSB.

The broadcast or transport channel for TV in countries that use NTSC or ATSC has a bandwidth of 6 MHz. To conserve bandwidth, SSB would be desirable, but the video signal has significant low frequency content (average brightness) and has rectangular synchronising pulses. The engineering compromise is Vestigial SideBand (VSB) transmission. In Vestigial SideBand (VSB), the full Upper SideBand (USB) of bandwidth W2 = 4.0 MHz is transmitted, but only W1 = 0.75 MHz of the Lower SideBand (LSB) is transmitted, along with a carrier. The carrier frequency is 1.25 MHz above the lower edge of the 6MHz wide channel. This effectively makes the system AM at low modulation frequencies and SSB at high modulation frequencies. The absence of the lower sideband components at high frequencies must be compensated for, and this is done in the IF amplifier.

Check Your Progress

10. What is demodulation?
11. Define the ways of demodulation.
12. What is synchronous detector?
13. State about the suppressed carrier transmission.
14. Define envelop detector.
15. What is simplest form of envelope detector?
16. Why square law detector is used?
17. What is sideband?
18. State the phasing method of SSB generation.
19. Define the term Independent SideBand (ISB).
20. What is Vestigial SideBand (VSB) technique?

2.4 ANSWERS TO 'CHECK YOUR PROGRESS'

NOTES

1. Amplitude Modulation (AM) involves the modulation of the amplitude of the carrier as analog sine wave. It occurs when a signal to be modulated is applied to a carrier frequency. The carrier frequency may be a radio wave or light wave. The amplitude of carrier frequency changes in accordance with the modulated signal, while the frequency of carrier does not change, and we get a complex wave.
2. The carrier frequency may be suppressed or transmitted at a relatively low level. This requires that the carrier frequency be generated, or otherwise derived, at the receiving site for demodulation. This type of transmission is known as Double SideBand - Suppressed Carrier (DSB-SC).
3. It is also possible to transmit a Single SideBand (SSB). The advantage is a reduction in analog bandwidth needed to transmit the signal. This type of modulation is known as Single SideBand-Suppressed Carrier (SSB-SC) and is ideal for Frequency Division Multiplexing (FDM).
4. The number of signal changes transmitted per unit of time is called the data rate of the modem. That rate is usually expressed in terms of a unit known as a baud. The baud is the number of times per second the line condition can switch from "1" to "0".

   C Shannon theorem states that the maximum capacity (bit rate) of a bandwidth limited transmission line with limited signal to noise ratio is given by C=W*log(1+S/N). Where C is the maximum capacity, W is the limited bandwidth and S/N is the power of the Signal-to-Noise ratio (S/N ratio).
5. Low level modulation is the technique, in which AM waves are generated in the initial stage of amplification, i.e., at a low power level. The generated AM signal is then amplified with the help of a number of amplifier stages.

   High level modulation is the technique, in which amplitude modulation takes place in the final stage of amplification and therefore modulation circuitry has to handle high power. For example, if the transmitter power is 1500 W and the modulation index is 1 then modulation power is 500 W. (33 per cent of transmitter power). The modulator circuitry must be able to deliver such a high power.
6. Modulation is required because there are several limitations associated with baseband transmission that can be overcome using modulation. There is a translation of the baseband signal during the modulation process which converts the low frequency signal to a high frequency signal. The shift in frequency is in proportion to the carrier's frequency.
7. The standard AM's carrier provides an amplitude reference which is lost in single or Double SideBand - Suppressed Carrier (DSB-SC) transmission. In the receiver, the Automatic Gain Control (AGC) responds to the carrier so that the reproduced audio level stays in a fixed proportion to the original modulation. Furthermore, when there is suppressed carrier transmission, there will not be any transmitted power when there is a pause in the

NOTES

modulation, so the AGC must respond to peaks of the transmitted power during peaks in the modulation. Such a situation will typically have what is referred to as fast attack, slow decay circuit that holds the AGC level for a second or more following peaks of this type, in between the program's short pauses or syllables.

8. The modulation index is indicative of the amount by which the modulated variable varies around its unmodulated level. It relates to carrier frequency variations:

$$h = \frac{\Delta f}{f_m} = \frac{f_\Delta \, | x_m(t) |}{f_m}$$

Where,

- f_m is highest frequency component present in the modulating signal $x_m(t)$
- Δf is the peak frequency deviation

In case of sine wave modulation, the modulation index is the ratio of the carrier wave's peak frequency deviation to the modulating sine wave's frequency.

9. Modulation is the overlaying of information or the signal onto an electronic or optical carrier waveform. If the message signal contains single frequency component, then the resulting modulating signal is called as single tone modulated signal. But if the message signal contains more than one frequency component, then the resulting modulated signal is termed as multitone modulated signal, therefore, the multitone modulation contains message signals which has more than one frequency component.

10. 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, therefore, 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).

11. There are several ways of demodulation depending on how parameters of the base-band signal, such as amplitude, frequency or phase are transmitted in the carrier signal. For example, for a signal modulated with a linear modulation like AM (Amplitude Modulation), we can use a synchronous detector. On the other hand, for a signal modulated with an angular modulation, we must use an FM (Frequency Modulation) demodulator or a PM (Phase Modulation) demodulator. Different kinds of circuits perform these functions.

12. A synchronous detector is a device that recovers information from a modulated signal by mixing the signal with a replica of the un-modulated carrier. This can be locally generated at the receiver using a phase-locked loop or other techniques. Synchronous detection preserves any phase

NOTES

information originally present in the modulating signal. Synchronous detection is a necessary component of any analog color television receiver, where it allows recovery of the phase information that conveys hue.

13. Suppressed carrier transmission is a special case in which the carrier level is reduced below that required for demodulation by a normal receiver. The carrier voltage $E_c \cos\omega_c t$ contains no information. The information is only contained in each of the two sidebands. Accordingly, carrier may be suppressed or eliminated. Further, the carrier takes up large portion of the total modulated power $\left(P_c = \frac{2}{3}P_t\right)$, therefore by suppressing the carrier, the power is saved. If the carrier is eliminated or suppressed, the system becomes Suppressed Carrier Double SideBand (DSB-SC) system. In this case, the carrier is reinserted in the radio receiver and the circuit becomes complicated as well as costly.
14. An envelope detector, sometimes called a peak detector, is an electronic circuit that takes a relatively high frequency amplitude modulated signal as input and provides an output, which is the demodulated envelope of the original signal.
15. The simplest form of envelope detector is the diode detector. A diode detector is simply a diode between the input and output of a circuit, connected to a resistor and capacitor in parallel from the output of the circuit to the ground. If the resistor and capacitor are correctly chosen, the output of this circuit should approximate a voltage-shifted version of the original (baseband) signal. A simple filter can then be applied to filter out the DC component.
16. A square law detector is a device that produces an output proportional to the square of some input. For example, in demodulating radio signals, a semiconductor diode can be used as a square law detector, providing an output current proportional to the square of the amplitude of the input voltage over some range of input amplitudes. A square law detector provides an output directly proportional to the power of the input electrical signal.
17. In radio communications, a sideband is a band of frequencies higher than or lower than the carrier frequency, that are the result of the modulation process. The sidebands carry the information transmitted by the radio signal. The sidebands comprise all the spectral components of the modulated signal except the carrier. The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). All forms of modulation produce sidebands.
18. The phasing method of SSB generation is based on the addition of two DSBSC signals, so phased that their upper sidebands are identical in phase and amplitude, whilst their lower sidebands are of similar amplitude but opposite phase. The two out-of-phase sidebands will cancel if added; alternatively, the in-phase sidebands will cancel if subtracted.

NOTES

19. Independent SideBand (ISB) is an AM single sideband mode which is used with some AM radio transmissions. Normally each sideband carries identical information, but ISB modulates two different input signals — one on the Upper SideBand (USB), the other on the Lower SideBand (LSB). This is used in some kinds of AM stereo, sometimes known as the Kahn system. ISB is a compromise between Double SideBand (DSB) and Single SideBand (SSB).
20. In case of SSB modulation, when a sideband is passed through the filters, the band pass filter may not work perfectly in practice. As a result of which, some of the information may get lost. Hence to avoid this loss, a technique is chosen, which is a compromise between DSB-SC and SSB, called as Vestigial SideBand (VSB) technique. The word vestige which means 'A Part' from which the name is derived. Both sidebands are not required for the transmission, as it is a waste. But a single band if transmitted, leads to loss of information. Hence, this technique has evolved. Vestigial SideBand Modulation or VSB Modulation is the process where a part of the signal called as vestige is modulated, along with one sideband.

2.5 SUMMARY

- Amplitude Modulation (AM) involves the modulation of the amplitude of the carrier as analog sine wave. It occurs when a signal to be modulated is applied to a carrier frequency. The carrier frequency may be a radio wave or light wave.
- The amplitude of carrier frequency changes in accordance with the modulated signal, while the frequency of carrier does not change, and we get a complex wave.
- The carrier frequency may be suppressed or transmitted at a relatively low level. This requires that the carrier frequency be generated, or otherwise derived, at the receiving site for demodulation. This type of transmission is known as Double SideBand - Suppressed Carrier (DSB-SC).
- It is also possible to transmit a Single SideBand (SSB). The advantage is a reduction in analog bandwidth needed to transmit the signal. This type of modulation is known as Single SideBand-Suppressed Carrier (SSB-SC) and is ideal for Frequency Division Multiplexing (FDM).
- The Vestigial SideBand (VSB) modulation type of analog modulation is almost like Single SideBand (SSB), except that the carrier frequency is preserved and one of the side bands is eliminated through filtering. Vestigial SideBand (VSB) transmission is usually found in television broadcasting.
- Amplitude modulation is rarely used individually as it is highly sensitive to the impacts of attenuation and line noise.
- Quadrature Amplitude Modulation (QAM) is based on the basic amplitude modulation. It improves the performance of the basic amplitude modulation. In this technique two carrier signals are transmitted simultaneously. The two carrier signals are at the same frequency with a 90 degrees phase shift.

NOTES

- The number of signal changes transmitted per unit of time is called the data rate of the modem. That rate is usually expressed in terms of a unit known as a baud. The baud is the number of times per second the line condition can switch from “1” to “0”.
- Data rate and transmission speed, which is expressed in terms of bits per second, usually are not the same, as several bits may be transmitted through the channel by the modem in each signal change (a few bits can be transmitted as one symbol).
- C Shannon theorem states that the maximum capacity (bit rate) of a bandwidth limited transmission line with limited signal to noise ratio is given by C=W*log(1+S/N). Where C is the maximum capacity, W is the limited bandwidth and S/N is the power of the Signal-to-Noise ratio (S/N ratio).
- Low level modulation: In this technique, AM waves are generated in the initial stage of amplification, i.e., at a low power level. The generated AM signal is then amplified with the help of a number of amplifier stages.
- High level modulation: Under this technique, amplitude modulation takes place in the final stage of amplification and therefore modulation circuitry has to handle high power. For example, if the transmitter power is 1500 W and the modulation index is 1 then modulation power is 500 W. (33 per cent of transmitter power). The modulator circuitry must be able to deliver such a high power.
- In the tuned circuit, each current pulse applied to the tuned circuit initiates a damped oscillation in it. The amplitude of the oscillation is proportional to the size of the current pulse and decay rate is proportional to the time constant of the circuit. Since a series of pulses are fed to the tuned circuit, each pulse will generate a complete sine wave proportional in amplitude to the size of applied pulse. This will be followed by the next sine wave, proportional to the size of the next applied pulse and so on.
- Single SideBand (SSB) modulation, which is a refinement of AM, helps to utilize electric power and bandwidth more efficiently. AM produces a modulated output signal which has twice the bandwidth of the original baseband signal.
- SSB controls this bandwidth doubling, and saves the power wasted on a carrier. Of course, this is achieved at the price of somewhat increased device complexity.
- SSB is also used for long distance telephone transmissions lines, as part of Frequency Division Multiplexing (FDM) technique. SSB technique is used in many applications as it offers many advantages. One of its many advantages is that it allows good quality signals by using very narrow bandwidth with relatively low power over longer distances.
- Modulation is required because there are several limitations associated with baseband transmission that can be overcome using modulation. There is a translation of the baseband signal during the modulation process which converts the low frequency signal to a high frequency signal. The shift in frequency is in proportion to the carrier’s frequency.

NOTES

- In case of transmission of baseband sound signals without modulation by more than one transmitter, the signals will all fall in the same frequency range - 0 to 20 kHz. This will result in all of the signals getting intermixed and leaving the receiver unable to separate one from the other.
- A baseband signal is a low frequency signal. Low frequency signals when transmitted are unable to go long distances getting heavily attenuated.
- Attenuation causes a reduction with increase in frequency of the transmitted signal, enabling the signal to travel longer distances.
- The frequency, of the signal to be transmitted, is increased when there is modulation of the signal. This leads to the communication range being increased.
- The modulation technique referred to as amplitude modulation (AM) is employed in the field of electronic communication, mainly for the purpose of information transmission by using a radio carrier wave.
- For implementation of amplitude modulation, the carrier wave's amplitude (signal strength) needs to be varied proportionate to the waveform which is to be transmitted.
- The standard AM's carrier provides an amplitude reference which is lost in single or Double SideBand - Suppressed Carrier (DSB-SC) transmission.
- In the receiver, the Automatic Gain Control (AGC) responds to the carrier so that the reproduced audio level stays in a fixed proportion to the original modulation.
- When there is suppressed carrier transmission, there will not be any transmitted power when there is a pause in the modulation, so the AGC must respond to peaks of the transmitted power during peaks in the modulation. Such a situation will typically have what is referred to as fast attack, slow decay circuit that holds the AGC level for a second or more following peaks of this type, in between the program's short pauses or syllables.
- The modulation index is indicative of the amount by which the modulated variable varies around its unmodulated level. It relates to carrier frequency variations:

$$h = \frac{\Delta f}{f_m} = \frac{f_\Delta\,|x_m(t)|}{f_m}$$

Where,

f_m is highest frequency component present in the modulating signal $x_m(t)$

Δf is the peak frequency deviation

- In case of sine wave modulation, the modulation index is the ratio of the carrier wave's peak frequency deviation to the modulating sine wave's frequency.
- The carrier component of the modulated wave has the similar amplitude as the unmodulated carrier, i.e., the amplitude of the carrier is unchanged; energy is either added or subtracted.

NOTES

- The modulated wave contains extra energy in the two sideband components. Consequently, the modulated wave contains more power in comparison to the carrier had before modulation happened.
- Because the amplitude of the sidebands depends on the modulation index V_m/V_c, hence it is anticipated that the total power in the modulated wave depends on the modulation index also. The total power in the modulated wave will be,

  $$P_t = \frac{V_{carr}^2}{R} + \frac{V_{LSB}^2}{R} + \frac{V_{USB}^2}{R} \text{ (rms)}$$

  Where all three voltages are (rms) values (√2 converted to peak), and *R* is the resistance, (e.g., antenna resistance), in which the power is dissipated.
- Modulation is the overlaying of information or the signal onto an electronic or optical carrier waveform. If the message signal contains single frequency component, then the resulting modulating signal is called as single tone modulated signal. But if the message signal contains more than one frequency component, then the resulting modulated signal is termed as multitone modulated signal, therefore, the multitone modulation contains message signals which has more than one frequency component.
- Amplitude modulation refers to a process that is used to vary the amplitude of high frequency carrier signal in accordance with the amplitude of the low frequency modulating or information signal, while the carrier signal's phase and the frequency are kept constant.
- In radio communications, a sideband is a band of frequencies higher than or lower than the carrier frequency, those are the result of the modulation process. The sidebands carry the information transmitted by the radio signal.
- The sidebands comprise all the spectral components of the modulated signal except the carrier.
- The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). When the message signal contains more than one frequency, then the corresponding modulation scheme is known as multitone modulation.
- Amplitude Modulation (AM) is a process by which the wave signal is transmitted by modulating the amplitude of the signal. The AM and is commonly used in transmitting a piece of information through a radio carrier wave.
- The modulators generate amplitude modulated waves. The two modulators that generate AM wave are Square Law Modulator and Switching Modulator.
- Switching modulator is similar to the square law modulator. The only difference is that in the square law modulator, the diode is operated in a non-linear mode, whereas, in the switching modulator, the diode must operate as an ideal switch.
- Demodulation is extracting the original information-bearing signal from a carrier wave.

NOTES

- 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, therefore, 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).
- There are several ways of demodulation depending on how parameters of the base-band signal, such as amplitude, frequency or phase are transmitted in the carrier signal. For example, for a signal modulated with a linear modulation like AM (Amplitude Modulation), we can use a synchronous detector.
- For a signal modulated with an angular modulation, we must use an FM (Frequency Modulation) demodulator or a PM (Phase Modulation) demodulator. Different kinds of circuits perform these functions.
- In electronics, a synchronous detector is a device that recovers information from a modulated signal by mixing the signal with a replica of the un-modulated carrier. This can be locally generated at the receiver using a phase-locked loop or other techniques.
- Synchronous detection preserves any phase information originally present in the modulating signal. Synchronous detection is a necessary component of any analog color television receiver, where it allows recovery of the phase information that conveys hue.
- Synchronous detectors are also found in some shortwave radio receivers used for audio signals, where they provide better performance on signals that may be affected by fading. To recover baseband signal, the synchronous detection technique is used.
- For synchronous demodulation, a mixer is used. The incoming signal is fed into the signal input of the mixer, and a local oscillator signal on the same frequency as the carrier of the incoming signal is fed into the other. This mixing process converts the carrier to a 0Hz signal and the sidebands to their base band frequency band, i.e., it reconstitutes the audio.
- The process of demodulation or detection consists in recovering the original modulating voltage from the modulated carrier voltage. The detection is a process reverse of the modulation. The detection process is achieved by mixing the carrier with the side band components carrying the intelligence, in a non-linear device.
- Suppressed carrier transmission is a special case in which the carrier level is reduced below that required for demodulation by a normal receiver. We know that the carrier voltage $E_c \cos\omega_c t$ contains no information. The information is only contained in each of the two sidebands. Accordingly, carrier may be suppressed or eliminated. Further, the carrier takes up large portion of the total modulated power $\left(P_c = \frac{2}{3} P_t\right)$, therefore by suppressing the carrier, the power is saved.

- In ordinary AM radio broadcast, carrier is allowed along with the two side bands and the system is referred as the double side band system.
- Carrier is necessary for the reproduction of modulating signal in the detector of radio receiver. Accordingly, in ordinary AM broadcast, carrier is allowed to propagate.
- If the carrier is eliminated or suppressed, the system becomes Suppressed Carrier Double SideBand (DSB-SC) system. In this case, the carrier is reinserted in the radio receiver and the circuit becomes complicated as well as costly. Hence, DSB-SC may be used in point-to-point communication. The circuit of suppressed carrier balanced modulator using transistors.
- An envelope detector, sometimes called a peak detector, is an electronic circuit that takes a relatively high frequency amplitude modulated signal as input and provides an output, which is the demodulated envelope of the original signal.
- The simplest form of envelope detector is the diode detector. A diode detector is simply a diode between the input and output of a circuit, connected to a resistor and capacitor in parallel from the output of the circuit to the ground. If the resistor and capacitor are correctly chosen, the output of this circuit should approximate a voltage-shifted version of the original (baseband) signal. A simple filter can then be applied to filter out the DC component.
- The process of separating the original information or 'Signal' from the 'Modulated Carrier'. In the case of 'Amplitude or Frequency Modulation' it involves a device, called a Demodulator or Detector, which produces a signal corresponding to the instantaneous changes in amplitude or frequency, respectively. This signal corresponds to the original modulating signal.
- A square law detector provides an output directly proportional to the power of the input electrical signal. In radio transmission, this process is a major function of a receiver, in order to retrieve the desired signal.
- Double-SideBand Suppressed Carrier (DSB-SC) transmission is transmission in which frequencies produced by Amplitude Modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level, ideally being completely suppressed.
- The name 'Suppressed Carrier' comes about because the carrier signal component is suppressed—it does not appear in the output signal.
- DSB-SC transmission is a special case of double-sideband reduced carrier transmission. It is used for radio data systems. This mode is frequently used in Amateur radio voice communications, especially on High Frequency (HF) bands.
- In radio communications, Single SideBand (SSB) modulation 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

NOTES

the bandwidth of which is twice the maximum frequency of the original baseband signal.

- In radio communications, a sideband is a band of frequencies higher than or lower than the carrier frequency, that are the result of the modulation process. The sidebands carry the information transmitted by the radio signal.
- The sidebands comprise all the spectral components of the modulated signal except the carrier. The signal components above the carrier frequency constitute the Upper SideBand (USB), and those below the carrier frequency constitute the Lower SideBand (LSB). All forms of modulation produce sidebands.
- DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal.
- A most important characteristic of any SSB generator is the amount of out-of-band energy it produces, relative to the wanted output.
- The Hartley modulator, named after R. V. L. Hartley, uses phasing to suppress the unwanted sideband. To generate an SSB signal with this method, two versions of the original signal are generated, mutually 90° out of phase for any single frequency within the operating bandwidth. Each one of these signals then modulates carrier waves (of one frequency) that are also 90° out of phase with each other.
- The phasing method of SSB generation is based on the addition of two DSBSC signals, so phased that their upper sidebands are identical in phase and amplitude, whilst their lower sidebands are of similar amplitude but opposite phase. The two out-of-phase sidebands will cancel if added; alternatively, the in-phase sidebands will cancel if subtracted.
- Independent SideBand (ISB) is an AM single sideband mode which is used with some AM radio transmissions.
- Normally each sideband carries identical information, but ISB modulates two different input signals — one on the Upper SideBand (USB), the other on the Lower SideBand (LSB). This is used in some kinds of AM stereo, sometimes known as the Kahn system.
- ISB is a compromise between Double SideBand (DSB) and Single SideBand (SSB) — the other is Vestigial SideBand (VSB). If the sidebands are out of phase with each other, then Phase Modulation (PM) of the carrier occurs. AM and PM together then create Quadrature Amplitude Modulation (QAM). ISB may or may not have the carrier suppressed.
- In case of SSB modulation, when a sideband is passed through the filters, the band pass filter may not work perfectly in practice. As a result of which, some of the information may get lost. Hence to avoid this loss, a technique is chosen, which is a compromise between DSB-SC and SSB, called as Vestigial SideBand (VSB) technique. The word vestige which means ‘A Part’ from which the name is derived.

- Both sidebands are not required for the transmission, as it is a waste. But a single band if transmitted, leads to loss of information. Hence, this technique has evolved.

NOTES

2.6 KEY TERMS

- **Amplitude Modulation (AM):** Amplitude Modulation (AM) involves the modulation of the amplitude of the carrier as analog sine wave. It occurs when a signal to be modulated is applied to a carrier frequency. The carrier frequency may be a radio wave or light wave.
- **Double SideBand - Suppressed Carrier (DSB-SC):** The carrier frequency may be suppressed or transmitted at a relatively low level. This requires that the carrier frequency be generated, or otherwise derived, at the receiving site for demodulation. This type of transmission is known as Double SideBand - Suppressed Carrier (DSB-SC).
- **SideBand-Suppressed Carrier (SSB-SC):** It is also possible to transmit a Single SideBand (SSB). The advantage is a reduction in analog bandwidth needed to transmit the signal. This type of modulation is known as Single SideBand-Suppressed Carrier (SSB-SC) and is ideal for Frequency Division Multiplexing (FDM).
- **Data rate:** The number of signal changes transmitted per unit of time is called the data rate of the modem. That rate is usually expressed in terms of a unit known as a baud. The baud is the number of times per second the line condition can switch from “1” to “0”.
- **Low level modulation:** In this technique, AM waves are generated in the initial stage of amplification, i.e., at a low power level. The generated AM signal is then amplified with the help of a number of amplifier stages.
- **High level modulation:** Under this technique, amplitude modulation takes place in the final stage of amplification and therefore modulation circuitry has to handle high power.
- **Multitone modulation:** If the message signal contains more than one frequency component, then the resulting modulated signal is termed as multitone modulated signal, therefore, the multitone modulation contains message signals which has more than one frequency component.
- **Demodulation:** Demodulation is extracting the original information-bearing signal from a carrier wave.
- **Demodulator:** 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, therefore, there are many types of demodulators.
- **Envelope detector:** An envelope detector, sometimes called a peak detector, is an electronic circuit that takes a relatively high frequency amplitude modulated signal as input and provides an output, which is the demodulated envelope of the original signal.

2.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. What is Amplitude Modulation (AM)?
2. Why is modulation essential?
3. Define the principle of amplitude modulation.
4. What is modulation index?
5. State about the power relation and multitone modulation.
6. How is the AM wave generated?
7. Define AM square law modulator.
8. Why is switching modulator used?
9. What does demodulation of AM refer to?
10. Define synchronous detection and non-linear demodulation.
11. What is suppress carrier AM demodulator?
12. State about the envelope detector.
13. What is the importance of square law demodulator?
14. Define DSB-SC and SSB modulation systems.
15. Elaborate on sideband and carrier power.
16. What is Independent SideBand (ISB) system?
17. Define Vestigial SideBand (VSB) modulation.

Long-Answer Questions

1. Briefly discuss the characteristic features and significance of Amplitude Modulation (AM).
2. Explain the necessity of modulation giving appropriate examples.
3. Discuss the principle of amplitude modulation with the help of relevant examples.
4. What is modulation index? How is it calculated? Explain giving examples.
5. Brief a note on power relation and multitone modulation giving examples.
6. Discuss the significance of AM wave generation and AM square law modulator with the help of appropriate examples.
7. Explain about switching modulator giving its significance.
8. Discuss the importance of demodulation of AM.
9. Elaborate on synchronous detection and nonlinear demodulation giving relevant examples.
10. Briefly discuss suppress carrier AM demodulator giving examples.
11. Explain in detail envelope detector and square law demodulator.

12. Discuss briefly about the DSB-SC and SSB modulation systems and sideband and carrier power with the help of examples.
13. Briefly discuss the method of generation and detection of DSB-SC and SSB.
14. Explain the characteristic features of Independent SideBand (ISB) system and Vestigial SideBand (VSB) modulation.

NOTES

2.8 FURTHER READING

Gurung, J.B. 2009. *Signals and Systems*. New Delhi: PHI Learning Pvt. Ltd.

Carlson, A Bruce and Paul B. Crilly. 2010. *Communication System*. Noida: Tata McGraw-Hill Education.

Lathi, B. P. and Zhi Ding. 2010. *Modern Digital and Analog Communication Systems*. Oxford: Oxford University Press.

Kennedy, George and Bernard Davis. 1993. *Electronic Communication Systems*. London: Glencoe/McGraw-Hill.

Roddy, Dennis and John G. Coolen. 1977. *Electronic Communications*. Noida: Pearson Education.

Greenwich, Edwin. 2012. *Analog Integrated Circuits*. Berlin: Springer Science & Business Media.

Millman, Jacob and Christos C. Halkias. 2001. *Integrated Electronics*. Noida: Tata McGraw-Hill Education.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey: Prentice-Hall.

Tanenbaum, Andrew. S. 2002. *Computer Networks*. New Jersey: Prentice-Hall PTR.

Basavaraj, B. and H.N. Shivashankar. 2004. *Basic Electronics*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 3 SAMPLING AND ANALOGUE PULSE MODULATION

NOTES

Structure

3.0 INTRODUCTION

General sound, video or data signals, which are also called baseband signals, are incompatible for over the media. This necessitates modulation, which amounts to impressing the modulating signal (the raw baseband signal) on the high-frequency carrier signal to obtain the modulated Radio-Frequency (RF) signal. This modulated signal is further amplified in the transmitter suitable enough to drive the antenna. At the receiver end, this signal is demodulated to obtain the original sound, video or data signals. There are three types of basic modulation techniques Amplitude Modulation (AM), Frequency Modulation (FM) and Phase Modulation (PM). In this unit, you will learn different modulation techniques involved in the generation of AM and FM waves, receivers, demodulators, and other associated concepts. In this unit, we will study in detail about sampling and analogue pulse modulation.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain the sampling theory, types of sampling and sampling analysis
- Describe the analogue pulse modulation and its types
- Discuss different modulation techniques, their merits and demerits

- Explain pulse amplitude modulation, pulse duration modulation and pulse position modulation
- State the signal to noise ratio in pulsed system

NOTES

3.2 SAMPLING THEORY AND SAMPLING ANALYSIS

A signal has three properties like voltage or amplitude, frequency and phase. Analog signals are continuous in time and have difference in voltage levels for different periods of the signal. Here, its main drawback is that the amplitude keeps on changing along with the period of the signal. This can be overcome by the digital form of signal representation. Here conversion of an analog form of the signal into digital form can be done using the sampling technique

Sampling is defined as "the process of obtaining discrete values for the instantaneous values of a continuous-time signal." A sample is a subset of data drawn from a larger set of continuous data in the time domain. When a source provides an analogue signal and that signal must be converted to a digital signal consisting of 1s and 0s, or High or Low, the signal must be discretized in time. This discretization of analog signal is termed as Sampling. Figure 3.1 epitomizes a continuous-time signal x(t) and a sampled signal $x_s(t)$. When x(t) is multiplied by a periodic impulse train, the sampled signal $x_s(t)$ is achieved.

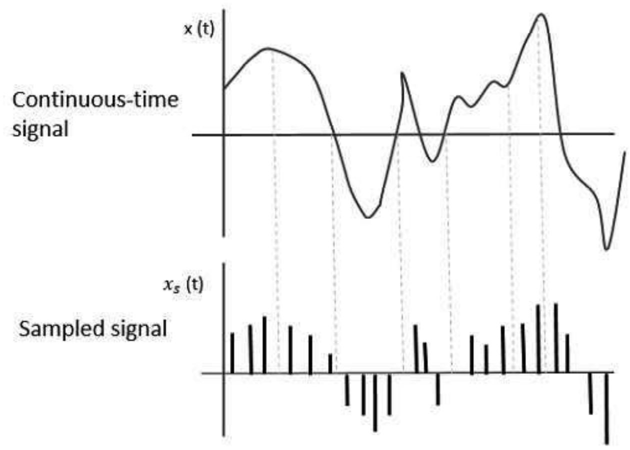


Fig 3.1 *Continuous-time signal x-t and a sampled signal x_s-t*

Sampling Rate

The interval between the samples should be maintained in order to discretize the signals. A sampling period T_s can be used to describe this gap.

Sampling Frequency = $1/T_s = f_s$

where, T_s is the sampling time and f_s is the sampling frequency or the sampling rate. The sampling rate refers to the number of samples obtained per second or for a specific set of values.

The sampling rate should be taken into account when reconstructing an analogue signal from a digitised source. The sampling rate should be such that neither the data in the message signal is lost nor is it over-lapped. As a result, a rate was established for this, known as the Nyquist rate.

Nyquist Rate

Assume a signal is band-limited, without a frequency component exceeding W Hertz. That means, W is the highest frequency. The sampling rate for this signal must be twice the highest frequency for successful reproduction of the original signal, that is,

$$f_S = 2W$$

where, f_S is the sampling rate and W is the highest frequency. Nyquist rate is the name given to this sampling rate. On the principle of this Nyquist rate, a statement known as the Sampling Theorem was formulated.

Sampling Theorem

The sampling theorem, often known as the Nyquist theorem, establishes the notion of sufficient sample rate in terms of bandwidth for band-limited functions. "A signal can be exactly recreated if it is sampled at a rate f_s that is larger than twice the maximum frequency W," according to the sampling theorem.

Consider a band-limited signal, or one whose value is non-zero between some –W and W Hertz, to better comprehend this sampling theorem. Such a signal is represented as:

$$x(f) = 0 \text{ for } |f| > W$$

The band-limited signal in frequency domain for the continuous-time signal x(t) can be visualized as Figure 3.2.

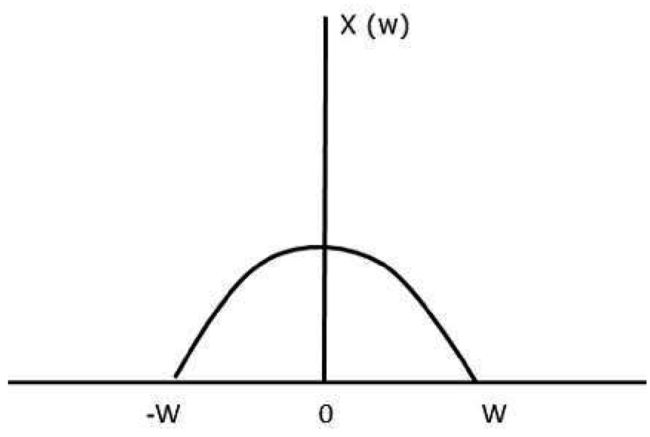


Fig 3.2 *Band Limited Signal*

There is a need for sampling frequency, that is, a frequency where no information is lost after sampling. For this, the Nyquist rate is required. It is the critical rate of sampling. The original signal can be retrieved if the signal x(t) is sampled above the Nyquist rate, but it cannot if it is sampled below the Nyquist rate.

NOTES

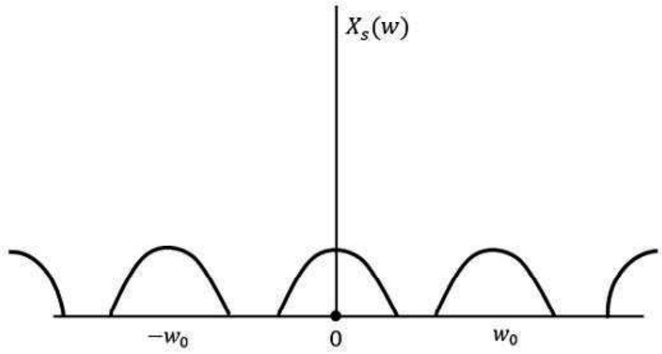


Fig 3.3 *Fourier transform of a signal $x_s(t)$ where $f_s > 2W$*

Figure 3.3 demonstrates the Fourier transform of a signal $x_s(t)$. Here, the information is reproduced without any loss. Because there is no mixing, recovery is feasible. The Fourier Transform of the signal $x_s(t)$ is

$$X_s(w) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(w - nw_0)$$

where T_s = Sampling Period and

$w_0 = 2\pi/Ts$

Let us have a look at what happens if the sampling rate is twice the highest frequency (2W), that means,

$f_s = 2W$

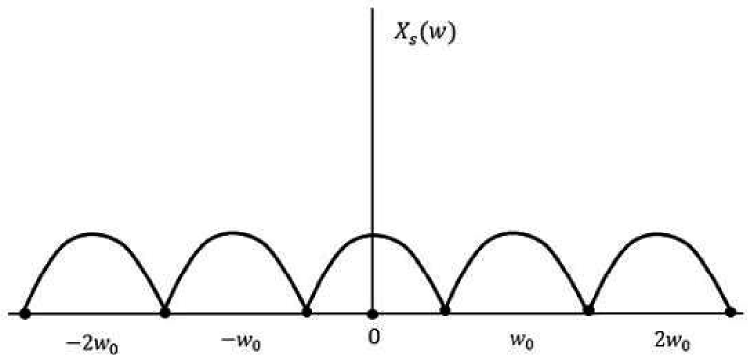


Fig 3.4 *Fourier transform of a signal $x_s(t)$ where $f_s = 2W$*

The result is shown in figure 3.4. The information is replaced without any loss. Hence, this is also a good sampling rate. Now, let us look at the condition,

$$f_s < 2W$$

The resultant pattern will look like figure 3.5. The over-lapping of data may be seen in this pattern, which results in information mixing up and loss. Aliasing is the term for the undesired phenomena of over-lapping.

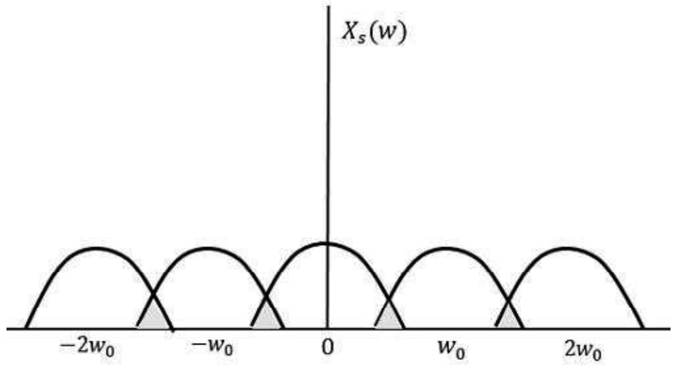


Fig 3.5 *Fourier transform of a signal $x_s(t)$ where $f_s < 2W$*

There are three types of sampling strategies in general:

(i) Instantaneous sampling

(ii) Natural sampling

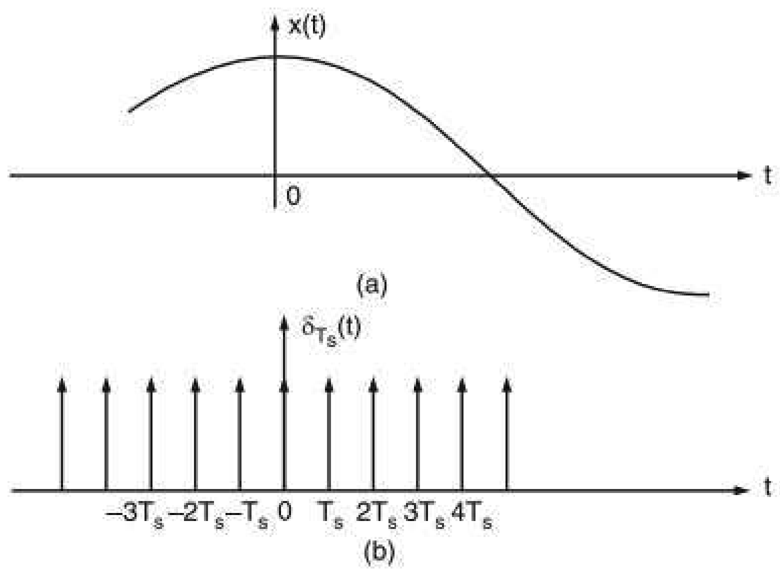
(iii) Flat top sampling

Instantaneous sampling is referred to as ideal sampling, whereas natural sampling and flat-top sampling are referred to as practical sampling methods.

3.2.1 Instantaneous or Ideal Sampling

In this sampling technique, the sampling function is a train of impulses and the principle used is known as multiplication principle. x(t) is the input signal (i.e., signal to be sampled) as shown in Figure 3.6(a). Figure 3.6(b) shows this sampling function. Figure 3.6(c) shows a circuit to produce instantaneous or ideal sampling. This circuit is known as the switching sampler. Figure 3.6(d) represents the sampled signal.

If the circuit just consists of a switch with a closing time 't' that approaches zero, the output g(t) of this circuit will only include the instantaneous value of the input signal x. (t). Because the pulse's width approaches zero, instantaneous sampling produces a train of impulses with a height equal to the input signal's instantaneous value x(t) at the sampling instant.



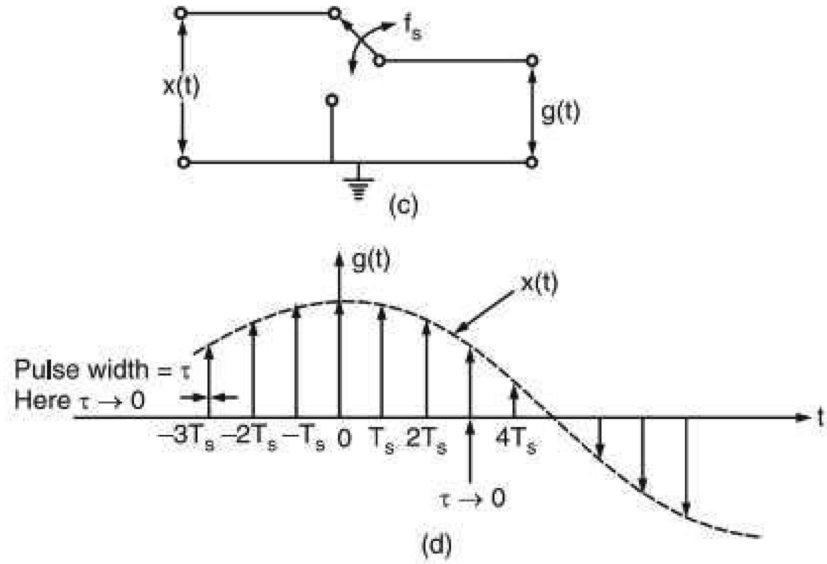


Fig 3.6 *(a) Baseband signal, (b) impulse train, (c) functional diagram of a switching sampler, (d) sampled signal*

The train of impulses may be represented as,

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

The waveform of this is known as the sampling function, and it is represented in Figure 6(b). The sampled signal g(t) is determined by multiplying x(t) and $\delta T_s(t)$. Thus,

$$g(t) = x(t) \cdot \delta_{T_s}(t)$$

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

Or

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

The preceding equation's Fourier transform of the optimally sampled signal can be written as:

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

3.2.2 Natural Sampling

As previously stated, instantaneous sampling produces samples with a width approaching zero, resulting in power content in the immediately sampled pulse that is minimal. As a result, this approach is unsuitable for transmission.

For the period of the sampling pulse, the pulse amplitude takes the shape of the analogue waveform in natural sampling. The sampled output's frequency spectrum differs from that of an ideal sample.

NOTES

Natural sampling is a practical method. In this type of sampling, the pulse has a definite width of τ. Consider the case of an analogue continuous-time signal x(t) sampled at f_s Hertz. It is assumed that f_s is greater than Nyquist rate in order to meet the sampling theorem. Consider a sampling function c(t), which is a train of periodic pulses with a width of τ and frequency f_s Hz. Figure 3.7 shows a functional diagram of a natural sampler.

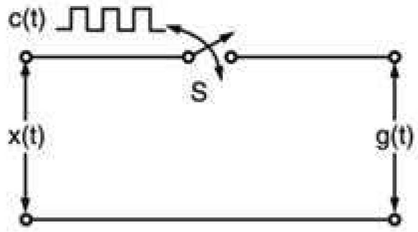


Fig 3.7 *A functional diagram of a natural sampler*

A sampled signal g(t) is generated with this natural sampler by multiplying the sampling function c(t) and the input signal x. (t). Now, when c(t) is high, the switch 'S' is closed, as seen in figure 3.7. Therefore,

g(t) = x(t) when c(t) = A

g(t) = 0 when c(t) = 0

where A is the amplitude of c(t). Figure 3.8(a), (b), and (c) show the waveforms of signals x(t), c(t), and g(t), respectively.

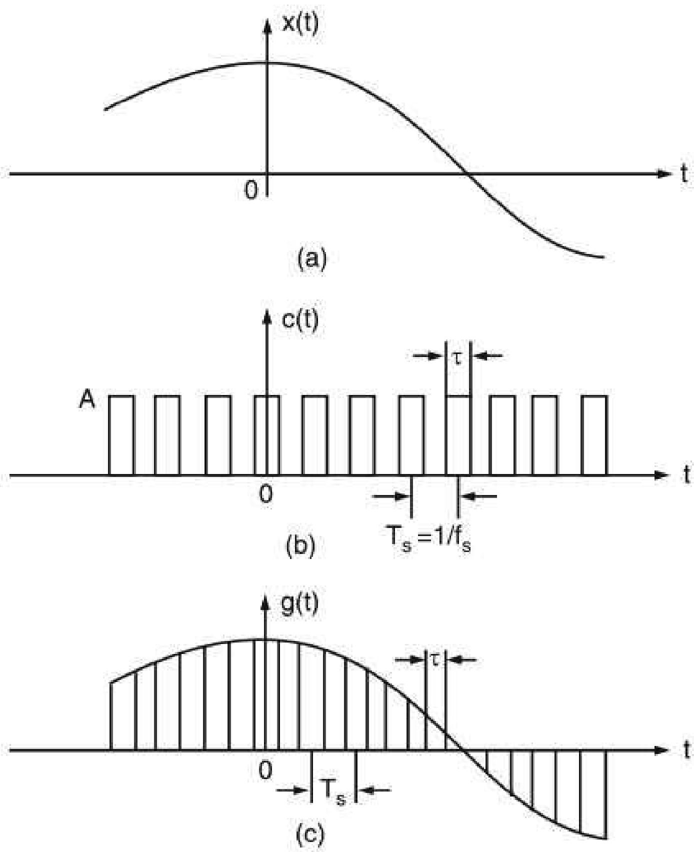


Fig 3.8 *(a) Continuous time signal x(t), (b) Sampling function waveform i.e., periodic pulse train, (c) Naturally sampled signal waveform g(t)*

The sampled signal g(t) can be quantitatively represented as:

$$g(t) = c(t) . x(t)$$

The periodic train of pulses with width t and frequency f_s is denoted by c(t).

NOTES

For each periodic waveform, the exponential Fourier series is written as:

$$c(t) = \sum_{n=-\infty}^{\infty} C_n e^{j2\pi nt/T_0}$$

Also, for the periodic pulse train of c(t),

$$T_0 = T_s = \frac{1}{f_s} = \text{period of } c(t)$$

or $$f_0 = f_s = \frac{1}{T_0} = \frac{1}{T_s} = \text{frequency of } c(t)$$

So,

$$c(t) = \sum_{n=-\infty}^{\infty} C_n . e^{j2\pi f_s nt} \text{ with } \frac{1}{T_0} = f_s$$

It should be noticed that because c(t) is a rectangular pulse train, C_n for this waveform will be stated as:

$$C_n = \frac{TA}{T_0} \sin c\,(f_n.T)$$

Here T = pulse width = τ

And f_n = harmonic frequency

But here, $f_n = nf_s$

Or,

$$f_s = \frac{n}{T_0} = nf_0$$

Hence,

$$C_n = \frac{\tau.A}{T_s} \cdot \sin c(f_{n.}\tau)$$

As a result, the Fourier series representation of c(t) is as follows:

$$c(t) = \sum_{n=-\infty}^{\infty} \frac{\tau.A}{T_s} \cdot \sin c(f_{n.}\tau)\, e^{j2\pi f_s \cdot nt}$$

Now, if we substitute the value of c(t) into the g(t) equation, we get:

$$g(t) = \frac{\tau A}{T_s} \cdot \sum_{n=-\infty}^{\infty} \sin c(f_{n.}\tau) .\, e^{j2\pi f_s nt} .x(t)$$

NOTES

For a naturally sampled signal g(t), this is the required time-domain representation. Let us now take the Fourier transform of the naturally sampled signal g(t) to obtain its frequency-domain representation:

$$G(f) = FT[g(t)]$$

or
$$G(f) = \frac{\tau A}{T_s} \sum_{n=-\infty}^{\infty} \sin c(f_{n.}\tau)\, FT[e^{j2\pi f_s nt}.x(t)]$$

Considering the Fourier transform's frequency-shifting property, which asserts that:

$$e^{j2\pi f_s n t}.x(t) \longleftrightarrow X(f - f_{s.}\, n)$$

Therefore,

$$G(f) = \frac{\tau A}{T_s} \cdot \sum_{n=-\infty}^{\infty} \sin c(f_{n.}\tau)\, X(f - f_s n)$$

Now, since f_n = nfs = harmonic frequency. Therefore,

$$G(f) = \frac{\tau A}{T_s} \cdot \sum_{n=-\infty}^{\infty} \sin c(nf_s.\tau)\, X\,(f - nf_s)$$

This equation demonstrates that the spectrum of x(t), i.e., X(f), is periodic in f_s and is represented by the sine function. The figure 3.9 depicts several arbitrary spectra for x(t) and their associated spectrum G(f).

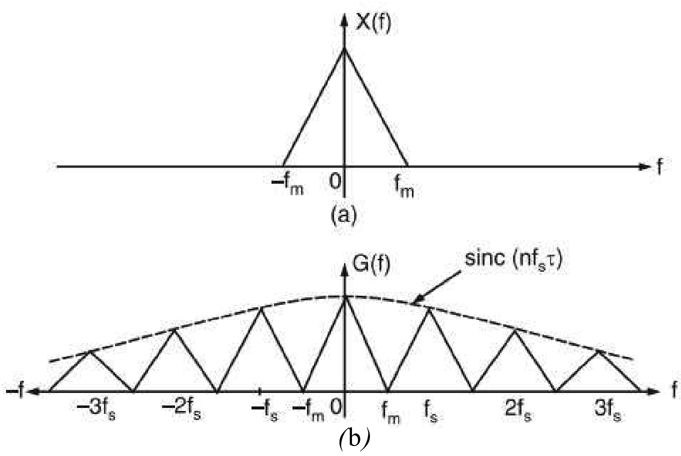


Fig 3.9 *(a) Spectrum of continuous-time signal x(t), (b) Spectrum of naturally sampled signal*

Flat Top or Rectangular Pulse Sampling

As with natural sampling, flat top or rectangular pulse sampling is also a practicable sampling approach. However, natural sampling is a little more complicated than flat top sampling. It is accomplished in a sample-and-hold circuit. Its goal is to sample and convert the continuously changing analogue input voltage to a sequence of steady voltage levels on a periodic basis. The input voltage is sampled using a narrow pulse and then maintained relatively steady until another sample is obtained. Flat-top sampling introduces less aperture distortion than natural sampling and can operate with a slower analog-to-digital converter.

At the commencement of sampling, the top of the samples stays unchanged and is equivalent to the instantaneous value of the baseband signal x(t). The duration or width of each sample is τ. The functional diagram of a sample and hold circuit, which is employed to produce the flat top samples, is shown in Figure 3.10(a).

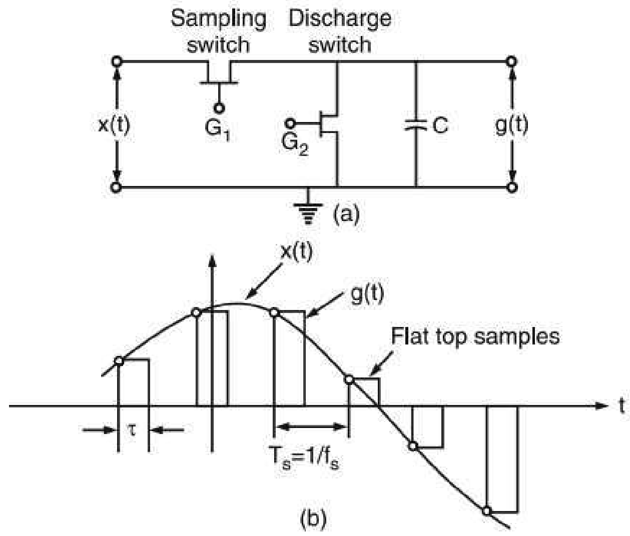


Fig 3.10 *(a) A sample and hold circuit to generate flat top samples*
(b) A general waveform of flat top sampling

The basic waveform of flat top samples is shown in Figure 3.10(b). It should be emphasized that only the pulse's starting edge represents the instantaneous value of the baseband signal x(t). Also, as shown in Figure 3.11, the flat top pulse of g(t) is mathematically identical to the convolution of an instantaneous sample and a pulse h(t).

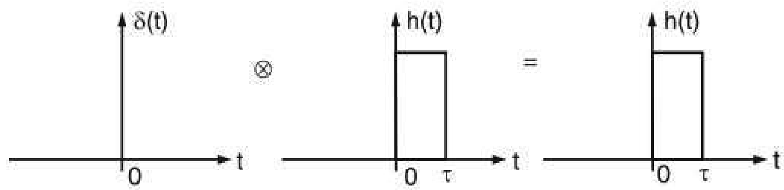


Fig 3.11 *Convolution of any function with delta function equal to that function*

The width of the pulse in g(t) is defined by the width of h(t), whereas the sampling instant is governed by the delta function. The starting edge of the pulse in Figure 3.10(b) depicts the point at which the baseband signal is sampled, and the width is defined by function h(t). Therefore, g(t) will be expressed as:

$$g(t) = s(t) \otimes h(t)$$

This equation has been explained in Figure 3.12.

We can now deduce from the property of the delta function that for any function f (t),

$$f(t) \otimes \delta(t) = f(t)$$

Flat top samples are obtained using this characteristic. It should be noted that the above equation is not used directly to produce flat top sampling; rather, a modified version of the above equation is used. As a result, s(t) is substituted for the delta function $\delta(t)$ in this updated equation. Note that $\delta(t)$ is a constant amplitude

NOTES

delta function, but s(t) is a train of impulses with variable amplitudes. This means that the signal s(t), which is sampled in real time, is convolved with the function h(t). As a result of convoluting s(t) and h(t), a pulse with a duration equal to h(t) but an amplitude defined by s(t) is obtained. The stream of impulses can now be quantitatively described as,

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

The multiplication of baseband signal x(t) and $\delta T_s(t)$ will yield the signal s(t). Thus,

$$s(t) = x(t) \cdot \delta_{T_s}(t)$$

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

Now, sampled signal g(t) is given as

$$g(t) = s(t) \otimes h(t)$$

or $$g(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)\, d\tau$$

or $$g(t) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT_s)\, \delta(\tau - nT_s)h\,(t-\tau)\, d\tau$$

or $$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \int_{-\infty}^{\infty} \delta(\tau - nT_s)h\,(t-\tau)\, d\tau$$

According to shifting property of delta function,

$$\int_{-\infty}^{\infty} f(t)\, \delta(t - t_0) = f(t_0)$$

Hence,

$$g(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\, h(t - nT_s)$$

For a flat top sampled signal, this equation defines the value of g(t) through sampled value $x(nT_s)$ and function $h(t-nT_s)$.

Now, we have

$$g(t) = s(t) \otimes h(t)$$

When both sides of the equation are Fourier transformed, we get

$$G(f) = S(f)\, H(f)$$

We know that S(f) is given as:

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

NOTES

Therefore,

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

As a result, the spectrum of a flat top sampled signal is:

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

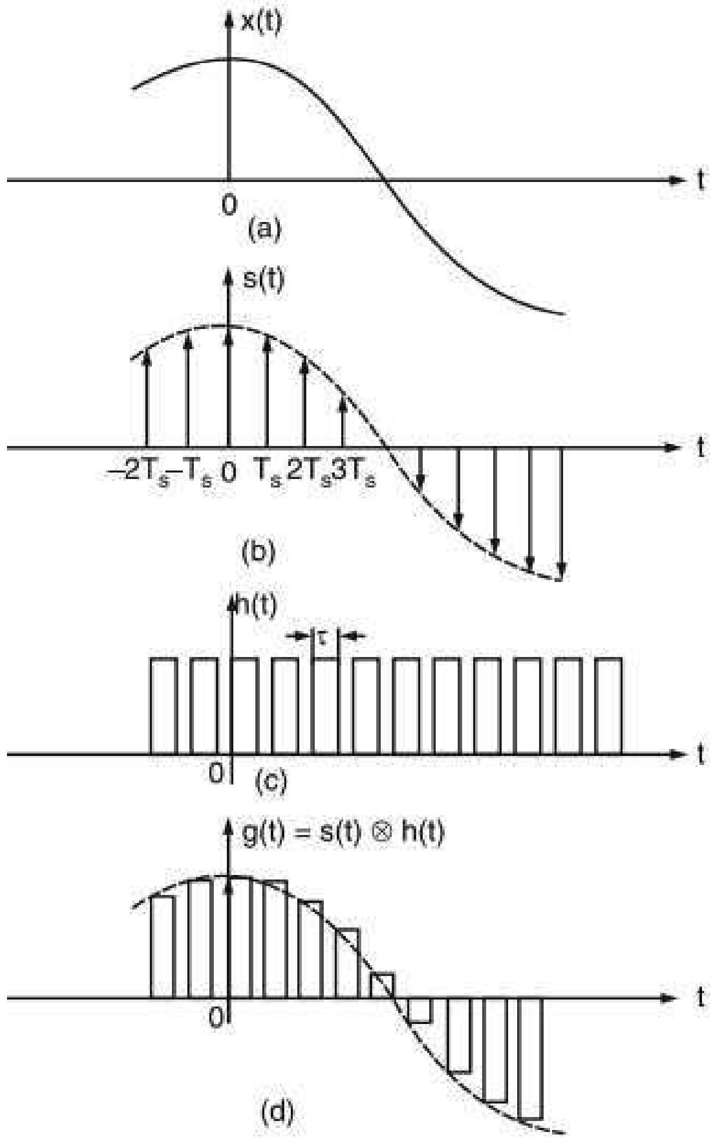


Fig 3.12 *(a) Baseband signal x(t), (b) Instantaneously sample signal s(t), (c) Constant pulse width function h(t), (d) Flat top sampled signal g(t) obtained through convolution of h(t) and s(t)*

Applications

There are few applications of sampling theorem. They are:

- To maintain sound quality in music recordings
- Sampling process applicable in the conversion of analog to discrete form
- Speech recognition systems and pattern recognition systems

- Modulation and demodulation systems
- Sensor data evaluation systems
- Radar and radio navigation system
- Digital watermarking and biometric identification systems, surveillance systems.

NOTES

3.2.3 Ideal Reconstruction

Reconstruction or interpolation is the technique of rebuilding a continuous time signal g(t) out of its samples. A signal g(t) band-limited to B Hz can be reconstructed (interpolated) precisely from its samples, as shown in Figure 3.13. This means not only that uniform sampling at above the Nyquist rate preserves all the signal information, but also that the original message can be reconstructed by simply running the sampled signal through an ideal low-pass filter with bandwidth B Hz. As seen from Eq. (3.1),

$$\bar{g}(t) = g(t)\delta_{T_s}(t)$$
$$= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} g(t)e^{jn2\pi f_s t} \qquad \ldots(3.1)$$

the sampled signal comprises of a component $(1/T_s)g(t)$, and to recover g(t) [or $G(f)$], the sampled signal

$$\bar{g}(t) = \sum g(nT_s)\delta(t - nT_s)$$

must be sent through an ideal low-pass filter of bandwidth B Hz and gain T_s. Such an ideal filter response has the transfer function

$$H(f) = T_s\,\Pi\left(\frac{\omega}{4\pi B}\right) = T_s\,\Pi\left(\frac{f}{2B}\right) \qquad \ldots(3.2)$$

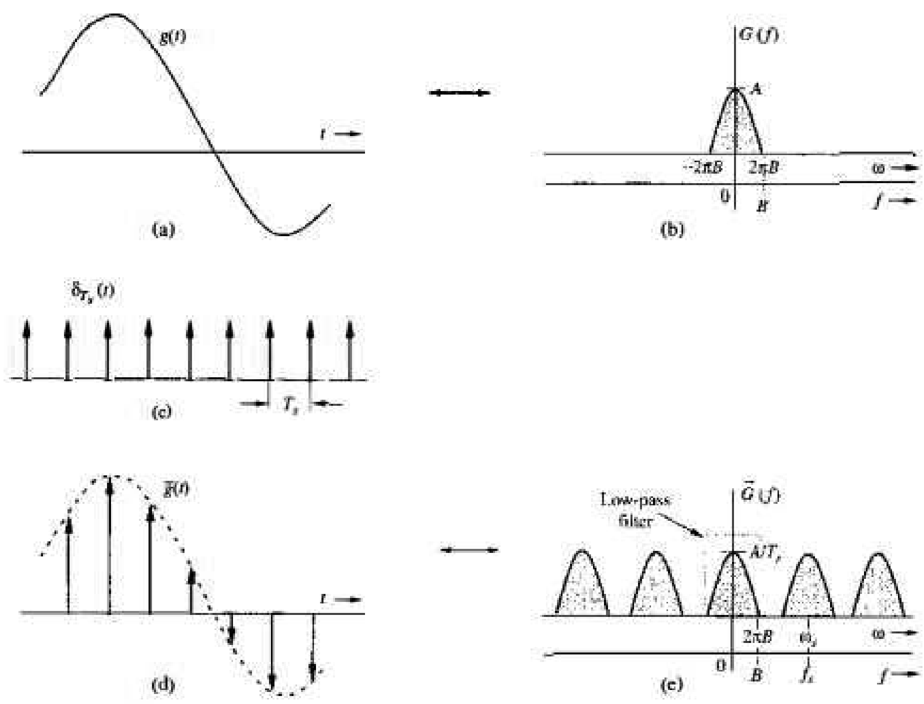


Fig 3.13 *Sampled signal and its Fourier spectra*

NOTES

To recover the analog signal from its uniform samples, the ideal interpolation filter transfer function established in Eq. (3.2) is presented in Figure 3.14(a). The impulse response of this filter, the inverse Fourier transform of $H(f)$, is

$$h(t) = 2BT_s \text{ sinc } (2\pi Bt) \quad \ldots(3.3)$$

Assuming the use of Nyquist sampling rate, that is, $2BT_s = 1$, then

$$h(t) = \text{sinc } (2\pi Bt) \quad \ldots(3.4)$$

This h(t) is shown in Figure 3.14(b). The fact that h(t) = 0 at all Nyquist sampling instants ($t = \pm n/2B$) except t = 0 is highly interesting. When the sampled signal $\bar{g}$ (t) is applied to the filter's input, the result is g(t). Because each sample in $\bar{g}$ (t) is an impulse, it generates a sine pulse with a height equal to the sample's strength, as seen in Figure 3.14. (c). When all of the sine pulses from all of the samples are added together, the result is g. (t). The impulse $g(kT_s)h(t - kT_s)$ is due to the k_{th} sample of the input $\bar{g}$ (t) with the filter output $g(kT_s)h(t - kT_s)$. Therefore, the filter output to $\bar{g}$ (t), which is g(t), may now be written as a sum.,

$$g(t) = \sum_k g(kT_s)h(t - kT_s)$$

$$= \sum_k g(kT_s) \text{ sinc } [2\pi B(t - kT_s)] \quad \ldots(3.5a)$$

$$= \sum_k g(kT_s) \text{ sinc } (2\pi Bt - k\pi) \quad \ldots(3.5b)$$

The interpolation formula epitomized in equation (3.5) is the interpolation formula, which provides values of g(t) between samples as a weighted sum of all the sample values.

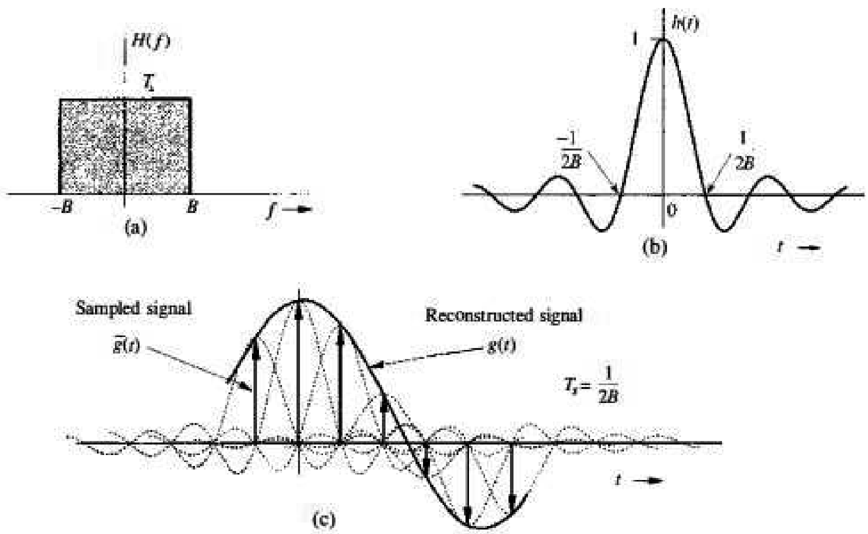


Fig 3.14 *Ideal interpolation*

3.2.4 Major Problems in Practical Sampling

Natural sampling is already discussed. Now, there should be a discussion about the major problems that are faced in practical sampling. These are listed below along with their remedies:

(i) Realizability of Reconstruction Filters

When a signal is sampled at the Nyquist rate of $f_s = 2B$ Hz, the spectrum $\bar{G}(f)$ is composed of $G(f)$ repeats with no gap between them, as illustrated in Figure 3.15(a). The sampled signal $\bar{g}(f)$ must be processed through an ideal low-pass filter to recover g(t) from $\bar{g}(f)$ (dotted region in Figure 3.15(a)). However, in fact, such a filter is unachievable; the closest approximation is an indefinite time delay in the response. This indicates that the signal g(t) can be retrieved with an indefinite time delay from its samples.

The difficulty can be solved by sampling the signal at a rate faster than the Nyquist rate ($f_s > 2B$). This gives $\overline{G}(f)$, which is made up of $G(f)$ repetitions with a limited band gap between the successive cycles, as illustrated in Figure 3.15(b). $G(f)$ can now be retrieved from $\bar{G}(f)$ by using a low-pass filter with a step-wise cut-off characteristic (dotted region in Figure 3.15(b)) $G(f)$. However, even in this instance, the filter gain must be zero beyond the first cycle of $G(f)$ (Figure 3.15(b)). As per theory, even this filter is difficult to realize. The only benefit is that the desired filter can be adequately simulated with a shorter time delay in this scenario. This proves that, even if the sampling rate is larger than the Nyquist rate, it is hard to retrieve a band-limited signal g(t) accurately from its samples in practice. However, as the sampling rate upsurges, the recovered signal tends to the desired signal in a closer way.

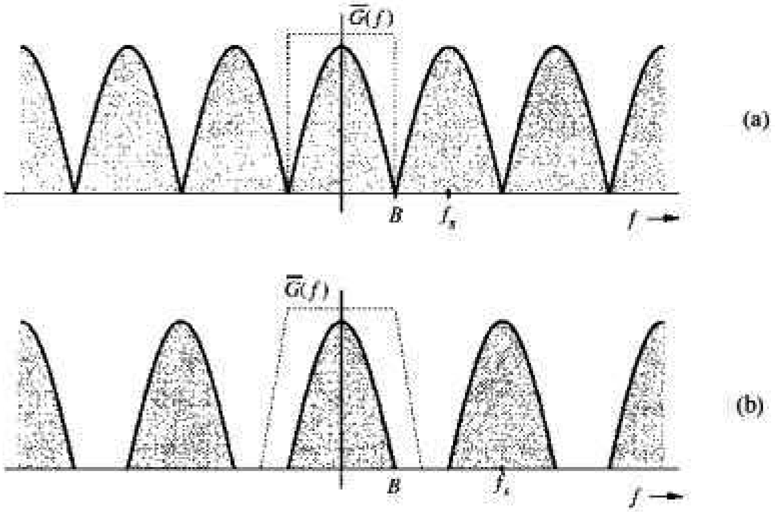


Fig 3.15 *Spectra of a sampled signal: (a) at the Nyquist rate; (b) above the Nyquist rote*

(ii) Treachery of Aliasing

There is one more major practical trouble in reconstructing a signal from its samples. On the presumption that the signal g(t) is band-limited, the sampling theorem was proved. All practical signals are time-limited; that is, they are of determinate duration or width. It may be shown that a signal cannot be both time- and band-limited at the same time. A signal that is time-limited cannot be band-limited, and vice versa

NOTES

(but a signal can be simultaneously non-time-limited and non-band-limited). Clearly, all practical signals, which are necessarily time-limited, are non-band-limited, as shown in Figure 3.16(a); have infinite bandwidth, and the spectrum $\bar{G}(f)$ consists of overlapping cycles of $G(f)$ repeating every f_s Hz (the sampling frequency), as illustrated in Figure 3.16(b). The spectral overlap is inescapable in this situation due to the infinite bandwidth, irrespective of the sampling rate. Interference between recurring spectral cycles is reduced but not eradicated by sampling at a greater rate. $\bar{G}(f)$ has no longer entire information about $G(f)$ due to the overlapping tails, therefore it is no longer possible, even theoretically, to retrieve g(t) exactly from the sampled signal $\bar{g}(f)$. When the sampled signal is processed through an ideal low-pass filter with cut-off frequency $f_s/2$ Hz, the output is not $G(f)$ but $G_a(f)$ (Figure 3.16(c)), which is a distorted version of $G(f)$ caused by two different reasons:

1. The loss of the tail of $G(f)$ beyond $|f| > f_s/2$ Hz.
2. The reappearance of this tail inverted or folded back onto the spectrum.

Note that the spectra cross at frequency $f_s/2 = 1/2T$ Hz, which is called the folding frequency. The spectrum may be viewed as if the lost tail is folding back onto itself at the folding frequency. For instance, a component of frequency $(f_s/2) + f_z$ shows up as, or "impersonates", a component of lower frequency $(f_s/2) - f_z$ in the reconstructed signal. Thus, the components of frequencies above $f_s/2$ reappear as components of frequencies below $f_s/2$. This tail inversion, known as spectral folding or aliasing is shown shaded in Figure 3.16 (b and c). Not only are all the constituents of frequencies above the folding frequency $f_s/2$ Hz lost during the aliasing process, but these same components resurface (aliased) as lower frequency components in Figure 3.16 (b or c). Such aliasing destroys the integrity of the frequency components below the folding frequency, $f_s/2$ as depicted in Figure 3.16(c).

NOTES

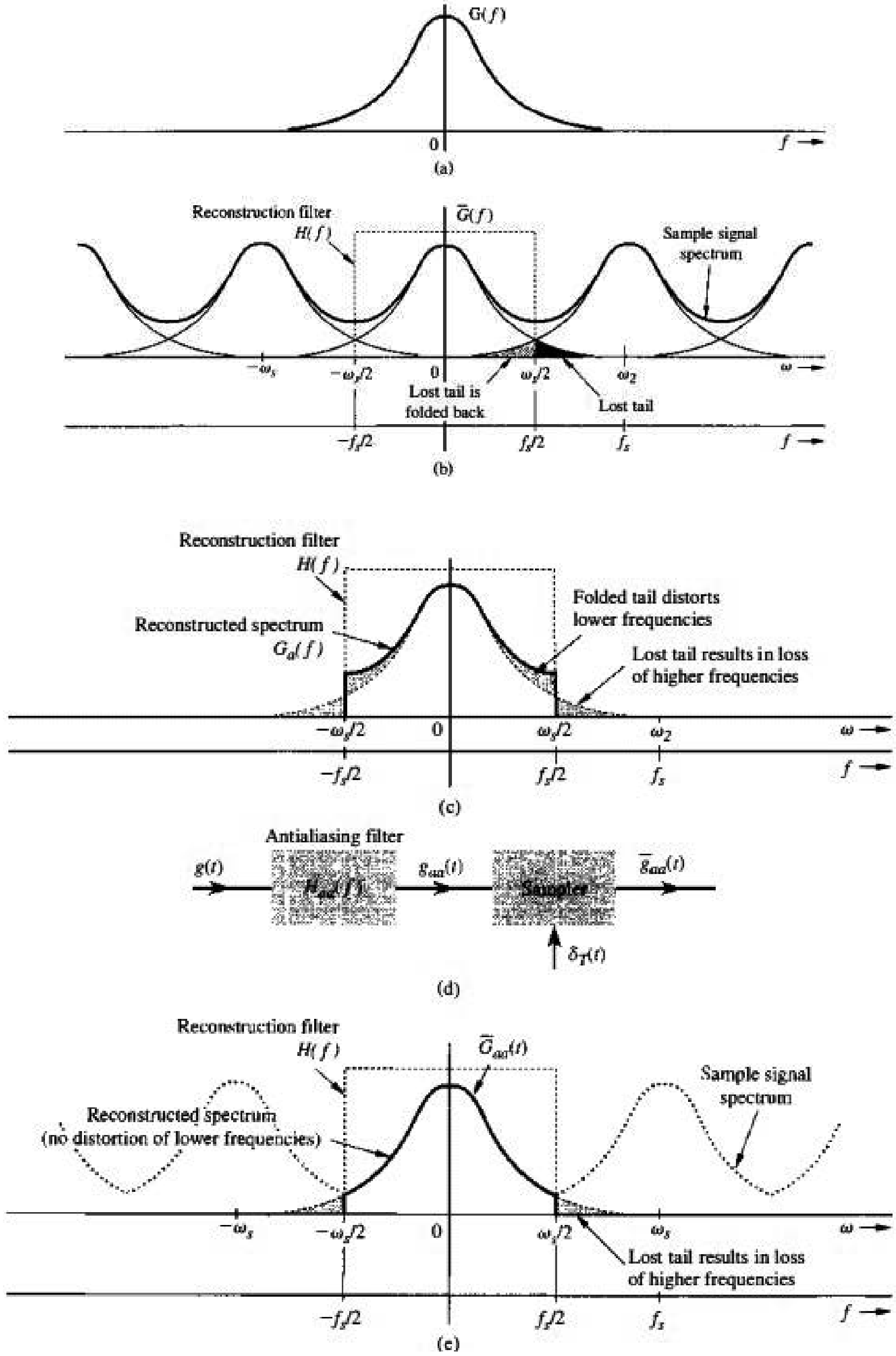


Fig 3.16 *Aliasing effect. (a) Spectrum of a practical signal g(t). (b) Spectrum of sampled g(t) . (c) Reconstructed signal spectrum. (d) Sampling scheme using antialiasing filter. (e) Sampled signal spectrum (dotted) and the reconstructed signal spectrum (solid) when anti-aliasing filter is used*

Defectors Eliminated: The Antialiasing Filter

The potential defectors are all the frequency components beyond the folding frequency ${f_s}/{2} = {1}/{2T}$ Hz. These components should be suppressed from g(t) before sampling g(t). Such suppression of higher frequencies can be accomplished

NOTES

by an ideal low-pass filter of cut-off $f_s/2$ Hz, as shown in Figure 3.16(d). This is called the antialiasing filter. Figure 3.16(d) also displays that antialiasing filtering is executed before sampling. Figure 3.16(e) shows the sampled signal spectrum and the reconstructed signal $G_{aa}(f)$ when the antialiasing scheme is used. An antialiasing filter essentially band-limits the signal g(t) to $f_s/2$ Hz. Only the constituents beyond the folding frequency $f_s/2$ Hz are lost in this method. These suppressed components are now unable to reemerge, causing frequency components below the folding frequency to be corrupted. Clearly, use of an antialiasing filter results in the reconstructed signal spectrum $G_{aa}(f) = G(f)$ for $|f| < f_s/2$. Thus, although the spectrum beyond $f_s/2$ Hz is lost, the spectrum for all the frequencies below $f_s/2$ remains intact. The effective aliasing distortion is cut in half owing to elimination of folding. As a result, antialiasing must be done prior to the signal is sampled.

An antialiasing filter also helps to reduce noise. Noise, generally, has a wideband spectrum, and without antialiasing, the aliasing phenomenon itself will cause the noise components outside the desired signal band to appear in the signal band. Antialiasing suppresses the entire noise spectrum beyond frequency $f_s/2$. The antialiasing filter is unattainable because it is an ideal filter. A steep-cutoff filter is cast-off in practice, which results in a greatly reduced remnant spectrum beyond the folding frequency $f_s/2$.

Check Your Progress

1. Define sampling.
2. What is Nyquist rate?
3. Name the types of sampling strategies.

3.3 ANALOGUE PULSE MODULATION

In analog modulation systems, some parameter of a sinusoidal carrier is varied according to the instantaneous value of the modulating signal. In Pulse modulation methods, the carrier is no longer a continuous signal but consists of a pulse train. Some parameter of which is varied according to the instantaneous value of the modulating signal.

The analog pulse modulation techniques are mainly classified into Pulse Amplitude Modulation, Pulse Duration Modulation/Pulse Width Modulation, and Pulse Position Modulation.

NOTES

3.3.1 Types of Analogue Pulse Modulation

In communication engineering, the message signal (also called baseband signal) is basically voice, video, etc., is technically known as *modulating signal*. Next is the high-frequency *carrier signal* (which is usually a sine wave), by which the modulating signal is shifted to RF (Radio-Frequency) band. By impressing suitably the modulating signal on to the high-frequency carrier signal, through the process of modulation, the transmitted signal, technically called *modulated signal* is obtained. In the transmitter, the modulated signal is amplified in a power amplifier to a level suitable to drive the antenna. In the receiver, the incoming RF modulated signal is ultimately translated back to the baseband signal (like voice, video, etc.) by the process called demodulation (which is reverse to modulation). Demodulation is done in the receivers.

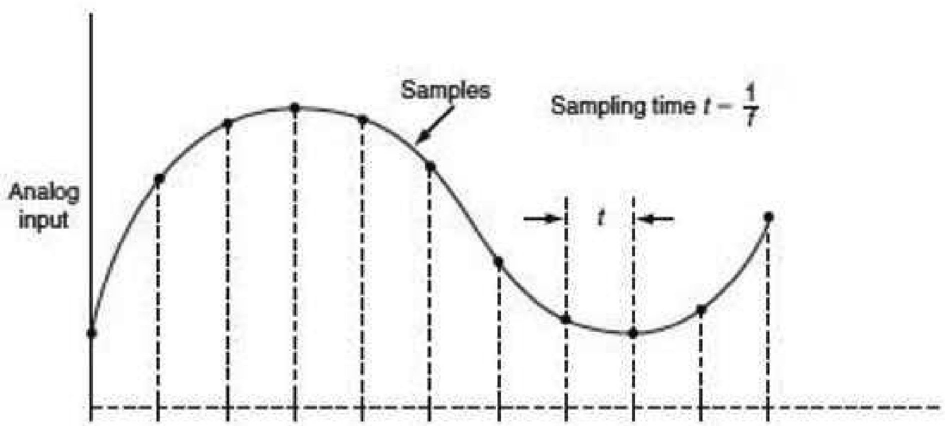
The process for converting continues time single into discrete time single is called Pulse modulation. Pulse code modulation is the technique for changing a message signal into binary pulse to represent the information to be transmitted. The binary pulses are the train of 0s or1s.

The main advantages for transmitting information by binary techniques, instead to continuous time single is reduction of noise interference, good ability of reproduction of signal and can achieve higher transmission bandwidth.

There are three basic forms of Analogue pulse modulation:

1. Pulse-amplitude modulation (PAM),
2. Pulse-width modulation (PWM),
3. Pulse-position modulation (PPM),

Figure shows an analog modulating signal and correspondent various singles in the form of PAM, PWM, and PPM modulators. In all three cases, using the Nyquist conditions, the analog signal is sampled by various sampling techniques with constant sampling time interval. For the proper sampling of analog single the sample rate must be at least two times the highest frequency component of the analog wave.



NOTES

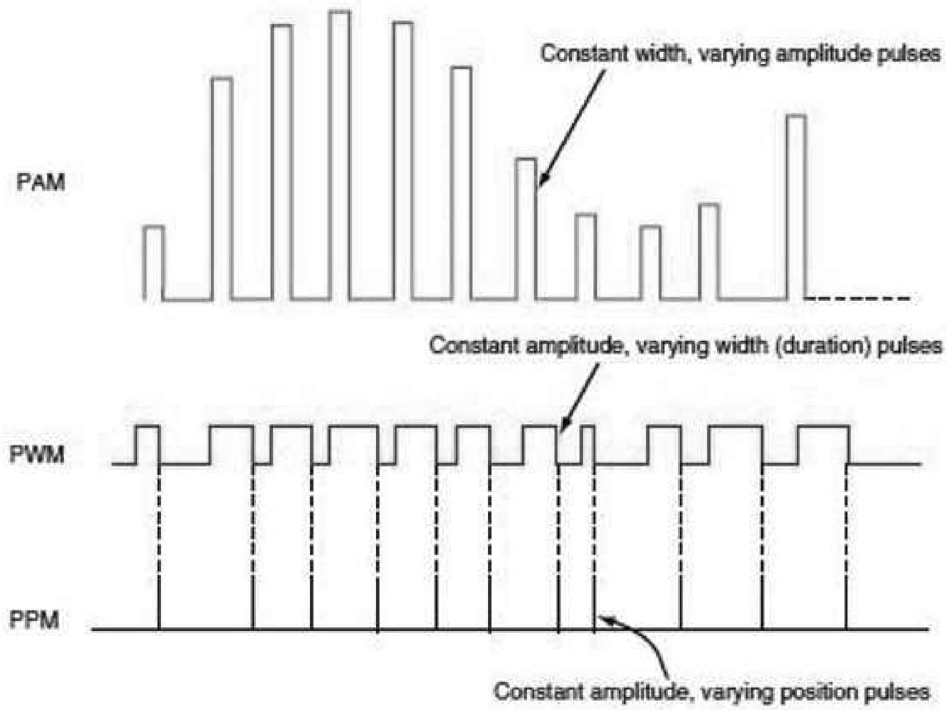


Fig 3.17 *Types of Pulse Modulation*

The need for modulation arises from the fact that the baseband signals, in their original form, are incompatible for transmission over the medium. Modulation is achieved through the modulating techniques. Modulation helps to:

- Reduce the antenna's height
- Avoid the undesired mixing of signals
- Increase the range of communications
- Multiplex the signals
- Make adjustments in the bandwidth
- Improve the quality of reception.

There are three forms of commonly employed modulating techniques: Amplitude Modulation (AM), Frequency Modulation (FM), Phase Modulation (PM).

Analysis of AM, FM and PM Waves

Modulation techniques are methods used to encode digital information in an analog communication system. Amplitude Modulation (AM), Frequency Modulation (FM), and Phase Modulation (PM) are the three different ways of modulating information onto a sinusoidal carrier. Though these are not the only methods of modulation, these are the most common ways of signal modulation.

All the three modulation techniques employ a carrier signal which has a single frequency to carry the intelligence (data) or information. For digital communications, this intelligence is in the form of 1s and 0s. While we are modulating the carrier, we are in fact changing its characteristics to correspond to 1s and 0s.

Amplitude Modulation: It amounts to imposing a signal on the amplitude of a frequency. Amplitude modulation has the undesired effect of modifying the frequency a bit as well, and is much more susceptible to environmental noise such as lightning. Amplitude modulation works well across much of the radio spectrum.

In digital communications, AM involves modifying the amplitude of the carrier to represent 1s and 0s. In the above example, a 1 is represented by the presence of the carrier for a predefined period of 3 cycles of carrier. Absence—or no carrier—indicates a 0. Loss of connection is read as 0s.

Advantages: Amplitude modulation is simple to design.

*Disadvantages***:** Noise spikes on transmission medium often interfere with the carrier signal.

Frequency Modulation: It amounts to imposing a signal by altering the frequency —it takes a very high frequency to modulate either audio or video signals—e.g. megahertz, so it eats up bandwidth.

In digital communication, it modifies the frequency of the carrier to represent 1s and 0s. In the above example, a 0 is represented by the original frequency of the carrier, and a 1 by a much higher frequency (where the cycles are spaced closer together).

Advantages: When compared to AM, FM has greater immunity to noise on transmission medium. In FM, a signal is always present and loss of signal is easily detected.

Disadvantages: It requires two frequencies and the detecting circuitry needs to recognize both the frequencies when signal is lost.

Phase Modulation: It involves altering the phase with a desired signal. It works well when the bandwidth of the desired signal is very small, e.g. digital communications, but reflections of phase modulated signals can be easily corrupted. It also needs some extra equipment.

In case of digital communications, PM modifies the phase of the carrier to represent a 1 or 0 by switching the carrier phase at every occurrence of a 1 bit, but remains unaffected for a 0 bit. The phase of the signal is measured relative to the phase of the preceding bit. The bits are timed to coincide with a specific number of carrier cycles (three in the present example = 1 bit).

Advantage: In this method, only one frequency used. It is easy to detect loss of carrier.

Disadvantages: To generate and detect phase changes complex circuitry is required.

Demodulation: It is the act of extracting or retrieving the original information-bearing signal from a modulated carrier wave. In other words, it is exactly the action reverse to modulation. A demodulator is an electronic circuit used to recover the information content from the modulated carrier wave. Modulator and demodulator are traditionally used terms in connection with radio transmission, but many other systems use many different kinds of modulators and demodulators.

NOTES

A **modem** is a contraction of the terms **mo**dulator and **dem**odulator.

3.3.2 Pulse Modulation Characteristics

Frequency modulation involves the modulation of the frequency of the analog sine wave as shown in Figure 3.18 where the instantaneous frequency of the carrier is deviated in proportion of the deviation of the modulated carrier with respect to the frequency of the instantaneous amplitude of the modulating signal. Simply, it may be said that it occurs when the frequency of a carrier is changed based upon the amplitude of input signal.

Unlike AM, the amplitude of carrier signal is unchanged. This makes FM modulation more immune to noise than AM and improves the overall signal-to-noise ratio of the communications system. Power output is also constant, differing from the varying AM power output. The amount of analog bandwidth necessary to transmit FM signal is greater than the amount necessary for AM, a limiting constraint for some systems.

The modulating index for FM is as follows:

$$\beta = f_p/f_m$$

where

β = Modulation index

f_m = frequency of the modulating signal, and

f_p = peak frequency deviation

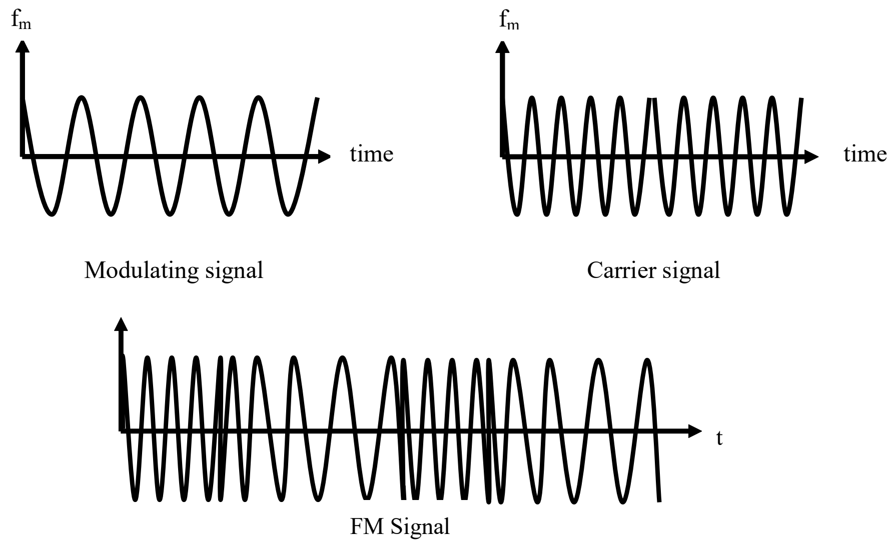


Fig. 3.18 *Frequency Modulation*

From Figure 3.18, it is inferred that the amplitude of the modulated signal always remains constant, irrespective of the frequency and the amplitude of modulating signal. It means that the modulating signal adds no power to the carrier in frequency modulation, unlike amplitude modulation. FM produces an infinite number of sidebands spaced by the in modulation frequency, f_m which is not in case of AM. Therefore, AM is considered as a linear process whereas FM as a nonlinear process. It is necessary to transmit all sidebands to reproduce a distortion-free signal. Ideally, the *bandwidth* of the modulated signal is infinite in this case. In general, the determination of the frequency content of an FM waveform is complicated, but when *b* is small, the bandwidth of the FM signal is:

$2f_m$.

NOTES

On the other hand, when β is large, the bandwidth is determined (empirically) as:

$2 f_m (1 + \beta)$.

Phase modulation

Phase Modulation (PM) is similar to frequency modulation. Instead of the frequency , the phase of the carrier wave changes. In PM, the phase of the carrier is made proportional to the instantaneous amplitude of the modulating signal.

Modulating index for PM is given as:

$\beta = \Delta\varphi$

where

$\Delta\varphi$ is the peak phase deviation in radians.

As in the case of angular modulation argument of sinusoidal is varied and therefore we will have the same resultant signal properties for frequency and phase modulation. A distinction in this case can be made only by direct comparison of the signal with the modulating signal wave, as shown in Figure 3.19.

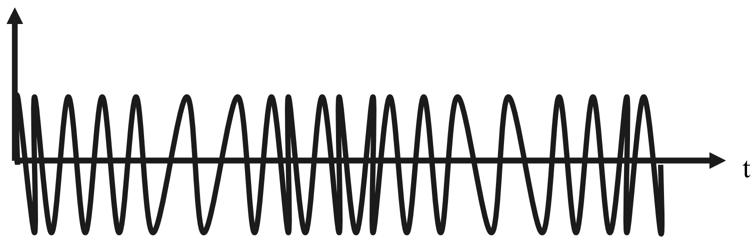


Fig. 3.19 *Phase Modulation*

Phase modulation and frequency modulation are interchangeable by selecting the frequency response of the modulator so that its output voltage will be proportional to integration of the modulating signal and differentiation of the modulating signal respectively. Bandwidth and power issues are same as in case of the frequency modulation.

Generation of FM Waves

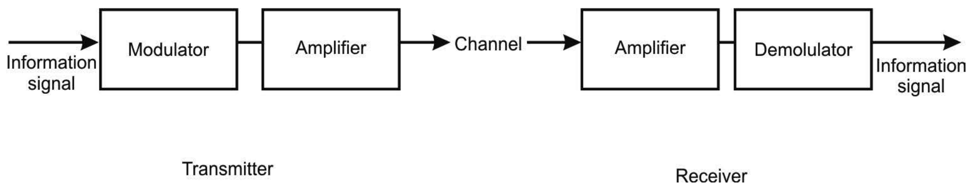
Basic system

The basic units of communications system are:

Transmitter: It picks up the input information signal and processes it prior to transmission. The transmitter first modulates the information onto a carrier signal, amplifies it and then broadcasts it over the channel

Channel: It is the medium that transports the modulated signal to the receiver. It is usually a wiring system like in cable TV or the Internet or space in case of wireless transmission.

Receiver: This sub-system takes the transmitted signal from the channel and processes it to retrieve the information signal. The receiver must be able to differentiate the signal from other signals that may be simultaneously using the same channel (this process is called tuning), amplify the signal for processing and demodulate (convert the carrier signal to original signal) to retrieve the information. It then processes the information for reception (for example, to be broadcast on a loudspeaker).



Freqency modulation

In order to facilitate electromagnetic transmission, the information signal is first converted from audio/video/digital format into an electric signal using a *transducer* after which the frequency of the signal is systematically modulated into a carrier signal to enable electromagnetic transmission.

A modulated carrier signal is used for two reasons:

- For reducing the wavelength that results in efficient transmission and reception. The optimum antenna size is ½ or ¼ of a wavelength. A usual audio frequency of 3000 Hz has a wavelength of 100 km and needs an effective antenna length of 25 km. By comparison, a typical FM carrier signal has100 MHz, and 3 m, and could use an antenna only 80 cm long.
- It allows simultaneous use of the same channel, and such process is called *multiplexing*. Each unique signal can be assigned a different carrier frequency (like radio transmissions) and still share the same channel.

The basic sine wave is like $V(t) = V_o \sin(2\pi f t + \phi)$ where the parameters are defined below:

$V(t)$ the voltage of the signal as a function of time.

V_o the amplitude of the signal (represents the maximum value achieved each cycle)

f the frequency of oscillation, the number of cycles per second (also known as Hz = 1 cycle per second).

ϕ is the phase of the signal, representing the starting point of the cycle.

Frequency modulation uses the information signal, $V_m(t)$ to vary the carrier frequency within some suitable range about its original value. The three signals in mathematical form:

- Information: $V_m(t)$
- Carrier: $V_c(t) = V_{co} \sin(2 p f_c t + \phi)$
- FM: $V_{FM}(\mathrm{t}) = V_{co} \sin(2\pi [f_c + (\Delta f/V_{mo}) V_m(t)] t + \phi)$

Here we have represented the carrier frequency term, with a time-varying frequency and also introduced a new term: Δf, *the peak frequency deviation*. In this form, the carrier frequency term: $f_c + (\Delta f/V_{mo})\, V_m\,(t)$ varies between the extremes of $f_c - \Delta f$ and $f_c + \Delta f$. The term Δf is referred to as the 'swing' in the frequency.

NOTES

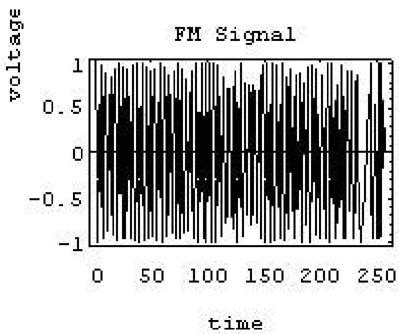
We can also define a modulation index for FM, analogous to AM:

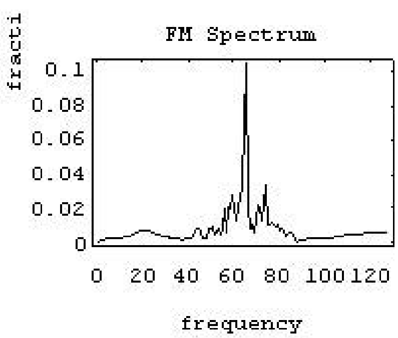
$$\beta = \Delta f/f_m, \text{ where } f_m \text{ is the maximum modulating frequency used.}$$

The modulation index, β, is a measure of the peak frequency deviation, Δf. In other words, β represents a way to express the peak deviation frequency as a multiple of the maximum modulating frequency, f_m, i.e. $\Delta f = \beta f_m$. For example, in a radio the audio signal to be transmitted ranges from 20 to 15,000 Hz . If the FM system used a maximum modulating index, β, of 5.0, then the frequency would 'swing' by a maximum of 5 × 15 kHz = 75 kHz above and below the carrier frequency.

FM spectrum: It shows the relative amounts of different frequency components in a signal, like in the display on the graphic-equalizer in your stereo showing the relative amounts of bass, midrange and treble. These correspond directly to the increasing frequencies (treble being the high frequency components).

Let us discuss the example of an audio spectrum to provide the frequency modulation:





In the example, the input information signal varies between 1 and 11 Hz. The carrier is at 65 Hz and the modulation index is 2. The individual sideband spikes are replaced by a more-or-less continuous spectrum. However, the extent of the sidebands is limited to $(\beta + 1) f_m$ (approximately) above and below. Here, that would be 33 Hz above and below, making the bandwidth about 66 Hz. We see the sidebands extend from 35 to 90 Hz, so our observed bandwidth is 65 Hz.

Here we have ignored the smooth humps at the extreme ends of the spectrum, because they are by-products of FM (in this example, there is no random noise). Since they consume only a small fraction of the total power, we can ignore these humps. In practice, the random noise would obscure them anyway.

FM performance

Bandwidth: As discussed, the bandwidth of an FM signal may be calculated using:

$$BW = 2\,(\beta + 1)\, f_m$$

where β is the modulation index and f_m is the maximum modulating frequency used.

NOTES

In FM, both the modulation index and the modulating frequency affect the bandwidth (whereas the bandwidth of AM signals depends only on the maximum modulation frequency).

Efficiency: The efficiency of a signal is the power in the side-bands as a fraction or percentage of the total. In FM signals, because of the higher number of side-bands produced, the efficiency is generally high. Remember that the conventional AM efficiency is limited to about 33 per cent efficiency to prevent distortion in the receiver when the modulation index was greater than 1. FM has no such analogous problem.

Although the side-band structure is somewhat complicated, it is safe to assume that the efficiency generally improves if we make the modulation index larger. But if we make the modulation index larger, then we have to make the bandwidth also larger. And this has its disadvantages. As a principle, a compromise between efficiency and performance is struck. The modulation index is normally limited in the range between 1 and 5, depending on the application.

Noise: FM systems can reject the noise far better than AM systems. Across the sprectrum, usually noise is spread uniformly. The amplitude of the noise varies randomly at these frequencies. The AM systems are very sensitive to random noise and the change in amplitude can actually modulate the signal and be picked up in the AM systems.

Inherently, the FM systems are immune to random noise. Since the noise is distributed uniformly in frequency and varies mostly in amplitude, FM receivers pick up virtually no interference.

In FM signals, efficiency and bandwidth both depend on the maximum modulating frequency as well as the modulation index.

The FM signal, compared to AM, has a higher efficiency, a larger bandwidth and higher immunity to noise.

Methods of FM Generation

Methods of FM generation method of FM are basically of two types-direct methods and indirect methods.

Direct Methods: Under direct methods there are two well-known methods, these are:

1. FM generation using FET veractor modulator
2. FM generation using veractor diode

1. Frequency modulation using FET reactance modulator

Figure 3.20 shows the basic circuit of FET reactance modulator. It behaves as reactance across terminals A-B, which may be connected across the tuned circuit of the oscillator to get FM output. The varying voltage (modulating voltage) V, across terminals A-B changes reactance of the FET. This change in reactance can be *inductive* or *captive*.

NOTES

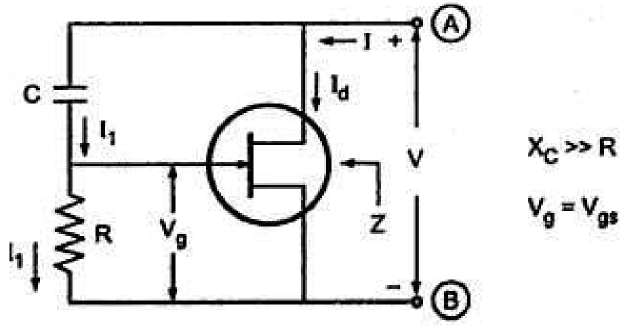


Fig. 3.20 *FET Reactance Modulator*

Neglecting the gate current, let the current through C and R be I_1. At the carrier frequency, the reactance of C is much larger than R. We can write equation for I_1 as,

$$I_1 = \frac{V}{R + \frac{1}{j\omega C}}$$

Since $j\omega C >> R,$ we can write the above equation as,

$$I_1 = j\omega CV$$

From the circuit, $V_g = I_1 R = j\omega CRV$

For the FET, $I_d = g_m V_{gs} = g_m V_g$

$$= j\omega CR\, g_m V$$

From the circuit, the impedance of the FET is,

$$Z = \frac{V}{I_d} = \frac{V}{j\omega g_m CRV} = \frac{1}{j\omega[g_m CR]} = \frac{1}{j\omega C_{rq}}$$

Here $C_{rq} = g_m CR$. Thus the impedance of FET is capacitive reactance. By varying the modulating voltage across FET, the operating point *gm* can be varied. Hence this varies C_{rq} and the change in the capacitance will change the frequency of the oscillator. If we connect inductance instead of capacitor, we get inductive reactance in the circuit

2. Frequency modulation using varactor diode

All the diodes exhibit small junction capacitance in the reverse biased condition. The varactor diodes are specially designed to explait and optimize this characteristic. The junction capacitance of the varactor diode changes when the reverse bias across it is varied The variations in capacitance of this diode are wide and linear. The varactor diodes provide the junction capacitance in the range of 1 to 200 pF. Figure 3.21 shows how varactor diode can be used to generate FM. L_1 and C_1 form the tank circuit of the carrier oscillator. The capacitance of the varactor diode depends upon the fixed bias Bet by R_1 and R_2 and the AF modulating signal. Either R_1 or R_2 is made variable so that the center carrier frequency can be adjusted over a narrow range. The Radio Frequency Choke (RFC) has high reactance at the canier frequency to prevent the carrier signal from getting into the modulating

NOTES

signal circuits. At positive going modulating signal adds to the reverse bias applied to the varactor diode D, which decreases its capacitance and increases the carrier frequency. A negative going modulating signal subtracts from the bias, increasing the capacitance, which decreases the carrier frequency.

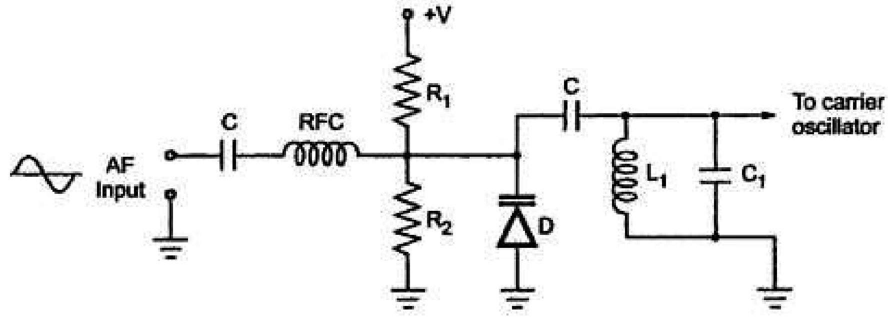


Fig. 3.21 *Varactor Diode for FM Generation*

The frequency of the LC oscillator changes due to temperature effects. Hence crystals are used in FM generators to provide frequency stability.

Automatic Frequency Correction

Need for Automatic Frequency Correction: The frequency of the oscillator is directly varied in direct FM transmitters. Hence such oscillators do not produce stable frequency. This problem can be overcome with the help of Automatic Frequency Correction (AFC).

The automatic frequency correction is incorporated in FM transmitter to keep carrier frequency stable. Figure 3.22 shows the block diagram of AFC circuit

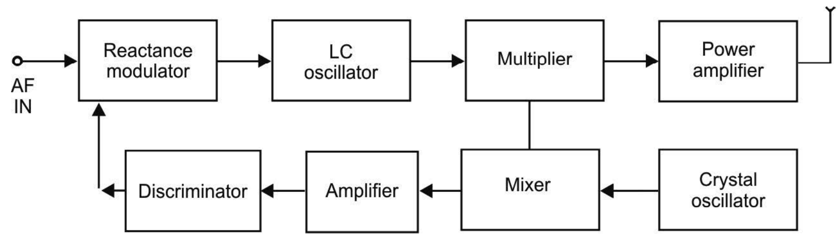


Fig 3.22 *Typical AFC Circuit*

The discriminator reacts is sensitive only to small changes in the carrier frequency but not to the frequency deviations in the carrier (since it is too fast).

Suppose the frequency of the carrier increases. This higher frequency is fed to the mixer for which the other input frequency is from the stable crystal oscillator. The discriminator is tuned to the correct frequency difference that is desired between the LC oscillator and the crystal oscillator, and since its input frequency is now somewhat higher, the discriminator will develop a positive d.c. voltage. This voltage is applied to the reactance modulator. The transconductance is increased by the positive voltage developed by the discriminator. This increases the equivalent capacitance of the reactance modulator and thereby decreases the oscillator frequency. Hence, frequency increase in the carrier frequency is lowered and brought to the correct value.

The correcting d.c. voltage developed by the discriminator may be applied to a varactor diode that is connected across the tank circuit of the oscillator and be used for AFC purposes.

In directly modulated FM transmitters, usually the frequency modulation is carried out at a lower frequency and with a smaller frequency deviation. Then passing this frequency modulated wave through frequency multiplier circuit, the desired carrier frequency and desired frequency deviation are achieved. But it may happen that the frequency multiplication required for getting the desired carrier frequency from a smaller carrier frequency may not be exactly the same as required for getting the desired frequency deviation from the original smaller deviation. In such situation, the mixer is used to get the final values.

NOTES

3.3.3 Pulse Amplitude Modulation (PAM)

It is an analog modulating scheme in which the amplitude of the pulse carrier varies relational to the instantaneous amplitude of the message signal.

There are several forms exist of a single modulation, pulse amplitude modulation (PAM) is one of forms of single modulation in this type of modulation data is transmitted through varying the amplitude of the pulses in consistent timed sequence of electromagnetic or electrical pulses.

Number of pulse amplitudes can be infinite, in the case of analog pulse amplitude modulation signals. Pulse amplitude modulation is mostly used in transmission of digital data, with non-baseband applications like Ethernet communication standard.

In pulse amplitude modulation, a signal is sampled at fixed intervals and which is made proportionate to the magnitude of the signal.

Some pulse amplitude modulation systems have the amplitude inversely proportional to instantaneous modulating signal. In some cases, the amplitude is directly proportional to instantaneous modulating signal.

Further the sampled pulses are fed to the receiving end by the channel, or sampled pulses are modulated using a carrier wave before transmission.

Pulse amplitude modulation are of two types.

1. **Single Polarity:** A direct current (DC) level which is fixed is added to the signal. This ensures single polarity signal and all pulses are positive.
2. **Double Polarity:** This sort of pulse amplitude modulation will have the output of modulating signal in *both positive* and *negative* ends.

The pulse amplitude modulated signal, will follow the amplitude of the original signal, as the signal traces out the path of the whole wave. In natural PAM, a signal sampled at the Nyquist rate is reconstructed, by passing it through an efficient Low Pass Frequency (LPF) with exact cutoff frequency. The following figures explain the Pulse Amplitude Modulation.

NOTES

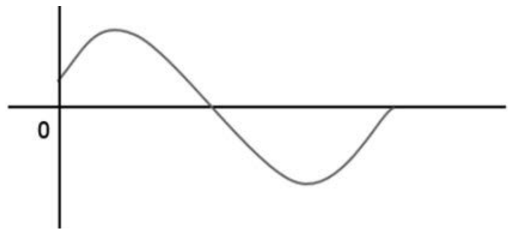


Fig. 3.23 *Pulse Amplitude Modulating Signal*

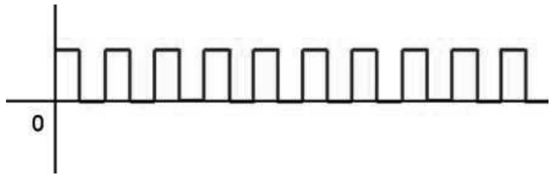


Fig. 3.24 *Carrier Pulse Train*

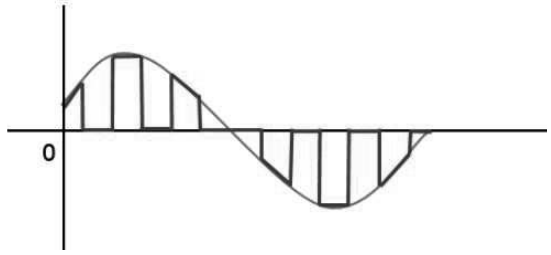


Fig. 3.25 *Natural PAM*

Though the PAM signal is passed through the low pass filter, it cannot recover the signal without distortion. Hence to avoid this noise, flat-top sampling is done as shown Figure 3.26.

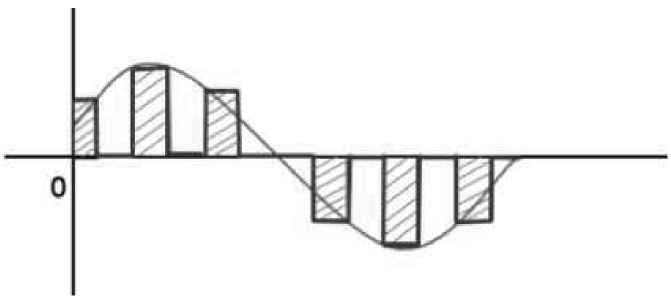


Fig. 3.26 *Flat-top PAM*

Flat-top sampling

It is the process in which sampled signal can be represented in pulses for which the amplitude of the signal cannot be changed with respect to the analog signal, to be sampled. The tops of amplitude remain flat. This process simplifies the circuit design.

3.3.4 Pulse Duration Modulation (PDM): Analysis, Generation and Recovery of PDM

Pulse Width Modulation (PWM) or Pulse Duration Modulation (PDM) or Pulse Time Modulation (PTM) is an analog modulating scheme in which the duration or

NOTES

width or time of the pulse carrier fluctuates proportional to the instantaneous amplitude of the message signal. In this method, the width of the pulse fluctuates but the amplitude of the signal remains constant. Amplitude limiters are used to make the amplitude of the signal constant. These circuits clip off the amplitude, to a preferred level and hence the noise is limited. Figure 3.27 explains the types of Pulse Width Modulations.

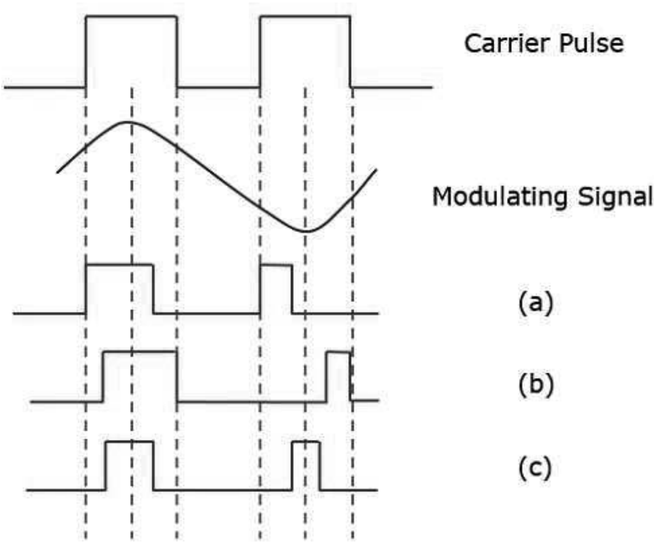


Fig. 3.27 *PWM with Its Variations*

There are three variations of PWM which are as follows:

- The leading edge of the pulse being constant, the trailing edge varies according to the message signal.
- The trailing edge of the pulse being constant, the leading edge varies according to the message signal.
- The center of the pulse being constant, the leading edge and the trailing edge varies according to the message signal.

3.3.5 Pulse Position Modulation (PPM): Analysis, Generation and Recovery

With the help of sampling theorem, a continuous time signal may be completely represented and recovered from the knowledge of samples taken uniformly. A sufficient number of signal samples must be acquired so that the original signal is properly represented in the samples. Also, the original signal should be recoverable or reconstructable entirely from its samples. The number of samples required is determined by the signal's maximum signal frequency. To recover the original signal from its samples, the low pass filter is used. Interpolation filter is the term for this. Reconstruction or interpolation is the process of rebuilding a continuous time signal x(t) from its samples.

Reconstruction is the process of creating an analog voltage (or current) from samples. A digital-to-analog converter takes a series of binary numbers and recreates the voltage (or current) levels that correspond to that binary number.

Then this signal is filtered by a low pass filter. This process is analogous to interpolating between points on a graph, but it can be shown that under certain conditions the original analog signal can be reconstructed exactly from its samples. Unfortunately, the conditions for exact reconstruction cannot be achieved in practice, and so in practice the reconstruction is an approximation to the original analog signal.

3.3.6 Comparison of PDM and PPM

In analogue modulation systems, the modulating signal's instantaneous value is used to change the characteristic of a sinusoidal carrier wave. The carrier in pulse modulation methods is no longer a continuous signal but a pulse train. Pulse modulation systems are divided into two categories, as shown below:

(i) Pulse Amplitude Modulation (PAM)

(ii) Pulse Time Modulation (PTM)

The amplitude of the carrier pulse train's pulses is altered according to the modulating signal in PAM. The figure 3.28 represents a PAM signal:

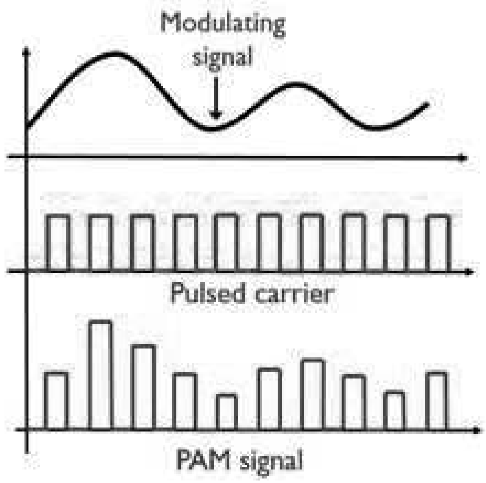


Fig 3.28 *Representation of PAM signal*

The amplitude of the pulses varies with regard to the amplitude of the analogue modulating signal, as in amplitude modulation, as seen in figure 3.28. However, unlike AM, the carrier wave is a pulse train instead of a continuous wave signal in this case. In addition, the carrier pulse's pulse timing is modified in PTM. There are two types of PTM, as listed below:

(i) Pulse Width/ Duration Modulation (PWM/ PDM)

(ii) Pulse Position Modulation (PPM)

In PDM, the carrier pulse train's width or duration is modified in response to the modulating signal. The pulse width modulated signal is depicted in Figure 3.29:

NOTES

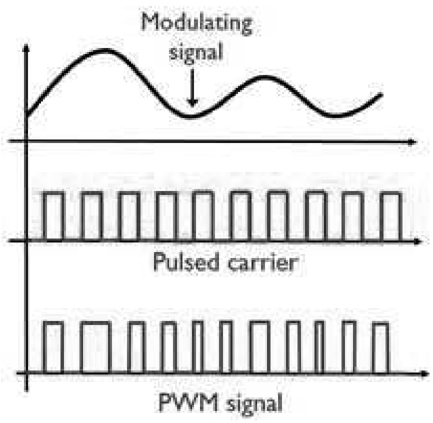


Fig 3.29 *Representation of PWM/ PDM signal*

One can see that, contrary PAM, the amplitude of the signal is consistent with this approach, while only the width varies. The PWM process is equivalent to frequency modulation in that the frequency of the pulses in the PWM signal varies as the width of the pulses varies. The position of carrier pulse train pulses is also altered in PPM. The figure 3.30 shows the PPM signal.

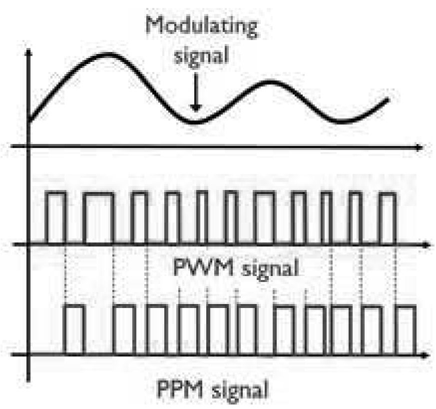


Fig 3.30 *Representation of PPM signal*

The pulse amplitude and pulse width are two constants that do not vary with the modulating signal's amplitude, but only the position does. It's worth noting that the pulse's position varies depending on the reference pulses. And all of these reference pulses are PWM pulses. PWM pulses' falling edges serve as the beginning point for PPM pulses.

Because a pulsed carrier is altered according to an analog message signal, these are known as analog pulse modulation techniques.

PAM, PDM and PPM can be compared according to the points given in Table 3.1.

NOTES

Table 3.1 *Comparison among PAM, PDM and PPM*

Sr. No.	Parameter	PAM	PDM	PPM
1.	Type of Carrier	Train of Pulses	Train of Pulses	Train of Pulses
2.	Variable Characteristic of the Pulsed Carrier	Amplitude	Width	Position
3.	Bandwidth Requirement	Low	High	High
4.	Noise Immunity	Low	High	High
5.	Information Contained in	Amplitude Variations	Width Variations	Position Variations
6.	Power efficiency (SNR)	Low	Moderate	High
7.	Transmitted Power	Varies with an amplitude of pulses	Varies with variation in width	Remains Constant
8.	Need to transmit synchronizing pulses	Not needed	Not needed	Necessary
9.	Bandwidth depends on	Bandwidth depends on the width of the pulse	Bandwidth depends on the rise time of the pulse	Bandwidth depends on the rise time of the pulse
10.	Transmitter power	Instantaneous transmitter power varies with the amplitude of the pulses	Instantaneous transmitter power varies with the amplitude and width of the pulses	Instantaneous transmitter power remains constant with the width of the pulses
11.	The complexity of generation and detection	Complex	Easy	Complex
12.	Similarity with other Modulation Systems	Similar to AM	Similar to FM	Similar to PM

3.3.7 Signal to Noise Ratio in Pulsed System (PAM, PDM and PPM)

Signal to noise ratio in pulse system $(SNR)_o$ is defined as the ratio of the average signal power (S) to the average quantizing noise power (N_q).

$$(SNR)_o = \frac{Average\,signal\;power}{Average\,quantizing\;noise\;power} = \left(\frac{S}{N_q}\right)_o$$

Here, average signal power, S = $\frac{A^2}{2}$

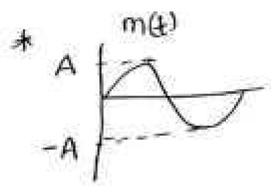


Fig 3.31

And average quantizing noise power, $N_q = q_e^2 = \frac{\nu^2}{12} = \frac{A^2}{3L^2}$

$$(quantizer\ step\ size, \nu = \frac{2A}{L} = \frac{2A}{2^n})$$

Now,
$$(SNR)_o = \frac{\frac{A^2}{2}}{\frac{A^2}{3L^2}} = \frac{3}{2}L^2$$

Expressing in decibels,
$$\left(\frac{S}{N_q}\right)_{odB} = 10\,log\left(\frac{S}{N_q}\right)_o$$
$$= 1.76 + 20 logL$$
$$= 1.76 + 6.02n \quad \text{(Since, } L = 2^n\text{)}$$

Check Your Progress

4. Why does the need for modulation arise?
5. Name commonly employed modulating techniques.
6. What is phase modulation?

3.4 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Sampling is defined as "the process of obtaining discrete values for the instantaneous values of a continuous-time signal." A sample is a subset of data drawn from a larger set of continuous data in the time domain.
2. The sampling rate should be such that neither the data in the message signal is lost nor is it over-lapped. As a result, a rate was established for this, known as the Nyquist rate.
3. There are three types of sampling strategies in general:
   i. Instantaneous sampling
   ii. Natural sampling
   iii. Flat top sampling
4. The need for modulation arises from the fact that the baseband signals, in their original form, are incompatible for transmission over the medium.
5. There are three forms of commonly employed modulating techniques: Amplitude Modulation (AM), Frequency Modulation (FM), Phase Modulation (PM).
6. In PM, the phase of the carrier is made proportional to the instantaneous amplitude of the modulating signal.

3.5 SUMMARY

NOTES

- Analog signals are continuous in time and have difference in voltage levels for different periods of the signal. Here, its main drawback is that the amplitude keeps on changing along with the period of the signal. This can be overcome by the digital form of signal representation.
- Sampling is defined as "the process of obtaining discrete values for the instantaneous values of a continuous-time signal." A sample is a subset of data drawn from a larger set of continuous data in the time domain.
- The interval between the samples should be maintained in order to discretize the signals.
- The sampling rate should be such that neither the data in the message signal is lost nor is it over-lapped. As a result, a rate was established for this, known as the Nyquist rate.
- The sampling theorem, often known as the Nyquist theorem, establishes the notion of sufficient sample rate in terms of bandwidth for band-limited functions. "A signal can be exactly recreated if it is sampled at a rate f_s that is larger than twice the maximum frequency W," according to the sampling theorem.
- Instantaneous sampling is referred to as ideal sampling, whereas natural sampling and flat-top sampling are referred to as practical sampling methods.
- In ideal sampling, the sampling function is a train of impulses and the principle used is known as multiplication principle.
- Natural sampling is a practical method. In this type of sampling, the pulse has a definite width of τ.
- Reconstruction or interpolation is the technique of rebuilding a continuous time signal g(t) out of its samples. A signal g(t) band-limited to B Hz can be reconstructed (interpolated) precisely from its samples.
- Noise, generally, has a wideband spectrum, and without antialiasing, the aliasing phenomenon itself will cause the noise components outside the desired signal band to appear in the signal band.
- In communication engineering, the message signal (also called baseband signal) is basically voice, video, etc., is technically known as *modulating signal*. Next is the high-frequency *carrier signal* (which is usually a sine wave), by which the modulating signal is shifted to RF (Radio-Frequency) band.
- The need for modulation arises from the fact that the baseband signals, in their original form, are incompatible for transmission over the medium. Modulation is achieved through the modulating techniques.
- In digital communications, AM involves modifying the amplitude of the carrier to represent 1s and 0s.
- In digital communication, FM modifies the frequency of the carrier to represent 1s and 0s.

NOTES

- In case of digital communications, PM modifies the phase of the carrier to represent a 1 or 0 by switching the carrier phase at every occurrence of a 1 bit, but remains unaffected for a 0 bit.
- Demodulation is the act of extracting or retrieving the original information-bearing signal from a modulated carrier wave. In other words, it is exactly the action reverse to modulation.
- Frequency modulation involves the modulation of the frequency of the analog sine wave.
- FM produces an infinite number of sidebands spaced by the in modulation frequency, f_m which is not in case of AM. Therefore, AM is considered as a linear process whereas FM as a nonlinear process. It is necessary to transmit all sidebands to reproduce a distortion-free signal.
- All the three modulation techniques employ a carrier signal which has a single frequency to carry the intelligence (data) or information.
- Phase Modulation (PM) is similar to frequency modulation. Instead of the frequency, the phase of the carrier wave changes. In PM, the phase of the carrier is made proportional to the instantaneous amplitude of the modulating signal.
- In order to facilitate electromagnetic transmission, the information signal is first converted from audio/video/digital format into an electric signal using a *transducer* after which the frequency of the signal is systematically modulated into a carrier signal to enable electromagnetic transmission.
- Frequency modulation uses the information signal, $V_m(t)$ to vary the carrier frequency within some suitable range about its original value.
- All the diodes exhibit small junction capacitance in the reverse biased condition. The varactor diodes are specially designed to explait and optimize this characteristic. The junction capacitance of the varactor diode changes when the reverse bias across it is varied. The variations in capacitance of this diode are wide and linear.
- Pulse Amplitude Modulation is an analog modulating scheme in which the amplitude of the pulse carrier varies relational to the instantaneous amplitude of the message signal.
- In pulse amplitude modulation, a signal is sampled at fixed intervals and which is made proportionate to the magnitude of the signal.
- Pulse amplitude modulation are of two types:
  1. Single Polarity: A direct current (DC) level which is fixed is added to the signal. This ensures single polarity signal and all pulses are positive.
  2. Double Polarity: This sort of pulse amplitude modulation will have the output of modulating signal in *both positive* and *negative* ends.
- Pulse Width Modulation (PWM) or Pulse Duration Modulation (PDM) or Pulse Time Modulation (PTM) is an analog modulating scheme in which the duration or width or time of the pulse carrier fluctuates proportional to the instantaneous amplitude of the message signal. In this method, the width of

NOTES

the pulse fluctuates but the amplitude of the signal remains constant. Amplitude limiters are used to make the amplitude of the signal constant.

- A digital-to-analog converter takes a series of binary numbers and recreates the voltage (or current) levels that correspond to that binary number. Then this signal is filtered by a low pass filter. This process is analogous to interpolating between points on a graph, but it can be shown that under certain conditions the original analog signal can be reconstructed exactly from its samples.

3.6 KEY TERMS

- **Analog signal:** An analog signal is any continuous signal for which the time-varying feature of the signal represents some other time-varying quantity, i.e. analogous to another time-varying signal.
- **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.

3.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the sampling theorem.
2. What is natural sampling?
3. Mention the applications of sampling theorem.
4. Differentiate between frequency modulation and phase modulation.
5. State the basic units of communications system.
6. Why is a modulated carrier signal used?
7. Why is there a need for automatic frequency correction?
8. State the types of pulse amplitude modulation.
9. Mention the variations of pulse width modulation.
10. What is signal to noise ratio in pulse system?

Long-Answer Questions

1. Explain flat top or rectangular pulse sampling.
2. Describe the types of modulation techniques.
3. Analyse the FM performance.
4. Discuss the direct method of frequency generation.

3.8 FURTHER READING

Gurung, J.B. 2009. *Signals and Systems*. New Delhi: PHI Learning Pvt. Ltd.

Carlson, A Bruce and Paul B. Crilly. 2010. *Communication System*. Noida: Tata McGraw-Hill Education.

Lathi, B. P. and Zhi Ding. 2010. *Modern Digital and Analog Communication Systems*. Oxford: Oxford University Press.

Kennedy, George and Bernard Davis. 1993. *Electronic Communication Systems*. London: Glencoe/McGraw-Hill.

Roddy, Dennis and John G. Coolen. 1977. *Electronic Communications*. Noida: Pearson Education.

Greenwich, Edwin. 2012. *Analog Integrated Circuits*. Berlin: Springer Science & Business Media.

Millman, Jacob and Christos C. Halkias. 2001. *Integrated Electronics*. Noida: Tata McGraw-Hill Education.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey: Prentice-Hall.

Tanenbaum, Andrew. S. 2002. *Computer Networks*. New Jersey: Prentice-Hall PTR.

Basavaraj, B. and H.N. Shivashankar. 2004. *Basic Electronics*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 4 MICROWAVE TRANSMISSION, COMMUNICATION AND TRANSFERRED ELECTRON DEVICES

Structure

4.0 INTRODUCTION

Microwave transmission is the transmission of information by electromagnetic waves with wavelengths in the microwave range (1 m - 1 mm) of the electromagnetic spectrum. Microwave signals are normally limited to the line-of-sight, so long-distance transmission using these signals requires a series of repeaters forming a microwave relay. It is possible to use microwave signals in over-the-horizon communications using tropospheric scatter, but such systems are expensive and generally used only in specialist roles.

A Gunn diode, also known as a Transferred Electron Device (TED), is a form of diode, a two-terminal semiconductor electronic component, with negative resistance, used in high-frequency electronics. It is based on the 'Gunn Effect' discovered in 1962 by physicist J. B. Gunn. Its largest use is in electronic oscillators to generate microwaves, in applications such as radar speed guns, microwave relay data link transmitters, and automatic door openers. Fading is the fluctuation of a signal's attenuation with numerous variables in wireless communications. Time, geographic location, and radio frequency are among the variables. Fading is frequently shown as a chaotic process. A communication channel that fades is known as a fading channel. Fading in wireless networks can be caused by multipath propagation (also known as multipath-induced fading), weather (especially rain), or shadowing from barriers impacting wave propagation (also known as shadow fading).

In this unit, you will learn about the microwave devices and communication, transferred electron devices and advantages and disadvantages of microwave transmission.

4.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain microwave devices and communication
- State the working of klystron and magnetron
- Describe the transferred electron devices
- Discuss the Gunn effect
- Explain the IMPATT diode and TRAPATT diode
- Describe the advantages and disadvantages of microwave transmission
- Discuss about the antennas used in microwave communication systems

4.2 MICROWAVE DEVICES AND COMMUNICATION

The transfer of information by electromagnetic waves with wavelengths in the microwave range (1 m - 1 mm) of the electromagnetic spectrum is known as microwave transmission. Because microwave signals are generally confined to line-of-sight transmission, long-distance transmission utilizing these signals necessitates the use of a microwave relay, which consists of a series of repeaters. Microwave transmissions can be used in over-the-horizon communications if tropospheric scatter is exploited, however such systems are expensive and are rarely used in specialized tasks.

Although an experimental 40-mile (64-kilometer) microwave telecommunication link across the English Channel was established in 1931, radar research during World War II enabled the capability for microwave communication to be used commercially. The Wireless Set No. 10 was introduced by the British Army during the war, and it used microwave relays to multiplex eight telephone channels over large distances. General Bernard Montgomery was able to keep in touch with his group headquarters in London because to a link across the English Channel. Microwave technology advanced quickly after WWII, resulting in the building of multiple transcontinental microwave relay systems in North America and Europe. These networks were used to transfer television signals for cross-country broadcast, and later, computer data, in addition to carrying thousands of phone calls at a time. During the 1970s and 1980s, communication satellites took over the television broadcast industry, and the introduction of long-distance fibre optic systems in the 1980s and especially the 1990s caused the relay networks to rapidly degrade, with the majority of them being abandoned.

New telecommunication technologies such as wireless networks and direct-broadcast satellites, which transmit television and radio straight into consumers' homes, have exploded in their use of the microwave spectrum in recent years. Larger line-of-sight linkages are once again common for connecting mobile phone towers, however they are rarely structured into long relay chains.

4.2.1 Klystron and Magnetron

Klystron: A vacuum tube used to vibrate and amplify microwave frequency signals is known as a Klystron (also known as a Klystron Tube or Klystron Amplifier). Russell and Sigurd Varian, two American electrical engineers, invented it. The kinetic energy of an electron beam is used in a klystron. In UHF, low-power klystrons are typically employed as oscillators, whereas high-power klystrons are typically utilized as output tubes.

A low-powered klystron can be configured in two ways. The first is a Reflex Klystron low-power microwave oscillator, and the second is a Reflex Klystron low-power microwave amplifier (Two Cavity Klystron or Multi Cavity Klystron).

Magnetron: A magnetron is a device that produces extremely powerful electromagnetic waves. It is a self-excited microwave oscillator and it is also referred to as a crossed field device. The electric and magnetic fields produced inside it are mutually perpendicular to each other, which is why it is called magnetron.

4.2.2 Travelling Wave Vibes

A travelling wave tube is a high-power amplifier that can amplify microwave signals over a wide frequency range. It is a sort of vacuum tube that can operate at frequencies ranging from 300 MHz to 50 GHz. Traveling wave tubes are non-resonant structures that allow the applied RF field to interact with the electron beam continuously throughout the tube's length. As a result, it has a greater working bandwidth.

4.2.3 Velocity Modulation: Basic Principles of Two Cavity Klystron and Reflex Klystron

Reflex Klystron: Oscillations will occur in a klystron if a portion of the output is used as feedback to the input cavity while maintaining the loop gain magnitude at unity. The phase shift of feedback path can be one cycle (2 cycles) or many cycles (multiple of 2 cycles). The cathode is where the electron beam is injected. Then there is the anode, which is also known as a focusing or accelerating anode. The electron beam is narrowed using this anode. The anode is coupled to the positive polarity of DC voltage source.

There is only one cavity in the reflex klystron, which is located near to the anode. For forward-moving electrons, this cavity serves as a buncher cavity, while for backward-moving electrons, it serves as a catcher cavity. The cavity gap is where the velocity and current modulation takes place. The distance 'd' is equal to the gap. The repeller plate is connected to the negative polarity of voltage source V.

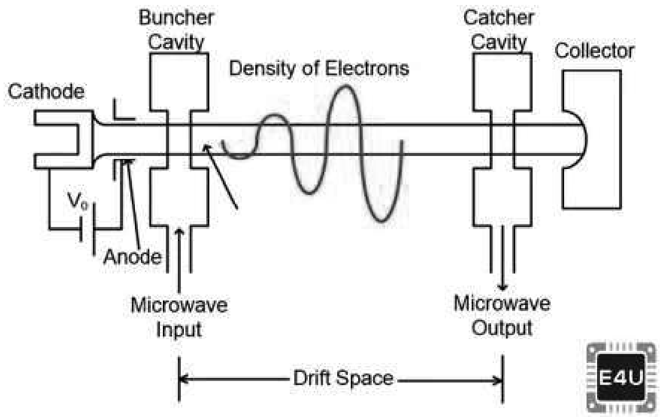


Fig. 4.1 *Klystron Tube*

Working Principle of Reflex Klystron

Reflex velocity and current modulation principle is used by Klystron. The cathode is where the electron beam is injected. The electron beam is accelerated as it travels through the anode. The electron travels through the tube at a constant speed until it reaches the cavity. In the cavity gap, electron velocity is varied, and these electrons try to reach the repeller. The repeller is linked to negative polarity of a voltage source. As a result, it opposes the force of electrons due to the same polarity. As electron kinetic energy drops, the electron pulls back into the cavity. On the way back, all electrons clumped together in one spot. It will eventually be 0 in the repeller space. Following that, due to the creation of the bunch, there will be current modulation. The energy of electrons is transformed into RF, which is then extracted from the cavity. The electron must bunch in the centre of the cavity gap for the klystron to operate at optimal efficiency.

Basic Principles of Two Cavity Klystron

The two cavity klystron works in the same way as the reflex klystron. The two-cavity klystron is constructed as indicated in Figure 4.2.

NOTES

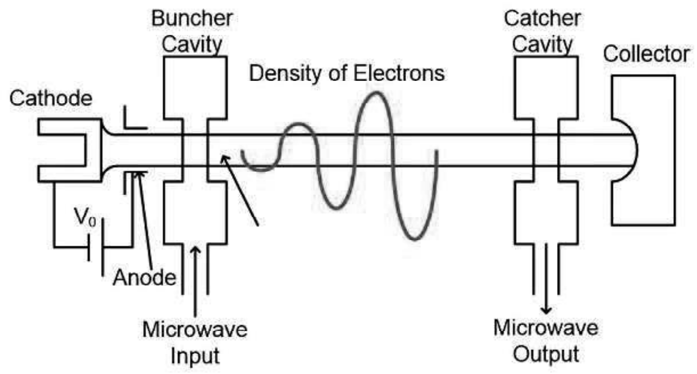


Fig. 4.2 *Two-Cavity Klystron*

Construction of Two-cavity Klystron

There are two cavities, as the name implies: One is a buncher cavity or input cavity, and the other is a catcher cavity or output cavity. Electrons are injected from the cathode and travel at a constant speed to the buncher cavity.

The input RF signal is sent to the input cavity, and the output signal is gathered in the output cavity. Both cavities have a gap, which is referred to as the microwave interaction zone.

The input RF signal in the first cavity modulates the velocity of electrons in the first cavity. This is referred to as velocity modulation. It expresses the electron bunching as well as travelling through the cavity of the catcher. The catcher cavity is where the current is modulated. The microwave fields deplete the kinetic energy of all electrons once they pass through the second cavity. Because of the reduced velocity, they will be retrieved by collectors.

4.2.4 Reflex Klystron Principles of Operation of Magnetrons

A magnetron is a device that produces a high-intensity electromagnetic wave. It is commonly referred to as a self-excited microwave oscillator. It is sometimes referred to a crossed-field gadget.

The name comes from the fact that the electric and magnetic fields produced inside the tube are mutually perpendicular to one another, causing them to cross each other's paths.

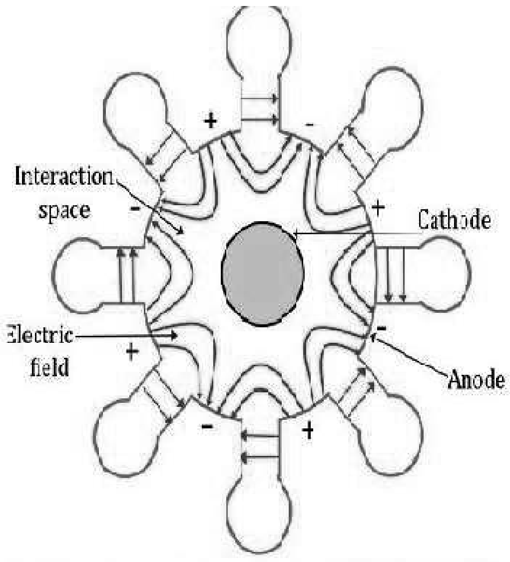
A magnetron is essentially a high-power vacuum tube with numerous cavities. Because an anode is present in the resonant cavity of the tube, it is also known as a cavity magnetron.

The operational concept of a magnetron is that high-power oscillations are formed when electrons interact with the electric and magnetic fields in the cavity.

Magnetrons work on a simple principle.

NOTES

Figure 4.3 shows a magnetron with 8 cavities.



Structure of Magnetron

Fig. 4.3 *Magnetron with 8 Cavities*

Magnetron and its Functions

A DC supply provides excitation to the magnetron's cathode, which results in the appearance of electrons.

When no RF input is available.

Case I: When there is no or no magnetic field

When there is no magnetic field, the electron that emerges from the cathode goes radially towards the anode. This is depicted in the Figure 4.4.

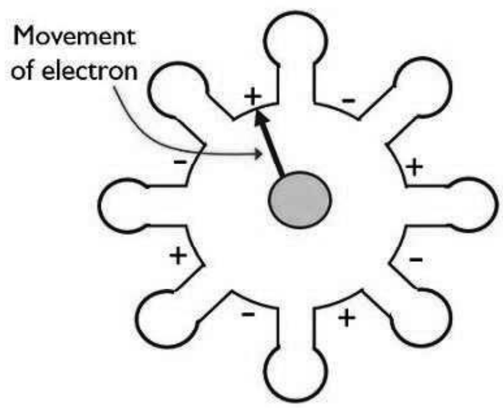


Fig. 4.4 *Absence of Magnetic Field*

This is because the travelling electron is unaffected by the magnetic field and travels in a straight line.

Case II: When a little magnetic field exists inside the magnetron, if a minor magnetic field exists inside the magnetron, the electron escaping from the cathode will deviate

NOTES

slightly from its straight course. As a result, the electron will move in a curved path from the cathode to the anode, as indicated in Figure 4.5.

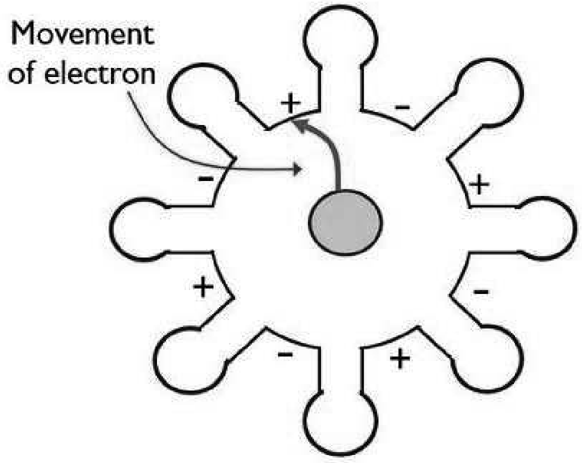


Fig. 4.5 *Presence of Small Magnetic Field*

The electron's mobility is caused by the effect of both electric and magnetic forces on it.

Case III: If the magnetic field is increased further, electrons emanating from the cathode will be strongly deflected by the magnetic field and will graze along the cathode's surface as shown in Figure 4.6.

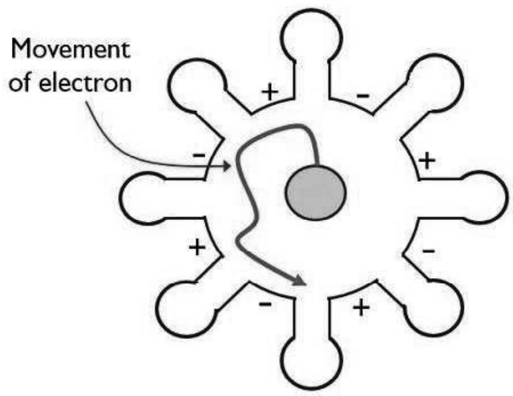


Fig. 4.6 *Presence of Large Magnetic Field*

When an RF field is present.

Case I: When an active RF input is applied to the magnetron's anode, oscillations are created in the magnetron's interaction space. When an electron is emitted from the cathode to the anode, it transfers its energy to the anode, causing it to oscillate. Favoured electrons are the type of electrons that fall into this category. The electrons will have a low velocity in this scenario and will take a long time to get from the cathode to the anode. This can be seen in Figure 4.7.

NOTES

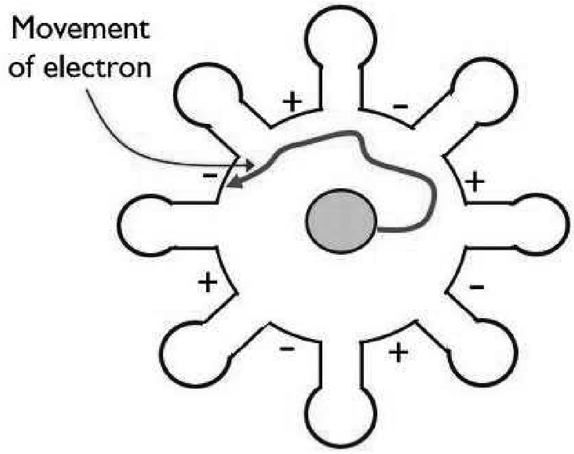


Fig. 4.7 When Moving Electron Release Energy to Oscillate

Case II: In the presence of RF input, a second condition develops. In this situation, the released electron from the cathode absorbs energy from the oscillations as it travels, increasing its velocity as a result.

The electrons will bounce back to the cathode despite reaching the anode, and these electrons are known as unfavoured electrons. The propagation of unfavourable electrons is shown in Figure 4.8.

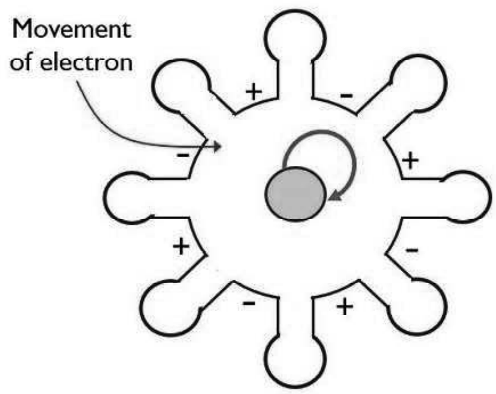


***Fig. 4.8** When Moving Electron Takes Energy from the Oscillations*

Case III: As the RF input is increased, the velocity of the electron emitted while travelling increases in order to catch up to the electron emitted earlier with a lower velocity.

As a result, all electrons that do not require energy from oscillations to move are referred to as preferred electrons. And these favoured electrons form an electron cloud or electron bunch as they travel from the cathode to the anode.

The phase focusing effect is the development of an electron bunch inside the tube.

The electron's orbit is restricted into spokes as a result of this. These spokes rotate according to a fractional value of electron emitted by the cathode till it reaches the anode, supplying energy to oscillations in the process.

NOTES

However, electrons emitted from the cathode region between spokes will absorb the field's energy and return to the cathode extremely fast. However, in relation to the energy provided to the oscillations, this energy is little. This is shown in Figure 4.9.

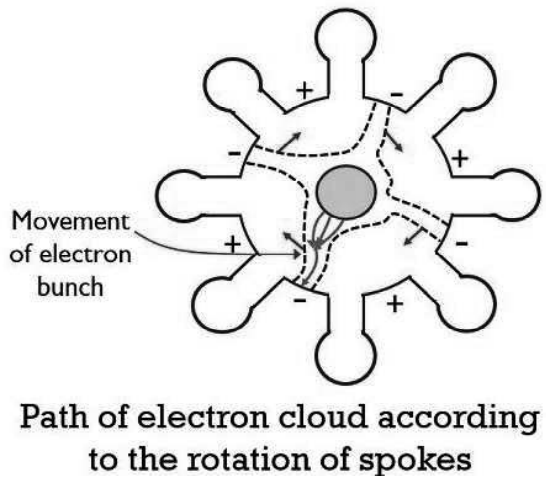


***Fig. 4.9** Path of Electron Cloud According to the Rotation of Spokes*

The field between the gaps in the cavity is enhanced by the flow of these favoured electrons inside the tube. This causes long-term oscillations inside the magnetron, resulting in high output power.

Frequency Pushing and Pulling

The term frequency pushing and pulling refers to the variation in the magnetron's oscillation frequency.

Variation in the velocity of electrons flowing from cathode to anode occurs when the voltage provided at the magnetron's anode is changed. As a result, the frequency of oscillations changes.

As a result, we may say that frequency pushing occurs when the magnetron's resonance frequency varies due to changes in the anode voltage.

The change in resonant frequency is occasionally caused by the magnetron's load impedance changing. When the change is purely resistive or reactive, the load impedance changes. Frequency pulling is the term for this type of frequency variation. This frequency volatility can be reduced with a stable power supply.

Check Your Progress

1. Define the term microwave transmission.
2. What is klystron?
3. How will you define Travelling Wave Tube (TWT)?
4. What do you mean by the magnetron?
5. What is frequency pushing and pulling?

NOTES

4.3 TRANSFERRED ELECTRON DEVICES

During the last few decades, the use of two-terminal semiconductor devices at microwave frequencies has increased. At higher microwave frequencies, the CW, average, and peak power outputs of these devices are significantly higher than those of the best power transistor. Negative resistance is a typical feature of all active two-terminal solid-state devices. Over a wide range of frequencies, the true part of their impedance is negative. The current through the resistance and the voltage across it are in phase in a positive resistance. The voltage drop across a positive resistance is positive, and the resistance dissipates a power of (2 R). However, in a negative resistance, the current and voltage are 180° out of phase. The voltage drop across a negative resistance is negative, and the power supply associated with the negative resistance generates a power of (2 R). In other words, positive resistances (passive devices) consume power, whereas negative resistances (active devices) generate power (active devices). The Transferred Electron Devices (TEDs) are discussed in this unit. Microwave transistors and Transferred Electron Devices (TEDs) have significant variances. Transistors use junctions or gates to operate, but TEDs are bulk devices with no junctions or gates. Most transistors are made of elemental semiconductors like silicon or germanium, whereas TEDs are made of compound semiconductors like Gallium Arsenide (GaAs), Indium Phosphide (InP), or Cadmium Telluride (CdTe). Transistors use 'Warm' electrons whose energy is not much higher than the thermal energy of electrons in the semiconductor (0.026 eV at ambient temperature), but TEDs use "hot" electrons whose energy is significantly higher than the thermal energy. The theory and technology of transistors cannot be applied to TEDs because of these basic distinctions.

4.3.1 Gunn Effect

In 1963, J. B. Gunn discovered the Gunn effect in an n-type GaAs bulk diode, an effect best explained by Gunn, who published multiple publications on his findings.

When the electric field is changed from zero to a threshold value, the carrier drift velocity increases linearly from zero to a maximum, according to Gunn's observations. The drift velocity is reduced and the diode exhibits negative resistance when the electric field exceeds the threshold value of 3000 V/cm for n-type GaAs. Figure 4.10 depicts the situation.

NOTES

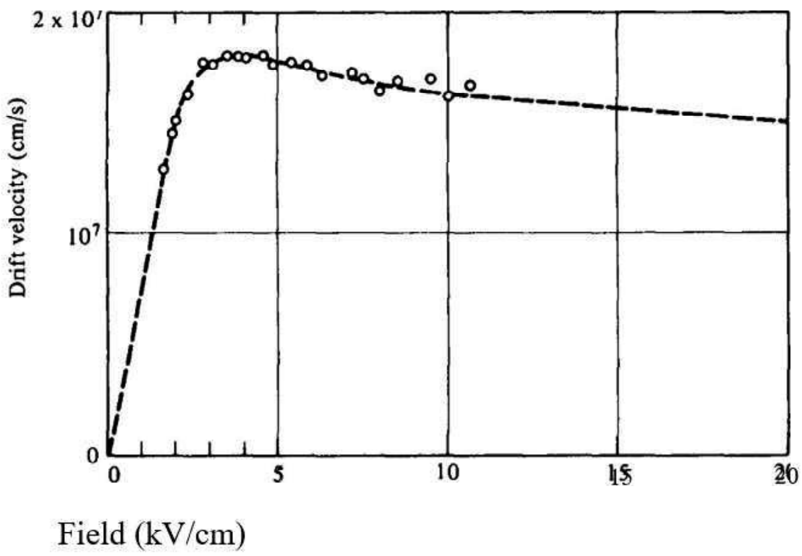


Fig. 4.10 *Electron Drift Velocity in n-type GaAs vs Electric Field.*

According to the Ridley—Watkins—Hilsum theory, a semiconductor's band structure must meet three characteristics in order for it to display negative resistance.

1. The separation energy between the bottoms of the lower and upper valleys must be several orders of magnitude more than the thermal energy (approximately 0.026 eV) at ambient temperature. This suggests that AE > 0.026 eV or AE > kT.
2. The gap energy between the conduction and valence bands must be lower than the separation energy between the valleys. Due to the production of hole-electron pairs, the semiconductor will break down and become extremely conductive before the electrons can begin to migrate to the upper valleys.
3. In the lower valley, electrons must have high mobility, small effective mass, and low density of state, whereas in the higher valley, electrons must have low mobility, big effective mass, and high density of state. In other words, electron velocities (dE/dk) in the lower valleys must be substantially higher than those in the upper valleys.

Silicon and germanium, the two most useful semiconductors, do not meet all of these requirements. These criteria are met by several compound semiconductors, such as Gallium Arsenide (GaAs), Indium Phosphide (InP), and Cadmium Telluride (CdTe). Others do not, including Indium Arsenide (InAs), Gallium Phosphide (Gap), and Indium Antimonide (InSb). A hypothetical current versus field characteristic of a two-valley semiconductor is shown in Figure 4.11

NOTES

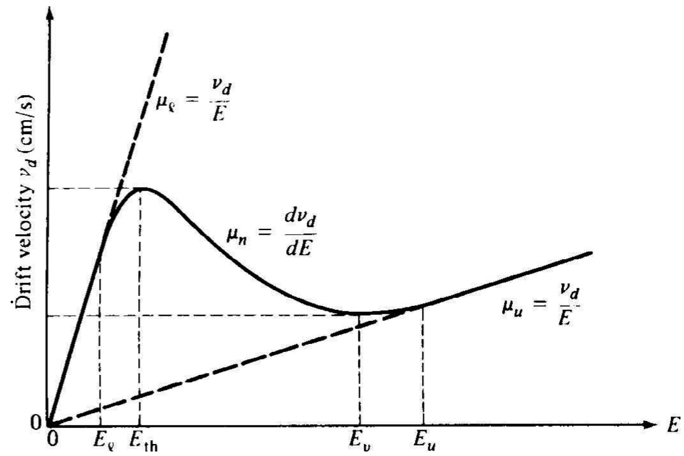


Fig. 4.11 *Electron drift velocity versus electric field.*

4.3.2 Principle of Operations

A high-field avalanche zone propagates through the diode, filling the depletion layer with a dense plasma of electrons and holes that are trapped in the low-field region behind it.

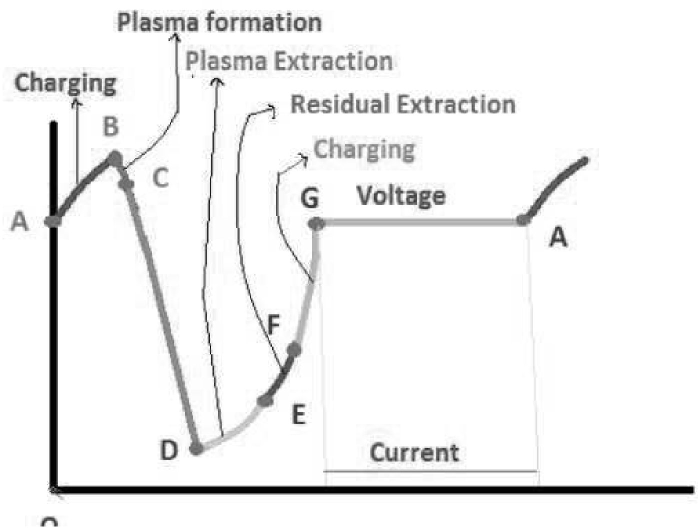


Fig. 4.12 *Typical Voltage Waveform for an Avalanche p+-n-n+ Diode*

Figure 4.12 shows a typical voltage waveform for an avalanche p+-n-n+ diode working in the TRAPATT mode with an assumed square wave current drive. A Point A has a homogeneous electric field throughout the sample that is large but less than the value required for avalanche breakdown.

J=6s represents the current density. The diode current is turned on at point A at this precise moment in time.

Because the charge carriers present are thermally generated, the diode charges up like a linear capacitor at first, pushing the electric field magnitude past the breakdown voltage.

The particle current exceeds the external current when a significant number of carriers are created, and the electric field is lowered throughout the depletion zone, causing the voltage to fall. The curve from point B to point C represents this part of the cycle.

The electric field is sufficiently large during this time interval for the avalanche to proceed, resulting in a concentrated plasma of electrons and holes. The voltage is reduced to D.

The holes created by the avalanche reach the p+ contact quickly and do not participate in the process, but the electrons are released into the N region and do not join with either the doner or the holes. The electrons drift over the N area at their maximum velocity, while current continues to flow through the external circuit through which they are travelling.

The ac voltage is at its negative peak when this current pulse reaches the cathode terminal, and the second delay of 90 degrees has occurred. The length of time is determined by the velocity and thickness of the highly doped N+ layer.

Because the overall plasma charge is considerable compared to the charge per unit time in the external current, removing the plasma takes a long time. Plasma is removed at point E. The voltage rises from point E to point F as the residual charge is eliminated. All of the charge generated inside has been eliminated at point F. The diode charged up like a fixed capacitor from point F to G. The diode current goes to zero for half a period at point G, and the voltage remains constant VA until the current returns and the cycle repeats.

4.3.3 Modes of Operation

Modes of Gunn Oscillation (1012/cm^2 (no L) 1014/cm^2)

The product of doping and length (n_oL) in most Gunn-effect diodes is larger than 10 12/cm^2. Gunn's observed mode, on the other hand, has a substantially lower product n_oL. The spacecharge perturbations in the specimen increase exponentially in space and time when the product of n_oL is more than 10 h^2/cm^2 in GaAs. As a result, a high-field domain forms and moves from the cathode to the anode.

Gunn detailed how Gunn oscillators behaved in various circuit setups. The period of oscillation is the time it takes for the domain to drift from the cathode to the anode when the circuit is mostly resistive or the voltage across the diode is constant. This is not a common mode for microwave applications. Negative conductivity devices, such as high-Q resonant microwave cavities, are typically used in resonant circuits. The frequency of a resonant circuit diode can be set to a range of around an octave without sacrificing efficiency.

The typical Gunn domain mode (or Gunn oscillation mode) is activated when the electric field is greater than the threshold field ($E > E_{th}$), as previously stated. As illustrated in Fig. 4.13, the high-field domain wanders along the specimen until it hits the anode or the low-field value falls below the sustaining field E_s required to maintain v_s. GaAs has a sustained drift velocity of 107 cm/s. The Gunn oscillation mode has three domain modes because the electron drift velocity changes with the electric field.

Mode of transit in the time domain (fL 107 cm/s). The high-field domain is stable when the electron drift velocity equals the sustaining velocity. In other terms, $C_d = v_s$ = 107 cm/s is the electron drift velocity.

The oscillation period is then equal to the transit time, To = T. Figure 4.13 depicts the situation (a). Because the current is only gathered when the domain reaches the anode, the efficiency is below 10%.

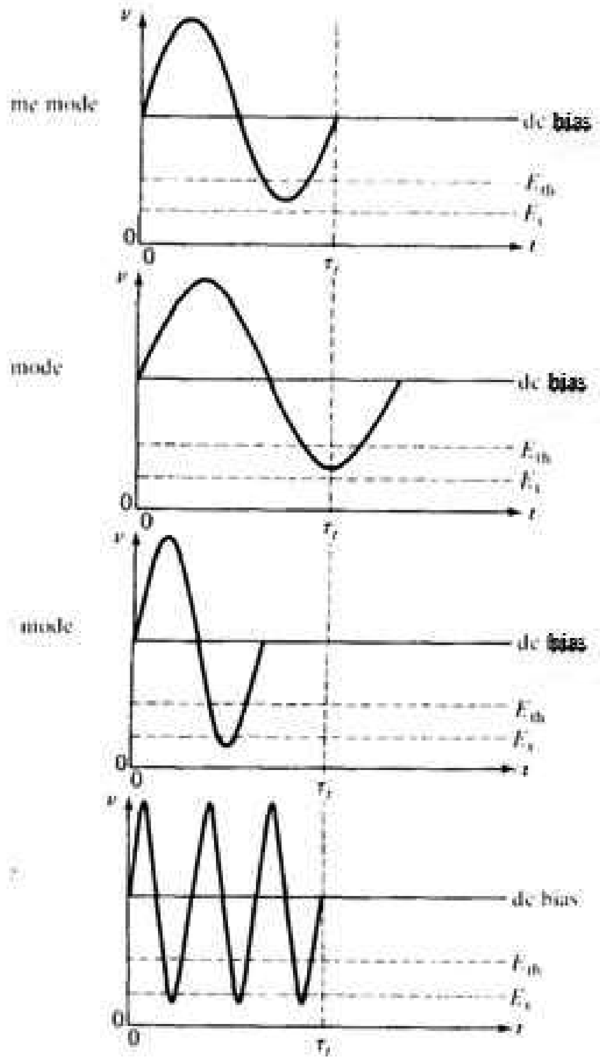


Fig. 4.13 *Gunn Oscillation*

Stable Amplification Mode (n_o' < 10^{12}/cm^2): When the device's n_oL product is smaller than about 10^{12}/cm^2, it exhibits amplification rather than spontaneous oscillation at the transit-time frequency. The negative conductance is used without domain creation, resulting in this condition. The number of carriers available for domain construction within the transit time is insufficient. As a result, signal amplification at the transit-time frequency is possible. Thin and Barber were the first to notice this mode. Uenohara further showed that there are different forms of amplification based on the device's fL product, as illustrated in Figure 4.14.

NOTES

The many modes of functioning of Gunn diodes can be categorised based on the times at which different activities take place.

There are some limitations to the LSA mode. It is extremely sensitive to changes in load, temperature, and doping levels. In order to avoid domain development, the RF circuit must also allow the field to grow up quickly.

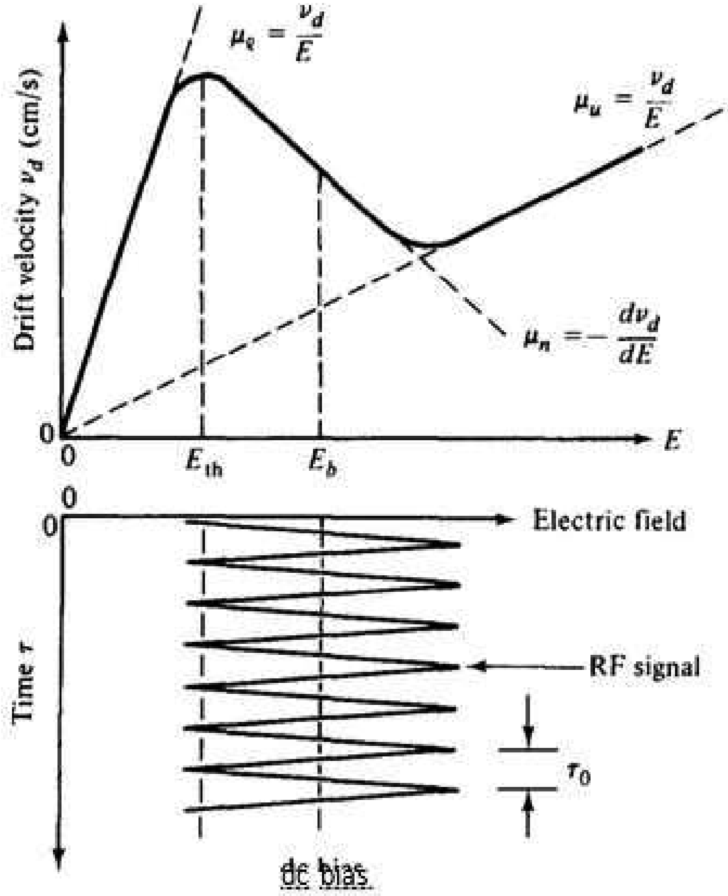


Fig. 4.14 *Amplification Mode*

LSA mode of operation

LSA oscillator can be simply written as

$$P = \eta (ME_{th}L)(n_0 e v_0 A)$$

Where h = dc-to-RF conversion efficiency (primarily a function of material and circuit considerations)

v_0 = operating voltage

I_o = operating current

M = multiple of the operating voltage above negative-resistance threshold voltage

E_{th} = threshold field (about 3400 V/cm)

L = device length (about 10 to 200 gm)

n_o = donor concentration (about 10^{15} e/cm^3)

e = electron charge (1.6 x 10 -19 C) v_o =average carrier drift velocity (about 10^7 cm/s)

A = device area (about 3 >< 10 -4 to 20 x 10 - ^{4} cm^2)

The intended operating frequency f_o determines n_o for an LSA oscillator, therefore peak power output is directly proportional to the volume (LA) of the device length L multiplied by the area A of the active layer for a correctly built circuit. It is impossible to raise active volume forever. This is related to electrical wavelength and skin-depth issues in the theoretical limit. However, available bias is limited in the practical limit because the coupling between the lower and higher valleys in Inp is weaker than in GaAs. The additional energy loss mechanism provided by the middle-valley energy level is necessary to avoid breakage caused by the high energies accumulated by the lower-valley electrons as a result of the weak coupling. Figure 4.15 illustrates this.

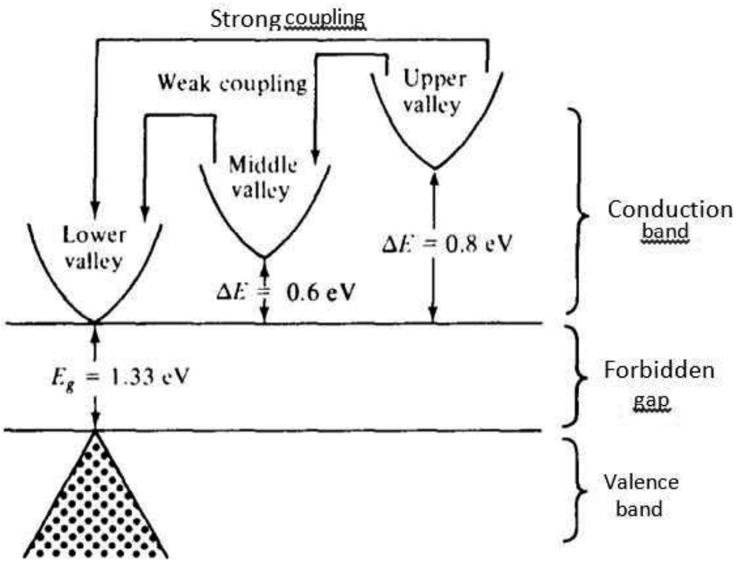


Fig. 4.15 *Three-valley-model energy level for Inp diode.*

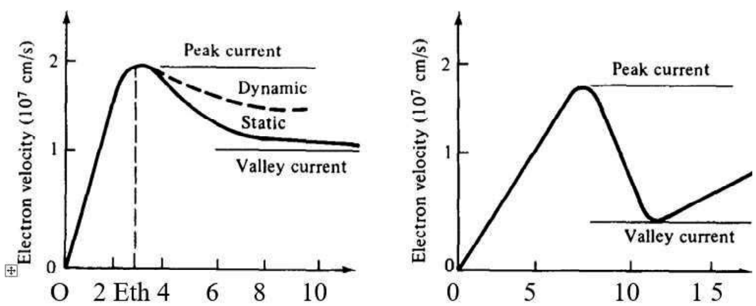


Fig. 4.16 *Electric field (kV/cm)*

To prevent collapse, the lower valley is lightly connected to the middle valley but strongly coupled to the top valley. This guarantees that electrons concentrate in the central valley under normal working conditions. The thermal excitation of electrons has less influence on Inp because the energy difference between the lower valley and the nearby energy levels is bigger, and the degradation of its peak-to-valley current ratio is nearly four times less than in GaAs.

4.3.4 IMPATT Diode

The IMPATT diode's full name is IMPATT ionisation Avalanche Transit-Time. This is a microwave diode with an extraordinarily high output power. At microwave frequencies, it is typically utilised as an amplifier and oscillator. The IMPATT diode operates at frequencies ranging from 3 to 100 GHz.

This diode generates negative resistance characteristics and hence functions as an oscillator for generating signals at microwave frequencies. This is due to the transit time effect as well as the impact ionisation avalanche effect. Single drift and double drift are the two varieties of IMPATT diodes that can be classified. P+NN+, P+NIN+, N+PIP+, and N+PP+ are single drift devices.

When the P+N junction of the P+NN+ device is linked in reverse bias, an avalanche breakdown occurs, causing the region of P+ to inject into NN+ at a saturation velocity. However, the holes injected from the NN+ area do not drift.

P+PNN+ is the greatest example of a double drift device. When the PN-junction is biased close to an avalanche breakdown, electrons drift via the NN+ zone and holes drift through the PP+ region, resulting in double drift devices.

IMPATT Diode Construction and Working

The IMPATT diode's construction is given below. This diode has four regions: P+-N-I-N+, P+-N-I-N+, P+-N-I-N+, P+-N- The PIN diode and IMPATT have the identical structure, however the IMPATT generates an avalanche current by using an extraordinarily high voltage gradient of about 400KV/cm. Various materials, such as Si, GaAs, InP, or Ge, are commonly utilised to construct it. However, GaAs is chosen because it has a lower noise behaviour.

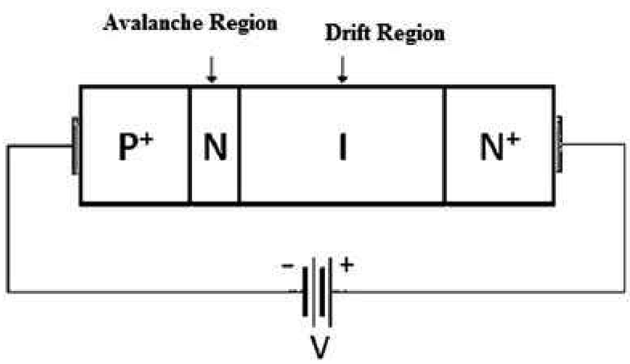


Fig. 4.17 *IMPATT Diode Construction*

This diode has a slightly different structure than a standard diode since a normal diode will break down in an avalanche situation. As a result of the massive

NOTES

amount of current generated, heat is generated within it. As a result, at microwave frequencies, structural deviation is mostly used to generate RF signals. When compared to other microwave diodes, the output of the IMPATT circuit is steady and relatively high.

IMPATT Diode Circuit: The IMPATT diode is used in the circuit below. This type of diode is typically employed at frequencies above 3 GHz. It has been shown that when a tuned circuit is given a voltage in the breakdown voltage zone approaching the IMPATT, oscillation occurs.

This diode, unlike other diodes, has a negative resistance and can generate a large amount of power, generally ten watts or more depending on the device. A current limiting resistor can be used to operate this diode from a supply. The value of this restricts current flow to the minimum amount required. To isolate the DC from the RF signal, the current is sent through an RF choke.

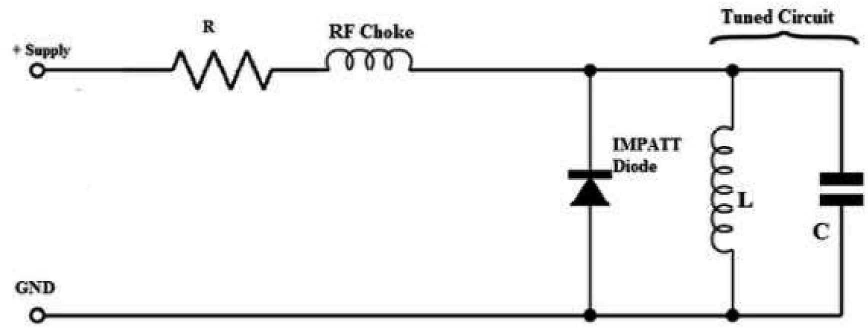


Fig. 4.18 *IMPATT Diode Circuit*

The IMPATT microwave diode is located outside of the tuned circuit, but it can also be found within a waveguide cavity that provides the required tuned circuit. The circuit will swing when the voltage supply is applied.

4.3.5 TRAPATT Diode

Trapped plasma avalanche triggered transit mode is abbreviated as TRAPATT.

It is a high-efficiency microwave generator that can operate at frequencies ranging from a few hundred megahertz to a few gigahertz.

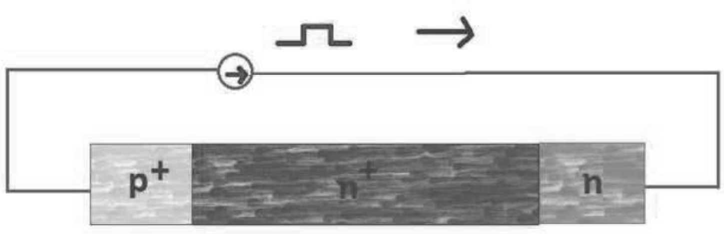


Fig. 4.19 *Silicon n+-p-p+ or p+-n-n+ diodes with a n type depletion region*

Silicon n+-p-p+ or p+-n-n+ diodes with a n type depletion region width ranging from 2.5 to 12.5 m are commonly used in high peak power diodes.

The depletion zone is usually doped in such a way that the diodes are well punched through.

At 2.5 to 7.5 m, the device P+ area is kept as thin as feasible.

The diameter of TRAPATT diodes ranges from 50 m for w operation to 750 m at lower frequencies for high peak power devices.

4.4 ADVANTAGES AND DISADVANTAGES OF MICROWAVE TRANSMISSION

Line-of-sight transmission is used in microwave transmission. The receiving station and the transmitter station must be visible to each other. Depending on the local geography, this places a restriction on the distance between stations. Due to the curvature of the Earth, the line of sight to the horizon is usually only 5 kilometres. Repeater stations must be built across the country so that the data stream can hop, skip, and leap. Microwaves are electromagnetic waves with frequencies ranging from 1 to 300 GHz. Microwaves are unidirectional, therefore they can be carefully concentrated when transmitted by an antenna. This necessitates the alignment of the transmitting and receiving antennas. The unidirectional property offers a clear benefit. An antenna pair can be aligned without interfering with another antenna pair that is aligned. Microwaves travel in a straight path. Because the antenna-mounted towers must be in direct sight of each other, towers that are far apart must be extremely tall. The curvature of the earth, as well as other obstructing impediments, prevent two short towers from communicating using microwaves. Because high-frequency microwaves cannot penetrate walls, repeaters are frequently required for long-distance communication. This mode of transmission employs a parabolic dish antenna and a horn antenna.

Microwaves operate at frequencies ranging from 3 to 10 GHz. Due to the huge bandwidth, they are able to transport large amounts of data.

Advantages of Microwave

- They do not require the purchase of right-of-way between towers and may carry large amounts of data due to their high operating frequencies.
- Low-cost land acquisition: Each tower takes up a little space.
- Signals with a short wavelength and high frequency necessitate a compact antenna.

Disadvantages of Microwave

- Solid objects, such as birds, rain, snow, and fog, cause attenuation.
- Reflected off of flat surfaces such as metal and water.
- Split (diffracted) around solid things.
- The atmosphere refracts the beam, causing it to be projected away from the receiver.

NOTES

4.4.1 Free Space Loss

The loss in signal intensity of an electromagnetic wave caused by a line-of-sight passage across open space (typically air) with no objects nearby to induce reflection or diffraction is known as Free-Space Path Loss (FSPL) in telecommunication. It excludes things like the gain of the antennas used at the transmitter and receiver, as well as any loss caused by technical flaws. The article on link budget contains a description of these losses.

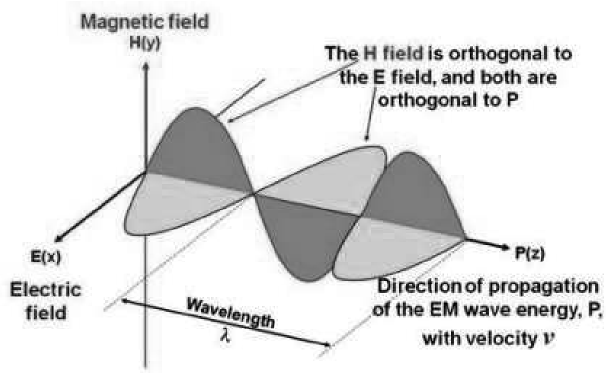


Fig. 4.20 Signal Intensity of an Electromagnetic Wave

Free-Space Path Loss Formula

The square of the distance between the transmitter and receiver, as well as the square of the radio signal's frequency, determines free-space path loss. The signal disperses with distance in any sort of wireless communication. As a result, the further away an antenna with a fixed area is from the transmitting antenna, the less signal power it receives. This is the most common mode of signal loss in satellite communications. A transmitted signal attenuates over distance because it is dispersed across a greater and larger area, even if no other sources of attenuation or damage are assumed. Free space loss is a type of attenuation that can be expressed as a ratio of the radiated power to the power received by the antenna or as a decibel value by taking 10 times the log of that ratio. Free space loss is zero for the ideal isotropic antenna.

$\frac{P_r}{P_t} = D_t D_r \left(\frac{\lambda}{4\pi d}\right)^2$ is the equation for FSPL.

Only in the distant field, where spherical spreading can be assumed, is this equation correct. It does not remain in close proximity to the transmitter.

Decibels of Free-Space Route Loss

In terms of dB: l is a convenient way to express FSPL.

NOTES

$$L_{dB} = 10\log\frac{P_t}{P_r} = 20\log\left(\frac{4\pi d}{\quad}\right)$$

$$\begin{aligned}\text{FSPL(dB)} &= 10\log_{10}\left(\left(\frac{4\pi}{c}df\right)^2\right)\\ &= 20\log_{10}\left(\frac{4\pi}{c}df\right)\\ &= 20\log_{10}(d) + 20\log_{10}(f) + 20\log_{10}\left(\frac{4\pi}{c}\right)\\ &= 20\log_{10}(d) + 20\log_{10}(f) - 147.55\end{aligned}$$

4.4.2 Propagation of Microwaves

Microwaves are electromagnetic waves with a frequency higher than that of television transmissions. The wavelength of these waves is only a few centimetres. These waves bend properly from the corners of the objects in their path due to their short wavelength. These are mostly utilised to generate fixed-direction beam signals. As a result, it is obvious that these waves are capable of detecting flying objects. If these are capable of detecting flying objects, they can also measure their speed. Microwaves travel at the same speed as light.

The transmission and receiving antennas must be able to see each other if these microwaves are to be used on Earth's surface. Repeaters are any antennas that are used between the transmission and receiving antennas. The purpose of a repeater is to boost the strength of a weak signal. Repeaters magnify the signal once more before passing it on to the next repeater in the chain. The expense of installing repeaters is too expensive. This technology cannot be used to cover the entire earth's surface. Another factor is because the earth's surface is covered in oceans. As a result, implementation is impossible. Communication satellites were utilised to overcome this constraint. In 1958, the United States launched the first communication satellite.

4.4.3 Atmospheric Effects of Propagation

The earth's atmosphere has properties that influence radio wave propagation. Because these effects occur at various points in the atmosphere, it is worthwhile to briefly explain the structure of the earth's atmosphere.

Given below are the zones of the world's environment:

- **Lower atmosphere** – This piece of the environment, which reaches out from the world's surface to somewhere in the range of 7 and 17 kilometers above it, is the most tempestuous in light of the fact that there is a great deal of warm development of air. The temperature of the air drops drastically as stature ascends because of sweeping cooling: when elevation rises, gases can extend because of an abatement in pressure, bringing down the temperature of the gas.

NOTES

- **Stratosphere** - The stratosphere begins in the troposphere and ends at a height of around 50 kilometres. This is where the ozone layer is found.
- **Mesosphere** — This layer of the atmosphere exists between 50 and 80-85 kilometres above the Earth's surface.
- **Ionosphere** - The ionosphere is a zone of highly charged particles that results from the ionisation of atmospheric atoms/molecules by solar radiation. It is located beyond the mesosphere and extends to 640+ km. Below 40 MHz, the ionosphere plays a significant influence in radio wave propagation.
- **Exosphere** — Exosphere goes beyond the mesosphere that extends for around 10,000 kilometres.

Refraction and reflection are the two most important atmospheric effects on radio wave transmission. The troposphere and the ionosphere are both capable of refraction. Tropospheric refraction occurs when the atmosphere's refractive index falls with altitude, causing waves to bend back toward the earth. Ionospheric refraction, on the other hand, is caused by the electrical characteristics of plasmas generated in the ionosphere as a result of atmospheric ionisation. If the frequency is low enough, reflection off the ionosphere is also feasible. We will call the former atmospheric refraction and the latter ionospheric propagation to distinguish between the two phenomena. Due to absorption by air molecules, water molecules, and precipitation, radio signals will be attenuated in the atmosphere (rain).

4.4.4 Fresnel Zone Problem

Blurring and blurring sources are an issue in the Fresnel Zone. Blurring is the variation in received signal brought about by changes in the transmission medium or pathways. Blurring is affected by various things. Blurring in a given circumstance is impacted by environmental conditions like downpour, lightning, etc. Blurring in a portable setting is subject to obstructions along the way that change over the long run. The sent sign is exposed to different transmission impacts because of these hindrances.

4.4.5 Ground Reflection

Two Ray Ground Reflection Model: In Friis propagation model, the line-of-sight (LOS) path between the transmitter and the receiver is considered. The expression for the received power becomes complex if the effect of reflections from the earth surface are included in the modeling. Moreover, a single reflected path is added to the line-of-sight path in the **two ray ground reflection model**, as shown in Figure 4.21. This model considers the phenomenon of reflection from the ground and the antenna heights above the ground. The ground surface is characterized by **reflection coefficient** which depends on the physical properties of the surface and the type of wave polarization.

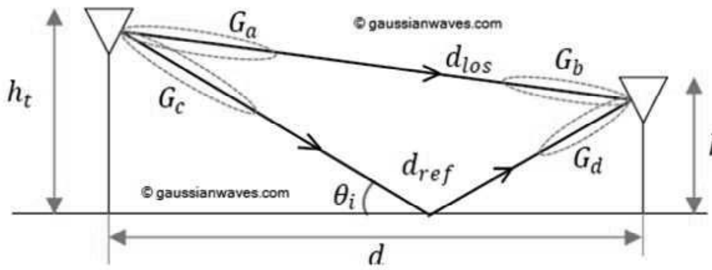


Fig. 4.21 *Two Ray Ground Reflection Model*

The received signal is made of two components, LOS ray that goes through free space from the transmitter and a reflected ray starting from the ground surface. The distances covered by LOS and reflected rays are determined by using the following equation.

$$d_{los} = \sqrt{d^2 + (h_t - h_r)^2}$$
$$d_{ref} = \sqrt{d^2 + (h_t + h_r)^2} \qquad \ldots\ldots(4.1)$$

Depending on the phase difference f between the LOS ray and reflected ray, the received signal may suffer constructive or destructive interference. Therefore, this model is also called as **two ray interference model**.

$$\phi = \frac{2\pi\,(d_{ref} - d_{los})}{\lambda} \qquad \ldots\ldots(4.2)$$

where λ is the radiating wave's wavelength, which may be determined from the transmission frequency. The power of received signal can be represented as under the large-scale assumption as

$$P_r = P_t \left[\frac{\lambda}{4\pi}\right]^2 \left| \frac{\sqrt{G_{los}}}{d_{los}} + R\frac{\sqrt{G_{ref}}e^{-j\phi}}{d_{ref}} \right|^2 \qquad \ldots..(4.3)$$

where $\sqrt{G_{los}} = \sqrt{G_a G_b}$ is the product of antenna field patterns in the LOS direction. $\sqrt{G_{los}} = \sqrt{G_c G_d}$ is the product of antenna field patterns in the reflected path is

NOTES

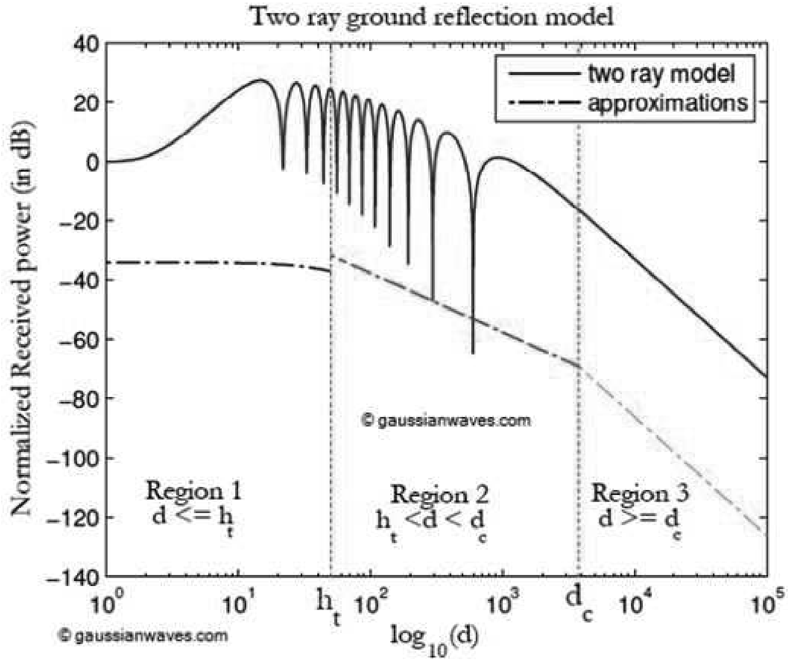


Fig. 4.22 *Distance vs Received Power for Two Ray Ground Reflection Model*

4.4.6 Fading Sources

Fading is the fluctuation of a signal's attenuation with numerous variables in wireless communications. Variables include time, geographic location and radio frequency. Fading is frequently shown as a chaotic process. A communication channel that fades is known as a fading channel. Fading in wireless networks can be caused by multipath propagation (also known as multipath-induced fading), weather (especially rain), or shadowing from barriers impacting wave propagation (also known as shadow fading).

Types of Fading Sources

Slow Versus Fast Fading: The terms slow and fast fading refer to the rate at which the magnitude and phase change is imposed by the channel on the signal changes. The coherence time is a measure of the minimum time required for the magnitude change or phase change of the channel to become uncorrelated from its previous value.

- **Slow Fading:** Slow fading occurs when the channel's coherence time is longer than the application's delay requirement. The channel's amplitude and phase change can be considered constant over the course of use in this regime. Shadowing occurs when a significant object, such as a hill or a huge structure, obscures the main signal route between the transmitter and the receiver, causing slow fading. According to the log-distance path loss model, the received power shift induced by shadowing is frequently described using a log-normal distribution with a standard deviation.

- **Fast Fading:** When the channel's coherence time is short compared to the application's delay requirement, fast fading happens. The amplitude and phase change imposed by the channel in this scenario changes significantly over time.

4.4.7 Detectors and Components

Detectors: Amplitude-Reasonable (AC) location is a valuable identification approach that improves on beneficiary plan while keeping a Steady Image Mistake Rate (SER). Accordingly, this examination considers AC identifier plan and SER investigation over channels utilizing M-Ary Plentifulness Shift Keying (MASK) regulation. We compute the SER of the ideal, close ideal, and imperfect AC identifiers, just as the cognizant, noncoherent, and heuristic AC indicators. Besides, for single and numerous getting receiving wire, the insightful SER of the heuristic not set in stone utilizing two particular ways. The main methodology yields a shut structure insightful articulation for the SER, which is likewise used to determine a straightforward recipe for the asymptotic SER at high sign to-commotion proportions. The subsequent methodology yields a shut structure insightful articulation for the SER, which is likewise used to determine a basic equation for the asymptotic SER at high Sign to-Clamor Proportions (SNRs). The SER of the AC and reasonable MASK identifiers are equivalent, particularly for high Rician K-factors and few getting receiving wires, as per the determined scientific and reenactment results. Besides, the outcomes uncover that the SER of the ideal AC locator is equivalent to the SER of the reasonable indicator. The best AC identifier intricacy, then again, is restrictively high, particularly at high SNRs. With the exception of the twofold ASK circumstance at low SNRs, the heuristic AC identifier significantly outflanks the ideal noncoherent locator as a rule. Moreover, the gathered discoveries uncover that the heuristic AC locator is stage commotion safe, beating the sound indicator in conditions when the framework is exposed to critical stage clamor.

Microwave Components: Microwave transistors and various types of diodes are examples of microwave components.

Microwave Transistors

Using microwave frequencies necessitates the development of special transistors. As a result, silicon n-p-n transistors with sufficient power at microwave frequencies have been created for microwave applications. They are typically 5 watts at a frequency of 3GHz with a gain of 5dB. Figure 4.23 depicts a cross-sectional view of such a transistor.

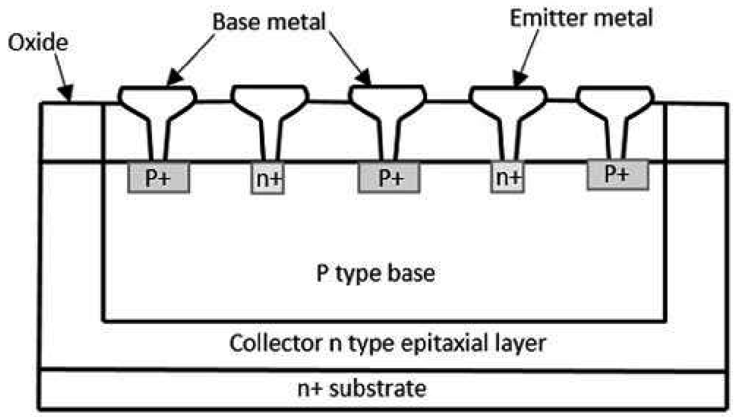


Fig. 4.23 n-p-n Silicon Double Diffused Epitaxial Transistor

Solid State Devices

Solid state devices can be classified as follows:

- **Depending upon their electrical behaviour** – Non-linear resistance, non-linear reactance, negative resistance type and controllable impedance type.
- **Depending upon their construction** – Point contact diodes, Schottky barrier diodes, metal oxide

The types of diodes which we have discussed have many uses such as amplification, detection, power generation, phase shifting, down conversion, up conversion, limiting modulation, switching, etc.

Varactor Diode

Varactor diode is a semi-conductor device in which the junction capacitance can be varied as a function of the reverse bias of the diode.

Schottky Barrier Diode

These diodes are mostly used for microwave detection and mixing. A semi-conductor pellet is mounted on a metal base. A spring loaded wire is connected with a sharp point to this silicon pellet. This can be easily mounted into coaxial or waveguide lines.

Gunn Effect Devices

J B Gunn discovered periodic fluctuations of current passing through the n-type GaAs specimen when the applied voltage exceeded a certain critical value. In these diodes, there are two valleys, L & U valleys in conduction band and the electron transfer occurs between them, depending upon the applied electric field. This effect of population inversion from lower L-valley to upper U-valley is called Transfer Electron Effect and hence these are called as Transfer Electron Devices.

4.4.8 Antennas Used in Microwave Communication System

Given below are the types of antennas used in microwave communication system:

1. **Reflector antennas:** Using a parabolic reflector, which is a curved area with a parabolic shape to direct radio waves, parabolic antennas are highly visible because of their dish shape.
2. These types of antennas are especially beneficial for applications that require high directivity and they are best used for point-to-point communication and in radio telescopes because of their high-end gain.
3. **Horn Antennas:** This type of antenna has a waveguide and flared end walls that form a structure resembling a megaphone. Horn antennas are used to transmit frequencies above 300MHz and ultra-high frequencies. They are also useful in calibrating antennas for other equipment such as automatic door openers because they have the ability to estimate the gain of other antennas near them.
4. **Lens Antennas:** A lens antenna directs or collects microwave radiation using a lens.
5. **Array antennas:** An array antenna is a high-gain antenna made up of a number of smaller antenna elements arranged in a grid.
6. **Leaky Wave Antenna:** A leaky wave antenna gets its radiation from a leaking transmission line.

Check Your Progress

6. What is the use of two-terminal semiconductor devices at microwave frequencies?
7. How will you define Gunn effect?
8. Define IMPATT diode.
9. Write a short note on TRAPATT diode.
10. How far can you see due to the curvature of the earth?
11. What affects blurring in Fresnel zone?
12. Define fading channel.
13. Name the types of fading sources.
14. What is varactor diode?
15. Define array antennas and leaky wave antenna.

4.5 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The transfer of information by electromagnetic waves with wavelengths in the microwave range (1 m - 1 mm) of the electromagnetic spectrum is known as microwave transmission.

2. A vacuum tube used to vibrate and amplify microwave frequency signals is known as a Klystron (also known as a Klystron Tube or Klystron Amplifier). Russell and Sigurd Varian, two American electrical engineers, invented it. The kinetic energy of an electron beam is used in a klystron. In UHF, low-power klystrons are typically employed as oscillators, whereas high-power klystrons are typically utilized as output tubes.
3. A travelling wave tube is a high-power amplifier that can amplify microwave signals over a wide frequency range. It's a sort of vacuum tube that can operate at frequencies ranging from 300 MHz to 50 GHz. Traveling wave tubes are non-resonant structures that allow the applied RF field to interact with the electron beam continuously throughout the tube's length. As a result, it has a greater working bandwidth.
4. A magnetron is a device that produces a high-intensity electromagnetic wave. It is commonly referred to as a self-excited microwave oscillator. It's sometimes referred to as a crossed-field gadget.
5. The term frequency pushing and pulling refers to the variation in the magnetron's oscillation frequency.
6. The use of two-terminal semiconductor devices at microwave frequencies has increased. At higher microwave frequencies, the CW, average, and peak power outputs of these devices are significantly higher than those of the best power transistor. Negative resistance is a typical feature of all active two-terminal solid-state devices.
7. When the electric field is changed from zero to a threshold value, the carrier drift velocity increases linearly from zero to a maximum, according to Gunn's observations. The drift velocity is reduced and the diode exhibits negative resistance when the electric field exceeds the threshold value of 3000 V/cm for n-type GaAs.
8. The IMPATT diode's full name is IMPATT ionisation Avalanche Transit-Time. This is a microwave diode with an extraordinarily high output power. At microwave frequencies, it is typically utilised as an amplifier and oscillator. The IMPATT diode operates at frequencies ranging from 3 to 100 GHz.
9. TRAPATT diode is a high-efficiency microwave generator that can operate at frequencies ranging from a few hundred megahertz to a few gigahertz.
10. Due to the curvature of the Earth, the line of sight to the horizon is usually only 5 kilometres.
11. Blurring in a given circumstance is impacted by environmental conditions like downpour, lightning, etc. Blurring in a portable setting is subject to obstructions along the way that change over the long run.
12. Fading is the fluctuation of a signal's attenuation with numerous variables in wireless communications. Time, geographic location, and radio frequency are among the variables. Fading is frequently shown as a chaotic process. A communication channel that fades is known as a fading channel. Fading in wireless networks can be caused by multipath propagation (also known as multipath-induced fading), weather (especially rain), or shadowing from barriers impacting wave propagation (also known as shadow fading).

13. Types of Fading Sources are:
    - Slow Fading
    - Fast Fading
14. A varactor diode is a semi-conductor device in which the junction capacitance can be varied as a function of the reverse bias of the diode.
15. Array antennas: An array antenna is a high-gain antenna made up of a number of smaller antenna elements arranged in a grid.

    Leaky Wave Antenna: A leaky wave antenna gets its radiation from a leaking transmission line.

4.6 SUMMARY

- The transfer of information by electromagnetic waves with wavelengths in the microwave range (1 m - 1 mm) of the electromagnetic spectrum is known as microwave transmission.
- New telecommunication technologies such as wireless networks and direct-broadcast satellites, which transmit television and radio straight into consumers' homes, have exploded in their use of the microwave spectrum in recent years.
- A vacuum tube used to vibrate and amplify microwave frequency signals is known as a Klystron (also known as a Klystron Tube or Klystron Amplifier). Russell and Sigurd Varian, two American electrical engineers, invented it. The kinetic energy of an electron beam is used in a klystron. In UHF, low-power klystrons are typically employed as oscillators, whereas high-power klystrons are typically utilized as output tubes.
- A magnetron is a device that produces extremely powerful electromagnetic waves. It's a self-excited microwave oscillator, and it's also referred to as a crossed field device. The electric and magnetic fields produced inside it are mutually perpendicular to each other, which is why it's called that. As a result, the two crosses cross each other.
- All electrons that do not require energy from oscillations to move are referred to as preferred electrons. And these favoured electrons form an electron cloud or electron bunch as they travel from the cathode to the anode.
- The term frequency pushing and pulling refers to the variation in the magnetron's oscillation frequency.
- The use of two-terminal semiconductor devices at microwave frequencies has increased. At higher microwave frequencies, the CW, average, and peak power outputs of these devices are significantly higher than those of the best power transistor. Negative resistance is a typical feature of all active two-terminal solid-state devices.
- A high-field avalanche zone propagates through the diode, filling the depletion layer with a dense plasma of electrons and holes that are trapped in the low-field region behind it.

- The IMPATT diode's full name is IMPATT ionisation Avalanche Transit-Time. This is a microwave diode with an extraordinarily high output power. At microwave frequencies, it is typically utilised as an amplifier and oscillator. The IMPATT diode operates at frequencies ranging from 3 to 100 GHz.
- TRAPATT diode is a high-efficiency microwave generator that can operate at frequencies ranging from a few hundred megahertz to a few gigahertz.
- The loss in signal intensity of an electromagnetic wave caused by a line-of-sight passage across open space (typically air) with no objects nearby to induce reflection or diffraction is known as Free-Space Path Loss (FSPL) in telecommunication. It excludes things like the gain of the antennas used at the transmitter and receiver, as well as any loss caused by technical flaws.
- Blurring in a given circumstance is impacted by environmental conditions like downpour, lightning, etc. Blurring in a portable setting is subject to obstructions along the way that change over the long run.
- Fading is the fluctuation of a signal's attenuation with numerous variables in wireless communications. Time, geographic location, and radio frequency are among the variables. Fading is frequently shown as a chaotic process.
- A communication channel that fades is known as a fading channel. Fading in wireless networks can be caused by multipath propagation (also known as multipath-induced fading), weather (especially rain), or shadowing from barriers impacting wave propagation (also known as shadow fading).
- A varactor diode is a semi-conductor device in which the junction capacitance can be varied as a function of the reverse bias of the diode.

4.7 KEY TERMS

- **Microwave transmission:** The transfer of information by electromagnetic waves with wavelengths in the microwave range (1 m - 1 mm) of the electromagnetic spectrum is known as microwave transmission.
- **Magnetron:** A magnetron is a device that produces a high-intensity electromagnetic wave. It is commonly referred to as a self-excited microwave oscillator. It's sometimes referred to as a crossed-field gadget.
- **Frequency pushing and pulling:** The term frequency pushing and pulling refers to the variation in the magnetron's oscillation frequency.
- **IMPATT Diode:** The IMPATT diode's full name is IMPATT ionisation Avalanche Transit-Time. This is a microwave diode with an extraordinarily high output power. At microwave frequencies, it is typically utilised as an amplifier and oscillator. The IMPATT diode operates at frequencies ranging from 3 to 100 GHz.
- **TRAPATT diode:** TRAPATT diode is a high-efficiency microwave generator that can operate at frequencies ranging from a few hundred megahertz to a few gigahertz.

NOTES

- **Array antennas:** An array antenna is a high-gain antenna made up of a number of smaller antenna elements arranged in a grid.
- **Leaky Wave Antenna:** A leaky wave antenna gets its radiation from a leaking transmission line.

4.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What do you mean by the microwave devices and communication?
2. Define the term magnetron.
3. What are the travelling wave vibes?
4. Give the basic principles of two cavity klystron.
5. Define the term Gunn effect.
6. Write a short note on IMPATT diode and TRAPATT diode.
7. Write advantages and disadvantages of microwave transmission.
8. What do you understand by the detectors and components?
9. How will you define the Antennas used in microwave communication systems?

Long-Answer Questions

1. Discuss the microwave devices and communication with the help of examples.
2. Explain the reflex klystron principles of operation of magnetrons. Give appropriate examples.
3. Elaborate on the transferred electron devices with the help relevant examples.
4. Differentiate between the IMPATT diode and TRAPATT diode with the help of examples.
5. Discuss the advantages and disadvantages of microwave transmission.
6. What type of antennas are used in microwave communication systems? Explain.

4.9 FURTHER READING

Gurung, J.B. 2009. *Signals and Systems*. New Delhi: PHI Learning Pvt. Ltd.

Carlson, A Bruce and Paul B. Crilly. 2010. *Communication System*. Noida: Tata McGraw-Hill Education.

Lathi, B. P. and Zhi Ding. 2010. *Modern Digital and Analog Communication Systems*. Oxford: Oxford University Press.

NOTES

Kennedy, George and Bernard Davis. 1993. *Electronic Communication Systems*. London: Glencoe/McGraw-Hill.

Roddy, Dennis and John G. Coolen. 1977. *Electronic Communications*. Noida: Pearson Education.

Greenwich, Edwin. 2012. *Analog Integrated Circuits*. Berlin: Springer Science & Business Media.

Millman, Jacob and Christos C. Halkias. 2001. *Integrated Electronics*. Noida: Tata McGraw-Hill Education.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey: Prentice-Hall.

Tanenbaum, Andrew. S. 2002. *Computer Networks*. New Jersey: Prentice-Hall PTR.

Basavaraj, B. and H.N. Shivashankar. 2004. *Basic Electronics*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 5 DIGITAL COMMUNICATIONS: PCM, CODES AND TELEGRAPHS

NOTES

Structure

5.0 INTRODUCTION

Digital electronics is a field of electronics involving the study of digital signals and the engineering of devices that use or produce them. This contrasts with analog electronics and analog signals. Digital electronic circuits are usually made from large assemblies of logic gates, often packaged in integrated circuits. Complex devices may have simple electronic representations of Boolean logic functions. Digital logic was the invention of George Boole in the mid-19th century.

In digital communications, transmission of information takes place in digital form. Digital transmission or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel. These could be copper wires, optical fibers, wireless communication channels, and storage media. The data is represented as an electro-magnetic signal like an electrical voltage, radio wave, microwave, etc.

NOTES

Data transmitted may be digital messages originating from a data source, for example, a computer or a keyboard. It may also be an analog signal such as a phone call or a video signal, digitized into a bit-stream for example using Pulse Code Modulation (PCM) or more advanced source coding (analog-to-digital conversion and data compression) schemes. This source coding and decoding is carried out by codec equipment.

In telecommunication, Data Signaling Rate (DSR), also known as gross bit rate, is the aggregate rate at which data passes a point in the transmission path of a data transmission system. The data signaling rate is the aggregate rate at which data pass a point in the transmission path of a data transmission system. In signal processing, a digital filter is a system that performs mathematical operations on a sampled, discrete-time signal to reduce or enhance certain aspects of that signal.

A delta modulation (DM or Δ-modulation) is an analog-to-digital and digital-to-analog signal conversion technique used for transmission of voice information where quality is not of primary importance. DM is the simplest form of Differential Pulse Code Modulation (DPCM) where the difference between successive samples is encoded into n-bit data streams. In delta modulation, the transmitted data are reduced to a 1-bit data stream.

A teleprinter (teletypewriter, teletype or TTY) is an electromechanical device that can be used to send and receive typed messages through various communications channels, in both point-to-point and point-to-multipoint configurations. Teleprinters could use a variety of different communication media.

Radiotelegraphy was the first means of radio communication. Fundamentally, in the radiotelegraphy, radio communication is done by means of Morse Code or other coded signals. The radio carrier is modulated by changing its amplitude, frequency, or phase in accordance with the Morse dot-dash system or some other code. At the receiver the coded modulation is recovered by an appropriate demodulator and the code groups are converted into the corresponding symbols. In many instances the symbols are generated by a computer and modem rather than with a manual telegraph key. Wireless telegraphy or radiotelegraphy is transmission of telegraph signals by radio waves.

In this unit, you will study about the advantages and disadvantages of digital communications, bit transmission, signaling rate, error probability, digital filtering, delta modulation, Pulse Code Modulation (PCM), PCM generation, binary coding, PCM bandwidth, PCM reception noise, quantization noise analysis, SIN ratio and channel capacity of PCM, error detection and correction codes, digital carrier systems, teleprinters and telegraphs circuits, and radio telegraphs transmitters.

5.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance, advantages and disadvantages of digital communications
- Know what bit transmission is

- Understand signaling rate and error probability
- Explain digital filtering and delta modulation
- Elaborate on Pulse Code Modulation (PCM) and PCM generation
- Describe binary coding and PCM bandwidth
- Define PCM reception noise and quantization noise analysis
- Comprehend on SIN ratio and channel capacity of PCM
- Explain the error detection and correction codes
- Understand about the digital carrier systems
- Discuss about teleprinters and telegraphs circuits
- Know the importance of radio telegraphs transmitters

5.2 DIGITAL COMMUNICATIONS

Till recently, the basic communication systems used were analog. But from the last decade, we see a remarkable growth in the field of digital communications. In digital communications, transmission of information takes place in digital form. The major reason of growth of digital communication recently is the availability of wide-band channels such as optical fibre, satellite channels, etc.

The reasons for development of digital communications are simply the advantages of digital with respect to analog. Some advantages of digital communication are listed below.

(i) Now digital formats can be used for different variety of signals. For example, we can transmit voice, video, photo by using the binary digits.

(ii) By using error correcting and error detecting codes, we can improve the noise performance of system.

But some disadvantages of digital communications are also there, such as:

(i) Now we need the wideband channels due to increased bandwith of signal.

(ii) In digital system, the systems are very complex.

Today, we use digital communication on the costly channels and complex systems but analog communication systems are still used and cannot be replaced completely.

The following are some of the differentiating features between analog and digital signals:

Analog Signal

1. Analog signals are continuous.
2. Analog signal can continuously vary.
3. The primary disadvantage of an analog signal is noise.
4. Sound waves are a continuous wave and as such are analog in the real world.

5. Analog signal required lesser bandwidth capacity than digital capacity.
6. Transmitting and receiving analog signals is usually less expensive.

Digital Signals

1. Digital signal is discrete.
2. Digital signal are based on 0s and 1s.
3. Noise is much easier to filter out of a digital signal.
4. PCs and all computers work using digital signals.
5. Digital signals require greater bandwidth capacity than analog signals.
6. Processing digital signals is more complex and the equipment employed is expensive.

Digital transmission or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel. These could be copper wires, optical fibers, wireless communication channels, and storage media. The data is represented as an electro-magnetic signal like an electrical voltage, radio wave, microwave, etc.

Analog communications stand for the transmission of continuously varying information signal, whereas digital communications is the transmission of discrete messages. The messages are either represented by a sequence of pulses by means of a line code (*baseband transmission*), or by a limited set of continuously varying wave forms (*passband transmission*), using a digital modulation method. The passband modulation and corresponding demodulation (also known as detection) is carried out by modem equipment. According to the common definition of digital signal, both baseband and passband signals representing bit-streams are considered as digital transmission, while an alternative definition only considers the baseband signal as digital, and passband transmission of digital data as a form of digital-to-analog conversion.

Data transmitted may be digital messages originating from a data source, for example, a computer or a keyboard. It may also be an analog signal such as a phone call or a video signal, digitized into a bit-stream for example using Pulse-Code Modulation (PCM) or more advanced source coding (analog-to-digital conversion and data compression) schemes. This source coding and decoding is carried out by codec equipment.

Asynchronous transmission uses start and stop bits to signify the beginning and ending bits. ASCII character would actually be transmitted using 10 bits e.g.: A "0100 0001" would become "1 0100 0001 0". The extra 1 (or 0 depending on parity bit) at the start and end of the transmission tells the receiver first that a character is coming and secondly that the character has ended. This method of transmission is used when data is sent intermittently as opposed to in a solid stream. In the present example, the start and stop bits are in bold. The start and stop bits must be of opposite polarity. This allows the receiver to recognize when the second packet of information is being sent.

Synchronous transmission uses no start and stop bits. Instead it synchronizes transmission speeds at both the receiving and sending ends of the transmission using clock signal(s) built into each component. A continuous stream of data is then sent between the two nodes. Since there are no start and stop bits, the data transfer rate is faster although more errors can occur, as the clocks will eventually get out of sync, and the receiving device would have the wrong time than that had been agreed in protocol for sending/receiving data, so some bytes could become corrupted (by losing bits). Ways to get around this problem include re-synchronization of the clocks and use of check digits to ensure the byte is correctly interpreted and received.

Although analog systems are less expensive in most cases, the digital communication systems offer higher efficiency, better performance, and greater flexibility. For example, in case of analog communication systems, Amplitude Modulated (AM) radio is a typical example, which can inexpensively communicate a band-limited analog signal from one location to another (point-to-point communication) or from one point to many (broadcast). An analysis of the radio receiver indicates that some residual error is *always* present in an analog system's output.

- ***Efficiency:*** 'The Source Coding Theorem' allows quantify the complexity of a given message source, and allows us to exploit that complexity by source coding (compression). In analog communication, the only parameters of interest are message bandwidth and amplitude. In case of analog communication, we cannot exploit signal structure to achieve a more efficient communication system.
- ***Performance:*** In digital communication systems, with the help of the Noisy Channel Coding Theorem, we can formulate error-correcting codes. This process can bring us close to error-free transmission. Even if we send information by a noisy channel, digital schemes are capable of filtering the errors to ensure error-free transmission while analog systems have limited capacity to overcome channel disturbances.
- ***Flexibility:*** Digital communication systems can transmit real-valued discrete-time signals, which could be analog ones obtained by analog-to-digital conversion, *and* symbolic-valued ones (computer data, for example). Any signal that can be transmitted by analog means can be sent by digital means, with the only issue being the number of bits used in A/D conversion (depending upon how accurately we need to represent signal amplitude). Images can be sent by analog means (commercial television), but better communication performance occurs when we use digital systems (HDTV). In addition to the ability of digital communications to transmit a wider variety of signals than analog systems, point-to-point digital systems can be organized into global (and beyond as well) systems that provide efficient and flexible information transmission. Even analog-based networks, such as the telephone systems, employ modern computer networking ideas rather than the purely analog systems of the past.

Consequently, with the increased speed of digital computers, the development of increasingly efficient algorithms, and the ability to interconnect computers to form a communications infrastructure, digital communication is now the best choice for many situations.

5.2.1 Advantages and Disadvantages of Digital Communications

Following are the advantages and disadvantages of digital communications.

NOTES

Advantages of Digital Communication

- In digital signals, the impact of noise interference, distortion is less.
- It facilitates video conferencing that saves time, money, and effort.
- It is easy to implement and less expensive.
- It is used in military applications.
- The correction and detection of errors are easy in digital communication, as there is a use of channel coding.
- As compared to analog signals, it is easy to save and retrieve digital signals.
- In digital signals, the configuring process is easy as compared to analog signals.
- There is a common encoding technique in most digital circuits, so for a number of processes, similar devices can be used.
- The probability of crosstalk is very less in digital communication.
- The implementation of hardware is more flexible in digital communication.
- In digital communication, to avoid signal jamming, the spread spectrum technique is used.
- It also facilitates audio conferencing to someone or a group of people in another location without traveling. Thus, it saves time, effort, and money.
- To maintain the secrecy of information, the signal processing functions like compression and encryption are employed in digital circuits.
- Digital communication is cheaper and simpler compared to analog signals because of the advancement of IC technologies.

Disadvantages of Digital Communications

- There is high power consumption in digital communication.
- There is a requirement for synchronization in the case of synchronous modulation.
- There is a sampling error.
- The most common limitation of digital communication is that it requires more transmission bandwidth. It is due to the higher data rate because of analog to digital conversion.
- Digital communication requires analog to digital conversion at a high rate.
- There can be a possibility of miscommunication if a user doesnot understand something.

5.2.2 Bit Transmission

Bit rate is the rate over network speed which is used to detect errors while transmitting data. The most popular method for detecting errors is inserting a parity bit alongside the data bits for a character. Receiving modems detect incorrect bit rate, which is also called parity bit. It requests the sending modem to retransmit the character.

NOTES

Table 5.1 *String of Data with Start and Stop Bits*

Stop Bit	Data Bits	Start Bit
0	1 0 0 0 0 0 1	0

Table 5.2 *String of Data with Inserted Parity Bit*

Stop Bit and Parity Bit	Data Bits	Start Bit
0	1 0 0 0 0 0 1	0

A modem is a bridge gap between digital and analog data transmission which allows the digital data to be transmitted/received over the telephone lines. In telecommunication and network computing, bit rate are sometimes written as data rate and are conveyed and processed per unit time. It is measured as bits per second (bit/s or bps).

Table 5.3 *The SI Prefix for Different Bits*

Bit	SI Prefix
Kilo	Kbit/s or Kbps
Mega	Mbit/s or Mbps
Giga	Gbit/s or Gbps
Tera	Tbit/s or Tbps

The net bit rate, also called useful bit rate, of a digital communication link, is the capacity of the physical layer protocol, such as: Framing Bits, Time Division Multiplex (TDM) and redundant Forward Error Checking (FEC).

The operating system indicates the connection speed of the network access technology device. The speed of connection of bit rate is determined by FEC and refers to physical layer net bit rate.

For example, the connection speed of IEEE (The Institute of Electrical and Electronic Engineers) 802.11, a wireless network, is the bit rate of 6 Mbit/s and 54 Mbit/s while the gross bit rate is between 12 Mbit/s and 72 Mbit/s which includes error correcting codes. The connection speed of ISDN (Integrated Services Digital Network) BSI (Basic Rate Interface) that is 2 B-Channels+ 1D Channel) of 64+64+16= 144 Kbit/s refers to the user data rates, while the line rate is 160 kbit/s. The connection speed of Ethernet 100 Base–TX (TRANSMIT) physical layer is 100 Mbit/s, which is also its gross bit rate.

Throughput and digital bandwidth consumption refers to the bit rate achieved in a computer network over a logical or physical communication link via network node. Throughput is affected by traffic load of network resources, whereas data transfer rate is achieved by average net bit rate of all protocol overheads, data packet transmissions.

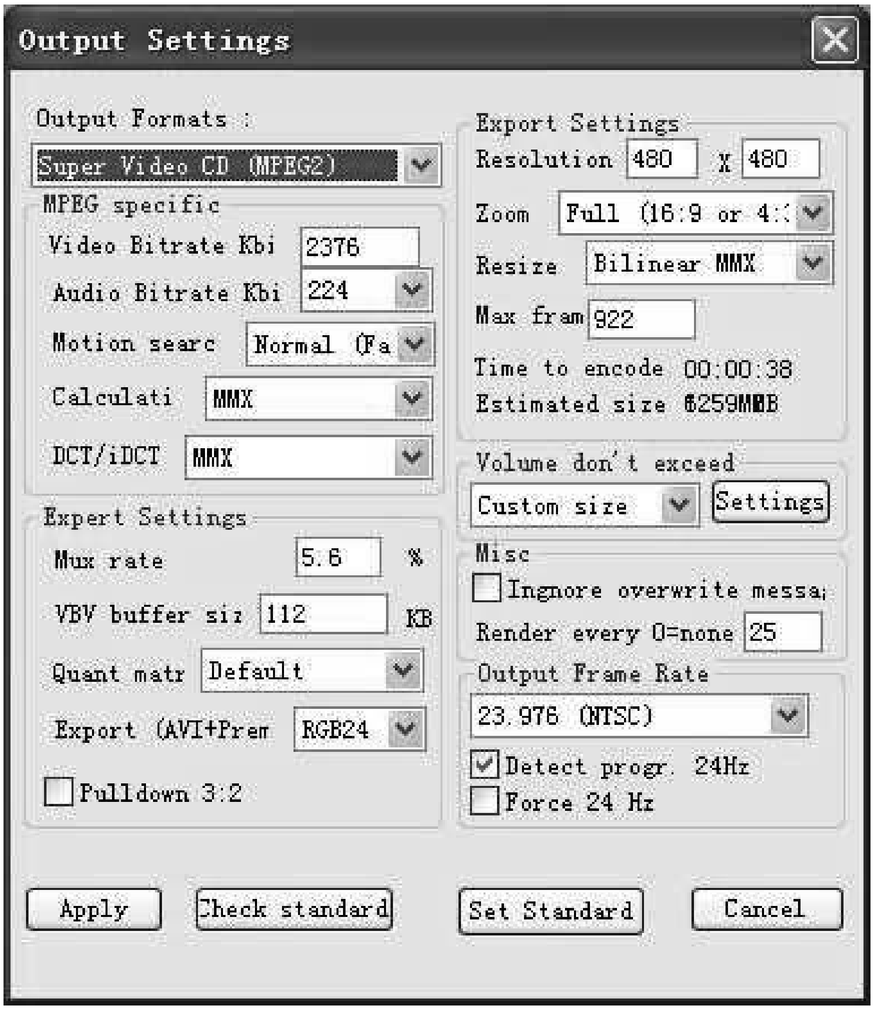


Fig. 5.1 *Output Setting of Bit Rate*

Figure 5.1 shows the video bit rate of the output MPEG file and audio bit rate, which are arranged for output setting. Motion search is set to achieve a balance between noise blurring and rendering speed around moving objects in the output file.

Multimedia Bit Rate

Bit rate is the number of bits used per unit to represent continuous medium such as audio or video following source coding (data compression) to the multimedia files. The size of the multimedia file is the product of bit rate (in bit/s) in bytes and the length of recording in seconds divided by eight. The bit rate is measured by input, which avoids interrupts with reference to streaming multimedia.

Fundamentals of Bit Rate

- The sample uses different number of bits.
- The data is encoded by different number of bits.
- The material is sampled at different frequencies.
- The information is digitally compressed by different algorithms.

Table 5.4 *The Reference Standards for Different Measurements of Bit Rates*

Bit Rate Measured	Reference Standard
32 Kbit/s	MW AM Quality
96 Kbit/s	FM Quality
128-160 Kbit/s	Bass Quality
192 Kbit/s	Digital Audio Broadcasting (DAB)
224-320 Kbit/s	CD Quality
800 bit/s	Recognizable Speech
8 Kbit/s	Telephone Quality
500 Kbit/s	Audio Format such as FLAC, WavPack, Monkey's Audio
1411.2 Kbit/s	Compact Disc Digital Audio Video (MPEG2)
16 Kbit/s	Video Phone Quality
128-384 Kbit/s	Business – Oriented Videoconferencing System Quality
1.25 Mbit/s	VCD Quality
5 Mbit/s	DVD Quality
15 Mbit/s	HDTV Quality
36 Mbit/s	HD DVD Quality
54 Mbit/s	Blue-Ray Disc Quality

NOTES

Properties of Bit Rate with Reference to Speed

The bit rate is fixed and uniform for a specific network. The gross bit rate is the number of bits transmitted per second by an ideal transmitter. It could be as high as 1 Mbit/s. The net bit rate means the number of useful bits carried per second. The latency of bit rate is the time interval between the instant of initialization, that is, transmission request and the actual start of transmission including all algorithms and parameters. It is used to calculate the speed of time to access the network, without getting error frames.

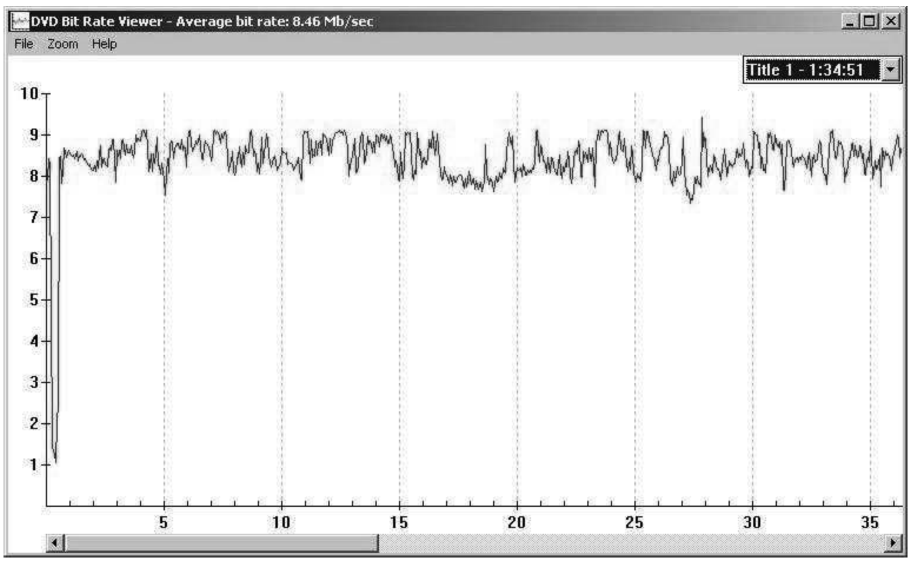


Fig. 5.2 *DVD Bit Rate via Network Speed*

The above Figure 5.2 shows the graph of DVD bit rate viewer as average bit rate according to network searching speed.

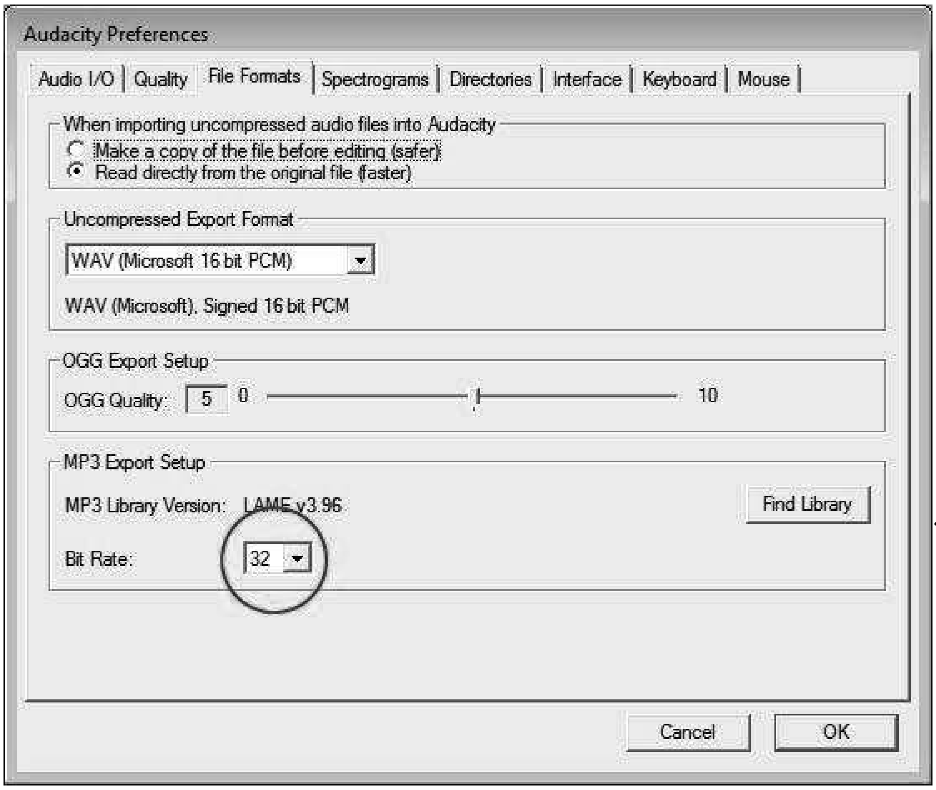


Fig. 5.3 *Audacity Preferences*

The above Figure 5.3 shows how the bit rate is set for MP3 (Music Player 3) Export set up as library version. Here, it is set as 32 Kbit/s. It is significant to export the messages as MP3s because they are smaller files as compared to WAV (Windows Media Audio) files.

5.2.3 Signaling Rate

In telecommunication, Data Signaling Rate (DSR), also known as Gross Bit Rate (GBR), is the aggregate rate at which data passes a point in the transmission path of a data transmission system. Fundamentally, the Data Signaling Rate (DSR) is the aggregate rate at which data pass a point in the transmission path of a data transmission system.

1. The DSR is usually expressed in bits per second.
2. The data signaling rate is given by $\sum_{i=1}^{m} \frac{\log_2 n_i}{T_i}$ where m is the number of parallel channels, n_i is the number of significant conditions of the modulation in the i-th channel, and T_i is the unit interval, expressed in seconds, for the i-th channel.
3. For serial transmission in a single channel, the DSR reduces to $(1/T) \log_2 n$; with a two-condition modulation, i.e., $n = 2$, the DSR is $1/T$, according to Hartley's law.
4. For parallel transmission with equal unit intervals and equal numbers of significant conditions on each channel, the DSR is $(m/T) \log_2 n$; in the case of a two-condition modulation, this reduces to m/T.

5. The DSR may be expressed in bauds, in which case, the factor $\log_2 n_i$ in the above summation formula should be deleted when calculating bauds.
6. In synchronous binary signaling, the DSR in bits per second may be numerically the same as the modulation rate expressed in bauds. Signal processors, such as four-phase modems, cannot change the DSR, but the modulation rate depends on the line modulation scheme, in accordance with condition 4. For example, in a 2400 bit/s 4-phase sending modem, the signaling rate is 2400 bit/s on the serial input side, but the modulation rate is only 1200 bauds on the 4-phase output side.

NOTES

Maximum Rate

The maximum user signaling rate, synonymous to gross bit rate or data signaling rate, is the maximum rate, in bits per second, at which binary information can be transferred in a given direction between users over the telecommunications system facilities dedicated to a particular information transfer transaction, under conditions of continuous transmission and no overhead information.

For a single channel, the signaling rate is given by,

$$SCSR = \frac{\log_2 n}{T},$$

Where SCSR is the Single-Channel Signaling Rate in bits per second, T is the minimum time interval in seconds for which each level must be maintained, and n is the number of significant conditions of modulation of the channel.

In the case where an individual end-to-end telecommunications service is provided by parallel channels, the Parallel Channel Signaling Rate (PCSR) is given by,

$$PCSR = \sum_{i=1}^{m} \frac{\log_2 n_i}{T_i}$$

Where PCSR is the total signaling rate for m channels, m is the number of parallel channels, T_i is the minimum interval between significant instants for the I-th channel, and n_i is the number of significant conditions of modulation for the I-th channel.

In the case where an end-to-end telecommunications service is provided by tandem channels, the end-to-end signaling rate is the lowest signaling rate among the component channels.

5.2.4 Error Probability

The probability of error in a digital communication system is mainly due to inter-symbol interference and bit errors during the digital transmission.

In digital transmission, the number of bit errors is the number of received bits of a data stream over a communication channel that have been altered due to noise, interference, distortion or bit synchronization errors.

The Bit Error Rate (BER) is the number of bit errors per unit time. The Bit Error Ratio (also BER) is the number of bit errors divided by the total number of

NOTES

transferred bits during a studied time interval. Bit error ratio is a unitless performance measure, often expressed as a percentage.

The **bit error probability p_e** is the expected value of the bit error ratio. The bit error ratio can be considered as an approximate estimate of the bit error probability. This estimate is accurate for a long-time interval and a high number of bit errors.

For example, assume that the following is transmitted bit sequence:

1 1 0 0 0 1 0 1 1

And the following is the received bit sequence:

0 1 0 1 0 1 0 0 1

The number of bit errors (the underlined bits) is, in this case, 3. The BER is 3 incorrect bits divided by 9 transferred bits, resulting in a BER of 0.3 or 30%.

Packet Error Ratio

The Packet Error Ratio (PER) is the number of incorrectly received data packets divided by the total number of received packets. A packet is declared incorrect if at least one bit is erroneous. The expectation value of the PER is denoted **packet error probability p_p**, which for a data packet length of *N* bits can be expressed as,

$$p_p = 1 - (1 - p_e)^N = 1 - e^{N \ln(1-p_e)}$$

Assuming that the bit errors are independent of each other. For small bit error probabilities and large data packets, this is approximately,

$$p_p \approx p_e N$$

Similar measurements can be carried out for the transmission of frames, blocks, or symbols.

5.2.5 Digital Filtering

Filters are a key feature of any signal processing or telecommunications system. The primary functions of a filter are:

(a) To restrict a signal to a certain frequency band or channel, as in an anti-aliasing filter or a radio/television channel selection.

(b) It is common practice in music coding to split up a signal into multiple smaller signals called 'Sub-Bands', which are then processed separately.

(c) To transform a signal's frequency spectrum, like in an audio-visual equalizer for example.

(d) Simulating a system's input-output relationship, such as a mobile communication channel, voice generation, musical instruments, telephone line echo, and room acoustics.

Digital filtering can be used to eliminate both hardware and software bottlenecks.; in the first case, the numerical processor is either a special-purpose chip or it is assembled out of a collection of digital integrated circuits that serve as the foundation for a digital filtering process – storage, delay, addition/subtraction and multiplication by constants. In contrast, a general-purpose mini-or micro-computer can also be programmed as a digital filter, in which case the numerical processor is the computer's CPU and memory.

NOTES

Signals that are then converted back to the analogue domain can be filtered by exhausting a *digital to analogue convertor* or worked with entirely in the digital domain (e.g., biomedical signals are digitized, filtered and analyzed to detect some abnormal activity).

The advantages of digital filters over analogue filters are as follows:

The numerical processor can easily be (re-)programmed to implement a number of different filters. The round-off error in a digital filter is the only factor that determines the filter's accuracy. This has two advantages:

- A priori, the algorithm's performance can be predicted based on how accurate it is.
- With the right design strategies, round-off error can be reduced, allowing digital filters to fulfil very strict magnitude and phase constraints (that would be nearly impossible to attain using analogue filters as a consequence of component tolerances and circuit noise).

The widespread use of mini- and micro-computers in engineering has greatly increased the number of digital signals recorded and processed. Power supply and temperature variations have no effect on a programme stored in a computer. Unlike digital circuitry, analogue circuitry is significantly more vulnerable to noise.

A digital filter is a mathematical technique that operates on a digital input signal to generate a digital output signal with the objective of filtering out unwanted data. It is implemented in hardware and/or software. Digital filters work with digitized analogue signals or just numbers in a computer memory that represent a variable. A simplified block design of a real-time digital filter with analogue input and output signals is shown in Figure 5.4.

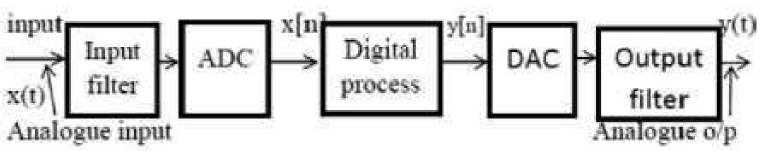


Fig. 5.4 *A Simplified Block Diagram illustrating the Operation of a Real-Time Digital Filter with Analogue Input and Output Signals*

Digital filters are used in a variety of applications, including data compression, biological signal processing, speech and image processing, data transmission, digital audio, and telephone echo cancellation. Several advantages of digital filters over analogue filters include the following:

1. Truly linear phase response.
2. Variables in the environment, such as temperature variations, have a negligible effect on performance.
3. When a programmable processor is used to design the filter, the frequency response can be altered spontaneously.
4. At extremely low frequencies, digital filters can be used.

NOTES

While digital filters provide a number of advantages over analogue filters, they do have a few disadvantages. Some main disadvantages are listed below:

1. Speed limitation: In comparison to analogue signals, the maximum bandwidth of signals processed in real time by digital filters is significantly less.
2. Finite word length effects: Digital filters are subject to noise sources, such as ADC and round off noise, which can result in instability.
3. Long design and development times: Digital filter design and development, particularly hardware development, can take substantially longer than analogue filter design and development.

Types of Filters

Filters can be roughly categorized into the following classes, depending on the shape of the filter equation and the implementation structure:

(a) Linear filters versus nonlinear filters.

(b) Time-invariant filters versus time-varying filters.

(c) Adaptive filters versus non-adaptive filters.

(d) Recursive versus non-recursive filters.

(e) Structures in direct-form, cascade-form, parallel-form, and lattice.

The primary focus of this section is on Linear Time Invariant (LTI) filters. This is a class of filters whose output is a linear combination of the input and whose coefficients are constant in time.

Filters can be described in the time or frequency domain using the following methods:

(a) Time Domain Input-Output Relationship

This relationship is stated by the difference equation that is used to express the output of a discrete time filter as a weighted sum of the input and preceding output samples. For instance, the difference equation for a first-order filter might be as follows:

$$y(m) = a\, y(m-1) + x(m) \qquad \text{... (5.1)}$$

Where $x(m)$ is the input to the filter, $y(m)$ denotes the output of the filter, and a denotes the filter coefficient.

(b) Impulse Response

A filter's reaction to an impulse input can be defined mathematically. For instance, the filter described in Equation (5.1) responds as follows to a discrete time impulse input at m=0:

$$y(m) = a^m \qquad m = 0,1,2,\ldots\ldots \qquad \ldots\ldots (5.2)$$

$$= 1,\ a,\ a^2, \ldots..$$

And it is assumed that $y(-1) = 0$

Impulse response is useful because:

(i) Because any signal can be viewed as the sum of a number of shifted and scaled impulses, the response of a linear filter to a signal is the sum of the responses to all the impulses that comprise the signal,

(ii) Because an impulse input contains all frequencies with equal energy, it excites a filter at all frequencies, and

(iii) Because an impulse input contains all frequencies with equal energy, an impulse input excites a filter at all frequencies.

(c) Transfer Function, Poles and Zeros

The ratio of the *z*-transforms of the filter output and input is the transfer function of a digital filter H(z):

$$H(z) = \frac{Y(z)}{X(z)} \qquad \text{..... (5.3)}$$

The transfer function of the filter in Equation (5.1), for example, is given by:

$$H(z) = \frac{1}{1-az^{-1}} \qquad \text{...... (5.4)}$$

The pole zero description of a filter is a valuable approach of acquiring understanding into its behavior. The roots of the denominator and numerator of the transfer function, respectively, are poles and zeros.

(d) Frequency Response

The frequency response of a filter explains how the filter alters the magnitude and phase of the frequencies of the input signal it receives. Performing the Fourier transform on the filter's impulse response or replacing the frequency variable $e^{j\omega}$ for the z variable $z = e^{j\omega}$ in the z-transfer function, as shown below, can be used to determine the frequency response of the filter:

$$H(z = e^{j\omega}) = \frac{Y(e^{j\omega})}{X(e^{j\omega})} \qquad \text{...... (5.5)}$$

It is possible to express the frequency response of a filter in terms of the filter's amplitude and phase responses, which are both complex variables.

Linear Time-Invariant Digital Filters

Linear Time-Invariant (LTI) filters are a form of filter whose output is a linear combination of the samples of the input signal with constant coefficients, as opposed to other types of filters. The linear property indicates that the filter response to a weighted sum of a number of signals is equal to the weighted sum of the filter responses of each individual signal in the weighted sum. The superposition principle is what we're talking about here. The term 'Time Invariant' refers to the fact that the filter coefficients, and hence the frequency response of the filter, remain constant during the course of its operation. The input-output relationship of a discrete-time linear filter in the time domain is described by the linear difference equation shown below in the graph.

$$y(m)=\sum_{k=1}^{N} a_k y(m-k)+\sum_{k=0}^{M} b_k x(m-k) \quad \text{… (5.6)}$$

NOTES

Where $\{a_k, b_k\}$ are the filter coefficients, and the output *y(m)* is a linear combination of the previous *N* output samples [$y(m-1), \ldots, y(m-N)$], the present input sample *x(m)* and the previous M input samples [$x(m-1), \ldots, x(m-M)$]. The characteristic of a filter is completely determined by its coefficients $\{a_k, b_k\}$.

The coefficients $\{a_k, b_k\}$ are constants computed to achieve a particular frequency response for a time-invariant filter. The z-transform of the difference Equation (5.6) yields the following filter transfer function:

$$H(z)=\frac{\sum_{k=0}^{M} b_k z^{-k}}{1-\sum_{k=1}^{N} a_k z^{-k}} \quad \text{… (5.7)}$$

The frequency response of this filter may be found by substituting the frequency variable $e^{j\omega}$ for the *z* variable, $\omega = e^{j\omega}$, in Equation (5.7).

$$H(e^{j\omega})=\frac{\sum_{k=0}^{M} b_k e^{-j\omega k}}{1-\sum_{k=1}^{N} a_k e^{-j\omega k}} \quad \text{… (5.8)}$$

Given that a signal is a weighted combination of a number of sine waves as a result of the Fourier transform, it follows from the superposition principle that linear filtering in the frequency domain can be thought of as a linear combination of the frequency constituents of the input multiplied by the signal's frequency response.

Filter Order

The order of a discrete-time filter is defined as the largest discrete-time delay employed in the filter's input-output equation. For instance, in Equations 5.6 or 5.7, the filter order is determined by the greater of the N or M numbers. The order of a continuous time filter is determined by the order of the highest differential term utilized in the filter's input-output equation.

Recursive and Non-Recursive Filters

The block diagram in Figure 5.5 illustrates an implementation of the linear time-invariant filter Equation (5.1). In Equation (5.7), the filter's transfer function is the ratio of two polynomials in the variable z, which can be expressed in cascade form as:

$$H(z) = H_1(z)\, H_2(z) \quad \text{…(5.9)}$$

Where $H_1(z)$ is the transfer function of an all-zero feed-forward filter as defined by:

$$H_1(z)=\sum_{k=0}^{M} b_k z^{-k} \quad \text{… (5.10)}$$

and $H_2(z)$ is the transfer function of a recursive feedback, all-pole filter as defined by:

NOTES

$$H_2(z)=\frac{1}{1-\sum_{k=1}^{N} a_k z^{-k}} \quad \text{… (5.11)}$$

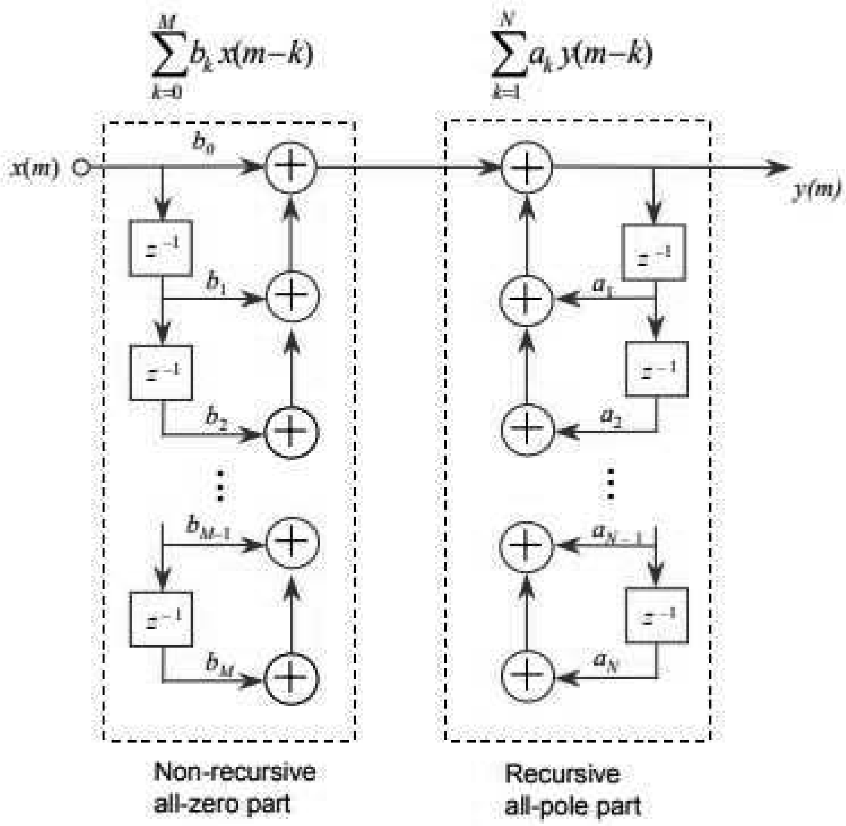


Fig. 5.5 *Illustration of a Direct-Form Pole-Zero IIR Filter*

Figure 5.5 illustrates a Direct-Form Pole-Zero IIR Filter, demonstrating that the output is the sum of two vector products: a combination of the input samples $[b_0,..., b_M]$ that is weighted $[x(m),..., x(M\text{-}1), x(M\text{-}1)]T$ plus a weighted sum of the output feedback values $[a_1,..., a_N]$ $[y(m\text{-}1), y(m\text{-}1),..., y(m\text{-}N)]$ T. T stands for transposition.

Non-Recursive or Finite Impulse Response (FIR) Filters

There is no feedback in a non-recursive filter, and its input-output relationship is given by

$$y(m)=\sum_{k=0}^{M} b_k x(m-k) \quad \text{…. (5.12)}$$

A non-recursive filter's output *y(m)* is a function of simply the input signal *x*, as shown in Figure 5.6 (*m*). A finite sequence of *M+1* samples, where *M* is the filter order, makes up the response of such a filter to an impulse. The filter is called a Finite-Duration Impulse Response (FIR) filter as a result. A non-recursive filter is also known as an all-zero filter, a feed-forward filter, or a Moving Average (MA) filter, which is a phrase that is commonly used in statistical signal processing literature.

Recursive or Infinite Impulse Response (IIR) Filters

A recursive filter contains feedback from output to input, and its output is generally a function of previous output samples as well as current and past input samples, as shown by the equation below:

$$y(m)=\sum_{k=1}^{N} a_k y(m-k)+\sum_{k=0}^{M} b_k x(m-k) \quad \dots (5.13)$$

The direct form method to Equation 5.10 is illustrated in Figure 5.5. A recursive filter, when activated by an impulse, produces an output that can be continued forever. The IIR (Infinite Duration Impulse Response) filter is often used to refer to a recursive filter as a result of this property. An IIR filter is often referred to by other names, such as feedback filters, pole zero filters, and Auto-Regressive-Moving-Average (ARMA) filters, the latter of which is a word that is frequently seen in the literature on statistical signal processing.

An IIR filter with discrete time characteristics has a *z*-domain transfer function equal to the ratio of two z-transform polynomials, as defined in Equation (5.7); it contains a number of poles corresponding to the denominator polynomial's roots, as well as a number of zeros corresponding to the numerator polynomial's roots. When comparing IIR and FIR filters, the fundamental difference is that an IIR filter is typically more compact than a FIR filter in that it can typically attain the appropriate frequency response with fewer coefficients. Because there are fewer filter coefficients, there is less storage space required as well as faster calculation and greater throughput. A consequence of this is that IIR filters are typically both more memory and computationally economical than FIR filters. Nonetheless, although a FIR filter is always stable, an IIR filter can become unstable (for example, if the IIR filter's poles are positioned outside of the unit circle), and caution must be exercised while constructing IIR filters in order to ensure that they are stable.

Difference between FIR and IIR Filters

S. No.	FIR	IIR
1.	Can have exactly linear phase response, i.e., do not distort the phase of the signal.	The phase responses, particularly at the band edges, are nonlinear.
2.	FIR filters realized non-recursively are always stable.	The stability of IIR filters cannot be guaranteed.
3.	The implications of implementing filters with a limited number of bits are less severe.	More severe consequences such as round off noise and coefficient quantization errors occur.
4.	Increase the number of coefficients required for sharp cutoff filters.	Reduced coefficients result in decreased processing time and storage.
5.	There is no analogue counterpart, but synthesis of filters with arbitrary frequency responses is much easier.	Analog filters can easily be converted to IIR digital equivalents.
6.	Algebraically more difficult to synthesize in the absence of CAD assistance.	Synthesis is less difficult.

Thus, a general rule for determining when to employ FIR or IIR is as follows:

- Use IIR when just sharp cutoff filters and high throughput are required, as IIR filters produce fewer coefficients than FIR filters.
- Use FIR when the number of filter coefficients is not excessive and, more importantly, when minimal or no phase distortion is needed.

Filter Design Steps

Filter design is the process of creating a signal processing filter that satisfies a set of contradicting requirements. The objective is to find a filter implementation that satisfies all of the conditions sufficiently well to be helpful. The filter design process can be thought of as an optimization problem in which each requirement contributes to a reduced error function. Certain aspects of the design process can be automated, but a skilled electrical engineer is typically required to achieve a satisfactory output. Digital filter design is a surprisingly hard subject. While filters are straightforward to understand and compute, the practical obstacles associated with their design and implementation are significant and the topic of advanced research. A digital filter is designed in five steps:

1. Filter requirements must be specified.
2. Determination of appropriate filter coefficients.
3. Use of a suitable structure to represent the filter (realization).
4. Examination of the effects of word length constraints on filter performance.
5. Filter implementation in software and/or hardware.

5.2.6 Binary Coding

Code is a symbolic representation of discrete information, which may be presented in the form of numbers, letters or physical quantities. The symbols used are the binary digits 0 and 1, which are arranged according to the rules of codes. These codes are used to communicate information to a digital computer and to retrieve messages from it. A code is used to enable an operator to feed data into a computer directly, in the form of decimal numbers, alphabets and special characters. The computer converts these data into binary codes and after computation, transforms the data into its original format, such as decimal numbers, alphabets and special characters.

When numbers, letters, or words are represented by a special group of symbols, this is called *encoding*, and the group of symbols is called a *code*. In Morse code, a series of dots and dashes represent alphabet, numerals and special characters.

Codes are broadly classified into five groups, viz., (i) Weighted Binary Codes, (ii) Non-weighted Codes, (iii) Error-detecting Codes, (iv) Error-correcting Codes, and (v) Alphanumeric Codes.

Weighted Binary Codes

Weighted binary codes obey their positional weighting principles. Each position of a number represents a specific weight. In a weighted binary code, the bits are multiplied by the weights indicated; the sum of these weighted bits gives the equivalent decimal digit.

NOTES

Straight binary coding is a method of representing a decimal number by its binary equivalent. The codes 8421, 2421, 5421 and 5211 are weighted binary codes. Each decimal digit is represented by a four-bit binary word, the three digits for the left being weighted. Table 5.5 consists of a few weighted 4-bit binary codes with their decimal numbers and complements.

Table 5.5 *Some Weighted 4-Bit Binary Codes*

Decimal Number	*8421*	*5421*	*2421*	*9's Complement of 2421 Code*
0	0000	0000	0000	1111
1	0001	0001	0001	1110
2	0010	0010	0010	1101
3	0011	0011	0011	1100
4	0100	0100	0100	1011
5	0101	1000	1011	0100
6	0110	1001	1100	0011
7	0111	1010	1101	0010
8	1000	1011	1110	0001
9	1001	1100	1111	0000

BCD or 8421 Code: The Binary Coded Decimal (BCD) uses the binary number system to specify the decimal numbers 0 to 9. It has four bits. The weights are assigned according to the positions occupied by these digits. The weights of the first (right-most) position is 2^0(1), the second 2^1 (2), the third 2^2(4), and the fourth 2^3 (8). Reading from left to right, the weights are 8-4-2-1, and hence it is called 8421 code.

The binary equivalent of 7 is $[111]_2$, but the same number is represented in BCD in 4-bit form as $[0111]_2$. Also, the numbers from 0 to 9 are represented in the same way as in the binary system, but after 9 the representations in BCD are different. For example, the decimal number 12 in the binary system is $[1100]_2$ but the same number is represented as [0001 0010] in BCD.

Example 1: Give the BCD Code for the decimal number 874.

Solution:

Decimal number $\rightarrow$ 874

BCD code $\rightarrow$ 1000 0111 0100

Hence, $(874)_{10} = (1000\ \ 0111\ \ 0100)_{BCD}$

Example 2: Give the BCD code equivalent for the decimal number $96 \cdot 42$.

Solution :

$$\begin{array}{llllll} \text{Decimal number} \rightarrow & 9 & 6 & 4 & 2 \\ & \downarrow & \downarrow & \downarrow & \downarrow \end{array}$$

BCD Code $\rightarrow$ 1001011001000010

Hence, $[96 \cdot 42]_{10} = [10010110 \cdot 01000010]_{BCD}$.

2421 Code: This is a weighted code; its weights are 2, 4, 2 and 1. A decimal number is represented in 4-bit form and the total weight of the four bits = 2 + 4 + 2 + 1 = 9. Hence, the 2421 code represents decimal numbers from 0 to 9. Upto 4, the 2421 code is the same as that in BCD; however, it varies for digits from 5 to 9. This code is also a self-complementing code i.e. the 9's complement of a number 'N' is obtained by complementing the 0s and 1s in the code word 'N'. For example, the 2421 code for 3 is 0011 and its natural complement 1100 gives 6 which is the 9's complement of 3. Table 1.3 gives the 2421 code of the decimal numbers and its complement. The bit combination 1101, when weighted by the reflective digits 2421, gives the decimal equivalent of $2\times1+4\times1+2\times0+1\times1=$ 2 + 4 + 0 + 1 = 7.

NOTES

Reflective Codes: A code is said to be *reflective* when the code for 9 is the complement of the code for 0, 8 for 1, 7 for 2, 6 for 3, and 5 for 4. While the 2421, 5211 and Excess-3 codes are reflective codes, the 8421 code is not. While finding the 9's complement, such as in 9's complement subtraction, reflectivity is desirable in a code.

Sequential Codes: A code can be said to be sequential when each succeeding code is one binary number greater than its preceding code. This greatly helps mathematical manipulation of data. While the 8421 and Excess-3 codes are sequential, the 2421 and 5421 codes are not.

Non-Weighted Codes

Non-weighted codes are codes that are not positionally weighted. This means that each position within a binary number is not assigned a fixed value. Excess-3 codes and Gray codes are examples of non-weighted codes.

Excess-3 Code: As the name indicates, the *Excess-3* represents a decimal number, in binary form, as a number greater than 3. An Excess-3 code is obtained by adding 3 to a decimal number. For example, to encode the decimal number 6 into an Excess-3 code, we must first add 3 in order to obtain 9. The 9 is then encoded in its equivalent 4-bit binary code 1001. The Excess-3 code is a self-complementing code, and this helps in performing subtraction operations in digital computers, especially in the earlier models. The Excess-3 code is also a reflective code.

Example 3: Convert $[643]_{10}$ into its Excess-3 code.

Solution:

$$\begin{array}{lccc} \text{Decimal number} & 6 & 4 & 3 \\ \text{Add 3 to each bit} & +3 & +3 & +3 \\ \text{Sum} \;\; \rightarrow & 9 & 7 & 6 \end{array}$$

Converting the above sum into its BCD code, we have

$$\begin{array}{llll} \text{Sum} & \rightarrow 9 & 7 & 6 \\ & \;\;\downarrow & \downarrow & \downarrow \\ \text{BCD} & \rightarrow 1001 & 0111 & 0110 \end{array}$$

Hence, the Excess-3 code for $[643]_{10}$ is 1001 0111 0110.

Table 5.6 lists the BCD, Excess-3 code and 9's complement representations for decimal digits. Note that both codes use only 10 of the 16 possible 4-bit code groups. The excess-3 code, however, does not use the same code groups. Its invalid code groups are 0000, 0001, 0010, 1101, 1110, and 1111.

NOTES

Table 5.6 Excess-3 Codes

Decimal	*8421(BCD) Code*	*Excess-3 Code*	*9's Complement*
0	0000	0011	1100
1	0001	0100	1011
2	0010	0101	1010
3	0011	0110	1001
4	0100	0111	1000
5	0101	1000	0111
6	0110	1001	0110
7	0111	1010	0101
8	1000	1011	0100
9	1001	1100	0011

Gray Code: The *Gray code* belongs to a class of codes called *minimum-change codes*, in which only one bit in the code group changes when moving from one step to the next. The Gray code is a *non-weighted code*. Therefore, it is not suitable for arithmetic operations but finds applications in input/output devices and in some types of analog-to-digital converters. The Gray code is a reflective digital code which has a special property of containing two adjacent code numbers that differ by only one bit. Therefore, it is also called a *unit-distance code*.

Table 5.7 shows the Gray code representation for the decimal numbers 0 to 15, together with the straight binary code.

Table 5.7 Gray Code

Decimal Numbers	*Binary Code*	*Gray Code*
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

Binary Coded Decimal

The Binary Coded Decimal (BCD) is a combination of four binary digits that represent decimal numbers. For example, the 8421 code is a type of binary coded decimal. It has four bits and represents the decimal digits 0 to 9. The numbers 8421 indicate the binary weights of the four bits. The ease of conversion between the 8421 code numbers and the familiar decimal numbers is the main advantage of this code. To express any decimal number in BCD, each decimal digit should be replaced by the appropriate four-bit code. Table 5.8 gives the binary and BCD codes for the decimal numbers 0 to 15.

NOTES

Table 5.8 *Decimal Numbers, Binary Equivalents and BCD*

Decimal Number	*Binary Number*	*Binary Coded Decimal (BCD)*
0	0000	0000
1	0001	0001
2	0010	0010
3	0011	0011
4	0100	0100
5	0101	0101
6	0110	0110
7	0111	0111
8	1000	1000
9	1001	1001
10	1010	0001 0000
11	1011	0001 0001
12	1100	0001 0010
13	1101	0001 0011
14	1110	0001 0100
15	1111	0001 0101

BCD Addition

BCD is a numerical code. Many applications require arithmetic operations. Addition is the most important of these because the other three operations, namely subtraction, multiplication and division, can be done using addition. The rule for addition of two BCD numbers is given below.

(i) Add the two numbers using the rules for binary addition.

(ii) If a four-bit sum is equal to or less than 9, it is a valid BCD number.

(iii) If a four-bit sum is greater than 9, or if a carry-out of the group is generated, it is an invalid result. Add 6 (0110_2) to the four-bit sum in order to skip the six invalid states and return the code to BCD. If a carry results when 6 is added, add the carry to the next four-bit group.

Example 4 : Add the following BCD numbers: (i) 1001 and 0100 and (ii) 00011001 and 00010100.

Solution: The following in the method for addition.

(i)

$$\begin{array}{r} 1\;0\;0\;1 \\ +\,0\;1\;0\;0 \\ \hline 1\;1\;0\;1 \\ +\,0\;1\;1\;0 \\ \hline \underbrace{0\;0\;0\;1}_{1}\;\;\underbrace{0\;0\;1\;1}_{3} \end{array}$$

(ii)

$$\begin{array}{r} 0\;0\;0\;1\quad 1\;0\;0\;1 \\ +\,0\;0\;0\;1\quad 0\;1\;0\;0 \\ \hline 0\;0\;1\;0\quad 1\;1\;0\;1 \\ +\,0\;1\;1\;0 \\ \hline \underbrace{0\;0\;1\;1}_{3}\quad \underbrace{0\;0\;1\;1}_{3} \end{array}$$

BCD Subtraction

Method I: Table 5.9 shows an algorithm for BCD subtraction. The 1's complement of the BCD subtrahend is entered into adder 1, and the complement (true) of the result is transferred to adder 2, where either a 1010 or 0000 is added, depending on the sign of the total result. Examples of a positive and negative total result are given in Table 5.9. Arrows indicate EAC (End-Around-Carry) or carry to the next decade.

NOTES

Table 5.9 *Algorithm for BCD Subtraction*

Decade Result	*Sign of Total Result*	
	(+) EAC = 1	*(–) EAC = 0*
	Transfer true results of adder 1	Transfer 1's complement of result of adder 1
C_n=1	0000 added in adder 2	1010 added in adder 2
C_n= 0	1010 added in adder 2	0000 added in adder 2

Total Result Positive:

```
              10^2     10^1  10^0     EAC indicates a (+) total result
       835    1000     0011  0101
      −274    1101     1000  1011 ←   1's Complement of 001001110100
      +561  1 0101     1011  0000
                          1     1
                       1100  0001     Transfer true output of adder 1
              0101     1100  0001
              0000     1010  0000
              0101  1  0110  0001
                    ↑————————————————Ignore this carry 1
                 5        6     1
```

Total Result Negative:

```
        10^2  10^1    10^0
   429  0100  0010    1001     No EAC indicates (–) total result
 − 476  1011  1000    1001 ←   1's Complement of 01000111 0110
  − 47  1111  1010    0010
          ↓      1      ↓
              1011
        0000  0100    1101     Transfer 1's Complement of adder 1 output
        0000  0000    1010
        0000  0100  1 0111
                    ↑——————— Ignore this carry
```

Method II: Another method in BCD subtraction is the addition of the 9's complement of the subtrahend to the minuend.

Example 5 : Subtract 748 from 983 using 9's complement method.

Solution:

9's complement of 748 = 999 − 748 = 251

$$\begin{array}{r} 999 \\ -\ 748 \\ \hline 251 \\ \hline \end{array}$$

Direct Method

```
  983          983
− 748        + 251   ← 9's Complement of 748
  235      1   234
                 1   ← EAC
               235
```

5.3 PULSE CODE MODULATION

NOTES

Pulse Code Modulation (PCM) is a digital representation of an analog signal where the magnitude of the signal is sampled regularly at uniform intervals, then quantized to a series of symbols in a digital (usually binary) code. PCM has been used in digital telephone systems and is also the standard form for digital audio and video in computers.

The following are the abbreviations and their expanded forms:

- **PCM:** Pulse Code Modulation
- **DPCM:** Differential Pulse Code Modulation
- **ADPCM:** Adaptive Differential Pulse Code Modulation

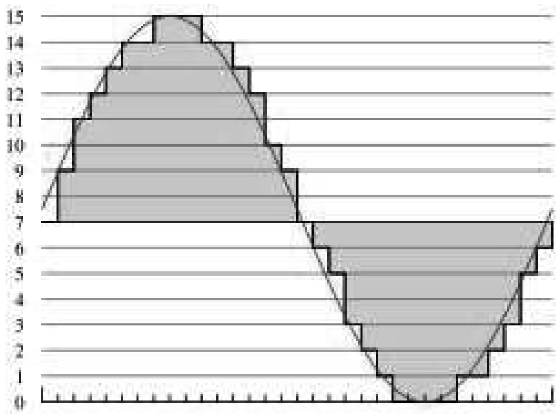


Fig. 5.6 *Sampling and Quantization of a Signal for 4-Bit PCM*

In Figure 5.6, a sine wave (curve) is sampled and quantized for PCM. The sine wave is sampled at regular intervals, as shown. For each sample, one of the available values is chosen by some algorithm or function. Several PCM streams could also be multiplexed into a larger aggregate data stream, generally for transmission of multiple streams over a single physical link. This technique is called time-division multiplexing, or TDM, and is widely used, notably in the modern public telephone systems. There are many ways to implement a real device that performs this task. In real systems, such a device is generally referred to as an ADC (Analog-to-Digital Converter) which produces a binary representation of the input, which would then be read by a processor of some sort.

Demodulation: The electronics involved in producing an accurate analog signal from the discrete data are similar to those used for generating the digital signal. These devices are DACs (Digital-to-Analog Converters), and operate similarly to ADCs. They produce on their output a voltage or current (depending on type) that represents the value presented on their inputs. This output would then generally be filtered and amplified for use.

Differential Pulse Code Modulation

Differential Pulse Code Modulation (DPCM) is a procedure of converting an analog into a digital signal in which an analog signal is sampled and then the difference between the actual sample value and its predicted value (predicted value is based on previous sample or samples) is quantized and then encoded forming a digital value.

NOTES

DPCM code words represent differences between samples unlike in PCM where code words represented a sample value. Basic concept of DPCM—coding a difference, is based on the fact that most source signals show significant correlation between successive samples so encoding uses redundancy in sample values which implies lower bit rate.

DPCM compression depends on the prediction technique, well-conducted prediction techniques lead to good compression rates, in other cases DPCM could mean expansion comparing to regular PCM encoding. Figure 5.7 shows DPCM encoder (transmitter).

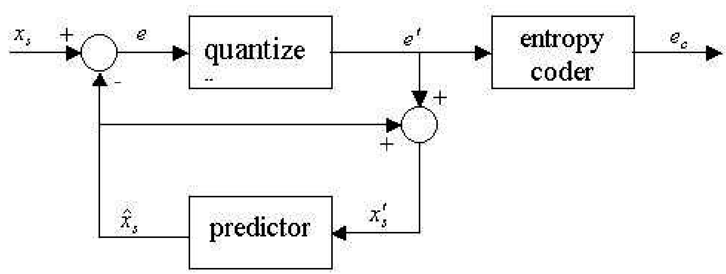


Fig. 5.7 *DPCM Encoder (Transmitter)*

DPCM conducted on signals with correlation between successive samples leads to good compression ratios.

Advantages of DPCM over PCM

(i) Very less number of levels are required to encode: 8 level PCM is converted into 4 level DPCM; so instead of 3 Bits now we require only 2 Bits for coding. So the required bandwith is reduced.

(ii) We see that designing of system is very easy. At the output we do not need any digital to analog conversion as in the case of PCM.

(iii) The correlation between successive samples is very good; so the output samples are far better than that of PCM.

Adaptive Differential Pulse Code Modulation

Adaptive Differential Pulse Code Modulation (ADPCM) is a waveform coding technique that attempts to code signals without any knowledge about how the signal was created. This implies that a waveform coder can be applied to other forms of data besides speech (e.g., video). In general, these coders are simple, with bit rates above 16 kbps. Anything lower degrades the reconstructed speech.

ADPCM is based on two principles of speech. Because there is a high correlation between consecutive speech samples, a relatively simple algorithm could be used to predict what the next sample might be, based on previous samples.

When the predicted sample was compared to the real sample, it was found that the resulting error signal had a lower variance than the original speech samples and could therefore be quantized with fewer bits. It was also found that no extra information about the predictor needed to be sent if the prediction was based on the quantized samples rather than on the incoming speech signal. The result was DPCM. Further studies showed that if the predictor and quantizer were made to be adaptive (i.e., that smaller samples are quantized using smaller steps and larger samples are quantized with larger steps), then the reconstructed speech more closely matched the original speech.

5.3.1 PCM Generation

Modulation is the process of varying one or more parameters of a carrier signal in accordance with the instantaneous values of the message signal. The message signal is the signal which is being transmitted for communication and the carrier signal is a high frequency signal which has no data, but is used for long distance transmission. There are numerous modulation techniques, which are classified according to the type of modulation employed. Among them all, the digital modulation technique used is Pulse Code Modulation (PCM). A signal is pulse code modulated to convert its analog information into a binary sequence, i.e., 1's and 0's. The output of a PCM resembles a binary sequence. Figure 5.8 shows an example of PCM output with respect to instantaneous values of a given sine wave.

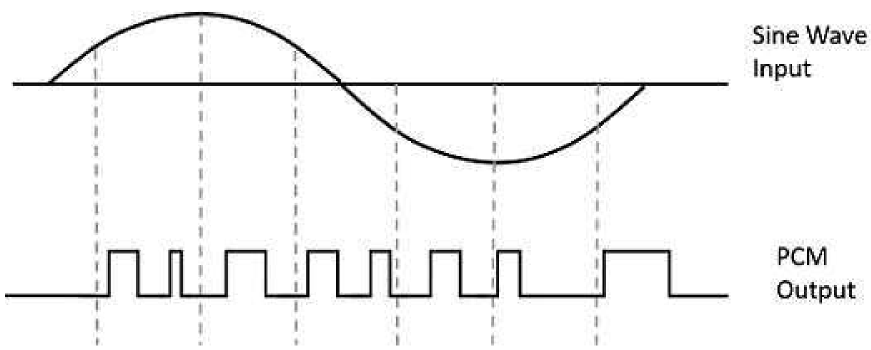


Fig. 5.8 *PCM Output with Respect to Instantaneous Value of Sine Wave*

Instead of a pulse train, PCM produces a series of numbers or digits. Each one of these digits, though in binary code, represent the approximate amplitude of the signal sample at that instant. In Pulse Code Modulation or PCM the message signal is represented by a sequence of coded pulses. This message signal is achieved by representing the signal in discrete form in both time and amplitude.

Basic Elements of PCM

The transmitter section of a pulse code modulator circuit consists of sampler, quantizer and encoder, which are performed in the analog-to-digital converter section. The low pass filter prior to sampling prevents aliasing of the message signal. The basic functions performed in the receiver section are regeneration of weakened signals, decoding, and reconstruction of the quantized pulse train. Following is the block diagram of PCM which represents the basic elements of both the transmitter and the receiver sections.

NOTES

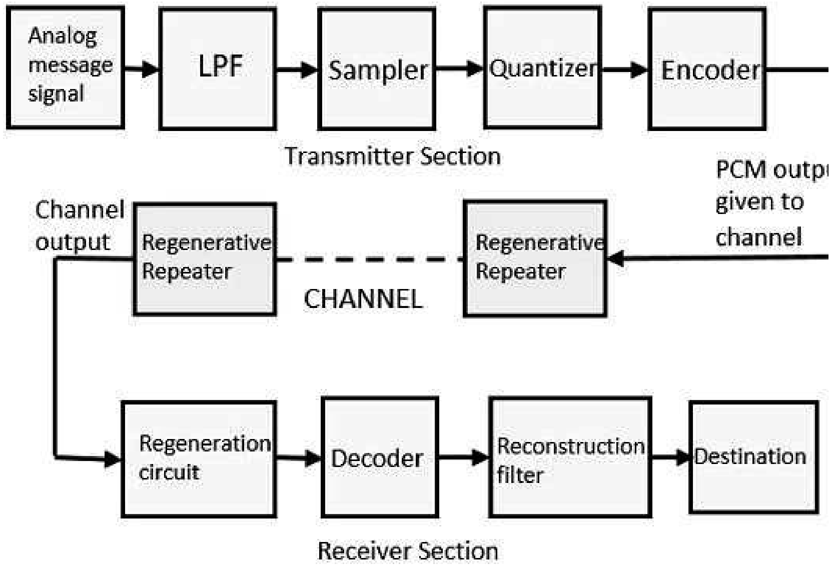


Fig. 5.9 *Block Diagram of PCM*

Low Pass Filter (LPF)

The function of low pass filter is to eliminate the high frequency components present in the input analog signal which is greater than the highest frequency of the message signal, to avoid aliasing of the message signal.

Sampler

Sampler is the technique which assist to collect the sample data at instantaneous values of message signal, so as to re-form the original signal. The sampling rate must be greater than twice the highest frequency component W of the message signal, in accordance with the sampling theorem.

Quantizer

Quantizing is a process in which excessive bits are getting reduced and the data is confined. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.

Encoder

The function of an encoder is to perform digitization of analog signal. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. LPF, Sampler, and Quantizer are three sections that acts as an analog to digital converter. Encoding minimizes the bandwidth used.

Regenerative Repeater

The function of regenerative repeater is to enhance the signal strength. The output of the channel also has one regenerative repeater circuit, whose role is to compensate the signal loss, reconstruct the signal, and also to increase its strength.

Decoder

The decoder circuit decodes the pulse coded waveform to reproduce the original signal. This circuit acts as the demodulator.

Reconstruction Filter

Once the digital-to-analog conversion is done by the regenerative circuit and the decoder, a low-pass filter is employed, called as the reconstruction filter whose function is to get back the original signal.

Hence, the Pulse Code Modulator circuit digitizes the given analog signal, codes it and samples it, and then transmits it in an analog form. This whole process is repeated in a reverse pattern to obtain the original signal.

5.3.2 Delta Modulation and PCM

Delta Modulation (DM or Δ-modulation) is an analog-to-digital and digital-to-analog signal conversion technique used to transmit voice data. In this method, the quality is not of primary importance. This is the simplest form of Differential Pulse-Code Modulation (DPCM) where the difference between successive samples is encoded into n-bit data streams. In DM, the transmitted data is reduced to a 1-bit data stream.

Its main features are:

- The analog signal is approximated with a series of segments.
- Each segment of the approximated signal is compared to the original analog wave to determine the increase or decrease in relative amplitude.
- This comparison determines the decision process for establishing the state of successive bits.
- Only the change of information—an increase or decrease of the signal amplitude from the previous sample—is transmitted whereas a no-change condition leaves the modulated signal to remain at the same 0 or 1 state of the previous sample.

Delta modulation must use oversampling techniques to achieve a high signal-to-noise ratio. In other words, the analog signal is sampled at a rate several times higher than the Nyquist rate.

Derived forms of delta modulation are continuously variable slope delta modulation, delta-sigma modulation, and differential modulation. The Differential Pulse Code Modulation or DPCM is the super set of DM.

Principle: Instead of quantizing the absolute value of the input analog waveform, delta modulation quantizes the difference between the present and the previous step, as shown in Figure 5.10.

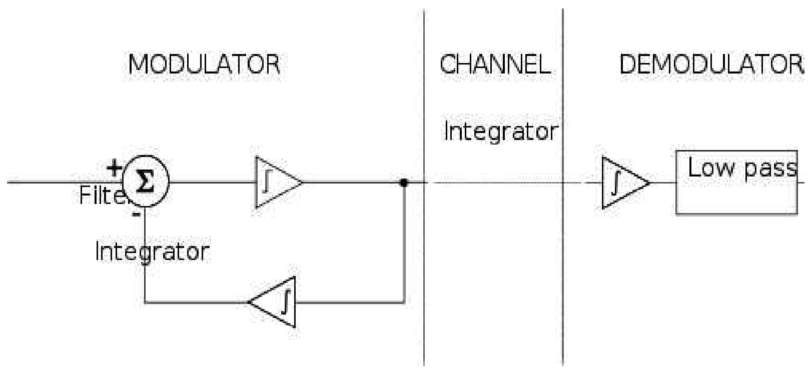


Fig. 5.10 Block Diagram of a Δ-Modulator and Demodulator

The modulator contains a quantizer that converts the difference between the input signal and the average of the previous steps. The quantizer, in its simplest form, can be realized with a comparator referenced to 0 (two levels quantizer), whose output is 1 if the input signal is positive or or 0 if the input is negative. This modulator is also known as a bit-quantizer as it quantizes only one bit at a time. The demodulator is simply an integrator that itself constitutes a low-pass filter (like the one shown in the feedback loop) whose output rises or falls with each 1 or 0 received.

Transfer Characteristics: The transfer characteristics of a DM system follow a signum function, as it quantizes only two levels and further one-bit at a time.

Output Signal Power: In DM there is no restriction on the amplitude of the signal waveform, because the number of levels is not fixed. On the other hand, there is a limitation on the slope of the signal waveform that must be observed if the slope overload is to be avoided. But, if the signal waveform changes slowly, there is no limit to the signal power that may be transmitted.

Bit-Rate: If the communication channel is of limited bandwidth, there is the possibility of interference in either DM or PCM (Pulse Code Modulation). Therefore, DM and PCM operate at same bit-rate.

Adaptive Delta Modulation (ADM): This is a modification of DM in which the step-size is not fixed. Rather, when slope overload occurs, the step-size becomes progressively larger. ADM reduces slope error at the expense of increasing quantizing error. This quantizing error can be reduced using a low pass filter.

Comparison of PCM and DM: Signal-to-noise ratio of DM is larger than signal-to-noise ratio of PCM. For an ADM, signal-to-noise ratio is comparable to signal-to-noise ratio of companded PCM.

5.3.3 PCM Bandwidth and PCM Reception Noise

Pulse Code Modulation (PCM) is a technique for digitally representing sampled analogue signals. It is the industry-standard digital audio format used in computers, compact discs, digital telephony, and other digital audio applications. A PCM stream samples the analogue signal's amplitude at regular intervals and quantizes each sample to the nearest value within a range of digital steps.

The bandwidth of a serial PCM transmission is determined by the bit rate and the form of the pulses. R is the bit rate.

$$R = nf_s$$

n = Number of Bits on the PCM Word.

$M = 2^n$; f_s = Sampling Rate.

For no aliasing: $f_s = 2B$, where B is the bandwidth of the analog signal that is to be converted.

Bandwidth of the PCM waveform is bounded by:

$$B_{PCM} \geq ½ (R) = ½ (nf_s)$$

Equality is obtained if a sin x/x pulse shape is used to generate the PCM waveform.

For one using a rectangular pulse with polar NRZ line codes:

$$B_{PCM} = R = nf_s \text{ (First Null Bandwidth)}$$

The bandwidth of the PCM signal has a bound given by:

$$B_{PCM} = nB$$

For appropriate values of n, the PCM signal's bandwidth will be much greater than the bandwidth of the analogue signal it replicates.

NOTES

5.3.4 PCM Reception Noise

The relationship between the recovered analogue peak signal power and the total average noise power is as follows:

$$\left(\frac{S}{N}\right)_{pk-out} = \frac{3M^2}{1+4(M^2-1)P_e}$$

and the ratio of average signal power to average noise power is as follows:

$$\left(\frac{S}{N}\right)_{out} = \frac{M^2}{1+4(M^2-1)P_e}$$

Where M is the number of quantized levels in the PCM system and P_e denotes the probability of bit error in the recovered binary PCM signal before it is transformed back to an analogue signal at the receiver DAC. P_e is minimal in most practical situations. As a consequence:

$$\left(\frac{S}{N}\right)_{pk-out} = 3M^2$$

and

$$\left(\frac{S}{N}\right)_{out} = M^2$$

Expressing in *dB*

$$\left(\frac{S}{N}\right)_{dB} = 6.02n + \alpha$$

where n is the number of bits in the PCM word, $\alpha = 4.77$ for the peak SNR and $\alpha = 0$ for the average SNR.

This is referred to as the 6 *dB* rule, and it highlights a significant performance property of PCM: each bit added to the PCM word results in an additional 6 *dB* improvement in SNR.

5.3.5 Quantization Noise Analysis

A fundamental control system is composed of the plant (the system being controlled), a sensor that detects the output variable (physical quantity) and converts it to an electrical quantity that the controller uses to drive the plant's actuator. Numerous noises and disturbances may develop in this closed loop system. Fundamentally, noises are categorized into two categories based on their behavior: systematic

NOTES

noise and random noise. Random noise encompasses all forms of disruptions and noises that exhibit nondeterministic behavior and are typically difficult to forecast their occurrence. Sensor noise, plant disturbances, and human mistake are all examples of this. Systematic noises, on the other hand, are those that can be foreseen, mathematically modelled, and compensated for during analysis and design. Along with all the components of a basic control system, digital control systems include some components that enable analysis in the digital realm. It comprises of ADCs and DACs, and the compensator may be implemented using a digital computer, depending on the system. The operation and analysis of ADCs and DACs is a well-documented and well-researched subject. ADCs and DACs rely heavily on quantization and sampling. The quantization noise is a type of systematic noise that occurs in digital control systems.

During Analog to Digital Conversion, the continuous (analogue) signal is 'Sampled' at specified time intervals and truncated (or quantized) to the nearest amplitude level. This is necessary because signal processing in the digital domain must occur in discrete time (hence the need for sampling) and the digital processor cannot have infinite resolution. As a result, the amount of quantization error (rounding off error) that happens in a system is proportional to the digital processor's resolution. To represent this resolution mathematically, we use the variable q to symbolize one LSB (Least Significant Bit). Thus, if a system has a greater resolution or a lower value of q, the quantization noise will be smaller. It is self-evident that the greatest quantization error will be half of one LSB, or $\pm q/2$. A quantizer is a device that performs quantization; it is seen in the block diagram format in Figure 5.11. Although the set of potential values for the quantizer's input may be endlessly big, the quantizer's output will always be a finite or countably infinite set of numbers. The quantizer's input signal is denoted by x, while its output signal is denoted by 'x'. Quantization noise is denoted by ρ, where denotes the difference between the output and the input. This could be modelled in a variety of ways. Quantizer input output waveforms resemble staircases, indicating unequivocally that quantization is a nonlinear operation. Analog-to-Digital Converters (ADCs) and digital-to-analog converter the quantization noise is a type of systematic noise that occurs in digital control systems.

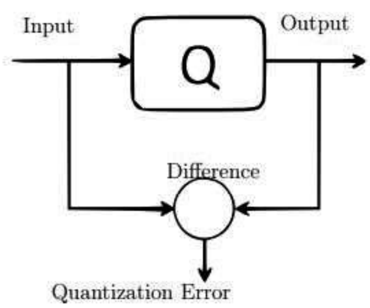


Fig. 5.11 *A Quantizer's Basic Representation*

Quantizing Theorem I and II

B. Widrow introduced the statistical theory of quantization in which he demonstrated how quantization noise in a digital control system or any other application involving

NOTES

digital signal processing could be modelled. According to the Quantizing Theorems I, if the CF (Characteristic Function, $\tilde{O}$) of the Probability Density Function (PDF) of the quantizer input x is band restricted (as determined by the Fourier transform (variable 'u') of the input signal PDF), then:

$$\tilde{O}_x(u) = 0; \ |u| > \pi/q$$

Then, the CF of the quantizer's input may be uniquely deduced from the CF of the quantizer's output x; the same is true for the PDF of the signals. In essence, the theorem implies that the PDFs of the Quantizer's input and output are uniquely connected. On the other hand, the Quantizing Theorem II establishes an essential and stronger result: the moment of the quantized variable (the quantizer's output) is equal to the sum of the input and a uniformly distributed noise with a zero mean and a mean square equal to $q^2/12$.

The following Figure 5.12 illustrates the quantization of an analogue signal. The blue line indicates an analogue signal, while the brown line indicates a quantized signal.

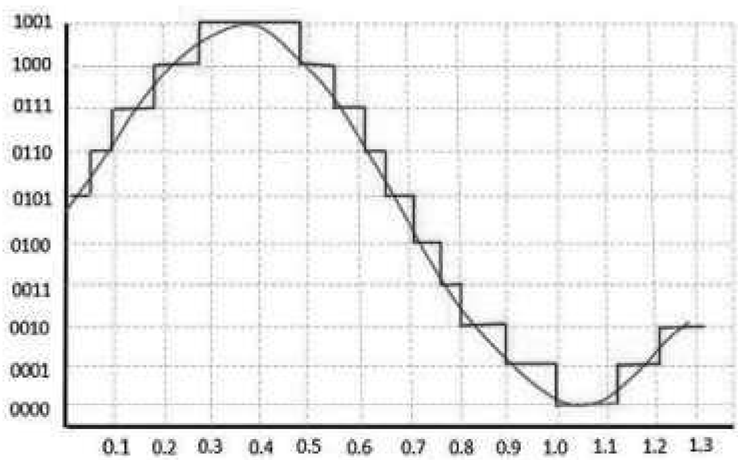


Fig 5.12. *Quantization Process*

Samples and quantization both result in the loss of information, but they are different. How good a Quantizer's output is depends on how many quantization levels it has been set up with. When the output is quantized, it has different amplitudes at each level, which are called representations or reconstructions. It is called a quantum or step-size when there is a space between two representation levels.

Quantization is classified into two categories.

- Uniform Quantization
- Non-Uniform Quantization

A uniform quantization is a quantization technique in which the quantization levels are consistently spaced. Non-uniform Quantization is a method of quantization in which the quantization levels are mismatched and the relationship between them is mostly logarithmic.

NOTES

Uniform Quantization

There are two distinct types of uniform quantization:

- Mid-Rise Type
- Mid-Tread type

The two types of uniform quantization are illustrated in the following Figure 5.13.

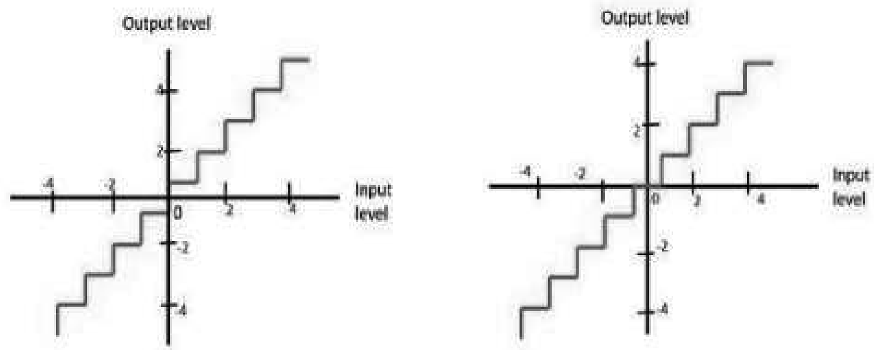


Fig. 5.13 *Mid-Rise Type and Mid-Tread Type Uniform Quantizations*

The Mid-Rise type is so named because the origin is located in the middle of a stair-case-like graph's ascending section. This type has an even number of quantization levels. The Mid-tread type is so named because the origin is located in the centre of a stair-case-like graph's tread. This type has an odd number of quantization levels. The mid-rise and mid-tread uniform quantizers are both symmetric about the origin.

Non-Uniform Quantization

The step size is not fixed in non-uniform quantization. It varies according to a particular law or the amplitude of the input signal. The following Figure 5.14 illustrates the properties of a non-uniform quantizer.

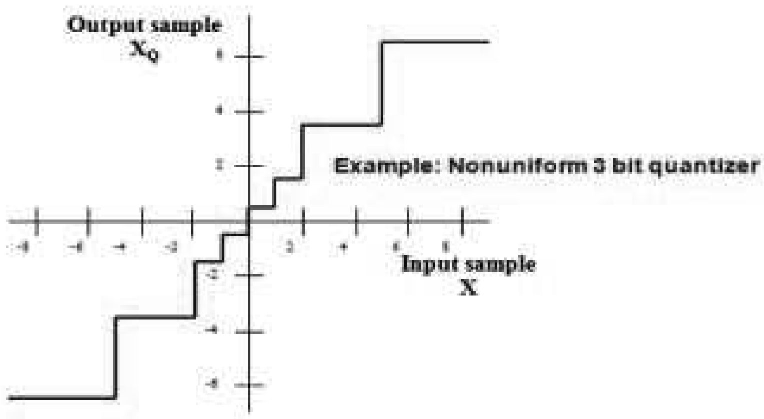


Fig. 5.14 *Properties of a Non-Uniform Quantizer*

NOTES

At low input signal levels, the step size is tiny. As a result, quantization is insignificant at these input levels. As a result, at low signal levels, the signal to quantization noise power ratio is enhanced. At high input levels, the step size is increased. As a result, the signal to noise power ratio remains nearly constant across the quantizer's dynamic range.

5.3.6 SIN Ratio and Channel Capacity of PCM

An important quantity in information theory and telecommunications engineering for defining theoretical upper bounds on channel capacity (or information transfer rate) in wireless communication systems, such as networks, is the Signal to Interference Plus Noise (SINR), also known as the Signal to Noise Plus Interference (SNIR). SINR is defined as the power of an interest signal divided by the total of the interference power (from all other interfering signals) and some background noise. The SINR equals the signal-to-interference ratio when the noise power term is zero (SIR). A mathematical model of wireless networks, such as cellular networks use the SINR, which is a more commonly used value than the SNR when it comes to simulating these networks.

To define the quality of wireless communications, the phrase "Signal-to-noise Ratio (SINR)" is commonly used in wireless communication. Route loss in wireless networks occurs when a signal loses energy as it travels farther. In contrast, the correct receipt of data in wired networks is determined by the presence of a wired path between the sender or transmitter and the receiver. Additional factors must be taken into account while designing a wireless network (e.g., the background noise, interfering strength of other simultaneous transmission). This component has been conceptualized using SINR.

For example, if a receiver is positioned at x in space (often on the plane), its associated SINR is given by:

$$SINR(x) = \frac{P}{I+N}$$

The incoming signal of interest is represented by P, while I represents the strength of all other (interfering) signals in the network and N represents a noise factor that can be either constant or unpredictable. As is the case with many other ratios in electronics and related domains, the SINR is frequently stated in decibels or dB.

Channel Capacity

Assume a source sends r messages every second and each message has an entropy of H bits. $R = r\,H$ bits/second is the information rate. One would intuitively conclude that when the information rate of a communication system grows, the number of errors per second will increase as well. Surprisingly, this is not true.

Shannon's Theorem

- Each communication system has a maximum data transmission rate C, referred to as the channel capacity.

NOTES

- If the information rate R is less than C, then employing intelligent coding techniques, one can reach arbitrarily reduced error probability.
- To achieve a decreased probability of mistake, the encoder must deal with larger blocks of signal data. This results in increased delays and computing requirements.

Thus, if $R \leq C$, transmission is possible in the presence of noise without error. Regrettably, Shannon's theorem is not a constructive proof; it only asserts the existence of such a coding scheme. As a result, the proof cannot be used to build a coding scheme capable of reaching the channel capacity. The converse of this theorem is also true: if $R > C$, no matter what coding technique is chosen, errors cannot be prevented.

Shannon-Hartley Theorem

Consider the band-limited Gaussian channel illustrated in Figure 5.15 running in the presence of additive Gaussian noise.

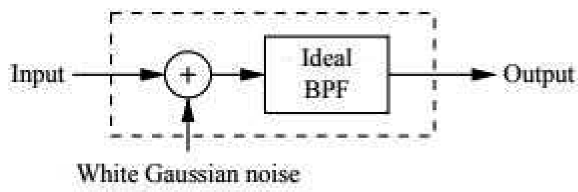


Fig. 5.15 *Band Limited Gaussian Channel*

According to the Shannon-Hartley theorem, the channel capacity is equal to

$$C = B \log_2(1 + S/N)$$

Where C denotes the channel's capacity in bits per second, B denotes the channel's bandwidth in Hertz, and S/N is the Signal-to-Noise ratio.

Although the theorem cannot be proved, it can be supported in part as follows:

Assume the received signal is accompanied by noise with an RMS value of σ and has been quantized with levels separated by $a = \lambda\sigma$. If λ is large enough, it should be able to distinguish the signal level with a low chance of error. Additionally, suppose each message is to be represented by a single voltage level. If there are to be M distinct messages, there must also be M distinct levels. Thus, the average signal power is as follows:

$$S = \frac{M^2 - 1}{12}(\lambda\sigma)^2.$$

As a result, the number of levels for a given average signal power is

$$M = \left(1 + \frac{12}{\lambda^2}\frac{S}{N}\right)^{1/2},$$

Where $N = \sigma^2$ is the noise power. Each communication conveys an equal amount of information if they are equally likely;

$$H = \log_2 M = \frac{1}{2}\log_2\left(1 + \frac{12}{\lambda^2}\frac{S}{N}\right) \text{ bits/message.}$$

NOTES

To determine the information rate, one must estimate the number of messages that a signal on a channel can carry per unit time. To illustrate the heuristic nature of the topic, imagine that the response of an ideal LPF with bandwidth B to a unit step has a rise time of 10–90 percent of $\tau = 0.44/B$. As a result, it is expected that with $T = 0.5/B \approx \tau$, one should be able to estimate the level reliably. The message rate is then increased.

$$r = \frac{1}{T} = 2B \text{ messages/s.}$$

The rate at which data is transmitted through a channel is therefore

$$R = rH = B\log_2\left(1 + \frac{12}{\lambda^2}\frac{S}{N}\right)$$

This is comparable to the Shannon-Hartley theorem if and only if these are equal to $\lambda = 3.5$. The Shannon-Hartley theorem states that with sufficiently advanced coding techniques, transmission at channel capacity can occur with arbitrarily tiny error.

The expression for the Gaussian channel's channel capacity makes intuitive sense:

- As the channel's bandwidth rises, it becomes easier to make more rapid changes to the information signal, increasing the information rate.
- As S/N grows, the information rate can be increased while still avoiding noise-induced mistakes.
- In the absence of noise, $S/N \to \infty$ and an infinite information rate are achievable, regardless of bandwidth.

As a result, bandwidth and SNR may be traded off. For instance, if S/N = 7 and B = 4 kHz, the channel capacity is $C = 12 \times 10^3$ bits/s. The channel capacity remains constant when the SNR is increased to S/N = 15 and B is reduced to 3 kHz.

However, when $B \to \infty$ the channel capacity does not become limitless, as noise power increases proportionately. If the spectral density of noise power is $\eta/2$, the total noise power equals $N = \eta B$, and so the Shannon-Hartley law holds true.

$$C = B\log_2\left(1 + \frac{S}{\eta B}\right) = \frac{S}{\eta}\left(\frac{\eta B}{S}\right)\log_2\left(1 + \frac{S}{\eta B}\right)$$
$$= \frac{S}{\eta}\log_2\left(1 + \frac{S}{\eta B}\right)^{\eta B/S}.$$

Noting that

$$\lim_{x\to 0}(1+x)^{1/x} = e$$

NOTES

and defining x as $x = S/\eta B$, the channel capacity increases indefinitely as B increases.

$$C_\infty = \lim_{B\to\infty} C = \frac{S}{\eta}\log_2 e = 1.44\frac{S}{\eta}.$$

This provides the greatest data transmission rate feasible for a system of a given power but without regard for bandwidth constraints. The power spectral density can be expressed as a function of the equivalent noise temperature using the formula $\eta = kT_e q$.

A few of the Gaussian channel's general properties can be presented. Assume binary digits are being transmitted at a rate equal to the channel capacity: $R = C$. If the average signal strength is S, the average energy per bit is $E_b = S/C$, where 1/C is the bit time.

With $N = \eta B$,

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b}{\eta}\frac{C}{B}\right).$$

Rearranging,

$$\frac{E_b}{\eta} = \frac{B}{C}(2^{C/B} - 1).$$

The Figure 5.16 given below illustrates this relationship graphically.

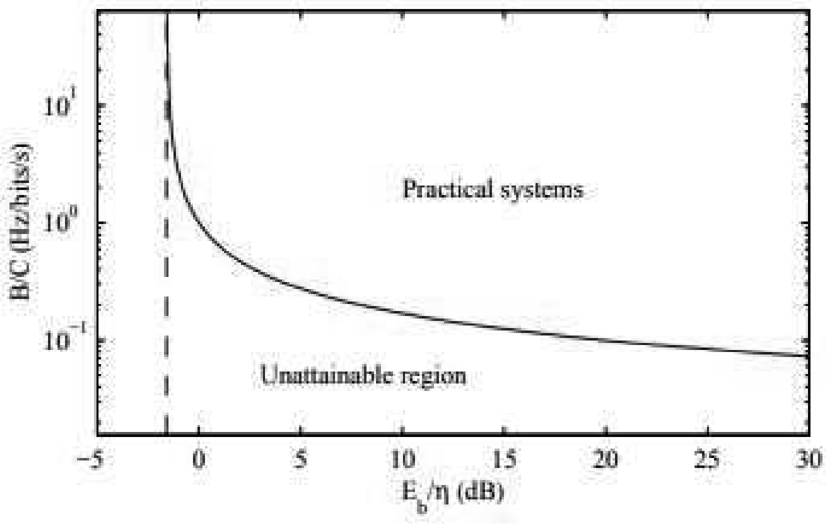


Fig. 5.16 *Graphical Relationship of Gaussian Channels*

As the asymptote is at $E_b/\eta = -1.59dB$, there is no error-free communication at any information rate below this value. This phenomenon is referred to as the Shannon limit.

NOTES

Check Your Progress

1. Define the term digital communications.
2. Differentiate between the asynchronous transmission and synchronous transmission.
3. What is Bit rate? Define multimedia bit rate.
4. In telecommunication, what is Data Signaling Rate (DSR)?
5. Why there is probability of error in a digital communication system?
6. What is digital filtering?
7. What are codes? How are they classified?
8. Define the terms Pulse Code Modulation (PCM) and Differential Pulse Code Modulation (DPCM).
9. What is Delta modulation?

5.4 CODES: ERROR DETECTION AND CORRECTION CODES

The memory and network technologies are not entirely reliable and are prone to errors while communicating data. Data processing and transmission systems experience errors due to several reasons. Some of the reasons are:

(a) Electrostatic interference can be caused from circuits or machines lying close by.

(b) Inductance and capacitance, loss in transmission due to leakages, impulses from static electricity in the atmosphere, etc., lead to distortion.

(c) Resistance to current in a cable can be caused due to attenuation.

Briefly, the main constraints are due to the physical medium that produces noise, distortion and attenuation or dropouts. However, the digital signals do not suffer from noise or distortion but they are susceptible to dropouts. The simplest way of detecting errors is to add up the sum of a certain number of words and embed the sum with the data. If the received data does not add up to the so-called checksum, the data has an error. However, this approach is good in detecting errors but this cannot reconstruct the original signal. Another way to minimize the effect of data loss is to use the interleaving process in which bits of each word are spread out between the bits of previous and subsequent words to reduce the chances that all the bits of any word are lost. This makes reconstruction of the data easier. Thus, the error detection and correction techniques involve the approaches to improve the ability to determine what the original data was after a loss or attenuation by using mathematical processes. A set of techniques that involves coding helps in detecting errors in stored or transmitted data, and rectify them. The use of several parity bits helps to detect any error, find if any bits are inverted, and incase they are, they should be re-inverted so that the original data is restored. Adding extra bits add to chances of detecting and correcting multiple errors. There are different methods depending upon Single Error Correction, Double Error Detection (SECDEC).

NOTES

Types of Errors

Binary data in the form of bits is susceptible to interference caused by the transmission media or hardware in a data communication environment. A signal can be distorted in many ways including the electrical characteristics of the transmitter and receiver and the characteristics of the transmission media. A transmission cable has inductance, capacitance and resistance. The inductance and capacitance tend to distort the shape of the signal while resistance causes the amplitude of the signal to reduce and therefore loss of power. Consequently, bit 1 may change to 0 or vice versa. If there is single bit change, it is treated as single bit error while if there are multiple bit changes, it is known as burst error.

Single Bit Errors: They are the errors that corrupt single bits of a transmission and change a single bit from 1 to 0 or vice versa in the data unit comprising of a byte, character or packet. Some of the frequent causes of such errors are power surges and other interference. For example, an ASCII code with 0 added to the left is transmitted. There are chances that the receiving device may receive another string of 8 bits due to corruption in a single bit of the transmitted 8 bits code and interpret it differently to take action. Figure 5.17 shows an example of such a situation wherein an ASCII code 00010101 for NAK (negative acknowledgement) is transmitted. If it is received as 00000101, it will mean ENQ (Enquiry). When the probability of error for both 1 and 0 are same, the channel is called a binary symmetric channel and the probability of error in binary digits is known as the Bit Error Rate (BER).

00010101		00000101
(NAK)		(ENQ)
Sender		Receiver

Fig. 5.17 *Representation of Single Bit Error*

The likelihood of corrupting data due to the single bit error is quite low in case of the serial data communication system as compared to the parallel data transmission. For example, if a sender is transmitting data at the rate of 1 Mbps, then each bit will last for $1/10^6$ seconds or 1 ìsecond in the transmission media including the hardware. In such a short time gap, the single bit will be quite independent from getting corrupted from noise. In case of parallel wire transmission, the 8 wires used to transmit 8 bits (1 byte) are most likely to be affected with a single bit error.

Burst Errors: When errors occur in multiple consecutive bits, then they are called burst errors. It occurs two or more bits in a data unit are corrupted and change their state from 0 to 1 or vice versa. For example, in computer networks, sometimes packets are lost or corrupted due to a burst error. A typical cause of burst errors is interference, often from lightning or electrical discharge which produces a spike for a very short span but quite higher than the packet transmission time. This causes the multiple bits to alter in the packet or data unit. The number of bits that get affected depends upon the duration of the noise and the transmission time of the number of bits of the transmitted packet. If a packet with a data rate of 10 Kbps is affected by a noise of duration of say 1/100 second, then it may force 100 bits to get corrupted. If the data rate is 1 Mbps for the same noise, then there will be a likelihood of 10,000 bits getting affected.

Detection

Error detection is the first step towards error correction and is simpler than error correction. There are four types of error detection mechanism. These are:

(i) Redundancy

It is the simplest type of error detection mechanism in which the same data string is transmitted twice. The receiving device performs a bit for bit mapping of both the received data strings to detect whether both the received data strings are the same or not. Any discrepancy in the received data strings indicates error in the string. Sending a bit stream twice is considered an accurate error detection system. However, this system involves quite a lengthy process because bit strings are transmitted twice and bit by bit mapping takes enormous time.

To overcome this problem, a group of bits are appended after the data string instead of repeating the entire data string. The process of introducing these extra bits is called *redundancy*. The redundant bits enable detection of error. The redundant bits are discarded when the veracity of the data string is ensured. In the redundancy process, the data string to be transmitted is passed through a device that analyses it and appends redundant bits at the sending side. The bit stream including the redundant bit is then passed over the transmission media. When the bit stream, along with redundant bits, is received at the receiving end, it is put in a device that examines the entire data stream. If the data stream passes the checking criteria, the checking device discards the redundant bits and passes the remaining bits for further processing of information.

There are three types of redundancy processes. These are: parity check, Cyclic Redundancy Check (CRC) and checksum.

Parity Check

Parity check using a single parity bit is the simplest and least expensive error detection method in which a parity bit is appended to a block of data, normally at the end of a 7-bit ASCII character. It consists of even parity and odd parity methods. Even parity method consists of one additional bit in such a manner that the total number of 1s in the resultant bit stream (the original information plus a parity bit) becomes even. Similarly, the odd parity method consists of addition of one bit in the orginal bit stream in such a manner that there are an odd number of 1s in the resultant bit stream. Figures 5.18 (a) and (b) explain the even and odd parity checks. In case of even parity, if a computer receives a character with an odd number of 1s, it then determines an error and requests for a retransmission. However, this method fails when two bits change and an error remains undetectable by the computer.

NOTES

Data stream	Parity Bit
1001100100111001	0

Fig. 5.18 (a) *One-Bit Even Parity*

Data stream	Parity Bit
1001100100111001	1

Fig. 5.18 (b) *One-Bit Odd Parity*

In parity bit methods, the operation of the receiver is simple as the receiver needs to count only the number of 1s in the received data stream with additional parity bit. If an even number of 1-valued bits is found with an odd parity method, the receiver will know that at least one bit error has happened. However, the chances of multiple bit errors in a frame would not be detected using single parity method. Hence, more robust error detection techniques are required. Parity check is a simple method of error detection and can be implemented by using only exclusive-OR (XOR) gates to generate the parity bit, which is easily added to the data using a shift register. The data stream is appended with bit 1 if it contains an odd number of 1s or bit 0 if it contains an even number of 1s at the transmitting side. At the receiving side, the parity bit is computed from the received data bits to obtain the data stream. Some examples of the even parity and the odd parity are given in Table 5.10.

Table 5.10 *Even Parity (Parity Bit is Appended Right Hand Side)*

Data stream	Parity bit code	Data stream after appending parity bit (word)
0000	0	0000**0**
0001	1	0001**1**
0010	1	0010**1**
0011	0	0011**0**
0101	0	0101**0**

Odd Parity (Parity Bit is Appended Right Hand Side)

Data stream	Parity bit code	Data stream after appending parity bit (word)
0000	0	0000**0**
0001	0	0001**0**
0010	0	0010**0**
0011	1	0011**1**
0101	1	0101**1**

Performance

It can be deduced from Table 5.10 that a minimum of two data bits need to be changed to move from one code word to another. Hence, such code words have a minimum distance of 2. This minimum distance is known as the *hamming distance*. The hamming distance of 2 enables a receiver to detect all single bit errors only in each code word. In case a data stream having two errors is detected by the receiver as another valid word, the receiver will detect no error. Thus errors in a data stream more than one bit cannot be detected.

Two-Dimensional Parity Check Method

NOTES

The two-dimensional parity check method is employed to detect any combination of two errors in a data stream or frame. However, this method cannot correct the errors. The two-dimensional parity check organizes the frame of bits in the form of a table and parity check bits are calculated for each row according to the simple parity check bit. Similarly, parity check bits are also calculated for all columns. They are transmitted along with the data at the transmitting side. At the receiving side, they are compared with the parity bits calculated on the received data.

The word 'DEAR' can be represented in ASCII form as 10001000 10001010 10000010 10100100 as the original data frame which is to be transmitted to the receiver. The 2-D parity for the same is calculated as follows:

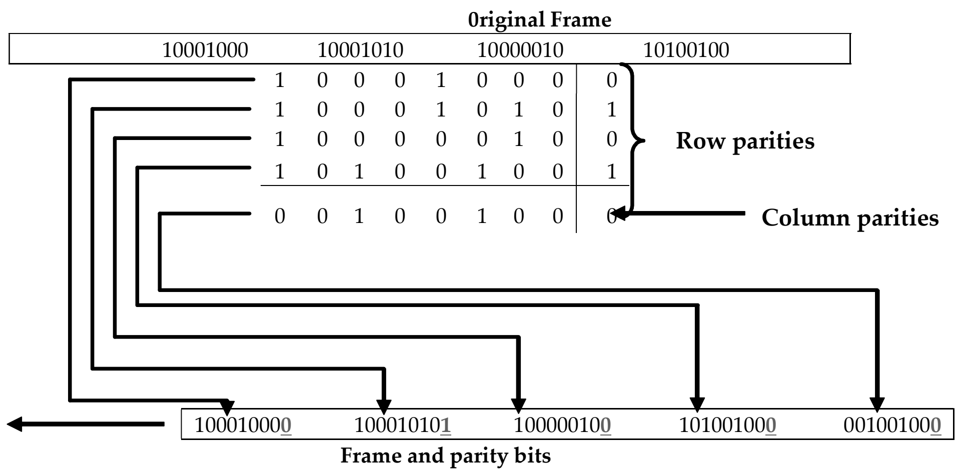


Fig. 5.19 *2-D Parity*

- Represent the original data in four rows and eight columns as shown in Figure 5.19.
- Deduce even parity for each row and column.
- Construct a frame with original data with row parities for individual characters and
- Append the column parities at the end of the data frame with row parities.
- The resultant frame is transmitted to the receiving side.

Performance

Two-dimension parity checking improves the likelihood of detecting burst errors. A 2-D parity check of *n* bits can detect a burst error of *n* bits. It is, however, difficult to detect one pattern of error when two bits in one frame are corrupted and two bits in exactly the same position in another frame are also corrupted; for example, if two frames 11011100 and 10111100 experience bit corruption in the first and second from last bits, the changed frames will be as 11001110 and 10111110. In such a case, the error cannot be detected by the 2-D parity check.

Forward Error Correction

The apparent quality of a communication channel is increased by two distinct approaches. These approaches are known as forward error-correction and ARQ. In FEC, redundant packets are transmitted along with source packets. If the number of lost packets is smaller than the number of redundant packets, then data is reconstructed without error. Though perfect recovery can be guaranteed, it still applies FEC. It maintains constant throughput and involves bounded time-delay too. The forward error-correction is also known as channel coding that is supported by digital signal processing. It improves the data reliability by a structure that priors the storage and transmission. The structure of this system is used to detect the error. Forward error-correction uses codes containing sufficient redundancy to prevent errors by detecting and correcting them at the receiving end without transmission of the original message. The redundancy of extra bits is ranged from a small percentage of extra bits to hundred percent redundancies with the number of error detecting bits roughly equaling the number of data bits. It is commonly used in satellite transmission.

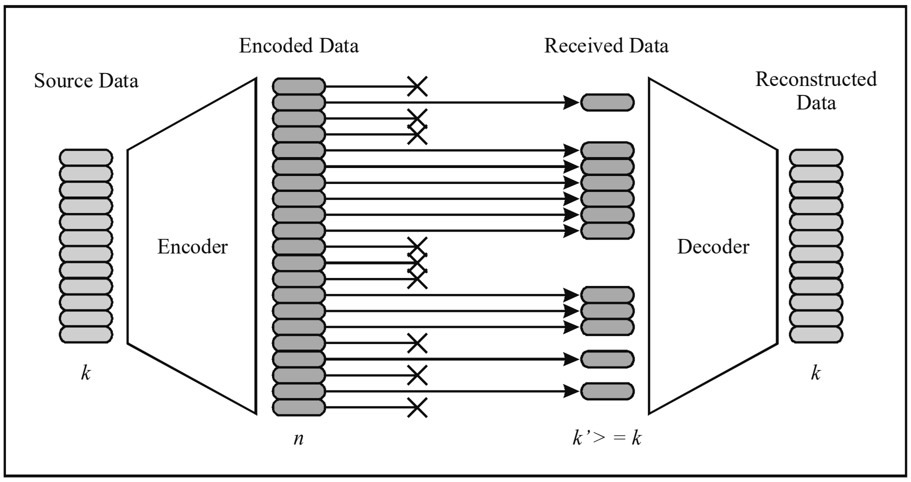


Fig. 5.20 *Forward Error-Correction with Coder and Encoder*

Figure 5.20 shows the alternative use of ARQ, which requires feedback. It is believed that this condition is not suitable for multicast and tolerance, which are only suitable for some applications. The source data first passes through an encoder that changes the source data into encoded data. After receiving the data by the decoder, it again changes into reconstructed data. The source data is represented by 'k' and is transformed into reconstructed data as 'k' too. FEC is a type of error correction which improves on simple error detection schemes by enabling the receiver to correct errors once they are detected. This reduces the need for retransmissions. It works by adding check bits to the outgoing data stream. Adding more check bits reduces the amount of available bandwidth and also enables the receiver to correct more errors. It is particularly well-suited for satellite transmissions, where bandwidth is reasonable but latency is significant. In data communication and networking, forward error-correction coding is used to send the data sequence to an encoder. Redundant or parity bits are inserted by encoder. This process is known as *codeword*. These codewords are transmitted to the

NOTES

receiver. The role of the receiver is to extract the original data sequence. Codes retain the little information about each individual code bit. This process prevents the removal of original data. The parity bits are increased for message delay. The most commonly used type of forward error-correction coding is known as *algebraic coding*, which was introduced by Claude Shannon who wrote the algebraic coding in his seminal Mathematical Theory of Communication in 1948. In this technique, parity bits are interspersed into the data sequence, whereas the role of the decoder is to apply an inverse of algebraic algorithm to identify and then make corrections for errors, which are generated by channel corruption. Convolution code, introduced in 1955, is known as another forward error-correcting technique and is processed with incoming bits in streams instead of blocks. Andrew Viterbi introduced a decoding technique in 1967 for decoding convolution codes. This decoding requires less memory to leave small number of sequences of data codes that are to be stored. Claude Berrou and his associates developed turbo code in 1993. The turbo code is known as the most powerful forward error-correction code.

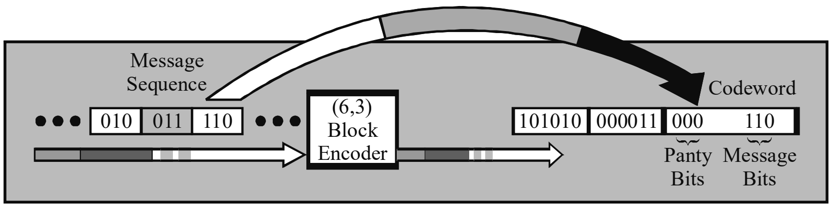


Fig. 5.21 *Message Sequence in Codeword*

Figure 5.21 shows an example of a (6, 3) block algebraic encoder. It produces a six-bit codeword for every three-bit message sequence. The 'parity bit' is taken as '000' and 'message bits' are taken as '110', which collectively produce 'codeword'. You also find that each six-bit output codeword is composed of the original three-bit message sequence and a three-bit parity sequence. This codeword format is known as systematic.

Working of Forward Error-Correction

In data communication and networking, a digital information source sends a data sequence comprising *k* bits of data to an encoder. The encoder inserts redundant or parity bits, thereby outputting a longer sequence of *n* code bits called a codeword. On the receiving end, codewords are used by a suitable decoder to extract the original data sequence. Codes are designated with the notation (n, k) according to the number of *n* output code bits and *k* input data bits. The ratio k/n is called the rate, R, of the code and is a measure of the fraction of information contained in each code bit; for example, each code bit produced by a (6, 3) encoder contains 1/2 bit of information. Another metric often used to characterize code bits is redundancy, expressed as (n–k)/n. Codes introducing large redundancy, i.e., large n–k or small k/n convey relatively little information per code bit. Codes that introduce less redundancy have higher code rates up to a maximum of 1 and convey more information per code bit. Large redundancy is advantageous because it reduces the likelihood that all of the original data will be wiped out during a single transmission. On the down side, the addition of parity bits will generally increase the transmission

NOTES

bandwidth or the message delay. For real-time applications, such as voice communications, the code-bit rate must be increased by a factor of n/k = 1/R to avoid a reduction in data throughput. Hence, for a given modulation scheme, the transmission bandwidth increases by that same factor n/k. If, however, the communication application does not require the real-time transfer of information, then additional message delay rather than increased bandwidth is the usual trade-off. Represented graphically, the general error-performance characteristics of most digital communication systems have a waterfall-shaped appearance. The importance of coding is possible when the system is viewed from the designer's perspective; for example, to obtain the same level of improved bit-error rate without the use of coding, a designer would have to achieve a larger signal-to-noise ratio, i.e., 12 decibels instead of 8 decibels. To do so would require the use of larger power supplies, bigger antennas or higher-quality components that introduce less noise.

Cyclic Redundancy Check

The technique of providing a data string that is added to information packets can be used to detect errors in the data packets. This error detection technique is used widely in computer networks. This technique is based on the redundancy concept. CRC is added to a packet frame at the data link layer in the case of OSI or TCP/IP network models. This is a technique of finding errors in data transmitted on a communications link. CRC is considered as one of the most reliable error detection methods and can detect more than 95 per cent of all errors. Among CRC techniques, CRC-16 standard code is the most commonly used code.

Unlike the checksum method which is based on addition, CRC implements a binary division in which a certain number of 1s and 0s are appended at the end of a data unit so that the number is exactly divisible by a second number which is predetermined. This is called as the CRC remainder or CRC. In the CRC technique, the sender and receiver agree to the fact that the data stream sent by the sending device will always be divisible by a common divisor. If the receiver gets a value which was not divisible by the common divisor, the receiving device will know that some error has occurred in the transmitted data. In CRC, a sequence of redundant bits, called *cyclic redundancy check bits*, are added to the end of the data stream in such a manner that the resulting frame becomes exactly divisible by a second, predetermined binary number. At the receiving side, the incoming frame is divided using the same number. If, no remainder exists, the data stream is assumed to be correct and is accepted. A remainder indicates that the data stream has been corrupted in transit and should be rejected. Thus, the data integrity of a received frame can be checked with the help of a polynomial algorithm, which is based on the content of the frame and the result is matched with the performance of the sender. It is then included in a (most often 16-bit) field appended to the frame. Hence, CRC codes are also known as *polynomial codes*. It uses a dividend polynomial, which is initially preset to 0, and the 1s and 0s of the data stream become the coefficients of the dividend polynomial. The division uses subtraction modulo 2 (no carries), and the remainder is transmitted as the error check field. The transmitted remainder is compared with its own computed remainder by the receiving station and the equal condition helps to indicate the absence of any

NOTES

error. The polynomial value is dependant on the usage of the protocol and code set. Mathematically, it can be explained as follows:

- For a *k* bit data stream to be transmitted, the sending end CRC generator generates an r-bit sequence. This is called the Frame Check Sequence (FCS). The data stream, along with the r-bit sequence, becomes a (k+r) bits frame that is actually transmitted.
- The redundancy *r*-bit FCS used by the CRC is derived by dividing the original data stream by a predetermined binary number. The remainder will be the CRC.
- The CRC is validated by using two conditions. In the first condition, it should possess exactly one bit less than the divisor. The second condition necessitates that when CRC is appended to the end of the data stream, the resulting data stream should be exactly divisible by the divisor.
- In CRC, the divisor is usually represented by a polynomial instead of 1s and 0s. A polynomial is represented as follows:

Polynomial	Bit string of polynomial	Binary representation
$X^7 + X^6 + X + 1$	$1. X^7 + 1.X^6 + 0.X^5 + 0.X^4 + 0.X^3 + 0.X^2 + 1. X + 1$	11000011
$X^5 + X^4 + X + 1$	$1.X^5 + 1.X^4 + 0.X^3 + 0.X^2 + 1. X + 1$	110011

The order of a polynomial is the power of the highest non-zero coefficient. The above examples are the polynomials of orders 7 and 5. On analysing the above polynomial, it is observed that both polynomials are indivisible by X and both are divisible by X + 1. These two conditions are used as the basis for choosing a polynomial as a divisor. Therefore, a polynomial should be selected to fulfill at least the following properties:

- All bit string should not have zero coefficients.
- It should be divisible by x+1.
- It should not be divisible by x.

Deriving the CRC

The receiving device and the sending device agree to use the same generating function P(x). The CRC is derived as follows:

- P(x) is the generator polynomial with r + 1 bits and M(x) is the message polynomial having *k* bits.
- Append *r* zero bits onto the right-hand side of the message so that it has *k* + *r* bits.
- Using modulo-2 division, divide the modified bit pattern by P(x). Modulo-2 arithmetic is performed with exclusive-OR operations which mean that 0 – 1=1, 1 – 1=0, 1 – 0=1 and 0-0=0.
- The last remainder is summed up to the modified bit pattern.

The above steps are described here with the help of an example. Consider a data stream 100100 which is six bits long. The sending and the receiving side agrees to select a divisor 1101 represented by the polynomial $X^3 + X^2 + 1$. As the divisor

NOTES

has 4 bits, 3 extra zeroes are added onto the end of the data stream. Subsequent to the Modulo 2 division, these zeroes will be replaced by the remainder. The Modulo 2 division is performed and shown in Figure 5.22.

			Quotient										
Steps			1	1	1	1	0	1					
1	Divisor	1101	1	0	0	1	0	0	0	0	0	Data plus extra 3 zeros	
	XOR operation		1	1	0	1							
2				1	0	0	0						
	XOR operation			1	1	0	1						
3					1	0	1	0					
	XOR operation				1	1	0	1					
4						1	1	1	0				
	XOR operation					1	1	0	1				
5							0	1	1	0			
6	XOR operation						**0**	**0**	**0**	**0**			
7								1	1	0	0		
	XOR operation							1	1	0	1		
8						**Remainder**			0	0	1		

Fig. 5.22 *Binary Division in CRC Derivation*

In step 5, the leftmost bit becomes 0 after the XOR operation. When the left becomes 0, you must use 0000 as divisor instead of the original divisor 1101. It should also be noted that 1101 divisor is divisible by x+1 (binary division) and is indivisible by *x*, therefore it can be used as divisor. Now, the 3 zeros as appended in the data stream are replaced by the remainder 001. Therefore, the frame that is transmitted becomes 100100001.

At the receiving side, a CRC checker executes the same Modulo 2 division on the frame and the received $k + r$ bit frame is divided by the same predetermined divisor 1101 and if it produces no remainder, it is assumed that no error has occurred during the transmission.

Some of the standard polynomials used as CRC generation are as follows:

Name	Polynomial	Application
CRC- 8	$X^8 + X^2 + X + 1$	ATM header
CRC - 10	$X^{10} + X^9 + X^5 + X^4 + X^2 + 1$	ATM AAL
ITU - 16	$X^{16} + X^{12} + X^5 + 1$	HDLC
ITU - 32	$X^{32} + X^{26} + X^{23} + X^{22} + X^{16} + X^{12} + X^{11} + X^{10} + X^8 + X^7 + X^5 + X^4 + X^2 + X + 1$	LANs

Performance

The advantage of CRC is that it requires only a shift register and a few XOR gates to perform the division because the division in CRC does not use standard arithmetic division and uses an exclusive-OR operation instead of the subtraction operation.

Codes for Error Detection

Communication between two hosts, whether over a telephone connection, across the Internet or as part of an Ethernet network gives rise to the possibility of messages not being correctly transmitted, i.e., the message can be corrupted during transit. The message may be corrupted through a number of different mechanisms. In all cases, the corruption needs to be detected and then appropriate actions are to be taken, i.e., data recovered, information resent, etc. Sometimes, the data can be corrupted. Errors are also known as *noise*. Attenuation occurs since the signal gradually becomes less strong as the distance over which it travels increases. This is due to the dissipation of energy and can be readily appreciated by two people communicating over increasing distances. Error-detecting codes include enough redundant information to permit the receiver to determine that an error occurred and have it request a retransmission. Error-detecting codes are preferable if errors do not occur frequently, as less redundant information needs to be sent. The first approach is useful when errors frequently occur, as it does not require that the data be retransmitted. Let us take an example in which the code for error detection has to be taken as follows:

0 = 000 **00**

1 = 001 **01**

2 = 010 **10**

3 = 011 **11**

4 = 100 **11**

5 = 101 **10**

6 = 110 **01**

7 = 111 **00**

In this example, two bits are added to each of three-bit numbers. This process can be categorized as follows:

Small Errors: 1-Bit Errors

If 001[01] becomes 000[01] due to a 1 bit error, by comparing the received code to the expected code, you can readily detect that an error has occurred, as [01] ≠ [00].

Larger Errors: 2-Bit Errors

The code has been arranged so that only those byte patterns which differ by three bits, for example, 000 and 111, 011 and 100 have been assigned for the same 2-bit error value, which enables you to detect 2-bit errors; for example, if 010 [10] becomes 100[10], you can expect to find a [11] ending, but instead, it is found as [10], hence you can easily understand that an error has occurred.

Huge Errors: 3-Bit Errors

In this case, the code is being collapsed; for example, 011[11] becomes 100[11] due to a 3 bit error. To find a [11] ending, and value [11] ending is found, hence the error is not detected.

In essence, the above 2-bit code is interesting as it enables us to correct small errors in addition to detecting errors; for example, let us assume that only 1 bit will be corrupted (000[00] becomes 100[00]). Obviously, an error has been detected as [00] and is not expected to find the sequence [100]. Let us further examine the bit sequences which have [00] as their correct ending, which include

NOTES

000 and 111. Assuming that only a 1-bit error has occurred, which might be an unsafe assumption, a received sequence could only have been a corrupted 000 (1-bit change) and not a corrupted 111 (2-bit change). Hence, the error is corrected. The 2-bit code is identified that permits us to detect up to a 2-bit error and to correct at most a 1-bit error.

A popular example is the hamming code detection method. Table 5.11 shows the hamming distances and theirs usage for detect errors.

Table 5.11 *Hamming Distance and their Uses*

Hamming Distance	Uses
1	A single error bit generates another valid codeword. Error detection and correction is impossible.
2	Two bit errors are needed before one codeword can be changed into another valid codeword. One bit error detection is possible; however, no error correction is possible.
3	Two bit errors can be detected, and it is possible to recover from a 1 bit error.
4	Three bit errors can be detected; however, it is still only possible to recover from 1 bit errors.
5	Four bit errors can be detected, and 2 bit errors can be readily recovered from.

Software implementations of error detection codes are considered to be slow compared to other parts of the data communication and networking system. This is especially true for powerful error detection codes, such as CRC. The fast software implementation of the Cyclic Redundancy Check (CRC), Weighted Sum Codes (WSC), one's-complement checksum, Fletcher checksum, CXOR checksum and block parity codes are frequently used in codes for error detection. Instruction count alone does not determine the fastest error detection code. Given the performance of various error detection codes, a protocol designer can choose a code with the desired speed and error detection power that is appropriate for Internet working and application. Error-correcting codes are widely used on wireless links that are noisy. However, they generate too large transmission overhead for reliable links, such as copper wire or fiber. Therefore, here error-detection codes are used. When error is detected, the data is retransmitted. The goal for error correcting codes is to add redundancy to the data so that the errors are not only detected but can be at the same time corrected without retransmission. For error-detecting codes, the goal is to only detect the errors with the minimal transmission overhead. They are based on polynomial code also known as Cyclic Redundancy Check (CRC). A k-bit frame is regarded as polynomial with coefficients 0 and 1 with terms from xk-1 to x0; for example, the value 110001 can be expressed in the following way:

$$110001 \rightarrow x5 + x4 + x0$$

There are two basic methods of error control for communication, both involving coding of the messages. With forward error correction, the codes are used to detect and correct errors. In a repeat request system, the codes are used to detect

NOTES

errors and, if there are errors, request a retransmission. Error detection is usually much simpler to implement than error correction and is widely used. Figure 5.23 shows the block diagram of data communication system that employs the error-correcting code.

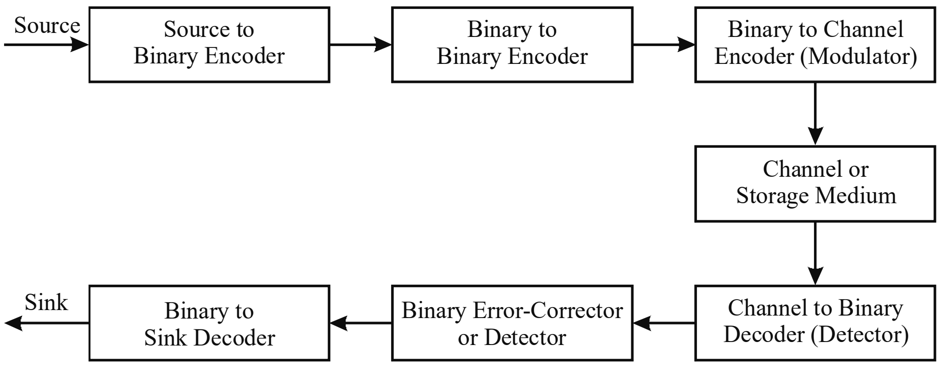


Fig. 5.23 *Block Diagram of Data Communication System Employing Error-Correcting Code*

Figure 5.23 shows that in data communication and networking, errors occur during the time of transmitting data from the host to the destination. The role of the modulator is to put restrictions that definitely causes a loss in channel capacity. The demodulator performs inverse operations of the modulator. It is associated with channel symbol and with the noise corrupted waveform received. The independent demodulation results in channel capacity. The encoder and decoder are implemented for error-correcting code.

5.5 DIGITAL CARRIER SYSTEMS

A carrier system is a type of telecommunications system that carries data, such as speech signals from a phone call or video signals from a television, for example. Typically, carrier systems transmit numerous channels of communication concurrently across a shared media. A digital carrier system is a type of communication system that encodes data using digital pulses rather than analogue signals. The digital carrier standards are as follows:

1. T-Carrier

T-carrier is a completely digital system that employs Pulse Code Modulation (PCM) and Time Division Multiplexing (TDM). The system utilizes four wires and is capable of bidirectional communication (two wires for receiving and two for sending at the same time). T_1 lines should be reserved for mission-critical applications requiring high bandwidth. T_1 lines perform optimally when the sites being connected are close in proximity (otherwise the cost is prohibitive).

2. E-Carrier

E-carrier, a European carrier, maintains a hierarchy of digital transmission standards. E-carrier is based on the original North America T-carrier digital carrier system; although the specifics are quite different with respect to signaling rates, framing convention, line coding technique, and PCM compounding technique. In many respects E-carrier is a significant upgrade over T-carrier. For instance, E-1 supports 30 DS-0 payload channels, compared to T_1's 24 channels, and the increased E-carrier levels capitalize on this advantage. Additionally, E-carrier provides non-intrusive signaling and control via two dedicated channels. As a result, E-carrier provides clear channel communication at a full 64Kbps per DS-0, in comparison to T-56 carrier's kbps data. The DS-0 (Digital Signal level Zero) is the fundamental building block of E-carrier, as it is with T-carrier and J-carrier, the Japanese version. E-carrier interleaves DS-0 channels at varying signaling rates using Time Division Multiplexing (TDM) to produce the services that comprise the European digital hierarchy.

3. SONET/SDH

SONET is a Time Division Multiplexing (TDM) protocol for transmission through optical fibres in the terrestrial United States. It has following properties:

- An optical telecommunications transport standard known as Synchronised Optical NETwork (SONET) was developed by the Exchange Carriers Standards Association (ECSA) for the American National Standards Institute (ANSI), which establishes industry standards for telecommunications and other industries in the United States and is now known as SONET.
- In Europe, the International Telecommunication Union's Telecommunication Standardization Sector employs a related standard, Synchronous Digital Hierarchy (SDH), which was developed by the International Telecommunication Union (ITU-T). SONET equipment is commonly utilized in North America, whereas SDH equipment is generally used throughout the rest of the world, including Europe.
- Both SONET and SDH are fibre optic transport protocols. The North American standard is SONET, or synchronous optical network.
- SDH is a comparable standard used throughout Europe and the rest of the world.
- SONET operates at the layer 1 level (Physical).
- SONET/SDH is particularly well-suited for conveying time-sensitive speech and video, but it is also utilized for high-speed data transmission.
- SONET is composed of an infinite number of frames.
- SONET is a high-speed data transmission protocol. Telephone companies have historically relied heavily on SONET, but this may be changing as new high-speed transmission technologies gain traction.
- SONET operates at a synchronous transport signal–level 1 (STS–1) transmission rate of 51.84 Mbps.

5.6 TELEPRINTERS AND TELEGRAPHS CIRCUITS

NOTES

Telegraph

Telegraphic devices transmit communications, referred to as telegrams, over great distances. The word "Telegraph" comes from two Greek words. "Tele" means "At a Distance" and "Gramma" means "Letter". When people refer to telegraphs, they usually mean electric telegraphs.

An electrical telegraph was a point-to-point text messaging system, used from the 1840s until the late 20th century when it was slowly replaced by other telecommunication systems. At the sending station switches connected a source of current to the telegraph wires. At the receiving station the current activated electromagnets which moved indicators, providing either a visual or audible indication of the text. It was the first electrical telecommunications system and the most widely used of a number of early messaging systems called telegraphs, that were devised to communicate text messages more rapidly than by physical transportation.

Electrical telegraph networks permitted people and commerce to transmit messages across both continents and oceans almost instantly, with widespread social and economic impacts. In the early 20th century, the telegraph was slowly replaced by teletype networks.

Samuel Morse independently developed and patented a recording electric telegraph in 1837. Morse's assistant Alfred Vail developed an instrument that was called the register for recording the received messages. It embossed dots and dashes on a moving paper tape by a stylus which was operated by an electromagnet. Morse and Vail developed the Morse code signaling alphabet.

The Morse telegraph circuit is seen in Figure 5.24. It is used to communicate between distant sites. Each instrument is described in detail below:

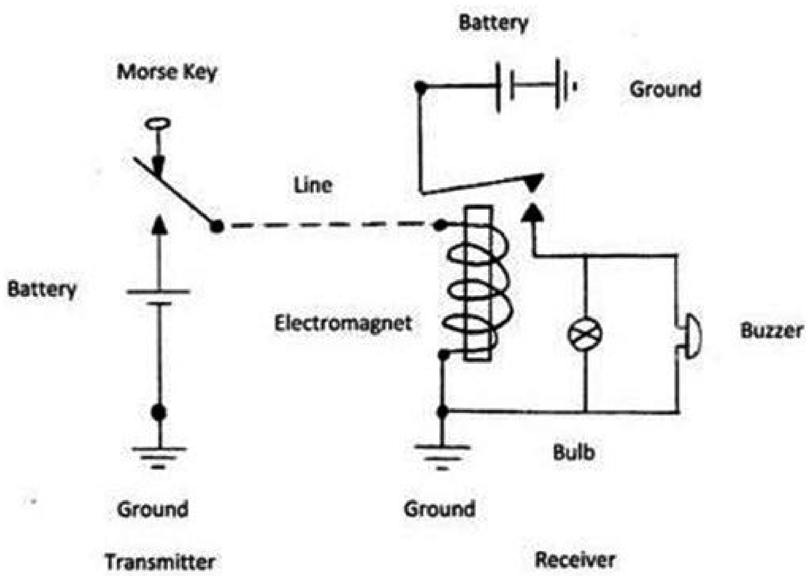


***Fig. 5.24** Morse Telegraph Circuit Diagram*

NOTES

i. **Telegraph Key:** This is the fundamental 'Transmitter'. A telegraph key is simply an electrical switch. This switch features a pair of contacts that are used to open and close the circuit. The key is ergonomically built to enable the user to make and break connections swiftly in order to send Morse code signals.

ii. **Morse Relay:** Morse relays are low-cost electromechanical amplifiers. It is constructed in such a way that it is particularly sensitive to electric current. A horseshoe electromagnet is designed in such a way that it attracts a delicate iron armature. When current runs through the coils of the electromagnet, a contact on the armature and a stationary post establish contact. These contacts are responsible for establishing and terminating the 'Local Circuit'.

iii. **Local Sounder:** The receiving instrument is referred to as the 'Sounder'. When the instrument's electromagnet is energized, an audible 'Click' is produced. Additionally, it emits a second 'Click' with a slightly different tone upon the current's termination. As a result, each current pulse makes a distinctive 'Click-Clack' sound. Local sounders are built for maximum volume and are not very sensitive to electrical current. When opposed to a relay, a local sounder requires up to twenty times the amount of current to activate.

iv. **Main Battery:** This is a battery constructed from a series of electrochemical cells. Several hundred volts may be necessary, depending on the line's length and the number of instruments cut into the wire. Around the turn of the century, these batteries were replaced with dynamos.

v. **Local Battery:** Typically, this was a single cell with a voltage of between 1 and 1.5 volts. It served only as a source of current for the local sounder. The local battery remained in use long into the twentieth century in the numerous depots and way stations that lacked alternating current power.

vi. **Telegraph Line:** Typically, an extremely thick iron wire was employed. To prevent corrosion, the wire was Galvanized. Using specially manufactured glass insulators, the telegraph line was suspended from wooden poles. Insulator design was crucial in minimizing leakage routes to ground. These leakage channels were dubbed 'Escapes', and they were capable of rendering a telegraph line inoperable during inclement weather.

Telegraphy and Longitude

The telegraph was very important for sending time signals to determine longitude, providing greater accuracy than previously available. Longitude was measured by comparing local time, for example, local noon occurs when the sun is at its highest above the horizon, with absolute time, i.e., a time that is the same for an observer anywhere on earth. If the local times of two places differ by one hour, the difference in longitude between them is 15° (360°/24h). Before telegraphy, absolute time could be obtained from astronomical events, such as eclipses, occultations or lunar distances, or by transporting an accurate clock (a chronometer) from one location to the other. The idea of using the telegraph to transmit a time signal for longitude determination was suggested by François Arago to Samuel Morse in 1837.

NOTES

Teleprinter

A teleprinter (teletypewriter, teletype or TTY) is an electromechanical device that can be used to send and receive typed messages through various communications channels, in both point-to-point and point-to-multipoint configurations. Initially they were used in telegraphy, which developed in the late 1830s and 1840s as the first use of electrical engineering, though teleprinters were not used for telegraphy until 1887 at the earliest. The machines were adapted to provide a user interface to early mainframe computers and minicomputers, sending typed data to the computer and printing the response. Some models could also be used to create punched tape for data storage (either from typed input or from data received from a remote source) and to read back such tape for local printing or transmission.

Teleprinters could use a variety of different communication media. These included a simple pair of wires; dedicated non-switched telephone circuits (leased lines); switched networks that operated similarly to the Public Telephone Network (TELEX or TELegraph EXchange); and radio and microwave links (Telex-On-Radio, or TOR). A teleprinter attached to a modem could also communicate through standard switched public telephone lines. This latter configuration was often used to connect teleprinters to remote computers, particularly in time sharing environments.

A teleprinter is, therefore, an electromechanical typewriter that may be used to send and receive typed messages via a variety of communication routes, ranging from a basic electrical connection, such as a pair of wires, to radio and microwave transmission. They may also act as a command line interface for early mainframe and minicomputer computers, transmitting written data to the computer with or without printed output and printing the computer's answer. The Figure 5.25 illustrates the teleprinter's block diagram, which depicts its various components:

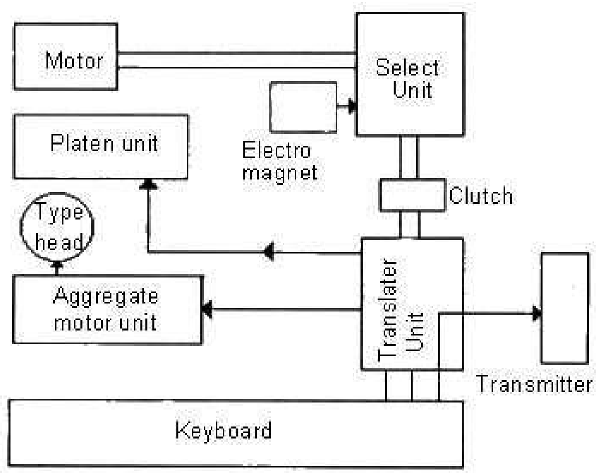


Fig. 5.25 *Block Diagram of Teleprinter*

The salient features of a teleprinter are listed below:

1. It is a telegraph transmitting receiving machine.
2. It resembles a typewriter because it has a typewriter like keyboard.

3. It is a mechanical device driven by electrical motors; recently electronic machine has been introduced, controlled by microprocessor.
4. It uses 5-unit code.
5. It works on start-stop principle.
6. It acts both as a transmitter and a receiver.
7. When used as a receiver, the signals received in the series form are converted into parallel. Then a detector converts it into the character and the printer prints it on the paper.
8. It also has the facility of local record.

5.7 RADIO TELEGRAPHS TRANSMITTERS

Wireless telegraphy, or **radiotelegraphy**, is the use of radio waves to transmit telegraph signals by radio waves. Radiotelegraphy transmits information using pulses of radio waves of two distinct lengths termed 'Dots' and 'Dashes' that spell out text messages, typically in Morse code. In a manual system, the sending operator taps on a button called a telegraph key to turn 'ON' and 'OFF' the transmitter, which generates radio wave pulses. At the receiver, the pulses are audible as beeps, which are converted back to text by a Morse code operator.

Radiotelegraphy was used for long-distance person-to-person commercial, diplomatic, and military text communication throughout the first half of the 20th century.

Wireless telegraphy or radiotelegraphy, commonly called CW (Continuous Wave), ICW (Interrupted Continuous Wave) transmission, or ON-OFF keying, and designated by the International Telecommunication Union as emission type A1A or A2A, is a radio communication method.

As per the Encyclopedia Britannica, "The radiotelegraphy refers to the radio communication by means of Morse Code or other coded signals. The radio carrier is modulated by changing its amplitude, frequency, or phase in accordance with the Morse dot-dash system or some other code. At the receiver the coded modulation is recovered by an appropriate demodulator and the code groups are converted into the corresponding symbols. In many instances the symbols are generated by a computer and modem rather than with a manual telegraph key".

In manual radiotelegraphy, the sending operator manipulates a switch called a telegraph key, which turns the radio transmitter ON and OFF, producing pulses of unmodulated carrier wave of different lengths called 'Dots' and 'Dashes', which encode characters of text in Morse code. At the receiving location, Morse code is audible in the receiver's earphone or speaker as a sequence of buzzes or beeps, which is translated back to text by an operator who recognizes and identifies Morse code. With automatic radiotelegraphy teleprinters at both ends use a code, such as the International Telegraph Alphabet No. 2 and produced typed text.

The ground was used as the return path for current in the telegraph circuit, to avoid having to use a second overhead wire.

By the 1860s, the telegraph was the standard way for sending most urgent commercial, diplomatic and military messages, and industrial nations had built continent-wide telegraph networks, with submarine telegraph cables allowing telegraph messages to bridge oceans. However, installing and maintaining a telegraph line linking distant stations was very expensive, and wires could not reach some locations, such as ships at sea.

Both electrostatic and electromagnetic induction were used to develop wireless telegraph systems. The most successful creator of an electromagnetic induction telegraph system was William Preece, chief engineer of Post Office Telegraphs of the General Post Office (GPO) in the United Kingdom. With this development, wireless telegraphy came to mean radiotelegraphy, Morse code transmitted by radio waves.

The radio telegraph transmitter handles the conversion of the audio signal into a radio signal and broadcasts it as a radio wave via an antenna.

Check Your Progress

10. Why is coding required?
11. State about the even parity and odd parity check methods.
12. Define about the carrier system.
13. What are the digital carrier standards?
14. What is telegraph? Who invented it?
15. Define the term teleprinter.
16. What is radiotelegraphy?

5.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. In digital communications, transmission of information takes place in digital form. Digital transmission or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel. These could be copper wires, optical fibers, wireless communication channels, and storage media. The data is represented as an electro-magnetic signal like an electrical voltage, radio wave, microwave, etc.
2. Asynchronous transmission uses start and stop bits to signify the beginning and ending bits. ASCII character would actually be transmitted using 10 bits e.g.: A "0100 0001" would become "1 0100 0001 0". The extra 1 (or 0 depending on parity bit) at the start and end of the transmission tells the receiver first that a character is coming and secondly that the character has ended. This method of transmission is used when data is sent intermittently as opposed to in a solid stream.

   Synchronous transmission uses no start and stop bits. Instead, it synchronizes transmission speeds at both the receiving and sending ends of the transmission

NOTES

using clock signal(s) built into each component. A continuous stream of data is then sent between the two nodes. Since there are no start and stop bits, the data transfer rate is faster although more errors can occur, as the clocks will eventually get out of sync, and the receiving device would have the wrong time than that had been agreed in protocol for sending/receiving data, so some bytes could become corrupted (by losing bits).

3. Bit rate is the rate over network speed which is used to detect errors while transmitting data. The most popular method for detecting errors is inserting a parity bit alongside the data bits for a character. Receiving modems detect incorrect bit rate, which is also called parity bit. It requests the sending modem to retransmit the character. Bit rates are sometimes written as data rate and are conveyed and processed per unit time. It is measured as bits per second (bit/s or bps).

   Multimedia bit rate is the number of bits used per unit to represent continuous medium such as audio or video following source coding (data compression) to the multimedia files. The size of the multimedia file is the product of bit rate (in bit/s) in bytes and the length of recording in seconds divided by eight. The bit rate is measured by input, which avoids interrupts with reference to streaming multimedia.

4. In telecommunication, Data Signaling Rate (DSR), also known as Gross Bit Rate (GBR), is the aggregate rate at which data passes a point in the transmission path of a data transmission system. Fundamentally, the Data Signaling Rate (DSR) is the aggregate rate at which data pass a point in the transmission path of a data transmission system.

   - The DSR is usually expressed in bits per second.
   - The data signaling rate is given by $\sum_{i=1}^{m} \frac{\log_2 n_i}{T_i}$ where m is the number of parallel channels, n_i is the number of significant conditions of the modulation in the i-th channel, and T_i is the unit interval, expressed in seconds, for the i-th channel.

5. The probability of error in a digital communication system is mainly due to inter-symbol interference and bit errors during the digital transmission.

   In digital transmission, the number of bit errors is the number of received bits of a data stream over a communication channel that have been altered due to noise, interference, distortion or bit synchronization errors.

   The Bit Error Rate (BER) is the number of bit errors per unit time. The Bit Error Ratio (also BER) is the number of bit errors divided by the total number of transferred bits during a studied time interval. Bit error ratio is a unitless performance measure, often expressed as a percentage.

6. Filters are a key feature of any signal processing or telecommunications system. Digital filtering can be used to eliminate both hardware and software bottlenecks.; in the first case, the numerical processor is either a special-purpose chip or it is assembled out of a collection of digital integrated circuits

that serve as the foundation for a digital filtering process – storage, delay, addition/subtraction and multiplication by constants.

A digital filter is a mathematical technique that operates on a digital input signal to generate a digital output signal with the objective of filtering out unwanted data. It is implemented in hardware and/or software. Digital filters work with digitized analogue signals or just numbers in a computer memory that represent a variable.

NOTES

7. Code is a symbolic representation of discrete information, which may be presented in the form of numbers, letters or physical quantities. The symbols used are the binary digits 0 and 1, which are arranged according to the rules of codes. These codes are used to communicate information to a digital computer and to retrieve messages from it.

   Codes are broadly classified into five groups, viz., (i) Weighted Binary Codes, (ii) Non-Weighted Codes, (iii) Error-Detecting Codes, (iv) Error-Correcting Codes, and (v) Alphanumeric Codes.

8. Pulse Code Modulation (PCM) is a digital representation of an analog signal where the magnitude of the signal is sampled regularly at uniform intervals, then quantized to a series of symbols in a digital (usually binary) code. PCM has been used in digital telephone systems and is also the standard form for digital audio and video in computers.

   Differential Pulse Code Modulation (DPCM) is a procedure of converting an analog into a digital signal in which an analog signal is sampled and then the difference between the actual sample value and its predicted value (predicted value is based on previous sample or samples) is quantized and then encoded forming a digital value.

9. Delta modulation (DM or Δ-modulation) is an analog-to-digital and digital-to-analog signal conversion technique used to transmit voice data. In this method, the quality is not of primary importance. This is the simplest form of Differential Pulse-Code Modulation (DPCM) where the difference between successive samples is encoded into n-bit data streams. In DM, the transmitted data is reduced to a 1-bit data stream.

10. A set of techniques that involves coding helps in detecting errors in stored or transmitted data and rectify them. Binary data in the form of bits is susceptible to interference caused by the transmission media or hardware in a data communication environment. The use of several parity bits helps to detect any error, find if any bits are inverted, and in case they are, they should be re-inverted so that the original data is restored. Adding extra bits add to chances of detecting and correcting multiple errors. There are different methods depending upon Single Error Correction, Double Error Detection (SECDEC).

11. Parity check using a single parity bit is the simplest and least expensive error detection method in which a parity bit is appended to a block of data, normally at the end of a 7-bit ASCII character. It consists of even parity and odd parity methods. Even parity method consists of one additional bit

NOTES

in such a manner that the total number of 1s in the resultant bit stream (the original information plus a parity bit) becomes even. Similarly, the odd parity method consists of addition of one bit in the original bit stream in such a manner that there are an odd number of 1s in the resultant bit stream.

12. A carrier system is a type of telecommunications system that carries data, such as speech signals from a phone call or video signals from a television, for example. Typically, carrier systems transmit numerous channels of communication concurrently across a shared media. A digital carrier system is a type of communication system that encodes data using digital pulses rather than analogue signals.

13. The digital carrier standards are as follows:

    T-Carrier: T-carrier is a completely digital system that employs Pulse Code Modulation (PCM) and Time Division Multiplexing (TDM). The system utilizes four wires and is capable of bidirectional communication (two wires for receiving and two for sending at the same time). T_1 lines should be reserved for mission-critical applications requiring high bandwidth. T_1 lines perform optimally when the sites being connected are close in proximity (otherwise the cost is prohibitive).

    E-Carrier: E-carrier, a European carrier, maintains a hierarchy of digital transmission standards. E-carrier is based on the original North America T-carrier digital carrier system; although the specifics are quite different with respect to signaling rates, framing convention, line coding technique, and PCM compounding technique. In many respects E-carrier is a significant upgrade over T-carrier.

    SONET/SDH: SONET is a Time Division Multiplexing (TDM) protocol for transmission through optical fibres in the terrestrial United States.

14. Telegraphic devices transmit communications, referred to as telegrams, over great distances. The word "Telegraph" comes from two Greek words. "Tele" means "At a Distance" and "Gramma" means "Letter". When people refer to telegraphs, they usually mean electric telegraphs.

    Samuel Morse independently developed and patented a recording electric telegraph in 1837. Morse's assistant Alfred Vail developed an instrument that was called the register for recording the received messages. It embossed dots and dashes on a moving paper tape by a stylus which was operated by an electromagnet. Morse and Vail developed the Morse code signaling alphabet.

15. A teleprinter (teletypewriter, teletype or TTY) is an electromechanical device that can be used to send and receive typed messages through various communications channels, in both point-to-point and point-to-multipoint configurations. A teleprinter is, basically, an electromechanical typewriter that may be used to send and receive typed messages via a variety of communication routes, ranging from a basic electrical connection, such as a pair of wires, to radio and microwave transmission. They may also act as a command line interface for early mainframe and minicomputer computers,

transmitting written data to the computer with or without printed output and printing the computer's answer.

16. Wireless telegraphy, or radiotelegraphy, is the use of radio waves to transmit telegraph signals by radio waves. Radiotelegraphy transmits information using pulses of radio waves of two distinct lengths termed 'Dots' and 'Dashes' that spell out text messages, typically in Morse code. In a manual system, the sending operator taps on a button called a telegraph key to turn 'ON' and 'OFF' the transmitter, which generates radio wave pulses. At the receiver, the pulses are audible as beeps, which are converted back to text by a Morse code operator.

5.9 SUMMARY

- In digital communications, transmission of information takes place in digital form.
- Digital transmission or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel. These could be copper wires, optical fibers, wireless communication channels, and storage media.
- The data is represented as an electro-magnetic signal like an electrical voltage, radio wave, microwave, etc.
- Analog communications stand for the transmission of continuously varying information signal, whereas digital communications are the transmission of discrete messages. The messages are either represented by a sequence of pulses by means of a line code (baseband transmission), or by a limited set of continuously varying wave forms (passband transmission), using a digital modulation method.
- The passband modulation and corresponding demodulation (also known as detection) is carried out by modem equipment. According to the common definition of digital signal, both baseband and passband signals representing bit-streams are considered as digital transmission, while an alternative definition only considers the baseband signal as digital, and passband transmission of digital data as a form of digital-to-analog conversion.
- Asynchronous transmission uses start and stop bits to signify the beginning and ending bits. ASCII character would actually be transmitted using 10 bits e.g.: A "0100 0001" would become "1 0100 0001 0". The extra 1 (or 0 depending on parity bit) at the start and end of the transmission tells the receiver first that a character is coming and secondly that the character has ended.
- Asynchronous transmission method of transmission is used when data is sent intermittently as opposed to in a solid stream.
- Synchronous transmission uses no start and stop bits. Instead, it synchronizes transmission speeds at both the receiving and sending ends of the transmission using clock signal(s) built into each component.

NOTES

- In synchronous transmission continuous stream of data is then sent between the two nodes. Since there are no start and stop bits, the data transfer rate is faster although more errors can occur, as the clocks will eventually get out of sync, and the receiving device would have the wrong time than that had been agreed in protocol for sending/receiving data, so some bytes could become corrupted (by losing bits).
- Bit rate is the rate over network speed which is used to detect errors while transmitting data. The most popular method for detecting errors is inserting a parity bit alongside the data bits for a character.
- Receiving modems detect incorrect bit rate, which is also called parity bit. It requests the sending modem to retransmit the character.
- A modem is a bridge gap between digital and analog data transmission which allows the digital data to be transmitted/received over the telephone lines.
- In telecommunication and network computing, bit rates are sometimes written as data rate and are conveyed and processed per unit time. It is measured as bits per second (bit/s or bps).
- The net bit rate, also called useful bit rate, of a digital communication link, is the capacity of the physical layer protocol, such as: Framing Bits, Time Division Multiplex (TDM) and redundant Forward Error Checking (FEC).
- Multimedia bit rate is the number of bits used per unit to represent continuous medium, such as audio or video following source coding (data compression) to the multimedia files. The size of the multimedia file is the product of bit rate (in bit/s) in bytes and the length of recording in seconds divided by eight. The bit rate is measured by input, which avoids interrupts with reference to streaming multimedia.
- In telecommunication, Data Signaling Rate (DSR), also known as Gross Bit Rate (GBR), is the aggregate rate at which data passes a point in the transmission path of a data transmission system.
- Fundamentally, the Data Signaling Rate (DSR) is the aggregate rate at which data pass a point in the transmission path of a data transmission system.
- The DSR is usually expressed in bits per second.
- The data signaling rate is given by $\sum_{i=1}^{m} \frac{\log_2 n_i}{T_i}$ where m is the number of parallel channels, n_i is the number of significant conditions of the modulation in the i-th channel, and T_i is the unit interval, expressed in seconds, for the i-th channel.
- For serial transmission in a single channel, the DSR reduces to $(1/T) \log_2 n$; with a two-condition modulation, i.e., $n = 2$, the DSR is $1/T$, according to Hartley's law.
- For parallel transmission with equal unit intervals and equal numbers of significant conditions on each channel, the DSR is $(m/T) \log_2 n$; in the case of a two-condition modulation, this reduces to m/T.

- The maximum user signaling rate, synonymous to gross bit rate or data signaling rate, is the maximum rate, in bits per second, at which binary information can be transferred in a given direction between users over the telecommunications system facilities dedicated to a particular information transfer transaction, under conditions of continuous transmission and no overhead information.
- The probability of error in a digital communication system is mainly due to inter-symbol interference and bit errors during the digital transmission.
- In digital transmission, the number of bit errors is the number of received bits of a data stream over a communication channel that have been altered due to noise, interference, distortion or bit synchronization errors.
- The Bit Error Rate (BER) is the number of bit errors per unit time. The Bit Error Ratio (also BER) is the number of bit errors divided by the total number of transferred bits during a studied time interval. Bit error ratio is a unitless performance measure, often expressed as a percentage.
- The bit error probability $\mathbf{p}_e$ is the expected value of the bit error ratio. The bit error ratio can be considered as an approximate estimate of the bit error probability. This estimate is accurate for a long-time interval and a high number of bit errors.
- Filters are a key feature of any signal processing or telecommunications system.
- Digital filtering can be used to eliminate both hardware and software bottlenecks.; in the first case, the numerical processor is either a special-purpose chip or it is assembled out of a collection of digital integrated circuits that serve as the foundation for a digital filtering process – storage, delay, addition/subtraction and multiplication by constants.
- In contrast, a general-purpose mini-or micro-computer can also be programmed as a digital filter, in which case the numerical processor is the computer's CPU and memory.
- A digital filter is a mathematical technique that operates on a digital input signal to generate a digital output signal with the objective of filtering out unwanted data. It is implemented in hardware and/or software. Digital filters work with digitized analogue signals or just numbers in a computer memory that represent a variable.
- The frequency response of a filter explains how the filter alters the magnitude and phase of the frequencies of the input signal it receives.
- Linear Time-Invariant (LTI) filters are a form of filter whose output is a linear combination of the samples of the input signal with constant coefficients, as opposed to other types of filters.
- The linear property indicates that the filter response to a weighted sum of a number of signals is equal to the weighted sum of the filter responses of each individual signal in the weighted sum.

NOTES

- Code is a symbolic representation of discrete information, which may be presented in the form of numbers, letters or physical quantities.
- The symbols used are the binary digits 0 and 1, which are arranged according to the rules of codes. These codes are used to communicate information to a digital computer and to retrieve messages from it.
- A code is used to enable an operator to feed data into a computer directly, in the form of decimal numbers, alphabets and special characters. The computer converts these data into binary codes and after computation, transforms the data into its original format, such as decimal numbers, alphabets and special characters.
- When numbers, letters, or words are represented by a special group of symbols, this is called encoding, and the group of symbols is called a code. In Morse code, a series of dots and dashes represent alphabet, numerals and special characters.
- Codes are broadly classified into five groups, viz., (i) Weighted Binary Codes, (ii) Non-Weighted Codes, (iii) Error-Detecting Codes, (iv) Error-Correcting Codes, and (v) Alphanumeric Codes.
- Pulse Code Modulation (PCM) is a digital representation of an analog signal where the magnitude of the signal is sampled regularly at uniform intervals, then quantized to a series of symbols in a digital (usually binary) code. PCM has been used in digital telephone systems and is also the standard form for digital audio and video in computers.
- Sampler is the technique which assist to collect the sample data at instantaneous values of message signal, so as to re-form the original signal. The sampling rate must be greater than twice the highest frequency component W of the message signal, in accordance with the sampling theorem.
- Quantizing is a process in which excessive bits are getting reduced and the data is confined. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.
- The function of an encoder is to perform digitization of analog signal. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. Encoding minimizes the bandwidth used.
- Delta modulation (DM or Δ-modulation) is an analog-to-digital and digital-to-analog signal conversion technique used to transmit voice data. In this method, the quality is not of primary importance. This is the simplest form of Differential Pulse-Code Modulation (DPCM) where the difference between successive samples is encoded into n-bit data streams. In DM, the transmitted data is reduced to a 1-bit data stream.
- A uniform quantization is a quantization technique in which the quantization levels are consistently spaced. Non-uniform Quantization is a method of quantization in which the quantization levels are mismatched and the relationship between them is mostly logarithmic.

- An important quantity in information theory and telecommunications engineering for defining theoretical upper bounds on channel capacity (or information transfer rate) in wireless communication systems, such as networks, is the signal to interference plus noise (SINR), also known as the signal to noise plus interference (SNIR).
- SINR is defined as the power of an interest signal divided by the total of the interference power (from all other interfering signals) and some background noise.
- The SINR equals the signal-to-interference ratio when the noise power term is zero (SIR).
- A mathematical model of wireless networks, such as cellular networks use the SINR, which is a more commonly used value than the SNR when it comes to simulating these networks.
- To define the quality of wireless communications, the phrase "Signal-to-noise Ratio (SINR)" is commonly used in wireless communication. Route loss in wireless networks occurs when a signal loses energy as it travels farther. In contrast, the correct receipt of data in wired networks is determined by the presence of a wired path between the sender or transmitter and the receiver.
- A set of techniques that involves coding helps in detecting errors in stored or transmitted data and rectify them. The use of several parity bits helps to detect any error, find if any bits are inverted, and in case they are, they should be re-inverted so that the original data is restored. Adding extra bits add to chances of detecting and correcting multiple errors. There are different methods depending upon Single Error Correction, Double Error Detection (SECDEC).
- Parity check using a single parity bit is the simplest and least expensive error detection method in which a parity bit is appended to a block of data, normally at the end of a 7-bit ASCII character. It consists of even parity and odd parity methods.
- Even parity method consists of one additional bit in such a manner that the total number of 1s in the resultant bit stream (the original information plus a parity bit) becomes even. Similarly, the odd parity method consists of addition of one bit in the original bit stream in such a manner that there are an odd number of 1s in the resultant bit stream.
- A carrier system is a type of telecommunications system that carries data, such as speech signals from a phone call or video signals from a television, for example. Typically, carrier systems transmit numerous channels of communication concurrently across a shared media. A digital carrier system is a type of communication system that encodes data using digital pulses rather than analogue signals.
- T-carrier is a completely digital system that employs Pulse Code Modulation (PCM) and Time Division Multiplexing (TDM). The system utilizes four wires and is capable of bidirectional communication (two wires for receiving

NOTES

and two for sending at the same time). T_1 lines should be reserved for mission-critical applications requiring high bandwidth. T_1 lines perform optimally when the sites being connected are close in proximity (otherwise the cost is prohibitive).

- E-carrier, a European carrier, maintains a hierarchy of digital transmission standards. E-carrier is based on the original North America T-carrier digital carrier system; although the specifics are quite different with respect to signaling rates, framing convention, line coding technique, and PCM compounding technique. In many respects E-carrier is a significant upgrade over T-carrier.
- Telegraphic devices transmit communications, referred to as telegrams, over great distances. The word "Telegraph" comes from two Greek words. "Tele" means "At a Distance" and "Gramma" means "Letter". When people refer to telegraphs, they usually mean electric telegraphs.
- Samuel Morse independently developed and patented a recording electric telegraph in 1837. Morse's assistant Alfred Vail developed an instrument that was called the register for recording the received messages. It embossed dots and dashes on a moving paper tape by a stylus which was operated by an electromagnet. Morse and Vail developed the Morse code signaling alphabet.
- A teleprinter (teletypewriter, teletype or TTY) is an electromechanical device that can be used to send and receive typed messages through various communications channels, in both point-to-point and point-to-multipoint configurations.
- Teleprinters could use a variety of different communication media. These included a simple pair of wires; dedicated non-switched telephone circuits (leased lines); switched networks that operated similarly to the Public Telephone Network (TELEX or TELegraph EXchange); and radio and microwave links (Telex-On-Radio, or TOR).
- A teleprinter attached to a modem could also communicate through standard switched public telephone lines. This latter configuration was often used to connect teleprinters to remote computers, particularly in time sharing environments.
- Wireless telegraphy, or radiotelegraphy, is the use of radio waves to transmit telegraph signals by radio waves.
- Radiotelegraphy transmits information using pulses of radio waves of two distinct lengths termed 'Dots' and 'Dashes' that spell out text messages, typically in Morse code.
- In a manual system, the sending operator taps on a button called a telegraph key to turn 'ON' and 'OFF' the transmitter, which generates radio wave pulses. At the receiver, the pulses are audible as beeps, which are converted back to text by a Morse code operator.

- As per the Encyclopedia Britannica, "The radiotelegraphy refers to the radio communication by means of Morse Code or other coded signals. The radio carrier is modulated by changing its amplitude, frequency, or phase in accordance with the Morse dot-dash system or some other code. At the receiver the coded modulation is recovered by an appropriate demodulator and the code groups are converted into the corresponding symbols. In many instances the symbols are generated by a computer and modem rather than with a manual telegraph key".

5.10 KEY TERMS

- **Digital transmission or digital communications:** Digital transmission or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel. These could be copper wires, optical fibers, wireless communication channels, and storage media.
- **Bit rate:** Bit rate is the rate over network speed which is used to detect errors while transmitting data. Bit rates are sometimes written as data rate and are conveyed and processed per unit time. It is measured as bits per second (bit/s or bps).
- **Data Signaling Rate (DSR):** In telecommunication, Data Signaling Rate (DSR), also known as Gross Bit Rate (GBR), is the aggregate rate at which data passes a point in the transmission path of a data transmission system.
- **Bit Error Rate (BER):** The Bit Error Rate (BER) is the number of bit errors per unit time.
- **Bit Error Ratio (BER):** The Bit Error Ratio (also BER) is the number of bit errors divided by the total number of transferred bits during a studied time interval. Bit error ratio is a unitless performance measure, often expressed as a percentage.
- **Digital filter:** A digital filter is a mathematical technique that operates on a digital input signal to generate a digital output signal with the objective of filtering out unwanted data. It is implemented in hardware and/or software.
- **Pulse Code Modulation (PCM):** Pulse Code Modulation (PCM) is a digital representation of an analog signal where the magnitude of the signal is sampled regularly at uniform intervals, then quantized to a series of symbols in a digital (usually binary) code.
- **Quantizer:** Quantizing is a process in which excessive bits are getting reduced and the data is confined. The sampled output when given to Quantizer, reduces the redundant bits and compresses the value.
- **Encoder:** The function of an encoder is to perform digitization of analog signal. It designates each quantized level by a binary code. The sampling done here is the sample-and-hold process. Encoding minimizes the bandwidth used.

5.11 SELF-ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. Why are digital communications used?
2. Give the advantages and disadvantages of digital communications.
3. Define bit transmission.
4. What is signaling rate?
5. What is error probability?
6. State about the digital filtering.
7. What is delta modulation?
8. Define Pulse Code Modulation (PCM) and PCM generation.
9. Why is binary coding done?
10. State about the PCM bandwidth and PCM reception noise.
11. Define quantization noise analysis.
12. What is SIN ratio and channel capacity of PCM?
13. What are error detection and correction codes?
14. Define digital carrier systems.
15. State about the teleprinters and telegraphs circuits.
16. What are radio telegraphs transmitters?

Long-Answer Questions

1. Briefly discuss the significance, types and characteristic features of digital communications giving appropriate examples.
2. Elaborate on the advantages and disadvantages of digital communications.
3. Discuss the concept of bit transmission with the help of examples.
4. Brief a detailed note on signaling rate, error probability and digital filtering giving appropriate examples.
5. Explain delta modulation giving examples.
6. Discuss in detail the significant features of Pulse Code Modulation (PCM) giving relevant examples.
7. Explain giving examples the PCM generation, PCM bandwidth and PCM reception noise.
8. Describe the significance of binary coding giving examples.
9. Elaborate on quantization noise analysis giving examples.
10. Discuss SIN ratio and channel capacity of PCM with the help of relevant examples.

11. What are codes? Explain the error detection and correction codes giving appropriate examples.
12. Briefly discuss the importance of digital carrier systems.
13. Discuss in detail the teleprinters, telegraphs circuits, and radio telegraphs transmitters giving examples.

NOTES

5.12 FURTHER READING

Gurung, J.B. 2009. *Signals and Systems*. New Delhi: PHI Learning Pvt. Ltd.

Carlson, A Bruce and Paul B. Crilly. 2010. *Communication System*. Noida: Tata McGraw-Hill Education.

Lathi, B. P. and Zhi Ding. 2010. *Modern Digital and Analog Communication Systems*. Oxford: Oxford University Press.

Kennedy, George and Bernard Davis. 1993. *Electronic Communication Systems*. London: Glencoe/McGraw-Hill.

Roddy, Dennis and John G. Coolen. 1977. *Electronic Communications*. Noida: Pearson Education.

Greenwich, Edwin. 2012. *Analog Integrated Circuits*. Berlin: Springer Science & Business Media.

Millman, Jacob and Christos C. Halkias. 2001. *Integrated Electronics*. Noida: Tata McGraw-Hill Education.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey: Prentice-Hall.

Tanenbaum, Andrew. S. 2002. *Computer Networks*. New Jersey: Prentice-Hall PTR.

Basavaraj, B. and H.N. Shivashankar. 2004. *Basic Electronics*. New Delhi: Vikas Publishing House Pvt. Ltd.