COMMUNICATION SYSTEMS

ABOUT THE AUTHOR

P Ramakrishna Rao obtained his BE (Telecommunications) degree from the Madras University. His ME degree on Microwave Engineering was obtained at the Indian Institute of Science, Bangalore, and his PhD degree in the field of Signal Processing from the Indian Institute of Technology Kanpur. He taught at the Regional Engineering College (now NIT), Trichy for over 25 years and retired as Professor in 1997. While at REC, Trichy, he had carried out six sponsored research projects funded by DOE and DRDO.

Professor Rao has over 25 research publications in reputed international journals and proceedings of international conferences. He served as a member of the National Radar Council's Working Group on Sonar and Underwater Electronics for a period of six years, from 1987 to 1993. He was the recipient of the first Academic Excellence Award of REC, Trichy and was a reviewer for three International Journals — IEEE Transactions on ASSP, IEE Part-F, and Electronics Letters (London). His areas of research interest include Signal Processing and Communications. Professor Rao has authored three books namely, *Signals and Systems*, *Analog Communication*, and *Digital Communication*, all published by Tata McGraw-Hill Education.

COMMUNICATION SYSTEMS

P Ramakrishna Rao
Former *Professor of ECE*
Regional Engineering College (now NIT Trichy)
Tiruchirapalli, Tamil Nadu

McGraw Hill Education (India) Private Limited
NEW DELHI

McGraw Hill Education Offices
New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

McGraw Hill Education (India) Private Limited

Published by the McGraw Hill Education (India) Private Limited
P-24, Green Park Extension, New Delhi 110 016

Communication Systems

Copyright © 2013, by McGraw Hill Education (India) Private Limited. No part of this publication may be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers. The program listings (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

This edition can be exported from India only by the publishers,
McGraw Hill Education (India) Private Limited

ISBN-13: 978-1-25-900685-2
ISBN-10: 1-25-900685-9

Vice President and Managing Director—MHE: *Ajay Shukla*

Head—Higher Education: Publishing and Marketing: *Vibha Mahajan*

Publishing Manager—SEM & Tech Ed.: *Shalini Jha*
Editorial Executive: *Koyel Ghosh*
Copy Editor: *Preyoshi Kundu*
Manager—Production Systems: *Satinder S Baveja*
Asst. Manager Production: *Anjali Razdan*
Production Executive: *Anuj K Shriwastava*

Asst. General Manager—Marketing: *Vijay Sarathi*
Sr. Product Specialist: *Tina Jajoriya*
Graphic Designer—Cover: *Meenu Raghav*

General Manager—Production: *Rajender P Ghansela*
Manager—Production: *Reji Kumar*

Information contained in this work has been obtained by McGraw Hill Education (India), from sources believed to be reliable. However, neither McGraw Hill Education (India) nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw Hill Education (India) nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw Hill Education (India) and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Typeset at The Composers, 260, C.A. Apt., Paschim Vihar, New Delhi 110 063 and printed at
Sanat Printers, 312 EPIP, HSIDC, Kundli, Sonipat Haryana.

Cover Printer: Sanat Printers

RACLCRDOLRYRR

In memory of my parents

CONTENTS

PREFACE

Over the past few decades, we have witnessed a remarkable progress in communication systems; both digital and analog. In keeping with these evolving technological advances of telecommunications, *Communication Systems* aims to make students get acquainted with these changes in a solid and illustrative manner. Although the book covers both analog and digital communication systems, greater emphasis has been placed on the latter by incorporating the most commonly known wireless digital technologies.

Communication Systems has evolved out of my lecture notes prepared during my time spent teaching undergraduate students in the Electronics and Communications Engineering (ECE) branch at Regional Engineering College (now National Institute of Technology) Tiruchirapalli, and Vasavi College of Engineering affiliated to Osmania University, Hyderabad. One of the goals in bringing out this book has been to present the subject in a clear, understandable, and logically organized manner and apply those concepts in solving exercise problems. Accordingly, every new concept is accompanied by one or more solved examples in order to emphasize the understanding of the concept, and students' ability to apply it.

The book is primarily intended to provide undergraduate students of ECE branch with a fairly good exposure to the basic principles of operation and analysis of digital communication systems along with analog communication systems. Utmost care has been taken to ensure that the book adequately covers all the topics included in the syllabi prescribed for analog communication and digital communication subjects by various Indian universities.

Salient Features

The following few lines talk about both content and pedagogical highlights of the textbook:

- Clear explanation of concepts using simple language and style
- Adequate in-depth coverage of pre-requisites like Signals and Systems and Probability and Random Processes by devoting three chapters for this purpose
- Focused and comprehensive coverage of digital baseband and band pass signaling and the noise performance of digital communication systems
- Relevant MATLAB examples at the end of each chapter
- Large number of appropriately selected problems at the end of each chapter, to enable students to apply the techniques learnt
- Excellent pedagogy to highlight every concept discussed in each chapter:
  296 Solved Examples
  284 Multiple-Choice Questions
  385 Review Questions
  220 Problems
  546 Diagrams

Scope of the Book

One-semester courses on 'Signals and Systems' and 'Probability and Random Processes' are essential prerequisites for Analog Communication and Digital Communication courses. In fact, most of the Indian universities do offer such courses before exposing the student to communication engineering. However, in order to cater to the needs of students who might not have an in-depth exposure to the topics in the prerequisite courses, three chapters (Chapters 2, 3 and 6) have been devoted in this text to adequately cover the necessary background material. This ensures that such students also will not have any difficulty in following Chapters 4, 5, 7, and 8 that deal with Analog Communication Systems and Chapters 9 to 15 that deal with Digital Communication Systems. The book will therefore be useful for

- all engineering undergraduate students specializing in ECE
- engineering undergraduate students of EEE, EI and CSE who have a one-semester introductory course on communication systems (Analog and Digital)
- students preparing for IETE and AMIE examinations
- students preparing for GATE and engineering services examinations
- practicing engineers as a reference text

Chapter Organization

Chapter 1 gives a qualitative explanation of the evolution of analog and digital communication systems, and discusses their basic components in a heuristic manner. **Chapters 2 and 3** deal with signals, their types and representation, geometrical aspects of signals, spectral analysis of signals, basics of LTI system theory and transmission of signals through systems. **Chapter 4** focusses with AM, DSB-SC, SSB-SC, VSB types of modulation, methods of generating modulated signals, methods of detection, and transmitters and receivers for these types of modulations. **Chapter 5** discusses the basic theory of FM and PM signals, their methods of generation, detection of FM signals, and FM transmitters and receivers. Probability theory, random processes, noise sources their characteristics and noise calculations are discussed in **Chapter 6**. The effect of noise on the performance of various types of AM systems and FM systems is discussed in **Chapter 7**. **Chapter 8** deals with low-pass sampling theorem, various types of sampling and reconstruction of a band-limited signal from its samples, band pass sampling theorem, quadrature sampling, generation of PAM, PDM and PPM signals and their detection. **Chapter 9** discusses waveform coding – PCM, DM, DPCM and ADPCM, time domain, frequency domain, and model-based speech compression techniques and digital multiplexing. Digital data transmission (binary and M-ary, baseband along with band pass) as well as reception are discussed in **Chapter 10**. In **Chapter 11,** a mathematical analysis of the performance of different types of digital communication systems in the presence of noise is presented. Information measure, entropy of a source, Shannon's theorem pertaining source coding, source coding for DM sources, channels, mutual information and channel capacity, Shannon's channel coding theorem, continuous sources and differential entropy, Shannon's information capacity theorem, and assessing the performance of some practical systems in the light of Shannon's theory, are discussed in **Chapter 12**. Various error-detecting and correcting codes, their effectiveness (performance) and implementation techniques for coding and decoding, are presented in **Chapter 13**. Spread spectrum communication systems, their ability to provide secure and reliable communication, CDMA using spread-spectrum techniques and generation of good PN codes are discussed in **Chapter 14**. Finally multipath fading and OFDM its usefulness and applications are given in **Chapter 15**.

How to Use the Book

The book has been so designed that it permits its use as a textbook for a variety of communication systems courses meant for UG programmes. Three different ways are indicated below:

- For a single-semester course in communication systems (Analog and Digital) meant for ECE students: *Chapters 1, 4, 5, noise part of Chapter 6 and Chapters 7 to 14.*
- For a two-semester course in communication systems (Analog and Digital) meant for ECE students:
  First semester: Chapters 1 to 8
  Second semester: Chapters 9 to 15
- For a one-semester course in communication systems, meant for non-ECE students :
  Chapter 1, selections from Chapters 4 and 5, Chapters 8 (noise performance of Analog Pulse Modulation may be omitted), 9 and 10 and selections from Chapter 13.

Web Supplements

Following web supplements may be accessed at *http://www.mhhe.com/rao/cs*:

- Solution manual
- Lecture slides

Acknowledgements

I am deeply indebted to many individuals for their help, guidance and support while writing this book. I am thankful to Dr. Shankar Prakriya, Professor, Department of Electrical Engineering, IIT Delhi, Prof K. Parvatheesam, former professor, Andhra University, and Prof. Prasad Gandikota, GITAM University for sparing time from their busy schedule to provide me with MATLAB examples. I would also like to thank Mr. Y. Rajani Kanth for typing the manuscript with meticulous care and generating all the line diagrams neatly and accurately.

The feedback from the following reviewers was extremely helpful in improving the quality and usefulness of the book and I am deeply indebted to all of them.

Satyabrata Jit	*Indian Institute of Technology (Banaras Hindu University) Varanasi, Uttar Pradesh*
Ashutosh Singh	*Harcourt Butler Technological Institute, Kanpur, Uttar Pradesh*
Mahesh Kumawat	*Shri Vaishnav Institute of Technology & Science, Indore, Madhya Pradesh*
Ekram Khan	*Aligarh Muslim University, Aligarh, Uttar Pradesh*
Sumit Kundu	*National Institute of Technology, Durgapur, West Bengal*
Sarat Kumar Patr	*National Institute of Technology, Rourkela, Orissa*
Poonam Singh	*National Institute of Technology, Rourkela, Orissa*
S N Singh	*National Institute of Technology, Jamshedpur, Jharkhand*
Sanjeev Gupta	*Dhirubhai Ambani Institute of Information and Communication, Gandhinagar, Gujarat*
Vijay Kumar Chakka	*Dhirubhai Ambani Institute of Information and Communication, Gandhinagar, Gujarat*
Sunil Uttamchand Nyati	*Shrama Sadhana Bombay Trust's College of Engineering and Technology, Bambhori, Jalgaon*
Anushka Deepak Kadage	*Shivaji University, Kolhapur, Maharashtra*
R Ramesh	*Sri Ramaswamy Memorial University, Kattankulathur, Tamil Nadu*
Ramesh Jayabalan	*P S Govindaswamy College of Technology, Coimbatore, Tamil Nadu*
G Nagarajan	*Pondicherry Engineering College, Puducherry*
K Jayanthi	*Pondicherry Engineering College, Puducherry*
R Vinod Kumar	*Salem College of Technology, Salem, Tamil Nadu*
Vineeta P Gejji	*Karnataka Law Society's Gogte Institute of Technology, Belgaum, Karnataka*
Ram Krishna Dasari	*Osmania University, Hyderabad, Andhra Pradesh*

I am thankful to the editorial team of Tata McGraw-Hill for encouraging me to write this book. The constant cooperation and helpful suggestions from the editor of this book, Ms. Koyel Ghosh and other members of her team in shaping this book to its present form has helped.

Finally, I am deeply indebted to my wife and children for their constant encouragement and excellent cooperation they have extended to me throughout.

P Ramakrishna Rao
Visakhapatnam

Feedback

Utmost care has been taken to make the book error-free; nevertheless, some errors might have crept in the book. Suggestions/constructive criticisms from the readers of this book are most welcome. Readers can write to me at *prakriya37@hotmail.com*.

P Ramakrishna Rao

Publisher's Note

It will encourage us to receive your comments/compliments/ideas on this present edition. Please write to us at *tmh.elefeedback@gmail.com* by mentioning the title and the author's name in the subject line. Report of any piracy related problems will be highly appreciated.

INTRODUCTION

1

> *"Ideas shape the course of history."*
>
> **John Maynard Keynes (1883–1946)**
> *British economist*

Learning Objectives

After going through this chapter, students will be able to

- *become familiar with the history of evolution of modern electrical communications,*
- *know the basic building blocks of an analog communication system as well as of a digital communication system, and the basic functions of each of the blocks,*
- *know how the electromagnetic spectrum is divided into different frequency bands, and the application areas of each of these bands,*
- *understand the characteristics and limitations of various communication channels,*
- *come to know the way different types of communication channels are modeled,*
- *know about communication resources and their significance and also understand the effect of noise, and the performance indices relevant for analog and digital communications, and*
- *become aware of the advantages and disadvantages of digital communication as compared to analog communication.*

Over the last few decades, various forms of modern electrical communications have permeated into every aspect of our daily lives. These include not only the telephone, radio, television, and FAX, but also the digital and multimedia communications making use of computer networking and cellular phones. By cutting down costs and the time required for information transmission, these modern communication systems have so radically changed our way of life and thinking and improved the quality of our lives, that we dread to imagine what our lives would be like without them.

1.1 HISTORICAL PERSPECTIVE

The era of electrical communication began with the successful commissioning, in 1844, of the first telegraphic link between Washington and Baltimore by Samuel Morse. Even though it is not widely known, it is a fact that this first electrical communication to be established, viz. telegraphy, is basically a digital communication. The telegraph code, invented by Morse and subsequently named after him as the 'Morse code', is indeed, a remarkable invention and reveals his inventive genius. It is a variable length binary code which used two symbols—*the dash* and *the dot*. The fact that he had assigned a single dot for representing 'e', the most frequently occurring letter of the English alphabet, bears testimony to his genius which recognized the need for coding efficiency a full century before Shannon talked about it in his information theory. Although binary codes with ‘0’ and ‘1’ as the two symbols were used by Gottfried Wilhelm Leibnitz [8] even in 1705, it was only in 1875 that a fixed-length five-digit binary code was designed by Baudot.

The invention of ‘telephone’ by Alexander Graham Bell in 1870 marked the beginning of the era of telephone communication. Although it had to be only over short distances in the beginning, subsequently the availability of the vacuum triodes for amplification of the telephone signal at regular intervals made it possible to have long-distance telephony. While the invention of the electromechanical Strowger switches led to the establishment of the ‘Automatic Exchanges’ which did not need any operators for establishing connection between the initiator of a call and the called party, the invention of the transistor paved the way for replacement of the electromechanical Strowger switches by solid-state switches which are faster and also are more reliable as they do not have any moving parts. Digital telephone switching and digital transmission of voice signals thus became possible in 1960s. In fact, the first T-1 carrier system of TDM multiplexed digital transmission, installed by Bell Telephone Laboratories, became operational in 1962. Approximately a decade later, this then paved the way for the Integrated Services Digital Network (ISDN) in which, a broad range of services like data, voice, and video are provided by digitizing and then integrating all of them using Time-Division Multiplexing (TDM) and transmitting in a digital form.

That wireless communication over long distances is possible by using electromagnetic waves became evident after Marconi demonstrated in 1901, the transmission of electromagnetic waves over a distance of about 2500 km. With the invention of the vacuum triode by De Forest in 1906 and AM superheterodyne receiver by Armstrong during World War I, AM Radio broadcasting became firmly established. Subsequently, around 1930, Armstrong invented frequency modulation and FM radio broadcasting also was started. Further, Zworykin of the United States of America demonstrated the first ‘Television System’ in 1929, and commercial TV broadcasting was started a few years later.

Telephone communication witnessed remarkable progress with the advent of wireless communications. Also the launching of the first communication satellite, the ‘Early Bird’, in 1965 marked the beginning of commercial satellite communications. Availability of communication satellites and optical fibers made it possible to transmit an extremely large number of telephone conversations simultaneously over very long distances.

The aforementioned developments like digital transmission of voice, ISDN, etc., would not have been possible but for the invention of PCM. Alec Reeves [12] invented PCM in 1937 itself. It was used in World War II to develop an encryption system for speech signals, and in fact, a 24-channel system was developed during the fag-end of the war. Extensive use of PCM, however, had to wait till transistors and VLSI technology became available to make the implementation of large-scale PCM-based systems possible and cost-effective. This started happening in late 1960s. In fact, it can be said that it was PCM which started the era of real digital communication. At this stage, one notable development that helped in the design of better PCM receivers, was the ‘Matched Filter’ developed by D.O North [13] in 1943 itself. North, in fact, developed it for optimum detection of known signals in the presence of noise – a problem that was of immense importance in the detection of targets using radar. Just as in the case of radar, in the case of detection of PCM signals too, the waveform of the transmitted signal, the presence or absence of which is to be detected, is known *a priori* and the actual shape of the waveform is not what is desired to be extracted from the received signal – only its presence (or absence) is to be correctly detected in the presence of noise. This similarity in the issues related to signal detection in the two cases made engineers to adopt the matched filter detection in PCM too, in order to optimize the receiver performance.

Attention then turned towards higher rate of transmission and better digital modulation techniques. The ‘adaptive equalization’ of Lucky [10] proposed in 1965, helped quite a bit in attaining higher transmission rates. The need for better digital modulation techniques became acute owing to another development around that time. Computers became all pervading especially after the Intel Corporation came out with its microprocessors in the early 1970s. Fast transmission of data over long distances with *almost zero probability of error* became a necessity. This goal of fast and reliable transfer of digital data was soon achieved partly by the use of adaptive equalization and partly by making use of the pioneering work done by G. Ungerboeck [17] in the area of channel coding and modulation. The concept of cellular telephony proposed by AT&T Bell Laboratories in 1968 [11] could not be implemented as a practical system till 1983 owing to non-availability of required technology. The newer generation cellular mobile communication systems also spurred research in better and more efficient modulation techniques. There was a gradual shift in the types of modulation used in these systems – from binary to QPSK, M-ary QPSK, etc., and the GMSK, OQPSK, and $\pi/4$ QPSK.

Parallel to these developments on the systems side, some very useful theoretical work was done, the outcome of which effectively addressed the problem of improvement of reliability of digital communication systems, which was mentioned earlier. Claude E. Shannon [16] of the Bell Laboratories, had, through his now famous paper of 1948, laid sound theoretical foundation for digital communications. Contrary to the belief held by the communication engineering fraternity at that time, he had stated and proved in his paper *that it is possible* to have error-less transmission of information over a channel even in the presence of noise. He showed that the condition required to be satisfied for this to be possible, is that the rate of transmission should be less than what is known as the *channel capacity*. This exciting possibility led to an unprecedented research activity in the area of error correcting codes during the next two to three decades, resulting in error correcting codes by Golay [6] and Hamming [7] around 1950, and the BCH codes [4]. All these codes come under the category of block codes. In these codes, a block of k information digits are mapped into n digits forming a codeword. This mapping is one-to-one but it cannot be implemented in real-time owing to the long processing time required. A totally new class of error-correcting codes, known as *convolutional codes*, originally invented in the 1960s by Forney [5] and later fully developed in 1970s by Forney, Viterbi, Wozencraft et al., could be implemented in real-time and were easy to implement. Further, in 1982, Ugerboeck [17] combined channel coding using convolution codes with modulation in the form of Trellis – coded modulation. Although use of these codes ensured almost error-free communication, they were far behind the optimum codes envisaged by Shannon insofar as their bit-energy-to-noise PSD ratio is concerned. The very powerful turbo codes [1,2] of Berrou, Glavieux and Thitimajshima proposed in 1993, however, came very close to optimum codes in the Shannon sense.

Research in the area of computer communications and computer networks, started in a very modest way using low speed transmission on voice-grade channels in the late 1950s, made significant progress between 1970 and 1995. One of the very important outcomes of this technique was intensely studied on 'ARPANET', a computer network developed specifically for research in the area of computer communication. This ARPANET was renamed as INTERNET in 1991 and put into commercial operation in 1994. Together with satellite communication, INTERNET has, over the last 16 years, revolutionized data, and voice and video communication. The computer has, over the last two decades, emerged as a means of communication. This convergence of computers and communications has revolutionized the field of communications. The inherent reliability and flexibility of digital communications and the availability of hardware at ever. decreasing prices, are making realization of even complex digital communication systems not only possible but also cost-effective. It is therefore not surprising that digital communications are steadily taking over the communications field.

1.2 BASIC BUILDING BLOCKS OF A COMMUNICATION SYSTEM

There are a very large number of different types of communication systems. So, by the term, 'communication system', one may be referring to an audio broadcasting system, a television broadcasting system, or a wireless computer network. As we are going to see later, these different communication systems do differ from each other in certain details. Nevertheless, there are certain basic functional blocks which one can find

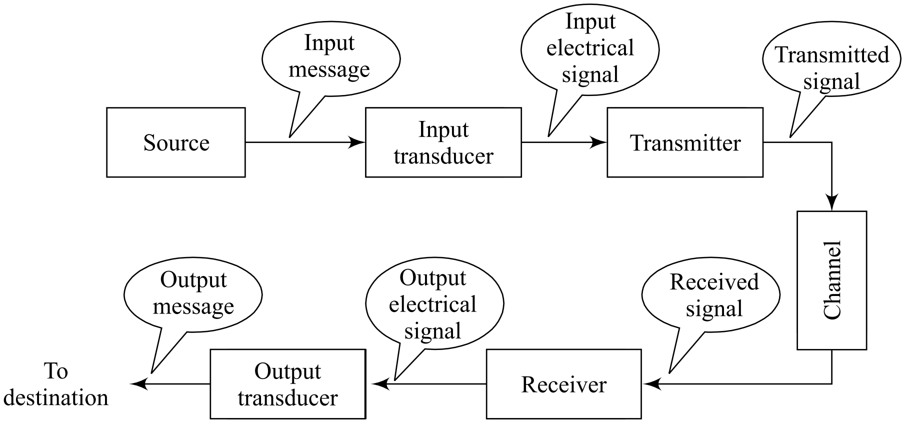


Fig. 1.1 *Block diagram of a general communication system*

in all of them. Our interest, at this stage, is to describe the working of a general communication system in terms of these basic functional blocks. Such a block diagram of a general communication system is shown in Fig.1.1. An appropriate interpretation of the function of each of the blocks of Fig.1.1 will make it represent practically any of the various types of communication systems in existence.

1. Baseband signal: The message to be transmitted originates from the source. This input message may be human voice, a still picture, or a video signal. If it is not electrical in nature, an appropriate transducer is used to convert it into an electrical signal. For instance, if the source produces human voice, which consists of pressure variations, the input transducer to be used is merely a microphone. The output electrical signal of this input transducer is called the *baseband signal*.

2. Transmitter: As it may not be possible to transmit baseband signal over long distances, the transmitter modifies the baseband signal so as to make it suitable for transmission over the channel. The transmitter consists of several subsystems and these take different forms depending upon the type of communication system under consideration. It may consist of baseband amplifiers, RF amplifiers and a modulator, or A/D converter, encoder and modulator.

3. Channel: The output of the transmitter, called the *transmitted signal*, goes through the channel to the receiver. Depending on the nature of the communication systems, the channel may take a variety of forms – it may be just a pair of twisted copper wires, a coaxial cable, a radio link, an optical fiber, or even a combination of some of these.

The transmitted signal and the received signal will, in general, be different. This is because of the modifications suffered by the transmitted signal during its passage through the channel. These modifications can be identified as the following:

(a) Distortion of the signal
(b) Attenuation of the signal
(c) Addition of noise

Distortion of the signal is due to frequency dependent gain/attenuation of the channel, multipath and Doppler shift. Such linear distortions can be eliminated, or at least partly removed by using at the receiver, an equalizer whose frequency response characteristic is complementary to that of the channel. Non-linear distortion is caused by channels whose attenuation is dependent upon the amplitude of the signal passing through them.

Attenuation of the signal during its passage through the channel may be caused by the 'power loss' in the case of wire-line channels, and the spreading and absorption that takes place in the case of free space and atmosphere. Attenuation increases with the length of the channel.

As the signal passes through the channel, it goes on getting corrupted by random interfering (i.e., undesired) signals and other electrical disturbances, all of which together are generally termed as *noise*. Thus, while the signal strength goes on decreasing with distance from the transmitter because of the channel attenuation, the noise goes on getting accumulated and hence becomes stronger and stronger. Noise may originate from external sources such as lightning discharges, radiation from automobiles ignition or fluorescent tubes or radiations from the sun or galactic sources. It may also originate from sources internal to the channel, such as the random movement of electrons in a conductor which gives rise to *thermal noise*, or the randomness inherent in the recombination or diffusion or partitioning of charged particles inside electronic devices. By proper design, one may reduce or even eliminate externally generated noise but internal noise can only be reduced but cannot be eliminated altogether. Noise is, in fact, the most important problem in communication, as it impairs the quality and rate of communication.

Communication channels may, as has been already mentioned, take a variety of forms. In certain communication systems, there may be a physical connection between the transmitter and the receiver, say, using a transmission line. In others, no such physical connectivity may exist. Based on this, channels may be broadly classified into the following two types.

(a) *Wire-line channels:* Telephone lines, coaxial cables, waveguides, optical fibers, etc.
(b) *Wireless channels:* Microwave radio, RF links, underwater acoustic channels, etc.

(a) Telephone lines: Twisted-pair of wires are extensively used for connecting telephone subscribers to the local telephone exchange. They provide a modest bandwidth of a few hundred kilohertz (kHz). However, they suffer from cross-talk interference and induced additive noise but provide good voice-grade communication over the frequency range of 300–3400 Hz, and have a fairly good linear response. Although no particular attention is paid to their phase response insofar as speech signals are concerned (since the ear is not sensitive to phase delay), in the case of data and image transmission, phase-delay variations do cause problems and so a linear phase response over the frequency range of interest, is a must. Hence, equalizers are used, which tend to keep a constant overall amplitude response and a linear phase response. Use of efficient modulation techniques, combined with adaptive equalization, can make it support data rates of the order of 20 kbps.

(b) Coaxial cables: With an outer hollow conductor and an inner conductor placed at its center, and a dielectric filling the intervening space, the coaxial cables are used for carrying signals of frequencies ranging from a few hundred kilohertz to about one gigahertz. They can provide a data rate of even 250–300 Mbps. The attenuation, of course, depends on the frequency, size and the type of dielectric filling, and a typical value may be around 200 db per 100 meters at 1 GHz. Thus, close spacing of repeaters becomes necessary. One helpful feature of the coaxial cables is that they are immune to interference and additive induced noise.

(c) Waveguides: Waveguides are typically used at frequencies ranging from about 1 GHz up to several hundreds of gigahertz. They can support large bandwidths of the order of a few gigahertz. Attenuation depends upon frequency, length, and material used for making the waveguide, and the internal coating used, if any. Rectangular copper waveguides of 25.4 mm × 12.7 mm cross section provide a typical attenuation of 0.11 dB per meter at 10 GHz. They are immune to interference and induced additive noise. However, they are quite expensive and are therefore used only for very short distances.

(d) Microwave radio: This is a wireless type of channel. Microwave radio operating in the line-of-sight (LOS) propagation mode, and using frequencies in the range of 1–300 GHz can be used for transmission of time-division multiplexed (TDM) digital data over long distances. This needs the antennas to be placed at considerable heights so as to get a clear line of sight between two successive repeaters. It gives a very large bandwidth and thereby permits high speed data transmission. Multipath fading is common in these terrestrial microwave LOS links unless sufficient care is taken while designing the links. Attenuation depends on meteorological conditions and fading can be heavy due to rain, snowfall or the passage of a cloud either along, or cutting across the line-of-sight path between two repeaters.

(e) Optical fiber: Optical fiber consists of a central '*core*' surrounded by another layer, called the *cladding*. Both core and cladding are made of silica while the *jacket*, which in turn surrounds the cladding and protects it, is made of plastic. The core carries electromagnetic waves at optical frequencies of the order of 10^{14} Hz and these waves are confined to the core by total internal reflection caused at the core-cladding interface owing to the difference in their refractive indices. Optical fibers support extremely large bandwidths – almost 10% of the center frequency, amounting to nearly 10^{13} Hz. Modern optical fibers provide very little attenuation in the order of 0.2 dB/km. They have a lot of advantages – immunity to interferences and induced noises, very small size and light weight, besides flexibility and ruggedness. Further, since they are made out of pure silica glass for which sand is the raw material, they are potentially low-cost wideband transmission lines.

4. Receiver: The main task of the receiver is to extract the baseband signal from the received signal. This output signal from the receiver should be as close an approximation as possible to the baseband signal at the input to the transmitter. The receiver does this by subjecting the received signal to a process called *demodulation*, and by removing or reducing the distortions and noise introduced during the passage of the transmitted signal through the channel.

The *output transducer* takes the electrical signal from the receiver, converts it into the same form as the input message at the transmitter and passes it on to the *destination* to which the original message from the source was intended.

1.3 ANALOG AND DIGITAL MESSAGES

The message signal to be transmitted, viz. the baseband signal, may be either an analog signal, or a digital signal. An analog signal is a continuous-time signal whose amplitude can take a continuum of values, for example, the temperature, the atmospheric pressure at a particular place, or the speech signal.

A digital message signal, on the other hand, is an ordered sequence of symbols drawn from a finite set of symbols, for example, a Morse-coded message which consists of an ordered sequence of two symbols, the dot and the dash. Another example is an ordered sequence of the two symbols, 0 and 1. We may choose to represent the symbol '0' by a pulse of zero amplitude or –V volts amplitude, and the symbol '1' by a pulse of +V volts amplitude. This is the familiar binary signal whose amplitude can take only one of a possible set of two values either 0 volts and +V volts, or –V volts and +V volts. If the finite set (of symbols) has M symbols where M is an integer and if we form an ordered sequence of symbols drawn from that set, the digital signal so constructed, is called a M-ary signal.

It may be noted that while it is possible to have an infinite number of different analog signals over a given finite interval of time, the number of different digital signals over the same interval of time is however, finite.

1.3.1 Analog and Digital Communication Systems

Communication systems may be broadly divided into two types – analog communication systems and digital communication systems. In analog communication systems, the baseband message, which is to be transmitted, is in the form of an analog signal. In digital communication systems, the baseband message, which is to be transmitted, is in the form of a digital signal. The baseband message signal fed to the transmitter may be in digital form either because the source has produced it in that form, or an analog waveform produced by the original source might have been sampled, quantized, and encoded to convert it into digital form before feeding to the transmitter as the baseband signal.

It may be noted that in a digital communication system, even though the baseband message signal is in digital form, the transmitted signal may still be an analog signal as in the case of a band pass digital communication system where, in the transmitted signal, we may use a sinusoid of one frequency to represent a binary '0' of the baseband digital signal, and a sinusoid of a different frequency to represent a binary '1' of the baseband digital signal.

1.4 AN OVERVIEW OF ANALOG COMMUNICATION SYSTEMS

Any analog communication system basically consists of a source, a transmitter, a channel and a receiver as shown in Fig. 1.2.

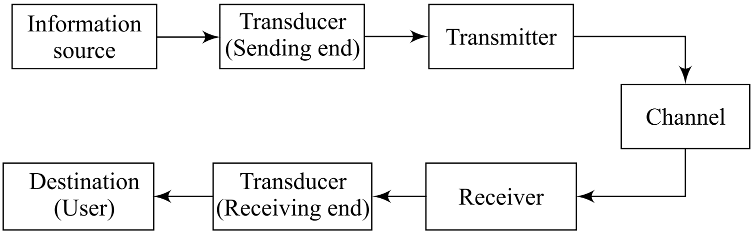


Fig. 1.2 *Basic elements of an analog communication system*

The information source produces information, which may be in the form of a speech, or music, or images. Since the output from the source of information is not an electrical signal, a suitable transducer is to be used to convert the information into an electrical signal so that the transmitter can handle it. If the source produces speech or music, the required transducer may be simply a microphone. For an image, it may take the form of a video camera that scans it and produces an electrical signal.

1. Transmitter: The job of the transmitter is to put the information bearing electrical signal (message) into a form suitable for transmission over the channel. Generally, the message signal is made to modulate a high frequency sinusoidal signal (generated in the transmitter), called the *carrier*. Modulation is a process by which one of the three characteristic parameters – amplitude, frequency, and phase of the carrier signal, is made to vary in accordance with the variations in the *amplitude* of the message signal. The message is thus *carried* by the carrier wave in the form of variations in its amplitude, frequency, or phase. It therefore amounts to translating the low-frequency message signal along the frequency scale. The resulting modulation is called *amplitude modulation* if carrier amplitude is the parameter which is varied; *frequency modulation* if the carrier frequency is the parameter which is varied; and *phase modulation*, if the carrier phase is the parameter which is varied. Through the frequency translation resulting from the modulation process, two things are achieved. First, since the size of an antenna has to be at least about 0.1λ for it to act as an efficient radiator of electromagnetic waves, it now becomes possible to have an antenna of reasonable size to radiate the modulated signal which has high frequency components only. Second, by using carrier signals of different frequencies for different transmitters, it would be possible to transmit several message signals simultaneously over the same physical channel, i.e., say, free space, without these signals interfering with each other. This process is called *multiplexing*. The transmitter not only translates the message into an appropriate high frequency band, it also sees to it that the power of the modulated signal which is ultimately fed to the antenna, is at an appropriate level.

2. Channel: The channel carries the output signal of the transmitter to the receiver. This output signal of the transmitter may have frequencies ranging from extremely low frequencies as those used in submarine communications to optical frequencies (typically 10^{15} Hz) as those used in optical communication systems. The bandwidths used may range from a few tens of Hertz to several hundreds of mega Hertz. The channel may take a variety of forms depending upon the frequency and the bandwidth of the signal and the application. It may be just a pair of twisted copper wires, a coaxial cable, a waveguide, an optical fiber, atmosphere and free space, or ocean water (as in the case of underwater communication systems). It may also be a combination of some of these.

3. Receiver: The primary job of the receiver is to recover the message signal from the transmitted signal received by it through the channel. Because of the attenuation caused by the passage through the channel, the received signal is generally very weak. So the receiver amplifies the received signal, and in case it is a modulated signal, it demodulates it. Because of the distortion suffered by the transmitted signal during its passage through the channel and because of the noise and interfering signals that have been added on to it, the demodulated signal at the output of the receiver will not be an exact replica of the message signal fed to the transmitter. The receiver may also contain circuitry intended to improve the signal-to-noise ratio (by filtering) and noise suppression. An appropriate transducer at the receiving end converts the electrical signal from the output of the receiver into a form suitable for the user at the destination. If the original message that was transmitted was speech or music, this receiving-end transducer would be just a loudspeaker.

1.5 AN OVERVIEW OF DIGITAL COMMUNICATION SYSTEM

In this section, a brief overview of a digital communication system will be given. In essence, the basic functions of each one of the elements that constitute a digital communication system making use of a block schematic diagram of the system for this purpose will be discussed.

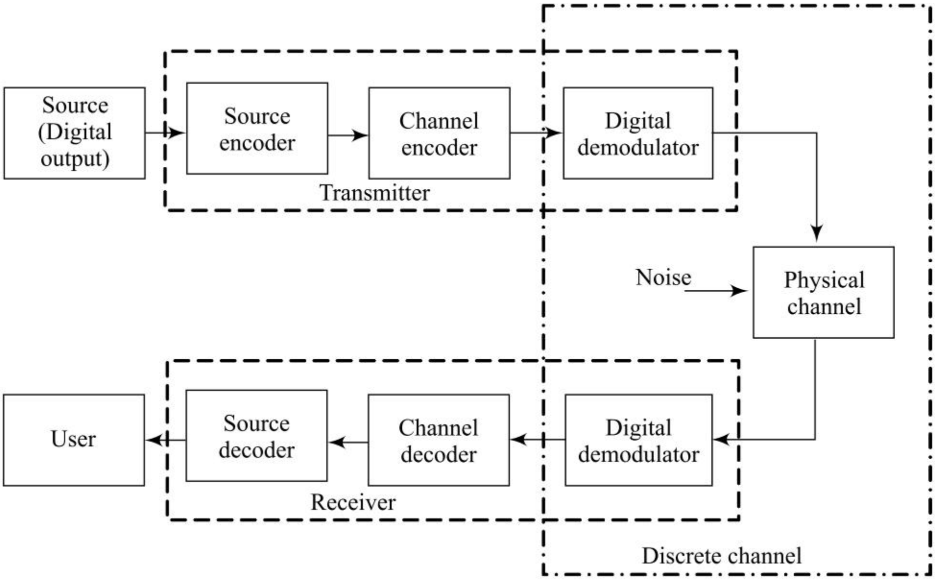


Fig. 1.3 *Block diagram of a digital communication system*

Note

(i) The source encoder, channel encoder, and digital modulator together constitute the transmitter.

(ii) The digital demodulator, channel decoder, and the source decoder together constitute the receiver.

(iii) The digital modulator, physical channel, and digital demodulator together constitute what is referred to as the 'Discrete Channel', as the input as well as the output of it are digital data.

1. Source: The source is where the information to be transmitted originates. This information/message may be available in digital form, as for instance, in the case of the output of a teletype system, or it may be available in an analog form. In the later case, it is sampled and digitized using an A/D converter to make the final source output to be digital in form.

2. Source encoder: Generally, the bit-stream at the source output will have considerable redundancy and so it will not be an efficient representation of the message or information given by the source from the point of view of the number of digits used. Perhaps, a fewer number of digits might suffice to convey the information. The source encoder therefore reduces the redundancy by performing a one-to-one mapping of its input bit-stream into another bit-stream at its output but with fewer digits. Thus, in a way, it performs data compression.

3. Channel encoder: Basically, the channel encoder is intended to introduce *controlled redundancy* into the bit-stream at its input in order to provide some amount of error-correction capability to the data being transmitted. This is needed because the data gets corrupted by the additive noise on the channel and this gives rise to the possibility of the channel decoder committing mistakes in the decoding of the data received from the channel. Redundancy helps in detecting erroneously decoded bits and makes it possible to correct the errors before passing on the data to the source decoder. As a crude example of how redundancy can help in error correction, consider the case of a channel encoder *triplicating* each digit in the binary bit-stream at its input. Since the channel noise affects the data randomly, it is very unlikely that all the three triplicated digits (corresponding to a particular data digit) would be affected. In fact, the probability of more than one out of the three triplicated digit being affected will be extremely low. So, if at all any one of the three triplicated bits is affected, by applying the 'majority rule', one can correct the affected digit and ensure correct decoding by the channel decoder.

From the foregoing, one should not jump to the conclusion that the channel encoder will undo the data compression achieved by the source coder and therefore nothing is gained by having the source coder. Which is better, removing the redundancy as far as possible before introducing some amount of *controlled* redundancy or, introducing the controlled redundancy straightaway without initially reducing the redundancy by using a source encoder?

4. Digital modulator: In most digital communication systems, the channel coder and the digital modulator are two separate entities. But in some cases, they may be combined. Here, for the sake of this overview of digital communication systems, we are considering them as two separate entities.

Since the physical channels are basically analog in nature, the digital modulator may take each binary digit at its input and map it, in a one-to-one fashion, into a continuous waveform. That is, a binary 'zero' at its input is mapped into a continuous signal $s_0(t)$ and a binary 'one' is mapped into another continuous signal, $s_1(t)$. This is called *binary modulation*. Alternatively, it may, each time, take a block of say 'b' binary digits at its input and map that block in a one-to-one fashion into one of a set of $2^b = M$, distinct continuous-time signals, $s_i(t)$, $i = 0, 1, 2, \ldots, (N - 1)$. This is called *M-ary digital modulation* ($M > 2$). Here, M distinct continuous-time signals are used because when a block of 'b' binary digits is considered; there will be 2^b *distinct* blocks, each containing 'b' binary digits. Normally, the number of binary digits transmitted over the channel per second is called the bit-rate and is denoted by R (in bits/sec). When M-ary digital modulation is used, if the channel bit rate is R, b/R seconds is the time available to transmit one of the M waveforms. Thus, while in binary modulation, the time available for each bit is $1/R$ sec, in M-ary modulation, the time available for transmitting each one of the M waveforms is b times that, i.e., (b/R) sec.

As you might have realized by now, all the operations done till now, the A/D conversion of the source output if it was in the analog form, the source coding, the channel coding and the digital modulation, are all only signal processing operations performed on the original message signal for achieving various definite objectives. So are the digital demodulation, channel decoding and source decoding operations in the receiver.

5. Physical channel: The digitally modulated signal is passed on to the physical channel, which is nothing but the physical medium through which the signals are transmitted. It may take a variety of forms – a pair of twisted wires, a coaxial cable, a waveguide, a microwave radio, or an optical fiber. A long-distance telephone call, for instance, may go through each one of these different forms of physical channels, at different stages of its passage from the place of origin to the final destination. During its passage through the channel, the signal gets corrupted by noise. This noise may be thermal noise originating from electronic circuits, or atmospheric noise, or man-made noise, or, as is generally the case, a combination of most, or all of them. We will discuss in some detail about the characteristics of different types of channels and their modelling in Section 1.8 of this chapter.

6. Digital demodulator: The digital demodulator of the receiver receives the noise corrupted sequence of waveforms from the channel and by inverse mapping, tries to give at its output, an estimate of the sequence of the binary (or M-ary) digits that were available at the input of the digital modulator at the transmitting end.

7. Channel decoder: The output sequence of digits from the digital demodulator are fed to the channel decoder. Using its knowledge of the type of coding performed by the channel encoder at the transmitting end, and using the redundancy introduced by the channel encoder, it produces as its output, the output of the source coder of the transmitter with as few errors as possible.

8. Source decoder: Using its knowledge of the type of encoding performed by the source encoder of the transmitter, the source decoder of the receiver *tries* to reproduce at its output, a replica of the output of the digital source at the transmitting end. It may not, however, be an exact replica of the source output. There may be some errors in the sense that some of the binary 1s produced by the source might be received by the user at the destination as 0s and vice versa. In a long sequence of binary digits transmitted, the *fractional* number of times such errors occur *on the average*, is referred to as the '*probability of error*'. A typical value of the probability of error may be, say, 1 in a hundred million, i.e., 10^{-8}.

If the original message signal from the source was in analog form, and if the user at the destination again needs it in the same form, a D/A converter is used to convert the binary bit stream at the output of the source decoder into an analog form.

1.6 ADVANTAGES AND DISADVANTAGES OF DIGITAL COMMUNICATION

Digital communication has several advantages over analog communication, as listed below.

1. **Complexity:** Digital communication systems are difficult to conceptualize but are simpler and easy to build, whereas analog communication systems are easy in conceptual terms but are complex and difficult to build.
2. **Cost:** In the earlier days, digital communication systems were quite expensive and so could not compete effectively with analog communication systems. But the costs of digital communication systems are coming down year after year because of improvements in VLSI technology and consequent availability of ICs at ever decreasing costs.
3. **Robustness:** Components and subsystems used in digital communication systems are more robust since they are inherently insensitive to variations in atmospheric conditions like temperature and humidity etc. and are not affected by mechanical vibrations. The problem of ageing of components and subsystems does not therefore arise as it does in the case of analog communication systems. Further, they do not need frequent adjustments.

4. **Storage and retrieval:** Storage and retrieval of voice, data or video at intermediate points (in the transmission path) is easy and is inexpensive in terms of storage space in the case of digital communications.
5. **Flexibility:** Digital communication offers considerable flexibility.
   (i) Data, voice and video can all be multiplexed using TDM and transmitted over the same channel.
   (ii) Signal processing and image processing operations like compression of voice and image signals, etc. can easily be carried out.
   (iii) Digital modulating waveforms, best suited for channels with known characteristics, can be chosen so as to make the system more tolerant to channel impairments.
   (iv) Adaptive equalization can be implemented.
6. **Effect of noise and interference:** Error-correction codes used in channel coding ensure fairly good protection against noise and interferences in the case of digital communications.
7. **Long-haul communication using a number of repeaters:** In long distance communication making use of a number of repeaters, digital communication scores over analog communication for the following reasons.

   In analog communication, at each repeater, the received signal is amplified and retransmitted. These amplifiers amplify the signal and noise components equally and also *add some more noise*. Because of this, after each repeater the SNR actually becomes lower and lower. So, when the signal reaches the destination after going through a large number of repeaters, it is dominated by noise.

   In digital communications however, if only we ensure that the signal power at the input of a repeater is reasonably large so as to ensure a very low probability of error while decoding, whatever may be the number of repeaters, the signal reaching the destination can be almost error-free. This is because, we make use of '*regenerative repeaters*' which consist of a receiver and a transmitter connected back to back as a package. So, the received signal is decoded to the baseband level at each repeater with practically no errors and this baseband signal is then used by the transmitter part of the repeater to transmit a strong error-free signal at a good power level.
8. PCM, the most popular digital communication, offers an *exponential* rate of exchange between bandwidth and $(\text{SNR})_D$, whereas in FM the $(\text{SNR})_D$ increases approximately only as the square of the bandwidth.
9. **Secrecy of communication:** In defense applications, secrecy and reliability of communication are of utmost importance. Very powerful encryption and decryption algorithms are available for digital data so as to maintain a high level of secrecy of communication. Spread-spectrum techniques of transmission, if used, will further enhance the secrecy of communication and protect it against eavesdropping or jamming by the enemy.

As against all the above stated advantages, digital communications suffer from the following disadvantages:

1. Digital communication systems generally need *more bandwidth* than analog communication systems.
2. *Digital components* generally consume more power as compared to *analog components*.

The advantages, however, far outweigh the disadvantages. That is why digital communication systems are steadily replacing the analog communication systems.

1.7 ELECTROMAGNETIC SPECTRUM, RANGES, AND APPLICATION AREAS

The available electromagnetic spectrum may be conveniently divided into 10 ranges. Depending upon the available propagation modes and their characteristics for each range of frequencies, any given range of frequencies is useful for certain specific types of communication. The ranges, their nomenclature and application areas are summarized in Table 1.1

Table 1.1 *Ranges of spectrum, nomenclature, and application areas*

S.No.	Frequency Range	Name Given	Areas of Application
1.	30 Hz–300 Hz	Extremely Low Frequencies (ELF)	Underwater communications
2.	300 Hz–3.0 kHz	Voice Frequency (VF)	Telephone
3.	3.0 kHz–30 kHz	Very Low Frequencies (VLF)	Navigation
4.	30 kHz–300 kHz	Low Frequency (LF)	Radio navigation
5.	300 kHz–3 MHz	Medium Frequencies (MF)	AM radio broadcasting
6.	3 MHz–30 MHz	High Frequencies (HF)	AM broadcasting, amateur radio, mobile
7.	30 MHz–300 MHz	Very High Frequencies (VHF)	T.V, FM broadcasting, and mobile communications
8.	300 MHz–3 GHz	Ultra High Frequencies (UHF)	T.V., radar, satellite communications
9.	3 GHz–30 GHz	Super High Frequencies (SHF)	Satellite communications, terrestrial microwave communications
10.	10^5 GHz–10^6 GHz	Optical Frequencies (OF)	Optical communications

1.8 CHANNEL TYPES, CHARACTERISTICS, AND MODELING

1. Channel modeling: In the analysis and design of communication systems, it will be necessary to model the channel as a system and incorporate into that model as many details of the electrical behavior of the channel as possible, so as to make it represent the actual situation as accurately as possible, subject to the constraints imposed by consideration of mathematical tractability, etc. Hence, from the point of view of modeling, it may be more convenient and appropriate to classify the channels as linear and non-linear channels, time-invariant and time-varying channels, and bandwidth-limited and power-limited channels, as these characteristics can easily be incorporated into the system used for modeling the channel.

Viewing the commonly used channels in the light of the above discussion, we find that the telephone channel is linear but bandwidth limited (bandwidth limited because, at any given time, it has to be shared by a very large number of users); the satellite channel is power limited, the mobile communication channel is time-varying and that the optical fiber channel is time-invariant. Thus, most of the commonly used communication channels can be generally represented by one of the following three models:

(a) *Additive Gaussian noise channel:* A channel model that is most extensively used is the additive Gaussian channel which portrays the channel as one which, as shown in Fig.1.4, simply attenuates the signal by a factor $\alpha (0 < \alpha < 1)$, and introduces 'additive noise', which itself is modeled as Gaussian

$$r(t) = \alpha s(t) + n(t) \tag{1.1}$$

The model is extremely simple and can be used to represent a large number of physical channels, and hence it is very widely used.

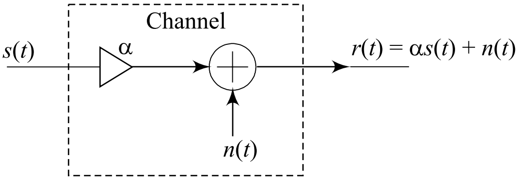


Fig. 1.4 *Additive Gaussian noise channel*

(b) *Bandwidth-limited linear channel:* As pointed out earlier, certain channels like the telephone channel are linear, but bandwidth limited. Such channels may be modeled as shown in Fig.1.5.

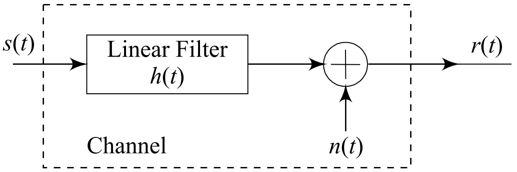


Fig. 1.5 *Bandwidth-limited linear channel*

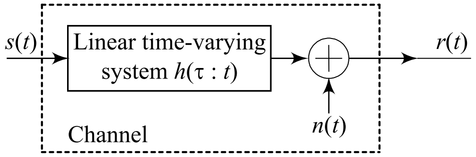


Fig. 1.6 *Linear time-variant channel*

These channels are time-invariant and so the filter shown in Fig.1.5 is an LTI system with an impulse response function, $h(t)$. Thus,

$$\begin{aligned} r(t) &= s(t) * h(t) + n(t) \\ &= \int_{-\infty}^{\infty} s(t-\tau)h(\tau)d\tau + n(t) \end{aligned} \tag{1.2}$$

(c) *Linear time-variant channels:* Channels like the under-water acoustic channels, some mobile communication channels and ionospheric scatter channels, in which the transmitted signal reaches the receiver through more than one path, and where these path lengths are varying with time, have, what is generally termed as 'time-varying' multipath propagation. Such channels are modeled using a time-varying system as shown in Fig. 1.6

In this model, $h(\tau : t)$ is the impulse response function of the time-variant linear system and represents the output at time t of the system which is at rest, when an impulse of unit strength is applied to it as input at time $(t - \tau)$. Thus,

$$r(t) = \int_{-\infty}^{\infty} h(\tau : t)s(t-\tau)d\tau + n(t) \tag{1.3}$$

1.9 NOISE, COMMUNICATION RESOURCES AND PERFORMANCE INDICES

As has been shown in Table 1.1, electromagnetic waves have frequencies over a wide range. This entire range of frequencies over which electromagnetic waves are available, is known as the '*spectrum*'. As indicated in that table, each range or band of frequencies is useful for certain applications. Since a very large number of users will be there for any specific application using a particular band of frequencies, there will be terrible interference caused to all the users and chaotic conditions will prevail, if there is no single controlling authority. Hence, allocation of frequencies for specific users is regulated by an international regulatory authority, which also specifies an *upper limit on the power* that can radiated and the bandwidth that can used at the frequency that is allocated to a user for a specific purpose. Thus, spectrum is a very precious commodity and must be used very sparingly by any system. This implies that communication systems should utilize minimum possible bandwidth while providing a certain assured level of quality of service. That is why, bandwidth is termed as one of the two *communication resources* available to the designer of a communication system. The other resource available to him is power, and this also has to be used to the minimum possible extent while achieving the objectives of a given communication system. All this only means that any system must be so designed that it makes *efficient use* of these two resources. In this connection, it must be noted that sometimes one of these two resources may be more important than the other. For example, in a typical telephone channel, there is a limitation on the bandwidth (limited to 3 kHz) and so it is at a premium, while there is no problem with respect to power. On the other hand, in satellite communication, the power of the transmitter on board the satellite is severely limited while the bandwidth is not.

For assessing the performance of a communication system (analog or digital) and for comparison of various systems, certain criteria are used. The most commonly used criteria are:

(a) Fidelity
(b) Bandwidth efficiency
(c) System complexity

The receiver gives, at its output, an estimate of the transmitted message signal and the criterion, 'fidelity', tells us how accurate that estimate is. *Bandwidth efficiency* refers to the extent to which the transmission bandwidth is utilized efficiently, by the communication system. The criterion, 'system complexity', has a significant bearing on the cost of the system and to estimate the complexity, an engineering judgment is needed that takes into account the level of sophistication of contemporary technology that can be used in the implementation of the communication system under consideration. Performance indices for measuring fidelity and bandwidth efficiency can be appropriately defined depending upon the type of system.

The function of an analog communication system is to make available, at the output of the demodulator, a replica of the baseband signal, i.e., the modulating signal. If there were to be no noise, achieving this objective will be an easy task. That is, perfect communication would be possible even with very little transmitted power. This is because, the received signal, although very weak, is not corrupted by noise and we may amplify it so as to bring the signal power to the desired level. Since our assumption that noise is totally absent makes these amplifiers also noise-free, the amplified message will be an exact replica of the transmitted message signal if the system is so designed that it does not produce any distortion. In practise, however, noise is always present and it corrupts the transmitted signal making the demodulated signal to be not an exact replica of the modulating signal. In fact, the demodulated signal can be shown to be the sum of two components – the desired signal and the noise. As mentioned earlier, 'fidelity' is the performance criterion that assesses the ability of the communication system to produce the desired signal faithfully at the output of the demodulator. Although one can think of various ways of measuring 'fidelity', the most widely used one, especially in cases where the output signal is in the form of the desired signal corrupted by additive noise, is the signal-to-noise ratio (SNR) at the output of the demodulator. It is defined as follows.

$$(SNR) \underline{\Delta} \frac{\text{Average power of signal component at output of the demodulator}}{\text{Average power of noise component at the demodulator output}}$$

As is evident from the above definition, *SNR* shows how strong the signal component is, in comparison with the noise component. SNR is usually expressed in decibels.

$$(SNR)_{\text{dB}} = 10 \log_{10}(SNR) = 10 \log_{10}(\text{Signal power}) - 10 \log_{10}(\text{Noise power})$$

Output *SNR* serves as a good performance index for analog communication systems in which faithful reproduction of the analog baseband modulating signal in the reconstructed waveform at the output of the demodulator is of paramount importance. Thus, other things like the transmitted power, transmission bandwidth and channel characteristics remaining the same, the system giving a higher output *SNR* is generally preferred over a system with a lower output *SNR*. This is because, for any specified output *SNR*, the former needs less transmitted power than the latter and so is able to utilize the communication resource, power, more efficiently. Generally, the ratio of the $(SNR)_d$ with carrier modulation to the $(SNR)_b$, the *SNR* with baseband transmission (i.e., direct transmission of the baseband signal without any modulation) is used as the performance index and is called the Figure of Merit (FOM).

The other performance index, bandwidth efficiency, indicates how well the system is utilizing the communication resource, bandwidth, and is defined as

$$\left.\begin{array}{c}\text{Bandwidth}\\ \text{Efficiency}\end{array}\right\} \underline{\Delta} = \frac{W}{B_T} = \frac{\text{Bandwidth of the baseband signal}}{\text{Bandwidth of the transmitted signal}}$$

(for analog communication

Obviously, the bandwidth efficiency of an analog communication system depends on the type of modulation employed. Of all the analog modulation schemes, Wideband FM (WBFM) has the least bandwidth efficiency while the single sideband modulation scheme has the highest (100%) bandwidth efficiency.

There is a fundamental difference between the demodulator in the receiver of an analog communication system and the source decoders in the receiver of a digital communication system insofar as the design objectives are concerned. The demodulator in an analog system is expected to faithfully extract the waveform of the baseband signal from the received signal. But, in the case of a digital communication system, extracting the baseband signal wave faithfully is not what is attempted, as it is not important. Instead, the decoder is expected to make a correct decision during each time-slot as to whether what was transmitted during that time slot was a binary 1 or a binary 0, in the case of binary transmission, and as to which one of the M possible symbols has been transmitted, if it were M-ary transmission. If the decoder is able to take correct decision in each time slot, the exact baseband message sequence can be generated in the receiver itself.

From the foregoing, it is evident that the performance index pertaining to the fidelity criterion has got to be different in the two cases (Analog and Digital). Clearly, a more useful and relevant performance index for fidelity in the case of digital communication systems is the Bit Error Rate (BER), or the Probability of Error, P_e.

The objective of a digital communication system is to transmit information at as fast a rate as possible while maintaining the probability of error within the limits admissible for the specific application for which the system has been designed. Further, it must do this using minimum possible bandwidth and power by maintaining a high bandwidth efficiency and power efficiency. Since a high rate of transmission requires a large bandwidth, bandwidth efficiency is defined as the bit-rate per unit bandwidth.

$$\therefore \quad \text{Bandwidth efficiency} \triangleq \frac{R_b}{B_T}$$

(For digital communication)

where R_b is the bit rate in bits per second and B_T is the transmission bandwidth utilized in Hertz. It must be noted that the transmission bandwidth, B_T, has to be appropriately defined, keeping the shape of the power spectrum of the digital signal being transmitted. Since probability of error *basically* depends upon the signal-to-noise ratio (although it can be reduced by appropriate channel coding too), for a white noise channel, a reasonable index of power efficiency may be taken to be the value of the ratio of E_b to N_0 required to achieve a specified probability of error at a specified bit rate R_b, where E_b is the bit energy and N_0 is the one-sided power spectral density of the white noise on the channel. Among other things, the type of modulation used does influence the bandwidth efficiency as well as the power efficiency. The source coding, by reducing the redundancy and therefore the number of bits required to represent the information given by the source, does help in improving the bandwidth efficiency of the system for any type of modulation employed. Similarly, channel coding, by giving the transmitted bit-stream the power to correct the errors caused in the data by the channel noise, helps to improve the power efficiency for any modulation that might be used, by reducing the (E_b/N_0) required for achieving a specified probability of error at a given rate of transmission.

Prior to the publication of Shannon's famous channel capacity theorem, it was considered impossible to have an absolutely error-free transmission over a noisy channel. In this theorem, Shannon has combined the effects of bandwidth and signal-to-noise ratio and related them to a quantity 'C', which he called as the channel capacity and which represents the maximum rate of transmission of information supported by a channel. For a channel of bandwidth B Hertz with additive noise spectral density $N_0/2$ (two-sided), he has shown that

$$C = B \log_2\left(1 + \frac{P}{N_0 B}\right) \text{bits/sec.}$$

where P is the average transmitted power. He has shown, contrary to the widely held opinion at that time, that it is possible to have *error-free transmission* over such a noisy channel as long as the rate of transmission R_b, is less than the channel capacity C. This theorem of Shannon implies that by employing appropriate type of channel coding, it is possible to transmit data at the rate of C bits/sec *without any error*. However, till date, no code has been invented that could achieve this. Hence, Shannon's channel capacity theorem sets a *fundamental limit* on the rate of error-free transmission. This channel capacity equation of Shannon further points to the possibility of a 'Bandwidth-SNR' trade-off. The equation says that one can achieve a certain value of C with a large B and a small P or vice versa.

References

1. Berrou, C., A. Glavieux, and P. Thitimajshima, Near Shannon-limit Error-correcting Coding and Decoding: Turbo Codes, *Proc. International conference on communication*, Geneva, Switzerland, pp. 1064–1070, May, 1993.
2. Berrou, C., A. Glavieux, and P. Thitimajshima, The Ten-year Old Turbo Codes are Entering Into Service, *IEEE Communication Magazine*, Vol. 41, No.8, pp. 110–116, 2003.
3. Bhattacharya, Amitabh, *Digital Communication*, Tata McGraw Hill, New Delhi, 2006.
4. Bose, R.C., and D.K. Ray Choudhary, On a Class of Error-correcting Binary Group Codes, *Information and Control*, Vol. 3, pp. 68–79, 1960.
5. Forney, G.D. Jr., Convolutional Codes-I: Algebraic Structures, *IEEE Transactions on Information Theory*, Vol.16, pp 720–738, 1970.
6. Golay, M.J.E., Notes on Digital Coding, *Proc. IRE*, pp. 657, June, 1949.
7. Hamming, R.W., Error-detecting and Error-corresponding Codes, *Bell System Technical Journal*, Vol. 29, pp. 147–160, 1950.
8. Haykin, Simon, *Digital Communications*, pp. 7, John Wiley & Sons, 2002.
9. Haykin, Simon, *Communication Systems*, John Wiley & Sons, 3rd Edition, 1995.
10. Lucky, R.W., Automatic Equalization for Digital Communications, *Bell System Technical Journal*, Vol. 44, pp. 547–588, 1965.
11. MacDonald, V.H., The Cellular Concept, *Bell System Technical Journal*, Vol. 58, No.1, pp. 15–43, 1979.
12. McMohan, M.A., The Making of a Profession – A Century of Electrical Engineering in America, *IEEE Press*, New York, 1984.
13. North, D.O., An Analysis of the Factors which Determine the Signal/Noise Discrimination in Pulsed Carrier Systems, *Proceedings of IEEE*, Vol. 51, pp. 1016–1027, 1963.
14. Proakis, John G., *Digital Communications*, 4th edition, McGraw-Hill International, 2001.
15. Reeves, A.H., That Past, Present and Future of PCM, *IEEE spectrum*, Vol. 15, No.5, pp. 58–63, 1975.
16. Shannon, C.E., A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol. 27, pp. 379–423, July, 1948.
17. Ungerboek, G., Channel Coding with Multilevel/Phase Signals, *IEEE Transactions on Information Theory*, Vol. 28, pp. 55–67, 1982.
18. Ungerboek, A Mathematical Theory of Communication, *Bell System Technical Journal*, Vol.27, pp. 623–656, Oct., 1948.
19. Ungerboek, Communication in the Presence of Noise, *Proc. of IRE*, Vol. 37, pp. 10–21, Jan. 1949.

SIGNALS, SIGNAL SPACES AND FOURIER SERIES

2

> *"It is not ignorance but knowledge which is the mother of wonder."*
>
> **Joseph Wood Krutch (1893–1970)**
> *American writer, critic, and naturalist*

Learning Objectives

After going through this chapter, students will be able to

- *determine the class to which a given signal belongs,*
- *mathematically and graphically represent some commonly used signals,*
- *determine the effect of certain time-domain operations like shifting, compressing and expanding of a given signal,*
- *understand the term, 'Signal space', visualize its geometrical structure, and be able to test and check whether a given set of signals forms an orthogonal set,*
- *derive a set of orthogonal signals from a given set of linearly independent signals, using Gram–Schmidt procedure,*
- *expand a given periodic signal in terms of its complex exponential/trigonometric Fourier series and sketch its magnitude and phase spectra, and*
- *find the auto-correlation of a given signal, or the cross-correlation and convolution of two given signals.*

2.1 INTRODUCTION

Communication, in general, involves transfer or transmission of a message/information from a source to a destination. This message may take a variety of forms – it may be an acoustic signal as in the case of speech, or may be a spatial distribution of brightness, as in the case of a still monochrome picture. Whatever may be its original form, it is converted into an electrical signal (a variation of electrical voltage with respect to time) by the use of appropriate instrumentation – a microphone in the case of the speech signal and a video camera in the case of the picture. We shall therefore assume that our signals are all electrical signals and that they are single-valued functions of time. Again, these signals may be either deterministic, or random. We shall consider random signals later; for now, we shall confine our attention to only deterministic signals.

Joseph Fourier (1768–1830) developed the theory for the study of signals in terms of their sinusoidal representation. Named after him as Fourier analysis, it is extensively used in various branches of science and engineering, as it offers an insight into the frequency content of a signal. Deterministic signals may be classified as periodic signals and aperiodic or non-periodic signals. Fourier series expansion provides information on the spectral content of a periodic signal while the Fourier transform provides this information in the case of a non-periodic signals. Fourier transform will be discussed in detail in the next chapter but Fourier series will be discussed in this chapter. This chapter will show that the Fourier series expansion of a periodic signal is just an orthogonal expansion of the signal obtained by taking certain specific complete sets of orthonormal signals as the basis sets. For this purpose, the concept of a signal space making use of the analogy between vectors and signals will be developed. This will enable us to talk of the norm of a signal, the angle between two signals, orthogonality of signals, component of one signal along another, and so on, paving the way for a discussion on Gram–Schmidt orthogonalization procedure. Once all these basic building blocks are in place, the concept of a complete orthonormal set is introduced. It is then shown that the complex- exponential Fourier series expansion of a periodic signal results when an orthogonal expansion of the signal is obtained using the set of complex-exponential signals as the basis set for the pertinent signal space. A useful feature of this approach to the introduction of Fourier series is the fact that in the process, the reader becomes familiar with Gram–Schmidt orthogonalization procedure as well as Schwarz's inequality, both of which are extremely useful in the study of communication engineering.

2.2 SIGNALS

All of us certainly have an intuitive idea of what a signal is, since signals play such an important role in our daily lives. When we speak, an acoustic signal, called speech, emanates and it is a function of the single independent variable, *time*. Similarly, when we look at a monochrome still image, the signal that we get from it is a variation of brightness or light intensity, I, from point to point. In other words, we have a signal here which is a function of two independent variables, i.e., x and y coordinates, since I is a function of x and y.

Thus, we may generalize the above discussion and say that a signal is a single-valued function of one or more independent variables and carries some information.

2.2.1 Types of Signals

Continuous-time signals and discrete-time signals *A signal is said to be a continuous-time signal, if its value is defined at all instants of time.* Here, we must realize that this definition has nothing to do with mathematical continuity of the waveform of a signal. Thus, even a rectangular waveform, which has discontinuities at regular intervals, is also a continuous-time signal, if the value of the waveform is defined even at all the discontinuities.

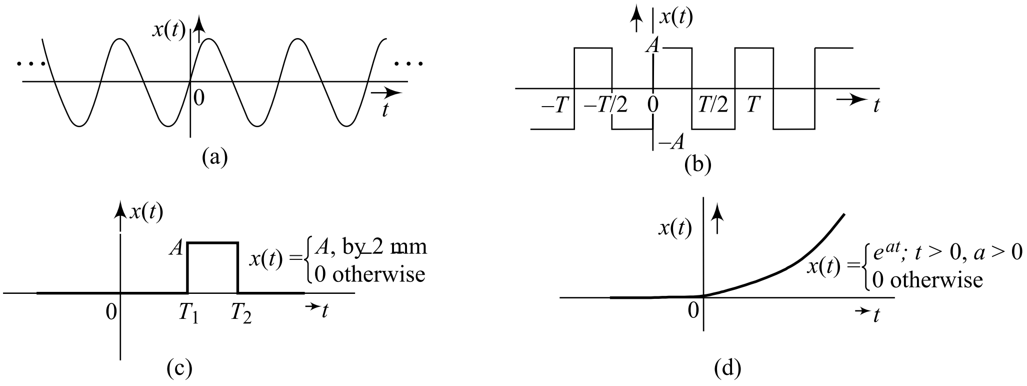


Fig. 2.1 *Examples of continuous-time signals*

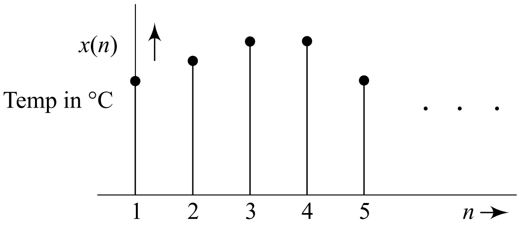


Fig. 2.2 *A discrete-time signal*

Discrete-time signals, on the other hand, are defined only at a discrete set of points in time. For example, if we record the temperature at a particular place every day at say 5 a.m., the data so recorded, represents a discrete-time signal, which is shown in Fig. 2.2.

It should be noted that the temperature between two successive values of n is not zero; *it is not known. Here, the parameter representing time, namely n, takes only integer values, i.e., time is discretized.* It should, however, be noted that the amplitude is not discretized and it can take a continuum of values.

Periodic and non-periodic signals *A continuous-time signal*, $x(t)$, *is said to be periodic in time if*

$$x(t) = x(t + mT) \tag{2.1}$$

for any t and any integer m. The smallest positive value of the constant T, satisfying the above relation, is called the fundamental period of the periodic signal $x(t)$.

Any continuous-time signal not satisfying Eq. (2.1) is said to be non-periodic.

Energy and power signals Let $x(t)$ be a current signal. Let this current be flowing through a resistance of R ohms. Then the instantaneous power delivered by the signal is $x^2(t)\cdot R$. If $x(t)$ is a voltage signal, the instantaneous power delivered is given by $x^2(t)/R$ watts. If we make the value of R equal to 1 ohm, irrespective of whether $x(t)$ is a voltage signal or current signal, the instantaneous power is simply $x^2(t)$ and this depends only on the signal. Hence, we define the power (instantaneous) associated with a signal $x(t)$ as simply $x^2(t)$. In case $x(t)$ is not purely real then the instantaneous power is represented by $|x(t)|^2$.

Thus, the total energy of a continuous-time signal $x(t)$ whether real valued, or complex valued is given by

$$E = \operatorname*{Lt}_{T\to\infty} \int_{-T}^{T} |x(t)|^2\, dt \tag{2.2}$$

Similarly the Average Power of $x(t)$ is given by

$$P_{\text{av}} = \operatorname*{Lt}_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2\, dt \tag{2.3}$$

Definitions

(i) *A signal whose total energy is finite and non-zero, is called an energy signal. If E is the energy,* $0 < E < \infty$.

(ii) *A signal whose average power is finite and non-zero is called a power signal. If* P_{av} *is the average power,* $0 < P_{\text{av}} < \infty$.

Note

(i) Obviously, since the averaging is done over the infinite time interval $-\infty < t < \infty$, *and the energy of the signal is finite,* ***an energy signal will have an average power which is zero****.*

(ii) Since the average power is finite and the averaging is done over the infinite time interval $-\infty < t < \infty$, ***a power signal will have infinite energy****.*

(iii) Signals with a 'finite' or 'asymptotically finite' duration in time, such as the ones given below, are energy signals.

(a) $x(t) = \begin{cases} A; & |t| \le T \\ 0; & \textit{otherwise} \end{cases}$

(b) $x(t) = Ae^{-a|t|};\ a > 0$

(iv) In general, all periodic signals are power signals. (But every power signal need not be a periodic signal)

(v) Every signal need not be either an energy signal or a power signal. A signal may be neither a power signal nor an energy signal.
Example: $x(t) = 5e^{-t}; -\infty < t < \infty$

Example 2.1 Check whether $x(t) = \begin{cases} Ae^{-t} & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}$ is an energy signal or a power signal

$$E = \int_0^\infty A^2 e^{-2t}\,dt = A^2 \int_0^\infty e^{-2t}\,dt = \frac{A^2}{-2} e^{-2t}\Big|_0^\infty = \frac{A^2}{2}$$

which is finite. Hence, $x(t)$ is an energy signal.

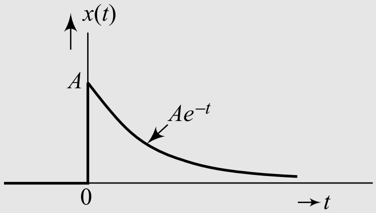


Fig. 2.3 *$x(t) = Ae^{-t}$*

Example 2.2 Is $x(t) = \cos 2\pi f_0 t$ an energy signal, or a power signal?

Solution The average power, P_{av} for this signal is given by

$$P_{\text{av}} = \underset{T\to\infty}{\text{Lt}} \frac{1}{2T} \int_{-T}^{T} x^2(t)dt$$

$$\underset{T\to\infty}{\text{Lt}} \frac{1}{2T} \int_{-T}^{T} \cos^2(2\pi f_0 t)dt = \underset{T\to\infty}{\text{Lt}} \frac{1}{4T} \int_{-T}^{T} (1 + \cos 4\pi f_0 t)dt$$

$$= \underset{T\to\infty}{\text{Lt}} \frac{1}{4T} \int_{-T}^{T} dt + \underset{T\to\infty}{\text{Lt}} \frac{1}{4T} \int_{-T}^{T} \cos 4\pi f_0 t\,dt$$

But the second integral is zero and so

$$P_{\text{av}} = \underset{T\to\infty}{\text{Lt}} \left[\frac{1}{4T} \times 2T\right] = \frac{1}{2}$$

Deterministic signals and random signals A signal whose value at any instant of time, $-\infty < t < \infty$, is known *apriori*, is called a deterministic signal. For example, $x(t) = 10 \cos 200\,\pi t$ is a deterministic signal since its value at any instant of time $-\infty < t < \infty$, can be determined.

As against this, there are some signals, which are random in nature, i.e., their values cannot be determined or predicted. Noise signals are examples of such signals. Such signals will be discussed in detail in Chapter 6.

Unit-step and unit-impulse functions

1. Unit-step function: This is denoted by $u(t)$ and is defined by the following:

$$u(t) \underline{\Delta} \begin{cases} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \quad (2.4)$$

$u(t)$ is diagrammatically represented as shown in Fig. 2.4. It is obvious that any signal $x(t)$, when multiplied by $u(t)$, retains without any change, only that part of $x(t)$ pertaining to non-negative values of time and the portion of the signal $x(t)$ corresponding to negative values of time, is reduced to zero.

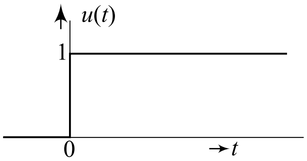


Fig. 2.4 *A unit-step function*

2. Unit-impulse function $\delta(t)$: This is not a function in the usual sense. In fact, it comes under the category of 'generalized functions', or 'distributions', and is defined by the following:

$$\int_{t_1}^{t_2} x(t)\delta(t)dt = \begin{cases} x(0) = x(t)\big|_{t=0} & \text{if } t_1 < 0 < t_2 \\ 0; & \text{otherwise} \end{cases} \tag{2.5}$$

where $x(t)$ is any function which is continuous atleast at $t = 0$.

Using the above definition, we can derive a number of important properties of the unit impulse function.

Property 1: The area under a unit impulse function is equal to one.

Proof Let $x(t) = 1$, this function is continuous at all points including $t = 0$.
Let $t_1 = -\infty$ and $t_2 = +\infty$. Then

$$\int_{-\infty}^{+\infty} 1 \cdot \delta(t)dt = 1 \Rightarrow \text{the area under } \delta(t) = 1.$$

Hence, the area under the unit impulse function is equal to one.

Property 2: The width of $\delta(t)$ along the time axis is zero.

Proof $\int_{0-\in}^{0+\in} 1 \cdot \delta(t)dt = 1$. Now let $\in \to 0$. However small $\in$ may become, since the range of integration, $-\in < t < \in$ includes $t = 0$, the area under the unit-impulse function still continues to be unity'. Hence, $\delta(t)$ has zero width along the time axis, around $t = 0$.

We may thus visualize $\delta(t)$ as being located at $t = 0$, having an area of 1 under it and occupying zero width along the time axis. Because of this, in diagrams, it is generally represented as shown in Fig. 2.5. The 1 marked at the arrowhead indicates that it is a unit impulse and that it has strength (area) of one.

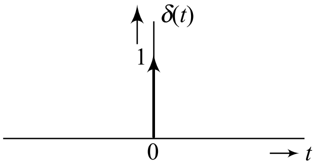


Fig. 2.5 *A unit impulse*

Since $\delta(t)$ represents a unit impulse occurring at $t = 0$, as per the usual notation, we represent a unit impulse located at $t = \tau$ by $\delta(t - \tau)$.

Property 3 (Sampling Property): From Eq. (2.5), we may now say that if a function $x(t)$ is continuous at $t = \tau$, then

$$\int_{t_1}^{t_2} x(t)\delta(t-\tau)dt = x(t)\big|_{t=\tau} = x(\tau)$$

for any t_1 and t_2 such that the interval t_1 to t_2 includes $t = \tau$.
But

$$\int_{t_1}^{t_2} x(\tau)\delta(t-\tau)dt = x(\tau)\int_{t_1}^{t_2} \delta(t-\tau)dt = x(\tau)$$

Therefore,
$$\int_{t_1}^{t_2} x(t)\delta(t-\tau)dt = \int_{t_1}^{t_2} x(\tau)\delta(t-\tau)dt,$$

for *any* $x(t)$ which is continuous at $t = \tau$ and for *any* t_1 and t_2, provided their interval includes $t = \tau$. Thus, we conclude

$$x(t)\delta(t-\tau) = x(\tau)\delta(t-\tau) \tag{2.6}$$

From Fig. 2.6 it is clear that when $\delta(t-\tau)$ multiplies the function $x(t)$ which is continuous at $t = \tau$, it just takes the sample of $x(t)$ at $t = \tau$ where the impulse is located, and produces an impulse of strength $x(\tau)$ located at $t = \tau$.

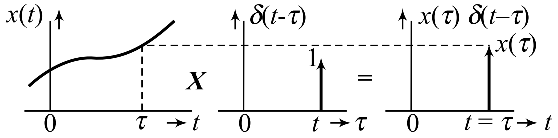


Fig. 2.6 *A diagrammatic representation of Eq. (2.6)*

For this reason, the above property represented by Eq. (2.6) is called the '*sampling property*' of an impulse function.

Property 4: This property, called the 'replication property' of an impulse function, states that if a function $x(t)$ is convolved with $\delta(t-\tau)$, a unit impulse located at $t = \tau$, then the function $x(t)$ gets shifted by τ sec and we get $x(t-\tau)$.

This is discussed in more detail and proved in Section 3.3 under properties of convolution.

Because of properties 1 and 2 above, the unit impulse function, $\delta(t)$, is usually visualized as the limiting case of a rectangular pulse $x_\Delta(t)$ of amplitude $1/\Delta$ and time duration Δ when the parameter Δ is allowed to tend to zero, as shown in Fig. 2.7. Note that the area under the rectangle remains equal to 1 even while $\Delta \to 0$.

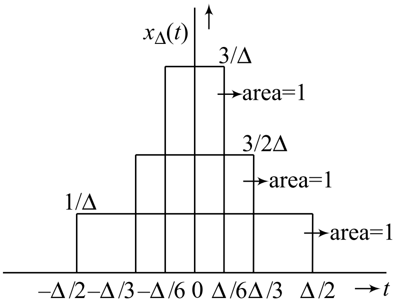


Fig. 2.7 *Unit impulse as limiting case of a rectangular pulse*

Relation between $u(t)$ and $\delta(t)$ There is an interesting and useful relationship between the unit-impulse function and the unit-step function. Consider

$$x(t) = \int_{-\infty}^{t} \delta(\lambda) d\lambda$$

Since the right-hand side of the above represents the area under the unit impulse function from $-\infty$ up to time t, if $t < 0$, the area will be zero. But if $t \ge 0$, the area is equal 1. Hence,

$$x(t) = \begin{cases} 1 & \text{for } t \ge 0 \\ 0 & \text{for } t < 0 \end{cases}$$

But this is precisely how we have defined $u(t)$.

$$\therefore \qquad u(t) = \int_{-\infty}^{t} \delta(\lambda) d\lambda \tag{2.7}$$

and

$$\frac{d}{dt} u(t) = \delta(t) \tag{2.8}$$

Some simple operations on signals Continuous-time signals may be subjected to several types of operations. These include addition and subtraction of signals, multiplication of signal by a constant, multiplication of two signals, convolution of two signals (discussed in Section 2.5), differentiation and integration of signal, shifting in time, and compressing/expanding a signal in time. Here, we shall briefly discuss only the last two – shifting in time and compression/expansion in time.

1. Shifting in time: Consider a continuous-time signal $x(t)$. Now consider the signal $x(t-t_0)$. At $t = t_1$, the function $x(t)$ takes the value $x(t_1)$. The function $x(t-t_0)$ too takes that value $x(t_1)$ when its argument takes the value t_1, i.e., when $t - t_0 = t_1$, or when $t = t_0 + t_1$. Thus, whatever happens to the signal $x(t)$ at $t = t_1$ happens to the signal $x(t-t_0)$ only at $t = t_0 + t_1$, i.e., after a delay of t_0 sec (if $t_0 > 0$).

Thus, if $t_0 > 0$, $x(t-t_0)$ *is a time-delayed version of* $x(t)$ *and* $x(t+t_0)$ *is a time advanced version of* $x(t)$.

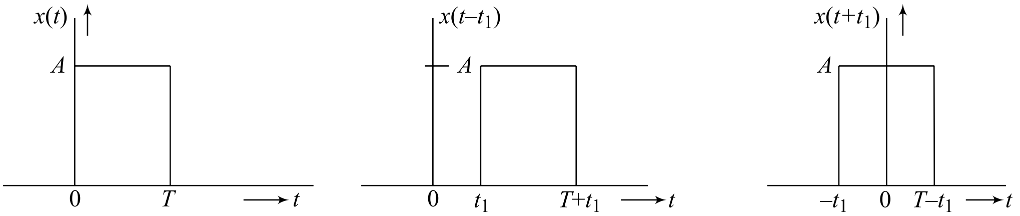


Fig. 2.8 *(a) x(t), (b) For $t_1 > 0$, $x(t - t_1)$ is obtained by shifting x(t) to the right by t_1 sec, (c) $x(t + t_1)$ is obtained by shifting x(t) to the left by t_1 sec, if $t_1 > 0$*

2. Compressing/expanding a signal in time (time scaling): Advancing the same arguments as above, it can be shown that if $x(t)$ is a continuous-time signal, then $x(at)$ represents a time-compressed version of $x(t)$ if 'a' is a positive number greater than 1 and a time-expanded version of $x(t)$, if 'a' is a positive number less than 1.

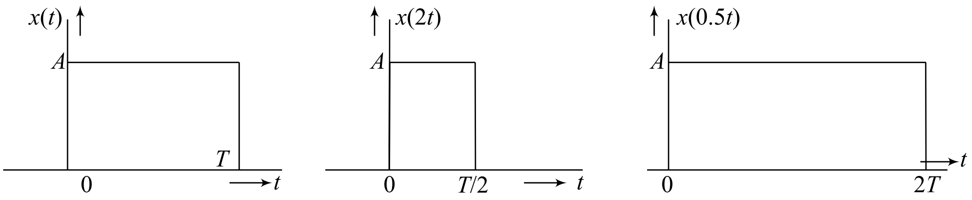


Fig. 2.9 *(a) x(t), (b) x(2t), a compressed version of x(t), (c) x(0.5t), an expanded version of x(t)*

Quite often, we will be performing time-shifting as well as time scaling of a signal. For example, consider $x(t)$ and $x(2t - 3)$. Then, to obtain $x(2t - 3)$ from $x(t)$, we should note that we have to do time shifting first *and then only* do the time scaling. This is because

$$x(t)|_{t\to(t-3)} = x(t-3) \quad \text{and} \quad x(t-3)|_{t\to 2t} = x(2t-3)$$

whereas

$$x(t)|_{t\to 2t} = x(2t) \quad \text{and} \quad x(2t)|_{t\to(t-3)} = x(2t-6) \neq x(2t-3)$$

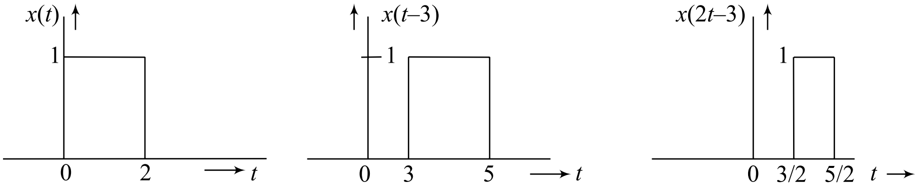


Fig. 2.10 *(a) x(t), (b) x(t – 3), (c) x(2t – 3)*

2.3 ANALOGY BETWEEN SIGNALS AND VECTORS

In this section, we will show that various concepts familiar to us in connection with vectors, can be extended to signals. Since the dot product of vectors gives a geometrical structure to the vector spaces by enabling us to talk about 'angle between vectors', 'orthogonality of vectors' and so on, we will define 'inner product' between signals (which is similar to the dot product between vectors) so as to enable us to talk about the 'distance between signals', 'orthogonality of signals', etc., as all these are essential for developing a geometrical structure for signal spaces. For this purpose, we will begin with a brief review of some of the basic and essential concepts in vector spaces in a manner that will make it easier for us to clearly bring out the analogy between vectors and signals.

As we know, a vector space **V**, like for example the 'Euclidian Space' is a set of vectors, whose elements, the vectors, satisfy certain conditions. The most important ones among these conditions are given below:

(a) The sum of any two vectors belonging to **V** is another vector which also belongs to **V**.
(b) There exists in **V**, a vector, called the zero vector, denoted by **O**), which is such that **A** + **O** = **A**, for any vector **A** belonging to **V**.
(c) For any vector **A** belonging to **V**, there exists another vector **B**, also belonging to **V** and such that **A** + **B** = the zero vector **O**.
(d) If any vector, say **A** belonging to **V** is multiplied by a scalar (a real number or complex number), the resultant vector also belongs to **V**.

We know that a vector **A** in the Euclidian space, such as the one shown in Fig. 2.11, can be expressed as:

$$\mathbf{A} = \boldsymbol{i}x_1 + \boldsymbol{j}x_2 + \boldsymbol{k}x_3 \tag{2.9}$$

Here, ***i***,***j*** and ***k*** are unit vectors along the *X*, *Y* and *Z* directions respectively. x_1, x_2 and x_3 are real numbers representing what we call as the coordinates of the vector **A** along the *X*, *Y* and *Z* directions.

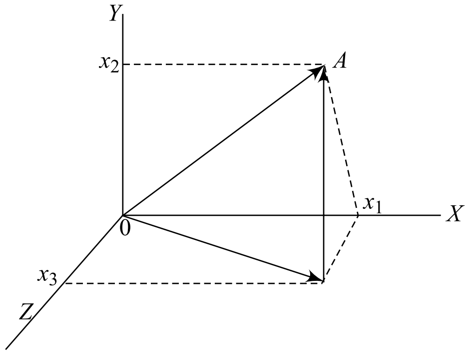


Fig. 2.11 *A vector* **A** *in the Euclidian space*

Note *A unit vector along the direction of some vector OB is obtained by dividing OB by its own magnitude.*

It is customary to represent the vector **A** as

$$\mathbf{A} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{2.10}$$

in terms of its coordinates along the *X*, *Y* and *Z* directions. Thus, the vector **A** may be assumed to be represented by the point *A* in the Euclidian space, since point *A* has x_1, x_2 and x_3 as its coordinates.

Using the representation of a vector by Eq. (2.9), we may represent the unit vectors ***i***,***j*** and ***k*** as

$$\boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.11}$$

Therefore, we may write the vector *A* of Eq. (2.1) as

$$\mathbf{A} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.12}$$

Equations (2.9) and (2.12), express **A** as a *linear combination* of the three unit vectors, with coordinates of **A** along *X*, *Y* and *Z*, viz., x_1, x_2 and x_3 as the weights for the linear combination.

2.3.1 Linearly Independent Vectors

The vectors ***i***,***j*** and ***k*** are said to be *linearly independent* in the sense that *none of them* can be expressed as a linear combination of the rest.

Example 2.3 Show that the following three vectors belonging to the Euclidian space are linearly independent.

$$\mathbf{X} = (1, 6, 5);\ \mathbf{Y} = (1, 1, 0),\ \text{and}\ \mathbf{Z} = (7, 5, 2)$$

Solution As we have already seen, if **X**, **Y** and **Z** are linearly independent, none of them should be a linear combination of the rest. In case they are not linearly independent, one of them at least, will be a linear combination of the other two. In that case, the determinant of the matrix formed by writing these vectors as the rows of a 3×3 matrix, will be zero. So, we shall check whether the matrix $[A]$ where

$$[A] = \begin{bmatrix} 1 & 6 & 5 \\ 1 & 1 & 0 \\ 7 & 5 & 2 \end{bmatrix}$$

is having a determinant of zero or not.

$$|A| = \begin{vmatrix} 1 & 6 & 5 \\ 1 & 1 & 0 \\ 7 & 5 & 2 \end{vmatrix} = 1(2-0) - 1(12-25) + 7(-5) = 2 + 13 - 35 = -20$$

Hence, $|A| \neq 0$ and so the given vectors are linearly independent.

Further, as can be easily seen, *any arbitrary vector*, say **OB**, or **B,** which belongs to the Euclidian space, can be expressed as a linear combination of the three unit vectors $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$. For example, if **B** has coordinates y_1, y_2 and y_3, then

$$\mathbf{B} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = y_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + y_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Since *any vector* belonging to this space can be generated as a linear combination of $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$, we say that the space (the Euclidian space) is *generated* by, or is *spanned by* these three vectors.

2.3.2 Basis Vectors

A set of vectors belonging to a vector space V(like the Euclidian space) which are linearly independent and which span (or generate) that space, are said to be forming a *basis set* for that space.

Thus, $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ form a basis set of vectors for the Euclidian space since they are linearly independent and also span the Euclidian space in the sense that *any vector* belonging to that space can be expressed as a linear combination of these three vectors.

Note *A basis set for a vector space is not unique, since there can be any number of basis sets (For example, rotation of the X, Y, Z axes creates a new set of basis vectors).*

Orthogonality of vectors The vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are not only linearly independent, they are also mutually orthogonal.

Definition Two vectors are said to be orthogonal to each other if their dot product is zero. A set of vectors are said to be an orthogonal set of vectors if any two *distinct* vectors from the set are orthogonal to each other. If, *in addition*, each one of the vectors of the set has a unit magnitude, the set is said to be *an orthonormal set*.

It is an easy matter to show that $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ form an orthonormal basis set of vectors for the Euclidian space.

Dimension The number of vectors in *any basis* set for a given space will be the same and this number is called the *dimension of that space*. It represents the minimum number of linearly independent vectors required to generate that vector space.

We may generalize the above concepts and visualize an n-dimensional vector space ($n > 3$) in which each vector will be a point in an n-dimensional space and can be represented by a column vector having the n-coordinates of that point as the entries. Thus, if α is a vector in an n-dimensional vector space,

$$\boldsymbol{\alpha} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} + \ldots + x_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}$$

where, $[1, 0, 0, \ldots, 0]^T$, $[0, 1, 0, \ldots, 0]^T$, …. $[0, 0, 0, \ldots, 1]^T$ are the n basis vectors (they are called *standard set of basis vectors*) for this n-dimensional space.

Example 2.4 Given two vectors, $\mathbf{X} = (0, 2, 1)$ and $\mathbf{Y} = (1, -2, 1)$, find the magnitude of each and check whether they are orthogonal. If they are not orthogonal, find the angle between them.

Solution Since $\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos \phi$, $\mathbf{A} \cdot \mathbf{A} = |\mathbf{A}|^2$

$$\therefore \quad |\mathbf{A}| = (\mathbf{A} \cdot \mathbf{A})^{1/2} \quad \therefore |\mathbf{X}| = (0 \times 0 + 2 \times 2 + 1 \times 1)^{1/2} = \sqrt{5}$$

$$|\mathbf{Y}| = [1 \times 1 + -2 \times -2 + 1 \times 1]^{1/2} = \sqrt{6}$$

$\mathbf{X} \cdot \mathbf{Y} = [0 \times 1 + 2 \times -2 + 1 \times 1] = -3$. Since $\mathbf{X} \cdot \mathbf{Y} \neq 0$, $\mathbf{X}$ and $\mathbf{Y}$ are not orthogonal.

$$\cos \phi = \frac{\boldsymbol{A} \cdot \boldsymbol{B}}{|\boldsymbol{A}||\boldsymbol{B}|}$$ Hence if θ is the angle between $\boldsymbol{X}$ and $\boldsymbol{Y}$ then

$$\theta = \cos^{-1}\left[\frac{\boldsymbol{X} \cdot \boldsymbol{Y}}{|\boldsymbol{X}||\boldsymbol{Y}|}\right] = \cos^{-1}\left[\frac{-3}{\sqrt{5} \times \sqrt{6}}\right]$$

or

$$\theta = \cos^{-1}\left[\frac{-3}{\sqrt{30}}\right] = \cos^{-1}\left[-\sqrt{\frac{3}{10}}\right]$$

2.3.3 Component of a Vector along Another Vector

Consider n non-zero orthogonal vectors, $\alpha_1, \alpha_2, \ldots \alpha_n$. consider the following linear combination of these n vectors:

$$\boldsymbol{\beta} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + x_3\boldsymbol{\alpha}_3 + \ldots + x_k\boldsymbol{\alpha}_k + \ldots x_n\boldsymbol{\alpha}_n$$

where $x_1, x_2, \ldots, x_n$ are some real numbers and β is the resultant vector.

Taking the dot product of both sides with the vector α_k, we get

$$\boldsymbol{\beta} \cdot \boldsymbol{\alpha}_k = x_1(\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_k) + x_2(\boldsymbol{\alpha}_2 \cdot \boldsymbol{\alpha}_k) + x_3(\boldsymbol{\alpha}_3 \cdot \boldsymbol{\alpha}_k) + \ldots + x_k(\boldsymbol{\alpha}_k \cdot \boldsymbol{\alpha}_k) + \ldots x_n(\boldsymbol{\alpha}_n \cdot \boldsymbol{\alpha}_k)$$

Since α_i, $i = 1, 2, 3, \ldots, n$ are orthogonal vectors, all the products on the RHS are zero, except $(\alpha_k \cdot \alpha_k)$, which we know, is equal to $|\alpha_k|^2$.

$$\therefore \quad \beta \cdot \alpha_k = x_k|\alpha_k|^2$$

Hence, $$x_k = \text{coordinate of } \beta \text{ along } \alpha_k = \frac{\boldsymbol{\beta} \cdot \boldsymbol{\alpha}_k}{|\alpha_k|^2} \tag{2.13}$$

and the component of β along $$\alpha_k = x_k \boldsymbol{\alpha}_k = \left(\frac{\boldsymbol{\beta} \cdot \boldsymbol{\alpha}_k}{|\alpha_k|^2} \right) \boldsymbol{\alpha}_k; \quad 1 \le k \le n \tag{2.14}$$

Example 2.5 For the two vectors **X** and **Y** of Example 2.4, find the component of **Y** along **X**.

Solution (**X** · **Y**) = –3 and |**X**| = $\sqrt{5}$, as obtained in the solution for Example 2.4.

$$\therefore \quad \text{component } \mathbf{Y} \text{ along } \mathbf{X} = \left(\frac{\boldsymbol{Y} \cdot \boldsymbol{X}}{|\boldsymbol{X}|^2} \right) \boldsymbol{X} = \left(\frac{-3}{5} \right)(0, 2, 1) = (0, -1.2, -0.6)$$

2.3.4 Signal Spaces

In digital communications, in general, one of a set of M ($M \ge 2$) possible signals, $s_i(t)$, $i = 1, 2, 3, \dots, M$, is transmitted every T sec. These M signals are known *a priori* to the receiver. What the receiver does not know, however, is, which one of the M signals has been transmitted during a given T sec period. The job of the receiver is then to correctly identify, during each T sec period, the transmitted signal, in the presence of noise. These M signals are continuous-time real-valued signals having a finite energy over a T sec period, i.e., the time period for which one of the M signals is transmitted.

Let us now consider the set S of all possible continuous-time signals having a finite energy over a period of T sec. These signals can easily be shown to satisfy all the four important conditions we have stated in the beginning, as the conditions required to be satisfied by a set of vectors, if they are to form a vector space. Any two signals, if added, will still give us a signal that is again continuous-time and having a finite energy over the period $[0, T]$. So the first condition of closure with respect to addition is satisfied since the sum signal again belongs to the set S. There exists a zero signal in S. It is a signal which has zero value over the entire interval $[0, T]$ and so is a continuous-time signal with zero energy over that interval. Thus, the second condition is satisfied. For energy signal, $s(t)$, over $[0, T]$, there exists in S, another signal $-s(t)$ over the same interval, and the sum of these two yields the zero signal. Hence, the third condition is satisfied. Finally, if any $s(t)$ belonging to S is multiplied by a real number (or a complex number), the resulting signal is also a continuous-time signal with some finite energy over the interval $[0, T]$, and thus belong to S. Thus, the set S of all continuous-time signals with finite energy, defined over the interval $[0, T]$, forms what we may call as a 'signal space', which is analogous to a 'vector space'. It is formed by a set of signals satisfying certain conditions, as stated earlier.

Now that the analogy between vectors and signals has been clearly established, we can extend the concept like 'linear independence', 'basis set', 'dimension', 'orthogonality', etc., associated with vectors to the signals too.

2.3.5 Linear Independence of Signals

A set of signals is said to be a linearly independent set of signals, provided none of them can be expressed as a linear combination of the rest.

Basis set for a signal space A basis set B for a signal space S, is a set of linearly independent signals which span (i.e., generate) the signal space S.

This means that the signals in a basis set should not only be linearly independent but should also be able to generate any signal belonging to S through a linear combination of some or all of them (basis signals).

Dimension of a signal space The dimension of a signal space S is the number of basis signals in *any basis set* for S.

2.3.6 Orthogonality of Signals

Just as we said that two distinct vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal if their dot product is zero, we now say that two distinct signals $s_1(t)$ and $s_2(t)$ are orthogonal to each other over an interval t_1 to t_2 if their inner product over that interval is zero, where we define their *inner product* as:

$$(s_1(t), s_2(t)) \triangleq \int_{t_1}^{t_2} s_1(t)s_2^*(t)dt \tag{2.15}$$

where $s_2^*(t)$ is the complex conjugate of $s_2(t)$. If we are dealing with signal spaces of only real signals, this complex conjugation in the RHS of Eq. (2.15) may be ignored.

Thus, signals $s_1(t)$ and $s_2(t)$ are said to be orthogonal, if their inner product $(s_1(t), s_2(t))$ is equal to zero.

2.3.7 Norm or Length of a Signal

The inner product permits us to give a geometrical structure to the 'signal space', just as the dot product did for a vector space. We can now talk not only of the orthogonality of signals but also of their 'lengths' and 'distance' between them, and so on. For this purpose, let us take the inner product of a signal *with itself*. This is usually represented by $\|s_i(t)\|^2$ (read as norm square of $s_i(t)$)

$$(s_i(t), s_i(t)) = \int_{t_1}^{t_2} s_i(t)s_i^*(t)dt = \int_{t_1}^{t_2} |s_i(t)|^2 dt \triangleq \|s_i(t)\|^2$$

$= E =$ Energy of the signal $s_i(t)$ over the interval t_1 to t_2.

$\therefore$ Norm or length of the finite energy signal $s_i(t) = \sqrt{E}$ (2.16)

2.3.8 Distance between Two Signals

This is a very useful quantity because if two signals $s_1(t)$ and $s_2(t)$ that are used for transmission by a transmitter, have a large distance separating them, the receiver is likely to commit less mistakes while performing statistical detection of the received signal in the presence of noise.

The distance between two signals $s_i(t)$ and $s_j(t)$, is defined as the norm of their difference, i.e.,

Distance between $s_i(t)$ and $s_j(t) = \|\{s_i(t) - s_j(t)\}\|$

$$= ((s_i(t) - s_j(t)), (s_i(t) - s_j(t)))^{1/2} = \left[\int_{t_1}^{t_2} |[s_i(t) - s_j(t)]|^2\, dt\right]^{1/2}$$

= Positive square-root of the energy of the difference signal over the interval t_1 to t_2

$$\therefore \quad \underset{s_i to s_j}{d} = \sqrt{E_{s_k}} \text{ where } s_k(t) \triangleq (s_i(t) - s_j(t)) \tag{2.17}$$

2.3.9 A Set of Orthogonal/Orthonormal Signals

A set S of *non-zero* signals $s_i(t)$, $i = 1, 2, 3, \ldots$ is said to be an *orthogonal set* over the interval t_1 to t_2 if

$$\int_{t_1}^{t_2} s_i(t)s_j^*(t)dt = \begin{cases} 0 & \text{for } i \neq j \\ \text{a number } k & \text{for } i = j \end{cases} \tag{2.18}$$

i.e., the inner-product of any two *distinct* signals must be zero.

In case, k in RHS of Eq. (2.18) is 1, then the signals are said to form an *orthonormal set* since every signal in the set has a norm, i.e., square-root of energy, equal to one.

Note *Any signal s(t) can be normalized so as to have a unit norm by dividing the signal by its own norm.*

i.e.,

$$\left\| \frac{s(t)}{\|s(t)\|} \right\| = 1 \tag{2.19}$$

A complete set of orthonormal signals A set S of orthonormal signals, all of which are defined over the interval t_1 to t_2 and having a finite energy over that interval, is said to be *a complete set of orthonormal signals*, if *any signal* defined over that interval and having a finite energy in that interval can be expressed, *without any error* as a linear combination of the members of the set S.

Some complete sets of orthogonal functions

1. The set of functions, $x_n(t) = e^{\frac{j2\pi nt}{T}}$, $n = 0, \pm 1, \pm 2, \ldots$, forms a complete set of orthogonal functions over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$.
2. The set of functions

$$\frac{1}{\sqrt{T}}, \sqrt{\frac{2}{T}} \cos \omega_0 t, \sqrt{\frac{2}{T}} \cos 2\omega_0 t, \ldots$$

$$\sqrt{\frac{2}{T}} \sin \omega_0 t, \sqrt{\frac{2}{T}} \sin 2\omega_0 t, \ldots \text{ with } \omega_0 \triangleq \frac{2\pi}{T}$$

   form a complete set of orthonormal functions over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$.
3. The set of Legendre Polynomials, $P_n(t)$, $n = 0, 1, 2, \ldots$, where

$$P_n(t) = \triangleq \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n$$

   forms a complete set of orthogonal functions.

Example 2.6 Show that the signals $x_n(t) = A \cos n\omega_0 t$, $n = 0, 1, 2, \ldots$, where, $\omega_0 \triangleq \frac{2\pi}{T}$, form a set of orthogonal functions over the interval $[0, T]$. Are they orthonormal? If they are not, obtain an orthonormal set.

Solution Consider any two distinct members of the given set, say, $A \cos m\omega_0 t$ and $A \cos n\omega_0 t$, where $m \neq n$. Then their inner product is given by

$$(A \cos m\omega_0 t, A \cos n\omega_0 t) = A^2 \int_0^T \cos m\omega_0 t \cdot \cos n\omega_0 t dt$$

$$= \frac{A^2}{2} \int_0^T \cos(m+n)\omega_0 t dt + \frac{A^2}{2} \int_0^T \cos(m-n)\omega_0 t dt$$

Since $m \neq n$, $m - n \neq 0$ $\quad \therefore$ let $(m-n) \triangleq k$, an integer.

Since $\cos(m+n)\omega_0 t$ has a frequency of $(m+n)f_0$ and $\cos(m-n)\omega_0 t$ has a frequency of $(m-n)f_0 = kf_0$, and since T is the period of $\cos\omega_0 t$, there will be an integer number of cycles of $\cos(m+n)\omega_0 t$ as well as $\cos(m-n)\omega_0 t$ in the integration interval $[0, T]$. Hence, both the above integrals are zero. Hence, $(A\cos m\omega_0 t, A\cos n\omega_0 t) = 0$ for $m \neq n$.

Therefore, the given set, viz., $x_n(t) = A\cos n\omega_0 t$, $n = 0, 1, 2, \ldots$ is an orthogonal set of signals.

If $m = n$, the first integral is zero but the second integral equals $A^2T/2$, which is the energy, E of the signal $A\cos n\omega_0 t$ for $n \neq 0$. For $n = 0$, the energy is simply $A^2T = 2E$.

Therefore, to normalize the orthogonal set and obtain an orthonormal set, we divide each member of this set by its own norm, i.e., by $\sqrt{2E}$ for $x_n(t)$ with $n = 0$ and by $\sqrt{E}$ for $x_n(t)$ with $n \neq 0$.

Thus, the normalized set $= y_n(t) = \dfrac{1}{\sqrt{T}}, \sqrt{\dfrac{2}{T}}\cos\omega_0 t, \sqrt{\dfrac{2}{T}}\cos 2\omega_0 t, \sqrt{\dfrac{2}{T}}\cos 3\omega_0 t, \ldots$

Example 2.7 Show that $f(t)$ is orthogonal to signals $\cos t, \cos 2t, \ldots, \cos nt, \ldots$ for all non-zero integer values of n over the interval 0 to 2π if

$$f(t) = \begin{cases} 1; & 0 \le t \le \pi \\ -1; & \pi \le t \le 2\pi \end{cases}$$

Solution $f(t)$ is plotted in Fig. 2.12.

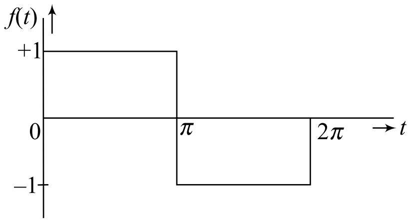


Fig. 2.12 *Waveform of f(t) of Example 2.7*

To show that $f(t)$ and the set of functions $\cos nt$, n an integer and $n \neq 0$, are orthogonal over the interval 0 to 2π, we show that their inner product, given by

$$(f(t), \cos nt) \triangleq \int_0^{2\pi} f(t)\cos nt\,dt \text{ equals zero.}$$

$$\int_0^{2\pi} f(t)\cos nt\,dt = \int_0^{\pi} 1\cdot\cos nt\,dt + \int_{\pi}^{2\pi} -1\cdot\cos nt\,dt$$

$$= \frac{1}{n}\sin nt\Big|_{t=0}^{\pi} - \frac{1}{n}\sin nt\Big|_{\pi}^{2\pi} = 0 \text{ for } n, \text{ an integer} \neq 0$$

Hence, the set of functions, $\cos nt$ for all non-zero integer values of n, will be orthogonal to the given $f(t)$.

Example 2.8 If $x_e(t)$ and $x_0(t)$ are respectively the even and odd parts of a signal $x(t)$, show that they are orthogonal over the interval $-T$ to T for any T.

Solution To show that $x_e(t)$ and $x_0(t)$ are orthogonal to each other over $-T$ to T for any T, we have to show that their inner product given by

$$(x_e(t), x_0(t)) = \int_{-T}^{T} x_e(t)x_0(t)\,dt \quad \text{equals zero}$$

Now, $x_e(t)\,x_0(t)$ is the product of an even signal and an odd signal, and hence it is odd.

$$\therefore \qquad \int_{-T}^{T} x_e(t)x_0(t)\,dt = 0 \qquad \text{for any } T$$

Hence, $x_e(t)$ and $x_0(t)$ are orthogonal over $(-T, T)$ for any T.

Note *Expansion of a signal x(t) using a complete set of orthogonal functions, is called the generalized Fourier series represenattion of the signal x(t).*

2.3.10 Component of One Signal along Another Signal

Consider n non-zero orthogonal signals, $s_1(t)$, $s_2(t)$, …, $s_n(t)$. Let $x(t)$ be a linear combination of these.

$$\therefore \quad x(t) = C_1 s_1(t) + C_2 s_2(t) + C_3 s_3(t) + \cdots + C_k s_k(t) + \ldots + C_n s_n(t)$$

where C_i, $i = 1$ to n are some real numbers.

Taking the inner-product of both sides with say $s_k(t)$, we get

$$(x(t), s_k(t)) = C_1(s_1(t), s_k(t)) + C_2(s_2(t), s_k(t)) + C_k(s_k(t), s_k(t)) + \ldots + C_n(s_n(t), s_k(t))$$

Since $s_i(t)$, $i = 1$ to n are orthogonal signals, all the inner products on the RHS of the above equation will be zero, except $(s_k(t), s_k(t))$, which is equal to $\|s_k(t)\|^2$ = Energy of $s_k(t)$.

$$\therefore \quad (x(t), s_k(t)) = C_k \|s_k(t)\|^2$$

Hence C_k = coordinate of $x(t)$ along $s_k(t) = \dfrac{(x(t), s_k(t))}{\|s_k(t)\|^2}; \quad 1 \le k \le n$ (2.20)

Therefore, the component of $x(t)$ along $s_k(t) = C_k\, s_k(t)$

$$\text{Component of } x(t) \text{ along } s_k(t) = \frac{(x(t), s_k(t))}{\|s_k(t)\|^2} s_k(t); \quad 1 \le k \le n \tag{2.21}$$

2.4 GRAM–SCHMIDT ORTHOGONALIZATION PROCEDURE

Given n linearly independent signals $x_i(t)$, $i = 1, 2, 3, \ldots, n$ belonging to a signal space S, this procedure enables us to derive from them a set of n orthogonal or orthonormal signals $s_i(t)$, $i = 1, 2, \ldots, n$, in S.

Procedure We first take $s_1(t) = x_1(t)$. Then we know that $s_2(t)$ should be orthogonal to $s_1(t)$. So, from $x_2(t)$, we subtract that part of it which is along $s_1(t)$. We call that $s_2(t)$. Hence,

$$s_2(t) = x_2(t) - \frac{(x_2(t), s_1(t))}{\|s_1(t)\|^2} \cdot s_1(t)$$

To obtain $s_3(t)$ from $x_3(t)$, we subtract from $x_3(t)$, its components along $s_1(t)$ as well as $s_2(t)$. The remaining part of $x_3(t)$ will therefore be orthogonal to both $s_1(t)$ as well as $s_2(t)$ and we call it $s_3(t)$.

$$\therefore \quad s_3(t) = x_3(t) - \left[\frac{(x_3(t), s_1(t))}{\|s_1(t)\|^2}\right] \cdot s_1(t) - \left[\frac{(x_3(t), s_2(t))}{\|s_2(t)\|^2}\right] \cdot s_2(t)$$

$$\therefore \quad s_3(t) = x_3(t) - \sum_{k=1}^{2} \left[\frac{(x_3(t), s_k(t))}{\|s_k(t)\|^2}\right] \cdot s_k(t)$$

We go on like this till we get $s_n(t)$. Hence, in general,

$$s_{m+1}(t) = x_{m+1}(t) - \sum_{k=1}^{m} \left[\frac{(x_{m+1}(t), s_k(t))}{\|s_k(t)\|^2}\right] \cdot s_k(t); m = 1 \text{ to } (n-1) \tag{2.22}$$

Once an orthogonal set of signals $s_i(t)$, $i = 1$ to n are obtained, to obtain an orthonormal set, we simply normalize each one of the orthogonal signals $s_i(t)$, $i = 1$ to n by dividing each one by its own norm. If we call the normalized set as $y_i(t)$, $i = 1$ to n, we have

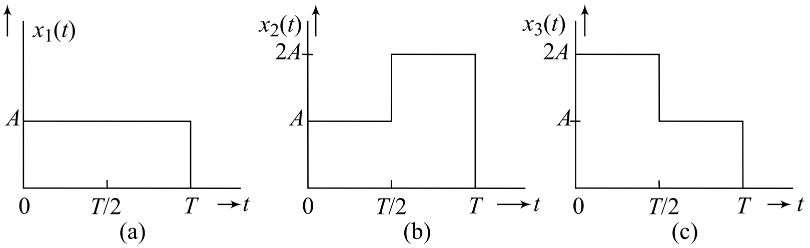
$$y_k(t) = \frac{s_k(t)}{\|s_k(t)\|} = \frac{s_k(t)}{\left[\int_{t_1}^{t_2} |s_k(t)|^2\, dt\right]^{1/2}} \tag{2.23}$$

Note *If the given signals $x_i(t)$, $i = 1$ to n are linearly independent, we will be able to derive again n orthonormal signals from them. If x_is are not linearly independent, we can obtain m orthonormal signals, $m < n$, where m is the dimension of the signal space generated by $x_i(t)$, $i = 1$ to n.*

Example 2.9 Given the three signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ as shown in Fig. 2.12, derive an orthonormal basis signal set for the signal space generated by them.

Solution We shall make use of Gram–Schmidt procedure to obtain an orthonormal set of signals from the given three signals.

First let us define $s_1(t) = x_1(t)$.



(a) (b) (c)

Fig. 2.13 *The given signals $x_1(t)$, $x_2(t)$ and $x_3(t)$*

Then,

$$s_2(t) = x_2(t) - \frac{(x_2(t), s_1(t))}{\|s_1(t)\|^2} \cdot s_1(t)$$

$$(x_2(t), s_1(t)) = (x_2(t), x_1(t)) = \int_0^T x_2(t)x_1(t)dt = \int_0^{T/2} A^2 dt + \int_{T/2}^{T} 2A^2 dt = \frac{3}{2}A^2T$$

$$\|s_1(t)\|^2 = \|x_1(t)\|^2 = (x_1(t), x_1(t)) = \int_0^T A^2 dt = A^2T$$

$$\therefore \qquad s_2(t) = x_2(t) - \left(\frac{3}{2}A^2T \cdot \frac{1}{A^2T}\right)x_1(t) = x_2(t) - \frac{3}{2}x_1(t)$$

The plot of $s_2(t)$ is shown in Fig. 2.14.

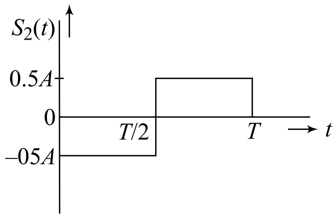


Fig. 2.14 *Signal $s_2(t)$*

$$s_3(t) = x_3(t) - \frac{(x_3(t), x_1(t))}{\|x_1(t)\|^2}x_1(t) - \frac{(x_3(t), s_2(t))}{\|s_2(t)\|^2}s_2(t)$$

$$(x_3(t), x_1(t)) = \int_0^T x_3(t)x_1(t)dt = \int_0^{T/2} 2A^2 dt + \int_{T/2}^{T} A^2 dt = \frac{3}{2}A^2T$$

$$(x_3(t), s_2(t)) = \int_0^T x_3(t)s_2(t)dt = \int_0^{T/2} 2A \times (-0.5A)dt + \int_{T/2}^{T} A \times (0.5A)dt = -\frac{0.5}{2}A^2T$$

$$\|s_2(t)\|^2 = (s_2(t), s_2(t)) = \int_0^{T/2} (-0.5A)(-0.5A)dt + \int_{T/2}^{T} (0.5A)(0.5A)dt = \frac{0.5A^2T}{2} = 0.25A^2T$$

$$\therefore \qquad s_3(t) = x_3(t) - \left(\frac{3}{2}A^2T\right)\left(\frac{1}{A^2T}\right)x_1(t) - \frac{0.25A^2T}{0.25A^2T} \cdot s_2(t) = x_3(t) - \frac{3}{2}x_1(t) - s_2(t)$$

A plot of the signal $\left[\frac{3}{2}x_1(t)+s_2(t)\right]$ is shown in Fig. 2.15. We find that it is such that $x_3(t)$ minus this signal gives us what is shown in Fig. 2.16. We find that it is equal to $-2\ s_2(t)$, i.e., $s_3(t) = -2\ s_2(t)$. Hence, they are not linearly independent (see the note under Gram–Schmidt procedure). This is due to the fact that the three given signals, $x_1(t)$, $x_2(t)$ and $x_3(t)$ are themselves not linearly independent. This is obvious from the fact that $x_1(t)$ can be obtained in terms of $x_2(t)$ and $x_3(t)$.

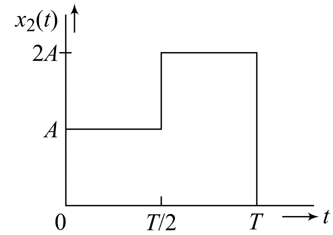


Fig. 2.15 *Signal $x_2(t)$*

Actually, $\quad x_1(t)=\frac{1}{3}[x_2(t)+x_3(t)]$

Hence the signal space generated by $x_1(t)$, $x_2(t)$ and $x_3(t)$ is not three dimensional – it is only two-dimensional and it has the orthogonal signals $s_1(t)$ and $s_2(t)$ as the basis signals. That they are orthogonal is clear from the fact their inner product, viz.,

$$(s_1(t), s_2(t))=\int_0^{T/2}(-0.5A)A\,dt+\int_{T/2}^{T}(0.5A)A\,dt=0$$

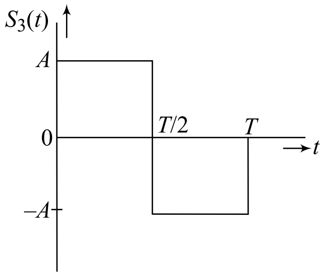


Fig. 2.16 *Signal $s_3(t)$*

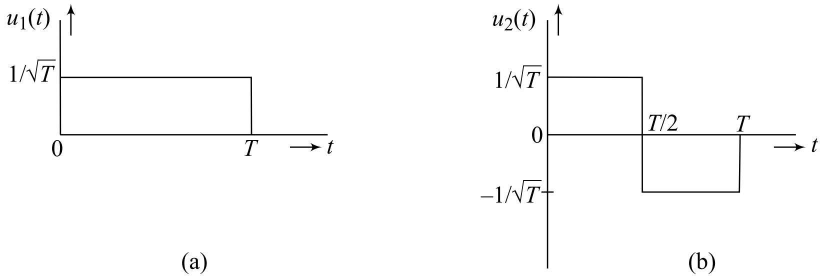
Now to normalize them, we have to find their norms

$$\|s_1(t)\|^2=(s_1(t), s_1(t))=\int_0^T s_1(t)dt=A^2T=\text{Energy } E$$

$$\|s_2(t)\|^2=(s_2(t), s_2(t))=\int_0^{T/2}(-0.5A)(-0.5A)dt+\int_{T/2}^{T}(0.5A)(0.5A)dt=\frac{0.5A^2T}{2}=0.25A^2T$$

∴ if $u_1(t)$ and $u_2(t)$ are the normalized versions of $s_1(t)$ and $s_2(t)$,

$$u_1(t)=\frac{s_1(t)}{\sqrt{A^2T}}=\frac{1}{\sqrt{E}}s_1(t)\quad\text{and}\quad u_2(t)=\frac{s_2(t)}{\|s_2(t)\|}=\frac{s_2(t)}{\sqrt{0.25A^2T}}=\frac{2s_2(t)}{\sqrt{E}}$$



(a) (b)

Fig. 2.17 *The orthonormal basis set*

Note that the energy of each of these signals is 1, as it should be.

Example 2.10 $u_1(t)$ and $u_2(t)$ have been obtained as the orthonormal basis set for the signal space of Example 2.9. Express the given signals, $x_1(t)$, $x_2(t)$ and $x_3(t)$ as linear combinations of $u_1(t)$ and $u_2(t)$, i.e., generate $x_1(t)$, $x_2(t)$ and $x_3(t)$ from $u_1(t)$ and $u_2(t)$, the two basis signals.

Solution Let $x_1(t) = a_1u_1(t) + a_2u_2(t)$

Then a_1 is the coordinate of $x_1(t)$ along $u_1(t)$. Hence, from Eq. (2.20), it is given by the inner-product of $x_1(t)$ with $u_1(t)$ divided by the norm square of $u_1(t)$, which of course, is 1.

$$\therefore \qquad a_1 = (x_1(t), u_1(t)) = \int_0^T A \cdot \frac{1}{\sqrt{T}} dt = \frac{A}{\sqrt{T}} \cdot T = A\sqrt{T}$$

$$a_2 = (x_1(t), u_2(t)) = \int_0^T x_1(t), u_2(t) dt = \int_0^{T/2} A\left(-\frac{1}{\sqrt{T}}\right) dt + \int_{T/2}^{T} A\left(\frac{1}{\sqrt{T}}\right) dt = 0$$

$$\therefore \qquad x_1(t) = A\sqrt{T} u_1(t) + 0 \cdot u_2(t)$$

Now, let $x_2(t) = b_1u_1(t) + b_2u_2(t)$

$$\therefore \qquad b_1 = (x_2(t), u_1(t)) = \int_0^T x_2(t), u_1(t) dt = \int_0^{T/2} A \cdot \frac{1}{\sqrt{T}} dt + \int_{T/2}^{T} 2A \cdot \frac{1}{\sqrt{T}} dt = \frac{3}{2} A\sqrt{T}$$

$$b_2 = (x_2(t), u_2(t)) = \int_0^T x_2(t), u_2(t) dt = \int_0^{T/2} A\left(-\frac{1}{\sqrt{T}}\right) dt + \int_{T/2}^{T} 2A \frac{1}{\sqrt{T}} dt = \frac{1}{2} A\sqrt{T}$$

$$\therefore \qquad x_2(t) = \frac{3}{2} A\sqrt{T} u_1(t) + \frac{1}{2} A\sqrt{T} u_2(t)$$

Let $x_3(t) = c_1u_1(t) + c_2u_2(t)$,

$$c_1 = (x_3(t), u_1(t)) = \int_0^T x_3(t), u_1(t) dt = \int_0^{T/2} 2A \cdot \frac{1}{\sqrt{T}} dt + \int_{T/2}^{T} A \frac{1}{\sqrt{T}} dt = \frac{3}{2} A\sqrt{T}$$

$$c_2 = (x_3(t), u_2(t)) = \int_0^T x_3(t), u_2(t) dt = \int_0^{T/2} 2A\left(-\frac{1}{\sqrt{T}}\right) dt + \int_{T/2}^{T} A \frac{1}{\sqrt{T}} dt = -\frac{1}{2} A\sqrt{T}$$

$$\therefore \qquad x_3(t) = \frac{3}{2} A\sqrt{T} u_1(t) - \frac{1}{2} A\sqrt{T} u_2(t)$$

Vector representation of signals/signal space diagrams While discussing the basics of vector spaces, we had said that a vector $\boldsymbol{A}$ in the Euclidian space could be represented as $[x_1, x_2, x_3]^T$ where x_1, x_2 and x_3 are respectively the coordinates of the vector $\boldsymbol{A}$ along the X, Y and Z orthogonal axes of the Euclidian space. The vector $\boldsymbol{A}$ itself could be written down as an orthogonal expansion as:

$$\boldsymbol{A} = x_1\boldsymbol{i} + x_2\boldsymbol{j} + x_3\boldsymbol{k} = [x_1, x_2, x_3]^T$$

where $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ are orthonormal basis vectors for the Euclidian space.

In Example 2.10, we had obtained $u_1(t)$ and $u_2(t)$ as an orthonormal basis set for the signal space S spanned by the three given signals $x_1(t)$, $x_2(t)$ and $x_3(t)$. We have also obtained the representation of the given signals in terms of the orthonormal basis signals $u_1(t)$ and $u_2(t)$ as:

$$x_1(t) = A\sqrt{T} u_1(t) + 0 \cdot u_2(t) = [A\sqrt{T}, 0]^T$$

$$x_2(t) = \frac{3}{2} A\sqrt{T} u_1(t) + \frac{1}{2} A\sqrt{T} u_2(t) = \left[\frac{3}{2} A\sqrt{T}, \frac{1}{2} A\sqrt{T}\right]^T$$

and

$$x_3(t) = \frac{3}{2} A\sqrt{T} u_1(t) - \frac{1}{2} A\sqrt{T} u_2(t) = \left[\frac{3}{2} A\sqrt{T}, -\frac{1}{2} A\sqrt{T}\right]^T$$

Hence, in analogy with the representation of the vector **A** as a point (in the Euclidian space) with coordinates x_1, x_2 and x_3 along $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ vectors, the signal $x_1(t)$ can be represented as a point P in the signal space S, with $A\sqrt{T}$ and 0 as its coordinates respectively along the $u_1(t)$ and $u_2(t)$ basis signal directions; signal $x_2(t)$ as a point Q in the signal space S, with coordinates $\frac{3}{2}A\sqrt{T}$ and $\frac{1}{2}A\sqrt{T}$ and the signal $x_3(t)$ as a point R with coordinates $\frac{3}{2}A\sqrt{T}$ and $-\frac{1}{2}A\sqrt{T}$. It is a geometrical representation or vector representation of signals in the signal space, and is generally known as the *signal-space diagram* of those signals which is shown in Fig. 2.7.

Remark *The orthonormal basis set $u_1(t)$ and $u_2(t)$ obtained from the given signal set $x_1(t)$, $x_2(t)$ and $x_3(t)$ by using Gram–Schmidt orthogonalization is not unique. This is because, if only we had initially set $s_1(t) = x_2(t)$ or $x_3(t)$ instead of $x_1(t)$ as we did, while applying Gram–Schmidt procedure, i.e., if we had ordered the signals differently, we would have got totally different sets of orthonormal basis signals. The representation of each of the signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ also would be different when the basis set is different. Hence, the signal-space diagram also is not unique.*

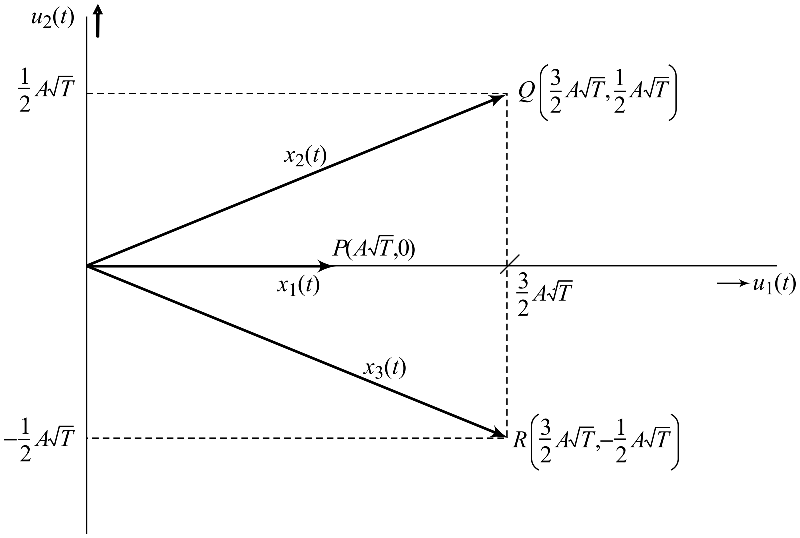


Fig. 2.18 *Signal-space diagram of signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ of Example 2.10*

Example 2.11 Determine the distance between the signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ of Example 2.10. If during each time slot, any one of these signals is transmitted, which type of error is more likely? **(a)** $x_1(t)$ being mistake for $x_2(t)$, **(b)** $x_1(t)$ being mistaken for $x_3(t)$, or **(c)** $x_2(t)$ being mistaken for $x_3(t)$?

Solution The distance between $x_1(t)$ and $x_2(t)$ in the signal space diagram is given by

$$\|[x_1(t) - x_2(t)]\| = \left[\left(\frac{3}{2}A\sqrt{T} - A\sqrt{T}\right)^2 + \left(\frac{1}{2}A\sqrt{T}\right)^2\right]^{1/2}$$

$$= \left[\frac{1}{4}A^2T + \frac{1}{4}A^2T\right]^{1/2} = \frac{1}{\sqrt{2}}A\sqrt{T}$$

Similarly, $$\|[x_1(t)-x_3(t)]\| = \left[\left(\frac{3}{2}A\sqrt{T}-A\sqrt{T}\right)^2+\left(-\frac{1}{2}A\sqrt{T}\right)^2\right]^{1/2}$$

$$=\left[\frac{1}{4}A^2T+\frac{1}{4}A^2T\right]^{1/2}=\frac{1}{\sqrt{2}}A\sqrt{T}$$

and $$\|[x_2(t)-x_3(t)]\| = \left[\left(\frac{3}{2}A\sqrt{T}-\frac{3}{2}A\sqrt{T}\right)^2+\left(\frac{1}{2}A\sqrt{T}+\frac{1}{2}A\sqrt{T}\right)^2\right]^{1/2}$$

$$=[A^2T]^{1/2}=A\sqrt{T}$$

Hence, $x_1(t)$ is equidistant from $x_2(t)$ and $x_3(t)$, the distance being $\frac{1}{\sqrt{2}}A\sqrt{T}$. The distance between $x_2(t)$ and $x_3(t)$ is, however, $A\sqrt{T}$ which is $\sqrt{2}$ times the distance between $x_1(t)$ and either $x_2(t)$ or $x_3(t)$.

Therefore, the probability of $x_2(t)$ being mistaken for $x_3(t)$ or vice versa, is much less than the probability of $x_1(t)$ being mistaken either as $x_2(t)$ or $x_3(t)$ (or either $x_2(t)$ or $x_3(t)$ being wrongly identified as $x_1(t)$).

Example 2.12 A *QPSK* modulator transmits one of the following four signals during each time slot of T sec.

$$x_i(t)=\sqrt{2P_s}\cos\left[\omega_0t+(2i-1)\frac{\pi}{4}\right];\ i=1,2,3,4$$

and $2\omega_0T=n\pi$ where n is an integer. **(a)** Draw the signal space diagram using orthonormal coordinates, and **(b)** Determine the maximum distance between any two signals

Solution The given signal may be written as

$$x_i(t)=\sqrt{2P_s}\left[\cos(2i-1)\frac{\pi}{4}\right]\cos\omega_0t-\sqrt{2P_s}\left[\sin(2i-1)\frac{\pi}{4}\right]\sin\omega_0t \text{ with } i=1,2,3,4$$

Since each $x_i(t)$ has a $\cos\omega_0t$ component and a $\sin\omega_0t$ component, and since these two are orthogonal over the interval $[0, T]$, without going through Gram–Schmidt orthogonalization procedure, let us straight away take the two *orthogonal* basis signals as, say, $\Psi_1(t)=\sqrt{2P_s}\cos\omega_0t;\ 0\le t\le T$ and $\Psi_2(t)=\sqrt{2P_s}\sin\omega_0t;\ 0\le t\le T$. These are orthogonal but not orthonormal. To normalize them, we have to divide each by its own norm.

$$\|\Psi_1(t)\|^2=(\Psi_1(t),\Psi_1(t))=\int_0^T\Psi_1^2(t)dt=\int_0^T 2P_s\cos^2\omega_0tdt$$

$$=2P_s\int_0^T\frac{1}{2}[1+\cos2\omega_0t]dt=P_sT \quad \text{since} \int_0^T\cos2\omega_0tdt=0$$

(since $2\omega_0T=n\pi$ where n is an integer)

$\therefore$ $$\|\Psi_1(t)\|=\sqrt{TP_s} \quad \therefore \quad u_1(t)=\text{Normalized basis signal}=\frac{\sqrt{2P_s}\cos\omega_0t}{\sqrt{TP_s}}$$

$\therefore$ $$u_1(t)=\sqrt{\frac{2}{T}}\cos\omega_0t \text{ and similarly, } u_2(t)=\sqrt{\frac{2}{T}}\sin\omega_0t.$$

Now,

$$
\begin{aligned}
x_i(t) &= \sqrt{2P_s}\left[\cos(2i-1)\frac{\pi}{4}\right]\cos\omega_0 t - \sqrt{2P_s}\left[\sin(2i-1)\frac{\pi}{4}\right]\sin\omega_0 t \\
&= \sqrt{P_s T}\left[\cos(2i-1)\frac{\pi}{4}\right]u_1(t) - \sqrt{P_s T}\left[\sin(2i-1)\frac{\pi}{4}\right]u_2(t) \\
&= \sqrt{P_s T}C_1 u_1(t) - \sqrt{P_s T}C_2 u_2(t),
\end{aligned}
$$

where $C_1 = \cos(2i-1)\frac{\pi}{4}$ and $C_2 = -\sin(2i-1)\frac{\pi}{4}$

Now, when

$$
\begin{aligned}
&i = 1, C_1 = \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}; C_2 = -\sin\frac{\pi}{4} = \frac{-1}{\sqrt{2}} \\
&i = 2, C_1 = \cos\left(\frac{3\pi}{4}\right) = -\frac{1}{\sqrt{2}}; C_2 = -\sin\frac{3\pi}{4} = \frac{-1}{\sqrt{2}} \\
&i = 3, C_1 = \cos\left(\frac{5\pi}{4}\right) = -\frac{1}{\sqrt{2}}; C_2 = -\sin\left(\frac{5\pi}{4}\right) = +\frac{1}{\sqrt{2}} \\
&i = 4, C_1 = \cos\left(\frac{7\pi}{4}\right) = \frac{1}{\sqrt{2}}; C_2 = -\sin\left(\frac{7\pi}{4}\right) = +\frac{1}{\sqrt{2}}
\end{aligned}
$$

The corresponding signal-space diagram is shown in Fig. 2.19.

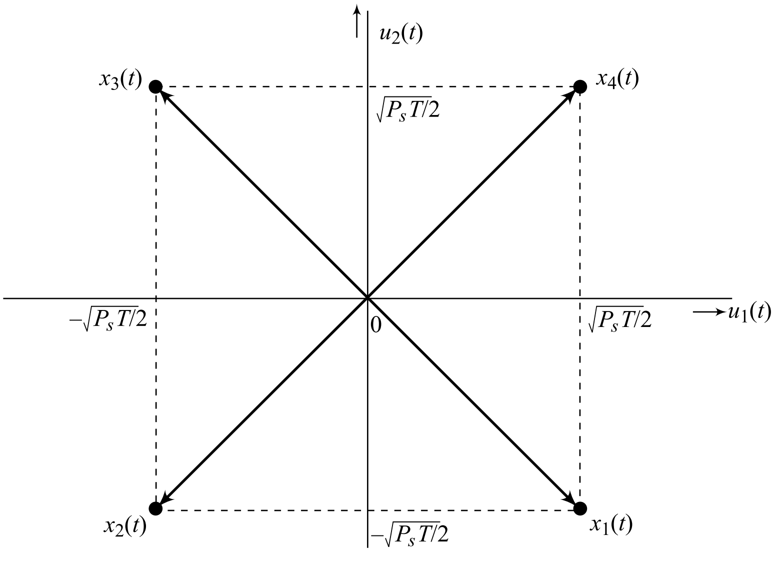


Fig. 2.19 *Signal-space diagram and signal vectors for the four signals of Example 2.12*

The maximum distance between any two vectors is equal to the distance between $x_3(t)$ and $x_1(t)$ or the distance between $x_2(t)$ and $x_4(t)$. This distance is equal to

$$d = 2 \times \sqrt{2} \cdot \sqrt{\frac{P_s T}{2}} = 2\sqrt{P_s T}$$

Example 2.13 **(a)** Show that the three functions given below are pairwise orthogonal over the interval [–2, 2], **(b)** Determine the value of the constant A that makes the set of functions an orthonormal set, and **(c)** Express the waveform $x(t) = \begin{cases} 1 & \text{for } 0 \le t \le 2 \\ 0 & \text{otherwise} \end{cases}$ in terms of the orthonormal set obtained in part (b).

(**University Examination Question**)

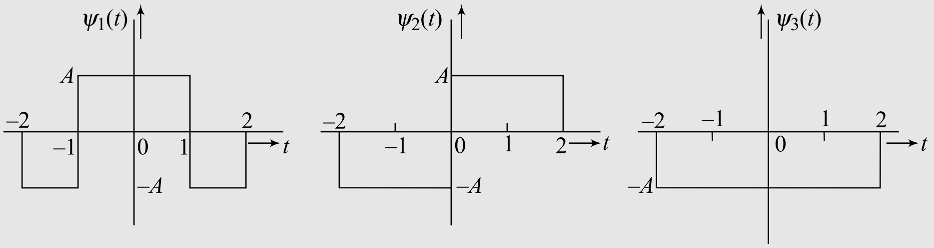


Fig. 2.20

Solution

(a) To show that the three functions $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$ are pairwise orthogonal, we have to show that pairwise, their innerproducts are zero, i.e.,

$$(\psi_1(t), \psi_2(t)) = (\psi_2(t), \psi_3(t)) = (\psi_3(t), \psi_1(t)) = 0$$

Now,

$$(\psi_1(t), \psi_2(t)) = \int_{-2}^{2} \psi_1(t)\psi_2(t)dt = \int_{-2}^{-1} \psi_1(t)\psi_2(t)dt + \int_{-1}^{0} \psi_1(t)\psi_2(t)dt + \int_{0}^{1} \psi_1(t)\psi_2(t)dt + \int_{1}^{2} \psi_1(t)\psi_2(t)dt$$

$$= \int_{-2}^{-1} (-A)(-A)dt + \int_{-1}^{0} (A)(-A)dt + \int_{0}^{1} (A)(A)dt + \int_{1}^{2} (-A)(A)dt = 0$$

In a similar way, it can be shown that $(\psi_2(t), \psi_3(t))$ and $(\psi_3(t), \psi_1(t))$ are each equal to zero. Hence, $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$ are pairwise orthogonal.

(b) If $\psi_i(t)$ is normalized, then $\|\psi_i(t)\|^2 = 1 = (\psi_i(t), \psi_i(t))$, for $i = 1, 2$ and 3.

Taking $(\psi_1(t), \psi_1(t)) = \|\psi_1(t)\|^2$, it is equal to

$$\int_{-2}^{2} \psi_1^2(t)dt = \int_{-2}^{-1} \psi_1^2(t)dt + \int_{-1}^{0} \psi_1^2(t)dt + \int_{0}^{1} \psi_1^2(t)dt + \int_{1}^{2} \psi_1^2(t)dt = 4A^2$$

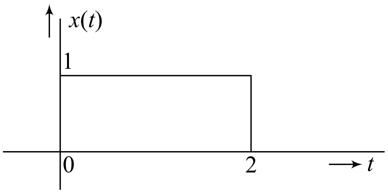
If $\psi_1(t)$ is normalized, this should be unity.

$\therefore$ $4A^2 = 1$ or $A = \pm\frac{1}{2}$. Let us take it as 1/2.

$\therefore$ A value of $A = 1/2$ will normalize $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$, and make them orthonormal functions.

(c) $x(t) = \begin{cases} 1 & \text{for } 0 \le t \le 2 \\ 0 & \text{otherwise} \end{cases}$

Therefore, $x(t)$ is shown below.



Since $x(t)$ is to be expressed in terms of $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$, let:

$$x(t) = K_1\psi_1(t) + K_2\psi_2(t) + K_3\psi_3(t)$$

where $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$ are normalized functions, i.e., $A = 1/2$.
Then from Eq. (2.13),

$$K_1 = (x(t), \psi_1(t)) = \int_{-2}^{2} x(t), \psi_1(t)dt = \int_0^1 \frac{1}{2}dt + \int_1^2 \left(-\frac{1}{2}\right)dt = 0$$

$$K_2 = (x(t), \psi_2(t)) = \int_0^2 \left(\frac{1}{2}\right) \cdot 1dt = 1$$

and

$$K_3 = (x(t), \psi_3(t)) = \int_0^2 \left(-\frac{1}{2}\right) \cdot 1dt = -1$$

$\therefore$

$$x(t) = 0 \cdot \psi_1(t) + 1 \cdot \psi_2(t) + (-1)\psi_3(t)$$

where $\psi_1(t)$, $\psi_2(t)$ and $\psi_3(t)$ are normalized functions (i.e., $A = 1/2$).

2.4.1 Schwarz's Inequality

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two vectors, we know that their dot product $\boldsymbol{A} \cdot \boldsymbol{B}$ is

$$\boldsymbol{A} \cdot \boldsymbol{B} = |\boldsymbol{A}||\boldsymbol{B}|\cos\theta \tag{2.24}$$

$\therefore$

$$\cos\theta = \frac{\boldsymbol{A} \cdot \boldsymbol{B}}{|\boldsymbol{A}||\boldsymbol{B}|} \tag{2.25}$$

Since $|\cos\theta| \le 1$ and the 'equal to' sign in Eq. (2.24) holds good only if $\theta = 0$, i.e., if $\boldsymbol{A} = \alpha\boldsymbol{B}$

$$\left|\frac{\boldsymbol{A} \cdot \boldsymbol{B}}{|\boldsymbol{A}||\boldsymbol{B}|}\right| \le 1; \text{ 'equal to' holding good only if } \boldsymbol{A} = \alpha\boldsymbol{B} \text{ where } \alpha \text{ is a constant}$$

i.e.,

$$|\boldsymbol{A} \cdot \boldsymbol{B}| \le |\boldsymbol{A}||\boldsymbol{B}|; \text{ with 'equal to' only if } \boldsymbol{A} = \alpha\boldsymbol{B} \tag{2.26}$$

From the analogy between signals and vectors, which we discussed in detail earlier, the 'dot product' of vectors can be replaced by the inner product of signals and the magnitude of a vector can be replaced by the 'norm' of a signal. Thus, if we consider two signals $s_1(t)$ and $s_2(t)$ belonging to some signal space, we may write an equation analogous to Eq. (2.26) as

$$|(s_1(t), s_2(t))| \le \|s_1(t)\|\|s_2(t)\|; \text{ the '=' sign holds good only if } s_2(t) = c\, s_1(t) \tag{2.27}$$

But

$$(s_1(t), s_2(t)) = \int_{-\infty}^{\infty} s_1(t)s_2(t)dt$$

if both the signals are real valued.
Further,

$$\|s_1(t)\| = [(s_1(t), s_1(t))]^{1/2} = \left[\int_{-\infty}^{\infty} s_1^2(t)dt\right]^{1/2}$$

$$\|s_2(t)\| = [(s_2(t), s_2(t))]^{1/2} = \left[\int_{-\infty}^{\infty} s_2^2(t)dt\right]^{1/2}$$

Substituting these in Eq. (4.24), we get

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2(t)dt\right|^2 \le \left[\int_{-\infty}^{\infty} s_1^2(t)dt\right]\left[\int_{-\infty}^{\infty} s_2^2(t)dt\right] \tag{2.28}$$

where the '=' sign holds true *if and only if* $s_2(t) = c\ s_1(t)$, where c is an arbitrary constant. This is what is referred to as '***Schwarz's Inequality***'. If the signals are complex valued, it takes the form

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)dt\right|^2 \le \left[\int_{-\infty}^{\infty} |s_1(t)|^2 dt\right]\left[\int_{-\infty}^{\infty} |s_2(t)|^2 dt\right] \tag{2.29}$$

where the '=' sign holds true if and only if $s_2(t) = cs_1^*(t)$.

As can easily be seen, Schwarz's inequality is only a generalization of the well-known relation that the dot product of two vectors will have a magnitude that is less than or equal to the product of the magnitudes of the individual vectors (Eq. (2.26)).

2.5 COMPLEX-EXPONENTIAL FOURIER SERIES

The expansion of a signal $x(t)$ using the exponential functions

$$x_n(t) = e^{\frac{j2\pi nt}{T}},\ n = 0, \pm1, \pm2, \ldots$$

is called the *complex-exponential Fourier series* expansion of $x(t)$.

Before proceeding further, we shall first show that this set of complex-exponential functions, is an orthogonal set over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$. For this, referring to Eq. (2.18), we have to show that

$$\int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}mt}\cdot\left(e^{j\frac{2\pi}{T}nt}\right)^* dt = \begin{cases} 0 \text{ if } m \ne n \\ a \text{ constant, if } m = n \end{cases}$$

Assume $m \ne n$.

$$\int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}mt}\cdot\left(e^{j\frac{2\pi}{T}nt}\right)^* dt = \int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}(m-n)t}\, dt$$

Since m and n are both integers, and $m \ne n$, $(m - n)$ will also be an integer, say k and $k \ne 0$.

$$\int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}(m-n)t}\, dt = \int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}kt}\, dt = \frac{T}{j2\pi k} e^{j\frac{2\pi}{T}kt}\Bigg|_{-T/2}^{T/2}$$

$$= \frac{T}{j2\pi k}[e^{j\pi k} - e^{-j\pi k}] = T\left(\frac{\sin \pi k}{\pi k}\right) = 0 \quad \text{since } k \text{ is an integer and } k \ne 0.$$

Assume $m = n$. Then $m - n = 0$ and hence

$$\int_{-T/2}^{T/2} e^{j\frac{2\pi}{T}mt}\cdot\left(e^{j\frac{2\pi}{T}nt}\right)^* dt = \int_{-T/2}^{T/2} 1\cdot dt = T, \text{ a constant}$$

Thus, $x_n(t) = e^{j\frac{2\pi}{T}nt}$, $n = 0, \pm1, \pm2, \ldots$, form a set of orthogonal functions and if we normalize them by multiplying each of them by $1/\sqrt{T}$, the resultant functions will be forming a set of orthonormal functions over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$. It can be shown that these sets are complete sets.

Since $x_n(t)$, $n = 0, \pm 1, \pm 2, \ldots$, form a complete set of orthogonal functions over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$, it should be possible to express any signal $x(t)$ over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$ as a linear combination of these complex-exponential functions. Hence, we may write

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j\frac{2\pi}{T}nt}$$

(n taking only integer values)

$$= \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; \quad -\frac{T}{2} \le t \le +\frac{T}{2}, \tag{2.30}$$

where $f_0 \triangleq \frac{1}{T}$.

That is, the expansion of $x(t)$ using this complete set of orthogonal functions will be valid only over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$.

However, if $x(t)$ is periodic with a period T, the expansion will be valid for all time: Hence, we write

$$\boxed{\underset{\substack{\text{Periodic}\\ \text{with period}\,T}}{x(t)} = \sum_{n=-\infty}^{n=\infty} c_n e^{j2\pi n f_0 t}; \quad f_0 = \frac{1}{T}; \quad -\infty < t < \infty} \tag{2.31}$$

Now, to determine the constants c_n, $n = 0, \pm 1, \pm 2, \ldots$, which are called the complex-exponential Fourier series coefficients of $x(t)$, we use Eq. (2.20) and write

$$c_n = \frac{\int_{-T/2}^{T/2} x(t) e^{-j2\pi n f_0 t}\, dt}{\int_{-T/2}^{T/2} |e^{j2\pi n f_0 t}|^2\, dt} = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi n f_0 t}\, dt \tag{2.32}$$

We may summarize the foregoing and state that if $x(t)$ is a periodic signal with period T, it can be represented by the complex-exponential Fourier series expansion as

$$\boxed{\begin{array}{l} x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; \quad f_0 \triangleq \frac{1}{T}; \quad -\infty < t < \infty, \\ \text{where} \\ c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi n f_0 t}\, dt \end{array}} \tag{2.33}$$

If $x(t)$ is not periodic, then the above expansion is valid only over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$.

From Eq. (2.33), it is clear that c_n's, are in general, complex numbers. Thus, we may write

$$c_n = |c_n| e^{j\theta_n} \tag{2.34}$$

where $|c_n|$ is the magnitude of c_n and θ_n is the angle of c_n.

A plot of $|c_n|$ vs n or nf_0, is called the amplitude spectrum of the signal x(t) and a plot of θ_n vs n or nf_0 is called the phase spectrum of x(t). The reason for calling these as amplitude spectrum and phase spectrum may be understood from the following discussion.

A close look at Eq. (2.33) reveals that the Fourier series expansion of a periodic signal $x(t)$ expresses it as a linear combination of an infinite number of complex exponentials with frequencies 0, $\pm f_1$, $\pm f_2$, $\pm f_3$, etc. Thus, it involves terms representing the D.C. component (zero frequency), the fundamental frequency component, the second harmonic frequency component and the other harmonic frequency components. That is, it consists of all the frequency components present in $x(t)$ and hence is called its 'spectrum' ; ***and this spectrum of the periodic signal, x(t), is a discrete spectrum, as it contains only certain discrete frequencies the zero frequency, the fundamental frequency f_0, and the other harmonic frequencies***.

In Eq. (2.33), if we substitute for c_n using Eq. (2.34), we get

$$x(t)=\sum_{n=-\infty}^{\infty}|c_n|e^{j(2\pi nf_0t+\theta_n)};\quad f_0\triangleq\frac{1}{T};\quad -\infty<t<\infty \tag{2.35}$$

Thus, $|c_n|$ represents the magnitude of the complex-exponential having a frequency of nf_0, while θ_n represents its initial phase (corresponding to $t = 0$). That is why, a plot of $|c_n|$ vs n (or nf_0) is called the magnitude spectrum of $x(t)$, while a plot of θ_n vs n (or nf_0) is called the phase spectrum of $x(t)$.

Thus, the spectrum of a continuous-time periodic signal is a discrete one.

Example 2.14 Determine the complex-exponential Fourier series expansion of the periodic signal shown in Fig. 2.21.

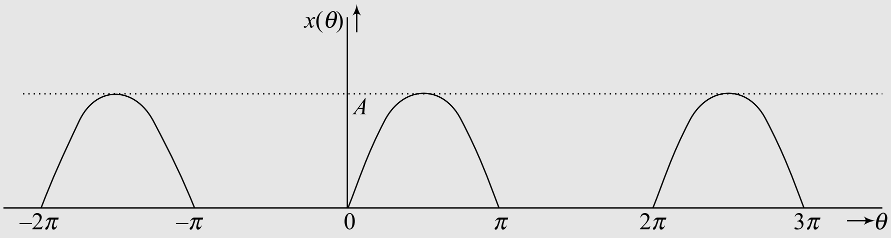


Fig. 2.21

Solution

$$x(\theta)=\begin{cases}A\sin\theta; & 0\le\theta\le\pi\\ 0; & \pi\le\theta\le 2\pi\end{cases}$$

The complex-exponential Fourier series expansion for the given signal may be written as

$$x(\theta)=\sum_{n=-\infty}^{\infty}c_ne^{jn\theta};\quad -\infty<\theta<\infty$$

where $$c_n=\frac{A}{2\pi}\int_0^{\pi}\sin\theta e^{-jn\theta}d\theta$$

Since $$\sin\theta=\frac{e^{j\theta}-e^{-j\theta}}{2j},\text{ we have}$$

$$c_n=\frac{A}{2\pi}\int_0^{\pi}\left(\frac{e^{j\theta}-e^{-j\theta}}{2j}\right)e^{-jn\theta}d\theta=\frac{A}{4\pi j}\int_0^{\pi}(e^{j(\theta-n\theta)}-e^{-j(\theta+n\theta)})d\theta$$

$$=\frac{A}{4\pi j}\left[\frac{e^{j(\theta-n\theta)}}{j(1-n)}+\frac{e^{-j(\theta+n\theta)}}{j(1+n)}\right]\Bigg|_0^{\pi}$$

$$= \frac{-A}{4\pi(1-n)} e^{j(1-n)\theta} \Big|_0^\pi + \frac{-A}{4\pi(1+n)} e^{-j(1+n)\theta} \Big|_0^\pi$$

$$= \frac{-A}{4\pi(1-n)} e^{j\pi(1-n)} + \frac{A}{4\pi(1-n)} - \frac{A}{4\pi(1+n)} e^{-j\pi(1+n)} + \frac{A}{4\pi(1+n)}$$

$$= \frac{A}{4\pi(n-1)}[e^{j\pi(1-n)} - 1] - \frac{A}{4\pi(1+n)}[e^{-j\pi(1+n)} - 1]$$

for n odd
$n \neq 1$ or -1 : $(1 - n)$ and $(1 + n)$ will both be even and hence $e^{-j\pi(1-n)}$ and $e^{-j\pi(1+n)}$ will both be equal to 1.

$\therefore$
$$\underset{n\,\text{odd}}{c_n} = 0$$

For n even: $(1 - n)$ and $(1 + n)$ will both be odd and hence $e^{j\pi(1-n)}$ and $e^{-j\pi(1+n)}$ will both be equal to –1.

$\therefore$
$$\underset{n\,\text{even}}{c_n} = \frac{-2A}{4\pi(n-1)} + \frac{2A}{4\pi(1+n)} = \frac{A}{\pi(1-n^2)}$$

For $n = 0$;
$$c_0 = \frac{A}{\pi}$$

For $n = 1$ the second term reduces to zero but the first term takes the form of zero divided by zero. Hence applying L'Hospital's rule to the first term, we get

$$c_1 = \frac{A}{4j}$$

For $n = -1$; the first term takes the value zero but the second term takes the form of zero by zero. Applying L'Hospital's rule to the second term, we get

$$c_{-1} = -\frac{A}{4j}$$

Hence, the complex exponential Fourier series expansion of the given waveform is

$$x(\theta) = \frac{A}{2}\sin\theta - \sum_{\substack{n=-\infty \\ n\,\text{even}}}^{\infty} \frac{A}{\pi(1-n^2)} e^{jn\theta} \quad]$$

Example 2.15 $x(t)$ is a periodic signal as shown in Fig. 2.22. Find its complex-exponential Fourier series expansion and plot its magnitude and phase spectra.

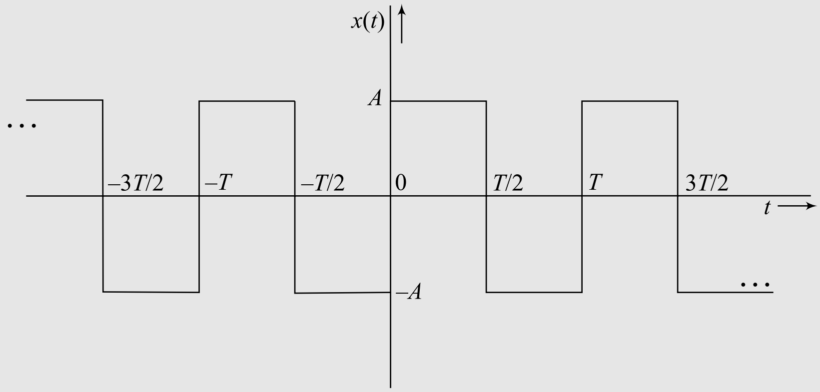


Fig. 2.22 *Signal of Example 2.15*

Solution $$c_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)dt = 0$$

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2\pi nf_0 t}\,dt = \frac{1}{T}\int_{-T/2}^{0} (-A)e^{-j2\pi nf_0 t}\,dt + \frac{1}{T}\int_{0}^{T/2} (A)e^{-j2\pi nf_0 t}\,dt$$

$$= \frac{-A}{T}\int_{-T/2}^{0} e^{-j2\pi nf_0 t}\,dt + \frac{A}{T}\int_{0}^{T/2} e^{-j2\pi nf_0 t}\,dt$$

$$= \frac{-A}{T}\frac{e^{-j2\pi nf_0 t}}{-j2\pi nf_0}\Bigg|_{-T/2}^{0} + \frac{A}{T}\frac{e^{-j2\pi nf_0 t}}{-j2\pi nf_0}\Bigg|_{0}^{T/2} = \frac{A}{j2\pi n}[1 - e^{j\pi n}] + \frac{A}{j2\pi n}[1 - e^{-j\pi n}]$$

$$= \frac{A}{j2\pi n}[2 - (e^{j\pi n} + e^{-j\pi n})] = \frac{A}{j\pi n}[1 - \cos \pi n]$$

But $$\cos n\pi = \begin{cases} 1 & \text{for } n \text{ even} \\ -1 & \text{for } n \text{ odd} \end{cases}$$

$\therefore$ $$c_n = \begin{cases} \dfrac{2A}{j\pi n} & \text{for } n \text{ odd} \\ 0 & \text{for } n \text{ even} \end{cases}$$

For the purpose of plotting the magnitude and phase spectra of $x(t)$, we shall assume $A = \pi$. The magnitude and phase spectra are plotted in Figs. 2.23 and 2.24 respectively.

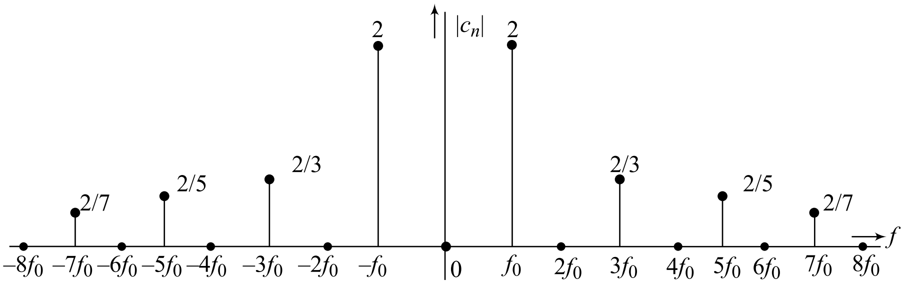


Fig 2.23 *Magnitude spectrum*

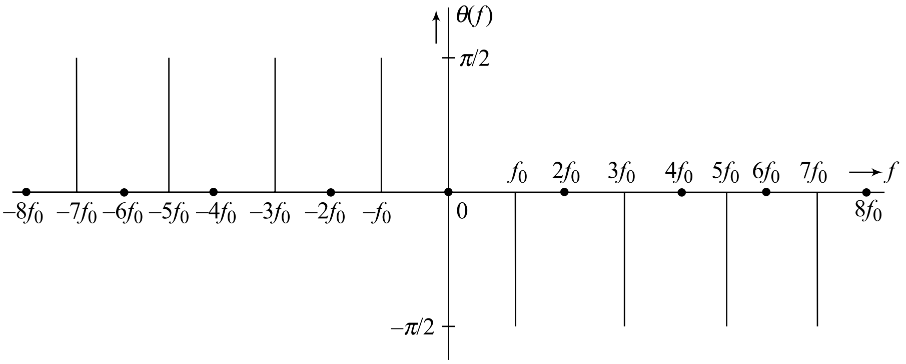


Fig. 2.24 *Phase spectrum*

2.5.1 Properties of Complex-Exponential Fourier Series Coefficients

We now give a list of important theorems and properties of the complex-exponential Fourier series coefficients (CEFSCs), c_ns. The reader is expected to provide the proof.

1. If $x(t)$ and $y(t)$ are two periodic signals with the same fundamental period, and if their CEFSCs are represented respectively by c_n^x and c_n^y, then the signal $z(t) = ax(t) + by(t)$ will have CEFSC's given by

$$\boxed{c_n^z = ac_n^x + bc_n^y} \tag{2.36}$$

This is called the linearity theorem.

2. If $x(t)$ is a periodic signal with fundamental frequency f_0 and if $y(t) \underline{\Delta} x(t-t_0)$, then

$$\boxed{c_n^y = e^{-j2\pi nf_0t_0}c_n^x} \tag{2.37}$$

This is called the time-shift theorem.

3. If $x(t)$ is a periodic signal with fundamental frequency f_0 and if $y(t) = e^{-j2\pi kf_0t}\cdot x(t)$, then

$$\boxed{c_n^y = c_{n-k}^x} \tag{2.38}$$

This is called the frequency-shift theorem.

4. $x(t)$ and $y(t)$ are periodic signals with the same fundamental period T, and if

$z(t) = x(t) * y(t) = \int\limits_{t_0}^{t_0+T} x(t-\tau)y(\tau)d\tau$, then

$$\boxed{c_n^z = T\cdot c_n^x\cdot c_n^y} \tag{2.39}$$

This is called the circular convolution theorem.

5. If $x(t)$ and $y(t)$ are periodic signals with the same fundamental period T, and if $z(t) = x(t)\cdot y(t)$, then

$$\boxed{c_n^z = c_n^x * c_n^y = \sum_{k=-\infty}^{\infty} c_k^x\cdot c_{n-k}^y} \tag{2.40}$$

This is called the multiplication theorem, or the modulation theorem.

6. If $x(t)$ is a periodic signal with fundamental frequency f_0 and if $y(t) = \frac{d}{dt}x(t)$, then

$$\boxed{c_n^y = j2\pi nf_0c_n^x} \tag{2.41}$$

This is known as the differentiation theorem

7. If $x(t)$ is a periodic signal with fundamental frequency f_0 and if $y(t) = x(at)$ where α is a non-zero real number, i.e., if $y(t)$ is a time-scaled version of $x(t)$, then

$$c_n^y = c_n^x \tag{2.42}$$

This is known as the scaling theorem. The above result implies that while the spacing between the spectral components is changed (i.e., it is now af_0 instead of f_0), the amplitudes of these spectral components remain unchanged.

8. Let $x(t)$ be a periodic signal with c_ns as its CEFSCs. Then

$$\text{Average power of } x(t) = \frac{1}{T}\int\limits_{-T/2}^{T/2} |x(t)|^2\,dt = \sum_{n=-\infty}^{\infty} |c_n|^2 \tag{2.43}$$

This is called the Parseval's theorem pertaining to the complex-exponential Fourier series. Since the average power of $c_ne^{j2\pi nf_0t}$ is equal to $|c_n|^2$, Eq. (2.43) merely states that the average power of a periodic signal is equal to the sum of the average powers of its orthogonal components.

9. If a periodic signal $x(t)$ is real-valued and its CEFSCs are represented by c_ns, then

$$\boxed{c_{-n} = c_n^*, \text{ if } x(t) \text{ is real valued}} \tag{2.44}$$

where the * indicates complex-conjugation.

10. If a periodic signal $x(t)$ with c_ns, as its CEFSCs is real-valued and even with respect to t, then c_ns are also real and are even with respect to n.

$$\boxed{c_n\text{s are real} \quad c_{-n} = c_n, \text{ if } x(t) \text{ is real an even}} \tag{2.45}$$

11. If a periodic signal $x(t)$ with c_ns, as its CEFSCs, is real-valued and has odd symmetry with respect to t, then c_ns are purely imaginary and have odd symmetry with respect to n.

$$\boxed{c_n\text{s purely imaginary} \quad c_{-n} = c_n^* = -c_n, \text{ if } x(t) \text{ is real and odd}} \tag{2.46}$$

2.6 TRIGONOMETRIC FOURIER SERIES

The expansion of a signal $x(t)$ using the complete set of orthonormal functions

$$\frac{1}{\sqrt{T}}, \sqrt{\frac{2}{T}}\cos\omega_0 t, \sqrt{\frac{2}{T}}\cos 2\omega_0 t, \ldots$$

$$\sqrt{\frac{2}{T}}\sin\omega_0 t, \sqrt{\frac{2}{T}}\sin 2\omega_0 t, \ldots \text{ with } \omega_0 \,\Delta\, \frac{2\pi}{T},$$

is referred to as the *trigonometric Fourier series expansion* of $x(t)$.

Before proceeding further, we shall first show that the above set is an orthonormal set over the interval $-\frac{T}{2}$ to $+\frac{T}{2}$. For this, we make use of Eq. (2.18).

1. First we shall show that all these functions have unit norm, i.e., that they have been normalized.

$$\int_{-T/2}^{T/2}\left(\frac{1}{\sqrt{T}}\cdot\frac{1}{\sqrt{T}}\right)dt = 1; \int_{-T/2}^{T/2}(\sqrt{2/T}\cos n\omega_0 t)^2\,dt = \frac{2}{T}\int_{-T/2}^{T/2}\left(\frac{1+\cos 2n\omega_0 t}{2}\right)dt$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} 1dt + \frac{1}{T}\int_{-T/2}^{T/2}\cos 2n\omega_0 t dt = 1 \quad \text{as the second integral is zero}$$

Further, $\displaystyle\int_{-T/2}^{T/2}(\sqrt{2/T}\sin n\omega_0 t)^2\,dt = 1$

Thus, we find that all these functions have been normalized.

2. We will now show that any two distinct functions in the above set are orthogonal to each other. For this, using Eq. (2.18), we find that

$$\int_{-T/2}^{T/2}\left(\frac{1}{\sqrt{T}}\sqrt{\frac{2}{T}}\cos n\omega_0 t\right)dt = \int_{-T/2}^{T/2}\left(\sqrt{\frac{1}{T}}\cos m\omega_0 t\right)\left(\sqrt{\frac{2}{T}}\cos n\omega_0 t\right)dt = 0$$

Also, $m \neq n$

$$\int_{-T/2}^{T/2}\left(\sqrt{\frac{2}{T}}\cos m\omega_0 t\right)\left(\sqrt{\frac{2}{T}}\sin n\omega_0 t\right)dt = 0 \quad \text{for any integer values of } m \;\&\; n.$$

and $\displaystyle\int_{-T/2}^{T/2}\left(\sqrt{\frac{2}{T}}\sin m\omega_0 t\right)\left(\sqrt{\frac{2}{T}}\sin n\omega_0 t\right)dt = 0 \quad \text{if } m \neq n$

Thus, the above set of functions is a set of mutually orthogonal functions. Further, since it is a complete set, we should be able to express any function $x(t)$ as a linear combination of these orthonormal functions. We may therefore write

$$x(t) = \alpha_0\left(\sqrt{\frac{1}{T}}\right) + \sum_{n=1}^{\infty} \alpha_n\left(\sqrt{\frac{2}{T}}\cos n\omega_0 t\right) + \sum_{n=1}^{\infty} \beta_n\left(\sqrt{\frac{2}{T}}\sin n\omega_0 t\right); -\frac{T}{2} \le t \le +\frac{T}{2}$$

and where $\omega_0(t) \triangleq \frac{2\pi}{T}$ (2.47)

Making use of Eq. (2.18) and noting that all the functions of the set are normalized

i.e., $$\int_{-T/2}^{T/2} |f_i(t)|^2\, dt = 1$$

We have

$$\alpha_0 = \int_{-T/2}^{T/2} x(t)\frac{1}{\sqrt{T}}\,dt = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} x(t)\,dt \tag{2.48}$$

$$\alpha_n = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt \tag{2.49}$$

and $$\beta_n = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt \tag{2.50}$$

If we now define

$$\alpha_0\frac{1}{\sqrt{T}} \triangleq a_0 \quad \text{then} \quad a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt \tag{2.51}$$

$$\alpha_n\sqrt{\frac{2}{T}} \triangleq a_n, \quad \text{then} \quad a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt \tag{2.52}$$

and $$\beta_n\sqrt{\frac{2}{T}} \triangleq b_n, \quad \text{then} \quad b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt \tag{2.53}$$

Using the above equations, Eq. (2.47) may now be written as

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n\cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t; -\frac{T}{2} \le t \le +\frac{T}{2}; \omega_0 \triangleq \frac{2\pi}{T} \tag{2.54}$$

If, however, $x(t)$ is periodic with a period T, then the above expansion of $x(t)$ is valid for all time, so that we may write

$$\boxed{\underset{\text{periodic}}{x(t)} = a_0 + \sum_{n=1}^{\infty} a_n\cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t; -\infty \le t \le \infty\,; \omega_0 \triangleq \frac{2\pi}{T}} \tag{2.55}$$

where

$$\begin{aligned} a_0 &= \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt \\ a_n &= \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt \\ b_n &= \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt \end{aligned} \tag{2.56}$$

2.6.1 Trigonometric and Complex-Exponential Fourier Series

The trigonometric Fourier series and the complex-exponential Fourier series are related.
For the CEFS, we had

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} = c_0 + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} c_n e^{jn\omega_0 t}$$

In the above equation, if we put

$$c_0 = a_0; \quad c_n = \frac{1}{2}(a_n - jb_n) \text{ and } c_{-n} = \frac{1}{2}(a_n + jb_n), \tag{2.57}$$

and simplify, we get

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

where

$$a_0 = c_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{jn\omega_0 t}\,dt\Big|_{n=0} = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt$$

$$a_n = c_n + c_{-n} = \frac{1}{T}\left[\int_{-T/2}^{T/2} x(t)e^{-jn\omega_0 t}\,dt + \int_{-T/2}^{T/2} x(t)e^{jn\omega_0 t}\,dt\right]$$

$$= \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt$$

$$b_n = [c_{-n} - c_n]\frac{1}{j} = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt$$

2.6.2 Symmetries of $x(t)$ and Computation of Fourier Series

When the periodic signal, $x(t)$, possesses certain symmetries, the computation of its Fourier series coefficients gets considerably simplified, as stated below.

1. The trigonometric Fourier series of a periodic signal $x(t)$ with even symmetry will consist only of cosinusoids, i.e., $b_n = 0$ for all n.
2. The trigonometric Fourier series of a periodic signal $x(t)$ with odd symmetry will consist only of sinusoids, i.e., $a_n = 0$ for all n.
3. A periodic function $x(t)$ with period T is said to be having rotational, or half-wave symmetry, if

   $x(t \pm T/2) = -x(t)$ for all t.

   All periodic signals with half-wave symmetry will have only odd harmonic components in their Fourier series expansion. (Prove this)

2.6.3 DIRICHLET's Conditions for Existence and Convergence of Fourier Series

From our discussion so far on Fourier series, it might appear that every periodic function can be expanded in the form of a Fourier series. However, this is not true.

We say that for a given $x(t)$, a Fourier series exists provided c_n is finite for all n, i.e., $|c_n| < \infty$.

Since

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-jn\omega_0 t}\,dt,$$

it follows that for the Fourier series to exist, $x(t)$ must satisfy the condition

$$\int_{-T/2}^{T/2} |x(t)|dt < \infty \tag{2.58}$$

The above condition for the existence of Fourier series is called the *Weak Dirichlet's Condition*.

It should however, be noted that existence of Fourier series does not guarantee their convergence at all points and that for convergence, the following conditions, known as *strong Dirichlet's conditions* must be satisfied:

1. $x(t)$ must be finite at all points.
2. $x(t)$ must have only a finite number of maxima and minima in one period.
3. $x(t)$ can have only a finite number of discontinuities and the discontinuities, if any, must be finite discontinuities.

Example 2.16 Find the trigonometric and complex-exponential Fourier series of the periodic signal shown in Fig 2.25.

Solution Here, $x(t) = At;\ 0 \le t \le 1$

(a) *Trigonometric Fourier series*

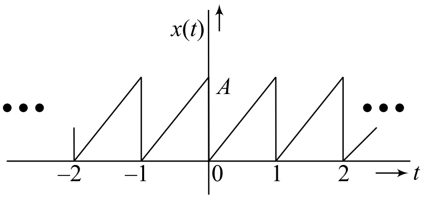


Fig. 2.25

$$\therefore \qquad a_0 = \frac{A}{1}\int_0^1 tdt = \frac{A}{2}$$

$$a_n = 2A\int_0^1 t\cos 2\pi ntdt = 0$$

$$b_n = 2A\int_0^1 t\sin 2\pi ntdt = \frac{-A}{\pi n}$$

(b) *Complex-exponential Fourier series*

$$\underset{n\neq 0}{c_n} = \frac{A}{1}\int_0^{T=1} te^{-j2\pi nt}dt = \frac{jA}{2\pi n}$$

$$c_0 = \frac{A}{1}\int_0^1 tdt = \frac{A}{2}$$

Example 2.17 For the periodic signal shown in Fig. 2.26, determine the **(a)** complex-exponential, and **(b)** trigonometric Fourier series.

Solution

(a) $x(\theta) = A\sin\theta;\ 0 \le \theta \le \pi$

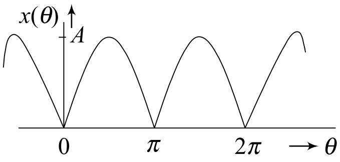


Fig. 2.26

$$\therefore \qquad x(\theta) = \sum_{n=-\infty}^{+\infty} c_n e^{jn\theta};\ -\infty < \theta < \infty$$

and $$c_n = \frac{A}{\pi}\int_0^{\pi} \sin\theta e^{-jn\theta}d\theta = \frac{A}{\pi}\int_0^{\pi}\frac{1}{j2}(e^{j\theta} - e^{-j\theta})e^{-jn\theta}d\theta$$

On simplification, this gives

$$c_n = \begin{cases} \dfrac{2A}{\pi(1-n^2)}; & n \text{ even} \\ 0 \qquad ; & n \text{ odd} \end{cases}$$

$$\therefore \qquad x(\theta) = c_0 + \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} c_n e^{jn\theta} = \frac{2A}{\pi} - \sum_{\substack{n=2 \\ (n\ \text{even})}}^{\infty} \frac{4A}{\pi(n^2-1)} \cos n\theta$$

$$\therefore \qquad x(\theta) = \frac{2A}{\pi} - \sum_{n=1}^{\infty} \frac{4A}{\pi(4n^2-1)} \cos 2n\theta$$

(b) *To find the trigonometric Fourier series*

$$x(\theta) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\theta + \sum_{n=1}^{\infty} b_n \sin n\theta$$

since $x(\theta)$ has even symmetry, $b_n = 0$ for all n.

Now, $$a_0 = \frac{A}{\pi}\int_0^{\pi} \sin\theta d\theta = \frac{2A}{\pi}$$

$$a_n = \frac{2}{\pi}\int_0^{\pi} x(\theta)\cos n\theta d\theta = \frac{2A}{\pi}\int_0^{\pi} \sin\theta \cos n\theta d\theta$$

Using the identity: $\sin\theta \cdot \cos n\theta = \frac{1}{2}[\sin(\theta + n\theta) + \sin(\theta - n\theta)]$

And by simplifying, we get

$$a_n = \begin{cases} \dfrac{4A}{\pi(1-n^2)} & ; \quad n\ \text{even} \\ 0 & ; \quad n\ \text{odd} \end{cases}$$

$$\therefore \qquad x(\theta) = \frac{2A}{\pi} + \sum_{n=1}^{\infty} a_{2n} \cos 2n\theta$$

$$= \frac{2A}{\pi} + \sum_{n=1}^{\infty} \frac{4A}{\pi(1-4n^2)} \cos 2n\theta$$

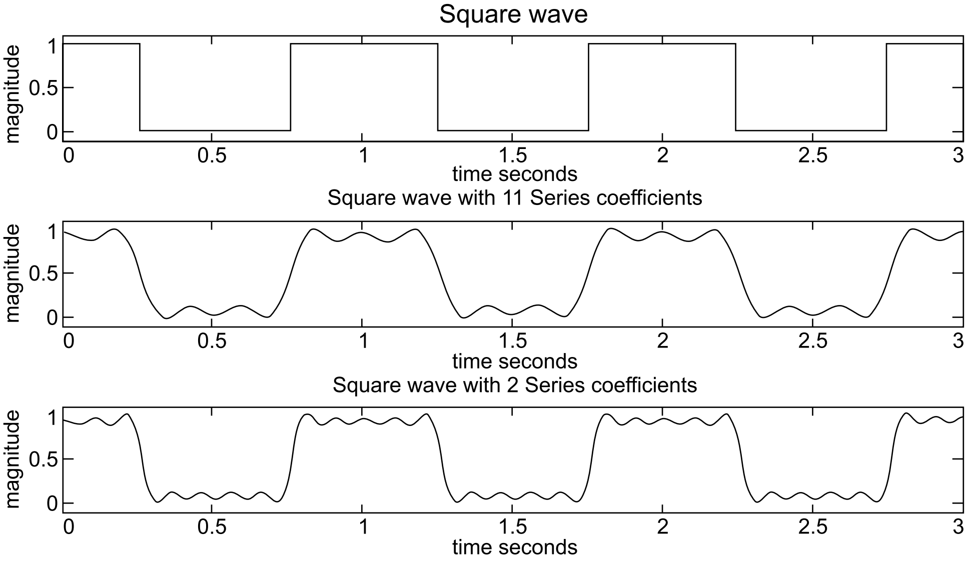
MATLAB Example 2.1 In this example, we study GIBB's phenomenon. For this, we plot the continuous-time Fourier series of a square wave and study the effect of truncation of Fourier series coefficients on the signal that is generated.

```
T=1; %We start with a square wave of time period 1 second
T=linespace(0, 3, 3000);  % We oversample by 1000 with 3 time periods
x = [ones (1, 250), zeros (1, 500) ones (1, 250)]; %The square wave
xp = kron (ones(1, 3), x); %The signal is repeated
subplot (311), plot(t, xp, ‘k’);
axis ([0, 3, -0.1, 1.1]);
title (‘Square wave’); xlabel (‘time, seconds’);
ylabel (‘magnitude’);

%First plot using only 11 Fourier Series Coefficients
xs1=zeros (1, 1000);
for k = -5:5
ek = cos (2*pi*k*t(1:1000)) + sqrt (-1) *sin (2*pi*k*t(1:1000));
xs1 = xs1+0.5*sinc (k/2)*ek;
end
```

```
subplot (312), plot (t, kron (ones (1, 3), real (xs1)), 'k');
title ('Square wave with 11 Series
Coefficients');xlabel ('time, seconds');
ylabel ('magnitude');
axis ([0, 3, -0.2, 1.2]);

%Repeat using 21 Fourier Series coefficients
xs1=zeros (1, 1000);
for k=-10 : 10
ek = cos (2*pi*k*t (1:1000)) + sqrt (-1)*sin (2*pi*k*t(1:1000));
xs1 = xs1+0.5*sinc (k/2) *ek;
end
subplot (313), plot (t, kron (ones (1, 3), real (xs1)), 'k');
title ('Square wave with 21 Series
coefficients');xlabel ('time, seconds');
ylabel ('magnitude');
axis ([0,3, - 0.2, 1.2]);
```



The figure shows the effect of truncation of Fourier Series on a square wave. The square wave (top) is plotted using only the coefficients for $|k| < 6$ (middle) and $|k| < 11$ (bottom). Note that the ripples do not decrease in magnitude, but only increase in frequency.

Summary

- A signal is a single-valued function of one or more variables and carries some information.
- A continuous-time signal is one whose value is defined at all instants of time. For example, a sine wave.
- A discrete-time signal is one whose values are defined only at a discrete set of points in time. For example, a sequence of numbers representing the temperature at a fixed time, taken on a daily basis.
- A signal $x(t)$ is said to be periodic in time with a period T if $x(t + mT) = x(t)$ for any t and any integer m.
- A signal whose total energy is finite and non-zero, is called an energy signal. For example, a rectangular pulse of finite duration: $x(t) = Ae^{-|t|/T}$

- A signal whose average power is finite and non-zero, is called a power signal. For example, a sine wave.
- (a) A unit impulse function is denoted by $\delta(t)$ and is defined by the following:

$$\int_{t_1}^{t_2} x(t)\delta(t)dt = \begin{cases} x(0) & \text{if } t_1 < 0 < t_2 \\ 0 & \text{otherwise} \end{cases}$$

  and where $x(t)$ is any function which is continuous at least at $t = 0$.

  (b) *Properties*:

  (i) Area under a unit impulse function is one.

  (ii) Width (along the time axis) of an impulse function is zero.

  (iii) If $x(t)$ is continuous at $t = \tau$, then

  $$x(t)\delta(t-\tau) = x(\tau)\delta(t-\tau) \text{ (Sampling property)}$$

  and $$x(t) * \delta(t-\tau) = x(t-\tau) \text{ (Replication property)}$$

- $u(t) = \int_{-\infty}^{t} \delta(\lambda)d\lambda$ and $\frac{d}{dt}u(t) = \delta(t)$
- A vector space is a set of vectors satisfying certain conditions, the most important of which are:
  - (a) When two vectors of the set are added, the resultant vector must again belong to the same set (closure property with respect to addition)
  - (b) Zero vector must belong to the set (Existence of zero vector)
  - (c) For every vector belonging to the set, the additive inverse must exist in the set.
  - (d) When any vector belonging to the set is multiplied by a scalar, the resultant vector must again belong to the same set.
- A signal space can be defined exactly on the same lines.
- A set of signals is said to be linearly independent if no signal of the set can be expressed as a linear combination of the rest.
- A basis set of signals for a signal space is a set of linearly independent signals which can, by various linear combinations, generate that signal space, i.e., every signal of the signal space must be capable of being expressed as a linear combination of the linearly independent basis signals.
- A basis set is not unique for a given signal space; i.e., for the same signal space there can be a number of basis signal sets.
- The number of signals in any basis set of a signal space is referred to as the dimension of that signal space.
- Just as two vectors of the Euclidian space are said to be orthogonal if their dot product is zero, we say that two signals are orthogonal if their *inner product* is zero.
- For the signal space S of all real-valued continuous-time signals having a finite energy over the interval $[0, T]$, a convenient inner product is

$$(x(t), y(t)) \underline{\Delta} \int_0^T x(t)y(t)dt$$

- The inner product of any signal $\in S$ with itself is called the norm-square of the signal and is equal to the energy of the signal over the interval $[0, T]$. Norm of $x(t)$ is

$$\|x(t)\| = \sqrt{\text{Energy of the signal}}$$

- A set of signals, $s_i(t), i = 1, 2, \ldots, n$ are said to be forming an orthonormal set of signals if $(s_i(t), s_j(t)) = \begin{cases} 1 & \text{for } j = i \\ 0 & \text{otherwise} \end{cases}$
- The component of a signal $s_1(t)$ along another signal $s_2(t)$, is given by

$$\left[\frac{(s_1(t), s_2(t))}{\|s_2(t)\|^2}\right] s_2(t)$$

- Given m signals, Gram–Schmidt orthogonalization procedure enables us to derive n orthonormal signals from them with $n \le m$.

- Schwarz's inequality

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)dt\right|^2 \le \left[\int_{-\infty}^{\infty} |s_1(t)|^2 dt\right]\left[\int_{-\infty}^{\infty} |s_2(t)|^2 dt\right]$$

  where the equality sign holds *if and only if* $s_2(t) = cs_1^*(t)$

- If $x(t)$ is a periodic signal with a period $T = 1/f_0$, then $x(t)$ can be written as

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi nf_0 t}; -\infty < t < \infty$$

  where c_ns are called the complex-exponential Fourier series coefficients and are given by

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2\pi nf_0 t}\, dt$$

Note *c_ns are in general complex numbers, even if $x(t)$ is a real-valued function.*

- (a) If $c_n = |c_n| e^{j\theta_n}$, a plot of $|c_n|$ vs. n(or nf_0) is called the magnitude spectrum of $x(t)$ and a plot of θ_n vs. n (or nf_0) is called the phase spectrum of $x(t)$.

  (b) The magnitude spectrum as well as the phase spectrum of a periodic continuous-time signal are discrete.

- (a) If $x(t)$ is periodic in t with a period $\tau = 1/f_0$, then

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t; -\infty \le t \le \infty$$

  where

$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)dt; \quad a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 tdt$$

  and

$$b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 tdt$$

  a_0, a_ns and b_ns are called Trigonometric Fourier series coefficients of $x(t)$.

  (b) For an $x(t)$ which has even symmetry, all b_ns are zero. For an $x(t)$ which has odd symmetry, all a_ns are zero. For an $x(t)$ which is symmetric about the time axis, $a_0 = 0$.

- (a) *Weak Dirichlet's condition*: For Fourier series to exist, a periodic function with period T must satisfy the condition

$$\int_{-T/2}^{T/2} |x(t)|dt < \infty$$

  (b) *Strong Dirichlet's condition*: The following conditions must be satisfied for the Fourier series of a periodic function $x(t)$ to converge:

    (i) $x(t)$ must be finite at all points.

    (ii) $x(t)$ must have a finite number of maximum and minimum in one period.

    (iii) $x(t)$ can have only a finite number of discontinuities and the discontinuities, if any, must be finite discontinuities.

References and Suggested Reading

1. Haykin, Simon, Barry Van Veen, *Signals and Systems*, 2nd Edition, John Wiley & Sons (Asia) Pvt. Ltd. 2004.
2. Lathi, B.P., *Signals, Systems and Communication*, B.S. Publications, 2001.

3. Roberts, M.J., *Signals and Systems*, Tata McGraw-Hill, 2003.
4. Rao, P. Ramakrishna, *Signals and Systems*, Tata McGraw-Hill, 2008.
5. Ziemer, R., W. Tranter, and D. Fannen, *Signals and Systems: Continuous and Discrete*, Prentice-Hall, 1998.

Review Questions

1. Define and give an example for each of the following
(a) Continuous-time signals, and (b) Discrete-time signals

2. Define and give an example for each of the following
(a) Energy signals, and (b) Power signals

3. Determine the values of the following integrals:

(a) $\int\limits_{t=-3}^{t=3} \delta(t-4)dt$, and (b) $\int\limits_{t=-2}^{t=2} e^{-5t}\delta(t-1)dt$

4. Define signal space. Give an example of a signal space.

5. Explain what is meant by a linearly independent set of signals.

6. Define the term 'basis set' for a signal space.

7. Define dimension of signal space.

8. How is orthogonality of two signals defined?

9. Explain the term 'norm of a signal'. What is its physical significance?

10. How do you find the distance between the signals $x(t)$ and $y(t)$ belonging to a certain signal space?

11. Explain the basic principle of Gram–Schmidt's orthogonalization procedure.

12. When do you say that the two signals $x(t)$ and $y(t)$ are orthogonal and/or orthonormal?

13. What is meant by a complete set of orthonormal functions. Give an example of such a set of functions.

14. State and explain Dirichlet's conditions for convergence of Fourier series.

15. Write down the complex-exponential Fourier series expansion of the signal $x(t) = 5 \cos 10\pi t$

Problems

1. Determine whether the following continuous-time signals are periodic or aperiodic. If they are periodic, determine their fundamental period.
(a) $x(t) = \cos 3t$, (b) $x(t) = e^{j\omega_0 t}$, (c) $x(t) = \cos^2 10\pi t$
(d) $x(t) = \sin^2 100\pi t + \sin 200\pi t$, (e) $x(t) = \cos tu(t)$, and (*f*) $x(t) = \sin 3t + \cos \pi t$

2. If $x(t)$ is as shown in Fig. P2.2, sketch and label each of the following signals:
(a) $x(t-3)$, (b) $x(2t)$, (c) $x(t/2)$,
(d) $x(-2t)$, and (e) $x(3t-2)$

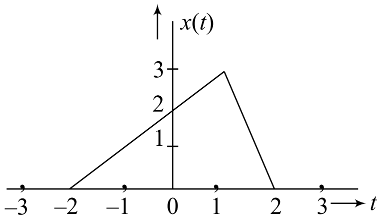


Fig. P2.2

3. Which of the following signals are power signals, and which of them are energy signals? Are there any signals which are neither power signals nor energy signals? Justify your answer in either case. For power/energy signals, find the average power or the total energy, whichever is appropriate.
(a) $(2 - e^{-3t})u(t)$, (b) $e^{j\omega_0 t}$, (c) $u(t-2) - u(t-4)$,
(d) $e^{-2t}\, u(t)$, (e) e^{-5t}, (f) $e^{-|t|}\Pi\left(\dfrac{(t+1)}{6}\right)$, and (g) $te^{-2t}u(t-1)$

4. The signal $x(t)$ given by

$$x(t) = \begin{cases} \dfrac{1}{2}[\cos \omega t + 1]; & -\pi \le \omega t \le \pi \\ 0 & ;\ \text{otherwise} \end{cases}$$

is called the raised cosine pulse, and is sketched in Fig. P2.4. Determine the total energy of this signal.

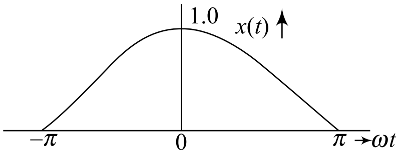


Fig. P2.4

5. For the signal $x(t)$ shown in Fig P2.5(a), determine the following using

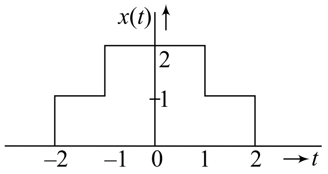


Fig. P2.5(a)

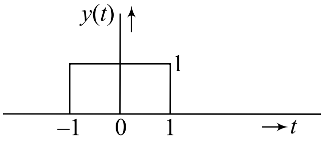


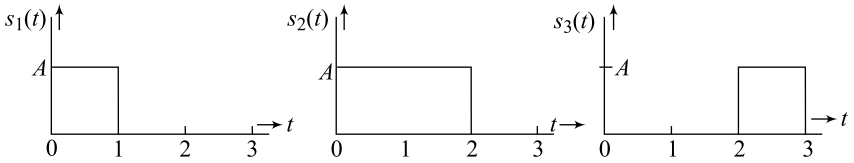
Fig. P2.5(b)

(a) a representation in terms of shifted versions of $u(t)$.
(b) a representation in terms of the rectangular pulse $y(t)$ and its scaled and shifted versions

6. Do the following vectors form a basis for R^3, the Euclidian space?
$\boldsymbol{X} = (1, 1, 0)$; $\boldsymbol{Y} = (3, 0, 1)$; $\boldsymbol{Z} = (5, 2, 1)$

7. Show that the vectors $\boldsymbol{X} = (1, 1, 0, 0)$; $\boldsymbol{Y} = (0, 0, 1, 1)$; $\boldsymbol{Z} = (1, 0, 0, 4)$ and $\boldsymbol{W} = (0, 0, 0, 2)$ form a basis for R^4. Find the coordinates of each of the standard basis vectors of R^4 in the ordered basis set ($\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{W}$).

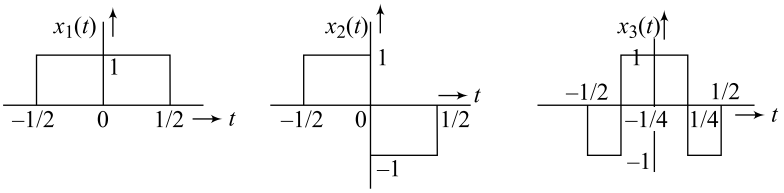
8. Signals $s_1(t)$, $s_2(t)$ and $s_3(t)$ are as shown in Fig. P2.8. Using Gram–Schmidt procedure, derive from them an orthonormal basis set of signals for the signal space spanned by $s_1(t)$ and $s_2(t)$ and $s_3(t)$.



(a) (b) (c)

Fig. P.2.8

9. Signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ are shown in Fig. P2.9.



(a) (b) (c)

Fig. P2.9

(a) Show that the signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ are orthogonal over the interval [−1/2, 1/2]
(b) If the signal $x(t) = 2 \sin 2\pi t$ is expanded in terms of these functions, find that representation. What is the integral squared error of this approximate representation of $2 \sin 2\pi t$?

10. $x_1(t), x_2(t), \ldots, x_n(t)$ are n mutually orthogonal signals defined over the interval $(-T, T)$. If a signal $y(t)$ is defined as

$$y(t) = \sum_{i=1}^{n} x_i(t)$$

show that the energy of the signal $y(t)$ over $(-T, T)$ is equal to the sum of the energies of $x_i(t)$s from $i = 1$ to n.

11. A signal $x(t)$ is as shown in Fig. P.2.11.
Show that $x(t)$ is orthogonal to the signals $\cos t$, $\cos 2t$, $\cos 3t$, …, $\cos nt$ for all integer values of n, $n \neq 0$, over the interval $(0, 2\pi)$.

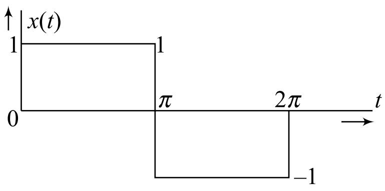


Fig. P2.11

12. Signal $x(t)=\begin{cases} t; & 0\le t\le 1 \\ 0; & \text{elsewhere}\end{cases}$

Expand $x(t)$ over the interval (0, 1) by

(a) Trigonometric Fourier series, and (b) Complex exponential Fourier series

13. Expand the periodic function $x(\theta)$ shown in Fig. P2.13 using trigonometric Fourier series.

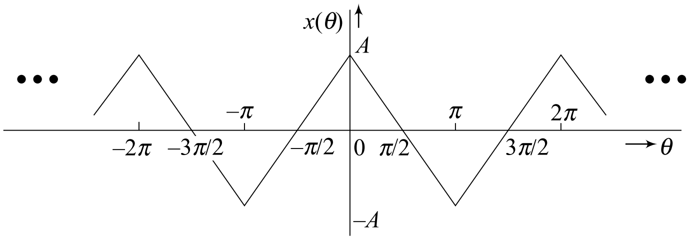


Fig. P2.13

14. Expand the periodic waveform $x(t)$ shown in Fig. P2.14 by complex exponential as well as trigonometric Fourier series.

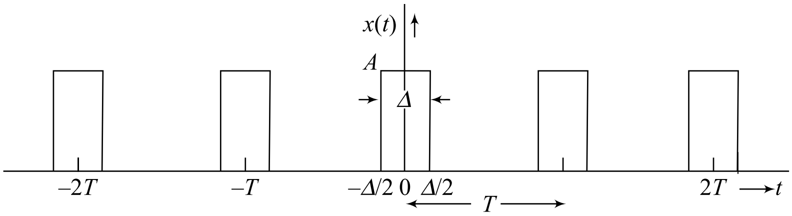


Fig. P2.14

15. For the periodic waveform shown in Fig. P2.15, determine the complex exponential and trigonometric Fourier series expansions.

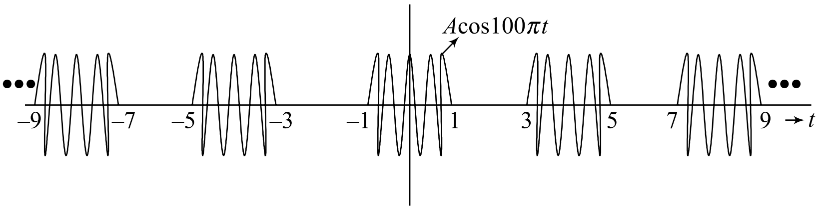


Fig. P2.15

16. Express the signal $x(t) = 2 + \sin\omega_0 t + 3\cos(\omega_0 t + \pi/4) + 2\cos 2\omega_0 t$ as the sum of complex exponentials and plot its magnitude and phase spectra.

17. (a) $x_n(t) = e^{j2\pi nt/T}$, where n takes all integer values from $-\infty$ to $+\infty$. Show that the functions $x_n(t)$s are orthogonal over any interval of T sec. Are they also orthonormal?

(b) Are the functions $\sin n\omega_0 t$ and $\cos m\omega_0 t$ orthogonal over the interval $(0, T)$, where $\omega_0 = 2\pi/T$? Are they orthonormal? If they are not, normalize them.

18. In Section 2.6, we stated that a periodic signal $x(t)$ having rotational, or half-wave symmetry will have only odd harmonics. Prove that statement. Also prove the converse of it, i.e., if a periodic signal $x(t)$ with period T has only odd harmonic components, then it has half-wave symmetry, so that $x(t - T/2) = -x(t)$ for any t.

Multiple-Choice Questions

1. The fundamental period T, of a periodic continuous-time signal $x(t)$, is

(a) the smallest positive constant satisfying the relation $x(t) = x(t + mT)$ for every t and any integer m
(b) the positive constant satisfying the relation $x(t) = x(t + mT)$ for every t and any integer m
(c) the largest positive constant satisfying the relation $x(t) = x(t + mT)$ for anY t and any integer m
(d) the smallest positive integer satisfying the relation $x(t) = x(t + mT)$ for any t and any m

2. The value of $\int_{-\pi/4}^{\pi/4} \cos \omega t \delta(\omega) d\omega$ is

(a) 0 (b) $\pi/2$ (c) $\sqrt{2}$ (d) 1

3. $e^{-t} u(t)$ is

(a) an energy signal (b) a power signal
(c) neither an energy signal nor a power signal (d) None of the above

4. $x(t) = \begin{cases} 1 & \text{for } 5 \le t \le 10 \\ -1 & \text{for } 10 \le t \le 15 \\ 0; & \text{otherwise} \end{cases}$. Then $x(t)$ can be expressed as

(a) $u(t + 5) - 2u(t + 10) + u(t + 15)$ (b) $u(t - 5) - u(t - 10) + u(t - 15)$
(c) $u(t - 5) - 2u(t - 10) + 2u(t - 15)$ (d) $u(t - 5) - 2u(t - 10) + u(t - 15)$

5. A set of signals $s_1(t)$, $s_2(t)$, …, $s_n(t)$ is said to be linearly independent only if

(a) the zero signal is one of the elements in the set
(b) the zero signal is not one of the elements in the set
(c) no linear combination of the signals in that set is equal to the zero signal unless all the coefficients of the linear combination are zero
(d) their linear combination is equal to zero without all the coefficients of the linear combination being equal to zero

6. A basis for a signal space S may be defined as a set of signals in S which

(a) span the space S (b) are mutually orthogonal
(c) are linearly independent (d) are linearly independent and span the space S

7. A basis set for an N-dimensional signal space will contain

(a) at least N signals (b) at the most N signals
(c) N signals (d) N^2 signals

8. The angle between the vectors $\boldsymbol{A} = (1, 0, 0)$ and $\boldsymbol{B} = (1, 1, 0)$ is

(a) 0° (b) 45° (c) 30° (d) 60°

9. S is the signal space of all continuous-time signals having a finite energy over the interval $[0, T]$. The norm of a signal $x(t)$ in S is

(a) square-root of the energy of $x(t)$ over $[0, T]$ (b) energy of the signal $x(t)$ over $[0, T]$
(c) square of the energy of $x(t)$ over $[0, T]$ (d) average power of $x(t)$ over the interval $[0, T]$

10. Any set S of orthogonal signals is a linearly independent set. This is

(a) true (b) false
(c) true only if S includes the zero signal (d) true only if S does not include the zero signals.

11. Consider the signal space S of all real-valued continuous-time functions defined over the interval $[0, T]$ and having a finite energy in that interval with an inner product defined as

$$(x(t), y(t)) \underline{\Delta} \int_0^T x(t)y(t)dt$$

In this space, the sequence of signals 1, $\cos \omega_0 t$, $\cos 2\omega_0 t$, …, where $\omega_0 = 2\pi/T$, form

(a) a complete orthonormal set (b) an orthonormal set
(c) an orthogonal set which is complete (d) an orthogonal set

12. From a given set of m signals, using Gram–Schmidt orthogonalization procedure, n orthonormal signals have been derived

(a) $n = m$ (b) $n < m$ (c) $n \leq m$ (d) $n \geq m$

13. The average power of the periodic signal $c_n e^{j2\pi n f_0 t}$ is

(a) c_n^2 (b) $|c_n^2|$ (c) $|c_n|^2$ (d) $c_n^2 e^{j4\pi n f_0 t}$

14. Parseval's theorem pertaining to Fourier series states that

(a) The signal $x(t)$ is equal to the sum of its components along each of the basis functions, $e^{j2\pi n f_0 t}$, $n = 0, \pm1, \pm2, \ldots$

(b) The average power of $x(t)$ is equal to the sum of the average powers of its components along each of the basis functions, $e^{j2\pi n f_0 t}$, $n = 0, \pm1, \pm2, \ldots$

(c) The energy of the signal $x(t)$ is equal to the sum of the energies of its components along each of the basis functions, $e^{j2\pi n f_0 t}$, $n = 0, \pm1, \pm2, \ldots$

(d) Energy of the signal may be obtained in the time-domain or from the frequency domain.

15. If $x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; -\infty < t < \infty$ and $f_0 \overset{\Delta}{=} \frac{1}{T}$ where T is the fundamental period of the periodic signal, $x(t)$, which is purely real-valued, then

(a) $c_n = -c_{-n}$ (b) $c_n = c_{-n}$ (c) $c_{-n} = c_n^*$ (d) $c_n = -c_n^*$

16. A periodic signal with fundamental period T, is said to possess 'Rotational symmetry', or 'Half-wave symmetry', if

(a) $x(t + T/2) = x(t)$ for any t (b) $x(t \pm T/2) = -x(t)$ for any t

(c) $x(t - T/2) = x(t)$ for any t (d) $x(t + T/2) = x(t - T/2)$ for any t

17. The Fourier series of a periodic signal $x(t)$ with period T will not converge if

(a) $x(t)|$ is not finite at all values of t (b) $x(t)$ has more than one maxima in one period T

(c) $x(t)$ is not continuous at all points (d) $x(t)$ is not a band-limited signal

18. The Fourier series expansion of the periodic signal $x(t) = |\sin 2\pi f_0 t|$ can have

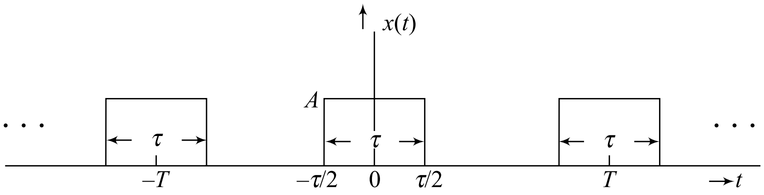
(a) only odd harmonics, i.e., components with frequency nf_0 where n is odd

(b) no dc component

(c) only even harmonics, i.e., components with frequency nf_0 where n is even

(d) both even and odd harmonics of the frequency f_0

19. In the discrete spectrum of the periodic signal $x(t)$ shown in Fig. M2.19, the harmonic component having zero amplitude is



$T = 10$ sec, $\tau = 0.2$ sec

Fig. M2.19

(a) fifth (b) tenth (c) fiftieth (d) twentieth

Key to Multiple-Choice Questions

1. (a)	2. (d)	3. (a)	4. (d)	5. (c)	6. (d)	7. (c)	8. (b)
9. (a)	10. (d)	11. (d)	12. (c)	13. (c)	14. (b)	15. (c)	16. (b)
17. (a)	18. (c)	19. (c)					

FOURIER TRANSFORM AND SIGNAL TRANSMISSION THROUGH SYSTEMS

3

"It is always the simple that produces the marvelous."

Amelia Barr (1831–1919)
British novelist

Learning Objectives

After going through this chapter, students will be able to

- *state the physical meaning of the Fourier transform of a signal,*
- *determine and plot the magnitude and phase spectra of any given Fourier transformable signal,*
- *determine the energy spectral density of any given energy signal, or the power spectral density of any given power signal,*
- *understand the meaning and significance of the terms: Linearity, Time-Invariance, as applied to systems,*
- *determine the impulse response, transfer function and step response of an LTI system given its electrical equivalent circuit and also comment on its stability,*
- *state the conditions required to be satisfied for distortionless transmission of a signal through an LTI system, and*
- *determine by applying Paley–Wiener criterion, whether a given transfer function is physically realizable or not.*

3.1 INTRODUCTION

In the previous chapter we had seen that any periodic signal satisfying Dirichlet's conditions could be expressed as a Fourier series and that such an expression is valid for all times. We had also observed that the Fourier series coefficients provide information regarding the frequency content of a periodic signal and that the spectrum of a periodic signal is a discrete spectrum. The Fourier series, being inherently periodic in nature, does not provide an appropriate tool for the representation of an aperiodic signal.

In this chapter, we will develop an appropriate mathematical tool for determining the spectral content of a non-periodic signal and study its properties as well as its applications in the analysis of LTI systems. We arrive at this tool, the Fourier transform, by starting with the complex-exponential Fourier series expansion of a periodic signal with a period T and then allowing T to tend to infinity. When we do this, we find that in the limit, the periodic signal becomes aperiodic and that the Fourier series expansion gives rise to the Fourier transform. This Fourier transform is a linear operator that maps a signal $x(t)$ satisfying certain conditions, into another function with the continuous variable 'ω' or 'f' as the independent variable. This frequency function gives an indication of the spectral content of the aperiodic signal $x(t)$ and gives a continuous spectrum. This transform is invertible and the inverse Fourier transform provides a representation of the signal $x(t)$ as a combination (integral) of weighted complex exponentials of all frequencies. The Fourier transform is an extremely useful mathematical tool and is extensively used in the analysis of LTI systems, cryptography, signal processing, etc.

Convolution, an operation on a pair of signals, assumes importance from the fact that an LTI system convolves a given input signal with its own impulse response function and gives the resultant signal as its response to the given input signal. Correlation is another operation on a given *pair* of signals and it reveals the degree of similarity between the two signals. It plays a very important role in the detection of known signals in the presence of noise. Radar, active sonar, and digital communications use correlation technique extensively. The correlation of a signal $x(t)$ with a shifted version of itself, shifted by a time interval τ, gives the auto-correlation function of the signal $x(t)$, and is a function of the shift τ. Frequency domain representation of the autocorrelation function, obtained by taking its Fourier transform, is called the Power Spectral Density (PSD), or the Energy Spectral Density (ESD) depending on whether $x(t)$ is a power signal or an energy signal. It represents how the signal power, or the energy, as the case may be, is distributed with respect to frequency.

Hilbert transform differs from other transforms like the Laplace transform or the Fourier transform in the sense that Hilbert transforming a signal does not bring about a change in the domain. The Hilbert transform of a time signal is also a time signal. Only thing that happens is that all the frequency components of the original signal suffer a phase shift of $-\pi/2$ radians. This property of the Hilbert transform makes it very useful in the representation of bandpass signals and bandpass systems.

In addition to the Fourier transform and the Hilbert transform, we will be presenting a brief review of the theory of linear time-invariant systems, as these play an important role in all communication systems. In the discussion on signal transmission through systems, one important topic that merits serious consideration is the distortionless transmission of a signal through an LTI system. So, in this chapter we will determine the conditions under which distortionless transmission is possible. Another aspect of signal transmission through systems that we consider in this chapter is the filtering action of the LTI systems.

3.2 CONTINUOUS-TIME FOURIER TRANSFORM

In the previous chapter, we had developed the continuous-time Fourier series as an orthogonal expansion and found that the complex-exponential Fourier series and the trigonometric Fourier series provide a powerful tool for determining the spectra of continuous-time periodic signals. Fourier series expansion, being inherently periodic in nature, does not provide an appropriate tool for the expansion of aperiodic, i.e., non-periodic signals. This is because, it gives the true representation of the aperiodic signal only for the interval over which the Fourier series expansion of the signal is made; outside this interval, it only repeats, even though the signal does not.

Consider a periodic signal $x(t)$ with a period $T = 1/f_0$. We know that in the limiting case as T tends to infinity, the periodic signal $x(t)$ becomes an aperiodic signal. Also, as the spectral lines in the discrete spectrum of the periodic signal with period T will be f_0 Hz apart where $f_0 = 1/T$, as T tends to infinity, while the signal itself becomes non-periodic, its spectrum becomes a continuous one. We shall proceed on these lines and derive the continuous-time Fourier transform as a limiting case of the Fourier series.

For the periodic signal, $x(t)$, we have the Fourier series expansion:

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \; ; -\infty < t < \infty \; ; \; f_0 \underline{\Delta} \frac{1}{T}, \tag{3.1}$$

where

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jn\omega_0 t} \, dt, \tag{3.2}$$

Now, as $T \to \infty$, $\omega_0 \to d\omega$, an infinitesimally small quantity so that $n\omega_0$ becomes a continuous variable, which we shall represent by ω. Then from the right-hand side of Eq. (3.2), it is clear that c_n becomes a function of ω. Hence, representing c_n as $c_n(\omega)$, we may re-write Eq. (3.2) as

$$Tc_n(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \, dt \tag{3.3}$$

Since the RHS (and therefore the LHS too) of Eq. (3.3) is a function only of ω (since we are integrating for all values of time), let us write the LHS simply as $X(\omega)$. Then

$$X(\omega) \underline{\Delta} \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \, dt \tag{3.4}$$

Further, form Eq. (3.1), it follows that

$$x(t) = \frac{1}{T} \sum_{\omega=-\infty}^{\infty} Tc_n(\omega) e^{j\omega t} \tag{3.5}$$

However,

$$\frac{1}{T} = \frac{\omega_0}{2\pi} \quad \text{and} \quad \frac{\omega_0}{2\pi} \to \frac{d\omega}{2\pi} \text{ as } T \to \infty$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} \, d\omega; -\infty < t < \infty \tag{3.6}$$

Equation (3.4) is called the Fourier transform equation and it transforms the time function $x(t)$ into $X(\omega)$, a function of the variable ω (or f). On the same lines, Eq. (3.6), which enables us to get back the time function $x(t)$ from the frequency function $X(\omega)$, is called the '*inverse Fourier transform*' equation. $x(t)$ is called the 'inverse Fourier transform' of $X(\omega)$. Together they are said to form a '*Fourier transform pair*'. Their relationship is symbolically represented using the following notation:

$$x(t) \xleftrightarrow{\text{FT}} X(f)$$

or

$$X(f) = \mathcal{F}[x(t)]$$

and

$$x(t) = \mathcal{F}^{-1}[X(f)]$$

For convenience, we use the frequency variable 'f' instead of ω in Eqs. (3.4) and (3.6) and write the Fourier and inverse Fourier transforms respectively as follows:

$$\boxed{X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} \, dt} \tag{3.7}$$

and

$$\boxed{x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi f t} \, df} \tag{3.8}$$

Note *There is a unique relationship between the signal x(t) and its Fourier transform, or spectrum X(f). For a given x(t) there is one and only one X(f) and for a given X(f), there is one and only one x(t).*

3.2.1 Existence and Convergence of a Fourier Transform

The Fourier transform $X(f)$ of a time function $x(t)$ is said to exist if $X(f)$ is finite, i.e., if $|X(f)| < \infty$.

Since
$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi f t}\,dt,$$

we have
$$|X(f)| = \left|\int_{-\infty}^{\infty} x(t)e^{-j2\pi f t}\,dt\right| \le \int_{-\infty}^{\infty} |x(t)e^{-j2\pi f t}|dt$$

Since
$$|e^{-j2\pi f t}| = 1 \text{ and } |x(t)e^{-j2\pi f t}| = |x(t)||e^{-j2\pi f t}|,$$

it follows that
$$\int_{-\infty}^{\infty} |x(t)|dt < \infty \tag{3.9}$$

is the condition required to be satisfied for $|X(f)|$ to be finite. Thus, Eq. (3.9) represents the condition to be satisfied for the existence of the Fourier transform of $x(t)$. It may however, be noted that this condition is a sufficient condition and not a necessary condition. This is because, as we shall see later, if we are prepared to allow 'singularity' functions, then it is possible to derive the Fourier transforms of even functions like the unit step, the sinusoid, etc., which are definitely not absolutely integrable (as required, according to Eq. (3.9)).

The Fourier transform integral given by Eq. (3.7) and the inverse Fourier transform integral given by Eq. (3.8) may not converge for all functions $x(t)$ and $X(f)$, respectively. As a detailed analysis of the convergence of these integrals is beyond the scope if this book, we simply state here that if a non-periodic signal $x(t)$ satisfies the Dirichlet conditions, then the pointwise convergence of the integral

$$\int_{-\infty}^{\infty} X(f)e^{j2\pi f t}\,df$$

is guaranteed for all values of t except those corresponding to discontinuities. The Dirichlet conditions are the following:

1. $x(t)$ should be absolutely integrable.
2. $x(t)$ should have only a finite number of maxima and minima in any finite interval of time.
3. In any finite interval of time, the number of discontinuities of $x(t)$ should be finite.
4. Discontinuities of $x(t)$, if any, should be finite discontinuities.

Most of the signals that we come across satisfy all the above conditions, except possibly the first one. However, as mentioned earlier, even if a signal $x(t)$ is not absolutely integrable, we can still Fourier transform it by permitting impulse functions. However, Fourier transforms of these signals do not converge.

3.2.2 Simple Properties of Fourier Transform

We will now give, without proof, a list of some simple, but very useful, properties of the Fourier transform. The reader is urged to supply the proof using Eqs. (3.7) and (3.8).

1. $X(0)$ is equal to the area under $x(t)$. This is because $X(f)=\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$

$$X(0)=\int_{-\infty}^{\infty} x(t)dt \text{ = area under the signal } x(t).$$

2. The Fourier transform $X(f)$ is, in general, a complex-valued function of frequency, even if the signal $x(t)$, is a real-valued one.
3. If $x(t)$ is real valued, then its Fourier transform $X(f)$, has Hermitian symmetry. That is

$$|X(-f)|=|X(f)| \text{ while } \angle X(-f)=-\angle X(f) \tag{3.10}$$

   This says that if $x(t)$ is real valued, the magnitude of $X(f)$ will have even symmetry while the phase of $X(f)$ will have odd symmetry.
4. (a) If the signal $x(t)$ has even symmetry, then its Fourier transform $X(f)$ is given by

$$X(f)=2\int_{0}^{\infty} x(t)\cos \omega t dt \tag{3.11}$$

   (b) If the signal $x(t)$ has odd symmetry, then its Fourier transform $X(f)$ is given by

$$X(f)=-2j\int_{0}^{\infty} x(t)\sin \omega t dt \tag{3.12}$$

3.2.3 Magnitude and Phase Spectra of Signals

As pointed out in Property 1 above, $X(f)$, the Fourier spectrum of a signal $x(t)$ is, in general, a complex-valued function of frequency. Hence, it will have a magnitude $|X(f)|$ and phase $\angle X(f)$, both of which are functions of frequency. For any signal $x(t)$, a plot of $|X(f)|$ vs. f is called the magnitude spectrum and a plot of $\angle X(f)$ vs. f is called the phase spectrum. We illustrate these concepts through the following example.

Example 3.1 $x(t)$ is a 'GATE' signal and is described by

$$x(t)=\begin{cases} A; & t \le |\tau/2| \\ 0; & \text{otherwise} \end{cases}$$

Determine and plot the magnitude and phase spectra of $x(t)$.

Solution The given signal is a rectangular pulse and its plot is as shown in Fig. 3.1.

This being a commonly used signal, it is given a special symbol.

$$x(t)=A\Pi(t/\tau) \quad \text{or} \quad A\,\text{rect}(t/\tau)$$

A in the above notation indicates that the rectangular pulse has an amplitude A; t indicates that the rectangular pulse is in time domain and τ indicates that the rectangular pulse has a total width of τ along the time axis.

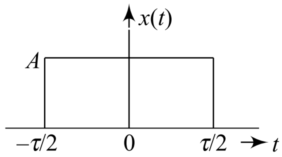


Fig. 3.1 *A rectangular pulse (GATE signal)*

Note *In this notation, it is always understood that the rectangular pulse is symmetrically situated with respect to the time origin, i.e., it extends from* $-\frac{\tau}{2}$ *to* $+\frac{\tau}{2}$

Now,

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} = \int_{-\tau/2}^{\tau/2} x(t)e^{-j2\pi ft}\,dt$$

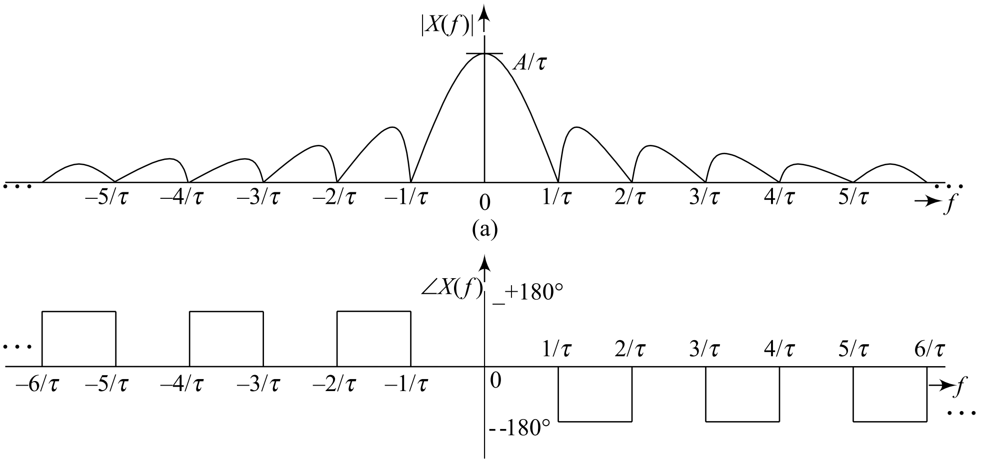
$$= 2\int_{0}^{\tau/2} A\cos\omega t dt = \frac{2A}{\omega}\sin\omega t\Big|_{0}^{\tau/2} = A\tau\left(\frac{\sin\pi f\tau}{\pi f\tau}\right)$$

If we define $\frac{\sin\pi\lambda}{\pi\lambda} \triangleq \text{sinc }\lambda$

We have

$$\boxed{X(f) = \mathcal{F}[A\Pi(t/\tau)] = A\tau\,\text{sinc } f\tau} \quad (3.13)$$

Plots of the magnitude and phase spectra of the signal *x*(*t*) are shown in the following Figs. 3.2(a) and 3.2(b), respectively.



(a)

(b)

Fig. 3.2 *(a) Magnitude spectrum of AΠ(t/τ), (b) Phase spectrum of AΠ(t/τ)*

In this example, *X*(*f*) which is equal to $A\tau$ sinc $f\tau$, is a purely real-valued function. However, this function changes its sign whenever the frequency '*f*' equals ±1/τ, ±2/τ, ±3/τ . . . . This change of sign is interpreted as a phase shift of 180°. Actually one need not distinguish between +180° phase shift and –180° phase shift. But, in Fig. 3.2(a) we have deliberately shown the +180° and –180° separately in order to emphasize the fact that *X*(*f*) must have Hermitian symmetry (i.e., magnitude spectrum should have even symmetry, and phase spectrum should have odd symmetry), since the given *x*(*t*) is purely real valued.

3.2.4 Physical Meaning of *X*(*f*) in Relation to the Signal *x*(*t*)

We shall now explore the physical meaning of the function *X*(*f*) in relation to the signal *x*(*t*). This we do by an appropriate physical interpretation of what the Parseval's theorem tells us. So, we shall first state and prove this theorem and then attempt to examine the significance of the function *X*(*f*).

Parseval's theorem This theorem is also known as *Rayleigh's theorem* pertaining to the Fourier transform. It states that if signals $x(t)$ and $y(t)$ have Fourier transforms $X(f)$ and $Y(f)$ respectively, then

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}df$$

where the overbar is used for representing complex conjugate.

Proof Since $Y(f)$ is the Fourier transform of $y(t)$, we have

$$y(t) = \mathcal{F}^{-1}[Y(f)] = \int_{-\infty}^{\infty} Y(f)e^{j2\pi ft}df$$

$$\therefore \qquad \overline{y(t)} = \int_{-\infty}^{\infty} \overline{Y(f)}e^{-j2\pi ft}df$$

Hence,

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} x(t)\left[\int_{-\infty}^{\infty} \overline{Y(f)}e^{-j2\pi ft}df\right]dt$$

Interchanging the order of the integrations in the RHS of the above, we get

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} \overline{Y(f)}\left[\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt\right]df$$

$$= \int_{-\infty}^{\infty} \overline{Y(f)}\cdot X(f)df$$

Thus,

$$\boxed{\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}df} \tag{3.14}$$

This is the general form of Parseval's theorem pertaining to the Fourier transform. A special form of this is obtained when $y(t)$ is the same as $x(t)$. In that case, Eq. (3.14) becomes

$$\boxed{\int_{-\infty}^{\infty} |x(t)|^2dt = \int_{-\infty}^{\infty} |X(f)|^2 df} \tag{3.15}$$

In the above equation, we know that the LHS represents the energy 'E' of the signal $x(t)$. Hence, Eq. (3.15) tells us that the function $|X(f)|^2$ when integrated for all frequencies, equals E. In other words, $|X(f)|^2$ denotes the energy density of the signal with respect to the frequency, at the frequency 'f'. *Hence, if we consider a specific frequency, f_0, and take a unit interval of frequency centered on f_0, then $|X(f_0)|^2$ represents the energy possessed by the signal in that unit interval frequency band around f_0.* The function $|X(f)|^2$ thus shows how the energy of the signal $x(t)$ is distributed with respect to frequency. Equation (3.15) further tells us that the energy of a signal may be calculated either in the time domain or in the frequency domain by using the RHS of the equation.

Example 3.2 If the signal $x(t) = Ae^{-t/T}u(t)$ is given as input to an ideal low pass filter whose cut-off frequency is $f_c = 1/2\pi T$, what percentage of the energy of $x(t)$ will be available at the output of the filter?

Solution We have to first find the spectrum of *X*(*f*) of the signal *x*(*t*). For this, we note

$$X(f)=\int_{-\infty}^{\infty} Ae^{-t/T}u(t)e^{-j2\pi ft}dt = A\int_{0}^{\infty} e^{\frac{-(1+j2\pi fT)t}{T}}\,dt$$

$$X(f)=\frac{AT}{1+j2\pi fT} \quad \text{and} \quad |X(f)|^2=\frac{A^2T^2}{1+4\pi^2f^2T^2}$$

Putting $2\pi fT = \tan\theta$ in the above and noting that $df = (1/2\pi T)\sec^2\theta\, d\theta$, we have

$$E_x = \text{Total energy in the signal } x(t) = \int_{-\infty}^{\infty}\frac{A^2T^2}{1+4\pi^2f^2T^2}df$$

$$=\int_{-\pi/2}^{\pi/2}\left(\frac{A^2T}{2\pi}\right)d\theta=\frac{A^2T}{2}$$

Now, when the signal *x*(*t*) is applied as input to a lowpass filter with $f_c = 1/2\pi T$, the filter passes on to the output side only those frequency components of *x*(*t*) which lie from $-f_c$ to $+f_c$. Hence, from Eq. (3.15) we know that the energy contained in the signal at the output of the filter is given by

$$E_0=\int_{f=-(1/2\pi T)}^{f=-(1/2\pi T)}\left(\frac{A^2T}{2\pi}\right)d\theta=\frac{A^2T}{4}$$

Thus, the percentage of the signal energy available at the output of the filter is given by *p*

where $$p=\frac{E_0}{E_x}\times 100\%=\frac{(A^2T/4)}{(A^2T/2)}\times 100 = 50\%$$

3.2.5 Fourier Transform Theorems

The Fourier transform theorems which we are going to discuss now will be very useful in finding the Fourier transforms of some complicated signals in terms of the Fourier transforms of simpler signals.

1. Linearity theorem: Fourier Transform is linear in the sense that it obeys the superposition and homogeneity principles.

If *x*(*t*) and *y*(*t*) are continuous-time signals with *X*(*f*) and *Y*(*f*) respectively as their Fourier Transforms, and if α and β are any two arbitrary constants, then

$$\boxed{\mathcal{F}[\alpha x(t)+\beta y(t)]=\alpha X(f)+\beta Y(f)} \tag{3.16}$$

Proof of this theorem is left as an exercise to the reader.

2. Time-delay theorem: This theorem gives us the Fourier transform of $x(t-\tau)$, the time-delayed version of an *x*(*t*) in terms of *X*(*f*), the Fourier transform of *x*(*t*). It says that

If $x(t)\xleftrightarrow{FT}X(f)$ Then, $x(t-\tau)\xleftrightarrow{FT}X(f)e^{-j2\pi f\tau}$

Proof $X(f)=\int_{-\infty}^{\infty}x(t)e^{-j2\pi ft}\,dt$

$$\therefore \qquad \mathcal{F}[x(t-\tau)]=\int_{-\infty}^{\infty}x(t-\tau)e^{-j2\pi ft}\,dt$$

Putting $t - \tau = t'$, $t = t' + \tau$ and $dt = dt'$

$$\therefore \qquad \mathcal{F}[x(t-\tau)] = \int_{-\infty}^{\infty} x(t')e^{-j2\pi f(t'+\tau)}\,dt'$$

$$= \left[\int_{-\infty}^{\infty} x(t')e^{-j2\pi ft'}\,dt'\right]e^{-j2\pi f\tau} = X(f)e^{-j2\pi f\tau}$$

$$\therefore \qquad \boxed{x(t-\tau) \xleftrightarrow{FT} X(f)e^{-j2\pi f\tau}} \tag{3.17}$$

Note *Since* $|X(f)e^{-j2\pi f\tau}| = |X(f)|$, *it follows that shifting of a signal along the time axis changes only the phase spectrum but not the magnitude spectrum.*

3. Modulation theorem: As mentioned earlier in Chapter 1, a message signal, $x(t)$, is made to modulate a high frequency sinusoidal carrier signal of frequency f_c in order to facilitate its transmission over long distances. One easy way of accomplishing this modulation is by multiplying the carrier signal with $x(t)$.

This theorem states that if $x(t) \xleftrightarrow{FT} X(f)$ then, $x(t)e^{j2\pi f_c t} \xleftrightarrow{FT} X(f - f_c)$.

Proof $$\mathcal{F}[x(t)e^{j2\pi f_c t}] = \int_{-\infty}^{\infty} \{x(t)e^{j2\pi f_c t}\}e^{-j2\pi ft}\,dt$$

$$= \int_{-\infty}^{\infty} x(t)e^{-j2\pi(f-f_c)t}dt = X(f - f_c)$$

$$\therefore \qquad \boxed{x(t)e^{j2\pi f_c t} = X(f - f_c)} \tag{3.18}$$

Equation (3.18) tells us that the spectrum of $x(t)e^{j2\pi f_c t}$ is just a frequency-shifted version of the spectrum of $x(t)$ itself. Suppose $x(t)$ is a low frequency signal having frequency components from 0 to W Hz. Let its spectrum be $X(f)$ as shown in Fig. 3.3(a). The actual shape of $X(f)$ assumed here has no particular significance. However, since $x(t)$ is a real-valued signal, as per Eq. (3.10), $X(f)$ must have a magnitude which has even symmetry. The spectrum of $x(t)e^{j\omega_c t}$, as given by Eq. (3.18), is plotted in Fig. 3.3(b).

In practice, we have to have a carrier signal which is real valued. So, instead of the complex-exponential signal, $x(t)e^{j2\pi f_c t}$ of frequency f_c, let us use $\cos 2\pi f_c t$ which is a real-valued signal.

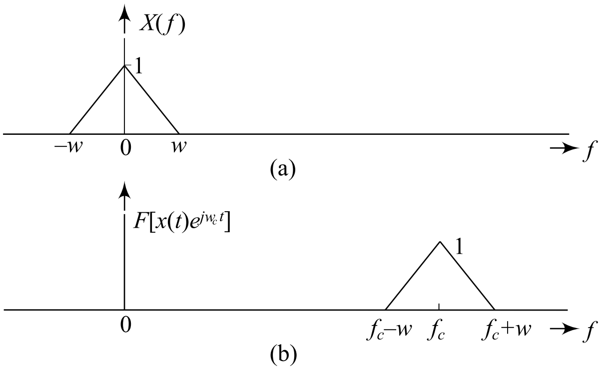


Fig. 3.3 *(a) Spectrum of x(t), (b) Spectrum of* $x(t)e^{j\omega_c t}$

From Eq. (3.18), we may write

$$x(t)e^{j\omega_c t} \xleftrightarrow{FT} X(f - f_c)$$

$$\therefore \qquad x(t)e^{-j\omega_c t} \xleftrightarrow{FT} X(f + f_c)$$

Adding these two and invoking the linearity property of the Fourier transform, we get

$$\therefore \qquad \boxed{x(t)\cos\omega_c t \xleftrightarrow{FT} \frac{1}{2}[X(f - f_c) + X(f + f_c)]} \tag{3.19}$$

Hence, the spectrum of $x(t)\cos\omega_c t$, the modulated signal, would appear as shown in Fig. 3.4.

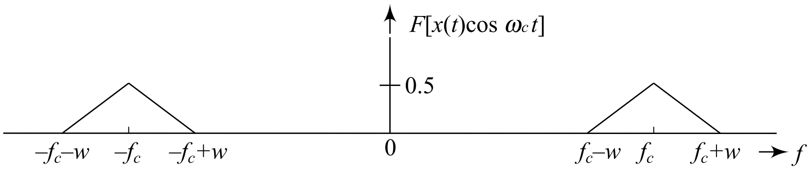


Fig. 3.4 *Spectrum of $x(t)\cos\omega_c t$*

It may be noted that whereas the spectrum of $x(t)\cos\omega_c t$ has even symmetry, that of $x(t)e^{j\omega_c t}$ does not have even symmetry. This is because, while $x(t)\cos\omega_c t$ is a real-valued function, $x(t)e^{j\omega_c t}$ is not.

4. Scaling theorem: This theorem deals with the effect on the spectrum of a signal when the signal is subjected to time scaling, i.e., compression or expansion in time. In Chapter 2, Section 2.2, while dealing with operations on signals, with reference to the time-scaling operation, we had observed that for a constant 'a', the signal $x(at)$ represents a time- compressed version of $x(t)$ if the constant $a > 1$ and a time-expanded version of $0 < a < 1$.

This theorem states that if $x(t) \xleftrightarrow{FT} X(f)$, then $x(at) \xleftrightarrow{FT} \frac{1}{|a|}X(f/a)$

Proof First, Let $a > 0$:
Putting $t' = at$, we have, $dt' = a\,dt$

$$\therefore \qquad \mathcal{F}[x(at)] = \frac{1}{a}\int_{-\infty}^{\infty} x(t')e^{-j2\pi(f/a)t'}dt' = \frac{1}{a}X(f/a)$$

Now consider the case of $a < 0$.
Putting $t' = at$, we have, $dt' = a\,dt$

$$\mathcal{F}[x(at)] = \frac{1}{a}\int_{+\infty}^{-\infty} x(t')e^{-j2\pi(f/a)t'}dt'$$

$$= \frac{1}{|a|}\int_{-\infty}^{\infty} x(t')e^{-j2\pi(f/a)t'}dt' = \frac{1}{|a|}X(f/a)$$

Combining the two cases, we may say

$$\boxed{x(at) \xleftrightarrow{FT} \frac{1}{|a|}X(f/a)} \tag{3.20}$$

Remark

(*i*) *For a > 1, x(at) represents a time-compressed version of the signal x(t) but X(f/a) represents a spectrum that has been expanded in frequency. Hence, compressing a signal in time results in an expansion of its spectrum.*

(*ii*) *For 0 < a < 1, x(at) represents a signal that is expanded in time. But X(f/a) represents a frequency compressed version of the spectrum. Hence expansion of a signal in time results in a compression of its spectrum.*

(*iii*) *Since compression in time leads to expansion in frequency and vice versa, a signal cannot be compressed/expanded simultaneously in time as well as in frequency.*

(*iv*) *If a < 0, there will be lateral inversion of the spectrum accompanied by compression or expansion, depending upon whether* |a| *is greater than or less than unity.*

Most of the readers would have experienced the manifestation of the above result in practice. A male voice recorded at some speed, would sound like a female voice if it is played back at a much higher speed. Similarly, a female voice recorded at some speed, would sound like a male voice, when played back at a much lower speed.

5. Duality theorem: This theorem enables us to write down the spectra of certain signals just by inspection, as illustrated in Example 3.3.

It states that if *X*(*f*) is the Fourier transform of a signal, *x*(*t*), the Fourier transform of *X*(*t*) is given by *x*(–*f*).

Proof $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$

Interchanging *t* and *f*,

$$X(t) = \int_{-\infty}^{\infty} x(f)e^{-j2\pi ft}df$$

Now, putting $f' = -f$

$$X(t) = -\int_{+\infty}^{-\infty} x(-f')e^{j2\pi f't}df' = \int_{-\infty}^{\infty} x(-f')e^{j2\pi f't}df' = \mathcal{F}^{-1}[x(-f)]$$

$$\therefore \qquad \boxed{X(t) \xleftrightarrow{FT} x(-f)} \tag{3.21}$$

6. Convolution theorem: This theorem tells us that the Fourier transform converts a time-domain convolution into a multiplication operation in the frequency domain. As it is much easier to compute a multiplication as compared to a convolution, this theorem enables us to use the Fourier transform to advantage in the computation of the output signal of a Linear Time-Invariant (LTI) system, since in these systems, the output signal is the convolution of the input signal with the impulse response, *h*(*t*), of the system.

Statement Let $x(t) \xleftrightarrow{FT} X(f)$, $y(t) \xleftrightarrow{FT} Y(f)$ and $z(t) = x(t) * y(t)$, where $*$ denotes convolution operation, then this theorem states that

$$Z(f) = \mathcal{F}[z(t)] = X(f) \cdot Y(f)$$

Proof $Z(f) = \int_{-\infty}^{\infty} z(t)e^{-j2\pi ft}dt = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x(\lambda)y(t-\lambda)d\lambda\right\}e^{-j2\pi ft}dt$

$$= \int_{-\infty}^{\infty} x(\lambda)\left\{\int_{-\infty}^{\infty} y(t-\lambda)e^{-j2\pi ft}dt\right\}d\lambda$$

$$= \int_{-\infty}^{\infty} x(\lambda)Y(f)e^{-j2\pi\lambda f}d\lambda \quad \text{(by applying time-delay theorem)}$$

$$= Y(f)\int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi\lambda f}d\lambda = Y(f)\cdot X(f)$$

$$\therefore \qquad \boxed{Z(f) = Y(f)\cdot X(f) \quad \text{if} \quad z(t) = x(t) * y(t)} \qquad (3.22)$$

7. Multiplication theorem: This theorem tells us that a time-domain product of two signals will be converted by the Fourier transform into the frequency-domain convolution of the Fourier transforms of the two signals. It states that if $x(t) \xleftrightarrow{FT} X(f)$, $y(t) \xleftrightarrow{FT} Y(f)$ and if $z(t) = x(t) \times y(t)$, then $Z(f) = X(f) * Y(f)$

Proof $Z(f) = \mathcal{F}[z(t)] = \int_{-\infty}^{\infty} z(t)e^{-j2\pi ft}dt$

But $$z(t) = x(t)\cdot y(t)$$

$$\therefore \qquad Z(f) = \int_{-\infty}^{\infty} \{x(t)\cdot y(t)\}\, e^{-j2\pi ft}dt$$

Now, writing $y(t)$ as the inverse Fourier transform of $Y(\lambda)$, where λ is a dummy frequency parameter,

$$Z(f) = \int_{-\infty}^{\infty} x(t)\,\{Y(\lambda)e^{j2\pi\lambda t}d\lambda\}\, e^{-j2\pi ft}dt$$

$$= \int_{-\infty}^{\infty} Y(\lambda)\left\{\int_{-\infty}^{\infty} [x(t)e^{j2\pi\lambda t}]\, e^{-j2\pi ft}dt\right\} d\lambda$$

Now, using the modulation theorem, we may write

$$Z(f) = \int_{-\infty}^{\infty} Y(\lambda)X(f-\lambda)d\lambda = X(f) * Y(f)$$

$$\therefore \qquad \boxed{Z(f) = X(f) * Y(f) \quad \text{if} \quad z(t) = x(t)\cdot y(t)} \qquad (3.23)$$

Example 3.3 Determine the energy contained in the signal $x(t) = 20$ sinc $10t$.

Solution We shall solve the problem by making use of Parseval's theorem. Earlier, we had seen (see Eq. (3.13)) that

$$A\Pi(t/\tau) \leftrightarrow A\tau \operatorname{sinc} f\tau.$$

Now, $A\tau$ sinc $f\tau$ is a frequency function and τ is a fixed time interval. We may, in order to use the duality theorem, write the corresponding time function as AW sinc Wt, by replacing the fixed time interval τ (of $A\tau$ sinc $f\tau$) by a fixed frequency interval, W, and by replacing the frequency variable f, by the time variable, t.

$\therefore$ Let $20 \operatorname{sinc} 10t = AW \operatorname{sinc} Wt$

Thus, $AW = 20$ and $W = 10$ $\quad\therefore\quad A = 2$

We know from the duality theorem that

$$AW \operatorname{sinc} Wt \leftrightarrow A\Pi(-f/w) = A\Pi(f/w)$$

Hence, $$20 \operatorname{sinc} 10t \xleftrightarrow{FT} 2\Pi(f/10) = X(f)$$

This $X(f)$ is a rectangular pulse in frequency domain with an amplitude of 2 and base width of 10.

∴ Applying Parseval's theorem,

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

$$= \int_{-5}^{5} 2^2\, df = 4 \times 10 = 40 \text{ units}$$

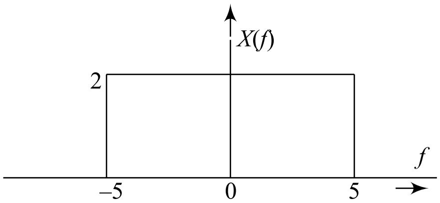


Fig. 3.5 *Fourier transform of x(t)*

Example 3.4 Find the Fourier transform of $x(t) = \begin{cases} \cos \pi t; & -\frac{1}{2} \le t \le \frac{1}{2} \\ 0 \quad ; & \text{otherwise} \end{cases}$

Solution We can solve it either by using the defining equation of the Fourier transform, or by using the convolution theorem.

(a) *By using the defining equation of Fourier transform*

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt = \int_{-1/2}^{1/2} \cos \pi t\, [\cos 2\pi ft - j \sin 2\pi ft] dt$$

$$= \int_{-1/2}^{1/2} \cos \pi t \cdot \cos 2\pi ft dt - j \int_{-1/2}^{1/2} \cos \pi t \cdot \sin 2\pi ft dt$$

In the above, the second integral is zero since cos πt is even while sin $2\pi ft$ is odd.

∴ $$X(f) = \int_{-1/2}^{1/2} \cos \pi t \cdot \cos 2\pi ft dt$$

$$= \frac{1}{2} \int_{-1/2}^{1/2} \{\cos \pi(2f+1)t + \cos \pi(2f-1)t\} tdt$$

$$= \frac{1}{2}\left[\frac{\sin \pi\,(2f+1)}{\pi(2f+1)}\Bigg|_{t=-1/2}^{t=1/2} + \frac{\sin \pi(2f-1)}{\pi(2f-1)}\Bigg|_{t=-1/2}^{t=1/2} \right]$$

$$= \frac{1}{2}\left[\text{sinc}\left(f + \frac{1}{2}\right) + \text{sinc}\left(f - \frac{1}{2}\right) \right]$$

(b) *Using the convolution theorem of Fourier transform*

The given $x(t)$ is shown in Fig. 3.6. As is clear from the figure, $x(t)$ may be viewed as the product of a signal $x_1(t) = \cos \pi t; -\infty < t < \infty$ and a window function $w(t) = \Pi(t/1)$ which has a value of 1 for $|t| \le ½$ and zero outside.

∴ $$x(t) = \cos \pi t \cdot \omega(t)$$

Hence, $$X(f) = \mathcal{F}[\cos \pi t] * W(f)$$

But $$\mathcal{F}[\cos \pi t] = \frac{1}{2}\left[\delta\left(f - \frac{1}{2}\right) + \delta\left(f + \frac{1}{2}\right) \right]$$

∴ $$X(f) = \frac{1}{2}\left[\delta\left(f - \frac{1}{2}\right) + \delta\left(f + \frac{1}{2}\right) \right] * W(f)$$

$$= \frac{1}{2}\left[W\left(f - \frac{1}{2}\right) + W\left(f + \frac{1}{2}\right) \right]$$

(Replication property of an impulse)

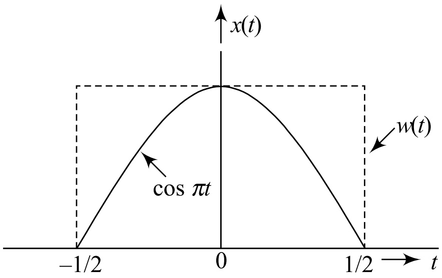


Fig. 3.6 *Given x(t)*

But $$w(t) = \Pi(t/1) \quad \therefore \quad W(f) = \text{sinc}\, f$$

$\therefore$ $$X(f) = \frac{1}{2}\left[\text{sinc}\left(f - \frac{1}{2}\right) + \text{sinc}\left(f + \frac{1}{2}\right)\right]$$

8. Differentiation-in-time theorem: This theorem enables us to straight away write down the Fourier transform of the derivative of a signal in terms of the Fourier transform of the signal itself. It states that if $x(t) \xleftrightarrow{FT} X(f)$, then $\dot{x}(t) = j2\pi\, fX(f)$

Proof $$x(t) = \mathcal{F}^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df \tag{3.24}$$

$\therefore$ $$\frac{dx(t)}{dt} = \frac{d}{dt}\left[\int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df\right] = \int_{-\infty}^{\infty}\{j2\pi fX(f)\}e^{j2\pi ft}df \tag{3.25}$$

Comparing Eqs. (3.24) with (3.25), we have

$$\boxed{\dot{x}(t) \xleftrightarrow{\text{FT}} j2\pi f\, X(f)} \tag{3.26}$$

n iterations of the above process yields

$$\boxed{\frac{d^n}{dt^n}x(t) \xleftrightarrow{\text{FT}} (j2\pi f)^n \cdot X(f)} \tag{3.27}$$

Remark

(*i*) *Here, it must be noted that even if x(t) is of finite energy and so is Fourier transformable, there is no guarantee that its derivatives will also be Fourier transformable.*

(*ii*) *The phase spectrum of* $\dot{x}(t)$ *is obtained by adding 90° to the phase spectrum of x(t) at all frequencies.*

(*iii*) *Multiplication of X(f) by 2πf clearly shows that differentiation accentuates high frequencies.*

Pertaining to Fourier transforms, an integration theorem also exists. But we can discuss it only a little later.

9. Differentiation-in-frequency theorem: This theorem can be considered as the dual of the differentiation-in-time theorem and it states that if *X*(*f*) is the Fourier transform of *x*(*t*), then the inverse Fourier transform of $\frac{d}{df}X(f)$ is given by $-j2\pi t\, x(t)$

Proof $$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt$$

$\therefore$ $$\frac{d}{df}X(f) = \int_{-\infty}^{\infty} x(t)\frac{d}{df}[e^{-j2\pi ft}]dt = \int_{-\infty}^{\infty}\{x(t)(-j2\pi t)\}e^{-j2\pi ft}dt$$

Comparing the LHSs and RHSs of the above two equations, we may state that

$$\boxed{(-j2\pi t)\,x(t) \xleftrightarrow{\text{FT}} \frac{d}{df}X(f)} \tag{3.28}$$

3.2.6 Fourier Transforms using Impulses

We shall now derive the Fourier Transforms of certain functions like the sine and cosine, the unit step, signum function, etc., which are not absolutely integrable. This we will do by using impulses.

1. Spectrum of an impulse function: $\mathcal{F}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t)e^{-j2\pi ft}dt$

Here, $e^{-j2\pi ft}$ is a complex-valued function of time which is continuous. Hence, from Eq. (2.5) of Section 2.2 (Chapter 2), i.e., the defining equation for a unit-impulse function, we find that

$$\int_{-\infty}^{\infty} e^{-j2\pi ft}\delta(t)\,dt = e^{-j2\pi ft}\Big|_{t=0} = 1$$

Hence, $$\boxed{\mathcal{F}[\delta(t)] = 1} \tag{3.29}$$

Equation (3.29) tells us that the spectrum of a unit impulse function $\delta(t)$ consists of all frequency components from $-\infty < f < \infty$ and that it has a value of unity at all frequencies, as shown in Fig. 3.7.

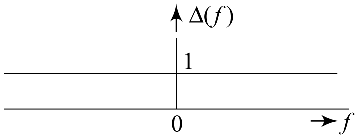


Fig. 3.7 *Spectrum of a unit impulse function in time, δ(t)*

2. Fourier transform of x(t) = 1: Applying duality theorem to the transform given by Eq. (3.29), we get

$$\mathcal{F}[1] = \delta(-f) = \delta(f) = \text{Unit impulse in the frequency domain}$$

∴ $$\boxed{1 \xleftrightarrow{\text{FT}} \delta(f)} \tag{3.30}$$

Since $\delta(t) \xleftrightarrow{\text{FT}} 1,$ if we apply the time-delay theorem, we get

$$\boxed{\delta(t-\tau) \xleftrightarrow{\text{FT}} e^{-j2\pi f\tau}} \tag{3.31}$$

3. Transform of $e^{j2\pi f_0 t}$: From modulation theorem, we know that

$$x(t)e^{j2\pi f_0 t} \xleftrightarrow{\text{FT}} X(f-f_0)$$

In the above, if we take $x(t)$ to be equal to 1,

$$\boxed{e^{j2\pi f_0 t} \xleftrightarrow{\text{FT}} \delta(f-f_0)} \tag{3.32}$$

4. Transform of cos $2\pi f_0 t$: We have noted that

$$e^{j2\pi f_0 t} \longleftrightarrow \delta(f-f_0),$$

$$e^{-j2\pi f_0 t} \longleftrightarrow \delta(f+f_0)$$

Combining the two and invoking the linearity theorem,

$$\boxed{\cos 2\pi f_0 t \xleftrightarrow{\text{FT}} \frac{1}{2}\left[\delta(f-f_0)+\delta(f+f_0)\right]} \tag{3.33}$$

The spectrum of cos $2\pi f_0 t$, as given Eq. (3.33) is shown in Fig. 3.8.

5. Transform of the signum function: The signum function in time, denoted by sgn(t), is defined as

$$\text{sgn}(t) = \begin{cases} 1 & \text{if } t > 0 \\ -1 & \text{if } t < 0 \end{cases} \tag{3.34}$$

We shall derive the Fourier transform of the signum function by making use of the differentiation theorem.

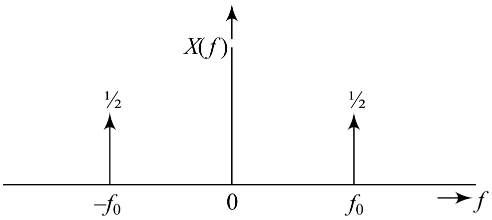


Fig. 3.8 *Spectrum of cos* $2\pi f_0 t$

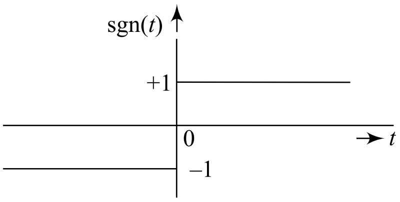


Fig. 3.9 *Signum function*

From Fig. 3.9, we find that

$$\frac{d}{dt}[\text{sgn}(t)] = 2\delta(t)$$

Therefore, from the differentiation theorem

$$\mathcal{F}\left[\frac{d}{dt}\text{sgn}(t)\right] = 2 = j2\pi f \mathcal{F}[\text{sgn}(t)] = j2\pi f S(f)$$

where, we have used $S(f)$ to denote the FT of sgn(t)

$$\therefore \qquad S(f) = \frac{1}{j\pi f} \qquad (3.35)$$

Note that at $f = 0$ the Fourier transform of sgn(t) appears to become infinitely large and therefore indeterminate, as per Eq. (3.35). However, noting that sgn(t) is an odd function of time and that the area under it must be zero, and recalling the result (see some simple properties of the Fourier transform) that $X(0)$ must be equal to the area under $x(t)$, we remove the indeterminacy at $f = 0$ by stipulating that $S(f) = 0$ at $f = 0$. Thus,

$$\boxed{\text{sgn}(t) \xleftrightarrow{\text{FT}} \begin{cases} \dfrac{1}{j\pi f}; & f \neq 0 \\ 0 \quad ; & f = 0 \end{cases}} \qquad (3.36)$$

6. Transform of *u*(*t*): From Fig. 3.9, it is clear that

$$1 + \text{sgn}(t) = 2u(t)$$

i.e.,

$$u(t) = \frac{1}{2}[1 + \text{sgn}(t)] \qquad (3.37)$$

Now taking Fourier transform on both sides, noting that the F.T of 1 is $\delta(f)$ and invoking the linearity theorem of the Fourier transform,

$$u(t) \xleftrightarrow{\text{FT}} U(f) = \frac{1}{2}\left[\delta(f) + \frac{1}{j\pi f}\right]$$

$$\therefore \qquad \boxed{U(f) = \frac{1}{2}\left[\delta(f) + \frac{1}{j\pi f}\right]} \qquad (3.38)$$

7. Integration theorem of Fourier transform: Now that we have derived the Fourier transform of a unit-step function, we are in a position to discuss the integration theorem.

This theorem states that if

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

$$Y(f) = \frac{1}{2}\left[X(f)\delta(f) + \frac{X(f)}{j\pi f}\right]$$

Proof Consider $x(t) * u(t)$. This is given by

$$x(t) * u(t) = \int_{-\infty}^{\infty} x(\tau)u(t-\tau)d\tau$$

But
$$u(t-\tau) = \begin{cases} 1 \text{ for } \tau < t \\ 0 \text{ for } \tau > t \end{cases}$$

$\therefore$
$$x(t) * u(t) = \int_{-\infty}^{t} x(\tau)d\tau = y(t)$$

$\therefore$
$$Y(f) = X(f)\cdot U(f) = \frac{1}{2}\left[X(f)\delta(f) + \frac{X(f)}{j\pi f}\right] \qquad \text{(From Eq. (3.38))}$$

Making use of the sampling property of the impulse function, we have

$$X(f)\delta(f) = X(0)\delta(f)$$

Hence,

$$\boxed{\int_{-\infty}^{t} x(\tau)d\tau \xleftrightarrow{\text{FT}} \frac{1}{2}\left[X(0)\delta(f) + \frac{X(f)}{j\pi f}\right]} \qquad (3.39)$$

Example 3.5 Find the Fourier transform of the signal $x(t)$ shown in Fig. 3.10(a).

Solution We shall use the differentiation theorem of Fourier transform to find the FT of $x(t)$.
From Fig. 3.10(c), we find that

$$\frac{d^2x(t)}{dt^2} = \ddot{x}(t) = -\delta(t+2) + 2\delta(t+1) - 2\delta(t-1) + \delta(t-2)$$

But, from the differentiation theorem of Fourier transform, we know that if

$$x(t) \xleftrightarrow{\text{FT}} X(f)$$

Then

$$\dot{x}(t) \xleftrightarrow{\text{FT}} j2\pi f X(f) \quad \text{and} \quad \ddot{x}(t) \xleftrightarrow{\text{FT}} -4\pi^2 f^2 X(f)$$

Hence,

$$\mathcal{F}[-\delta(t+2) + 2\delta(t+1) - 2\delta(t-1) + \delta(t-2)] = -4\pi^2 f^2 X(f)$$

But, the LHS of the above is

$$= -e^{-j4\pi f} + 2e^{-j2\pi f} - 2e^{j2\pi f} + e^{j4\pi f}$$

$$\therefore \quad 4\pi^2 f^2 X(f) = (e^{j4\pi f} - e^{-j4\pi f}) - 2(e^{j2\pi f} - e^{-j2\pi f})$$

$$= 2j\sin 4\pi f - 4j\sin 2\pi f$$

$$\therefore \quad X(f) = \frac{-1}{j2\pi^2 f^2}[\sin 4\pi f - 2\sin 2\pi f]$$

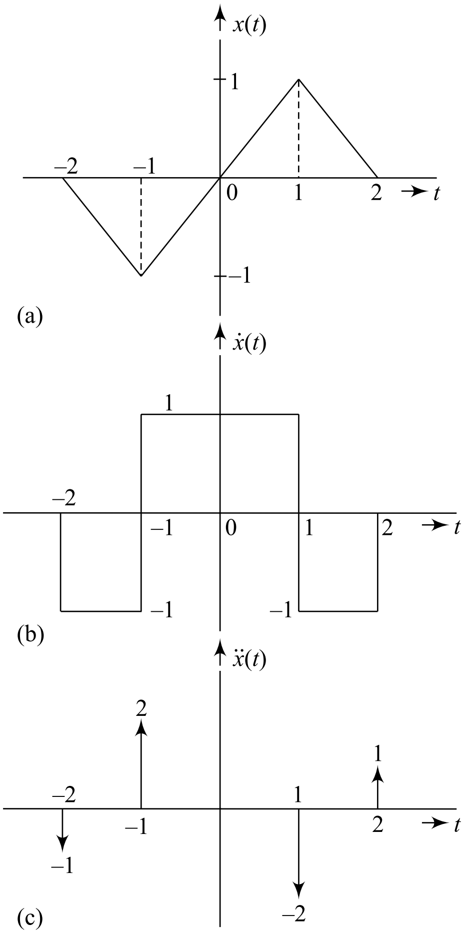


Fig. 3.10 *(a) x(t), (b), (c) ẍ(t)*

Example 3.6 Find the signal $f(t)$ if its Fourier transform $F(\omega)$ is as shown in Figs. 3.11(a) and (b).

Solution We know that

$$F(\omega) = |F(\omega)|\, e^{j\theta(\omega)}$$

Here, $|F(\omega)| = \pi$ for $|\omega| \le W$ and $\theta(\omega) = \begin{cases} \pi/2 & \text{for } \omega<0 \\ -\pi/2 & \text{for } \omega>0 \end{cases}$

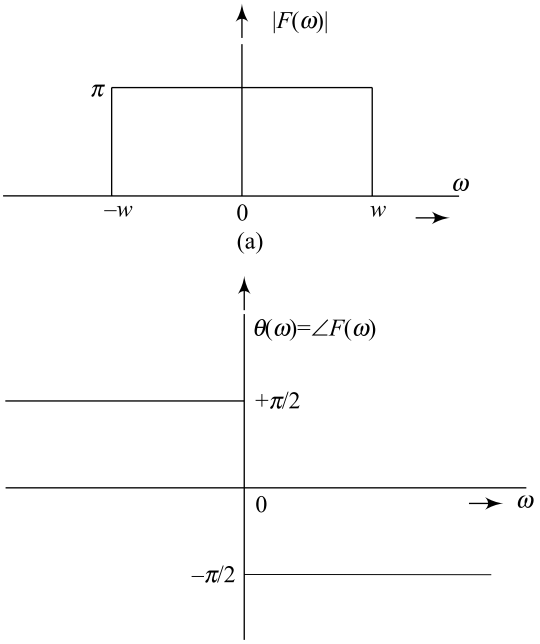
The inverse Fourier transform of $F(\omega)$, say $f(t)$, is given by

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega) e^{j2\pi ft}\, d\omega$$

$$\frac{1}{2\pi}\left[\int_{-w}^{0} \pi e^{j\pi/2}\cdot e^{j\omega t}\, d\omega + \int_{0}^{w} \pi e^{-j\pi/2}\cdot e^{\omega t}\, d\omega\right]$$

On simplification, this gives

$$f(t) = \left[\frac{1-\cos Wt}{t}\right]$$



(a)

(b)

Fig. 3.11

Example 3.7 If $x(t) \xleftrightarrow{\text{FT}} X(f)$, find the Fourier transforms of the following signals in terms of $X(f)$:

(a) $x(t-2)e^{jt}$ (b) $x(1-t)$ (c) $x\left(\frac{t}{2}-2\right)$

Solution

(a) Let $x_1(t) = x(t-2)e^{jt}$. Then $x(t-2) \xleftrightarrow{\text{FT}} X(f)e^{-j4\pi f}$ and $x_1(t) = x(t-2)e^{j2\pi(1/2\pi)t}$

$$\therefore \quad x_1(t) \xleftrightarrow{\text{FT}} X\left(f - \frac{1}{2\pi}\right)e^{-j4\pi\left(f-\frac{1}{2\pi}\right)}$$

(From modulation theorem)

(b) Let $x_2(t) = x(1-t)$. Now, $x(t) \xleftrightarrow{\text{FT}} X(f)$. Hence,

$$x(-t) \xleftrightarrow{\text{FT}} X(-f) \quad \text{and} \quad x(-t+1) \xleftrightarrow{\text{FT}} X(-f)e^{-j2\pi f}$$

(Time-delay theorem)
(From scaling theorem)

(c) Let $x_3(t) = x\left(\frac{t}{2}-2\right)$. Since $x(t) \xleftrightarrow{\text{FT}} X(f)$

From time-delay theorem, we have $x(t-2) \xleftrightarrow{\text{FT}} X(f)e^{-j4\pi f}$

and from scaling theorem, $x\left(\frac{t}{2}-2\right) \xleftrightarrow{\text{FT}} 2X(2f)e^{-j8\pi f}$

Example 3.8 Find the Fourier transform of $x(t)=\left(\dfrac{1}{1+t^2}\right)$.

Solution We know that $e^{-|t|}\xleftrightarrow{\text{FT}}\dfrac{2}{1+4\pi^2 f^2}$

Now, applying duality theorem

$$\frac{2}{1+4\pi^2 t^2}\xleftrightarrow{\text{FT}} e^{-|f|}$$

If we let $y(t)=\dfrac{2}{1+4\pi^2 t^2}$ and $y(at)=\dfrac{2}{1+t^2}=2x(t)$

$$y(at)=\frac{2}{1+4\pi^2 a^2 t^2}=\frac{2}{1+t^2}$$

$$\therefore \qquad 4\pi^2 a^2 = 1 \text{ and } a=\frac{1}{2\pi}$$

$$\therefore \quad \text{If } Y(f)=e^{-|f|},\quad 2X(f)=\mathcal{F}[y(at)]=\frac{1}{a}Y(f/a)$$

$$\therefore \quad 2X(\text{f}) = 2\pi Y(2\pi f) \qquad \therefore \quad X(f)=\pi e^{-|\omega|}$$

Example 3.9 Find the Fourier transform of the periodic GATE waveform shown in Fig. 3.12.

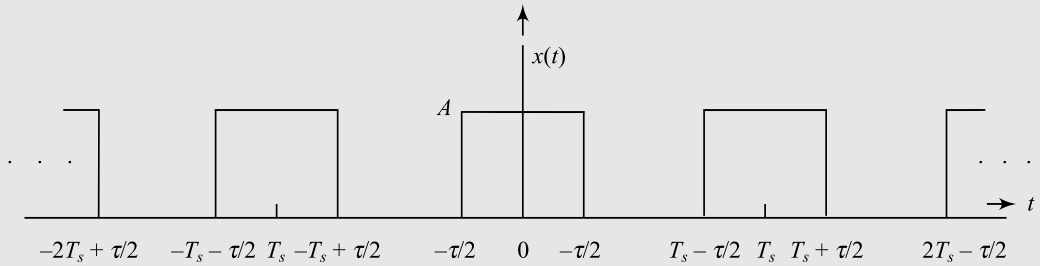


Fig. 3.12 *Periodic GATE signal*

Solution The periodic GATE signal is obtained by a periodic repetition of the GATE signal shown in Fig. 3.1 at regular intervals of say, T_s seconds where $T_s > \tau$. If the GATE signal of Fig. 3.1 is represented by $p(t)$, then we may write

$$x(t)=\sum_{k=-\infty}^{+\infty} p(t-kT_s)$$

Since $x(t)$ is a periodic signal, we can use its Fourier series expansion and write

$$x(t)=\sum_{k=-\infty}^{\infty} p(t-kT_s)=\sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t})$$

where $f_s \triangleq \dfrac{1}{T_s}$ and $c_n=\dfrac{1}{T_s}\displaystyle\int_{-T_s/2}^{T_s/2} x(t)e^{-j2\pi n f_s t}\,dt$

Since $\frac{\tau}{2} < T_s/2$ and $p(t) = 0$ for $|t| \geq \tau/2$, we may write

$$c_n = \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} x(t)e^{-j2\pi n f_s t}\,dt = \frac{1}{T_s}\int_{-\infty}^{\infty} p(t)e^{-j2\pi n f_s t}\,dt = f_s P(nf_s)$$

where $P(nf_s) = P(f)|_{f=nf_s}$ and $P(f) = F[p(t)]$

But we know that $P(f) = A\tau\,\text{sinc}\,f\tau$ (see Example 3.1)

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t} = f_s \sum_{n=-\infty}^{\infty} P(nf_s)e^{j2\pi nf_s t}$$

Taking Fourier transform on both sides, applying linearity theorem and recalling that

$F[e^{j2\pi nf_s t}] = \delta(f - nf_s)$, we get

$$X(f) = \sum_{n=-\infty}^{\infty} Af_s\tau\,\text{sinc}(nf_s\tau)\,\delta(f - nf_s)$$

$$= Af_s\tau \sum_{n=-\infty}^{\infty} \text{sinc}\,(nf_s\tau)\delta(f - nf_s)$$

3.3 CONVOLUTION AND CORRELATION

3.3.1 Convolution and Correlation of Signals

Before we proceed further with the various Fourier transform theorems, it is necessary for us to discuss about two important operations – convolution and correlation of two signals, in detail. A study of convolution of two signals is important because we deal mostly with linear time-invariant systems and these systems produce an output signal by convolving the input signal with their own impulse response. Similarly, correlation operation assumes importance because the correlation operation performed on a pair of signals, reveals the degree of similarity between the two signals. It is an operation which is widely used in communication engineering and radars.

Convolution The convolution of two continuous-time signals $x(t)$ and $y(t)$, represented by the notation $x(t) * y(t)$ is given by

$$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)\,d\tau \tag{3.40}$$

By a change of variable, the above integral, generally referred to as the convolution integral, may also be written as

$$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(t-\tau)y(\tau)\,d\tau \tag{3.41}$$

In Eqs. (3.40) and (3.41), τ is a dummy variable; and since the integration is performed for all values of τ, the result of integration is a function only of '*t*', and is denoted as $z(t)$.

Properties of convolution

1. Commutative property: $x(t) * y(t) = y(t) * x(t)$

2. Associative property: $[x(t) * y(t)] * z(t) = x(t) * [y(t) * z(t)]$

3. Distributive property: $x(t) * [y(t) + z(t)] = x(t) * y(t) + x(t) * z(t)$

4. Linearity Property: If $x(t) * y(t) = w(t), x(t) * z(t) = r(t)$ and if a and b are any two arbitrary constants, then

$$x(t) * [ay(t) + bz(t)] = aw(t) + br(t)$$

5. $x(t) * \delta(t - \tau) = x(t - \tau)$ . . . **(Replication property of $\delta(t)$):**
As we know, $\delta(t - \tau)$ is a unit-impulse function located at $t = \tau$. This property tells us that when $x(t)$ is convolved with a unit impulse located at $t = \tau$, the function $x(t)$ is shifted by τ sec (to the right, if $\tau > 0$).

Proof We know that $$x(t) * y(t) = \int_{-\infty}^{\infty} x(\lambda) y(t - \lambda) d\lambda$$

$$\therefore \quad x(t) * \delta(t - \tau) = \int_{-\infty}^{\infty} x(\lambda) \delta(t - \tau - \lambda) d\lambda$$

But we know from the defining equation of a delta function, that the above integral is simply equal to $x(t - \tau)$ [see Eq. (3.5)] This is a very useful result and is used quite often.

6. If $z(t) = x(t) * y(t)$, then $\dot{z}(t) = \dot{x}(t) * y(t) = x(t) * \dot{y}(t)$. This may easily be proved using the 'differentiation theorem of Fourier transform', which we are going to discuss a little later.

Example 3.10 Given $x(t) = 5 \cos t$ and $y(t) = 2e^{-|t|}$, find $x(t) * y(t)$.

Solution Let $z(t) = x(t) * y(t)$

$$\therefore \quad z(t) = \int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau$$

But $y(t) = 2e^{-|t|}$ so that $$y(t - \tau) = 2e^{-|t-\tau|} = \begin{cases} 2e^{t-\tau} & \text{for } t < \tau \\ 2e^{-(t-\tau)} & \text{for } t > \tau \end{cases}$$

$$\therefore \quad z(t) = \int_{-\infty}^{t} (5 \cos \tau)\, 2e^{-(t-\tau)} d\tau + \int_{t}^{\infty} (5 \cos \tau)\, 2e^{(t-\tau)} d\tau$$

$$= 5\{\cos t + \sin t] + 5[\cos t - \sin t] = 10 \cos t; \quad \infty < t < \infty$$

Correlation between two continuous-time energy signals Correlation is an operation between two signals and it gives us the degree of similarity between the two signals. In Eq. (2.21) of Chapter 2, we had shown that the component of $x(t)$ along the signal $y(t)$, may be written as

$$\frac{\int_{t_1}^{t_2} y^*(t) x(t) dt}{\int_{t_1}^{t_2} |y(t)|^2 dt} \cdot y(t) \tag{3.42}$$

Note *In case x(t) and y(t) are real-valued signals, the complex conjugation, represented by the symbol '*' in the above equations, may be ignored.*

If the two energy signals, $x(t)$ and $y(t)$ are such that

$$\int_{t_1}^{t_2} y^*(t) x(t) dt = 0 \tag{3.43}$$

we say that the two signals are orthogonal to each other and there is no similarity between them in the interval t_1 to t_2. But, we are generally interested in their similarity over the entire interval from $-\infty$ to $+\infty$ and therefore, we may think of using the following integral:

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt \tag{3.44}$$

But there is a problem in straight away using the above equation. To understand this problem, consider $x(t)$ and $y(t)$ shown in Figs. 3.13 and 3.14. The two signals are exactly identical, except that $y(t)$ is a time-delayed version of $x(t)$.

Fig. 3.13 *Signal x(t)*

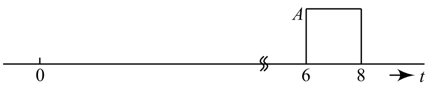


Fig. 3.14 *Signal y(t)*

If we straight away apply Eq. (3.44) to them, the integral reduces to zero forcing us to conclude that there is no similarity between them! But we know that they are exactly similar, but for the time delay.

Hence, to overcome the above problem, let us introduce a sliding or lag, parameter τ, and modify Eq. (3.44) as follows:

$$\int_{-\infty}^{+\infty} x(t)y^*(t-\tau)dt \tag{3.45}$$

Since the integration is performed over the entire range of values of t, the above integral yields a function of only τ, the lag parameter. Hence, let us write

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} x(t)y^*(t-\tau)dt \tag{3.46}$$

τ has been called the *sliding parameter*, or *lag parameter* because, if $\tau > 0$, as τ increases, $y(t)$ slides along the time axis to the right, and if $\tau < 0$, as τ increases, $y(t)$ slides to the left. Thus, in Eq. (3.46) we are keeping $x(t)$ fixed and sliding $y(t)$ and for each value of the sliding parameter τ, we are finding out the area under the product of $x(t)$ and the shifted $y(t)$. Obviously, $R_{xy}(\tau)$ takes a maximum value when the shifted version of $y(t)$ has maximum overlap with $x(t)$. For the $x(t)$ and $y(t)$ shown in Figs. 3.13 and 3.14, this happens when $\tau = -6$. By putting $(t - \tau) = \lambda$, Eq. (3.46) may be rewritten as

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)dt \tag{3.47}$$

Here, $R_{xy}(\tau)$ is called the cross-correlation between the signals $x(t)$ and $y(t)$, for a lag of τ sec. If $x(t)$ and $y(t)$ have some similarity as in the case of the signal shown in Figs. 3.13 and 3.14, then $R_{xy}(\tau)$ will be non-zero at least for some values of lag parameter τ. If however $R_{xy}(\tau)$ is zero for all values of τ, it means that $x(t)$ and $y(t)$, have no similarity and we can say that the two signals have no correlation or that they are un-correlated.

Symmetry properties of cross-correlation From Eq. (3.47), if $x(t)$ and $y(t)$ are complex-valued signals,

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)dt \tag{3.48}$$

Replacing τ by $-\tau$ in Eq. (3.47), we have

$$R_{xy}(-\tau) = \int_{-\infty}^{\infty} x(t-\tau)y^*(t)dt$$

$$= \int_{-\infty}^{\infty} x(t)y^*(t+\tau)dt \tag{3.49}$$

But
$$R_{yx}(\tau) = \int_{-\infty}^{\infty} y(t+\tau)x^*(t)\,dt \tag{3.50}$$

Comparing Eq. (3.49) with Eq. (3.50), we find that

$$\boxed{R_{yx}(\tau) = R_{xy}^*(-\tau) \quad \text{for complex-valued signals}} \tag{3.51}$$

In case the two signals are real valued, it is clear that

$$\boxed{R_{yx}(\tau) = R_{xy}(-\tau) \quad \text{for real-valued signals}} \tag{3.52}$$

When $y(t)$ is the same as $x(t)$, the correlation is of a signal $x(t)$ with itself and therefore it is called as 'auto-correlation' and denoted by $R_{xx}(\tau)$ or simply $R_x(\tau)$. Thus,

$$\left.\begin{aligned} R_{xx}(\tau) &= \int_{-\infty}^{\infty} x(t)x(t-\tau)\,dt \\ &= \int_{-\infty}^{\infty} x(t+\tau)x(t)\,dt \end{aligned}\right\} \text{if } x(t) \text{ is real-valued} \tag{3.53}$$

and

$$\left.\begin{aligned} R_{xx}(\tau) &= \int_{-\infty}^{\infty} x(t)x^*(t-\tau)\,dt \\ &= \int_{-\infty}^{\infty} x(t+\tau)x^*(t)\,dt \end{aligned}\right\} \text{if } x(t) \text{ is complex-valued} \tag{3.54}$$

From the above, we find that

$$\boxed{R_{xx}(-\tau) = R_{xx}(\tau) \text{ if } x(t) \text{ is real}} \tag{3.55}$$

and
$$\boxed{R_{xx}(-\tau) = R_{xx}^*(\tau) \text{ if } x(t) \text{ is complex}} \tag{3.56}$$

Thus, if $x(t)$ is real valued, its auto-correlation function has even symmetry. But if $x(t)$ is complex valued, then its auto-correlation function has Hermitian symmetry.

Another important property of the auto-correlation is

$$\boxed{R_{xx}(0) > |R_{xx}(\tau)|;\ \tau \neq 0} \tag{3.57}$$

(*For a rigorous proof, refer to Ref.* 3.)

Equation (3.57) says that the auto-correlation of a signal $x(t)$ takes the maximum value for zero lag. This is obvious, since the overlap between $x(t)$ and $x(t + \tau)$ is maximum when $\tau = 0$. Further in that case, $R_{xx}(0)$ represents the energy of the signal $x(t)$.

Correlation between two continuous-time power signals The discussion so far was confined to correlation of continuous-time energy signals. But periodic signals of the deterministic type and random signals are not energy signals – they are power signals. Since the energy of these signals over an infinite time interval is not finite it would be more appropriate to define the cross-correlation of two power signals, $x(t)$ and $y(t)$, as

$$R_{xy}(\tau) \underline{\Delta} \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y^*(t-\tau)\,dt \tag{3.58}$$

and auto-correlation of $x(t)$ as

$$R_{xx}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x^*(t-\tau)\,dt \tag{3.59}$$

However, in the case of power signals that are periodic deterministic signals, the average over an infinite interval and average over one period will be the same. Hence, if $x(t)$ and $y(t)$ are periodic with period T_0, we may write

$$R_{xy}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t)y^*(t-\tau)\,dt \tag{3.60}$$

$$R_{xx}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t)x^*(t-\tau)\,dt \tag{3.61}$$

Remark *Since $x(t)$ and $y(t)$ both have a period T_0, the integrands of Eqs. (3.60) and (3.61) are also periodic with period T_0. Thus, auto-correlation of a periodic signal, and cross-correlation of two periodic signals with the same period will be periodic with the same period.*

Example 3.11 Find $R_{xx}(\tau)$ if $x(t) = e^{-t}u(t)$.

Solution $R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt = \int_{-\infty}^{\infty} e^{-t}u(t)e^{-(t-\tau)}u(t-\tau)\,dt$

(a) *When $\tau > 0$*

$$u(t)u(t-\tau) = \begin{cases} 1 & \text{for } t > \tau \\ 0 & \text{for } t < \tau \end{cases}$$

$\therefore$
$$R_{xx}(\tau) = \int_{\tau}^{\infty} e^{-t}e^{-(t-\tau)}dt = e^{\tau}\int_{\tau}^{\infty} e^{-2t}dt = \frac{1}{2}e^{-\tau}u(\tau)$$

(b) *When $\tau < 0$*

$$u(t)u(t-\tau) = \begin{cases} 1 & \text{for } t \ge 0 \\ 0 & \text{for } t < 0 \end{cases}$$

$\therefore$
$$R_{xx}(\tau) = \int_{0}^{\infty} e^{-t}e^{-(t-\tau)}dt = \int_{-\infty}^{\infty} e^{-2t}dt = \frac{1}{2}e^{\tau}u(-\tau)$$

Combining the two results, we may write

$$R_{xx}(\tau) = \frac{1}{2}e^{-|\tau|}$$

Example 3.12 If $x(t) = A\cos(\omega_0 t + \theta)$, find $R_{xx}(\tau)$.

Solution $R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} A^2\cos(\omega_0 t + \theta)\cos[\omega_0(t-\tau)+\theta]\,dt$

Put $\phi = \omega_0 t + \theta$

$$\therefore \quad R_{xx}(\tau) = \frac{A^2}{T_0} \cdot \frac{1}{\omega_0} \int_{-\pi+\theta}^{\pi+\theta} \cos\phi[\cos\phi\cos\omega_0\tau + \sin\phi\sin\omega_0\tau]d\phi$$

$$= \frac{A^2}{4\pi}\cos\omega_0\tau \int_{-\pi+\theta}^{\pi+\theta}(1+\cos 2\phi)d\phi + \frac{A^2}{4\pi}\sin\omega_0\tau \int_{-\pi+\theta}^{\pi+\theta} \sin 2\phi \, d\phi$$

$$= \frac{A^2}{4\pi}\cos\omega_0\tau\,[\pi+\theta+\pi-\theta] = \frac{A^2}{2}\cos\omega_0\tau$$

Therefore, if $x(t)$ is periodic with a period of $T_0 = \left(\frac{2\pi}{\omega_0}\right)$, we find that $R_{xx}(\tau)$ is periodic in τ with the same period T_0.

3.3.2 Relationship between Convolution and Correlation

There is a close resemblance between convolution and correlation operations. In view of this, we shall examine the relationship between the two. For this purpose, consider two signals, $x(t)$ and $y(t)$.

Correlation:
$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y(t-\tau)dt \tag{3.62}$$

Convolution:
$$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau \tag{3.63}$$

If we now define $\therefore$
$$w(t) \underline{\Delta}\, x(t) * y(-t) = \int_{-\infty}^{\infty} x(\tau)y(\tau-t)d\tau \tag{3.64}$$

Then, replacing the dummy variable τ in Eq. (3.64) by u, we have

$$w(t) = \int_{-\infty}^{\infty} x(u)y(u-t)du \tag{3.65}$$

In the cross-correlation Eq. (3.62), if t is replaced by u

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(u)y(u-\tau)du \tag{3.66}$$

A comparison of Eqs. (3.65) and (3.66) reveals that

$$R_{xy}(\tau) = [x(t) * y(-t)]\big|_{t\to\tau} \tag{3.67}$$

and
$$R_{xx}(\tau) = [x(t) * x(-t)]\big|_{t\to\tau} \tag{3.68}$$

We had earlier seen that the Fourier transform provides a very powerful tool for the computation of convolution. From the above equations, it is clear that it can as well be used for the computation of correlation too.

Note *Since we are replacing y(t) by y(–t) in the convolution in order to get the cross-correlation, if y(t) has even symmetry with respect to 't', the two operations of convolution and correlation become one and the same.*

3.3.3 Energy Spectral Density (ESD)

Consider two energy signals $x(t)$ and $y(t)$. Let $z(t) = x(t) * y(t)$ and let their cross-correlation for a lag τ *be* $R_{xy}(\tau)$. Then,

$$z(t) = x(t) * y(t) \tag{3.69}$$

$$Z(f) = X(f) \cdot Y(f) \tag{3.70}$$

Then from Eq. (3.67), we know that

$$\mathcal{F}[R_{xy}(\tau)] = \mathcal{F}[x(t)] \cdot \mathcal{F}[y(-t)] \tag{3.71}$$

i.e.,

$$\mathcal{F}[R_{xy}(\tau)] = X(f) \cdot Y(-f) \underline{\Delta} S_{xy}(f) \tag{3.72}$$

In a similar manner, we have

$$\mathcal{F}[R_{yx}(\tau)] \underline{\Delta} S_{yx}(f) = Y(f) \cdot X(-f) \tag{3.73}$$

∴ If $R_{xx}(\tau)$ is the auto-correlation function of $x(t)$, then

$$\mathcal{F}[R_{xx}(\tau)] = X(f) \cdot X(-f) = |X(f)|^2 \tag{3.74}$$

But, from Parseval's theorem for Fourier transform, we know that (refer to Eq. (3.15)) $|X(f)|^2$ represents the energy density of $x(t)$ with respect to frequency and is called the 'Energy Spectral Density' (ESD). It shows how the energy of $x(t)$ is distributed with respect to frequency, and is denoted by $S_{xx}(f)$.

∴

$$\mathcal{F}[R_{xx}(\tau)] = S_{xx}(f) = |X(f)|^2 \tag{3.75}$$

The above equation tells us that for an energy signal, $x(t)$, its auto-correlation function $R_{xx}(\tau)$ and its energy spectral density (ESD) denoted by $S_{xx}(f)$, are a Fourier transform pair.

i.e.,

$$\boxed{R_{xx}(\tau) \xleftrightarrow{\text{FT}} S_{xx}(f)} \tag{3.76}$$

This relationship is generally referred to as the auto-correlation theorem and may be derived directly as follows.

$$\mathcal{F}[R_{xx}(\tau)] = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f\tau} d\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [x(t)x(t-\tau)dt] e^{-j2\pi f\tau} d\tau$$

$$= \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \cdot \int_{-\infty}^{\infty} x(t-\tau) e^{-j2\pi f(\tau - t)} d\tau$$

Putting $\lambda = (t - \tau)$

$$\mathcal{F}[R_{xx}(\tau)] = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \cdot \int_{-\infty}^{\infty} x(\lambda) e^{-j2\pi(-f)\lambda} d\tau$$

$$= X(f) \cdot X(-f) = |X(f)|^2 = S_{xx}(f)$$

Since $R_{xx}(\tau) \xleftrightarrow{\text{FT}} S_{xx}(f)$, we have

$$\int_{-\infty}^{\infty} S_{xx}(f) e^{j2\pi f\tau} df = R_{xx}(\tau)$$

Putting $\tau = 0$ on both sides of the above equation

$$R_{xx}(0) = \text{ACF for zero lag} = \int_{-\infty}^{\infty} S_{xx}(f) df$$

= Area under the energy spectral density function
= Energy of $x(t)$.

∴

$$\boxed{R_{xx}(0) = \text{Energy of } x(t), \text{an energy signal}} \tag{3.77}$$

3.3.4 Power Spectral Density (PSD)

In the foregoing, we have considered energy signals which have a finite amount of energy over $-\infty < t < \infty$, and we have shown that the Fourier transform of the auto-correlation of such signals gives the energy spectral density.

Now, we shall consider signals that do not have a finite energy over the interval $-\infty < t < \infty$. Periodic signals and random signals come under this category. Since we cannot talk about the Fourier transforms of such signals, if $x(t)$ is a deterministic power signal, let us take a segment of it of duration T seconds. This segment will have a finite amount of energy. Specifically, let

$$x_T(t) \,\underline{\Delta} \begin{cases} x(t); & \text{for } |t| \le T/2 \\ 0 \;\; ; & \text{otherwise} \end{cases} \tag{3.78}$$

Thus, $x_T(t)$ is a finite energy signal and hence, is Fourier transformable.

Let
$$x_T(t) \xleftrightarrow{\text{FT}} X_T(f) \tag{3.79}$$

We know that $|X_T(f)|^2$ represents the ESD of the signal $x_T(t)$. Since the duration of the signal $x_T(t)$ is T seconds, we may define the average power spectral density of $x_T(t)$ as

$$P_{x_T x_T}(f) = \frac{|X_T(f)|^2}{T} \tag{3.80}$$

and the average power spectral density of $x(t)$ as

$$P_{xx}(f) = \underset{T\to\infty}{\text{Lt}} \left[\frac{|X_T(f)|^2}{T}\right] \tag{3.81}$$

Recalling that the auto-correlation function of a real-valued power signal has been defined as

$$R_{xx}(\tau) = \underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau)\,dt$$

and taking the Fourier transform of the above on both sides

$$\begin{aligned}
\mathcal{F}[R_{xx}(\tau)] &= \mathcal{F}\left[\underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau)\,dt\right] \\
&= \underset{T\to\infty}{\text{Lt}} \left[\mathcal{F}\left\{\frac{1}{T}\int_{-\infty}^{\infty} x_T(t)x_T(t-\tau)\,dt\right\}\right] \\
&= \underset{T\to\infty}{\text{Lt}} \left[\frac{1}{T}\mathcal{F}\left\{\int_{-\infty}^{+\infty} x_T(t)x_T(t-\tau)\,dt\right\}\right] = \underset{T\to\infty}{\text{Lt}} \frac{1}{T}\mathcal{F}[R_{x_T x_T}(\tau)] \\
&= \underset{T\to\infty}{\text{Lt}} \left[\frac{|X_T(f)|^2}{T}\right] = P_{xx}(f) \qquad \text{(From Eq. (3.81))}
\end{aligned}$$

∴ Power spectral density of a power signal is the Fourier transform of the auto-correlation of the signal.

$$\boxed{R_{xx}(\tau) \xleftrightarrow{\text{FT}} P_{xx}(f)} \tag{3.82}$$

Example 3.13 Find the ACF and ESD of the signal $x(t) = e^{-t}u(t)$.

Solution We know, from Eq. (2.119) that

$$R_{xx}(\tau) = [x(t) * x(-t)]\big|_{t\to\tau}$$

$\therefore$ $$S_{xx}(f) = X(f)\cdot X(-f) = |X(f)|^2$$

But $$X(f) = \mathcal{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt = \int_{0}^{\infty} e^{-(1+j\omega)t}dt = \frac{1}{1+j\omega}$$

$\therefore$ $$|X(f)|^2 = \frac{1}{1+\omega^2} = S_{xx}(f), \text{ the ESD of } x(t).$$

Thus, $$R_{xx}(\tau) = \mathcal{F}^{-1}[S_{xx}(f)] = \mathcal{F}^{-1}\left[\frac{1}{1+\omega^2}\right] = \frac{1}{2}e^{-|\tau|}$$

Example 3.14 Find the ACF and ESD of the signal $x(t) = A\Pi(t/2T)$.

Solution $$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)\,dt$$

and $$R_{xx}(\tau) = [x(t) * x(-t)]\big|_{t\to\tau}$$

Since $x(t)$ has even symmetry,

$$R_{xx}(\tau) = [x(t) * x(t)]\big|_{t\to\tau}$$

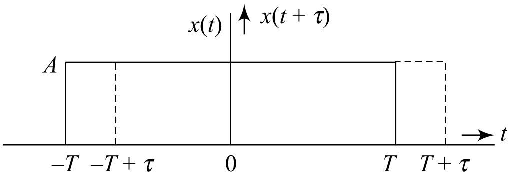


Fig. 3.15 *x(t) of Example 3.14*

$\therefore$ $$S_{xx}(f) = \mathcal{F}[x(t) * x(t)] = |X(f)|^2 = 4A^2T^2 \operatorname{sinc}^2(2fT)$$

$\therefore$ $$R_{xx}(f) = \mathcal{F}^{-1}[4A^2T^2 \operatorname{sinc}^2(2fT)] = A^2[2T - |t|]$$

$R_{xx}(\tau)$ may however be determined directly as

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t) * x(-t)\,dt = \int_{-T+\tau}^{T} A^2 dt = A^2(2T - \tau)$$

However, since $x(t)$ is real valued, $R_{xx}(\tau)$ must have even symmetry with respect to τ.

$\therefore$ $$R_{xx}(\tau) = A^2(2T - |\tau|)$$

Hence, $R_{xx}(\tau)$ is a triangular waveform as shown in Fig. 3.16.

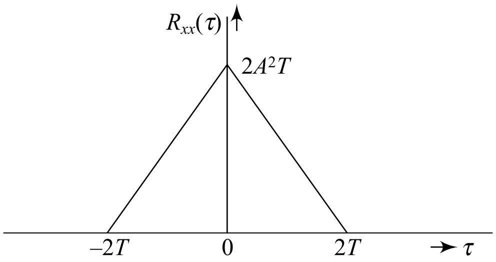


Fig. 3.16 *$R_{xx}(\tau)$ of AP(t/2T)*

Properties of power spectral density

1. $P_{xx}(f)$ of a signal is always non-negative, since

$$P_{xx}(f) = \underset{T\to\infty}{\text{Lt}}\left[\frac{|X_T(f)|^2}{T}\right]$$

2. $P_{xx}(f)$ is the Fourier transform of $R_{xx}(\tau)$
3. The total area under the PSD curve of a signal equals the average power of the signal

since $\therefore$ $$R_{xx}(\tau) = \mathcal{F}^{-1}[P_{xx}(f)] = \int_{-\infty}^{\infty} P_{xx}(f)e^{j2\pi f\tau}\,df$$

$\therefore$ $$R_{xx}(0) = \int_{-\infty}^{\infty} P_{xx}(f)df = \text{area under the PSD curve}$$

But $$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)x(t-\tau)\,dt$$

$\therefore$ $$R_{xx}(0) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x^2(t)\,dt = \text{Average power of } x(t)$$

$\therefore$ $$\boxed{\int_{-\infty}^{\infty} P_{xx}(f)\,df = \text{Average power } P_{\text{av}} \text{ of } x(t)} \tag{3.83}$$

4. The power spectral density of a real-valued power signal $x(t)$ is an even function of frequency, i.e., $P_{xx}(-f) = P_{xx}(f)$, if $x(t)$ is real valued.

Proof We know that for a real-valued power signal $x(t)$

$$|X_T(-f)| = |X_T(f)|$$

$$P_{xx}(-f) = \underset{T\to\infty}{\text{Lt}}\left[\frac{|X_T(-f)|^2}{T}\right] = \underset{T\to\infty}{\text{Lt}}\left[\frac{|X_T(f)|^2}{T}\right] = P_{xx}(f)$$

Thus, $P_{xx}(f)$ is an even function of frequency.

Output ESD and PSD of Linear time-invariant (LTI) systems An LTI system is characterized by its impulse response function, $h(t)$, in the sense that for any given input signal $x(t)$, the output signal $y(t)$ is given by

$$y(t) = x(t) * h(t)$$

Taking Fourier transform on both sides, we have

$$Y(f) = X(f)\cdot H(f) \tag{3.84}$$

This equation clearly shows that the spectrum of the input signal gets modified during its passage through the LTI system. We shall now examine the way the ESD of the output signal is related to the ESD of the input signal $x(t)$, when $x(t)$ is an energy signal. We shall also examine how the PSD of the output signal is related to the PSD of the input signal when the input signal is a power signal.

Relationship between input and output ESDs Let $x(t)$ be an energy signal. From Eq. (3.84), we have

$$|Y(f)|^2 = |X(f)\cdot H(f)|^2 = |X(f)|^2\cdot|H(f)|^2$$

But we know that $|Y(f)|^2$ and $|X(f)|^2$ represent, respectively, the ESDs of output and input signals.

$\therefore$ $$|Y(f)|^2 = S_{yy}(f) = |X(f)|^2\cdot|H(f)|^2 = S_{xx}(f)\cdot|H(f)|^2$$

Hence, $$\boxed{S_{yy}(f) = |H(f)|^2\cdot S_{xx}(f)} \tag{3.85}$$

Relation between input and output PSDs Let us now assume that the input signal $x(t)$, is a power signal. Also, let

$$x_T(t) = w(t)\cdot x(t) \tag{3.86}$$

where $w(t)$ is a rectangular window function defined by

$$w(t) = \begin{cases} 1 & \text{for } |t| \le T \\ 0 & \text{otherwise} \end{cases} \tag{3.87}$$

Then, we know that

$$P_{xx}(f) = \underset{T\to\infty}{\text{Lt}}\left[\frac{|X_T(f)|^2}{T}\right]$$

where $$X_T(f) = \mathcal{F}[x_T(t)],$$

and that $$P_{yy} = \underset{T\to\infty}{\text{Lt}}\left[\frac{|Y_T(f)|^2}{T}\right]$$

In the above equation, $$Y_T(f) = H(f)\cdot X_T(f)$$

Then, it follows from the above expression for $P_{xx}(f)$ and $P_{yy}(f)$, that

$$\boxed{P_{yy}(f) = |H(f)|^2 \cdot P_{xx}(f)} \tag{3.88}$$

Example 3.15 $x(t) = e^{-t/\tau}u(t)$ is applied as input to an L-section high pass RC filter with a time constant of τ sec. Find the ESD of the output of the filter. Express the output signal energy as a percentage of the input signal energy.

Solution Transfer function $H(f)$ of the RC filter is

$$H(f) = \frac{j2\pi fRC}{1+j2\pi fRC} = \frac{j\omega\tau}{1+j\omega\tau} \text{ since } \tau = \text{RC}$$

$$\therefore \quad |H(f)|^2 = \frac{\omega^2\tau^2}{1+\omega^2\tau^2}$$

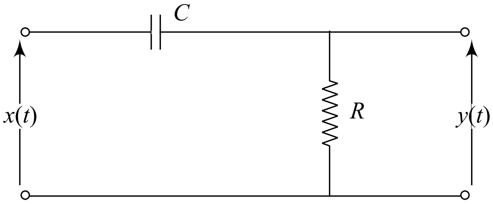


Fig. 3.17 *An L-section high pass RC filter*

$\therefore$ If input energy spectral density is $S_{xx}(f)$, and output ESD is $S_{yy}(f)$, then

$$S_{yy}(f) = \frac{\omega^2\tau^2}{1+\omega^2\tau^2}\cdot S_{xx}(f)$$

But $$S_{xx}(f) = |X(f)|^2, \quad \text{where} \quad X(f) = \int_{-\infty}^{\infty} e^{-t/\tau}u(t)e^{-j\omega t}dt = \frac{\tau}{1+j\omega\tau}$$

$$\therefore \quad |X(f)|^2 = \frac{\tau^2}{1+\omega^2\tau^2} \quad \text{and} \quad E_x = \int_{-\infty}^{\infty}|X(f)|^2 df = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\tau^2}{1+\omega^2\tau^2}d\omega$$

Putting $\omega\tau = \tan\theta$, $1+\omega^2\tau^2 = \sec^2\theta$ and $d\omega = \frac{1}{\tau}\sec^2\theta\, d\theta$

$$\therefore \quad E_x = \text{Energy in the input signal } = \frac{\tau^2}{2\pi}\int_{-\pi/2}^{+\pi/2}\frac{1}{\tau}\,d\theta = \frac{\tau}{2}$$

$$S_{xx}(f) = |X(f)|^2|H(f)|^2 = \frac{\omega^2\tau^4}{(1+\omega^2\tau^2)^2}$$

$$\therefore \quad E_y = \text{Total energy at the output of the filter } \int_{-\infty}^{\infty} S_{xx}(f)df$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\omega^2\tau^4}{(1+\omega^2\tau^2)^2}d\omega$$

Substituting tan θ for $\omega\tau$ and performing the above integration, we get $E_y = \tau/4$

$$\therefore \qquad \frac{E_y}{E_x}\times 100\% = \frac{\tau/4}{\tau/2}\times 100 = 50\%$$

Example 3.16 The signal $x(t) = 10\cos(4\pi\times 10^3)t$ is given as input to an L-section low pass RC filter having 3 db cut-off frequency of 10^3 Hz. Determine and sketch the output PSD.

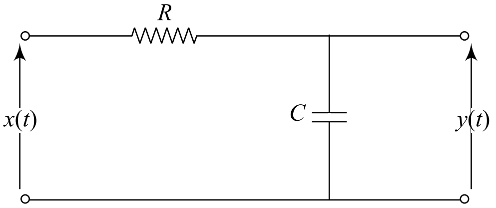


Fig. 3.18 *An L-section RC low pass filter*

Solution First, let us find $P_{xx}(f)$, i.e., the PSD of the input signal. But

$$P_{xx}(f) = \mathcal{F}[R_{xx}(\tau)]$$

Now to find $R_{xx}(\tau)$

$$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} A^2\cos(\omega_0 t+\theta)\cos[\omega_0(t-\tau)+\theta]\,dt$$

Putting $\phi = \omega_0 t + \theta$ and performing the above integration, we get

$$R_{xx}(\tau) = \frac{A^2}{2}\cos\omega_0 t = \frac{10^2}{2}\cos(4\pi\times 10^3\tau)$$

$$\therefore \qquad P_{xx}(f) = \mathcal{F}[50\cos(4\pi\times 10^3)\tau] = 25[\delta(f-f_0)+\delta(f+f_0)]$$

where $f_0 = 2\times 10^3\,\text{Hz}$

Therefore, $P_{yy}(f) = |H(f)|^2\cdot P_{xx}(f) = |H(f)|^2\cdot 25[\delta(f-f_0)+\delta(f+f_0)]$

But $H(f) = \dfrac{1}{1+j\omega RC}$

$$|H(f)|^2 = \frac{1}{1+\omega^2R^2C^2}$$

Since 3-db frequency for an RC low pass filter $= \dfrac{1}{2\pi RC}$

$$\frac{1}{2\pi RC} = 10^3 \quad \therefore R^2C^2 = \frac{1}{4\pi^2\times 10^6}$$

$$\therefore \qquad P_{yy}(f) = \frac{25}{1+\omega^2\left(\dfrac{1}{4\pi^2\times 10^6}\right)}[\delta(f-f_0)+\delta(f+f_0)]$$

$$= 5[\delta(f-f_0)+\delta(f+f_0)]$$

(From sampling property of delta functions)

where $f_0 = 2\times 10^3\,\text{Hz}$

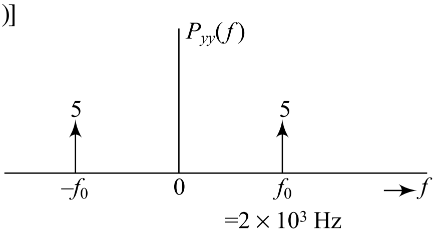


Fig. 3.19 *A sketch of $P_{yy}(f)$*

3.4 HILBERT TRANSFORM

3.4.1 Definition and Frequency-domain Representation

The Hilbert transform, $\hat{x}(t)$, *of a signal x(t), is defined as the signal obtained by convolving x(t) with* $1/(\pi t)$.

∴
$$\hat{x}(t) \underline{\Delta}\, x(t) * \frac{1}{\pi t} \tag{3.89}$$

i.e.,
$$\hat{x}(t) = \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau = \int_{-\infty}^{\infty} \frac{x(t-\tau)}{\pi\tau} d\tau \tag{3.90}$$

Note

(*i*) *This definition of Hilbert transform is applicable to all signals that are Fourier transformable.*

(*ii*) *Since all applications of Hilbert transform are concerned with real-valued signals, we shall henceforth assume that x(t) is real valued.*

The effect on $x(t)$, of Hilbert transforming it, is best understood in the frequency domain. Taking the Fourier transform on both sides of Eq. (3.89), and denoting the Fourier transform of $\hat{x}(t)$ as $\hat{X}(f)$, we have

$$\hat{X}(f) = X(f) \cdot \mathcal{F}\left[\frac{1}{\pi t}\right]$$

From Eq. (3.36), we have

$$\text{sgn}(t) \xleftrightarrow{\text{FT}} \frac{1}{j\pi f} \tag{3.91}$$

Hence, from the duality theorem of Fourier transforms,

$$\frac{1}{j\pi t} = \text{sgn}(-f) = -\text{sgn}(f), \text{ since sgn}(f) \text{ is an odd function of } f.$$

∴
$$\frac{1}{\pi t} \xleftrightarrow{\text{FT}} -j\,\text{sgn}(f) \tag{3.92}$$

Going back to Eq. (3.91), we therefore have

$$\hat{X}(f) = -j\,\text{sgn}(f)X(f) \tag{3.93}$$

But
$$\text{sgn}(f) = \begin{cases} 1 & \text{for } f > 0 \\ -1 & \text{for } f < 0 \end{cases}$$

∴
$$\frac{\hat{X}(f)}{X(f)} = \begin{cases} -j & \text{for } f > 0 \\ j & \text{for } f < 0 \end{cases} \tag{3.94}$$

Since $\hat{X}(f)$ is the spectrum of $\hat{x}(t)$ while $X(f)$ is the spectrum of $x(t)$, it follows from Eq. (3.94) that the effect of Hilbert transforming a signal $x(t)$ is merely to give a phase shift of –90° to all of the positive frequency components of $x(t)$ and a phase shift of +90° to all of its negative frequency components. Further, since

$$|-j| = |j| = 1$$

we have

$$|\hat{X}(f)| = |X(f)| \tag{3.95}$$

i.e., *Hilbert transform does not alter the magnitude spectrum.*

From Eq. (3.89), it is clear that we may visualize the Hilbert transform $\hat{x}(t)$ of a signal $x(t)$ to be the output of linear time-invariant system with an impulse response function

$$h(t) = \frac{1}{\pi t}, \tag{3.96}$$

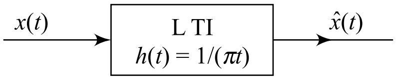
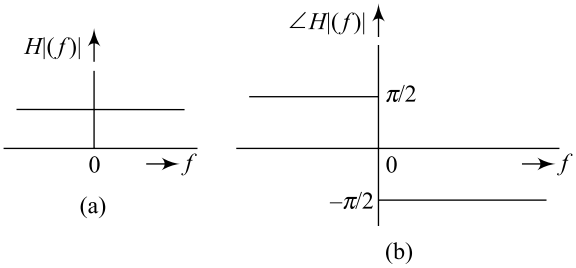


Fig. 3.20 *Hilbert transformer*

and whose input signal is $x(t)$. Such an LTI system, called the *Hilbert transformer*, will have a transfer function $H(f)$ given by

$$H(f) = \frac{\hat{X}(f)}{X(f)} = -j\,\mathrm{sgn}(f) = \begin{cases} -j & \text{for } f > 0 \\ \ \ j & \text{for } f < 0 \end{cases} \tag{3.97}$$



(a)

(b)

Fig 3.21 *(a) Magnitude response of a Hilbert transformer, (b) Phase response of a Hilbert transformer*

Properties of Hilbert Transform

1. Hilbert transform does not change the domain of a signal.
2. Hilbert transform does not alter the amplitude spectrum of a signal.
3. If $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$, then $\hat{x}(t) \xleftrightarrow{\text{HT}} -x(t) = \hat{\hat{x}}(t)$.

Proof $\hat{x}(t) \xleftrightarrow{\text{FT}} -jSgn(f)X(f)$.

Hence

$$\hat{\hat{x}}(t) \xleftrightarrow{\text{FT}} \{-jSgn(f)\}\{-jSgn(f)\}X(f)$$

$\therefore$

$$F[\hat{\hat{x}}(t)] = -\mathrm{sgn}^2(f) \cdot X(f) = -X(f)$$

$\therefore$

$$\boxed{\hat{\hat{x}}(t) = -x(t)} \tag{3.98}$$

4. A signal and its Hilbert transform are orthogonal to each other.

i.e.,

$$\int_{-\infty}^{\infty} x(t)\hat{x}(t)dt = 0$$

Proof of the above property is left as an exercise for the readers.

Remarks

From the –90° phase-shift property of Hilbert transform, it follows that

(i) $\sin \omega_0 t \xleftrightarrow{HT} -\cos \omega_0 t$ *(3.99)*

(ii) $\cos \omega_0 t \xleftrightarrow{HT} \sin \omega_0 t$ *(3.100)*

5. If $x(t)$ is a low pass signal and $y(t)$ is a high pass signal, and if their spectra are non-overlapping, then

$$x(t)y(t) = x(t) \cdot \hat{y}(t) \tag{3.101}$$

This property is extremely useful in communication engineering and may be proved as follows:

Proof Let $z(t) = x(t) \cdot y(t)$

Taking the Fourier transform of the above on both sides,

$$Z(f) = X(f) * Y(f) = \int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)d\lambda$$

$$\therefore \qquad \hat{Z}(f) = \mathcal{F}[\hat{z}(t)] = -j\,\text{sgn}(f)Z(f)$$

$$= -j\int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)\text{sgn}(f)d\lambda$$

Hence, $$\hat{z}(t) = \mathcal{F}^{-1}[\hat{Z}(f)] = -j\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)\text{sgn}(f)e^{j2\pi ft}d\lambda df$$

$$= -j\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(\lambda)e^{j2\pi\lambda t}\cdot Y(f-\lambda)e^{j2\pi(f-\lambda)t}\,\text{sgn}(f)\,dfd\lambda \qquad (3.102)$$

Here, $x(t)$ is a low pass signal, band limited to say, W Hz, Hence, the range of values of λ for which $X(\lambda)$ is non-zero, are $|\lambda| \le W$. But, $y(t)$, being a high pass signal, the range of values of 'f' for which $Y(f)$ is non-zero are typically $|f| >> W$. Hence, in the integral on the RHS of Eq. (3.102), we will be interested in small values of the variable λ and only very large values of the variable 'f'. Hence $(f - \lambda)$ in it may be replaced by 'f' without any error (as the spectra of $x(t)$ and $y(t)$ are non-overlapping) and we may re-write Eq. (3.102) as

$$\hat{z}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(\lambda)e^{j2\pi\lambda t}\cdot Y(f)e^{j2\pi ft}[-j\,\text{sgn}(f)]\,df\,d\lambda$$

$$= \int_{-\infty}^{\infty} X(\lambda)e^{j2\pi\lambda t}d\lambda \cdot \int_{-\infty}^{\infty} Y(f)e^{j2\pi ft}[-j\,\text{sgn}(f)]df$$

$$= x(t)\cdot\hat{y}(t)$$

Hence, if $x(t)$ is a low pass signal and $y(t)$ is high pass signal, and if their spectra are non-overlapping, then

$$x(t)y(t) = x(t)\cdot\hat{y}(t)$$

3.4.2 Analytic Signal or Pre-envelope

If $x(t)$ *is a real-valued signal, its analytic signal or pre-envelope is defined as*

$$x_+(t) \triangleq x(t) + j\hat{x}(t) \qquad (3.103)$$

The analytic signal, or the pre-envelope of $x(t)$ is thus a complex-valued signal, with $x(t)$ itself as its real part and the Hilbert transform of $x(t)$, as its imaginary part. It plays an important role in the representation of band pass signals and in the analysis of band pass systems.

The importance of the analytic signal stems from the nature of its spectrum. If we take the Fourier transform of both sides of Eq. (3.103) and denote the Fourier transform of $x_+(t)$ by $X_+(f)$, we have

$$X_+(f) = X(f) + j[-j\,\text{sgn}(f)X(f)]$$

$$= \begin{cases} X(f) + X(f); & f > 0 \\ X(f) - X(f); & f < 0 \end{cases}$$

$$\therefore \quad X_+(f) = \begin{cases} 2X(f) & \text{for } f > 0 \\ 0 & \text{for } f < 0 \end{cases} \qquad (3.104)$$

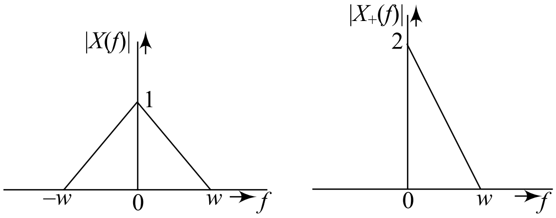


Fig. 3.22 *(a) Magnitude spectrum of* $x(t)$*, (b) Magnitude spectrum of* $x_+(t)$

3.4.3 Complex-envelope Representation of Band Pass Signals

A band pass signal is one whose spectrum is non-negligible only in a band of frequencies, occupying a width of say $2W$ Hz around a certain frequency f_c called the centre frequency with $W << f_c$. We come across band pass signals quite frequently in communication engineering. For example, a typical double sideband amplitude modulated audio broadcast signal occupies a bandwidth of about 10 KHz centered on a carrier frequency of, say, a few megahertz.

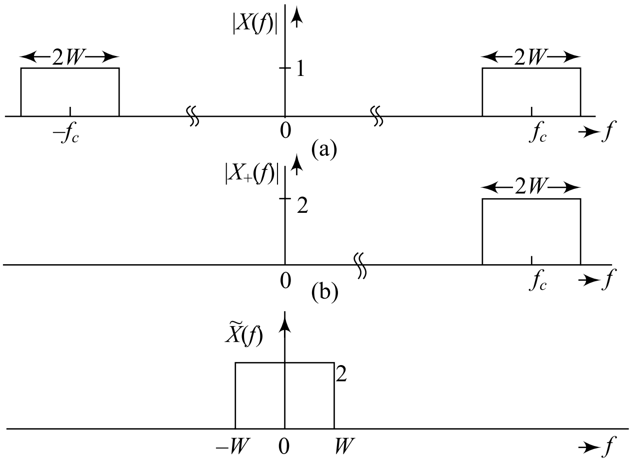


Fig. 3.23 *(a) Amplitude spectrum of the band pass signal x(t), (b) Amplitude spectrum of pre-envelope of x(t), (c) Amplitude spectrum of complex envelope of x(t)*

Consider a real-valued band pass signal with amplitude spectrum as shown in Fig. 3.23(a). The amplitude spectrum of the pre-envelope of $x(t)$ is shown in Fig. 3.23(b). If the pre-envelope of signal $x(t)$ is $x_+(t)$, then shifting its spectrum to the left along the frequency scale by an amount of f_c is equivalent to multiplying $x_+(t)$ by $e^{-j2\pi f_c t}$ (from the modulation theorem of FT). That is,

if $$\tilde{X}(f) = X_+(f + f_c) \tag{3.105}$$

then $$\tilde{x}(t) = x_+(t)e^{-j2\pi f_c t} \tag{3.106}$$

Hence, $$x_+(t) = \tilde{x}(t)e^{j2\pi f_c t} \tag{3.107}$$

Now, since $x_+(t) = x(t) + j\hat{x}(t)$, we have

$$x(t) = \text{Re}[x_+(t)] = \text{Re}[\tilde{x}(t)e^{j2\pi f_c t}] \tag{3.108}$$

Because of Eq. (3.108), $\tilde{x}(t)$ *is called the complex-envelope of the band pass signal* $x(t)$. Note that while $x(t)$ is band pass signal, its complex-envelope $\tilde{x}(t)$ is a complex-valued low pass signal. The reason for calling this low pass signal, $\tilde{x}(t)$ as the complex envelope of real-valued band pass signal, $x(t)$ is as follows. Suppose

$$x(t) = a(t)\cos[\omega_c t + \theta(t)], \tag{3.109}$$

where $a(t)$ and $\theta(t)$ are real-valued low pass signals. Then, we may write

$$x(t) = a(t)\cos[\omega_c t + \theta(t)] = \text{Re}[\{a(t)e^{j\theta(t)}\}e^{j\omega_c t}] \tag{3.110}$$

In Eq. (3.110), $\{a(t)e^{j\theta(t)}\}$ is obviously the complex envelope with $e^{j\omega_c t}$ being the complex carrier. A comparison of Eqs. (3.108) and (3.110) reveals that

$$\tilde{x}(t) = a(t)e^{j\theta(t)} \tag{3.111}$$

The complex-envelope representation of a band pass signal is a very convenient tool that is widely used in the representation of radar and sonar signals as well as in the analysis of band pass systems.

3.4.4 In-phase and Quadrature Component Representation

Using complex envelope, we shall now derive the 'in-phase and quadrature component' representation of a real-valued band pass signal $x(t)$ with center frequency f_c. Let $\tilde{x}(t)$ be the complex envelope of $x(t)$. Since $\tilde{x}(t)$ is complex-valued function, let

$$\tilde{x}(t) = x_I(t) + jx_Q(t) \tag{3.112}$$

Since $\tilde{x}(t)$ is a low pass signal of bandwidth, say, W, $x_I(t)$ and $x_Q(t)$ are also low pass signals of the same bandwidth W, but are real valued. From Eq. (3.108), we have

$$x(t) = \text{Re}[\{\tilde{x}(t)e^{j\omega_c t}] = \text{Re}[\{x_I(t) + jx_Q(t)\}\{\cos\omega_c t + j\sin\omega_c t\}]$$

$$\therefore \quad \boxed{x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t} \tag{3.113}$$

This representation of the band pass signal $x(t)$, is called the *canonical representation of x(t)*. The low pass real-valued signal, $x_I(t)$ is called the 'in-phase' component of the band pass signal $x(t)$, while the real-valued low pass signal, $x_Q(t)$, is called the 'quadrature' component of the band pass signal, $x(t)$. This is because, while $x_I(t)$ multiplies $\cos\omega_c t$, $x_Q(t)$ multiplies $\sin\omega_c t$ which is in phase quadrature with the carrier signal $\cos\omega_c t$.

In the foregoing discussion, we have used three different representations of the real-valued band pass signal, $x(t)$, with center frequency f_c. These different representations are:

$$x(t) = a(t)\cos[\omega_c t + \theta(t)] \tag{3.114}$$

$$x(t) = \text{Re}[\tilde{x}(t)e^{j\omega_c t}] = \text{Re}[x_+(t)] \tag{3.115}$$

and

$$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \tag{3.116}$$

The entities used in these three representations are obviously related. By expanding RHS of Eq. (3.114) and comparing with RHS of Eq. (3.116), we get

$$\left.\begin{aligned} x_I(t) &= a(t)\cos\theta(t) \\ x_Q(t) &= a(t)\sin\theta(t) \end{aligned}\right\} \tag{3.117}$$

By writing $\cos[\omega_c t + \theta(t)]$ of eqn. (3.114) as $\text{Re}[e^{j\{\omega_c t + \theta(t)\}}]$ and comparing with RHS of Eq. (3.115), we get

$$\tilde{x}(t) = \text{complex envelope} = a(t)e^{j\theta(t)} \tag{3.118}$$

so that

$$a(t) = |\tilde{x}(t)| \tag{3.119}$$

Further, from Eq. (3.107), we have

$$x_+(t) = \text{pre-envelope of } x(t) = \tilde{x}(t)e^{j2\pi f_c t}$$

$$\therefore \quad |x_+(t)| = |\tilde{x}(t)| = a(t) \tag{3.120}$$

Also from Eq. (3.119) we have

$$a(t) = [x_I^2(t) + x_Q^2(t)]^{1/2} \tag{3.121}$$

and

$$\theta(t) = \tan^{-1}\left[\frac{x_Q(t)}{x_I(t)}\right] \tag{3.122}$$

Example 3.17 If $x(t)$ is an energy signal, show that $x(t)$ and $\hat{x}(t)$ are orthogonal to each other over the interval $-\infty < t < \infty$.

Solution To show that $x(t)$ and $\hat{x}(t)$ are orthogonal over $-\infty < t < \infty$ we have to prove that

$$\int_{-\infty}^{\infty} x(t)\hat{x}(t)\,dt = 0$$

From the generalized Parseval's theorem of Fourier transform (see Eq. (3.14)), we have

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}\,dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}\,df$$

where the overbar indicates complex conjugation.

If $x(t)$ is real valued, $\hat{x}(t)$ is also real valued since it is after all obtained by convolving $x(t)$ with $1/(\pi t)$.

$$\therefore \quad \int_{-\infty}^{\infty} x(t)\hat{x}(t)dt = \int_{-\infty}^{\infty} x(t)\overline{\hat{x}(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{\hat{X}(f)}df$$

$$= \int_{-\infty}^{\infty} X(f)\,[j\,\mathrm{sgn}(f)\overline{X(f)}]\,df = j\int_{-\infty}^{\infty} \mathrm{sgn}(f)\,|X(f)|^2 df$$

However, since sgn(f) is an odd function of 'f', while $|X(f)|^2$ is an even function of 'f', the integrand in the last integral is odd and hence the integral is zero.

$$\therefore \quad \int_{-\infty}^{\infty} x(t)\hat{x}(t)dt = 0$$

Example 3.18 Find the Hilbert transform of the rectangular pulse $x(t) = A\Pi(t/\tau)$.

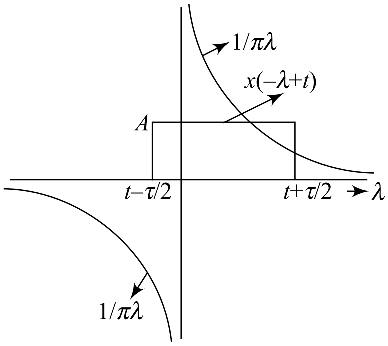


Fig. 3.24 *Convolution of x(t) with 1/(πt)*

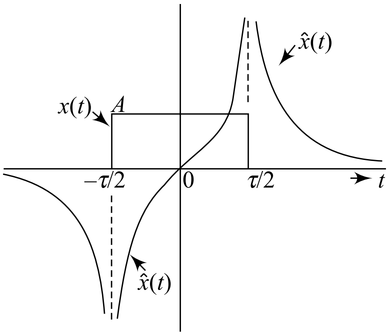


Fig. 3.25 *Signal x(t) and its Hilbert transform*

Solution

$$\hat{x}(t) = x(t) * \frac{1}{\pi t} = A\int_{-\infty}^{\infty} \frac{\Pi(t-\tau)}{\pi\lambda}d\lambda$$

$$= A\int_{t-\tau/2}^{t+\tau/2} \frac{1}{\pi\lambda}d\lambda$$

$$\therefore \quad \hat{x}(t) = \frac{A}{\pi}\left[\log_e\left|\frac{t+\tau/2}{t-\tau/2}\right|\right]$$

Figure 3.24 shows $x(t)$ and its Hilbert transform $\hat{x}(t)$. Note that $\hat{x}(t)$ goes to $-\infty$ and $+\infty$ at the two points of discontinuity of the signal $x(t)$.

Example 3.19 Given a signal $x(t) = A\Pi(t/T)\cos(\omega_c t + \theta)$, find (a) its analytic signal, (b) spectrum of its analytic signal, (c) complex envelope, and (d) the natural envelope, $a(t)$. Assume that $f_c T >> 1$.

Solution $\hat{x}(t) = A\Pi(t/T)\sin(\omega_c t + \theta)$ (From Eq. (3.101))

(a) Analytic signal or Pre-envelope of $x(t)$

$$= x_+(t) = x(t) + j\hat{x}(t) = A\Pi(t/T)[\cos(\omega_c t + \theta) + j\sin(\omega_c t + \theta)]$$

$$= A\Pi(t/T)e^{i(\omega_c t+\theta)}$$

(b) $$X_+(f) = \begin{cases} ATe^{j\theta}\text{sinc}(f - f_c)T; & f > 0 \\ 0; & f < 0 \end{cases}$$

(c) $\tilde{x}(t)$ = Complex envelope of $x(t) = x_+(t)e^{-j2\pi f_c t}$

$$= A\Pi(t/T)e^{j\theta}$$

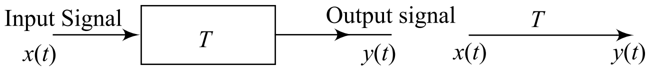
(d) Natural envelope of $x(t)$, i.e., $a(t) = |\tilde{x}(t)| = A\Pi(t/T)$.

3.5 SIGNAL TRANSMISSION THROUGH SYSTEMS

3.5.1 Review of LTI System Theory

We may define a system as an entity which acts on one or more inputs, or excitations, and produces one or more responses, or outputs. *We shall, however, confine our attention to single-input, single-output systems only*.

A system is generally represented diagrammatically as shown in Fig. 3.26(a) or (b).



(a) (b)

Fig. 3.26 *(a) and (b) Diagrammatic representation of a system*

Systems may be broadly classified into

1. Continuous-time systems
2. Discrete-time systems

Continuous-time systems take a continuous-time signal as input and produce another continuous-signal as output. Discrete-time systems, similarly, take a discrete-time signal as input, act upon it and produce another discrete-time signal as output. Each of these, in turn, may be further classified into the following types:

(a) Static (i.e., memory less) or dynamic (with memory)
(b) Linear or non-linear
(c) Time varying or time invariant

(a) Static and dynamic systems *A system is said to be static, memory less, or instantaneous, if its present output is determined entirely by the present input only.*

As an example, we may consider a continuous-time system with input-output relationship given by an algebraic equation such as

$$y(t) = Ax(t) + B$$

where A and B are constants. ***Among electrical systems, all purely resistive networks, however complicated they may be are 'static systems' only***.

Definition *A system is said to be dynamic, or a 'system with memory', if its present output depends for its value not only on the present input, but also on some past inputs.*

As an example consider a system represented by the differential equation:

$$RC\frac{dy(t)}{dt}+y(t)=x(t)$$

The value of $y(t)$ is dependent on the value of $x(t)$ not only at the instant t, but also on the *initial conditions*. It is the energy storage element C, the voltage across which cannot change instantaneously, that makes this circuit a dynamic system.

Remark *As a generalization, we may say that static systems have their input–output relation described by algebraic equations while dynamic systems have their input–output relation described through differential equations. Also, all purely resistive networks are static systems whereas those with energy storage elements like inductors and capacitors are dynamic systems.*

(b) Linear and non-linear systems

Definition *A continuous-time dynamic system is said to be 'at rest' or in the 'ground state', if all of its energy storage elements are devoid of any stored energy.*

Definition *Let T be a continuous-time system which is at rest. Let an input signal $x_1(t)$ given to T result in an output signal $y_1(t)$; and an input $x_2(t)$ result in an output of $y_2(t)$. Then the system T is said to be linear, if for any pair of arbitrary constants a_1 and a_2, an input of $a_1x_1(t)+a_2x_2(t)$ given to the system T results in an output of $a_1y_1(t)+a_2y_2(t)$.*

Any continuous-time system not satisfying the above condition is said to be a non-linear system.

Remark *A linear system should basically satisfy the properties of superposition and homogeneity. The above definition takes care of both these.*

(c) Time varying and time-invariant systems Time invariance is the property of a system which makes the behavior of the system independent of time.

Definition *Let y(t) be the response of a continuous time system T to an arbitrary input signal x(t). The system T is said to be time invariant or 'fixed' if, for any value of the real constant τ, it gives a response of $y(t-\tau)$ for an input of $x(t-\tau)$.*

If this condition is not satisfied, T is said to be a time-varying system.

Example 3.20 A certain continuous-time system, is described by the following input-output relation:

$$y(t) = x(2t)$$

Is this system

(a) static or dynamic?

(b) linear or non-linear?

(c) fixed or time varying?

Justify your answer.

Solution

(a) Since $y(t) = x(2t)$, the output, at any instant of time t_1 depends for its value on the present input for $t_1 = 0$, on future values of input for $t_1 > 0$ and on past values of input for $t_1 < 0$. Hence, the system is not static.

(b) $x(t) \xrightarrow{T} x(2t)$

$\therefore \quad x_1(t) \xrightarrow{T} x_1(2t) = y_1(t)$ and $x_2(t) \xrightarrow{T} x_2(2t) = y_2(t)$

Then $[a_1x_1(t) + a_2x_2(t)] \xrightarrow{T} [a_1x_1(2t) + a_2x_2(2t)]$
$= a_1y_1(t) + a_2y_2(t)$

$\therefore$ it is a linear system.

(c) $x(t) \xrightarrow{T} x(2t) = y(t)$

$\therefore \quad x(t-\tau) \xrightarrow{T} x(2t-\tau) \neq y(t-\tau)$ since
$y(t-\tau) = x(2t-2\tau)$

$\therefore$ the system is not time invariant.

Example 3.21 Show that an ideal differentiator with input $x(t)$ and output $y(t)$ related by $y(t) = \frac{dx(t)}{dt}$, is a linear time-invariant system.

Solution We are given that $x(t) \xrightarrow{T} y(t) = \frac{dx(t)}{dt}$

Hence, if $x_1(t) \xrightarrow{T} y_1(t)$ then $y_1(t) = \frac{dx_1(t)}{dt}$

and if $x_2(t) \xrightarrow{T} y_2(t)$, then $y_2(t) = \frac{dx_2(t)}{dt}$

Also if $[a_1x_1(t)+a_2x_2(t)]$ is given as the input,

$$[a_1x_1(t) + a_2x_2(t)] \xrightarrow{T} y(t) = \frac{d}{dt}[a_1x_1(t) + a_2x_2(t)] = a_1\frac{dx_1(t)}{dt} + a_2\frac{dx_2(t)}{dt}$$

$$\therefore \quad y(t) = a_1y_1(t) + a_2y_2(t)$$

Hence, the system T, i.e., the ideal differentiator, is a linear system.
To show that it is time invariant, consider

$$x_1(t) = x(t-\tau)$$

Then $x_1(t) \xrightarrow{T} y_1(t) = \frac{dx(t-\tau)}{dt}$

Put $t - \tau = \lambda$

$\therefore dt = d\lambda$ and $\frac{dx(t-\tau)}{dt} = \frac{dx(\lambda)}{d\lambda} = y(\lambda) = y(t-\tau)$

$\therefore$ the ideal differentiator is time invariant.

Causality A *system is said to be a 'causal system' or a 'non-anticipatory system' if its output at any instant of time depends for its value only on the input at that instant and the previous instants but not on the input at future instants.*

This means that a causal system is one which cannot anticipate what the future values of input would be and respond to those inputs now itself.

Thus, all physically realizable real-time systems must be causal.

Note *Henceforth we shall be discussing only about linear time-invariant systems, i.e., LTI systems. Hence, unless otherwise specified, whenever we use the term 'system' it should be understood that we are referring only to a linear time-invariant (LTI) system.*

Impulse response, h(t), of LTI systems

Definition The impulse response, $h(t)$, of an LTI system is defined as the response of the system to a unit impulse given to it as input, when the system is in ground state.

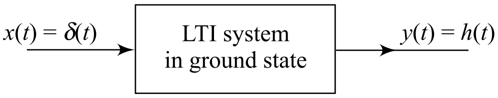


Fig. 3.27 *Impulse response of a system*

Impulse response characterization of LTI systems

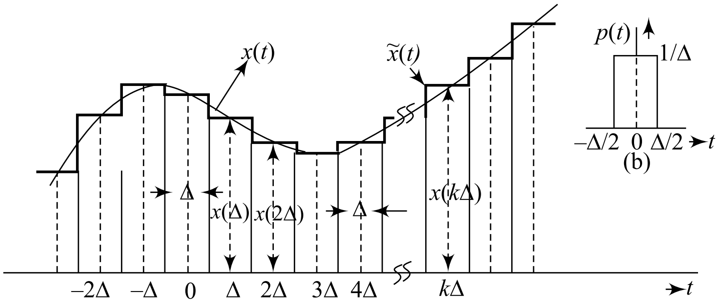


Fig 3.28 *(a) Signal x(t) and its approximation, (b) Rectangular pulse p(t)*

In Fig. 3.28(a), $x(t)$ is some arbitrary continuous-time signal and $\tilde{x}(t)$ is its approximation. It is clear that $\tilde{x}(t)$ approaches $x(t)$ as Δ tends to zero. Referring to the above two figures, we may write

$$\tilde{x}(t) = \sum_{k=-\infty}^{\infty} x(k\Delta)p(t - k\Delta)\Delta \tag{3.123}$$

$$x(t) = \operatorname*{Lim}_{\Delta\to 0} \tilde{x}(t) = \operatorname*{Lim}_{\Delta\to 0}\left[\sum_{k=-\infty}^{\infty} x(k\Delta)p(t - k\Delta)\Delta\right] \tag{3.124}$$

But

$$\operatorname*{Lt}_{\Delta\to 0} p(t) = \delta(t)\text{, a unit impulse located at } t = 0.$$

Further, $k\Delta$ becomes a continuous variable, say τ, as $\Delta \to 0$. Also, Δ itself may be represented by $d\tau$.

$$\therefore \qquad x(t) = \operatorname*{Lim}_{\Delta\to 0} \tilde{x}(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \tag{3.125}$$

Let us now give this signal $x(t)$ as input to an LTI system T in ground state and with impulse response $h(t)$. Then, we know that

$$x(t) \xrightarrow{T} y(t) = T[x(t)]$$

$$\therefore \qquad y(t) = T\left[\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau\right] = \int_{-\infty}^{\infty} x(\tau)T[\delta(t-\tau)]d\tau$$

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

$$\boxed{y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau} \tag{3.126}$$

The integrals in Eq. (3.126) are called *convolution integrals*, or the *superposition integrals*.

From Eq. (3.126), we find that the knowledge of $h(t)$, the impulse response of the system T, would enable us to calculate the output, $y(t)$, of the system for any given input signal, $x(t)$. *Hence, we can say that an LTI system is completely characterized by its impulse response function h(t).*

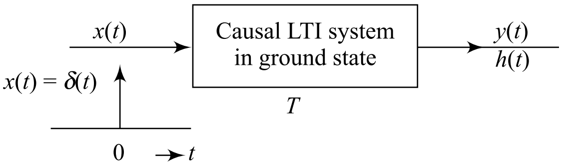


Fig. 3.29 *Impulse response of a causal system*

Causality and Impulse Response Let T be an LTI system which is in ground state. Let a unit impulse function, $\delta(t)$ be applied to T as input at $t = 0$. $\therefore$ For $t < 0$, $x(t) = 0$ and because the system is in ground state, the output $y(t)$, which we know, is $h(t)$, must be zero for all $t < 0$, since the system, being causal, cannot produce an output in anticipation of an input which is going to be applied at $t = 0$. At $t = 0$, the unit impulse is applied and therefore for $t \geq 0$, the output, $h(t)$, need not be zero.

$$\therefore \quad \boxed{\text{For a causal LTI system, } h(t) = 0 \text{ for } t < 0} \tag{3.127}$$

In the light of Eq. (3.127), the convolution integrals, for a *causal* LTI system, can be written as

$$\boxed{y(t) = \int_{0}^{\infty} x(t-\tau)h(\tau)d\tau = \int_{-\infty}^{t} x(\tau)h(t-\tau)\,d\tau} \tag{3.128}$$

Step response of LTI systems

Definition *The step response, g(t), of an LTI system T, is defined as the response of T to a unit-step function applied as input to T at t = 0, with the system T in ground state.*

Since $\delta(t)$ and $u(t)$ are related as

$$u(t) = \int_{-\infty}^{t} \delta(\lambda)\,d\lambda$$

It follows that for a LTI system, the step response and impulse response are related through the following equation:

$$g(t) = \int_{-\infty}^{t} h(\lambda)d\lambda \quad \therefore \quad h(t) = \frac{dg(t)}{dt} \tag{3.129}$$

Example 3.22 An RC low pass filter is shown in Fig. 3.30. Find its impulse response and step response.

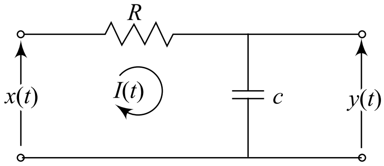


Fig. 3.30 *RC low pass filter*

Solution Since $i(t) = C\dfrac{dy(t)}{dt}$, we may write the mesh equation as

$$RC\frac{dy(t)}{dt} + y(t) = x(t)$$

Let us first find the impulse response and then make use of Eq. (3.8) to find $g(t)$.

To find the impulse response $h(t)$, put $x(t) = \delta(t)$ in the above differential equation and assume the system to be in ground state(see definition of $h(t)$). Then taking the Laplace transform on both sides of the differential equation, we get

$$RC[sY(s) - y(0^-)] + Y(s) = 1 \quad \text{for} \quad t > 0$$

$$\therefore \qquad Y(s) = \frac{1/RC}{s + 1/RC}$$

Now, taking the inverse Laplace transform of the above

$$y(t) = \begin{Bmatrix} \text{Response of the system to a unit impulse} \\ \textit{function} \text{ when the system in ground state} \end{Bmatrix} = h(t) = \frac{1}{RC}e^{-t/RC} \cdot u(t).$$

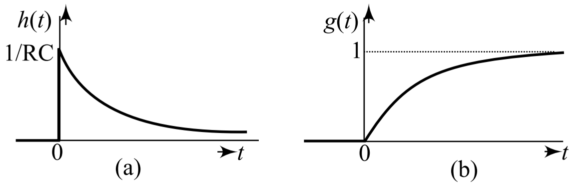
To find the step response, we note that

$$g(t) = \int_{-\infty}^{t} h(\lambda)\,d\lambda = \int_{-\infty}^{t} \frac{1}{RC} e^{-\lambda/RC} u(\lambda)\,d\lambda$$

$$\frac{1}{RC}\int_{0}^{t} e^{-\lambda/RC}\,d\lambda = [1 - e^{-t/RC}]\,u(t)$$

Alternatively, we may determine $g(t)$ by putting $x(t) = u(t)$ in the differential equation of the system, and solving it assuming the initial condition to be zero. Once $g(t)$ is obtained, we can differentiate it with respect to time to get $h(t)$.

The impulse response and the step response of the given low pass RC filter are plotted in Figs. 3.31(a) and (b), respectively.



(a) (b)

Fig. 3.31 *(a) Impulse response, (b) Step response of an RC low pass filter*

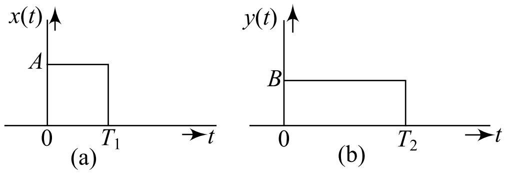
Example 3.23 A particular LTI system has $h(t) = e^{-2t}u(t)$. Determine its output signal $y(t)$ corresponding to an input signal $x(t) = u(t)$.

Solution $y(t) = x(t) * h(t) = \int_{0}^{\infty} e^{-2t} u(t) u(t-\tau)\,d\tau$

Since $u(\tau) = 0$ for $\tau < 0$ and $u(t-\tau) = 0$ for $\tau > t$

$$y(t) = \int_{0}^{t} e^{-2t}\,d\tau = -\frac{1}{2}e^{-2t}\Big|_{\tau=0}^{t} = \frac{1}{2}(1 - e^{-2t})$$

Example 3.24 If $x(t)$ and $y(t)$ are as shown in Figs. 3.32(a) and (b), determine graphically, the signal $z(t) = x(t) * y(t)$.



(a) (b)

Fig. 3.32

Solution

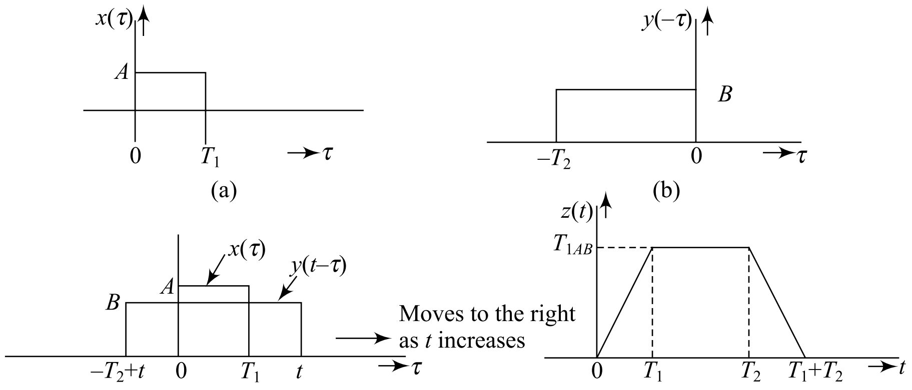


Fig. 3.33

$z(t) = \int_{-\infty}^{\infty} x(\tau) y(t-\tau)\,d\tau$ = Area under the product of $x(\tau)$ and $y(t - \tau)$ for ant t.

From Fig. 3.33(c), the following points are evident:

1. When $t \le 0$, the product of $x(\tau)$ and $y(t - \tau)$ is zero, as there is no overlap of the two.
2. As t increases beyond zero, the overlap and hence the area under the product increases linearly with t. This continues till $t = T_1$; and at this value of t, the area under the product i.e., $z(t)$, takes the maximum value equal to ABT_1.
3. As t increases beyond T_1, the overlap area and hence $z(t)$ will remain constant till $t = T_2$. When this value is reached, the left-side edge of $y(t - \tau)$ coincides with the y-axis, and any further increase in t beyond $t = T_2$ will make the overlap area to linearly decrease with time.
4. When t reaches the value $T_1 + T_2$, the left-side edge of $y(t - \tau)$ coincides with the right-side edge $x(\tau)$. Hence the overlap area and hence $z(t)$ becomes zero and remain at the at that value for all $t > (T_1 + T_2)$.
5. Signal $z(t)$ will have a trapezoidal shape in this case, the height of the trapezium being ABT_1 (since $T_1 < T_2$). The total base width of the trapezium = $T_1 + T_2$.
6. In case $T_1 = T_2 = T$, $z(t)$ will have a triangular waveform with height equal to ABT and base width equal to $2T$.

Example 3.25 The input $x(t)$ and the corresponding output $y(t)$ of a causal LTI system T are as shown in Figs. 3.34(a) and (b), respectively. Find the impulse response function $h(t)$ of the system.

Solution We know that for an LTI system, if $x(t) \xrightarrow{T} y(n)$, then $\dot{x}(t) \xrightarrow{T} \dot{y}(n)$

In this problem, $x(t) = 2u(t-3)$. Therefore, $\dot{x}(t) = 2\delta(t-3)$.

Since the system T is causal and since $y(t)$ is increasing linearly with time from $t = 3$ with a gradient of 1,

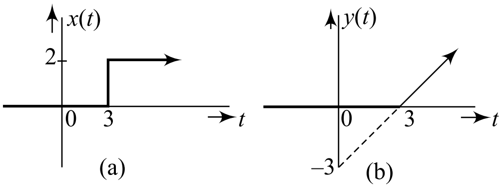
$$\dot{y}(t) = u(t-3)$$

$$\therefore \qquad 2\delta(t-3) \xrightarrow{T} u(t-3)$$

or

$$\delta(t) \xrightarrow{T} \frac{1}{2}u(t) \qquad (\text{ system is LTI})$$

$\therefore$ Impulse-response function $h(t) = \dfrac{1}{2}u(t)$.



(a) (b)

Fig. 3.34 *(a) Input x(t), (b) Output y(t)*

Stability One important way of defining stability of a system is in terms of '*bounded-input, bounded-output*' stability criterion or the BIBO criterion.

A signal $x(t)$ is said to be a bounded signal, bound to a value M, where M is a finite real positive number, provided the magnitude of $x(t)$ never exceeds M, i.e., provided $|x(t)| \le M$ for all 't', $-\infty < t < \infty$.

Criterion *The BIBO stability criterion says that a system T is stable if every bounded input given to it results in an output signal which is also bounded.*

Using the above criterion, we shall now derive the conditions required to be satisfied by an LTI system with impulse response $h(t)$, if it is to be stable in the BIBO sense.

Theorem *An LTI system T with impulse response h(t), is stable in the BIBO sense iff h(t), is absolutely integrable.*

i.e.,

$$\text{iff} \int_{-\infty}^{\infty} |h(t)|dt < \infty$$

Proof

1. *The forward implication* which states that system T is stable if its impulse response function is absolutely integrable.

Let $x(t)$ be any *arbitrary* bounded signal, bound to M, a positive finite real number. Let $x(t)$ be given as input to T. Then we know that $y(t)$, the output is given by

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

$$\therefore \qquad |y(t)| = \left| \int_{-\infty}^{\infty} x(t-\tau)h(\tau)\,d\tau \right| \le \int_{-\infty}^{\infty} |x(t-\tau)h(\tau)|\,d\tau$$

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

Since the maximum possible value of $|x(t-\tau)|$ for any τ, is M, we may write

$$\int_{-\infty}^{\infty} |x(t-\tau)||h(\tau)|\,d\tau \le M \int_{-\infty}^{\infty} |h(\tau)|\,d\tau$$

$$\therefore \qquad |y(t)| \le M \int_{-\infty}^{\infty} |h(\tau)|\,d\tau$$

Since M is finite and the integral of the absolute value of $h(t)$ is also given to be finite, it follows that $|\,y(t)\,|$ is always less than or equal to some finite positive real number. Hence, a bounded signal is obtained as the output for any arbitrary bounded input signal. Hence, T is stable in the BIBO sense.

2. The reverse implication states that an LTI system T with impulse response $h(t)$ cannot be stable in the BIBO sense if $h(t)$ is not absolutely integrable.

To prove this, we choose a particular $x(t)$ which is known to be a bounded signal, give it as input to T and show that if $h(t)$ is not absolutely integrable, then the resulting output signal $y(t)$ cannot be a bounded signal, i.e., that T cannot be a stable system in the BIBO sense.

Consider

$$x(\tau) = \begin{cases} 1 & \text{if } h(-\tau) > 0 \\ -1 & \text{if } h(-\tau) < 0 \\ 0 & \text{if } h(-\tau) = 0 \end{cases} \tag{3.130}$$

Since $|x(t)|$ is either 1 or zero, $x(t)$ is obviously a bounded signal, bound to a value 1. When $x(t)$ is given as input to T, the output is given by

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

$$\therefore \qquad y(0) = \int_{-\infty}^{\infty} x(\tau)h(-\tau)\,d\tau = \int_{-\infty}^{\infty} |h(-\tau)|\,d\tau \qquad \text{(From Eq. (3.130))}$$

$$y(0) = \int_{-\infty}^{\infty} |h(-\tau)|\,d\tau = \int_{-\infty}^{\infty} |h(\tau)|\,d\tau$$

But

$$\int_{-\infty}^{\infty} |h(\tau)|\,d\tau \text{ is not finite (given)}$$

$y(0)$ is, thus, not finite. $\therefore$ $y(t)$ is not a bounded signal even though $x(t)$ is. Hence, T is not stable.

An LTI system with impulse response $h(t)$ is stable in the BIBO sense iff $\int_{-\infty}^{\infty} |h(t)|\,dt < \infty$

Example 3.26 Examine the stability of the system shown in Fig. 3.35.

Solution By writing down Kirchhoff's mesh equations for the two loops, and eliminating $i_1(t)$ and $i_2(t)$, we get the differential equation as

$$\mathrm{LC}\frac{d^2y(t)}{dt^2}+\frac{L}{R}\frac{dy(t)}{dt}+y(t)=x(t)$$

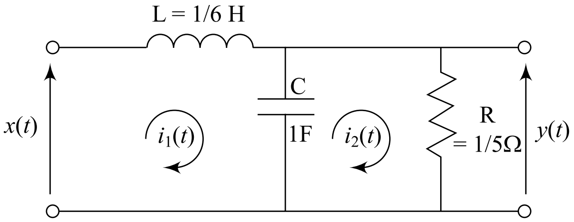


Fig. 3.35

In order to examine the BIBO stability, we have to first find $h(t)$, of the system. Let us replace $x(t)$ by $\delta(t)$ in the above differential equation, take the Laplace transform on both sides, and assume zero initial conditions.

$$LC[s^2Y(s)-\cancel{sy}(0^-)-\cancel{\dot{y}}(0^-)]+\frac{Ls}{R}Y(s)-\cancel{\frac{L}{R}}y(0^-)+Y(s)=1$$

i.e.,
$$Y(s)\left[LCs^2+\frac{L}{R}s+1\right]=1$$

$\therefore$
$$Y(s)=\frac{1}{LCs^2+\frac{L}{R}s+1}=\frac{1}{\frac{1}{6}s^2+\frac{5}{6}s+1}=\frac{6}{s^2+5s+6}$$

$$=\frac{-6}{s+3}+\frac{6}{s+2}$$

Now, taking the inverse Laplace transform on both sides,

$$y(t)=h(t)=6[e^{-2t}-e^{-3t}]u(t)$$

$\therefore$
$$\int_{-\infty}^{\infty}|h(t)|dt=6\int_{0}^{\infty}|e^{-2t}-e^{-3t}|dt=6\int_{0}^{\infty}(e^{-2t}-e^{-3t})dt=1$$

$\therefore$ $\int_{-\infty}^{\infty}|h(t)|dt$ is finite and the given system is stable in the BIBO sense.

Eigensignals of a system Suppose we give a sinusoidal signal of some frequency as input to a linear amplifier. The output signal is also a sinusoidal signal of the same frequency but perhaps with an amplitude and phase different from those of the input signal. But suppose we now give a rectangular waveform, or any non-sinusoidal waveform as the input signal and observe the output waveform. We find that the output waveform is not exactly similar *in shape* to the input waveform – the leading and trailing edges will not be vertical and there will be a droop in the tops of the pulses. Why was the output waveform having exactly the same shape as the input waveform when the input was a sinusoidal signal, and not when the input was a rectangular waveform? The answer is that a sinusoidal signal of any frequency is an 'eigensignal' of the linear amplifier, while the rectangular waveform signal is not.

Definition *An eigensignal of a system T is a signal, which when given as input to the system T, gives rise to an output signal which is essentially the same as the input signal except for a change in the amplitude and possibly a shift in time.*

Complex exponentials as Eigensignals of LTI systems Consider a stable LTI system T, with an impulse response $h(t)$. Since the system is stable, its $h(t)$ is absolutely integrable and, therefore, has a Fourier transform. Now, assume that a complex exponential signal

$$x(t) = e^{j\omega_0 t},$$

where ω_0 may be any real number, is given as input to the LTI system T. Let the corresponding output signal be $y(t)$.

Then, $$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau = \int_{-\infty}^{\infty} e^{j\omega_0(t-\tau)}h(\tau)d\tau$$

$\therefore$ $$y(t) = e^{j\omega_0 t}\int_{-\infty}^{\infty} h(\tau)e^{-j\omega_0\tau}d\tau = e^{j\omega_0 t}H(f)\Big|_{f=f_o}, \tag{3.131}$$

where $H(f) = F[h(t)]$ and is called the '*transfer function*', or, '*frequency response function*' of the system T. $H(f_0)$ which is the value of the complex-valued frequency function $H(f)$ at the frequency f_0, the input signal frequency, is in general, a complex number.

Thus, from Eq. (3.131), we find that when the complex exponential of some arbitrary frequency f_0 is given as input to an LTI system with some $h(t)$, which of course is absolutely integrable so as to make T stable, but otherwise arbitrary, the output is equal to a complex number $H(f_0)$ times the input signal.

Hence a complex exponential of any frequency is an eigensignal of any LTI system.

Once $e^{j\omega_0 t}$ is known to be an eigensignal, it is an easy matter to prove that all sinusoids whatever may be their frequency, are eigensignals of *all* LTI systems.

Transfer function of an LTI system If the input and output signal waveforms are of the same shape, but their amplitudes are different, it makes sense to take the ratio of output to input and this ratio, which yields a complex number in general, may be called as the complex gain of the system. From Eq. (3.131), we may write

$$\left.\frac{y(t)}{x(t)}\right|_{x(t)=e^{j\omega_0 t}} = H(f_0) \tag{3.132}$$

$H(f_0)$ thus represents the complex gain of the system at the frequency f_0. Thus $H(f)$ is the complex gain as a function of frequency and is therefore called the '*frequency response function*' or '*transfer function*' of the system. Since $H(f)$ is in general, complex, we may write

$$H(f) = |H(f)|\,e^{j\theta(f)} \tag{3.133}$$

In Eq. (3.133), $|H(f)|$ represents the magnitude of the gain of the system as a function of frequency, while $\theta(f)$ represents the phase shift (introduced by the system) as a function of frequency.

Hence, a plot of $|H(f)|$ vs. f is called the magnitude response of the system, and a plot of $\theta(f)$ vs. f is called the phase response of the system.

Example 3.27 For the LTI system described by the differential equation

$$\frac{dy(t)}{dt} + 6y(t) = x(t)$$

Determine the impulse response function and plot the magnitude and phase responses

Solution Taking the Laplace transform on both sides with $x(t)$ equal to $\delta(t)$ and all initial conditions as zero, we get

$$sY(s) + 6Y(s) = X(s) = 1$$

$$Y(s)[s+6]=1 \qquad \therefore\ Y(s)=\frac{1}{s+6}$$

$$h(t)=y(t)=e^{-6t}u(t) \qquad \therefore\ H(f)=\frac{1}{6+j2\pi f}$$

$$|H(f)|=\frac{1}{\sqrt{36+4\pi^2 f^2}} \text{ and } \theta(f)=-\tan^{-1}\left(\frac{2\pi f}{6}\right)=-\tan^{-1}\left(\frac{\pi}{3}f\right)$$

The magnitude and phase responses are as shown in Figs. 3.36(a) and (b), respectively.

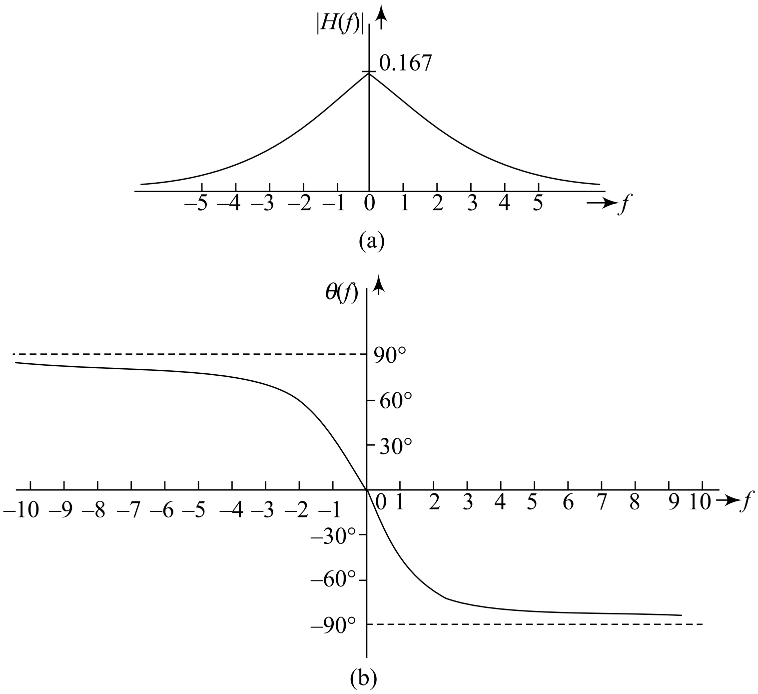


Fig. 3.36 *(a) Magnitude response, (b) Phase response*

3.5.2 Signal Transmission through LTI Systems

In this section, we shall discuss two specific aspects of transmission of a signal through an LTI system, i.e., undistorted transmission of a signal through an LTI system and filtering action of LTI systems: By undistorted transmission, we mean that the signal, during its passage through the system, does not suffer any distortion, except possibly a change in its amplitude and a time delay. By filtering we mean changing of the spectrum of the input signal in some desired manner by passing the signal through an LTI system.

1. Distortionless transmission through an LTI system: From our discussion on eigensignals in the previous section, it should not be concluded that only an eigensignal can pass through an LTI system without distortion. While an *eigensignal* can pass through *any* LTI system without distortion, *any signal* can pass through an LTI system without distortion *provided the system satisfies certain conditions*. We will now see what those conditions are.

As stated earlier, in distortionless transmission, the shape of output signal waveform is exactly the same as that of the input signal except possibly for a change in its amplitude and some time delay. Hence, for such systems

$$y(t) = Ax(t - \tau) \tag{3.134}$$

In the above equation, A represents the amplification (or attenuation) and τ represents the time delay. Taking the Fourier transform on both sides of Eq. (3.134), we get

$$Y(f) = AX(f)e^{-j2\pi f\tau} \tag{3.135}$$

But we know that $y(t) = x(t) * h(t)$

and hence $$Y(f) = X(f) \cdot H(f)$$

or $$\frac{Y(f)}{X(f)} = H(f) = \text{Transfer function of the system}$$

∴ From Eq. (3.135), we have

$$\frac{Y(f)}{X(f)} = H(f) = Ae^{-j2\pi f\tau} \tag{3.136}$$

From Eq. (3.136), it follows that for a distortionless transmission system

(a) The amplification/attenuation, given by $|H(f)|$, is a constant, independent of frequency.

(b) The phase-shift, or phase-delay, given by $\theta(f) = \angle H(f) = -2\pi f\tau$, is proportional to frequency.

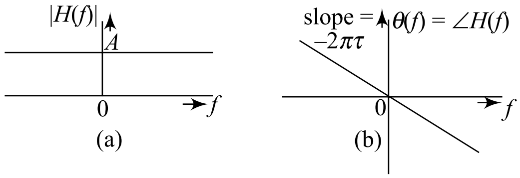


Fig. 3.37 *(a) Magnitude response, (b) Phase response of a distortionless transmission system*

From the above, it follows that the magnitude response and phase response of a distortionless transmission system (LTI) can be depicted as shown in Figs. 3.36(a) and (b) respectively.

However, no physical system can have a constant gain and a linear phase response for all frequencies. A physical system may, however, fulfill the above two requirements, at least approximately, over some range of frequencies – the gain may fall and the phase response may not be linear, outside this range of frequencies.

No practical signal can extend in time from minus infinity to plus infinity. All practical signals must have only finite duration. This implies that their spectrum must extend from minus infinity of frequency to plus infinity of frequency. Although a signal may have its frequency components extending from –∞ to +∞, fortunately, the amplitude of these frequency components become insignificantly small beyond some frequency. In other words, most of the energy of the signal is contained in some finite bandwidth.

From the foregoing discussion, we realize that gain of a system falling outside some range of frequencies, and spectrum of a signal too becoming insignificant beyond some frequency, underscore the need for defining terms like '*system bandwidth*' and '*signal bandwidth*', and then re-interpret the conditions for distortionless transmission of a signal through an LTI system in terms of these two.

Signal bandwidth Even if the spectrum of a signal extends theoretically up to infinity, we define its bandwidth as the width of only that part of its spectrum which contains *some specified* percentage (say 95%) of the total energy of the signal. Note that even though we generally draw a two-sided spectrum (in which the frequency refers to the frequency of a complex exponential and not of a co-sinusoid), the bandwidth is always specified in terms of positive frequency only, i.e., frequency of co-sinusoids. These concepts are illustrated in Fig. 3.38.

Thus, for the signal whose magnitude spectrum is shown in Fig. 3.38, f_0 is called the signal bandwidth, if

$$\frac{\int_{-f_0}^{f_0} |X(f)|^2 df}{\int_{-\infty}^{\infty} |X(f)|^2 df} = 0.95$$

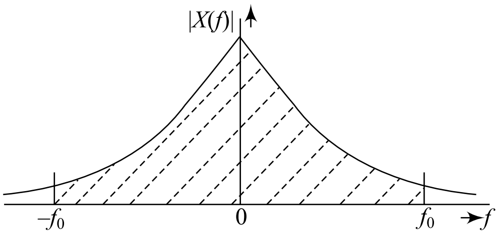


Fig. 3.38 *Two-sided spectrum of an arbitrary signal, the bandwidth of which is specified as f_0 Hz*

System bandwidth As stated earlier, for any physical system, the magnitude response characteristic cannot be absolutely flat for all frequencies because of the ever present parasitic capacitances across the output terminals, which tend to reduce the gain of the system at very high frequencies. Figure 3.39 shows the frequency response of a system. Theoretically, this response extends up to infinite frequency, as the response is going down to zero only asymptotically. Note that the gain takes a maximum value and is fairly constant over a certain frequency range, and falls off on either side. One way of defining the system bandwidth in such a case is to identify the frequency range over which the frequency response does not fall below 0.707 of the maximum value and call it as the system bandwidth. In Fig. 3.39, it is the frequency range from f_l to f_h. This bandwidth $(f_h - f_l)$, is called the half-power bandwidth, or the 3-db bandwidth.

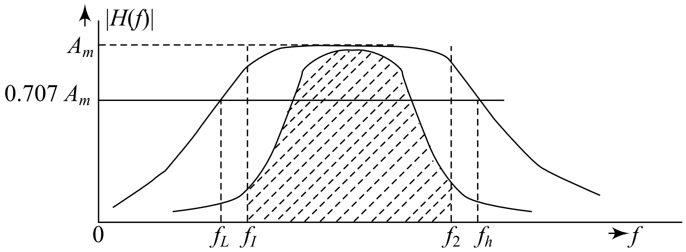


Fig. 3.39 *Frequency response characteristic of a system and the spectrum of an input signal*

In Fig. 3.39, the signal bandwidth f_1 to f_2 lies within the system bandwidth, f_e to f_h. Thus, all the significant frequency components of the signal experience almost the same gain, as the frequency response characteristic of the system is fairly flat from f_1 to f_2. Hence, insofar as this signal is concerned, there will be almost distortionless transmission of it through this system, provided the phase response of the system is linear over the range of frequencies of interest, i.e., f_1 to f_2. Insofar as the linear phase response requirement is concerned it can be shown that this imposes a constraint on the system that its $h(t)$ must be symmetrical about $t = \tau$, the time delay introduced by the system and that $h(t)$ must be maximum at $t = \tau$.

Example 3.28 An LTI system is a distortionless transmission system with gain A which is independent of frequency and with a constant time delay τ. Show that its $h(t)$ must be symmetrical about $t = \tau$ and that it has the maximum value at that point.

Solution Since T is a distortionless transmission system, we know that its transfer function can be written down as

$$H(f) = |H(f)| e^{-j2\pi f\tau} = Ae^{-j2\pi f\tau}$$

Taking the inverse Fourier transform on both sides,

$$h(t) = \int_{-\infty}^{\infty} Ae^{-j2\pi f\tau} \cdot e^{j2\pi ft} df$$

$$= A\int_{-\infty}^{\infty} e^{-j2\pi f(\tau-t)} df = A\int_{-\infty}^{0} e^{-j2\pi f(\tau-t)} df + A\int_{0}^{\infty} e^{-j2\pi f(\tau-t)} df$$

$$= A\int_{0}^{\infty} e^{j2\pi f(\tau-t)} df + A\int_{0}^{\infty} e^{-j2\pi f(\tau-t)} df = 2A\int_{0}^{\infty} \cos 2\pi f(\tau - t)$$

If we put $t = \tau + t_1$ where, t_1 is an arbitrary real number

$$h(t)\big|_{t=\tau+t_1} = 2A\int_{0}^{\infty} \cos 2\pi f t_1 df$$

Similarly,

$$h(t)\big|_{t=\tau-t_1} = 2A\int_{0}^{\infty} \cos 2\pi f t_1 df$$

$\therefore$

$$h(t)\big|_{t=(\tau+t_1)} = h(t)\big|_{t=(\tau-t_1)}$$

Thus, $h(t)$ has even symmetry about $t = \tau$. Also, since

$$h(t) = 2A\int_{0}^{\infty} \cos 2\pi f(\tau - t)\, df$$

and since $\cos 2\pi f(\tau - t) = \cos 0 = 1$ for $t = \tau$, $h(t)$ takes the maximum value at $t = \tau$.

2. Filtering action of LTI systems: A filter is a system which is specifically designed to modify the spectrum of any input signal in some desired manner. A properly designed LTI system can work as a filter, as may be seen from the following.

Let T be an LTI system with impulse response, $h(t)$. Let $x(t)$ be given as input signal to T and let the corresponding output signal be $y(t)$. Then, we know that

$$y(t) = x(t) * h(t)$$

where * denotes linear convolution operation. Taking Fourier transform of the above on both sides,

$$Y(f) = X(f) \cdot H(f) \tag{3.137}$$

Thus, the spectrum $X(f)$ of the input signal is modified by the transfer function $H(f)$ of the system to give us the spectrum $Y(f)$, of the output signal $y(t)$. From Eq. (3.137) we may write

$$|Y(f)|\, e^{j\theta_y(f)} = |X(f)| e^{j\theta_x(f)} \cdot |H(f)|\, e^{j\theta_H(f)} \tag{3.138}$$

$\therefore$

$$|Y(f)| = |X(f)| \cdot |H(f)| \tag{3.139}$$

and

$$\theta_y(f) = \theta_x(f) + \theta_H(f) \tag{3.140}$$

Equations (3.139) and (3.140) show how the transfer function of the system modifies magnitude spectrum and the phase spectrum of the input signal respectively. It must, however, be noted that $|H(f)|$ and $\theta_H(f)$ of a stable, causal LTI system cannot be specified independently, as the real and imaginary parts of $H(f)$ of such a system are Hilbert transforms of each other.

Ideal filters Applications arise, quite often, wherein we will be interested in transmitting not the entire spectrum of a signal, but only certain frequency band/bands in it. We make use of filters for this purpose.

The bands of frequencies transmitted through a filter without any appreciable attenuation are called the *pass bands* and the bands of frequencies which are highly attenuated, are called *stop bands* of the filter. Depending on the type of filter, there may be one or more pass bands and stop bands.

A filter which transmits, without any attenuation, all frequencies of the input signal that are less than a certain frequency, called the cut-off frequency and rejects all frequencies above it, is called a *low pass filter*. A filter whose stop band is below a certain cut-off frequency and its pass band above that frequency, is called a *high pass filter*. Other types of filters which are of interest are the *band pass filter,* which passes a certain specified band of frequencies from say f_1 to f_2 and rejects all other frequencies, and the *band stop* or *band rejection filter* which eliminates all frequencies within a certain specified band and passes all other frequencies.

1. Ideal low pass filter: Consider an ideal low pass filter with a pass band gain A, pass band width B Hz and a linear phase response with a slope of $-2\pi\tau$. Then its transfer function is

$$H(f) = A\prod(f/2B)e^{-j2\pi f\tau} \tag{3.141}$$

Taking the inverse Fourier transform, we get its impulse response function as

$$h(t) = 2AB\text{sinc } 2B(t - \tau) \tag{3.142}$$

The magnitude response, phase response and impulse response functions of this ideal LPF are shown in Fig. 3.39.

2. Ideal high pass filter: Consider an ideal high pass filter with pass band gain A, cut-off frequency f_c and time delay t_0 sec. Its transfer function may be written as

$$H(f) = |H(f)|e^{j\theta(f)} \tag{3.143}$$

where $|H(f)| = A[1 - \prod(f/2f_c)]$ (3.144)

and $\theta(f) = -2\pi f t_0$ (3.145)

The magnitude response and phase response of this ideal high pass filter (HPF) are shown in Fig. 3.40.

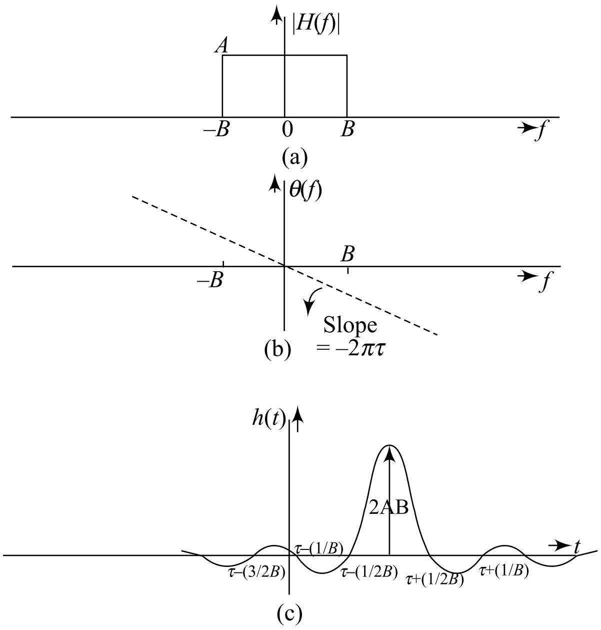


Fig. 3.40 *(a) Magnitude response of an ideal LPF, (b) Phase response of an ideal LPF, (c) Impulse response h(t) of an ideal LPF*

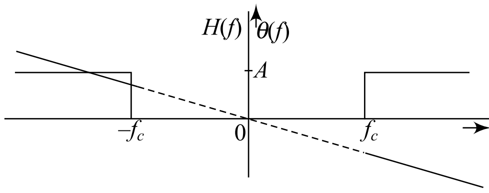


Fig. 3.41 *Magnitude and phase response of an ideal HPF*

Example 3.29 Find the impulse response $h_{HP}(t)$ of an ideal high pass filter with a pass band gain of A, cut-off frequency of f_c Hz and a linear phase response with a slope of $-2\pi t_0$.

Solution The transfer function $H(f)$ of an ideal HPF, we may write as

$$H(f) = |H(f)|e^{j\theta(f)}$$

where $|H(f)| = A[1 - \Pi(f/2f_c)]$

and $\theta(f) = -2\pi t_0 f$

$$\therefore \qquad h(t) = \mathcal{F}^{-1}[H(f)] = \mathcal{F}^{-1}\{[A - A\Pi(f/2f_c]e^{-j2\pi t_0 f}\}$$

$$= \mathcal{F}^{-1}[Ae^{-j2\pi t_0 f} - A\Pi(f/2f_c e^{-j2\pi t_0 f}]$$

$$= A\mathcal{F}^{-1}[e^{-j2\pi t_0 f}] - A\mathcal{F}^{-1}\left[\Pi(f/2f_c)\right] * \mathcal{F}^{-1}[e^{-j2\pi t_0 f}]$$

$$= A\delta(t - t_0) - A[(2f_c \text{sinc } 2f_c t)] * \delta(t - t_0)$$

$$\therefore \qquad h(t) = A\delta(t - t_0) - 2Af_c \text{sinc } 2f(t - t_0)$$

Since the sinc function extends in time from –∞ to +∞, the ideal HPF is also not a causal system and hence, is not physically realizable.

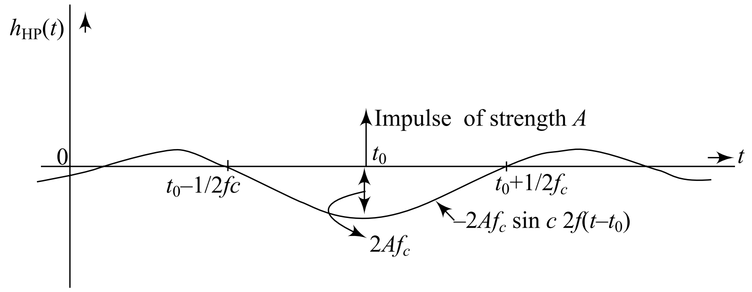


Fig. 3.42 *Impulse response of an ideal HPF of pass band gain A and cut-off frequency* f_c

3. Ideal band pass filter: Consider an ideal band pass filter (BPF) with pass band from f_1 to f_2, pass band gain A and a time delay t_0 sec.

Let $(f_2 - f_1)$ = B Hz and $f_1 f_2 = f_0^2$

Then $H(f)$ of the ideal BPF may be written down as

$$H(f) = |H(f)|e^{j\theta(f)}$$

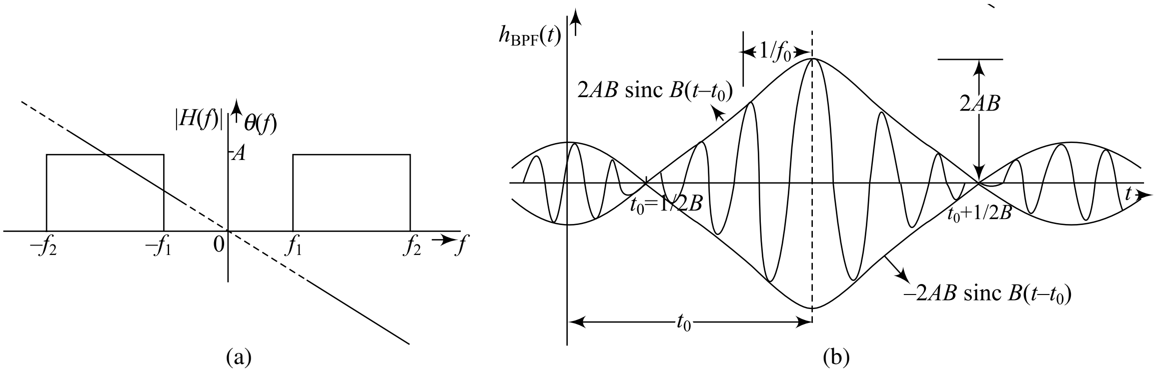
where $$|H(f)| = A\Pi[(f + f_0)/B] + A\Pi[(f - f_0)/B] \tag{3.146}$$

and $$\theta(f) = -2\pi f t_0 \tag{3.147}$$

Taking the inverse Fourier transform of $H(f)$, we get

$$h(t) = 2AB[\text{sinc } B(t - t_0)] \cos 2\pi\ f_0(t - t_0) \tag{3.148}$$

The magnitude and phase responses of this ideal BPF are shown in Fig. 3.43(a) while its impulse response function is shown in Fig. 3.43(b).



(a) (b)

Fig. 3.43 *(a) Magnitude and phase response of an ideal BPF, (b) Impulse response of an ideal BPF*

Example 3.30 Determine the impulse response function, $h(t)$ of an ideal BPF with pass band gain of A and pass band bandwidth of B Hz centered on f_0 Hz and having a linear phase response.

Solution We have $f_2 - f_1 = B$ Hz and $f_1 f_2 = f_0$.
We may write

$$H(f) = |H(f)|e^{j\theta(f)}$$

where $$|H(f)| = A\Pi[(f + f_0)/B] + A\Pi[(f - f_0)/B]$$

and $$\theta(f) = -2\pi f t_0$$

Here, $-2\pi t_0$ is the gradient of the linear phase response.

$$\begin{aligned} \therefore \quad h(t) &= \mathcal{F}^{-1}\{[A\Pi\{(f + f_0)/B\} + A\Pi\{(f - f_0)/B\}]e^{-j2\pi f t_0}\} \\ &= \mathcal{F}^{-1}[A\Pi\{(f + f_0)/B\}e^{-j2\pi f t_0} + A\Pi\{(f - f_0)/B\}e^{-j2\pi f t_0}] \\ &= A\{(B\operatorname{sinc} Bt)\}e^{-j2\pi f t_0} * \delta(t - t_0) + A\{(B\operatorname{sinc} Bt)\}e^{j2\pi f t_0} * \delta(t - t_0) \\ &= \{A[B\operatorname{sinc} B(t - t_0)]e^{-j2\pi\, f_0(t-t_0)} + [B\operatorname{sinc} B(t - t_0)]e^{j2\pi f_0(t-t_0)}\} \\ &= 2AB\operatorname{sinc} B(t - t_0)]\cos 2\pi\, f_0(t - t_0) \end{aligned}$$

Like the ideal LPF and ideal HPF, the ideal BPF too is non-causal and hence not physically realizable.

Note *It may be noted that the impulse response functions of all these ideal filters have sinc functions in them. Hence, their h(t)s extend from t = –∞ to t = +∞. Thus, h(t) ≠ 0, for t < 0 for all these filters and hence they are non-causal and cannot be physically realized.*

3.5.3 Paley–Wiener Criterion for Physical Realizability

Till now, we have been discussing the question of physical realizability of an LTI system only in terms of its impulse-response function, $h(t)$, being equal to zero for all negative values of time, i.e., in the time domain. But, we will generally be facing the problem of determining the physical realizability, or otherwise, of an LTI system, given its transfer function, as in the case of filters.

Paley–Wiener criterion can be used to test whether a system, with a given magnitude response, $|H(f)|$ is physically realizable or not. It states that a square integrable magnitude response function $|H(f)|$ is physically realizable if

$$\int_{-\infty}^{\infty} \frac{|\log_e |H(f)||}{1 + f^2} df < \infty \tag{3.149}$$

From this, it is clear that any magnitude response which is equal to zero continuously over a range of frequencies, cannot be realized physically since $|\log_e |H(f)||$ becomes infinitely large for such a case. It may be noted that every ideal filter, LPF or HPF or BPF or BSF, has its magnitude response staying at zero values over a certain continuous range of frequencies, i.e., over the entire stop bands. Hence, they are not physically realizable.

Further, Eq. (3.149) suggests that the magnitude response of a physically realizable system cannot rise or fall suddenly, as is the case with all the ideal filters. Suppose, for instance, that

$$|H(f)| = Ae^{-|f|}$$

The magnitude response is decreasing here at a rate corresponding to an exponential order. From Eq. (3.149), we find that this magnitude response does not violate the Paley–Wiener criterion and so is physically realizable. But suppose

$$|H(f)| = Ae^{-f^2}$$

The rate of change of the response, in this case, is more than the exponential order; and we find that when this magnitude response is substituted in Eq. (3.149), it violates the Paley–Wiener criterion and therefore it is not causal, i.e., it is not physically realizable.

Thus, this criterion enables us to determine directly, *without going into the time domain*, whether a given magnitude response function is physically realizable or not.

3.5.4 System Bandwidth and Rise Time

Whenever pulses with steep leading and trailing edges are transmitted through, let us say, a cable, or a pair of wires, we find that the pulses obtained at the receiving end will have leading and trailing edges with finite slopes. The steepness of say, the leading edge, is expressed in terms of what is called the *rise time*, which is the time taken by the pulse to rise from 10% of its final value to 90% of its final value. Thus, even though the input pulses may have zero rise time, the output pulses have a non-zero, finite rise time. This is due to the fact that while a pulse with very steep leading and trailing edges has considerable high frequency components, the cable or pair of wires used for transmission, has poor magnitude response at those high frequencies; that is, the poor bandwidth of the cable or transmission lines, is responsible for the non-zero rise time of the output pulses. We shall therefore examine the relationship between bandwidth and rise time. For this purpose, we shall model the leading edge of the input pulse by a unit-step function and the cable or transmission line by a low pass filter, say, a first-order RC low pass filter, or an ideal LPF.

Example 3.31 Find the relation between bandwidth and the rise time of a pulse in the case of the first-order RC low pass filter shown in Fig. 3.44.

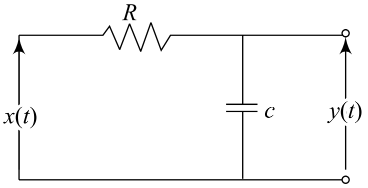


Fig. 3.44 *First-order RC low pass filter*

Solution The differential equation is

$$RC\frac{dy(t)}{dt} + y(t) = x(t)$$

Taking Fourier transform on both sides, we get

$$\frac{Y(f)}{X(f)} = H(f) = \frac{1}{1 + j2\pi f\text{RC}} \tag{3.150}$$

In Example 3.22, we have shown that the unit-step response, $g(t)$ of the system, is given by

$$g(t) = (1 - e^{-t/RC})\, u(t) \tag{3.151}$$

The 3-db frequency f_c of this filter is such that

$$|H(f)|_{f=f_c} = \frac{1}{\sqrt{2}}$$

This gives $$f_c = B = \frac{1}{2\pi RC} \quad \therefore \quad RC = \frac{1}{2\pi B}$$

Substituting this for RC in Eq. (3.151), we get

$$g(t) = (1 - e^{-2\pi Bt})u(t) \tag{3.152}$$

Referring to Fig. 3.44,

$$g(t)\big|_{t=t_1} = (1 - e^{-2\pi Bt_1}) = 0.9 \tag{3.153}$$

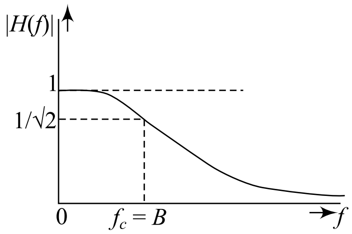


Fig. 3.45 *Magnitude response of a first-order RC low pass filter*

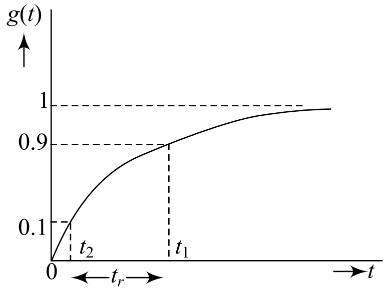


Fig. 3.46 *Unit-step response of a first-order RC low pass filter*

$$g(t)\big|_{t=t_2} = (1 - e^{-2\pi Bt_2}) = 0.1 \tag{3.154}$$

From Eqs. (3.153) and (3.154), we have

$$e^{-2\pi Bt_1} = 0.1 \quad \text{and} \quad e^{-2\pi Bt_2} = 0.9$$

$$\therefore \qquad e^{2\pi B(t_1 - t_2)} = 0.9$$

Taking logarithm to the base e on both sides,

$$(t_1 - t_2) = t_r = \text{Rise-time} = \frac{\log_e 9}{6.28B} = \frac{0.35}{B}$$

Thus,

$$\boxed{t_r = \frac{0.35}{B} \text{ for a first-order } RC \text{ low pass filter}} \tag{3.155}$$

Rise time with an ideal LPF We will now model the cable or transmission line as an ideal LPF with cut-off frequency B Hz. Let the magnitude response be as shown in Fig. 3.47.

Without loss of generality, we shall further assume that the time delay τ, of the ideal LPF is zero. Then, from Eq. (3.142), we have

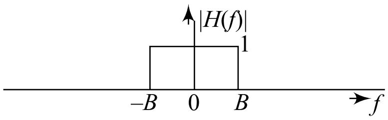


Fig. 3.47 *Magnitude response of an ideal LPF*

$$h(t) = 2B \operatorname{sinc} 2Bt \tag{3.156}$$

The step response $g(t)$ is given by

$$g(t) = \int_{-\infty}^{t} h(\lambda)\, d\lambda = \int_{-\infty}^{t} 2B \operatorname{sinc} 2B\lambda\, d\lambda \tag{3.157}$$

Put $\tau = 2B\lambda$ we get

$$g(t) = \int_{-\infty}^{0} \operatorname{sinc} \tau\, d\tau + \int_{0}^{2Bt} \operatorname{sinc} \tau\, d\tau \tag{3.158}$$

But

$$\int_{-\infty}^{0} \operatorname{sinc} \tau\, d\tau = \int_{0}^{\infty} \operatorname{sinc} \tau\, d\tau = \frac{1}{2} \tag{3.159}$$

The other integral in Eq. (3.158) has to be evaluated numerically or by referring to the table of 'sine integral function', $Si(\theta)$, where

$$Si(\theta) \underline{\Delta} \int_{0}^{\theta} \frac{\sin x}{x}\, dx \tag{3.160}$$

Since

$$\operatorname{sinc} \tau \underline{\Delta} \frac{\sin \pi\tau}{\pi\tau},$$

Putting $x = \pi\tau$, we get

$$Si(\theta) \underline{\Delta} \int_{0}^{\theta} \frac{\sin x}{x}\, dx = \pi \int_{0}^{\theta/\pi} \operatorname{sinc} \tau\, d\tau \tag{3.161}$$

A plot of the above sinc integral function, $S_i(\theta)$, is given in Fig. 3.48.

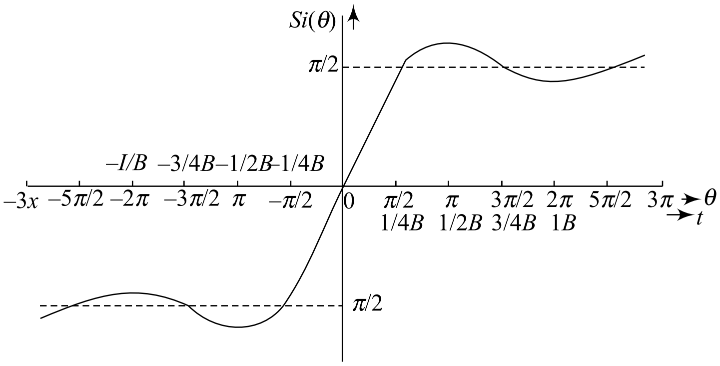


Fig. 3.48 *Sinc integral function*

From Eq. (3.160), we have

$$\frac{1}{\pi} Si(\theta) = \int_{0}^{\theta/\pi} \text{sinc}\, \tau \, d\tau \tag{3.162}$$

Now, reverting to Eq. (3.158), and recalling that our interest is in evaluating $g(t)$

$$g(t) = \frac{1}{2} + \int_{0}^{2Bt} \text{sinc}\, \tau \, d\tau = \frac{1}{2} + \frac{1}{\pi} Si(\theta), \tag{3.163}$$

$\therefore$ putting $\dfrac{\theta}{\pi} = 2Bt,$ we have

$$t = \frac{\theta}{2\pi B} \tag{3.164}$$

Using Eqs. (3.163) and (3.164), we plot $g(t)$ vs. t; and this is shown in Fig. 3.49.

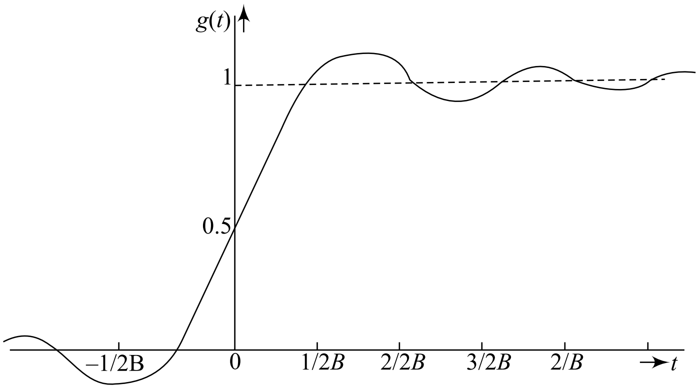


Fig. 3.49 *Response of an ideal LPF to a unit step*

Now, let us find the slope of $g(t)$ at $t = 0$. From Eq. (3.163)

$$\left.\frac{d}{dt} g(t)\right|_{t=0} = \frac{d}{dt}\left[\frac{1}{2} + \int_{0}^{2Bt} \text{sinc}\, \tau \, d\tau\right]\Bigg|_{t=0} = 2B \tag{3.165}$$

Approximating the portion of $g(t)$ between $g(t) = 0$ to $g(t) = 1$ to a straight line, we find the slope of this to be $2B$. Hence, the time taken to increase from $g(t) = 0$ to $g(t) =1$ is $\left(\frac{1}{2B}\right)$.

Thus, the time required by $g(t)$ to increase from a value of 0.1 to 0.9, which is the rise time t_r, is given by

$$t_r \cong \frac{0.8}{2B} \cong \frac{0.4}{B}$$

$$\therefore \qquad \boxed{t_r = \frac{0.4}{B} \quad \text{for an ideal LPF of bandwidth } B \text{ Hz}} \qquad (3.166)$$

MATLAB Example 3.1 In this example, we demonstrate how the Fourier transform of a signal can be calculated using MATLAB. We use as an example the sum of two sinusoids that are closely spaced in frequency—one at 1.2566 rad/s and the other at 2.5133 rad/s.

```
Ts = 0.02;              %A suitable time period that satisfies the aliasing
condition is used.
t = 0 : Ts : 100 ;      %A suitable time interval is defined
f1 = 0.2 ; f2 = 0.4 ;   %the two frequencies are defined
x = cos(2*pi*f1*t) + cos(2*pi*0.4*t) ; % The signal is
generated
% We now proceed to estimate the Fourier Transform using the FFT, first
% using small number of points and then using larger number of points
N = 256 ; % Number of FFT points
ws = 2*pi/(N*Ts) ;      %Frequency separation obtained
fp = 0 : N/2 ;
y = Ts*fft(x(1 : N)) ; %The fft is the algorithm for fast
computation of the DFT,
% which is used to find the transform of at a finite number
of points.
subplot (211), plot (fp*ws, abs(Y(fp + 1)), 'K') ;     %The transform
is plotted for N points
axis ([0, 15, 0, 3]) ;
xlabel ('Frequency, rad. per sec.') ; ylabel ('Magnitude')
title ('Magnitude of Transform') ;
% Now estimate the transform using larger number of points
N = 1024 ; % Number FFT points
ws = 2*pi/(N*Ts) ; % Frequency separation obtained
fp = 0 : N/2 ;
y = Ts*fft (x(1 : N)) ;
subplot (212), plot (fp*ws, abs(y(fp + 1)), 'k') ; % The transform
is plotted for N = 1024 points
axis ([0, 15, 0, 12]) ;
xlabel ('Frequency, rad. per sec.') ; ylabel ('Magnitude')
title ('Magniude of Transform') ;
```

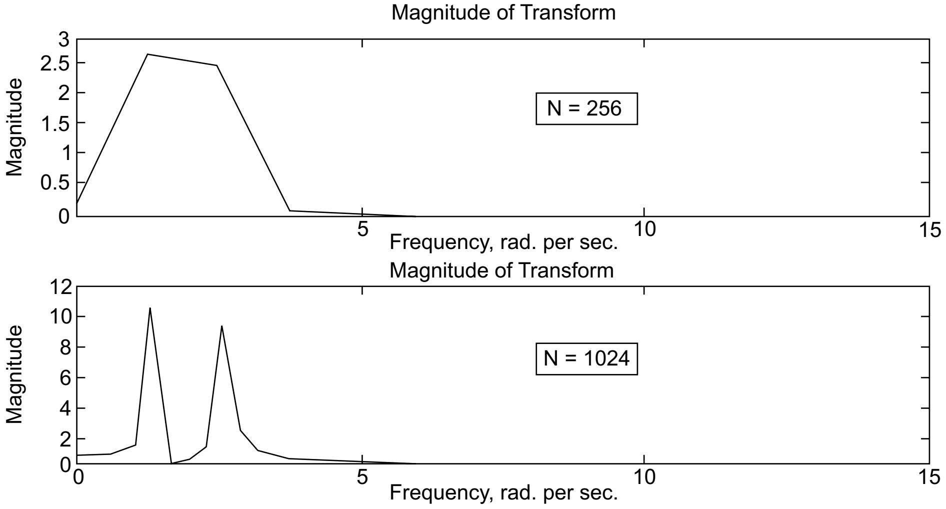


Fig. 3.50

The plot depicts magnitude of the transform obtained using $N = 256$ (top) and $N = 1024$ (bottom) points in the evaluation. As the number of points becomes infinite, impulses will be observed at the correct frequencies.

MATLAB Example 3.2 A periodic signal $x(t)$ with period $T_o = 6$ is defined by $x(t) = \Pi(t/2)$ for *abs t* $< = 3$. This signal is passed through an LTI system with an impulse response given by

$$h(t) = \{e^{-t/2} \quad \text{for} \quad 0 \leq t \leq 4$$
$$= 0 \text{ otherwise}$$

Determine the discrete spectrum of the output signal numerically using MATLAB.

Matlab Program

```
%
% Generation of x(t) signal
%
clc
df = 0.01;
fs = 10;
ts = 1/fs
t = [-8:ts:8];
%
% Generation of periodic signal
%
x = zeros(size(t));
x(11:30) = ones(size(x(11:30)));
x(71:90) = ones(size(x(71:90)));
x(132:151) = ones(size(x(132:151)));
subplot (2,2,1)
plot(t,x)
grid on
xlabel (‘time’);
ylabel (‘amplitude’);
title (‘Periodic Signal’)
ylim ([0,1.25]);
%
```

```
% Generation of impulse response
%
h = zeros(size(t));
h(82:120) = exp(-t(82:120)/2)
subplot (2,2,2)
plot(t,h)
grid on
xlabel(‘time’);
ylabel(‘Amplitude’);
title (‘Impluse Response’);
%
% Transfer function
%
H = fft(h)/fs;  % frequency resolution
f = [0:df:fs];
H1 = fftshift(H)    % rearrange H
subplot (2,2,3)
stem (t,abs(H1))
xlabel (‘Frequency’)
grid on
y = x.*H1
subplot (2,2,4)
stem (t,abs(y))
grid on
xlabel (‘Frequency’);
title (‘discrete spectrum of output signal’);
```

Results

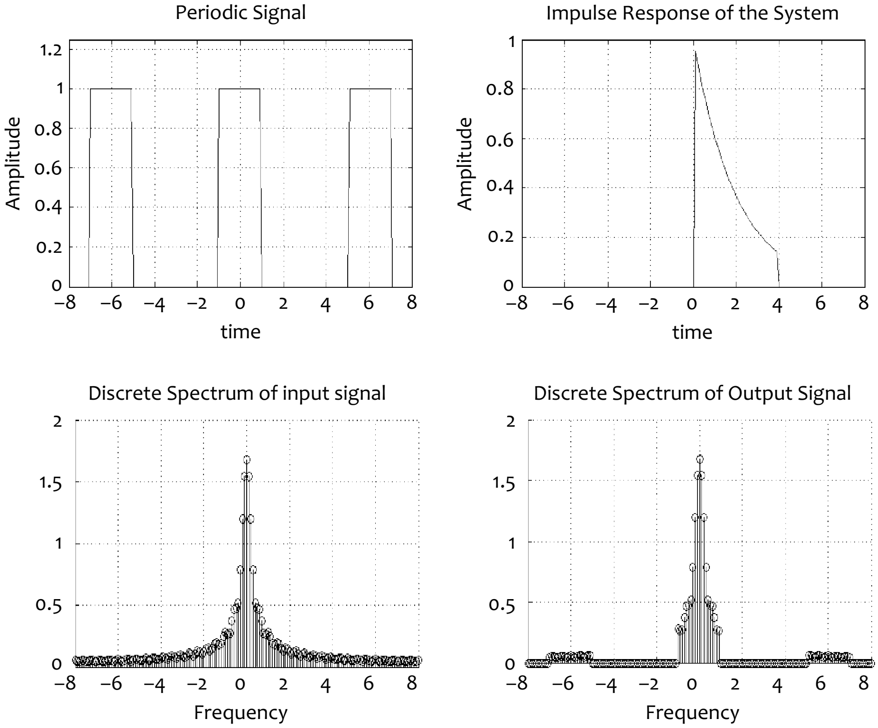


Fig. 3.51

MATLAB Examples 3.3 Plot the magnitude spectrum and phase spectrum of the non-periodic signal shown in Fig. 3.52.

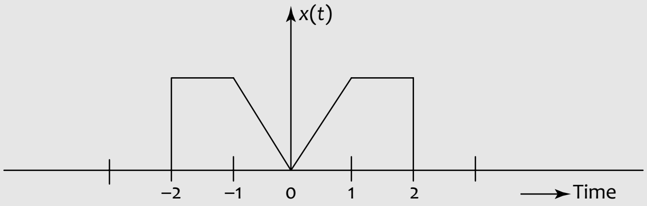


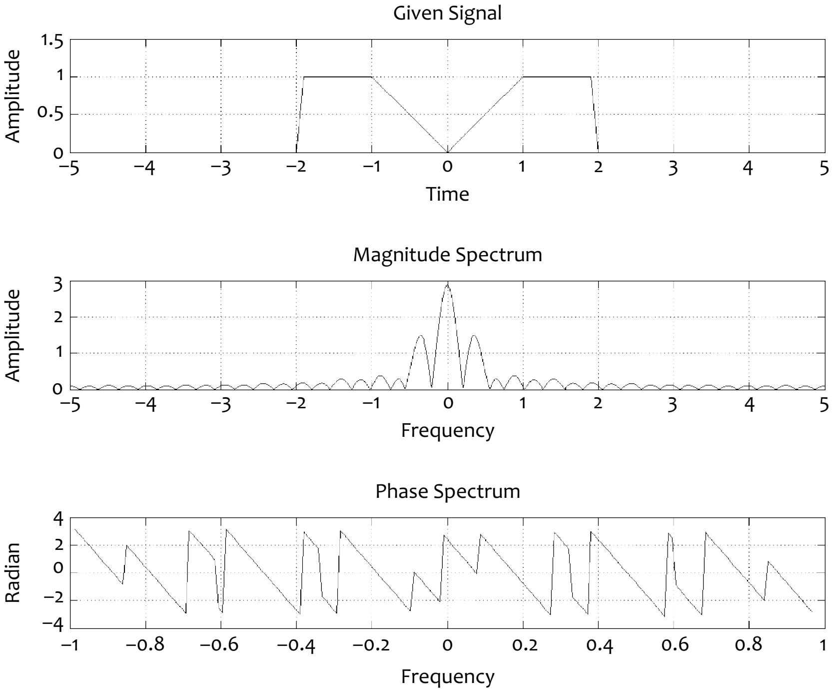
Fig. 3.52

MATLAB Program

```
clc
df = 0.01;
fs = 10;   % sampling frequency
ts = 1/fs  % sampling time
t = [-5:ts:5]   % time scale
%
% Generation of nonperiodic signal
%
x = zeros(size(t));
x(32:41) = ones(size(x(32:41)));
for i= 1:1:10
x(41+i)=1-0.1*i;
end
for i = 1:1:10
x(51+i) = 0.1*i;
end
x(61:70) = ones(size(x(61:70)));
subplot (3,1,1)
plot(t,x)
ylim([0 1.5]);
grid on
xlabel ('Time');
ylabel ('Amplitude');
title('Given Signal');
%
% Finding magnitude spectrum and phase spectrum of the nonperiodic signal
%
[X,x1,df1] = fftseq(x,ts,df);
X1 = X/fs;
f = [0:df1:df1*(length(x1)-1)]-fs/2;
subplot (3,1,2)
plot(f,fftshift(abs(X1)));
grid on
xlabel ('frequency');
ylabel ('amplitude');
title ('Magnitude Spectrum');
subplot (3,1,3)
plot(f(412:612),fftshift(angle(X1(412:612))))
```

```
grid on
xlabel (‘frequency’);
ylabel (‘radian’);
title (‘Phase Spectrum’);
```

Results


Given Signal

Magnitude Spectrum

Phase Spectrum

Fig. 3.53

MATLAB Example 3.4 As discussed already, complex sinusoids are eigenfunctions of an LTI filter, and real sinusoids are not in general eigenfunctions. In this example, we demonstrate that real sinusoids are eigenfunctions of linear time invariant filters when the impulse response of the LTI filter is real and even (conjugate symmetric if complex) when a windowed sinsusoid is the input, the output is plotted.

```
t=4*pi*[0 : 100] * 0.01; x=sin(t);
h = [0.5 1 0.5];
y = conv(x, h);
ys=y(2 : length (t) + 1);
%This effectively makes the impulse response h(-1) = 0.5, h(0) = 1 and %h(1) = 0.5
%since advancing the output effectively advances the impulse response %by 1.
%Note that this makes the impulse response real and even.
%plot(t, x, t, ys)
z = [x.*, ys. *];
%stem(ti, xi, 'o', ti, xil, 's', ti, xic, 'x');
h=stem(t(26 : 75), z(26 : 75, :), 'k');
set (h(1), 'MarkerFaceColor', 'black')
set (h(2), 'MarkerFaceColor', 'black', "Marker', 'square')
title ('Input and Output of the LTI system');
```

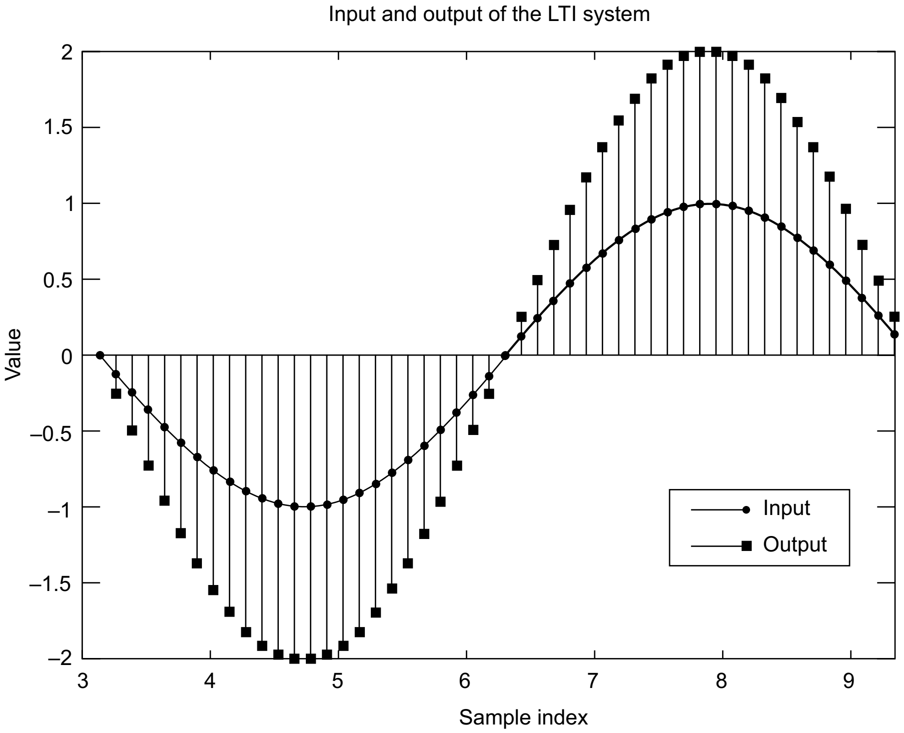


Fig. 3.54 *The input and the output of an LTI filter with sinusoidal input are compared for the special case when the impulse response is real and even*

```
ylabel ('value') ; xlabel ('sample index') ;
legend ('Input ' , 'output')
%It is clear that the output is a scaled version of the input Note % that for
LTI filters that are not conjugate symmetric, a real sinusoid %will not be an
eigenfunction, as is readily verified.
```

MATLAB Example 3.5 In this exmple, we demonstrate the fact that periodic correlation of a real continuous time signal that is bandlimited can be efficiently computed using the Fourier series.

```
a1=0.5+j/sqrt(2);
a0=0.1;
a2=-1/sqrt(2)-0.1*sqrt(-1);
T=1;
t=linspace(0,1,100);
```

one period of a periodic signal $x(t)$ is calculated

```
e1=cos(2*pi*t/T)+sqrt(-1)*sin(2*pi*t/T);
e2=cos(2*pi*2*t/T)+sqrt(-1)*sin(2*pi*2*t/T);
x1 = a0*ones(1,length(t)) + a1*e1 + a1'*e1'.' + a2*e2 + a2'*e2'.';
```

We wish to compute the periodic correlation of the periodic signal $x(t)$
(one period of which is given by % $x1$) with itself. we do so using the Fourier series.

```
cor = sum([(abs(a0)^2)*ones(1,length(t)); (abs(a1)^2)*2*real(e1);
(abs(a2)^2)*2*real(e2)]);
```

```
subplot(211),plot(t,x1,'k')
grid on
xlabel('time, seconds');
ylabel('magnitude');
title('One Time period of the periodic signal');
subplot(212),plot(t,cor,'k')
grid on
xlabel('lag');
ylabel('magnitude');
title('Plot of the correlation');
```

Notice the symmetry in the correlation. What is this due to?

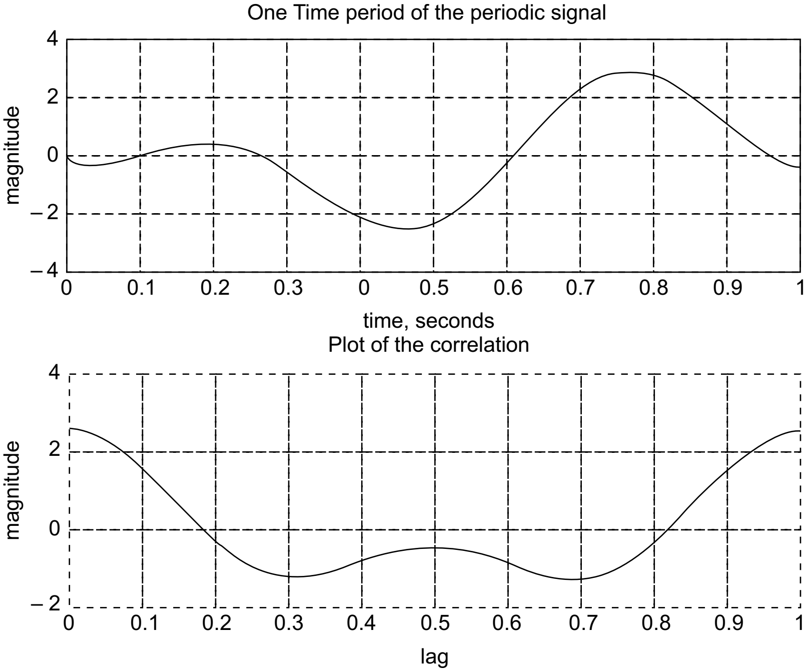


Fig. 3.55 *The figure depicts one time period of a continuous-time periodic signal (above), and its periodic correlation (below)*

MATLAB Example 3.6 In this example, we demonstrate that while a Hilbert transformer is difficult to build in general, a discrete-time equivalent is easily built, especially so when the range of frequencies of the input signals is small. We consider a simple case of length 3.

$h = [–4, 0, 4]$;

It is clear that this is a very poor approximation to a Hilbert tansformer. We will, however, use a sinusoidal input, so the magnitude imperfections do not matter.

```
N=100;
x = cos(2*pi*[0:999]/N);
```

Observe the output over one time period

```
y = 4*x(N:2*N-1)-4*x(N+1:2*N);
```

Note that the impulse response has been assumed to be centered so that

```
h(-1) =-4, h(0)=0 and h(1)=4
```

```
w = linspace (-pi,pi,100);
resp = 2*4*sqrt(-1)*sin(w);
subplot(211),plot(w,abs(resp),'k',w,angle(resp),'--k')
title('Magnitude and Phase responses');
xlabel('Frequency, randians/sec');
ylabel('Magnitude/angle');
axis([-pi,pi,-9,9])
legend('magnitude','angle');
subplot(212),plot(1:N,x(N+1:2*N),'k',1:N,y,'--k');
title('Output for Sinusoidal Input')
xlabel('sample')
ylabel('value');
legend('input','output')
```

Notice that the input and the output are Hilbert transform pairs—they are 90° phase shifted versions of each other.

It is just that the magnitude scaling is imperfect. For a small band of frequencies around this frequency, the magnitude of the output is approximately the same, and the distortion should be insignificant. Use of a filter of longer length will help reduce this distortion (make the frequency response of the filter close to the ideal Hilbert Transormer).

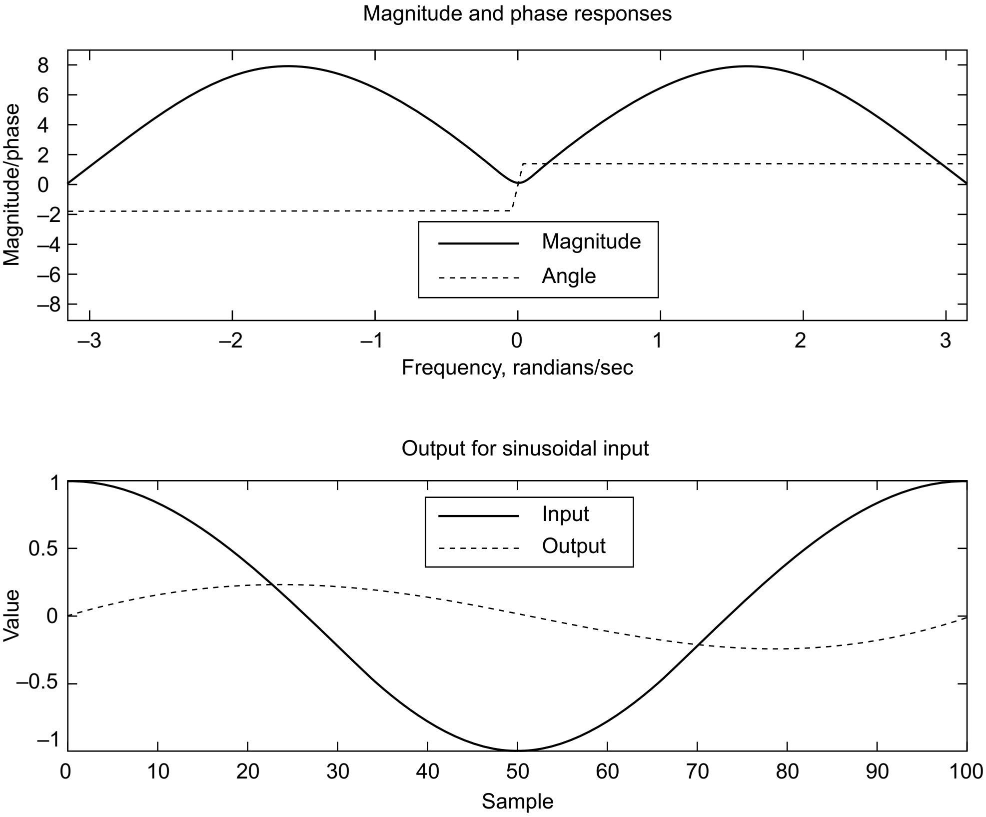


Fig. 3.56 *Figure depicts the magnitude and phase response of the filter chosen (above). Also plotted are the input and output of the filter (below). Note that they are Hilbert Transform pairs, except that the output needs to be scaled*

Summary

- Fourier transform of $x(t) = X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt$

  Inverse Fourier transform of $X(f) = x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi f t}\,df$

- (a) *Condition for the existence of the FT of x(t)* : $\int_{-\infty}^{\infty} |x(t)|dt < \infty$

  (b) *Dirichlet's conditions for convergence of FT and IFT*:
    - (i) $x(t)$ should be absolutely integrable.
    - (ii) $x(t)$ should have only a finite number of maxima and minima in any finite interval of time.
    - (iii) In any finite interval of time, the number of discontinuities of $x(t)$ should be finite.
    - (iv) Discontinuities of $x(t)$, if any, should be finite discontinuities.

- Properties of Fourier transform:

  (a) If $x(t) \xleftrightarrow{\text{FT}} X(f)$, then $X(0) = \int_{-\infty}^{\infty} x(t)\,dt$ = Area under $x(t)$.

  (b) $X(f)$ is in general, a complex function of frequency, even if $x(t)$ is a real-valued function.

  (c) If $x(t)$ is real-valued, $X(f)$ will have Hermitian symmetry.

- *Magnitude and phase spectra of an x(t):* A plot of $|X(f)|$ vs f is called the magnitude spectrum of $x(t)$. A plot of $\angle X(f)$ vs. f is called the phase spectrum of $x(t)$.

- *Fourier transform theorems:*

  (a) $\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2\, df =$ Energy of $x(t)$: Parseval's theorem

  (b) $[\alpha x(t) + \beta y(t)] \xleftrightarrow{\text{FT}} \alpha X(f) + \beta Y(f)$: Linearity theorem

  (c) $x(t-\tau) \xleftrightarrow{\text{FT}} X(f)e^{-j2\pi f\tau}$: Time-delay theorem

  (d) $x(t)e^{j2\pi f_c t} \xleftrightarrow{\text{FT}} X(f - f_c)$. : Modulation theorem

  (e) $x(at) \xleftrightarrow{\text{FT}} \dfrac{1}{|a|}X(f/a)$: Scaling theorem

  (f) $X(t) \xleftrightarrow{\text{FT}} x(-f)$: Duality theorem

  (g) If $z(t) = x(t) * y(t)$, then $Z(f) = Y(f) \cdot X(f)$: Convolution theorem

  (h) If $z(t) = x(t) \cdot y(t)$, then $Z(f) = X(f) * Y(f)$: Multiplication theorem

  (i) $\dot{x}(t) \xleftrightarrow{\text{FT}} j2\pi\, f X(f)$: Differentiation-in-time theorem

  (j) $-j2\pi t\; x(t) \xleftrightarrow{\text{FT}} \dfrac{d}{df}X(f)$: Differentiation-in-frequency theorem

  (k) $\int_{-\infty}^{t} x(\tau)d\tau \xleftrightarrow{\text{FT}} \dfrac{1}{2}\left[X(0)\delta(f) + \dfrac{X(f)}{j\pi f}\right]$: Integration theorem

- The convolution of two continuous-time signals, $x(t)$ and $y(t)$, is given by

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

- The correlation $R_{xy}(\tau)$ between two continuous-time energy signals is given by

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y^*(t-\tau)\,dt = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)\,dt$$

- The correlation between two continuous-time power signals is given by

$$R_{xy}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)y^*(t-\tau)\,dt = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t+\tau)y^*(t)\,dt$$

- The auto-correlation of a periodic signal $x(t)$ with period T_0 is given by

$$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)x^*(t-\tau)\,dt$$

- Relationship between convolution and correlation:

$$R_{xx}(\tau) = [x(t) * x(-t)]\big|_{t\to\tau}$$

- *ESD of an energy signal:* If $x(t)$ is an energy signal, its energy spectral density (ESD) is given by the FT of its auto-correlation function

$$R_{xx}(\tau) \xleftrightarrow{\text{FT}} S_{xx}(f)$$

and

$$R_{xx}(0) = \text{Energy of } x(t) = \int_{-\infty}^{\infty} S_{xx}(f)\,df$$

- The power spectral density (PSD) of a power signal, $x(t)$ is the FT of its ACF

$$R_{xx}(\tau) = \underset{T\to\infty}{\text{Lt}}\ \frac{1}{T}\int_{-T/2}^{T/2} x(t)x(t-\tau)\,dt \ = \text{ACF of } x(t)$$

and

$$P_{xx}(f) = \text{PSD of } x(t) = \text{F}[R_{xx}(\tau)]$$

- Properties of PSD:
  (a) $P_{xx}(f)$, the PSD of a signal $x(t)$, is always non-negative.
  (b) $P_{xx}(f)$ is the Fourier transform of $R_{xx}(\tau)$
  (c) The total area under the PSD curve of a signal equals the average power of the signal.
  (d) PSD of a real-valued power signal, $x(t)$, is an even function of frequency.
- Relationship between input and output spectral densities of an LTI system:
  (a) ESD: $S_{yy}(f) = |H(f)|^2 \cdot S_{xx}(f)$
  (b) PSD: $P_{yy}(f) = |H(f)|^2 \cdot P_{xx}(f)$
- (a) *Hilbert Transform*: $x(t) \xleftrightarrow{\text{FT}} \hat{x}(t) = x(t) * \dfrac{1}{\pi t}$
  (b) *Properties of Hilbert transform:*
    (i) Hilbert transform does not change the domain of a signal.
    (ii) Hilbert transform does not alter the amplitude spectrum of a signal.
    (iii) If $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$, then $\hat{x}(t) \xleftrightarrow{\text{HT}} -x(t)$. $\therefore \hat{\hat{x}} = -x(t)$.
    (iv) A signal and its Hilbert transform are orthogonal to each other.
- (a) *Analytic signal*: If $x(t)$ is a real-valued signal, its analytic signal, or pre-envelope is defined as: $x_+(t)\ \underline{\Delta}\ x(t) + j\hat{x}(t)$
  (b) *Spectrum of analytic signal*: $X_+(f) = \begin{cases} 2X(f) & \text{for } f > 0 \\ 0 & \text{for } f < 0 \end{cases}$
- Different representations of band pass signals:
  (a) $x(t) = a(t)\cos[\omega_c t + \theta(t)]$, where $a(t)$ and $\theta(t)$ are low pass signals – Envelope and phase representation.
  (b) $x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t$: $x_I(t)$ and $x_Q(t)$ are In-phase and Quadrature low pass signals.

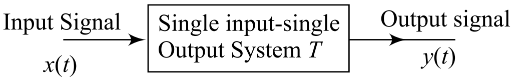
  Here, $a(t) = \sqrt{x_I^2(t) + x_Q^2(t)}$

and $$\theta(t) = \tan^{-1}\left[\frac{x_Q(t)}{x_I(t)}\right]$$

(c) $x(t) = \text{Re}[\tilde{x}(t)e^{j2\pi f_c t}] = \text{Re}[x_+(t)]$

where $\tilde{x}(t)$ = complex envelope of $x(t) = a(t)e^{j\theta(t)}$

- A *system* may be defined as an entity which acts on one or more inputs (or excitations) and produces one or more responses.



$$x(t) \xrightarrow{T} y(t)$$

- *Continuous-time systems*: these are defined as those systems which take continuous-time signals as input and produce continuous-time signals as output.
- *Discrete-time systems*: It takes a discrete-time signals as input and produces another discrete-time signal as output.
- *Static systems*: A system is said to be static or memoryless, or instantaneous, if its present output is determined entirely by its present input only. Static systems have their input-output relation described by algebraic equations.
- *Dynamic systems*: A system is said to be dynamic, or with memory, if its present output depends for its value not only on the present input, but also on some past inputs. Dynamic systems have their input-output relation described by different equations.
- *Linear and non-linear systems*: Let *T* be a continuous-time system which is at rest (i.e., all its energy storage elements are devoid of any stored energy). Let an input signal x_1(t) given to *T* result in an output signal $y_1(t)$; and an input $x_2(t)$ result in an output $y_2(t)$. Then the system *T* is said to be linear if for any pair of arbitrary constants $a_1$and a$_2$, an input of $[a_1x_1(t)+ a_2x_2(t)]$ given to the system *T* results in an output of $[a_1y_1(t)+ a_2y_2(t)]$. A continuous-time system not satisfying the above condition is said to be 'non-linear'.
- *Time-invariant and time-varying systems*: Let $x(t) \xrightarrow{T} y(t)$. Then *T* is said to be a time-invariant system if for any real number τ, $x(t-\tau) \xrightarrow{T} y(t-\tau)$. If this condition is not satisfied, the system is said to be time-varying.
- *Causal systems*: A system is said to be a 'causal system', or a 'non-anticipatory' system, if its output at any instant of time depends for its value only on the input at that instant and some previous instants, but not on the input at future instants.

Note *All physically realizable real-time systems must be causal.*

- *Impulse response, h(t) of an LTI system:* The impulse response, *h*(*t*), of an LTI system is the response of the system to a unit impulse given to it as input when it (the system) is at rest.
- *Complete characterization of an LTI system:* The *h*(*t*) of an LTI system completely characterizes the system in the sense that a knowledge of *h*(*t*) enables us to determine the response of the system for any arbitrary specified input.

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

These integrals are known as convolution integrals.

- *h(t) of a causal system:* For a causal LTI system, *h*(*t*) = 0 for *t* < 0.
- *Relation between step response and impulse response*: If *g*(*t*) is unit step response, then $h(t) = \dfrac{dg(t)}{dt}$.
- *Bounded signal:* A signal, *x*(*t*), is said to be a bounded signal, bound to a value *M*, where *M* is a positive real number, provided the magnitude of *x*(*t*) never exceeds *M*.
- *BIBO stability criterion:* A system *T* is stable in the Bounded-input, Bounded-output sense, provided every bounded input given to it results in an output signal that also bounded.
- *BIBO stability theorem:* An LTI system T is stable in the BIBO sense iff its impulse response, *h*(*t*) is absolutely integrable, i.e., iff $\int_{-\infty}^{\infty} |h(t)|dt < \infty$.

- *Eigen signals of an LTI system:* An Eigen signal of a system T is a signal, which, when given as input to the system, gives rise to an output signal which is essentially the same as the input signal except for a change in the amplitude and possibly a shift in time. For an LTI system, a complex exponential of any frequency is an Eigen signal.
- *Transfer function or frequency response of an LTI system:* The transfer function, or the frequency response, $H(f)$ of an LTI system is the Fourier transform of its impulses response. It is the ratio of the output to the input of the LTI system when the input is the Eigen signal $\exp(j2\pi ft)$.
- Magnitude and phase responses of a system:
  A plot of $|H(f)|$ vs. frequency is called the magnitude response.
  A plot of $\angle H(f)$ vs. frequency is called the phase response.
- For any real system, magnitude response will have even symmetry and the phase response will have odd symmetry.
- *Condition for distortionless transmission:* For distortionless transmission through LTI system, the system's magnitude response, $|H(f)|$ should be a constant, independent of frequency and its phase response $\theta(f) = \angle H(f)$ should be proportional to frequency.
- *Ideal low pass filter:* For an ideal LPF with pass band gain A, pass band bandwidth B and a linear phase response with a slope of $-2\pi\tau$, impulse response $h(t) = 2AB \operatorname{sinc} 2B(t-\tau)$ and $H(f) = A\Pi(f/2B)e^{-j2\pi f\tau}$
- *Ideal band pass filter:* For an ideal BPF with passband from f_1 to f_2, pass band gain A and a time delay t_0 sec,
  $$H(f) = A\Pi[(f+f_0)/B] + A\Pi[(f-f_0)/B];\ \theta(f) = -2\pi f t_0$$
  and $$h(t) = 2AB\,[\sin c\ B\,(t-t_0)]\cos 2\pi f_0(t-t_0)$$
- *Paley–Wiener criterion:* It permits us to determine the physical realizability, or otherwise of an LTI system directly from the transfer function $H(f)$ of the system.
  It says that an LTI system with a given $|H(f)|$ which is square integrable, is physically realizable if
  $$\int_{-\infty}^{\infty} \frac{|\log_e |H(f)||}{(1+f^2)}\,df < \infty$$
- *Rise time and bandwidth:*
  (a) For a first-order RC low pass filter, rise time $t_r = \dfrac{0.35}{B}$ where B is its half-power bandwidth.
  (b) For an ideal LPF of bandwidth B Hz, the rise time $t_r = \dfrac{0.4}{B}$

References and Suggested Reading

1. Carlson, Bruce *et al.*, *Communication Systems – An Introduction to signals and Noise in Electrical Communication*, 4th Edition, McGraw-Hill International Edition, 2002.
2. Haykin, Simon, and Barry Van Veen, *Signals and Systems*, 2nd Edition, John Wiley & Sons (Asia), 2004.
3. Ramakrishma Rao, P., *Signals and Systems*, Tata McGraw-Hill, 2008.
4. Ziemer, R, W.H. Tranter, and D. Ronald Fannin, *Signals and Systems: Continuous and Discrete*, 2nd Edition, Macmillan Publishing Company, New York, 1990.

Review Questions

1. If $X(f)$, the Fourier transform of $x(t)$, has Hermitian symmetry, comment on the nature of $x(t)$?

2. Write down the Fourier transform of $x(t) = 5\sin(\omega_0 t + \theta)$.

3. If $X(f)$ is the Fourier transform of $x(t)$, what does $|x(f)|^2$ represent in relation to the signal $x(t)$?

4. Sketch the magnitude and phase spectra of the signal $x(t) = 20\sin(50\pi t + 45°)$.

5. Explain the usefulness of the convolution theorem of Fourier transform in determining the convolution of two continuous-time signals.

6. Explain, graphically, the difference between convolution and correlation of two continuous-time signals.

7. Show that the ACF of two real-valued continuous-time signals is an even function of τ, the lag parameter.

8. Show that the power spectral density of a power signal, $x(t)$, is the Fourier transform of its auto-correlation function.

9. Derive the relation between the output signal and input signal power spectral densities of an LTI system.

10. Sketch the magnitude and phase responses of a Hilbert transformer.

11. Show that $\hat{\hat{x}}(t) = -x(t)$

12. Define the 'analytic signal' of a real-valued signal $x(t)$.

13. Define and explain the significance of the 'Complex Envelope' of a real-valued band pass signal $x(t)$.

14. Define the terms 'static system' and 'dynamic system' and give one example for each of these.

15. How do you define 'linearity property' of a system?

16. The input-output relationship of a system is as shown in Fig. R3.16.
   (a) Is this system linear? Justify your answer.
   (b) Is this system static or dynamic? Why?

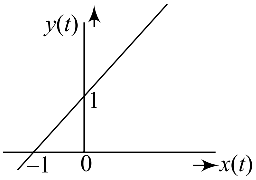


Fig. R3.16

17. The impulse response of a certain system is $h(t) = A\Pi\left(\frac{t-T/2}{T}\right)$. Is this system static or dynamic? Why? What is the input-output relationship for the system?

18. Define the 'time-invariance' property of a continuous-time system? Give examples of a time-varying system.

19. Define the following terms:
   (a) Impulse response
   (b) Causality, as applied to systems

20. What is the relationship between 'impulse response' and 'step response' of an LTI system?

21. What is the condition on the impulse response of an LTI system for the system to be stable in the BIBO sense?

22. Define the terms 'Eigensignal' and 'Transfer Function' of a stable LTI system.

23. State the two conditions required to be satisfied by an LTI system for an input signal to pass through it without any distortion?

24. Why are the ideal LPF, HPF and BPF not physically realizable?

Problems

1. Find the Fourier transforms of the following signals:
   (a) $x(t) = e^{-3t}u(t-2)$
   (b) $x(t) = e^{-2|t|}$
   (c) $x(t) = 2te^{-2t}u(t)$
   (d) $x(t)$ shown in Fig. P3.1
   (e) $x(t) = [\exp\{j2\pi(t-1) - (t-1)\}]u(t-1)$

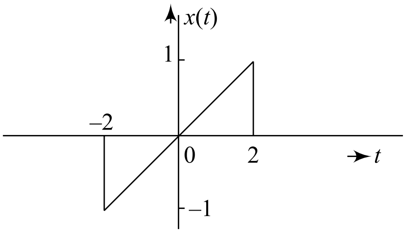
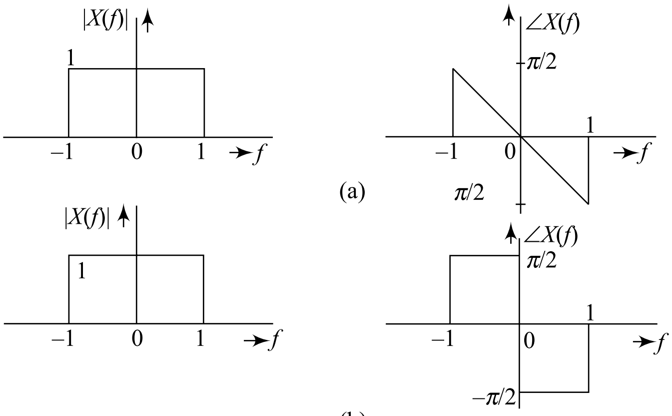


Fig. P3.1

2. Find the signal $x(t)$ whose Fourier transform $X(f)$ is given in (i) Fig. P3.2(a) and (ii) Fig. P 3.2(b).



(a)

(b)

Fig. P3.2

3. Use Parseval's theorem to calculate the energy in the signal $x(t) = 4\text{ sinc }40t$.

4. Calculate the energy contained in the signal in problem 3 for $|f| \le \dfrac{3}{2\pi}$. Express it as a percentage of the total energy of the signal.

5. Find the convolution of $x(t) = 5\Pi(t/4)$ with $y(t) = 5\Pi(t/4)$.

6. Find the Fourier transform of $z(t) = 100\Lambda(t/8)$ where $100\,\Lambda(t/8)$ is a triangular pulse symmetrical about the $t = 0$ axis and having a peak amplitude of 100 and a total base width of 8 sec. (*Hint:* Use the result of Problem 20 and the convolution theorem of Fourier transform).

7. Given that $X(f)$ is the Fourier transform of $x(t)$, find the Fourier transforms of the following:

(a) $y(t) = 2x(3t - 2)$ (b) $y(t) = x\left(\dfrac{t}{2} - 1\right)e^{j200\pi t}$ (c) $y(t) = x(1 - 2t)$

8. Find the Fourier transforms of the signals shown in Figs. P3.8(a) to (e).

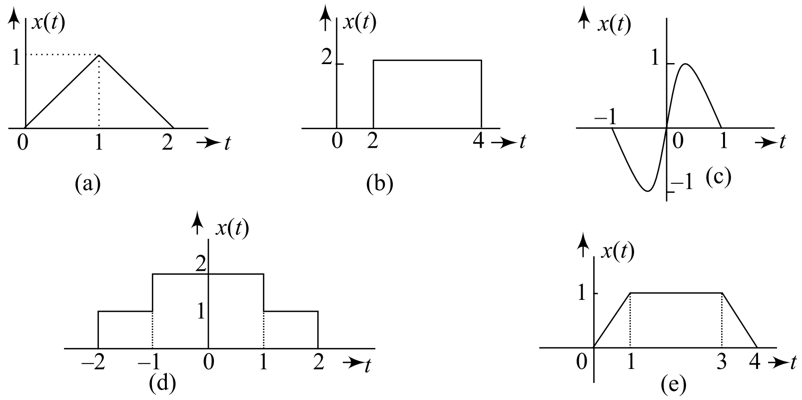


Fig. P3.8

9. Find $x(t)$ if its Fourier transform $X(f)$ is given by

(a) $X(f) = \dfrac{j2\pi f}{(1 + j2\pi f)^2}$

(b) $X(f) = \dfrac{2\sin^2\omega}{\omega^2}$

(c) $X(f) = 5\left[\dfrac{\text{sinc }4f}{(1 + j2\pi f)}\right]$

(d)

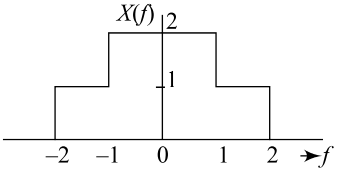


Fig. P3.9

10. If the signal shown in Figs. P3.3(a), (b) and (d) of Problem 8 are multiplied by cos $50\pi t$, determine and sketch the magnitudes of the Fourier transforms of the resulting signals.

11. Determine the Fourier transform of the $x(t)$ shown in Fig. P3.11.
  (a) By applying time-domain differentiation theorem.
  (b) By identifying $x(t)$ as having been obtained by the convolution of $\Pi(t/T)$ with itself and scaling down the magnitude by T and then applying convolution theorem.

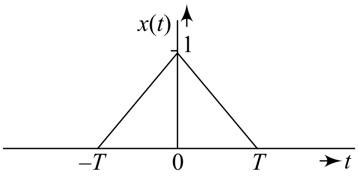


Fig. P3.11

12. Find $R_{xy}(\tau)$ given that $x(t) = 3\cos\omega_0 t$ and $y(t) = 2\cos\omega_0 t$.

13. Determine the ACF $R_{xx}(\tau)$ for the signal $x(t)$ of Fig. P.3.13. Take the FT and determine its power spectral density.

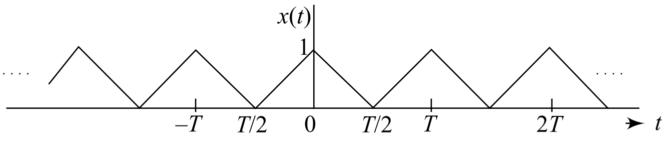


Fig. P3.13

14. Find the cross-correlation $R_{xy}(\tau)$ of the periodic signals $x(t)$ and $y(t)$ shown in Figs. P3.13 and P3.14.

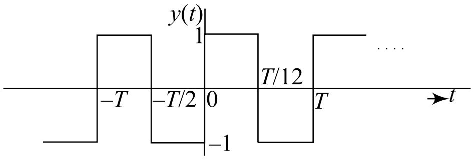


Fig. P3.14

15. Find the ACF $R_{xx}(\tau)$ and energy spectral density, $S_{xx}(f)$ of the rectangular pulse shown in Fig. P3.15.

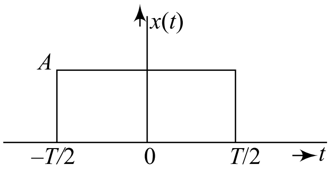


Fig. P3.15

16. Determine the transfer function of an LTI system T if the system is to give as its output the cross-correlation between the input $x(t)$ and the function $z(t)$ given by

$$z(t) = \begin{cases} 10e^{2t} & \text{for } t < 0 \\ 0 & \text{for } t > 0 \end{cases}$$

17. For the signals shown in Fig. P3.17 determine the following:

(a) $x(t) * y(t)$ (b) $R_{xy}(\tau)$ (c) $x(t) * z(t)$ (d) $R_{xz}(\tau)$
(e) $x(t) * w(t)$ (f) $R_{xw}(\tau)$ (g) $y(t) * z(t)$ (h) $R_{yz}(\tau)$
(i) $y(t) * w(t)$ (j) $R_{yw}(\tau)$ (k) $z(t) * w(t)$ (l) $R_{zw}(\tau)$

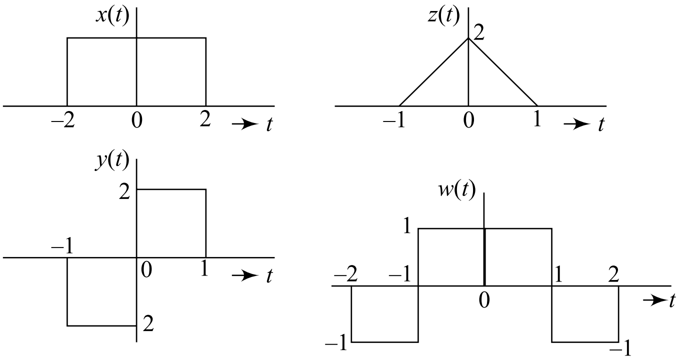


Fig. P3.17

18. Find the power spectral density of $x(t) = 10 \cos 20\pi t$. What will be the power spectral density of $10 \sin 20\pi t$?

19. Referring to Fig. P3.19, determine (a) $R_{yy}(\tau)$, (b) $R_{yx}(\tau)$ in terms of $R_{xx}(\tau)$.

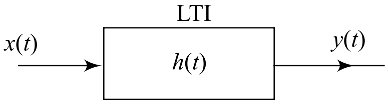


Fig. P3.19

20. The power spectral density of a certain signal is given by

$$P_{xx}(f) = \frac{4}{4 + 4\pi^2 f^2}$$

What is the rms value of the signal?

21. If $x(t) = (1/t)\sin t$, show that $\hat{x}(t) = (1/t)(1 - \cos t)$.

22. $x_1(t)$ and $x_2(t)$ are two narrowband signals centered on the same carrier frequency, f_c. If $x_3(t) = x_1(t) + x_2(t)$, show that $\tilde{x}_3(t) = \tilde{x}_1(t) + \tilde{x}_2(t)$ where $\tilde{x}_i(t)$ is the complex envelope of $x_1(t)$?

23. Find the Hilbert transforms of the following signals and show in each case that the signal and its transform are orthogonal:

(a) $x(t) = \sin \omega_0 t$ (b) $x(t) = 5 \cos 60\,\pi t \cdot \cos 6 \times 10^4\,\pi t$

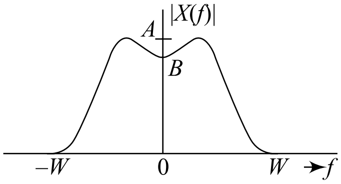


Fig. P3.24

24. $X(f)$ shown in Fig. P3.24 is the Fourier transform of a signal $x(t)$ and is real. Determine and sketch the spectrum of each of the following signals:

(a) $y(t)=\frac{1}{2}[x(t)+j\hat{x}(t)]$ (b) $z(t)=[x(t)+j\hat{x}(t)]e^{j2\pi f_c t}$ where $f_c >> W$

(c) $w(t)=[x(t)-j\hat{x}(t)]e^{-j2\pi f_c t}$ where $f_c >> W$

25. Sketch the signals

(a) $x(t)=200\,\text{sinc}\,(200t)\cos 2\pi 10^4\, t$

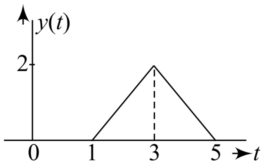
(b) $y(t)=[x(t)+j\hat{x}(t)]$ where $x(t)$ is as given in part (a).

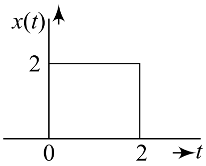
(c) Determine $\tilde{x}(t)$, the complex envelope of $x(t)$ and sketch its spectrum.

26. Determine whether the following systems with the given input-output relationship are linear or non-linear, static or dynamic, time invariant or time varying and causal or non-causal.

(a) $y(t) = x(t + 5)$ (b) $y(t) = 2x(t) + 3$ (c) $y(t) = x(t^2)$

(d)





(e) $y(t) = |x(t)|$ (f) $y(t)=\int_{-\infty}^{t} x(\lambda)\,d\lambda$ (g) $y(t)=\frac{d}{dt}x(t)$ (h) $y(t)=5x(2t+3)$

(i) $y(t)=\log_e|x(t)|$

27. Determine whether the following systems are static or dynamic, linear or non-linear, time invariant or time varying, causal or non-causal.

(a) $2\frac{dy(t)}{dt}+4y(t)+6=3x(t)$

(b) $2\frac{dy(t)}{dt}+10y(t)=2x(t)$

(c) $3\frac{dy(t)}{dt}+ty(t)=4x(t)$

(d) $5\frac{dy(t)}{dt}+2y^2(t)=3x(t)$

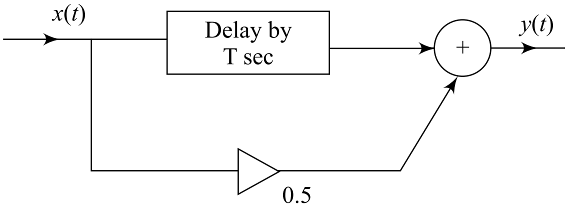


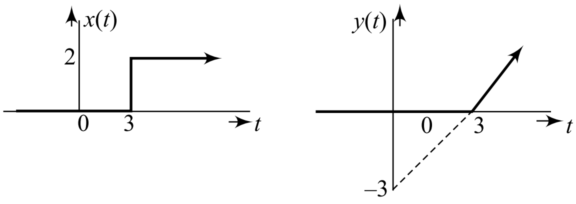
Fig. P3.29

28. Show that an ideal integrator is an LTI system.

29. Obtain the impulse response and the transfer function $H(f)$ of the system shown in Fig. P3.29. Plot its magnitude response:

30. Find the impulse responses of the systems given in Problem 2(a) and (b).

31. The input $x(t)$ and the corresponding output $y(t)$ of a causal LTI system T are as shown in Fig. P3.31. Find the impulse response function $h(t)$ of the system



(a) Input $x(t)$ (b) Output $y(t)$

Fig. P3.31

32. An LTI system has an impulse response of $e^{-t}\cos(100\,\pi t)u(t)$. Determine the output of the system for an input of $x(t)=\cos(100\,\pi t)u(t)$.

33. Show that a sinusoid/co-sinusoid of any frequency is an Eigen signal of any LTI system.

34. An LTI system has an impulse response $h(t)=e^{-t}u(t)$. For an input $x(t)=10\Pi(t/4)$, determine the output.

35. Find the frequency response, $H(f)$, of the system described by $\frac{d^3y(t)}{dt^3}+0.5\frac{d^2y(t)}{dt^2}+0.75\frac{dy(t)}{dt}+2y(t)=x(t)$

36. For the system shown in Fig. P3.36, find the frequency response.

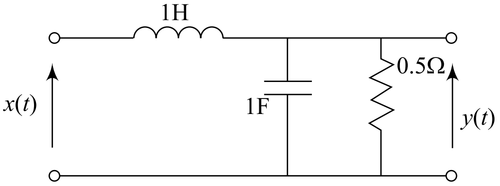


Fig. P3.36

37. The impulse response of a system is $h(t) = 10e^{-3t}u(t)$. Find and plot the response of the system to an input $x(t) = \Pi\left(\frac{t-1}{2}\right)$.

Multiple-Choice Questions

1. Strictly speaking, which one of the following signals is not Fourier transformable?
(a) $e^{-|t|}$ (b) rect(t/τ) (c) tr(t/τ) (d) $\sin \omega_0 t$

2. If the signal $x(t)$ is real valued and its Fourier transform is $X(f)$ then
(a) $X(f)$ is real valued (b) $|X(f)| = |X(-f)|$
(c) $X(f)$ has even symmetry (d) $X(f)$ has odd symmetry

3. If $x(t) = 10$ rect $(t/2)$, the zero-frequency value of its spectrum is given by
(a) 10 (b) 5 (c) 2 (d) 20

4. Shifting a time signal along the time axis causes
(a) a change in the amplitude spectrum (b) a change in both amplitude and phase spectrum
(c) a change only in the phase spectrum (d) no change in amplitude as well as phase spectrum

5. If $x(t) = 10$ sinc $5t$, the energy contained in the signal is
(a) 100 (b) 50 (c) 10 (d) 20

6. If $y(t) \underline{\Delta} x(t) * \delta(t - \tau)$, $Y(f)$ is given by
(a) $X(f)e^{+j2\pi f\tau}$ (b) $X(f)e^{-j2\pi f\tau}$
(c) $X(f - f_c)$ where $f_c \underline{\Delta} 1/\tau$ (d) It is not Fourier transformable

7. The Fourier transform of t sinc $10t$ is equal to
(a) $\frac{1}{j20\pi}[\Pi(f/10)]$ (c) $\frac{f}{20\pi}[\Pi(f/10)]$
(b) $\frac{1}{20j\pi}[\delta(f+5) - \delta(f-5)]$ (d) $\frac{j}{20\pi}[\delta(f+5) - \delta(f-5)]$

8. If $y(t) = x(2-t)$, $Y(f)$ is given by
(a) $X(-f)e^{-j4\pi f}$ (b) $X(f)e^{-j4\pi f}$ (c) $X(-f)e^{j4\pi f}$ (d) $X(f)e^{j4\pi f}$

9. The Fourier transform of $e^{at}u(-t)$ is
(a) $\frac{1}{a - j\omega}$ (b) $\frac{1}{-a + j\omega}$ (c) $\frac{1}{a + j\omega}$ (d) $\frac{-1}{a - j\omega}$

10. $x(t) = 10$ sinc $2t$ and $y(t) = \cos 200\,\pi t$. The spectrum $z(f)$ if $z(t) \underline{\Delta} x(t) \cdot y(t)$ is given by
(a) $10\,[\text{sinc}\,2(f - 100) + \text{sinc}\,2(f + 100)]$ (b) $5\left[\Pi\left(\frac{(f+100)}{2}\right) + \Pi\left(\frac{(f-100)}{2}\right)\right]$
(c) $10\left[\Pi\left(\frac{(f+200)}{2}\right) + \Pi\left(\frac{(f-200)}{2}\right)\right]$ (d) $5\left[\Pi\left(\frac{(f+200)}{2}\right) + \Pi\left(\frac{(f-200)}{2}\right)\right]$

11. It is possible to compute the cross-correlation $R_{xy}(\tau)$ between two signals $x(t)$ and $y(t)$ directly from their convolution provided

(a) $x(t)$ has even symmetry (b) $x(t)$ has odd symmetry
(c) $y(t)$ has odd symmetry (d) $y(t)$ has even symmetry

12. If $x(t) = 10\Pi(t/10)$, $S_{xx}(f)$ is a

(a) sinc function (b) triangular function
(c) sinc-square function (d) rectangular function

13. $x(t) = 5\Pi(t/10)$, the maximum value of $R_{xx}(\tau)$ is

(a) 250 (b) 50 (c) 500 (d) 25

14. $x(t) = 10\Pi(t/10)$, the maximum value of $S_{xx}(0)$ is

(a) 100 (b) 1000 (c) 500 (d) 5000

15. $x(t) = 10\Pi(t/10)$. The total area under the $S_{xx}(f)$ curve is

(a) 1000 (b) 500 (c) 100 (d) 10000

16. The signal $e^{-t}u(t)$ is applied as input to an L-section RC low pass filter with time constant equal to 1. The energy spectral density at the output of the filter at the 3db cut-off frequency of the filter is

(a) 1 (b) 0.5 (c) 0.25 (d) 1.5

17. If $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$, then their Fourier transforms are related as

(a) $\hat{X}(f) = j\,\text{sgn}(f)X(f)$ (b) $\hat{X}(f) = j\,\text{sgn}(f)X(f)$
(c) $\hat{X}(f) = -j\,\text{sgn}(f)X(-f)$ (d) $\hat{X}(f) = j\,\text{sgn}(f)X(-f)$

18. If $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$; $\hat{X}(f) = \mathcal{F}[\hat{x}(t)]$, and $\hat{X}(f) = |\hat{X}(f)|e^{j\hat{\theta}(f)}$ then

(a) $|\hat{X}(f)| = -|X(f)|$ (b) $\hat{X}(f) = \overline{X(f)}$
(c) $|\hat{X}(f)| = |X(f)|$ and $\hat{\theta}(f) = 90°$ (d) $|\hat{X}(f)| = |X(f)|$ and $\hat{\theta}(f) = -90°$

19. $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$, then $\hat{\hat{x}}(t)$ equals

(a) $-x(t)$ (b) $x(t)$ (c) $\overline{x(t)}$ (d) $x(-t)$

20. $\cos 20\pi\,t \cdot \cos 2000\pi\,t \triangleq x(t)$. Then $\hat{x}(t)$ is

(a) $\sin 20\pi\,t \cdot \sin 2000\pi\,t$ (b) $\sin 20\pi\,t \cdot \cos 2000\pi\,t$
(c) $\cos 20\pi\,t \cdot \sin 2000\pi\,t$ (d) None of these

21. If $x_+(t)$ is the analytic signal corresponding to the real-valued signal $x(t)$, and if $X_+(f) = \mathcal{F}[x_+(t)]$, then $X_+(f)u(-f)$ is given by

(a) 0 (b) $2X(f)$ (c) $2X(-f)$ (d) None of these

22. If $x(t) \xleftrightarrow{\text{HT}} \hat{x}(t)$ and $y(t) \triangleq \hat{x}(t) - jx(t)$, then $Y(f)u(-f)$ is

(a) 0 (b) $2X(f)$ (c) $-2X(f)$ (d) None of these

23. If $\tilde{x}(t)$ is the complex envelope of a real-valued band pass signal $x(t)$ and if $\tilde{x}(t) = x_1(t) + jx_2(t)$ then $x(t)$ is given by

(a) $x_1(t)\sin\omega_c t - x_2(t)\cos\omega_c t$ (b) $x_1(t)\cos\omega_c t + x_2(t)\sin\omega_c t$
(c) $x_1(t)\cos\omega_c t - jx_2(t)\sin\omega_c t$ (d) $x_1(t)\cos\omega_c t - x_2(t)\sin\omega_c t$

24. If $x(t)$, a real-valued band pass signal is given by $x(t) = (10\cos 20\pi t)\cos[20000\,\pi t + \pi/4]$, then the magnitude of its analytic signal $x_+(t)$ is

(a) $10\cos 20\pi t$ (b) 10 (c) $10\sin 20\pi t$ (d) 5

25. (Choose the incorrect answer). A system composed of purely resistive networks is

(a) dynamic (b) linear (c) time-invariant (d) static

26. The system with $y(t) = x(3t)$ is

(a) static (b) linear (c) fixed (d) causal

27. The LTI system with $h(t) = e^{-t}, -\infty < t < \infty$, is

(a) Causal and stable (c) non-causal and stable
(b) Causal and unstable (d) non-causal and unstable

28. An LTI system has $h(t) = 5\delta(t-2)$

(a) It is unstable (b) It is delayer

(c) It amplifies and delays the input signal (d) It samples the input signal.

29. Two LTI systems with impulse responses $h_1(t)$ and $h_2(t)$ are connected in series (cascade). the impulse response of the overall system is

(a) $h_1(t)+h_2(t)$ (b) $\dfrac{h_1(t)h_2(t)}{h_1(t)+h_2(t)}$ (c) $h_1(t)*h_2(t)$ (d) $h_1(t)\cdot h_2(t)$

30. When two LTI systems with impulse responses $h_1(t)$ and $h_2(t)$ are connected in parallel, the impulse response of the overall system is

(a) $h_1(t)+h_2(t)$ (b) $\dfrac{h_1(t)h_2(t)}{h_1(t)+h_2(t)}$ (c) $h_1(t)*h_2(t)$ (d) $h_1(t)\cdot h_2(t)$

31. $x_1(t) = A[u(t)-u(t-10)]$ and $x_2(t) = B[u(t)-u(t-5)]$. $x_1(t)*x_2(t)$ is a

(a) triangular pulse (b) rectangular pulse (c) trapezoidal pulse (d) sinc pulse

32. $x_1(t) = A[u(t)-u(t-5)]$ and $x_2(t) = B[u(t)-u(t-5)]$. $x_1(t)*x_2(t)$ is a

(a) triangular pulse (b) rectangular pulse (c) trapezoidal pulse (d) sinc pulse

33. The transfer function *H(f)* of the *RC* low pass filter shown in Fig. M3.33, is given by

(a) $\dfrac{1}{1+j2\pi f RC}$ (b) $\dfrac{1/\text{RC}}{1+j2\pi f RC}$

(c) $\dfrac{1}{1-j2\pi f RC}$ (d) $j2\pi f RC$

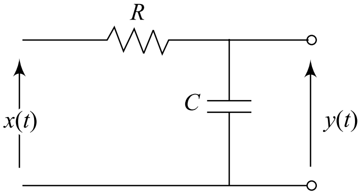


Fig. M3.33

34. The 3db cut-off frequency for the filter of Question 9 is

(a) $\dfrac{1}{RC}$ (b) RC (c) $\dfrac{1}{2\pi RC}$ (d) None of these

35. Two continuous-time LTI systems, each with an impulse response function $h(t) = \dfrac{\sin(at)}{at}$, are connected in cascade. Then, the impulse response of the overall system is

(a) $k\dfrac{\sin(at)}{at}$ (b) $k\left[\dfrac{\sin(at)}{at}\right]^2$

(c) $\dfrac{\sin bt}{bt}$, with *b* not necessarily equal to *a* (d) None of these

36. If $*$ denotes convolution operation and over bar denotes complex conjugation, the relation $y(t) = \int_{-\infty}^{\infty} x(\tau)\overline{x(t+\tau)}\,d\tau$ can be expressed as

(a) $x(t)*\overline{x(t)}$ (b) $x(t)*\overline{x(-t)}$ (c) $x(-t)*\overline{x(-t)}$ (d) None of these

37. A signal $x(t) = [\sin(\pi t)/(\pi t)]^2$ is passed through an LTI system with impulse response $h(t) = \sin(2\pi t)/(\pi t)$. The output *y*(*t*) of the system is

(a) *x*(*t*) (b) cannot be of the form of *x*(*t*)

(c) of the form of a sinc pulse (d) None of the above

38. When the input to an LTI system is a unit step function, the output is a bounded signal. Which of the following inferences is correct?

(a) The system is not necessarily stable (b) The system is not definitely stable

(c) The system is definitely unstable (d) None of the above

39. Signal transmission through an LTI system cannot be distortionless unless

(a) |*H*(*f*)| is constant for all frequencies and phase-shift is proportional to frequency.

(b) |*H*(*f*)| remains constant and phase-shift is proportional to frequency at least over the signal bandwidth

(c) |*H*(*f*)| and phase shift are both independent of frequency

(d) |*H*(*f*)| and phase shift are both proportional to frequency

40. An LTI system has a gain independent of frequency and produces a time delay of τ sec for all frequencies. Which of the following statements is true?

(a) It produces phase distortion
(b) Its phase-shift vs. frequency relationship is linear
(c) It produces a constant phase shift for all frequencies
(d) None of the above

41. An LTI system with flat magnitude response, is producing a constant time delay of τ sec for all frequencies. If $h(t)$ is the impulse response of the system,

(a) $h(t)$ takes a maximum value at $t = \tau/2$
(b) $h(t)$ takes a minimum value at $t = \tau/2$
(c) $h(t)$ takes a maximum value at $t = \tau$
(d) $h(t)$ takes a minimum value at $t = \tau$

42. The signal $x(t) = 10 \text{ sinc } 20t$ is applied as the input signal to an LTI system. The minimum bandwidth over which the gain of the system should be constant and the phase response should be linear, for distortionless transmission of the signal, is

(a) 5 Hz (b) 10 Hz (c) 20 Hz (d) None of these

43. The impulse response, $h(t)$ of an ideal LPF having transfer function $H(f) = A\Pi(f/2B)e^{-j2\pi f\tau}$ is given by

(a) $A \text{ sinc } 2B(t-\tau)$ (b) $2\,AB \text{ sinc } B(t-\tau)$ (c) $AB \text{ sinc } 2B(t-\tau)$ (d) $B \text{ sinc } 2B(t-\tau)$

44. For a first-order RC low pass filter with 3-db bandwidth of B Hz, the 10% to 90% rise time is given by

(a) $\frac{0.35}{B}$ (b) $\frac{2.5}{B}$ (c) $\frac{4.5}{B}$ (d) $\frac{3.5}{B}$

45. For an ideal LPF of bandwidth B Hz, the 10% to 100% rise time is given approximately by

(a) $\frac{3.5}{B}$ (b) $\frac{2.5}{B}$ (c) $\frac{4.0}{B}$ (d) $\frac{0.40}{B}$

Key to Multiple-Choice Questions

1. (d)	2. (b)	3. (d)	4. (c)	5. (d)	6. (b)	7. (d)	8. (a)
9. (a)	10. (b)	11. (d)	12. (c)	13. (a)	14. (d)	15. (a)	16. (c)
17. (b)	18. (d)	19. (a)	20. (c)	21. (a)	22. (c)	23. (d)	24. (a)
25. (a)	26. (b)	27. (d)	28. (c)	29. (c)	30. (a)	31. (c)	32. (a)
33. (a)	34. (c)	35. (a)	36. (b)	37. (a)	38. (a)	39. (b)	40. (b)
41. (c)	42. (b)	43. (c)	44. (a)	45. (d)			

AMPLITUDE MODULATION

4

"You are who you are and what you are because of what has gone into your mind. You can change who you are and what you are by changing what goes into your mind."

Zig Ziglar (born 1926 – present)
American author, salesman, and motivational speaker

Learning Objectives

After going through this chapter, students will be able to

- *have a clear idea of the meaning of modulation and the need for modulation,*
- *give the time-domain representation, spectrum and the methods of generation and detection of amplitude modulated signals,*
- *explain the operation of an envelope detector, the types of distortions that can arise and the reason behind each type of distortion,*
- *give the time-domain and frequency-domain representation as well as the methods of generation and detection of DSB-SC, SSB-SC and VSB signals,*
- *list the key specifications for AM audio broadcast transmitters, state the merits and demerits of high-level and low-level modulation in AM transmitters, and draw their block diagrams and explain their working,*
- *explain clearly the problems like image frequency interference and adjacent channel interference as well as the effect of the choice of the IF, in the case of AM superheterodyne receivers,*
- *draw the block diagrams of AM superheterodyne receivers, and SSB-SC transmitters and receivers, and*
- *explain the principle of frequency division multiplexing (FDM), its implementation and the AT&T FDM hierarchy.*

4.1 INTRODUCTION

Communication basically involves transmission of information from one point to another. The information bearing signals which are to be transmitted, may be in the form of speech, music or image signals. These signals cannot be transmitted directly and need some pre-processing. This pre-processing needed for making them suitable for transmission is called '***modulation***'.

4.1.1 What is Modulation?

This modulation process consists of varying from instant to instant, one of the parameters of a high frequency sinusoidal signal, called the ***carrier signal*** in accordance with the instantaneous amplitude of the information bearing message signal. In general, a sinusoidal carrier wave may be represented by

$$c(t) = A_c \cos(2\pi f_c t + \theta) \tag{4.1}$$

There are three parameters associated with the carrier signal. These are A_c, the amplitude; f_c, the frequency; and θ, the phase. Depending on which one of these parameters is varied in the modulation process in accordance with the message signal amplitude, the modulation is called 'amplitude modulation', 'frequency modulation', or 'phase modulation'.

4.1.2 Need for Modulation

1. Antenna Size Long distance communication is invariably by the propagation of electromagnetic waves through the atmosphere, or free-space. This requires efficient radiation of electromagnetic waves from an antenna. The information-bearing signals like speech, etc., are basically low frequency signals. For instance, a speech signal may typically have frequency components from a few hundred hertz up to a maximum of 10 kHz. We know that for an antenna to efficiently radiate a signal fed to it, the physical size of the antenna has to be at least of the order of 0.1 λ where λ is the wavelength of the signal fed to it. Even if we consider the highest frequency component of speech, viz. 10 kHz, the minimum length required for the antenna works out to 3 km, which is definitely not practicable. Hence, we have to raise the frequency of the information-bearing signal, speech, to a level at which an antenna of reasonable size can efficiently radiate it. This process of translating a low frequency information-bearing signal to a high frequency slot is achieved by modulation. Modulation is necessary not only from the point of view of having an antenna of reasonable size to radiate the modulated signal. It is essential because of various other reasons too, as noted below.

2. Selecting the desired signal Consider a high frequency carrier modulated by a low frequency information-bearing signal, say, a speech signal, being radiated by a transmitting antenna. The receiving antenna may be tuned to that particular carrier frequency so that only the desired speech signal is received and all other modulated signals reaching the receiving antenna are rejected. But, if there is no modulation and if we assume that several transmitting stations are simultaneously radiating a number of different speech signals, since all speech signals occupy the same spectrum, how are we going to select one particular speech signal in which we are interested and reject all the others?

3. Multiplexing Multiplexing is the technique used for transmitting several information-bearing signals simultaneously over the same physical channel. Modulation process makes it possible to multiplex several message signals and transmit them simultaneously by using different carrier frequencies for the various message signals.

4.2 AMPLITUDE MODULATION

First, let us clearly state the terminology and the notation that is widely used in literature and adopted here. The message signal which is used for modulating the carrier signal, is called the '*modulating signal*', or the 'message signal' and is denoted by $x(t)$. The signal that results after the modulation process, is referred to as the *modulated signal* and is denoted by $x_c(t)$. The carrier signal is denoted by $c(t)$.

Amplitude modulation is the earliest and one of the most widely used type of modulation. Its main virtue is the simplicity of its implementation.

Definition Amplitude Modulation (AM) is that type of modulation in which, the amplitude of the carrier is changed from instant to instant in such a way that at any instant of time, the ***change*** in the peak amplitude of the carrier from its unmodulated value, is directly proportional to the instantaneous amplitude of the modulating signal.

4.2.1 Time-Domain Description

Let $x(t)$ be the modulating signal with a peak amplitude of say A_m. We shall, for convenience, assume here that $x(t)$ has been so normalized that $|x(t)| \le 1$. Then, from the above definition and Eq. (4.1), the amplitude modulated signal may be expressed as

$$x_c(t) = A\cos(\omega_c t + \theta)$$

where
$$A = A_c + A_m x(t)$$

Here, A_c is the peak amplitude of the unmodulated carrier, A_m is the peak amplitude of the modulating signal and $x(t)$ is the normalized modulating signal, i.e., $|x(t)| \le 1$.

Let
$$m \triangleq \frac{A_m}{A_c}$$

Then,
$$A = A_c + A_m x(t) = A_c[1 + mx(t)]$$

Hence,

$$x_c(t) = A_c[1 + mx(t)]\cos(\omega_c t + \theta)$$

Without loss of generality, we may take $\theta = 0$ so that

$$\boxed{x_c(t) = A_c[1 + mx(t)]\cos\omega_c t} \tag{4.2}$$

where m is called the '*modulation index*' or the 'depth of modulation' and is defined as the ratio of peak amplitude of the modulating signal to the peak amplitude of the unmodulated carrier. It is a constant and is such that $0 \le m \le 1$. Instead of being expressed as a fraction, the depth of modulation may also be expressed as a percentage. Since $|x(t)| \le 1$, if $m > 1$, then $[1 + mx(t)]$ can become negative near the negative peaks of $x(t)$ and it results in a situation called '*over modulation*'. Over modulation is always to be avoided since, as we are going to see later, it leads to a distorted version of the message after the demodulation in the receiver. Hence the restriction that the modulation index 'm' should always be between 0 and 1.

In Eq. (4.2), the factor $A_c[1 + mx(t)]$ is the peak amplitude of the modulated carrier wave or, the amplitude of the envelope at the instant t. The change from the unmodulated peak value is $A_c mx(t)$ which is proportional to $x(t)$.

4.2.2 Single-Frequency Message Signal

For simplicity, let us assume for a moment that our message signal, $x(t)$ is a single frequency given by

$$x(t) = \cos\omega_m t; \qquad \omega_m = 2\pi f_m \tag{4.3}$$

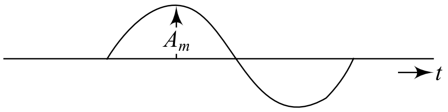
Then, from Eq. (4.2), we get the modulated signal as

$$\begin{aligned} x_c(t) &= A_c[1 + m\cos\omega_m t]\cos\omega_c t \\ &= A_c\cos\omega_c t + mA_c\cos\omega_c t \cdot \cos\omega_m t \end{aligned}$$

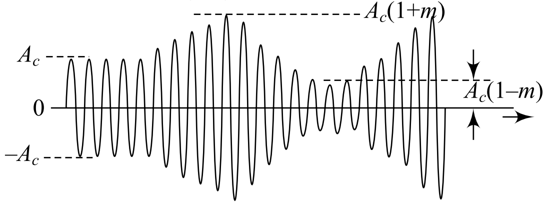
$$\boxed{x_c(t) = A_c\cos\omega_c t + \frac{1}{2}mA_c\cos(\omega_c + \omega_m)t + \frac{1}{2}mA_c\cos(\omega_c - \omega_m)t} \tag{4.4}$$

Thus, when the carrier signal of frequency f_c is amplitude modulated by a modulating signal of frequency f_m, the modulated signal has three frequency components – the carrier frequency component represented in Eq. (4.4) by the first term, i.e., $A_c\cos\omega_c t$, the upper side-frequency component having a frequency of $(f_c + f_m)$ and represented in Eq. (4.4) by the second term, i.e., $\frac{1}{2}mA_c\cos(\omega_c + \omega_m)t$, and the lower side-frequency component having a frequency of $(f_c - f_m)$ and represented by the third term, i.e., $\frac{1}{2}mA_c\cos(\omega_c - \omega_m)t$. They are called upper and lower side frequencies because they are on either side of the carrier frequency component and displaced from it by the same interval of frequency, i.e., f_m.

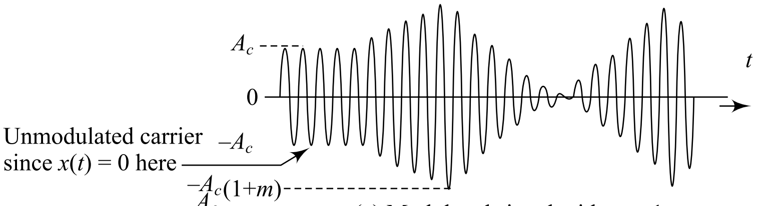
Equation (4.4) permits us to draw a phasor diagram for the AM signal when the modulating signal is a single tone. This phasor diagram is shown in Fig. 4.2(b).



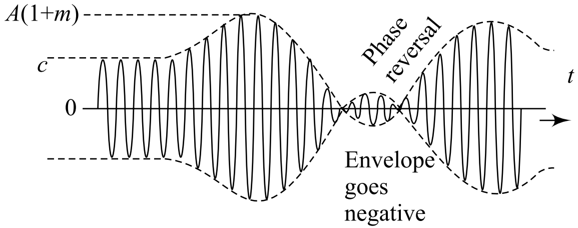
(a) Modulating signal $x(t) = \sin \omega_m t$



(b) Modulated signal with $m = 0.5$



(c) Modulated signal with $m = 1$



(d) Over-modulated signal with $m = 1.2$

Fig. 4.1 *Amplitude modulation with different values of m*

As the carrier component has a frequency of f_c, if we consider the phasor corresponding to this component as our reference, the upper side-frequency component having an amplitude of $\frac{mA_c}{2}$ and a frequency of $(f_c + f_m)$ will appear to be rotating at a frequency of f_m in the counter-clockwise direction, with respect to the carrier phasor. The lower side-frequency component having an amplitude of $\frac{1}{2}mA_c$ and a frequency of $(f_c - f_m)$ will appear to be rotating in the clockwise direction at a frequency of f_m, with respect to the carrier phasor.

From Eq. (4.4), we may also obtain the amount of power in the carrier component and in each of the side-frequency components. We have

Power in the carrier component = $\frac{1}{2}A_c^2$ = say P_c

Power in the upper side-frequency component = $\frac{1}{8}m^2A_c^2$

Power in the lower side-frequency component = $\frac{1}{8}m^2A_c^2$

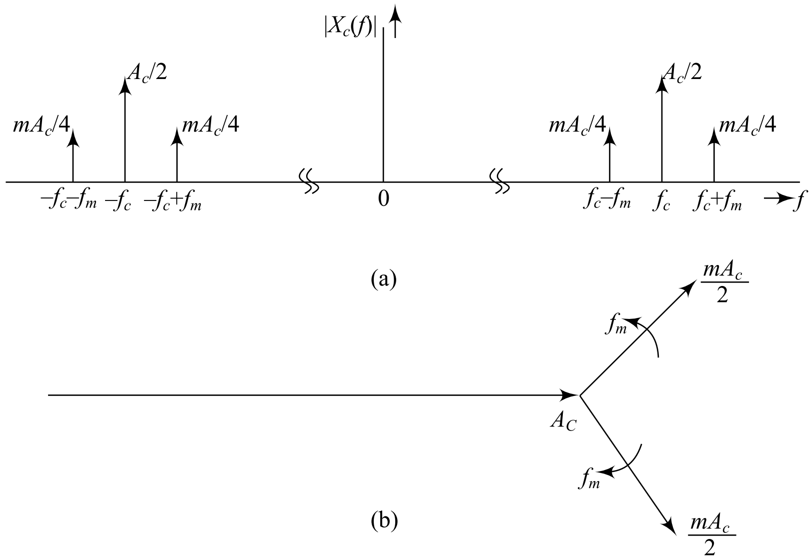


Fig. 4.2 *(a) Spectrum of an AM signal with single-tone modulation, (b) Phasor diagram of a single-tone modulated AM signal*

$$\therefore \qquad \boxed{\text{total power in the AM signal} = P_T = P_c\left[1+\frac{m^2}{2}\right]} \qquad (4.5)$$

Remark

Equation (4.5) tells us that for a single-tone amplitude modulated signal, even with a modulation index of m = 1, the maximum possible value, the carrier component constitutes two-thirds of the total power in the modulated signal. As the carrier component does not carry any information and as the message, or the information-bearing signal x(t), which in this case has been assumed to be $\cos\omega_m t$*, can be completely recovered from any one of the two side-frequency components, the carrier power in the AM signal is a waste. The carrier only helps in carrying the message, but is ultimately rejected in the receiver after the message is recovered. Hence, it is preferable to reduce or even eliminate the power in the carrier component of the modulated signal.*

If instead of a single tone, the message signal $x(t)$ consists of several frequency components, say f_{m1}, f_{m2}, f_{m3}, each one of these will produce a corresponding upper side-frequency component and a lower side-frequency component. Thus, in addition to the carrier component $A_c\cos\omega_c t$, there will be three upper side-frequency components:

$\frac{1}{2}m_1A_c\cos(f_c+f_{m1})$, $\frac{1}{2}m_2A_c\cos(f_c+f_{m2})$ and $\frac{1}{2}m_3A_c\cos(f_c+f_{m3})$ and three lower side-frequency components $\frac{1}{2}m_1A_c\cos(f_c-f_{m1})$, $\frac{1}{2}m_2A_c\cos(f_c-f_{m2})$ and $\frac{1}{2}m_3A_c\cos(f_c-f_{m3})$. Here, m_1, m_2 and m_3 represent the modulation indices for the three components and their values depend on the amplitudes of the frequency components with frequencies f_{m1}, f_{m2} and f_{m3} relative to the carrier amplitude A_c. The overall modulation index m of such an $x(t)$ is then given by (refer to Eq. (4.5))

$$m = \sqrt{m_1^2 + m_2^2 + m_3^2} \tag{4.6}$$

Example 4.1 A sinusoidal carrier signal of peak amplitude 5 V and frequency 100 kHz is amplitude modulated by a 5 kHz signal of peak amplitude 3 V. What is the modulation index? Draw the two-sided spectrum of the modulated signal.

Solution $x_c(t) = A_c[1 + m\cos\omega_m t]\cos\omega_c t = 5[1 + mx(t)]\cos\omega_c t$

$$= 5\cos\omega_c t + m5x(t)\cos\omega_c t$$

$\therefore$ $A_c = 5$ and $5m = 3$ since $|x(t)| \le 1$

$\therefore$ $m = 3/5 = 0.6$

The side frequencies are

$$(-f_c - f_m), (-f_c + f_m); (f_c - f_m) \text{ and } (f_c + f_m)$$

i.e., –105kHz, –95 kHz; 95 kHz and 105 kHz.

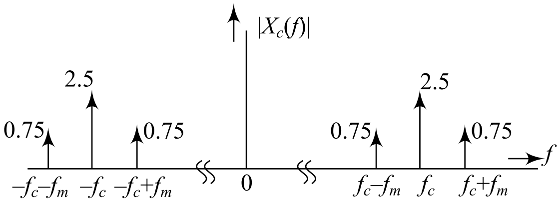


Fig. 4.3 *Spectrum of AM signal of Example 4.1*

The relationship between the message $x(t)$, carrier $c(t)$ and the AM signal $x_c(t)$, is diagrammatically illustrated in Fig. 4.4. For the purpose of this figure, it is assumed that $x(t) = \sin\omega_m t$. Time-domain as well as frequency-domain representations are given for all the signals.

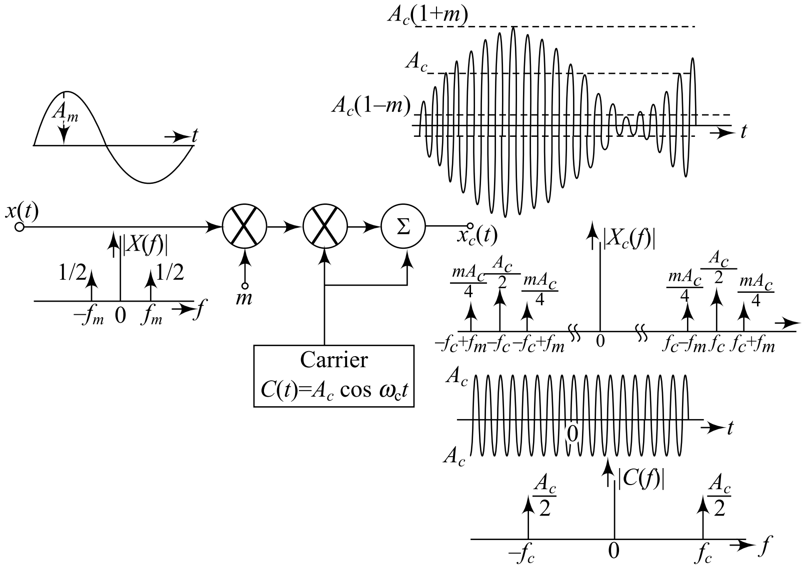


Fig. 4.4 *Amplitude modulation: Waveforms and spectra of message, carrier and modulated signal*

Example 4.2 A carrier wave of frequency 10 MHz and peak value of 10 V is amplitude modulated by a 5 kHz sine wave of amplitude 6 V. Determine the modulation index and draw the one-sided spectrum of the modulated wave. **(JNTU, May 2007)**

Solution Peak value of the modulating signal = A_m = 6 V
Peak value of the carrier signal = A_c = 10 V

Modulation index = $m = \dfrac{A_m}{A_c} = \dfrac{6}{10} = 0.6$

The AM signal may be represented as

$$x_c(t) = A_c[1 + m\cos\omega_m t]\cos\omega_c t$$

where A_c = 10v, m = 0.6, $\omega_m = 2\pi f_m = 2\pi \times 5 \times 10^3$ rad/sec and

$$\omega_c = 2\pi f_c = 2\pi \times 10 \times 10^6 \text{ rad/sec}$$

$x_c(t)$ may therefore be written in an expanded form as

$$\begin{aligned} x_c(t) &= A_c\cos\omega_c t + mA_c(\cos\omega_c t \cdot \cos\omega_m t) \\ &= A_c\cos\omega_c t + \frac{1}{2}mA_c\cos(\omega_c + \omega_m)t + \frac{1}{2}mA_c\cos(\omega_c - \omega_m)t \end{aligned}$$

Thus, $x_c(t)$ is made up of three frequency components – the carrier component having a frequency of 10 MHz and a peak amplitude A_c of 10 V, an upper side-frequency component having a frequency of (10 MHz + 5 kHz) and a peak amplitude of 0.3 × 10 = 3 V, and a lower side-frequency component with a frequency of (10 MHz–5 kHz) and an amplitude of 3 V. Thus, the one-sided spectrum is as shown in Fig. 4.5.

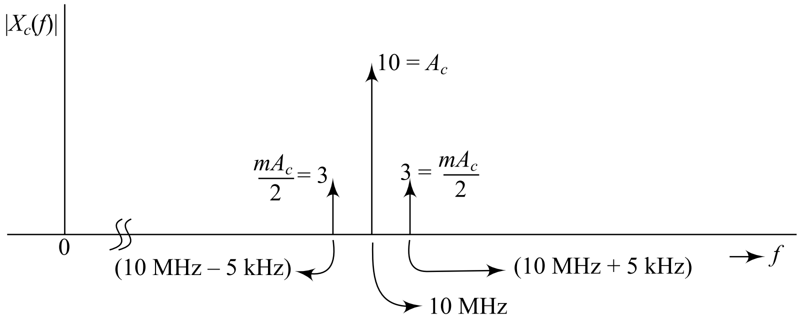


Fig. 4.5 *One-sided spectrum of AM signal of Example 4.2*

4.2.3 Frequency-Domain Description

As the next step, if we consider a modulating signal $x(t)$ which has its spectrum extending from 0 Hz to f_m Hz, then instead of side frequencies, we have to deal with sidebands – an upper sideband (USB) extending from f_c to $(f_c + f_m)$ Hz and a lower sideband (LSB) extending from $(f_c - f_m)$ Hz to f_c Hz. Let the message signal, $x(t)$, have an amplitude spectrum as shown in Fig. 4.6.

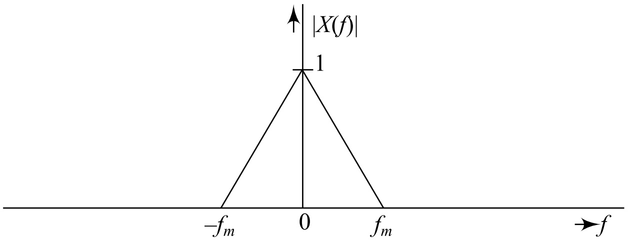


Fig. 4.6 *Amplitude spectrum of the message signal x(t)*

From Eq. (4.2), we have

$$\begin{aligned} x_c(t) &= A_c[1 + mx(t)]\cos\omega_c t \\ &= A_c\cos\omega_c t + mA_c x(t)\cos\omega_c t \end{aligned} \tag{4.7}$$

Since $\cos\omega_c t \xleftrightarrow{\text{F.T}} \frac{1}{2}[\delta(f+f_c)+\delta(f-f_c)]$,

Taking the FT of Eq. (4.7) on both sides

$$X_c(f)=\frac{A_c}{2}[\delta(f+f_c)+\delta(f-f_c)]+\frac{mA_c}{2}[X(f-f_c)+X(f+f_c)] \tag{4.8}$$

Here, we have made use of the FT pair

$$x(t)\cos\omega_c t \xleftrightarrow{\text{F.T}} \frac{1}{2}[X(f-f_c)+X(f+f_c)]$$

A plot of $X_c(f)$, the spectrum of the amplitude modulated signal $x_c(t)$ is as shown in Fig. 4.7 [Note that $X(f–f_c)$ is $X(f)$ shifted to the right by f_c and $X(f+f_c)$ is $X(f)$ shifted to the left by f_c].

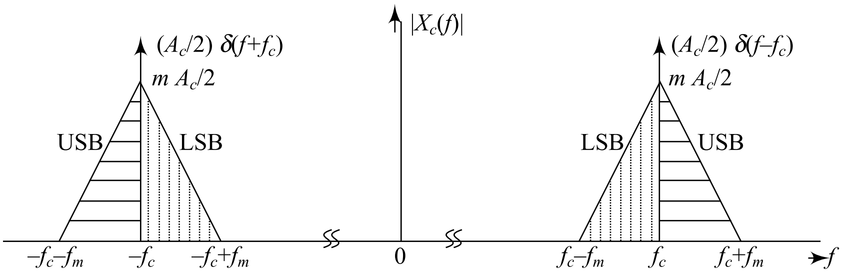


Fig. 4.7 *Amplitude spectrum of an AM signal*

Thus, if the maximum frequency component in the message, $x(t)$, is f_m, the amplitude modulated signal has a bandwidth of $2f_m$. Transmitters in audio broadcasting radio stations employ AM and they handle audio frequencies up to about 5 kHz. Thus, two such stations whose service areas have an overlap, must have a separation of at least 10 kHz in their carrier frequencies.

Thus,

$$\boxed{\text{Bandwidth of AM signal} = 2W} \tag{4.9}$$

where W Hz is the highest frequency component in $x(t)$, the message signal.

Example 4.3 A carrier, amplitude modulated to a depth of 50% by a sinusoid, produces side frequencies of 5.005 MHz and 4.995 MHz. The amplitude of each side frequency is 40 V. Find the frequency and amplitude of the carrier signal.

Solution Upper side frequency = $f_c+f_m=5005$ kHz

Lower side frequency = $f_c-f_m=4995$ kHz

Adding these two, $2f_c$ = 10,000 kHz $\quad\therefore f_c$ = 5000 kHz = 5 MHz

If carrier is $A_c \cos \omega_c t$, m is the modulation index and f_m is the modulating signal frequency, we can write the AM signal as

$$\begin{aligned} x_c(t) &= A_c[1+m\cos\omega_m t]\cos\omega_c t \\ &= A_c\cos\omega_c t+\frac{mA_c}{2}\cos(\omega_c+\omega_m)t+\frac{mA_c}{2}\cos(\omega_c-\omega_m)t \end{aligned}$$

$\therefore$ Side-frequency amplitude $= \dfrac{mA_c}{2} = 40, \quad \therefore A_c = 80/m = 80/0.5 = 160$ V

Hence, the carrier amplitude is 160 V and its frequency is 5 MHz.

Example 4.4 If all AM broadcasting stations handle audio frequencies of up to 5 kHz, how many AM broadcasting stations can be accommodated from 1 MHz to 1.5 MHz of the medium waveband?

Solution We know that the bandwidth occupied by an AM signal is equal to twice the highest audio frequency in its modulating signal.

$\therefore$ bandwidth required for each station = 2 × 5 kHz = 10 kHz

Bandwidth available = 1.5 MHz – 1.0 MHz = 500 kHz

$\therefore$ Number of stations that can be accommodated = 500/10 = 50.

4.2.4 Carrier and Sideband Components of Power in an AM Signal

From Eq. (4.2), the average power in an amplitude-modulated signal, $x_c(t)$, is given by

$$\left\langle x_c^2(t) \right\rangle = \frac{1}{2}\left[\left\langle A_c^2 \{1 + mx(t)\}^2 \right\rangle\right]$$

where the symbol <z> is used to represent the average value of z.

$$\therefore \qquad \left\langle x_c^2(t) \right\rangle = \frac{1}{2}\left[A_c^2 + 2m\left\langle x(t)\right\rangle A_c^2 + m^2 A_c^2 \left\langle x^2(t)\right\rangle\right]$$

Assuming $\left\langle x(t)\right\rangle = 0$, which is quite justifiable, since the dc component of the $x(t)$ is anyhow blocked by a capacitor in the detector stage of the receiver,

$$\left\langle x_c^2(t)\right\rangle = \text{ Average Power of } x_c(t) = \frac{1}{2}\left[A_c^2 + m^2 A_c^2\left\langle x^2(t)\right\rangle\right] \tag{4.10}$$

In the above equation, $\frac{1}{2}A_c^2$ represents the carrier component of power and

$$\frac{1}{2}m^2 A_c^2\left\langle x^2(t)\right\rangle = \text{Average total sideband power} \tag{4.11}$$

Since $x(t)$ is assumed to have been normalized so that $|x(t)| \le 1$, the maximum average power in $x(t)$, i.e., the maximum value of $\left\langle x^2(t)\right\rangle$, can be unity. The maximum possible sideband power is therefore obtained by putting $m = 1$ and $\left\langle x^2(t)\right\rangle = 1$. This works out to $\frac{1}{2}A_c^2$.

Thus, the average power of the AM signal under the above conditions of $m = 1$ and $\left\langle x^2(t)\right\rangle = 1$ is given by

$$\boxed{\left\langle x_c^2(t)\right\rangle = \frac{1}{2}A_c^2 + \frac{1}{2}A_c^2 \quad \text{for } m = 1 \text{ and } \left\langle x^2(t)\right\rangle = 1} \tag{4.12}$$

where the first term is the average power of the carrier component and the second term is the maximum possible value of the average total power of the two sidebands.

Thus, even when the sideband average power is maximized, the carrier power constitutes 50% of the total average power of an AM signal. If the modulating signal is a single tone, its average power $\left\langle x^2(t)\right\rangle$ is only ½ and in that case, the maximum value of the average power in the sidebands obtained by putting m = 1 in R.H.S. of Eq. (4.10) is

$$\boxed{\underset{\text{single-tone } x(t)}{\left\langle x_c^2(t)\right\rangle} = \frac{1}{2}A_c^2 + \frac{1}{4}A_c^2 \quad \text{for } m = 1 \text{ and } \left\langle x^2(t)\right\rangle = 1} \tag{4.12a}$$

Hence, in this case, as already shown earlier, the carrier component of power constitutes as much as 66.6% of the total average power of the AM signal.

The carrier component does not carry any information. Only the sidebands carry the message information. In fact, as mentioned earlier, the message can be recovered from just one sideband. From the foregoing, it is clear that amplitude modulation suffers from the following two disadvantages:

1. At least 50% of the transmitted power is the carrier power and it is a waste since carrier component does not carry any information. *So it is wasteful in power*.
2. While one sideband with a bandwidth of f_m is enough to recover the message, AM transmits the carrier plus both the sidebands, occupying a bandwidth of $2f_m$. *Thus, it is wasteful in bandwidth too*.

Example 4.5 When unmodulated carrier alone is transmitted, the antenna current is 9 A. When sinusoidal modulation is present, the antenna current is found to be 11 A. What is the percentage of modulation used?

Solution From Eq. (4.5), we have

$$P_T = P_c\left[1 + \frac{m^2}{2}\right]$$

We have $P_T = 11^2 \times r$ and $P_c = 9^2 \times r$, where r is the radiation resistance of the antenna.

$$\therefore \qquad \frac{P_T}{P_C} = \frac{11^2}{9^2} = 1 + \frac{m^2}{2}$$

$$\therefore \qquad m = \sqrt{2\left(\frac{121-81}{81}\right)} = \sqrt{\frac{80}{81}} = 0.994$$

Example 4.6 It is found that a radio transmitter is radiating a total power of 100 kW when the modulation index is 0.8. What is the carrier power being radiated by the transmitter? What is the sideband power?

Solution $P_T = 100 \times 10^3 = P_c\left[1 + \dfrac{0.64}{2}\right]$

$$\therefore \qquad P_c = \frac{100 \times 10^3}{1.32} = 75.8 \text{ kW}$$

∴ The carrier power being radiated = 75.8 kW

The total sideband power radiated = (100 – 75.8) kW = 24.2 kW.

Example 4.7 A certain transmitter (AM) is radiating 132 kW when a certain audio sine wave is modulating it to a depth of 80% and 150 kW when a second sinusoidal audio wave also modulates it simultaneously. What is the depth of modulation for the second audio wave?

Solution $P_{T_1} = P_c\left[1 + \dfrac{0.64}{2}\right] = 1.32P_c = 132 \text{ kW}$

$$\therefore \qquad P_c = 100 \text{ kW}$$

Let the modulation index of the second sinusoid be *m*

$$\therefore \qquad P_{T_2} = 150\times 10^3 = 100\times 10^3\left[1+0.32+\frac{m^2}{2}\right]$$

$$\therefore \qquad 50\times 10^3 = 100\times 10^3\left[0.32+\frac{m^2}{2}\right]$$

$$\therefore \qquad (50-32)\times 10^3 = m^2\times 50\times 10^3$$

$$\text{or} \qquad m^2 = \left(\frac{18}{50}\right) = 0.36 \qquad \therefore m = 0.6$$

Example 4.8 Determine the overall percentage of modulation in the above example when both the sinusoidal audio signals are simultaneously modulating the carrier.

Solution $m = \sqrt{m_1^2 + m_2^2} = \sqrt{0.64+0.36} = 1$

∴ overall percentage of modulation = 100%

Example 4.9 An AM transmitter of 1 kW power is fully modulated. Calculate the power transmitted, if it is transmitted as SSB. **(JNTU Sep., 2007)**

Solution When fully modulated, the total power of an AM signal is

$$P_T = P_c\left[1+\frac{m^2}{2}\right] = P_c\left[1+\frac{1}{2}\right] = \frac{3}{2}P_c = 1\text{ kW}$$

where P_c is the average power of the unmodulated carrier.

∴ carrier component of power in the AM signal = P_c = 2/3 kW.

Total sideband power in the AM signal (with 100% modulation) = $\left(1-\frac{2}{3}\right)\text{kW} = \frac{1}{3}\text{ kW}$

∴ power in each sideband = $\frac{1}{6}$ kW

It is this amount of power which will be transmitted if a single sideband is transmitted.

Example 4.10 An AM transmitter has an unmodulated carrier power of 10 kW. It can be modulated by a sinusoidal modulating voltage to a maximum depth of 40%, without overloading. If the maximum modulation index is reduced to 30%, what is the extent up to which the unmodulated carrier power can be increased without overloading?

Solution It is given that the unmodulated carrier power = P_c = 10 kW

Maximum depth of modulation without overloading = 40% $\quad \therefore m_1 = 0.4$

Total power in the AM signal = $P_T = P_C\left[1+\frac{m_1^2}{2}\right] = 10^4\times 1.08 = 10.8\text{ kW}$

∴ To avoid overloading, we have to see that the total power in the AM signal does not exceed 10.8 kW.

When the percentage of modulation is 30%, $m_2 = 0.3$

Now, let P_C^1 be the max. unmodulated carrier power that would make the total power in the AM signal to reach the value 10.8 kW.

$$\therefore \qquad P_T = 10.8 \times 10^3 = P_C^1\left[1 + \frac{(0.3)^2}{2}\right] = 1.045 P_C^1$$

$$\therefore \qquad P_C^1 = \frac{10.8 \times 10^3}{1.045} = 10.33 \text{ kW}$$

Hence, with a modulation index of 0.3, the unmodulated carrier power can be increased up to 10.33 kW without overloading.

Example 4.11 Calculate the percentage power saving when the carrier and one of the sidebands are suppressed in an AM wave modulated to a depth of (a) 100% (b) 50%. **(JNTU, May, 2007)**

Solution Since nothing has been mentioned about the modulating signal waveform, let us assume that it is sinusoidal. Then if P_T is the total power in the AM signal and P_c is the power in the carrier, we know that

$$P_T = P_c\left[1 + \frac{m^2}{2}\right] \text{ when the carrier and both the sidebands are transmitted.}$$

If the carrier and one sideband are suppressed, the total power is

$$P_T' = P_c \frac{m^2}{4}$$

$$\therefore \text{ \% saving in power} = \frac{100(P_T - P_T')}{P_T}$$

$$= \frac{100\left\{P_c\left[1 + \frac{m^2}{2}\right] - P_c\frac{m^2}{4}\right\}}{P_c\left[1 + \frac{m^2}{2}\right]} = \frac{100\left[1 + \frac{m^2}{4}\right]}{\left[1 + \frac{m^2}{2}\right]}$$

(a) When $m = 1$ corresponding to 100% modulation

$$\text{\% saving in power } = 100\left[\frac{1 + 1/4}{1 + 1/2}\right] = 83.3\%$$

(b) When $m = 0.5$ corresponding to 50% modulation

$$\text{\% saving in power } = \frac{(1 + 0.25/4)}{(1 + 0.25/2)} = 94.4\%$$

Example 4.12 Determine the maximum power efficiency of an AM modulator.

Solution Power efficiency of an AM modulator is given by

$$\eta = \frac{\text{Total power in the information bearing sidebands}}{\text{Total power in the modulated signal}}$$

We know that for AM, $P_T = P_c\left[1+\frac{m^2}{2}\right]$ for single-tone modulation

$$= P_c\left[1+m^2\overline{x^2}\right] \text{ for a general modulating signal } x(t)$$

$$\therefore \qquad \eta = \frac{P_c m^2\overline{x^2}}{P_c[1+m^2\overline{x^2}]} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}},$$

where $\overline{x^2}$ is the average power in the message signal and $0 \le m \le 1$.

$$\eta = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} = \frac{1}{1+\frac{1}{m^2\overline{x^2}}} \qquad (\text{Since } m^2\overline{x^2} = 0 \text{ is ruled out.})$$

$$\therefore \qquad \eta_{\max} = \frac{1}{1+\left(\frac{1}{m^2\overline{x^2}}\right)_{\min}}$$

Now $\frac{1}{m^2\overline{x^2}}$ takes minimum value when $m^2\overline{x^2}$ takes the maximum value.

$m_{\max} = 1$ and $\overline{x^2_{\max}} = 1$ since $|x(t)| \le 1$

$$\therefore \qquad \eta_{\max} = \frac{1}{1+1} = 0.5$$

4.2.5 Effect of Over Modulation

A diode detector, or an envelope detector (which is extensively used in all AM broadcast receivers) as we will be seeing later, tries to extract the envelope from an amplitude modulated wave. As the envelope follows the variations in the amplitude of the modulating signal, when the dc component is subtracted or removed from the envelope signal, ideally the modulating signal is obtained (see Fig. 4.1(a) and Eq. (4.2)). All this is true only when the envelope of the modulated signal truly follows the variation in amplitude of the modulating signal, i.e., as long as the modulation index is between 0 and 1. However, when the modulation index exceeds 1, i.e., under over-modulated conditions, as can be seen from Fig. 4.1(c), near the negative peak of the modulating signal, the envelope does not follow the variations of the amplitude of the modulating signal. Hence, under these conditions, the output of the envelope detector gives a distorted version of the modulating signal. *Therefore, over modulation should always be avoided.* It may also be noted from the Fig. 4.1(c) that when over modulation takes place, the recovered signal from the detector will be the $|e(t)|$ where $e(t)$ is the envelope of the modulated signal.

We have till now assumed, as we did while drawing Fig. 4.1, that the modulating signal is a single tone. In practice, it will never be a tone signal. When it is some complex waveform signal, the amplitude modulated signal will be as shown in Fig. 4.8.

In a case like this, we define two indices of modulation:

1. The positive peak modulation index $\triangleq \frac{A_{c_{\max}} - A_c}{A_c}$
2. The negative peak modulation index $\triangleq \frac{A_c - A_{c_{\min}}}{A_c}$

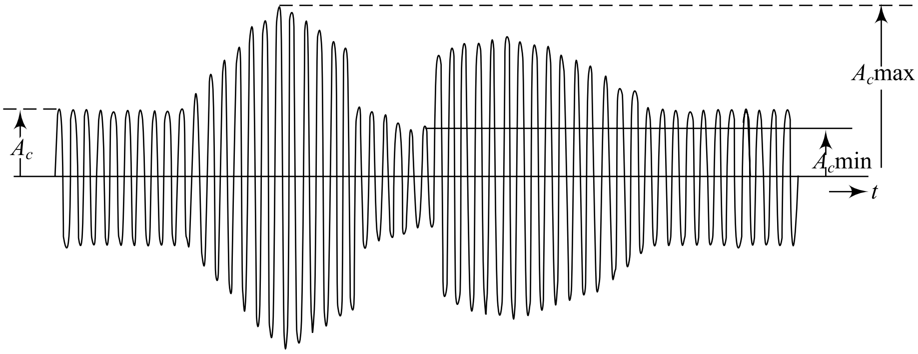


Fig. 4.8 *Modulated signal when the modulating signal is some complex waveform*

In the above, A_c represents the peak amplitude of the unmodulated carrier wave, $A_{c\,\max}$ represents the maximum value, and $A_{c\,\min}$ represents the minimum value of the peak amplitude of the carrier wave with modulation.

4.2.6 Measurement of Modulation Index

A straightforward method of measuring the percentage of modulation is to observe the modulated waveform on the screen of an oscilloscope by applying the amplitude modulated signal to the *Y*-deflection circuit of the scope. If it is sinusoidal modulation, a measurement of $A_{c\,\max}$ and $A_{c\,\min}$ (see Fig. 4.6) will give us the value of percentage of modulation as

$$\text{Modulation percentage} = \frac{A_{c\max} - A_{c\min}}{A_{c\max} + A_{c\min}} \times 100\%$$

However, there is an alternative method, known as the *trapezoid method* for determining the modulation index. It is a better method as it reveals distortions, if any, in the modulation process and is also applicable for complex modulating signals. The method involves connecting the modulated signal to the vertical deflection circuit and the modulating signal to the horizontal deflecting circuit. If care is taken to preserve their correct phases, we get a trapezoid displayed on the screen of the oscilloscope. Some of the possible shapes of the display are shown in Fig. 4.9.

Example 4.13 An AM (Double sideband plus full carrier) signal waveform is as shown in Fig. 4.10.

(a) Determine the modulation index *m*.

(b) Write down the expression for the modulated signal.

(c) Determine the total power, carrier power, and sideband power.

Solution

(a) $A_c[1+m] = 100 \qquad \therefore 2A_c = 160$

$A_c[1-m] = 60$

$\therefore \; A_m = 100 - A_c = 20 \qquad \therefore m = \frac{A_m}{A_c} = 0.25$

(b) $x_c(t) = A_c[1 + mx(t)]\cos\omega_c t$

$= 80[1 + 0.25\cos\omega_m t]\cos\omega_c t$

(c) $P_c = \frac{1}{2}A_c^2 = \frac{1}{2}\times 6400 = 3200 \text{ W}$

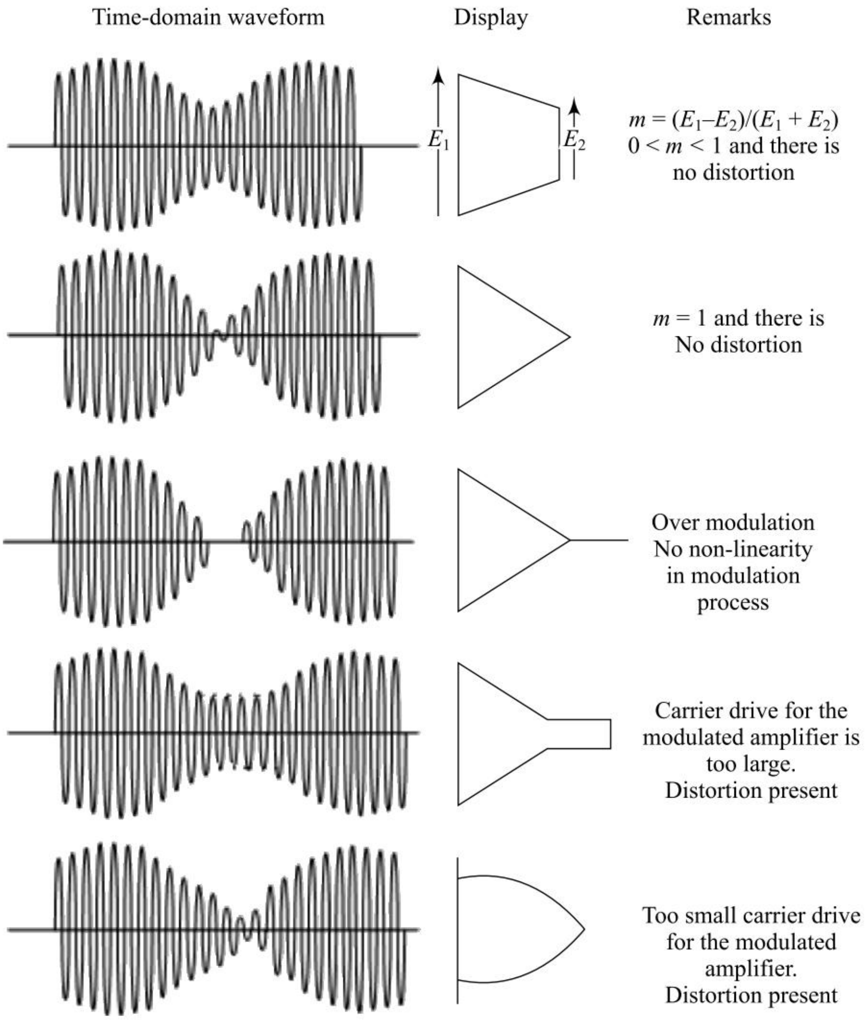


Fig. 4.9 *Trapezoidal patterns under different conditions*

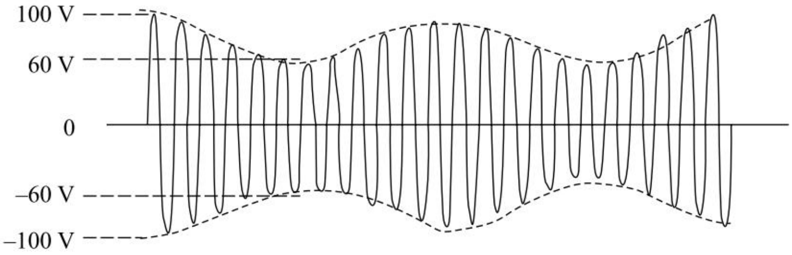


Fig. 4.10

$$\therefore\ P_T = 3200\left[1+\left(\frac{1}{4}\right)^2/2\right] = 3200+\frac{3200}{32} = 3300\text{ W}$$

$\therefore$ sideband power = 100 W

Example 4.14 A modulating signal consists of a symmetrical triangular wave having zero dc component and a peak-to-peak voltage of 12 V. It is used to amplitude modulate a carrier of peak voltage 10 V. Calculate the modulation index and the ratio of the side lengths (L_1/L_2) of the corresponding trapezoidal pattern.

(**JNTU, May, 2007**)

Solution A sketch of the modulated signal is shown in Fig. 4.11.

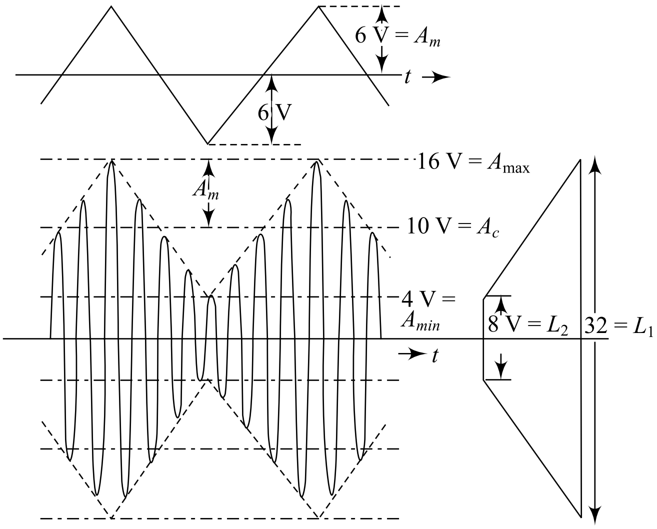


Fig. 4.11 *Modulated signal*

Modulation index

$$m = \frac{A_{\max}-A_{\min}}{A_{\max}+A_{\min}} = \frac{16-4}{16+4} = \frac{12}{20} = 0.6$$

$$L_1 = 2(A_c + A_m) = 16\times 2 = 32$$

$$L_2 = 2(A_c - A_m) = 2(10-6) = 8$$

$$\therefore\ \frac{L_1}{L_2} = \frac{32}{8} = 4$$

4.2.7 Generation of Amplitude Modulated (AM) Signals

There are a variety of methods available for generating amplitude modulated signals. However, amplitude modulators may be classified into the following types, depending on the technique used

1. Modulators using non-linear devices
2. Modulators using product devices
3. Modulators using switching devices

4.2.8 Modulators using Non-Linear Devices (Square Law Modulators)

Let a device have a non-linear relation between its input and output which can be represented by

$$e_{\text{out}} = a_0 + a_1 e_{\text{in}} + a_2 e_{\text{in}}^2 \tag{4.13}$$

where the constants a_0, a_1, a_2 depend on the shape of the input-output characteristic of the device. Suppose we make

$$e_{\text{in}}(t) = x(t) + E_c \cos\omega_c t \tag{4.14}$$

where $x(t)$ is the modulating signal with $|x(t)| \le 1$; and $E_c \cos\omega_c t$ is the carrier signal.

Substituting for $e_{\text{in}}(t)$ in Eq. (4.13) using Eq. (4.14), we get

$$e_{\text{out}}(t) = \left(a_0 + \frac{E_c^2}{2}\right) + a_1 x(t) + a_2 x^2(t) + \frac{E_c^2}{2}\cos 2\omega_c t + a_1 E_c\left[1+\left(\frac{2a_2}{a_1}\right)x(t)\right]\ \cos\omega_c t$$

In the above, the first term is a dc term which can always be suppressed by using a coupling capacitor. The second term $a_1x(t)+a_2x^2(t)$ is a low frequency term having frequency components near those of the modulating signal. The third term is a very high frequency term which is at twice the carrier frequency. The last term $a_1E_c\left[1+\left(\frac{2a_2}{a_1}\right)x(t)\right]\cos\omega_c t$ is the amplitude modulated signal (see Eq. (4.2)) and so is the useful term. To separate out this and reject the second and third terms, we need to simply use a band pass filter centered on f_c and having a bandwidth equal to *twice* that of the modulating signal $x(t)$.

A modulator of this type may easily be realized by making use of the non-linear relation between the gate voltage and the drain current of an FET as shown in Fig. 4.12.

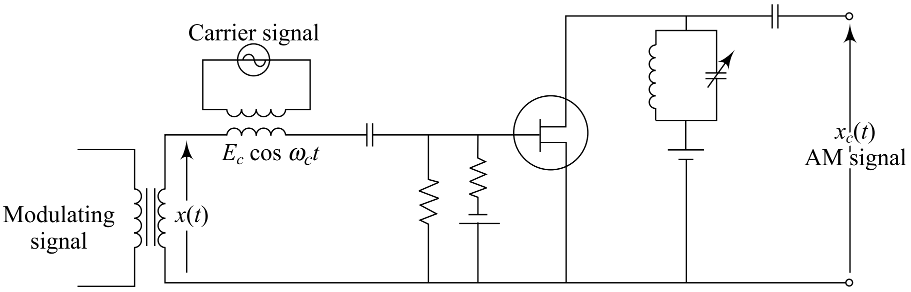


Fig. 4.12 *A square-law amplitude modulator*

In the above modulator, the tank circuit connected between the drain and source is tuned to the carrier frequency f_c and it is ensured that it has a reasonably low Q to give a bandwidth that is twice the modulating signal bandwidth. At the same time, the Q will be large enough to satisfactorily reject the modulating signal component as well as the components having frequencies that are multiples of the carrier frequency. The method of separation of the useful last term of e_{out} from the rest can perhaps be better understood by going into the frequency domain. For this, let us take the Fourier transform of $e_{\text{out}}(t)$.

$$E_{\text{out}}(f)=\left(a_0+\frac{E_c^2}{2}\right)\delta(f)+a_1X(f)+a_2[X(f)*X(f)]+\frac{E_c^2}{4}[\delta(f+2f_c)+\delta(f-2f_c)]$$

$$+\frac{a_1E_c}{4}[\delta(f+f_c)+\delta(f-f_c)]+a_2E_c[X(f+f_c)+X(f-f_c)] \tag{4.15}$$

A sketch of $E_{\text{out}}(f)$ is shown in Fig. 4.13.

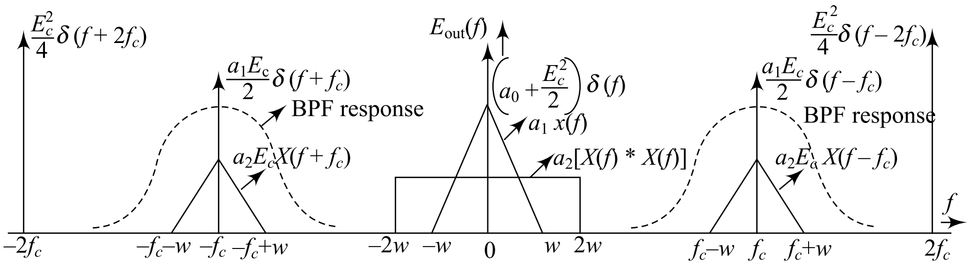


Fig. 4.13 *Spectrum of $e_{\text{out}}(t)$ of the square law modulator*

4.2.9 Modulators using Product Devices

These are based on Eq. (4.2) which states that an amplitude modulated signal is given by

$$x_c(t) = A_c[1 + mx(t)]\cos\omega_c t$$
$$= A_c\cos\omega_c t + mA_c x(t)\cos\omega_c t$$

The amplitude modulated signal $x_c(t)$, can therefore be obtained from an arrangement as shown in Fig. 4.14. $mx(t)$ and $A_c\cos\omega_c t$ are multiplied in the analog signal multiplier and then $A_c\cos\omega_c t$ is added to it to obtain $x_c(t)$. The analog signal multiplier, or, the product device used here can easily be realized using what is generally referred to as the 'variable transconductance multiplier', which is a differential amplifier in which the gain, which depends upon the transconductance of the transistor, is varied in accordance with one of the signals to be multiplied, by allowing it to control the total emitter current of the differential amplifier. Thus, when the other signal to be multiplied is applied to the differential amplifier input, its differential output will be proportional to the product of the two signals. The other part of Fig. 4.14 may, of course, be realized using an op-amp.

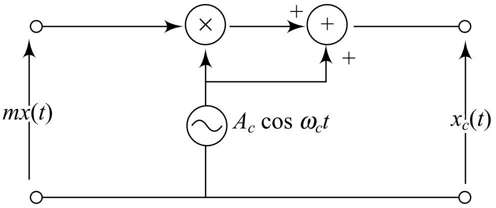


Fig. 4.14 *Block diagram of a product modulator*

4.2.10 Modulators using Switching Devices

These modulators make use of a switch, which may be a diode or a transistor. This switch allows current to flow through the load (a tank circuit tuned to the carrier frequency) in the form of truncated sinusoidal pulses occurring at regular intervals of $(1/f_c)$, where f_c is the carrier frequency. If these current pulses are made to vary with the amplitude of the modulating signal, it is possible to get an amplitude modulated wave across the load.

Switching modulator using a diode

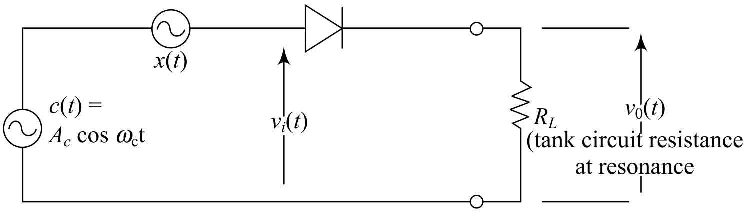


Fig. 4.15 *A switching modulator using a diode*

If we assume that

1. the forward resistance of the diode is extremely small compared to R_L, and
2. $|x(t)| \le 1$ and $A_c >> 1$

then we may state that

$$v_0(t) = \begin{cases} v_i(t) \text{ whenever } A_c\cos\omega_c t > 0 \\ 0 \quad \text{otherwise } (A_c\cos\omega_c t < 0) \end{cases} \tag{4.16}$$

Since

$$v_i(t) = x(t) + A_c\cos\omega_c t \tag{4.17}$$

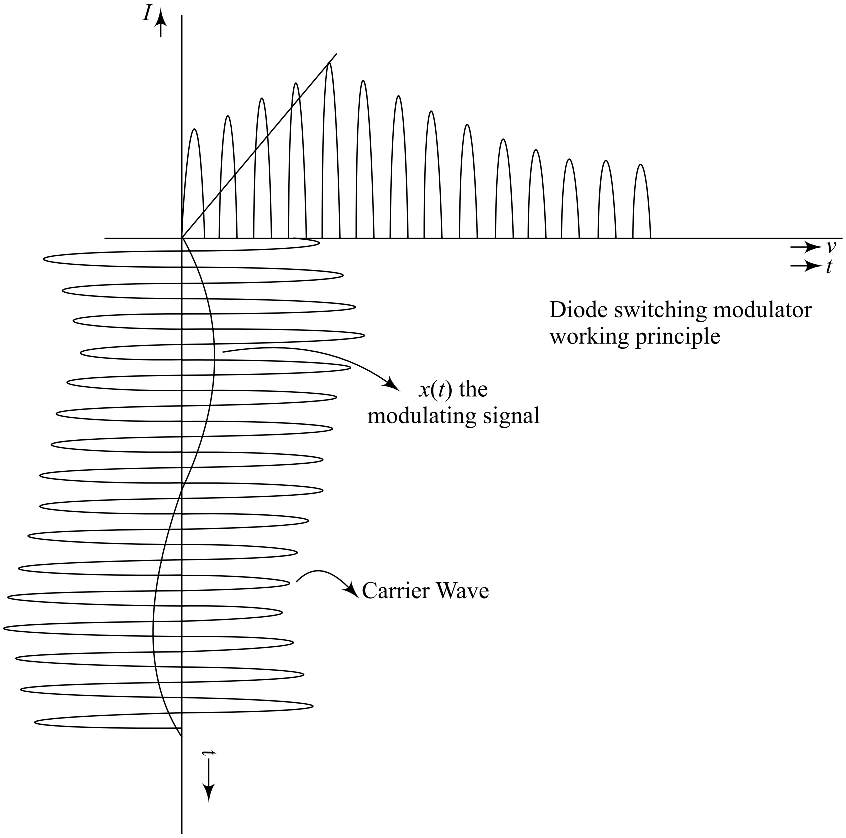


Fig. 4.16 *Working principle of a diode switching modulator*

it means that

$$v_0(t) = v_i(t)g(t) = [x(t) + A_c \cos \omega_c t]g(t) \tag{4.18}$$

where $g(t)$ is a gate waveform with a period $T_0 = (1/f_c)$ as shown in Fig. 4.17.

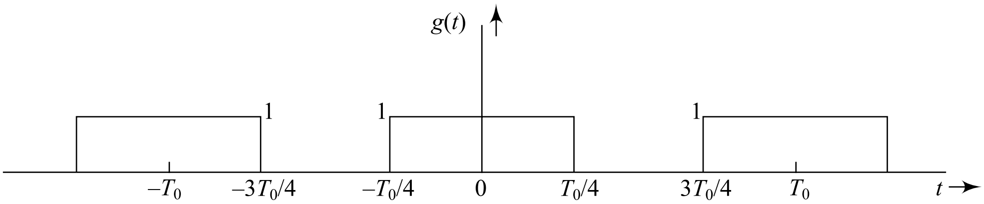


Fig. 4.17 *The periodic gate waveform g(t)*

The periodic gate waveform of Fig. 4.17 may be expanded using trigonometric Fourier series.

Let

$$g(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_c t + \sum_{n=1}^{\infty} b_n \sin n\omega_c t$$

Then we know that $a_0 = ½$ since $g(t)$ has an amplitude of 1 with a duty cycle of 0.5. Further, $b_n = 0$ for all n because of the even symmetry of $g(t)$. Also,

$$a_n = \frac{2}{T_0}\int_{-T_0/2}^{+T_0/2} x(t)\cos n\omega_c t dt = \frac{4}{T_0}\int_{0}^{T_0/2} x(t)\cos n\omega_c t dt$$

$$= \frac{4}{T_0}\int_{0}^{T_0/4} \cos n\omega_c t dt = \frac{4}{T_0}\left(\frac{1}{n\omega_c}\right)\sin n\omega_c t\Big|_0^{T_0/4}$$

$$= \frac{2}{n\pi}\left[\sin\left(\frac{n\pi}{2}\right)\right] = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{2}{n\pi}(-1)^{\frac{n-1}{2}} & \text{if } n \text{ is odd} \end{cases}$$

$$\therefore \qquad g(t) = \frac{1}{2} + \sum_{n=1}^{\infty}\left[\frac{2}{n\pi}\sin\left(\frac{n\pi}{2}\right)\right]\cos n\omega_c t$$

$$\therefore \qquad g(t) = \frac{1}{2} + \frac{2}{\pi}\cos\omega_c t - \frac{2}{3\pi}\cos 3\omega_c t + \frac{2}{5\pi}\cos 5\omega_c t - \frac{2}{7\pi}\cos 7\omega_c t + \dots \tag{4.19}$$

Substituting in Eq. (4.18) for $g(t)$ using Eq. (4.19) and for $v_i(t)$ using Eq. (4.17), and rejecting the constant terms and terms involving only the modulating signal frequencies as well as $2f_c$ and above (since the tank circuit constituting the load is tuned to f_c and has a bandwidth of $2W$, where W is the band limiting frequency of the modulating signal), we get

$$v_0(t) = \frac{A_c}{2}\left[1 + \frac{4}{\pi A_c}x(t)\right]\cos\omega_c t \tag{4.20}$$

From its form, we can easily recognize that $v_0(t)$ is an amplitude modulated signal, the carrier component being $A_c/2$, the carrier frequency being f_c and the modulation index being

$$m = \frac{4}{A_c\pi} \tag{4.21}$$

Equation (4.21) implies that the peak amplitude of the carrier, viz., A_c, must be small in order to have a value of m close to unity. However, A_c must be quite large compared to 1 as otherwise the assumptions made by us for this analysis will be violated.

Transistor switching modulator or collector-modulated class-C amplifier A transistor switching modulator, or a collector modulated class-C amplifier is shown in Fig. 4.18.

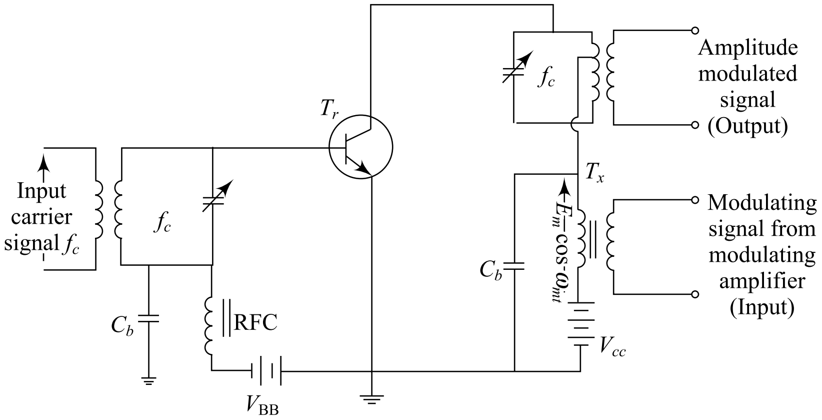


Fig. 4.18 *A collector-modulated class-C amplifier*

The base bias supply V_{BB} reverse biases the base-emitter junction beyond cut off and the transistor T_r works under class-C conditions. The input carrier signal level is so adjusted that the conduction angle for the collector current is approximately 120° which gives good power efficiency for the class-C amplifier while allowing reasonable output power.

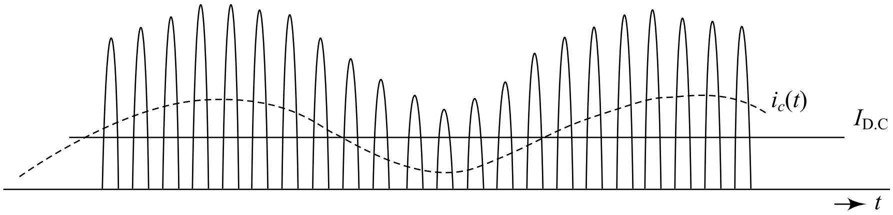


Fig. 4.19 *Collector current pulses in a class-C collector modulated amplifier*

There will be one current pulse for each RF cycle. These current pulses excite the tank circuit on the collector side, which is tuned to a frequency of f_c. Thus, across the tank circuit, we get a sinusoidal RF voltage at carrier frequency, the peak amplitude of the sinusoid varying in accordance with the modulating signal. If the carrier drive is adjusted to be sufficiently large, collector current pulses exist even at the trough of the modulating signal voltage. The average value of the current over a modulating signal cycle is marked as I_{DC} and it is the direct current drawn from the collector supply voltage V_{CC} The average value of these current pulses over each RF cycle (i.e., carrier cycle) will, however, be varying from one RF cycle to the next. This component of current is marked in Fig. 4.19 as $i_c(t)$.

Let the final stage of the modulating signal amplifier produce a message signal

$$e_m = E_m \cos \omega_m t \tag{4.22}$$

in the collector circuit through the modulating transformer, T_X, as shown in Fig. 4.14.

If

$$m \triangleq \frac{E_m}{V_{\text{CC}}} = \text{modulation index} \tag{4.23}$$

$$i_c(t) = I_{\text{DC}}(1 + m \cos_m t) \tag{4.24}$$

$P_T(t) \triangleq$ Total power input into the collector circuit (averaged over an RF cycle)

$$= V_{\text{CC}}[1 + m \cos \omega_m t]\, i_c(t) \tag{4.25}$$

But

$$i_c(t) = I_{\text{DC}}(1 + m \cos_m t) \qquad \text{(From Eq. (4.24))}$$

Thus,

$$P_T(t) = V_{\text{CC}}[1 + m \cos \omega_m t]\, I_{\text{DC}}[1 + m \cos \omega_m t] \tag{4.26}$$

We define

$$P_B \triangleq V_{\text{CC}} \cdot I_{\text{DC}} \tag{4.27}$$

Then P_B represents the dc power supplied by the V_{CC} supply to the collector circuit.

$$\therefore \quad P_T(t) = P_B[1 + 2m \cos \omega_m t + m^2 \cos^2 \omega_m t]$$

$$\therefore \quad P_T(t) = P_B\left[1 + \frac{1}{2}m^2 \cos 2\omega_m t + 2m \cos \omega_m t + \frac{1}{2}m^2\right] \tag{4.28}$$

When we average $P_T(t)$ *over a modulating signal cycle*, the second and third terms on the RHS of Eq. (4.28) vanish.

$$\therefore \quad P_{T\text{av}} = P_B\left[1 + \frac{m^2}{2}\right] \tag{4.29}$$

P_{Tav} represents the total average power supplied to the collector circuit. Of this, $P_B = V_{CC} \cdot I_{DC}$ represents the power supplied by the V_{CC} supply. The remaining part, viz. $P_B m^2/2$ is supplied by the final stage of the modulating amplifier.

If η denotes the collector circuit efficiency (η is generally about 80 to 90%, i.e., 0.8 to 0.9), then ηP_{Tav} = Total average power in the amplitude modulated output signal = P_0

$$P_0 = \eta P_{Tav} = \eta P_B + \eta P_B \frac{m^2}{2} \tag{4.30}$$

ηP_B = carrier component of $P_0 = P_c$

$\eta P_B \frac{m^2}{2} = P_C \frac{m^2}{2}$ = Total sideband power component of P_0.

From the foregoing, it is clear that the carrier component of the output AM signal is generated from the power drawn from the V_{CC} supply and the total sideband power of the output AM signal is derived from the power supplied by the modulating signal, i.e., from the final stage of the modulating signal amplifier.

Example 4.15 Referring to Fig. 4.8, if $A_{c\ \max} = 75$ $A_{c\ \min} = 15$, determine the following assuming sinusoidal modulating signal (a) m, (b) carrier power and total sideband power, and (c) amplitude and phase of the additional carrier to be added in order to have m = (i) 50% , (ii) 90%.

Solution

(a) $m = \frac{A_{c\max} - A_{c\min}}{A_{c\max} + A_{c\min}} = \frac{75-15}{75+15} = \frac{60}{90} = 66.7\%$

(b) $P_T = P_c\left[1 + \frac{m^2}{2}\right]$. Here $P_c = \frac{A_c^2}{2}$ but $A_c = 15 + \frac{75-15}{2} = 45$

$\therefore \quad P_c = \frac{45^2}{2} = \frac{2025}{2} = 1012.5$ W

$\therefore$ Sideband power (total) $= P_c \cdot \frac{m^2}{2} = \frac{45^2}{2} \times \frac{4}{9} \times \frac{1}{2} = 225$ W

(c) $x_c(t) = 45\left[1 + \frac{2}{3}\cos\omega_m t\right]\cos\omega_c t + A\cos\omega_c t$

$$= (45 + A)\left[1 + \left(\frac{30}{(45+A)}\right)\cos\omega_m t\right]\cos\omega_c t$$

(i) $\frac{30}{45+A} = 0.5 \qquad \therefore 60 = 45 + A \qquad \therefore A = 15$

$\therefore$ carrier to be added = $15\angle 0°$

(ii) $\frac{30}{45+A} = 0.9 \qquad \therefore A = -11.67$

$\therefore$ carrier to be added = $11.67\angle 180°$

Example 4.16 A transistor class-A amplifier working with an efficiency of 20% is collector-modulating a transistor class-C power amplifier working with a collector-circuit efficiency of 60%. The class-C power amplifier transistor is dissipating 24 W when the modulation depth is 80%. (a) What is the carrier power in the output modulated wave? (b) What will be the class-C power amplifier collector dissipation for 100% modulation? (c) What should be the modulating amplifier transistor rating in watts for this depth of modulation? (d) What is the overall efficiency of the circuit (including class-C and class-A power amplifiers)?

Solution Let dissipation in the transistor (class-C power amplifier) be P_d

At $m = 0.6$: $$P_d = P_{\text{in}} - P_0 = P_{\text{in}}[1 - \eta] = 0.4P_{\text{in}} = 24$$

$\therefore$ $$P_{\text{in}} = \frac{24}{0.4} = 60 \text{ W}$$

But $$P_{\text{in}} = P_c\left[1 + \frac{0.6^2}{2}\right] = 1.18P_c = 60$$

$\therefore$ $$P_c = \frac{60}{1.18} = 50.85 \text{ W}$$

This represents the power required to be supplied to the class-C amplifier in order to produce the carrier component in the output amplitude modulated signal.

$\therefore$ carrier component of output AM signal = 50.85 × 0.6 = 30.5 W.

$\therefore$ carrier power in the output modulated wave is 30.5 W.

(a) $m = 1$

$P_0 = 30.5(1 + 0.5) = 45.75$ W = Total output power with $m = 1$.

$\therefore$ the corresponding input power = $P_{in} = 45.75/0.6 = 76.25$ W

$\therefore$ $P_d = 76.25 - 45.75 = 30.5$ W

($m = 1$)

(b) $m = 1$ The AF output to give $m = 1$ is supplied by the power required to generate the output sideband power with $m = 1$. This is given by

$$P_{SB} = P_c \times 0.5 = 50.85 \times 0.5 = 25.425 \text{ W}$$

(c) The class-A power amplifier transistor undergoes maximum dissipation when it is delivering zero output power. Under this condition, the dissipation equals the input power to the class-A amplifier.

$\therefore$ $$P_{d\max} = P_{\text{in}} = \frac{25.425}{0.2} = 127.125 \text{ W}$$

(d) Overall efficiency at $m = 0.6$

$$\eta_{m=0.6} = \frac{\text{Total output power with } m = 0.6}{\text{Total input power (for the class-C and class-A amplifiers) with } m = 0.6}$$

$$= \frac{36}{50.85 + 127.125} = \frac{36}{180} = 0.2$$

$\therefore$ the overall efficiency = 20%

Example 4.17 A carrier signal $A_c \cos\omega_c t$ and a modulating signal $x(t) = \cos\omega_m t$ are applied in series to a diode switching modulator. What should be the carrier amplitude, A_c, if the AM signal at the output is to have a modulation index of 85%? Assume that the diode acts as an ideal switch.

Solution From Eq. (4.21), we have

$$m = \frac{4}{A_c\pi}$$

$$A_c = \frac{4}{m\pi} = \frac{4}{0.85\pi} = 1.498 \text{ V}$$

Example 4.18 A collector-modulated class-C power amplifier is giving an amplitude modulated signal of average power 100 W at the output, while operating with a collector-circuit efficiency of 80%. Assuming the modulation index to be 0.8, find (a) the power to be supplied by the modulating amplifier, and (b) the dissipation in the transistor.

Solution $P_0 = P_c\left(1+\frac{m^2}{2}\right) = 100\text{ W} \qquad \therefore P_c = \frac{100}{1.32} = 75.75\text{ W}$

$\therefore$ The output sideband power = 100 – 75.75 = 24.25 W.

(a) Since the power supplied by the modulating amplifier gets converted into the output sideband power and since the efficiency of the class-C modulated amplifier is 80%, we have

$$\left.\begin{array}{l}\text{Power to be supplied by the}\\ \text{modulating amplifier}\end{array}\right\} = 24.25\times\frac{1}{0.8} = 30.3\text{ W}$$

(b) Let the dissipation in the transistor be P_D with 80% modulation

$$P_D = P_{\text{in}} - P_0 = P_0\left(\frac{1}{\eta}\right) - P_0 = P_0\left(\frac{1-\eta}{\eta}\right) = 100\times\frac{0.2}{0.8} = 25\text{ W}$$

4.3 DEMODULATION OF AM SIGNALS

In order to send the message signal across to the destination, the transmitter modulates a carrier signal with the message signal and transmits the modulated signal through the channel. At the receiving end, the message signal is recovered from the modulated signal through a process called ‘demodulation’ or ‘detection’, and the carrier signal, which, as we know, does not carry any information, is rejected. *Thus, demodulation is the process of recovering the message signal from a modulated signal.*

There are several techniques available in principle, for demodulation of amplitude modulated signals. These are:

1. Coherent/synchronous detection
2. Square law detection
3. Envelope detection

Of these three, the simplest and by far the most widely used one is the ‘Envelope Detector’. Hence, after discussing the principle of the first two, we shall discuss the third one in detail.

4.3.1 Coherent/Synchronous Detection

The modulated signal which is received is given by

$$x_c(t) = A_c[1+mx(t)]\cos\omega_c t \qquad \text{(Refer to Eq. (4.2))}$$

As shown in Fig. 4.20, coherent/synchronous detection consists of:

1. Generating the carrier signal, correct in frequency and phase, at the receiver.
2. Multiplying $x_c(t)$, the received signal, by this locally generated carrier signal.
3. Low pass filtering the above product of the two signals.

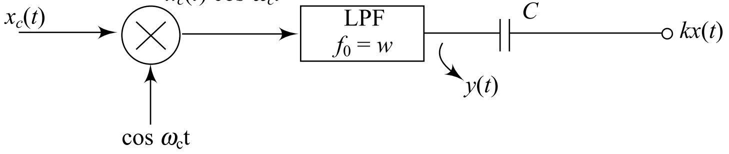


Fig. 4.20 *Block diagram of a coherent detector for AM signals*

$$\therefore \quad x_c(t)\cos\omega_c t = A_c\cos^2\omega_c t + mA_c x(t)\cos^2\omega_c t$$

$$= \frac{A_c}{2} + \frac{A_c}{2}\cos 2\omega_c t + \frac{mA_c}{2}x(t)[1+\cos 2\omega_c t]$$

If the highest frequency component present in $x(t)$ is W Hz, let the cut-off frequency of the low pass filter be W Hz. Then at the output of the low pass filter (LPF), we will have

$$y(t) = \frac{A_c}{2} + \frac{mA_c}{2}x(t)$$

The dc component represented by $\frac{A_c}{2}$ is blocked by the coupling capacitor, C, and at the output, we get $m\frac{A_c}{2}x(t)$, which is a scaled version of the message signal.

It is not an easy thing to generate in the receiver, a carrier signal of the correct frequency and which is in phase with the carrier of the received signal. We will be discussing in more detail about this problem when we deal with detection of double sideband suppressed carrier (DSB-SC) signal. It should suffice to state here that synchronous detection, though theoretically possible, is never used in practice for the detection of AM waves because of the above problem, and the availability of simple diode detectors (envelope detectors).

4.3.2 Square Law Detection

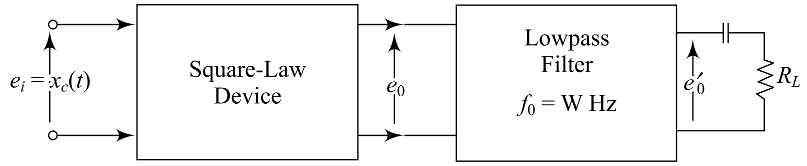


Fig. 4.21 *Square law detector*

Let the square law device/circuit have an input-output relation given by

$$e_0 = a_0 + a_1 e_i + a_2 e_i^2 \tag{4.31}$$

where e_0 is the output signal and e_i is the input signal.

But $$e_i = x_c(t) = A_c[1+mx(t)]\cos\omega_c t$$

$\therefore$ substituting this in Eq. (4.31), we have

$$e_0 = a_0 + a_1 A_c[1+mx(t)]\cos\omega_c t + a_2 A_c^2[1+mx(t)]^2\cos^2\omega_c t$$

$$= \left(a_0 + \frac{a_2 A_c^2}{2}\right) + [a_1 A_c + ma_1 A_c x(t)]\cos\omega_c t + \frac{a_2}{2}A_c^2\cos 2\omega_c t$$

$$+ a_2 A_c^2 mx(t) + a_2 A_c^2 mx(t)\cos 2\omega_c t + \frac{a_2}{2}m^2 A_c^2 x^2(t) + \frac{a_2}{2}m^2 A_c^2 x^2(t)\cos 2\omega_c t$$

Since the low pass filter has a cut-off frequency $f_0 = W$ Hz which is very small compared to the carrier frequency f_c, the output of the low pass filter will be

$$e'_0 = \left(a_0 + \frac{a_2}{2}A_c^2\right) + a_2 A_c^2 mx(t) + \frac{a_2}{2}m^2 A_c^2 y(t)$$

where $y(t)$ is the signal consisting of all frequency components of $x^2(t)$ which have frequencies less than or equal to W Hz, the cut-off frequency of the low pass filter. The first-term $\left(a_0 + \frac{a_2}{2}A_c^2\right)$ representing the dc component may be blocked by using a coupling capacitor. The next term $a_2 A_c^2 mx(t)$ is the desired signal and passes through the LPF. However, since $y(t)$ and $x(t)$ have overlapping spectra, the final output across the load will not be the message signal alone; there will be distortion due to the last term. To keep this distortion low compared to the desired signal term, viz. the second term, one has to ensure that $|mx(t)|$ is reasonable small compared to 1 so that the last term becomes negligible compared to the second.

Example 4.19 A signal $v(t) = [1 + m(t)]\cos\omega_c t$ is detected using a square law detector whose input-output relationship is $v_0 = v_{\text{in}}^2$. If the Fourier transform of the signal $m(t)$ is constant at the value M_0 from $-f_m$ to $+f_m$, sketch the Fourier transform of the output of the square law detector in the frequency range $-f_m < f < f_m$. **(GATE Exam 1998)**

Solution The square law device of the square law detector has an input–output relationship $v_0 = v_{\text{in}}^2$

∴ when $v(t)$ is given as input to this square law device,

$$v_0(t) = v^2(t) = [1 + m(t)]^2 \cos^2\omega_c t$$

$$= \frac{1}{2} + m(t) + \frac{1}{2}m^2(t) + \frac{1}{2}\cos 2\omega_c t + m(t)\cos 2\omega_c t + \frac{1}{2}m^2(t)\cos 2\omega_c t$$

In a square law detector, the square law device will be followed by an LPF whose cut-off frequency is the highest frequency available in the modulating signal $m(t)$. Since the signal $m(t)$ has its spectrum extending from $-f_m$ up to $+f_m$, the highest modulating signal frequency and hence the cut-off frequency of the LPF in the detector, is f_m Hz.

∴ when $v_0(t)$ is low pass filtered with this LPF, its output is

$$v_D(t) = \frac{1}{2} + m(t) + \frac{1}{2}\overline{m^2(t)}$$

Note *All the other components are rejected by the LPF.*

where $\overline{m^2(t)}$ represents that part of $m^2(t)$ made up of frequency components from $-f_m$ to $+f_m$. This is because $m^2(t)$ will have components having frequencies from $-2f_m$ to $+2f_m$ as is going to be evident from what follows.

Consider $F\left[\frac{1}{2} + m(t) + m^2(t)\right]$. This is

$$\frac{1}{2}\delta(f) + M(f) + \frac{1}{2}F[m(t)\cdot m(t)]$$

$$= \frac{1}{2}\delta(f) + M(f) + \frac{1}{2}[M(f) * M(f)]$$

Before sketching $v_D(f)$, let us see the shape of $[M(f)*M(f)]$

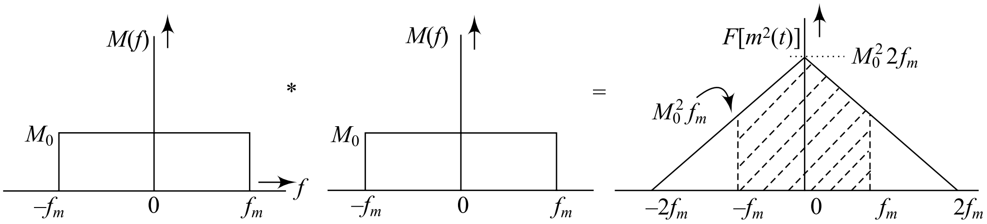


Fig. 4.22 *Signals for Example 4.19*

Because the LPF has a cut-off frequency of f_m, only that part of the spectrum of $m^2(t)$ which lies between $-f_m$ and $+f_m$ will have to be considered (it is shown shaded in the spectrum of $m^2(t)$).

$$\therefore\ V_D(f) = \frac{1}{2}\delta(f) + M(f) + \text{ That part of } \mathcal{F}[m^2(t)] \text{ which is from } -f_m \text{ to } f_m$$

When we sketch this, we get the spectrum as shown in Fig. 4.23.

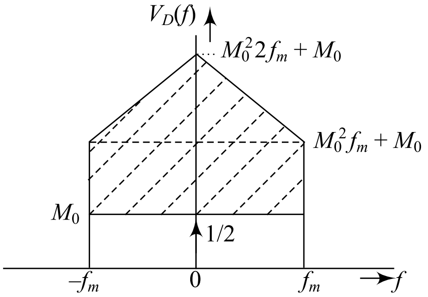


Fig. 4.23 *Fourier transform of the output*

4.3.3 Envelope Detector

We know that the envelope of an amplitude-modulated signal follows the variations in amplitude of the message, or the modulating signal, if the modulation is without distortion. The diode detector, or the envelope detector tries to extract the envelope of the received amplitude-modulated signal, and that is why it is called the envelope detector. The envelope detector circuit is very simple and inexpensive as it consists of a diode and a few resistors and capacitors; and if properly designed, gives an output that is a very good approximation of the message signal. The basic circuit of an envelope detector is shown in Fig. 4.24.

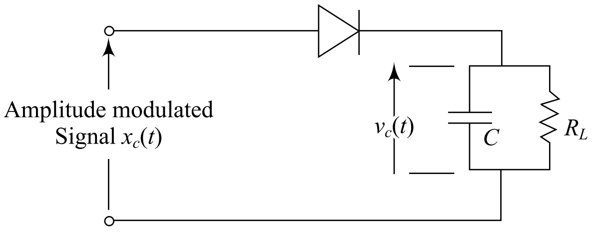


Fig. 4.24 *Basic circuit of an envelope detector*

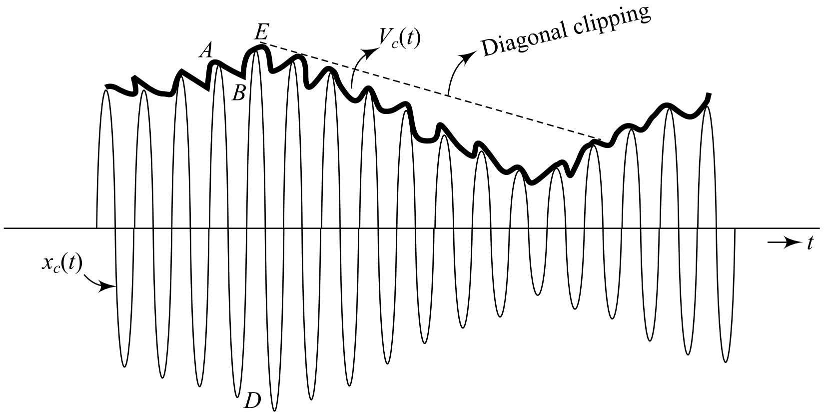


Fig. 4.25 *Working of an envelope detector*

Principle of operation During the positive half-cycle of the RF, the diode is forward biased and it conducts, charging the capacitor C. At the peak of an RF cycle, say point A, the capacitor gets charged to that peak value. Then onwards, the RF voltage of the AM wave decreases very fast. As the voltage across C cannot decrease that fast, the AM wave voltage will be less than the capacitor voltage and so the diode is reverse biased and it stops conducting. So the charging of the capacitor stops and it starts discharging through the resistor R_L. While this process is going on, the RF voltage of the AM wave goes through the portion ADB. At point B, the instantaneous voltage across the capacitor and the RF voltage of the AM wave are equal. After this instant corresponding to B, while the RF voltage is trying to increase further, the voltage across the capacitor is trying to decrease further. Hence the diode is again forward biased and it starts conducting, charging the capacitor. This charging of the capacitor continues till the peak of RF cycle at which the diode stops conducting. This cycle of events will go on repeating in all the subsequent RF cycles. The voltage across the capacitor therefore follows the variations as shown by the thick line in Fig. 4.21. It is readily seen that $v_c(t)$, the voltage across the capacitor approximately follows the envelope of $x_c(t)$, the AM signal. Low pass filtering the $v_c(t)$ removes the RF component in it and the message signal can be recovered by blocking the DC component using a coupling capacitor.

In the above explanation, certain conditions, not explicitly mentioned, have been assumed to be satisfied. These conditions are:

1. *That the charging of the capacitor takes place almost instantaneously so that the voltage across the capacitor can almost follow the portion of the RF cycle from B to E.* If the source resistance for $x_c(t)$ is R_s and the forward resistance of the diode is R_f, the charging time constant is $(R_s + R_f)\ c \approx R_sC$ since R_f is generally very small compared to R_s. Then for the above condition to be satisfied, it is required that

$$R_sC << \frac{1}{f_c} \tag{4.32}$$

2. *That the time constant for the discharge of the capacitor, viz. R_LC, should be quite large compared the period of the RF*

i.e.,
$$R_LC >> \frac{1}{f_c} \tag{4.33}$$

unless this condition is satisfied, the capacitor voltage, $v_c(t)$, will not able to follow the envelope of the AM wave during the rising portion of the envelope.

3. *That the discharge time constant, although quite large compared to the RF period, $(1/f_c)$, it is nevertheless small compared to the period of the modulating signal.*

i.e.,
$$R_LC << \left(\frac{1}{f_m}\right) \tag{4.34}$$

where f_m is the frequency of the modulating signal. In case the modulating signal is not single tone, f_m should be taken as the frequency of the highest frequency component present in the modulating signal. If this condition is not satisfied, then, during the time when the envelope is decreasing, the capacitor voltage, $v_c(t)$ cannot follow the envelope and we get a severely distorted version of the modulating signal as $v_c(t)$ the output of the envelope detector. This distortion, referred to as 'diagonal clipping', is shown in Fig. 4.19.

All the three conditions stated above may be combined as

$$\boxed{R_sC << \frac{1}{f_c} << R_LC << \frac{1}{f_m}} \tag{4.35}$$

How diagonal clipping can be avoided As we will be interested in using the maximum possible value of R_LC that would still allow us to avoid diagonal clipping, we shall now derive such an upper limit for R_LC for the case of sinusoidal modulation.

Let f_m be the frequency of the highest frequency component present in the modulating signal $x(t)$, and let it produce a modulation index m. Note that we are considering a sort of worst-case condition.

From Eq. (4.2), we may write the expression for the envelope of the AM signal as

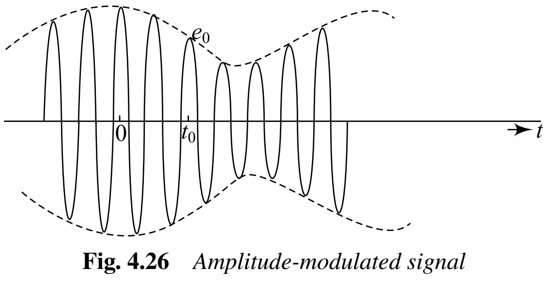


Fig. 4.26 *Amplitude-modulated signal*

$$e(t) = A_c[1 + m\cos\omega_m t] \tag{4.36}$$

∴ Rate of change of the envelope at $t = t_0$ is given by

$$\left.\frac{de(t)}{dt}\right|_{t=t_0} = -m\omega_m A_c \sin\omega_m t_0 \tag{4.37}$$

Magnitude of the envelope at $t = t_0$ is given by

$$e_0 = A_c[1 + m\cos\omega_m t_0] \tag{4.38}$$

At any time t ($t > t_0$), the voltage across the capacitor is given by

$$e_c(t) = e_0 e^{-(t-t_0)/R_L C}$$

∴ Rate of change of the capacitor voltage $= \dfrac{de_c(t)}{dt}$

$$= \frac{-e_0}{R_L C} e^{-(t-t_0)/R_L C}$$

As the capacitor commences discharging at $t = t_0$, the maximum rate of change of the capacitor voltage occurs at $t = t_0$.

∴ Maximum rate of change of the capacitor voltage $= \left.\dfrac{de_c(t)}{dt}\right|_{t=t_0} = \dfrac{-e_0}{R_L C}$ (4.39)

To avoid diagonal clipping, we have to ensure that the maximum rate of *fall* of capacitor voltage is always greater than or equal to the maximum rate of *fall* of the envelope.

∴
$$\frac{e_0}{R_L C} \ge A_c m\omega_m \sin\omega_m t_0 \tag{4.40}$$

If we now substitute for e_0 in the above equation by using Eq. (4.40), we have

$$\frac{A_c[1 + m\cos\omega_m t_0]}{R_L C} \ge A_c m\omega_m \sin\omega_m t_0$$

or

$$R_L C \le \frac{1}{\omega_m\left[\dfrac{m\sin\omega_m t_0}{1 + m\cos\omega_m t_0}\right]} \tag{4.41}$$

For the above inequality, the worst-case condition arises when the right-hand side takes a minimum value. This happens when t_0 is such that

$$\left[\frac{m\sin\omega_m t_0}{1 + m\cos\omega_m t_0}\right]$$

takes a maximum value. By differentiating the above expression, we find that it takes a maximum value when

$$\cos\omega_m t_0 = -m$$

∴ corresponding to this worst-case condition,

$$\boxed{R_L C \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]} \tag{4.42}$$

Equation (4.42) gives the maximum value of the discharge time constant that can be used for given values of modulation index and the frequency of the maximum frequency component in the modulating signal, without causing diagonal clipping.

Example 4.20 A simple diode detector uses a load resistance of 400 kilo-ohms. Across this resistance, there is a 100 pf capacitor. If the maximum modulation depth of the input amplitude modulated signal is 75%, what is the maximum frequency of the modulating signal that can be detected without diagonal clipping?

Solution $R_L C = 400 \times 10^3 \times 100 \times 10^{-12} = 4 \times 10^{-5}$ sec

∴ from Eq. (4.42), we have

$$R_L C \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]$$

$$\therefore \qquad 4 \times 10^{-5} \le \frac{1}{6.28 f_m}\left[\frac{\sqrt{1-(0.75)^2}}{0.75}\right]$$

$$\therefore \qquad f_m \le \frac{10^5}{4 \times 6.28}\left[\frac{0.6614}{0.75}\right] = 3510 \text{ Hz}$$

$$\therefore \qquad f_m \le 3510 \text{ Hz} \quad \therefore \text{ maximum frequency} = 3510 \text{ Hz}$$

4.3.4 Practical Diode Detector

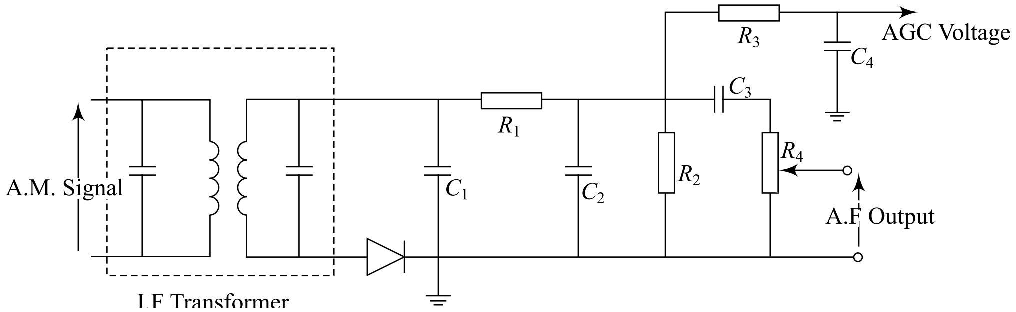


Fig. 4.27 *A practical diode detector*

The circuit of a practical diode detector is shown in Fig. 4.27. In this circuit, C_1 and C_2 are provided for RF bypass. Their values are such that their reactances are negligible at the carrier frequency (here, intermediate frequency) and extremely high at the audio frequencies. This is to ensure that while they provide good filtering of RF, they do not shunt the load resistance of the diode. C_3 is a coupling condenser and is meant for blocking the dc component while having negligible reactance (as compared to R_4) for audio frequencies. R_3 and C_4 act as a filter for audio frequencies so that almost pure dc voltage is available for AGC.

Negative peak-clipping in a diode detector The conditions based on which the values of the capacitors C_1, C_2, C_3 and C_4 are chosen, have been stated above. Although these conditions can never be fully fulfilled in practice, in our analysis of the diode detector, we shall make the following simplifying assumptions:

1. C_1 and C_2 act as short circuits for the carrier (intermediate) frequency and as open circuits for dc and audio frequencies.
2. Capacitors C_3 and C_4 act as perfect short circuits for the entire range of audio frequencies and as open circuits for the dc components.

Keeping in view the above assumptions, if we look at the circuit of Fig. 4.27, we find that the loads presented to the diode at dc and at audio frequencies are different. This difference in the loads sets a limit to the maximum value of the modulation index, m, of the incoming modulated wave. As we will see, a distortion, referred to as '*negative peak clipping*', happens if the received AM signal has a modulation index greater than a certain limit which is determined by the dc and audio loads. From Fig. 4.27, we find that

The detector load for dc $\underline{\Delta}$ $R_{\text{DC}} = R_1 + R_2$ (4.43)

The detector load for audio frequencies $\underline{\Delta}$ $R_{\text{AC}} = R_1 + (R_2 \| R_3 \| R_4)$ (4.44)

where $(R_2 \| R_3 \| R_4)$ denotes parallel combination of R_2, R_3 and R_4.

Since $R_2 > (R_2 \| R_3 \| R_4)$, it follows that the ac load (i.e., at audio frequencies) of the detector is always less than the dc load.

$$\therefore \qquad R_{\text{AC}} < R_{\text{D.C}} \tag{4.45}$$

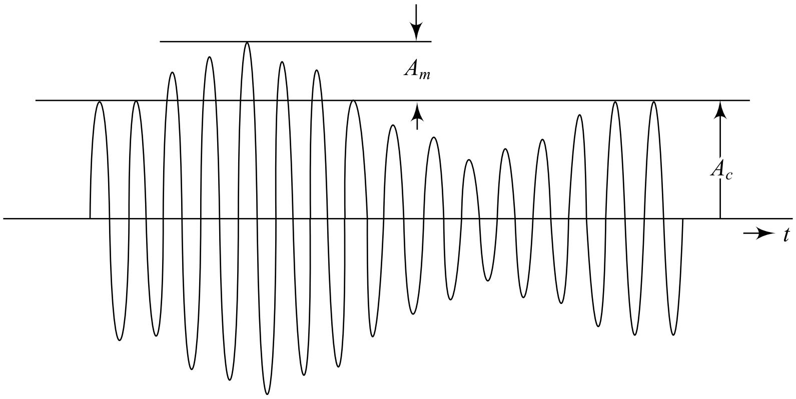


Fig. 4.28 *Received AM signal*

Modulation index of the received AM signal = $m = \dfrac{A_m}{A_c}$.

Modulation index for the diode current $\underline{\Delta}$ $m_d = \dfrac{I_m}{I_c}$

But $$I_m = \frac{A_m}{R_{\text{AC}}} \quad \text{and} \quad I_c = \frac{A_c}{R_{\text{DC}}}$$

where

A_m = Peak of the audio component of the envelope

A_c = Peak of the unmodulated carrier wave

$$\therefore \qquad m_d = \frac{I_m}{I_c} = \frac{A_m/R_{\text{AC}}}{A_c/R_{\text{DC}}} = \left(\frac{A_m}{A_c}\right)\left(\frac{R_{\text{DC}}}{R_{\text{AC}}}\right) = m\left(\frac{R_{\text{DC}}}{R_{\text{AC}}}\right) \tag{4.46}$$

$$\therefore \qquad m_{\max} = m_{d\max}\left(\frac{R_{\text{AC}}}{R_{\text{DC}}}\right)$$

But $$m_{d\max} \le 1$$

$$\therefore \qquad \boxed{m_{\max} = \left(\frac{R_{AC}}{R_{DC}}\right)} \tag{4.47}$$

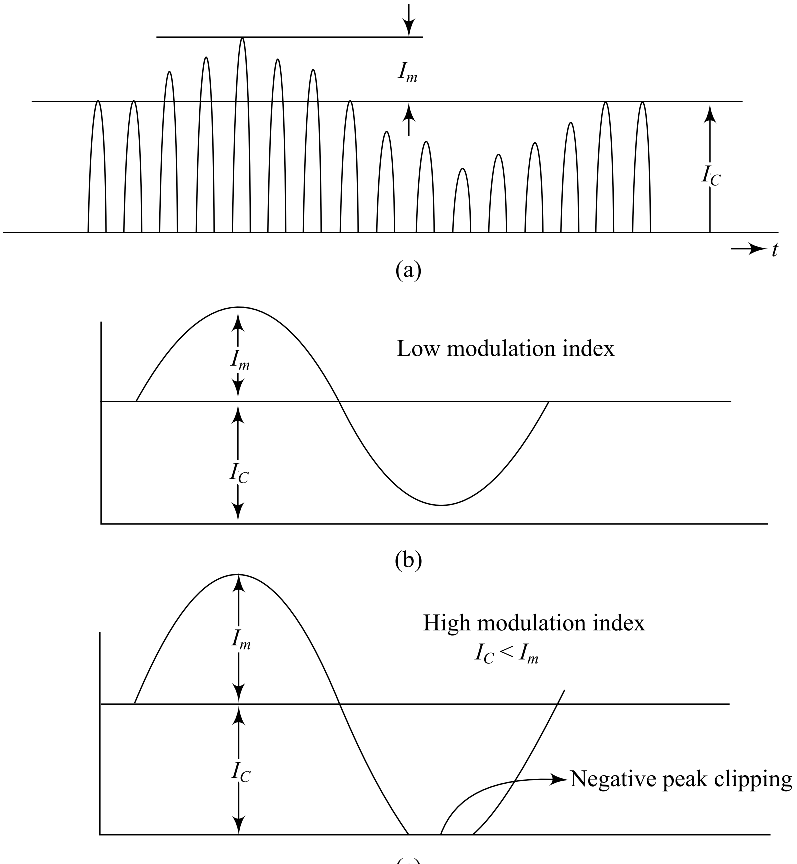


Fig. 4.29 *(a) Actual diode current, (b) dc and audio components of current for low value of m, (c) dc and audio components of current for high value of m*

and $$\boxed{\text{negative peak clipping occurs if} \quad m > \left(\frac{R_{\text{AC}}}{R_{\text{DC}}}\right)} \tag{4.48}$$

The above result is only an approximation since the assumptions 1 and 2 at the beginning of this derivation are not valid at all audio frequencies. C_1 and C_2 may act almost like short circuits at the intermediate frequency but they will not be acting as perfect open circuits at all audio frequencies. At the higher audio frequencies like say 10 kHz, these capacitors will have a finite reactance and this shunts the load. Further, the coupling condenser C_3 does not provide a reactance that is negligible compared to R_4, at the lower audio frequencies. Thus, the detector load for audio frequencies is not a pure resistance as has been assumed; instead, it will be an impedance with a capacitive reactance component. For an excellent discussion on the performance of a diode detector, the reader may refer to *Electronics and Radio Engineering* by F. E. Terman, McGraw-Hill Inc.

Example 4.21 The output of a diode envelope detector is fed through a dc blocking capacitor to an amplifying stage which has an input resistance of 10 kΩ, determine the maximum depth of sinusoidal modulation the detector can handle without negative peak clipping. **(JNTU Sep. 2007)**

Solution The blocking capacitor is meant to block the dc voltage present across the diode load resistance of 5 kΩ from reaching the input to the amplifier. Its value will be such that at even the lowest audio frequencies its reactance will be negligible compared to the input resistance of the amplifier. So, while the dc load for the diode is 5 kΩ, its ac load is the parallel combination of 5 kΩ and 10 kΩ.

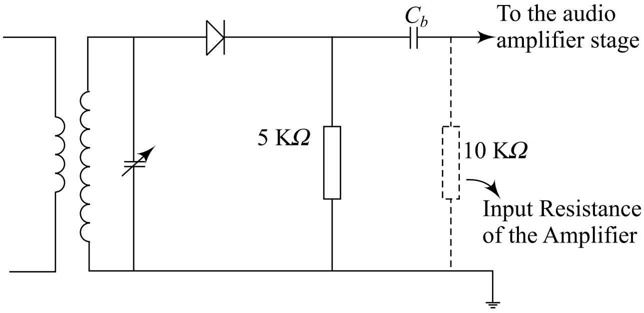


Fig. 4.30

i.e.,
$$R_{\text{ac}} = \frac{5\times 10}{(5+10)}\ \text{k}\Omega = 3.3\ \text{k}\Omega$$
$$R_{\text{DC}} = 5\ \text{k}\Omega$$

From Eq. (4.46), we know that the maximum value of the modulation index of the input AM signal which still does not cause negative peak clipping, is given by

$$m_{\max} = \frac{R_{\text{ac}}}{R_{\text{DC}}} = \frac{3.3}{5} = 0.66$$

Example 4.22 A signal $x_c(t) = 5[1+2\cos\omega_m t]\cos\omega_c t$ is to be demodulated. Check whether some of the following detectors can be used: (a) An envelope detector, (b) A square law detector, and (c) A synchronous detector or coherent detector.

Solution $x_c(t) = 5[1+2\cos\omega_m t]\cos\omega_c t$

This is an over-modulated AM signal. Hence, the envelope will be distorted and an envelope detector cannot be used.

Let us check whether a square law detector can be used.

$y_c(t)$ = output of the square law device = $ax_c^2(t)$

$$= a\left\{25[1+4\cos\omega_m t + 4\cos^2\omega_m t]\left[\frac{1}{2}(1+\cos 2\omega_c t)\right]\right\}$$

$$= \left\{\frac{25}{2} + 50\cos\omega_m t + \frac{50}{2}\left(1+\cos 2\omega_m t\right) + \frac{25}{2}\cos 2\omega_c t + 50\cos 2\omega_c t\cdot\cos\omega_m t + 50\cos^2\omega_m t\cdot\cos 2\omega_c t\right\}a$$

If the dc component is blocked by a coupling condenser and the high frequency components are removed by using an LPF of cut-off frequency f_m after the square law device, the final output will be $z(t) = a\cdot 50\cos\omega_m t$, which is proportional to the modulating signal.

Hence, a square law detector can be used.

Now, let us check whether a synchronous demodulator can be used. Recall that in synchronous demodulation, we first multiply the received modulated signal by the locally generated carrier signal and then pass the product through an LPF having a cut-off frequency of W Hz, the bandwidth of the modulating signal.

$$x_c(t)\cos\omega_c t = 5[1+2\cos\omega_m t]\cos^2\omega_c t = \frac{5}{2}[1+2\cos\omega_m t][1+2\cos\omega_c t]$$

$$= \frac{5}{2} + \frac{5}{2}\cos 2\omega_c t + 5\cos\omega_m t + \frac{5}{4}\cos(2\omega_c + \omega_m)t + \frac{5}{4}\cos(2\omega_c - \omega_m)t$$

$\therefore$ The output of the LPF = $z(t) = \left(5\cos\omega_m t + \frac{5}{2}\right)$

The dc component, i.e., 5/2, can be rejected by using a coupling condenser, and the output will then be only the message signal.

Hence, either a square law detector, or a synchronous detector, may be used, but not the envelope detector.

4.4 AM BROADCAST TRANSMITTERS

These transmitters handle speech, music, etc., and each of them is meant to serve a large number of receivers. Hence they use AM with full carrier. They generally handle audio frequencies up to 5 kHz and use carrier frequencies in the medium waveband of 550 kHz to 1605 kHz, or in the short waveband of 3 MHz to 30 MHz. Transmitters operating in the medium waveband primarily depend on ground wave propagation and because of the attenuation inherent in this mode of propagation, they have a limited service area. Those operating in the short waveband primarily depend on the sky wave propagation and so cover a very large area. AM broadcast transmitters use carrier powers of the order of 1 kW to 100 kW, or more. Since these transmitters are meant for voice communication and audio frequencies up to 5 kHz are used, and since AM with both sidebands and full carrier is used, the bandwidth required is 10 kHz. Thus, carrier frequencies are allocated with 10 kHz separation between adjacent channels. This necessitates very good carrier frequency stability, as otherwise the side frequencies of one channel are likely to drift into the adjacent channel and cause interference. That is why stringent stability requirements are imposed on the carrier frequencies, making it mandatory for the carrier frequency drift to be not more than ±20 parts per million, i.e., less than ±0.02% of the assigned carrier frequency.

4.4.1 High-Level and Low-Level Modulation

In a transmitter, modulation of the carrier may be performed either at a low carrier power level or at a high carrier power level. In the former case, it is called '*low-level modulation*' while in the latter case, it is called '*high-level modulation*'. As the modulated signal is produced at a low carrier level in the case of low-level modulation, the modulated signal so produced will have to be raised to the required power level using a chain of power amplifiers. As the modulated signal occupies certain bandwidth, these power amplifiers will have to be necessarily either class-A or class-AB tuned power amplifiers; and these will have very low efficiencies. In the case of high-level modulation, however, the carrier signal produced by an oscillator is first amplified using a series of tuned power amplifiers, which in this case can be class-C power amplifiers (with very high power efficiency) since the signal to be amplified is a sine wave. The final stage of this class-C power amplifiers chain may be plate modulated, or collector modulated, depending on whether a vacuum triode, or a transistor is used as the device. As shown in the analysis of a collector modulated class-C power amplifier, the total sideband power in the modulated signal so generated, will be derived from the final stage of the amplifier chain used for the power amplification of the modulating signal. Thus, unlike the low-level modulation case, the modulating signal power required in this case can be very high. For example, if a transmitter which is to radiate 10 kW of average power of the modulated signal employs high-level modulation and if the modulation index is say 0.8, the total sideband power will be 2424 W; and if the modulated class-C amplifier has a power efficiency of 85%, the final stage of the modulating amplifier will have to deliver about 2.85 kW of modulating signal power.

AM transmitters are generally categorized into two types – those with high-level modulation and those with low-level modulation.

Definition Modulation of the carrier by the message signal may be performed at any point beyond the oscillator buffer stage up to and including the final power amplifier. If the modulating message signal is

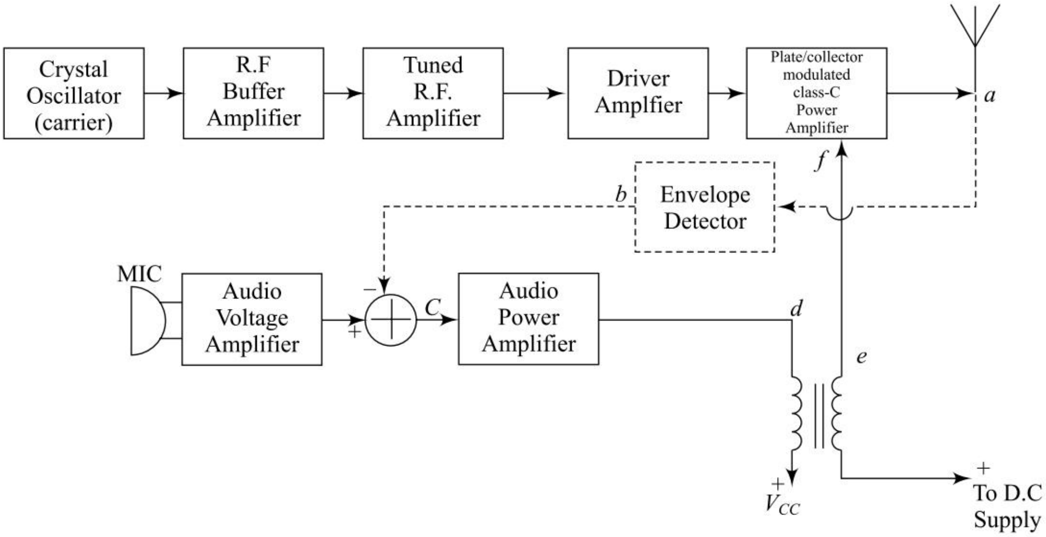


Fig. 4.31 *AM transmitter with high-level modulation*

introduced in series with the collector/plate supply voltage of the final power amplifier stage so that it becomes a collector/plate modulated class-C amplifier, the modulation is referred to as high-level modulation. On the other hand, if the modulating message signal is introduced beyond the buffer at any point up to and including the base of the final power amplifier, the modulation is referred to as low-level modulation.

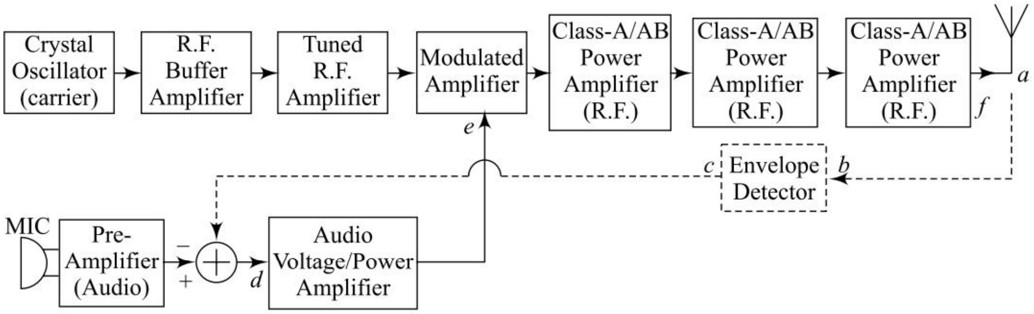


Fig. 4.32 *AM transmitter with low-level modulation*

To summarize, the advantages and disadvantages of these two types of modulation are:

Advantages of low-level modulation

1. The modulation circuit is relatively simple as the power levels to be handled are low.
2. The power required to be supplied by the modulating signal amplifier, is very low. Hence, it is especially useful when the modulating signal is a video signal, since it is difficult to get large amounts of video power, as in the case of TV transmitters.

Disadvantages of low-level modulation Since the modulated signal is generated at a low power level, class-A or class-AB tuned power amplifiers will have to be used to raise the power of this signal to the required level. These have very low efficiencies.

Advantages of high-level modulation As the modulation is performed at a high power level of the carrier, there is no need to use class-A or class-AB tuned power amplifiers. Since class-C power amplifiers are used for raising the power level of the carrier, the efficiency is quite high.

Disadvantages of high-level modulation Large amounts of modulating signal power will be needed.

4.4.2 Carrier Frequency Stability

As mentioned earlier in the beginning of this section, in order to avoid causing interference to the adjacent channels, it is absolutely necessary that the carrier frequency is extremely stable and the carrier frequency drift, if any, is not more than 20 parts per million. To achieve this level of carrier frequency stability, only crystal oscillators must be used to generate the carrier. Further, it is necessary to

1. ensure that the oscillator is not loaded and the impedance coming across its output does not change. For this purpose, a buffer has to be used as shown in Figs. 4.31 and 4.32. It must have a very high input impedance and a low output impedance.
2. keep the crystal used in the carrier oscillator circuit at a constant temperature, as temperature variations can cause frequency drift.
3. ensure that the dc supply voltages for the crystal oscillator circuit are absolutely steady, since variations in these voltages can cause frequency drifts.

Neutralization Apart from carrier frequency stability, another thing that needs special mention in connection with transmitter, is the need for neutralization of the RF amplifiers. Whether it is a vacuum tube, or a transistor that is used as the active device for the amplifier, it will have inter-electrode capacitances. It is the base-collector (or grid-plate capacitance in the case of vacuum tubes) inter-electrode capacitance which causes stability problem for the RF amplifiers, because at these frequencies, even the very small inter-electrode capacitance (generally of the order of a few pico-farads) will have small enough reactance to provide a good feedback path from collector to base (plate to grid in the case of vacuum tubes). This positive feedback can cause parasitic oscillations in the RF amplifiers. These oscillations will generally be at much higher frequencies than the carrier. They will distort the carrier signal waveform, and so will have to be avoided. The technique adopted is to neutralize the positive feedback by deliberately providing a negative feedback in equal measure – hence the name neutralization for all the different methods using this approach. Among the various neutralization methods available, Hazeltine method and Rice method are worth mentioning.

1. Hazeltine method: We first note that points *A* and *B* of the collector tank circuit in Fig. 4.33 are 180° out of phase. To neutralize the feedback from the collector (point *B*) to the base through the capacitance C_{cb}, we connect another capacitor C_N between the point *A* and the base. We then adjust it to a value equal to C_{cb}.

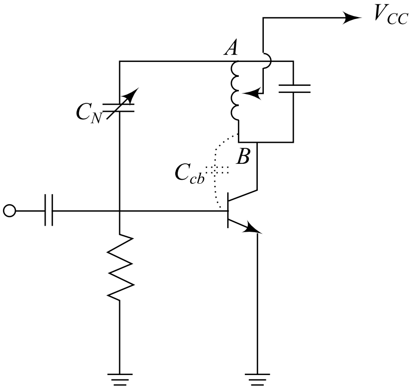


Fig. 4.33 *Hazeltine method of neutralization*

2. Rice method: The same principle is used in this method too. The only difference is that now two points which are 180° out of phase *on the base side* are used, as shown in Fig. 4.34. Because the center tap of the transformer secondary is earthed, points *A* and *B* are always 180° out of phase. Since the inter-electrode capacitance C_{cb} is connecting the point *C* to the base, i.e., point *B*, point *A* which is 180° out of phase with *B* is connected by us through the neutralizing capacitor C_N to the same point *C* and C_N is adjusted to be equal to C_{cb}, so as to neutralize the effect of C_{cb}.

4.4.3 Feedback in Transmitters

Negative feedback is invariably provided in AM broadcast transmitters with a view to improve their performance. The AM signal fed to the antenna should ideally have, as its envelope, (after removal of the dc component), the message signal as available at the output of the *audio voltage amplifier*. This will be the case only if there is no distortion produced in the audio power amplifiers, the modulation characteristic of the modulator is exactly linear and in case low level modulation is employed, if the class-A/AB tuned power amplifiers do not cause any distortion of the envelope.

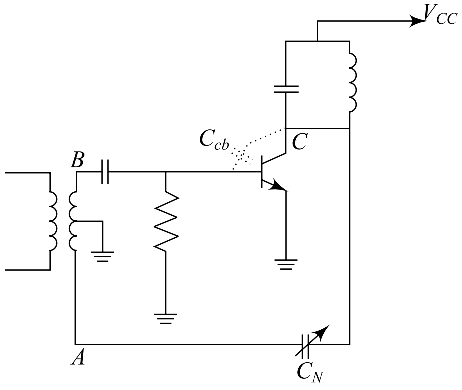


Fig. 4.34 *Rice neutralization*

This negative feedback is provided as shown in Figs. 4.31 and 4.32. The AM signal to be radiated is picked up at the point ‘a’, its envelope is extracted and the dc component is removed in order to obtain, what in an ideal situation should be the undistorted message signal. This is then added to the output of the audio voltage amplifier in such a way that it subtracts from the voltage amplifier output, as shown in Figs. 4.31 and 4.32. The loop a-b-c-d-e-f thus acts as the feedback loop. To avoid oscillations which will be caused if the feedback turns positive, it should be ensured that the loop gain $|A\beta| < 1$ for all the audio frequency components.

This negative feedback improves the performance of the transmitter as it reduces the distortion of the envelope of the radiated signal by making it closely resemble the message signal. It reduces the noise and power frequency hum also.

4.5 AM BROADCAST RECEIVERS

Historically, the earliest AM receivers were crystal, regenerative and super-regenerative receivers. However, they were soon superseded by the ‘tuned radio frequency’ (TRF) receivers, which continued to be quite popular till about the beginning of the World War II. However, the superheterodyne type of receiver, actually invented by Major Armstrong some time during World War I, became popular by about mid-1930s because of its far superior performance, and now it forms the standard structure of not only AM broadcast receivers, but also FM broadcast receivers, TV receivers and even radar receivers.

We shall discuss the TRF receiver first, and then discuss the superheterodyne receiver in some detail.

4.5.1 Tuned Radio Frequency (TRF) Receiver

As shown in Fig. 4.35, a TRF receiver simply consists of a chain of two or three single-tuned RF amplifiers, all of them tuned to the same frequency, followed by a detector, an audio voltage amplifier and an audio power amplifier that feeds the loudspeaker.

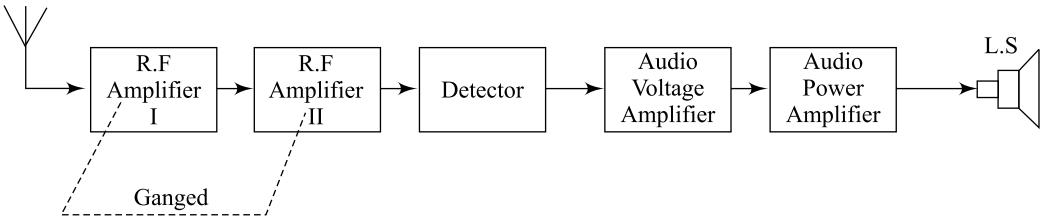


Fig. 4.35 *A tuned radio frequency receiver*

These TRF receivers are quite simple and inexpensive. But they suffer from several severe disadvantages, chief one among which is poor 'adjacent channel selectivity'. Because of this, when the receiver is tuned to a particular station, say of carrier frequency f_c, signals radiated by stations operating on adjacent channels having carrier frequencies of $f_c \pm 10$ kHz, are also received, although they are attenuated to some extent. *This is called adjacent channel interference*. This problem gets aggravated if the receiver is to be tuned over a wide frequency range, as the Qs of the tuned circuits go on changing when the receiver (i.e., the tuned RF amplifiers) is tuned to different frequencies. The adjacent channel selectivity is of course lowest when the receiver is tuned to the highest end of its frequency range.

Further, as all the amplification (of the received signal) required for proper operation of the detector, has to be at the signal frequency, there exists the possibility of instability of the RF amplifiers. Also, it has to be ensured that the RF amplifiers are all tuned to exactly the same frequency as the receiver is tuned to different stations.

The superheterodyne receiver, which we are going to discuss next, overcomes all the above problems.

4.5.2 Superheterodyne AM Broadcast Receivers

Principle of superheterodyne receivers Almost all the gain of a TRF receiver is obtained in the RF amplifiers, *at signal frequency*; and this gain varies quite a bit as the receiver is tuned to different stations. In a superheterodyne receiver, by a process of mixing, the message bearing received AM signal, whatever may be its carrier frequency, is converted into an AM signal carrying the same message signal *at a fixed carrier frequency* called the '*intermediate frequency*' (IF), which is lower than the lowest carrier frequency covered by the receiver. About 70–75% of the gain of the receiver is obtained through amplification at this fixed frequency IF by using a fixed-tuned high gain amplifier, called the IF amplifier. This signal is then detected and the extracted message signal is then amplified and fed to the loudspeaker. This way, the superheterodyne receiver overcomes all the disadvantages of the TRF receiver.

The block diagram of an AM superheterodyne broadcast receiver is shown in Fig. 4.36.

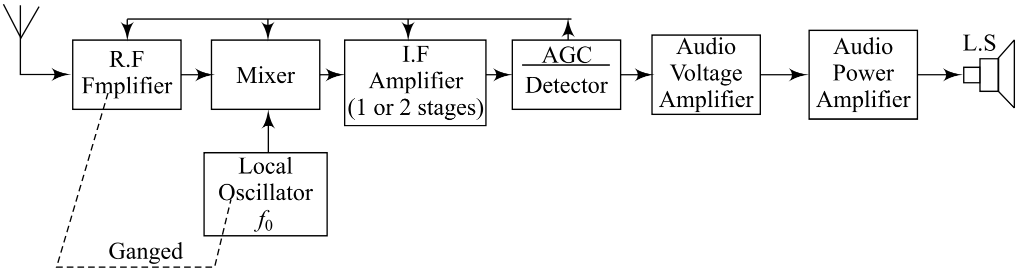


Fig. 4.36 *Block diagram of an AM superheterodyne broadcast receiver*

We shall now discuss briefly the salient features and the functions of each block in the above block diagram.

1. RF amplifier: It is a tuned voltage amplifier that selects and amplifies the signal induced in the antenna having a carrier frequency corresponding to the frequency to which it is tuned. Its bandwidth is 10 kHz. It is not designed to give a high gain and its main functions are:

(a) to ensure that the receiver has a good overall signal-to-noise ratio. If RF amplifier is not used, the mixer, which inherently is a noisy stage will be the first stage in the receiver. As the overall noise figure depends to a very large extent on the noise figure of the first stage, this will not be a desirable arrangement.

(b) to give good image frequency rejection and IF rejection capability to the receiver.

(c) to give some amount of adjacent channel selectivity.

2. Local oscillator: This is an LC oscillator which produces a sinusoidal signal of frequency f_0 which is such that $f_0 - f_c$ = IF, the predetermined fixed frequency called the intermediate frequency, where f_c is the frequency of the carrier of the station to which the receiver is tuned (this is the frequency to which the RF amplifier is tuned). The receiver may be tuned to any frequency from 550 kHz to 1605 kHz. But whatever may be the frequency to which the receiver is tuned, the local oscillator frequency tracks it in such a way as to always maintain the local oscillator frequency above the signal frequency by an amount of 455 kHz, the usual IF used in AM broadcast receivers. This is achieved by using ganged variable capacitors for tuning the tank circuits of the RF amplifier and the local oscillator and also by using appropriate tracking techniques, as discussed later. The LO frequency, f_0 can be, theoretically speaking, higher or lower than the signal frequency f_c by an amount of IF. But, for reasons discussed in detail later, it is always kept higher than the signal frequency.

3. Mixer: The received AM signal with a carrier frequency f_c, amplified by the RF amplifier, is fed as one of the inputs to the mixer, the other input signal being the output of the local oscillator, a sinusoidal signal of frequency $f_0 = f_c + f_{\text{if}}$. Mixing is a non-linear process and it results in generation of the sum and difference frequency components in addition to the original frequency components of the two input signals. The output circuit of the mixer – a tank circuit tuned to the difference frequency, i.e., the intermediate frequency, rejects all other frequency components. Thus, the output of the mixer is an AM signal whose carrier frequency is the intermediate frequency f_{if} (455 kHZ) and which is modulated by the original message signal.

Thus the mixer and local oscillator convert the received AM signal with a carrier frequency f_c into another AM signal with f_{if} as the carrier frequency. The modulation present on the original carrier is simply transferred on to the new carrier, which is the intermediate frequency. The mixer output circuit, of course, is designed to have a 3 db bandwidth of 10 kHz to accommodate all the side frequencies of the AM signal.

4. IF amplifier(s): One or two stages of IF amplifiers are generally sued. These are fixed-tuned voltage amplifiers of high gain. These IF amplifiers provide a 3 db bandwidth of 10 kHz centered on the intermediate frequency. They provide good sensitivity and selectivity to the receiver.

5. Detector: This extracts the modulating signal from the AM signal. In commercial AM broadcast receivers, envelope detectors are used and they require a minimum of at least 1 volt amplitude for proper operation. They are designed so as to provide linear operation and avoid distortions – particularly the distortion due to diagonal clipping and negative peak clipping.

As shown in Fig. 4.27 the envelope detector can be used to provide a dc voltage of appropriate polarity for automatic gain control, i.e., AGC. As shown in Fig. 4.36, this voltage is used for biasing the preceding stages so as to control their gains and thus provide AGC.

6. Automatic gain control (AGC): An arrangement for automatic gain control, or AGC, is necessary in radio receivers for the following reasons:

(a) When the receiver is tuned from one station to another, difference in signal strengths of the two stations causes an unpleasantly loud output, if from a weak station, we are moving to a strong one, unless we initially keep the volume control very low before changing the tuning from one station to another. Changing the volume control every time before attempting to retune the receiver is however, cumbersome.

(b) Even if we are not retuning to another station, signal strength from the station to which the receiver is tuned can go on fluctuating due to signal fading, causing corresponding fluctuations in the audio output from the receiver.

The points noted above underscore the need for keeping the audio output power from the receiver somewhat constant when the input RF signal level changes because of any one of the two reasons listed above. This calls for an arrangement by which the overall gain of the receiver can be made to automatically vary when the signal strength changes, in such a manner as to keep the audio output reasonably constant. Such an arrangement is called *automatic gain control* or AGC.

In receivers, automatic gain control is achieved by producing an AGC voltage form the detector circuit as shown in Fig. 4.27. This AGC voltage will be high for stronger RF input signals and low for weaker signals. We therefore apply this as a bias voltage to the RF amplifiers, mixer and the IF amplifier stages in such a way that it reduces their gain of these stages by reducing their transconductance.

This type of arrangement is called 'simple AGC'. However, there is one serious difficulty with this 'simple AGC'. Even weak RF input signals also produce some AGC voltage, though it may be small. So, while reducing the receiver gain for stronger RF signals, it reduces the receiver gain to some extent even for weak RF signals. This is undesirable.

7. Delayed AGC: To overcome this disadvantage of a simple AGC, what is referred to generally as the 'delayed AGC', may be used. It allows the AGC action to commence only after the input RF signal level reaches a predetermined level, as shown in Fig. 4.37, which depicts the AGC characteristics.

Delayed AGC is generally obtained by having a separate diode rectifier circuit for producing the AGC voltage, and applying a positive bias of predetermined value to the cathode of that diode so that it conducts and produces the AGC voltage only after the RF input to the receiver is sufficiently large.

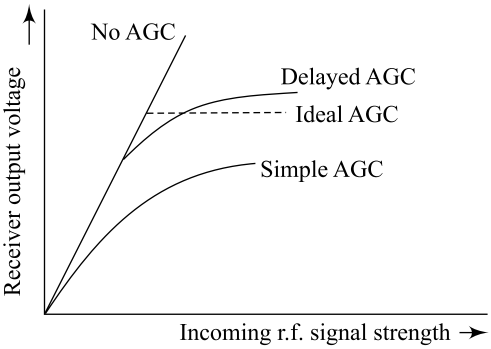


Fig. 4.37 *AGC characteristics*

8. Audio Voltage and Audio Power Amplifiers: The demodulator output is the message signal. But it is very weak and cannot be used directly to actuate a loudspeaker. So the audio signal coming out from the detector stage is first amplified using a voltage amplifier stage to raise it to a level at which it can drive a class-A audio power amplifier which is the next stage. This power amplifier is designed to have minimum distortion and a 3 db bandwidth of at least 5 kHz. It is a transformer coupled to a loudspeaker. This output transformer is also called the matching transformer since it provides good matching between the high output impedance of the power amplifier and the low impedance of the loudspeaker.

9. Choice of local oscillator frequency: It was remarked earlier that, theoretically, the local oscillator frequency, f_0, can be either greater than, or less than the carrier frequency f_c of the received signal and that what is required is only that the difference between the two should be equal to the fixed value of the intermediate frequency, f_{if}, of the receiver. We also said that for certain practical reasons, it is chosen to be higher than the f_c. We shall now examine this question.

Consider an AM superhet receiver meant for the medium waveband, which we shall take as extending from 555 kHz to 1605 kHz. Let us assume that the IF fixed for the receiver is 455 kHz.

(a) $f_0 > f_c \qquad \therefore\ f_0 = f_c + f_{\text{if}}$

Since f_c ranges from 555 kHz to 1605 kHz,

f_0 ranges from (555 + 455) kHz to (1605 + 455) kHz.

i.e., from 1010 kHz to 2060 kHz.

$\therefore\ f_{0\,\max} = 2060$ kHz and $f_{0\,\min} = 1010$ kHz

Since the frequency of the oscillator is *inversely proportional to the square root* of the tank circuit capacitance, if $C_{\max}$ and $C_{\min}$ are the maximum and minimum values of the gang condenser (oscillator section) used for tuning the oscillator, we have

$$\left(\frac{C_{\max}}{C_{\min}}\right) = \left(\frac{f_{0\max}}{f_{0\min}}\right)^2 = \left(\frac{2060}{1010}\right)^2 = (2.04)^2 = 4.16$$

This ratio is quite practicable.

Note *Even if the vanes on the rotor of the variable air condenser are completely out, the capacitance will not be zero because of the parasitic capacitances, which generally will be of the order of a few tens of pico-farads. Hence, $C_{min} \neq 0$.*

(b) $f_0 < f_c$

$\therefore\ f_0 = f_c - f_{if}$

$\therefore$ In this case, $f_{0\,\max} = (1605 - 455)$ kHz = 1150 kHz

$f_{0\,\min} = (555 - 455)$ kHz = 100 kHz

$$\therefore \qquad \left(\frac{C_{\max}}{C_{\min}}\right) = \left(\frac{f_{0\max}}{f_{0\min}}\right)^2 = \left(\frac{1150}{100}\right)^2 = 132.25$$

This is an impractical value since it implies that $C_{\max}$ should be of the order of a few thousand pico-farads!

Thus, this practical difficulty forces us to choose the local oscillator frequency to be higher than the signal frequency to which the receiver is tuned.

10. Adjacent channel selectivity: Medium frequency and high frequency bands are used for AM broadcasting and channel allocation is made using a 10 kHz separation between adjacent channels. Spectrum crowding does not permit a larger spacing between channels.

When a receiver is tuned to a particular station, adjacent channel interference occurs due to the inability of the receiver to totally reject the signal at the adjacent channel frequency. Thus, from the adjacent channel selectivity point of view, an ideal situation is one in which the RF sections of the receiver have a frequency selectivity characteristic of the shape shown in Fig. 4.38. However, no practical filter can give such a frequency response.

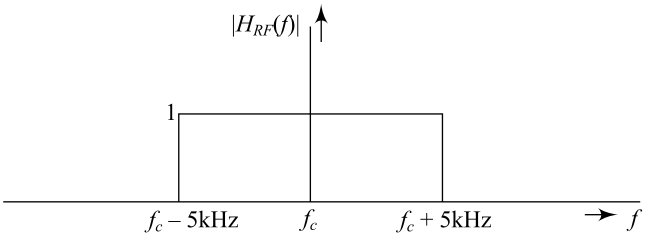


Fig. 4.38 *Ideal selectivity curve*

Further, in the RF sections, uniformly good adjacent channel selectivity cannot be maintained over the entire frequency range covered by the receiver. When the receiver is tuned to a station operating near the lower end of the AM band, say, 600 kHz, a signal from another station operating on the adjacent channel, i.e., at a frequency of 610 kHz, can be effectively suppressed since 10 kHz is *not* a very small fraction of 600 kHz. However, when the receiver, and hence the RF amplifiers, are tuned to a station at the higher end of the MW band, say 1600 kHz, an adjacent channel signal of 1610 kHz will not be very much attenuated. Hence, as we move towards the higher end of the receiver's frequency range, the adjacent channel selectivity provided by the RF amplifiers becomes progressively poorer. Such a problem does not arise in the case of IF amplifiers since these are fixed tuned and always operate at a center frequency of f_{if}, the intermediate frequency (455 kHz), whatever may be the station to which the receiver is tuned – whether it is at the lower end, or the higher end of the tuning range of the receiver. That is why almost all the adjacent channel selectivity desired for a superheterodyne receiver, is sought to be obtained from the IF stages. A good receiver is expected to give an adjacent channel selectivity of at least 60 to 80 db.

An ideal selectivity curve for the IF stages is also the same as the one shown in Fig. 4.38, except that f_c in it has to be replaced by f_{if}, the intermediate frequency. This shape may be approximated by using any of the following techniques:

(a) By using 3 or 4 identically tuned IF stages with the inter-stage transformers loosely coupled: We know that the overall frequency response of a number of cascaded amplifier stages is the product of the responses of individual stages. The skirts become sharper as we multiply and it is so arranged that an overall 3 dB bandwidth of 10 kHz is obtained, as shown in Fig. 4.39.

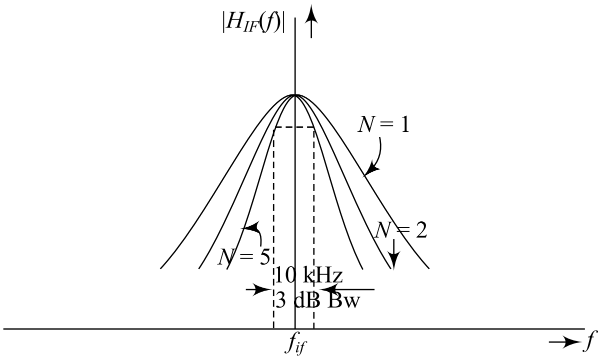


Fig. 4.39 *Selectivity curves for stages 1, 2, and 5*

(b) By using three or more stagger tuned IF stages: Three or more odd number (N) of loosely coupled IF stages may be used, to give an overall response that is reasonably flat, but has a ripple, and has fairly sharply falling skirts, as shown in Fig. 4.40. As N increases, the ripple amplitude becomes smaller and the skirts become sharper.

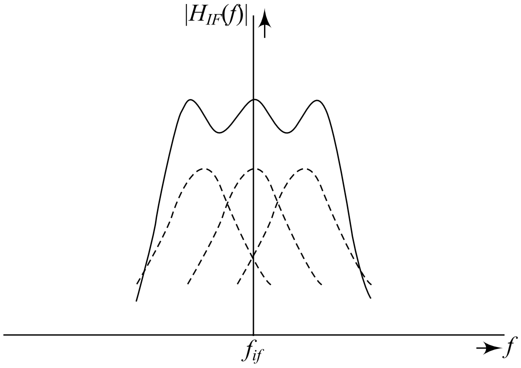


Fig. 4.40 *Selectivity curve for an IF amplifier comprising three stagger-tuned stages*

(c) By using stages with over-coupled inter-stage transformers: When transformers are over-coupled, we know that double hump appears in the frequency response. The first stage is loosely coupled while the next two inter-stage transformers are tightly coupled. The overall response exhibits three humps-one at the center and one each on either side of it.

Nowadays, high frequency op-amp-based single-chip IF amplifiers are available.

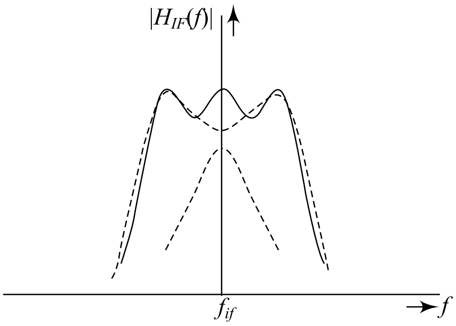


Fig. 4.41 *Using over-coupled transformers in the IF stage*

11. Image frequency rejection and image frequency rejection ratio: Suppose the receiver is tuned to a station with a carrier frequency f_c. Then the tuned circuits in the RF stage are tuned to the signal frequency f_c and the local oscillator frequency f_0 will therefore be $(f_c + f_{\text{if}})$. Now, if there is another station operating with a carrier frequency of $(f_0 + f_{\text{if}}) = (f_c + 2f_{\text{if}})$ and if that signal passes through the RF stage even in a slightly attenuated condition, in the mixer, it will also produce an output at the intermediate frequency since it also differs from the local oscillator frequency by f_{if}. So, this undesired signal also gets amplified in the IF stages along with the desired signal and causes interference at the destination. Hence, if a receiver is tuned to a desired signal having a carrier frequency f_c, *the signal with a carrier frequency of* $(f_c + 2f_{\text{if}})$ *can cause interference and it is called the image signal for the desired signal with carrier frequency* f_c. This image signal should not therefore be allowed to reach the input of the mixer stage. Of course, it is not possible to completely eliminate it, but it should be attenuated heavily in the RF stage. To what extent it can be attenuated, will depend on

(a) the Q of the tuned circuits in the RF stages (higher the values of Q, better is the image frequency rejection).

(b) the value of the IF for the receiver (higher the value of the IF, better is the image frequency rejection).

(c) whether the desired signal is close to the lower end or the higher end of the tuning range of the receiver. For fixed values of Q and IF, image rejection is better when the desired signal is at the lower end of the tuning range.

The extent to which the image frequency signal is rejected by the receiver is generally expressed in terms of what is referred to as the 'Image Frequency Rejection Ratio (IFRR)', which is defined as follows:

$$\boxed{\text{IFRR} \;\underline{\Delta}\; 10\log_{10}\left|\frac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f_c')}\right|^2} \tag{4.49}$$

where f_c is the desired carrier frequency to which the receiver is tuned, f_c' is the corresponding image frequency, i.e., $f_c' = f_c + 2(\text{IF})$ and $H_{\text{RF}}(f)$ is the transfer function of the RF amplifier.

The dependence of image rejection capability of a receiver on the above quantities follows from the off-resonance behavior of a parallel resonant circuit. We shall now examine this briefly.

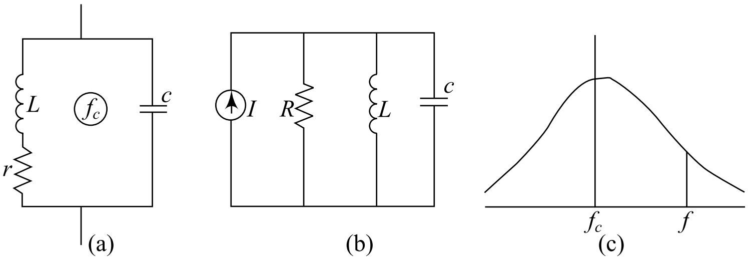


Fig. 4.42 *(a) A parallel resonant circuit, (b) Its equivalent circuit, (c) Its frequency response*

Let R be the equivalent parallel resistance which takes care of the small series resistance r associated with the coil of inductance L. Then

$$\text{Admittance at resonance } Y_{f_c} = \frac{1}{R}$$

At some frequency $f \neq f_c$, the admittance of the circuit is given by

$$Y_f = \frac{1}{R} + j\left(\omega C - \frac{1}{\omega L}\right) \tag{4.50}$$

If Q is the figure of merit of the tuned circuit at resonance, and if

$$\frac{1}{Y_{f_c}} \underline{\Delta}\, A_c \quad \text{and} \quad \frac{1}{Y_f} \underline{\Delta}\, A \tag{4.51}$$

We may write

$$\frac{A_c}{A} = 1 - jQ\left(\frac{\omega_c}{\omega}\right) + j\left(\frac{\omega}{\omega_c}\right)Q$$

$$= 1 + jQ\left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right) \tag{4.52}$$

$$\therefore \qquad \left|\frac{A}{A_c}\right|^2 = \frac{1}{1 + Q^2\left(\dfrac{\omega}{\omega_c} - \dfrac{\omega_c}{\omega}\right)^2} = \frac{1}{1 + x^2Q^2} \tag{4.53}$$

where $$x \triangleq \left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right) \tag{4.54}$$

$$\therefore \qquad \left|\frac{A}{A_c}\right| = \frac{\text{off-resonance response}}{\text{response at resonance}} = \frac{1}{\sqrt{1+x^2Q^2}} \tag{4.55}$$

Equation (4.55) clearly brings out the dependence of the degree of image rejection on the values of Q, f_{if} and f_c. Obviously $|A/A_c|$ decreases as Q increases. If we consider the image frequency, $\omega = \omega_c + 2\omega_{\text{if}}$, when ω_{if} is large, value of x will be higher and so $|A/A_c|$ will decrease for larger values of IF. For a given ω_{if}, x increases as ω_c is decreased and so again the image rejection will be better since $|A/A_c|_{\omega=\omega_c+2\omega_{\text{if}}}$ decreases. Note that Eq. (4.55) is for the case of a single stage of RF amplifier. *For multistage case, the relative responses get multiplied.*

From Eqs. (4.49) and (4.55), it follows that

$$\text{IFRR} \triangleq 20 \log_{10}\sqrt{1+x^2Q^2}\text{ , where } \quad x = \left(\frac{\omega'}{\omega_c} - \frac{\omega_c}{\omega'}\right)$$

$$\omega' \text{ being } 2\pi \text{ times the image frequency.} \tag{4.56}$$

Example 4.23 An AM superheterodyne broadcast receiver is tuned to 600 kHZ. If the Q of its single-stage RF amplifier tank circuit is 60 and the IF (for the receiver) is 455 kHz, determine the image rejection of the receiver in dB. In case it has a two-stage RF amplifier with identical tank circuits, what will be the image rejection?

Solution

$$\left|\frac{A}{A_c}\right| = \frac{1}{\sqrt{1+x^2Q^2}}; \qquad x = \left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right)$$

For the image signal, $\omega = \omega_c + 2\omega_{\text{if}}$ $\quad \therefore f = 600\times 10^3 + 2\times 455\times 10^3$

$\therefore$ Image frequency = 1510×10^3 Hz when $f_c = 600 \times 10^3$ Hz

$$\therefore \quad x = \left(\frac{1510}{600} - \frac{600}{1510}\right) = 2.5166 - 0.3973 = 2.1193$$

$$\therefore \quad 1 + x^2Q^2 = 1 + 4.49 \times (60)^2 = 16.165 \qquad \therefore \sqrt{1+x^2Q^2} = \sqrt{16,165} = 127.14$$

$$\therefore \quad \left|\frac{A}{A_c}\right| = \frac{1}{\sqrt{1+x^2Q^2}} = \frac{1}{127.14}$$

$$\therefore \quad \left|\frac{A_c}{A}\right|_{\text{dB}} = \text{Image rejection in dB} = 20 \log_{10} 127.14 = 42 \text{ dB}$$

If a two-stage RF amplifier is used, image rejection = 84 dB.

Example 4.24 When a superheterodyne receiver is tuned to 555 kHz, its local oscillator provides the mixer with an input at 1010 kHz. What is the image frequency? The antenna at the receiver is connected to the mixer via a tuned circuit whose loaded Q is 40. What will be the rejection ratio for the calculated image frequency?

Solution We know that

Image frequency $f' = f_c + 2f_{\text{if}}$ and local oscillator frequency $f_0 = f_c + f_{\text{if}}$

$\therefore\ f_0 = f_c + f_{\text{if}} = 1010\text{ kHz} = 555\text{ kHz} + f_{\text{if}}\text{ kHz}$

$\therefore\ f_{\text{if}}$ = Intermediate frequency = 455 kHz

$\therefore$ image frequency = $f' = 555\text{ kHz} + 2\times 455\text{ kHz} = 1465\text{ kHz}$

$\therefore$ from Eq. (4.55), we have

IFRR (Image Frequency Rejection Ratio) = $10\log_{10}\left|\dfrac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f')}\right|^2 = 20\log_{10}\sqrt{1+x^2Q^2}$

But $$x = \left(\frac{f'}{f_c} - \frac{f_c}{f'}\right) = \left(\frac{1465}{555} - \frac{555}{1465}\right) = 2.2608$$

and $$Q^2 = 40\times 40 = 1600$$

$\therefore$ $$\text{IFRR} = 20\log_{10}\sqrt{1+(2.26.8)^2 40^2} = 20\log_{10} 90.5 = 39.132\text{ dB}$$

Example 4.25 In a broadcast superheterodyne receiver having no RF amplifier, the loaded Q of the antenna coupling circuit is 100. If the IF is 455 kHz, determine (a) the image frequency and its rejection ratio for tuning at 1100 kHz, and (b) the image frequency and its rejection ratio for tuning at 25 MHz.

Solution Image frequency $f' = f_c + 2f_{\text{if}}$

(a) Since the receiver is tuned to a frequency of 1100 kHz

$$f_c = 1100\text{ kHz}$$

It is given that f_{if}, the intermediate frequency is

$$f_{\text{if}} = 455\text{ kHz}$$

$\therefore\ f'$, the image frequency = $f_c + 2f_{\text{if}} = (1100+910)\text{ kHz} = 2010\text{ kHz}$

From Eq. (4.56), we have

$$\text{Image Frequency Rejection Ratio (IFRR)} = 10\log_{10}\left|\frac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f')}\right|^2\text{ dB}$$

$$= 20\log_{10}\sqrt{1+x^2Q^2}$$

Here, $$x = \left(\frac{f'}{f_c} - \frac{f_c}{f'}\right) = \left(\frac{2010}{1100} - \frac{1100}{2010}\right) = (1.8272 - 0.5472) = 1.28$$

and $$Q^2 = 100\times 100 = 10^4$$

$\therefore$ $$\text{IFRR} = 20\log_{10}\sqrt{1+(1.28)^2 10^4} = 20\log_{10} 128 = 42.14\text{ dB}$$

(b) Assuming the same Q for the antenna coupling circuit when the receiver is tuned to 25 MHz, we have, in this case

$$f_c = 25\text{ MHz};\quad f_{\text{if}} = 455\text{ kHz}$$

$\therefore$ $$f' = \text{Image frequency} = f_c + 2f_{\text{if}} = 25\text{ MHz} + 910\text{ kHz} = 25{,}910\text{ kHz}$$

$$x = \left(\frac{f'}{f_c} - \frac{f_c}{f'}\right) = \left(\frac{25.910}{25} - \frac{25}{25.910}\right) = (1.0364 - 0.9648) = 0.0716$$

$\therefore$ $$\text{IFRR} = 20\log_{10}\sqrt{1+10^4(0.0716)^2} = 20\log_{10}\sqrt{52.2656} = 17.18\text{ dB}$$

4.5.3 Double Spotting

Suppose the carrier frequency of the desired station is f_{s1} and the receiver (i.e., the RF amplifiers) are tuned to this frequency. For this dial setting, the local oscillator frequency $f_{01} = (f_{s1} + f_{\text{if}})$. Now, suppose we go down the tuning range of the receiver. The local oscillator frequency also goes down. At some setting of the receiver tuning dial, the local oscillator frequency will take the value $f_{02} = (f_{s1} - f_{\text{if}})$. Then, with this dial setting also, the signal f_{s_1} will be received, although with reduced strength, since f_{02} and f_{s1} differ by the IF. This phenomenon of a desired signal f_s being received at two different dial settings of the receiver, is known as '*double spotting*'. It must be noted that with the dial setting such that the local oscillator frequency is f_{02}, the RF amplifiers are tuned to a signal frequency $f_{s2} = (f_{02} - f_{\text{if}})$ and that the signal frequency f_{s1} which is equal to $(f_{02} + f_{\text{if}})$ is just the image frequency of f_{s2}. *Thus, the cause for the occurrence of double spotting is the same as the one for the occurrence of image interference* and the steps to be taken to avoid it are the same – improving the Q of the RF amplifiers and choosing the largest possible value for the intermediate frequency.

Therefore we will now discuss the various factors affecting the choice of the value of the intermediate frequency, f_{if}, of the receiver.

4.5.4 Choice of the Value of IF

Following are the factors governing the choice of the IF of a superheterodyne receiver:

1. The IF should be outside the tuning range of the receiver; except in certain special types of receivers, it is generally chosen lower than the lowest frequency covered by its tuning range. Hence, for an AM broadcast receiver, it should be less than 550 kHz.
2. A lower value of IF improves the selectivity of the receiver and reduces the adjacent channel interference.
3. A higher value of IF makes the frequency difference between the desired station of frequency f_c to which the receiver is tuned, and its image frequency $(f_c + 2f_{\text{if}})$, larger. Hence, the image frequency rejection is improved.

Because of the conflicting requirements as stated above, the choice of the value of IF is generally a matter of compromise. Hence, it is generally chosen to be the highest possible value which is lower than the lowest frequency in the tuning range of the receiver.

Typical values of IF are 455 or 465 kHz for AM broadcast receivers, 9.7 MHz for the FM broadcast receivers, 26 MHz for the video channel of VHF band TV receivers and 41 MHz to 46 MHz for the video channel of UHF band TV receivers.

4.5.5 Tracking and Alignment

In a superheterodyne receiver, the tuning capacitors of the RF amplifier and the local oscillator are ganged, i.e., the rotating plates of both these variable capacitors are mounted on a common shaft so as to have only one tuning control for the receiver. But we know that the difference between the local oscillator frequency and the frequency to which the RF amplifier is tuned, should be equal to the IF and should be maintained at that value irrespective of the station to which the receiver is tuned, i.e., irrespective of the position of the shaft of the tuning capacitor. This means that the local oscillator frequency should track the frequency to which the receiver is tuned and keep itself always above the latter by an amount equal to the IF. This is achieved as follows.

For single-band receivers, the plates of the variable capacitor of the local oscillator section are made smaller than those of the RF amplifier section, in order to make the local oscillator frequency to be above the frequency to which the RF amplifier is tuned. In order to keep this difference equal to the fixed IF of the receiver for all positions of the rotor shaft of the ganged capacitor, i.e., for proper tracking, the rotor mounted plates of the oscillator section are suitably *segmented.*

For superheterodyne receivers covering the medium wave as well as the short wavebands, the two sections of the ganged condenser are made exactly identical. The minimum and maximum values of capacitance in each section being about 50 pf and 500 pf, respectively. The inductance in the local oscillator tank circuit is made slightly smaller than the one used in the RF amplifier tuned circuit so as to keep the local oscillator frequency higher than the frequency to which the RF amplifier is tuned. In addition, small variable capacitors, a padder (C_P) and trimmer (C_T), may be used in the local oscillator tuned circuit. *Padder is the name given to the capacitor connected in series with the variable tuning condenser while trimmer is the name given to the one connected across the tuning condenser*. If a padder alone or a trimmer alone is used, it leads to what is generally referred to as the *two-point tracking*, wherein the LO frequency and the frequency to which the receiver is tuned, differ exactly by the correct value of the IF of the receiver only at two frequencies in the tuning range of the receiver, one located near the lower end of the range and the other near the upper end. In between these two points at which tracking is perfect, the difference between the LO frequency and receiver tuning frequency will not differ exactly by the IF and we say there is a small 'tracking error'. This tracking error can be adjusted to be small by means of the padder or the trimmer, as the case may be. Using a padder as well as a trimmer will give a three-point tracking. These various conditions are shown in Figs. 4.43 (a), (b), and (c).

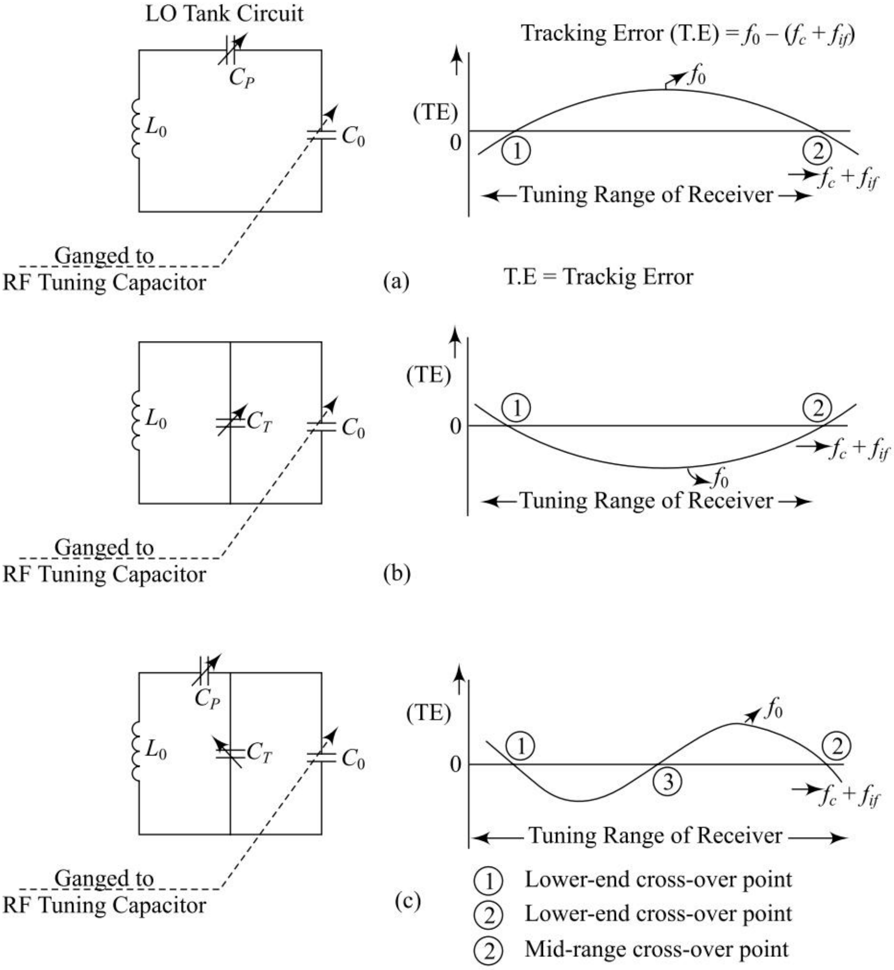


Fig. 4.43 *Local oscillator tank circuit with padder and trimmer connections. Also shown are tracking curves for two-point and three-point tracking*

The local oscillator tank circuit's inductance is first adjusted to give perfect tracking when the receiver is tuned to the middle of the band. The receiver is then tuned to a frequency near the high frequency end of its tuning range and the trimmer is adjusted to obtain the correct oscillator frequency that gives exact IF. Next, the receiver is tuned to a frequency near the lower end of its tuning range and now the padder is adjusted to get the correct oscillator frequency that gives the exact IF. These steps are then repeated approximately three or four times to achieve the correct tracking. It may be noted in this connection that to a large extent, the trimmer capacitor determines the higher-end crossover point while the padder capacitor determines the lower-end crossover point. The mid-range crossover point is determined by the inductance L_0.

4.5.6 Double Heterodyne Receivers

It has already been explained earlier that *for good image rejection a high value of IF is required* and that *for good sensitivity and selectivity, a low value of IF is required*. Hence, the choice of IF value is generally based on a compromise between these conflicting requirements.

For the reception of AM signals in the medium wave and short wave band, usage of 455 kHz or 465 kHz as the IF does not cause any problems since at these signal frequencies 455 kHz is large enough to give a good image rejection and at the same time it is small enough to give a good adjacent channel selectivity even though adjacent channels are separated only by 10 kHz. At higher signal frequencies, as are used in FM, an IF of 10.7 MHz is large enough to give good image rejection but would have been too large to give a bandwidth of say 10 kHz as required for AM. However, since the adjacent channel separation for FM is 200 kHz, it is possible to get the required values of Qs using L and C, to get an IF bandwidth of 200 kHz at a center frequency of 10.7 MHz.

But if we consider VHF communication receivers which have high signal frequencies but need an IF bandwidth of only 10 kHz, problems arise in the choice of IF. The high signal frequencies require a high IF for adequate rejection of image signals. However, a bandwidth of 10 kHz centered on a high value of IF would necessitate filters with extremely high values of Q – like those that can be obtained only from crystal filters.

However, this problem posed by high signal frequencies and small adjacent channel separation, may be solved by the use of double heterodyne, or double conversion receivers *that can give good image rejection as well as good selectivity*.

The idea is simple – *use a high first IF to get good image rejection and a low second IF to get good gain* (*sensitivity*) *and adjacent channel selectivity*, *by resorting to double conversion*. Sometimes the first IF is chosen higher than the signal frequency upper limit and the LO frequency is chosen to be IF – signal frequency). In that case, the filter in the output of the first mixer selects the sum frequency. The LO for the second mixer is generally a crystal oscillator. Since the second IF is chosen quite low, the second IF amplifier is designed to give almost all the required sensitivity (gain).

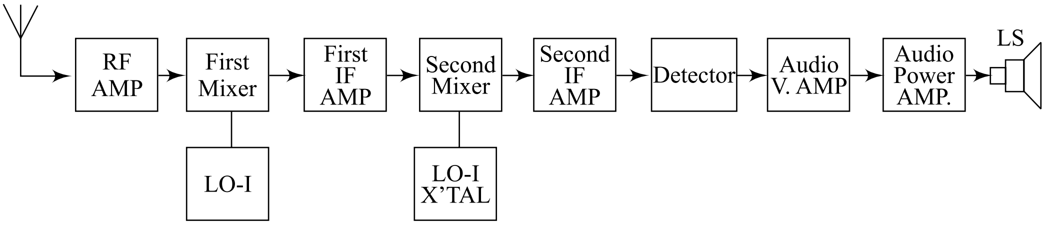


Fig. 4.44 *Double heterodyne receiver*

4.5.7 Receiver Parameters and Characteristics

When we discuss the performance of a receiver, the most important parameters that need to be considered are its *sensitivity, selectivity, fidelity and noise figure*, although there are many others which also influence its performance. We shall now discuss these parameters in some detail.

1. Sensitivity: Whatever may be the transmitted power, because of the losses in the transmission path, the signal received by a receiving antenna will generally be exceedingly weak. The signal power at the input terminals of the receiver may be of the order of pico-watts (10^{-12} W) or less (or, the voltage may be a few micro-volts or less). However, the loudspeaker needs about 1 W of audio power to be applied to it for satisfactory operation. In fact, the envelope detector of the AM broadcast receiver itself requires an AM signal voltage of at least 1 V for proper demodulation. Thus, a considerable voltage (RF and IF) amplification is needed before the demodulation and again a considerable audio voltage and power amplification is needed after detection. This overall gain determines the 'sensitivity' of the receiver, since *the sensitivity of a receiver is expressed as the signal voltage required to be applied to the receiver input to obtain some specified standard output power*. For AM broadcast receivers, it has been defined as follows.

Definition 'The sensitivity of an AM broadcast receiver is the amplitude of a carrier wave modulated to 30% by a 400 Hz tone, which, when applied to the input of the receiver through a standard artificial antenna, produces an output of 0.5 W in a resistance of appropriate value connected in the place of the loudspeaker.'

Note *The artificial antenna, comprising an inductance of 20 micro-henries in series with a 200 pf capacitor, is used to simulate the standard wire antenna of a broadcast receiver.*

The sensitivity, as defined above, naturally depends on the frequency of the applied carrier. Hence it is generally given as a curve as shown in Fig. 4.45. Since most of the gain of the receiver is obtained in the IF stage, the gain of this stage plays a key role in determining the sensitivity of the receiver. Since this gain is obtained at a constant frequency, the sensitivity of the receiver is, *to a large extent*, independent of the signal frequency.

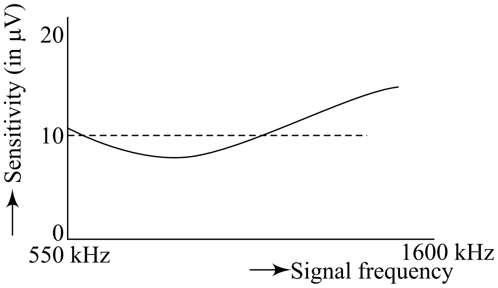


Fig. 4.45 *Typical sensitivity curve of a standard AM broadcast receiver*

2. Selectivity: The selectivity of a receiver represents the ability of the receiver to distinguish between the desired signal to which the receiver is tuned and the other signal frequencies.

It is expressed as the ratio of the signal voltage (*i.e., a carrier modulated to 30% by a 400 Hz tone*) *required at the input to the receiver to produce a standard output when the frequency of the carrier of the signal voltage is slightly away from the desired carrier frequency (i.e., the one to which the receiver is tuned), to the signal voltage required to be applied as input to produce the same standard output when the signal voltage is at the desired carrier frequency*. This ratio of the signal voltages is generally expressed in decibels. When the input signal voltage is at the desired frequency, the input voltage required to produce the standard output takes a minimum value and increases on either side, as we move away from desired frequency to which the receiver has been tuned. The selectivity is also expressed by means of a selectivity curve. Figure 4.46 shows the typical selectivity curve of a receiver.

The 3-dB bandwidth of the selectivity curve tells us whether all the side frequencies are getting through or not. If it is less than 10 kHz, the high frequency components of the modulating signal (which appear at the edges of the sidebands) are getting rejected and that the received message is getting distorted. *The selectivity at 10 kHz off resonance on either side represents the adjacent channel selectivity.*

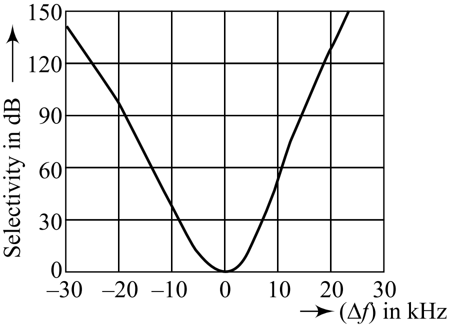


Fig. 4.46 *Typical selectivity curve of an AM broadcast receiver. '0' on X-axis corresponds to desired frequency*

3. Fidelity: Ideally a receiver should be able to give at its output, a signal that is an exact replica of the modulating signal. A good receiver therefore, should be able to do this with very little distortion.

The output signal may be a distorted version of the modulating signal because of some or all of the following reasons:

(a) Inter-modulation frequency components may be generated when the desired signal mixes with an interfering signal in a non-linear way.

(b) Inter-modulation frequency components may be produced even by the non-linearities, if any, present in the detector stage.

(c) Distortion due to suppression of the high frequency components of the modulating signal, can take place if the IF bandwidth of the receiver is inadequate for the audio bandwidth being handled by the transmitter. For instance, if the transmitter is handling audio frequencies up to 5 kHz, the IF bandwidth required is 2 × 5 kHz = 10 kHz. But suppose, the IF bandwidth is only 6 kHz, then all the frequency components of the modulating signal which are above 3 kHz will be suppressed, causing distortion.

(d) Distortion of the message signal can arise also due to the IF amplifier frequency response being not constant over its bandwidth of 10 kHz.

(e) Poor low frequency and/or high frequency response as well as non-flat mid-band gain of the audio voltage and power amplifiers will also cause distortion of the message signal.

The term 'fidelity' denotes how faithfully the receiver is able to reproduce the modulating or message signal at its output; and is generally expressed in the form of a characteristic as shown in Fig. 4.47. For plotting this curve, a carrier signal which is 30% modulated by an audio-modulating tone, is applied as input to the receiver and its relative response is plotted for various values of the frequency of the modulating tone, taking the response for 400 Hz modulating tone as the reference (0 dB).

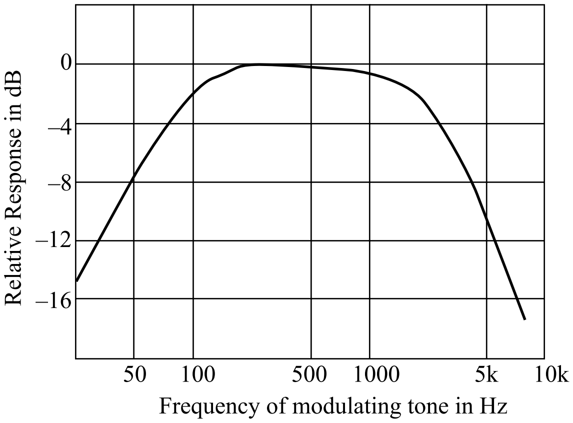


Fig. 4.47 *Typical fidelity curve of a standard broadcast receiver*

4. Noise figure: The noise figure of a two-port network indicates the amount of noise power internally generated in the network. In the case of a receiver, the received signal, given as input, has signal and noise components; this noise being the additive noise contributed by the channel. Now, when it passes through the receiver, the receiver's internally generated noise gets added. Hence, the noise figure of a receiver indicates to what extent the receiver degrades the received signal's signal-to-noise ratio, since we have defined the noise figure as the ratio of the input signal-to-noise ratio to the output signal-to-noise ratio.

i.e.,
$$\text{NF} = \frac{(S/N)\text{ input}}{(S/N)\text{ output}}$$

AM broadcast receivers generally have noise figures of the order of about 5 to 10 dB.

The noise figure of a receiver is an important parameter since it determines the smallest signal power that it can receive without making the output signal get drowned in noise.

4.6 DOUBLE SIDEBAND SUPPRESSED CARRIER (DSB-SC) MODULATION

While discussing the carrier and sideband components of power in an AM signal, it was shown in Section 4.2 that even with $m = 1$, a large portion (66.67% in the case of tone modulation) of the total average power of the AM signal lies in the carrier component. Since the information in the message (i.e., modulating signal) is contained only in the sidebands and not in the carrier, and since the carrier is anyhow filtered out and rejected in the receiver, the carrier component of power in an AM signal is a waste. Further, the AM signal occupies a bandwidth of $2W$ where the modulating signal is of bandwidth W. In fact, the information contained in the message is completely available in any one of the two sidebands and can be recovered in the receiver even if just one sideband alone occupying a bandwidth of W is transmitted. Thus, the AM is wasteful in power as well as bandwidth.

A modulation process in which the modulated signal contains no carrier component and has only the two sidebands, is called 'Double Sideband Suppressed Carrier Modulation', or simply, 'DSB-SC Modulation'. Before we discuss how such a signal may be produced, let us see what happens if the carrier component of an AM signal is removed. For this, referring to Eq. (4.7), if we ignore the first term which represents the carrier, we get

$$x_c(t) = A_c x(t) \cos \omega_c t \tag{4.57}$$

(We have absorbed m into the amplitude factor A_c)

From the above equation, it is clear that a DSB-SC signal can be generated easily just by taking the product of the carrier and modulating signals.

Since the carrier component is totally absent in the DSB-SC signal, demodulating it for recovering the message signal, $x(t)$, requires complex receiving equipment. Hence, unlike AM, it cannot be used for broadcasting purposes. Since both the sidebands are present, it requires a bandwidth of $2W$, just like the AM. Hence, it is not used even in carrier telephony and point-to-point radio communication since SSB-SC, i.e., single sideband suppressed carrier is more preferable in such applications because it offers saving of power as well as bandwidth. It is in forming the chrominance signal in the NTSC and PAL color television systems that the DSB-SC has found its greatest use. This again is mainly because of the quadrature multiplexing (about which we will be discussing later) possibility that DSB-SC offers. Further, generation of a DSB-SC signal constitutes the first step in the generation of SSB-SC signal using the filter method. Hence, we shall discuss, in some detail, the time-domain and frequency-domain representation, as well as the methods of generation of DSB-SC signals.

4.6.1 Time-Domain Representation of DSB-SC Signals

From Eq. (4.57), we have

$$\underset{\text{(DSB-SC)}}{x_c(t)} = x(t)A_c\cos\omega_c t \tag{4.58}$$

Since $x(t)$ multiplies the carrier signal $A_c\cos\omega_c t$, whenever $x(t)$ changes sign, the DSB-SC modulated signal suffers a 180° carrier phase reversal. Such a thing does not happen in AM unless over-modulation takes place. Further, as may be inferred from the waveforms of Fig. 4.48, simple envelope detector using a diode cannot be used for recovering the message signal from a DSB-SC signal.

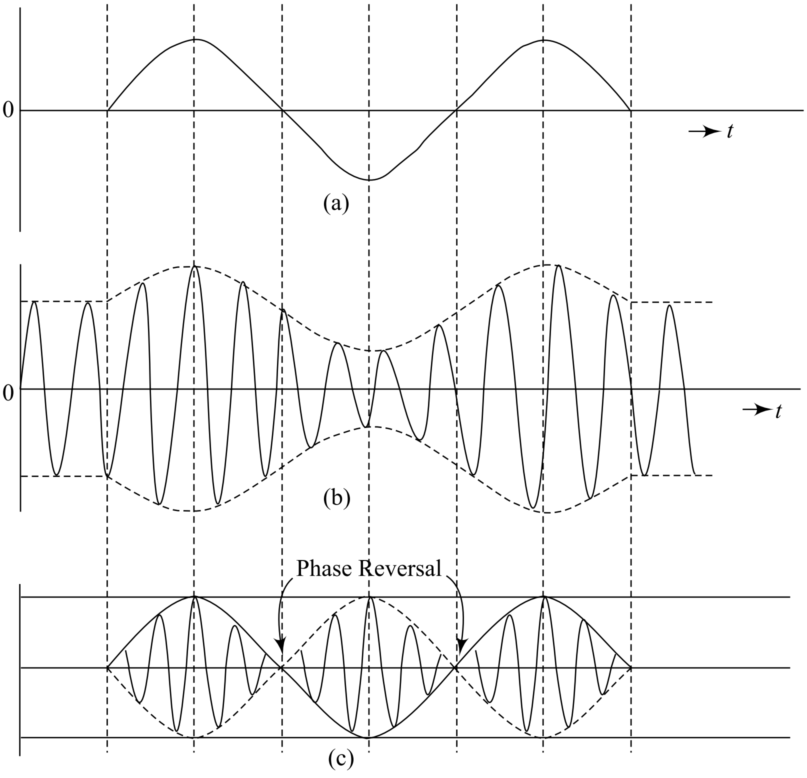


Fig. 4.48 *Waveform of (a) Modulating signal, (b) AM signal (m < 1), (c) the DSB-SC signal (product of x(t) and c(t))*

Power in a DSB-SC signal From Eq. (4.57), we have

$$x_c(t) = A_c x(t)\cos\omega_c t$$

We know that the average power of $x_c(t)$ is given by

$$P_{x_c} = \underset{T\to\infty}{\text{Lt}}\frac{1}{T}\int_{-T/2}^{T/2} x_c^2(t)\,dt$$

$$= \underset{T\to\infty}{\text{Lt}}\frac{1}{T}\int_{-T/2}^{T/2}\frac{1}{2}A_c^2x^2(t)\left[1+\cos 2\omega_c t\right]dt$$

$$= \frac{A_c^2}{2}\underset{T\to\infty}{\text{Lt}}\frac{1}{T}\int_{-T/2}^{T/2}x^2(t)dt + \frac{A_c^2}{2}\underset{T\to\infty}{\text{Lt}}\frac{1}{T}\int_{-T/2}^{T/2}\cos 2\omega_c t dt$$

The second integral is zero since it is the area under a cosine curve. Further,

$$\underset{T\to\infty}{\text{Lt}}\ \frac{1}{T}\int_{-T/2}^{T/2} x^2(t)dt = P_x = \text{Average power in the modulating signal}$$

$$\therefore \qquad \boxed{P_{x_c} = \frac{A_c^2}{2}\cdot P_x} \qquad (4.59)$$

4.6.2 Frequency-Domain Representation of DSB-SC Signals

For the sake of this discussion let the magnitude spectrum of the message or modulating signal be as shown in Fig. 4.49. Its shape has no particular significance except that it should have even symmetry since $x(t)$, the modulating signal, is real valued.

Taking the Fourier transform on both sides of Eq. (4.57), we have

$$X_c(f) = \mathcal{F}[x(t)\cdot A_c \cos\omega_c t] = \frac{1}{2}A_c[X(f-f_c)+X(f+f_c)] \qquad (4.60)$$

Figure 4.50 gives a plot of $|X_c(f)|$ making use of the $X(f)$ that we have assumed earlier.

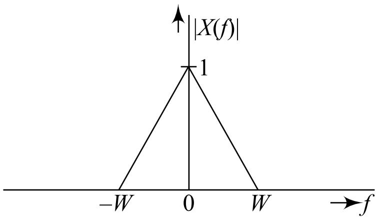


Fig. 4.49 *Spectrum of the modulating signal*

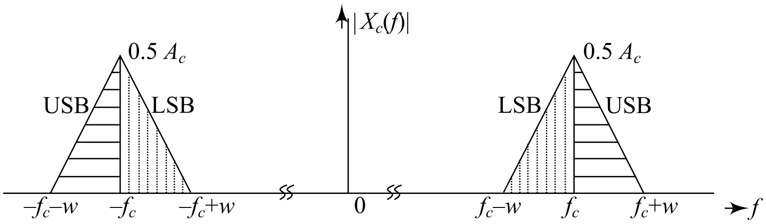


Fig. 4.50 *Amplitude spectrum of a DSB-SC signal*

From the above spectrum of a DSB-SC signal, it is clear that the signal contains both the sidebands and therefore needs a bandwidth of $2W$, just like the AM signal, but the carrier frequency component is not present. Because of this, all the average power of the DSB-SC signal resides in its two sidebands only.

Example 4.26 The modulating signal in an AM-SC system is a multiple-tone signal by $m(t) = A_1\cos\omega_1 t + A_2\cos\omega_2 t + A_3\cos\omega_3 t$. The signal $m(t)$ modulates a carrier $A_c\cos\omega_c t$. Plot the single-sided spectrum and find the bandwidth of the modulated signal. Assume that $\omega_3 > \omega_2 > \omega_1$ and $A_1 > A_2 > A_3$. **(JNTU Sept., 2007)**

Solution AM-SC system is nothing but DSB-SC system. We know that in DSB-SC modulation, the modulated signal is simply the product of the modulating signal $m(t)$ and the carrier. Hence, the modulated signal $x_c(t)$ is given by

$$\begin{aligned}x_c(t) &= m(t)[A_c\cos\omega_c t] = [A_1\cos\omega_1 t + A_2\cos\omega_2 t + A_3\cos\omega_3 t]\,A_c\cos\omega_c t\\ &= A_1A_c\cos\omega_c t\cos\omega_1 t + A_2\,A_c\cos\omega_c t\cos\omega_2 t + A_3\,A_c\cos\omega_c t\cos\omega_3 t\\ &= \frac{1}{2}A_1A_c[\cos(\omega_c+\omega_1)t + \cos(\omega_c-\omega_1)] + \frac{1}{2}A_2\,A_c[\cos(\omega_c+\omega_2)t + \cos(\omega_c-\omega_2)]\\ &\quad + \frac{1}{2}A_3\,A_c[\cos(\omega_c+\omega_3)t + \cos(\omega_c-\omega_3)]\end{aligned}$$

Noting that $\omega_3 > \omega_2 > \omega_1; A_1 > A_2 > A_3$ and $\omega_c >> \omega_1, \omega_2$ and ω_3, we may plot the one-sided spectrum of $x_c(t)$ as shown below.

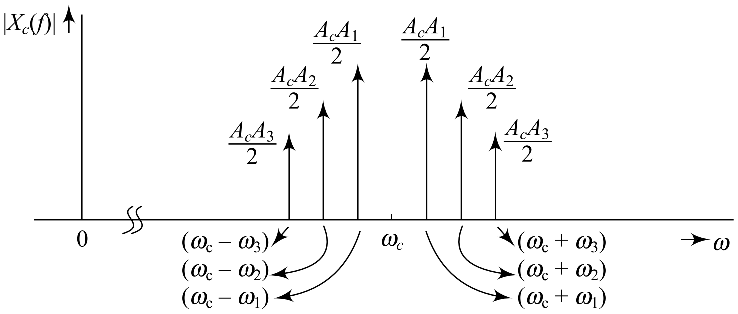


Fig. 4.51

Example 4.27 The signal $x(t) = \text{sinc}(10^5 t)$ is used for DSB-SC modulating a carrier signal having a frequency of 10 MHz. Determine the bandwidth of the modulated signal and sketch its spectrum.

Solution We know that if the modulating signal has a bandwidth of W Hz, then the DSB-SC wave has a bandwidth $2W$ Hz.

To determine W, we take the FT of $x(t)$

$$X(f) = 10^{-5}\Pi(10^{-5}f) = 10^{-5}\Pi(f/10^5)$$

A plot of this is shown below in Fig. 4.52.

Hence, the spectrum of the DSB-SC modulated signal is as shown in Fig. 4.53.

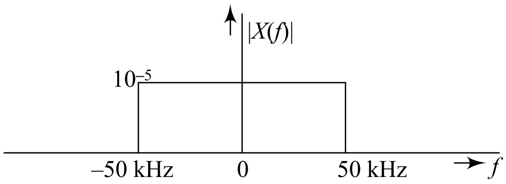


Fig. 4.52

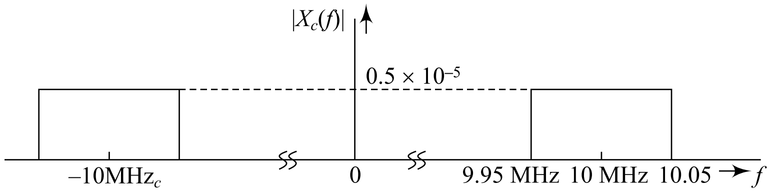


Fig. 4.53 *Spectrum of $x_c(t)$*

4.6.3 Generation of DSB-SC Signals

1. Balanced modulator: We had seen that an AM signal may be written as

$$x_c(t) = A_c[1 + mx(t)]\cos\omega_c t$$

Suppose we now consider two AM signals identical in all respects except that the message signals in the two cases are 180° out of phase. We may write them as

$$x_{c_1}(t) = A_c[1 + mx(t)]\cos\omega_c t \tag{4.61}$$

and

$$x_{c_2}(t) = A_c[1 - mx(t)]\cos\omega_c t \tag{4.62}$$

Subtracting x_{c_2} from x_{c_1}, we have

$$x_{c_3}(t) = x_{c_1}(t) - x_{c_2}(t) = 2mA_c x(t)\cos\omega_c t \tag{4.63}$$

We recognize that $x_{c_3}(t)$, so obtained is a DSB-SC signal. The above analysis suggests that to generate a DSB-SC signal using a carrier signal *c*(*t*) and a message signal, *x*(*t*), we need to have two identical AM generating circuits, to which the carrier *c*(*t*) is applied in the same phase but the message signal is fed 180° out of phase and we take difference of the output signal of the two amplitude modulators. A simple circuit realization of the above is illustrated in Fig. 4.54 in which we have used two identical amplitude modulators using the non-linearity of FETs (see Fig. 4.10).

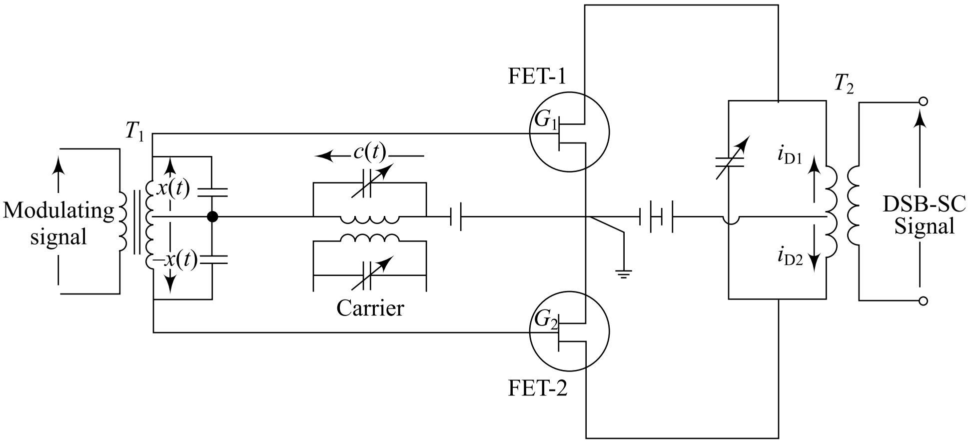


Fig. 4.54 *A balanced modulator using FETs*

As can be seen from the above circuit diagram, the carrier signal is applied to the gates G_1 and G_2 of the two identical FETs in the same phase. The modulating signal, however, is applied to G_1 and G_2 in opposite phase, since the modulating signal developed across the two halves of the secondary of the transformer T_1 will be 180° out of phase with respect to each other.

Suppose the modulating signal is not applied and the carrier alone is applied. Since the carrier signals at G_1 and G_2 are in phase, the carrier components of i_{D1} and i_{D2} which flow in opposite directions through the primary of the output transformer, do not induce any carrier component of voltage on the secondary side of transformer T_2. The carrier is thus eliminated. Because of the symmetry of this circuit, it is called a balanced modulator.

To show that the balanced modulator produces an output signal which is a DSB-SC signal, we proceed exactly in the same way as we did for the analysis of the circuit of the amplitude modulator of Fig. 4.8.

Let $$i_{D_1} = a_0 + a_1 e_{g_1} + a_2 e_{g_1}^2$$

(Non-linear relationship between the gate voltage and the drain current of the FET)

But $$e_{g_1} = x(t) + A_c \cos\omega_c t$$

$$\therefore \quad i_{D_1} = a_0 + a_1[x(t) + A_c \cos\omega_c t] + a_2[x(t) + A_c \cos\omega_c t]^2$$

$$= \left[a_0 + \frac{a_2}{2}A_c^2\right] + a_1 x(t) + a_1 A_c \cos\omega_c t + a_2 x^2(t) + \frac{a_2}{2}A_c^2 \cos 2\omega_c t + 2a_2 x(t) A_c \cos\omega_c t \quad (4.64)$$

Since the two FETs are identical and are operating under identical conditions,

Let $$i_{D_2} = a_0 + a_1 e_{g_2} + a_2 e_{g_2}^2$$

But $$e_{g_2} = -x(t) + A_c \cos\omega_c t$$

$$\therefore \qquad i_{D_2} = \left[a_0 + \frac{a_2}{2}A_c^2\right] - a_1 x(t) + a_1 A_c \cos\omega_c t + a_2 x^2(t) + \frac{a_2}{2}A_c^2 \cos 2\omega_c t - 2a_2 x(t) A_c \cos\omega_c t$$

$$\therefore \qquad (i_{D_1} - i_{D_2}) = 2a_1 x(t) + 4a_2 A_c x(t) \cos\omega_c t$$

The output tank circuit, which is tuned to the carrier frequency, eliminates the first term and a voltage $e_0(t)$ where

$$e_0(t) = kx(t)\cos\omega_c t \tag{4.65}$$

where k is a constant, is induced in the secondary of transformer T_2. $e_0(t)$, being proportional to the product of the modulating signal and the carrier signal, is a DSB-SC signal.

The two FETs will not, in practise, be *exactly* identical. The centre taps on the secondary of the transformer T_1 and the primary of the transformer T_2, may not be *exact* centre taps. The degree of suppression of the carrier depends on to what extent these conditions are met.

2. Ring modulator (Balanced modulator using diodes): One popular type of balanced modulator (especially in telephone circuits), is the ring modulator shown in Fig. 4.55. It consists of four diodes, all pointing in the same direction and forming a ring, because of which it is named as a ring modulator. A square-wave carrier of frequency, say, f_c, is used for switching these diodes, and is applied between the center taps of the secondary of transformer T_1 and the center tap of the primary of transformer T_2, as shown in the figure. First let us assume that modulating signal is absent.

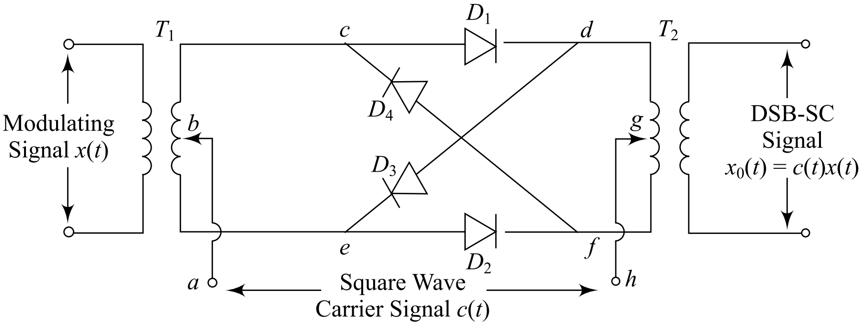


Fig. 4.55 *A ring modulator*

During the positive half-cycle of the carrier wave, let us say when 'a' is positive with respect to 'h', the diodes D_1 and D_2 are forward biased and the diodes D_3 and D_4 are reverse biased. Carrier component of current flows through the paths *a-b-c-d-h* and *a-b-e-f-g-h*. During the next half-cycle, i.e., negative half-cycle of the carrier, diodes D_3 and D_4 are forward biased and diodes D_1 and D_2 are reverse biased. Carrier signal now drives a current through the paths *h-g-d-e-b-a* and *h-g-f-c-b-a*. If the center taps b and g are true center taps and the diodes are identical in their behavior, in both the half-cycles of the carrier, no carrier component of voltage is induced either in the primary of T_1 or in the secondary of T_2 since in each of the half-cycles, equal and opposite currents flow through the secondary winding of T_1 and the primary winding of T_2. Thus, no carrier component will be produced in the output.

It may be noted that just like the case of the DSB-SC modulator of Fig. 4.44, here also, whenever the modulating signal induced in one-half of the secondary of T_1 adds to the carrier, the modulating signal induced in the other half subtracts from the carrier. Thus, DSB-SC signals with carrier frequencies of f_c, $3f_c$, $5f_c$, $7f_c$, etc., are produced in the output because the square wave carrier signal (as shown in Example 2.4) has only the fundamental and odd harmonics. Thus, if the modulating signal has a spectrum as shown in Fig. 4.56 with a bandwidth of W Hz, then the spectrum of the ring modulator output would be as shown in Fig. 4.57.

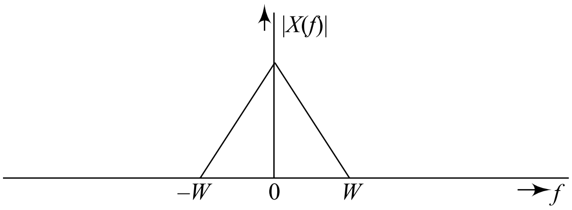


Fig. 4.56 *Spectrum of the modulating signal*

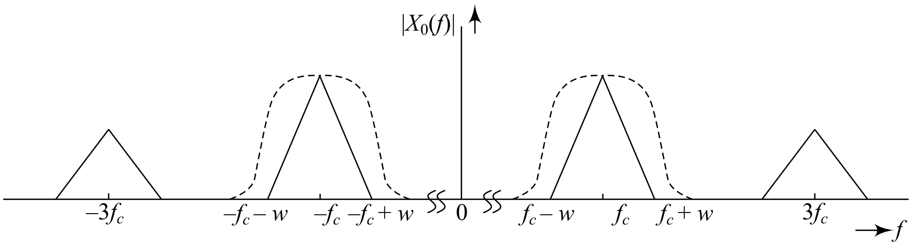


Fig. 4.57 *Spectrum of the output of the ring modulator*

Since we are interested only in the DSB-SC signal corresponding to the fundamental frequency component of the carrier, by using a bandpass filter of center frequency f_c and bandwidth $2W$, at the output of the ring modulator, we can obtain the desired DSB-SC modulated signal.

Example 4.28 For the balanced ring modulator circuit, the carrier input frequency f_c = 500 kHz and the modulating input signal frequency ranges from 0 to 5 kHz. Determine the output frequency range and the output frequency for a single modulating signal input frequency of 3.4 kHz. (**JNTU Sept., 2007**)

Solution Referring to Figs. 4.55 and 4.56, and noting that, as mentioned underneath Fig. 4.57, we are interested in only the fundamental frequency component of the carrier, at the output of the ring modulator, we get only a DSB-SC signal centered on the fundamental carrier frequency component f_c = 500 kHz. This is because of the bandpass filter with center frequency f_c and bandwidth equal to twice the maximum frequency (W) of the modulating signal, used at the output of the ring modulator.

Thus, the output will contain frequency components ranging from (f_c – W) to (f_c + W), i.e., from 495 kHz to 505 kHz.

For a single modulating signal of frequency 3.4 kHz, the output of the ring modulator (with the above mentioned band pass filter) will have only two frequency components (500 – 3.4) kHz and (500 + 3.4) kHz, i.e., 496.6 kHz and 503.4 kHz.

4.6.4 Detection of DSB-SC Signals

In the case of AM, we could use a simple envelope detector to extract the message signal from the modulated signal because the envelope of the modulated signal, in the absence of over-modulation, was found to be a replica of the modulating signal. Figure 4.43(c) clearly shows that such a situation does not exist in the case of DSB-SC modulated signals. For these signals, we go in for coherent or synchronous detection, the basic principle of which we have briefly discussed in Section 4.3.

If $x(t)$ is the modulating signal and $A_c \cos\omega_c t$ is the carrier signal, we know that the DSB-SC modulated signal formed by these two is given by

$$x_c(t) = A_c x(t) \cos\omega_c t$$

If this is the signal received by the receiver, let us say we generate a carrier signal (in the receiver) having exactly the same frequency and phase as possessed by the suppressed carrier of the received signal. Then, we may multiply the received $x_c(t)$ by this locally generated carrier to get

$$x_c(t)\cos\omega_c t = A_c x(t)\cos^2\omega_c t = A_c x(t)\left[\frac{1+\cos 2\omega_c t}{2}\right]$$

$$=\frac{1}{2}A_c x(t)+\frac{1}{2}A_c x(t)\cos 2\omega_c t \qquad (4.66)$$

The first term here is a quantity proportional to $x(t)$ and is hence the desired signal, while the second term is a very high frequency component which can easily be removed by low pass filtering. Synchronous detection, or coherent detection may therefore be represented by the following block diagram:

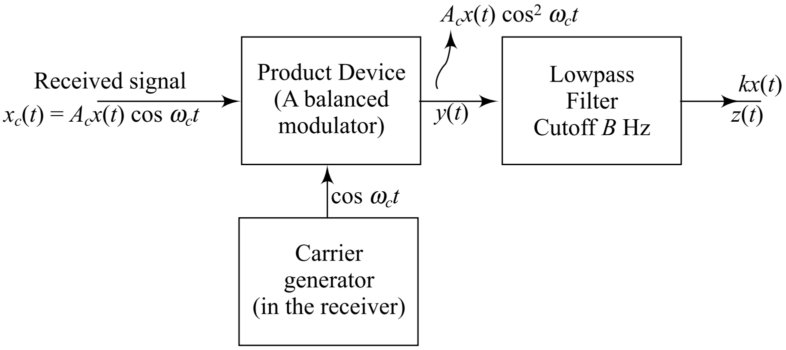


Fig. 4.58 *A coherent detector*

Effect of phase error of the locally generated carrier Let us now examine the effect of any deviation in the phase of the locally generated carrier. So, if the received DSB-SC modulated signal is

$$x_c(t) = A_c x(t)\cos\omega_c t$$

let the locally generated carrier be $\cos(\omega_c t+\theta)$, where θ is the phase error. With reference to Fig. 4.35, the output $y(t)$ of the product device will now be

$$y(t) = A_c x(t)\cos\omega_c t\cdot\cos(\omega_c t+\theta)$$

$$\therefore \qquad y(t)=\frac{1}{2}A_c x(t)\cos(2\omega_c t+\theta)+\frac{1}{2}A_c x(t)\cos\theta \qquad (4.67)$$

The low pass filter following the product device has a cut-off frequency B Hz, where $W \le B \le (2f_c - W)$, W being the bandwidth of $x(t)$. The first term on the RHS of Eq. (4.67) has frequency components around $2f_c$ while the second term is proportional to the modulating signal $x(t)$ provided θ is constant. Hence, the first term gets eliminated by the low pass filter and we have:

$$z(t)=\frac{1}{2}A_c x(t)\cos\theta \qquad (4.68)$$

Thus, the signal output of the coherent detector is $\left(\frac{A_c}{2}\right)x(t)\cos\theta$ instead of $x(t)$. This has the following consequences:

1. Even if θ remains constant, which of course, is not true in practice, the cos θ factor tends to reduce the output message signal.
2. The phase of the received signal goes on varying with time in a random fashion because of the changes in the channel conditions. Thus, the phase deviation from the correct value of the locally generated carrier, namely θ, goes on changing randomly. This random variation of θ and consequently of cos θ, has the effect of producing distortion in the recovered message signal.

The foregoing simple analysis and the subsequent discussion clearly bring out the need to maintain the locally generated carrier signal always in frequency and phase synchronism with the frequency and phase of the suppressed carrier in the received DSB-SC signal. We shall now discuss Costas receiver, or Costas loop and the squaring loop systems which accomplish this task.

Example 4.29 Consider the wave obtained by adding a non-coherent carrier $A_c \cos(\omega_c t + \phi)$ to the DSB-SC wave, $m(t)\cos\omega_c t$, where $m(t)$ is the message waveform. This waveform is applied to an ideal envelope detector. Find the resulting detector output. Evaluate the output for (a) $\phi = 0$, and (b) $\phi \neq 0$ and $|m(t)| << A_c/2$ **(JNTU, May, 2007)**

Solution The input to the envelope detector is given by

$$\begin{aligned} y(t) &= [A_c \cos(\omega_c t + \phi) + m(t)\cos\omega_c t] \\ &= [A_c \cos\phi + m(t)]\cos\omega_c t - [A_c \sin\omega_c t]\sin\omega_c t \end{aligned} \qquad \text{(i)}$$

Hence, $y(t)$ may be put in the polar form $y(t) = R(t)\cos[\omega_c t + \theta(t)]$, where $R(t)$ is the envelope and $\theta(t)$ is the phase angle. It is this $R(t)$ which an ideal envelope detector extracts and gives as output.
Thus,

$$\begin{aligned} y(t) &= R(t)(\cos\omega_c t)\cos\theta(t) - R(t)(\sin\omega_c t)\sin\theta(t) \\ &= [R(t)\cos\theta(t)]\cos\omega_c t - [R(t)\sin\theta(t)]\sin\omega_c t \end{aligned} \qquad \text{(ii)}$$

Comparing Eqs. (i) and (ii),

$$R(t)\cos\theta(t) = [A_c \cos\phi + m(t)]$$

and

$$R(t)\sin\theta(t) = A_c \sin\phi$$

Hence,

$$\begin{aligned} R(t) &= \sqrt{R^2(t)\cos^2\theta(t) + R^2(t)\sin^2\theta(t)} \\ &= \sqrt{[A_c \cos\phi + m(t)]^2 + [A_c \sin\phi]^2} \end{aligned}$$

∴

$$R(t) = \sqrt{A_c^2 + m^2(t) + 2m(t)A_c \cos\phi}$$

and

$$\theta(t) = \tan^{-1}\left[\frac{A_c \sin\phi}{A_c \cos\phi + m(t)}\right]$$

1. When $\phi = 0$

$$R(t) = \sqrt{A_c^2 + 2A_c m(t) + m^2(t)} = \sqrt{[A_c + m(t)]^2} = A_c + m(t)$$

Thus, when $\phi = 0$, the output of the envelope detector is $z(t) = A_c + m(t)$, where A_c is a dc component and $m(t)$ is the message signal.

2. When $\phi \neq 0$, and $|m(t)| << A_c/2$

In this case,

$$R(t) = \sqrt{A_c^2 + m^2(t) + 2m(t)A_c \cos\phi} \approx \sqrt{A_c^2 + 2A_c m(t)\cos\phi}$$

(Since $m^2(t)$ can be neglected in comparison with A_c)

∴

$$R(t) \cong \sqrt{A_c^2 + 2A_c m(t)\cos\phi} = A_c\sqrt{1 + \frac{2m(t)}{A_c}\cos\phi}$$

Now

$$A_c >> 2|m(t)| \qquad \therefore \frac{2m(t)}{A_c}\cos\phi << 1$$

Now, we know that when $x << 1$, $\sqrt{1+x}$ can be approximated by $\left(1+\frac{1}{2}x\right)$

$\therefore$ In this case the output of the envelope detector will be

$$z(t) \cong A_c\left[1+\frac{m(t)}{A_c}\cos\phi\right] = A_c + m(t)\cos\phi$$

Thus the output is again having a dc component of A_c plus the attenuated version (provided ϕ is constant) of the message signal $m(t)$ since $\cos\phi < 1$ (as $\phi \neq 0$). Further, if ϕ varies with time, the message signal $m(t)$ is multiplied $\phi(t)$ and so the message component at the output of the envelope detector will be only a mutilated version of the actual message.

4.6.5 Costas Loop

It consists of two coherent detectors. A voltage controlled oscillator initially adjusted to operate at the correct suppressed carrier frequency, f_c, assumed to be known *a priori*, supplies the 'locally generated carrier' to the two coherent detectors – to one of them directly and to the other through a –90° phase shifter. The former coherent detector which is supplied $\cos\omega_c t$ directly as the locally generated carrier, is called the 'Inphase channel' or I-channel, while the one to which $\sin\omega_c t$ is applied as the local carrier, is called the 'Quadrature channel' or the *Q*-channel. Both the coherent detectors are fed with the same received DSB-SC signal $A_c x(t)\cos\omega_c t$.

Suppose the carrier phase error is zero, i.e., $\theta = 0$. Then the output of the I-channel is $\frac{1}{2}A_c x(t)$ while that of the *Q*-channel is zero. The *I*-channel output is taken as the demodulated signal. Now, suppose there is a carrier phase error of θ. Then the *I*-channel output is $\frac{1}{2}A_c x(t)\cos\theta$ while that of the *Q*-channel is $\frac{1}{2}A_c x(t)\sin\theta$. As shown in Fig. 4.59, both these outputs are fed to the phase discriminator, which consists of a product device followed by a low pass filter. For θ values that are quite small, we know that $\cos\theta \approx 1$ and $\sin\theta \approx \theta$. Thus, the output of the product device in the phase discriminator is of the form $A_c^2 x^2(t)\theta$. The low pass filter, which has a very low cut-off frequency of the order of a few Hertz, gives a dc voltage proportional to θ at its output since variations in θ will be very slow compared to variations in $x^2(t)$.

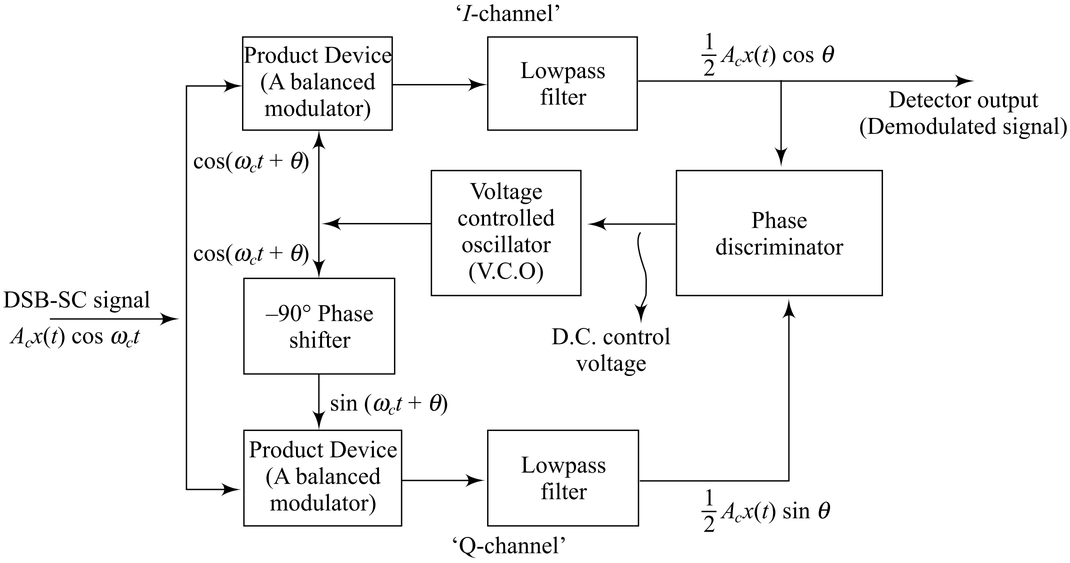


Fig. 4.59 *Costas receiver or Costas loop*

Thus, we obtain a control dc voltage which has the same polarity (positive or negative) as θ and is proportional to it. This changes the VCO output in such a way as to minimize θ by locking it to f_c. The phase error is thus kept very small.

The Costas loop thus provides a good practical solution to the 'phase synchronism' problem encountered in coherent detection. However, it suffers from one major drawback – the 180° phase ambiguity for the demodulated signal, i.e., the output of the loop. To understand what is meant by this 180° phase ambiguity, suppose that the phase of the modulating signal in the DSB-SC signal is reversed so that the received signal is $-A_c x(t)\cos\omega_c t$ instead of $A_c x(t)\cos\omega_c t$. Since the output of the product device in the phase discriminator is given by $A_c^2 x^2(t)\theta$, it is insensitive to the polarity of the modulating signal. Thus, when the loop is working and is locked to the carrier frequency, one cannot be sure whether it has got locked in such a way as to give a demodulated output of $x(t)$ or $-x(t)$. When the $x(t)$ is an audio signal, one need not bother about this 180° phase ambiguity as our ear is not sensitive to it. However, if $x(t)$ is polar data that can take positive and negative values, the phase ambiguity can cause serious problems, as a binary 1 may be detected as a '0' and vice versa. Another disadvantage with Costas loop is that the phase control of the loop ceases if there is no modulation. This is not a serious problem as the lockup establishes very fast.

4.6.6 Squaring Loop

Unlike the Costas loop, the squaring loop extracts the carrier signal of correct frequency and phase from the received DSB-SC signal itself.

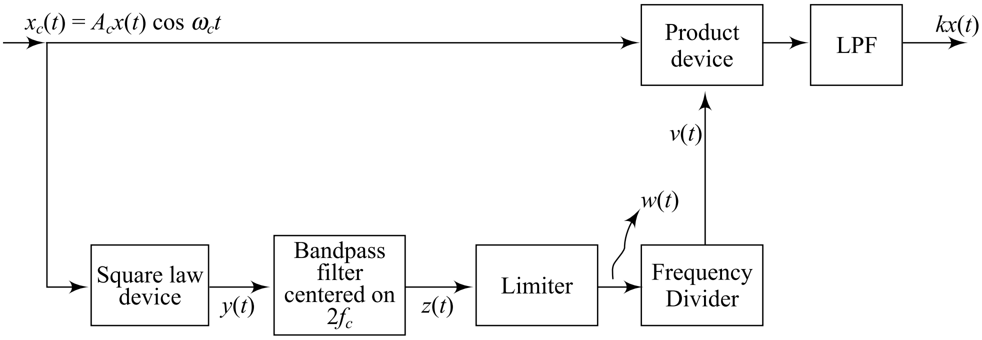


Fig. 4.60 *Squaring loop for carrier recovery*

$$y(t) = x_c^2(t) = A_c^2 x^2(t)\cos^2\omega_c t = \frac{1}{2}A_c^2 x^2(t)[1+\cos 2\omega_c t]$$

$$z(t) = \frac{1}{2}A_c^2 x^2(t)\cos 2\omega_c t$$

The variations with respect to time, of the peak amplitude of $\cos 2\omega_c t$, caused by the multiplication by $x^2(t)$, are removed by the limiter to give an output $w(t)$, where

$$w(t) = k_1\cos 2\omega_c t$$

The frequency divider circuit then gives an output $v(t)$, where

$$v(t) = k_2\cos\omega_c t$$

This $v(t)$, which represents the missing carrier signal correctly in frequency and phase, is then used for coherent detection by multiplying the received DSB-SC signal with it using a product device (a balanced modulator) and then low pass filtering this product using a low pass filter with a cut-off frequency B Hz such that $W < B < (2f_c - W)$, where W Hz is the band limiting frequency of the modulating signal.

Just like the Costas loop, the squaring loop also suffers from the disadvantage of 180° phase ambiguity insofar as the demodulated signal $x(t)$ is concerned.

4.6.7 Quadrature Carrier Multiplexing of DSB-SC Signals (QAM)

Quadrature carrier multiplexing (also called Quadrature Amplitude Modulation or QAM), is a technique which enables us to transmit simultaneously over the same physical channel, two different message signals $x_1(t)$ and $x_2(t)$ having spectra that occupy the same bandwidth, using a single carrier frequency. The carrier signals DSB-SC modulated by the two messages have the same frequency, but differ in phase by 90°. Thus, the modulated signals may be represented by

$$x_{c_1}(t) = A_c x_1(t) \cos \omega_c t$$

and

$$x_{c_2}(t) = A_c x_2(t) \sin \omega_c t$$

We may transmit the multiplexed signal:

$$x_c(t) = x_{c_1}(t) + x_{c_2}(t) = A_c[x_1(t)\cos \omega_c t + x_2(t) \sin \omega_c t]$$

over the channel. This signal $x_c(t)$ occupies a bandwidth of only W Hz, even though $x_1(t)$ and $x_2(t)$ individually have a bandwidth of W Hz each. This is because the spectra of $x_1(t)$ and $x_2(t)$ completely overlap. Although their spectra completely overlap, signal $x_1(t)$ and $x_2(t)$ can be recovered from the multiplexed signal $x_c(t)$ by coherent detection as shown in Fig. 4.61, wherein the balanced modulators act as product devices.

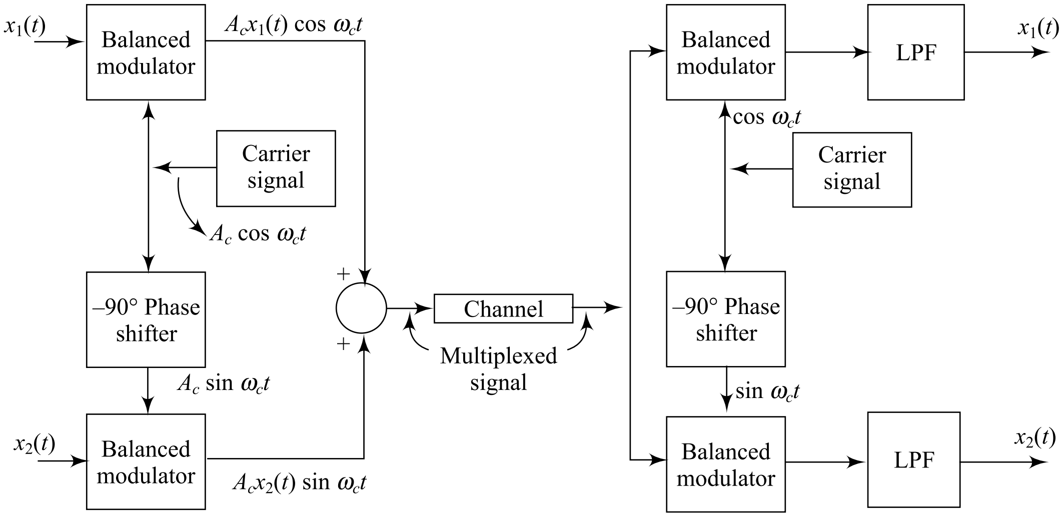


Fig. 4.61 *Quadrature carrier multiplexed system*

At the receiving end, the message signals $x_1(t)$ and $x_2(t)$ are recovered in the following manner:

$$x_c(t) = A_c x_1(t) \cos \omega_c t + A_c x_2(t) \sin \omega_c t$$

$$\therefore \quad x_c(t) \cos \omega_c t = \frac{1}{2} A_c x_1(t)[1 + \cos 2\omega_c t] + \frac{1}{2} A_c x_2(t) \sin 2\omega_c t$$

Subsequent low pass filtering of $x_c(t)\cos \omega_c t$ removes the high frequency component $\frac{1}{2} A_c x_1(t) \cos 2\omega_c t$ as well as $\frac{1}{2} A_c x_2(t) \sin 2\omega_c t$, leaving a signal $\frac{1}{2} A_c x_1(t)$ which is proportional to $x_1(t)$ at the output of the LPF. In a similar way, $x_2(t)$ is obtained by multiplying $x_c(t)$ by $\sin \omega_c t$ and then low pass filtering the product.

It is, of course, necessary that the $\cos\omega_c t$ generated in the receiver be in frequency and phase synchronism with the missing carrier in the multiplexed signal that is received. For this purpose, a Costas receiver may be used, or else, a low-level pilot carrier may be transmitted along with the multiplexed signal.

Quadrature carrier multiplexing reduces the number of subcarriers used besides reducing the bandwidth requirement of the multiplexed signal.

Example 4.30 What is the effect of a frequency error $\Delta\omega$ in the angular frequency of the locally generated carrier on the coherently demodulated signal in the case of DSB-SC?

Solution Let the received DSB-SC signal be $x_c(t) = A_c(\cos\omega_c t)x(t)$.
Let the locally generated carrier be $\cos(\omega_c + \Delta\omega)t$. Then, output of the product device is (refer to Fig. 4.52)

$$y(t) = A_c x(t)\cos\omega_c t\cdot\cos(\omega_c + \Delta\omega)t$$

$$= \frac{1}{2}A_c x(t)[\cos(\Delta\omega)t + \cos(2\omega_c + \Delta\omega)t]$$

and $$z(t) = \frac{1}{2}A_c x(t)\cos(\Delta\omega)t \text{ = Demodulated signal.}$$

$\Delta\omega$ will generally be quite small compared to ω_c; but it can be comparable to W, the highest frequency component in $x(t)$. Thus, a beat frequency is produced, giving rise to serious distortion.

4.7 SINGLE SIDEBAND MODULATION

In the previous section, we had discussed in detail about DSB-SC modulation. Because of the absence of any carrier component in the modulated signal, the DSB-SC, of course, offers some saving of power. However, both the sidebands are present although, as stated earlier, from the point of transmission of information one would have sufficed. Thus, it does not offer the maximum possible power saving. Moreover, as both the sidebands are present, it requires a bandwidth of $2W$, i.e., twice the maximum frequency in the message signal, same as in AM.

So we shall proceed to the next logical step of suppressing not only the carrier, but also one of the sidebands, so as to maximize the saving in transmitted power as well as bandwidth required for transmission. This leads us to what is called the Single Sideband Suppressed Carrier or SSB-SC modulation.

4.7.1 Frequency-domain and Time-domain Representation of SSB-SC Signals

In Fig. 4.50, we had sketched the amplitude spectrum of a typical DSB-SC signal. Figure 4.62 shows the same with a scaled version of the amplitude spectrum of the message signal itself superimposed on it.

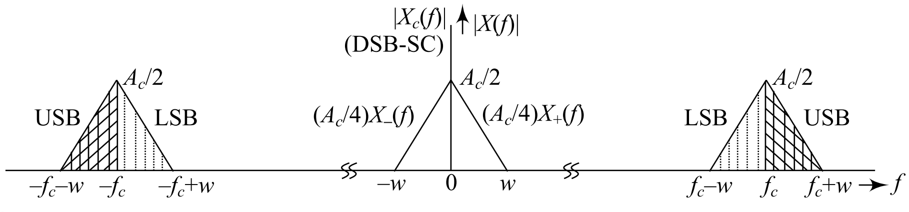


Fig. 4.62 *Amplitude spectrum of DSB-SC signal and Amplitude spectrum of the message signal (scaled)*

From Fig. 4.62, we may draw the spectra of the USSB-SC signal, i.e., the SSB-SC signal in which only the upper sideband is present, and of the LSSB-SC signal in which only the lower sideband is present, as shown in Fig. 4.63(a) and (b), respectively.

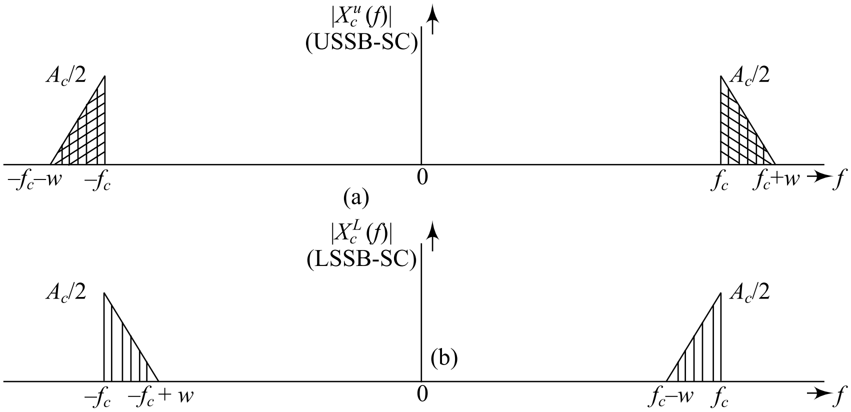


Fig. 4.63 *(a) Spectrum of a USSB-SC signal, (b) Spectrum of a LSSB-SC signal*

The message spectrum shown in Fig. 4.63 may be visualized as the sum of $\frac{A_c}{4}X_+(f)$ and $\frac{A_c}{4}X_-(f)$ where $\frac{A_c}{4}X_+(f)$ is the positive frequency part and $\frac{A_c}{4}X_-(f)$ is the negative frequency part (From Eq. (2.154) of Section 2.8). We know that

$$\mathcal{F}^{-1}\left[\frac{A_c}{4}X_+(f)\right]=\frac{A_c}{4}X_+(t) \tag{4.69}$$

and
$$\mathcal{F}^{-1}\left[\frac{A_c}{4}X_-(f)\right]=\frac{A_c}{4}X_-(t) \tag{4.70}$$

where
$$x_+(t)=x(t)+j\hat{x}(t) \tag{4.71}$$

is the pre-envelope of $x(t)$ for positive frequencies,
and

$$x_-(t)=x(t)-j\hat{x}(t) \tag{4.72}$$

is the pre-envelope of $x(t)$ for negative frequencies.
Then, from Fig. 4.62, we may write
If spectrum of the USSB-SC signal $=X_c^u(f)$,

then
$$X_c^u(f)=\frac{A_c}{4}[X_+(f-f_c)+X_-(f+f_c)] \tag{4.73}$$

Taking the inverse Fourier transform on both sides of the above, we get

$$\begin{aligned}x_c^u(t)=\text{USSB-SC signal}&=\frac{A_c}{4}x_+(t)e^{-j\omega_c t}+\frac{A_c}{4}x_-(t)e^{j\omega_c t}\\&=\frac{A_c}{4}[x(t)+j\hat{x}(t)]e^{-j\omega_c t}+\frac{A_c}{4}[x(t)-j\hat{x}(t)]e^{+j\omega_c t}\\&=\frac{A_c}{4}\left[x(t)\left\{e^{j\omega_c t}+e^{-j\omega_c t}\right\}\right]-j\frac{A_c}{4}\left[\hat{x}(t)\left\{e^{j\omega_c t}-e^{-j\omega_c t}\right\}\right]\end{aligned}$$

$\therefore$
$$x_c^u(t)=\frac{A_c}{2}[x(t)\cos\omega_c t-\hat{x}(t)\sin\omega_c t] \tag{4.74}$$

Equation (4.74) represents the general form of a USSB-SC signal. The corresponding expression for an LSSB-SC signal may be derived by proceeding in a similar way and the result is

$$\therefore \qquad x_c^L(t) = \frac{A_c}{2}[x(t)\cos\omega_c t + \hat{x}(t)\sin\omega_c t] \qquad (4.75)$$

Example 4.31 A carrier $c(t) = A_c\cos\omega_c t$ is USSB-SC modulated by a modulating signal $x(t) = \cos\omega_m t$. Write down the expression for the modulated signal. Sketch its spectrum.

Solution From Eq. (4.73), we have

$$\begin{aligned} x_c^u(t) &= \frac{A_c}{2}[x(t)\cos\omega_c t - \hat{x}(t)\sin\omega_c t] \\ &= \frac{A_c}{2}[\cos\omega_m t\cdot\cos\omega_c t - \sin\omega_m t\cdot\sin\omega_c t] \\ &= \frac{A_c}{2}\cos(\omega_c + \omega_m)t \end{aligned}$$

$$\therefore \qquad x_c^u(t) = \frac{A_c}{4}[\delta\{f-(f_c+f_m)\} + \delta\{f+(f_c+f_m)\}]$$

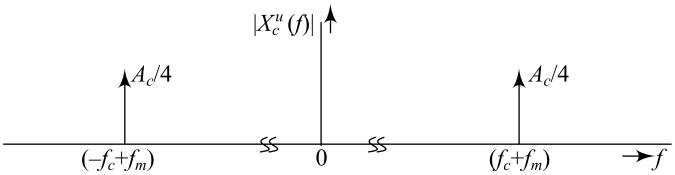


Fig. 4.64 *Spectrum of the USSB-SC modulated signal of Example 4.31*

4.7.2 Methods of Generation of SSB-SC Signals

There are *mainly* two methods of generation of SSB-SC modulated signals. These are:

1. Filter method or balanced modulator-filter method: In this method, we first generate a DSB-SC signal and then filter out from it the unwanted sideband.

2. Phasing method: This method of generation of SSB-SC signals is based on direct implementation of Eqs. 4.73 and 4.74 depending on whether a USSB-SC signal or LSSB-SC signal is needed.

There is also another method, known as the '*Third Method*' or the '*Weaver's Method*', which is a variant of the phasing method.

Filter method of generating SSB-SC signals As mentioned earlier, in this method, a DSB-SC signal is first generated using a balanced modulator and the unwanted sideband is suppressed using an appropriate filter. Though the method may appear very simple and straightforward from the above description, there are some practical difficulties one encounters while implementing the filtering which will be discussed now.

Suppose the modulating signal $x(t)$ has a spectrum as shown in Fig. 4.65. Then the DSB-SC signal will have a spectrum as shown in Fig. 4.66.

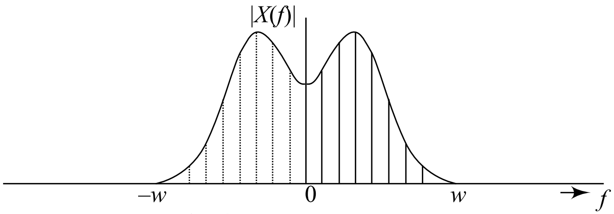


Fig. 4.65 *Spectrum of x(t)*

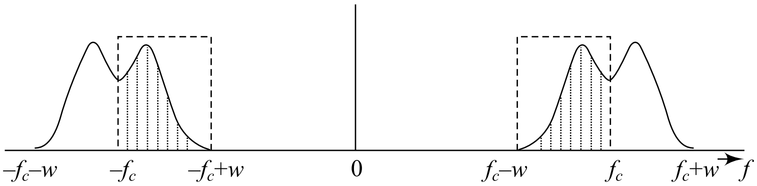


Fig. 4.66 *Spectrum of the DSB-SC signal. The dotted lines show the pass band of the filter for obtaining LSSB-SC signal*

To obtain an LSSB-SC signal from the DSB-SC signal, the pass band of the filter must extend from ($f_c - W$) to $+f_c$ and must suddenly change over to the stop band without any transition band, if the unwanted (USSB) sideband is to be fully removed and if the desired (LSSB) sideband is to suffer no distortion. However, we know that such a filter cannot be realized in practice. If the above two conditions of removing the unwanted sideband fully and causing no distortion to the desired sideband are to be fulfilled using a practical filter with a finite transition band, then it is easy to see that the spectrum of the modulating signal should have a gap near the zero frequency, i.e., it should have a spectrum $X(f)$ whose shape is somewhat as shown in Fig. 4.67.

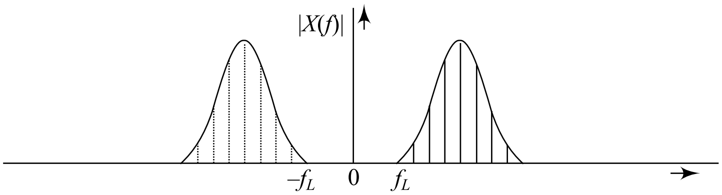


Fig. 4.67 *Spectrum of a modulating signal with an energy gap*

with a gap between $-f_L$ and $+f_L$. With this type of modulating signal, the DSB-SC signal will have a spectrum as shown in Fig. 4.68. As may be seen from the figure, it is now possible to make use of a practical narrow band pass filter with a finite transition bandwidth for suppressing the unwanted upper sideband of the DSB-SC signal.

The filter – a narrow band pass filter, has to have almost constant gain over a bandwidth W covering the lower sideband and can have a transition bandwidth from $(f_c - f_L)$ to $(f_c + f_L)$, i.e., a bandwidth of $2f_L$.

Fortunately, voice signals have practically no energy up to about 300 Hz, i.e., these signals possess a spectrum of the type shown in Fig. 4.67, with f_L = 300 Hz. However, if f_c is say 10 MHz, the transition bandwidth of the filter, which is now $2f_L$ = 600 Hz, will be extremely small compared to f_c. Hence an extremely high value of Q is needed for the filter. To overcome this difficulty, a very low carrier frequency, like 100 kHz, is used for generating the DSB-SC signal so that the required Q value of the filter is practically attainable at least with crystal filters. After suppressing the unwanted sideband, the carrier frequency is raised to the required level by mixing this SSB-SC signal of a low frequency carrier with a high frequency signal generated by a crystal oscillator, as shown in the block diagram of Fig. 4.69.

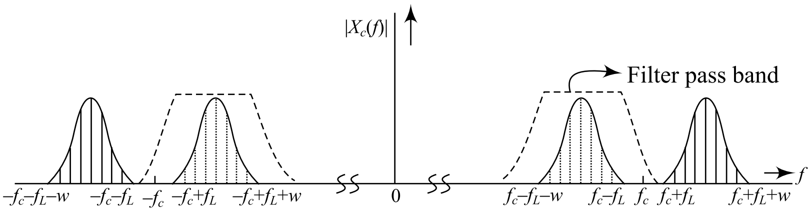


Fig. 4.68 *Spectrum of a DSB-SC signal when the modulating signal has a gap in its spectrum, as shown in Fig. 4.40. Filter pass band is shown in dotted lines*

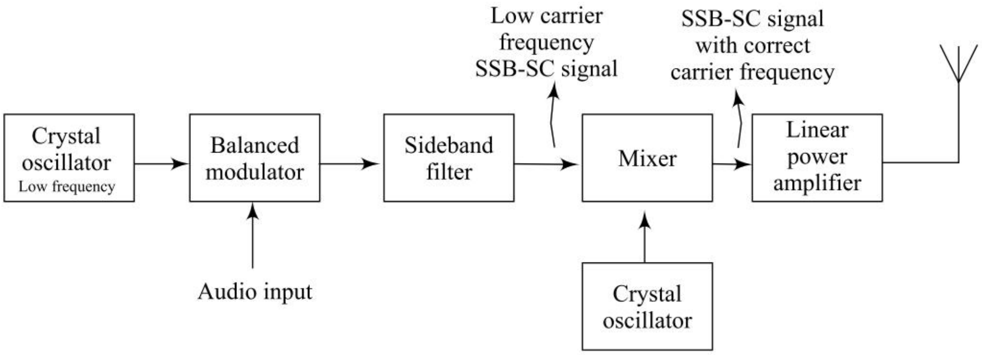


Fig. 4.69 *Block diagram of an SSB-SC transmitter*

With regard to the above block diagram, the following points may be noted.

1. For changing over from LSSB-SC to USSB-SC signals, the sideband filter is not changed; instead, a different crystal is used in the crystal oscillator used for generation of the low frequency carrier.
2. After raising the carrier frequency to the required level, the signal power is raised to the required level by using class-A or class-AB linear power amplifiers.
3. The sideband filter must *attenuate* the unwanted sideband at least up to 60 dB relative to the desired sideband.

Alternatively, we may use a two-stage SSB-SC modulator in order to overcome the problem with the design of the sideband suppression filter. A block diagram showing the essential details of this method is given in Fig. 4.70.

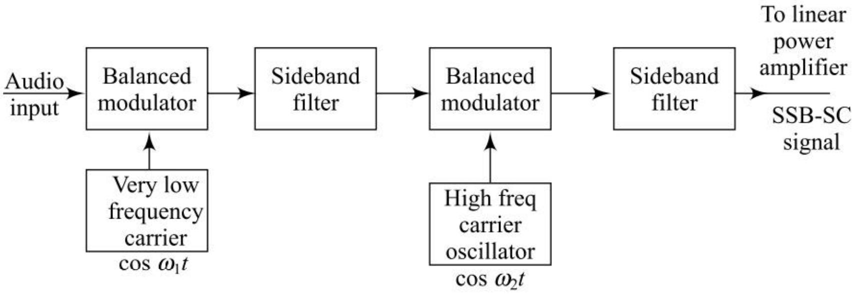


Fig. 4.70 *A two-stage SSB-SC modulator*

The first carrier frequency f_1 is chosen to be very low so that the design of the first sideband filter is simplified. The SSB-SC signal from the first stage, which is now the baseband signal for the second stage, has a gap of approximately $2f_1$ Hz in its spectrum and so the design of the second filter also does not cause any problem.

Phasing method of generating SSB-SC signals This method is based on direct implementation of Eq. 4.74 (or Eq. 4.75) which gives the time-domain representation of USSB-SC (or LSSB-SC) signal. To produce $x(t)\cos\omega_c t$, we need one balanced modulator to which we have to feed the modulating signal $x(t)$ and the carrier oscillator output, viz. $\cos\omega_c t$, directly, as shown in Fig. 4.71. To produce $\hat{x}(t)\sin\omega_c t$, we need to have a second balanced modulator, to which we apply $\hat{x}(t)$ obtained by passing $x(t)$ through a –90° phase shifter and $\sin\omega_c t$ obtained by passing the carrier oscillator output through a –90° phase shifter.

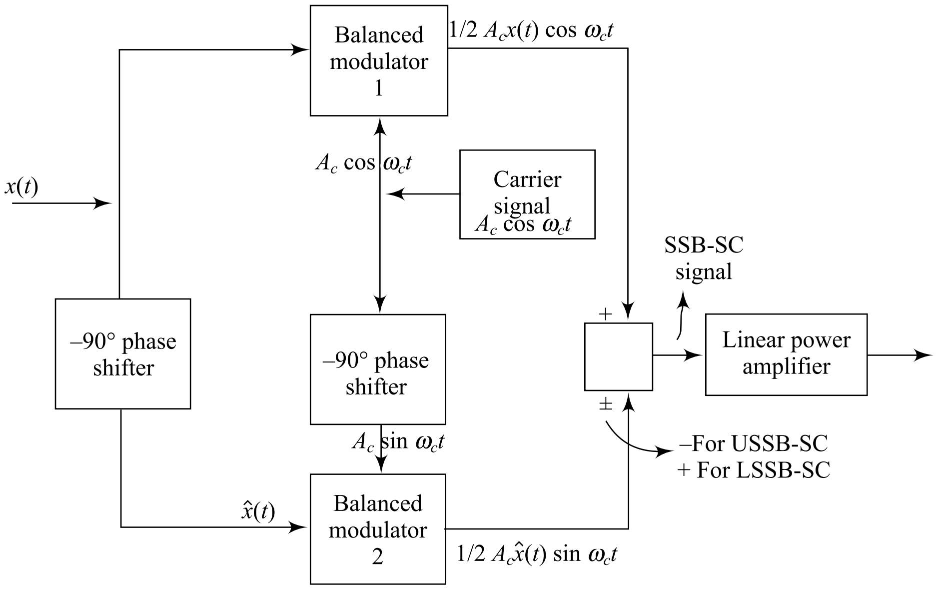


Fig. 4.71 *Phasing method of generating SSB-SC signal*

The carrier being a single frequency signal, the –90° phase shifter for it is a very simple circuit. But the modulating signal $x(t)$ will have several frequency components in it and hence the –90° phase shifter used for it should produce an exact –90° phase shift for every frequency component and further it should have the same gain for all these frequency components, i.e., it should be a Hilbert transformer. This is a complex circuit and is generally expensive.

Comparison of filter method and phasing method

1. The filter method needs costly sideband suppression filters. Although the phasing method does not need these filters, it needs wideband –90° phase shifters which are not easy to realize.
2. Being stable, the filter method does not need constant attention and adjustment. Phasing method, however, needs constant adjustments.
3. In the filter method, the unwanted sideband is suppressed quite effectively; almost to –60 dB relative to the desired sideband. In phasing method suppression of the unwanted sideband is not that effective. This is because of the wideband phase shifters not producing exact –90° phase shift for all frequencies. A deviation of even 2° in the phase shift from the ideal –90° would cause that particular side frequency to be suppressed only to about 20 to 25 dB relative to the corresponding side frequency in the desired sideband.
4. In the filter method, it is not very easy to change from USSB-SC to LSSB-SC and vice versa. In the phasing method, however, it is quite easy to change over from USSB-SC to LSSB-SC and vice versa.
5. Changing the carrier frequency in the case of the filter method is cumbersome as it involves changing the sideband suppression filters and crystals in the local oscillators in the mixer stages, and then re-tuning all the stages. In the phasing method, it is quite easy to change the carrier frequency.
6. The filter method can be successfully implemented only for modulating signals having a gap of a few hundred hertz near the origin, in their spectra. There is, however, no such restriction in the case of the phasing method.

Weaver's method This method, a variant of the phasing method, was invented in 1950 by D.K. Weaver. It avoids the need for wideband phase shifters, which are difficult to construct and are expensive, and instead, uses an AF subcarrier at an audio frequency, say f_0.

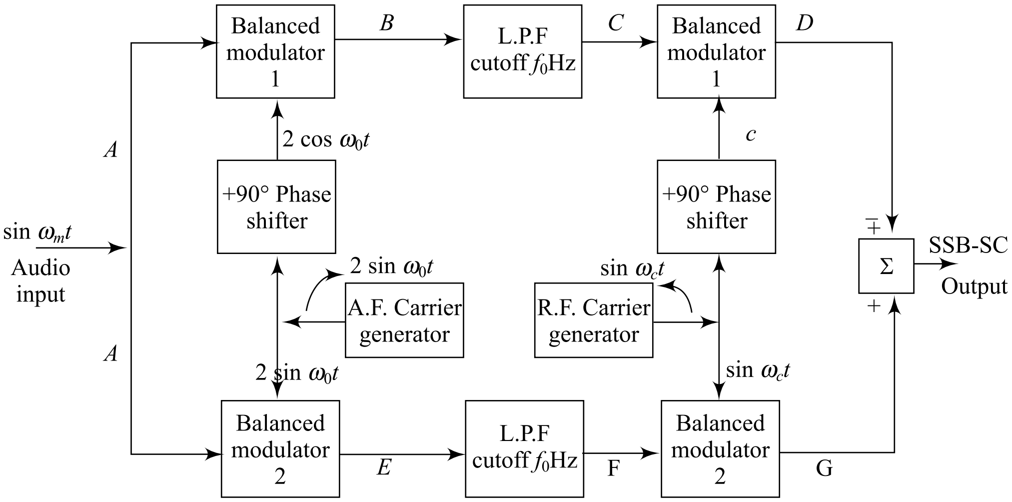


Fig. 4.72 *Weaver's method of generation of SSB-SC signals*

A: $\sin\omega_m t$

B: $2\sin\omega_m t\sin\omega_0 t$

C: $\sin(\omega_m-\omega_0)t$

D: $\sin(\omega_m-\omega_0)t\cos\omega_c t$

$$=\frac{1}{2}[\sin(\omega_c-\omega_0+\omega_m)t-\sin(\omega_c+\omega_0-\omega_m)t]$$

E: $2\sin\omega_m t\sin\omega_0 t$

F: $\cos(\omega_m-\omega_0)t$

G: $\sin\omega_c t\cos(\omega_m-\omega_0)t$

$$=\frac{1}{2}[\sin(\omega_c+\omega_0-\omega_m)t+\sin(\omega_c-\omega_0+\omega_m)t]$$

$\therefore$ D + G: $\sin(\overline{\omega_c-\omega_0}+\omega_m)$

–D + G: $\sin(\overline{\omega_c+\omega_0}+\omega_m)$

Thus, D + G gives USSB-SC signal with (f_c-f_0) as the carrier frequency; and –D + G gives LSSB-SC signal with (f_c+f_0) as the carrier frequency.

Advantages of Weaver's Method

1. No need for any sideband suppression filters.
2. No need for any wideband phase shifters.
3. As the phase shifters used are for a single frequency, they are extremely simple and inexpensive.
4. No need for frequent adjustments.
5. Easy to change over from USSB-SC to LSSB-SC and vice versa.

4.7.3 Detection of SSB-SC Signals

SSB-SC signals can be demodulated using coherent detection, as shown in Fig. 4.73.

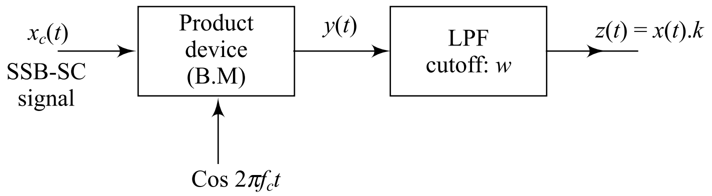


Fig. 4.73 *Coherent detection of SSB-SC signals*

Let the SSB-SC signal be $x_c(t)$. Then from Eqs. (4.74) and (4.75),

$$x_c(t) = \frac{A_c}{2} x(t)\cos\omega_c t \pm \hat{x}(t)\sin\omega_c t\cos\omega_c t$$

$$\therefore \quad y(t) = x_c(t)\cos\omega_c t$$

$$= \frac{A_c}{2} x(t)\cos^2\omega_c t \pm \frac{A_c}{2}\hat{x}(t)\sin\omega_c t\cos\omega_c t$$

$$= \frac{A_c}{4} x(t) + \frac{A_c}{4}\cos 2\omega_c t \pm \frac{A_c}{4}\hat{x}(t)\sin 2\omega_c t$$

The second and third terms are high frequency terms and will be rejected by the LPF whose cut-off frequency is W Hz, the band-limiting frequency of $x(t)$.

Hence,
$$z(t) = \frac{1}{4} A_c x(t) = k \cdot x(t) \tag{4.76}$$

The above analysis assumes that the locally generated carrier signal used for feeding to the product device, is in phase and frequency synchronism with the missing carrier component of the received SSB-SC signal.

A frequency-domain interpretation of coherent detection of SSB-SC signals is given in Fig. 4.74(a) and (b).

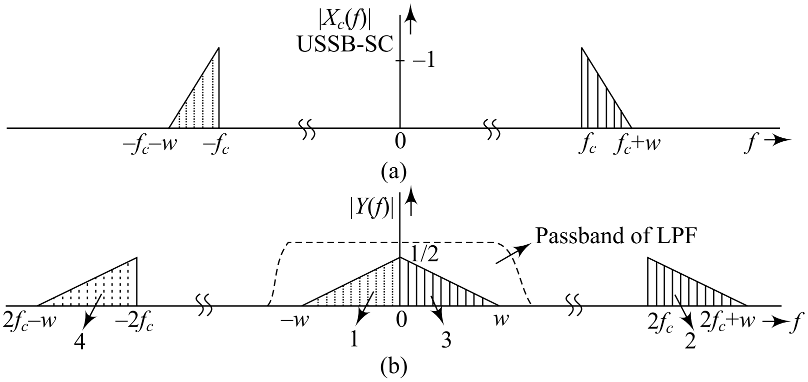


Fig. 4.74 *Frequency-domain interpretation of coherent detection of a USSB-SC signal*

Figure 4.74(a) shows the spectrum of the USSB-SC signal, $x_c(t)$.

Now,
$$y(t) = x_c(t)\cdot\cos 2\pi f_c t$$

$$\therefore \quad Y(f) = X_c(f) * \frac{1}{2}[\delta(f - f_c) + \delta(f + f_c)]$$

$$= \frac{1}{2} X_c(f - f_c) + \frac{1}{2} X_c(f + f_c) \tag{4.77}$$

1 and 2 of Fig. 4.74(b) represent $\frac{1}{2}X(f-f_c)$ while 3 and 4 represent $\frac{1}{2}X(f+f_c)$. Thus, Fig. 4.74(b) represents $Y(f)$. The part of this from $f = -W$ to $f = +W$, which represents the spectrum of the modulating signal, $x(t)$, can be separated from the rest of the spectrum $Y(f)$ by using an LPF whose gain is constant in its pass band from $-W$ to W. In time domain terms, this amounts to extracting $x(t)$ from $y(t)$.

As mentioned earlier, the time-domain analysis as well as the frequency-domain analysis of coherent detection which show that the message signal, $x(t)$, can be recovered without any distortion, assume that the locally generated carrier signal used in the coherent detection process, is in frequency and phase synchronism with the missing carrier in the received SSB-SC signal.

4.7.4 Effect of Phase and Frequency Errors of the Local Carrier

1. When the local carrier has a phase error θ:

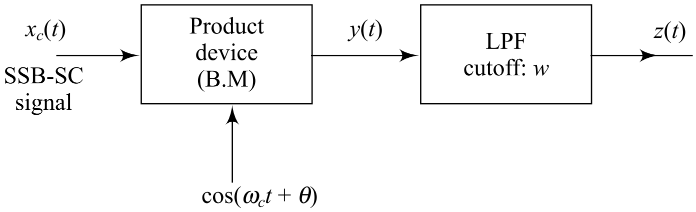


Fig. 4.75 *Effect of phase error of the local carrier*

Let the received SSB-SC signal be represented by

$$x_c(t) = \frac{A_c}{2}[x(t)\cos\omega_c t \mp \hat{x}(t)\sin\omega_c t]$$

whereas we know, the minus sign applies for USSB-SC signals and plus sign for the LSSB-SC signals

$$\therefore \quad y(t) = x_c(t)\cdot\cos(\omega_c t+\theta)$$

$$= \frac{A_c}{4}[x(t)\{\cos(2\omega_c t+\theta)+\cos\theta\}] \mp \frac{A_c}{4}[\{\hat{x}(t)\sin 2\omega_c t\cos\theta\} - \{\hat{x}(t)\overline{1-\cos 2\omega_c t}\sin\theta\}]$$

So, after low pass filtering

$$z(t) = \frac{A_c}{4}[\cos\theta x(t) \pm (\sin\theta)\hat{x}(t)]$$

Note that in this equation, the plus sign applies for USSB-SC.
Taking Fourier transform on both sides

$$Z(f) = \frac{A_c}{4}[X(f)\cos\theta \pm (-j\,\mathrm{sgn}\,f)X(f)\sin\theta]$$

But
$$\mathrm{sgn}\,f = \begin{cases} 1 & \text{for } f > 0 \\ -1 & \text{for } f < 0 \end{cases}$$

$\therefore$ on simplification,

$$Z(f) = \frac{A_c}{4}X(f)e^{\pm j\theta} \tag{4.78}$$

where the negative sign applies for USSB.

Equation (4.78) tells us that all frequency components of $x(t)$ suffer a *constant* phase shift of θ irrespective of their frequency. Obviously it would lead to phase distortion. Also, as θ varies randomly with time due to channel variations, it means that all frequency components suffer the same phase shift which goes on varying randomly with time. This type of severe phase distortion may not be of much concern insofar as audio signals are concerned, since the human ear is not sensitive to phase distortion. But if $x(t)$ is a video signal, phase distortion cannot be tolerated at all as our eyes are very sensitive to any phase changes.

2. When the local carrier oscillator has a frequency error: Let the local carrier oscillator have a frequency $(f_c + \Delta f)$ where f_c is the frequency of the missing carrier signal in the SSB-SC signal. Then referring to Fig. 4.75, we will have $\cos 2\pi(f_c + \Delta f)t$ in the place of $\cos(\omega_c t + \theta)$ shown therein.
In that case,

$$y(t) = \frac{1}{2}A_c \cos 2\pi(f_c + \Delta f)t[x(t)\cos 2\pi f_c t \pm \hat{x}(t)\sin \omega_c t]$$

$$= \frac{1}{4}A_c x(t)[\cos 2\pi(\Delta f)t + \cos 2\pi(2f_c + \Delta f)t] - \frac{1}{4}A_c \hat{x}(t)[\sin 2\pi(2f_c + \Delta f)t - \sin 2\pi(\Delta f)t]$$

Since the LPF has a cut-off frequency $W << f_c$, we have

$$z(t) = \frac{1}{4}A_c[x(t)\cos 2\pi(\Delta f)t \pm \hat{x}(t)\sin 2\pi(\Delta f)t] \tag{4.79}$$

(With positive sign for USSB)

This is a very interesting result because we now find from the above equation that $z(t)$ is not $x(t)$ at all. Far from being so, it is actually a SSB-SC signal for which $x(t)$ is the modulating signal and (Δf) is the carrier signal. From Eq. (4.79), it is clear that when (Δf) is close to zero, $z(t)$ is approximately proportional to $x(t)$. In fact, a frequency error of more than a few Hz results in unacceptable levels of distortion in the output of the coherent detector. This places severe constraint on the local oscillator generating the carrier.

For this reason, sometimes a pilot carrier at low power level is inserted into the SSB-SC signal before it is transmitted. At the receiving end, a technique, referred to as 'Homodyne Detection', is resorted to. This is shown in Fig. 4.76.

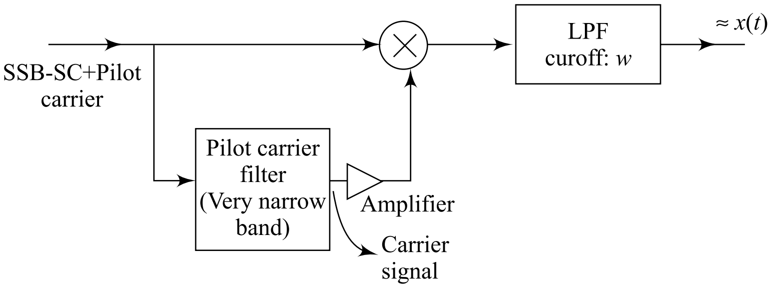


Fig. 4.76 *Homodyne detection*

Equation (4.79) shows that if the transmitted signal is a USSB-SC signal and (Δf) is positive, then the detected signal is an LSSB-SC signal with $x(t)$ SSB-SC modulating the (Δf). Alternatively, if the transmitted signal is an LSSB-SC signal and (Δf) is negative, then also the detected signal is an LSSB-SC signal. If there is an energy gap in the spectrum of the modulating signal $x(t)$, as would be the case if $x(t)$ is a speech signal, then the effect of LSSB-SC modulation of (Δf) by $x(t)$ is to reduce all frequency components of the speech signal $x(t)$ by (Δf). The effect of this is to reduce the energy gap in the spectrum of detected signal. On the other hand, if the transmitted signal is an LSSB-SC signal and (Δf) is positive (or USSB-SC is transmitted and (Δf) is negative), from Eq. (4.79), we find that the demodulated signal is a USSB-SC signal with (Δf) as carrier and $x(t)$ as the modulating signal. If $x(t)$ is a speech signal, this amounts to increasing the frequency of all frequency components of $x(t)$ by (Δf). This manifests as an increase in the energy gap of $x(t)$ obtained as the detected signal. For speech signals, this does not cause very severe distortion provided (Δf) is less than about ±10 Hz. In the case of music, translation in frequency of all the frequency components will result in severe distortion and therefore even if (Δf) is less than ±10 Hz, it will still be unacceptable. In the case of video signal there will be no energy gap at all. Hence the detected signal will be a highly distorted version of the original modulating signal $x(t)$ and this cannot be tolerated at all.

Example 4.32 A synchronous detection of SSB signal shows phase and frequency discrepancy. Consider that

$$s(t)=\sum_{i=1}^{N}[\cos\omega_c t\cos(\omega_i t+\phi_i)-\sin\omega_c t\sin(\omega_i t+\phi_i)]$$

is an SSB signal. This signal is multiplied by the locally generated carrier $\cos\omega_c t$ and then passed through a low pass filter.

(a) Prove that the modulating signal can be completely recovered if the cut-off frequency of the filter is $f_N < f_0 < 2f_c$.

(b) Determine the recovered signal when the multiply signal is $\cos[\omega_c t+\phi]$.

(c) Determine the recovered signal when the multiplying signal is $\cos[(\omega_c+\Delta\omega)t]$, given $\Delta\phi << \phi_i$, where $\omega_c = 2\pi f_c$ and $\Delta\omega = 2\pi\Delta f$ **(JNTU, Sept., 2007)**

Solution Synchronous detection is another name for coherent detection. Although it has not been explicitly mentioned, this question assumes that $f_N > f_{N-1} > f_{N-2} .... > f_1$, where f_1 to f_N are the frequencies of the N single-tone modulating signals, whose sum, viz.,

$$x(t)=\sum_{i=1}^{N}\cos(\omega_i t+\phi_i)$$

is the modulating signal for the given SSB-SC signal.

(a) *When the local carrier oscillator has no frequency or phase error, i.e., it is* $\cos\omega_c t$ When we use this for coherent detection, $s(t)$ is multiplied by this $\cos\omega_c t$ and then the product is low pass filtered.

$$s(t)\cos\omega_c t=\sum_{i=1}^{N}[\cos^2\omega_c t\cos(\omega_i t+\phi_i)-\sin\omega_c t\cos\omega_c t\sin(\omega_i t+\phi_i)]$$

Replacing $\cos^2\omega_c t$ by $\frac{1}{2}[\cos 2\omega_c t+1]$ and $\sin\omega_c t\cos\omega_c t$ by $\frac{1}{2}[\sin 2\omega_c t]$

$$s(t)\cos\omega_c t=\frac{1}{2}\sum_{i=1}^{N}\{\cos(\omega_i t+\phi_i)[1+\cos 2\omega_c t]-\sin(\omega_i t+\phi_i)[\sin 2\omega_c t]\}$$

When we low pass filter this using an LPF whose cut-off frequency f_0 is greater than the highest modulating signal frequency f_N but less than $2f_c$, we get output of the coherent detector

$$=z(t)=\frac{1}{2}\sum_{i=1}^{N}\cos(\omega_i t+\phi_i)=x(t)$$

Since all the other terms represent close to $2f_c$ and are not passed by the LPF.

Thus, the modulating signal can be completely recovered in this case.

(b) *When the multiplying signal (i.e., local carrier signal) is* $\cos[\omega_c t+\phi]$ Proceeding exactly as in the above case,

$$s(t)\cos\omega_c t=\frac{1}{2}\sum_{i=1}^{N}[\cos(\omega_i t+\phi_i)\{\cos(2\omega_c t+\phi)+\cos\phi\}-[\sin(\omega_i t+\phi_i)]\{\sin(2\omega_c t+\phi)-\sin\phi\}]$$

When this is low pass filtered by an LPF whose cut-off frequency f_0 is such that $f_N < f_0 < 2f_c$

$$z(t) = \text{output of the filter } = \frac{1}{2}\sum_{i=1}^{N}\{\cos\phi\cos(\omega_i t+\phi_i)+\sin\phi\sin(\omega_i t+\phi_i)\}$$

This $z(t)$ can be shown to be $x(t)$ with all its frequency components given a constant phase shift of ϕ. Thus, there will be severe phase distortion.

(c) *When the multiplying signal is* $\cos[(\omega_c+\Delta\omega)t]$ Proceeding exactly as in the above two cases, it can be shown that

$$z(t) = \text{output of the filter } = \frac{1}{2}\sum_{i=1}^{N}[\cos(\Delta\omega)t\cos(\omega_i t+\phi_i)+\sin(\Delta\omega)t\sin(\omega_i t+\phi_i)]$$

This $z(t)$ is not $x(t)$ at all. It is an SSB-SC signal for which the $x(t)$ is the modulating signal and (Δf) is the carrier signal. If $\Delta\omega \approx 0$, then $\sin(\Delta\omega)t \approx 0$ and $\cos(\Delta\omega)t \approx 1$ and so $z(t) \approx \frac{1}{2}x(t)$; otherwise it represents a highly distorted version of $x(t)$.

Example 4.33 What is known as a phase shift SSB-SC demodulator, is shown in Fig. 4.77. Show that it demodulates an SSB-SC signal.

Solution Consider the following SSB-SC modulated signal:

$x_c(t) = x(t)\cos\omega_c t - \hat{x}(t)\sin\omega_c t$, which is a USSB signal, where, $\hat{x}(t)$ is the Hilbert Transform of the message signal, $x(t)$.

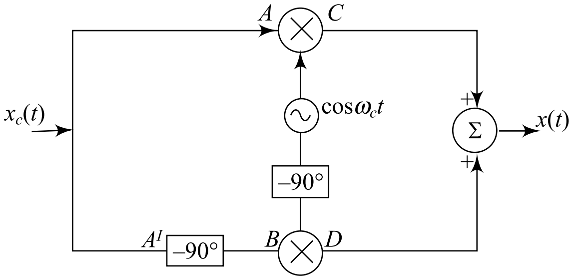


Fig. 4.77 *Phase-shift SSB-SC detector*

Signal at A = Signal at A' = $x_c(t)$

$\therefore$ signal at C: $x_c(t)\cos\omega_c t = x(t)\cos^2\omega_c t - \frac{1}{2}\hat{x}(t)\sin 2\omega_c t$

$\therefore$ signal at B: $\hat{x}_c(t) = x(t)\sin\omega_c t + \hat{x}(t)\cos\omega_c t$. (Property of H.T)

$\therefore$ signal at D: $\hat{x}_c(t)\sin\omega_c t = x(t)\sin^2\omega_c t + \frac{1}{2}\hat{x}(t)\sin 2\omega_c t$

Thus, the output of the phase-shift demodulator = Signal at C + Signal at D

$$= x_c(t)[\cos^2\omega_c t + \sin^2\omega_c t] = x(t) = \text{ Message signal}$$

The given system does act as a demodulator for SSB-SC.

4.7.5 Applications of SSB-SC Modulation

From the foregoing discussion, it is clear that SSB-SC modulation cannot be used for transmission of music and video signal and that it may be used only for transmission of speech signals since they have an energy gap around the origin, in their spectra. Because it conserves power as well as bandwidth, it is ideally suited for simultaneous transmission of a very large number of telephone speech signals by the use of what is called 'Frequency-Division Multiplexing, or simply 'FDM'. Hence the usefulness of SSB-SC can be summarized as follows.

1. Point-to-Point speech communication but not for audio broadcasting in which millions of receivers may be interested in what is being broadcast by a single transmitter. This is because, though SSB-SC transmission saves on power as well as bandwidth and thereby reduces the cost of the transmitter, the SSB-SC receivers are quite complex and expensive. It just does not make sense to make millions of receivers expensive only to save a little on the cost of a transmitter.
2. Transmission of a very large number of telephone conversations simultaneously over the same physical channel by using FDM.
3. As the carrier and one of the sidebands are suppressed, for the same average transmitted power, compared to the AM, SSB-SC gives more signal power at the destination. Further, since it occupies only half of the bandwidth required for AM, for the same power spectral density of white noise on the channel, the noise power entering an SSB receiver is half of the noise power entering an AM receiver. Thus, assuming that the noise added by the internal circuitry of the two receivers is the same, the output signal-to-noise ratio for an SSB-SC receiver will be far better compared to that of an AM receiver. We will be discussing this aspect in more quantitative terms in Chapter 7.

4.7.6 Frequency Division Multiplexing

Multiplexing refers to the technique used for simultaneous transmission of a number of different message signals over the same physical channel. There are mainly two important methods used for multiplexing – Frequency Division Multiplexing (FDM) and Time Division Multiplexing (TDM). We will be discussing about TDM in detail later. In FDM, we assign specific non-overlapping bandwidth slots for the various messages and then transmit the combined signal. The fact that different message signals occupy different non-overlapping frequency slots is made use of at the receiving end for separating them and recovering the individual messages.

In telephony, intelligibility being the sole criterion, the bandwidth of a speech signal is limited only to 3.2 kHz in order to conserve the spectrum. Hence, when telephone messages are FDM-ed, each of the messages is assigned a bandwidth of 4 kHz in order to provide for guard bands in the multiplexed signal. These guard bands facilitate the recovery of the individual messages by making the specifications for the band pass filters used for making them less stringent. SSB-SC modulation is used to translate each message signal to the 4 kHz bandwidth slot assigned to it. Thus, if N telephone message signals are to be FDM-ed, as shown in Fig. 4.78, N subcarriers, each differing from its adjacent one by 4 kHz, are used. These subcarriers are LSSB-SC modulated by the telephone message signals. Before modulation, each telephone message is first passed through a low pass filter to ensure that it is strictly band limited to 3.2 kHz. After LSSB-SC modulation, the modulated signal is passed through a band pass filter. The ith message channel, having a subcarrier frequency of f_{ci}, will have a BPF whose pass band extends from (f_{ci} – 4 kHz) to f_{ci}.

In Fig. 4.78, the multiplexed signal is fed directly to the channel. For long distance transmission of the multiplexed signal, however, a main carrier is modulated by this multiplexed baseband signal before being fed to the channel. Correspondingly, at the receiving end of the channel, a carrier demodulator retrieves the multiplexed baseband signal which is then fed simultaneously to all the BPFs which separate the various SSB-SC subcarrier modulated message signals. These are then coherently demodulated using the various subcarriers and the detected message signals are then passed through LPFs and recovered.

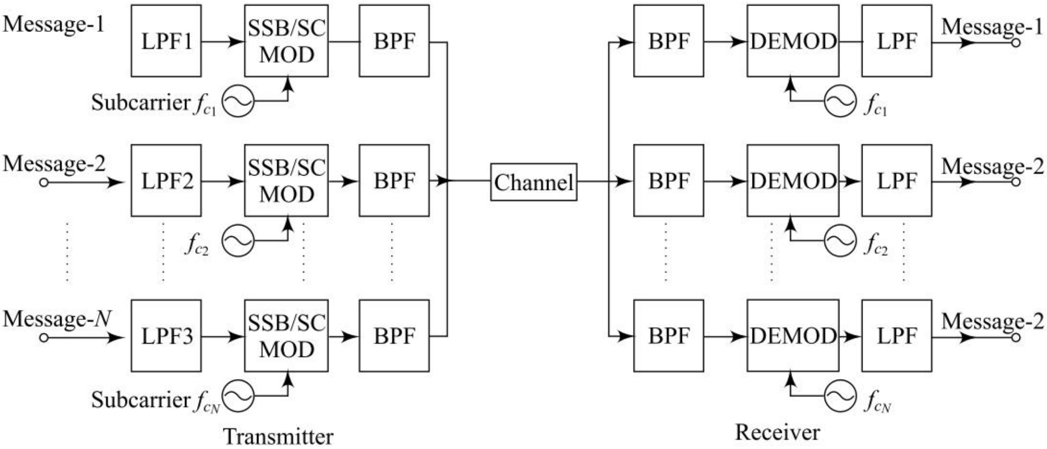


Fig. 4.78 *An FDM system*

A practical FDM system will have several stages of multiplexing. In the first stage of multiplexing, 12 telephone voice messages are multiplexed to form what is generally called as the 'Basic Group'. The subcarriers used for forming this group have frequencies 64 kHz, 68 kHz, 72 kHz, . . ., 104 kHz. With LSSB-SC modulation, the 12 telephone messages are translated into frequency bands (or slots) of 60–64 kHz, 64–68 kHz, . . . ,104–108 kHz. Thus, the basic group carrying 12 telephone messages, occupies a bandwidth of 48 kHz. In the second stage of multiplexing, five such Basic Groups are FDM-ed to form what is known as a 'Super Group', which occupies the frequency range of 312 kHz to 552 kHz which are combined to form 'Master Groups' and these are in turn multiplexed to form 'Very Large Groups'. Table 4.1 shows the AT&T FDM hierarchy.

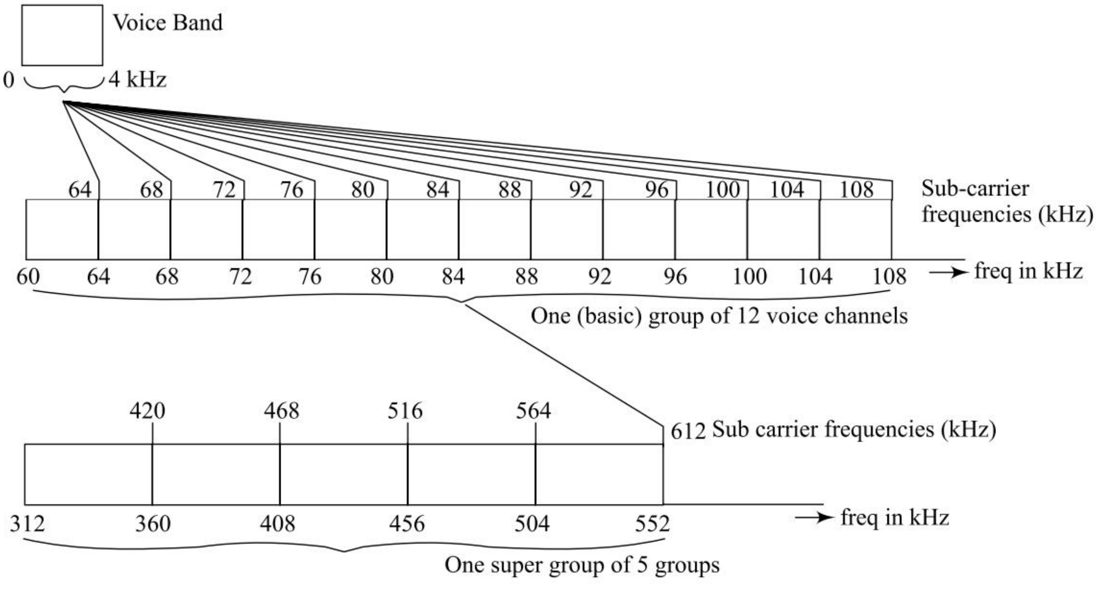


Fig. 4.79 *Illustrating the formation of groups and super groups*

Table 4.1 *AT&T FDM hierarchy*

Type of Group	Frequency Range	Bandwidth	Number of telephone channels
Group	60 – 108 kHz	48 kHz	12
Super Group	312 – 552 kHz	240 kHz	60
Master Group	564 – 3084 kHz	2.52 MHz	600
Very Large Group (Jumbo Group)	0.5 – 17.5 MHz	17 MHz	3600

4.7.7 Independent Sideband Transmission (ISB)

A variant of SSB-SC transmission is the Independent sideband transmission in which two sidebands are transmitted with reduced/no carrier. The two sidebands, however, carry different speech signals and hence the name – Independent sideband transmission. It thus doubles the capacity of the communication channel and is therefore used for point-to-point communication in areas with high traffic density.

Carrier signal generated by a crystal oscillator is applied as input to two balanced modulators simultaneously. To one of these, say balanced modulator-I (BM-I) a speech signal A is applied. To the other balanced modulator, speech signal–B is applied. One sideband filter suppresses the lower sideband in the DSB-SC signal produced at the output of BM-1 while another sideband filter suppresses the upper sideband of the DSB-SC signal at the output of BM-2. Thus at the output of one sideband filter, we have the upper sideband while at the output of the other sideband filter we have the lower sideband. These two have the same carrier, but they carry different speech signals.

4.8 SINGLE SIDEBAND TRANSMISSION AND RECEPTION

Single sideband transmission has, as mentioned earlier, several advantages over AM. But there are some disadvantages too. Because its bandwidth is only half of that of AM, it conserves the spectrum. In addition, smaller bandwidth implies less channel noise and less susceptibility to selective fading. However, the receiver complexity makes it unsuitable for broadcast purposes.

SSB transmission may be with no carrier at all, or with a pilot carrier, i.e., a re-inserted carrier with reduced power.

4.8.1 SSB-SC Transmitter

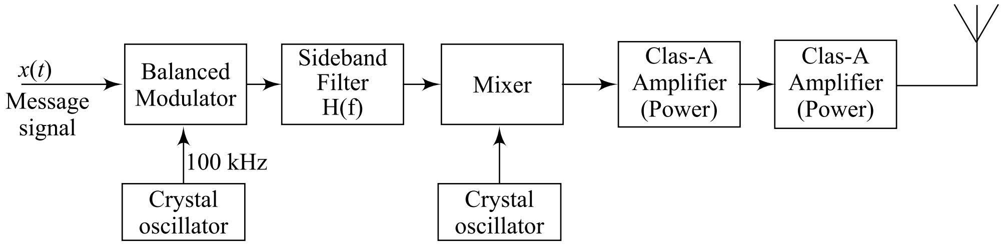


Fig. 4.80 *SSB-SC transmitter using filter method of generating the SSB-SC signal*

As mentioned earlier, it is possible to use the filter method for generation of an SSB signal only if the message signal spectrum has a hole near the origin, i.e., if the message has no frequency components from 0 to say, about 200 Hz. Fortunately, voice signal satisfies this condition. Further, initially, to generate the SSB-SC signal, a low frequency carrier, usually 100 kHz, is used, in order to ease the stringent requirements on the sideband suppression filter which is to attenuate the unwanted sideband at least by 40 dB. With a 200 Hz hole on either side of the origin of the two-sided spectrum of the message signal, the output of the balanced modulator will be as shown in Fig. 4.81.

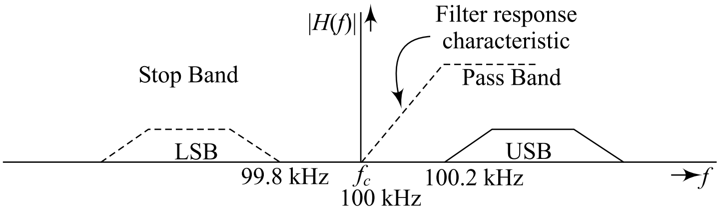


Fig. 4.81 *DSB-SC signal spectrum and the sideband filter response characteristic*

If the lower sideband is to be suppressed, the filter response can change over from pass band to stop band over the frequency interval 100.2 kHz to 100 kHz, i.e., a transition bandwidth of 200 Hz at a carrier frequency of 100 kHz. Even now, the Q of the sideband filter will have to be of the order of several thousands – a value that ordinary RLC filters cannot provide.

4.8.2 Sideband Filters

The Q-value required for the sideband suppression filter depends on (i) the center frequency (ii) the amount of attenuation needed for the unwanted sideband, and (iii) the transition bandwidth available. It is given by

$$Q = \frac{f_c}{(\Delta f)}[0.25\sqrt{A}] \tag{4.80}$$

where

f_c = center frequency
Δf = transition bandwidth permitted
$A = \text{antilog}\left[\frac{|\text{Attenuation required in dB}|}{20}\right]$

As discussed earlier, if $f_c = 100 \times 10^3$ Hz, Δf = 200 Hz and A = 40 dB, the Q value required works out to be 12,500. Such high Q values can be attained only by using special filters like mechanical filters, crystal filters or surface acoustic wave, or SAW filters. The Q-values these filters can provide are: SAW filters – well over 30,000; crystal filters – around 20,000; mechanical filters – around 10,000; ceramic filters – around 2500, LC filters – up to 500.

4.8.3 Raising the Carrier Frequency and Power

Once the unwanted sideband is removed by the sideband suppression filter, the carrier frequency and power will have to be raised to the required levels.

The frequency of the crystal oscillator (used as the local oscillator for the mixer) is suitably chosen so that the frequency at the output of the mixer (say, the difference frequency) is the correct carrier frequency at which the SSB-SC signal is to be finally radiated. But before the signal is fed to the antenna, its power level must be raised to the required level. As the modulation has already taken place, the power cannot be raised using high efficiency class-C amplifiers. Instead, only class-A amplifiers will have to be used in order to avoid distortion of the modulated signal.

4.8.4 Pilot Carrier SSB Transmitter

The main advantage of SSB-SC is that because of the absence of the carrier and one of the sidebands, all the transmitted power is in the message bearing signal and the transmission bandwidth is halved. But the absence of the carrier in the received signal makes it necessary to have a complex receiver circuit for recovering the message. Hence, in order to reduce the complexity of the receiver to some extent while maintaining, to a very large extent the two main advantages of SSB-SC, a pilot carrier SSB is used in which a reduced low-frequency carrier signal (10%) is again added to the SSB-SC signal before the mixer stage, where the carrier frequency is raised to its final value.

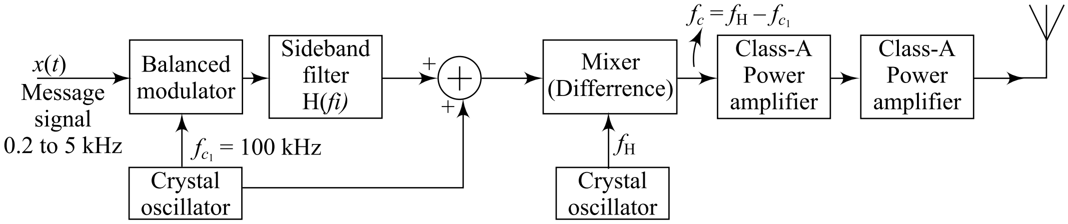


Fig. 4.82 *Block diagram of a pilot-carrier SB transmitter*

4.8.5 SSB Receivers

SSB-SC receivers We had seen in Section 4.5 that for the detection of SSB-SC signals, we have to resort to coherent detection which involves multiplication of the received SSB-SC by a locally generated carrier signal, and that ideally, this signal should be in frequency and phase synchronism with the suppressed carrier of the SSB-SC signal. Hence in the receiver we employ a highly stable oscillator, preferably a crystal oscillator and give its output either directly, or after frequency division, as one of the inputs to a product device (a balanced modulator), the other input to it being SSB-SC signal derived from the received signal after due processing so as to make its suppressed carrier have a frequency exactly equal to that of this stable oscillator or a sub-multiple of it.

Since SSB signals have a very small bandwidth (5 kHz for each sideband) very good adjacent channel selectivity is a must for these receivers. Further, since HF band is generally used for point-to-point communication using SSB modulation, the required adjacent channel selectivity can be obtained only by resorting to double conversion (refer to the introduction for the section on double heterodyne receivers, given in this chapter). Figure 4.83 shows the block diagram of a communication receiver meant to receive SSB-SC signals in the HF range by employing double conversion.

Since these communication receivers are generally designed to receive either the upper sideband, or the lower sideband, or both the sidebands (in the case of ISB transmission), a bandwidth of 10 kHz is provided. As the tuning range covered by these HF. communication receivers is from 3 MHz to 30 MHz, the first IF is generally 2.2 MHz (slightly below the lower end of the tuning range) to give a good image signal rejection and the second IF is 200 kHz, low enough to give good adjacent channel selectivity and making it easy to design the second IF amplifier to give a large gain.

The II IF amplifier output is detected to obtain an AGC voltage which is applied to the RF and IF amplifiers. It is also used to prompt the squelch circuit to make the audio amplifier inoperative in case the strength of the received signal is very weak. This is done in order to avoid annoying sounds being produced by the loudspeaker in the absence of a strong desired signal.

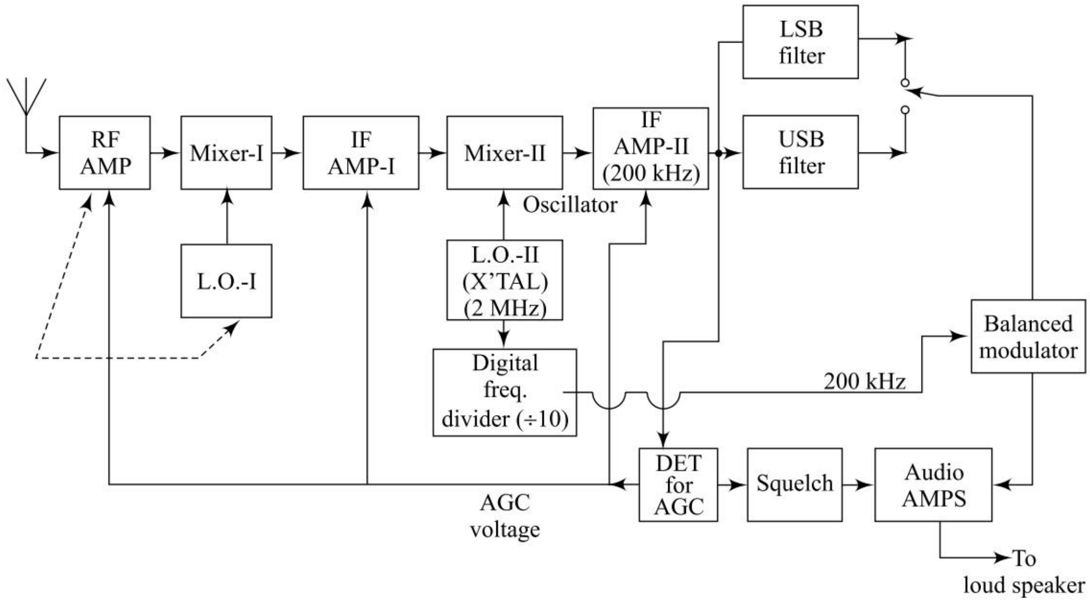


Fig. 4.83 *Block diagram of an SSB-SC receiver*

SSB-pilot carrier receivers These receivers are of the double-conversion type and they make use of the pilot carrier to ensure frequency synchronization with the transmitted carrier.

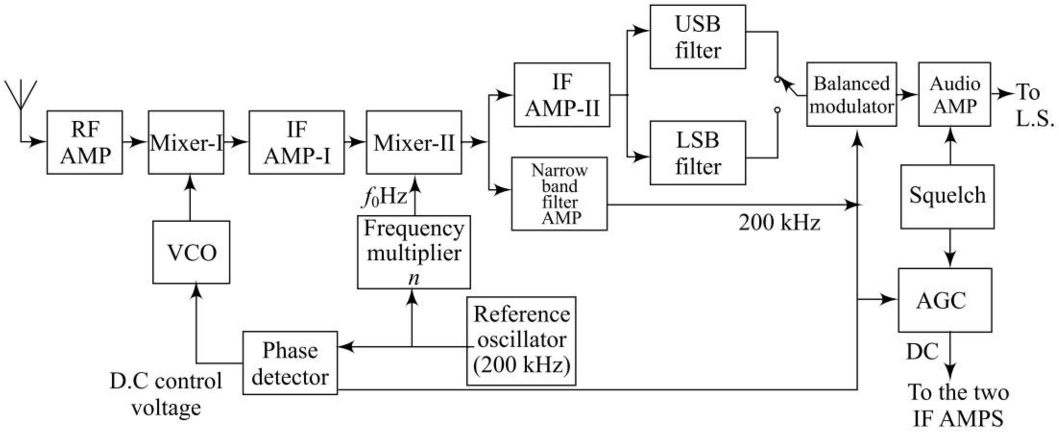


Fig. 4.84 *Block diagram of a pilot-carrier SSB-receiver*

The mixer-I will produce an output which is a pilot carrier SSB signal with the pilot-carrier at the first IF, viz. IF-I. A stable reference oscillator, a crystal oscillator, produces 200 kHz carrier signal. The frequency multiplier produces an output signal at $f_0 = (n \times 200)$ kHz. The SSB signal with pilot carrier at IF-I which is the output of the first IF amplifier, is mixed with this signal at a frequency of f_0 (coming from the frequency multiplier) in mixer-II. The values of IF-I and n are so chosen that at the output of this second mixer, we get the SSB signal with its pilot carrier at 200 kHz, i.e., $(f_0 - \text{IF-I}) = 200$ kHz. This output of mixer-II is fed simultaneously to IF amplifier-II and a very narrow-band filter and amplifier. The output of this NB filter and amplifier is the 200 kHz pilot carrier only, as the narrow band filter has its pass band centered on 200 kHz and it is so narrow that the sideband is rejected. This 200 kHz signal from the NB filter and amplifier, is fed to the balanced modulator which is the output of IF amplifier-II. So, this product device (which is followed by a low pass filter) acts as a coherent detector whose output is the modulating audio signal. After voltage and power amplification, this goes to the loudspeaker.

4.8.6 ISB Transmitter

As explained in Section 4.7.7, an ISB transmission is one in which two sidebands are transmitted with either a pilot carrier or no carrier. The two sidebands, however, carry different speech signals.

As shown in Fig. 4.85, a low frequency carrier, of 100 kHz, is applied as input to two balanced modulators BM-I and BM-II simultaneously. These balanced modulators give DSB-SC signals. BM-I is given message-I while BM-II is given message-II. The crystal filter following BM-I produces a USSB-SC signal while the crystal filter following BM-II gives an LSSB-SC filter. These two signals, as well as a reduced carrier signal of 100 kHz, are given to an adder whose output is a pilot carrier ISB signal. The carrier frequency is then raised to the desired final carrier frequency value using a mixer and a crystal oscillator with a frequency of f_0 which is 100 kHz higher than the final carrier frequency desired, i.e., f_c. The power is then raised to the required level using a few stages of tuned linear class-A power amplifiers before taking it to the transmitting antenna.

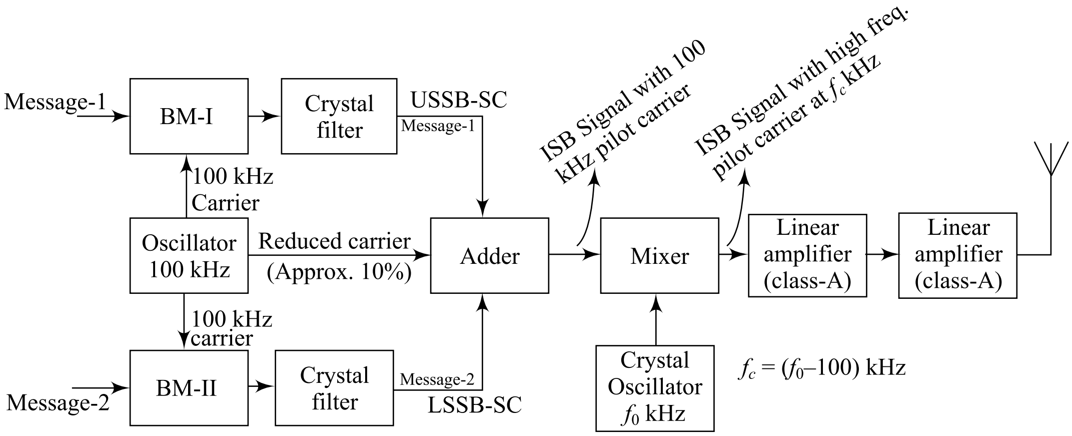


Fig. 4.85 *Block schematic diagram of a pilot carrier ISB transmitter*

4.8.7 ISB Receiver

ISB receivers are double-conversion superheterodyne receivers. The received signal, consisting of the two independent sidebands and the pilot carrier, is amplified by an RF stage and then fed to a mixer (mixer-I) to which the LO-I output is also given. The first IF amplifier, IF amp-I, amplifies the signal and feeds it to Mixer-II to which the output of the second local oscillator, LO-II, is also given. The LO-II frequency is so chosen that at the output of Mixer-II, the ISB signal will have a pilot carrier of frequency 100 kHz. As shown in Fig. 4.86, the IF amp-II output is simultaneously applied to (i) very narrow band filter which extracts the 100 kHz carrier signal, (ii) A USB filter which extracts the channel-A SSB signal, and (iii) An LSB filter which extracts the channel-B SSB signal.

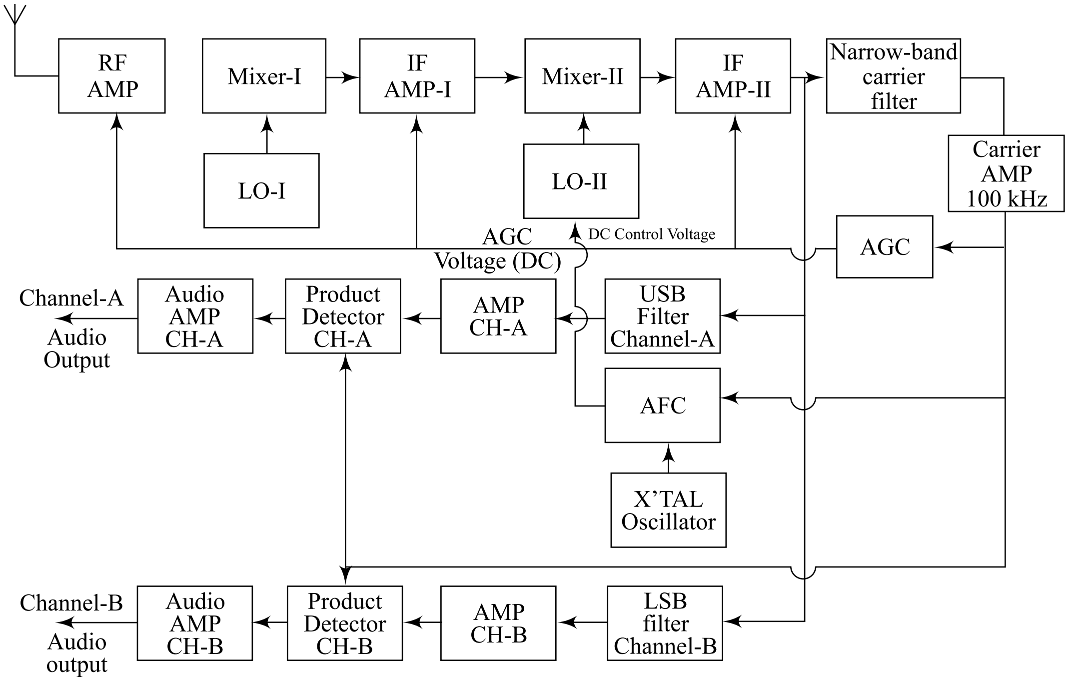


Fig. 4.86 *Block diagram of a pilot carrier ISB receiver*

The 100 kHz carrier from the narrow band carrier filter is amplified and fed simultaneously to the AGC circuit, the AFC circuit and the channel-A and channel-B detectors. The AGC voltage (dc) produced by the AGC circuit is applied to the RF and IF amplifiers as bias voltage to automatically control the gain. The output of the second IF amplifier, comprising the pilot carrier, the upper sideband (containing message of channel-A) and the lower sideband (containing message of channel-B), is applied simultaneously to the USB filter and the LSB filter. The amplified 100 kHz carrier and the amplified USB signal are fed to the product detector, the output of which is amplified by an audio amplifier to get the channel-A message. The channel-B message is similarly obtained from the audio amplifier of channel-B. In order to ensure that the carrier frequency at the output of mixer-II is always maintained at 100 kHz, an AFC circuit is used. The output of the 100 kHz carrier amplifier and the output of a 100 kHz crystal oscillator are both fed to this AFC. Using these two, the AFC circuit produces a dc control voltage which adjusts the LO-II frequency in such a way as to keep the carrier frequency of the output of mixer-II at 100 kHz.

4.9 COMPATIBLE SINGLE SIDEBAND (CSSB) SYSTEM

Single sideband transmission has several advantages over the conventional amplitude modulation (AM) transmission. It needs less average transmitted power and its transmission bandwidth is only half that of AM. Smaller bandwidth implies conservation of spectrum as well as less susceptibility to frequency-selective fading. The biggest drawback with SSB, however, is the complexity (and hence the cost) of the receiver, and this makes it unsuitable for broadcasting. In fact, the only reason for extensive use of the conventional AM for broadcasting purposes in the medium and shortwave bands is the extreme simplicity of its receiver.

The above considerations have naturally motivated researchers to explore the possibility of generating SSB signals that could be received using the existing domestic AM receivers which employ the simple envelope detector. Such SSB signals are referred to as compatible SSB signals. It has been found that although an exactly compatible SSB signal in the usual sense is not feasible, an approximately compatible SSB signal can, however, be produced. By ***approximately compatible***, we mean that its envelope may not be an exact replica of the modulating signal and its spectrum also may differ slightly from the one we have for the conventional SSB signal.

Kahn has described a modulation technique that produces an approximately compatible SSB signal. It is generated by making a modulating signal to modulate not only the amplitude (conventional AM) but also the phase of the carrier. The inspiration for this approach to generate a compatible SSB signal appears to have come from the fact that when one of the side frequency components of a single-tone modulated AM signal is suppressed, the resultant signal is one which is modulated both in amplitude as well as phase. To see why it is so, look at the phasor diagram of a single-tone modulated AM signal, given in Fig. 4.2(b). Suppose we remove the phasor corresponding to the lower side-frequency component, which is rotating in the clockwise direction relative to the carrier at the rate of f_m revolutions per second, where, f_m is the frequency of the single-tone modulating signal. The tip of the resultant phasor lies on the circumference of the circle drawn with a radius of $mA_c/2$ and with its center at the tip of the carrier phasor. So the amplitude of the resultant varies between $A_c(1 - m/2)$ and $A_c(1 + m/2)$ and its phase varies between $\tan^{-1}(m/2)$ and $-\tan^{-1}(m/2)$. Thus, the SSB signal obtained by the removal of the lower side-frequency component, has amplitude as well as phase modulation at the modulating signal frequency f_m. Amplitude variation between $A_c(1 - m/2)$ and $A_c(1 + m/2)$ at the modulating signal frequency will produce an envelope that somewhat resembles the modulating signal. (Recall that in conventional AM, the carrier amplitude varies between $A_c(1 - m)$ and $A_c(1 + m)$ at the modulating signal frequency). Thus, just as in AM, an envelope detector will be able to extract that envelope.

However, the modulation technique described by Kahn does not give an SSB signal in the real sense. This is because, with a single-tone modulating signal, what is obtained is not a single side-frequency but an infinite number of side-frequency components having frequencies $f_c + f_m$, $f_c + 2f_m$, $f_c + 3f_m$, ... But all these side-frequency components are only on one side of the carrier frequency. In this sense it resembles an SSB signal. A mathematical analysis of this modulation technique is highly complicated because of the non-linearities involved. But experimental studies have confirmed the above comments.

If this approximately compatible SSB system is used for speech or music as AM is, then the low frequency fundamental components of the audio modulating signal, along with quite a good number of harmonic components, will fall within the allowed spectrum of the SSB signal. The higher harmonics, which fall outside the allowed spectrum, may not be able to cause much of adjacent channel interference since the energy content in the higher harmonics will be very small. The high frequency fundamental components of the audio modulating signal will have very few of their harmonic components within the allowed bandwidth of the SSB signal. The other higher harmonic components, which fall outside will not be able to cause much of adjacent channel interference because at these higher audio frequencies, even the fundamental component will not have much of energy as the energy in an audio signal is heavily concentrated in the lower frequencies.

Thus, while there can be some signal distortion owing to the harmonics of the low frequency audio components falling within the allowed SSB bandwidth, the adjacent channel interference caused by the higher harmonic components is not expected to be a matter of serious concern. In fact, experimental investigations of the performance of the compatible SSB system with ordinary domestic AM receivers being used for reception, have shown that there is not much of a difference between conventional AM and the approximately compatible SSB except in adjacent channel interference, in which the conventional AM was found to be superior.

4.10 COMMUNICATION RECEIVERS

Although these have a basic structure similar to the ones used in our household, they are highly sophisticated and versatile and can be operated only by technical people and not by ordinary public. While it can do in a much better way what our household receiver does, it can do many other things too. For instance, it can detect and display the individual frequency components in an FM signal. In fact, modern communication receivers can receive AM, FM and SSB signals. The large number of special features built into it can make it possible for it to give an exceptionally high quality of reception. Some of the special features that it has are:

1. Band spreading or fine tuning
2. Double conversion
3. Variable sensitivity and selectivity
4. Built-in Beat Frequency Oscillator (BFO) action
5. Tuning calibration facility
6. Noise limiters
7. Squelch or muting
8. Automatic frequency control
9. Availability of a built-in meter

The basic block diagram of a communication receiver is shown in Fig. 4.87.

As can be seen from Fig. 4.87, the basic structure of a communication receiver is the same as that of an AM superheterodyne receiver. However, since the communication receiver has to receive SSB signals too, there are some additional circuits. To receive FM signals, it is to be provided with broadband IF stages, an FM demodulator and an amplitude limiter. These are not shown in the block diagram of Fig. 4.87 and have to be switched in when FM reception is desired.

We will now discuss briefly about the special features of communication receivers, which were listed above.

1. Band spreading or fine tuning: Communication receivers will generally have two stages of high sensitivity and low noise RF amplifiers. Fine tuning facility enables the receiver to distinguish between two transmitted signals whose carriers are very close. This feature is absolutely essential in a communication receiver. In the earlier days, it used to be provided by means of a trimmer capacitor kept in parallel with the main tuning capacitor, and it could be varied by means of a separate knob labeled as the fine tuning knob. Nowadays, however, fine tuning is performed by frequency synthesis.

2. Double conversion: Communication receivers have to receive signals over a wide frequency range. For effective image rejection, a high value of intermediate frequency is required. However, for good sensitivity and selectivity, a low value of IF is preferable. Hence, communication receivers employ double conversion technique (see Section 4.5.6 for details) in order to achieve good image rejection as well as good sensitivity and selectivity.

All communication receivers employ delayed AGC (see Section 4.5.2).

3. Variable selectivity and sensitivity: The range of received signal strengths that a communication receiver has to cope up with, can be of the order of 100 dB. No AGC system can effectively function over such a wide range of received signal strength. In fact, there exists the risk of the last IF amplifier and the detector getting overloaded or even damaged when a strong signal appears while we are in the process of tuning the receiver. To prevent such an eventuality, a 'sensitivity control' is provided to manually set the bias applied to the RF amplifiers and thereby control the upper limit of the receiver sensitivity that the AGC can allow. Thus, a variable sensitivity is achieved in a communication receiver through the use of the sensitivity control knob.

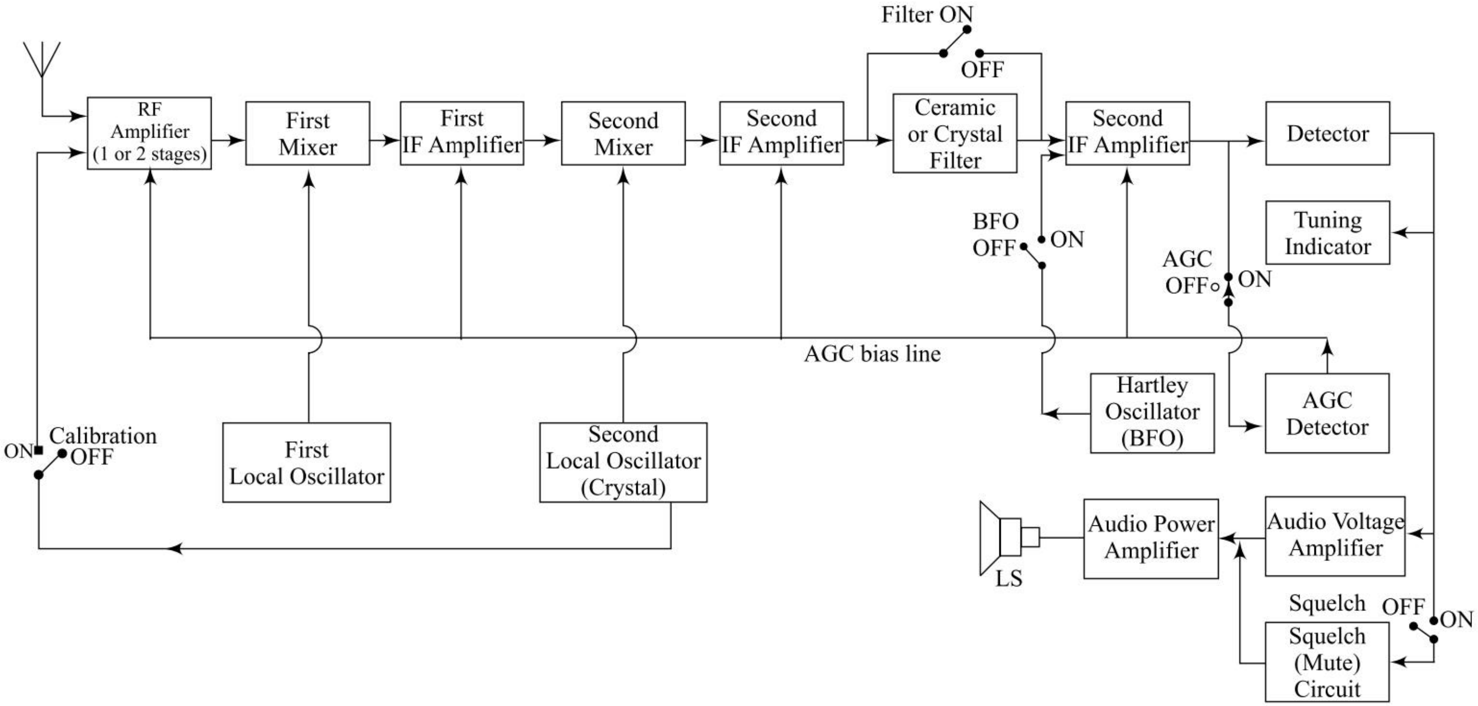


Fig. 4.87 *Basic block diagram of a communication receiver*

As has been explained in Section 4.5.2, adjacent channel selectivity of a receiver depends on the value of the IF and the bandwidth of the IF stage. For good adjacent channel selectivity, it is preferable to have a low value of IF and an IF bandwidth that is just sufficient to accommodate the bandwidth of the signal being received so as to prevent out-of-band noise from entering the detector stage. The signal bandwidth can have a large variation since the communication receiver has to receive a variety of signals. For instance, an AM signal may need a bandwidth of 10 kHz while a radio telephony signal does not need anything above 300 Hz. Hence, in a communication receiver, a provision is made to allow us to choose the bandwidth of the second IF stage (which has a low value of IF) to have one of the preset values – 1, 2, 4, 6, 8, 10 or 12 kHz, depending on the type of signal to be received. A knob, labeled as the 'variable selectivity' knob, allows us to switch into the circuit ceramic or crystal filters of appropriate bandwidth from among the seven values mentioned earlier. In addition, for further suppression of undesired nearby (frequency-wise) signals, a ***notch filter***, also called a ***wave trap*** is provided. The notch frequency of this filter can be adjusted to be on one or the other side of the second IF pass band. All these features enable the communication receiver to have a very good blocking capability, i.e., ability to suppress undesired spurious signals which might be close to the desired signal frequency.

4. Built-in beat frequency oscillator (BFO) action: Transmission of wireless telegraphy is in the form of an interrupted carrier wave, i.e., a carrier wave that is transmitted only during the transmission of a dash or a dot. However, the ordinary envelope detector does not give any indication of the presence, or the absence, of the carrier wave. The so-called BFO available in a communication receiver comes in handy for receiving the wireless telegraph signals. What is actually available in a communication receiver is, in fact, not a BFO, but only a simple Hartley Oscillator that oscillates at a frequency which is 1 kHz (or 400 Hz) above the second IF. When interrupted carrier wave reception is desired, this oscillator is switched into the circuit and the AGC is switched off. In the detector stage, the oscillator output beats with the second IF to give an audible 1 kHz tone whenever the carrier is present, i.e., whenever a dash or a dot is transmitted.

5. Tuning calibration facility: Communication receivers have provision internally to tuning calibration. By throwing a switch, the output of a non-sinusoidal crystal oscillator operating at 500 kHz, is given as input to the RF stage. The BFO is switched in. Because the harmonics of the non-sinusoidal oscillator are spaced at regular intervals of 500 kHz, whistles will now be heard at 500 kHz spacing in the tuning dial. The tuning pointer, or the cursor may be moved appropriately to the correct position using these whistles. This calibration process has to be carried out periodically.

6. Noise limiters: Impulse noise created by lightning and ignition systems of motor vehicles, adds to the modulation envelope of an amplitude modulated signal and produces highly annoying impulse noises through the loudspeaker. To avoid this, communication receivers have an arrangement which cuts off the detector diode and thus silences the receiver for the duration of the impulse. This is achieved by using a diode and a differentiating circuit to produce a negative spike voltage whenever the impulse noise makes the received signal amplitude to cross a prescribed level. It is necessary to make this circuit inoperative while receiving wireless telegraph signals, as otherwise, it will interfere with the reception of these signals.

7. Squelch or muting: In wireless communication systems used by organizations such as Police, Ambulance and Coastal rescue services, the receiver must be continuously monitored in order not to miss any SOS calls. As these signals are received sporadically, there will be no signal at the receiver input for most part of the time. The AGC, therefore, makes the receiver sensitivity very high. Noise will therefore be highly amplified and highly annoying sounds will be produced at the output of the receiver. As the receiver has to be continuously monitored, the person doing that work will get tired in no time. So, a squelch or mute circuit is provided which uses the AGC bias to make the audio amplifiers inoperative in the absence of any input signal to the receiver.

8. Automatic frequency control (AFC): For SSB signal reception, the local oscillator stability is extremely important. Hence, when a communication receiver is used for SSB signal reception, an AFC must be provided unless, of course, the communication receiver uses a frequency synthesizer for obtaining the local oscillator signals (For details of an AFC circuit, see Section 4.6.2).

9. Availability of a built-in meter: A meter is provided in any communication receiver and the primary purpose of it is to indicate the strength of the received signal. For this, it measures the collector current of an IF amplifier to which AGC voltage is given. It actually measures the voltage across the resistance-capacitance parallel combination in the emitter circuit (of the IF amplifier) which is meant for producing some self bias. Since a strong received signal produces a large AGC voltage and thus reduces the collector current of the IF amplifier, the meter has its zero on the right extreme. When used like this, the meter is referred to as an S-meter (Signal meter).

4.11 VESTIGIAL SIDEBAND MODULATION

In television, two message signals need to be transmitted – video, or the picture signal, and audio or the sound signal. TV transmitters employ amplitude modulation for the video signal and frequency modulation for the sound signal. The video signal that they handle occupies a bandwidth of 5 MHz. If ordinary AM with carrier and both the sidebands, is employed, the modulated signal, i.e., the TV signal which is transmitted, will occupy a huge bandwidth, viz. 10 MHz, which is impractical. However, to reduce this bandwidth requirement, it is not possible to employ SSB transmission, for the following reasons:

1. If we employ SSB-SC, the receiver becomes quite complex and expensive, as we have to use coherent detection.
2. Even if one sideband and the carrier are to be transmitted in order to make the receiver simpler, difficulties arise in the transmitter. The phasing method of generation, as we know, does not give the high level of suppression of the unwanted sideband required for commercial TV broadcasting.
3. The filter method of generation requires, as has already been discussed, a hole in the low frequency part of the spectrum from zero hertz up to at least a few hundred Hertz. However, video signals will not have such a hole in their spectra. In fact, they are generally quite rich in dc and low frequency components. Thus, it is not possible to employ even the filter method.
4. Further, the phase response of the band pass filters used in the filter method will not be linear near the pass band edges. This will make the received video signal to be distorted. Since the eye is quite sensitive to phase, this cannot be tolerated.

Since the use of AM (both the sidebands plus the carrier) as well as SSB with pilot carrier is ruled out because of the above reasons, what is known as vestigial sideband modulation is used. In this, in addition to the carrier and one sideband, a part, or what may be called the ‘vestige’ of the other sideband is also transmitted. That is why it is called ‘Vestigial Sideband Modulation’, or VSB modulation.

Consider a video signal $x(t)$. Let its spectrum be as shown in Fig. 4.88(a) when we feed this and a carrier signal $\cos 2\pi f_c t$ to a balanced modulator let us say we get a signal $y(t) = x(t) \cdot \cos 2\pi f_c t$. The spectrum $Y(f)$ of $y(t)$ is given by

$$Y(f) = \frac{1}{2}[X(f+f_c) + X(f-f_c)] \tag{4.81}$$

$y(t)$ is a DSB-SC signal and its spectrum $Y(f)$ is shown in Fig. 4.88(b). This $y(t)$ is say appearing at the receiving end as $z(t)$. In order to recover $x(t)$ from $z(t) = y(t)$, we multiply it by the carrier signal (coherent detection).

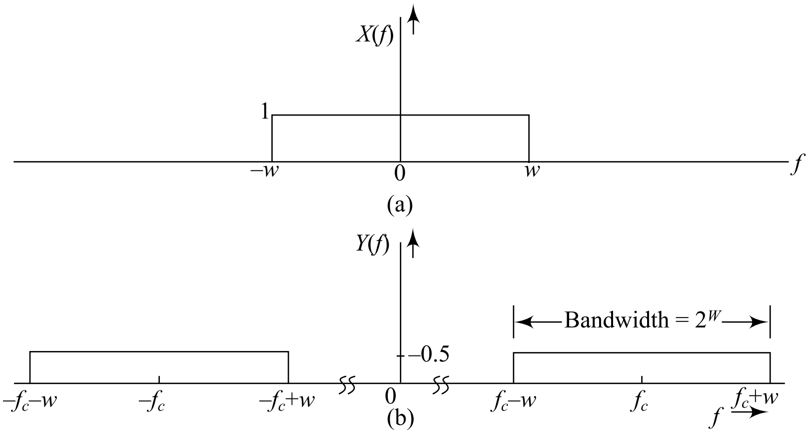


Fig. 4.88 *(a) Spectrum of x(t), (b) Spectrum of y(t)*

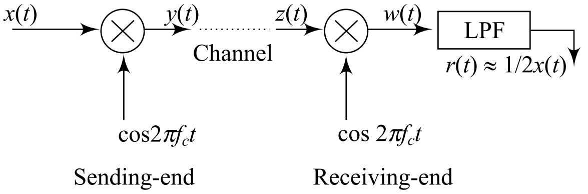


Fig. 4.89 *Coherent reception of y(t)*

$$\therefore \quad w(t) = y(t)\cdot\cos 2\pi f_c t = x(t)\cos^2 2\pi f_c t = \frac{1}{2}x(t)[1+\cos 4\pi f_c t]$$

$$\therefore \quad w(t) = 0.5x(t) + 0.5x(t)\cos 4\pi f_c t$$

$$\therefore \quad W(f) = 0.5X(f) + \frac{0.5}{2}[X(f+2f_c) + X(f-2f_c)]$$

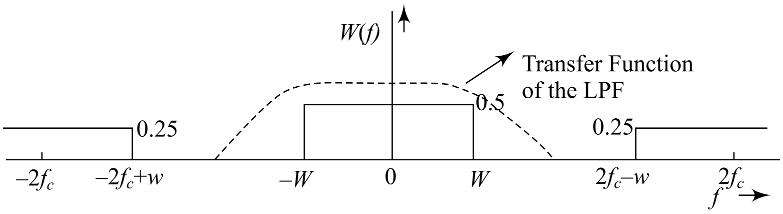


Fig. 4.90 *Spectrum of w(t) and transfer function of the LPF*

As shown in Fig. 4.90, the LPF which has a cut-off frequency of *w*, will reject the high frequency components centered around $2f_c$ and will pass only 0.5 *x*(*t*), whose spectrum is from –*W* to *W*. Hence, we are able to recover, at the receiving end a signal *r*(*t*) which is a scaled version of *x*(*t*).

For the assumed message signal *x*(*t*), we have until now, considered only the DSB-SC transmission and reception. As shown in Fig. 4.79(b), the required bandwidth for this is 2*W*. But this will be too large when *x*(*t*) is a video signal. However, as we observed earlier, we cannot use SSB-SC transmission. So, let us now consider the vestigial sideband transmission, in which we transmit the carrier and one sideband plus a *vestige* of the other sideband as shown in Fig. 4.91(a).

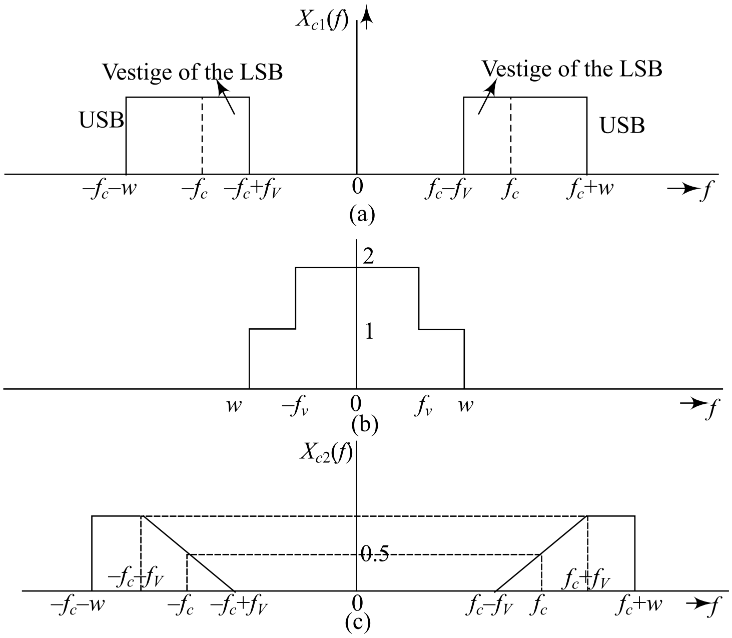


Fig. 4.91 *(a) Full upper sideband plus a portion of LSB transmitted, (b) The demodulated signal resulting from (a), (c) USB and LSB suitably shaped to avoid the distortion shown in (b).*

Figure 4.91(b) clearly brings out the need to suitably shape the USB and the LSB of the transmitted signal in order to recover the message signal $x(t)$ without distortion after the frequency translation that takes place in the demodulator. Figure 4.91(c) shows an appropriate way of shaping the USB and LSB of the transmitted signal so as to avoid distortion in the demodulated signal. This shaping may be considered to be done by a filter, called the vestigial sideband filter, or VSB filter, as shown in Fig. 4.92.

4.11.1 Frequency-Domain Representation of a Vestigial Sideband (VSB) Signal

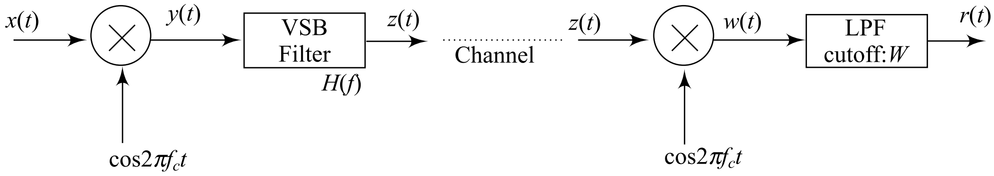


Fig. 4.92 *Generation of VSB signal and recovery of x(t)*

Let the message signal $x(t)$ have a spectrum $X(f)$. Then the spectrum at the input to the VSB filter is $Y(f)$ given by

$$Y(f) = \frac{1}{2}[X(f+f_c) + X(f-f_c)] \tag{4.82}$$

Since the VSB filter has a transfer function of $H(f)$, the spectrum at the output of the VSB filter, which is the frequency-domain representation of the VSB signal, is given by

$$Z(f) = H(f)\cdot Y(f) = \frac{1}{2}H(f)[X(f+f_c) + X(f-f_c)] \tag{4.83}$$

4.11.2 Transfer Function of the VSB Filter

The spectrum of $w(t)$ is

$$W(f)=\frac{1}{2}H(f)[X(f+f_c)+X(f-f_c)]*\frac{1}{2}[\delta(f+f_c)+\delta(f-f_c)]$$

$$=\frac{1}{4}H(f+f_c)[X(f+2f_c)+X(f)]+\frac{1}{4}H(f-f_c)[X(f-2f_c)+X(f)]$$

since the LPF has a cut-off frequency of W and $W << f_c$, terms like $X(f+f_c)\ H(f+f_c)$ and $X(f-2f_c)\ H(f-f_c)$ vanish because of the low pass filtering. We may therefore write the spectrum $R(f)$ of the demodulated signal $r(t)$ as

$$R(f)=\frac{1}{4}[H(f+f_c)+H(f-f_c)]X(f)$$

But this demodulated signal must be proportional to $x(t)$

i.e., $R(f)=kX(f);\quad -W\le f\le W$

This means that

$$H(f+f_c)+H(f-f_c)=\text{a constant};\quad -W\le f\le W$$

The choice of this constant is purely arbitrary and let us take it as *unity*.

$$\therefore\quad H(f+f_c)+H(f-f_c)=1;\quad -W\le f\le W \qquad (4.84)$$

Since the shape of the spectrum of $y(t)$, the DSB-SC signal, is as shown below in Fig. 4.93(a), it follows that the shape of the transfer function $H(f)$ of the VSB filter should be as shown in Fig. 4.93(b) so that $[H(f).Y(f)]$ which is $Z(f)$, will have the desired shape as shown in Fig. 4.93(c).

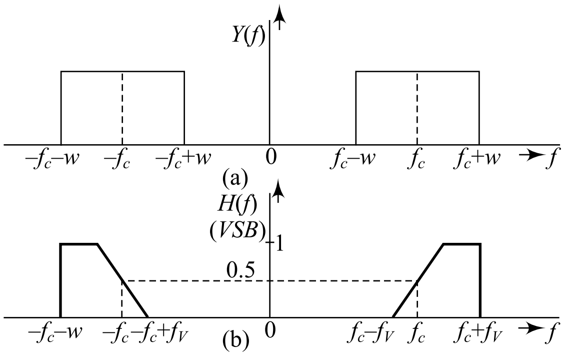


Fig. 4.93 *(a) DSB signal y(t), (b) H(f), the transfer function of the VSB filter (In (a), the part outlined with dark lines is the spectrum of the VSB signal.)*

$H(f+f_c)$ is obtained by shifting $H(f)$ to the left along the frequency axis by an amount of f_c, while $H(f-f_c)$ is obtained by shifting $H(f)$ to the right by an amount of f_c. These are shown in Figs. 4.94(a) and (b).

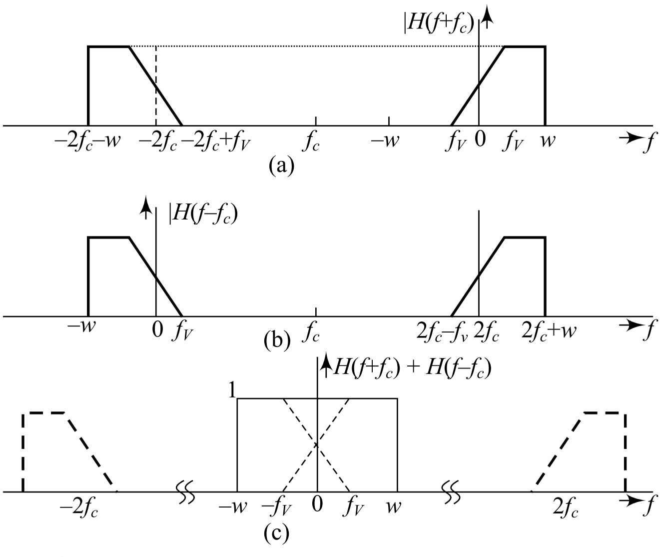


Fig. 4.94 *Sketches of (a) $H(f+f_c)$, (b) $H(f-f_c)$, (c) $H(f+f_c)+H(f-f_c)$*

4.11.3 Time-Domain Representation of the VSB Signal

An analytical expression for the VSB signal (i.e., the time-domain representation of the VSB signal), may be obtained by taking the inverse Fourier transform of its spectrum

$$\therefore \qquad z(t) = \underset{\text{VSB}}{x_c\,(t)} = F[H(f)\{X(f+f_c)+X(f-f_c)\}]$$

$$= \int_{-\infty}^{\infty} H(f)[X(f+f_c)+X(f-f_c)]\,e^{j2\pi ft}df$$

$$\therefore \qquad \underset{\text{VSB}}{x_c\,(t)} = \int_{-\infty}^{\infty} H(f)X(f+f_c)\,e^{j2\pi ft}df + \int_{-\infty}^{\infty} H(f)X(f-f_c)\,e^{j2\pi ft}df \tag{4.85}$$

Before proceeding further with determining the inverse Fourier transform of $Z(f)$, the spectrum of the VSB signal, let us define a frequency function $H_v(f)$ as follows:

$$H_v(f) = H(f_c) - H(f-f_c) = H(f+f_c) - H(f_c) \quad \text{for } -W \le f \le W \tag{4.86}$$

where $\qquad H(f_c) = H(f)\big|_{f=f_c}$

Since $H(f_c) = \dfrac{1}{2}$, we find from Eq. (4.85) and Fig. 4.94(a) that $H_v(f)$ has a shape as shown in Fig. 4.95.

Figure 4.95 clearly brings out the fact that the function $H_v(f)$ is an odd function of frequency.

i.e., $\qquad H_v(-f) = -H_v(f) \qquad (4.87)$

Now, reverting to Eq. (4.85), and making the following substitutions, i.e.,

$$\alpha = (f - f_c)$$

and $\qquad \beta = (f + f_c)$

we get

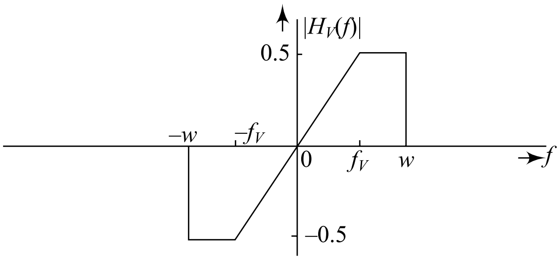


Fig. 4.95 *The function $H_V(f)$ of Eq. (4.85)*

$$\underset{\text{VSB}}{x_c\,(t)} = \int_{-\infty}^{\infty} H(\alpha+f_c)X(\alpha)\,e^{j2\pi(\alpha+f_c)t}d\alpha + \int_{-\infty}^{\infty} H(\beta-f_c)X(\beta)\,e^{j2\pi(\beta-f_c)t}d\beta \tag{4.88}$$

But from Eq. (4.86), we find that

$$H(\alpha+f_c) = H(f_c) + H_V(\alpha)$$

and $\qquad H(\beta-f_c) = H(f_c) - H_V(\beta)$

Substituting these in Eq. (4.88), using the fact that $H(f_c) = 0.5$ and simplifying, we get

$$\underset{\text{VSB}}{x_c\,(t)} = x(t)\cos 2\pi f_c t - g(t)\sin 2\pi f_c t \tag{4.89}$$

where

$$g(t) \;\underline{\Delta}\; -2j\int_{-\infty}^{\infty} X(f)H_V(f)\,e^{j2\pi ft}df \tag{4.90}$$

As is to be expected, if $f_V \to 0$, $\underset{\text{VSB}}{x_c\,(t)} \to \underset{\text{SSB-SC}}{x_c\,(t)}$

i.e., $\quad g(t) \to \hat{x}(t)$ since $H_V(f) \to \text{sgn}(f)$ as $f_V \to 0 \qquad$ (see Fig. 4.86)

4.11.4 Spectrum of Transmitted TV Signal and Receiver Response

From the above discussion on the frequency-domain and time-domain representation of VSB signals, the reader should not conclude that the signal transmitted by a TV transmitter will have a spectrum as shown by the product of $H_v(f)$ and $Y(f)$ in Figs. 4.93(a) and (b).

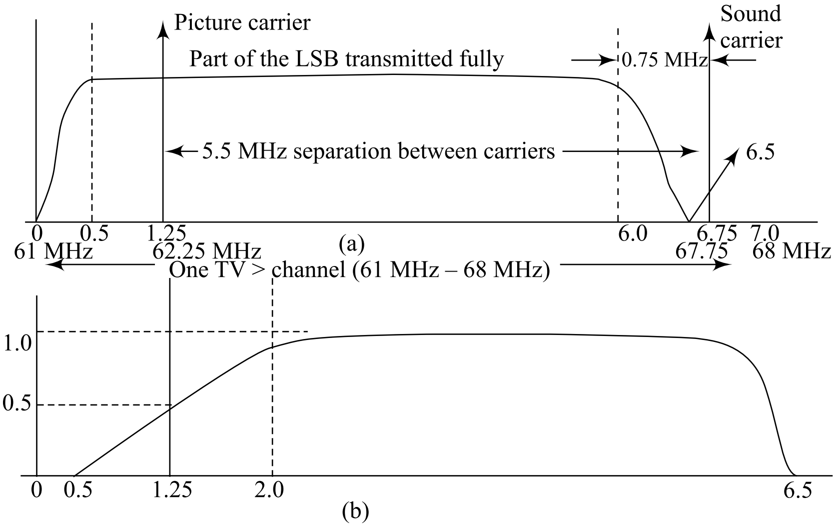


Fig. 4.96 *(a) Spectrum of the transmitted TV Signal (CCIR-B, Monochrome), (b) Typical response characteristic of the video amplifier in the receiver*

In fact, in practice, the transmitted signal will have *full carrier*, *full* upper sideband and a *part* of the lower sideband, as shown in Fig. 4.96(a). This type of spectrum for the transmitted signal is obtained by asymmetrically tuning the tank circuits of the linear amplifiers used in the transmitters video channel for raising the power level after the modulation process. Because they are tuned asymmetrically, while the upper sideband is transmitted in full, only 0.75 MHz width of the lower sideband is transmitted in full and the rest of it transmitted only partly, as shown clearly in Fig. 4.96(a).

The fully transmitted part of the lower sideband leads to the effect shown in Fig. 4.91(b). This distortion is avoided by shaping the *response characteristic of the receiver* as shown in Fig. 4.96(b), on the lines suggested in Fig. 4.91(c). The picture detector in the receiver therefore gets a VSB signal with full carrier although what has been transmitted is *not* a VSB signal. It must be noted here that while two-sided spectra are shown in Fig. 4.91, only one-sided spectrum and response are shown in Fig. 4.96(a) and (b), respectively. Also note that the partly transmitted part of the LSB which is more than 0.75 MHz away from the picture carrier, is totally rejected by the receiver as the receiver response is zero for these frequencies.

4.11.5 Detection of VSB Signals

Although we had, in the analysis leading to the determination of the transfer function $H_V(f)$ of the VSB filter, assumed a coherent detector, in actual practice, in TV receivers, it is not possible to have such a detector, as it is quite complex and makes the TV receiver quite expensive. So, although it results in some distortion of the demodulated signal, the TV receiver uses only a simple envelope detector. It is for this reason that the transmitter transmits full carrier in addition to one full sideband and a vestige of the other sideband. We shall now briefly analyze the action of the envelope detector and examine how this distortion may be reduced.

Since the detector input is the VSB signal, plus the carrier, let us scale the VSB signal of Eq. (4.89) by a factor '*m*' ($0 < m < 1$), the modulation index and then add the carrier term $\cos 2\pi f_c t$, to get

$$s(t) = [1 + mx(t)]\cos 2\pi f_c t - mg(t)\sin 2\pi f_c t \tag{4.91}$$

As the envelope detector extracts the envelope of this signal given to it as input, the detector output is

$$a(t) = \{[1 + mx(t)]^2 + [mg(t)]^2\}^{1/2}$$

$$a(t) = [1 + mx(t)]\left\{1 + \left[\frac{mg(t)}{1 + mx(t)}\right]^2\right\}^{1/2} \tag{4.92}$$

$[1 + mx(t)]$ being the correct envelope term, the other one, viz., $\left[\frac{mg(t)}{1 + mx(t)}\right]^2$ is the distortion term.

Hence, to reduce the distortion due to demodulation by an envelope detector, we have to either reduce the modulation index '*m*' or reduce $g(t)$ by increasing the width of the vestige of the LSB. In commercial TV, we do both. That is the reason why the width of the vestige of the LSB is as high as 0.75 MHz (see Fig. 4.96(a)).

4.12 COMPARISON OF VARIOUS VARIETIES OF AMPLITUDE MODULATION

Table 4.2 *Comparison of various varieties of amplitude modulation*

System	Useful part of transmitted power	BW	Carrier suppression	Sideband suppression	Figure of merit	Receiver complexity	Applications
AM + Both SB	Low	2*W* Hz	No	No	$\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$	Simple	Audio broadcasting
DSB-SC	Good	2*W* Hz	Yes	No	1	Complex	Quadrature multiplexing, point-to-point communication
SSB-SC	Very good	*W* Hz	Yes	Carrier and one sideband are suppressed	1	Complex	Used for long-haul point-to-point communication
VSB	Moderate	$W <$ B.W. $< 2W$	No	One sideband is fully transmitted while the other is partially transmitted	-	Simple	TV broadcasting

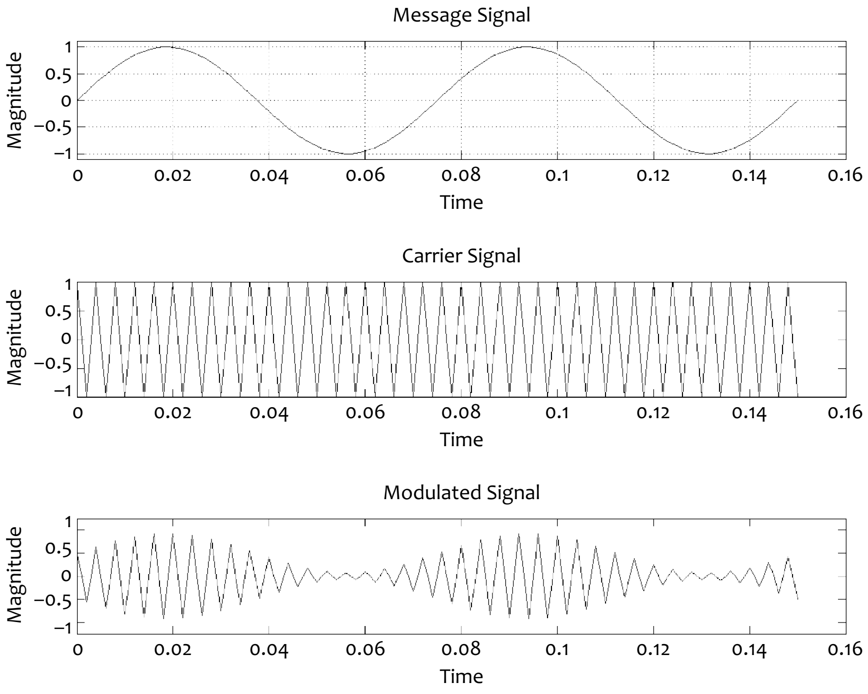
MATLAB Example 4.1 Determine the spectra of the message signal $m(t)$ and the amplitude-modulated signal $x_c(t)$ (AM with carrier + both side bands) and plot them. Plot also the waveform of the message signal. Carrier signal is cos $(2*\pi*250*t)$ and modulation index $m = 0.85$. The message signal is a sinusoidal signal of 6.67 Hz.

MATLAB program

```
%
% Amplitude modulation
%
t0 = 0.15; %    signal duration
f = 1/0.15;
ts = 0.001;%    sampling interval
fc = 250;  %    carrier frequency
fs = 1/ts; %    sampling frequency
t = [0:ts:t0];  % time vector
a = 0.85;  %    modulation index
df = 0.5;  %    required frequency resolution
%
% Generation of message signal
%
m1 = sin(2*pi*2*f*t)
figure(1)
subplot (3,1,1)
plot(t,m1)
n = length(m)
grid on
xlabel(‘time’);
ylabel(‘Magnitude’);
title (‘Message signal’);
ylim([-1.1 1.1]);
%
% Generation of carrier signal
%
c = cos(2*pi*fc.*t);% carrier signal
subplot (3,1,2)
plot(t,c);
title (‘Carrier Signal’);
xlabel (‘time’)
%
% Generation of modulated signal and spectrum
%
[M,m,df1] = fftseq(m,ts,df);          % Fourier transform
M = M/fs;                             % scaling
f = [0:df1:df1*(length(m)-1)]-fs/2;   % frequency vector
u = (1+a*m);
u = u(1:151).*c;                       % modulated signal
```

```
subplot (3,1,3)
plot(t,u)
xlabel (‘time’);
title (‘Modulated signal’);
ylim ([-1.2 1.2])
%
% Generation of frequency spectrum of message signal
%
[U,u,df1] = fftseq(u,ts,df); % Fourier transform
U = U/fs                     % scaling
% frequency spectrum of message signal
figure (2)
subplot (1,2,1)
plot(f,abs(fftshift(M)));
xlabel (‘Frequency’);
title (‘Spectrum of Message signal’);
subplot (1,2,2)
plot(f,abs(fftshift(U)));
xlabel(‘Frequency’);
title (‘Spectrum of Modulated signal’);
```

Results



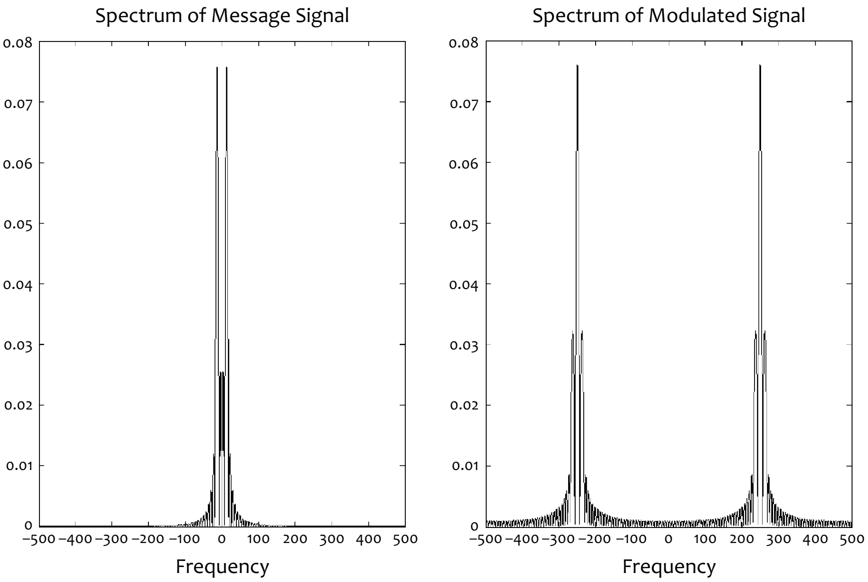


Fig. 4.97

MATLAB Example 4.2 Using MATLAB 9 generate an amplitude-modulated wave and detect it using the simple envelope detector shown in Fig. 4.98:

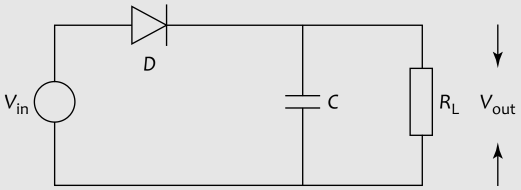


Fig. 4.98

Show the waveforms of the modulating signal, amplitude-modulated signal and output of the detector for carrier signal angular velocity of 50 radians/s and modulating signal angular velocity of 1 radian/s.

1. $R_LC = 2\pi/10$ and modulation index alpha = 0.5
2. $R_LC = 2\pi/3$ and alpha = 0.9 (in this case diagonal clipping should take place)

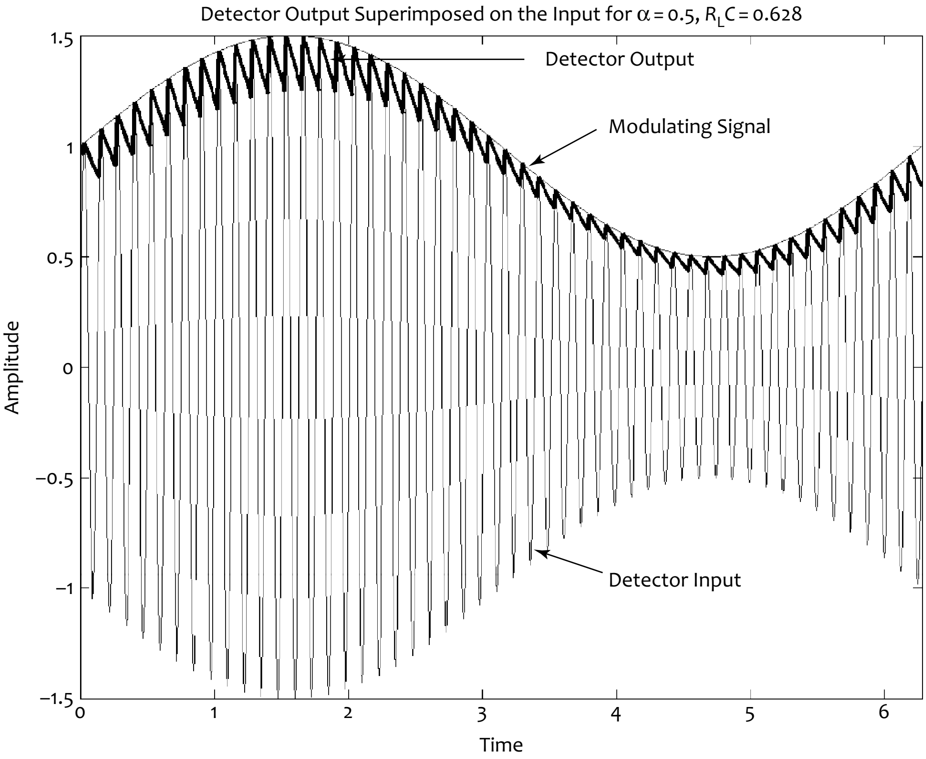
MATLAB Program

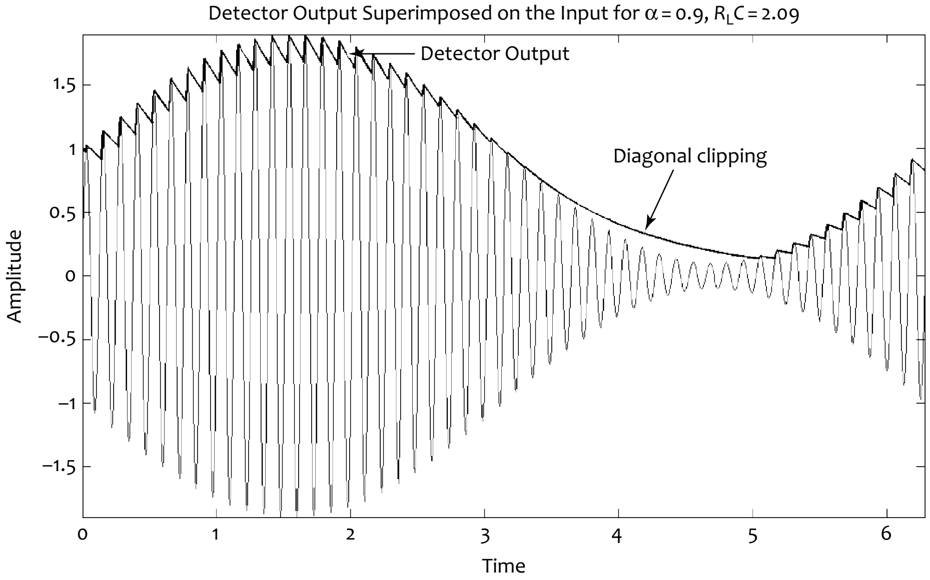
```
%
% Envelope.m detects AM waveform
%
% Part 1 of the problem
RL C = 2*pi/10;
alpha = 0.5;
Dt = 2*pi/1000;
W = 50;
global RLC, alpha, W, Dt;
t = 0:2*pi/1000:2*pi;
%
% Allocation of memory for input and output arrays
%
Vin = zeros(1,1001);
Vout = zeros(1,1001);
```

```
%
% Define input array
%
V = 1+alpha*sin(t); % modulating signal
Vin = (1+alpha*sin(t)).*sin(W*t);
%
% First point of output is the initial value of the envelope
%
Vout(1) = 1;
%
% Compute output over all points
%
for i = 2:1001
    if Vin(i) > Vout(i-1)
         Vout(i) = Vin(i)
    else
         Vout(i) = Vout(i-1)*exp(-Dt/RL C);
    end
end
% Plot input then pause
figure (1)
plot(t,Vin);
hold on
plot(t,Vout,'k','LineWidth',2);
hold on
plot(t,V,'g');
axis ([0 2*pi -1-alpha 1+alpha]);
title ('Figure 1: Detector output superimposed on the input for \alpha = 0.5,
RLC = 0.628');
xlabel('time');
ylabel('Amplitude');
legend ('Detector output','Input to the detector','Modulating Signal',0)
%
% Part II of the problem
%
RC = 3*pi/10;
alpha = 0.9;
Dt = 2*pi/1000;
W = 50;
global RLC, alpha, W, Dt;
t = 0:2*pi/1000:2*pi;
%
% Allocation of memory for input and output arrays
%
Vin = zeros(1,1001);
Vout = zeros(1,1001);
%
% Define input array
%
V = (1+alpha*sin(t))              % Modulating signal
Vin = (1+alpha*sin(t)).*sin(W*t);
%
% First point of output is the initial value of the envelope
```

```
%
Vout(1) = 1;
%
% Compute output over all points
%
  for i = 2:1001
  if Vin(i) > Vout(i-1)
     Vout(i) = Vin(i)
  else
    Vout(i) = Vout(i-1)*exp(-Dt/RL C);
  end
end
% Plot input then pause
figure (2)
plot(t,Vin);
hold on
plot(t,Vout,'k','LineWidth',2);
axis ([0 2*pi -1-alpha 1+alpha]);
title ('Figure 2: Detector output superimposed on the input for \alpha = 0.9,
R_{L}C = 2.09');
xlabel('time');
ylabel('Amplitude');
legend ('Detector output','Input to the detector','Modulating Signal',0)
```

Results





Detector Output Superimposed on the Input for $\alpha = 0.9$, $R_L C = 2.09$

Fig. 4.99

Summary

- Modulation is the process of translating a low frequency information bearing signal to a high frequency slot.
- Modulation is necessary for (a) keeping the antenna size small, (b) making it possible for the receiver to select the desired message signal, (c) multiplexing and transmitting several information bearing signals simultaneously.
- In continuous-wave modulation, the amplitude, frequency, or the phase of a high frequency sinusoidal signal, called the carrier, is changed in accordance with the variations in the amplitude of the message signal.
- *Amplitude Modulation or AM (carrier plus both sidebands):* It is that type of modulation in which the amplitude of the carrier is changed from instant to instant in such a way that at any instant of time, the *change in the peak amplitude* of the carrier from its unmodulated value is directly proportional to the instantaneous amplitude of the message/modulating signal.
- *Time-domain description of AM*: $x_c(t) = A_c[1 + mx(t)]\cos\omega_c t$, where, $x(t)$ is the message or modulating signal, $A_c \cos\omega_c t$ is the unmodulated carrier signal and 'm' is the modulation index, whose value lies between 0 and 1, i.e., $0 \le m \le 1$ and $|x(t)| \le 1$. For single-tone message signal, $x(t) = \cos\omega_m t$ so that

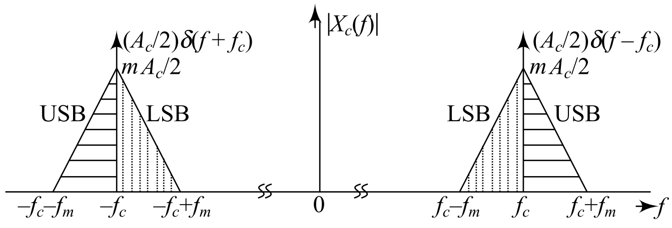
$$x_c(t) = A_c[1 + m\cos\omega_m t]\cos\omega_c t = A_c\cos\omega_c t + \frac{mA_c}{2}\cos(\omega_c + \omega_m)t + \frac{mA_c}{2}\cos(\omega_c - \omega_m)t$$

- *Frequency-domain description of AM:*

$$X_c(f) = \frac{A_c}{2}[\delta(f + f_c) + \delta(f - f_c)] + \frac{mA_c}{2}[X(f - f_c) + X(f + f_c)]$$

where for single-tone modulation, $X(f) = \frac{1}{2}[\delta(f + f_m) + \delta(f - f_m)]$

- *Amplitude spectrum of an AM signal:*



- *Carrier and sideband power components in AM:*
  (a) When a general message signal $x(t)$, with $|x(t)| \le 1$ is used:

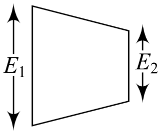
  $$\overline{x_c^2(t)} = \text{Average power in an AM signal } = \frac{1}{2}A_c^2[1+m^2\overline{x^2(t)}]$$

  where $\overline{x_c^2(t)}$ = Average power of the message signal.

  (b) For single-tone modulation,

  $$\overline{x_c^2(t)} = \text{Average power in an AM signal } = \frac{1}{2}A_c^2\left[1+\frac{m^2}{2}\right]$$

- *Trapezoidal pattern:* When $0 \le m \le 1$ and there is no distortion,



$$m = \left(\frac{E_1 - E_2}{E_1 + E_2}\right)$$

- *Generation of AM:*
  (a) By the use of non-linear devices
  (b) By the use of product devices
  (c) By the use of switching devices
- *Plate/collector–modulated class-C amplifier:* Total average power in the AM output signal = $P_0 = \eta P_{T\text{av}}$

  $$= \eta P_B + \eta P_B\left(\frac{m^2}{2}\right); \text{ where } P_{T\text{av}} = \text{total average power}$$

  Supplied to the collector/plate circuit: P_B = Power supplied by the V_{CC} or E_{bb} supply; η = plate-circuit efficiency of the modulated class-C amplifier and m = modulation index.

  $$\therefore \qquad P_0 = \text{carrier power + total sideband power}$$

  Carrier power is supplied by $E_{\text{bb}}/V_{\text{cc}}$ supply and sideband power is supplied by the final stage of the modulating amplifier.
- *Detection of AM signals:* An AM signal may be detected by (a) Coherent detection, (b) Square law detection, or (c) Envelope detection.
  (a) Coherent detector: The received signal is multiplied by a locally generated carrier signal and the product is low pass filtered using an LPF with a cut-off frequency W Hz, the baseband signal bandwidth.
  (b) Square law detector: The received AM signal is fed to a square law device and then its output is low pass filtered using an LPF with a cut-off frequency of W Hz.
  (c) Envelope detector: If there is no distortion in the modulation process, the envelope of an AM signal follows the variations in amplitude of the message signal. The diode/envelope detector tries to extract the envelope of the received AM signal. The detector consists of a diode in series with a parallel combination of R_L and C, to which the AM signal is applied. The output is taken across the parallel combination of R_L and C. It should be seen that

  $$R_sC << \frac{1}{f_c} << R_LC << \frac{1}{f_m}$$

  where R_s is the source resistance, and f_m is the highest modulating signal frequency.
- *Distortions in envelope detection*
  (a) Diagonal clipping: To avoid this, it must be ensured that

  $$R_LC \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]$$

(b) Negative-peak clipping: To avoid this, it must be ensured that

$$m_{\max} \leq \left(\frac{R_{\text{AC}}}{R_{\text{DC}}}\right); \quad R_{\text{AC}} = \text{AC load res. of the envelope detector}$$

$$R_{\text{DC}} = \text{DC load res. of the envelope detector}$$

- *Disadvantages of AM:* As the information contained in the message is completely available in any one of the two sidebands, it can be recovered even if just one sideband alone, occupying a bandwidth of W is transmitted. *Thus, AM is wasteful in power as well as bandwidth.*
- *DSB-SC:* An amplitude modulation process, in which the modulated signal contains no carrier components and has only two sidebands, is called double sideband suppressed carrier modulation.

$$\underset{\text{DSB-SC}}{x_c(t)} = A_c x(t) \cos \omega_c t$$

For $x(t)$ which is single tone: $\underset{\text{DSB-SC}}{x_c(t)} = \frac{A_c}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$

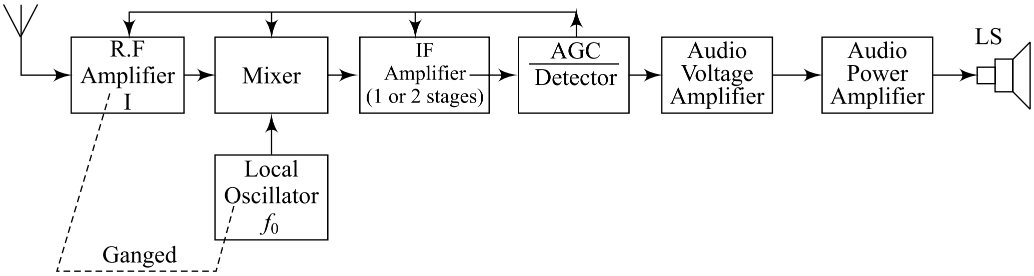
- *Generation of DSB-SC:* Since a balanced modulator multiplies the two signals given to it, it can be used for generating DSB-SC signals by giving $A_c \cos \omega_c t$, the carrier signal and $x(t)$, the message signal, as the two inputs to it. Or else, a ring modulator may be used.
- DSB-SC signals can be detected only by synchronous or coherent detection only.
- The locally generated carrier signal used for coherent detection, has to be in frequency and phase synchronism with the carrier in the received sidebands. For this purpose, a 'Costas Loop' or a 'Squaring-loop' may be used.
- Quadrature carrier multiplexing, or quadrature amplitude modulation, is a technique by which two different message signals, $x_1(t)$ and $x_2(t)$, having spectra occupying the same bandwidth, can be transmitted simultaneously over the same physical channel, using the same carrier frequency.
- A transmitter has to generate the carrier, raise its power level, process the message signal and raise its power level, and modulate the carrier at an appropriate power level.
- *Functions of a receiver:* To pick up any desired signal, amplify it, extract the message signal by demodulating the picked up signal and amplifying the message signal and operate the output device like a loudspeaker.
- *Classification of receivers:* They are classified in different ways: (a) According to the type of modulation of the received signal, (b) According to the frequency range of operation, and (c) According to the configuration of the receiver – TRF, superhet, etc.
- *AM broadcast transmitters:* Use audio frequencies up to 5 kHz, operate in MW band from 550 kHz to 1650 kHz and in SW band from 3 MHz to 30 MHz. MW band transmitters primarily depend upon ground wave propagation, while SW band transmitters depend upon skywave propagation. Carrier powers of 1 kW to 100 kW are used. Carrier frequency stability of the order of ±0.02% is mandatory. Adjacent carrier separation is 10 kHz since carrier and both sidebands are transmitted.
- *(a) High-level modulation:* In an AM transmitter, if the modulating message signal is introduced in series with the collector/plate supply voltage of the final RF power amplifier, the modulation is referred to as high-level modulation.

  (b) Low-level modulation: In an AM transmitter, if the modulating signal is introduced beyond the buffer at any point up to and including the grid/base of the final RF power amplifier, the modulation is referred to as low-level modulation.

  (c) Advantages and disadvantages: High-level modulation permits the use of class-C RF power amplifiers which are highly efficient. But it requires very large amounts of message signal power. Low-level modulation compels us to use class-A or AB type of RF power amplifiers (which are inefficient) after the modulation stage. But it does not need large amounts of message signal powers.
- *Neutralization of RF amplifiers:* RF stages in a transmitter need to be provided with neutralization circuits to prevent them from oscillating. Hazeltine and Rice methods of neutralization are quite common.

- *Negative feedback in AM broadcast transmitters:* Negative feedback is generally provided in all AM broadcast transmitters. This is done by taking a small portion of the AM signal given to the antenna, envelope detecting it and feeding the resulting audio message signal in series with the output of the audio *voltage amplifier* so as to oppose it in order to give negative feedback. This reduces the distortion of the envelope of the radiated AM signal and also reduces the noise and power frequency hum.
- *TRF receivers:* A tuned radio frequency receiver (TRF receiver) consists of RF amplifiers, a detector and audio voltage and power amplifiers. It is one of the earliest types of receiver and has very poor adjacent channel selectivity.
- *Principle of superheterodyne receiver:* In a superheterodyne receiver, the received RF signal is converted into another RF signal carrying the same message signal, but having a fixed carrier frequency called the intermediate frequency (IF) which is lower than the lowest carrier frequency covered by the receiver. Most of the gain of the receiver is obtained at the IF. This is then detected and the message signal is amplified.

  (a) Superheterodyne broadcast receiver:



- *Why local oscillator frequency f_0 is kept greater than carrier frequency f_c:* In a superheterodyne receiver, the difference between f_0 and f_c should be equal to f_{if} of the receiver. Therefore, f_0 may be greater than f_c or less than f_c. But it is always arranged to be greater than f_c as otherwise, the tuning capacitor range required will be far greater than what can be obtained in practice.
- *Adjacent channel selectivity:* When a receiver is tuned to a particular station, adjacent channel signal also will be picked up to some extent due to the inability of the receiver to totally reject it. This selectivity depends mostly on the shape of the IF amplifier's response and to some extent on the shape of the RF amplifier's response. In a good receiver, adjacent channel selectivity should be of the order of 60 to 80 dB. For this purpose, the IF amplifier response is shaped appropriately by using three or more stagger-tuned stages, or three or more identically tuned IF stages with loose coupling of the inter-stage transformers.
- *Image frequency:* If a receiver with intermediate frequency f_{if} is tuned to a carrier frequency f_c, the corresponding image frequency is $f' = (f_c + 2f_{\text{if}})$
- *Image frequency rejection ratio (IFRR):* IFRR $\underline{\Delta}\, 10\log_{10}\left|\frac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f')}\right|^2$. Its value depends upon the value of the loaded Q of the tuned circuits of the RF stages, the value of the IF of the receiver (higher the better) and on whether f_c is close to the lower end or the higher-end of the tuning range of the receiver. It should be at least 40 dB.
- *Double spotting:* The phenomenon of a desired signal f_s being received at two different dial settings of the receiver, is known as double spotting. The cause is poor image rejection.
- *Choice of IF:*

  (a) IF should be outside the tuning range of the receiver.

  (b) Lower value of IF reduces adjacent channel interference.

  (c) Higher value of IF improves image rejection.

  Usual values: 455 to 465 kHz for AM receivers and 10.7 MHz for FM receivers.
- *Tracking:* In a superheterodyne receiver, ideally, the local oscillator frequency should always keep itself above the carrier frequency f_c to which the receiver is tuned by an amount equal to IF. This is referred to as tracking. In practise, perfect tracking cannot be achieved exactly over the entire tuning range of the receiver.

- *Two-point tracking:* Perfect tracking is obtained only at two frequencies over the tuning range and at the other frequencies the difference between f_0 and f_c is kept as close as possible to the correct IF. For this purpose, a 'padder capacitor' in series with the tuning capacitor, or a 'trimmer capacitor' in parallel with the tuning capacitor are used. These are small variable capacitors.
- *Three-point tracking:* It is possible to get perfect tracking at three points over the tuning range of the receiver and only a small error at all other points, by the use of both a padder and a trimmer.
- *Double heterodyne receivers:* In VHF communication receivers requiring an IF bandwidth of only 10 kHz, double heterodyning is used in order to get good selectivity as well as good image rejection. The first IF is chosen high to get good image rejection and the second IF is chosen low to get good adjacent selectivity.
- *Receiver parameters*
  (a) Sensitivity, (b) Selectivity, (c) Fidelity, and (d) Noise Figure.
- *Single sideband suppressed carrier (SSB-SC) modulation:* It is an amplitude modulation process in which the carrier as well as one of the sidebands is suppressed and only one sideband is transmitted.
- *Frequency domain representation of SSB-SC signals:*

  (a) USSB-SC signal: $X_C^U(f) = \frac{A_c}{2}[X_+(f - f_c) + X_-(f + f_c)]$; $\begin{matrix} X_+(f) = F[x_+(t)] \\ x_+(t) = x(t) + j\hat{x}(t) \end{matrix}$

  (b) LSSB-SC signal: $X_C^L(f) = \frac{A_c}{2}[X_-(f - f_c) + X_+(f + f_c)]$; $\begin{matrix} X_-(f) = F[x_-(t)] \\ x_-(t) = x(t) - j\hat{x}(t) \end{matrix}$

- *Time-domain representation of SSB-SC signals:*

$$x_C^U(t) = \text{ USSB-SC signal } = \frac{A_c}{2}[x(t)\cos\omega_c t - \hat{x}(t)\sin\omega_c t]$$

and $$x_C^L(t) = \text{ LSSB-SC signal } = \frac{A_c}{2}[x(t)\cos\omega_c t + \hat{x}(t)\sin\omega_c t]$$

- *Generation of SSB-SC signals:* There are three methods:
  (a) Filter method: In this method, first a DSB-SC signal is generated. From this the unwanted sideband is suppressed using a filter. Its advantages are: (i) very stable, and (ii) used in commercial circuits.
  (b) Phasing method: In this, two balanced modulators, BM_1 and BM_2 are used. BM_1 is fed with $A_c\cos\omega_c t$ and $x(t)$ and its output is the product of these two. BM_2 is fed with $A_c\sin\omega_c t$ and $\hat{x}(t)$. Its output is the product of these two. Output of BM_2 is either added or subtracted from output of BM_1. Addition gives LSSB-SC while subtraction gives USSB-SC signal. Its advantages are: (i) used by radio amateurs, and (ii) needs frequent adjustment.
  (c) Weaver's method or Third method: It is a variant of the phasing method and obviates the need for wideband 90° phase shifters by using 4 BMs.
- *Detection of SSB-SC signals:*
  (a) If the locally generated carrier has a phase error θ, the detector output will be $x(t)$ with all its frequency components shifted by θ. Hence severe phase distortion results.
  (b) If the locally generated carrier has a frequency error (Δf), then the detector output will not be $x(t)$. Instead, it will be an SSB-SC signal with $x(t)$ as modulating signal and (Δf) as the carrier.
- *Applications and advantages of SSB-SC transmission:*
  (a) Useful for point-to-point communication for speech but not for audio broadcasting.
  (b) Bulk transmission of telephone conversations using FDM.
  (c) Gives better (S/N) at the destination as compared to AM.
- *Frequency Division Multiplexing (FDM):* A technique used for simultaneous transmission of a number of different message signals over the same physical channel. For this, different message signals are, by frequency translation, made to occupy different non-overlapping frequency slots and this multiplexed signal is transmitted. At the receiving end, the messages are separated by using BPFs, demodulators and lowpass filters.

- *Typical FDM hierarchy*

Type of group	Frequency range	Bandwidth	Number of telephone channels
Group	60 – 108 kHz	48 kHz	12
Super group	312 – 552 kHz	240 kHz	60
Master group	564 – 3084 kHz	2.52 MHz	600
Very large group (Jumbo group)	0.5 – 17.5 MHz	17 MHz	3600

- *SSB-SC transmitters:* Since a 200 Hz wide hole exists near the origin in the spectrum of an audio signal, filter method can be used for generating the SSB-SC signal. Initially, a low carrier frequency of 100 kHz is used to make the filter's requirements less stringent even when 40–60 dB suppression of unwanted sideband is to be achieved. High *Q* filters such as SAW filters, Crystal filters, mechanical filters and ceramic filters are used for sideband suppression. After sideband suppression the carrier frequency is raised to the required level using a crystal oscillator and a mixer.
- *SSB-SC receivers:* Since HF band is used for point-to-point communication using SSB, and since SSB signal bandwidth is only 5 kHz, it is necessary to use double heterodyne receivers. The first IF is generally 2.2 MHz and the second IF is 200 kHz.
- *Vestigial sideband modulation:* Since use of AM as well as SSB is not possible for video transmission, what is known as vestigial sideband modulation is used. In this, in addition to the carrier and one sideband, a part, or what is generally called as a vestige of the other sideband is also transmitted. It is used for TV transmission. A VSB signal may be detected using an envelope detector. The distortion in the detected signal will be small, if the depth of modulation is small.

References and Suggested Readings

1. Blake, Roy, *Electronic Communication Systems*, 2nd Edition, Thomson-Delmar, Singapore, 2005.
2. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems*, 4th Edition (International edition), McGraw-Hill, Singapore, 2002.
3. G. Proakis, John, and M. Salehi: 'Communication Systems Engineering', Second Edition, Prentice-Hall of India, Indian Reprint, 2006.
4. Haykin, Simon, *Communication Systems*, 3rd Edition, John Wiley & Sons, 1995.
5. J. Schoenbeck, Robert, *Electronic Communications – Modulation and Transmission*, 2nd Edition, Prentice Hall Career & Technology, Englewood Cliffs, New Jersey, USA, 1992.
6. Kennedy and Davis, *Electronic Communication Systems*, 4th Edition, McGraw-Hill International Editions, 1992.
7. Roddy, Dennis, and John Coolen, *Electronic Communications*, 3rd Edition (Eastern Economy Edition), Prentice-Hall of India Pvt. Ltd, New Delhi, 1994.
8. Taub, Herbert, and Donald Schilling, *Principles of Communication Systems*, Tata McGraw-Hill, Kogakusha International Student Edition, 1971.
9. Terman, F. E., *Electronic and Radio Engineering*, 4th Edition, McGraw-Hill, 1955.
10. W. Couch II, Leon, *Digital and Analog Communication Systems*, 6th Edition, Pearson Education.

Review Questions

1. Define 'amplitude modulation'.
2. What is modulation index? What happens if it is greater than unity?
3. A carrier signal $A_c \cos\omega_c t$ is amplitude modulated by a message signal $A_m \cos\omega_m t$, where, $A_m < A_c$. (a) Write down the expression for the modulated signal; (b) Write down the expression for the carrier component and the side-frequency components; (c) Draw the phasor diagram of the modulated signal.

4. From the expression for the amplitude modulated signal of Question 3 above, write down the expression from the RMS value of the modulated signal.
5. Sketch the spectrum of an AM signal assuming sinusoidal modulation with a modulation index of m ($m < 1$).
6. A carrier signal is sinusoidally modulated to a depth of $m = 0.8$. What percentage of the total power of the modulated signal is in the two sidebands?
7. State one important advantage and one important disadvantage of AM. Where is AM used?
8. What is diagonal clipping? How can it be avoided?
9. State how a DSB-SC signal may be generated.
10. Assuming sinusoidal modulation, sketch the spectrum of a DSB-SC signal for some m ($m < 1$).
11. How can a DSB-SC signal be demodulated?
12. Name one practical application in which DSB-SC modulation is put to use.
13. Briefly explain quadrature carrier multiplexing.
14. Define 'high-level modulation' and 'low-level modulation', and discuss the advantages and disadvantages of each.
15. Draw the block schematic diagram of an AM broadcast transmitter and explain the function of each block.
16. Explain the neutralization techniques adopted in the RF amplifiers of a transmitter.
17. Draw the block schematic diagram of a TRF type of AM broadcast receiver. Explain its functioning and its deficiencies.
18. What is the basic principle of a superheterodyne broadcast receiver? How does it overcome the limitations noted in the case of a TRF receiver?
19. Draw the block schematic diagram of a superheterodyne AM broadcast receiver and with its help, explain the working of the receiver.
20. Taking the case of a medium wave band superheterodyne AM broadcast receiver, explain why the local oscillator frequency is arranged to be above *and not below* the signal frequency.
21. What is meant by an image signal? What are the steps generally taken to minimize image signal interference?
22. With reference to a superheterodyne broadcast receiver, explain what is meant by tracking. How is it ensured?
23. Distinguish between two-point and three-point tracking.
24. Discuss the factors governing the choice of IF for a superheterodyne receiver.
25. Justify the following statements:
    (a) Good image signal suppression requires that the IF be high.
    (b) Good adjacent channel selectivity can be obtained by choosing a low value of IF.
26. Define and explain the terms: 'Sensitivity', 'Selectivity', and 'Fidelity'. What are the various factors that influence these parameters?
27. Discuss the advantages and disadvantages of SSB-SC transmission.
28. In the filter method of generation of an SSB-SC signal, why do we have to use a low frequency carrier initially?
29. In the filter method of generation of an SSB-SC signal, why is it necessary that the message signal should have a hole near the origin in its spectrum?
30. How is an SSB-SC signal demodulated?
31. With reference to SSB-SC signal modulation, discuss the effect of an error in the locally generated carrier signal's (a) frequency, (b) phase.
32. State the applications of SSB transmission.
33. Draw the spectrum of an LSSB-SC signal. Write down an expression for this spectrum in terms of that of the message signal.
34. How does the two-stage SSB-SC modulation overcome the problems associated with the design of the sideband suppression filter?
35. Critically compare the filter method and the phasing method of generation of SSB-SC signals.
36. With the help of block schematic diagram, clearly explain homodyne detection of an SSB signal transmitted with a pilot carrier.
37. Explain briefly the basic principle of FDM.
38. Explain why SSB transmission even with a pilot carrier is not feasible in the case of TV.
39. With the help of a neat block schematic diagram, explain the working of an SSB-SC transmitter.

40. Clearly explain the need for the following: 'SSB transmitters use a low-frequency carrier initially'.
41. In SSB-SC transmitters using filter method of generation of the SSB signal, sideband filters have to be used for suppression of the unwanted sideband. What type of filters are used and why?
42. Draw the block schematic of a pilot-carrier SSB transmitter.
43. With the help of a neat block schematic diagram, explain the working of an SSB-SC receiver.
44. Draw the block diagram of a pilot-carrier SSB receiver and explain its working.
45. Sketch the typical spectrum of the VSB signal that is given as input to the video detector of a TV receiver.
46. Write down an expression for the time-domain representation of a VSB signal.
47. Sketch the spectrum of typical TV signal.
48. Sketch the typical response characteristic of the video amplifier section of a TV receiver.
49. What are the steps taken in commercial TV broadcasting to ensure that the distortion arising in the detected video signal owing the use of an envelope detector is within tolerable limits?

Problems

1. An AM signal is given by $x_c(t) = [30 + 9\cos 2000\pi t + 12\cos 3000\pi t]\cos 2\pi \times 10^5 t$.
   (a) Sketch the spectrum of the modulated signal.
   (b) Determine the effective modulation index.
   (c) Determine the carrier power and total sideband power.
2. A class-C collector modulated class-C amplifier is producing an AM signal at its output with a carrier component of power equal to 50 Watts. The modulating amplifier is a class-A power amplifier. If the class-C amplifier has an efficiency of 75% and the class-A amplifier has an efficiency of 40%, determine (a) the total input d.c. power for the two amplifiers, and (b) the dissipation in each of the devices used for the class-C and the class-A amplifiers, for modulation indices of (i) 40% and (ii) 100%.
3. A square law device has an input-output relation given by $e_0 = a_1 e_{\text{in}} + a_2 e_{\text{in}}^2$. To this device, we give an input signal which is the sum of the message signal, $x(t) = 0.3\cos 2\pi 50t + 0.4\cos 2\pi\, 150t$ and a carrier signal of frequency 5 kHz. The output signal $e_0(t)$ is then subjected to band pass filtering. What should be the center frequency and the bandwidth of this BPF if the output of the filter is to be an AM signal?
4. The square law device of Problem 3 is now proposed to be used for detection of an AM signal given by $e_{\text{in}}(t) = x_c(t) = A_c[1 + mx(t)]\cos 2\pi f_c t$ (a) determine $e_0(t)$, and (b) What are the conditions to be satisfied if the message signal $x(t)$ is to be recovered?
5. A carrier signal of frequency, f_c, is DSB-SC modulated using the message signal $x(t) = 10\,\text{sinc}\, 2 \times 10^3 t$. The resulting modulated signal is to be demodulated using a coherent detector whose locally generated carrier may be assumed to be in perfect synchronism with that of the modulator. Determine the lowest value of f_c for which the coherent detector output yields $x(t)$.
6. $x(t) = (\cos 2\pi \times 500t + 2\cos 2\pi \times 1000t)$ DSB-SC modulates the carrier $c(t) = 50\cos 2\pi \times 10^5 t$. Find the expressions for the USSB-SC and LSSB-SC components of the modulated signal, and sketch their spectra.
7. A message signal $x(t)$ is positive for all t. This message DSB-SC modulates a carrier signal. Show that an envelope detector can be used to demodulate this DSB-SC signal.
8. A carrier of frequency f_c = 100 kHz is DSB-SC modulated by a message signal $x(t) = \cos 2000\pi t + 2\cos 4000\pi t$ to give a modulated signal $x_c(t) = 50x(t)\cos 2\pi \times 10^5 t$.
   (a) Sketch the spectrum of $x_c(t)$, the modulated signal.
   (b) Find the average powers of all the frequency components in $x_c(t)$.
9. A superheterodyne receiver has an IF of 460 kHz. Its RF amplifier is tuned to an incoming signal of carrier frequency 700 kHz. If at this frequency the tuned circuit of the RF amplifier has a Q of 60, determine the image frequency rejection in dB.
10. A double conversion receiver is tuned to an incoming signal of 25 MHz at which frequency its tank circuit has a Q of 65. The receiver is using a first IF of 1.5 MHz and a second IF of 150 kHz. Calculate (in decibels) the image frequency rejection. Make reasonable assumptions, if necessary.

11. A carrier signal of frequency $f_c = 10^5$ Hz is LSSB-SC modulated by a message signal given by $x(t) = \cos 2000\pi t + 2\cos 4000\pi t + 3\cos 6000\pi t$. Sketch the two-sided spectrum of the modulated signal. If the carrier peak amplitude, $A_c = 50$, what is the average power of the modulated signal? What is its bandwidth?
12. For the carrier signal and the message signal given in Problem 10, determine the time-domain expression for the USSB-SC signal by first determining $\hat{x}(t)$ and sketch its two-sided spectrum.
13. A message signal $x(t)$ having a bandwidth of 5 kHz has been normalized so that $|x(t)| \le 1$ for all t. This normalized message, having an average power of 1 W modulates the carrier signal

$$c(t) = 20\cos 2\pi f_c t$$

Determine the average power in the modulated signal if the modulation is
(a) SSB-SC (b) DSB-SC (c) AM with a modulation index $m = 0.8$
14. A two-stage SSB-SC modulator is shown in Fig. 4.64. The message signal, $x(t)$, is a voice signal with frequency components from 0.3 kHZ to 3.5 kHz. If the carrier frequency f_1 is 100 kHz and the high frequency oscillator frequency, f_2, is 5 MHz, and if the final output signal is to be a USSB-SC signal, specify the details of the two sideband filters.
15. Equation (4.74) gives the time-domain representation of a USSB-SC signal in terms of the message signal $x(t)$, its Hilbert transform $\hat{x}(t)$ and the carrier frequency f_c. Using that equation, derive the expression for the message signal $x(t)$ in terms of the USSB-SC signal $x_c^u(t)$, its Hilbert transform and the carrier frequency. This expression for $x(t)$ suggests a method of demodulating $x_c^u(t)$. Draw the block schematic diagram of such a demodulator.
16. A scrambler is a system used for privacy of communication. In the two-stage SSB-SC generator of Fig. 4.60, assume that the message signal has an amplitude spectrum as shown in Fig. P4.1; that the first oscillator frequency $f_1 >> W$, that the first sideband filter passes only the upper sideband, that the second sideband filter is a low pass filter with a cutoff frequency of W, and that the second oscillator frequency $f_2 = f_1 + W$. Show that the two-stage SSB-SC generator now works as a scrambler by determining and sketching the spectrum of its output signal. Show that the same set-up may be used for unscrambling this output signal.

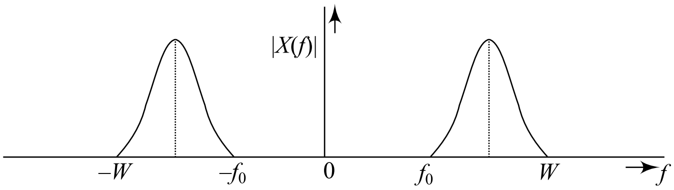


Fig. P4.16

17. An SSB transmitter uses a set-up of the form shown in Fig. P4.17 to generate the SSB signal using filter method.

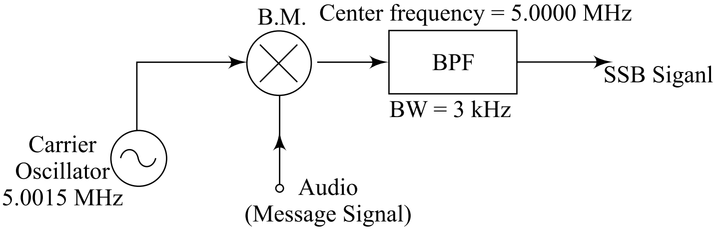


Fig. P4.17

For the values given in the figure, determine
(a) Whether the lower sideband or the upper sideband will be produced.
(b) The carrier frequency value if the other sideband is to be produced.

Multiple-Choice Questions

1. In an amplitude modulated wave obtained by sinusoidal modulation of the carrier, the positive peak amplitude of the RF is varying between 12 V and 4 V. The modulation index and the unmodulated carrier amplitude are respectively
(a) 1/3, 8 V (b) 0.5, 8 V (c) 0.5, 4 V (d) 1/3, 4 V

2. An amplitude modulated wave is given by

$$x_c(t) = 10\cos 1200\pi t + 40\cos 1400\pi t + 10\cos 1600\pi t$$

The modulating signal frequency and modulation index are
(a) 200 Hz, 0.5 (b) 400 Hz, 0.25 (c) 200 Hz, 0.25 (d) 400 Hz, 0.5

3. To save transmitted power, the carrier of an AM signal obtained by sinusoidal modulation to a depth of modulation equal to 1, has been recovered. The percentage saving in power is
(a) 33.33 (b) 50 (c) 66.66 (d) 100

4. A collector modulated class-C amplifier is drawing 50 W from the V_{cc} supply. If an output AM wave with 100% modulation is obtained, the average power supplied by the final modulating power amplifier stage is
(a) 50 W (b) 16.66 W (c) 33.33 (d) 25 W

5. When sinusoidally modulated, the RMS value of the current in the antenna of an AM transmitter increases 15% over its unmodulated value. The modulation index is
(a) 0.6 (b) 0.8 (c) 0.5 (d) 0.707

6. Two sinusoidal signals are simultaneously modulating a carrier, the modulation indices being 0.3 and 0.4. The overall modulation index is
(a) 0.5 (b) 0.1 (c) 0.7 (d) 0.12

7. When the modulation index is halved, it is found that the antenna current (RMS value) is also halved. The type of modulation used is
(a) AM (carrier plus both sidebands) (b) Single sideband plus carrier
(c) SSB-SC (d) Vestigial sideband

8. In an AM transmitter employing low-level modulation, the amplifiers following the modulator stage have to be
(a) frequency multipliers (b) linear tuned class-A or class-AB amplifiers
(c) class-C amplifiers (d) class-B amplifiers

9. The advantages of base modulation over collector modulation of a class-C amplifier is
(a) better linearity of the modulation characteristic (b) better efficiency of the class-C modulated amplifier
(c) it requires lower modulating signal power (d) it gives more output power

10. An RF amplifier of a superheterodyne receiver
(a) helps in image signal suppression (b) improves the adjacent channel selectivity
(c) makes it easier to align the receiver (d) improves the fidelity of the receiver considerably

11. In an AM broadcast superheterodyne receiver the local oscillator frequency is arranged to be higher than the incoming signal frequency in order to
(a) provide better image rejection
(b) make tracking easier
(c) produce the correct intermediate frequency, since a lower LO frequency will not permit generation of correct IF
(d) enable us to cover the required tuning range with the practically possible ratio of maximum to minimum values of the variable capacitors

12. A low IF will
(a) improve the image signal rejection capability of the receiver
(b) improve adjacent channel selectivity
(c) make it difficult to get good sensitivity for the receiver
(d) improve the fidelity of the receiver

13. The occurrence of double spotting indicates
(a) that the IF is too high
(b) that the selectivity is poor

(c) that image rejection capability of the receiver is inadequate
(d) that the local oscillator frequency is less than that of the incoming signal

14. Double conversion superheterodyne receivers use
(a) a high first IF and a lower second IF
(b) a low first IF and a higher second IF
(c) a low IF for the first as well as the second IF stages
(d) a high IF for both the first and second IF stages

15. Harmonic generators use
(a) class-A amplifiers (b) class-AB amplifiers
(c) class-B amplifiers (d) class-C amplifiers

16. The most noisy stage of an AM broadcast receiver is
(a) the RF stage (b) the mixer stage (c) the IF stage (d) the audio stage

17. The noise figure of a superheterodyne receiver is mostly controlled by
(a) the RF stage (b) the mixer stage (c) the IF stage (d) the audio stage

18. A superheterodyne AM broadcast receiver has an IF of 455 kHz. If it is tuned to a frequency of 700 kHz, the image frequency is
(a) 1610 kHz (b) 1155 kHz (c) 245 kHz (d) 210 kHz

19. The stage contributing significantly to the sensitivity of a superheterodyne AM broadcast receiver is the
(a) RF stage (b) mixer stage (c) IF stage (d) detector stage

20. A high value of IF for a superheterodyne receiver
(a) improves image frequency rejection (b) improves the selectivity
(c) improves the sensitivity (d) improves the fidelity

21. For broadcasting, AM is preferred to SSB because
(a) AM signal is easy to generate (b) AM gives better signal-to-noise ratio
(c) SSB receivers are complex and expensive (d) AM transmitters do not need expensive filters

22. In filter method of generation of SB-SC, the type of filters that can be used are
(a) LC filter (b) crystal filters (c) RC filters (d) active filters

23. In the filter method of generation of SSB-SC, in order to make the filter specifications less stringent
(a) it is ensured that the modulating signal has no high frequency components
(b) a high frequency carrier is used initially for generating the DSB-SC signal
(c) only those modulating signals which have a high dc and low frequency content are used
(d) a low frequency carrier is used initially for generating the DSB-SC signal

24. The ‘third method’, or the Weaver’s method, has the following advantage over the ‘Phasing method’:
(a) It does not need wideband 90° phase-shifters
(b) It gives better carrier stability
(c) It gives much better suppression of the unwanted sideband
(d) It does not need frequent adjustments

25. SSB-SC modulation is not used for audio broadcasting because
(a) it is difficult to generate SSB-SC signals
(b) it makes the receiver circuit quite complex and expensive
(c) SSB-SC modulation cannot be used for speech signals

26. Vestigial sideband modulation is generally used for
(a) TV broadcasting (b) point-to-point communications
(c) telemetering (d) stereo broadcasting

27. In a diode detector circuit, if the ac load for the diode is very much smaller than the dc load, it can result in
(a) poor sensitivity of the receiver (b) poor AGC
(c) diagonal clipping (d) negative peak clipping

28. An SSB-SC signal may be demodulated using a
(a) diode envelope detector (b) synchronous detector
(c) ratio detector (d) None of these

Key to Multiple-Choice Questions

1. (b)	2. (a)	3. (c)	4. (d)	5. (b)	6. (a)	7. (c)	8. (b)
9. (c)	10. (a)	11. (d)	12. (b)	13. (c)	14. (a)	15. (d)	16. (b)
17. (a)	18. (a)	19. (c)	20. (a)	21. (c)	22. (b)	23. (d)	24. (a)
25. (b)	26. (a)	27. (d)	28. (b)				

ANGLE MODULATION

5

> *"The only way of finding the limits of the possible is by going beyond them into the impossible."*
>
> **Arthur C. Clarke (1917–2008)**
> *British science fiction author, inventor, and futurist*

Learning Objectives

After going through this chapter, students will be able to

- *understand the concept of angle-modulation and the difference between frequency modulation and phase modulation,*
- *derive expressions for frequency modulated and phase modulated waves,*
- *draw the block diagrams of the direct and indirect methods of generation of WBFM signals and explain the operation of various types of FM detectors,*
- *derive the expression for the spectrum of an angle-modulated wave for a single-tone modulating signal and can find the effective bandwidth of the modulated signal,*
- *recognize the key specifications for an FM audio broadcast transmitter, draw its block diagram and explain its working, and*
- *draw the block diagram of an FM stereo transmitter and receiver and explain their working.*

5.1 INTRODUCTION

In Chapter 4, we had considered amplitude modulation, wherein, the carrier signal amplitude is changed in accordance with the variation in amplitude of the message signal. As had been stated there, this is only one way of modulating the carrier signal. Instead of the amplitude, if the frequency of the carrier is varied in accordance with the variations of the amplitude of the modulating signal, we call it 'frequency modulation'; and if it is the phase of the carrier that is changed as per the variations of the amplitude of the modulating signal, we call it 'phase modulation'. Since both of these ultimately vary the phase angle of the carrier signal, although in different ways, and are closely related, both of these modulations are together referred to as 'Angle Modulation'.

Amplitude modulation is sometimes referred to as a linear modulation, although, strictly speaking, it is not a linear one. Angle modulation, as we are going to see, is however, highly non-linear. This makes the analysis of angle modulation quite involved for a general class of modulating signals, thus forcing us to go in for an approximate analysis. Further, an angle modulated signal has *theoretically* an infinite bandwidth even for a single-tone modulating signal, thus compelling us to talk of its "effective bandwidth", a finite bandwidth within which a large percentage (generally more than 98%) of the average power of the modulated signal lies. This effective bandwidth of an angle modulated signal is very much larger than that of an AM signal for the same modulating signal bandwidth. Also the complexity of implementation is generally much more for angle modulation as compared to the AM. But, it has two great advantages which make it very attractive for certain applications.

1. As we are going to see in Chapter 7 which discusses the noise performance of amplitude and frequency modulation systems, frequency modulation systems have in general, better noise immunity as compared to the AM systems. Further FM systems offer a BW-to-(S/N) trade-off which makes it possible to operate an FM transmitter at a relatively low power and still maintain the required (S/N) ratio at the destination provided we are prepared to pay the price for it in terms of larger transmission bandwidth.
2. Unlike in AM where the transmission bandwidth increases in proportion to the message signal bandwidth, in the case of FM, the transmission bandwidth is, by and large, unaffected by the message bandwidth.

These advantages make FM extremely useful for high fidelity broadcasting of music and a few other applications.

Our discussion in this chapter will be mainly focused on the various modulation and demodulation techniques and the theories behind them. While it is true that modulation and demodulation are the most important operations taking place at the transmitter and receiver respectively, it is however the various processes taking place at the transmitter and at the receiver all of which together make it possible to have good communication. Hence, towards the end of this chapter, we will discuss FM transmitters and receivers in some detail. Of course, it is neither necessary, nor possible to go into the details at circuit level. We will confine the discussion only to the block schematic diagrams level.

5.2 ANGLE MODULATED SIGNALS

Consider a carrier signal

$$c(t) = A_c \cos \omega_c t \tag{5.1}$$

When this carrier is angle modulated, the modulated signal may be represented by

$$x_c(t) = A_c \cos \theta(t)$$

$$\therefore \qquad \theta(t) = \omega_c t + \phi(t) \tag{5.2}$$

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] \tag{5.3}$$

$\phi(t)$, the change in phase of the modulated signal from its unmodulated value (i.e., $\omega_c t$), is called the phase deviation.

5.2.1 Phase Modulation

In phase modulation the phase deviation, $\phi(t)$ is varied in such a way that at any instant of time, t, it is proportional to the instantaneous amplitude of the modulating signal, $x(t)$.

Hence,
$$\phi(t) = k_p x(t) \tag{5.4}$$

where
$$k_p \triangleq \text{ phase deviation constant}$$

It represents the change in phase angle per unit amplitude of the modulating signal $x(t)$ and has the units of radians per volt.

The phase modulated signal may therefore be written as

$$\underset{P.M}{x_c(t)} = A_c \cos[\omega_c t + k_p x(t)] \tag{5.5}$$

5.2.2 Frequency Modulation

In understanding ‘frequency modulation’, the concept of ‘instantaneous frequency’ plays a very vital role. As our concept of frequency itself is that it represents the number of full cycles completed *per second*, the term, '*instantaneous frequency*' may, at first, sound a little odd. But, when the term, 'instantaneous speed' does not sound odd even though speed is defined in much the same way as frequency has been, as the distance covered in a unit time, why should 'instantaneous frequency' sound odd? When speed v, is varying with time and is denoted by $v(t)$, a function of time, we know that the distance, $s(t)$ covered in say t seconds, is given by

$$s(t) = \int_0^t v(\alpha)d\alpha$$

and $$v(t) = \frac{ds(t)}{dt} = \text{speed at the instant } t$$
$$= \text{rate of change of distance with respect to time.}$$

Exactly in the same way since $\theta(t) = \omega t$ and $\omega = 2\pi f$, if the frequency is varying with respect to time, we write

$$\theta(t) = 2\pi\int_0^t f(\alpha)d\alpha \ = \text{Phase angle at the instant } t$$

and $$f(t) = \text{Frequency at the instant } t = \frac{1}{2\pi}\frac{d\theta(t)}{dt}$$

Thus, instantaneous frequency of a signal is defined as 1/2π times the rate of change of its phase angle.

Definition In frequency modulation, the instantaneous frequency of the modulated wave changes in such a way that at any instant, the change from the unmodulated carrier frequency is directly proportional to the instantaneous amplitude of the modulating signal, $x(t)$.

But $$x_c(t) = A_c \cos\theta(t) \ = \text{modulated signal}$$

Therefore, its instantaneous frequency $f_i(t)$ is given by

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = \frac{1}{2\pi}\frac{d}{dt}[\omega_c t + \phi(t)] = f_c + \frac{1}{2\pi}\frac{d}{dt}\phi(t) \tag{5.6}$$

From the definition of frequency modulation given above, the change in $f_i(t)$ from f_c, the unmodulated carrier frequency, called the frequency deviation, should be proportional to the amplitude of $x(t)$.
Thus, from Eq. (5.6), we have

$$\frac{1}{2\pi}\frac{d}{dt}\phi(t) = k_f x(t) \tag{5.7}$$

∴ $$f_i(t) = f_c + k_f x(t) \tag{5.8}$$

where k_f represents the change in instantaneous frequency for a unit amplitude of the modulating signal with units of Hertz/volt, and is referred to as the '*frequency deviation constant*'.

and $$\theta(t) = \int_0^t 2\pi f_i(t)dt + \phi_0 \tag{5.9}$$

where ϕ_0 is a constant reference phase, generally taken as zero without any loss of generality.

Hence, the FM signal may be represented as

$$\underset{FM}{x_c(t)} = A_c \cos\left[\int_0^t 2\pi f_c dt + \int_0^t 2\pi k_f x(\alpha)d\alpha\right]$$

Thus, $$\underset{FM}{x_c(t)} = A_c \cos\left[\omega_c t + 2\pi k_f \int_0^t x(\alpha)d\alpha\right] \tag{5.10}$$

From Eqs. (5.5) and (5.10), we find that $\phi(t)$ of Eq. (5.3) is given by

$$\phi(t) = \begin{cases} k_p x(t) & \text{for PM} \\ 2\pi k_f \int_0^t x(\alpha)\, d\alpha & \text{for FM} \end{cases} \tag{5.11}$$

The above equation clearly brings out the different ways adopted by PM and FM to change $\theta(t)$ using the modulating signal $x(t)$. It also clearly shows that a phase modulator can indeed be used for producing frequency modulation and vice versa. If the message signal, $x(t)$, is integrated and given as the modulating signal to a phase modulator, the output modulated signal will be a frequency modulated signal. Conversely, if the message signal $x(t)$ is differentiated and then fed as the modulating signal to a frequency modulator, the modulated signal that we get would be a phase modulated signal.

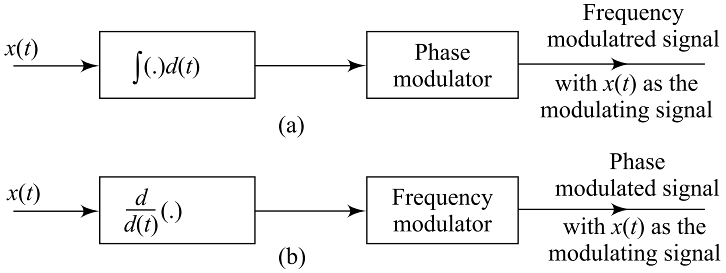


Fig. 5.1 *(a) Frequency modulation using a phase modulator, (b) Phase modulation using a frequency modulator*

Example 5.1 An angle-modulated signal is given by

$$x_c(t) = 6\cos[2\pi \times 10^7 t + 0.2\sin(10^4)\pi t]$$

(a) if $x_c(t)$ is a phase-modulated signal with $k_p = 5$ rad/volt; and
(b) if $x_c(t)$ is a frequency-modulated signal with $k_f = 5 \times 10^2$ Hz/volt
In each case, determine the modulating signal $x(t)$.

Solution

$$\underset{\text{FM}}{x_c(t)} = A_c \cos\left[\omega_c t + 2\pi k_f \int_0^t x(\alpha)\, d\alpha\right] \qquad \text{(From Eq. (5.10))}$$

$$\underset{\text{PM}}{x_c(t)} = A_c \cos[\omega_c t + k_p x(t)] \qquad \text{(From Eq. (5.5))}$$

(a) *For PM*

Compare the above equation for $\underset{\text{PM}}{x_c(t)}$ with the given equation of the angle modulated signal. If we take $x(t) = A_m \sin 10^4 \pi t$, it means that $k_p \cdot A_m = 0.2$
But k_p is given to be 5. $\therefore$ $5A_m = 0.2$ or $A_m = 0.2/5 = 0.04$
$\therefore$ the modulating signal $x(t)$ in this case is

$$x(t) = 0.04 \sin 10^4 \pi t$$

(b) *For FM*

Compare $\underset{\text{FM}}{x_c(t)}$ as given by Eq. (5.10) with the given $x_c(t)$.

Let the modulating signal be $A_m \cos 2\pi \times 5 \times 10^3 t$.

$$\therefore \qquad 2\pi k_f \int_0^t A_m \cos 10^4 \pi\alpha d\alpha = \left(\frac{k_f A_m}{f_m}\right) \sin 10^4 \pi t, \text{ where } f_m = 5 \times 10^3$$

From the expression for $x_c(t)$, we therefore have

$$\frac{k_f A_m}{f_m} . \sin 10^4 \pi t = 0.2 \sin 10^4 \pi t$$

$$\therefore \qquad \frac{k_f A_m}{f_m} = 0.2. \text{ Substituting values of } k_f \text{ and } f_m, \text{ we get}$$

$$A_m = \frac{0.2 \times 5 \times 10^3}{5 \times 10^2} = 2$$

$\therefore$ the message signal, in the case of FM is $x(t) = 2 \cos 2\pi \times 5 \times 10^3 t$

Example 5.2 The message signal shown in the following figure phase modulates a carrier signal $A_c \cos \omega_c t$, where f_c = 1 MHz. If a maximum frequency deviation of 80 kHz is needed, determine the value of the phase constant k_p to be used by the modulator. With this value of k_p, what will be the range of variation of the carrier frequency?

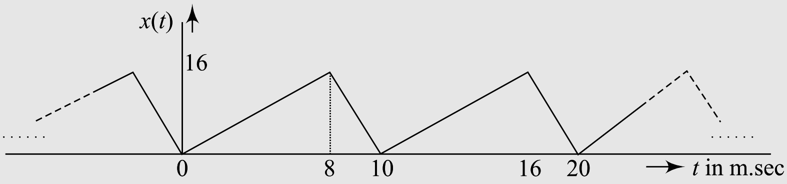


Fig. 5.2 *Signal for Example 5.2*

Solution The modulated signal $x_c(t)$ is given by

$$x_c(t) = A_c \cos(\omega_c t + k_p x(t))$$

$$\therefore \text{ instantaneous frequency, } \qquad f_i = \frac{1}{2\pi} \frac{d}{dt}[\omega_c t + k_p x(t)]$$

$$\therefore \qquad f_i = f_c + \frac{1}{2\pi} k_p \frac{d}{dt} x(t)$$

$$\therefore \qquad (f_i - f_c) = \text{Maximum frequency deviation}$$

$$= \frac{1}{2\pi} k_p \left| \frac{d}{dt} x(t) \right|_{\max}$$

$$\therefore \qquad \left| \frac{d}{dt} x(t) \right|_{\max} = \frac{16}{2 \times 10^{-3}} \qquad \text{(From the waveform of } x(t))$$

$$= 8000 \text{ v/sec}$$

$$\therefore \text{ maximum frequency deviation } = \frac{1}{2\pi} \cdot k_p \cdot 8000 = 80 \times 10^3$$

$$\therefore \qquad k_p = \frac{80 \times 10^3 \times 2\pi}{8000} = 20\pi \text{ rad/volt}$$

From $t = 0$ to $t = 8$ m.sec $\qquad \frac{d}{dt}x(t) = 2\text{v/m.sec} = 2000\text{v/sec}$

$\therefore$ during this period, frequency deviation

$$= \frac{1}{2\pi} k_p \frac{d}{dt} x(t) = \frac{1}{2\pi} \times 20\pi \times 2000 = 20 \text{ kHz}$$

Hence, from 0 m.sec to 8 m.sec, the frequency of the modulated signal is

1000 kHz + 20 kHz = 1020 kHz

From 8 m.sec to 10 m.sec, the frequency deviation is negative and has a value of 80 kHz. Hence, during this period the frequency of the modulated wave is

1000 kHz – 80 kHz = 920 kHz

The frequency of the modulated signal varies between 920 kHz and 1020 kHz.

5.2.3 Angle Modulated Signals for Some Simple Modulating Waveforms

Sinusoidal modulating signal

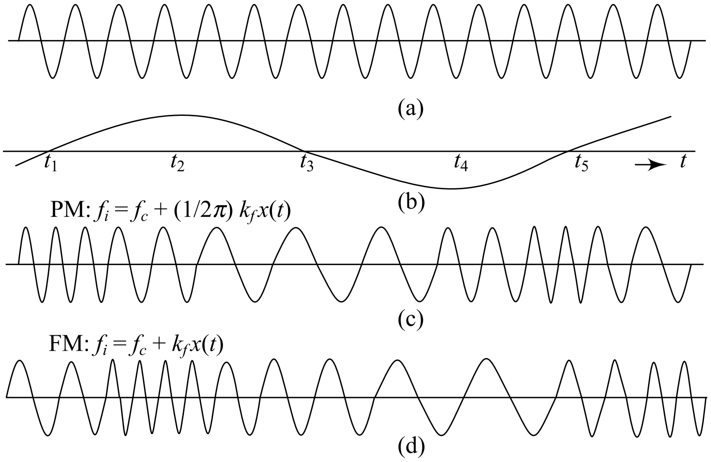


Fig. 5.3 *(a) Carrier signal, (b) Modulating sinusoidal signal, (c) Phase modulated signal, (d) Frequency modulated signal*

Unit-step function

(See Fig. 5.4 on next page)

5.2.4 Modulation Indices for FM and PM

For a single-tone message signal

Let $\qquad x(t) = A_m \cos(2\pi f_m t) \qquad (5.12)$

Then from Eq. (5.11) we have, for PM

$$\phi(t) = k_p x(t) = k_p A_m \cos(2\pi f_m t) \qquad (5.13)$$

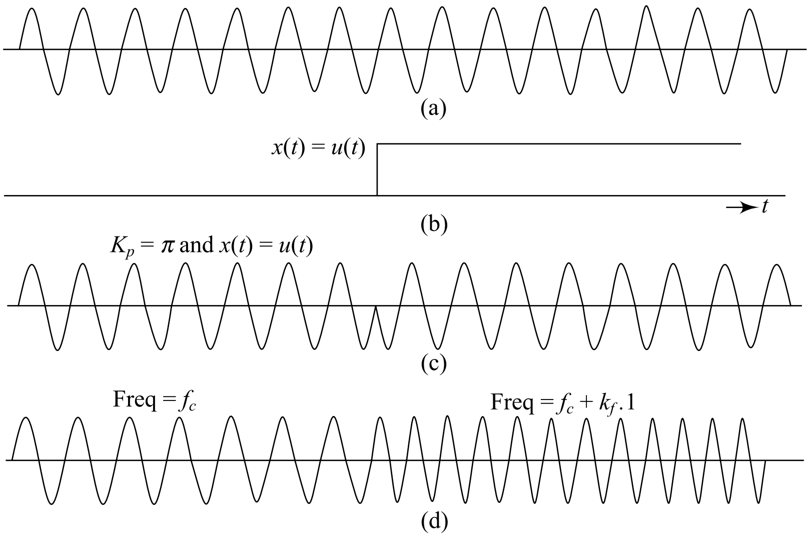


Fig. 5.4 *(a) Carrier signal, (b) Unit-step modulating signal, (c) Phase modulated signal, (d) Frequency modulated signal*

and for FM

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha)d\alpha = \frac{k_f A_m}{f_m}\sin(2\pi f_m t) \tag{5.14}$$

Hence, from Eqs. (5.5) and (5.10), we may write the modulated signals as

$$\text{PM:} \qquad x_c(t) = A_c \cos[2\pi f_c t + k_p A_m \cos(2\pi f_m t)] \tag{5.15}$$

$$\text{FM:} \qquad x_c(t) = A_c \cos\left[2\pi f_c t + \frac{k_f A_m}{f_m}\sin(2\pi f_m t)\right] \tag{5.16}$$

If we now define

$$\beta_p \triangleq \text{ Modulation index for PM} = k_p A_m \tag{5.17}$$

and

$$\beta_f \triangleq \text{ Modulation index for FM } = \frac{A_m k_f}{f_m} \tag{5.18}$$

Then the corresponding modulated signals may be written as

$$\underset{\text{PM}}{x_c(t)} = A_c \cos[2\pi f_c t + \beta_p \cos(2\pi f_m t)] \tag{5.19}$$

$$\underset{\text{FM}}{x_c(t)} = A_c \cos[2\pi f_c t + \beta_f \sin(2\pi f_m t)] \tag{5.20}$$

Note

(*i*) *Since k_p, the phase deviation constant represents, as pointed out earlier (see Eq. (5.4)), the phase deviation produced in the carrier per unit amplitude of the modulating signal, the parameter β_p of Eq. (5.17) represents the maximum phase deviation.*

(*ii*) *From Eq. (5.8), it is clear that $k_f A_m$ represents the peak frequency deviation. Referring to Eq. (5.18) then β_f, the modulation index for FM represents the ratio of the peak frequency deviation to the frequency of the modulating single-tone signal. This ratio is called the 'deviation ratio'.*

For a general modulating signal Having seen the physical meaning of the modulation indices β_{p} for PM, and β_f for FM, in the case of a single-tone modulating signal, we may now extend the concept of modulation index to a general modulating signal by defining β_p and β_f as follows:

$$\beta_p \triangleq k_p \max[|x(t)|] = (\Delta\phi)_{\max} \tag{5.21}$$

and

$$\beta_f \triangleq \frac{k_f \max[|x(t)|]}{W} = \frac{(\Delta f)_{\max}}{W} \tag{5.22}$$

where W represents the bandwidth of the modulating signal, $(\Delta\phi)_{\max}$ represents the peak phase deviation for PM and $(\Delta f)_{\max}$, the peak frequency deviation for FM.

Example 5.3 An FM transmitter has a frequency deviation constant of 100 Hz/volt. To the modulator of this transmitter, a sinusoidal modulating signal of r.m.s. value 2 V and a frequency of 1 kHz, is applied. Determine the peak frequency deviation and the deviation ratio.

Solution Peak amplitude of the modulating signal = $2\sqrt{2}$ V
Deviation constant k_f of the modulator = 100 Hz/volt.
$\therefore$ peak frequency deviation = $2\sqrt{2} \times 100 = 200\sqrt{2}$ Hz

$$\text{Deviation ratio } = \left[\frac{\text{Peak frequency deviation}}{\text{modulating signal frequency}}\right]$$

$$= \frac{200\sqrt{2}}{1000} = \frac{\sqrt{2}}{5}$$

Example 5.4 A frequency modulated signal is given by

$$x_c(t) = 10\cos[2\pi \times 10^8 t + 5\sin 2\pi \times 200t]$$

Determine (a) the carrier frequency, (b) the modulating signal frequency, (c) the peak frequency deviation, and (d) the modulation index.

Solution

(a) f_c = 100 MHz = 10^8 Hz

(b) f_m = 200 Hz

(c) $\beta_f = \dfrac{\text{Peak req. deviation}}{\text{modulating signal frequency}} = 5$

$\therefore$ peak frequency deviation = $5 f_m = 5 \times 200 = 1$ kHz

(d) Modulation index $\beta_f = 5$ as stated in (c)

Example 5.5 An FM transmitter is operating with the maximum frequency deviation of 75 kHz. What will be the modulation index if a sinusoidal signal is used for modulation and it has a frequency of (a) 100 Hz, and (b) 20 kHz.

Solution

(a) $\beta_f = \dfrac{\text{Frequency deviation}}{\text{Modulating signal frequency}} = \dfrac{75 \times 10^3}{100} = 750$

(b) $\beta_f = \dfrac{\text{Frequency deviation}}{\text{Modulating signal frequency}} = \dfrac{75 \times 10^3}{20 \times 10^3} = 3.75$

Example 5.6 An FM signal with single-tone modulation has a frequency deviation of 15 kHz and a bandwidth of 50 kHz. Find the frequency of the modulating signal.

Solution From Carson's rule, BW $= 2(\beta_f + 1)f_m = 50 \times 10^3$

$$\therefore \qquad 2\beta_f + 2 = \frac{50 \times 10^3}{f_m}, \text{ but } \beta_f = \frac{\Delta f}{f_m} = \frac{15 \times 10^3}{f_m}$$

$$\therefore \qquad \frac{2 \times 15 \times 10^3}{f_m} + 2 = \frac{50 \times 10^3}{f_m}$$

Multiplying throughout by f_m, we get

$$2f_m = 50 \times 10^3 - 30 \times 10^3 = 20 \times 10^3$$

$$\therefore \qquad f_m = 10 \times 10^3 = 10 \text{ kHz}$$

Example 5.7 A signal $x(t) = 5\cos 20\pi \times 10^3 t$ angle modulates a carrier signal $A_c \cos \omega_c t$. Determine the modulation index and the bandwidth of the modulated signal for (a) an FM system with $k_f = 12$ kHz/volt, and (b) a PM system with $k_p = 1.0$ rad/volt.

Solution

(a) β_f = Modulation index $= \left(\frac{k_f \cdot A_m}{f_m}\right) = \frac{12 \times 10^3 \times 5}{10^4} = 6$

$\therefore$ bandwidth $\quad B_T = 2(k_f \cdot A_m + f_m) = 2(\beta_f + 1)f_m \quad$ (Carson's rule)

$$= 2 \times 7 \times 10 \times 10^3 = 140 \text{ kHz}$$

(b) β_p = Modulation index = $k_p.A_m = 1 \times 5 = 5$

$\therefore$ bandwidth $\quad B_T = 2(k_p \cdot A_m + 1)f_m = 2(\beta_p + 1)f_m$

$$= 2 \times 6 \times 10^4 = 120 \text{ kHz}$$

Example 5.8 A phase modulator with k_p= 4 rad/v is fed with a sine wave modulating signal of 3 V peak amplitude and 2 kHz frequency. What is the peak frequency deviation produced in the carrier frequency?

Solution The phase deviation $\phi(t)$ produced by the modulating signal

$$= k_p x(t) = 4 \times 3 \sin 2\pi \times 2 \times 10^3 t$$

$$\therefore \qquad \phi(t) = 12 \sin 4\pi \times 10^3 t$$

If the modulated signal = $x_c(t) = A_c \cos[\omega_c t + \varphi(t)]$, the instantaneous frequency f_i is given by

$$f_i = \frac{1}{2\pi}\frac{d}{dt}[\omega_c t + \phi(t)] = f_c + \frac{1}{2\pi}\frac{d}{dt}[\phi(t)]$$

$$f_i = f_c + \frac{1}{2\pi}\frac{d}{dt}[12 \sin(4\pi \times 10^3)t] = f_c + 24 \times 10^3 \cos(4\pi \times 10^3)t$$

$\therefore$ peak frequency deviation of the carrier is

$$\Delta f = 24 \times 10^3 = 24 \text{ kHz}$$

Example 5.9 A modulating signal $x(t)$ with a trapezoidal waveform as shown is used for (a) frequency modulating a carrier signal of 2 MHz frequency with a frequency deviation constant, k_f of 4 kHz/volt, and (b) phase modulating a carrier with a phase deviation constant k_p of 4 rad/V. In each of these cases, find the maximum instantaneous frequency of the modulated signal.

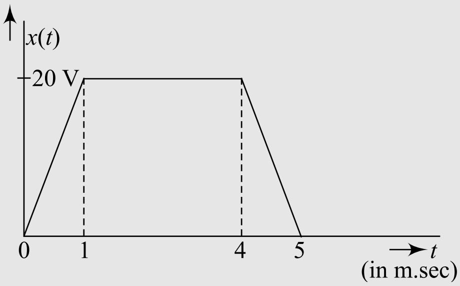


Fig. 5.5 *Signal for Example 5.9*

Solution

(a) Instantaneous frequency $f_i = f_c + k_f \cdot x(t)$

$\therefore$ maximum instantaneous frequency = $(f_i)_{\max}$

$= f_c + k_f \cdot [x(t)]_{\max} = 2 \times 10^6 + 4 \times 10^3 \times 20 = 2.08$ MHz

(b) Instantaneous frequency $f_i = f_c + \dfrac{1}{2\pi}\dfrac{d}{dt}[\varphi(t)]$ where $f(t)$ = phase deviation = $k_p.x(t)$

$$\therefore \qquad (f_i)_{\max} = f_c + \frac{k_p}{2\pi}\left[\frac{d}{dt}x(t)\right]_{\max}$$

Now $$\frac{d}{dt}x(t) = \frac{20}{1 \times 10^{-3}} = 20{,}000 \text{ in the interval 0 to 1 m.sec.}$$

It is zero from 1 m.sec to 4 m.sec and – 20,000 from 4 m.sec to 5 m.sec. Beyond 5 m.sec, it is always zero.

Hence,

$$\left[\frac{d}{dt}x(t)\right]_{\max} = 20{,}000$$

$$\therefore \qquad (f_i)_{\max} = f_c + \frac{k_p}{2\pi}\left[\frac{d}{dt}x(t)\right]_{\max} = 2 \times 10^6 + \frac{4}{2\pi} \times 20 \times 10^3 = 2012.74 \text{ kHz}$$

$$= 2.01274 \text{ MHz}$$

This value is obtained in the interval 0 m.sec to 1 m.sec.

Example 5.10 A particular modulated signal is given by

$$x_c(t) = 2\cos\omega_c t + 0.4\cos 2\pi f_m t \cdot \sin\omega_c t$$

Comment on the nature/type of modulation.

Solution $$x_c(t) = 2\cos\omega_c t + 0.4\cos 2\pi f_m t \cdot \sin\omega_c t$$

$$= \sqrt{2^2 + (0.4\cos 2\pi f_m t)^2}\ \cos[\omega_c t + \theta(t)]$$

where $$\theta(t) = \tan^{-1}\left[\frac{0.4\cos 2\pi f_m t}{2}\right] \cong 0.2\cos 2\pi f_m t$$

Here, we have made use of the approximation that $\theta \approx \theta$ when θ is quite small.

Now, $$\sqrt{2^2 + (0.4\cos 2\pi f_m t)^2} = 2\sqrt{1 + 0.08\cos^2 2\pi f_m t}$$

Since $|\cos 2\pi f_m t| \le 1$, $0.2\cos^2 2\pi f_m t << 1$. Hence, we will use the approximation that

$$\sqrt{1+x} \approx \left(1+\frac{1}{2}x\right) \quad \text{if x << 1}$$

$$\therefore \quad x_c(t) = 2\left[1+\frac{0.08}{2}\cos^2 2\pi f_m t\right]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

$$= 2\left[1+\frac{0.08}{2}\{1+\cos 4\pi f_m t\}\right]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

$$= 2[1.02 + 0.02\cos 4\pi f_m t]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

while $\cos[\omega_c t - 0.2\cos 2\pi f_m t]$ indicates angle modulation, the peak amplitude of this angle-modulated signal, which is $2[1.02 + 0.02\cos 4\pi f_m t]$ indicates amplitude modulation.

Thus, the given $x_c(t)$ is having a combination of amplitude modulation and angle modulation.

5.3 NARROWBAND ANGLE MODULATION

There exists considerable similarity between narrowband angle modulation and AM. We will be examining this aspect in this section.

Referring to Eq. (5.3), we know that an angle modulated signal could be represented by

$$x_c(t) = A_c\cos[\omega_c(t) + \phi(t)]$$

whereas stated in Eq. (5.11),

$$\phi(t) = k_p x(t) \quad \text{for PM}$$

and

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha)d\alpha \quad \text{for FM}$$

Expanding the above equation for $x_c(t)$, we get

$$x_c(t) = A_c[\cos\omega_c(t)\cos\phi(t) - \sin\omega_c(t)\sin\phi(t)] \tag{5.23}$$

Now, if $\phi(t)$ is quite small, say $\phi(t) \le 0.2$ radians, we may make the following approximations:

$$\cos\phi(t) \approx 1 \quad \text{and} \quad \sin\phi(t) \approx \phi(t)$$

Substituting these in the expression for $x_c(t)$, we get

$$x_c(t) = A_c[\cos\omega_c(t) - \phi(t)\sin\omega_c(t)] \tag{5.24}$$

Let us now consider single-tone modulation and let

$$x(t) = A_m\cos\omega_m t$$

5.3.1 Case of Phase Modulation

For this case $\phi(t) = k_p x(t)$

Hence, referring to Eq. (5.24), we have

$$\therefore \quad x_c(t) = A_c[\cos\omega_c(t) - \phi(t)\sin\omega_c(t)] = A_c[\cos\omega_c(t) - k_p A_m\cos\omega_m(t)\cdot\sin\omega_c(t)]$$

But

$$\cos\omega_m(t)\cdot\sin\omega_c(t) = \frac{1}{2}[\sin(\omega_c+\omega_m)t + \sin(\omega_c-\omega_m)t]$$

$$\therefore \quad x_c(t) = A_c\cos\omega_c(t) - \frac{A_c k_p A_m}{2}[\sin(\omega_c+\omega_m)t + \sin(\omega_c-\omega_m)t]$$

$$\therefore \quad \underset{\text{NBPM}}{x_c(t)} = A_c\cos\omega_c(t) + \frac{k_p A_c A_m}{2}[\cos\{(\omega_c+\omega_m)t + \pi/2\} + \cos\{(\omega_c-\omega_m)t + \pi/2\}] \tag{5.25}$$

Equation (5.25) shows that the narrowband phase-modulated signal too has three components – the carrier component represented by the first term, the upper side-frequency component represented by the second term, and the lower side-frequency component represented by the third term, just like an amplitude modulated wave. Further, just like AM, the narrowband angle modulated signal also has a bandwidth of $2f_m$, where f_m is the highest modulating signal frequency. However, there is a difference – the two side-frequency components are shifted in phase by 90° relative to the carrier component, as may be seen from Eq. (5.25) and the phasor diagram shown in Fig. 5.6.

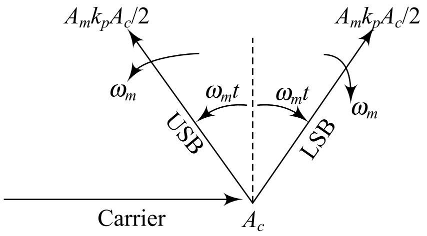


Fig. 5.6 *Phasor diagram of a narrowband phase modulated signal*

5.3.2 Case of Frequency Modulation

For FM,

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha)d\alpha = 2\pi k_f A_m \int_0^t \cos\omega_m \alpha d\alpha$$

$$= \frac{2\pi k_f A_m}{\omega_m}\sin\omega_m t = \left(\frac{k_f A_m}{f_m}\right)\sin\omega_m t$$

∴ substituting this in Eq. (5.24), we get

$$x_c(t) = A_c\left[\cos\omega_c(t) - \left(\frac{k_f A_m}{f_m}\right)\sin\omega_m(t)\sin\omega_c(t)\right]$$

$$= A_c\cos\omega_c(t) + \frac{A_c}{2}\left(\frac{k_f A_m}{f_m}\right)[\cos(\omega_c+\omega_m)t - \cos(\omega_c-\omega_m)t] \tag{5.26}$$

Hence, the components of an NBFM signal may be represented by the following phasor diagram:

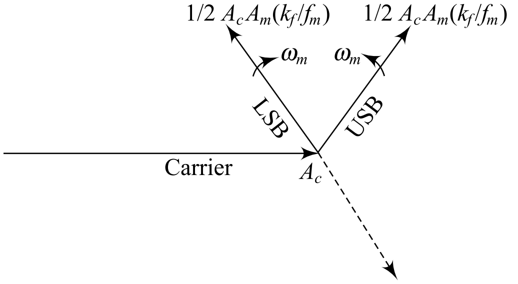


Fig. 5.7 *Phasor diagram of a narrowband frequency modulated signal*

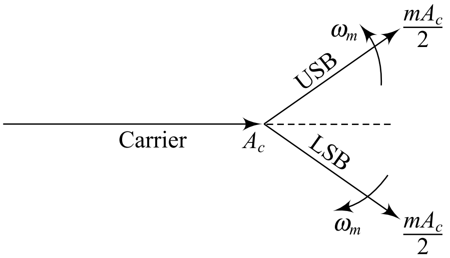


Fig. 5.8 *Phasor diagram of a single-tone modulated AM signal*

It may be instructive to compare the above phasor diagrams with that of a single-tone modulated AM signal shown in Fig. 5.8.

For a single-tone modulated AM signal,

$$x_c(t) = A_c[1 + m\cos\omega_m(t)]\cos\omega_c(t)$$

$$= A_c\cos\omega_c(t) + \frac{mA_c}{2}[\cos(\omega_c+\omega_m)t + \cos(\omega_c-\omega_m)t]$$

5.3.3 Spectrum of a Narrowband FM Signal

Making use of Eq. (5.26), we may draw the two-sided spectrum of a narrowband FM signal as shown in Fig. 5.9.

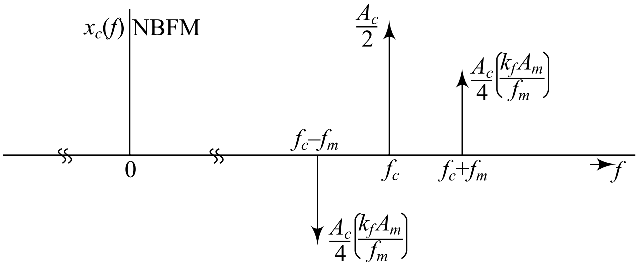


Fig. 5.9 *Two-sided spectrum of an NBFM signal*

Example 5.11 A single-tone signal of 5 kHz frequency modulates a carrier of 90 MHz, and produces a frequency deviation of 50 kHz. Find the peak value of the angle of phase advance/retardation produced by this signal. Also determine the deviation that would be produced by a signal of equal amplitude and of 1000 Hz frequency.

Solution From Eq. (5.20), we have

$$\underset{\text{F.M}}{x_c(t)} = A_c \cos[\omega_c t + \beta_f \sin 2\pi f_m t]$$

∴ phase advance/retardation produced at any instant t is given by

$$\phi(t) = \beta_f \sin 2\pi f_m t$$

Obviously, the maximum value of this is $\beta_f \triangleq \left(\frac{\Delta f}{f_m}\right)$

∴ in the first case, $\beta_{f_1} = \frac{50 \times 10^3}{5 \times 10^3} = 10$ radians

In the second case, Δf remains the same as the amplitude of the new modulating signal is the same as that of the previously used modulating signal.

∴ in the second case, $\beta_{f_2} = \frac{50 \times 10^3}{1 \times 10^3} = 50$ radians.

5.3.4 Generation of Narrowband PM/FM

Equation (5.24) tells us that a narrowband angle modulated signal can be represented as

$$x_c(t) = A_c[\cos \omega_c(t) - \phi \sin \omega_c t]$$

where

$$\phi(t) = \begin{cases} k_p x(t) & \text{for PM} \\ \left[\int_0^t x(\alpha)d\alpha\right] 2\pi k_f & \text{for FM} \end{cases}$$

Hence, as per this equation, an angle-modulated narrowband signal may be generated by means of an arrangement shown in Fig. 5.10.

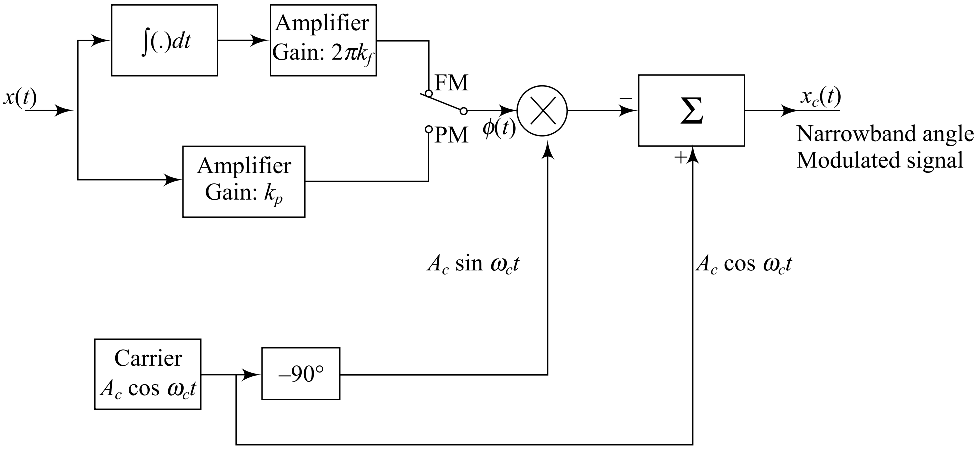


Fig. 5.10 *Generation of narrowband angle modulated signal*

As we shall be seeing later, one important method of generation of wideband FM, which in fact, is of much interest to us, viz., the Armstrong method, or the indirect method of generation of WBFM, is based on generation of narrowband FM as per the arrangement shown above.

5.4 SPECTRUM OF AN ANGLE-MODULATED SIGNAL

Non-linearities inherently present in angle modulation process make the derivation of the spectrum of an angle-modulated signal mathematically intractable except when the modulating signal is a simple one, like a sinusoid. We shall therefore derive the spectrum for an angle-modulated signal when the modulating signal is a sinusoid and then try to extend this result for the case of slightly more complex modulating signals.

5.4.1 Spectrum for Single-Tone Modulation

We had seen that an angle-modulated wave could be represented as (refer to Eq. (5.3))

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)]$$

where

$$\phi(t) = \begin{cases} k_p x(t) & \text{for PM} \\ 2\pi k_f \int_0^t x(\alpha)d\alpha & \text{for FM} \end{cases}$$

1. For FM: As we have assumed single-tone modulating signal,

Let:
$$x(t) = A_m \cos \omega_m t \tag{5.27}$$

$\therefore$
$$\phi(t) = 2\pi k_f \int_0^t A_m \cos \omega_m \alpha d\alpha = \left(\frac{A_m k_f}{f_m}\right) \sin \omega_m t \tag{5.28}$$

Let us now define

$$\beta_f \triangleq \frac{A_m k_f}{f_m} = \frac{\text{Peak frequency deviation}}{\text{modulating signal frequency}} \tag{5.29}$$

$$\beta_f \triangleq m_f = \text{ modulation index for FM}$$

2. For PM: For the case of PM let the single-tone modulating signal be represented by

$$x(t) = A_m \sin \omega_m t \tag{5.30}$$

$$\phi(t) = k_p x(t) = k_p A_m \sin \omega_m t \tag{5.31}$$

We now define

$$\beta_p \triangleq \text{modulation index for PM} \triangleq k_p A_m \triangleq m_p \tag{5.32}$$

Since k_p represents the phase deviation for unit amplitude of the modulating signal and A_m represents the peak amplitude of the modulating signal, β_p obviously denotes the peak phase deviation.
We thus find that

$$\phi(t) = \beta \sin \omega_m t \tag{5.33}$$

wherein for FM

$$\beta = \beta_f = \frac{A_m k_f}{f_m} \quad \text{and} \quad x(t) = A_m \cos \omega_m t \tag{5.34}$$

and for PM

$$\beta = \beta_p = k_p A_m \quad \text{and} \quad x(t) = A_m \sin \omega_m t \tag{5.35}$$

So, henceforth, we shall put

$$\phi(t) = \beta \sin \omega_m t$$

and suitably interpret for PM and FM, so that the analysis becomes common for the two.
We know that

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] = A_c \cos[\omega_c(t) + \beta \sin \omega_m(t)]$$

$$\therefore \quad x_c(t) = A_c \operatorname{Re}[e^{j\omega_c t} \cdot e^{j\beta \sin \omega_m t}] \tag{5.36}$$

In the RHS of the above, $e^{j\beta \sin \omega_m t}$ is a periodic function of time and its period is

$$T = \left(\frac{1}{f_m}\right) \tag{5.37}$$

Since the function is periodic, it can be expanded as a Fourier series and the expansion will be valid for all time.

$$\therefore \text{ let} \quad e^{j\beta \sin \omega_m t} = \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_m t} \tag{5.38}$$

where

$$c_n = \frac{\omega_m}{2\pi} \int_{-\pi/\omega_m}^{+\pi/\omega_m} e^{j\beta \sin \omega_m t} e^{-jn\omega_m t} dt$$

If we put

$$x = \omega_m t \tag{5.39}$$

Then,

$$dt = \left(\frac{1}{\omega_m}\right) dx; \text{ when } t = \pi/\omega_m, x = +\pi$$

and when

$$t = -\pi/\omega_m, x = -\pi$$

$$\therefore \quad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j(nx - \beta \sin x)} dx \tag{5.40}$$

The above integral is a function of n and β and is known as the 'Bessel Function' of the first kind of order n with β as its argument. It is denoted by $J_n(\beta)$. It cannot be evaluated in closed form. However, it has been extensively tabulated for various values of n, the order, and β, the argument.

$$\therefore \qquad c_n = J_n(\beta) \tag{5.41}$$

Substituting this in Eq. (5.38), we have

$$e^{j\beta \sin \omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta) e^{jn\omega_m t} \tag{5.42}$$

Now, substituting this in the RHS of Eq. (5.36), we get

$$x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \sum_{n=-\infty}^{+\infty} J_n(\beta) e^{jn\omega_m t} \right]$$

$$\therefore \qquad x_c(t) = A_c \sum_{n=-\infty}^{+\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t \tag{5.43}$$

Equation (5.43) enables us to expand the angle-modulated signal $x_c(t)$ in terms of its carrier and side-frequency components. The carrier component is given by $A_c J_0(\beta) \cos \omega_c t$ corresponding to $n = 0$. The upper side-frequency components with frequencies $(\omega_c + \omega_m), (\omega_c + 2\omega_m), (\omega_c + 3\omega_m), \ldots$ are obtained by putting $n = 1, 2, 3, \ldots$ and the lower side-frequency components having frequencies of $(\omega_c - \omega_m), (\omega_c - 2\omega_m), (\omega_c - 3\omega_m), \ldots$ are obtained by putting $n = -1, -2, -3, \ldots$. Thus, even for this simple case of single-tone modulating signal, the angle-modulated signal actually has an infinite number of side frequency components and an infinite bandwidth. However, fortunately it is possible to define what is called an effective bandwidth which is finite, because for any β, $J_n(\beta)$ tends to zero as n tends to infinity, making the amplitudes of the higher side-frequency components negligibly small (see Fig. 5.11).

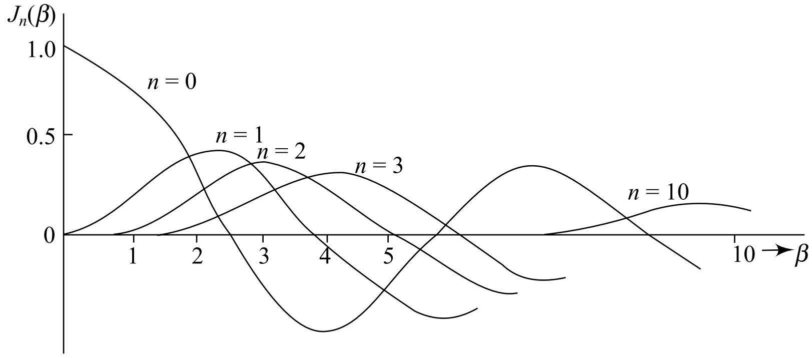


Fig. 5.11 *$J_n(\beta)$ for various values of n*

An infinite series expansion of $J_n(\beta)$ is given by

$$J_n(\beta) = \sum_{k=0}^{\infty} \frac{\left(\frac{\beta}{2}\right)^{n+2k} (-1)^k}{k!(n+k)!} \tag{5.44}$$

However, for small values of β, $J_n(\beta)$ may be approximated by

$$J_n(\beta) = \frac{\left(\frac{\beta}{2}\right)^n}{n!} \tag{5.45}$$

Some useful properties of $J_n(\beta)$ are given in Table 5.1.

Table 5.1 *Useful properties of $J_n(\beta)$*

S. No.	Property	S. No.	Property
1.	$J_0(0) = 1$	5.	$J_n(\beta) = J_{-n}(\beta)$ if n is even
2.	$J_n(0) = 0$, if n is a non-zero integer	6.	$J_n(\beta) = -J_{-n}(\beta)$ if n is odd
3.	$J_n(\beta) = J_n(-\beta)$ if n is even	7.	$J_n(\beta) \to 0$ as $n >> \beta$
4.	$J_n(\beta) = -J_n(-\beta)$ if n is odd		

Expanding the RHS of Eq. (5.43) term by term, and noting the fact that

$$J_{-n}(\beta) = +J_n(\beta) \text{ for } n \text{ even}$$

and

$$J_{-n}(\beta) = -J_n(\beta) \text{ for } n \text{ odd}$$

we get

$$\begin{aligned} x_c(t) = A_c J_0(\beta)\cos\omega_c t + A_c[&J_1(\beta)\{\cos(\omega_c+\omega_m)t - \cos(\omega_c-\omega_m)t\} \\ &+J_2(\beta)\{\cos(\omega_c+2\omega_m)t + \cos(\omega_c-2\omega_m)t\} \\ &+J_3(\beta)\{\cos(\omega_c+3\omega_m)t - \cos(\omega_c-3\omega_m)t\} \\ &+ \;\ldots\ldots.. \\ &\text{(an infinite number of such terms)} \end{aligned} \tag{5.46}$$

Note that for an angle modulated signal, $A_c\,J_0(\beta)$ is the amplitude of the carrier, $A_c\,J_1(\beta)$ is the amplitude of the first side-frequency, $A_c\,J_2(\beta)$ is the amplitude of the second side-frequency, and so on. Figure 5.12 shows the amplitude spectra of an FM signal for single-tone modulation for different modulation indices. It may be noted that unlike in AM, the amplitude of the carrier component in the modulated signal varies with the modulation index. This is because the value of $J_0(\beta)$ goes on changing with the value of β (see Fig. 5.11) and may be positive, zero, or even negative. In fact, for values of β like $\beta \cong 2.3$ for which $J_0(\beta)$ has zero crossings, the carrier component completely vanishes in the modulated signal.

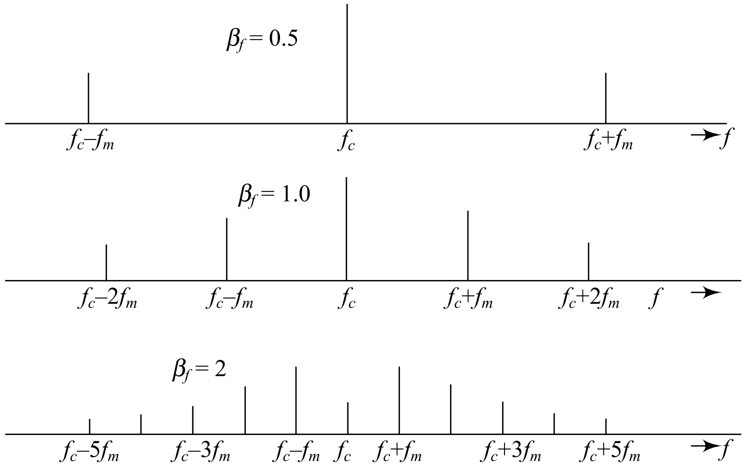


Fig. 5.12 *Amplitude spectra of an FM signal with single-tone modulation for different modulation indices*

Note *For these sketches, f_m is decreased while keeping $A_m k_f$ constant to get larger values of β.*

5.4.2 Spectrum of an Angle-Modulated Signal for a Periodic Message Signal

In the foregoing discussion, we have studied the spectrum of an angle-modulated signal when the modulating signal was a single-tone signal and found that the spectrum contains an infinite number of side-frequencies. We shall now determine the spectrum of an angle-modulated signal when the modulating signal is a periodic signal. We know that an angle-modulated signal can be represented as

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] \tag{5.47}$$

Let us assume that the modulation is phase modulation and that the modulating signal $x(t)$ is a periodic wave with a period $T_0 = 1/f_0$.

$$\therefore \quad x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] = A_c \cos[\omega_c(t) + \beta_p x(t)]$$

$$= A_c \,\text{Re}[e^{j\omega_c t} \cdot e^{j\beta_p x(t)}] \tag{5.48}$$

Since $x(t)$ is periodic with a period T_0, $e^{j\beta_p x(t)}$ is also periodic with the same period. Hence, we may expand this function as a complex-exponential Fourier series.

Let
$$e^{j\beta_p x(t)} = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; \quad -\infty < t < \infty$$

where

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j\beta_p x(t)} \cdot e^{-j2\pi n f_0 t} dt \tag{5.49}$$

$$x_c(t) = A_c \,\text{Re}\left[e^{j\omega_c t} \cdot \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t} \right]$$

$$= A_c \sum_{n=-\infty}^{\infty} |c_n| \cos\{2\pi(f_c + nf_0)t + \angle c_n\} \tag{5.50}$$

Example 5.12 Find the spectrum of a phase modulated signal when the modulating signal is a periodic square wave as shown in Fig. 5.13.

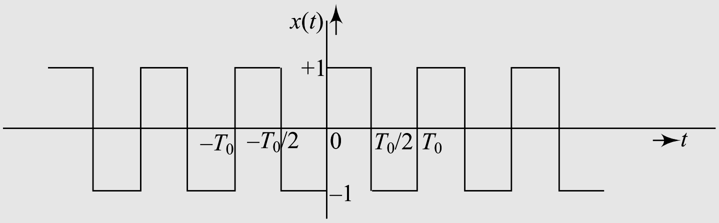


Fig. 5.13 *A square wave modulating signal*

Solution Let β_p be the modulation index.

Then
$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] = A_c \cos[\omega_c(t) + \beta_p x(t)]$$

$$\therefore \quad x_c(t) = A_c \,\text{Re}[e^{j\omega_c t} \cdot e^{j\beta_p x(t)}]$$

Since $x(t)$ is a periodic square wave with a frequency of $f_0 = 1/T_0$, we may expand $e^{j\beta_p x(t)}$, which is also periodic with the same period, using complex-exponential Fourier series.

Let $$e^{j\beta_p x(t)} = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; \quad -\infty < t < \infty$$

$$\therefore \quad c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j\beta_p x(t)} \cdot e^{-j2\pi n f_0 t} dt$$

Evaluating the integral and simplifying the result, we get

$$c_n = \begin{cases} 0 & \text{for } n \text{ even} \\ 2n\pi \sin\beta_p & \text{for } n \text{ odd} \end{cases}$$

$$\therefore \quad e^{j\beta_p x(t)} = \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{+\infty} 2n\pi \sin\beta_p e^{-j2\pi n f_0 t}; \quad -\infty < t < \infty$$

$$\therefore \quad x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{+\infty} 2n\pi \sin\beta_p e^{-j2\pi n f_0 t} \right]$$

$$= A_c \operatorname{Re}\left[2\pi \sin\beta_p \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{+\infty} n e^{j(\omega_c t - 2\pi n f_0 t)} \right]$$

$$= 2\pi(\sin\beta_p) A_c \left[\sum_{\substack{n=-\infty \\ n \text{ odd}}}^{+\infty} n \cos(\omega_c - 2\pi n f_0) t \right]$$

5.5 POWER OF AN ANGLE-MODULATED SIGNAL AND EFFECTIVE BANDWIDTH

As we had already seen, an angle-modulated signal may be represented as

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)]$$

Hence, the average power in the angle-modulated signal is

$$\left\langle x_c^2(t) \right\rangle = \left\langle A_c^2 \cos^2[\omega_c t + \phi(t)] \right\rangle$$

$$= \frac{1}{2} A_c^2 + \frac{1}{2} A_c^2 \left\langle \cos[2\omega_c t + 2\phi(t)] \right\rangle$$

$$\therefore \quad \left\langle x_c^2(t) \right\rangle = \frac{1}{2} A_c^2 \tag{5.51}$$

Remark *Equation (5.51) shows that an angle-modulated wave has a constant average power, since the RHS of the equation is a constant, independent of time.*

From the analysis in the previous section leading to the spectrum of an angle-modulated signal, it appears as though the bandwidth occupied by an angle-modulated signal is infinitely large. Strictly, form a theoretical point of view, this is correct. But, as has been pointed out earlier, since the amplitude of the nth side-frequency component is $A_c J_n(\beta)$ and as

$$J_n(\beta) \to 0 \quad \text{as} \quad n \to \infty \qquad \text{(Refer to Table 5.1)}$$

most of the power of the angle-modulated signal resides in the carrier component and some finite number of side-frequency components. This enables us to define what is called the '*effective bandwidth*' of an angle-modulated signal by considering only those sidebands which have a significant portion of the total power of the modulated signal.

Following the above argument, we define the '*effective bandwidth*' of an angle-modulated signal as the bandwidth occupied by those *minimum number* of first k side-frequency components, which along with the carrier component, have at least 98% of the total power of the modulated signal.

$$\left.\begin{array}{l}\text{Now, power in the first } k\\ \text{side frequency components}\\ \text{and the carrier}\end{array}\right\} = \frac{1}{2}A_c^2 \sum_{n=-k}^{k} J_n^2(\beta)$$

$$\therefore \qquad \frac{\dfrac{1}{2}A_c^2 \displaystyle\sum_{n=-k}^{k} J_n^2(\beta)}{\dfrac{1}{2}A_c^2} = 0.98 = J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta)$$

$\therefore$ k must be so chosen that it is the smallest integer satisfying

$$J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta) \geq 0.98 \tag{5.52}$$

Table 5.2 *A short table of Bessel functions (values of $J_n(\beta)$ for various values of n and β)*

n	β=0.1	β = 0.2	β = 0.5	β = 1.0	β = 2.0	β = 5.0	β = 8.0	β = 10.0
0	0.997	0.990	0.938	0.765	0.224	–0.178	0.172	–0.246
1	0.050	0.100	0.242	0.440	0.577	–0.328	0.235	0.043
2	0.001	0.005	0.031	0.115	0.353	0.047	–0.113	0.255
3				0.020	0.129	0.365	–0.291	0.058
4				0.002	0.034	0.391	–0.105	–0.220
5					0.007	0.261	0.186	–0.234
6					0.001	0.131	0.338	–0.014
7						0.053	0.321	0.217
8						0.018	0.223	0.318
9						0.006	0.126	0.292
10						0.001	0.061	0.207
11							0.026	0.123
12							0.010	0.063
13							0.003	0.029
14							0.001	0.012
15								0.004
16								0.001

By referring to the Bessel function tables (see Table 5.2) we find that for any given β, the value of k satisfying Eq. (5.52) is approximately equal to the integer part of (1+β). For example, for $\beta = 1$, $n = 2 = (\beta + 1)$; for $\beta = 2$, $n = 3 = (\beta + 1)$, and so on. Since $\lfloor(\beta+1)\rfloor$ side-frequency components are to be considered, the transmission bandwidth B_T for angle-modulated signals with modulation index β, is given by (for single-tone modulation)

$$B_T = \text{Transmission bandwidth} = 2\lfloor(\beta+1)\rfloor f_m \tag{5.53}$$

where $\lfloor x \rfloor$ is used to denote the nearest integer value of x and f_m is the frequency of the single-tone modulating signal. The above formula is generally referred to as '*Carson's Rule*'.

It may be noted that in Eq. (5.53), β has to be taken as β_p for phase modulation and β_f for frequency modulation.

Since $\beta_p = k_p \cdot A_m$

and $\beta_p = (k_f \cdot A_m)/f_m$

where A_m denotes the peak amplitude of the single-tone modulating signal, we may rewrite Carson's Rule as

$$B_T = \begin{cases} 2(k_p \cdot A_m + 1)f_m & \text{for PM} \\ 2(k_f \cdot A_m + f_m) & \text{for FM} \end{cases} \tag{5.54}$$

where k_p is the phase deviation constant and k_f is the frequency deviation constant.

When a non-sinusoidal modulating signal is used, as generally is the case, Carson's rule is extended to this case by modifying it as follows.

$$B_T = 2(\lfloor\beta\rfloor + 1)W \tag{5.55}$$

where W is the bandwidth of modulating signal, $x(t)$, and $\lfloor\beta\rfloor$ is the nearest integer value of β which is the modulation index defined as

$$\beta = \begin{cases} k_p \max\left[|x(t)|\right] & \text{for PM} \\ \dfrac{k_f \max\left[|x(t)|\right]}{W} & \text{for FM} \end{cases} \tag{5.56}$$

5.5.1 Relationship between PSD of an FM Wave and the PDF of Its Modulating Signal

There exists an interesting and useful relationship between the power spectral density of an FM wave and the amplitude probability density function of its modulating signal and we shall now derive this in a heuristic way.

Let $x(t)$ be the modulating signal, frequency modulating a carrier signal of peak amplitude A_c and frequency f_c. Let $x(t)$ have an amplitude probability density function (PDF) given by $f_X(x)$.

From Eqs. (5.6) and (5.7), we have the instantaneous frequency f_i of the modulated signal at the instant t, given by

$$f_i = f_c + \frac{1}{2\pi}k_f x(t) = f_c + (\Delta f)$$

where f_c is the carrier frequency, k_f is the frequency deviation constant and Δf is the frequency deviation produced by $x(t)$ at the instant t.

Since $f_X(x)$ is the amplitude probability density function of $x(t)$, it follows that

$$P[x \le x(t) < (x+dx)] = f_X(x)dx$$

$$= \text{Probability of } x(t) \text{ lying between } x \text{ and } (x+dx).$$

But, we know (from the expression for f_i) that when $x(t)$ lies between x and $(x + dx)$, f_i lies between $f_c + \frac{1}{2\pi}k_f x$ and $f_c + \frac{1}{2\pi}k_f(x+dx)$.

Let $P_{x_c}(f)$ be the power spectral density of the frequency modulated wave. Then the area under this PSD curve is equal to the total average power of the modulated signal, and is equal to $A_c^2/2$. So, the fraction of the power of the modulated signal within the frequency interval $f = \mu$ to $f = \mu + d\mu$, is given by

$$\frac{P_{x_c}(\mu)d\mu}{(A_c^2/2)}$$

From the equation for the instantaneous frequency we know that when f the frequency of the FM wave lies between μ and $(\mu + d\mu)$, correspondingly, the value of $x(t)$ lies between some x_1 and $x_1 + dx$. The fractional time for which $x(t)$ lies between x_1 and $(x_1 + dx)$ is given by $f_x(x_1)dx$; where $f_x(x_1)$ is the value of $f_X(x)$ at $x = x_1$. Now, making the reasonable assumption that the fractional power of the modulated signal between frequencies μ and $(\mu + d\mu)$ is directly proportional to the fractional time for which $x(t)$ lies between x_1 and $(x_1 + dx)$, we have

$$\frac{P_{x_c}(\mu)d\mu}{(A_c^2/2)} = K_1 f_X(x_1)dx$$

where K_1 is a constant of proportionality to dx and A_c is constant, we may write

$P_{x_c}(f) = Kf_X(x)$, where K is a constant.

Thus, we have the important result that $P_{x_c}(f)$, *the PSD of an FM signal is directly proportional to* $f_X(x)$, *the amplitude probability density function of its modulating signal* $x(t)$.

5.5.2 Effective Bandwidth of a Gaussian Modulated FM Signal

Though not exactly Gaussian, the amplitude density function of many of the signals that we come across in practise can be approximated to Gaussian density. Determining the effective bandwidth of a Gaussian modulated signal therefore assumes importance. We shall now proceed with this, making use of the above result.

If $x(t)$, the modulating signal, has a Gaussian probability density function, it follows from the earlier result that the power spectral density $P_{x_c}(f)$ also is going to be Gaussian. Since the total area under a PSD curve is to be equal to the average power, and since in our case, it is $A_c^2/2$, the two-sided PSD of $x_c(t)$ may be written as

$$P_{x_c}(f) = \frac{A_c^2}{4\sqrt{2\pi}(\Delta f)_{\text{rms}}}[e^{-(f-f_c)^2/2(\Delta f)_{\text{rms}}^2} + e^{-(f+f_c)^2/2(\Delta f)_{\text{rms}}^2}]$$

A sketch of this is shown in Fig. 5.14.

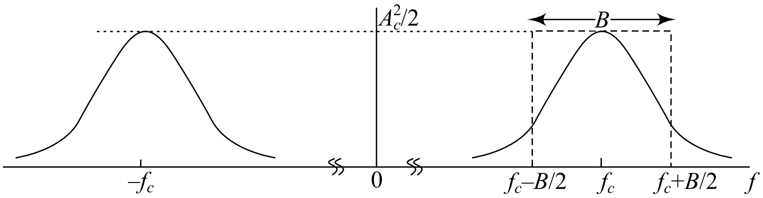


Fig. 5.14 *PSD of a Gaussian modulated FM signal*

Defining the effective bandwidth, B as usual, as that bandwidth, within which 98% of the average power of the modulated signal is available, we may write

$$0.98\frac{A_c^2}{2}=\frac{A_c^2}{4\sqrt{2\pi}(\Delta f)_{\text{rms}}}\cdot\int_{f_c-B/2}^{f_c+B/2} e^{-(f-f_c)^2/2(\Delta f)_{\text{rms}}^2}df$$

Let $\mu \triangleq (f-f_c)$ $\therefore df=d\mu$. When $f=f_c-\frac{B}{2}$, $\mu=-B/2$ and when $f=f_c+\frac{B}{2}$, $\mu=B/2$
∴ RHS becomes

$$\frac{A_c^2}{2\sqrt{2\pi}(\Delta f)_{\text{rms}}}\cdot\int_{-B/2}^{B/2} e^{-\mu^2/2(\Delta f)_{\text{rms}}^2}d\mu=\frac{A_c^2}{\sqrt{2\pi}(\Delta f)_{\text{rms}}}\cdot\int_{0}^{B/2} e^{-\mu^2/2(\Delta f)_{\text{rms}}^2}d\mu$$

If we now put $y=\frac{\mu}{\sqrt{2}(\Delta f)_{\text{rms}}}$, $dy=\frac{d\mu}{\sqrt{2}(\Delta f)_{\text{rms}}}$

∴

$$0.98\frac{A_c^2}{2}=\frac{A_c^2}{2}\frac{2\sqrt{2\pi}(\Delta f)_{\text{rms}}}{\sqrt{2\pi}(\Delta f)_{\text{rms}}}\cdot\int_0^{\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}} e^{-y^2}dy$$

$$=\frac{2}{\sqrt{\pi}}\int_0^{\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}} e^{-y^2}dy=erf\left[\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}\right]$$

From error function tables,

$$\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}=1.645 \qquad \therefore B=4.6(\Delta f)_{\text{rms}}$$

5.5.3 Comparison between FM and PM

1. Equation (5.54) clearly brings out the difference between phase modulation and frequency modulation. On the RHS of this equation, f_m just adds to k_fA_m which is the peak frequency deviation, in the case of FM. But in the case of PM, f_m multiplies $(1+k_pA_m)$. Thus, increase in f_m, the modulating signal frequency, will have very little effect on the transmission bandwidth in the case of FM, while it will have a very significant effect (on the transmission bandwidth) in the case of phase modulation.
2. Increasing the amplitude of the modulating signal, on the other hand, will have same effect on the transmission bandwidth in the case of both PM and FM.

As pointed out in point 1, the bandwidth of an FM signal is practically unaffected by an increase in the modulating signal frequency. This property, coupled with the fact that FM signals are relatively unaffected by the additive noise on the channel, makes frequency modulation eminently suited for broadcasting of high quality music which necessitates handling of audio frequencies up to even 15 kHz. That is why commercial FM broadcasting uses audio frequencies up to 15 kHz. (AM broadcasting on the other hand, handles audio frequencies up to only 5 kHz). In order to get a good signal-to-noise ratio at the destination, these FM broadcasting stations use modulation indices (i.e., β values) of the order of at least 5 (see Chapter 9). As per Carson's rule, therefore, a transmission bandwidth of at least 180 kHz is needed for FM broadcasting. In practise, a bandwidth of 200 kHz is provided.

Example 5.13 Equation (5.54), which gives the transmission bandwidth, B_T, of an FM signal as $B_T=2(\beta_f+1)f_m$ for single-tone modulation. This equation, known as Carson's rule, was derived on the basis that the 'effective bandwidth', B_T, has at least 98% of the total average power of the FM signal. Instead of Carson's rule, sometimes we use the equation $B_T'=(2\beta_f+1)f_m$. Determine the percentage of average power of an FM signal contained in it, assuming $\beta_f=1$.

Solution

(a) As per Carson's rule, $B_T = 2(1+1)f_m = 4f_m$

The average power in a bandwidth up to k side frequencies expressed as a fraction of the total average power of the FM signal is given by Eq. (5.52) as

$$J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta)$$

For $\beta = \beta_f = 1$, $k = 4$ since $B_T = 4f_m$

From Table 5.2 of $J_n(\beta)$ for various values of n and β, if we compute, we get

$$J_0^2(\beta) + 2\sum_{n=1}^{4} J_n^2(\beta) = 0.999683 \quad \text{for } k = 4, \beta = 1.$$

$\therefore$ % Power in $B_T = 99.9683$

(b) If we use the approximate formula $B_T' = (2\beta_f + 1)f_m$,

with $\beta_f = 1$, we get $B_T' = 3f_m$ $\therefore$ $k = 3$ in this case.

Then
$$J_0^2(\beta) + 2\sum_{n=1}^{3} J_n^2(\beta) = 0.999675$$

$\therefore$ % Power in $B_T = 99.9675$

Example 5.14 In an FM system, with a modulating signal frequency of 600 Hz and a peak modulating voltage of 3.6 V, the modulation index is 60. Find the frequency deviation constant and the peak frequency deviation. If the modulating signal frequency is reduced to 400 Hz while the modulating voltage is simultaneously increased to 4 V, what is the value of the modulation index?

Solution $\beta = \dfrac{k_f \cdot A_m}{f_m} = 60 = \dfrac{k_f \times 3.6}{600} \qquad \therefore k_f = \dfrac{60 \times 600}{3.6} = 10^4 \text{ Hz/V}$

$\therefore$ peak frequency deviation = 3.6×10^4 = 36 kHz

In the second case, $\beta = \dfrac{k_f \cdot 4}{400} = \dfrac{10^4 \times 4}{400} = 100$

Example 5.15 Compute the bandwidth requirement for the transmission of an FM signal having a frequency deviation of 75 kHz and an audio bandwidth of 10 kHz. **(JNTU Sept., 2007)**

Solution Frequency deviation Δf = 75 kHz

Audio bandwidth = 10 kHz $\therefore$ β_f = modulation index = 75/10 = 7.5

$\therefore$ maximum audio frequency = f_m = 10 kHz = W

$\therefore$ using Carson's Rule, the required bandwidth is given by

$$\text{Bandwidth} = 2(\beta_f + 1)W = 2(7.5 + 1)10 \times 10^3 = 170 \text{ kHz.}$$

Example 5.16 An FM radio link has a frequency deviation of 30 kHz. The modulating frequency is 3 kHz. Calculate the bandwidth needed for the link. What will be the bandwidth if the deviation is reduced to 15 kHz? **(JNTU Sept., 2007)**

Solution

In the first instance, $\Delta f_1 = 30$ kHz and $f_m = 3$ kHz

$$\therefore \qquad \beta_{f_1} = \text{Modulation index } = \frac{\Delta f_1}{f_m} = \frac{30}{3} = 10$$

$\therefore$ by Carson's rule, the required bandwidth is

$$BW_1 = 2(\beta_{f_1} + 1)f_m = 2(10+1) \times 3 \times 10^3 = 66 \text{ kHz}$$

If now the deviation is reduced to 15 kHz,

$$\therefore \qquad \beta_{f_2} = \frac{\Delta f_2}{f_m} = \frac{15}{3} = 5$$

$\therefore$ by Carson's rule, the bandwidth required now is

$$BW_2 = 2(\beta_{f_2} + 1)f_m = 2(5+1) \times 3 \times 10^3 = 36 \text{ kHz}$$

Example 5.17 A signal $x(t)$, whose Fourier transform $X(f)$ is shown in Fig. 5.15 is normalized so that $|x(t)| \le 1$. This signal is to be transmitted using FM with a frequency deviation constant k_f = 60 kHz per volt. What will be the bandwidth required for transmission?

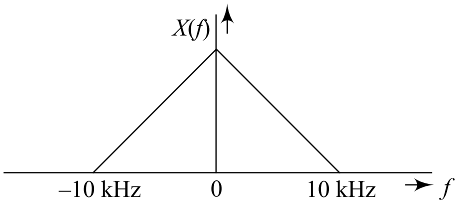


Fig. 5.15 *Spectrum of the signal of Example 5.17*

Solution Here, the bandwidth W of the modulating signal is

$$W = 10^4 \text{ Hz}$$

k_f is given to be 60 kHz/volt and A_m = 1 V since $|x(t)| \le 1$

$$\therefore \qquad \beta_f = \frac{k_f \cdot A_m}{W} = \frac{60 \times 10^3 \times 1}{10 \times 10^3} = 6 \,.$$

$$\therefore \qquad B_T = 2(\beta_f + 1)W = 2 \times 7 \times 10^4 = 140 \text{ kHz}$$

Example 5.18 An angle-modulated signal is of the form

$$x_c(t) = 50\cos[2\pi \times 10^7 \times t + 5 \sin 2\pi \times 1.5 \times 10^3 t]$$

(a) If $x_c(t)$ is a frequency-modulated signal, find the modulation index and the transmission bandwidth required.

(b) If $x_c(t)$ is a phase-modulated signal, find the modulation index and the transmission bandwidth required.

(c) In part (a), if the frequency of the modulating signal is doubled, what will be the modulation index and the transmission bandwidth?

(d) In part (b), if the frequency of the modulating signal is doubled, what will be the modulation index and the transmission bandwidth?

Solution

(a) $\beta_f = 5$ and $B_T = 2(5+1)1500 = 18$ kHz since $5 = \left(\dfrac{k_f A_m}{1500}\right) = \beta_f$

(b) $\beta_p = 5$ and $B_T = 2(5+1)1500 = 18$ kHz

(c) $\beta_f = 2.5$ and $B_T = 2(2.5+1)3000 = 21$ kHz $\left(\text{since } \dfrac{k_f A_m}{3000} = 2.5\right)$

(d) $\beta_p = k_p \cdot A_m = 5$ and is not affected by the doubling of f_m.

$B_T = 2(\beta_p + 1) f_m = 2 \times 6 \times 3000 = 36000 \text{ Hz} = 36 \text{ kHz}$

Example 5.19 An angle-modulated signal has the form

$$v(t) = 100\cos[2\pi f_c t + 4\sin 2000\pi t] \quad \text{where } f_c = 10 \text{ MHz}$$

(a) Determine the average transmitted power. (b) Determine the peak phase deviation. (c) Determine the peak frequency deviation. (d) Is this FM or a PM signal? Explain. **(JNTU, May, 2007)**

Solution

(a) Average transmitted power $= \frac{(100)^2}{2} = 500W$

(b) Peak phase deviation:

Since 4 sin 2000 πt represents the phase deviation at any instant '*t*' and since sin 2000πt has a peak value 1, the peak phase deviation is equal to 4 radians.

(c) Peak frequency deviation: The instantaneous frequency is given by

$$f_i(t) = \frac{1}{2\pi}\left\{\frac{d}{dt}[2\pi f_c t + 4\sin 2000\pi t]\right\}$$

$$= f_c + \frac{4}{2\pi} \cdot 2000\pi \cdot \cos 2000\pi t = f_c + 4000\pi \cdot \cos 2000\pi$$

$\therefore$ the frequency deviation at the instant '*t*' is 4000 cos 2000πt and the peak frequency deviation is 4000 Hz.

(d) It can be considered to be PM signal with $\beta_p = 4$ and a modulating signal of sin 2000πt **OR** it can be considered to be an FM signal with $\beta_f = 4$ and a modulating signal of cos 2000πt.

Example 5.20 An FM wave with modulation index $\beta = 1$ is transmitted through an ideal band pass filter with mid-band frequency f_c and bandwidth $5f_m$, where f_c is the carrier frequency and f_m is the frequency of the sinusoidal modulating wave. Determine the amplitude spectrum of the filter output.

Solution From Eq. (5.46), the spectrum of an FM wave $x_c(t)$, with β as the modulation index, is given by

$$\begin{aligned} x_c(t) = A_c J_0(\beta)\cos\omega_c t &+ A_c[J_1(\beta)\{\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t\}] \\ &+ A_c[J_2(\beta)\{\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)t\}] \\ &+ A_c[J_3(\beta)\{\cos(\omega_c + 3\omega_m)t - \cos(\omega_c - 3\omega_m)t\}] \\ &+ \ldots\ldots \end{aligned}$$

Even though theoretically the side-frequency components are infinite in number on either side of f_c, only the carrier component and the first two side-frequency components on the two sides of f_c fall within the pass band of the BPF.

From Table 5.2 which gives the values of $J_n(\beta)$ for some values of β and n = 0, 1, 2, 3, etc., we find that $J_0(1) = 0.765$, $J_1(1) = 0.44$, $J_2(1) = 0.115$. Thus, the signal at the output of the filter is given by

$$\begin{aligned} y(t) = A_c[0765\cos\omega_c t &+ 0.44\{\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t\} \\ &+ 0.115\{\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)t\}] \end{aligned}$$

The spectrum of $y(t)$ is as shown in Fig. 5.16.

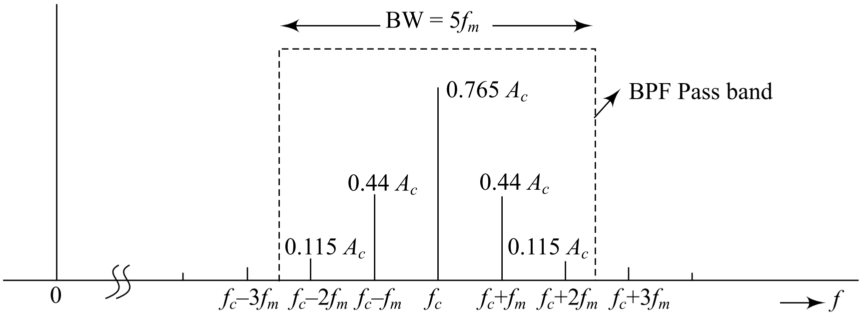


Fig. 5.16 *Amplitiude spectrum for Example 5.20*

Example 5.21 Express the carrier power as a fraction or percentage of the total power in an FM signal being transmitted with $\beta_f = 2$

Solution With $\beta_f = 2, J_0^2(\beta_f) = (0.224)^2 = 0.050176$

Total average power in the FM signal $= \frac{1}{2}A_c^2$, if A_c is the peak amplitude of the unmodulated carrier (see Eq. (5.51))

The average power in the carrier component $= \frac{1}{2}A_c^2 J_0^2(\beta_f)$ since the peak amplitude of the carrier component in an WBFM signal is given by

$$A_c J_0(\beta_f) \qquad \text{(See Eq. (5.46))}$$

$$\therefore \qquad \frac{\text{Carrier Power}}{\text{Total Power}} = \frac{\frac{1}{2}A_c^2 J_0^2(\beta_f)}{\frac{1}{2}A_c^2} = J_0^2(\beta_f) = 0.050176$$

As a percentage, it is just 5.0176%

Example 5.22 A carrier signal $A_c \cos\omega_c t$ is angle modulated by the sum of two single tones (sinusoids) of frequencies f_1 and f_2 with modulation indices β_1 and β_2, respectively. The modulated signal is

$$x_c(t) = A_c \cos[\omega_c t + \beta_1 \sin\omega_1 t + \beta_2 \sin\omega_2 t]$$

Derive an expression for its spectrum.

Solution The given angle-modulated signal may be written as

$$x_c(t) = A_c \,\text{Re}[e^{j\omega_c t} \cdot e^{j\beta_1 \sin\omega_1 t} \cdot e^{j\beta_2 \sin\omega_2 t}]$$

From Eq. (5.42), we have

$$e^{j\beta_1 \sin\omega_1 t} = \sum_{n=-\infty}^{\infty} J_n(\beta_1) e^{jn\omega_1 t} \quad \text{and} \quad e^{j\beta_2 \sin\omega_2 t} = \sum_{n=-\infty}^{\infty} J_m(\beta_2) e^{jm\omega_2 t}$$

Substituting these in the above equation for $x_c(t)$, we have

$$x_c(t) = A_c \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} J_n(\beta_1) J_m(\beta_2) \cos[\omega_c t + n\omega_1 t + m\omega_2 t]$$

Thus the spectrum will have the carrier component, side frequencies of the type $\cos[\omega_c t \pm n\omega_1 t]$ and $\cos[\omega_c t \pm m\omega_2 t]$ and also of the type $\cos[(\omega_c t \pm n\omega_1 \pm m\omega_2)t]$.

5.6 GENERATION OF WIDEBAND ANGLE-MODULATED SIGNALS

5.6.1 Indirect or Armstrong Method

An important method for generation of a wideband angle modulated signal is to first generate a narrowband angle-modulated signal using the narrowband angle modulator shown in Fig. 5.10 and then convert the narrowband signal into a wideband signal. This method is known as the '*indirect method*' of generation of wideband FM and PM signals. It is also known as '*Armstrong method*'.

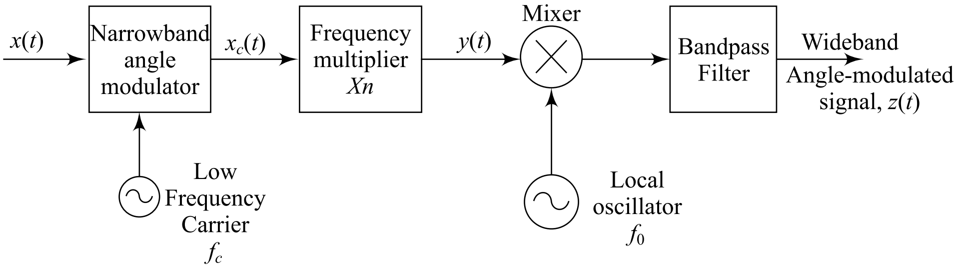


Fig. 5.17 *Armstrong or indirect method of generation of wideband angle-modulated signals*

Figure 5.17 shows the block schematic diagram of the indirect method of generation of wideband angle-modulated signals. As shown in the figure, the first stage is a narrowband angle modulator of the type shown in Fig. 5.10. The modulating signal $x(t)$ and a low frequency carrier signal produced by a crystal oscillator are given as input signals and it uses these two signals to produce a narrowband angle-modulated signal with a carrier frequency of f_c. A low frequency carrier is used for producing the narrowband signal. The next stage is a frequency multiplier used for converting the narrowband signal into a wideband signal and it raises the carrier frequency from f_c to nf_c. The frequency multiplier stage consists of a non-linear device whose output is tuned to the desired harmonic of f_c. Generally a class-c amplifier whose output circuit is a tank circuit tuned to nf_c serves as an '*Xn*' frequency multiplier. The collector current pulses of class-c amplifier have a conduction angle of about 100° to 200° and are quite rich in harmonics. Quite often this frequency multiplier stage consists of the cascade connection of several doublers and/or triplers.

Although the output signal of the frequency multiplier stage is certainly a wideband angle-modulated signal, the carrier frequency, nf_c, of this wideband signal will not in general be the correct desired carrier frequency at which the wideband signal is to be transmitted. Hence, we use a mixer to which we connect the output of a local oscillator having an appropriately chosen frequency f_0 and if necessary, a chain of frequency multipliers, in order to finally get a carrier frequency which is the desired carrier frequency. As the mixer produces the sum frequency and the difference frequency, a bandpass filter which has a center frequency equal to either the sum frequency, or the difference frequency (whichever is needed) and whose pass band is adequate to accommodate the effective bandwidth of the wideband signal, is used.

If the narrowband angle-modulated signal is represented as

$$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)]\ ;\ \phi(t) \text{ is small}$$

then the output of the *Xn* frequency multiplier will be

$$y(t) = A_c \cos[n\omega_c(t) + n\phi(t)]$$

Note *The frequency multiplier multiplies the instantaneous frequency $\omega_i(t)$ which is given by*

$$\omega_i(t) = \omega_c + \frac{d}{dt}\phi(t)$$

If the BPF selects the difference frequency generated by the mixer,

$$z(t) = A_c \cos[(n\omega_c - \omega_0)t + n\phi(t)] \tag{5.57}$$

Since we can choose n and f_0, by an appropriate choice of these two, we can ensure that the wideband angle-modulated signal $z(t)$ has the desired carrier frequency.

Advantages and disadvantages of indirect method

1. As crystal oscillators are used for obtaining the carrier frequency, it (the carrier frequency) is very stable.
2. Since the narrowband FM is generated by a phase modulator, a long chain of frequency multipliers will have to be used to bring the frequency deviation to the required level.

Example 5.23 In a wideband FM generator using the indirect method, the narrowband FM signal initially generated has a carrier frequency of 200 kHz and a frequency deviation of 49 Hz. Choose appropriate values for the local oscillator frequency for the mixer and the frequency multiplication required before and after the mixer if the final WBFM signal is to have a carrier frequency of 91.2 MHz and the standard frequency deviation of 75 kHz.

Solution Final carrier frequency = $f_{c4} = 91.2 \times 10^6$ Hz

Initial carrier frequency = $f_{c_1} = 200 \times 10^3$ Hz

$$\therefore \qquad \left.\begin{matrix}\text{Frequency multiplication needed}\\ \text{for the carrier frequency}\end{matrix}\right\} = \frac{91.2\times10^6}{200\times10^3} = 456$$

Initial frequency deviation = 49 Hz = $(\Delta f)_1$

Final frequency deviation = 75×10^3 Hz = $(\Delta f)_3$

$$\therefore \qquad \left.\begin{matrix}\text{Frequency multiplication needed}\\ \text{for the carrier frequency}\end{matrix}\right\} = \frac{(\Delta f)_3}{(\Delta f)_1} = \frac{75\times10^3}{49} = 1530.6$$

If we use frequency multiplication of 1530 at one go, the frequency deviation attains the correct value of 75 kHz but the carrier frequency becomes 200 kHz × 1530 = 306 MHz, which is too high a value.

Hence, we shall split the frequency multiplication and perform it in two stages – one before the mixer and the other after the mixer. *The mixer does not change the frequency deviation but can be used for reducing the carrier frequency to a value which when subjected to multiplication by the second stage of frequency multipliers, will give the specified final carrier frequency.*

Since frequency multipliers are generally either doublers or triplers, and since 64 × 24 = 1536 ≈1530, the overall frequency multiplication that we require, let us first subject the NBFM signal to a frequency multiplication of 64.

$$\therefore \qquad f_{c_2} = f_{c_1} \times 64 = 200 \times 10^3 \times 64 = 12.8 \text{ MHz}$$

$$(\Delta f)_2 = (\Delta f)_1 \times 64 = 49 \times 64 = 3.136 \text{ kHz}$$

$$(\Delta f) \text{ at the output of the mixer} = (\Delta f) \text{ at the input to the mixer} = (\Delta f)_2 = 3.136 \text{ kHz}$$

Final carrier frequency required = 91.2 MHz = f_{c_4}

$\therefore$ carrier frequency f_{c_3} at the output of the mixer = $\dfrac{91.2\times10^6}{24} = 3.8$ MHz

For the mixer, the input carrier frequency = f_{c_2} = 12.8 MHz

Hence, the local oscillator frequency of the mixer = (12.8–3.8) MHz = 9 MHz

Figure 5.18 shows the WBFM generator along with the carrier frequencies and frequency deviation at the various stages.

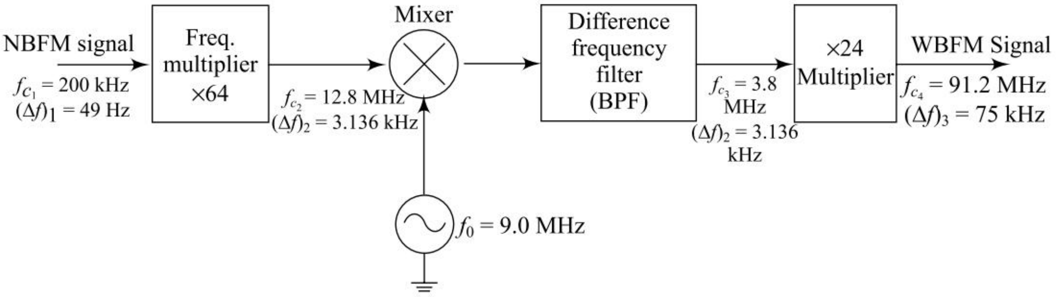


Fig. 5.18 *Indirect method of generation of WBFM of Example 5.23*

Example 5.24 In a WBFM generator of the Armstrong type shown in Fig. 5.15, the initial low frequency carrier is of 200 kHz frequency. The maximum frequency deviation range is from 100 Hz to 15 kHz, and if the final maximum frequency deviation, and the carrier frequency are to be 75 kHz and 102.4 MHz, respectively, choose an appropriate multiplier and the mixer oscillator frequency.

Solution

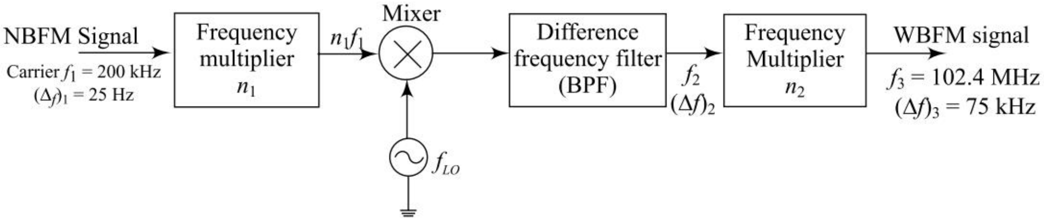


Fig. 5.19 *WBFM generator of Example 5.24*

Since the mixing operation changes the frequency but not the frequency deviation, and since frequency multipliers change both the frequency as well as the deviation, we shall use the frequency multipliers to get the required ratio of frequency deviation (from 25 Hz to 75 kHz) and try to get the final carrier frequency of 102.4 MHz by an appropriate choice of f_{LO}, the frequency of the local oscillator.

$$\therefore \text{ total frequency multiplication needed} = n_1 n_2 = \frac{(\Delta f)_3}{(\Delta f)_1} = \frac{75\times10^3}{25} = 3000$$

Now, $\dfrac{f_3}{n_2} = f_2 = f_{LO} - n_1 f_1$ or $f_3 = n_2 f_{LO} - n_1 n_2 f_1$

$\therefore$ $$102.4\times10^6 = n_2 f_{LO} - 3000\times200\times10^3$$

$\therefore$ $$n_2 f_{LO} = (102.4+600)\times10^6 = 702.4 \text{ MHz}$$

Now, let us choose $n_2 = 100$ and $n_1 = 30$, so that $n_1 n_2 = 3000$

$\therefore$ $$f_{LO} = \frac{702.4\times10^6}{n_2} = \frac{702.4\times10^6}{100} = 7.024 \text{ MHz}$$

Hence, $$f_2 = f_{LO} - n_1 f_1 = 7.024\times10^6 - n_1 f_1 = (7.024-6)\times10^6 \text{ Hz}$$

$\therefore$ $$f_2 = 1.024 \text{ MHz}, \text{ and } f_3 = f_2\times n_2 = 102.4 \text{ MHz}, \text{ as required.}$$

5.6.2 Direct Method of Generation of WBFM

The basic approach of the direct method is quite simple. Here, we vary the frequency of an LC oscillator in accordance with the variations in the amplitude of the message or the modulating signal. This is accomplished by placing an additional reactance across the tank circuit of the oscillator and making this reactance to vary with the amplitude of the message. There are two methods for creating this variable reactance. One is to use a 'varicap', or a varactor diode and the other is to use a reactance modulator.

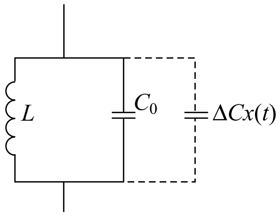


Fig. 5.20 *Variable capacitor across the tank circuit*

Basic principle Let c_0 be the tank circuit capacitance in the absence of any modulation and let

$$f_c = \frac{1}{2\pi\sqrt{LC_0}} \tag{5.58}$$

where f_c is the unmodulated carrier frequency. Let $\Delta Cx(t)$ be the capacitance produced across the tank circuit by the varicap or the reactance modulator.

$\therefore$ total tank circuit capacitance = $C(t) = C_0 + \Delta Cx(t)$; $|x(t)| \le 1$

The instantaneous frequency of the oscillator at the instant 't' is given by

$$f_i(t) = \frac{1}{2\pi\sqrt{LC(t)}} = \frac{1}{2\pi\sqrt{LC_0\left[1+\left(\frac{\Delta C}{C_0}\right)x(t)\right]}}$$

Then, using Eq. (5.58), we may write $f_i(t)$ as

$$f_i(t) = f_c \frac{1}{\sqrt{1+\left(\frac{\Delta C}{C_0}\right)x(t)}} \tag{5.59}$$

Since $|x(t)| \le 1$ and $\left(\frac{\Delta C}{C_0}\right)$ is generally very small,

$\therefore$ $$\left(\frac{\Delta C}{C_0}\right)x(t) \triangleq \in << 1 \tag{5.60}$$

If we now make use of the approximation that

$$\frac{1}{\sqrt{1+\in}} \approx \left(1-\frac{1}{2}\in\right) \tag{5.61}$$

we get

$$f_i(t) = f_c \frac{1}{\sqrt{1+\left(\frac{\Delta C}{C_0}\right)x(t)}} \approx f_c\left[1-\left(\frac{\Delta C}{2C_0}\right)x(t)\right] \tag{5.62}$$

The approximation of Eq.(5.61) is accurate up to 1% for $\left(\frac{\Delta C}{C_0}\right) \le 0.013$

i.e., for $f_c\left(\frac{\Delta C}{2C_0}\right)x(t) \le 0.0065 f_c$

But $f_c\left(\frac{\Delta C}{2C_0}\right)x(t) = \Delta f =$ Peak frequency deviation (see Eq. (5.62))

This amounts to saying that Eq. (5.62) is accurate up to 1% if

$$\left(\frac{\Delta f}{f_c}\right) \le 0.0065 \tag{5.63}$$

Noting that f_c, the unmodulated carrier frequency for FM should be in the VHF range in the 88–108 MHz band, let us take f_c to be typically 100 MHz. Then Eq. (5.63) means that (Δf), the peak frequency deviation that can be obtained has to be limited to

$$(\Delta f) \le 0.0065 \times 10^8 \text{ Hz}$$

or

$$(\Delta f) \le 650 \text{ kHz} \tag{5.64}$$

So, it turns out that this is not at all a restriction since the frequency deviation that we need in practise (75 kHz) is much smaller than 650 kHz.

Thus, we can obtain a WBFM signal by producing a variable capacitor that varies according to the amplitude variations of the modulating signal. This variable capacitor can be realized either by using a varactor diode or by means of a reactance modulator. What is important is that the direct method of generation of WBFM needs only simple circuits that do not involve any frequency multipliers, etc. However, the direct method of generation uses LC oscillators for the carrier generation and these have poor frequency stability. Hence, in order to meet the stringent specifications regarding the carrier frequency stability for transmitters, it becomes necessary to use some Automatic Frequency Control or AFC arrangement in conjunction with these LC oscillators. These are discussed in detail in Chapter 6 which deals with AM and FM transmitters and receivers. We shall now discuss briefly, the two methods – one using a varactor diode and the other using a reactance modulator, for obtaining the capacitance that varies according to the amplitude variations of the modulating signal.

5.6.3 Using a Varactor Diode

The modulating signal $x(t)$ is given in series with the reverse bias for the varactor diode and the diode itself is placed across the tank circuit of an *LC* oscillator. Figure 5.21 shows a tuned-collector *LC* oscillator across whose tank circuit, a varactor diode is connected. The RFC (RF choke) together with the bypass capacitor C_b ensures that the RF from the oscillator does not enter the modulating signal circuit. The coupling condenser, C_c, is of such a small value that it works like a perfect open circuit for the modulating signal frequencies while offering negligible reactance to the RF signal. It also prevents the dc bias supply of the varactor from reaching the oscillator.

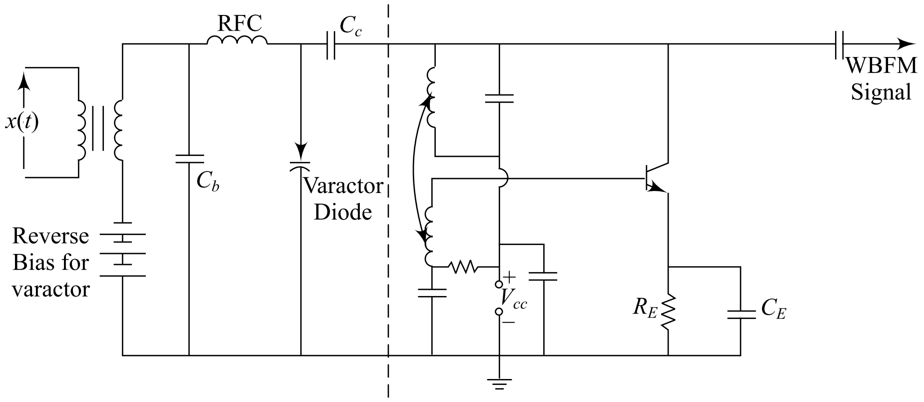


Fig. 5.21 *Typical arrangement for putting the varactor diode across the tank circuit of an LC oscillator*

5.6.4 Using a Reactance Modulator

Principle of operation Let the RF voltage generated by the LC oscillator be applied across the terminals A-A.

In the analysis that follows, we shall make the following assumptions:

1. $i_1 << i_d$
2. $X_c >> R$

The gate voltage $$e_g = i_1 R = \frac{e}{(R - jX_c)} \cdot R \tag{5.65}$$

The drain current $$i_d = g_m \cdot e_g = g_m \cdot \frac{e \cdot R}{(R - jX_c)} \tag{5.66}$$

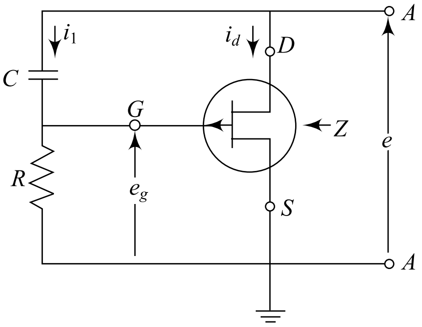


Fig. 5.22 *A FET-based reactance modulator*

Because of our first assumption that $i_1 << i_d$, we may write the impedance seen across the terminals A-A as

$$z = \frac{e}{i_d} = \frac{e(R - jX_c)}{e \cdot g_m \cdot R} = \frac{R - jX_c}{g_m \cdot R} = \frac{1}{g_m}\left[1 - j\frac{X_c}{R}\right] \tag{5.67}$$

But since $X_c >> R$ (second assumption), we may write

$$z \approx -j\frac{X_c}{R \cdot g_m} \tag{5.68}$$

The above equation shows that the impedance z is a capacitive reactance given by

$$X_{eq} = \frac{X_c}{g_m \cdot R} = \frac{1}{2\pi f_c C g_m R} = \frac{1}{2\pi f_c C_{eq}} \tag{5.69}$$

where f_c is the frequency of the oscillator voltage and

$$C_{eq} = g_m \cdot \mathrm{R} \cdot \mathrm{C} \tag{5.70}$$

Hence, the tank circuit of the oscillator, which is connected across the terminals *A-A* will effectively find a capacitance C_{eq} across the terminals. Thus, if we want to make this $C_{eq} = \Delta C \cdot x(t)$ of Fig. 5.20, we should make g_m to vary according to the variations in the amplitude of the message signal, $x(t)$.

Recalling that

$$g_m \triangleq \left.\frac{\partial i_d}{\partial e_g}\right|_{v_{ds}\ \text{constant}} \tag{5.71}$$

all that needs to be done to make g_m proportional to $x(t)$, is to operate the FET in that part of its transfer characteristic where $i_d = Ke_g^2$ (so that $g_m \alpha\, e_g$) and place the message signal $x(t)$ in series with the gate bias voltage.

Note that although Fig. 5.23 shows a FET based reactance modulator, the foregoing analysis is equally applicable to a BJT-based reactance modulator. Figure 5.23 shows a BJT-based reactance modulator used in conjunction with a Colpitt's oscillator for generating wideband FM.

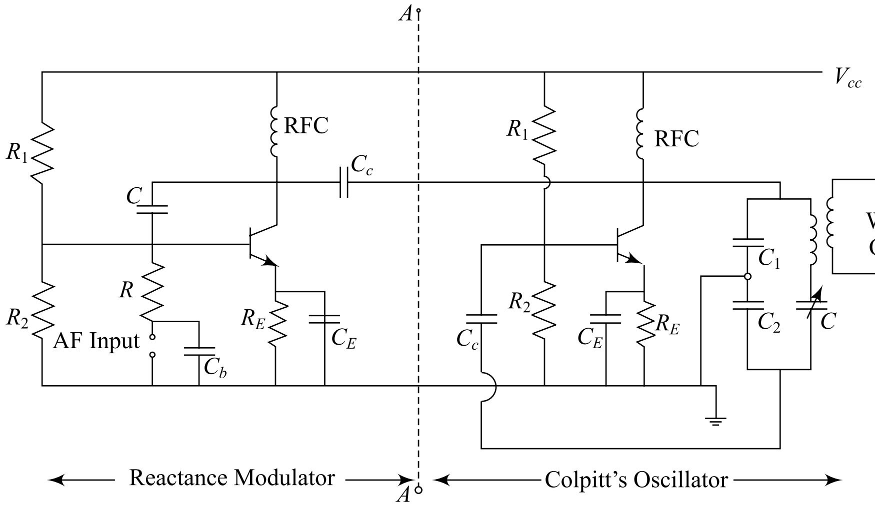


Fig. 5.23 *Direct method of generation of wideband FM using a reactance modulator*

While the varactor diode can only present a capacitive reactance across the tank circuit of the oscillator, a reactance modulator can offer a capacitive, or an inductive reactance across the oscillator tank circuit. For instance, if the positions of *R* and *C* are interchanged in the reactance modulator circuit if Fig. 5.22, it can be shown that the circuit to the left of terminals *A*-*A* will appear as an inductive reactance.

Reactance modulator offers better stability than the varactor diode circuit. However, it suffers from the disadvantage that the input impedance is very small. This is because of the small values of *R* and X_c of the series *RC* circuit. Further, at very high frequencies of the oscillator, the X_c becomes small (even if we make *C* quite small) and hence *R* also has to be made small, reducing the input impedance to such low values as to make the circuit unworkable. So, to realize the carrier frequencies required for FM, it becomes necessary to use frequency multipliers while working the oscillator at a low frequency, typically less than about 5 MHz.

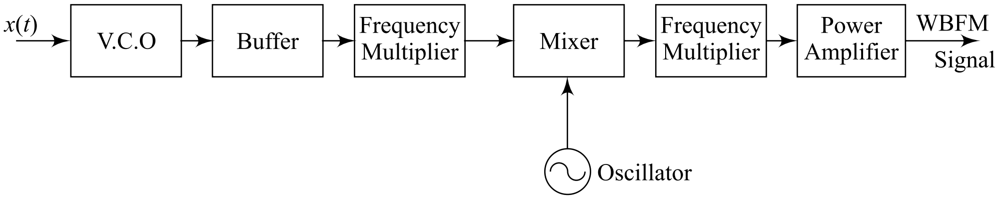


Fig. 5.24 *Varactor diode direct method of generation of a WBFM signal*

Fig. 5.25 *Reactance-modulator method of generation of WBFM signal*

As mentioned earlier at the beginning of this section, both the methods suffer from the disadvantage that the carrier frequency is obtained from an *LC* oscillator and not a crystal oscillator. Hence, the carrier frequency stability will be poor. This makes it necessary to use some automatic frequency control wherein the carrier frequency of the WBFM signal is controlled by a crystal oscillator. Details of the AFC circuit are given in Chapter 7 in which transmitter details are discussed.

5.6.5 Comparison of Narrowband and Wideband FM

Wideband FM typically has a maximum deviation of 75 kHz and makes use of audio frequencies up to 15 kHz. Thus, it is eminently suitable for high quality music broadcasting, since FM has considerable immunity for additive noise. The bandwidths occupied by these wideband FM signals are, of course, large and are of the order of 200 kHz.

Narrowband FM (strictly speaking, it is not NBFM as defined earlier) on the other hand is used for FM mobile communication systems operating in the VHF band and used by the police department, by the taxis and for ship-to-shore communication. Unlike music, speech requires only intelligibility but not high quality. Hence, audio frequencies in the range of 30 Hz to about 3 kHz or 5 kHz would be quite sufficient. Even for these speech (or telephone) quality audio frequency ranges, the bandwidth required for these so-called NBFM communication systems may be of the order of 25 to 30 kHz since frequency deviations of the order of 10 to 15 kHz are used in order to get at least some degree of noise immunity, as the NBFM in the strict sense, is no better than conventional AM insofar as noise performance is concerned. These point-to-point FM mobile communication systems operating in the VHF band also make use of pre-emphasis and de-emphasis in order to get good SNR at the destination.

5.7 EFFECTS OF CHANNEL NON-LINEARITIES ON FM SIGNALS

We shall now briefly discuss the effect of passing an FM signal

$$x_c(t) = A_c \cos[2\pi f_c t + \phi(t)] \tag{5.72}$$

where

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha) d\alpha \tag{5.73}$$

through a memoryless channel having a non-linear input-output relation such as

$$e_0(t) = a_0 + a_1 e_i(t) + a_2 e_i^2(t) + a_3 e_i^3(t) \tag{5.74}$$

where $e_i(t)$ and $e_0(t)$ represent, respectively, the input and output voltages, while a_0, a_1, a_2 and a_3 are constants. Replacing $e_i(t)$ by $x_c(t)$ in Eq. (5.74), expanding the terms on the right-hand side and rearranging the terms, we get

$$\begin{aligned} e_0(t) = &\left(a_0 + \frac{1}{2}a_2 A_c^2\right) + \left(a_1 A_c + \frac{3}{4}a_3 A_c^3\right)\cos[2\pi f_c t + \phi(t)] \\ &+ \frac{1}{2}a_2 A_c^2 \cos[4\pi f_c t + 2\phi(t)] + \frac{1}{4}a_3 A_c^3 \cos[6\pi f_c t + 3\phi(t)] \end{aligned} \tag{5.75}$$

Thus, the channel output, $e_0(t)$ consists of a dc component and three FM signals with f_c, $2f_c$ and $3f_c$ as their carrier frequencies. The FM signal with carrier frequency f_c, which is the desired component, can be separated out from the rest by using a BPF with center frequency f_c and bandwidth equal to

$$2\left[k_f\,|x(t)|_{\max}+W\right] \tag{5.76}$$

where W is the bandwidth of the message signal, $x(t)$. The BPF output is

$$\left(a_1A_c+\frac{3}{4}a_3A_c^3\right)\cos[2\pi f_ct+\phi(t)] \tag{5.77}$$

Thus, we find that the effect of the non-linearity of the channel is only to change the amplitude of the FM signal, which, of course, does not cause any problem, as it is no distortion. *On the other hand, instead of an FM signal, if we had passed an AM signal through the same channel, it would have got terribly distorted.* This therefore, indicates the advantage in using FM when the channel includes devices like say, the TWT amplifier which generally has a non-linear input-output relation when it is operating at its power limit.

However, channel non-linearities of a type which produce phase changes with signal amplitude changes, will create problems, and it should be ensured that such non-linearities are very small.

Example 5.25 The tank circuit of a 0.5 MHz LC oscillator has an inductance of 1 mH connected across a capacitor. The output of this oscillator is frequency modulated by an FET reactance modulator consisting of a series connection of a 1500 W resistor and a 10 pf capacitor, with the capacitor connected between the gate and drain of the FET. The message signal varies the mutual conductance of the FET by ±0.6 mA/volt, find the peak frequency deviation that is produced.

Solution $c_{eq}=g_mRC=0.6\times10^{-3}\times1500\times10\times10^{-12}$

$= 9$ pf $= \Delta C$

$\therefore$ peak frequency deviation = $(f.\Delta C)/2C_0$

$$=\frac{0.5\times10^6\times9\times10^{-12}}{2\times C_0}$$

where C_0 = tank circuit capacitance

$\therefore$ since $f_0=\dfrac{1}{2\pi\sqrt{LC_0}},\quad C_0=\dfrac{1}{4\pi^2\times f_0^2\times L}$

$$C_0=\frac{1}{4\pi^2\times25\times10^{10}\times10^{-3}}=10^{-10}\text{F}\,.$$

$\therefore$ peak frequency deviation $=\dfrac{0.5\times10^6\times9\times10^{-12}}{2\times10^{-10}}=22.5$ kHz

5.8 DETECTION OF FM SIGNALS

Since in FM the carrier frequency is changed in accordance with the amplitude of the modulating signal, FM signal demodulation is essentially one of frequency-to-amplitude conversion. There are several FM demodulators – the slope detector, the phase discriminator of Foster and Seeley, the ratio detector, the FM feedback detector, the quadrature FM detector, the zero-crossing detector and the phase-locked loop detector. The slope detector, historically the earliest and also the simplest of all, is, of course, no longer in use; but once the principle of it is understood, it is easy to understand the phase discriminator and ratio detector. So we shall first briefly discuss the principle of the slope detector.

5.8.1 Slope Detector

In the mixer stage of the receiver, the carrier frequency of the received signal is changed to a fixed frequency called the intermediate frequency, f_{if}, which has a value of 10.7 MHz in the case of standard FM broadcast receivers. Hence, the FM signal arriving at the input to the discriminator (from the IF stage) is having a carrier frequency of f_{if}. The slope detector simply consists of a resonant circuit tuned to a frequency f_0 which is slightly more than f_{if} followed by an envelope detector. The f_0 is so chosen that f_{if} falls in the middle of the range of frequencies over which the response of the resonant circuit is almost linear. This region is from f_{min} to f_{max} as shown in Fig. 5.26.

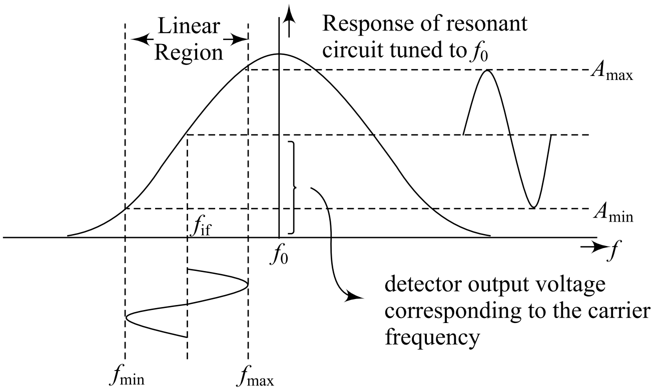


Fig. 5.26 *Principle of slope detector*

As can be seen from Fig. 5.26, the frequency variations of the input FM signal are converted into corresponding changes in voltage at the output of the detector. This frequency-to-voltage conversion will be linear to the extent that the region marked ‘linear region’ is really linear. Thus, the slope detector converts the FM signal into an AM signal with carrier frequency of f_{if} = 10.7 MHz and modulating signal the same that the FM signal was carrying. The AM signal can be detected and the modulating signal extracted by using a conventional envelope detector as shown in Fig. 5.27.

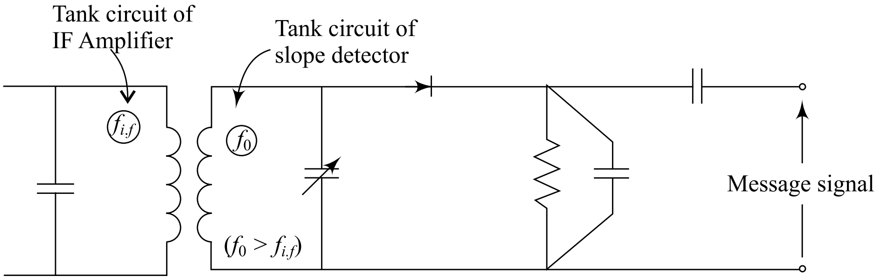
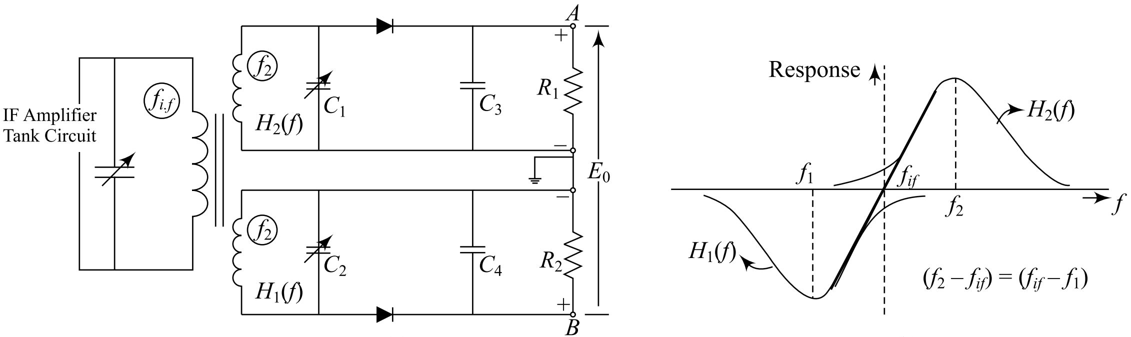


Fig. 5.27 *Frequency-to-amplitude converter followed by an envelope detector*

Although it is simple and inexpensive, the slope detector suffers from one serious disadvantage, viz., non-linearity in the frequency-to-amplitude conversion. This non-linearity arises from the fact that the response curve of the resonant circuit can be considered to be linear only over a very small region.

5.8.2 Dual-Slope Detector or Balanced Discriminator

To overcome the problem of non-linearity encountered in the simple slope detector discussed earlier, Foster and Seeley proposed the dual-slope detector. This makes use of two resonant circuits with identical responses but with slightly different resonant frequencies. The technique used in order to obtain a larger linear range is illustrated in Fig. 5.28(b).



(a) (b)

Fig. 5.28 *(a) Dual slope detector circuit, (b) Technique for larger linear range*

When the incoming signal frequency is equal to the IF, the responses of $H_1(f)$ and $H_2(f)$ will be equal and so the voltages developed across R_1 and R_2 will be equal. From the way D_1 and D_2 are connected, terminals A and B will be at the same potential with respect to the ground and so E_0, the potential difference between them is zero. If the incoming signal has a frequency above the IF, the response of $H_2(f)$ will be more and that of $H_1(f)$ will be less (when compared to what it was when incoming signal frequency was IF). Hence, the voltage drop across R_1 will be greater than the voltage drop across R_2. Hence, terminal A will be at a higher potential than terminal B with respect to ground and $E_0 \neq 0$. If the incoming signal has a frequency less than the IF, response $H_1(f)$ will be more than the response $H_2(f)$, causing B to be at a higher potential than A. Thus, the frequency variations of the incoming FM signal are converted into corresponding variations in the amplitude of E_0. Therefore E_0 will be the modulating signal assuming the overall response (see Fig. 5.28(b)) to be perfectly linear between f_1 and f_2.

5.8.3 Foster–Seeley Discriminator

Originally developed as a subsystem of an automatic frequency control unit, this FM detector is known as Foster–Seeley discriminator, Phase-shift discriminator and center-tuned discriminator.

A tank circuit consisting of a center-tapped inductance L_2 and capacitor C_2 is inductively coupled to the inductance L_1 of the tank-circuit of the IF stage. The diodes D_1 and D_2 and the elements R_3, C_3 and R_4, C_4 are connected to this secondary side tank circuit as shown in Fig. 5.29. Further, a large RF coupling capacitor C and a large RF choke are connected as shown in the figure.

The primary and secondary tank circuits are tuned to the same frequency – the IF, which is the carrier frequency for the FM signal being fed to the discriminator. At the RF, the circuit comprising C, L and C_4 is effectively coming across L_1. Since the reactance of the RF choke L far exceeds the reactances of C and C_4, the voltage across the choke L, say V_L, is practically equal to the voltage across the primary, i.e., V_P.

$$\therefore \qquad V_L \cong V_P \tag{5.78}$$

If M is the mutual inductance between the primary and secondary windings, the voltage induced in the secondary, viz., V_s, is given by

$$V_s = \pm j\omega M I_p \tag{5.79}$$

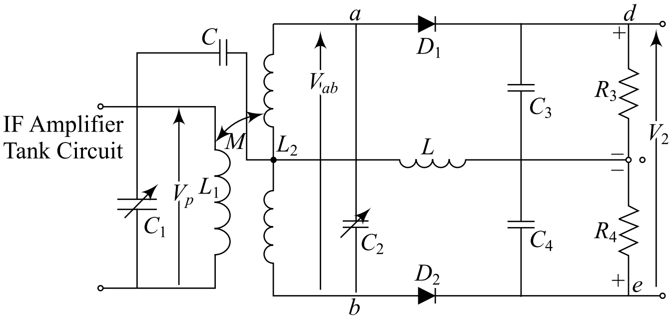


Fig. 5.29 *Foster–Seeley discriminator*

The direction of winding of the secondary determines whether the positive or the negative sign is to be used. I_p in Eq. (5.79) above, denotes the current flowing through the primary winding L_1, and is given by

$$I_P \cong \frac{V_P}{j\omega L_1} \tag{5.80}$$

While writing Eq. (5.80), we have assumed that the secondary side load impedance reflected into the primary, as well as the resistance of the primary coil is negligible, *since the Q-factors of the primary and secondary are large and the mutual inductance M is small.*

Taking the negative sign in RHS of Eq. (5.79) and substituting in it for I_p using Eq. (5.80), we get

$$V_s = -\frac{M}{L_1}V_p \tag{5.81}$$

This induced voltage V_s produces a voltage drop V_{ab} across the capacitor C_2 given by

$$V_{ab} = \frac{V_s(1/j\omega C_2)}{R_2 + j\omega L_2 + (1/j\omega C_2)} = \frac{MV_p}{L_1}\left[\frac{1}{\left\{\left(\frac{\omega}{\omega_c}\right)^2 - 1\right\} - j\omega C_2 R_2}\right] \tag{5.82}$$

Hence, when the frequency f of the incoming FM signal is equal to the IF, i.e., f_c, then

$$V_{ab} = j\left[\frac{M}{L_1\omega C_2 R_2}\right]V_p \tag{5.83}$$

i.e., V_{ab} leads V_p by 90°.

The voltage V_{a0} applied to diode D_1 is given by

$$V_{ao} = \frac{1}{2}V_{ab} + V_L = \frac{1}{2}V_{ab} + V_P \tag{5.84}$$

The voltage V_{bo} applied to diode D_2 is given by

$$V_{bo} = -\frac{1}{2}V_{ab} + V_L = -\frac{1}{2}V_{ab} + V_P \tag{5.85}$$

Hence, when $f = f_c$, i.e., when there is no modulation for the incoming signal, the phasor diagram will be as shown in Fig. 5.30(a).

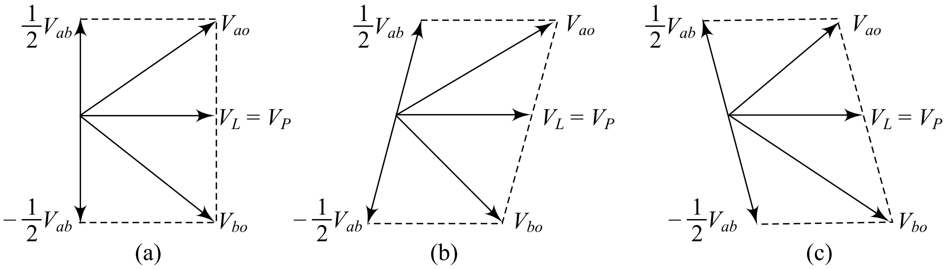


Fig. 5.30 *Phasor diagrams showing voltages across D_1 and D_2 for (a) $f = f_c$, (b) $f > f_c$, (c) $f < f_c$*

Thus $V_{ao} = V_{bo}$. The diode D_1 charges capacitor C_3 and diode D_2 charges capacitor C_4. Neglecting the diode drops and assuming R_3C_3 and R_4C_4 to be large compared to $(1/f_c)$, we may say that C_3 and C_4 will be charged to the peak values of the voltage V_{ao} and V_{bo} respectively. From Fig. 5.30(a), we find that when $f = f_c$, $|V_{ao}| = |V_{bo}|$. Hence,

$$V_{do} = V_{eo}$$

and therefore,

$$V_2 = 0$$

From Eq. (5.82), we find that the phasor diagrams for $f > f_c$ and $f < f_c$ will be as shown in Figs. 5.25(b) and (c) respectively and that

1. For $f > f_c$: $|V_{ao}| > |V_{bo}|$ $\therefore$ V_2 is positive and equal to $|V_{ao}| - |V_{bo}|$
2. For $f < f_c$: $|V_{ao}| < |V_{bo}|$ $\therefore$ V_2 is negative and equal in magnitude to $|V_{bo}| - |V_{ao}|$

For the Foster–Seeley discriminator, if we plot the frequency response around f_c, we will get the S-shaped curve similar to the one shown in Fig. 5.28(b); and the frequency-to-amplitude conversion is fairly linear if the discriminator is properly designed. However, this discriminator responds to amplitude variations also, as is evident from Eqs. (5.83) and (5.85). *Hence, if this discriminator is used, it must be preceded by a limiter stage.*

5.8.4 Ratio Detector

The ratio detector is a modified version of the Foster–Seeley discriminator, the modifications being such as to make it unresponsive to the amplitude variations of the incoming FM signal while responding in the same way as the Foster–Seeley circuit for the input signal's frequency variations.

The circuit of a ratio detector is shown in Fig. 5.31.

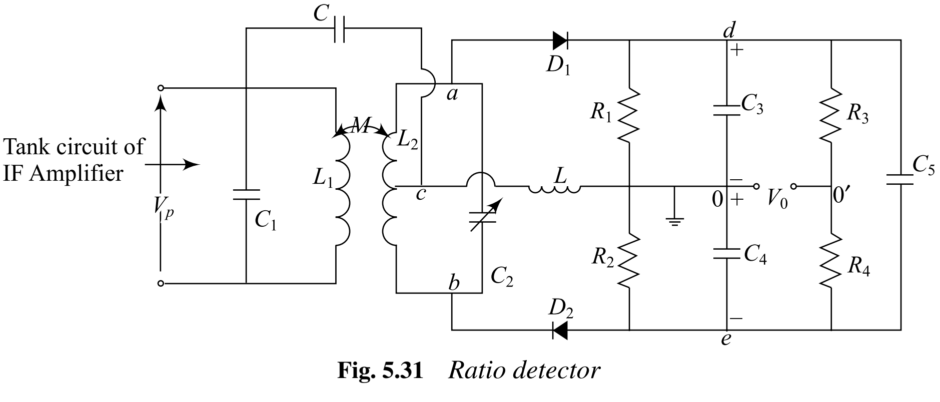


Fig. 5.31 *Ratio detector*

It may be noted that the ratio detector circuit is essentially the same as that of the Foster–Seeley discriminator except for the following three modifications:

1. Diode D_2 is reversed in direction.
2. A large capacitor C_5 is connected across the output voltage of the two diodes.
3. The output voltage of the detector is drawn across 0′ and 0.

5.8.5 Frequency-to-Amplitude Conversion

In the Foster–Seeley discriminator, we had seen that the output voltage is equal to the difference between the output voltages of the two diodes and that it varies in amplitude according to the amount of deviation in frequency of the input FM signal from the unmodulated carrier frequency. We shall now show that although the output voltage is now taken between the terminals 0′ and 0, it is still proportional to the difference in the diode output voltages.

Output voltage

$$V_0 = V_{0'e} - V_{0e}$$

But since

$$R_3 = R_4, V_{0'e} = \frac{1}{2}(V_{de})$$

∴

$$V_0 = \frac{V_{de}}{2} - V_{0e} = \frac{V_{d0} + V_{0e}}{2} - V_{0e} = \frac{V_{d0} - V_{0e}}{2} \tag{5.86}$$

V_{do} is the dc output voltage of diode D_1 and V_{oe} is the dc output voltage of diode D_2. Thus, just like in the Foster–Seeley circuit, here also, the output voltage is proportional to the difference between the diode output voltages. The only difference is that whereas it was equal to the difference between the diode output voltages for the Foster–Seeley discriminator, in the case of the ratio detector, it is *one-half* of it. However, here too, the output voltage amplitude varies in accordance with the amount of deviation of the input signal frequency from the unmodulated carrier frequency – just like the Foster–Seeley discriminator.

5.8.6 Response to Amplitude Variations

We shall now show, in a qualitative manner, how the ratio detector responds to changes in the amplitude of the incoming FM signal so as to make its output unaffected by these amplitude variations.

It is the capacitor C_5 that makes the ratio detector's output to be unaffected by amplitude variations. This it does in two ways. Primarily the large time constant associated with it does not permit the voltage across it to change quickly. Thus, irrespective of changes in the amplitude of the incoming signal, it tries to maintain a constant voltage V_{de}. That is the sum of the two diode output voltages is kept constant even while the difference between them changes as the frequency of the incoming signal changes. Secondly, it helps in bringing into play as amplitude-dependent damping of the primary and secondary tank circuits in such a way as to offset the effect of any increase or decrease of the amplitude. For example, if the amplitude tries to increase suddenly, a larger charging current tends to flow into the capacitor C_5. But, as its voltage and therefore the load voltage cannot increase suddenly, it amounts to having a low value of load presented to the secondary side tank circuit and its Q-factor is lowered. Because of the reflected load, the primary side tank circuit also will have its Q-factor lowered. These changes in Q will lower the IF amplifier gain and therefore the amplitude of the incoming signal fed to the detector gets reduced automatically. Similarly, when the amplitude of incoming signal decreases suddenly, the loading on the tank circuit will decreases their Q-factor will increase and the IF amplifier gain will increase, which in turn will tend to increase the amplitude of the signal fed to the discriminator.

Because of the above reasons, the ratio detector does not respond to sudden changes in the amplitude of the incoming FM signal such as those caused by the additive noise on the channel. *Slow fading of the signal, however, does cause the ratio detector output voltage to change accordingly.*

5.8.7 Quadrature FM Detector

In this, a quadrature signal is first generated from the received FM signal (i.e., output of the IF amplifier or the limiter) by passing it through a delay line or a phase-shift network. This delay line/phase-shift network is so designed that at carrier frequency it gives a phase shift of 90° while giving a group delay of say some t_1 sec. As shown in Fig. 5.32, this quadrature signal is multiplied by the FM signal given as input to the delay line/phase-shift network and the product is low pass filtered with an LPF having a cut-off frequency of W Hz, which is the message bandwidth.

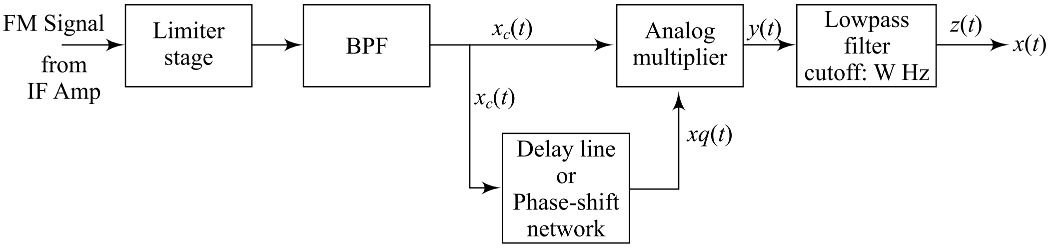


Fig. 5.32 *A quadrature FM detector*

Let the FM signal from the IF amplifier fed to the delay-line and the analog multiplier be represented as

$$\left.\begin{aligned} x_c(t) &= A_c\cos[\omega_c t+\phi(t)] \\ \text{where,}\quad \phi(t) &= 2\pi k_f\int_0^t x(\alpha)d\alpha \end{aligned}\right\} \tag{5.87}$$

k_f being the deviation constant and $x(t)$, the normalized message signal. Then the quadrature signal is

$$\begin{aligned} x_q(t) &= A_c\cos\lfloor\omega_c t-90^\circ+\phi(t-t_1)\rfloor \\ &= A_c\sin[\omega_c t+\phi(t-t_1)] \end{aligned} \tag{5.88}$$

$\therefore$ output of the analog multiplier is given by

$$y(t)=\sin[\omega_c t+\phi(t-t_1)]\cos[\omega_c t+\phi(t)]$$

$$\therefore \qquad \begin{aligned} y(t) &= [\sin\omega_c t\cos\phi(t-t_1)+\cos\omega_c t\sin\phi(t-t_1)][\cos\omega_c t\cos\phi(t)-\sin\omega_c t\sin\phi(t)] \\ &= \frac{1}{2}\sin 2\omega_c t\cos\phi(t)\cos\phi(t-t_1)+\cos^2\omega_c t\cdot\cos\phi(t)\sin\phi(t-t_1) \\ &= \sin^2\omega_c t\cdot\sin\phi(t)\cos\phi(t-t_1)-\frac{1}{2}\sin 2\omega_c t\cdot\sin\phi(t)\sin\phi(t-t_1) \end{aligned}$$

Writing $(1+\cos 2\omega_c t)/2$ for $\cos^2\omega_c t$ and $(1-\cos 2\omega_c t)/2$ for $\sin^2\omega_c t$, and canceling all high frequency terms involving $\sin 2\omega_c t$ and $\cos 2\omega_c t$, we get the output of the low pass filter as

$$z(t)=K_1\sin[\phi(t)-\phi(t-t_1)] \tag{5.89}$$

where K_1 is a constant.

If $[\phi(t)-\phi(t-t_1)]$ is very small, say very much less than π radians (this will be the case, since t_1, the group delay, will generally be quite small), then we may make the approximation

$$\sin[\phi(t)-\phi(t-t_1)]\cong[\phi(t)-\phi(t-t_1)] \tag{5.90}$$

again,

$$\frac{d\phi(t)}{dt}\approx\left[\frac{[\phi(t)-\phi(t-t_1)]}{t_1}\right] \qquad \text{(meaning of a derivative)} \tag{5.91}$$

This is true when t_1 is quite small, in fact so small that $\phi(t)$ does not change much between $(t - t_1)$ and t.

$\therefore$
$$z(t) = K_1[\phi(t) - \phi(t - t_1)] = K_1 t_1 \frac{d\phi(t)}{dt} = K_2 \frac{d\phi(t)}{dt} \tag{5.92}$$

where K_2 includes t_1.

But
$$\frac{d\phi(t)}{dt} = 2\pi k_f x(t) \quad \text{(from Eq. (5.87))} \tag{5.93}$$

$\therefore$
$$z(t) = K_3 x(t) \tag{5.94}$$

where
$$K_3 = K_2 \cdot 2\pi k_f$$

$\therefore$ the output of the LPF is proportional to the message signal. Thus, the set-up of Fig. 5.27 acts as an FM detector. In fact, even though several approximations have been made in the above analysis, the quadrature FM detector provides better linearity than the Foster–Seeley discriminator. Hence, it is used in some of the expensive FM receivers as it gives a better audio quality.

5.8.8 Zero-Crossing FM Detector

An FM detector with an excellent linear relation between input frequency and output voltage, is the zero-crossing FM detector. In this detector, a hard-limiter first converts the incoming FM signal into a rectangular waveform. A mono-stable multivibrator which is designed to get triggered by the rising edges of this rectangular waveform produces rectangular pulses of fixed duration τ as shown in Fig. 5.33(d).

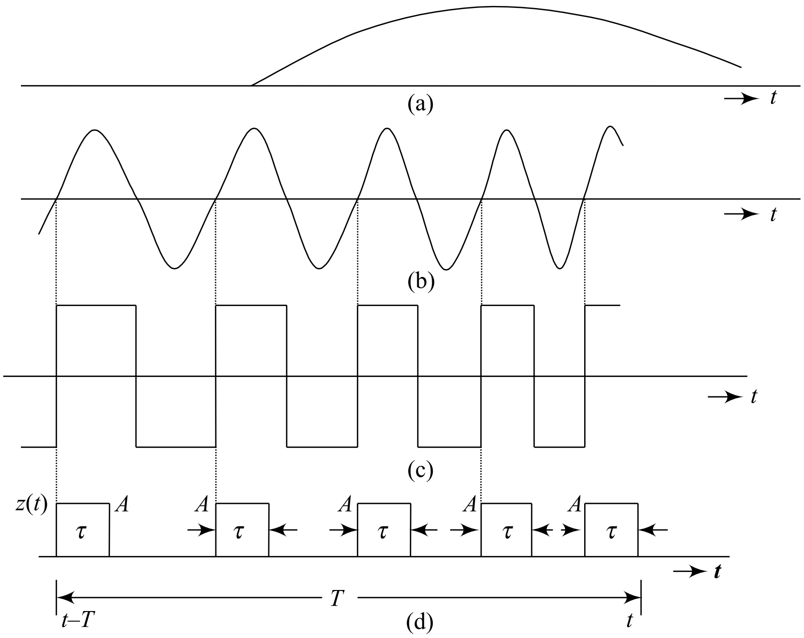


Fig. 5.33 *Waveforms to illustrate the principle of working of a zero-crossing detector (a) Modulating signal (assumed to be singe-tone), (b) FM signal, (c) Hard-limited FM signal, (d) Output of the mono-stable multivibrator*

If this waveform in (d) is integrated for a period of T seconds such that

$$\frac{1}{f_c} << T << \frac{1}{W}$$

where f_c = unmodulated carrier frequency of the FM signal given as input to the detector (i.e., IF)

and W = Bandwidth of the message signal

Then,

$$\frac{1}{T}\int_{t-T}^{t} z(\lambda)d\lambda = \frac{nA\tau}{T} \tag{5.95}$$

where n is the number of zero-crossings which is proportional to the (instantaneous) frequency. Thus, the integrator output is proportional to the frequency. A practical form of a balanced zero crossing detector is illustrated in Fig. 5.34.

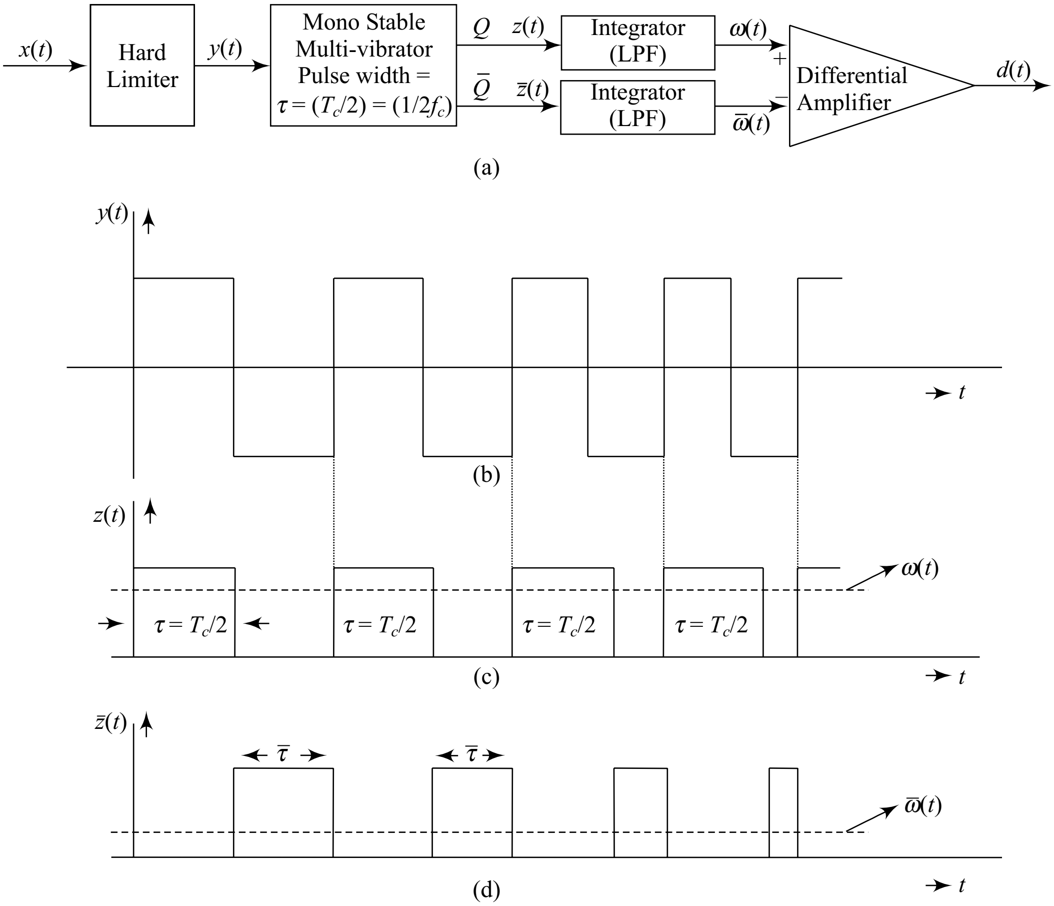


Fig. 5.34 *Balanced zero-crossing detector (a) Circuit, (b), (c), and (d) Waveforms*

When there is no modulation, $y(t)$ will have 50% duty cycle and hence $\omega(t) = \bar{\omega}(t)$ so that $d(t) = 0$. As the frequency increases above f_c, $\omega(t)$ increases while $\bar{\omega}(t)$ decreases. Hence, $d(t)$ is positive and increases with frequency deviation above f_c. When frequency decreases below f_c, $d(t)$ is negative and its amplitude increases with frequency deviation below f_c. Practical balanced zero-crossing FM detectors can have better than 0.1% linearity and they can operate up to even 10 MHz. Higher operating frequencies may be obtained by resorting to frequency division after the hard limiter.

5.8.9 Phase-Locked Loop (PLL) Detector

There exists one disadvantage with all the FM demodulation methods described earlier. All these methods have the same bandwidth as the bandwidth occupied by the FM signal, which, of course, is very much more than the bandwidth of the message signal. Thus, these demodulators pass on all the noise contained in the bandwidth of the FM signal.

Using feedback to reduce this bandwidth and thereby reduce the noise power at the output of the demodulator is one way of tackling the problem. Such an approach leads to what is known as an FM demodulator with feedback (FMFB). A demodulator using this approach is the phase-locked loop (PLL).

The block diagram of an arrangement that uses a PLL for FM demodulation is shown in Fig. 5.35.

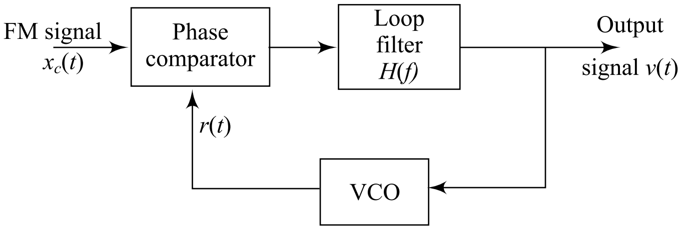


Fig. 5.35 *Block diagram of a phase-locked loop*

As can be seen from this figure, it is a feedback system, in fact, a negative feedback system comprising a phase comparator and a loop filter with a VCO in the feedback path. The phase comparator is just a product device. The loop filter has a high gain and a passband from 0 Hz to W Hz. The VCO is a voltage-controlled oscillator, whose output is a sine wave, the frequency of which is determined by the control voltage given as input to it. In fact, for our purpose here, any system that can generate an FM signal can be used as the VCO.

For a mathematical analysis of the system, let us assume that the VCO has been initially so adjusted that

1. it produces a sine wave with a frequency exactly equal to the unmodulated carrier frequency of the incoming FM signal, when there is no control voltage applied to it; and
2. the sine wave signal that it generates under the condition stated above has a 90° phase difference with the carrier signal of the incoming FM signal.

Accordingly, let us assume that the incoming FM signal, $x_c(t)$ is given by

$$x_c(t) = A_c \sin[2\pi f_c t + \phi(t)] \tag{5.96}$$

where

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha)d\alpha$$

Let the loop filter output be $v(t)$. Since this controls the frequency of the VCO output signal $r(t)$, the frequency of $r(t)$ is given by

$$f_r(t) = f_c + k_v v(t) \tag{5.97}$$

and

$$r(t) = A_r \cos[2\pi f_c t + \phi_r(t)] \tag{5.98}$$

where

$$\phi_r(t) = 2\pi k_v \int_0^t v(\tau)d\tau \tag{5.99}$$

Since the phase comparator multiplies the two signals given to it,

$$\begin{aligned} x_c(t)\cdot r(t) &= A_c A_r \sin[2\pi f_c t + \phi(t)]\cos[2\pi f_c t + \phi_r(t)] \\ &= \frac{1}{2}A_c A_r\{\sin[4\pi f_c t + \phi(t) + \phi_r(t)]\sin[\phi(t) - \phi_r(t)]\} \end{aligned}$$

The filter eliminates the high frequency component at the frequency of $2f_c$.

$\therefore$ Output of the loop filter = $v(t) = \frac{1}{2}A_c A_r \sin[\phi(t) - \phi_r(t)]$ (5.100)

If the PLL is in the phase-locked condition,

$$\phi_e(t) \triangleq \phi(t) - \phi_r(t) \tag{5.101}$$

will be very small.

Hence, we may make the approximation $\sin\phi_e \cong \phi_e$

$$v(t) = \frac{1}{2}A_c A_r \phi_e \tag{5.102}$$

and

$$\phi_e(t) = \phi(t) - 2\pi k_v \int_0^t v(\tau)d\tau \qquad \text{(From Eqs. (5.99) and (5.101))}$$

$$\therefore \quad \frac{d}{dt}\phi_e(t) + 2\pi k_v v(t) = \frac{d}{dt}\phi(t); \qquad v(t) = \int_0^\infty \phi_e(t)h(t-\tau)d\tau \tag{5.103}$$

(since $v(t)$ is the output of the LTI filter with $h(t)$ as its impulse response)

Because of the approximation we made, that $\sin\phi_e \cong \phi_e$, we are now getting a *linear* differential equation relating ϕ, ϕ_e and $v(t)$. Using this, we may now draw the linearized version of the PLL as shown in Fig. 5.36.

Taking the Fourier transform of the differential equation, we have

$$j2\pi f\Phi_e(f) + 2\pi k_v \Phi_e(f)H(f) = j2\pi f\Phi(f)$$

$$\Phi_e(f) = \frac{\Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)}$$

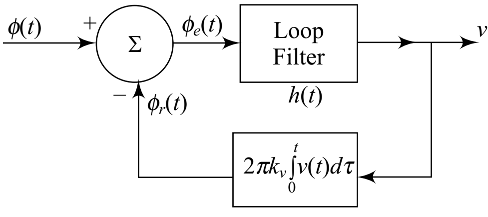


Fig. 5.36 *Linearized equivalent circuit of the PLL*

But $\qquad V(f) = \Phi_e(f)H(f), \qquad$ (since $v(t) = \phi_e(t) * h(t)$)

$$V(f) = \frac{H(f)\cdot\Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)} \tag{5.104}$$

If the gain of the loop filter is high enough so that

$$\left(\frac{k_v}{jf}\right)H(f) >> 1 \text{ for } |f| < W \tag{5.105}$$

Then,

$$V(f) = \left(\frac{jf}{k_v}\right)\Phi(f) = \left(\frac{j2\pi f}{2\pi k_v}\right)\Phi(f) \tag{5.106}$$

$$\therefore \qquad v(t) = \left(\frac{1}{2\pi k_v}\right)\frac{d}{dt}\phi(t) = \frac{d}{dt}\left[2\pi k_f \int_0^t x(\alpha)\,d\alpha\right]\cdot\left[\frac{1}{2\pi k_v}\right]$$

$$\therefore \qquad v(t) = \left(\frac{k_f}{k_v}\right)x(t) \tag{5.107}$$

v(*t*) is proportional to the modulating signal, *x*(*t*), and hence *v*(*t*) is the demodulated signal. If *X*(*f*) = 0 for $|f| \geq W$ Hz, *V*(*f*) also is zero for $|f| \geq W$. ∴ *H*(*f*) can be made equal to zero for all *f* such that $|f| \geq W$. That is noise at the output of the loop filter will be limited to only the message bandwidth, unlike in the case of the demodulators discussed earlier. As we are going to see later in Chapter 9, there is a 'threshold effect' for FM in the sense that if the signal-to-noise ratio at the input to an FM detector is less than a certain critical value, called the 'threshold', the output of the receiver will be only noise. We are going to see in that chapter that a PLL may be used as the FM detector to lower the threshold.

Example 5.26 Show that *d*(*t*), the output of the balanced zero-crossing detector shown in Fig. 5.29(a) is approximately proportional to the amplitude of the normalized modulating signal *x*(*t*) of the input FM signal, $x_c(t)$.

Solution Referring to Fig. 5.29, let the integrating period be *T* where,

$$\frac{1}{f_c} << T << \frac{1}{W}, \quad \text{as stated in Eq. (5.94)}$$

Then

$$\langle\omega(t)\rangle = \frac{1}{T}\int_{t-T}^{t} z(\lambda)d\lambda = \frac{nA\tau}{T} \quad \text{as per Eq. (5.95)}$$

where $\langle\omega(t)\rangle$ is the average value of *x*(*t*) over a period *T*.

$$\langle\omega(t)\rangle = \frac{nA\tau}{T} = \frac{\lfloor f_c + k_f x(t)\rfloor TA\tau}{T} = A\tau[f_c + k_f x(t)]$$

where $[f_c + k_f x(t)] = f_i$ = instantaneous frequency = *n*, the number of *positive going zero-crossings* of *y*(*t*).

But
$$\tau = \frac{T_c}{2} = \frac{1}{2f_c} \quad \therefore \langle\omega(t)\rangle = A\tau[f_c + k_f x(t)] = \frac{A}{2} + \frac{A}{2}\left(\frac{\Delta f}{f_c}\right) \tag{5.108}$$

$\overline{\tau}$ changes with the amplitude of *x*(*t*) and is given by

$$\overline{\tau} = \left[\frac{1}{f_c + k_f x(t)_c} - \frac{1}{2f_c}\right] = \frac{1}{f_c}\left[\frac{1}{1 + (k_f/f_c)x(t)} - \frac{1}{2}\right] \approx \frac{1}{f_c}\left(1 - \frac{1}{2}\frac{k_f}{f_c}x(t) - \frac{1}{2}\right)$$

$$= \frac{1}{2f_c}\left(1 - \left(\frac{k_f}{f_c}\right)x(t)\right)$$

$$\langle \bar{\omega}(t) \rangle = \frac{nA\bar{\tau}}{T} = \frac{[f_c + k_f x(t)]TA\left(\frac{1}{2f_c}\right)\left[1-\left(\frac{k_f}{f_c}\right)x(t)\right]}{T} = \frac{A}{2}\left[1-\left(\frac{k_f}{f_c}\right)^2 x^2(t)\right]$$

$$\therefore \qquad \langle d(t) \rangle = \frac{nA}{T}[\tau - \bar{\tau}] = \left[1-\left(\frac{\Delta f}{f_c}\right)x^2(t)\right]\left(\frac{\Delta f}{f_c}\right)$$

Since $\left(\frac{\Delta f}{f_c}\right) << 1$ and $x(t)$ has been normalized such that $|x(t)| \le 1$,

$$\langle d(t) \rangle \cong \left(\frac{\Delta f}{f_c}\right) = \left(\frac{k_f}{f_c}\right)x(t) \qquad \therefore \langle d(t) \rangle \propto x(t)$$

5.9 FM BROADCASTING

FM radio broadcasting for speech and music makes use of the 88 MHz –108 MHz band. The peak frequency deviation is to be 75 kHz, audio frequencies up to 15 kHz are handled and the bandwidth is 200 kHz, i.e., two adjacent carriers are to have a separation of 200 kHz. The transmitters employ pre-emphasis – i.e., boost the high frequency components of the message or baseband signal in order to improve the signal to noise ratio at the destination. FM broadcast receivers are of the superheterodyne type, the intermediate frequency being 10.7 MHz.

Irrespective of the carrier frequency of the signal to which the receiver is tuned, owing to the gang-tuning of the RF amplifier and the local oscillator, the carrier frequency at the mixer output is always the intermediate frequency of 10.7 MHz. Since it operates at a constant frequency, the IF amplifier is designed to give a large gain. Although the transmitted FM signal has constant amplitude, it gets corrupted by the additive noise in the channel and the received signal has small random variations in its amplitude. These are removed in the receiver by the limiter stage. A balanced discriminator extracts the message or the baseband signal from the FM signal at the output of the limiter. In monophonic receivers, the discriminator output will be just the audio. This is amplified, de-emphasized for removing the extra boost given to the higher audio frequencies before transmission, low pass filtered for removing the out-of-band noise, if any, and then finally fed to the loudspeaker.

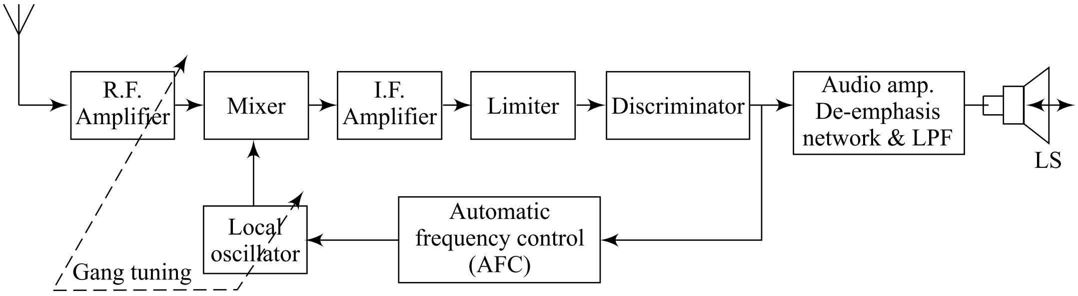


Fig. 5.37 *Block diagram of superheterodyne FM broadcast receiver*

5.9.1 Capture Effect

Suppose there is an interfering signal having a frequency *close* to the desired signal to which we have tuned the receiver, and that the interfering signal is quite weak compared to the desired signal. If it were to be AM, in the receiver output, we will be getting not only the desired signal but also the interfering one, the latter as a sort of weak background noise. But, in the case of FM, the situation will be totally different – only the relatively strong desired signal will be received and the weak interfering signal will be suppressed to a very large extent. *This phenomenon is called 'Capture Effect', since the stronger signal virtually captures the receiver.*

This phenomenon may be explained as follows. Let the desired signal have a carrier of peak amplitude A and frequency ω_c. Let the interfering signal have a frequency $(\omega_c+\Delta\omega)$ and a peak amplitude B. For our analysis here, the modulations of the desired and interfering signals may be totally ignored, as they do not play any part. The received signal may be written as

$$\begin{aligned} r(t) &= A\cos\omega_c t + B\cos(\omega_c+\Delta\omega)t \\ &= (A + B\cos\Delta\omega t)\cos\omega_c t - (B\sin\Delta\omega t)\sin\omega_c t \end{aligned}$$

Hence, $r(t)$ may be written as

$$r(t) = R(t)\cos[\omega_c t + \theta(t)]$$

where

$$R(t) = [(A + B\cos\Delta\omega t)^2 + (B\sin\Delta\omega t)^2]^{1/2}$$

and

$$\theta(t) = \tan^{-1}\frac{B\sin\Delta\omega t}{A + B\cos\Delta\omega t}$$

Neglecting $B\cos\Delta\omega t$ in comparison with A as $A >> B$,

$$\theta(t) \cong \tan^{-1}\frac{B\sin\Delta\omega t}{A}$$

In the case of FM, the amplitude $R(t)$ of the received signal $r(t)$ is of no consequences. q(t), the phase deviation of the desired carrier signal, caused by the interfering signal, is however, important, as it produces an output in the receiver. But

$$\theta(t) \cong \tan^{-1}\left[\left(\frac{B}{A}\right)\sin\Delta\omega t\right] \approx 0 \quad \text{if } A >> B$$

So, the stronger the desired signal, relative to the interfering signal, the better is the suppression of the interfering signal. It may be noted here that the interfering signal need not be only an undesired carrier or modulated signal. It may be made up of just *noise* frequency components closed to the desired carrier frequency. Thus, capture effect suppresses noise too.

5.10 FM TRANSMITTERS AND RECEIVERS

FM broadcasting has been assigned the 88 MHz–108 MHz frequency band. All transmitting stations are to ensure that the unmodulated carrier frequency is within ±2 kHz of the assigned carrier frequency. With a maximum frequency deviation of ±75 kHz and a maximum audio frequency of 15 kHz, the signal occupies a bandwidth of 180 kHz (Carson's rule). Further, a guard band of 20 kHz is provided to ensure interference-free communication in the service area. Since the FM band of 88 MHz–108 MHz is in the VHF band, it is on line-of-sight propagation that FM broadcasting depends. The primary service area is determined largely by (i) the effective radiated power, and (ii) the height of the antenna; and may be up to about 80 km.

FM broadcast transmitters handle audio frequencies up to 15 kHz. Because of this and the relative immunity enjoyed by FM with regard to additive noise of the channel, FM broadcast transmitters are particularly useful for transmission of high quality music. These transmitters use carrier frequencies of 88 MHz–108 MHz in the VHF band, and make use of powers of the order of 100 kW.

As discussed earlier, a WBFM signal may be generated either by the indirect method, or the direct method.

5.10.1 FM Transmitter Based on Indirect Method of Generation of WBFM (Armstrong Method)

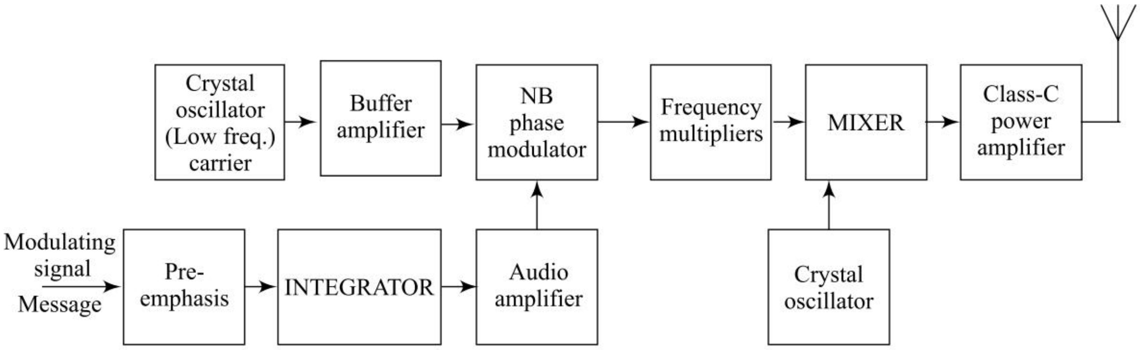


Fig. 5.38 *FM transmitter based on the indirect method*

As explained earlier, initially a low frequency carrier is used. A narrowband phase modulator of the type shown in Fig. 5.10 is used. The signal is subjected to pre-emphasis, integrated, amplified and used as the signal for modulating the low frequency carrier. Frequency multiplier chain and mixer are used to obtain the required values of final carrier frequency and peak deviation. A chain of class-C amplifiers is used to raise the power of the modulated signal to the required value.

5.10.2 FM Transmitter Based on Direct Method of Generation

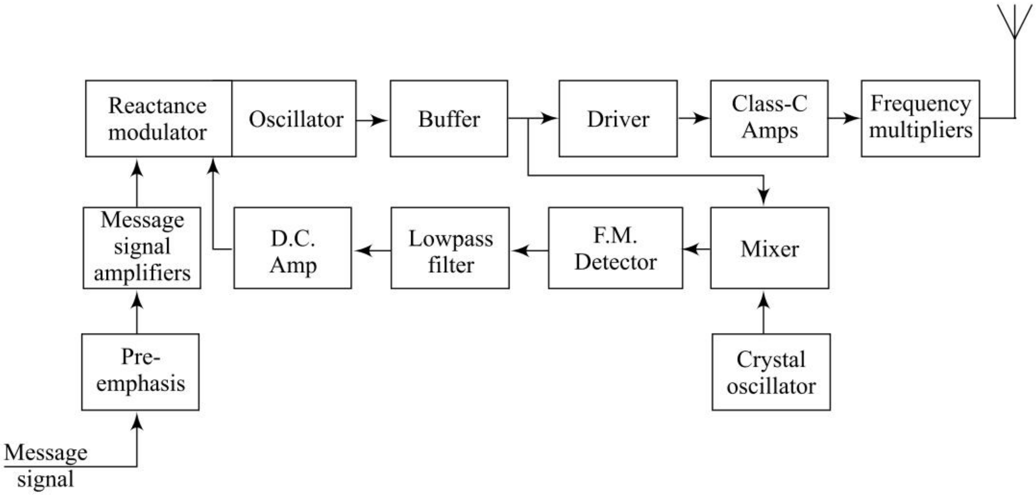


Fig. 5.39 *Block diagram of an FM transmitter using direct method of generation of WBFM*

As mentioned in Section 5.6, the direct method of generation of WBFM has the disadvantage that the unmodulated carrier signal is not generated by a crystal oscillator and therefore, is not very stable. Hence, a frequency stabilization circuit is a must. One such arrangement is shown in Fig. 6.26. In this, the modulated signal is taken from the output of the buffer and fed to a mixer to which, the output of a crystal oscillator also is given. If the transmitter is to operate with a carrier frequency of f_c, the crystal oscillator frequency f_0 is so chosen that $f_d \triangleq (f_0 - f_c)$ is reasonably small. This difference frequency signal from the mixer is applied to a balanced discriminator which is so adjusted that it gives zero output when its input signal has a frequency exactly equal to f_d. The low pass filter following this FM discriminator has a very low cutoff frequency and removes the modulating signal component. The output of this filter will be zero if the carrier frequency is exactly f_c and will be a dc voltage of appropriate sign depending on whether the carrier frequency is above or below the correct value f_c. This dc voltage is used to modify the bias applied to the reactance modulator in such a way as to bring the oscillator unmodulated carrier frequency to the correct value.

5.10.3 FM Receivers

Just like the AM broadcast receivers, FM broadcast receivers are also superheterodyne receivers. Their tuning range, i.e., the standard VHF FM broadcast band, is 88 MHz–108 MHz. The standard value of the intermediate frequency for these receivers is 10.7 MHz. Figure 5.40 shows a block schematic diagram of a typical mono-aural FM broadcast receiver.

The tuned circuits of the RF stage and the local oscillator are ganged and so when the RF stage tuning is varied from one end to the other, things are so arranged that the local oscillator frequency varies form 98.7 MHz to 118.7 MHz so that when we take the difference frequency at the output of the mixer, the FM signal obtained has always a carrier frequency of 10.7 MHz, i.e., the intermediate frequency, irrespective of the frequency of the station to which the receiver is tuned.

The RF amplifier stage is generally a double-tuned low noise dual-gate MOSFET cascade amplifier with high values of input and output impedances. *In FM receivers, image rejection does not pose a problem*. This is because, the image signal, which is 2 × IF Hz away from the frequency of the desired signal, is always outside the tuning range of the receiver irrespective of whether the receiver is tuned to a station near the lower-end, or the upper-end of the tuning range (Note that $88+2 \times 10.7 = 109.4$ MHz and $108–2 \times 10.7 = 86.6$ MHz, both of which are outside the tuning range).

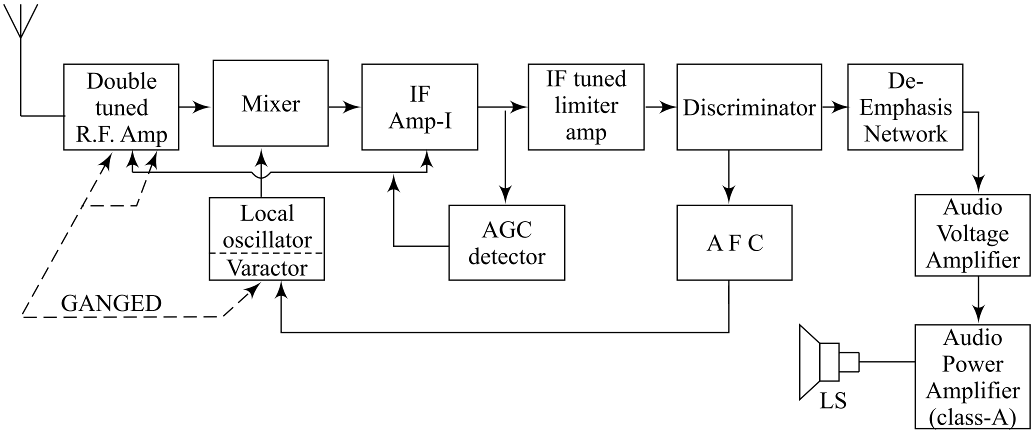


Fig. 5.40 *Block diagram of a superheterodyne FM broadcast receiver*

Generally two or three high gain IF amplifier stages are employed and one of them is used as an amplitude limiter to remove the additive noise which causes amplitude variations. These IF amplifiers are designed to have a bandpass characteristic with a flat response in the 180 kHz passband centered on 10.7 MHz.

Amplitude limiting action too may be obtained in an IF stage either by including back-to-back connected diodes in the input tuned circuit of the IF amplifier, or by designing the IF stage to be driven to saturation and cut off, depending upon whether low-level or high-level limiting is desired.

The discriminator may be a dual-slope discriminator or a ratio detector – its main function being to convert frequency variations of its input signal into corresponding amplitude variations, with the output voltage remaining at zero volts when the input signal frequency is exactly equal to the IF.

As stated earlier, pre-emphasis AND de-emphasis are used in all FM communication systems in order to ensure a good SNR at the destination. The message signal is deliberately distorted at the transmitter before using it for modulation, by passing it through a pre-emphasis network, which boosts up the high frequency components. The post-detection noise power spectrum increases as the square of the frequency, as we will be seeing in Chapter 7 when we discuss the noise performance of FM systems. To remove the distortion introduced by the pre-emphasis network, the output of the discriminator in the receiver is passed through a de-emphasis network which de-emphasizes the high frequency components so as to restore the original relative amplitudes of the various frequency components of the message signal. In that process of de-emphasizing, while the message spectrum is restored to its original shape, the high frequency noise components at the output of the discriminator get reduced and so the SNR at the destination is improved.

The audio voltage and power amplifiers then raise the power level of the audio signal so that it can actuate the loudspeaker. As FM handles audio up to 15 kHz, and so is mostly used for high quality music broadcasting, these audio amplifiers should have flat frequency response form very low audio frequencies up to 15 kHz so as not to introduce any distortion. The audio power amplifier must, of course, be a class-A amplifier.

Generally, AFC is provided to keep the frequency of the local oscillator at the value that produces the correct intermediate frequency. If the average value of the intermediate frequency differs from the center frequency of the dual-slope discriminator then, a dc voltage will be developed at the output of the discriminator. The polarity of this dc voltage will depend on the direction of deviation of the IF with respect to the center frequency of the discriminator. This dc voltage is extracted from the discriminator and is applied to the varactor diode across the tank circuit of the local oscillator in such a way as to change the local oscillator frequency in the right direction so that it gives the correct value of intermediate frequency. It is thus ensured that slight frequency drifts of the local oscillator do not cause any deterioration of the performance of the receiver.

5.10.4 FM Stereo Broadcasting

In monophonic transmission of music, the output from only one microphone is used. But in stereophonic transmission outputs from two different microphones, kept at different locations on the stage, are used for transmission. We call the outputs from the two microphones as message signals $x_L(t)$, the left message signal, and $x_R(t)$, the right message signal, and each of these occupies a bandwidth of 15 kHz. In an FM stereo transmitter, using the $x_L(t)$ and $x_R(t)$, we first produce the sum signal $[x_L(t) + x_R(t)]$ and the difference signal $[x_L(t) - x_R(t)]$, as shown in Fig. 5.41. The sum signal is passed through the pre-emphasis network and then without any further processing, is taken to an adder where a pilot tone of 19 kHz is added to it. On the other hand, the difference signal, after being passed through the pre-emphasis network, is used for DSB-SC modulating a 38 kHz carrier obtained by doubling the 19 kHz pilot carrier. The DSB-SC signal so generated is added to the sum signal and the pilot carrier. The output of this adder, consisting of the sum signal, the pilot carrier and the DSB-SC signal, is used as the baseband signal for frequency modulating the final carrier used for transmission.

From the foregoing, it is clear that functionally, the receiver should first recover the baseband signal (whose spectrum is shown in Fig. 5.42). So up to the discriminator stage, there is no difference between a stereophonic FM receiver and a monophonic FM receiver. The above spectrum clearly indicates the various functions that the stereo FM receiver should perform to get $x_L(t)$ and $x_R(t)$ separately. All these are shown in the block diagram of the receiver given in Fig. 5.43.

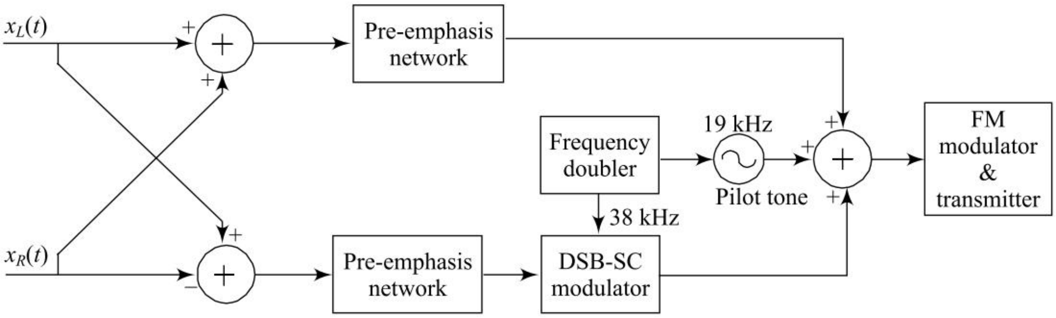


Fig. 5.41 *FM stereo transmitter*

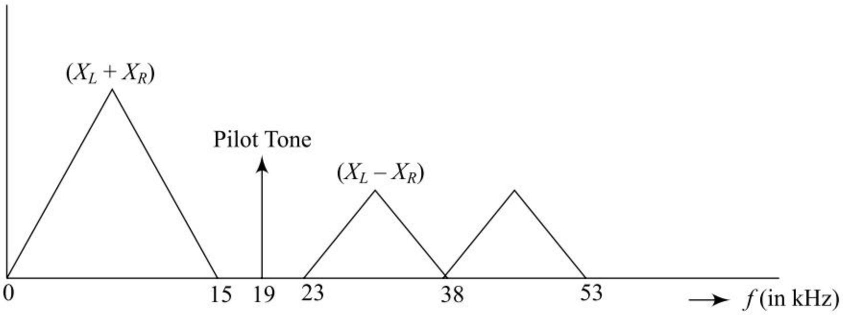


Fig. 5.42 *One-sided spectrum of the baseband signal used for frequency modulating the final carrier*

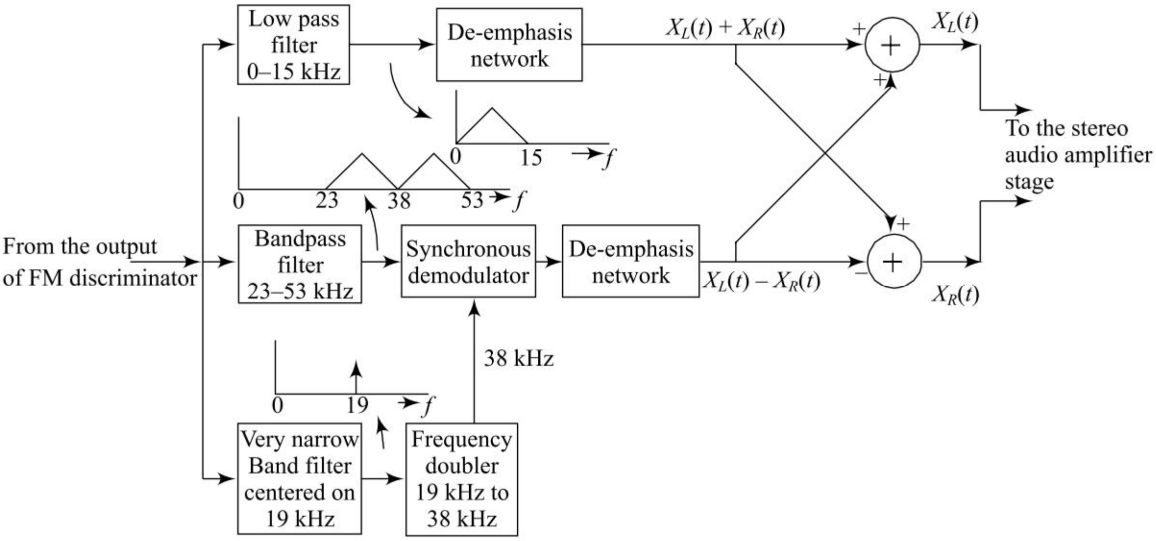


Fig. 5.43 *FM stereo receiver after the discriminator stage*

The output of the discriminator is the baseband signal whose spectrum is as in Fig. 5.42. This is fed simultaneously to a low pass filter, a bandpass filter and a narrowband filter centered on 19 kHz, to separate out the three component signals comprising the baseband – the sum signal, the DSB-SC signal containing the difference signal and then the pilot tone of 19 kHz frequency. The sum signal, after de-emphasis, serves as the audio signal for the monophonic FM receiver whose post-discriminator bandwidth is only 15 kHz. Thus, a monophonic FM receiver also can receive the audio from a stereophonic FM transmitter and this audio signal is the sum signal. The stereophonic receiver, however, makes use of the sum and difference signals to obtain $x_L(t)$ and $x_R(t)$ separately as shown in Fig. 5.43. These are then fed to the (stereo) audio amplifier and they finally drive the dual loudspeakers.

MATLAB Example 5.1 (Frequency Modulation)

Message signal is sin(2*pi*10*t), carrier signal cos (2*pi*200*t).

Type of modulation: FM, Frequency deviation constant kf = 50

Using Matlab, do the following:

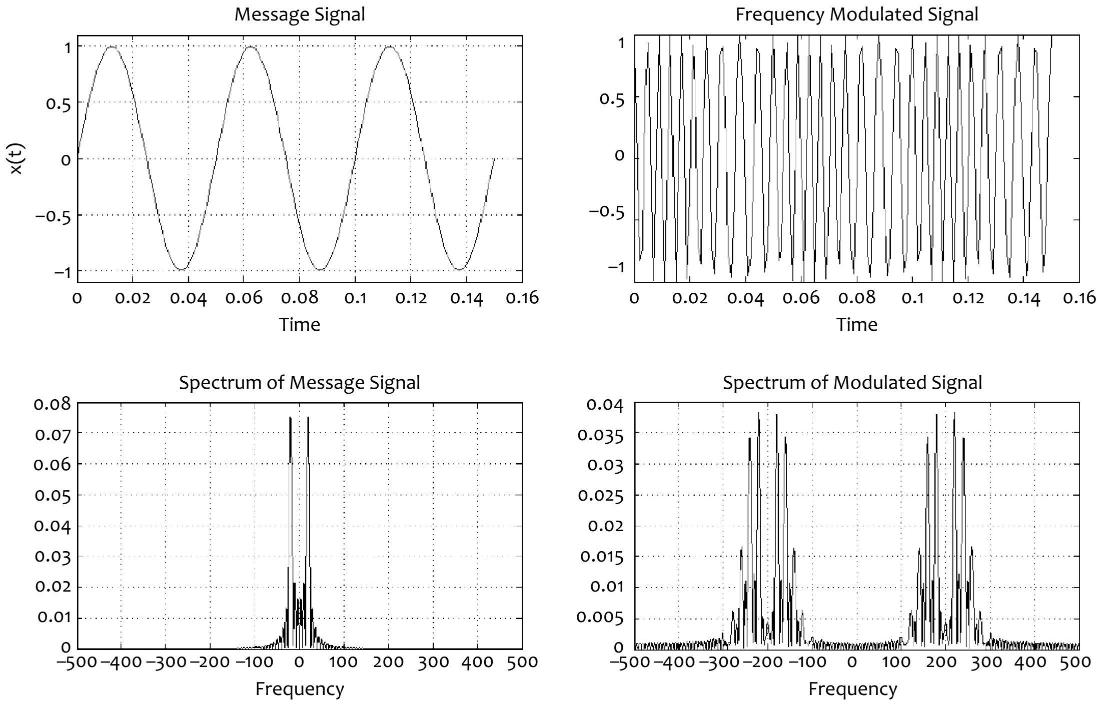
(a) Plot the message signal.
(b) Plot the modulated signal.
(c) Determine and plot the spectrum of the message signal.
(d) Plot the spectrum of the frequency-modulated signal.

MATLAB Program

```
%
% This program calls the function 'fftseq' to solve the problem
%
t0 = 0.15;      % signal duration
ts = 0.001;     % sampling interval
f = 10 hz;
fc = 200;       % carrier frequency
kf = 50;        % modulation index
fs = 1/ts;      % sampling frequency
t = [0:ts:t0];  % Time vector
df = 0.25       % frequency resolution
%
% Message signal
%
m = sin(2*pi*2*f*t)
% plot of the message signal
subplot (2,2,1)
plot(t,m)
grid on
ylim([-1.1 1.1]);
xlabel ('time');
ylabel ('x(t)');
title ('Message Signal');
% integral of m
int_m(1) = 0;
for i = 1:length(t)-1
  int_m(i+1) = int_m(i)+m(i)*ts;
end
%
% Finding the Fourier transform of the m signal
%
[M,m,df1] = fftseq(m,ts,df);          % Fourier Transform
M = M/fs;                              % Scaling
f = [0:df1:df1*(length(m)-1)] - fs/2; % Frequency Vector
%
% Generation of modulated signal
```

```
%
u = cos(2*pi*fc*t+2*pi*kf*int_m)      % modulated signal
subplot (2,2,2)
plot(t,u(1:length(t)));
xlabel ('time');
title ('Modulated Signal');
%
% Finding the Fourier transform of the modulated signal u
%
[U,u,df1] = fftseq(u,ts,df);          % modulated signal
U = U/fs                              % scaling
% Plots of magnitute of message and modulated signal in frequency domain
length(f)
length (M)
length (U)
subplot (2,2,3)
plot(f,abs(fftshift(M)));
xlabel ('frequency');
title ('Spectrum of Message Signal')
grid on
subplot (2,2,4)
plot (f,abs(fftshift(U)));
xlabel ('Frequency');
title ('Spectrum of Modulated Signal')
grid on
```

Results



Message Signal

Frequency Modulated Signal

Spectrum of Message Signal

Spectrum of Modulated Signal

Fig. 5.44

Summary

- Frequency modulation and phase modulation are together known as angle-modulation.
- FM and PM both change the phase angle, but in different ways.
- PM: $x_c(t) = A_c \cos[\omega_c t + k_p x(t)]$ where k_p is called the phase deviation constant.
- FM: $x_c(t) = A_c \cos\left[\omega_c t + k_f 2\pi \int_0^t x(\alpha)d\alpha\right]$ where k_f is called the frequency deviation constant.
- When $x(t)$, the modulating signal = $A_m \cos 2\pi f_m t$,
  - (a) $\underset{\text{PM}}{x_c(t)} = A_c \cos[2\pi f_c t + k_p A_m \cos 2\pi f_m t]$
  - (b) $\underset{\text{FM}}{x_c(t)} = A_c \cos\left[2\pi f_c t + \left(\frac{k_f A_m}{f_m}\right)\sin 2\pi f_m t\right]$
  - (c) β_p = Modulation index for PM $\underline{\Delta}$ $k_p A_m$
  - (d) β_f = Modulation index for FM $\underline{\Delta}$ $\frac{k_f A_m}{f_m}$
- Deviation Ratio $\underline{\Delta}\left[\frac{\text{Peak frequency deviation}}{\text{Modulating signal frequency}}\right]$
- (a) If $\phi(t)$, the phase deviation is less than or equal to 0.2 radian, it is called Narrowband Angle Modulation.
  - (b) Bandwidth of an NB angle-modulated signal = $2f_m$.
- (a) If $x(t)$ is integrated and fed as the modulating signal to a phase modulator, an FM signal is obtained.
  - (b) If $x(t)$ is differentiated and fed as the modulating signal to a Frequency modulator, a PM signal is obtained.
- (a) An angle modulated signal has, theoretically, an infinite bandwidth, even for a single-tone modulating signal.
  - (b) The bandwidth within which at least 98% of the average power of an angle modulated signal is contained, is called the 'Effective Bandwidth' of the angle modulated signal.
- (a) The average power of an angle-modulated signal is $P_{\text{av}} = \frac{1}{2}A_c^2$ where A_c is peak amplitude of the carrier.
  - (b) Carson's Rule for effective bandwidth for single-tone modulation.

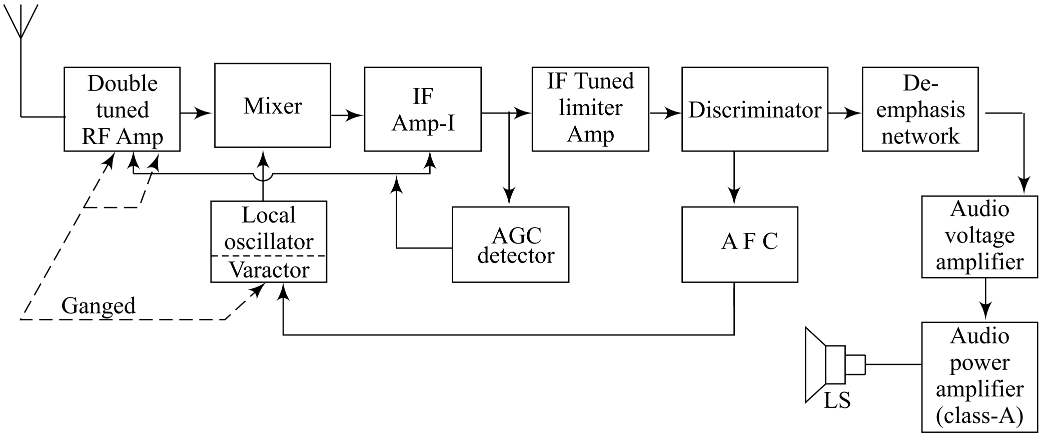
$$B_T = \begin{cases} 2(k_p A_m + 1)f_m & \text{for PM} \\ 2(k_f A_m + 1)f & \text{for FM} \end{cases}$$

  - (c) Carson's Rule for a general modulating signal:

$$B_T = 2(\lfloor \beta \rfloor + 1)W, \text{ where } \beta \underline{\Delta} \begin{cases} k_p \max|x(t)| & \text{for PM} \\ \frac{k_f \max|x(t)|}{W} & \text{for FM} \end{cases}$$

- (a) B_T of an FM signal is practically unaffected by an increase in the modulating signal frequency.
  - (b) B_T of a PM signal increases almost linearly with the increase of the modulating signal frequency.
- WBFM may be generated either by the indirect (or Armstrong method), or by the direct method.
- (a) Indirect method gives a WBFM signal with good frequency stability, but needs a number of frequency multipliers.
  - (b) Direct method needs AFC unit for stabilizing the frequency but does not need frequency multipliers.
- An FM signal may be demodulated by using a Foster–Seeley detector, a Ratio detector, a Quadrature detector, a Zero-crossing detector, or a Phase-Locked Loop (PLL).
- WBFM is used for high quality music broadcasting in the 88 MHz to 108 MHz band, using a maximum frequency deviation of 75 kHz, a bandwidth of about 180 kHz and a carrier separation of 200 kHz. Typical IF for FM is 10.7 MHz.

- A superheterodyne receiver for FM has a limiter stage after the IF amplifier stage to remove the small random variations in the amplitude of the FM signal, caused by the additive noise.
- *FM broadcast transmitters:* Handle audio frequencies up to 15 kHz. Used mostly for high quality speech and music. Operate in the VHF frequency band from 88 MHz to 108 MHz. Depend on line-of-sight propagation and so service area is limited to about 40–80 km. Carrier frequency stability of ±2 kHz needed. Maximum frequency deviation is ±75 kHz. Adjacent carrier separation is 200 kHz. Power of the order of 100 kW are used.
- *FM transmitter based on indirect method:* Refer to Fig. 5.38.
- *FM transmitter based on direct method:* Refer to Fig. 5.39. Note that it is imperative to make use of a carrier frequency stabilization circuit for FM transmitters based on the direct method.
- FM broadcast receiver block diagram:



- *Limiting:* In an FM receiver, the amplitude variations of the received FM signal, caused by noise, etc., are removed by using amplitude limiters. Amplitude limiting action may be obtained in an IF stage by including back-to-back connected diodes in the input tuned circuit of the IF amplifier.
- *Pre-emphasis and de-emphasis:* All FM communication systems use pre-emphasis at the transmitter and de-emphasis at the receiver, to improve SNR at the destination. Pre-emphasis consists of boosting the high frequency components of the message signal before modulation and de-emphasis attenuates the high-frequency components of the message signal obtained in the receiver at the output of the discriminator, so that the distortion of the message signal, introduced by the pre-emphasis, is removed.

References and Suggested Reading

1. Haykin, Simon, *Communication Systems*, John Wiley & Sons, New York, 4th Edition, 2001.
2. Taub, and Schilling, *Principles of Communication Systems*, Tata McGraw-Hill, New Delhi, 1995.
3. Carlson, Bruce *et. al.*, *Communication Systems*, McGraw-Hill International Edition, 4th Edition, 2002.
4. Proakis, J.G., and M. Salehi, *Communication Systems Engineering*, 2nd Edition, Prentice-Hall of India, New Delhi, 2006.
5. Clarke, K.K., and D.T. Hess, *Communication Circuits: Analysis and Design*, Addison-Wesley, Mass, 1971.
6. Couch, L.W. II, *Digital and Analog Communication Systems*, 4th Edition, Macmillan, New York, 1993.
7. Ziemer, R.E., and W.H, Tranter, *Principles of Communications: Systems, Modulation and Noise*, Houghton, Miffin, Boston, 1990.
8. Schoenbeck, Robert J., *Electronic Communications – Modulation and Transmission*, 2nd Edition, Prentice Hall Career & Technology, Englewood Cliffs, New Jersey, USA, 1992.
9. Roddy, Dennis, and John Coolen, *Electronic Communications*, 3rd Edition (Eastern Economy Edition), Prentice-Hall of India Pvt. Ltd., New Delhi, 1994.
10. Kennedy, and Davis, *Electronic Communication Systems*, 4th Edition, McGraw-Hill International Edition, 1992.

Review Questions

1. Define frequency modulation.
2. Define phase modulation.
3. Derive an expression for the time-domain representation of a frequency modulated signal.
4. Explain how you would use a phase modulator for obtaining a frequency modulated signal.
5. Sketch the waveform of a phase modulated signal assuming $k_p = \pi/2$ and $x(t) = u(t)$.
6. Define the term, 'modulation index' for FM in the case of single-tone modulation and for a general modulating signal.
7. By deriving the necessary expressions, show that a narrowband angle-modulated signal and an AM signal have similar forms. (Assuming single-tone modulation). Draw the phasor diagrams for both the cases.
8. By drawing the block schematic diagram, show how a narrowband angle-modulated signal may be generated.
9. Assuming single-tone modulation, derive an expression for the spectrum of an angle-modulated signal.
10. Making use of the Bessel function tables, sketch the spectrum of an angle-modulated signal for f_m = 5 kHz and β = smallest value of β for which the carrier component vanishes. Sketch the two-sided spectrum up to the 3rd side-frequency component.
11. Using the expression for the spectrum of an angle-modulated signal for single-tone modulation by a tone of frequency f_m, show that the transmission bandwidth of the modulated signal is given by $B_T = 2(\beta + 1)f_m$, where β is the modulation index.
12. Define 'Effective Bandwidth' of an angle-modulated signal.
13. Explain how the transmission bandwidth changes with respect to changes in the modulating signal frequency in the case of PM and FM.
14. With the help of a neat block schematic diagram, explain the indirect method of generation of WBFM signals.
15. Explain the reactance modulator method of generation of WBFM. Why is it necessary to use AFC in this method of generation?
16. Explain the working of a Foster–Seeley detector for FM.
17. How is a phase-locked loop (PLL) useful in detecting FM signals?
18. With a neat block diagram, briefly explain the principle of working of a superheterodyne FM broadcast receiver.
19. Why is a limiter stage used in the superheterodyne FM broadcast receiver? Explain the principle of working of the limiter. Sketch the transfer function of a hard limiter.
20. Explain the working of an FM broadcast transmitter employing the direct method of generation of WBFM by drawing the block diagram. In particular, explain how the drift of the carrier is countered.
21. Explain the working of a FM broadcast transmitter employing the indirect method of generation of WBFM by drawing the block diagram.
22. Draw the block schematic diagram of a FM broadcast receiver, and explain its working.
23. With the help of block schematic diagrams and sketches of the spectra of appropriate signals, explain the principle of stereo FM transmission and reception.

Problems

1. Sketch the waveforms of the resulting modulated signal when a high frequency sinusoidal carrier signal is modulated by the modulating signal shown in Fig. P5.1, if the modulation is (a) frequency modulation, and (b) phase modulation.

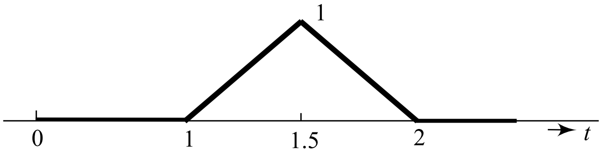


Fig. P5.1

2. An FM signal is of the form $x_c(t) = 75\cos[2\pi \times 5 \times 10^6 t + 6\sin 200\pi t]$
   - (a) What is the modulating signal frequency?
   - (b) What is the carrier frequency?
   - (c) Determine the peak frequency deviation
   - (d) Determine the deviation ratio
   - (e) Determine the modulation index.
   - (f) Determine the average power of this FM signal.
   - (g) What is the (effective) bandwidth of this FM signal?
3. A message signal, $x(t) = 100\,\text{sinc}\,2000t$ frequency modulates a carrier signal $c(t) = 200\cos 2\pi \times 10^8 t$, with a modulation index of 5.
   - (a) Write down an expression for $x_c(t)$, the modulated signal.
   - (b) What is the peak frequency deviation?
   - (c) What is the average power of the modulated signal?
   - (d) What is the bandwidth of this modulated signal?
4. The carrier signal $c(t) = 200\cos 2\pi \times 10^8 t$ is phase modulated by the message signal, $x(t) = 2\cos 2\pi \times 10^3 t$, the peak phase deviation being $\pi/5$.
   - (a) What is the bandwidth of this PM signal?
   - (b) Sketch the magnitude spectrum of the modulated signal up to frequencies lying within the bandwidth calculated in (a).
5. $x_1(t)$ and $x_2(t)$ are two modulating or message signals and $x_1(t) + x_2(t) = x_3(t)$. When $x_1(t)$ modulates the carrier $c(t)$, the modulated signal is $x_{1c}(t) \times x_{2c}(t)$ and $x_{3c}(t)$ are similarly defined, using the same carrier.
   - (a) When the modulation is AM, show that the modulation is linear in the sense that it obeys superposition principle, by proving that $x_{3c}(t) = x_{1c}(t) + x_{2c}(t)$
   - (b) When the modulation is angle modulation, show that the modulation is not linear, i.e., that in this case, $x_{3c}(t) \neq x_{1c}(t) + x_{2c}(t)$
6. An NBFM signal with a carrier frequency of 200 kHz and peak frequency deviation of 21.3 Hz is to be used to produce a WBFM signal of carrier frequency about 100 MHz and peak frequency deviation of 75 kHz, using frequency multipliers, a mixer, etc. as shown in the Fig. P5.6. Determine N_1, N_2 and f_c to achieve the desired result. Note that the multipliers should comprise either doublers or triplers, or a combination of these two.

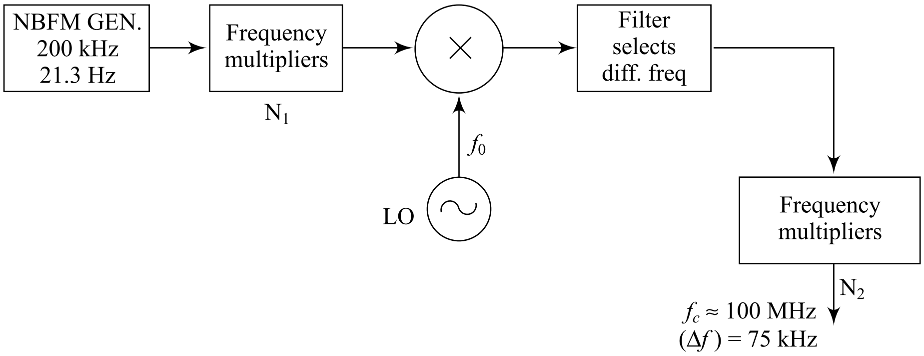


Fig. P5.6

7. An FM signal is represented by

$$x_c(t) = 50\cos\left[2\pi f_c t + 50\int_o^t x(\tau)\,d\tau\right]$$

where the modulating signal $x(t)$ is as shown in Fig. P5.7.

(a) Write down the expression for the instantaneous frequency and sketch it.
(b) What is the value of the deviation constant?
(c) What is the peak frequency deviation?

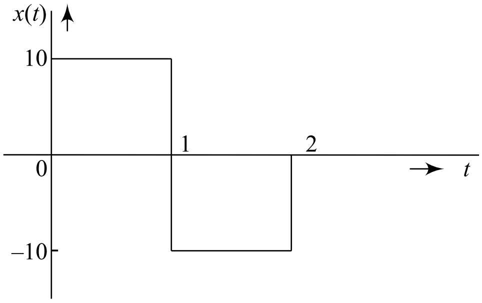


Fig. P5.7

8. An angle-modulated signal is given to be

$$x_c(t) = 75\cos[2\pi\times10^7 t + 6\sin 2\pi\times 2\times10^3 t]$$

(a) If it is an FM signal, what are its frequency deviation constant, modulation index β_f and transmission bandwidth?
(b) If it is a phase-modulated signal, what are its phase deviation constant, modulation index β_p and transmission bandwidth?
(c) For each of the above cases, determine the pertinent values when f_m, the message frequency, is increased to 4×10^3 Hz.

9. A sinusoidal carrier of 150 MHz frequency and 1 V peak amplitude is frequency modulated by a 2 kHz sinusoidal modulating signal, producing a peak frequency deviation of 10 kHz. Using the Bessel function tables, sketch the amplitude spectrum of the modulated signal up to ten side frequencies. Using Carlson's rule, determine the bandwidth of the modulated signal.

10. An NBFM signal generated with a carrier frequency of 100 kHz, and a frequency deviation of 30 Hz, is applied to a frequency multiplier chain consisting of five doublers and then a frequency multiplier chain consisting of three triplers. Assuming the modulating signal to be a 2 kHz tone, determine the frequency deviation and the modulation index at the end of the doubler chain and at the end of the tripler chain.

11. Explain how a square law device may be used for increasing the frequency deviation of an FM signal.

12. Figure P5.12 shows an arrangement used frequently as an FM demodulator at microwave frequencies. The delay line produces a delay of T sec that corresponds to $\pi/2$ radians phase shift at the carrier frequency f_c. The FM signal $x_c(t)$ may be taken to be

$$x_c(t) = A_c\cos[2\pi f_c t + \beta_f\sin(2\pi f_m t)];\quad \beta_f < 1$$

Assuming $T < \dfrac{1}{f_m}$ so that $\cos 2\pi f_m T \approx 1$, show that the output signal is proportional to the modulating signal.

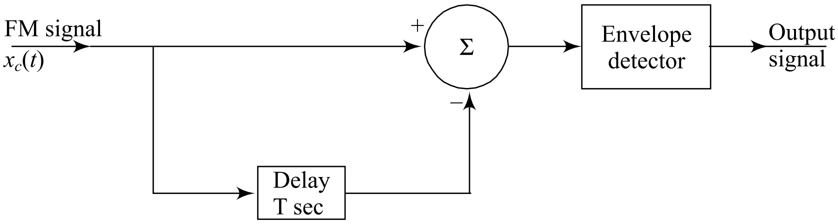


Fig. P5.12

13. An FM transmitter using the direct method of generation of WBFM, is using a varactor diode modulator which produces a frequency deviation of 2.5 kHz per volt. The maximum deviation produced by the modulator is 360 Hz. The modulator is followed by a buffer and a tripler, doubler and tripler for frequency multiplication.
(a) Can this transmitter produce a 6 kHz peak deviation at the output?
(b) If the final carrier frequency is to be 180 MHz, what should be the oscillator frequency?
(c) What is the audio voltage to be applied to the varactor to obtain the full deviation at the output?

Multiple-Choice Questions

1. For fixed values of the phase deviation constant and the amplitude of the single-tone modulating signal, the modulation index for phase modulation
   (a) increases with modulating signal frequency f_m
   (b) decreases with increasing values of f_m
   (c) is not dependent on f_m
   (d) increases with f_m up to a certain value of f_m and then decreases.

2. In frequency modulation by a single-tone modulating signal, the frequency deviation constant and the modulating signal frequency are both doubled. The modulation index will be
   (a) quadrupled (b) unchanged (c) doubled (d) 0.25 times the previous value

3. To produce frequency modulation using a phase modulator
   (a) the message signal must be integrated and then used for modulation
   (b) the message signal must be differentiated and then used for modulation
   (c) the phase modulated signal must be integrated
   (d) the phase modulated signal must be differentiated

4. If phase modulation is to be produced using a frequency modulator
   (a) the message signal must be integrated and then used for modulation
   (b) the message signal must be differentiated and then used for modulation
   (c) the frequency modulated signal must be integrated
   (d) the frequency modulated signal must be differentiated

5. In phase modulation by a single-tone modulating signal, the phase deviation constant is doubled and the modulating signal frequency is halved. The modulation index is
   (a) halved (b) quadrupled (c) doubled (d) unchanged

6. $x(t)$, a message signal, angle-modulates a carrier $A_c \cos \omega_c t$. The modulated signal is $A_c \cos[\omega_c t + \phi(t)]$. If it is phase modulation, $\phi(t)$ is
   (a) $2\pi k_p \int_0^t x(\alpha)\,d\alpha$ (b) $\frac{2\pi k_p}{W}$ (c) $2\pi k_p x(t)$ (d) $k_p x(t)$

Note *k_p is phase deviation constant.*

7. For a frequency modulated signal, the modulation index is doubled. The average power of the modulated signal is
   (a) quadrupled (b) doubled (c) unaltered (d) None of these

8. For a WBFM signal, when the frequency of the single-tone modulating signal is doubled, the transmission bandwidth
   (a) doubles
   (b) does not change
   (c) increases slightly but does not become double
   (d) reduces considerably since the deviation ratio is halved

9. In commercial FM broadcasting, the audio frequency range handled is only up to
   (a) 15 kHz (b) 5 kHz (c) 3.5 kHz (d) 10.7 kHz

10. For wideband phase modulation, when the frequency of the single-tone modulating signal is doubled, the transmission bandwidth
   (a) does not change at all (b) doubles
   (c) increases slightly but does not double (d) reduces slightly

11. The transmission bandwidth required for commercial FM broadcasting is
   (a) 75 kHz (b) 10 kHz (c) 200 kHz (d) 220 kHz

12. The standard intermediate frequency used in the superheterodyne FM receiver is
   (a) 88 MHz (b) 455 MHz (c) 15 MHz (d) 10.7 MHz

13. A narrowband FM signal has a carrier frequency of f_c and a frequency deviation of (Δf). The signal is passed through a frequency doubler. The new carrier frequency and deviation are

(a) $(2f_c, \Delta f)$ (b) $(2f_c, 2\Delta f)$ (c) $(2f_c, 1/2\,\Delta f)$ (d) $(f_c, 2\Delta f)$

14. An FM signal having a carrier frequency of 12 MHz and a frequency deviation of 3.2 kHz is given to a mixer along with a local oscillator signal of frequency 10 MHz. The filter following the mixer allows only the difference frequency. The new values of carrier frequency and deviation are

(a) (2 MHz, 3.2 kHz) (b) (2 MHz, 0.53 kHz) (c) (2 MHz, 2.67 kHz) (d) (2 MHz, 0.64 kHz)

15. A narrowband FM signal is generated using a phase modulator. The maximum deviation at the output of a phase modulator is about

(a) ±250 Hz (b) ±1 kHz (c) ± 1 MHz (d) ± 25 Hz

16. The type of reactance that a reactance modulator presents to the tank circuit of the oscillator can be

(a) only capacitive (b) only inductive
(c) either capacitive or inductive (d) neither capacitive nor inductive

17. A reactance modulator is presenting capacitive reactance to the oscillator. To make it offer inductive reactance, we have to

(a) interchange the positions of *C* and *R*
(b) replace *C* by *L*
(c) making $R >> (1/\omega_c)$
(d) Reactance modulator cannot be made to present an inductive reactance

18. In a stereo FM transmitter, the difference signal $[x_L(t) - x_R(t)]$ modulates a 38 kHz tone. The type of modulation employed is

(a) AM (b) DSB-SC (c) SSB-SC (d) FM

19. The Foster–Seeley discriminator responds to the input FM signal's

(a) amplitude variations only (b) amplitude as well as frequency variations
(c) frequency variations only (d) variations neither in amplitude nor in frequency

20. The ratio detector responds to the input FM signal's variations in

(a) amplitude only (b) frequency only
(c) both amplitude and frequency (d) neither amplitude nor frequency

21. The noise figure of a superheterodyne receiver is mostly controlled by

(a) RF stage (b) the mixer stage (c) IF stage (d) the audio stage

22. In FM broadcasting, the peak frequency deviation and the maximum audio frequency handled, are respectively

(a) 75 kHz; 10 kHz (b) 75 kHz; 15 kHz (c) 200 kHz; 10 kHz (d) 75 kHz; 5 kHz.

Key to Multiple-Choice Questions

1. (c)	2. (b)	3. (a)	4. (b)	5. (c)	6. (d)	7. (c)	8. (c)
9. (a)	10. (b)	11. (c)	12. (d)	13. (b)	14. (a)	15. (d)	16. (c)
17. (a)	18. (b)	19. (b)	20. (b)	21. (a)	22. (b)		

PROBABILITY, RANDOM PROCESSES AND NOISE

6

> *"I didn't fail 2000 times, I learned 2000 ways how not to make a light bulb."*
>
> **Thomas Edison (1847–1931)**
> *American inventor and businessman*

Learning Objectives

After going through this chapter, students will be able to

- *thoroughly revise all the key concepts in probability and random processes,*
- *apply the results of probability and random processes part of the chapter to the study of noise as well as the study of noise performance of analog and digital communication systems,*
- *understand how noise degrades the quality of communication,*
- *list the various sources of noise and describe the characteristics of the noise generated by each of those sources,*
- *determine the thermal noise voltage across a two-terminal network of only resistors, or of resistors and reactive elements connected in some manner,*
- *calculate the noise equivalent bandwidth of a given filter, the equivalent noise resistance of amplifiers/systems, and equivalent noise figure (or noise temperature) of a number of two-port networks connected in cascade, and*
- *calculate the inphase and quadrature components of a band pass noise process given its envelope and phase angle representation.*

6.1 INTRODUCTION

Probability theory lays the foundation for a study of random processes and both of them are inextricably connected with communication engineering.

The two most important entities in the study of communication engineering are 'noise' and 'signal'. Noise is unpredictable in nature and any quantitative study of it requires modeling of it by a random process. Any useful signal also is unpredictable, in nature because if it was not so and was absolutely predictable, then the receiver could know it a priori and there would have been no need to transmit it.

When a signal passes through a channel, it suffers several changes. Some of these changes are caused by phenomena which are deterministic in nature and can therefore be eliminated. Linear and non-linear distortion and inter-symbol interference come under this category. On the other hand, phenomena like fading, etc., are essentially non-deterministic and have to be modeled as random processes.

It is not proposed, and it is also not possible, to cover these topics of probability and random processes in an exhaustive manner in this chapter. As the reader must have been exposed to these topics earlier, we propose to adopt a review-like approach. The review of probability theory will be limited to cover only those areas that are essential for understanding random processes.

The function of a communication system is to make available, at the destination, a signal originating at a distant point. This signal is called the *desired signal*. But, unfortunately, during its passage through the channel and the front-end of the receiver, this desired signal gets corrupted by a number of undesired signals. All these undesired signals, put together, constitute what is referred to as the *noise*. This noise is mostly random (i.e., unpredictable) in nature, but it can, at times have deterministic components as well – like the power supply hum and certain oscillations. These deterministic components, however, can be eliminated by proper shielding and introduction of notch filters, etc. Hence, in this chapter, we will be concentrating only on the random components constituting the noise – their types, origins, mean-squared values, and spectral contents etc.

In analog communication, the additive channel noise adds directly to the amplitude of the transmitted waveforms. As will be seen later in Chapter 7, this modifies, to some extent, even the frequency and phase of the waveform transmitted through the channel. Thus, irrespective of the type of modulation used, the waveform of the output signal at the detector of the receiver is not going to be exactly the same as the waveform of the modulating signal used at the transmitter. In digital communication too, channel noise can occasionally induce decoding errors which make the decoded sequence at the receiving end to be different, in some locations, from the baseband sequence used at the transmitter.

Thus, channel noise entering the receiver degrades the performance of communication systems. We cannot remove this noise by filtering. As the channel noise is often of very large bandwidth – much more than that of the transmitted signal, we can at the most prevent the out-of-band noise (i.e., noise outside the signal bandwidth) from entering the receiver, but it is not possible to remove the inband noise.

From the foregoing, it is evident that a study of the origin and the characteristics of various types of noise is very essential in order to take all possible steps to reduce its effect.

6.2 BASICS OF PROBABILITY

Modern probability theory is based on the following three axioms:

1. $P(A) \geq 0$ where A is any event.
2. $P(S) = 1$ where S is the 'certain' event.
3. If events A and B are mutually exclusive, i.e., if $A \cap B = \{\phi\}$, where $\{\phi\}$ is the null set, then $P(A \cup B) = P(A) + P(B)$.

In the above, $P(E)$ is to be read as 'probability of the event E'. An event itself is defined in terms of the outcomes of a random experiment, i.e., an experiment whose outcome cannot be predicted with certainty. Tossing a coin, throwing a die, and randomly picking a card out of a deck of playing cards, are all examples of random experiments. Each of these experiments has certain possible outcomes, called the elementary outcomes – '*head*' and '*tail*' for the tossing of a coin; 1, 2, 3, 4, 5, and 6 for the throwing of a die, and each one of the 52 cards in the deck of playing cards. The set of all possible outcomes is referred to as the 'sample space' and is denoted by S. *Events are the subsets of sample space*. For example, for the random experiment of 'throwing a die', while 1, 2, 3, 4, 5 and 6 are the *elementary outcomes*, and can be considered as events, one may also define 'events' using subsets of these elementary outcomes, Thus, we may consider 'even' and 'odd' as events – event 'even' being associated with the subset {2, 4, 6} and event 'odd' being associated with the subset {1, 3, 5}. Thus, in general, events are subsets of S and we assign a non-negative number $P(E)$, $0 \leq P(E) \leq 1$, for each event in such a way that axioms 1 to 3 above are satisfied.

A sample space may be '*discrete*' or '*non-discrete*'. It is said to be discrete if the number of elements in it, i.e., the number of elementary outcomes for the experiment are finite, or countably infinite. Otherwise, it is called a 'non-discrete' sample space. In all the random experiments considered above, the sample space is discrete. But, suppose our random experiment is to randomly choose an instant say between 9 a.m. and 10 a.m. for making a telephone call. For this experiment, the sample space is 'non-discrete'.

When we consider an experiment with a non-discrete sample space, we get into problems. It is not possible to consider every subset of this sample space as an event and assign probabilities to each of them without violating the 'axioms'. To overcome this problem, we consider as events only those subsets of S which belong to what is called the $\sigma-$ field, $\mathcal{B}$, defined on S as follows:

1. $S \in \mathcal{B}$
2. If the event $A \in \mathcal{B}$ then $\overline{A}$ also belongs to $\mathcal{B}$. ($\overline{A}$ is complement of A)
3. If any A and B belong to $\mathcal{B}$, then $A \cup B \in \mathcal{B}$

The three entities S, $\mathcal{B}$, and P where P is the probability measure, together constitute what is generally referred to as the '***probability space***'

From the three axioms listed in the beginning of Section 6.2, it is possible to derive the following basic properties of P:

1. $P(A) = 1 - P(\overline{A})$
2. $P(\phi) = 0$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6.3 CONDITIONAL PROBABILITY

Let A and B be two events defined on the same probability space with individual probabilities $P(A)$ and $P(B)$. Then $P(A \mid B)$, i.e., the conditional probability of A given B, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \begin{cases} \dfrac{P(AB)}{P(B)}; & P(B) \neq 0 \\ 0 \quad ; & \text{otherwise} \end{cases} \tag{6.1}$$

Remark

(i) If A and B are mutually exclusive, i.e., if $A \cap B = 0$; $P(A|B) = 0$.

(ii) If $P(A|B) = P(A)$, i.e., the occurrence of B does not affect the probability of A, events A and B are said to be statistically independent. In the case,

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A) \Rightarrow P(AB) = P(A) \cdot P(B). \tag{6.2}$$

Example 6.1 A die is thrown and you are told that the outcome is even. Then what is the probability that the result is 2?

Solution Let us denote the event 'even' by B and the event that the outcome is 2 by A. Then since B is said to have occurred if either 2, or 4 or 6 has turned up, $P(B) = \frac{3}{6} = \frac{1}{2}$

Since $$A \cap B = A, P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

$\therefore$ $P(2 \mid \text{even}) = 1/3$

Total probability theorem Let the events $A_1, A_2, \ldots, A_n$ belong to the same probability space and let them be such that

$$A_1 \cup A_2 \cup A_b \ldots \cup A_n = S$$

If B is any arbitrary event, also belonging to the same probability space, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)$$
$$= \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

This is called the total probability theorem

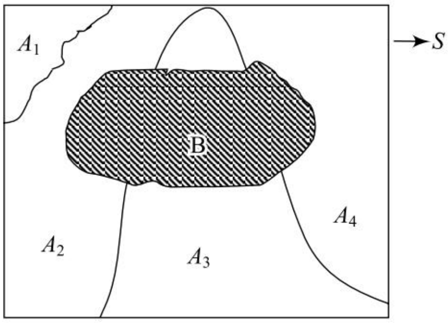


Fig. 6.1 *Partitioning of S*

Bayes' theorem Bayes' Theorem, or Bayes' rule enables us to find the conditional probability of A_i given B, in terms of the conditional probabilities of B given A_i, $i = 1$ to n.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

$$\therefore \qquad P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \qquad (6.3)$$

This is called Bayes' theorem.

Statistical independence As stated earlier, two events A and B are said to be *statistically independent* if

$$P(AB) = P(A) \cdot P(B)$$

The three events A, B and C are said to be statistically independent if the following two conditions are satisfied:

1. $P(AB) = P(A) \cdot P(B)$; $P(BC) = P(B) \cdot P(C)$ and $P(AC) = P(A) \cdot P(C)$
2. $P(ABC) = P(A) \cdot P(B) \cdot P(C)$

In general, n events, $A_1, A_2, \ldots, A_n$ are said to be independent if for every $k < n$ the events $A_1, A_2, \ldots, A_k$ are independent and further, if

$$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2) \ldots P(A_n)$$

Example 6.2 There are 5 boxes B_1, B_2, B_3, B_4, and B_5 containing compact fluorescent lamps. Each box contains 1000 lamps. It is known that B_1 has 5%, B_2 has 20%, B_3 has 3%, B_4 has 10%, and B_5 has 14% defective units. If a box is selected at random and randomly a lamp is picked out of it, what is the probability of that this lamp so picked, is defective? If the lamp so picked is found to be defective, what is the probability that it was picked from box B_1?

Solution Since the box has been randomly chosen,

$$P(B_1) = P(B_2) = P(B_3) = P(B_4) = P(B_5) = 1/5 = 0.2$$

$\therefore$ probability of the picked lamp being defective =

$$P(D) = P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3) + P(D|B_4)P(B_4) + P(D|B_5)P(B_5)$$
$$= 0.2[0.05 + 0.2 + 0.03 + 0.1 + 0.14] = 0.104$$

$\therefore$ the probability of the picked up lamp being defective = 0.104

Now, given that the lamp that is picked is defective the probability of its having been taken from box B_1 is say, $P(B_1|D)$

$$P(B_1|D) = \frac{P(B_1 \cdot D)}{P(D)} = \frac{P(D|B_1)P(B_1)}{P(D)}$$

But $P(D|B_1) = 0.05$, $P(D) = 0.104$ and $P(B_1) = 0.2$

$$P(B_1|D) = \frac{0.05 \times 0.2}{0.104} = \frac{0.01}{0.104} = 0.0961$$

6.4 RANDOM VARIABLES

Definition A real random variable is a mapping of the outcomes of a random experiment to the real line and satisfying the following two conditions:

1. $\{\boldsymbol{X} \le x\}$ i.e., $\{\boldsymbol{X}(\xi) \le x\}$ is an event for every real number x.
2. $P\{\boldsymbol{X}(\xi) = +\infty\} = 0 = P\{\boldsymbol{X}(\xi) = -\infty\}$

The mapping referred to in the above definition, is therefore having S, the set of all outcomes as its domain and R, the set of real numbers, as its range. If ξ represents an outcome, $\boldsymbol{X}(\xi)$ is used to denote the number, the random variable, assigned to the outcome ξ, and $\boldsymbol{X}$ denotes the rule according to which each ξ is allotted a real number. However, for simplicity of notation, we use $\boldsymbol{X}$ instead of $\boldsymbol{X}(\xi)$ to denote the number assigned to ξ. The ambiguity, if any, caused by this, may be resolved easily from the context.

For example, in the random experiment of tossing of a coin, we may assign the number 1 for the outcome '*heads*' and the number '0' for the outcome '*tails*'. Then

$$\boldsymbol{X}(\text{heads}) = 1 \text{ and } \boldsymbol{X}(\text{tails}) = 1$$

Suppose x denotes some real number. As x is given various values along the real line, the elements of S that constitute the set $\{\boldsymbol{X} \le x\}$ also change because, after all, $\{\boldsymbol{X} \le x\}$ represents a subset of S consisting of all the outcomes ξ which are such that $\boldsymbol{X}(\xi) \le x$. Thus $\{\boldsymbol{X} \le x\}$ is a set of outcomes. As mentioned in the definition of a random variable, we demand that the mapping be such that this set is an event for every x.

A complex random variable $\boldsymbol{Z}$ is given by

$$\boldsymbol{Z} = \boldsymbol{X} + j\boldsymbol{Y} \tag{6.4}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are real random variables.

Definition The Cumulative Distribution Function (CDF) of a random variable $\boldsymbol{X}$ is denoted by $F_X(x)$ and is defined by

$$F_X(x) \underline{\Delta} P\{\boldsymbol{X} \le x\} \tag{6.5}$$

To make the notation simpler, we shall use $F(x)$ instead of $F_X(x)$. We shall therefore be representing the CDF of a random variable $\boldsymbol{Y}$ by $F(y)$.

Example 6.3 For the random experiment of tossing of a coin, if we define a random variable $\boldsymbol{X}$ by saying that

$$\boldsymbol{X}(\xi = \text{Heads}) = 1 \text{ and } \boldsymbol{X}(\xi = \text{Tails}) = 0$$

The CDF of the random variable (r.v) $\boldsymbol{X}$ will be as shown in Fig. 6.2, since $P\{\text{Heads}\} = P\{\text{Tails}\} = 0.5$. Since $F(x) = P\{\boldsymbol{X} \le x\}$, when $x < 0$, $F(x) = 0$. when $x \ge 0$, but less than one, $F(x) = 0.5$ and for $x \ge 1$, $F(x) = 1$. Thus, we get a staircase type of CDF.

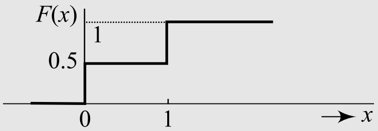


Fig. 6.2 *Cumulative distribution function of* ***X***

Properties of CDF

1. $F(x)$ lies between 0 and 1; i.e., $0 \le F(x) \le 1$
2. $F(\infty) = 1$ and $F(-\infty) = 0$
3. $F(x)$ is a non-decreasing function of x.
4. $F(x)$ is continuous from the right; i.e., $\lim\limits_{\in \to 0} F(x + \in) = F(x); \in > 0$
5. $F(b) - F(a) = P[a < \boldsymbol{X} < b]$
6. $P[\boldsymbol{X} = x_1] = F(x_1) - F(x_1^-)$, where, $F(x_1^-) \underline{\Delta} \lim\limits_{\in \to 0} F(x_1 - \in); \in > 0$

6.4.1 Types of Random Variables

Random variables are categorized as discrete random variables, continuous random variables and mixed-type random variables, based upon the type of CDF.

A random variable, whose CDF is having a staircase shape is called a *discrete random variable*. A random variable with a CDF which is a continuous function of x is called a *continuous random variable*. A random variable which is neither a discrete random variable nor a continuous random variable is called a *mixed random variable*.

Definition The Probability Density Function (PDF) of a random variable $\boldsymbol{X}$ is defined as the derivative with respect to x of its CDF, viz., $F_x(x)$

$$\therefore \qquad f_x(x) = \frac{dF_x(x)}{dx} \tag{6.6}$$

If $\boldsymbol{X}$ is a discrete random variable, we know that its $F_X(x)$ will be of the staircase type. Hence, as shown in Fig. 6.3, its probability density function (PDF) will be zero everywhere except at the points of discontinuity, where it will have impulses.

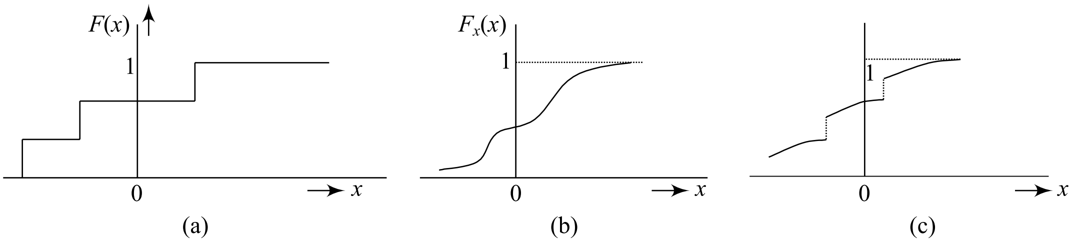


Fig. 6.3 *(a) CDF of a discrete random variable , (b) CDF of a continuous random variable, (c) CDF of a mixed random variable*

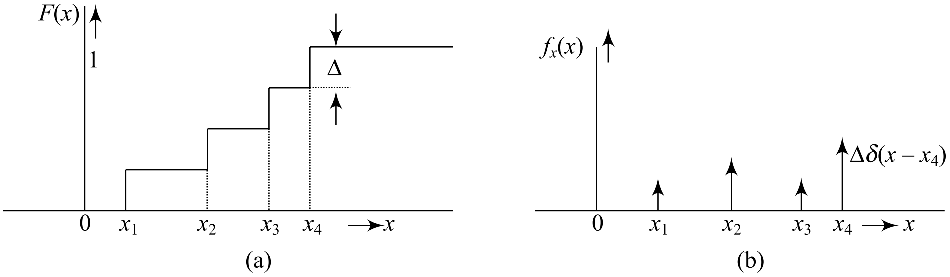


Fig. 6.4 *(a) CDF of a discrete random variable, (b) PDF of a discrete random variable*

The PDF, $f_X(x)$ of a continuous random variable $\boldsymbol{X}$ will be a continuous function of x. The PDF of a mixed random variable involves impulses but need not necessarily be zero between any two consecutive impulses.

Properties of PDF

1. Since the CDF is a non-decreasing function of x, its derivative, $f_X(x)$, will be non-negative, i.e., $f_X(x) \geq 0$.
2. The area under any probability density function will be unity, i.e., $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
3. $\int_{x_1}^{x_2} f_X(x)\,dx = P[x_1 < \boldsymbol{X} \leq x_2]$
4. $F_X(x) = \int_{-\infty}^{x} f_X(\alpha)\,d\alpha$

In the case of a discrete random variable, since the derivative of the CDF results in impulses, it is more appropriate to talk in terms of probability masses, $p_i = P[\boldsymbol{X} \leq x_i]$. In this case, $p_i \geq 0$ for all i and $\sum_i p_i = 1$

6.4.2 Some Useful Random Variables

In what follows, we give the distributions or density functions of a number of continuous and discrete random variables which are useful in the study of communication engineering (analog and digital)

1. Continuous random variables

(a) Uniform random variable A random variable $\boldsymbol{X}$ is called a uniform random variable if its probability density function $f_X(x)$ is given by

$$f_X(x) = \begin{cases} \dfrac{1}{(x_2 - x_1)}; \ x_1 \leq x \leq x_2; & -\infty < x_1 < x_2 < \infty \\ 0 & \text{otherwise} \end{cases} \tag{6.7}$$

Such a random variable is generally denoted by $\boldsymbol{U}(x_1, x_2)$. The cumulative distribution function CDF of this random variable is as shown in Fig. 6.5(b) and is given by

$$F_X(x) = \begin{cases} 1 & \text{for } x \geq x_2 \\ \dfrac{(x - x_1)}{(x_2 - x_1)} & \text{for } x_1 \leq x \leq x_2 \\ 0 & \text{for } x < x_1 \end{cases} \tag{6.8}$$

The uniform random variable is used to model a continuous random variable, about which we have no other knowledge except for the finite range over which its values are spread. Such a situation arises in the case of a sinusoid whose phase is random. We model its phase by a uniformly distributed random variable, its range of values being from 0 to 2π.

(b) Gaussian or normal random variable The random variable $\boldsymbol{X}$ is said to be a Gaussian or normal random variable with mean 'm' and variance σ^2 if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} \tag{6.9}$$

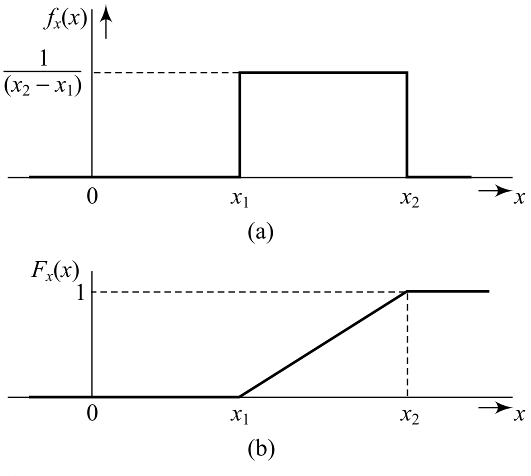


Fig. 6.5 *(a) PDF of a uniformly distributed random variable, (b) CDF of a uniformly distributed random variable*

This density function has a shape as shown in Fig. 6.6 and is symmetric with respect to $x = m$. If σ^2 is large, the values of $\boldsymbol{X}$ are more spread out around the mean value and if it is small, the values are more concentrated near the mean value. Since the density function is completely determined by the two parameters – the mean and the variance, it is generally denoted by $N(m, \sigma^2)$. A Gaussian random variable with zero mean and unit variance, is called the standard normal random variable and is denoted by $N(0, 1)$

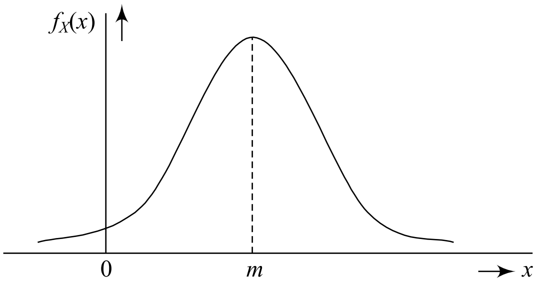


Fig. 6.6 *Density function of a Gaussian random variable with mean m and variance σ^2*

Gaussian distribution function, $F_X(x)$ is given by

$$F_X(x) = P[\boldsymbol{X} \le x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-m)^2/2\sigma^2} dy$$

$$= \int_{-\infty}^{\left(\frac{x-m}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{6.10}$$

The Gaussian density function is the most extensively used one in communication engineering. This is because thermal noise, which is a major source of noise in communications, is, Gaussian in nature.

For a $N(0, 1)$ random variable, Eq. (6.10) reduces to

$$g_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = P[\boldsymbol{X} \le x] \tag{6.11}$$

In communications engineering, the so-called 'tail probability' of a Gaussian random variable is the one, which one has to determine frequently while calculating error probabilities. So it is given a special symbol $Q(x)$ called the Q-function and is given by

$$Q(x) = 1 - g_X(x) = P[X > x] \tag{6.12}$$

This Q-function, which is extensively tabulated, has the following important properties:

$$Q(-x) = 1 - Q(x) \tag{6.13}$$

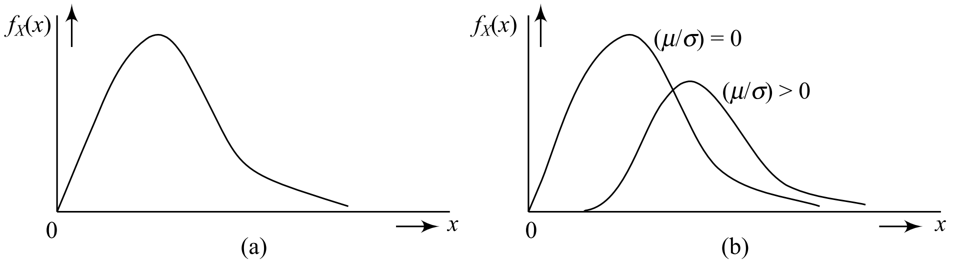
$$Q(0) = \frac{1}{2} \tag{6.14}$$

$$Q(\infty) = 0 \tag{6.15}$$

(c) Rayleigh random variable A random variable $\boldsymbol{X}$ is said to be having Rayleigh distribution with parameter σ^2 if its density function is

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/2\sigma^2}; & x \ge 0 \\ 0; & x < 0 \end{cases} \tag{6.16}$$

The Rayleigh distributed random variable has a mean value of $\sigma\sqrt{\frac{\pi}{2}}$ and a variance of $\left(2 - \frac{\pi}{2}\right)\sigma^2$. The shape of Rayleigh density function is shown in Fig. 6.7(a).



(a) (b)

Fig. 6.7 *(a) Rayleigh density function, (b) Ricean density function*

If a band pass signal has identically distributed Gaussian zero-mean random processes as its inphase and quadrature components, it can be shown that its envelope will have Rayleigh distribution. This density function is extensively used in the study of fading communication channels.

(d) Ricean random variable (Rice distribution) A random variable $\boldsymbol{X}$ is said to be a Ricean random variable with parameters μ and σ^2, if its probability density function $f_X(x)$ is of the form

$$f_X(x) = \left[\frac{1}{\sigma^2}\, x e^{-(x^2+\mu^2)/2\sigma^2}\right] \cdot I_0\left(\frac{\mu x}{\sigma^2}\right) \tag{6.17}$$

where

$$I_0(\alpha) \,\Delta\, \frac{1}{\pi}\int_0^{\pi} e^{\alpha \cos\theta}\, d\theta \tag{6.18}$$

i.e., $I_0(\alpha)$ is the modified Bessel function of the first kind and zeroth order.

The shape of Rice density function (see Fig. 6.6(b)) is somewhat similar to that of Rayleigh density function. In fact, as can be seen from Eqs. (6.16) and (6.17), Rice density function simplifies into Rayleigh density function when the parameter $\mu = 0$.

If a band pass signal has Gaussian random processes with the same variance but *different* non-zero mean values as its inphase and quadrature components, it can be shown that its envelope will have Ricean distribution. Ricean distribution, just like Rayleigh distribution, is widely used in the study of fading channels. The sum of a sinusoid and a narrowband noise can be shown to have Ricean distribution for its envelope.

2. Discrete random variables

(a) Bernoulli random variable A discrete random variable, $\boldsymbol{X}$, is said to be a Bernoulli random variable provided it takes the values 1 and 0 with probabilities of P and (1-P). This random variable is quite useful in modeling a binary data generator and also in modeling the error pattern in the received binary data when the channel introduces random errors.

(b) Binomial random variable A discrete random variable, $\boldsymbol{X}$**,** is said to be a Binomial random variable with parameters n and p if

$$P[\boldsymbol{X} = k] = \binom{n}{k} p^k q^{n-k};\ 0 \le k \le n \tag{6.19}$$

In fact, this gives the number of 1s in a sequence of 1s and 0s generated by n independent Bernoulli trials. Therefore, it may be used to model the total number of erroneous bits in the received data when a sequence of n bits is transmitted over a channel having a bit-error probability of p.

(c) Poisson random variable Poisson distribution is used to model phenomena such as ‘shot noise’ in electron devices and ‘radio active decay’.

The Poisson random variable represents the number of times a particular event occurs in an interval t_0 seconds, given that the probability of its occurrence once in a small interval Δt is $\lambda(\Delta t)$.

The Poisson frequency function is given by

$$P(k, t_0) = e^{-\lambda \tau_0}\left[\frac{(\lambda t_0)^k}{k!}\right] \tag{6.20}$$

= Probability of occurrence of the event k number of times in a time interval of t_0 seconds.

When $t_0 = 1$ second, Eq. (6.20) reduces to

$$P(k, 1) = e^{-\lambda}\left[\frac{(\lambda)^k}{k!}\right] \tag{6.21}$$

The binomial model becomes approximately the same as the Poisson model when the two parameters n and p of the binomial model are such that n is very large, p is very small and np remains constant.

Example 6.4 A fair coin is tossed 5 times. What is the probability of

(a) 'Heads' appearing two times?
(b) 'Heads' appearing atleast once?
(c) 'Heads' appearing twice and 'Tails' appearing thrice?
(d) 'Tails' appearing at the most once?

Solution

(a) From Eq. (6.19), we have

$$P[\boldsymbol{X} = 2] = 5_{C_2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{5!}{2!3!}\cdot\frac{1}{4}\cdot\frac{1}{8} = \frac{5}{16}$$

(b) $P[\boldsymbol{X} \geq 1] = 1 - P[X = 0] = 1 - 5_{C_0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^5 = \frac{31}{32}$

(c) $P[\boldsymbol{X} : 2 \text{ heads and } 3 \text{ tails}] = 5_{C_2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \frac{5}{16}$

(d) $P[\boldsymbol{X} \ngtr 1] = P[\boldsymbol{X} = 0] + P[\boldsymbol{X} = 1] = 5_{C_2}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0 + 5_{C_4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 = \frac{6}{32}$

Example 6.5 Determine the mean of a Poisson random variable.

Solution Mean of

$$\boldsymbol{X} = E[\boldsymbol{X}] = \sum_{x=0}^{\infty} xP(\boldsymbol{X} = x)$$

$$= \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x\cdot\frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{(Since the term corresponding to } x = 0 \text{ is zero)}$$

$$= \sum_{x=1}^{\infty} x\cdot\frac{e^{-\lambda}\lambda^x}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda}\left[1 + \frac{\lambda}{1} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right]$$

By expanding the summation in the above step.

$$1+\frac{\lambda}{1}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots=e^{\lambda}$$

$$\therefore \qquad E[X]=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda$$

Example 6.6 Gaussian density function is given as

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-m)^2/2\sigma^2}$$

Show that $\frac{1}{\sqrt{2\pi\sigma^2}}$ is a normalization factor required to make the total area under the density function equal to 1.

Solution Let $(x-m)\underline{\Delta}\, z \quad \therefore\ dx=dz$

$$f_X(z)=\left[\int_{-\infty}^{\infty}e^{-z^2/2\sigma^2}dz\right]\frac{1}{\sqrt{2\pi\sigma^2}}$$

Now, let $A\,\underline{\Delta}\int_{-\infty}^{\infty}e^{-z^2/2\sigma^2}dz$. Then it is enough if we show that

$$A=\sqrt{2\pi\sigma^2}$$

$\therefore$ consider

$$A^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(z^2+W^2)/2\sigma^2}dzdw$$

If we now put $z = r\cos\theta$ and $w = r\sin\theta$, $dzdw = rdrd\theta$ and $(z^2 + w^2) = r^2$. Hence, we get

$$A^2=\int_0^{2\pi}\int_0^{\infty}e^{-r^2/2\sigma^2}rdrd\theta=\left[\int_0^{2\pi}d\theta\right]\left[\int_0^{\infty}e^{-r^2/2\sigma^2}rdr\right]$$

Now, put $\frac{r^2}{2\sigma^2}\,\underline{\Delta}\, v$. Then, $dv=\frac{2rdr}{2\sigma^2}=\frac{rdr}{\sigma^2}$

$$\therefore \qquad A^2=2\pi\sigma^2\int_0^{\infty}e^{-v}dv=2\pi\sigma^2 \qquad \therefore\ A=\sqrt{2\pi\sigma^2}$$

Hence, the factor $\frac{1}{\sqrt{2\pi\sigma^2}}$ in $f_X(x)$ is a normalization factor.

6.5 FUNCTIONS OF A SINGLE RANDOM VARIABLE

Consider a function $g(x)$ of the real variable x. Let us also consider a random variable $\boldsymbol{X}$ whose range is included in the domain of $g(x)$. Then, for every outcome, ξ, of the random experiment, $\boldsymbol{X}(\xi)$ is a real number which is in the domain of the function $g(x)$. Thus, we may talk of the function $g(\boldsymbol{X})$, a function of the random variable $\boldsymbol{X}$. If we can call this as another random variable $\boldsymbol{Y}$, then

$$\boldsymbol{Y}=g(\boldsymbol{X}) \tag{6.22}$$

We can then talk of the CDF, $F_Y(y)$ of the random variable $\boldsymbol{Y}$.

$$F_Y(y)=P[\xi\in S:g(\boldsymbol{X}(\xi))\le y] \tag{6.23}$$

Now, for $\boldsymbol{Y}$ to be a random variable for every y, the set of values of x such that $g(x) \le y$ must consist of the unions and intersections of a countable number of intervals. This means that for every y,

$$Y = g(x)$$

must have a countable number of solutions. Then only

$$g(\boldsymbol{X}(\xi)) \le y$$

will be an event. If the function $g(x)$ belongs to such a class, and further, if at every $x_i = g^{-1}(y)$, a derivative exists for the function $g(x)$, and the derivative is not zero, then it can be shown that the density function of $\boldsymbol{Y}$ is given by

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \tag{6.24}$$

Example 6.7 If $\boldsymbol{Y} = \boldsymbol{X}^2$, find $f_Y(y)$ in terms of $f_X(x)$.

Solution Let us consider the set of values $\{x_i\}$ of x which are such that for a given y, $g(x_i) \le y$ for all i.

For $y < 0$, there does not exist any value of x for which $g(x) < y$, i.e $x^2 < y$. So let us consider only $y \ge 0$. For this case $x^2 \le y$ is true for

$$-\sqrt{y} \le x \le \sqrt{y}.$$

$$\therefore \quad F_Y(y) = P[-\sqrt{y} \le \boldsymbol{X} \le \sqrt{y}] = P[\boldsymbol{X} \le \sqrt{y}] - P[\boldsymbol{X} \le -\sqrt{y}]$$

$$\therefore \quad F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \qquad ; y > 0$$

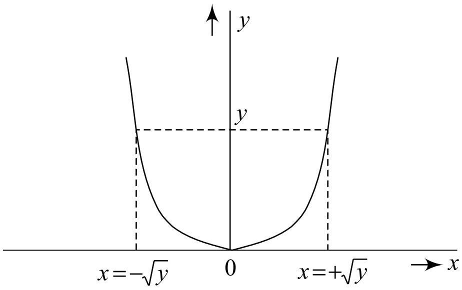


Fig. 6.8 $y = x^2$

To get the corresponding density function, $f_Y(y)$, we have

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] & \text{for } y > 0 \\ 0 & \text{for } y < 0 \end{cases}$$

6.6 MEAN, VARIANCE, AND CHARACTERISTIC FUNCTION

6.6.1 Mean

The mean, or the expected value of a random variable $\boldsymbol{X}$ with the density function $f_X(x)$, is defined as

$$E\{\boldsymbol{X}\} \,\underline{\Delta}\, \int_{-\infty}^{\infty} x f_X(x)\,dx \tag{6.25}$$

The expected value, or mean, will be just a number and it is generally denoted by either m_X or η_X.
For a discrete random variable, we had already seen that

$$f_X(x) = \sum_i p_i \delta(x - x_i) \tag{6.26}$$

Substituting this for $f_X(x)$ in Eq. (6.25), we get

$$E\{\boldsymbol{X}\} = \sum_i p_i x_i, \quad \text{where} \quad p_i = P[\boldsymbol{X} = x_i] \tag{6.27}$$

Example 6.8 A random variable $\boldsymbol{X}$ has a density function $f_X(x)$ given by

$$f_X(x) = 2e^{-2x}u(x)$$

Find the expected value of this random variable.

Solution $\eta_X = \int_{-\infty}^{\infty} x \cdot 2e^{-2x}u(x)\,dx = \int_{0}^{\infty} 2xe^{-2x}dx = \frac{1}{2}$

Example 6.9 A loaded die produces the numbers 1, 2, 3, 4, 5 and 6 with probabilities 0.10, 0.12, 0.12, 0.14, 0.20 and 0.32, respectively. Find the mean value.

Solution $\eta = \sum_{i=1}^{6} i \cdot p_i = (1 \times 0.10) + 2(0.12) + 3(0.12) + 4(0.14) + 5(0.20) + 6(0.32)$

$$= 4.18$$

6.6.2 Variance

The variance of a random variable $\boldsymbol{X}$ with expected value η_X, is defined as

$$\text{Var}[\boldsymbol{X}] = \sigma^2 = E[(\boldsymbol{X} - \eta_X)^2] = E[\boldsymbol{X}^2] - 2\eta_X E[\boldsymbol{X}] + \eta_X^2 = E[\boldsymbol{X}^2] - \eta_X^2$$

$$\therefore \qquad \sigma^2 = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2 \tag{6.28}$$

Since $E[(X - \eta_X)^2] \geq 0,$ it follows that

$$E[\boldsymbol{X}^2] \geq \{E[\boldsymbol{X}]\}^2$$

Discrete random variable: In this case,

$$\sigma^2 \underline{\Delta} \sum_i p_i (x_i - \eta_X)^2 \quad \text{where,} \quad p_i = P[\boldsymbol{X} = x_i] \tag{6.29}$$

Note *The positive square-root of the variance is referred to as the 'standard deviation'.*

Example 6.10 Find the mean value and the variance of a random variable $\boldsymbol{X}$ which is uniformly distributed between $x = a$ to $x = b$.

Solution

(a) Mean $\eta_X = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \int_a^b x \cdot \frac{1}{(b-a)}dx = \frac{b+a}{2}$

(b) Variance $\sigma^2 = \int_{-\infty}^{\infty} E[\boldsymbol{X} - \eta_X]^2 dx = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2$

$$= \left[\left\{\int_a^b x^2 \cdot \frac{1}{(b-a)}dx\right\} - \frac{(b+a)^2}{4}\right] = \frac{(b-a)^2}{12}$$

$$\therefore \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

Example 6.11 Find the variance of a Bernoulli random variable.

Solution A Bernoulli random variable takes the values 1 and 0 with probabilities p and $(1 - p)$

$$\therefore \qquad \eta_X = 1 \cdot p + 0 \cdot (1 - p) = p$$

$$E[\boldsymbol{X}^2] = p \cdot 1^2 + (1 - p) \cdot 0^2 = p$$

$$\therefore \qquad \sigma^2 = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2 = p - p^2 = p(1 - p)$$

As noted earlier, the positive square-root of the variance of a random variable $\boldsymbol{X}$ is referred to as its 'standard deviation' and is denoted by σ_{x}. The standard deviation provides a measure of the spread of the values of the random variable $\boldsymbol{X}$ with respect to its mean value η_x. Since

$$\sigma_x^2 = E[\boldsymbol{X}^2] = \{E[X]\}^2$$

$$\sigma_x = \text{Standard deviation} = \sqrt{E[\boldsymbol{X}^2] - \{E[X]\}^2} = \sqrt{\overline{\boldsymbol{X}^2} - \eta_x^2} \tag{6.30}$$

where $\overline{X^2} - E[\boldsymbol{X}^2]$

Statement of Chebyshev's inequality

Note *Chebyshev is spelt by some authors as Tchebycheff.*

Let $\boldsymbol{X}$ be a continuous random variable with mean η_x and standard deviation σ_x. Let k be any positive number. Then Chebyshev's inequality says that

$$P[|\boldsymbol{X} - n_x| \geq k\sigma_x] \leq \frac{1}{k^2} \tag{6.31}$$

Physically this means that the probability of a random variable $\boldsymbol{X}$ taking a value that is more than k standard deviations away (on either side) from its mean value η_x, will not be greater than $(1/k^2)$.

Proof Let $\boldsymbol{Y} \triangleq \boldsymbol{X} - \eta_x$, and $k\sigma_x \triangleq a$
Then

$$E[\boldsymbol{Y}^2] = \int_{-\infty}^{\infty} y^2 f_y(y)dy \geq \int_{-\infty}^{\infty} y f_y(y)dy + \int_{a-\epsilon}^{\infty} y^2 f_y(y)dy \tag{6.32}$$

But we know that whenever $|\boldsymbol{X} - \eta_x| \geq k\sigma_x$, $y^2 \geq a^2$
i.e., whenever $a \leq |y| < \infty$, $y^2 \infty a^2$. Hence, Eq. (6.32) may be rewritten as

$$E[\boldsymbol{Y}^2] = a^2 \left[\int_{-\infty}^{-a} f_y(y)\,dy + \int_{a-\epsilon}^{-\infty} f_y(y)\,dy \right] \tag{6.33}$$

But

$$\int_{-\infty}^{-a} f_y(y)\,dy = P[\boldsymbol{Y} \leq -a]$$

and

$$\int_{a-\epsilon}^{\infty} f_y(y)\,dy = P[\boldsymbol{Y} \geq a]$$

Hence, Eq. (6.33) may be written as

$$\frac{1}{a^2}E[\boldsymbol{Y}^2] \geq P[\boldsymbol{Y} \leq -a] + P[\boldsymbol{Y} \geq a]$$

i.e.,
$$\frac{\sigma_x^2}{a^2} = \frac{1}{k^2} \geq P[|\boldsymbol{Y}| \geq a]$$

i.e.,
$$P|\boldsymbol{X} - \eta_x| \geq [k\sigma_x] \leq \frac{1}{k^2} \tag{6.34}$$

Properties of Mean and Variance Let c be a constant and $\boldsymbol{X}$ be a random variable with mean η_X. Then mean will have the following properties:

1. $E[c\boldsymbol{X}] = c\,E[\boldsymbol{X}] = c\eta_X$
2. $E[c] = c$
3. $E[\boldsymbol{X} + c] = E[\boldsymbol{X}] + c = \eta_X + c$

If the random variable $\boldsymbol{X}$ has a variance σ_X^2 and if c is a constant, the following are the properties of variance:

1. $\text{Var}[c \cdot \boldsymbol{X}] = c^2\text{Var}[\boldsymbol{X}] = c^2\sigma_X^2$
2. $\text{Var}\,[c] = 0$
3. $\text{Var}[\boldsymbol{X} + c] = \text{Var}[\boldsymbol{X}] = \sigma_X^2$

6.6.3 Characteristic Function of a Random Variable

Definition The characteristic function of a random variable $\boldsymbol{X}$ is denoted by $\phi_X(\omega)$ and is defined as

$$\phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}dx \tag{6.35}$$

If we now define
$$\Phi(s) \triangleq \int_{-\infty}^{\infty} f_X(x)e^{sx}dx$$

So that
$$\Phi(j\omega) = \phi_X(\omega) \text{ of Eq. (6.35)}$$

Taking the first derivative of $\Phi(s)$ with respect to s, we get

$$\Phi^{(1)}(s) = \int_{-\infty}^{\infty} xf_X(x)e^{sx}dx = E[\boldsymbol{X}e^{s\boldsymbol{X}}] \tag{6.36}$$

If we take the nth derivative with respect to s, we get

$$\Phi^{(n)}(s) = \int_{-\infty}^{\infty} (x)^n f_X(x)e^{sx}dx = E[(\boldsymbol{X})^n e^{s\boldsymbol{X}}] \tag{6.37}$$

If we put $s = 0$ in Eqs. (6.36) and (6.37), we find that

$$\Phi^{(1)}(0) = \text{First derivative of } \Phi(s) \text{ with respect to } s \text{ at the origin} = E[\boldsymbol{X}] \tag{6.38}$$

$$\Phi^{(n)}(0) = n^{\text{th}} \text{ derivative of } \Phi(s) \text{ with respect to } s \text{ at the origin} = E[\boldsymbol{X}^n] \tag{6.39}$$

i.e., derivatives of various orders of the moment generating function $\Phi(s)$ at the origin give the moments of various orders for the random variable $\boldsymbol{X}$.

Note *Thus, the characteristic function of a random variable **X** helps us in determining moments of various order for **X** in an easy manner.*

Discrete random variable If $\boldsymbol{X}$ is a discrete random variable which takes values x_i with probabilities $p_i = 1, 2, 3, \ldots$, then Eq. (6.28) reduces to

$$\phi_X(\omega) = \sum_i p_i e^{j\omega x_i} \tag{6.40}$$

Example 6.10 Show that the characteristic function of a Gaussian random variable $\boldsymbol{X}$ with mean value m and variance σ^2, is given by

$$\phi_X(\omega) = e^{(jm\omega - 0.5\sigma^2\omega^2)}$$

Solution The density function $f_X(x)$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2}$$

Let us transform the random variable $\boldsymbol{X}$ into another random variable $\boldsymbol{Y}$ by putting

$$\boldsymbol{Y} = (\boldsymbol{X} - m) / \sigma$$

Then, $E[\boldsymbol{Y}] = \frac{1}{\sigma} E[\boldsymbol{X}] - \frac{m}{\sigma} = 0.$ $\quad\therefore$ mean of $\boldsymbol{Y}$ is zero

$$\text{Var}[\boldsymbol{Y}] = E[\boldsymbol{Y}^2] - \{E[\boldsymbol{Y}]\}^2 = E[\boldsymbol{Y}^2] = E\left[\frac{(\boldsymbol{X} - m)}{\sigma^2}\right] = 1$$

$\therefore$ $\boldsymbol{Y} = N(0, 1)$; and its density function is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy$$

$$\Phi_Y(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \cdot e^{sy} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(sy - y^2/2)} dy$$

But $sy - y^2/2 = \frac{s^2}{2} - \frac{1}{2}(y - s)^2$. $\quad\therefore$ substituting this in the RHS of the above equation,

$$\Phi_Y(s) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left[\frac{s^2}{2} - \frac{1}{2}(y-s)^2\right]} dy = e^{s^2/2}\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-s)^2/2)} dy\right] = e^{s^2/2} \tag{6.41}$$

But $\boldsymbol{Y} = \frac{\boldsymbol{X} - m}{\sigma}$ or $\boldsymbol{X} = \sigma\boldsymbol{Y} + m$

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega(\sigma y + m)} dx = e^{j\omega m} \int_{-\infty}^{\infty} f_X(x) e^{j\omega\sigma y} dx = E[e^{j\omega\sigma \boldsymbol{Y}}] e^{j\omega m} \tag{6.42}$$

Since $\Phi_X(\omega) = E[e^{j\omega \boldsymbol{X}}]$, we may write $E[e^{j\omega\sigma \boldsymbol{Y}}] = \Phi_{\sigma y}(\omega)$

But $\Phi_Y(\omega) = e^{-\omega^2/2}$ from Eq. (6.41)

$$\Phi_{\sigma Y}(\omega) = e^{-\sigma^2\omega^2/2}$$

$\therefore$ from Eq. (6.42), we have

$$\Phi_X(\omega) = e^{j\omega m} \cdot e^{-\sigma^2\omega^2/2} = e^{(j\omega m - 0.5\sigma^2\omega^2)}$$

Example 6.11 Find the characteristic function of a Bernoulli random variable.

Solution The Bernoulli random variable takes the values 1 and 0 with probabilities p and $(1 - p)$, respectively

$\therefore$
$$\Phi_X(\omega) = p \cdot e^{j\omega \cdot 1} + (1 - p) \cdot e^{j\omega \cdot 0} = 1 + p[e^{j\omega} - 1]$$

6.7 FUNCTIONS OF TWO RANDOM VARIABLES

Consider two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined on the same probability space. The cumulative distribution functions, $F_X(x)$ and $F_Y(y)$ defined as

$$F_X(x) = P[\boldsymbol{X} \le x] \text{ and } F_Y(y) = P[\boldsymbol{Y} \le y]$$

are called marginal distribution functions and the corresponding density functions are called marginal density functions.

We may now define joint, or bivariate distribution function $F_{XY}(x, y)$ or $F(x, y)$ of the two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$F(x, y) \underline{\Delta} P[\boldsymbol{X} \le x, \boldsymbol{Y} \le y] \tag{6.43}$$

The joint density function may be defined as

$$f(x, y) \underline{\Delta} \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{6.44}$$

Then, since $F(x, -\infty) = F(-\infty, y) = 0$ and $F(+\infty, +\infty) = 1$, we have

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha, \beta) d\alpha d\beta \tag{6.45}$$

The marginal distribution functions and density functions can be obtained from the joint distribution and density functions respectively as follows:

$$F_x(x) = F(x) = F_{XY}(x, \infty) \quad \text{and} \quad F_y(y) = F(y) = F_{XY}(\infty, y) \tag{6.46}$$

and
$$f(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy; \quad f(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx \tag{6.47}$$

1. Discrete random variables: If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two discrete random variables defined on a certain probability pace, and if they take values x_i and y_k with probabilities p_i and q_k, respectively and given by

$$p_i = P[\boldsymbol{X} = x_i] \quad \text{and} \quad q_k = P[\boldsymbol{Y} = y_k] \tag{6.48}$$

Then their joint probability p_{ik} is given by

$$p_{ik} = P[\boldsymbol{X} = x_i \text{ and } \boldsymbol{Y} = y_k] \tag{6.49}$$

Of course, just like the marginal probabilities, the joint probabilities also add up to a value 1.

i.e.,
$$\sum_i \sum_k p_{ik} = 1 \tag{6.50}$$

Also,
$$p_i = \sum_k p_{ik} \quad \text{and} \quad p_k = \sum_i p_{ik} \tag{6.51}$$

Conditional CDFs and conditional PDFs Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random variables defined on the same probability space. The conditional CDF of $\boldsymbol{Y}$ given $\boldsymbol{X} \le x$, a real number, is denoted by $F_Y(y|\boldsymbol{X} \le x)$ and is defined by

$$F_Y(y|\boldsymbol{X} \le x) \underline{\Delta} \frac{P\{\boldsymbol{X} \le x, \boldsymbol{Y} \le y\}}{\boldsymbol{P}\{\boldsymbol{X} \le x\}} = \frac{F(x, y)}{F_X(x)} \tag{6.52}$$

$$\therefore \qquad f_Y(y|\boldsymbol{X} \le x) = \frac{\dfrac{\partial}{\partial y}[F(x,y)]}{F_X(x)} \tag{6.53}$$

The PDF of $\boldsymbol{Y}$ given $\boldsymbol{X} = x$, is represented by $f_{y|x}(y|x)$ or, $f(y|x)$ and is given by

$$f(y|x) = \frac{f(x, y)}{f(x)} \tag{6.54}$$

Similarly,

$$f(x|y) = \frac{f(x, y)}{f(y)} \tag{6.55}$$

If the random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent,

$$f(y|x) = \frac{f(x, y)}{f(x)} = f(y) \quad \therefore \quad f(x, y) = f(x) \cdot f(y) \tag{6.56}$$

2. Discrete random variables: If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete type of random variables with $P[\boldsymbol{X} = x_i] = p_i$ and $P[\boldsymbol{Y} = y_k] = q_k$,

and $P\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_k\} = p_{ik}$ with say $i = 1$ to M and $k = 1$ to N, then

$$P\{\boldsymbol{Y} = y_k | \boldsymbol{X} = x_i\} = \frac{P\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_k\}}{P\{\boldsymbol{X} = x_i\}} = \frac{p_{ik}}{p_i} \tag{6.57}$$

(a) *Conditional mean and variance* The conditional mean of the random variable $\boldsymbol{Y}$ given that $\boldsymbol{X} = x$, is represented by $\eta_{y|x}$ and is given by

$$\eta_{y|x} = E[\boldsymbol{Y}|x] = \int_{-\infty}^{\infty} y f(y|x)\, dy \tag{6.58}$$

The conditional variance is represented by $\sigma^2_{y|x}$ and is given by

$$\sigma^2_{y|x} = E[(\boldsymbol{Y} - \eta_{y|x})^2 | x] = \int_{-\infty}^{\infty} (y - \eta_{y|x})^2 f_{y|x}(y|x)\, dy \tag{6.59}$$

(b) *Independence, uncorrelatedness and orthogonality*

(i) If two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent,

$$f(y|x) = f(y); \quad f(x|y) = f(x) \quad \text{and} \quad f(x, y) = f(x) \cdot f(y)$$

(ii) The covariance of the two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as

$$\begin{aligned} C_{XY} &= E[(\boldsymbol{X} - \eta_X)(\boldsymbol{Y} - \eta_Y)] \\ &= E[\boldsymbol{XY}] - E[\boldsymbol{X}]E[\boldsymbol{Y}] \end{aligned} \tag{6.60}$$

The correlation coefficient ρ_{XY} of two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is

$$\rho_{XY} \triangleq \frac{C_{XY}}{\sigma_X \sigma_Y} \tag{6.61}$$

The random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be uncorrelated if their covariance is zero, i.e.,

$$C_{XY} = 0, \text{ i.e., when } \rho_{XY} = 0 \text{ or } E[\boldsymbol{XY}] = E[\boldsymbol{X}]E[\boldsymbol{Y}] \tag{6.62}$$

(iii) Random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be orthogonal, if

$$E[\boldsymbol{XY}] = 0$$

6.8 JOINTLY GAUSSIAN RANDOM VARIABLES

Definition Two random variables X and Y are said to be jointly Gaussian, if their joint density function is of the form

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right\}\right] \quad (6.63)$$

Properties

1. If X and Y are jointly Gaussian, then (a) they are individually Gaussian, and (b) the conditional densities $f(x|y)$ and $f(y|x)$ are also Gaussian.
2. If X and Y are individually Gaussian, they need not necessarily be jointly Gaussian.
3. Jointly Gaussian random variables are completely characterized by their mean vector and covariance matrix.
4. For jointly Gaussian random variables, uncorrelatedness implies statistical independence.

6.9 CENTRAL LIMIT THEOREM

The central limit theorem states that if $(X_1, X_2, \ldots, X_n)$ are independent random variables with means $(m_1, m_2, \ldots, m_n)$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots, \sigma_n^2)$, then the cumulative distribution function of the random variable

$$\sum_{i=1}^{n}\left(\frac{X_i - m_i}{\sqrt{n}\,\sigma_i}\right)$$

converges to that of a Gaussian random variable having a mean of zero and a variance of 1.

In case, the n random variables are not only independent, but are also identically distributed with mean of each = m and variance of each = σ^2, then, the CDF of their mean converges to the CDF of a Gaussian random variable having a mean of m and a variance of (σ^2/n).

It is as a consequence of central limit theorem that the sum of the noises produced by a very large number of independent sources tends to have Gaussian distribution.

Example 6.12 Random variable $Y = \sin X$, where X is uniformly distributed between $-\pi/2$ to $+\pi/2$. Find the density function of Y.

Solution $\Phi_Y(\omega) = E[e^{j\omega y}] = E[e^{j\omega \sin x}] = \int_{-\infty}^{\infty} e^{j\omega \sin x} f_X(x)\,dx$

Since X is uniformly distributed over $-\pi/2$ to $+\pi/2$, we have

$$f_X(x) = \begin{cases} \frac{1}{\pi}; & -\frac{\pi}{2} \le x \le \frac{\pi}{2} \\ 0; & \text{otherwise} \end{cases}$$

$$\therefore \qquad \Phi_Y(\omega) = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{j\omega \sin x}\,dx$$

Since $y = \sin x$, $dy = \cos x dx$ and when $x = -\pi/2$, $y = -1$ and when $x = \pi/2$, $y = 1$

$$\therefore \qquad \Phi_Y(\omega) = \frac{1}{\pi}\int_{-1}^{1} e^{j\omega y}\frac{1}{\sqrt{1-y^2}}\cdot dy \quad \text{But} \quad \Phi_Y(\omega) = E[e^{j\omega y}] = \int_{-\infty}^{\infty} e^{j\omega y} f_Y(y)\,dy$$

$\therefore$ we find that $f_Y(y) = \dfrac{1}{\pi\sqrt{1-y^2}}$ for $|y| \le 1$ and zero otherwise.

Example 6.13 X and Y are two independent zero-mean Gaussian random variables with variance σ^2. We define another pair of random variables r and θ in terms of X and Y as follows:

$$r = \sqrt{X^2 + Y^2}; \theta = \tan^{-1}(Y/X) \text{ where } |\theta| < \pi$$

Obtain the joint density function of r and θ. Also obtain their marginal densities.

Solution $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$ and $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2}$

Since X and Y are given to be independent random variables, their joint density is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

Now, we are given that $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$

$\therefore$ one solution is

$$x_1 = r\cos\theta \text{ and } y_1 = r\sin\theta$$

Then the Jacobian $$J(r, \theta) = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial y_1}{\partial r} & \frac{\partial y_1}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

$\therefore$ $$f_{r,\theta}(r, \theta) = r f_{xy}(x_1, y_1) = \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2}; 0 < r < \infty \;\&\; |\theta| < \pi$$

This is their joint density function. To obtain the marginal density function of r we integrate $f_{r,\theta}(r, \theta)$ for all values of θ from $-\pi$ to $+\pi$. Similarly, to get the marginal density of θ, we integrate $f_{r,\theta}(r, \theta)$ w.r.t. r from $r = 0$ to $r = \infty$.

$\therefore$ $$f_r(r) = \int_{-\pi}^{\pi} f_{r,\theta}(r, \theta)\, d\theta = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}; 0 < r < \infty$$

$$f_\theta(\theta) = \int_0^{\infty} f_{r,\theta}(r, \theta)\, dr = \frac{1}{2\pi}; |\theta| < \pi$$

$\therefore$ r has Rayleigh density function while θ is uniformly distributed between $-\pi$ and $+\pi$. Also, since $f_{r,\theta}(r, \theta) = f_r(r) \cdot f_\theta(\theta)$, we find that r and θ are statistically independent.

6.10 RANDOM PROCESSES

Earlier in Section 6.4, we had defined a random variable X as the rule according to which we could assign a real number to each outcome, ξ, of a random experiment. Thus, we define the random variable as a function of ξ and denoted it by $X(\xi)$, or simply, by X. Now, if with every outcome ξ of a random experiment, we associate a time signal instead of a number, we get a family of time signals, each one associated with one outcome ξ and this family of time signals is called a random process and is shown in Fig. 6.9.

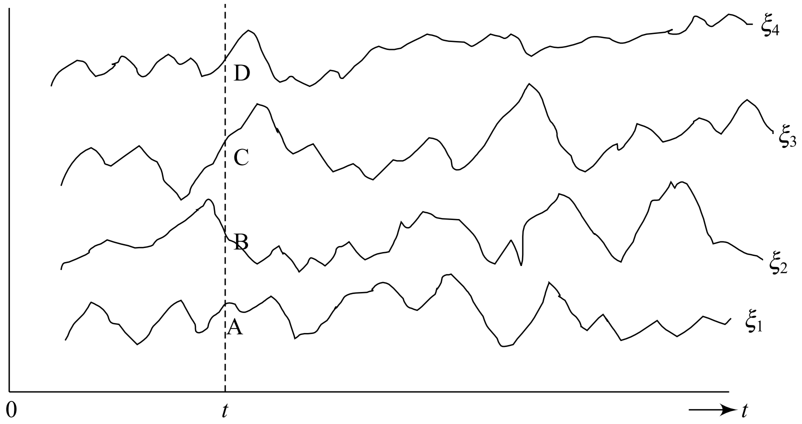


Fig. 6.9 *Random process as family of time signals, one for each* ξ

Thus, the random process is a function of two variables, time t and outcome ξ, and is therefore denoted by $\boldsymbol{X}(t,\xi)$. However, for notational simplicity, we generally omit the ξ and represent a random process simply by $\boldsymbol{X}(t)$.

Now, if $t \in R$, the set of all real numbers, the random process is called a continuous random process and if $t \in I$, the set of all integers, the process is called a discrete random process. We shall be discussing continuous random processes only. Hence, unless specifically stated otherwise, by a 'random process', we mean only a continuous random process.

From the foregoing, it is clear that

1. When ξ is fixed and t is a variable, $\boldsymbol{X}(t, \xi)$ represents a single time signal corresponding to that ξ, or what is generally called, a single realization of the process;
2. When t is fixed and ξ is a variable, $\boldsymbol{X}(t, \xi)$ represents a set of real numbers (as shown in Fig. 6.8), one for each ξ and hence $\boldsymbol{X}(t, \xi)$ in this case, is just a random variable;
3. When both t and ξ are fixed, $\boldsymbol{X}(t, \xi)$ represents a mere number;
4. When both t and ξ are variables, $\boldsymbol{X}(t, \xi)$ represents a family of time signals and is a random process.

A simple example of a random process is perhaps a sinusoid with a random phase.

6.10.1 First and Second Order Statistics

Since the random process becomes a random variable when t is fixed, we can talk about the distribution and density functions of a process in terms of those of a random variable. For a particular fixed value of t, $\boldsymbol{X}(t)$ is a random variable and its distribution is

$$F(x,t) = P[\boldsymbol{X}(t) \le x] \tag{6.64}$$

The derivative with respect to x of this first-order distribution function, $F(x,t)$ of the process $\boldsymbol{X}(t)$,

$$f(x,t) = \frac{\partial}{\partial x}[F(x,t)] \tag{6.65}$$

is referred to as the first-order density function of the process $\boldsymbol{X}(t)$.

On the same lines, we define the joint distribution function of the random variables $\boldsymbol{X}(t_1, \xi)$ and $\boldsymbol{X}(t_2, \xi)$ obtained by considering the process at the two fixed instants of time t_1 and t_2, as the second-order distribution function and it is

$$F(x_1, x_2; t_1, t_2) = P[\boldsymbol{X}(t_1) \le x_1, \boldsymbol{X}(t_2) \le x_2] \tag{6.66}$$

The second-order density function is

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2}{\partial x_1 \partial x_2}[F(x_1, x_2; t_1, t_2)] \tag{6.67}$$

Of course, as usual, we must have the first-order statistics from the second-order statistics; i.e.,

$$f(x_1, t_1) = [F(x_1, \infty; t_1, t_2)] \quad \text{and} \quad f(x_1, t_1) = \int_{-\infty}^{\infty} f(x_1, x_2; t_1, t_2)\, dx_2$$

Mean Proceeding on the same lines, we define the mean of the random process $\boldsymbol{X}(t)$ as the mean of the random variable $\boldsymbol{X}(t)$

$$\therefore \qquad \eta_X(t) = E[\boldsymbol{X}(t)] = \int_{-\infty}^{\infty} x \cdot f(x, t)\, dx \tag{6.68}$$

So, the mean of $\boldsymbol{X}(t)$ is a deterministic function of time and at any instant of time t_0, it equals the mean of the random variable $\boldsymbol{X}(t_0)$.

Auto-correlation The auto-correlation $R_X(t_1, t_2)$ of a random process $\boldsymbol{X}(t)$ is a *deterministic function* of two variables t_1 and t_2 and is defined as the expected value of the product of the random variables $\boldsymbol{X}(t_1)$and $\boldsymbol{X}(t_2)$.

$$\begin{aligned} R_X(t_1, t_2) &= E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)] \text{ if } \boldsymbol{X}(t) \text{ is a real process} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2)\, dx_1 dx_2 \end{aligned} \tag{6.69}$$

and if $\boldsymbol{X}(t)$ is a complex valued process,

$$R_X(t_1, t_2) = E[\boldsymbol{X}(t_1)\boldsymbol{X}^*(t_2)] \tag{6.70}$$

where the * indicates complex-conjugation

Auto-covariance The auto-covariance of the process $\boldsymbol{X}(t)$ is the co-variance of the two random variables $\boldsymbol{X}(t_1)$ and $\boldsymbol{X}(t_2)$ and is denoted by $C_X(t_1, t_2)$

$$\begin{aligned} \therefore \qquad C_X(t_1, t_2) &= E\{[\boldsymbol{X}(t_1) - \eta_X(t_1)][\boldsymbol{X}(t_2) - \eta_X(t_2)]\} \text{ for a real process} \\ &= R_X(t_1, t_2) - \eta_X \eta_Y \text{ for a real process} \end{aligned}$$

$$\text{and} \qquad C_X(t_1, t_2) = E\{[\boldsymbol{X}(t_1) - \eta_X(t_1)][\boldsymbol{X}(t_2) - \eta_X(t_2)]^*\} \text{ for a complex process}$$

$$\therefore \qquad C_X(t_1, t_2) = R_X(t_1, t_2) - \eta_X(t_1)\eta_X^*(t_2) \text{ for a complex process} \tag{6.71}$$

Note

(i) $R_X(t, t) = E[|\boldsymbol{X}(t)|^2]$ = *Average power in* $\boldsymbol{X}(t)$.

(ii) $C_X(t, t) = E[|\boldsymbol{X}^2(t)|] - \{E[\boldsymbol{X}(t)]\}^2$ = *variance of* $\boldsymbol{X}(t)$.

Cross-correlation and cross-covariance If we have two random processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$, their cross-correlation is defined as

$$R_{XY}(t_1, t_2) = E[X(t_1)\boldsymbol{Y}^*(t_2)] = R_{YX}^*(t_2, t_1) \tag{6.72}$$

The cross-covariance of the two processes is defined as

$$\begin{aligned} C_{XY}(t_1, t_2) &= E\{[\boldsymbol{X}(t_1) - \eta_X(t_1)][\boldsymbol{Y}(t_2) - \eta_Y(t_2)]^*\} \\ \therefore \qquad &= R_{XY}(t_1, t_2) - \eta_X(t_1)\eta_Y^*(t_2) \end{aligned} \tag{6.73}$$

(In Eqs. (6.72) and (6.73), complex conjugation can be ignored if the processes are real-valued).

6.10.2 Independent Processes

The two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be statistically independent processes if the set of random variables $\{\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)\}$ and $\{\boldsymbol{Y}(t_1'), \boldsymbol{Y}(t_2'), \ldots \boldsymbol{Y}(t_n')\}$ are mutually independent for all values of $t_1, t_2, \ldots, t_n, t_1', t_2', \ldots t_n'$ and all integer values of n.

Uncorrelated processes Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be uncorrelated processes if

$$C_{XY}(t_1, t_2) = 0 \text{ for all values of } t_1 \text{ and } t_2 \tag{6.74}$$

Orthogonal processes Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be orthogonal processes if

$$R_{XY}(t_1, t_2) = 0 \text{ for all values of } t_1 \text{ and } t_2 \tag{6.75}$$

Note *If two processes are orthogonal and in addition if any one of them, or both have zero mean, then the two processes will be uncorrelated.*

Example 6.14 $\boldsymbol{X}(t) = A\cos(\omega t + \phi)$, where ϕ is a random variable uniformly distributed between $-\pi$ and $+\pi$. Determine the mean and auto-correlation of $\boldsymbol{X}(t)$.

Solution

(a) $$\eta_X(t) = E[\boldsymbol{X}(t)] = \int_{-\infty}^{\infty} \boldsymbol{X}(t)\cdot f_\phi(\phi)\,d\phi$$

$$= \frac{A}{2\pi}\int_{-\infty}^{\infty} \cos(\omega t + \phi)\,d\phi = 0$$

(b) $$R_X(t_1, t_2) = E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)] = E[A^2\cos(\omega t_1 + \boldsymbol{\phi}) + \cos(\omega t_2 + \boldsymbol{\phi})]$$

$$= A^2 E[\cos(\omega t_1 + \boldsymbol{\phi}) + \cos(\omega t_2 + \boldsymbol{\phi})]$$

$$= \frac{1}{2}A^2 E[\cos\omega(t_1 - t_2) + \cos(\omega t_1 + \omega t_2 + 2\phi)]$$

$$= \frac{1}{2}A^2\cos\omega(t_1 - t_2)$$

Example 6.15 If $\boldsymbol{X}(t) = \boldsymbol{a}e^{j\omega t}$, determine its auto-correlation.

Solution $$R_X(t_1, t_2) = E[X(t_1)X^*(t_2)] = E[\boldsymbol{a}e^{j\omega t_1} \,.\, \boldsymbol{a}^* e^{-j\omega t_2}]$$

$$= E[|\boldsymbol{a}|^2]e^{j\omega(t_1 - t_2)}$$

6.11 STATIONARITY, AUTO-CORRELATION AND POWER SPECTRUM

As we have seen till now, the statistical properties of a random process – like its mean, auto-correlation, etc., are in general dependent upon time. However, there is an important class of random processes, whose statistical properties are independent of time. These processes are called *stationary processes*.

There are different levels of stationarity – strict-sense stationarity, *k*th-order stationarity, wide-sense stationarity, etc.

Definition A strict-sense stationary process $\boldsymbol{X}(t)$ is one whose density function of any order is independent of time; i.e.,

$$f_X(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) = f_X(x_1, x_2, \ldots, x_n; t_1 + \in, t_2 + \in, \ldots, t_n + \in) \tag{6.76}$$

For any integer n and any real number $\in$

If Eq. (6.72) is true only up to $n \le K$, then the process $\boldsymbol{X}(t)$ is said to be K-th order stationary.

Strict stationarity is a very restrictive condition and most of the processes are not stationary in the strict sense. Wide-sense stationarity, on the other hand, is a less restrictive one, and is satisfied by many of the processes of interest.

Definition A random process $\boldsymbol{X}(t)$ is said to be wide-sense stationary (i.e., WSS), if it satisfies the following conditions:

1. Its mean, $\eta_X(t) = E[\boldsymbol{X}(t)]$ is independent of time.
2. Its auto-correlation function $R_X(t_1, t_2)$ is a function only of $\tau = (t_1 - t_2)$ and not of t_1 and t_2 individually.

$$\therefore \qquad R_X(t_1, t_2) = R_X(\tau) = E[\boldsymbol{X}(t+\tau)\boldsymbol{X}^*(t)]$$

When $\tau = 0$, $R_X(0) = E[|X(t)|^2]$ = average power of $\boldsymbol{X}(t)$, and the power is independent of time.
As we will henceforth be dealing only with WSS processes, unless specifically stated otherwise, the term 'process' would be assumed to mean a WSS process only.

Properties of auto-correlation function

1. It is deterministic.
2. It takes maximum value when $\tau = 0$.
3. $R_X(0)$ = average power of the process.
4. For real, process $\boldsymbol{X}(t)$, $R_X(-\tau) = R_X(+\tau)$, i.e., $R_X(\tau)$ has even symmetry.
5. For a complex process $R_X(-\tau) = R_X^*(\tau)$ where * denotes complex conjugation.

Example 6.16 Show that the random process $\boldsymbol{X}(t)$ $X(t) = A\cos(\omega t + \phi)$, where ϕ is a random variable uniformly distributed over $-\pi$ to $+\pi$, is WSS.

Solution It has already been shown in Example 6.14 that $\eta_X(t) = 0$ and hence, is independent of time. It has also been shown that its auto-correlation $R_X(t_1, t_2)$ is given by

$$R_X(t_1, t_2) = \frac{1}{2}A^2 \cos\omega(t_1 - t_2)$$

$\therefore$ $R_X(t_1, t_2)$ is a function only of $t_1 - t_2 = \tau$ and not of the individual values of t_1 and t_2.
Thus, $\boldsymbol{X}(t)$ satisfies the two conditions required to be satisfied by a process to be WSS.

6.11.1 Ergodicity

We have seen that the mean η_X of a random process is given by the ensemble average $E[\boldsymbol{X}(t)]$ of the process. Referring to Fig. 6.8, the ensemble average $E[\boldsymbol{X}(t)]$ is the mean of the values A, B, C and D, i.e., the average of the values of $\boldsymbol{X}(t, \xi)$ at a fixed t and for all possible values of ξ. Hence, to find the ensemble average of the process $\boldsymbol{X}(t)$, we should have all of its realizations available to us. Since the auto-correlation also involves ensemble average, its determination also requires all the realizations of $\boldsymbol{X}(t)$ to be available. *In fact, determination of any statistical average of a process requires that all the realizations of it be available.*

However, in practice, whenever we observe a random process, it is only one realization of it which we observe. In practice, therefore, it is not possible for us to have all the realizations of the process – i.e., *it is not possible in practice to determine the ensemble average of a process*. The only thing we can possibly do is to try to determine the time-average of the single realization that we observe. Even this single realization also we can observe only for a limited period of time, certainly not from minus infinity to plus infinity. However, it is pertinent to examine whether we can *at least estimate* the ensemble averages from the time averages. ***Random processes, for which the time averages equal the ensemble averages are known as 'ergodic processes'***. However, it must be noted that a process may be ergodic for statistics up to a particular order only. For instance, the process may be ergodic in mean but may not be ergodic in auto-correlation.

Example 6.17 Show that the process $X(t) = A_c \cos(\omega_0 t + \theta)$ where θ is uniformly distributed over $-\pi$ to $+\pi$, is ergodic in mean and auto-correlation.

Solution We have already seen in Example 6.16 that it is WSS and that

$$\eta_X(t) = E[X(t)]$$

and that

$$R_X(\tau) = \frac{1}{2} A_c^2 \cos \omega\tau$$

Now, we shall find the η_X and $R_X(\tau)$ by time averaging and show that we get the same result for η_X and $R_X(\tau)$.

$$\eta_X = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} A_c \cos(\omega_0 t + \theta)\, dt = \lim_{T\to\infty} \frac{A_c}{T\omega_0} \left[\sin(\omega_0 t + \theta) \Big|_{-T/2}^{T/2} \right] = 0$$

$\therefore$ it is ergodic in mean

$$R_X(\tau) = \lim_{T\to\infty} \frac{A_c^2}{T} \int_{-T/2}^{T/2} \cos\{\omega_0(t+\tau) + \theta\} \cos(\omega_0 t + \theta)\, dt$$

$$= \lim_{T\to\infty} \frac{A_c^2}{T} \left[\cos^2\theta \int_{-T/2}^{T/2} \cos\alpha \cos\beta dt \right] - \lim_{T\to\infty} \frac{A_c^2}{2T} \sin 2\theta \left[\int_{-T/2}^{T/2} \sin\alpha \cos\beta dt \right]$$

$$- \lim_{T\to\infty} \frac{A_c^2}{2T} \sin 2\theta \left[\int_{-T/2}^{T/2} \cos\alpha \sin\beta dt \right] + \lim_{T\to\infty} \frac{A_c^2}{T} \sin^2\theta \left[\int_{-T/2}^{T/2} \sin\alpha \sin\beta dt \right]$$

where $\alpha \triangleq \omega_0(t+\tau)$ and $\beta = \omega_0 t$

Replacing $\sin^2\theta$ of the last term by $(1 - \cos^2\theta)$ and simplifying, all the terms vanish except

$$\lim_{T\to\infty} \frac{A_c^2}{2T} \int_{-T/2}^{T/2} \cos(\alpha - \beta)\, dt$$

But this equals $A_c^2 \cos \omega_0 \tau$

Hence, the expected value and time average value of $[X(t+\tau)X(t)]$ are same.

$\therefore$ the given $X(t)$ is ergodic in auto-correlation.

6.11.2 Power Spectral Density of a Random Process

Definition The power spectral density, or simply, the power spectrum of a random process (real or complex) is the Fourier transform of its auto-correlation. (This is generally referred to as Wiener–Khinchin theorem)

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \tag{6.77}$$

$\therefore$

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df \tag{6.78}$$

Since in general, $R_X(-\tau) = R_X^*(\tau), S_X(f)$, the PSD is always a real-valued function of frequency. Further, if $X(t)$ is a real process, $R_X(\tau)$ is real and also even with respect to τ. Hence, its Fourier transform $S_X(f)$ will also be real and even.

If $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are two processes, we define their cross-correlation and cross-spectral density as follows:

$$\text{Cross-correlation} = R_{XY}(\tau) \underline{\Delta} E[\boldsymbol{X}(t+\tau)\boldsymbol{Y}^*(t)] \tag{6.79}$$

We now define the cross-power spectrum, or cross-spectral density $S_X(f)$ of $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ as the Fourier transform of $R_{XY}(\tau)$.

i.e,
$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j2\pi f\tau}d\tau \tag{6.80}$$

and
$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(f)e^{j2\pi f\tau}df \tag{6.81}$$

Since $R_{XY}(-\tau) = R_{YX}^*(\tau)$, the cross-spectral density is, in general, a complex function of f even if both the processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are real-valued.

Example 6.18 Determine the power spectrum of the processes

$\boldsymbol{X}(t) = A_c \cos(\omega_0 t + \theta)$; θ is uniformly distributed over $(-\pi, \pi)$.

Solution In Example 6.17, we obtained the ACF of this $\boldsymbol{X}(t)$ as

$$R_X(\tau) = \frac{A_c^2}{2}\cos\omega\tau$$

∴ the power spectrum, which is the Fourier transform of $R_X(\tau)$ is

$$S_X(f) = \frac{A_c^2}{2}\int_{-\infty}^{\infty}\cos\omega_0 te^{-j2\pi f\tau}d\tau$$

But we know that $\mathcal{F}[\cos\omega_0\tau] = \frac{1}{2}[\delta(f+f_0)+\delta(f-f_0)]$

∴
$$S_X(f) = \frac{A_c^2}{4}[\delta(f+f_0)+\delta(f-f_0)]$$

6.11.3 Gaussian Processes

A random process $\boldsymbol{X}(t)$ is said to be a Gaussian random process, if the random variables $\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots., \boldsymbol{X}(t_n)$ are jointly Gaussian for all n and all $t_1, t_2, \ldots, t_n$.

Properties of Gaussian processes

1. A Gaussian process is completely described by its mean and auto-correlation.
2. If a Gaussian process is wide sense stationary, then it is stationary in the strict sense too.
3. If a Gaussian process is given as input to an LTI system, the output process also is Gaussian.
4. If two processes which are jointly Gaussian are uncorrelated, then they are statistically independent.

Gaussian processes are very important in communication engineering. This is mainly because of the fact that thermal noise, which plays a key role in communications, can be closely modeled by a Gaussian process. In addition, some of the information sources also can be modeled as Gaussian processes.

White Noise Process

Definition A process $\boldsymbol{X}(t)$ whose power spectral density is a constant for all frequencies, is called a *white process.*

The PSD of a white process is sketched in Fig. 6.10. As shown in the figure, it has a constant value $N_0/2$ for all frequencies.

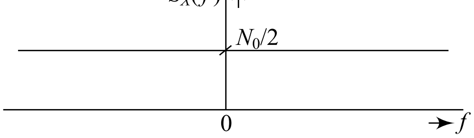


Fig. 6.10 *Power spectral density of a white process*

Since the area under any PSD curve is equal to the total average power of the process, a constant PSD makes a white process to have an infinite average power. Thus in practice, there cannot be any source producing a perfect white process. Every so-called white process has a power spectral density that tends towards zero at some frequency, although it might remain constant (or almost constant) up to that frequency. Although no source can produce a white process, the concept of a 'white process' is, nevertheless, quite useful. This is because, if the PSD is constant up to a very high frequency which is far beyond the frequencies at which any practical communication system operates, then, insofar as our communication systems are concerned, we can safely assume that the PSD of the process is absolutely constant, i.e., the process is a white process. It is in this sense that we say that thermal noise is white, although we know that its PSD tends to fall off beyond approximately 10^{12} Hz.

Auto-correlation of a white process Since the auto-correlation is the inverse Fourier transform of the power spectral density, a white process with a PSD of $N_0/2$ will have an auto-correlation of

$$R_n(\tau) = \frac{N_0}{2}\delta(\tau) \tag{6.82}$$

Since the auto-correlation is an impulse function, it means that no two samples of a white process will have any correlation, however close (in terms) the two samples may be. That is why we call the white process as '*white noise*'.

6.12 LTI SYSTEMS WITH RANDOM PROCESSES AS INPUTS

In this section, we will be discussing how the mean and auto-correlation of the output process may be determined in terms of those of the input process and the impulse response of the LTI system. Of particular interest is the relationship between the PSD of the output process and PSD of the input process. Before we can talk about the power spectrum of the output, it is of course necessary to examine whether the output process will also be stationary if the input process is.

Definition Two processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be jointly stationary if they are individually stationary and if their cross-correlation $R_{XY}(t_1, t_2)$ is a function only of $\tau = (t_1 - t_2)$ and not individually of t_1 and t_2.

Let us give a stationary process $\boldsymbol{X}(t)$ as input to an LTI system with impulse response, $h(t)$. Let the output process be $\boldsymbol{Y}(t)$. We shall now show that the input and output processes are jointly stationary and that

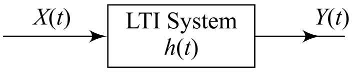


Fig. 6.11 *An LTI system with a random process as input*

1. Mean of the output = $\eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)\,dt$ a constant independent of time. (6.83)
2. Cross-correlation of input and output processes $= R_{XY}(\tau) = R_X(\tau) * h(-\tau)$ (6.84)
3. Correlation of output process $= R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$ (6.85)

1. We know that

$$\boldsymbol{Y}(t) = \boldsymbol{X}(t) * h(t) = \int_{-\infty}^{\infty} \boldsymbol{X}(t-u)h(u)\,du$$

Taking the expectation on both sides, we have

$$E[\boldsymbol{Y}(t)] = \eta_Y(t) = E\left[\int_{-\infty}^{\infty} \boldsymbol{X}(t-u)h(u)\,du\right]$$

$$= \int_{-\infty}^{\infty} E[\boldsymbol{X}(t-u)]h(u)\,du$$

Since $\boldsymbol{X}(t)$ is stationary,

$$= E[\boldsymbol{X}(t-u)] = E[\boldsymbol{X}(t)] = \eta_X = \text{a constant}$$

$$\therefore \qquad \boxed{E[\boldsymbol{Y}(t)] = \eta_X \int_{-\infty}^{\infty} h(t)\,dt = \text{a constant independent of } t.}$$

2. $R_{XY}(t_1, t_2) \underline{\Delta}\, E[\boldsymbol{X}(t_1)\boldsymbol{Y}(t_2)]$

$$\therefore \qquad R_{XY}(t_1, t_2) = E\left[\boldsymbol{X}(t_1)\int_{-\infty}^{\infty}\boldsymbol{X}(t_2-u)h(u)\,du\right]$$

$$= \int_{-\infty}^{\infty} E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2-u)]h(u)\,du$$

If we now put $u = -\lambda$,

$$R_{XY}(t_1, t_2) = \int_{-\infty}^{\infty} R_X(t_2 - t_2 + u)h(u)\,du = \int_{-\infty}^{\infty} R_X(\tau + u)h(u)\,du$$

$$= \int_{-\infty}^{\infty} R_X(\tau - \lambda)h(\lambda)\,d\lambda = R_X(\tau) * h(-\tau)$$

$$\therefore \qquad \boxed{R_{XY}(t_1, t_2) = R_X(\tau) * h(-\tau)} \tag{6.86}$$

$\therefore$ the cross-correlation is a function only of $\tau = (t_1 - t_2)$

3. To find the auto-correlation of the output process

$$R_Y(t_1, t_2) = E[\boldsymbol{Y}(t_1)\boldsymbol{Y}(t_2)]$$

But

$$\boldsymbol{Y}(t_1) = \int_{-\infty}^{\infty} \boldsymbol{X}(u)h(t_1-u)\,du$$

$$\therefore \qquad R_Y(t_1, t_2) = E\left[\left\{\int_{-\infty}^{\infty}\boldsymbol{X}(u)h(t_1-u)\,du\right\}\boldsymbol{Y}(t_2)\right]$$

$$= \int_{-\infty}^{\infty} E\{\boldsymbol{X}(u)\boldsymbol{Y}(t_2)\}h(t_1-u)\,du$$

$$= \int_{-\infty}^{\infty} R_{XY}(u-t_2)h(t_1-u)\,du$$

If we put $(t_1 - u) = \lambda$, we get $u = t_1 - \lambda; = -d\lambda$

$$\therefore \qquad R_Y(t_1, t_2) = \int_{-\infty}^{\infty} R_{XY}(t_1 - t_2 - \lambda)h(\lambda)\,d\lambda$$

$$\therefore \qquad R_Y(\tau) = \int_{-\infty}^{\infty} R_{XY}(\tau - \lambda)h(\lambda)\,d\lambda = R_{XY}(\tau) * h(\tau) \tag{6.87}$$

$\therefore R_Y(t_1, t_2)$, the auto-correlation of the output process $\boldsymbol{Y}(t)$ is a function only of $\tau = (t_1 - t_2)$, but not individually of t_1 and t_2. As we have already shown that its mean is independent of t, it means that the process $\boldsymbol{Y}(t)$ is stationary (WSS). Further, we have shown that $R_{XY}(t_1, t_2) = R_{XY}(\tau)$. Hence, it follows that the input and output processes are jointly stationary.

Substituting for $R_{XY}(\tau)$ in Eq. (6.87) using Eq. (6.86), we get

$$\boxed{R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)}$$

6.12.1 Input and Output Spectra and Cross-Power Spectrum

Equations (6.83), (6.84) and (6.85) give us the output mean, the output auto-correlation and the input-output cross-correlation respectively in terms of the input quantities and the impulse response of the LTI system.

Now, to get the relationships in the frequency domain, let us take the Fourier transforms of these equations.

1. From Eq. (6.83), we have

$$\eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)\,dt = \eta_X \qquad \text{(Area under the impulse response)}$$

But we know that

$$\int_{-\infty}^{\infty} h(t)\,dt = \left[\int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\,dt\right]\Bigg|_{f=0} = H(f)\Big|_{f=0} = H(0)$$

$$\therefore \qquad \boxed{\eta_Y = \eta_X H(0)} \tag{6.88}$$

2. From Eq. (6.84), we have

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

Taking Fourier transform on both sides and noting that

$$\mathcal{F}[h(-\tau)] = H^*(f) \tag{6.89}$$

$$\boxed{S_{XY}(f) = S_X(f)\cdot H^*(f)} \tag{6.90}$$

3. From Eq.(6.85), we have

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$

$\therefore$ taking Fourier transform of this on both sides

$$S_Y(f) = S_X(f)\cdot H(f)\cdot H^*(f)$$

$$\therefore \qquad \boxed{S_Y(f) = S_X(f)\cdot |H(f)|^2} \tag{6.91}$$

This is a very important result and is used quite frequently in communication engineering.

Example 6.19 An ideal differentiator is an LTI system. If a WSS process $X(t)$ of mean η_X and auto-correlation $R_X(\tau)$ is given as input to it, determine the mean and the power spectrum of the output.

Solution From Eq. (6.88), we have $\eta_Y = \eta_X H(0)$

But $H(f)$ of an ideal differentiator $= j2\pi f$

$\therefore$ $H(0) = 0$. It then follows that $\eta_Y = 0$

From Eq. (6.91), we have

$$S_Y(f) = S_X(f)\cdot |H(f)|^2$$

$$|H(f)|^2 = H(f)H^*(f) = j2\pi f\cdot(-j2\pi f) = 4\pi^2 f^2$$

$$S_Y(f) = 4\pi^2 f^2\cdot S_X(f)$$

Here, $S_X(f) = \mathcal{F}[R_X(\tau)]$

$\therefore$ $S_y(f) = 4\pi^2 f^2[\mathcal{F}\{R_X(\tau)\}] = 4\pi^2 f^2 S_X(f)$

6.13 REPRESENTATION OF BAND-LIMITED AND BAND PASS PROCESSES

6.13.1 Band-Limited Processes

In the case of a deterministic signal $x(t)$ which is low pass and band limited to say W Hz, i.e., $X(f) = 0$ for $|f| \geq W$, we know from the low pass sampling theorem for deterministic signals, that if $x(t)$ is sampled at regular intervals of T_s where $T_s \leq \frac{1}{2W}$, the samples so obtained completely represent the band limited deterministic signal $x(t)$ and that in fact $x(t)$ can be expanded as follows in terms of these samples and an infinite set of sinc functions displaced in time with respect to each other by T_s.

$$x(t) = \frac{2W}{f_s} \sum_{k=-\infty}^{+\infty} x(kT_s) \operatorname{sinc} 2W(t - kT_s);\ f_s = 1/T_s, -\infty < t < \infty \tag{6.92}$$

In the particular case when $T_s = \frac{1}{2W}$, this equation reduces to

$$x(t) = \sum_{k=-\infty}^{+\infty} x(k/2W) \operatorname{sinc} (2Wt - k);\ -\infty < t < \infty \tag{6.93}$$

The equality sign in the above equations holds at all instants of time, i.e., it holds point-wise.

Since the signals as well as noise that we have to deal with in communications are random processes, it will be of interest to examine whether a band-limited low pass process also could be represented by its samples, or, in short, whether a similar low pass sampling theorem exists in the case of random processes too. Fortunately there is a *similar* theorem applicable to stationary low pass band-limited processes, and it states as follows.

Theorem If $\boldsymbol{X}(t)$ is a stationary low pass process which is band-limited to W Hz, i.e., if $S_X(f) = 0$ for $|f| \geq W$ Hz, and if it is sampled at regular intervals of T_s where $T_s = \frac{1}{2W}$, then

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \operatorname{sinc} 2W(t - kT_s)\right|^2\right] = 0 \tag{6.94}$$

Equation (6.94) implies that under the conditions stated in the theorem, $\boldsymbol{X}(t)$ is equal, *in the mean-square sense*, to

$$\sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \operatorname{sinc} 2W(t - kT_s)$$

Proof To prove Eq. (6.94), let us first expand the LHS of it. Writing down term by term and *assuming the process $\boldsymbol{X}(t)$ to be real*, we get

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \operatorname{sinc} 2W(t - kT_s)\right|^2\right] = E[\boldsymbol{X}^2(t)] - 2\sum_{k=-\infty}^{+\infty} E[\boldsymbol{X}(T)\boldsymbol{X}(kT_s)] \operatorname{sinc} 2W(t - kT_s)$$

$$+ \sum_k \sum_l E[\boldsymbol{X}(kT_s)\boldsymbol{X}(lT_s)] \operatorname{sinc} 2W(t - kT_s) \operatorname{sinc} 2W(t - lT_s)$$

But $E[\boldsymbol{X}^2(t)] = R_X(0)$ and $E[\boldsymbol{X}(kT_s)\boldsymbol{X}(lT_s)] = R_X(kT_s - lT_s)$.

Now, if we put $m = l - k$, $l = m + k$ and LHS of Eq. (6.94) equal

$$R_X(0) - 2\sum_{k=-\infty}^{+\infty} R_X(t - kT_s)\,\text{sinc}\,2W(t - kT_s)$$

$$+\sum_k\sum_m R_X(-mT_s)\,\text{sinc}\,2W(t - kT_s)\,\text{sinc}\,2W(t - kT_s - mT_s) \tag{6.95}$$

But $\sum_k\sum_m R_X(-mT_s)\,\text{sinc}\,2W(t - kT_s)\,\text{sinc}\,2W(t - kT_s - mT_s)$

$$= \sum_k \text{sinc}\,2W(t - kT_s)\sum_m R_X(-mT_s)\,\text{sinc}\,2W(t - kT_s - mT_s) \tag{6.96}$$

However, since $\boldsymbol{X}(t)$ is a real process, its ACF has even symmetry

$$\therefore \qquad R_X(-mT_s) = R_X(mT_s) \tag{6.97}$$

Further, since $\boldsymbol{X}(t)$ is band limited to W Hz, it means that its ACF, $R_X(t)$, which is a deterministic function, has a Fourier transform, $S_X(f)$, which equals zero for all $|f| \geq W$. Hence, as per the low pass sampling theorem for band limited deterministic signals, using Eq. (6.93), we may expand $R_X(t)$ in terms of its samples, as follows

$$R_X(t) = \sum_{m=-\infty}^{+\infty} R_X(mT_s)\,\text{sinc}\,2W(t - mT_s) \tag{6.98}$$

$$\therefore \qquad R_X(t - kT_s) = \sum_{m=-\infty}^{+\infty} R_X(mT_s)\,\text{sinc}\,2W(t - kT_s - mT_s) \tag{6.99}$$

$\therefore$ using Eqs. (6.97) and (6.99), the RHS of Eq. (6.96) may be written as

$$= \sum_{k=-\infty}^{\infty} R_X(t - kT_s)\,\text{sinc}\,2W(t - kT_s)$$

Hence, Eq. (6.95) may be modified as

$$R_X(0) - 2\sum_{k=-\infty}^{+\infty} R_X(t - kT_s)\,\text{sinc}\,2W(t - kT_s) + \sum_{k=-\infty}^{\infty} R_X(t - kT_s)\quad \text{sinc}\,2W(t - kT_s)$$

$$\therefore \qquad \text{LHS of Eq. (6.90)} = R_X(0) - \sum_{k=-\infty}^{+\infty} R_X(t - kT_s)\quad \text{sinc}\,2W(t - kT_s) \tag{6.100}$$

Now, $R_X(t)$ is a deterministic signal which is band limited to W Hz since its FT, $S_X(f) = 0$ for $|f| \geq W$ Hz. But we know from the low pass sampling theorem for deterministic signals that

$$R_X(t) = \sum_{k=-\infty}^{+\infty} R_X(kT_s)\,\text{sinc}\,2W(t - kT_s) \tag{6.101}$$

In Eq. (6.101), it is assumed that the sampling at regular intervals of $T_s = \dfrac{1}{2W}$ is done in such a manner that there is a sample taken at $t = 0$ sec. Instead, if we have a sample $t = t_0$ and every T_s sec on either side of it, Eq. (6.101) gets modified and from that modified equation we may write (by putting $t = 0$) as

$$R_X(0) = \sum_{k=-\infty}^{+\infty} R_X(t_0 - kT_s)\,\text{sinc}\,[2W(t_0 - kT_s)]$$

Since t_0 can take any value, we may write

$$R_X(0) = \sum_{k=-\infty}^{+\infty} R_X(t - kT_s)\,\text{sinc}\,[2W(t - kT_s)]$$

Substituting this on the RHS of Eq. (6.95), we get

$$E\left|x(t) - \sum_{k=-\infty}^{+\infty} X(t - kT_s)\,\text{sinc}\,[2W(t - kT_s)]\right|^2 = 0$$

Thus, the theorem is proved.

6.13.2 Band Pass Processes

In Section 2.10.5, we had discussed in detail, the inphase and quadrature component representation of a deterministic band pass signal $x(t)$

i.e., $$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \tag{6.102}$$

where $X_I(t)$ and $X_Q(t)$ are low pass signals.

Let $X(t)$ be a stationary, zero-mean band pass process, whose power spectral density is of the form shown in Fig. 6.12.

We encounter such a band pass noise process when say white noise is filtered by a band pass filter. Similarly to the band pass signal case, we shall now define two low pass processes $X_I(t)$ and $X_Q(t)$, where

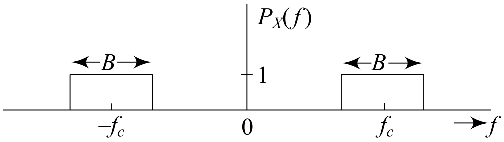


Fig. 6.12 *Power spectral density of a band pass process X(t)*

$$X_I(t) = X(t)\cos 2\pi f_c t + \hat{X}(t)\sin 2\pi f_c t \tag{6.103}$$

and $$X_Q(t) = \hat{X}(t)\cos 2\pi f_c t - X(t)\sin 2\pi f_c t \tag{6.104}$$

Here $\hat{X}(t)$ is the Hilbert transform of $X(t)$.

From Eqs. (6.103) and (6.104), it is easy to verify that

$$X_I(t)\cos 2\pi f_c t - X_Q(t)\sin 2\pi f_c t = X(t) \tag{6.105}$$

When $X(t)$ is a stationary, zero-mean band pass process, the inphase component process $X_I(t)$ and the quadrature component process $X_Q(t)$ have certain very important properties which we state below without proof. The proofs can be found from the references given at the end of the chapter.

Properties of $X_I(t)$ and $X_Q(t)$

1. $X_I(t)$ and $X_Q(t)$ are zero-mean, low pass, jointly stationary processes. If in addition $X(t)$ is Gaussian, then $X_I(t)$ and $X_Q(t)$ will be jointly Gaussian.
2. The inphase and quadrature components $X_I(t)$ and $X_Q(t)$ have the same average power as the process $X(t)$ itself.

   i.e., $$P_{X_I} = P_{X_Q} = P_X = \int_{-\infty}^{\infty} P_X(f)\,df \tag{6.106}$$

3. $X_I(t)$ and $X_Q(t)$ have identically the same power spectral density. This is obtained by shifting the positive frequency portion of $P_X(f)$ to the left by f_c, shifting the negative frequency portion of $P_X(f)$ to the right by f_c and then adding these two. Since the total area under a PSD curve gives the average power, from this, it is clear that the total average powers are the same for the inphase and quadrature components as well as the process $X(t)$ itself.

6.13.3 PDF of the Envelope of Narrowband Zero-Mean Gaussian Noise

Consider a narrowband zero-mean Gaussian noise process $n(t)$ centered on frequency f_c and having a variance σ^2. Using Eq. (6.108), we may represent it in terms of the inphase and quadrature components as

$$n(t) = n_I(t)\cos 2\pi f_c t - n_Q(t)\sin 2\pi f_c t \tag{6.107}$$

or, alternatively we may use the envelope and phase representation and write it as

$$n(t) = r(t)\cos[2\pi f_c t + \theta(t)] \tag{6.108}$$

where

$$r(t) = \sqrt{n_I^2(t) + n_Q^2(t)} \text{ and } \theta(t) = \tan^{-1}\left[\frac{n_Q(t)}{n_I(t)}\right] \tag{6.109}$$

We know that in these equations, $n_I(t)$, $n_Q(t)$, $r(t)$ and $\theta(t)$ are all low pass processes. The probability distribution of the envelope $r(t)$ can be obtained from those of $n_I(t)$ and $n_Q(t)$. For this purpose, let us take a snapshot of $n_I(t)$ and $n_Q(t)$ at some fixed instant of time. $n_I(t)$ and $n_Q(t)$ are zero-mean jointly Gaussian processes (see the properties of these processes) having the same average power σ^2. The random variable N_I and N_Q obtained by taking the snapshot of $n_I(t)$ and $n_Q(t)$ respectively, are zero-mean independent Gaussian random variables, each having a variance of σ^2. Since they are independent, their joint density function will be the product of their individual density functions. So we may write the joint density function as

$$f_{N_I N_Q}(n_I, n_Q) = \frac{1}{2\pi\sigma^2}\exp\left[\frac{n_I^2 + n_Q^2}{2\sigma^2}\right] \tag{6.110}$$

As we are interested in the density function of the envelope, i.e., in $f_{r,\theta}(r, \theta)$, we have to transform the coordinate system from the rectangular coordinate system used by the inphase and quadrature component representation into the circular, i.e., (r, θ) coordinate system. For this purpose, let us define

$$n_I = r\cos\theta \quad \text{and} \quad n_Q = r\sin\theta \tag{6.111}$$

Then the Jacobian of the coordinate system transformation is given by

$$J(r, \theta) = \begin{vmatrix} \dfrac{\partial n_I}{\partial r} & \dfrac{\partial n_I}{\partial \theta} \\ \dfrac{\partial n_Q}{\partial r} & \dfrac{\partial n_Q}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \tag{6.112}$$

Substituting in Eq. (6.110) for n_I aad n_Q by making use of Eq. (6.111) and using the above Jacobian

$$f_{r,\theta}(r, \theta) = \frac{r}{2\pi\sigma^2}\exp\left[-\frac{r^2}{2\sigma^2}\right] \tag{6.113}$$

Since $f_{r,\theta}(r, \theta)$ is independent of θ, it means that the random variables r and θ are statistically independent. It then follows that $f_\theta(\theta)$ is uniformly distributed over 0 to 2π and so has a value of $(1/2\pi)$ in that interval.

$$\therefore \quad f_\theta(\theta) = \begin{cases} \dfrac{1}{2\pi}; & 0 \le \theta \le 2\pi \\ 0; & \text{otherwise} \end{cases} \tag{6.114}$$

Since r and θ are independent,

$$f_{r,\theta}(r, \theta) = f_r(r) \cdot f_\theta(\theta) \tag{6.115}$$

Thus, $$f_r(r) = \begin{cases} \dfrac{r}{\sigma^2} \exp\left[-\dfrac{r^2}{2\sigma^2}\right]; & 0 \le r \le \infty \\ 0; & r < 0 \end{cases} \tag{6.116}$$

This density function is referred to as the *Rayleigh density function* and it is sketched in Fig. 6.7(a).

Thus, the envelope of a zero-mean Gaussian narrowband noise process is Rayleigh distributed, and the phase is uniformly distributed over 0 to 2π.

6.13.4 Envelope of Sine Wave plus Zero-Mean Gaussian Narrowband Noise

Consider the process *x*(*t*) given by

$$x(t) = A \cos 2\pi f_c t + n(t) \tag{6.117}$$

where *n*(*t*) is a zero-mean Gaussian narrowband noise with variance σ^2 and centered on the frequency f_c. Using the inphase and quadrature component representation of *n*(*t*),

$$\begin{aligned} x(t) &= A \cos 2\pi f_c t + n_I(t) \cos 2\pi f_c t - n_Q(t) \sin 2\pi f_c t \\ &= [A + n_I(t)] \cos 2\pi f_c t - n_Q(t) \sin 2\pi f_c t \\ &= n_I'(t) \cos 2\pi f_c t - n_Q(t) \sin 2\pi f_c t \end{aligned} \tag{6.118}$$

where $$n_I'(t) \triangleq A + n_I(t) \tag{6.119}$$

Since $n_I(t)$ is a zero-mean Gaussian low pass process with a variance of σ^2, and *A* is a constant, $n_I'(t)$ is a Gaussian low pass process with a mean of *A* and a variance of σ^2. Further, like $n_I(t)$, it is also statistically independent of the zero-mean Gaussian low pass process $n_Q(t)$.

x(*t*) of Eq. (6.120) may be represented in the envelope and phase form as

$$x(t) = r(t) \cos [2\pi f_c t + \psi(t)]$$

where $$r(t) = \sqrt{[n_I'(t)]^2 + n_Q^2(t)} \quad \text{and} \quad \psi(t) = \tan^{-1}\left[\frac{n_Q(t)}{n_I(t)}\right]$$

Let a snapshot of the low pass process $n_I'(t)$, $n_Q(t)$, $r(t)$ and $\psi(t)$ at a fixed instant of time give us the random variables $N_I'(t)$, N_Q, *R* and ψ, respectively. The Gaussian random variable $N_T'(t)$ has a mean of *A* and is independent of the Gaussian zero-mean random variable N_Q. So we may write their joint density function as

$$\begin{aligned} f_{N_I'N_Q}(n_I', n_Q) &= \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(n_I' - A)^2 + n_Q^2}{2\sigma^2}\right] \\ &= \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(n_I^2 + n_Q^2}{2\sigma^2}\right] \end{aligned} \tag{6.120}$$

In the present case, the Jacobian for transformation of the coordinate system is

$$J(r, \psi) = \begin{vmatrix} \dfrac{\partial n_I'}{\partial r} & \dfrac{\partial n_I'}{\partial \psi} \\ \dfrac{\partial n_Q}{\partial r} & \dfrac{\partial n_Q}{\partial \psi} \end{vmatrix} = \begin{vmatrix} \cos \psi & -r \sin \psi \\ \sin \psi & r \cos \psi \end{vmatrix} = r$$

∴ substituting ($r \cos \psi - A$) for n_I and $r \sin \psi$ for n_Q in Eq. (6.120) and using the above Jacobian, we get

$$\begin{aligned} f_{r,\psi}(r, \psi) &= \frac{r}{2\pi\sigma^2} \exp\left[-\frac{(r \cos \psi - A)^2 + (r \sin \psi)^2}{2\sigma^2}\right] \\ &= \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2 + A^2 - 2Ar \cos \psi}{2\sigma^2}\right] \end{aligned} \tag{6.121}$$

In the above equation, because of the presence of the term $2\,Ar\cos\psi$, it is not possible to express $f_{r,\psi}(r,\ \psi)$ as the product of two functions, one only of r and the other only of ψ. Hence, in this case the random variables R and ψ are not statistically independent. Hence, in order to get the marginal density function $f_R(r)$ in which we are interested, we have to integrate the joint density function $f_{R,\ \psi}(r,\ \psi)$ with respect to ψ for all values of ψ, i.e.,

$$\begin{aligned} f_R(r) &= \int_0^{2\pi} f_{R,\psi}\,d\psi \\ &= \int_0^{2\pi} \frac{r}{2\pi\sigma^2}\exp\left[-\frac{r^2+A^2-2Ar\cos\psi}{2\sigma^2}\right]d\psi \\ &= \frac{r}{2\pi\sigma^2}\exp\left[-\frac{r^2+A^2}{2\sigma^2}\right]\int_0^{2\pi}\exp\left[-\frac{Ar\cos\psi}{2\sigma^2}\right]d\psi \end{aligned} \tag{6.122}$$

But, we know that

$$=\frac{1}{2\pi}\int_0^{2\pi}\exp[x\cos\psi]\,d\psi = I_0(x) \tag{6.123}$$

where $I_0(x)$ is the modified Bessel function of the first kind and of order zero. Hence, we may write

$$f_R(r) = \frac{r}{\sigma^2}\exp\left[-\frac{r^2+A^2}{2\sigma^2}\right]I_0\left(\frac{Ar}{\sigma^2}\right) \tag{6.124}$$

This density function is, as we know, called the Rician density function (see Section 6.4.2) and has been sketched in Fig. 6.7(b) for two values of A. for $A = 0$, it reduces to Rayleigh density function.

Thus, the envelope of a sinusoid plus zero-mean Gaussian narrowband noise will have Rician distribution.

Mobile radio channels exhibit fading because of multipath effects (see Chapter 15). These channels are modeled either as Rician fading channels or as Rayleigh fading channels depending upon respectively the presence or absence of fixed scatterers in the multipath environment of the mobile radio.

6.14 NOISE – SOURCES AND CLASSIFICATION

As shown in Fig. 6.13, noise may be broadly classified, depending on the location of the sources, into two types – *external noise* and *internal noise*. Note that these terms, external and internal, are used with reference to the receiver. External noise may, in turn, be divided into *atmospheric noise*, *extraterrestrial noise*, and *man-made noise*. Internal noise is mainly of two types – *thermal noise* and *shot noise*.

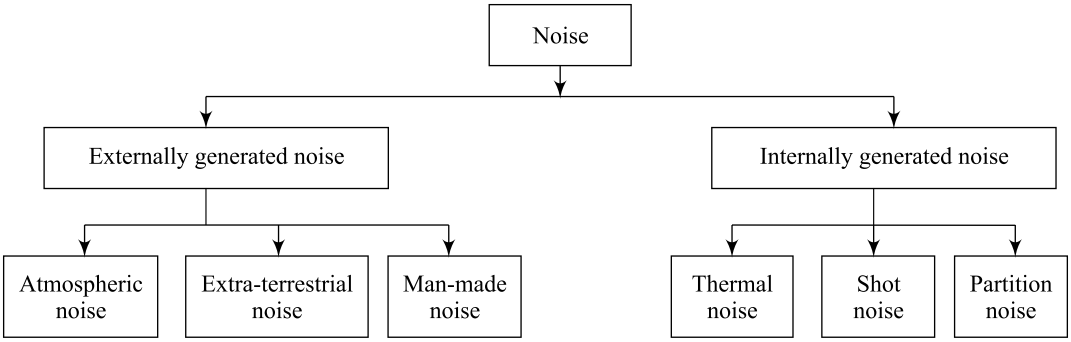


Fig. 6.13 *Types of noise*

6.14.1 Atmospheric Noise

Atmospheric noise (also referred to as 'static') arises from lightning discharges (cloud-to-cloud, or cloud-to-earth), caused by thunder storms. Lightnings are heavy electrical current discharges, running into thousands of amperes, and are accompanied by intense radiation of electromagnetic waves over a broad spectrum of frequencies. Different frequency bands of these electromagnetic waves propagate via the usual modes of propagation like the ground wave and sky wave just like ordinary radio waves and corrupt the desired signal. Atmospheric noise has frequency components extending from very low frequencies up to hundreds of megahertz and its intensity varies with frequency as well as time of the day. Further, it has been experimentally observed that during the daytime, its intensity decreases with frequency up to about 2 MHz and that there exists a relative peak of intensity around 10 MHz. It has a relative dip at around 2 MHz. During night. time also its intensity decreases with frequency but has generally higher values (than those obtained during daytime) at all frequencies. Its intensity during the night becomes very low or insignificant; beyond about 10 MHz.

From the foregoing it is clear that the disturbance caused by atmospheric noise is more severe in the medium wave band as compared to the short-wave band; and it is very little in the case of VHF and UHF bands that are used for television.

6.14.2 Extra-Terrestrial Noise

This has two components:

1. Solar Noise
2. Galactic Noise

1. Solar Noise: Our sun, being a gaseous body with very high surface temperatures (in excess of 6000° C), radiates considerable amount of noise, whose intensity has been observed to have a cyclic variation with an 11-year period, called the 11-year sun-spot cycle.

2. Galactic Noise: All the stars are also hot gaseous bodies and they too radiate noise. The radiation reaching the earth from each individual star may be very small compared to that from our sun, because of their very large distance. But they are large in numbers and are spread all over the sky, making their overall contribution *not* insignificant. In addition, the suns of other galaxies and our own 'Milky Way' also radiate noise. This noise, called the 'galactic noise', is almost uniformly intense from all parts of the sky but is slightly more intense in the direction of our Milky Way.

The extra-terrestrial radiation has spectral components from a few megahertz to about a few gigahertz. However, only those components which have frequencies above 20 MHz pass through the ionosphere and reach the earth. Further, those with frequencies above approximately 1.5 GHz are absorbed by hydrogen in the interstellar space. *Thus, extra-terrestrial noise can cause disturbance to communications in the frequency range of 20 MHz to 1.5 GHz.*

6.14.3 Man-made Noise

Automobile ignition, aircraft ignition, fluorescent lamps, sparking at the brushes of electric motors, etc., radiate electromagnetic waves that cause disturbance to communications, especially in the range of 1 MHz to 500 MHz. Because of the nature of its origin, this noise is more intense in urban areas than in rural areas. However, it must be noted that noise emanating from these sources can travel considerable distances.

6.15 THERMAL NOISE

We know that at any temperature above 0° K, the free electrons in a conductor possess kinetic energy and so will be in random motion because of collisions with the lattice. This random motion of electrons is equivalent to a random current flow with in the conductor, and this creates a random voltage across the conductor. This random voltage across a conductor arising from the random motion of free electrons inside it because of thermal agitation, is called *thermal noise*. It is also known as *Johnson noise*. This thermal noise voltage fluctuates randomly about a mean value of zero.

Analyzing the thermal agitation of the free electrons by using quantum mechanics, it has been shown that at a temperature of $T°$ K, the power spectral density of the thermal noise across a conductor having a resistance of R ohms, is given by

$$P(f) = \frac{2Rh|f|}{\left(e^{\frac{h|f|}{kT}} - 1\right)} \text{ volts}^2\text{/Hertz} \tag{6.125}$$

where

h = Planck's constant = 6.6×10^{-34} Joule-sec,
k = Boltzmann's constant = 1.38×10^{-23} J/°K

In Eq. (6.125),

$$e^{\frac{h|f|}{kT}} = 1 \text{ at } |f| = 0$$

and it goes on increasing as $|f|$ increases. Further, its rate of increase will be greater than that of the numerator. Thus, $P(f)$, power spectral density of thermal noise, has a maximum value at $f = 0$ and it goes on decreasing as $|f|$ increases. The maximum value of $P(f)$, occurring at $f = 0$, can be obtained by using L'Hosptal's rule and is given by

$$P(f)\big|_{f=0} = 2KTR \text{ volts}^2\text{/Hz} \tag{6.126}$$

Although $P(f)$ decreases as $|f|$ increases, the rate of decrease at normal room temperature is so small that it may safely be assumed to be remaining constant at the value 2 kTR even up to frequencies of the order of 10^{12} to 10^{13} Hz, as its value drops only by 10% from its zero frequency value even at a frequency of 2000 GHz. As this frequency is far more than the frequencies and bandwidths used in any of our ordinary communication systems, for all practical purposes, we can safely assume that the power spectral density (PSD) of thermal noise is constant and independent of frequency and that it has a value given by

$$P(f) = 2kTR \text{ Volt}^2\text{/Hz} \tag{6.127}$$

Note

It must be noted that P(f), as given in Eq. (6.105) represents the two-sided power spectral density as shown in Fig. 6.14.

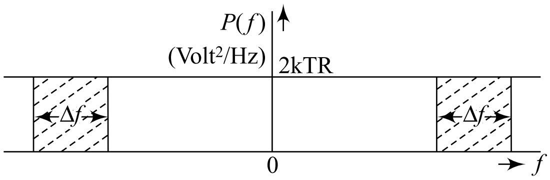


Fig. 6.14 *PSD of thermal noise*

Since thermal noise has a PSD which is almost a constant, it has all frequency components from minus infinity to plus infinity, in equal measure. Such a noise is called *white noise*. Since its PSD is constant, its ACF (inverse FT of PSD) is an impulse function in time. This indicates that any two samples of white noise, however close they may be in time, are uncorrelated. However, it must be noted that no physical noise source can be a white noise source, since white noise implies infinite noise power (area under PSD curve).

We may now determine the r.m.s. value of the noise voltage across a resistor of R ohms at a temperature of $T°$ K over a bandwidth of Δf. From Fig. 6.14, we find that

$$\text{Mean-squared value of noise in } R = (2\Delta f)\, 2kTR = 4\, kTR\Delta f \text{ volts}^2 \tag{6.128}$$

$$\boxed{\left.\begin{array}{l}\text{Hence r.m.s. value of voltage}\\ \text{across the resistor } R\end{array}\right\} = 2\sqrt{kTR\Delta f} \text{ volts}} \tag{6.129}$$

From the foregoing, it is clear that insofar as noise calculations are concerned, we may model a resistor of R ohms at temperature $T°$ K as follows.

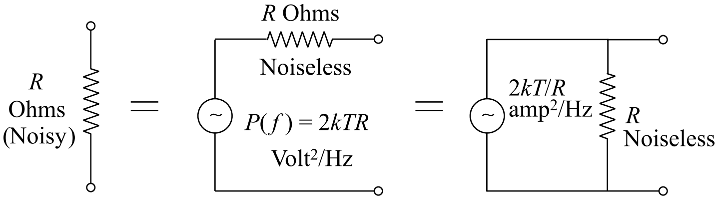


Fig. 6.15 *Modeling a noisy resistor*

We shall now make use of the noise model of a resistor shown in Fig. 6.15 to obtain the noise equivalent circuits of resistances in series and in parallel.

6.15.1 Resistors in Series and in Parallel

Series connection

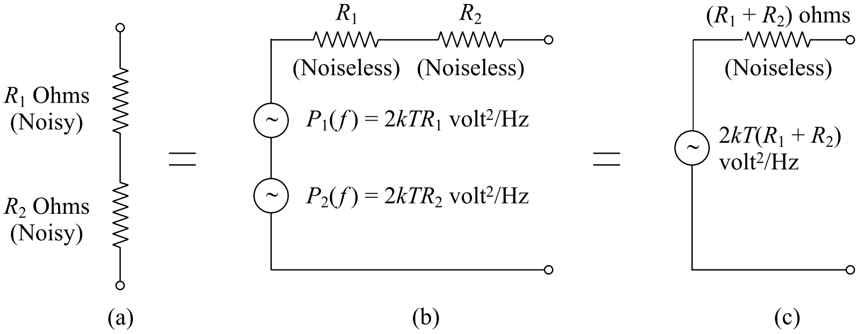


Fig. 6.16 *Resistors in series*

Superposition of PSDs In Fig. 6.16, let resistor R_1 produce noise voltage $n_1(t)$ and R_2 produce noise voltage $n_2(t)$. Then the total power of the sum process $[n_1(t) + n_2(t)]$ is given by

$$P_{12} = E[n_1(t) + n_2(t)]^2 = E[n_1^2(t)] + E[n_2^2(t)] + 2E[n_1(t)n_2(t)]$$

But since the noise processes produced in R_1 and R_2 are independent and zero-mean processes,

$$E[n_1(t)n_2(t)] = 0$$

Further,

$$E[n_1^2(t)] = P_1 = \text{Average power of the noise process } n_1(t)$$

and $$E[n_2^2(t)] = P_2 = \text{Average power of the noise process } n_2(t)$$

$\therefore$ $$P_{12} = P_1 + P_2$$

Thus, it is their powers (or the mean squared values which get added, and not the voltages. *This means that, as shown in Fig. 6.16, in the equivalent circuit, it is the noise power spectral densities to which superposition principle applies – not to the noise voltages produced by the two resistors.*

Parallel connection

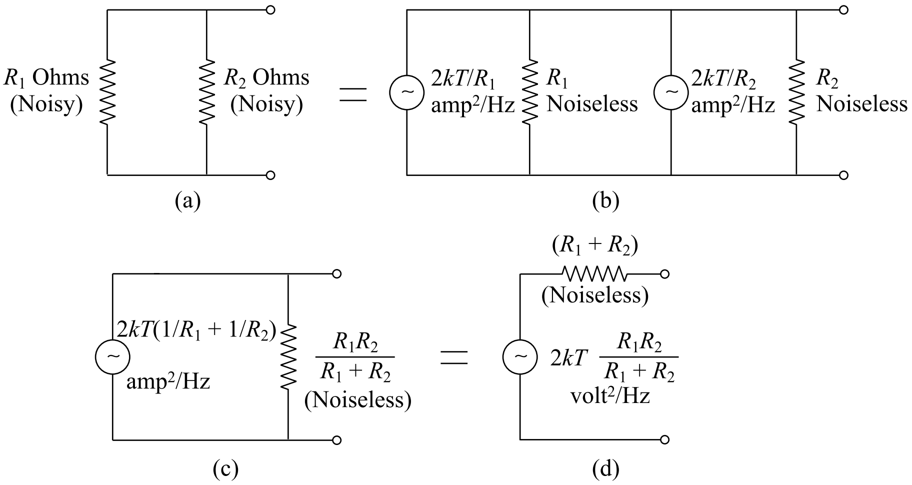


Fig. 6.17 *Resistances in parallel*

Note

(i) When two resistors are in series, it is their noise powers, or their noise power spectral densities (PSDs) in volt²/Hz, which can be added, and not their noise voltages.

(ii) When two resistors are in parallel, it is their noise powers, or their noise PSDs in amp²/Hz, which can be added – not their individual noise currents.

Thus, in a circuit with multiple noise sources which are independent, the principle of superposition applies not to the r.m.s. voltages or currents of the sources, but only to their mean-squared values or power spectra. As has been already shown, the justification for the above two statements stems from the fact that the two noise sources are independent and hence uncorrelated and further the noise has zero mean.

Example 6.20 Find the r.m.s. value of the thermal noise voltage across a resistor of 1 MΩ at a temperature of 27° C if the measurement is made with an instrument having a bandwidth of 10^4 Hz.

Solution From Eq. (6.129), we have

$$\begin{aligned} e_{\text{r.m.s}} &= \sqrt{4kTR(\Delta f)} \text{ volts} \\ &= \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 10^6 \times 10^4} \\ &= \sqrt{12 \times 1.38 \times 10^{-11}} = 12.868\ \mu\text{V} \end{aligned}$$

From the above result, the reader may wonder why we should bother about the thermal noise at all, if their r.m.s. values are typical a few micro-volts. However, if we see the signal voltage levels at the front-end of a receiver, they will also have typically values of the same order. If the resistance considered in the example is the input resistance of the front-end of the receiver, it means that we have a situation where the signal and noise have approximately the same levels of magnitude at the input of an amplifier – not a desirable situation, as the amplifier is likely to add some more noise while amplifying the input signal and noise by the same factor.

Example 6.21 A 10 kΩ and a 20 kΩ resistors are both at a room temperature of 27° C. For a 100 kHz bandwidth, determine the r.m.s. value of the thermal noise voltage across (a) each one of them, (b) their series combination, and (c) their parallel combination.

Solution

(a) (i) Across the 10 kΩ resistor

From Eq. (6.129), we have

$$e_{\text{r.m.s}} = \sqrt{4kTR(\Delta f)} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 10^4 \times 10^5}$$

$$\therefore \quad e_{\text{r.m.s}} = 4.07\ \mu\text{V}$$

(ii) Across the 20 kΩ resistor

$$e_{\text{r.m.s}} = \sqrt{4kTR(\Delta f)} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 2 \times 10^4 \times 10^5}$$

$$\therefore \quad e_{\text{r.m.s}} = 5.75\ \mu\text{V}$$

(b) With the two resistors in series

$$e_{\text{r.m.s}} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 3 \times 10^4 \times 10^5} = 7.04\ \mu\text{V}$$

It may also be found out as

$$e_{\text{r.m.s}} = \sqrt{(4.07)^2 + (5.75)^2} = 7.04\ \mu\text{V}$$

(c) With the two resistors in parallel

Resistance of parallel combination $\dfrac{10 \times 20}{(10+20)} = 6.67\ \text{k}\Omega$

$$\therefore \quad e_{\text{r.m.s}} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 6.67 \times 10^3 \times 10^5} = \sqrt{110.4 \times 10^{-13}}$$

$$\therefore \quad e_{\text{r.m.s}} = 3.32\ \mu\text{V}$$

6.15.2 Thermal Noise and Reactive Circuits

Pure reactive circuit elements like inductances and capacitances do not dissipate any power and do not produce thermal noise. A lossy reactive element like an inductance which can be represented by pure inductance in series with a resistance, or a lossy capacitor, i.e., a capacitor in which dielectric loss takes place and which can be represented by a pure lossless capacitor in shunt with a resistance, do generate thermal noise. *While calculating the thermal noise in circuits containing reactive elements, we should, however, consider the effect of the reactive elements on the shape of the noise power spectrum.*

Example 6.22 A resistor R ohms at a temperature of T° K is connected across a pure capacitor of C Farads. Determine the r.m.s. value of noise voltage across the capacitor C.

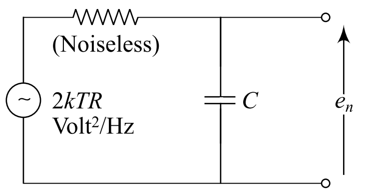


Fig. 6.18 *Noise equivalent circuit of Example 6.22*

Solution Representing the resistance R by its noise equivalent circuit, we have

For the RC low pass filter of Fig. 6.18, the transfer function is given by

$$H(f) = \frac{1}{1 + j\omega CR}$$

$$\therefore \qquad |H(f)|^2 = \frac{1}{1+4\pi^2 f^2 C^2 R^2}$$

$$\therefore \qquad P_0(f) = \left.\begin{array}{l}\text{PSD of the noise voltage}\\ \text{across the capacitor } C\end{array}\right\} = 2kTR\,|H(f)|^2 = \frac{2kTR}{1+4\pi^2 f^2 C^2 R^2}\ \text{V}^2/\text{Hz}$$

To find the r.m.s. value of the noise voltage across the resistor, we first determine P_0, the average noise power across the output by integrating the power spectral density of the output noise across the capacitor, i.e., $P_0(f)$ over the entire frequency range from $f = -\infty$ to $f = +\infty$.

$$\therefore \qquad P_0 = \int_{-\infty}^{\infty} \left(\frac{2kTR}{1+4\pi^2 f^2 C^2 R^2}\right) df$$

Substituting $2\pi fCR = \tan\theta$, and integrating,

$$P_0 = \int_{-\pi/2}^{\pi/2} \left(\frac{2kTR}{\sec^2\theta}\right)\cdot\left(\frac{\sec^2\theta}{2\pi CR}\right) d\theta = \frac{kT}{C}$$

$\therefore$ r.m.s. value of the noise voltage across the capacitor $= e_{\text{r.m.s.}} = \sqrt{\dfrac{kT}{C}}$ volts.

This result appears a bit surprising because, the r.m.s. value of the output noise voltage is independent of R, although the r.m.s. value of the thermal noise voltage across the resistance, over any bandwidth, is proportional to $\sqrt{R}$. Actually, what happens is, as the value of R increases, even though the input noise voltage power spectrum increases proportional to R, the bandwidth over which noise is allowed to pass through the RC low pass filter goes on decreasing with R as the cut-off frequency is inversely proportional to R. Thus, the noise power available at the output, and hence the r.m.s. value of the noise voltage across the output terminals, is independent of the value of R.

Example 6.23 The input circuit of an RF amplifier is a tuned circuit comprising a coil having a resistance of r ohms and inductance of L Henries connected across a capacitor of C farads. Determine the r.m.s. value of the thermal noise voltage across the input terminals of the amplifier at resonance.

Solution Let the tuned circuit be at resonance. Consider a small bandwidth Δf around the resonance frequency. r.m.s. value of the thermal noise voltage across the r ohms resistance over a bandwidth of (Δf) around the resonance frequency is given by e_{rms}, where

$$e_{\text{rms}} = \sqrt{4kTr(\Delta f)} \qquad (6.130)$$

$$\therefore \qquad \left.\begin{array}{l}\text{At resonance, the r.m.s. value}\\ \text{of voltage across the capacitor } C\\ \text{over a bandwidth of } \Delta f\end{array}\right\} = \sqrt{4kT(Q^2 r)(\Delta f)}$$

where Q is the magnification factor of the tank circuit at resonance and is assumed here to remain constant over a small interval of frequency, Δf.

But $Q^2 r = R_d$, the dynamic resistance of the tank circuit at resonance.

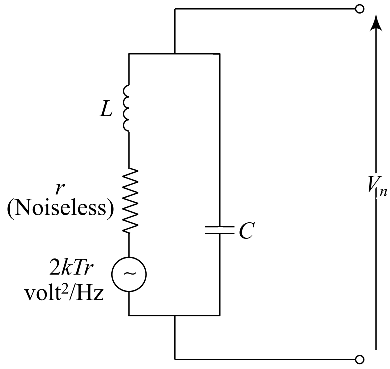


Fig. 6.19 *Tuned circuit of Example 6.23*

$$\therefore \qquad \left.\begin{array}{l}\text{At resonance, input noise voltage}\\ \text{(r.m.s. value) for the r.f amplifier}\end{array}\right\} = \sqrt{4kTR_d(\Delta f)} \qquad (6.131)$$

Equation (6.131) represents an interesting result, as it tells us that insofar as thermal noise at resonance across the tank circuit is concerned, it is the dynamic resistance R_d of the tank circuit at resonance, which appears to be producing the noise.

Example 6.24 A parallel circuit resonates at 90 MHz and its capacitor C is 30 pF. The Q of the tuned circuit is 50 and the circuit is at a temperature of 17°C. Calculate the r.m.s. value of the noise voltage in a bandwidth of 20 kHz around the resonance frequency?

Solution The equivalent series resistance r of the tuned circuit $= r = \dfrac{X_c}{Q}$

$$= \frac{1}{2\pi \times 90 \times 10^6 \times 30 \times 10^{-12} \times 50} = 1.17\ \Omega$$

where X_c is the reactance of C at resonance

The effective equivalent resistance for the tuned circuit, at resonance

$$= R_d = Q^2 r = (50)^2 \times 1.17 = 2925\ \Omega$$

$\therefore$ r.m.s. value of the noise voltage across the tuned circuit

$$= \sqrt{4kTR_d(\Delta f)} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 20 \times 10^3 \times 2925}$$

$$= \sqrt{96876 \times 10^{-17}} = 98.42 \times 10^{-8}\ \text{V} = 0.9842\ \mu\text{V}$$

6.15.3 Available Noise Power

Maximum power transfer theorem tells us that maximum power will be delivered by the source to the load resistance R_L of Fig. 6.20, when R_L equals R, the source resistance. Under this condition, the load is said to be matched to the source and the power delivered to R_L under matched conditions, is given by

$$\left(\frac{V}{2R}\right)^2 \cdot R = \frac{V^2}{4R} \triangleq \text{Available power from the source}$$

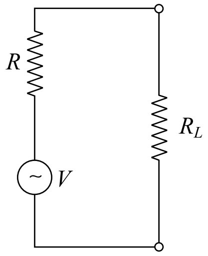


Fig. 6.20 *Maximum power transfer*

Considering a resistor of R ohms as a thermal noise source as shown in Fig. 6.21, we have

$$\text{Available noise power} = \left(\frac{\sqrt{4kTR(\Delta f)}}{2R}\right)^2 \cdot R = kT(\Delta f)$$

$$\therefore \quad \boxed{\text{Available noise power} = kT(\Delta f)} \qquad (6.132)$$

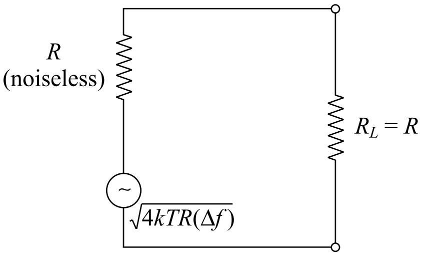


Fig. 6.21 *Maximum noise power transfer from a resistor*

Noise temperature of a source The noise temperature of a source is defined as T where

$$T \triangleq \frac{p}{k(\Delta f)} \qquad (6.133)$$

where p is the available power from the source in a bandwidth (Δf) Hz. It may be noted here that the source may be a thermal noise source or it may be some other type. If it is thermal type, T will be the temperature of that source. If it is not thermal type, T may not have anything to do with the actual temperature of the source.

6.15.4 White Noise

A noise in which all frequency components from $f = -\infty$ to $f = +\infty$ are present in equal measure, i.e., whose power spectral density remains constant for all frequencies and is independent of frequency, as shown in Fig. 6.22, is called *white noise*.

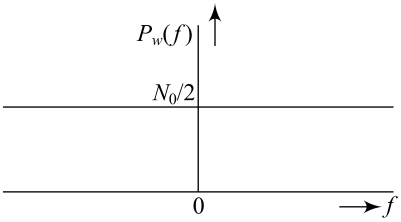


Fig. 6.22 *PSD of white noise*

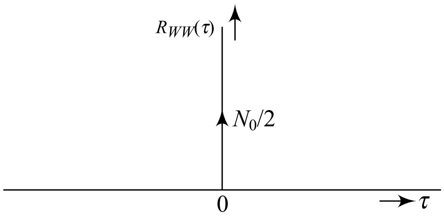


Fig. 6.23 *ACF of white noise*

Thus, the auto-correlation function of white noise is given by

$$R_{WW}(\tau) = \mathcal{F}^{-1}\left[\frac{N_0}{2}\right] = \frac{N_0}{2}\delta(\tau) \tag{6.134}$$

The fact that the auto-correlation function is an impulse implies that if we take two samples of white noise, however close the two samples may be, they are uncorrelated. Thus, we find that white noise is perfectly random.

We know that the total area under the PSD curve of any signal gives the average power of that signal. Since the PSD of white noise remains constant for all frequencies from $f = -\infty$ to $f = +\infty$, the area under its PSD curve is infinity. This means that a white noise source must be producing an infinite average power, which is of course impossible in practice. Thus, there cannot be any physical source producing exact white noise. *However, 'white noise' is very useful conceptually and is easy to deal with mathematically.*

Note *We had stated earlier that 'thermal noise' although not exactly white, can be regarded as white for all practical purposes since its PSD remains almost flat even up to 10^{12} Hz – frequencies far beyond those used by any conventional communication system. Further, since by its very nature, it is the aggregate of the noise components produced by the independent random movements of a very large number of charged carriers in a conductor, from central limit theorem, we conclude that thermal noise is Gaussian and zero mean. Thermal noise is thus a zero-mean White (approximately) Gaussian noise.*

Example 6.25 White noise with PSD of $N_0/2$ is filtered using an ideal LPF whose cut-off frequency is f_c Hz. What is the maximum rate at which the output of the LPF can be sampled, if the samples so obtained are to be uncorrelated?

Solution

PSD of noise at output of LPF = $|H(f)|^2$ × PSD of white noise = $P_n(f)$

Hence PSD of noise at the output of LPF, viz., $P_n(f)$ will be as shown in Fig. 6.24(c). The ACF of the noise at the output of LPF, viz., $R_{nn}(\tau)$ which is the inverse Fourier transform of $P_n(f)$, is a sinc function and is shown in Fig. 6.24(d).

Since this auto-correlation function goes through zero values at regular intervals of $\left(\dfrac{1}{2f_c}\right)$ seconds, the minimum sampling interval should be $\left(\dfrac{1}{2f_c}\right)$ for the samples to be uncorrelated. Hence, the sampling of the output noise of the LPF should be done at a frequency of $2f_c$ samples per second for the samples to be uncorrelated.

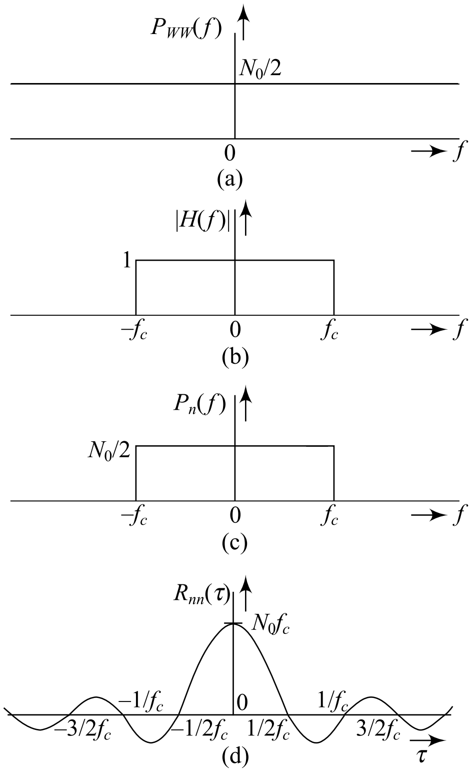


Fig. 6.24 *(a) PSD of white noise, (b) Magnitude response of the ideal LPF, (c) PSD of noise at output of LPF, (d) ACF of output noise,* $R_{nn}(\tau)$

6.16 SHOT NOISE

In the previous section, we had considered, in some detail, 'thermal noise', which is one of the important constituents of internal noise. Another important source of internal noise is what is called the *shot noise*. This is produced in electronic devices such as vacuum and semiconductor diodes, photo-diodes, transistor, etc. It is due to the random emission of electrons from the cathode in the case of vacuum tubes and due to the inherent randomness in the diffusion of minority carriers and drift of majority carriers across the junction in the case of semiconductor devices.

Let us consider the case of simple vacuum diode with plane, parallel electrodes. The cathode of this device emits electrons due to a process called 'thermionic emission'. When the cathode is kept at a constant temperature, the number of electrons emitted per second, on the average, remains the same and if the anode is given a sufficiently large positive potential with respect to the cathode, the tube operates in what is called the '*temperature-limited condition*'. When the tube is operated in this condition, all the electrons emitted by the cathode ultimately reach the anode and the number of electrons reaching the anode per second, is limited only by the rate of emission of electrons by the cathode, i.e., limited by the temperature of the cathode and not by the voltage applied to the anode. Under this condition, an electron emitted from the cathode surface gets accelerated towards the anode and ultimately reaches it after a brief interval, called the 'transit-time', i.e., time taken by the electron to travel the distance between the cathode and anode. Under normal temperature-limited conditions, this transit time will be extremely small, of the order of a micro-microsecond.

Let us now follow the motion of one such electron emitted by the cathode. Since the initial velocity with which it is emitted is extremely small compared to the final velocity it acquires before reaching the anode, we will assume that the initial velocity is zero. Then, due to the uniform electric field between the cathode and the anode, it gets accelerated and its velocity goes on increasing linearly with time. Since an electron is a charged particle, its movement creates current and as its velocity increases uniformly with time, the current contributed by the electron also increases linearly with time. Finally, when the electron reaches the anode the current drops down to zero. Thus, the waveform of the current created by a single electron will be as shown in Fig. 6.25. The maximum value attained by the current must be $(2q/\tau)$ where q is the charge of an electron (1.6×10^{-19} coulomb) and τ is the transit-time, since the area of the triangular current pulse must be equal to q.

Suppose a steady current of 1 mA is flowing through the diode under temperature-limited condition. A current of 1 mA means that on the average 6×10^{15} electrons are reaching the anode per second. We have deliberately used the word 'average' because 6×10^{15} electrons per second does not necessarily mean that *exactly* 6×10^9 electrons reach the anode every micro-second, or that exactly 6×10^3 electrons reach the anode every 10^{-12} sec. The actual number may fluctuate about these values because the number of electrons emitted per second from the cathode goes on varying randomly with a mean value which is a constant and dependent upon the temperature of the cathode. Hence, the waveform of the currents from individual electrons when a large number of the emitted electrons are considered will be as shown in Fig. 6.26, where t_1, t_2, t_3, etc., are random instants of time. The average number of such random instants per second is however, constant.

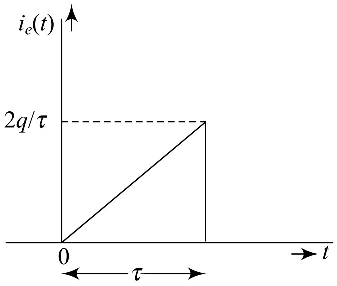


Fig. 6.25 *Current waveform created by a single electron*

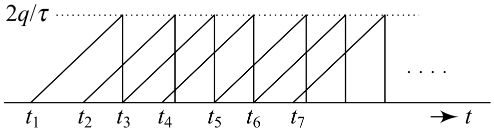


Fig. 6.26 *Waveforms of currents contributed by randomly emitted electrons*

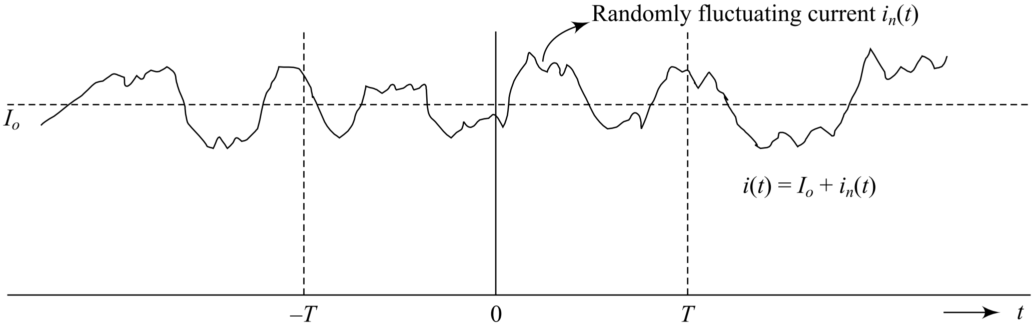


Fig. 6.27 *Anode current waveform obtained by summing up the various triangular current pulses of Fig. 6.26*

Since the transit time, τ, is extremely small, we may approximate each triangular current pulse of area q by an impulse of strength q. In Fig. 6.26 then we will have impulses of strength q occurring at random instants t_1, t_2, t_3, etc. As $i_e(t)$, the current pulse due to a single electron, is a finite energy signal and is therefore Fourier transformable, let

$$i_e(t) \xleftrightarrow{F \cdot T} I_e(f) \tag{6.135}$$

But

$$i_e(t) \approx q\delta(t)$$

Now, to find $P_{nI}(f)$, the power spectral density of the diode current component $i_n(t)$, we note that $i_n(t)$ is a random signal and so is not Fourier transformable. We shall therefore follow the approach adopted in Section 2.7 for determining the PSD of deterministic power signals. Accordingly, let us consider a signal $i_{nT}(t)$ defined as

$$i_{nT}(t) \underline{\Delta} \begin{cases} i_n(t); & |t| \le T \\ 0; & \text{otherwise} \end{cases} \tag{6.136}$$

$i_{nT}(t)$ is thus a segment of $i_n(t)$ and is of duration $2T$. Hence $i_{nT}(t)$ is a finite energy signal. Let

$$i_{nT}(t) \xleftrightarrow{F \cdot T} I_{nT}(f)$$

Following the arguments similar to those of Section 2.7 and recognizing that here $i_{nT}(t)$ is a segment of one realization of the random signal $i_n(t)$, we write the expression for the PSD of $i_n(t)$ as

$$P_{nI}(f) = \underset{T\to\infty}{\text{Lt}}\ E\left[\frac{|I_{nT}(f)|^2}{2T}\right], \tag{6.137}$$

where the symbol E is used to indicate the ensemble average since $|I_{nT}(f)|^2$ changes from one realization to another.

Assuming that on the average N electrons arrive at the anode per second, $2TN$ electrons arrive in $2T$ seconds and we may write

$$P_{nI}(f) = \underset{T\to\infty}{\text{Lt}}\ E\left[\frac{|I_{nT}(f)|^2}{2T}\right] = \underset{T\to\infty}{\text{Lt}}\ (2TN)E\left[\frac{|I_e(f)|^2}{2T}\right] \cong q^2 n$$

But $qN = I_0$, the dc current in the anode circuit.

$$\therefore \qquad \boxed{P_{nI}(f) = I_0 q \text{ amp}^2/\text{Hz}} \tag{6.138}$$

Thus, the PSD of the anode current is independent of 'f'. This indicates that shot noise is a white noise process. However, it must be remembered that this is only an approximation, since we have approximated the triangular current pulses by impulses of current by considering τ the transit time to be negligibly small. If we do not make that approximation and use the Fourier transform of triangular pulse instead of that of an impulse function, and proceed with the derivation, we will find that $P_{nI}(f)$ is not independent of frequency and that it falls off slowly with increasing frequencies. However, its rate of decrease with frequency is so low that it is, for all practical purposes, constant up to frequencies of the order of a few hundred megahertz.

Note *$P_{nI}(f)$, as given in Eq. (6.138) is a two-sided power spectrum. Hence, over a bandwidth of (Δf) Hz, the mean-squared value of the shot noise current is given by*

$$I_n^2 = 2I_0 q(\Delta f)\ amp^2 \tag{6.139}$$

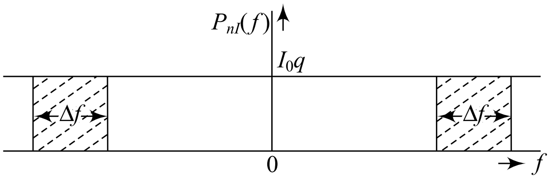


Fig. 6.28 *Showing that $P_{nI}(f)$ given by Eq. (6.138) is a two-sided power spectrum*

6.16.1 Shot Noise in Space-Charge Limited Diodes

For a vacuum diode operating in the *space-charge limited region* of its characteristic, the randomness in the number of electrons arriving at the anode is somewhat smoothened out due to the presence of a thick cloud of electrons near the cathode surface. Hence, in this case, Eq. (6.138) is modified as

$$P_{nI}(f) = \alpha I_0 q \text{ amp}^2/\text{Hz} \tag{6.140}$$

In Eq. (6.140), α is a 'space-charge smoothing factor' whose value depends on the density of the space-charge and may vary from 0.01 to 1. It is given by (Ref. 2)

$$\alpha = \frac{1.28\,kT_c g_d}{qI_0} \tag{6.141}$$

where k = Boltzmann's constant = 1.38×10^{-23} J/°K
T_c = cathode temperature in °K
g_d = Dynamic conductance of the diode = Rate of change of plate current with plate voltage
q = Charge of an electron = 1.6×10^{-19} coulomb

6.16.2 Shot Noise in Semiconductor Diodes

Shot noise arises in the case of semiconductor diodes also, because of the random nature of the number of minority carriers diffusing across the junction and also of the generation and recombination of holes and electrons. An analysis of the shot noise in semiconductor diodes yields a somewhat similar equation

$$I_n^2 = 2(I + 2I_0)q(\Delta f)\ \text{amp}^2 \tag{6.142}$$

In Eq. (6.142), I is the dc current flowing across the *p-n* junction, expressed in amperes and I_0 is the reverse saturation current in amperes. This equation, however, is applicable only at low frequencies and low injection currents.

Partition Noise In multi-electrode devices like the vacuum triodes and pentodes as well as the bipolar junction transistors, one more type of noise, known as 'partition noise', is generated. In triodes and pentodes, it arises due to the random distribution of the electrons emitted by the cathode between the grids and the anode or plate; and in the case of transistors, due to the random distribution between the base and collector, of the charged carriers injected into the base region.

In supreheterodyne radio receivers, it is this partition noise which makes the mixer stage the most noisy one.

Example 6.26 A vacuum diode operating in the temperature-limited region and carrying a direct current of I_0 amperes, with a resistance of R ohms connected across it through a coupling capacitor is used as a noise source.

(a) Determine the PSD of the output noise neglecting the effect of the coupling condenser.

(b) Find the ratio of mean-squared value of the thermal noise to the mean-squared value of the total noise at the output.

Solution With a direct current of I_0 amps flowing through it, the mean-squared value of the shot-noise generated by the diode in temperature-limited condition is given by

$$I_{sh}^2 = 2I_0 q(\Delta f)\ \text{amp}^2 \qquad \text{(From Eq. (6.139))}$$

The thermal noise generated by the resistance of R ohms has a mean-squared value given by

$$I_{th}^2 = \frac{4kT(\Delta f)}{R}\ \text{amp} \qquad \text{(From Fig. 6.15)}$$

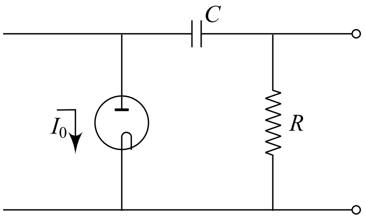


Fig. 6.29

$\therefore$ PSD of the total output noise = $\left(2I_0 q + \dfrac{4kT}{R}\right)$ amp^2/Hz

and
$$\left(\frac{\text{Mean-squared value of the thermal noise}}{\text{Mean-squared value of total noise}}\right)=\frac{4kT(\Delta f)/R}{\left(2I_0q+\frac{4kT}{R}\right)\Delta f}$$
$$=\left(\frac{4kT}{4kT+2I_0qR}\right)$$

6.17 NOISE EQUIVALENT BANDWIDTH OF A FILTER

Definition Let T be an arbitrary filter, with transfer function $H(f)$. *The noise-equivalent bandwidth of this filter T is defined as the bandwidth B of an ideal low pass filter whose pass band gain is* $|H(0)|=|H(f)||_{f=0}$, *such that when a white noise source of power spectral density* $N_0/2$ *is applied as input, the ideal LPF gives the same output power as the filter T under consideration*

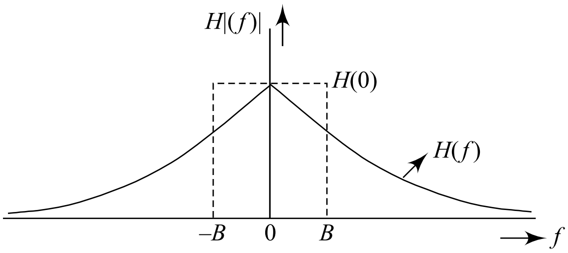


Fig. 6.30 *Transfer functions of T and the ideal LPF of pass band gain H(0) and bandwidth B Hz*

With white noise of PSD equal to $N_0/2$ applied as input:

(a) Output noise power of the filter = $\frac{N_0}{2}\int_{-\infty}^{\infty}|H(f)|^2\,df=N_0\int_{0}^{\infty}|H(f)|^2\,df$

(With $h(t)$ real-valued, $|H(f)|$ must have even symmetry.)

(b) Output noise power of the ideal LPF $=\frac{N_0}{2}\int_{-B}^{B}|H(0)|^2\,df$

$$=N_0B|H(0)|^2$$

$$N_0B|H(0)|^2=N_0\int_0^{\infty}|H(f)|^2\,df$$

∴ B = Noise-equivalent bandwidth of the filter T

$$=\frac{\int_0^{\infty}|H(f)|^2\,df}{|H(0)|^2}$$

∴ noise equivalent bandwidth $B=\frac{\int_0^{\infty}|H(f)|^2\,df}{|H(0)|^2}$, for a filter with transfer function $H(f)$ (6.143)

Example 6.27 Determine the noise equivalent bandwidth of the RC low pass filter shown in Fig. 6.31.

Solution For the RC low pass filter,

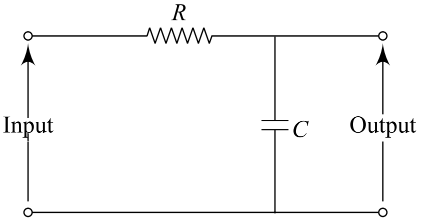


Fig. 6.31

$$H(f)=\frac{1}{1+j\omega CR} \quad \therefore \quad |H(f)|^2=\frac{1}{1+4\pi^2 f^2C^2R^2}$$

$\therefore$ when white noise of PSD equal to $N_0/2$ is applied as input to the RC low pass filter, the output noise power is

$$P_1=\frac{N_0}{2}\int_{-\infty}^{\infty}\frac{1}{1+4\pi^2 f^2C^2R^2}df=N_0\int_{0}^{\infty}\frac{1}{1+4\pi^2 f^2C^2R^2}df$$

Put $2\pi fCR=\tan\theta \quad \therefore \quad 2\pi CRdf=\sec^2 d\theta$

$$\therefore \qquad P_{no}=N_0\int_{0}^{\pi/2}\frac{(1/2\pi CR)\sec^2\theta}{\sec^2\theta}d\theta=\frac{N_0}{4RC}$$

When white noise of PSD equal to $N_0/2$ is applied as input to an ideal LPF of bandwidth B and pass band gain $= H(0) = 1$, the corresponding output noise power is

$$P_2=2B\left(\frac{N_0}{2}\right)\cdot|H(0)|^2=N_0B$$

$$\therefore \qquad N_0B=\frac{N_0}{4RC}, \text{ i.e. } B=\frac{1}{4RC}$$

Example 6.28 If zero-mean white noise of two-sided PSD η/2 *W*/Hz is applied as input to the low pass RC filter of Fig. 6.31, determine and sketch the PSD and auto-correlation function of the filtered noise.

Solution The transfer function *H*(*f*) of this filter is

$$H(f)=\frac{1}{1+j\omega RC}$$

The PSD of the input white noise process is

$$P_X(f)=\eta/2$$

$\therefore$ from Eq. (6.87), we know

$P_Y(f)$ = power spectral density of the output noise process

$$=|H(f)|^2\cdot P_X(f)=\frac{\eta/2}{1+(\omega RC)^2}$$

Taking the inverse Fourier transform of $P_Y(f)$, we get

$$R_Y(\tau)=\frac{\eta}{4RC}e^{-\left(\frac{|\tau|}{RC}\right)}$$

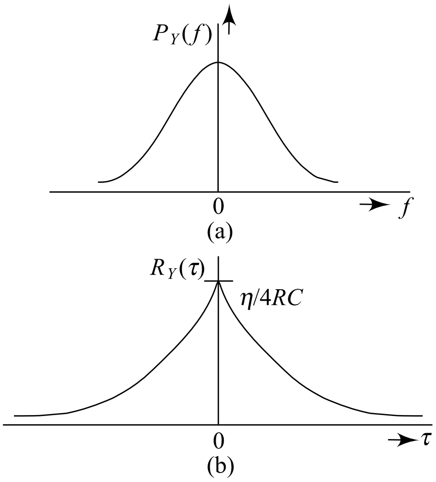


Fig. 6.32 *(a) $P_Y(f)$ of output noise process (b) IFT of $P_Y(f)$*

Example 6.29 A parallel resonant circuit resonant at 100 MHz has a capacitance of 20 pF. If the *Q*-factor of the circuit at resonance is 40, and the circuit temperature is 17°C, what is the equivalent noise bandwidth of the tuned circuit?

Solution Effective or equivalent series resistance r of the tuned circuit = $r = \dfrac{X_c}{Q}$

$$= \frac{1}{2\pi \times 100 \times 10^{6} \times 20 \times 10^{-12} \times 40} = \frac{1}{2\pi \times 8 \times 10^{-12}} = 1.9894\ \Omega$$

$\therefore$ effective parallel resistance $= Q^2 \cdot r = 1600 \times 1.9894 = 3183\ \Omega = R_d$

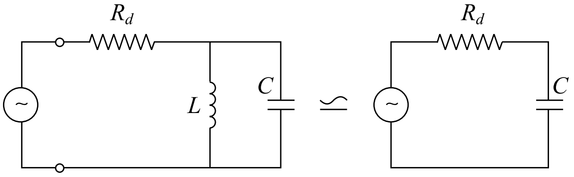


Fig. 6.33 *Noise model of the parallel resonant circuit and its approximate equivalent circuit beyond resonance frequency*

Since this is an RC low pass filter, the noise equivalent bandwidth is given by

$$B_N = \frac{1}{4R_dC} = \frac{10^{12}}{4 \times 3483 \times 20}$$

$$= \frac{10^6}{25464}$$

$$= 39.27\ \text{MHz}$$

Example 6.30 Determine the noise equivalent bandwidth of a normalized low pass Butterworth filter of order 2.

Solution The squared-magnitude response of a Butterworth filter of order n is given by

$$|H_n(f)|^2 = \frac{1}{1 + (f/B)^{2n}}$$

where B is the 3 dB cut-off frequency. Hence, for a normalized second-order Butterworth filter, putting $n = 2$ and $B = 1$, we get

$$|H_2(f)|^2 = \frac{1}{1 + f^4}$$

$\therefore$
$$|H_2(f)|^2\big|_{f=0} = \frac{1}{1} = 1 = |H_2(0)|^2$$

$\therefore$ the noise equivalent bandwidth of a second-order Butterworth filter is given by (refer to Eq. (6.142)

B_N = Noise equivalent bandwidth

$$= \frac{\int_0^{\infty} |H_2(f)|^2 df}{|H_2(0)|^2} = \int_0^{\infty} \frac{1}{1 + f^4} df \qquad \text{since } |H_2(0)| = 1$$

But
$$\int_0^{\infty} \frac{x^{m-1}}{1 + x^n} dx = \frac{\pi/n}{\sin\,(m\pi/n)} \quad \text{for } n > m > 0 \qquad \text{(Refer to Appendix A)}$$

$\therefore$
$$\int_0^{\infty} \frac{1}{1 + f^4} df = \frac{\pi/4}{\sin(\pi/4)} = \frac{1}{\sqrt{2}} \cdot \frac{\pi}{4} \qquad \therefore B = \text{Noise equivalent bandwidth} = \frac{1}{\sqrt{2}} \cdot \frac{\pi}{4}$$

6.17.1 Equivalent Noise Resistance

In noise calculations, it is often quite convenient to represent the noise arising from a device or a whole system like say a radio receiver, by the thermal noise generated *by a fictitious resistance* R_{eq} *at room temperature connected at the input of the device or system, with the device or the system itself considered as totally noiseless*. The idea is that R_{eq} connected at the input would produce at the output of the noiseless device / system, the same mean-squared value of noise as is being produced by the noisy device / system itself.

Example 6.31 For a two-stage amplifier with the following details, calculate the equivalent input noise resistance:
First stage: Voltage gain 12: Input resistor 500 Ω; Equivalent noise resistance 1000 Ω; Output resistor 30 kΩ
Second stage: Voltage gain 20: Input resistor 90 kΩ; Equivalent noise resistance 10 kΩ; Output resistor 500 kΩ

Solution We shall start from the output side of stage 2 and work backwards.

Step 1: A resistance of 500 k at the output of the 2nd stage is equivalent, insofar as noise contribution is concerned, to a resistor of $\frac{500 \times 10^3}{(20)^2} = 1.25\text{ k}$ at the input of the 2nd stage.

Step 2: The resistor of the first stage (30 k) and the input resistor of the second stage (90 k) are in parallel and this parallel combination is in series with the noise equivalent resistance (10 k) of the second stage and the 1.25 kΩ obtained in step 1.

i.e.,
$$\left(\frac{30 \times 90}{30 + 90} + 10 + 1.25)\right) \times 10^3\ \Omega = 33.75\text{ k}\Omega$$

Step 3: The resistance of 33.75 kΩ obtained at the output of the first stage will be equivalent, insofar as noise contribution is concerned, to a resistor of $\frac{33.75 \times 10^3}{(10)^2} = 337.5\ \Omega$ connected at the input of the first stage. But this stage already has at its input, a 500 Ω input resistor and a 1000 Ω noise equivalent resistance of the first stage. Hence, the total noise resistance at the input of the first stage amplifier is

$$R_{eq} = 500 + 1000 + 337.5 = 1837.5\ \Omega$$

6.18 NOISE FIGURE AND EQUIVALENT NOISE TEMPERATURE OF TWO-PORT NETWORKS

6.18.1 Signal-to-Noise Ratio

As mentioned earlier in the discussion on the result of Example 6.20, in communication engineering the values of signal and noise are individually not of much significance. It is their relative strength that matters. Hence, we will always be interested in the ratio of signal power to noise power rather than the signal power alone or the noise power alone. Thus, we define the Signal-to-Noise Ratio (*SNR*) as

$$\boxed{SNR \triangleq \frac{\text{Signal power}}{\text{Noise power}}} \tag{6.144}$$

Note that SNR is a ratio of powers and not of voltages. We may talk about the *SNR* at the input or the output of an amplifier. It is generally more convenient to express the *SNR* in decibels rather than as just a ratio.

$$\boxed{(SNR)_{\text{dB}} = 10 \log_{10}\left[\frac{\text{Signal power}}{\text{Noise power}}\right]} \tag{6.145}$$

Modification of SNR by an amplifier Consider an amplifier with a *power gain G*. Let this amplifier have an input *SNR* of $(S/N)_i$.

$$\text{Input } SNR\left(\frac{S}{N}\right)_i = \frac{\text{Input Signal Power}}{\text{Input Noise Power}} = \left(\frac{S_i}{N_i}\right)$$

The amplifier amplifies both the signal power as well as the noise power by the same factor, G. Further, since the amplifier contains some noise producing elements like resistors and electron devices, it produces some *additional* noise power, say N_a, *at the output*. Hence, at the output side, we have

$$\text{Signal Power} = G \cdot S_i$$

$$\text{Noise Power} = (G \cdot N_i + N_a)$$

$$\text{Thus, the } SNR \text{ at the output} = \left(\frac{S}{N}\right)_0 = \frac{G \cdot S_i}{(G \cdot N_i + N_a)} < \frac{S_i}{N_i} \tag{6.146}$$

Therefore, for any amplifier, or, for that matter, for any two-port network with some noise producing active/passive elements in it, the *SNR* at the output will always be less than the *SNR* at the input, i.e., there is a deterioration of the signal-to-noise ratio. *Thus, an amplifier does not improve the signal-to-noise ratio, it only degrades it.*

Noise figure In Eq. (6.146), if the noise power at the output contributed by the amplifier / linear two-port network due to the noise generated within, viz., N_a, were to be zero, i.e., if the amplifier was totally noise-free, then output *SNR* would have been equal to the input *SNR*. The measure of how noisy an amplifier is, can therefore be obtained from the ratio of the input *SNR* to the output *SNR*. This ratio will have a value of 1 if the amplifier/two-port linear network is totally noise-free and a value greater than unity otherwise. How large the ratio is compared to unity would give us an indication of how noisy the amplifier/two-port linear network is.

This ratio is called the '*noise figure*' of the amplifier.

$$\boxed{F = \text{Noise Figure} \triangleq \frac{(S/N)_i}{(S/N)_0} = \frac{N_0}{GN_i} = \frac{GN_i + N_a}{GN_i}} \tag{6.147}$$

With regard to the 'noise figure', there are a few points that need to be noted. These are:

1. If we consider the ratio of N_0 to GN_i at a single frequency, then the noise figure so obtained is called the *Spot Noise Figure*. The frequency at which it is valid should also be stated along with the spot noise figure, as the value would be different at different frequencies.
2. If the total noise powers (over the entire bandwidth that is of interest to us) at the output and input are considered, then the ratio of N_0 to GN_i gives what is called the *integrated noise figure*.
3. The integrated noise figure is the one most generally used, firstly because it is more realistic and secondly because it can be measured more easily. However, it is the *Single frequency noise figure*, or the *Spot noise figure* which is most easily computed.
4. Power spectral density represents power as a function of frequency. Hence the spot noise figure can be obtained as function of frequency by taking the ratio of power spectral densities of N_0 and GN_i.
5. We know that the maximum noise power that a two-port network can deliver to a load can be obtained only under matched conditions, and since an amplifier amplifies the noise power available at its input terminals, the noise figure is defined only in terms of available noise powers, so that mismatches, if any, are automatically taken care of.

Available output and internal noise powers in terms of F

1. From Eq. (6.147), we have available output noise power $N_0 = F \cdot GN_\text{i}$. Now making use of Eq. (6.107) for N_i, we have

$$\begin{aligned} N_0 &= \text{Available output noise power} \\ &= FGkT_0(\Delta f) \end{aligned} \tag{6.148}$$

where T_0 is the room temperature. In RHS of Eq. (6.148), $GkT_0(\Delta f)$ is the component of the output noise power obtained by amplification of the available input noise power $kT_0(\Delta f)$. If the amplifier had been noise-free, the output noise power would have been only this component, i.e., $GkT_0(\Delta f)$. However, due to the noise internally generated in the amplifier, it is increased by a factor $F(>1)$.

2. From Eq. (6.148),

$$\frac{N_0}{G} = FkT_0(\Delta f)$$

This is the total output noise including the internally generated noise, referred to the input. Of this, $kT_0(\Delta f)$ is the available noise power at the input terminals because of the source. Hence, the internally generated noise, referred to the input, is given by

$$N_a' = \text{Internally generated noise referred to input} = (F-1)\,kT_0(\Delta f) \qquad (6.149)$$

Example 6.32 An amplifier has a noise figure of $F = 12$ dB. Express the internally generated component of the output noise power as a fraction of the available output noise power.

Solution From Eq. (6.149), internally generated noise, referred to the input $= (F-1)kT_0(\Delta f)$

From Eq. (6.148), available output noise power = $FGKT_0(\Delta f)$

Now, internally generated noise referred to the output = $G(F-1)kT_0(\Delta f)$

∴ internally generated component of output noise power/available output noise power

$$= \frac{(F-1)GkT_0(\Delta f)}{FGkT_0(\Delta f)} = \frac{(F-1)}{F}$$

Here, F is in the form of a ratio of the SNRs and not in decibels.

∴ we should convert the given value of F into a ratio.

For this, we note that $F(\text{in dB}) = 10\log_{10}\left[\frac{(S/N)_i}{(S/N)_o}\right] = 12$

$$\therefore \qquad \frac{(S/N)_i}{(S/N)_o} = 10^{1.2} = 15.85$$

$$\therefore \qquad \frac{F-1}{f} = \frac{14.85}{15.85} = 0.9369$$

$$\therefore \qquad \frac{\text{Internally generated component of noise power}}{\text{Available ouput noise power}} = 0.9369$$

Example 6.33 The available output noise power from an amplifier is 80 nW, the available power gain of the amplifier being 40 dB and the equivalent noise bandwidth being 25 MHz. Calculate the noise figure, assuming T_0 to be 27° C.

Solution From Eq. (6.148), we know that the available output noise power N_0 is given by

$$N_0 = FGkT_0(\Delta f)$$

where T_0 is the room temperature and given to be 27° C = 300° K.

$$\therefore \qquad F = \frac{N_0}{GkT_0(\Delta f)} = \frac{80\times 10^{-9}}{10^4 \times 1.38\times 10^{-23}\times 25\times 10^6 \times 300}$$

$$= \frac{2318}{300} = 7.7267$$

$$\therefore \qquad 10\log_{10} F = 10\log_{10} 7.7267 = 8.879 \text{ dB}$$

6.18.2 Equivalent Noise Temperature

Although the noise figure F gives a good measure of the degree of noisiness of a device, amplifier, or any two-port linear network, there is one disadvantage with it. We know that it is equal to one for a noise-free network and that the greater the value of F, the noisier the amplifier / network is. Thus, for low-noise microwave devices and amplifiers, the value of F is very close to one. It then becomes difficult to compare the 'noisiness' of two low-noise amplifiers by comparing their noise figures. A good alternative in such cases is to use what is called the *equivalent noise temperatures* of these amplifiers. Since this also tells us how noisy a device or circuit is, it must be related to the noise figure F. We shall now define the term *noise equivalent temperature* and then see how it is related to F.

Definition The equivalent noise temperature of a device or a two-port linear network is a fictitious temperature T_e which is such that the available noise power at that temperature, viz., $kT_e(\Delta f)$ is equal to the internally generated noise power of the device or the two-port network referred to its input.

From Eq. (6.147), we have

$$F = \frac{GN_i + N_a}{GN_i}$$

From the above definition, it is clear that N_a in the RHS of the above can be replaced by $G[kT_e(\Delta f)]$. Further, we know that

$$N_i = kT_0(\Delta f); \quad T_0 = \text{Room Temp}$$

Hence,

$$F = \frac{GN_i + N_a}{GN_i} = \frac{GkT_0(\Delta f) + GkT_e(\Delta f)}{GkT_0(\Delta f)} = \frac{T_0 + T_e}{T_0}$$

$$\therefore \quad \boxed{F = 1 + \left(\frac{T_e}{T_0}\right)} \tag{6.150}$$

or

$$\boxed{T_e = (F - 1)T_0} \tag{6.151}$$

From Eq. (6.151), it is clear as to why the use of T_e is preferable for low noise devices/amplifiers. The small difference between F and 1 for these low noise amplifiers is magnified by getting multiplied by T_0, the room temperature in degrees Kelvin (i.e., nearly 300).

6.18.3 Noise Figure of Amplifiers in Cascade

In communication engineering, quite often we come across a number of amplifiers or two-port networks connected in cascade. It then becomes necessary to determine the overall noise figure of the cascade connection in terms of the noise figures of the individual amplifiers or two ports.

In this connection, let us recapitulate the following

1. From Eq. (6.125), we have

$$F = \frac{GN_i + N_a}{GN_i} = \frac{\text{Actual output noise power}}{\text{Noise output power if the amplifier is noise-free}} \tag{6.152}$$

2. If we have an amplifier with noise figure F, available power gain G, and an available input noise power $kT_0(\Delta f)$, its output noise power (total) will be $FG\,kT_0(\Delta f)$ since F is defined with reference to available noise power. Also, the noise power internally generated by the amplifier, when referred to the input, is given by $(F–1)\,kT_0(\Delta f)$ Eq. (6.149).

Now consider the cascade connection of two amplifiers as shown in Fig. 6.34.

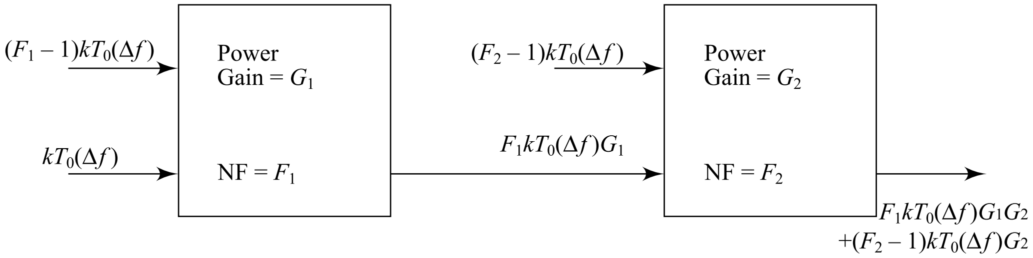


Fig. 6.34 *Cascade connection of two amplifiers*

Then the overall noise figure F is given by

$$F = \frac{\text{Actual output noise power}}{\text{Output noise power assuming the amplifiers to be noise-free}}$$

$$= \frac{F_1kT_0(\Delta f)G_1G_2 + (F_2 - 1)kT_0(\Delta f)G_2}{kT_0(\Delta f)G_1G_2}$$

$$\therefore \qquad F = F_1 + \frac{(F_2 - 1)}{G_1} \tag{6.153}$$

This is known as *Frii's formula*. It may be extended to any number of amplifiers connected in cascade.

$$\boxed{F = F_1 + \frac{(F_2 - 1)}{G_1} + \frac{(F_3 - 1)}{G_1G_2} + \frac{(F_4 - 1)}{G_1G_2G_3} + \dots} \tag{6.154}$$

Here, F_1, F_2, F_3, . . . are the noise figures and G_1, G_2, G_3, . . . are the available power gains of the first, second and third amplifiers, etc.

6.18.4 Improvement of Overall Noise Figure

1. From Eq. (6.154), it is clear that *if the available power gain G_1 of the first amplifier is quite large, the overall noise figure F of the cascade connection will be approximately equal to the noise figure of the first system* in the cascade connection.
2. Since our objective is to have a low overall noise figure, *it becomes necessary to choose a system with high power gain and low noise figure as the first stage* in a chain of cascade amplifiers. In a superheterodyne radio receiver, as already mentioned earlier, the mixer stage is the most noisy. That is why it is always preferable to precede it with a high gain RF amplifier having a low noise figure, so that the overall noise figure is not allowed to be affected by the presence of the noisy mixer stage.

6.18.5 Equivalent Noise Temperature of Cascaded Amplifiers

Let the individual stages have equivalent noise temperatures T_{e1}, T_{e2}, T_{e3}, . . . and available power gains G_1, G_2, G_3, ... Let the room temperature be T_0. If the equivalent noise temperature of the cascade connection is say T_e, then from Eqs. (6.131) and (6.128), we have

$$1 + \frac{T_e}{T_0} = 1 + \frac{T_{e1}}{T_0} + \frac{T_{e2}}{G_1T_0} + \frac{T_{e3}}{G_1G_2T_0} + \dots$$

$$\therefore \qquad \boxed{T_e = T_e + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1G_2} + \dots} \tag{6.155}$$

Example 6.34 A source with an internal resistance of 50 Ω and an internal emf of 6 µV is supplying the signal voltage to an amplifier that has an input resistance of 75 Ω. The amplifier has an equivalent noise resistance of 1470 Ω. For a noise bandwidth of 5 kHz, calculate the output (S/N) ratio in dB at room temperature of 290° K.

Solution The signal voltage V_s developed across the input resistance of 75 Ω is the signal voltage actually available at the input of the amplifier. Hence, we will use the Thevenin's equivalent circuit of this.

Mean-squared value of the noise voltage

$$= 4kT_0(R_{\text{th}} + R_{\text{eq}})(\Delta f)$$

$$= 4 \times 1.38 \times 10^{-23} \times 290 \times 5 \times 10^3 (30 + 1470)$$

$$= 12 \times 10^{-14} \text{ V}^2$$

Mean-squared value of the signal voltage $= (3.6)^2 \times 10^{-12} \text{ V}^2$

$$= 12.96 \times 10^{-12} \text{ V}^2$$

$$\text{Hence } (S/N) \text{ ratio} = \frac{12.96 \times 10^{-12}}{12 \times 10^{-14}} = 108$$

$$\therefore \quad \left(\frac{S}{N}\right)_{\text{dB}} = 10 \log_{10} 108 = 20.3 \text{ dB}$$

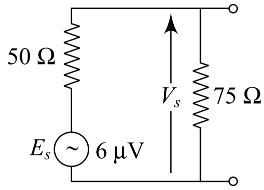


Fig. 6.35 *Signal source voltage circuit*

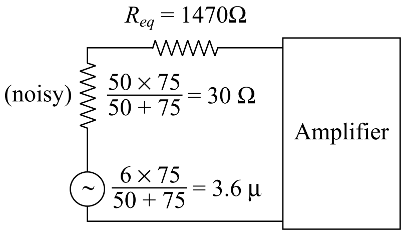


Fig. 6.36 *Input circuit of the amplifier. Noise equivalent resistance of the amplifier is included*

Example 6.35 Determine (a) the noise figure *F* of the amplifier of Example 6.34 in dB. (b) Also determine its equivalent noise temperature.

Solution

(a) We know from Eq. (6.152) that the noise figure *F* is given by

$$F = \frac{\text{Actual output noise power}}{\text{Output noise power assuming the amplifiers to be noise-free}}$$

In our case, actual noise output power =

$$= [4kT_0(R_{\text{th}} + R_{\text{eq}})(\Delta f)]G = G \times 12 \times 10^{-14} \text{volt}^2$$

where R_{eq} is the equivalent noise resistance of the amplifier referred to the input, and G is the available power gain.

Noise output power assuming the amplifier to be noise free

$$= [4kT_0R_{th}(\Delta f)]G = G \times 2.4 \times 10^{-15}$$

$$\therefore \quad F = \frac{G \times 12 \times 10^{-14}}{G \times 2.4 \times 10^{-15}} = 50 = 10 \log_{10} 50 = 16.9 \text{ dB}$$

Note

As can be seen from the above steps, F is in fact, given by

$$F = \frac{R_{th} + R_{eq}}{R_{th}} = 1 + \frac{R_{eq}}{R_{th}} = 1 + \frac{1470}{30} = 1 + 49 = 50$$

(b) Since $F = 1 + \frac{T_e}{T_0}$, we have $T_e = (50 - 1)\,290 = 14{,}210° \text{ K}$

Example 6.36 A low-noise amplifier of equivalent noise temperature 30° K and 20 dB available power gain precedes a microwave receiver which has a noise figure of 25 dB. What is the overall noise equivalent temperature if the room temperature is 27° C?

Solution $T_e = T_{e_1} + \frac{T_{e_2}}{G_1}$ (see Eq. (6.154))

$T_{e_1} = 30°\ K$, $G_1 = 20 \text{ dB} = 100$;

F_2 = Noise figure of the microwave receiver = 25 dB = $10^{2.5}$ = 316.228

∴ Equivalent noise temperature of the microwave receiver = $(F_2 - 1)T_0$
= (315.228) × (273 + 27) = 315.228 × 300 = 945.684° K

∴ Overall noise equivalent temperature = $T_e = T_{e_1} + \frac{T_{e_2}}{G_1}$

$$= 30° + \frac{316.228}{100} = 33.16228° \text{ K}$$

6.18.6 Equivalent Noise Temperature and Noise Figure of a Lossy Line

Let us consider a lossy transmission line of power loss L where L is the ratio of input power to the output power. Let it be terminated on both sides by R_0 ohms, its characteristic resistance as shown in Fig. 6.37.

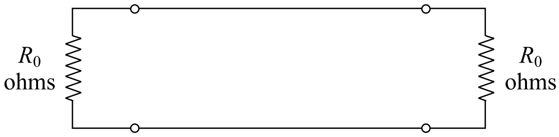


Fig. 6.37 *Lossy transmission line terminated in its R_0*

For simplicity, let us assume that the line and the resistances of R_0 ohms each are all at the ambient temperature T_0° K. The lossy line acts as a thermal source and if T_e is its equivalent noise temperature, and $g_L = 1/L$ is its gain, then from the way T_e has been defined as the internally generated noise power available at the output referred to the input, we may write

$$\left.\begin{matrix}\text{Available internally generated} \\ \text{noise power at the output}\end{matrix}\right\} = kT_e g_L(\Delta f)$$

But this must be equal to the total noise power available at the output minus the noise power generated in the R_0 at the input side and made available at the output-end of the line.

Since the noise power contributed by input side R_0 and made available at the output end is given by $kT_0 g_L(\Delta f)$, if the total available noise power at the output is $kT_0(\Delta f)$, we may write

$$kT_e g_L(\Delta f) = kT_0(\Delta f) - kT_0 g_L(\Delta f)$$

∴ $$T_e = \left(\frac{T_0}{g_L} - T_0\right) = T_0(L - 1)$$

∴ $$T_e = T_0(L - 1)$$

But we know, from Eq. (6.151) that $T_e = T_0(F - 1)$

Hence, combining the above two equations, we get

$$\boxed{F = \text{Noise figure of the lossy line} = L}$$

Example 6.37 In TV receivers, the antenna is often mounted on a tall mast and a long lossy cable is used to connect the antenna to the receiver. To overcome the effect of the lossy cable, a pre-amplifier is mounted on the antenna as shown in Fig. 6.38(a).

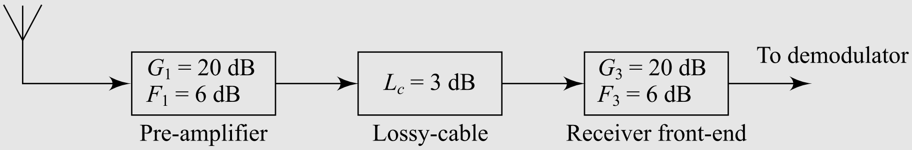


Fig. 6.38(a) *Antenna with pre-amplifier*

(a) Find the overall noise figure of the system.

(b) Find the overall noise figure of the system if the pre-amplifier is omitted and the gain of the front-end is increased by 20 dB. **(VTU, March, 2001)**

Solution We know, from the derivation given above, that for a lossy cable, the noise figure (ratio) equals its power loss. So, in our case,

$$F_c = \text{Noise figure of the lossy cable} = L\ (\text{ratio}) = 2$$

(a) Applying Frii's formula for the overall noise figure,

$$F = F_1 + \frac{(F_C - 1)}{G_1} + \frac{(F_3 - 1)}{G_1 \cdot G_C}$$

where $G_C = (1/LC)$ = gain of the cable

$$\therefore \qquad F = 3.981 + \frac{(2-1)}{100} + \frac{(39.8-1)}{(100 \times 1/2)}$$

$$= 4.767 = 6.782 \text{ dB}$$

$6 \text{ dB} = 10^{0.6} = 3.981$

$3 \text{ dB} = 10^{0.3} = 2$

$16 \text{ dB} = 10^{1.6} = 39.8$

$60 \text{ dB} = 10^{6} = 10^{6}$

$20 \text{ dB} = 10^{2} = 100$

(b) When the pre-amplifier is omitted and the gain of the front-end is increased by 20 dB, the system configuration is as given in Fig. 6.38(b).

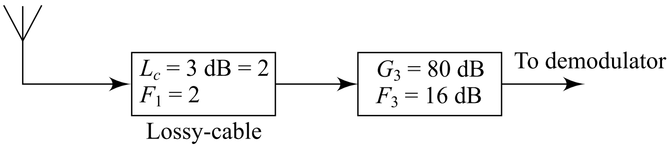


Fig. 6.38(b) *System configuration for part (b)*

$\therefore$ overall noise-figure F is now given by

$$F = F_C + \frac{(F_3 - 1)}{(1/LC)} = 2 + \frac{(39.8-1)}{(1/2)} = 79.6$$

$$\therefore \qquad F_{\text{dB}} = 10 \log_{10} 79.6 = 19.01 \text{ dB}$$

Example 6.38 A satellite receiving system consists of a low noise amplifier (LNA) that has a gain of 47 dB and a noise temperature of 120° K, a cable with a loss of 6.5 dB and the main receiver with a noise factor of 7 dB. Calculate the equivalent noise temperature of the overall system referred to the input for the following system connections:

(a) LNA at the input followed by the cable connecting to the main receiver.

(b) The input direct to the cable, which is then connected to the LNA, which, in turn, is connected to the main receiver. **(VTU, February, 2002)**

Solution As the value of the ambient temperature, T_0 is not given, let us conveniently assume it as 17° C = 290° K.

(a) For the first case, configuration is as follows

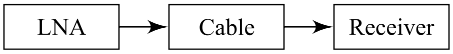


Fig. 6.39(a) *First configuration for Example 6.38*

$G_1 = 47 \text{ dB} = 10^{4.7}$
$= 50118.72$
$T_{e_1} = 120° \text{K}$

$L_C = \text{loss} = 6.5 \text{ dB}$
$= 4.4668 = 1/G_c$
$\therefore\ F_C = 4.4668$
$T_{e_c} = (F_c - 1) = 1005.38$

$F_3 = 7\text{dB}$
$= 10^{0.7} = 5.0118$
$T_{e_c} = (F_3 - 1)T_0$
$= 4.0118 \times 290 = 1163.442$

Applying Frii's formula for the overall equivalent noise temperature T_e (see Eq. (6.154))

$$T_e = T_{e_1} + \frac{T_{e_c}}{G_1} + \frac{T_{e_3}}{G_1 \cdot G_C} = 120 + \frac{1005.38}{50118.72} + \frac{1163.442}{(50118.72/4.4668)}$$
$$= 120 + 0.0200 + 0.10369 = 120.12° \text{K}$$

(b) For the second case, the configuration is as follows

Fig. 6.39(b) *Second configuration for Example 6.38*

$L_C = loss = 6.5 \text{ dB}$
$= 10^{0.65} = 4.4668$
$F_C = 10^{0.65} = 4.4668$
$T_{e_c} = 1005.38; G_C = 1/L_C$

$G_1 = 47 \text{ dB} = 10^{4.7}$
$= 50118.72$
$T_{e_1} = 120° \text{K}$

$F_3 = 7 \text{ dB} = 10^{0.7} = 5.0118$
$T_{e_c} = (F_3 - 1)T_0$
$= 4.0118 \times 290 = 1163.442$

$\therefore$ applying Frii's formula for the overall equivalent noise temperature T_e, we have

$$T_e = T_{e_c} + \frac{T_{e_1}}{G_C} + \frac{T_{e_3}}{G_C G_1} = 1005.38 + \frac{120}{(1/4.4668)} + \frac{1163.442}{(50118.72/4.4668)}$$
$$= 1005.38 + 536.016 + 0.10369 = 1541.499° \text{K}$$

Note *From the definition of the equivalent noise temperature of a two-port, we know that it is the temperature T_e which is such that the available noise power at that temperature, viz., $kT_e(\Delta f)$, is equal to the internally generated noise power of the two-port, referred to its input.*

Example 6.39 A coil having an inductance of 2 H and an internal resistance of 1 ohm is shunted by a capacitor of 2 F. Determine the power density spectrum of the thermal noise the network terminals.

(GATE Exam, 1991)

Solution Thermal noise is produced in the circuit only by the resistance of the coil. We shall therefore draw the equivalent circuit as shown in Fig. 6.40.

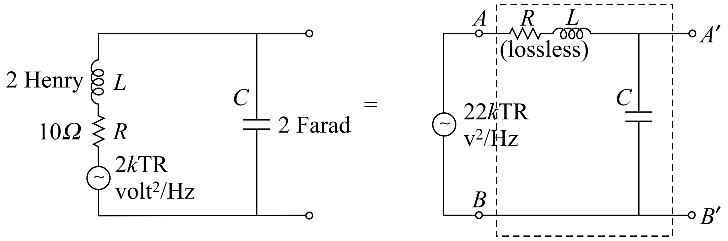


Fig. 6.40 *Circuit for Example 6.39*

The two-port network shown inside the dotted line box has input power spectral density of $2kTR$ V^2/Hz. Hence, its output power spectral density (PSD) will be

$$\text{Output PSD} = (\text{Input PSD}) \times |H(f)|^2$$

where $H(f)$ is the transfer function of the two-port.

Now,

$$H(f) = \frac{\dfrac{1}{j\omega C}}{R + j\omega L + \left(\dfrac{1}{j\omega C}\right)} = \frac{1}{(1-\omega^2 LC) + j\omega RC}$$

$$\therefore \qquad |H(f)|^2 = \frac{1}{(1-\omega^2 LC)^2 + \omega^2 R^2 C^2}$$

Substituting the values of R, L, and C, we get

$$|H(f)|^2 = \frac{1}{(1-4\omega^2)^2 + 4\omega^2} = \frac{1}{1+16\omega^4 - 4\omega^2}$$

$\therefore$ output PSD = $2kTR|H(f)|^2$ V^2/Hz

$$= \frac{2kT}{(1+16\omega^4 - 4\omega^2)} \quad \text{since } R = 1\Omega$$

6.18.7 Measurement of Noise Figure

In Section 6.16, of this chapter, we had shown that a temperature-limited vacuum diode carrying a plate current of I_0 amperes generates shot noise current component whose mean-squared value is given by Eq. (6.139) as

$$I_n^2 = 2I_0 q(\Delta f)\,\text{amp}^2$$

Thus, the temperature-limited vacuum diode can be used as a noise source. As stated earlier, this noise is not exactly white, but has a flat spectral density even up to a few hundred megahertz. Hence, for most of our communication systems, for which the carrier frequencies are in the RF, this source can be regarded as a white noise source. For microwave communication systems, one may make use of noise generators which use a fluorescent tube placed inside a waveguide as a noise source.

A simple setup for the measurement of the noise figure F of a two-port network, using a temperature-limited vacuum diode as the noise source, is shown in Fig. 6.41.

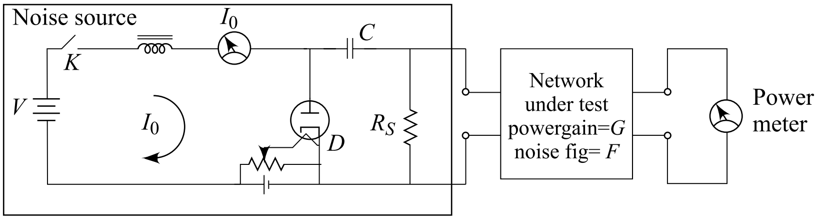


Fig. 6.41 *Measurement of noise figure*

Let us assume that the value of C is such that its effect can be ignored.

Initially, we make $I_0 = 0$ by opening the switch K and note the power meter reading. Let it be P_1. Then P_1 is the output noise power with input noise power being only the thermal noise generated by the R_S which is actually the parallel combination of the output resistance of the source and the input resistance of the 2-port network under test. Hence, from Eqs. (6.113) and (6.152), we have

$$P_1 = 4kTR_S(\Delta f)GF \qquad (6.156)$$

With the switch K now closed and with a diode direct current of I_0, let the output noise power be P_2. Now, this P_2 is caused by an input noise power consisting of shot-noise and thermal noise.

$$\therefore \quad P_2 = 4kTR_S(\Delta f)GF + 2qI_0R_S^2G(\Delta f) \qquad (6.157)$$

$$\therefore \quad \frac{P_2}{P_1} = 1 + \frac{2qI_0R_S^2G(\Delta f)}{4kTR_S(\Delta f)GF} = 1 + \frac{qI_0R_S}{2kTF}$$

Note *In Eq. (6.156), the first term represents the output noise power including the amplified thermal noise power given to the input and the noise power at the output due to the internally generated noise. Hence, to get the total output noise power when both thermal noise and shot-noise are present at the input, we merely add to the first term the output shot noise power which is G times the input shot noise power.*

Let us now adjust the cathode temperature of the diode (by adjusting the filament voltage) so that P_2 / P_1 becomes 2. Let the new plate current under this condition be I_0'. Then

$$2 = 1 + \frac{qI_0'R_S}{2kTF}, \text{ i.e.,} \quad F = \frac{qI_0'R_S}{2kT}$$

Now, if $T = 290°$ K,

$$\therefore \quad F = \frac{1.6 \times 10^{-19} I_0'R_S}{2 \times 1.38 \times 10^{-23} \times 290} = 20I_0'R_S$$

$$\therefore \quad F = 20I_0'R_S$$

Thus, we will be able to determine the value of the noise figure of the two-port network from the values of I_0' and R_S.

6.19 NARROWBAND NOISE REPRESENTATION

In any communication system, the front-end of the receiver will be designed to have a bandwidth just equal to the bandwidth of the transmitted signal. For example, in the case of AM audio broadcasting, 5 kHz being the maximum audio frequency handled by the transmitter, the transmitted amplitude modulated signal occupies a bandwidth of 10 kHz, five kilohertz on either side of the carrier. Hence, the front-end of an AM broadcast receiver is designed to have a bandwidth of just 10 KHz. While a smaller than the required bandwidth for the front-end of the receiver results in distortion of the received signal, a larger than required bandwidth would only allow more noise power to enter the receiver without any increase in the signal power.

If the channel noise is modeled as a zero-mean white Gaussian Process, and the front-end of the receiver is modeled as a narrowband filter with center frequency f_c, the received noise will then be a narrowband noise process with center frequency f_c and its PSD will be somewhat as shown in Fig. 6.42.

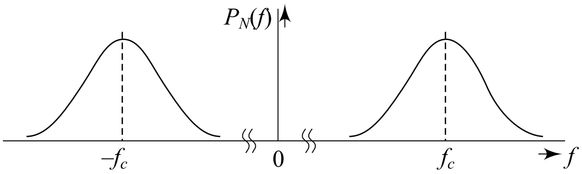


Fig. 6.42 *PSD of noise entering the receiver*

Earlier, in Section 6.13.2, we had shown that it is possible to represent a narrowband signal, $x(t)$, with center frequency f_c in terms of its inphase and quadrature components as

$$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \qquad \text{(Refer to Eq. (6.108))}$$

where the lowpass signal $x_I(t)$ and $x_Q(t)$ were respectively called the inphase and quadrature components of the signal $x(t)$.

In the present case, we are not dealing with a narrow band deterministic signal $x(t)$; instead, we are dealing with a narrowband noise – a narrowband random process $n(t)$. However, we may proceed exactly the same way as we did in Chapter 2 for the deterministic signal case and write

$$\text{Pre-envelope of } n(t) = n_+(t) = n(t) + j\hat{n}(t) \qquad (6.135)$$

and

$$\text{complex envelope of } n(t) = \tilde{n}(t) = n_+(t)\exp[-j2\pi f_c t] \qquad (6.136)$$

Let the low pass complex process $\tilde{n}(t)$ be represented as

$$\tilde{n}(t) = n_I(t) + jn_Q(t) \qquad (6.137)$$

Since $\tilde{n}(t)$ is a low pass process of bandwidth, say W Hz, $n_I(t)$ and $n_Q(t)$ are also low pass of the same bandwidth.

From Eq. (6.135), we have

$$n(t) = \text{Re}[n_+(t)] = \text{Re}[\tilde{n}(t)e^{j\omega_c t}] \qquad \text{(From Eq. (6.136))}$$

Now, substituting for $\tilde{n}(t)$ using Eq. (6.137), we get

$$\therefore \quad n(t) = \text{Re}[\{n_I(t) + jn_Q(t)\}\{\cos\omega_c t + j\sin\omega_c t\}]$$

$$\therefore \quad \boxed{n(t) = n_I(t)\cos\omega_c t - n_Q(t)\sin\omega_c t} \qquad (6.138)$$

RHS of Eq. (6.138) is called the inphase and quadrature component representation of the narrowband noise process $n(t)$ centered on f_c.

There are a few extremely useful properties associated with the inphase and quadrature components, viz., $n_I(t)$ and $n_Q(t)$. These are stated and the pertinent proofs are given in the sections on probability and random processes.

MATLAB Example 6.1 Generate a discrete time sequence of $N = 2000$ independent identically distributed (uniformly) random numbers in the interval [–1/2, 1/2]. Compute the autocorrelation R_x of the sequence $\{X_n\}$. Find the power spectrum of $\{X_n\}$ by finding the DFT of R_x using FFT. [Note: $R_x(k)$ and $S_x(f)$ have to be computed for each value of k and f respectively at least some 10 to 20 times and the average of all the values of the $R_x(k)$ for each k and of $S_x(f)$ for each f must be taken.]

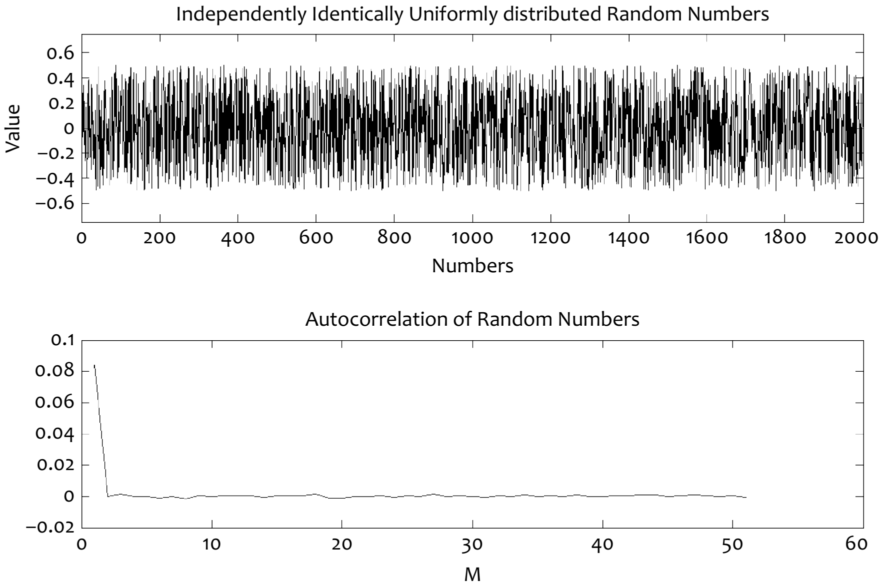
(a) Plot $R_x(k)$ and $S_x(f)$.
(b) Bandpass filter the white Gaussian noise $\{Xn\}$ using bandpass filter.
(c) Determine and plot the autocorrelation and the power spectrum of the output noise.

MATLAB Program

```
clc
N = 2000;                       % Number of samples
M = 50;
Nxav = zeros(1,M+1);
Sxav = zeros(1,M+1);
for i = 1:10   % takes the ensemble average over ten realizations
X = rand(1,N)-(1/2);            % Generate a uniform number sequence on (–1/2,1/2)
    Nx = Nx_est(X,M);           % autocorrelation of x
    Sx = fftshift(abs(fft(Nx))) % power spectrum of x
    Nxav = Nxav+Nx;
    Sxav = Sxav+Sx;
end;
Nxav = Nxav/10;
Sxav = Sxav/10;
figure (1)
subplot (2,1,1)
plot(X)
xlabel('Numbers')
title ('Independently Identically uniformly distributed random numbers');
subplot (2,1,2)
plot(Nxav);
title ('Autocorrelation of random numbers');
xlabel ('M');
figure (2)
subplot (3,1,1)
f = –0.5:1/M:0.5
plot (f,Sxav)
title ('Power spectrum');
xlabel ('Frequency')
%
% Bandpass filter (BPF) the white Gaussian noise {Xn} using
% a BPF response as given
% generation of white noise
%
for i = 1:2:N
    [X1(i) X1(i+1)] = gengauss;
    [X2(i) X2(i+1)] = gengauss;
end
```

```
A = [1 -0.9];
B = 1;
Xc = filter(B,A,X1);                  % in-phase component
Xs = filter(B,A,X2);                  % quadratic component
fc = 2000/pi;
for i = 1:N
    band_pass_process(i) = Xc(i)*cos(2*pi*fc*i)-Xs(i)*sin(2*pi*fc*i);
end
%
% Determine the autocorrelation and the spectrum of bandpass process
%
M = 50;
bpp_autocorr= Nx_est(band_pass_process,M);
bpp_spectrum =fftshift(abs(fft(bpp_autocorr)));
subplot (3,1,2)
plot(bpp_autocorr)
title ('Autocorrelation of Gaussian noise band pass process');
xlabel('M')
subplot (3,1,3)
plot(f,bpp_spectrum)
title ('Spectrum of Gaussian noise band pass process');
xlabel ('frequency');
```

Results



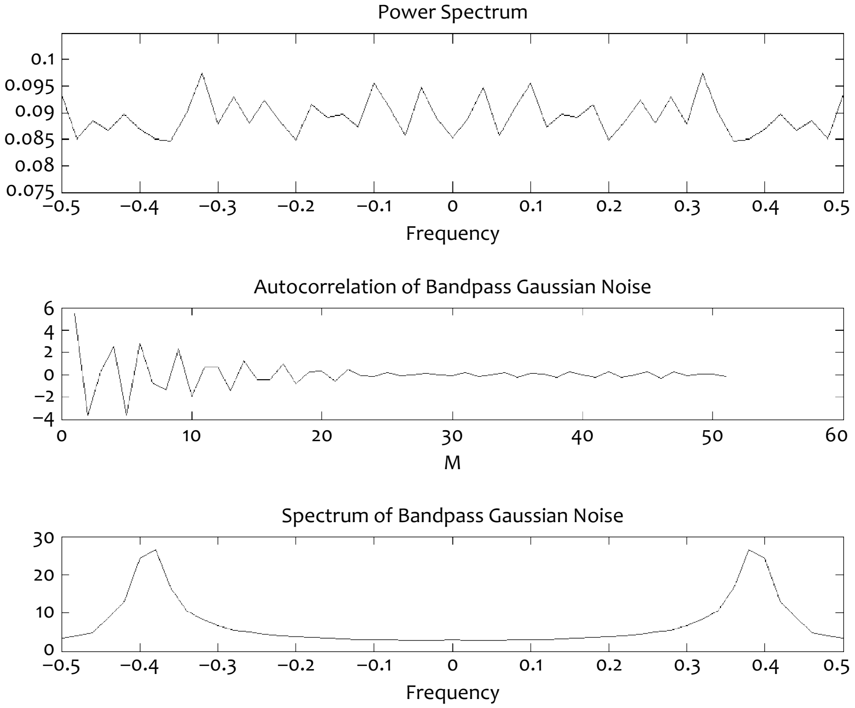


Fig. 6.43

Summary

- Modern probability theory is based on the following axioms:
  (a) If A is an event, $P(A) \geq 0$
  (b) If S is the certain event, $P(S) = 1$
  (c) If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$
- Sample space is the set of all possible outcomes of a random experiment.
- Events are defined in terms of subsets of the sample space forming a Borel field σ.
- Probability is a non-negative number less than or equal to one which is assigned to an event and it has to satisfy certain conditions.
- Conditional probability of A given B is $P(A|B) = \dfrac{P(AB)}{P(B)}$ where $P(AB)$ = Probability of joint occurrence of A & B and $P(B) \neq 0$.
- Bayes' theorem: $P(A_i|B) = \dfrac{P(B/A_i)P(A_i)}{\sum_{i=1}^{n} P(B/A_i)P(A_i)}$
- (a) Events A and B are said to be independent events if $P(AB) = P(A) \cdot P(B)$
  (b) In general, n events, $A_1, A_2, \ldots, A_n$ are said to be independent, if for every $k < n$, the events $A_1, A_2, \ldots, A_k$ are independent and further, if

$$P(A_1, A_2, A_3, \ldots, A_n) = P(A_1)P(A_2)\ldots P(A_n)$$

- A real random variable is a mapping of the outcomes of a random experiment to the real line and satisfying the following two conditions:
  (a) $\{\boldsymbol{X} \le x\}$, i.e., $\{\boldsymbol{X}(\xi) \le x\}$ is an event for $\forall$ real number x.
  (b) $P\{\boldsymbol{X}(\xi) = +\infty\} = P\{\boldsymbol{X}(\xi) = -\infty\} = 0$
- The cumulative distribution function, CDF of a random variable $\boldsymbol{X}$ is denoted by $F_X(x)$ and is defined as: $F_X(x) \underline{\Delta}\, P\{\boldsymbol{X} \le x\}$
- *Properties of CDF*
  (a) $F_X(x)$ lies between 0 and 1
  (b) $F_X(\infty) = 1$ and $F_X(-\infty) = 0$
  (c) $F_X(x)$ is a non-decreasing function of x
  (d) $F_X(x)$ is continuous from the right
  (e) $F_X(b) - F_X(a) = P[\text{a} < \boldsymbol{X} \le b]$
- Random variables are of three types – continuous, discrete and mixed types. Random variables whose CDF is a continuous function is called a continuous random variables. A random variable whose CDF has a staircase shape is called a discrete random variable. A random variable which is neither discrete, nor continuous, is called a mixed random variable.
- The probability density function PDF is defined as $f_X(x) = \frac{d}{dx}[F_X(x)]$
- *Properties of PDF*
  (a) $f_X(x) \ge 0$
  (b) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
  (c) $\int_{x_1}^{x_2} f_X(x)\,dx = P[x_1 < \boldsymbol{X} \le x_2]$ and
  (d) $F_X(x) = \int_{-\infty}^{x} f(\alpha)\,d\alpha$
- (a) Uniform random variable is one whose PDF is constant over a certain interval or range of x.

$$\therefore \qquad f_X(x) = \begin{cases} \dfrac{1}{(x_2 - x_1)}; & x_1 \le x \le x_2 \\ 0 & \text{elsewhere} \end{cases}$$

  (b) A Gaussian random variable is one having a PDF of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2}$$

  where σ^2 = variance and m = mean value of the random variable $\boldsymbol{X}$.
  (c) A Rayleigh random variable is one which has PDF $f_X(x)$ given by

$$f_X(x) = \begin{cases} \left(\dfrac{x}{\sigma^2}\right) e^{-x^2/2\sigma^2}; & x \ge 0 \\ 0; & x < 0 \end{cases}$$

  (d) A Rician random variable is one which has a PDF of the form

$$f_X(x) = \left[\frac{1}{\sigma^2}\, x e^{-(x^2+\mu^2)/2\sigma^2}\right] \cdot I_0\left(\frac{\mu x}{\sigma^2}\right)$$

  where $I_0(\alpha)$ is the modified Bessel function of the first kind and zeroth order.
- A Bernoulli random variable is a discrete random variable which takes the values 1 and 0 with probabilities of P and $(1 - P)$.

- A discrete random variable ***X*** is said to be a binomial random variable with parameters *n* and *p* if
  $P[\boldsymbol{X} = k] = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}; 0 \le k \le n$
- If ***X*** is a random variable and if ***Y*** = g(***X***), then $f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|}$
- The 'mean' or 'expected value' of a r.v. ***X*** is $E\{\boldsymbol{X}\} = \int_{-\infty}^{\infty} x f_X(x)\,dx$, if ***X*** is continuous.
- If ***X*** is discrete random variable, $E\{\boldsymbol{X}\} = \sum_i p_i x_i$, where $p_i = P[\boldsymbol{X} = x_i]$
- The variance of a continuous random variable ***X*** with expected value η_x is defined as
  Variance of ***X*** = $\text{Var}[\boldsymbol{X}] = \sigma_X^2 = E[(\boldsymbol{X} - \eta_X)^2] = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2$
- For a discrete random variable ***X***, $\sigma_X^2 = \sum_i p_i (x_i - \eta_X)^2$ where $p_i = P[\boldsymbol{X} = x_i]$
- The positive square root of variance is called 'Standard Deviation'.
  The characteristic function of a continuous random variable ***X*** is defined as

$$\phi_X(\omega) \,\underline{\Delta}\, \int_{-\infty}^{\infty} f_X(x) e^{j\omega x}\,dx$$

- The characteristic function of a discrete random variable ***X*** which takes values x_i, $i = 1, 2, \ldots$ with probabilities p_i, is given by

$$\phi_X(\omega) = \sum_i p_i e^{j\omega x_i}$$

- The joint, or bivariate distribution function $F_{X,Y}(x, y)$ is

$$F_{X,Y}(x, y) \,\underline{\Delta}\, P[\boldsymbol{X} \le x, \boldsymbol{Y} \le y]$$

- The joint density function of two random variables ***X*** and ***Y*** is

$$f(x, y) \,\underline{\Delta}\, \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

- $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha, \beta)\,d\alpha\,d\beta;\ F_X(x) = F_{X,Y}(x, \infty)$

  $F_Y(y) = F_{X,Y}(\infty, y);\quad f(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\,dy \quad \text{and} \quad f(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\,dx$
- (a) if two random variables ***X*** and ***Y*** are statistically independent,

$$f(y|x) = f(y); f(x|y) = f(x) \quad \text{and} \quad f(x, y) = f(x) \cdot f(y)$$

  (b) Random variables ***X*** and ***Y*** are said to be uncorrelated if their covariance is zero, i.e., if

$$C_{XY} \,\underline{\Delta}\, E[(\boldsymbol{X} - \eta_X)(\boldsymbol{Y} - \eta_Y)] = 0$$

  Then $E[\boldsymbol{XY}] = E[\boldsymbol{X}]E[\boldsymbol{Y}]$

  (c) Two random variables ***X*** and ***Y*** are said to be orthogonal if *E*[***XY***] = 0
- Two random variables ***X*** and ***Y*** are said to be jointly Gaussian if

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right\}\right]$$

- The 'Central Limit Theorem' says that the sum of *n* independent random variables will have a CDF that converges to the CDF of a Gaussian random variable.

- A random process is a function of two variables – time t and outcome ξ and is denoted by $\boldsymbol{X}(t, \xi)$.
- The *mean* of a random process at the instant $t = t_1$ is defined as the expected value (or mean) of the random variable $\boldsymbol{X}(t_1)$.
- The *ACF* of a random process is defined as the expected value of the product of $\boldsymbol{X}(t_1)$ and $\boldsymbol{X}(t_2)$

$$R_X(t_1, t_2) \triangleq E[\boldsymbol{X}(t_1) \cdot \boldsymbol{X}(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2)\, dx_1 dx_2$$

- The *auto co-variance* of a random process $\boldsymbol{X}(t)$ is

$$C_X(t_1, t_2) = E[\{\boldsymbol{X}(t_1) - \eta_X(t_1)\}\{\boldsymbol{X}(t_2) - \eta_X(t_2)\}]$$

- The *auto-correlation* of a random process $\boldsymbol{X}(t)$ is $R_X(t_1, t_2) = E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)]$
  The average power in $\boldsymbol{X}(t) = R_X(t, t) = E[|\boldsymbol{X}(t)|^2]$
  Variance of $\boldsymbol{X}(t) = C_X(t, t) = E[\boldsymbol{X}^2(t)] - \{E[\boldsymbol{X}(t)]\}^2$
- (a) *Cross covariance*: $C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - \eta_X(t_1)\eta_Y(t_2)$
  (b) *Cross correlation*: $R_{XY}(t_1, t_2) = E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)]$
- *Independent Processes*: Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be statistically independent, if the set of random variables $\{\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)\}$ and $\{\boldsymbol{Y}(t_1'), \boldsymbol{Y}(t_2'), \ldots \boldsymbol{Y}(t_n')\}$ are mutually independent for all values of $t_1, t_2, \ldots, t_n$ and $t_1', t_2', \ldots t_n'$ and all integer values of n.
- *Uncorrelated Processes*: $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be uncorrelated process if $C_{XY}(t_1, t_2) = 0$ for all values of t_1 and t_2.
- *Orthogonal Process*: Processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be orthogonal processes if $R_X(t_1, t_2) = 0$ for all t_1 and t_2.
- If $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are orthogonal processes and, in addition, if either (or both) of them has zero mean, then they are uncorrelated.
- *Stationarity*: Random processes, whose statistical properties like mean, ACF, etc., are independent of time, are called stationary processes.
- *WSS*: A process $\boldsymbol{X}(t)$ is said to be stationary in the Wide sense, if its mean, i.e., $E[\boldsymbol{X}(t)]$ is independent of time and if its ACF $R_X(t_1, t_2)$ is such that it is a function only of $(t_2 - t_1)$ and not individually, of t_1 and t_2.
- *Ergodicity*: Random processes for which the time averages equal the ensemble averages, are known as Ergodic processes.
- *Wiener–Khinchine theorem*: The PSD of a random process is the Fourier transform of its auto-correlation.
- *Gaussian random process*: $\boldsymbol{X}(t)$ is a Gaussian random process if the random variables $\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)$ are jointly Gaussian for all values of $t_1, t_2, \ldots, t_n$ and all integer values of n.
- *White noise process*: A process $\boldsymbol{X}(t)$ whose PSD is a constant for all frequencies, is called a white noise process.
- *ACF of a white noise process*: For a white noise process with a PSD of $N_0/2$, the ACF is

$$R_n(\tau) = \frac{N_0}{2}\delta(t)$$

- *LTI systems with random inputs*: If $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are respectively the input and output processes for an LTI system, then:
  (a) Mean of the output $= \eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)\, dt$
  (b) Input and output cross-correlation $= R_{XY}(\tau) = R_X(\tau) * h(-\tau)$
  (c) Correlation of output process $= R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$
  (d) From (iii), it follows that $S_Y(f) = S_X(f) \cdot |H(f)|^2$
- *Low pass sampling theorem for random processes*: If $\boldsymbol{X}(t)$ is a stationary process which is band limited, i.e., if $S_X(f) = 0$ for $|f| \geq W$ Hz and if it is sampled at regular intervals of T_s where $T_s = 1/2W$, then

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s)\,\text{sinc}\, 2W(t - kT_s)\right|^2\right] = 0$$

- *Canonical representation of band pass processes*: A stationary band pass process $\boldsymbol{X}(t) = \boldsymbol{R}(t)\cos[\omega_c t + \boldsymbol{\theta}(t)]$ can be represented in the canonical form, or the inphase and quadrature component form as
$$\boldsymbol{X}(t) = \boldsymbol{X}_I(t)\cos\omega_c t - \boldsymbol{X}_Q(t)\sin\omega_c t$$
where $\boldsymbol{X}_I(t)$ is the inphase component and $\boldsymbol{X}_Q(t)$ is the quadrature component. Both $\boldsymbol{X}_I(t)$ and $\boldsymbol{X}_Q(t)$ are low pass processes, and $R(t) = \sqrt{\boldsymbol{X}_I^2(t) + \boldsymbol{X}_Q^2(t)}; \theta(t) = \tan^{-1}\left[\frac{X_Q(t)}{X_I(t)}\right]$. Further, $\overline{\boldsymbol{X}^2(t)} = \overline{\boldsymbol{X}_I^2(t)} = \overline{\boldsymbol{X}_Q^2(t)}$.
- Noise degrades the performance of communication system.
- Noise sources may be internal to the communication system or may be external to it. Atmospheric noise, Extra-terrestrial noise, and man-made noise are due to external sources while thermal noise, shot noise and partition noise are due to internal sources
- Disturbance caused by atmospheric noise is more severe in the medium wave band as compared to the short wave band, and it is very little in the VHF and UHF bands that are used for television.
- Extra-terrestrial noise can cause disturbance to communications in the frequency range 20 MHz to 1.5 GHz.
- Man-made noise causes disturbance to communications in the 1 MHz–500 MHz frequency range.
- Random motion of electron in a conductor cause thermal noise (also known as Johnson noise). It has zero mean value and has an almost flat spectral density even up to 200 GHz. Hence, for all practical purposes it can be considered as a zero-mean white noise.
- For thermal noise, $P(f) = 2KTR\ \text{V}^2/\text{Hz}$; where K is Boltzmann's constant, T is absolute temperature in °K and R is the resistance in Ohms.
$$\left.\begin{matrix}\text{r.m.s. value of noise voltage across a resistor} \\ \text{of } R \text{ ohms in a bandwdith of } \Delta f \text{ Hz, at T}^\circ k\end{matrix}\right\} = 2\sqrt{KTR(\Delta f)}\ \text{V}.$$
- When two resistors are in series, it is their noise power spectral densities in volt^2/Hz which can be added but not their noise voltages.
- When two resistors are in parallel, it is their noise power spectral densities in amp^2/ Hz which can be added, but not their noise currents.
- A noise whose PSD is flat and independent of frequency, is called 'white' noise. If $(N_0/2)$ is its PSD, then its ACF $= (N_0/2)\ \delta(\tau)$, an impulse. This means that however closely (in time) we may take two samples of a white noise process, the two samples will be un-correlated.
- Shot noise which arises in electron devices, is due to the random emission of electrons from the cathode in the case of vacuum tubes and due to the inherent randomness in the diffusion of minority carriers and the drift of majority carriers across the junction in the case of semiconductor devices.
- Shot noise is approximately a white noise process with a two-sided power spectral density of $I_0 q$ amp^2/Hz, where I_0 is the average current through the device and q is the magnitude of the charge of the charged particles in motion. Its r.m.s value $= \sqrt{2I_0 q(\Delta f)}$ amp, where (Δf) is the bandwidth over which the current is considered.
- In multi-electrode electron devices like triodes, pentodes, BJTs, etc., partition noise is generated due to the random distribution of electrons (or charged carriers) between the various electrodes – grid and plate in the case of triodes and base and collector in the case of a BJT.
- The noise equivalent bandwidth of a filter with transfer function $H(f)$ is defined as the bandwidth B of an ideal LPF whose pass band gain is $H(0)$ such that when a white noise source of PSD $= N_0/2$ is applied as input, the ideal LPF gives the same output noise power as the filter under consideration.
- The equivalent noise resistance R_{eq} of a device or a system is that value of resistance which when connected at the input of the device or the system, with the system itself considered noiseless, produces at its output a mean squared value of the noise which is the same as what is being produced by the device/system itself.
- SNR = Signal-to-noise ratio $\underline{\Delta}\ \frac{\text{Signal Power}}{\text{Noise Power}}$
$$(\text{SNR})_{\text{dB}} = 10\log_{10}\left[\frac{\text{Signal Power}}{\text{Noise power}}\right]$$

- F = Noise Figure $\triangleq \dfrac{\text{SNR at input}}{\text{SNR at output}}$

  For any practical device/system, it is always greater than unity. The closer the value of F is to unity, the better.
- Available output noise power = $FGkT_0(\Delta f)$

  $$\left.\begin{array}{l}\text{Internally generated noise of a system referred to}\\ \text{the input of that system}\end{array}\right\} = (F-1)kT_0(\Delta f)$$
- For low noise amplifiers and devices, it is more convenient to use noise temperature instead of noise figure.

  Definition The equivalent noise temperature of a device, or a two-port linear network, is a fictitious temperature T_e which is such that the available noise power at that temperature, viz., $kT_e(\Delta f)$ is equal to the internally generated noise power of the device or the two-port network, referred to its input.
- $T_e = (F-1)T_0$, where T_e is noise equivalent temperature of a device/network whose noise figure is F and T_0 is the room temperature.
- Frii's formula for noise figure of amplifiers in cascade

  $$F = F_1 + \frac{(F_2-1)}{G_1} + \frac{(F_3-1)}{G_1G_2} + \frac{(F_4-1)}{G_1G_2G_3} + \ldots$$

  For the overall noise temperature of amplifiers in cascade:

  $$T_e = T_e + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1G_2} + \ldots$$
- A band pass noise $n(t) = R(t)\cos[\omega_c t + \theta_n(t)]$ can be represented in the canonical form or the inphase and quadrature components form as

  $$n(t) = n_i(t)\cos\omega_c t - n_q(t)\sin\omega_c t$$

  where $n_i(t)$ and $n_q(t)$, called the inphase and quadrature components respectively, are such that they are low pass processes and

  $$\overline{n_i^2(t)} = \overline{n_q^2(t)} = \overline{n^2(t)}$$

References and Suggested Reading

1. Davenport Jr., W.B., and W.L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.
2. Freeman, J.J., *Noise*, John Wiley & Sons, Inc, New York, 1958.
3. Gray, R.M. and Davisson, L.D., *Random Processes*, Prentice-Hall, Englewood Cliffs, New Jersey, 1986.
4. Hancock, John C., *An Introduction to the Principles of Communication Theory*, McGraw-Hill, New York, 1961.
5. Haykins, Simmon, *Communication Systems*, John Wiley & Sons, New York, 4th Edition, 2001.
6. Proakis, John G. and Masoud Salehi, *Communication Systems Engineering*, 2nd Edition, Eastern Economy Edition, Prentice-Hall of India, New Delhi, 2006.
7. Papoulis, A. and Unnikrishna Pillai, S, *Probability, Random Variables and Stochastic Processes*, 4th Edition, Tata-McGraw-Hill, New Delhi, 2002.
8. Roddy, Dennis and John Coolen, *Electronic Communications*, 3rd Edition (Eastern Economy Edition), Prentice-Hall of India, New Delhi, 1994.
9. Stark, H. and J.W. Woods, *Probability, Random Processes and Estimation Theory for Engineers*, 2nd Edition, Prentice-Hall, New Jersey, 1994.

Review Questions

1. What are the constituents of a probability space?

2. Explain the need for introducing the σ-field as an element of the probability space.

3. The probability, P, assigned to an event must satisfy certain conditions. What are they?

4. State Bayes' theorem for conditional probability.

5. When do you say that two events are independent?
6. Define a random variable. Give an example.
7. What is meant by the Cumulative Distribution Function (CDF) of a random variable?
8. State the properties of the CDF and PDF of a random variable.
9. Distinguish between discrete, continuous, and mixed type of random variables.
10. If a normal random variable has a mean of m and variance of σ^2, what is its density function and what is the area under its density function curve?
11. Define the terms 'mean' and 'variance' of a random variable.
12. Define 'characteristic function' of a random variable $\boldsymbol{X}$. How is it useful?
13. Define 'joint distribution function' of two random variables.
14. Explain the meaning of 'conditional PDF of $\boldsymbol{Y}$ given $\boldsymbol{X}$'.
15. When do you say the random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent? Uncorrelated? Orthogonal?
16. Define joint Gaussianity of two random variables.
17. State the properties of jointly Gaussian random variables.
18. Explain what the 'central limit theorem' states and comment on the importance of the theorem.
19. Define the term 'random process'.
20. Interpret what a random process represents when (a) time variable is fixed, and (b) outcome ξ is fixed.
21. Define 'first and second-order distribution functions' of a random process.
22. Explain what you understand by the terms 'mean', 'auto-correlation', and 'auto co-variance', of any random process.
23. When do you say two random processes are independent? Is it when it is (a) uncorrelated, or (b) when it is orthogonal?
24. Distinguish between strict-sense stationarity and wide-sense stationarity with regard to a random process.
25. State the properties of the auto-correlation function of a stationary process.
26. What is ergodicity?
27. Define the 'power spectrum' of a random process and state its properties.
28. What is a Gaussian process? State some of its properties.
29. Sketch the PSD and ACF of a white noise process.
30. What do you understand from the statement: 'When a stationary random process is applied as input to an LTI system, the input and output processes are jointly stationary'?
31. State the 'Sampling theorem' for stationary low pass band-limited processes.
32. How are the average powers of the 'inphase' and 'quadrature' components related to the average power of a band pass process?
33. In a communication scenario, what is meant by 'noise'?
34. Name the important components of external noise and internal noise.
35. In which bands of the electromagnetic spectrum is communication affected by atmospheric noise? Why?
36. What are the sources of 'galactic noise'? What is the range of frequencies over which this noise has its spectral components?
37. What is the origin of thermal noise? Comment on its power spectral density.
38. Explain the meaning of the term 'available noise power'.
39. What is 'white noise'? Sketch the PSD and ACF of white noise. Why is it not possible to have a 'white noise' source in practice?
40. How does 'shot-noise' originate? Comment the power spectrum of shot-noise current.
41. What is meant by partition noise?
42. Define and explain the term 'noise equivalent bandwidth of a filter'.
43. What is meant by 'equivalent noise resistance' of an amplifier?
44. The 'signal-to-noise ratio' at the output of an amplifier is given to be 200. What is its value in decibels?
45. Define and explain the terms 'noise figure' and 'noise temperature' of a two-port network? How are they related?
46. Explain clearly, why in a super heterodyne receiver, it is preferable to have an RF amplifier with high gain to be the first stage instead of a mixer.

Problems

1. Event $A = \{3 \le x \le 6\}$ and event $B = \{4 \le x \le 7\}$. Find $A \cup B$, $A \cap B$.

2. A box contains 5 red balls numbered 1, 2, 3, 4, 5 and 3 black balls numbered 1, 2, 3. Our random experiment is to randomly pick one ball from the box. What are the outcomes involved in the following events?
   (a) A = a ball with an odd number
   (b) B = a black ball with number greater than 1
   (c) C = a ball bearing a number less than 3

3. Given that AB = null set, show that $P(A) \le P(\overline{B})$

4. Prove that: $P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$

5. A and B are two disjoint events. What conditions should be fulfilled for them to be independent?

6. Show that $P(AB/C) = P(A/BC)P(B/C)$.

7. A source produces the binary digits 0 and 1 with probabilities 0.4 and 0.6 respectively. The channel over which these digits are transmitted has an error probability of 0.3.
   (a) What is the probability of a 1 being obtained at the output of the channel?
   (b) If a 1 has been obtained at the output, what is the probability that it is due to the source giving a 1 to the channel?

8. Ram is to make a telephone call at some random instant in the interval (0, 20) in seconds. What is the probability of his making the call in the 10 sec to 18 sec interval? What is the probability of his making the call in the 10 to 18 sec interval given that he did not make the call up to the end of the 8th second?

9. There are three sections – A, B and C of a class. In a test, 25% of the students from A section, 10% of the students from the section B and 15% of the students from the C section, failed. Two answer scripts are randomly picked from those of a randomly selected section. (a) What is the probability that both the answer scripts belonged to failed students? (b) Assuming that both scripts belonged to failed students, what is the probability that these were from section A?

10. A Gaussian random variable $\boldsymbol{X}$ has zero mean and a variance of 2. Find the probability $P[2 \le x \le 3]$. Also, find $P[2 \le x \le 3 \text{ Given } \boldsymbol{X} \ge 1]$.

11. Find the mean, variance and the density function of random variable $\boldsymbol{Y}$ given that $\boldsymbol{Y} = 3\boldsymbol{X} + 6$ and that $\boldsymbol{X}$ is Gaussian with $\eta_X = 2$ and $\sigma_X^2 = 3$.

12. Determine the CDF and PDF of $\boldsymbol{Y}$ given that $\boldsymbol{Y} = 2\boldsymbol{X} + 3$ and that $f_X(x) = 2e^{-x}u(x)$.

13. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are zero-mean Gaussian random variables with a variance of σ^2 for each. Assuming them to be independent, determine the density function of the random variable $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$

14. $\boldsymbol{X}$ and $\boldsymbol{Y}$ are zero-mean, identically distributed Gaussian random variables with a variance of σ^2 for each. Determine the probability density function of the random variable $\boldsymbol{Z} = \sqrt{\boldsymbol{X}^2 + \boldsymbol{Y}^2}$.

15. Find whether the function $f(t) = \sin 2\pi f_0 t$ can be the auto-correlation function of a random process. Irrespective of whether your answer is yes or no, give reasons.

16. When $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are jointly stationary, we know that $R_{XY}(t_1, t_2) = R_{XY}(\tau)$ where τ is $(t_1 - t_2)$. Show that $R_{XY}(-\tau) = R_{YX}(\tau)$. How are $S_{XY}(\tau)$ and $S_{XY}(\tau)$ related?

17. If $S_X(f)$ is the power spectrum of a stationary random process, $\boldsymbol{X}(t)$, find the PSDs of the following processes:
   (a) $\boldsymbol{X}(t - T)$ where T is a constant
   (b) $\boldsymbol{X}(t) - \boldsymbol{X}(t - T)$

18. In Section 6.12, we have proved that when the input process to an LTI system is stationary, the output process too is stationary. Is the converse of this also true? Why or why not?

19. A white noise process of PSD = $N_0/2$ is the input to an ideal LPF having a cut-off frequency of 2 kHz. If uncorrelated samples are required, at what rate should the output of the filter be sampled?

20. A zero-mean white Gaussian noise with power spectral density $N_0/2$ is passed through an ideal band pass filter of center frequency f_c and bandwidth $2W$. If the output process is $\boldsymbol{n}(t)$, determine:
   (a) the density function of the envelope of $\boldsymbol{n}(t)$.
   (b) the density function of the envelope of the process

$$\boldsymbol{X}(t) = A \cos 2\pi f_0 t + \boldsymbol{n}(t), \text{ where } A \text{ is a constant.}$$

21. Thermal noise voltage (r.m.s) across a resistor has been found to be 10 micro-volts at a temperature of 27° C and over some bandwidth B Hz. What will be the r.m.s. thermal noise voltage at (a) 77° C with bandwidth B Hz, and (b) 77° C with a bandwidth $2B$ Hz?

22. Determine the mean-squared value of the noise voltage across a resistor of 20 kΩ at a temperature of 27° C over a noise bandwidth of 20 kHz.

23. Three resistors of resistance values 10 k, 20 k and 30 k are at a temperature of 27° C. Determine the r.m.s. value of the noise voltage over a bandwidth of 1 MHz when (a) they are all connected in series, and (b) when they are all connected in parallel.

24. A parallel tuned circuit has a capacitor of 1500 pF and is tuned to 2 MHz. It has a Q-factor of 90. What is the r.m.s. noise voltage across the tuned circuit at a temperature of 27° C if the voltage is measured over a bandwidth of 10 kHz?

25. If I_0 = 10 mA and D is in temperature limited condition, determine the r.m.s. noise voltage across the terminals aa'. Assume room temperature of 27° C.

26. An instrument used for measuring noise voltages has an input impedance that is effectively equivalent to a resistor of 100 kΩ in parallel with a capacitance of 0.1 μF. What is the noise equivalent bandwidth of instrument?

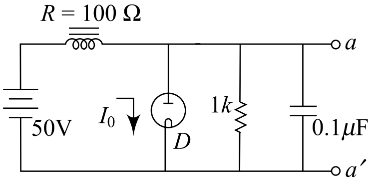


Fig. P8.1

27. A signal source having an internal resistance of 300 Ω and an internal e.m.f. of 10 μV is connected to the input of an amplifier. The amplifier has an input resistance of 1200 Ω and equivalent noise resistance of 300 Ω. For a noise bandwidth of 2 kHz and a room temperature of 27° C, determine (a) the output (S/N) ratio in db, and (b) the noise figure of the amplifier.

28. In Problem 7, what should be the internal e.m.f. of the signal source if an output signal-to-noise ratio of only 20 dB is required?

29. In a particular superheterodyne radio receiver, the antenna circuit comprising a tank circuit tuned to the incoming signal is coupled inductively to the input of the mixer stage. The coupling provides a step-up ratio of 10:1 and also provides perfect matching with the 12 kΩ input resistance of the mixer stage. If this stage has noise equivalent resistance of 80 kΩ, what should be the r.m.s. value of the signal voltage induced in the antenna to give an (S/N) of 20 dB? Assume a room temperature of 27° C and an effective noise bandwidth of 10 kHz.

30. Consider a receiving system consisting of an antenna with a leading cable having a loss factor $L = 1.5$ dB $= F_1$, an RF pre-amplifier with a noise figure of $F_2 = 7$ dB and a gain of 20 dB, followed by a mixer with a noise figure of $F_3 = 10$ dB and a conversion gain of 8 dB, and finally, an integrated-circuit. If amplifier with a noise figure $F_4 = 6$ dB and a gain of 60 dB.

(a) Find the overall noise figure and noise temperature of the system.

(b) Find the noise figure and noise temperature of the system with pre-amplifier and cable interchanged. (VTU August, 2002).

Multiple-Choice Questions

1. If A and B are two events, $P(A \cup B)$ equals

(a) $P(A) + P(B)$
(b) $P(A) + P(B) + P(A \cap B)$
(c) $P(A) + P(B) - P(A \cap B)$
(d) $P(A) + P(B) - P(A \mid B)$

2. A and B are two events and $P(A \mid B) = 0$. Then

(a) B is a certain event
(b) A is an impossible event
(c) A and B are independent
(d) A and B are mutually exclusive

3. Box A contains 4 white balls and six red balls. Box B contains 8 white balls and two red balls. One of the boxes is randomly selected and a ball is randomly picked from it. If the ball so picked up is a red ball, the probability that it would have been picked up from box A is

(a) 0.75 (b) 0.6 (c) 0.8 (d) 0.25

4. The figure shows the distribution function of a random variable $\boldsymbol{X}$. The probability of the random variable $\boldsymbol{X}$ taking a value between 2.5 and 4.0 is

(a) 2/3 (b) 1/2
(c) 1/3 (d) 2/9

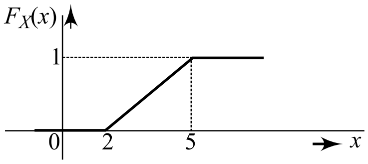


Fig. M6.4

5. A zero-mean band pass signal has identically distributed Gaussian processes as its inphase and quadrature components. The envelope of the band pass process has

(a) Gaussian distribution (b) Ricean distribution.
(c) Rayleigh distribution. (d) Uniform distribution

6. The variance σ^2 of a random variable $\boldsymbol{X}$ is given by

(a) $E[X^2]$ (b) $\{E[X]\}^2$ (c) $E[X^2] - \{E[X]\}^2$ (d) $E[X^2] + \{E[X]\}^2$

7. A random variable is uniformly distributed between 3 and 6. Its variance is

(a) 0.75 (b) 0.25 (c) 0.5 (d) 1

8. The variance of a Bernoulli random variable is

(a) p^2 (b) $(1-p)^2$ (c) $p(1-p)$ (d) $(1+p)^2$

9. $\boldsymbol{X}$ is a random variable with variance σ_x^2. The variance of $(\boldsymbol{X} + a)$ where a is a constant is

(a) $(\sigma_x + a)^2$ (b) σ_x^2 (c) $(\sigma_x^2 + a^2)$ (d) $(\sigma_x^2 - a^2)$

10. The density function $f_X(x)$ of a discrete random variable $\boldsymbol{X}$ is given by

$$f_X(x) = 0.2\delta(x-1) + 0.2\delta(x-2) + 0.4\delta(x-3) + 0.15\delta(x-4) + 0.15\delta(x-5)$$

The mean value of $\boldsymbol{X}$ is:

(a) 2.5 (b) 3.2 (c) 2.8 (d) 3.0

11. The variance σ_x^2 of $\boldsymbol{X}$ in the above question is

(a) 1.65 (b) 2.6 (c) 1.1 (d) 3.2

12. The characteristic function of a random variable that takes the values 1 and 0 with probabilities of 0.6 and 0.4 is

(a) $0.6 + (e^{j\omega} - 1)$ (b) $0.6 - (e^{j\omega} - 1)$ (c) $1 - 0.6(e^{j\omega} - 1)$ (d) $1 + 0.6(e^{j\omega} - 1)$

13. If random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent, then $f_{XY}(x, y)$ is equal to

(a) $f_X(x) + f_Y(y)$ (b) $f_X(x) \cdot f_Y(y)$ (c) $f_X(x) * f_Y(y)$ (d) $f_X(x) - f_Y(y)$

14. Random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are such that $E[\boldsymbol{X} \cdot \boldsymbol{Y}] = E[\boldsymbol{X}]E[\boldsymbol{Y}]$. The random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are

(a) statistically independent (b) orthogonal
(c) uncorrelated (d) Nothing can be concluded

15. When ξ is fixed, $\boldsymbol{X}(t, \xi)$ represents

(a) a random variable (b) a single realization of the random process
(c) a real number (d) a family of time signals

16. Two random processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ are such that $R_{XY}(t_1, t_2) = 0$ for all t_1 and t_2 and further one of them has zero mean. The processes are

(a) uncorrelated but not orthogonal (b) orthogonal but not uncorrelated
(c) statistically independent and orthogonal (d) orthogonal and uncorrelated

17. Auto-correlation function $R_X(\tau)$ of a stationary process $\boldsymbol{X}(t)$ is

(a) a deterministic function with maximum value at $\tau = 0$
(b) a deterministic function which is periodic
(c) a stationary random process
(d) a periodic stationary process

18. The power spectrum, $S_X(f)$, of a random process $\boldsymbol{X}(t)$ is a

(a) real-valued function of frequency with even symmetry
(b) complex-valued function of 'f' with conjugate symmetry
(c) real-valued function of 'f' with even symmetry if $\boldsymbol{X}(t)$ is real valued
(d) real-valued function of 'f' with odd symmetry

19. A process is said to be an ergodic process if
(a) its ensemble averages are different from time averages
(b) it is not stationary
(c) ensemble averages are same as time averages
(d) it is neither continuous, nor discrete

20. For two Gaussian processes to be statistically independent, it is enough if
(a) they are orthogonal
(b) they are uncorrelated
(c) they are orthogonal and one of them has zero-mean
(d) they are uncorrelated and both are zero-mean

21. If a zero-mean Gaussian process is given as input to an LTI system, the output of the LTI system is
(a) a zero-mean Gaussian process
(b) a Gaussian process but not necessarily of zero mean
(c) a zero-mean process but not necessarily Gaussian
(d) not necessarily zero-mean or Gaussian as it depends on the nature of $h(t)$ of the system

22. A stationary random process with a mean of 2 is passed through an LTI system with $h(t) = 2e^{-2t}u(t)$. The mean of the output process is
(a) 4 (b) 0.5 (c) 2 (d) 1

23. A white noise process with power spectral density of $N_0/2$, is given as input to an LTI system with $h(t) = 2e^{-2t}u(t)$. The PSD of the output process is
(a) $\dfrac{2N_0}{4-\omega^2}$ (b) $\dfrac{4N_0}{4+\omega^2}$ (c) $\dfrac{N_0}{4+\omega^2}$ (d) $\dfrac{2N_0}{4+\omega^2}$

24. The effect of atmospheric noise is most severe in
(a) medium wave band (b) shortwave band (c) VHF band (d) microwave region

25. Extra-terrestrial noise can cause disturbance to communications in the frequency range
(a) Below 100 kHz (b) 100 kHz to 10 MHz
(c) 15 MHz to 1.5 GHz (d) Above 1.5 GHz

26. Man-made noise can cause disturbance to communications especially in the frequency range
(a) below 1 MHz (b) 1 MHz to 500 MHz
(c) 500 MHz to 5 GHz (d) above 5 GHz

27. The power spectrum of thermal noise is flat almost up to
(a) 100 kHz (b) a few MHz (c) a few GHz (d) 10^{12} to 10^{13} Hz

28. The two-sided power spectral density of thermal noise is
(a) kTR volt^2/Hz (b) $2kTR$ volt^2/Hz (c) $4kTR$ volt^2/Hz (d) $I_0\,q$ amp^2/Hz

29. The r.m.s. value of the thermal noise voltage across a resistor of $R\Omega$ at a temperature of T°K measured over a bandwidth of (Δf) Hz is
(a) $2\sqrt{kTR(\Delta f)}$ (b) $4kTR(\Delta f)$ (c) $kTR(\Delta f)$ (d) None of the above

30. The r.m.s. value of the thermal noise voltage across resistors R_1 and R_2 are 3 micro-volts and 4 micro-volts respectively**.** The r.m.s. value of the thermal noise across their series combination is
(a) 10 micro-volts (b) 7 micro-volts (c) 5 micro-volts (d) None of the above

31. When the temperature (in °K) of a resistor is doubled, the r.m.s. value of the noise voltage across it is
(a) doubled (b) halved (c) quadrupled (d) 1.414 times its previous value

32. Given a resistance of R ohms at $T°$ K, the available noise power from it over a bandwidth of (Δf) Hz is
(a) $kT(\Delta f)$ (b) ½ $kT(\Delta f)$ (c) $4kTR(\Delta f)$ (d) $2kTR(\Delta f)$.

33. White noise is filtered using an ideal LPF of cutoff frequency 1 kHz. The frequency at which the output noise of the filter should be sampled in order to get totally uncorrelated samples is
(a) 1 kHz (b) 500 Hz
(c) 2 kHz (d) Not possible to get uncorrelated samples

34. The two-sided PSD of the shot-noise generated by a vacuum diode operating in the temperature-limited region and carrying a direct current of I_0 amps is

(a) $2 I_0 q$ amp^2/Hz
(b) $I_0 q$ amp^2/Hz
(c) ½ $I_0 q$ amp^2/Hz
(d) None of the above

35. For the same direct current I_0 flowing through it as in a temperature-limited vacuum diode, a vacuum diode operating in the space-charge limited region

(a) does not produce any shot noise
(b) produces less shot-noise
(c) produces more shot-noise
(d) produces the same amount of shot-noise

36. The noise equivalent bandwidth of an L-section RC-low pass filter

(a) increases with the time-constant RC
(b) decreases with the time-constant RC
(c) does not depend upon the time-constant RC
(d) None of the above

37. Temperature and bandwidth remaining constant, the available noise power from a resistor of R ohms

(a) is independent of R
(b) increases with R
(c) decreases with R
(d) None of the above

38. An amplifier

(a) improves the signal-to-noise ratio
(b) does not alter the signal-to-noise ratio
(c) degrades the signal-to-noise ratio
(d) None of the above

39. When a number of amplifiers are connected in cascade, the overall noise figure is approximately equal to

(a) the noise figure of the most noisy amplifier
(b) the noise figure of the least noisy amplifier
(c) the sum of noise figures of all the amplifiers
(d) the noise figure of the first amplifier

Key to Multiple-Choice Questions

1. (c)	2. (d)	3. (a)	4. (b)	5. (c)	6. (c)	7. (a)	8. (c)
9. (b)	10. (d)	11. (a)	12. (d)	13. (b)	14. (c)	15. (b)	16. (d)
17. (a)	18. (c)	19. (c)	20. (b)	21. (a)	22. (d)	23. (d)	24. (a)
25. (c)	26. (b)	27. (d)	28. (b)	29. (a)	30. (c)	31. (d)	32. (a)
33. (c)	34. (b)	35. (b)	36. (b)	37. (a)	38. (c)	39. (d)	

NOISE PERFORMANCE OF AM AND FM SYSTEMS

7

> *"Achievement seems to be connected with action. Successful men and women keep moving. They make mistakes, but they don't quit."*
>
> **Conrad Hilton (1887–1979)**
> *American hotelier*

Learning Objectives

After going through this chapter, students will be able to

- *explain the way the channel and the receiver are modeled for a study of the noise performance of the system,*
- *analyze the noise performance of FM and the various types of AM systems and compare them,*
- *understand how an FM system offers the possibility of power bandwidth trade-off and why there is a limit for this trade-off,*
- *understand how and why it is possible to improve the destination SNR for FM systems by employing pre-emphasis and de-emphasis, and*
- *explain the 'threshold effect' in FM receivers, and the various methods of threshold extension.*

7.1 INTRODUCTION

In Chapter 6, we had discussed the various types of noise, their sources and characteristics, and noted that thermal noise and shot noise are both white insofar as the frequencies and bandwidths used in practical communication systems are concerned. In Chapter 4, we had studied the methods of generation and demodulation of various types of amplitude modulated signals like AM, DSB-SC, SSB-SC, etc. Similarly, in Chapter 5, we studied angle-modulation, and discussed the modulation and de-modulation methods for FM and PM. Again in Chapter 6, we reviewed probability and random processes and discussed 'noise'.

In the present chapter, we will make use of the material covered in the previous six chapters and examine, by deriving the necessary expressions, the noise performance of a continuous wave (CW) communication systems. From these results, we will not only be able to compare the various CW communication systems on the basis of their noise performance, but also use them for communication system design. For CW communication systems, a convenient and useful parameter for assessing the noise performance of any modulation-demodulation scheme, is the destination signal-to-noise ratio $(S/N)_D$, i.e., the ratio of the average signal power to the average noise power at the output of the receiver. For determining this $(S/N)_D$ for different modulation and de-modulation schemes, we must have the models for the pertinent signals, channel noise and receiving systems. We already have the mathematical representation of message signals and modulated signals. We will model the channel noise as zero mean, white Gaussian noise with a two-sided power spectral density (PSD) of η/2. For each type of modulation and de-modulation that we take up, we shall use an appropriate receiver structure and then model it suitably for the purpose of this analysis.

After studying the noise performance of various types of amplitude modulation systems and frequency modulation systems, we shall, towards the end of the chapter, discuss a few related topics like improvement in the noise performance of frequency modulation (FM) systems by the use of pre-emphasis at the transmitter and de-emphasis at the receiver, threshold effect in FM receivers and the threshold extension techniques.

7.2 DESTINATION SNR OF A BASEBAND SYSTEM

The baseband system is one in which the baseband signal is directly sent over the channel without any carrier and modulation, The receiver too, does not have any demodulator and is modeled as an ideal low pass filter with a cut-off frequency of W Hz, which is the bandwidth of the baseband signal.

As mentioned earlier, let the power spectral density of the zero-mean white noise on the channel be η/2. Hence, the shaded area of Fig. 7.1 represents the average noise power lying within the bandwidth of the baseband and corrupting the signal.

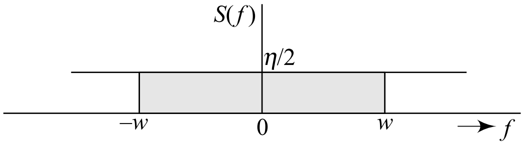


Fig. 7.1 *Noise power corrupting the baseband signal*

$$\therefore \quad N = \begin{Bmatrix} \text{Average noise power} \\ \text{within the bandwidth } 2W \end{Bmatrix} = 2W \cdot \eta/2 = \eta W$$

Hence, if we denote the average signal power at the receiver input as S_R, since the receiver is modeled as an ideal LPF with cut-off frequency equal to W, the destination signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} \underline{\Delta}\, \gamma \tag{7.1a}$$

$\left(\frac{S}{N}\right)_D$ of other systems are generally compared with this.

7.3 MODEL FOR LINEAR MODULATION SYSTEMS

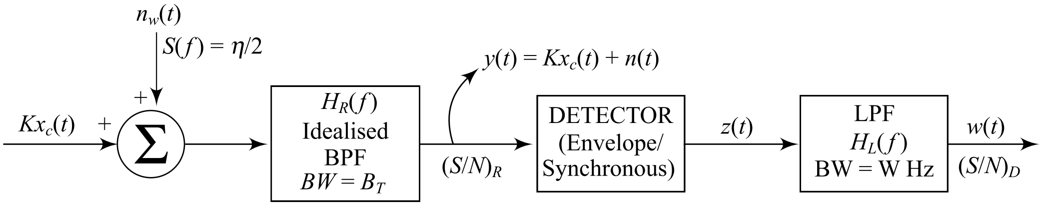


Fig. 7.2 *Model for linear modulation systems*

$x_c(t)$ is the transmitted modulated signal. k is the attenuation factor so that $K \cdot x_c(t)$ is the received signal assumed to be having a carrier frequency $f_c = f_{i \cdot f}$, the intermediate frequency. $n_W(t)$ is the zero-mean white Gaussian noise of the channel with a two-sided power spectral density of η/2. It gets added to the received signal. $H_R(f)$ is an ideal band pass filter which is used for modeling the combined effect of the RF amplifier and the IF amplifier. It has a bandwidth B_T, the transmission bandwidth of the modulated signal and is also the bandwidth of the front-end of the receiver. This bandwidth B_T is $2W$ for AM and DSB-SC and is W for SSB-SC. Further, the filter is assumed to have the bandwidth B_T centered on $f_c = f_{i \cdot f}$, the intermediate frequency, and the gain of the filter is unity in its pass band. The output of this filter will have a signal component $Kx_c(t)$ and noise component $n(t)$ where $n(t)$ is Gaussian and zero mean, but not white. It is a band pass noise having an average power = $(\eta/2) \cdot (2B_T) = \eta B_T$. Thus the input to the detector block in the model is

$$y(t) = Kx_c(t) + n(t)$$

The signal-to-noise ratio at the input to the detector is denoted by $(S/N)_R$. The LPF, shown as the last block, is an ideal unit-gain LPF with a cut-off frequency of W Hz, which is the bandwidth of the message signal $x(t)$, and is used to model the LPF which follows the analog signal multiplier of a synchronous detector, or the response characteristic of the audio amplifiers following an envelope detector. The destination SNR is denoted by $(S/N)_D$.

7.3.1 Figure of Merit

To facilitate comparison of the various types of modulation systems, we generally define a 'Figure of Merit' of a system as

$$\text{Figure of Merit} \triangleq \frac{(S/N)_D}{(S/N)_C} = \frac{\text{(destination signal-to-noise ratio)}}{\text{channel signal-to-noise ratio}}$$

where the 'channel signal-to-noise ratio' is defined as

$$\left(\frac{S}{N}\right)_C = \text{Channel SNR} \triangleq \frac{\text{Average power of the modulated signal}}{\text{Average power of noise in the message bandwidth } (W)}$$

$$= \frac{S_R}{2W(\eta/2)} = \frac{S_R}{\eta W} = \gamma$$

$$\therefore \quad \text{figure of merit} = \left[\left(\frac{S}{N}\right)_D \cdot \frac{1}{\gamma}\right] \tag{7.1b}$$

Thus, in fact the Figure of Merit of a particular modulation system is the ratio of the destination SNR with that modulation, to the destination SNR for baseband transmission. Higher the figure of merit as compared to 1, the better it is.

7.3.2 Pre-detection signal-to-noise ratio

$$y(t) = \text{Detector input} = Kx_c(t) + n(t) \tag{7.2}$$

where $n(t)$ is a band pass noise with an average power of ηB_T where $B_T = 2W$ or W, depending on the type of modulation

$$S_R = \text{Average signal power at the input to the detector}$$

$$= K^2 \overline{x_c^2(t)} \tag{7.3}$$

where the overbar on $x_c^2(t)$ denotes its average value.

$$N_R = \text{Average noise power at the input to the detector}$$

$$= \overline{n^2(t)} = (\eta/2)(2B_T) = \eta B_T \tag{7.4}$$

$$\therefore \quad \text{pre-detection signal-to-noise ratio} = \left(\frac{S}{N}\right)_R = \frac{S_R}{N_R} = \frac{S_R}{\eta B_T} \tag{7.5}$$

But, we have already seen that in the case of baseband transmission, the destination SNR is given by

$$\gamma = \frac{S_R}{\eta W} \tag{7.6}$$

Hence, we may express the pre-detection SNR, viz., $(S/N)_R$ as

$$\left(\frac{S}{N}\right)_R = \frac{S_R}{\eta B_T} = \frac{S_R}{\eta W}\left(\frac{W}{B_T}\right) = \gamma\left(\frac{W}{B_T}\right) \tag{7.7}$$

With this, we are now ready to proceed with the determination of the destination signal-to-noise ratios for various linear modulation schemes.

7.4 $(S/N)_D$ FOR SSB-SC SYSTEMS

For SSB-SC systems, the detector is a synchronous detector and the transmitted signal is given by

$$x_c(t) = \frac{1}{2}A_c[x(t)\cos\omega_c t \mp \hat{x}(t)\sin\omega_c t] \qquad \text{(refer to Eqs. (4.73) and (4.74))} \tag{7.8}$$

with negative sign for the USSB and positive sign for the LSSB.

But $$S_R = K^2\overline{x_c^2(t)} = \frac{1}{4}K^2A_c^2\overline{\left[x^2(t)\cos^2\omega_c t \mp 2x(t)\hat{x}(t)\sin\omega_c t\cos\omega_c t + \hat{x}^2(t)\sin^2\omega_c t\right]}$$

$$= \frac{1}{2}\frac{1}{4}K^2A_c^2\overline{x^2(t)} + \frac{1}{2}\left(\frac{1}{4}\right)K^2A_c^2\overline{\hat{x}^2(t)}$$

But $\overline{x^2(t)} = \overline{\hat{x}^2(t)}$ since the Hilbert transform does not alter the power.

$$\therefore \qquad S_R = \frac{1}{4}A_R^2\overline{x^2(t)} \tag{7.9}$$

where $A_R = KA_c$

Now, input to the synchronous detector is given by

$$y(t) = Kx_c(t) + n(t) \tag{7.10}$$

If we assume that it is a USSB system and substitute for $x_c(t)$ using Eq. (7.8), and further, if we substitute for the band pass noise $n(t)$ in Eq. (7.10) by its inphase and quadrature component representation, we get

$$y(t) = \frac{1}{2}A_R x(t)\cos\omega_c t - \frac{1}{2}A_R\hat{x}(t)\sin\omega_c t + n_i(t)\cos\omega_c t - n_q(t)\sin\omega_c t \tag{7.11}$$

In the synchronous detector, $y(t)$ is multiplied by $\cos\omega_c t$ and low pass filtered to give $w(t)$ (see Fig. 7.2)

$$\therefore \qquad z(t) = y(t)\cos\omega_c t = \frac{1}{2}A_R x(t)\cos^2\omega_c t - \frac{1}{4}A_R\hat{x}(t)\sin 2\omega_c t + n_i(t)\cos^2\omega_c t - \frac{1}{2}n_q(t)\sin 2\omega_c t$$

When this $z(t)$ is low pass filtered using an ideal LPF with a cut-off frequency of W Hz, all high frequency components are rejected.

$$w(t) = \frac{1}{4}A_R x(t) + \frac{1}{2}n_i(t) \tag{7.12}$$

In the above, the first term is the signal term and the second term is the noise term.

$$\left(\frac{S}{N}\right)_D = \frac{(1/4)^2A_R^2\overline{x^2(t)}}{(1/4)\overline{n_i^2(t)}} = \frac{(1/4)A_R^2\overline{x^2(t)}}{\eta B_T} = \frac{S_R}{\eta B_T} = \frac{S_R}{\eta W} = \gamma \tag{7.13}$$

(Since $B_T = W$ for SSB)

In the above, we have made use of the properties of the inphase and quadrature components of a zero-mean band pass process, that $\overline{n_i^2(t)} = \overline{n_q^2(t)} = \overline{n^2(t)}$ and that $\overline{n^2(t)} = \eta B_T$ from Eq. (7.4). Since it is an SSB-SC system, $B_T = W$ and so $\overline{n_i^2(t)} = \eta W$.

∴ $$\boxed{\left(\frac{S}{N}\right)_{\substack{D\\ \text{SSB-SC}}} = \gamma} \tag{7.14a}$$

Thus, the 'Figure of Merit' for an SSB-SC system is (From Eqs. (7.14) and (7.7)

$$\underset{\text{SSB-SC}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C}$$

∴ $$\text{figure of merit for SSB-SC} = \left(\frac{\gamma}{\gamma}\right) = 1 \tag{7.14b}$$

7.5 DSB-SC SYSTEMS

For a DSB-SC system a coherent or synchronous demodulator will be used and the modulated signal is given by

$$x_c(t) = A_c x(t) \cos \omega_c t \tag{7.15}$$

∴ received signal = $Kx_c(t)$

and $$Kx_c(t) = KA_c x(t) \cos \omega_c t = A_R x(t) \cos \omega_c t \tag{7.16}$$

∴ $$S_R = \text{Received signal power} = K^2\, \overline{x_c^2(t)} = \frac{1}{2}\, \overline{x^2(t)} \cdot A_R^2 \tag{7.17}$$

Also, $$B_T = 2W \tag{7.18}$$

$$\text{Input to the detector} = y(t) = A_R x(t) \cos \omega_c t + n(t) \tag{7.19}$$

But $$n(t) = n_i(t) \cos \omega_c t - n_q(t) \sin \omega_c t$$

∴ $$y(t) = [A_R x(t) + n_i(t)] \cos \omega_c t - n_q(t) \sin \omega_c t \tag{7.20}$$

The synchronous detector multiplies $y(t)$ by $\cos \omega_c t$

∴ $z(t)$ = output of the multiplier in the detector

$$= [A_R x(t) + n_i(t)] \cos^2 \omega_c t - n_q(t) \sin \omega_c t \cos \omega_c t$$

$$= \frac{1}{2}[A_R x(t) + n_i(t)] + \frac{1}{2}[A_R x(t) + n_i(t)] \cos 2\omega_c t - \frac{1}{2} n_q(t) \sin 2\omega_c t$$

∴ $w(t)$, the output of the low pass filter is given by

$$w(t) = \frac{1}{2}[A_R x(t) + n_i(t)] \tag{7.21}$$

In this, the message signal component is $\frac{1}{2} A_R x(t)$ and the noise component is $\frac{1}{2} n_i(t)$.

∴ $$\left(\frac{S}{N}\right)_D = \frac{A_R^2\, \overline{x^2(t)}}{\overline{n_i^2(t)}} \tag{7.22}$$

But $\frac{1}{2} A_R^2\, \overline{x^2(t)}$ = received signal power = S_R (from Eq. (7.17) for a DSB-SC signal)

and $$\overline{n_i^2(t)} = \overline{n^2(t)} = \eta B_T = 2\eta W$$

(Since $n(t)$ is zero mean, $n_i(t)$ and $n(t)$ will have the same variance.)

∴ substituting these values in Eq. (7.22), we get

$$\left(\frac{S}{N}\right)_D = \frac{2S_R}{2\eta W} = \frac{S_R}{\eta W} = \gamma$$

$$\therefore \qquad \boxed{\left(\frac{S}{N}\right)_{D \atop \text{DSB-SC}} = \gamma} \qquad (7.23)$$

Thus, the 'Figure of Merit' for a DSB-SC system is (from Eqs. (7.7) and (7.23)

$$\underset{\text{(DSB-SC)}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{\gamma}{\gamma} = 1 \qquad (7.23a)$$

Example 7.1 A DSB-SC signal is transmitted over a noisy channel, with the power spectral density of the noise being as shown in Fig. 7.3(a). The message bandwidth is 4 kHz, and the carrier frequency is 200 kHz. Assuming that the average power of the modulated wave is 10 W, find the output signal-to-noise ratio of the receiver.

Solution Here, the additive noise on the channel is *not* white. It has a triangular-shaped two-sided power spectral density as shown in Fig. 7.3(b). The DSB-SC signal has a bandwidth of 8 kHz since the message signal bandwidth is given to be 4 kHz. Hence, the receiver front-end bandwidth is also 8 kHz and is centered on 200 kHz, the carrier frequency. Thus, the two-sided power spectrum of the band pass noise entering the receiver is as shown by the shaded area in Fig. 7.3(b).

The value of the noise PSD at 200 kHz, i.e., the height at 200 kHz is equal to 0.5×10^{-6} *W*/Hz (from similar triangles). Hence, we may compute the area of each of the trapezoidal shaded portion as the area of a rectangle of width 8 kHz and height 0.5×10^{-6} *W*/Hz.

∴ average power of the band pass noise entering the receiver = $\overline{n^2(t)}$

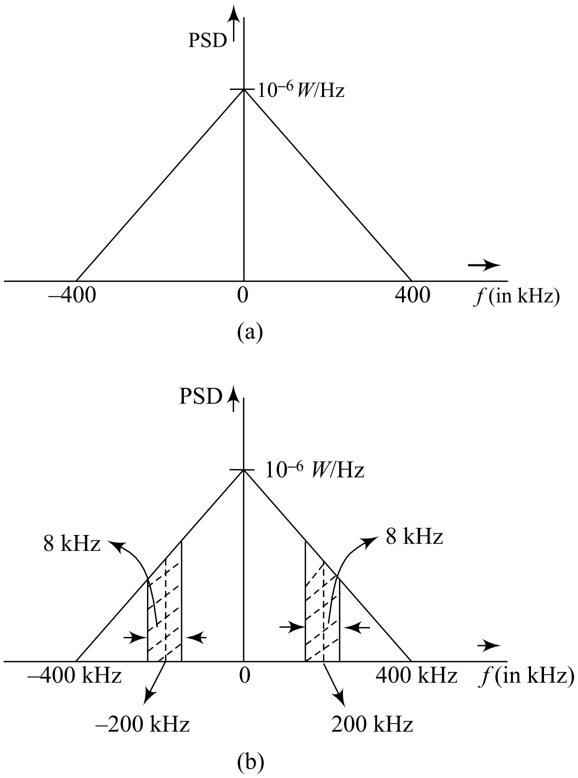
= Total area of the shaded portion

$$= 2 \times \frac{1}{2} \times 10^{-6} \times 8 \times 10^3 = 8 \times 10^{-3}\, W$$

But from the properties of band pass noise, we know that

$$\overline{n^2(t)} = \overline{n_i^2(t)}$$

$$\therefore \quad \overline{n_i^2(t)} = 8 \times 10^{-3}\, W$$



(a)

(b)

Fig. 7.3 *(a) Noise PSD for Example 7.1, (b) Two-sided noise power spectrum for Example 7.1*

Average power of the received DSB-SC signal $= \frac{1}{2} A_R^2 \overline{x^2(t)}$ (From Eq. (7.17)

$$\therefore \quad S_R = \text{Received signal power} = \frac{1}{2} A_R^2 \overline{x^2(t)} = 10\text{ W}$$

But from Eq. (7.22), we know that

$$\left(\frac{S}{N}\right)_D = \text{Destination SNR} = \frac{A_R^2\,\overline{x^2(t)}}{\overline{n_i^2(t)}} = \frac{20}{8\times 10^{-3}} = 2.5\times 10^3$$

$$\left(\frac{S}{N}\right)_D = 2.5\times 10^3 \quad \text{or} \quad \left(\frac{S}{N}\right)_D \text{ in dB} = 10\ \log_{10} 2.5\times 10^3 = 33.97\text{ dB}$$

7.6 AM SYSTEMS

In the case of AM systems, the carrier as well as both the sidebands are transmitted, and so the transmission bandwidth B_T is

$$B_T = 2W \tag{7.24}$$

where, of course, W is the bandwidth of $x(t)$, the message signal. The transmitted signal, $x_c(t)$ is given by

$$x_c(t) = A_c[1 + mx(t)]\cos \omega_c t \tag{7.25}$$

where m, $0 \le m \le 1$, is the modulation index and $x(t)$ is the *normalized* message signal assumed to be *zero mean* and normalized so that $|x(t)| \le 1$. An AM signal can be detected using a synchronous detector or an envelope detector. In practice, however, only an envelope detector is used for AM. For arriving at the $(S/N)_D$ of an AM system, we shall first assume a synchronous demodulator and then derive the expression assuming an envelope detector.

7.6.1 AM System with a Synchronous Detector

$$\text{The received signal} = Kx_c(t) = KA_c[1 + mx(t)]\cos \omega_c t$$

$$= A_R[1 + mx(t)]\cos \omega_c t \tag{7.26}$$

$$\therefore \qquad S_R = \text{Average received signal power} = K^2\,\overline{x_c^2(t)}$$

$$\therefore \qquad S_R = A_R^2\,\overline{[1 + mx(t)]^2 \cos^2 \omega_c t}$$

Since $x(t)$ is zero mean, the above expression reduces to

$$S_R = \frac{1}{2}A_R^2\left[1 + m^2\,\overline{x^2(t)}\right] \tag{7.27}$$

$y(t)$, the input to the synchronous detector is given by

$$y(t) = A_R[1 + mx(t)]\cos \omega_c t + n(t) \tag{7.28}$$

Replacing $n(t)$ in the above by its inphase and quadrature component representation, we get

$$y(t) = \{A_R[1 + mx(t)] + n_i(t)\}\cos \omega_c t - n_q(t)\sin \omega_c t \tag{7.29}$$

The synchronous detector multiplies this by the carrier, i.e., cos $\omega_c t$.

$$\therefore \qquad z(t) = \{A_R[1 + mx(t)] + n_i(t)\}\cos^2 \omega_c t - \frac{1}{2}n_q(t)\sin 2\omega_c t \tag{7.30}$$

The low pass filter removes all the high frequency components as its cut-off frequency is W. Hence, replacing $\cos^2 \omega_c t$ by $\frac{1}{2}(1 + \cos 2\omega_c t)$ and then rejecting all the terms representing high frequency components, we get

$$w(t) = \frac{1}{2}\{A_R[1 + mx(t)] + n_i(t)\} \tag{7.31}$$

In the above equation, $\frac{1}{2}A_R$ represents a dc component, $\frac{1}{2}A_R m x(t)$ represents the message signal component and $\frac{1}{2}n_i(t)$ represents the noise component. In the receiver, anyhow, the dc component at the output of the detector will be blocked by using a blocking capacitor. So, we ignore the dc component of $w(t)$. Then

$$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} = \frac{S_D}{N_D} \tag{7.32}$$

But
$$\overline{n_i^2} = \overline{n^2} = \eta B_T = 2\eta W = N_D \tag{7.33}$$

(Since $n(t)$ is of zero mean, variances of $n_i(t)$ and $n(t)$ will be equal.)

Also, $S_R = \frac{1}{2}A_R^2\left[1 + m^2\,\overline{x^2(t)}\right]$ (From Eq. (7.27))

∴ we may write

$$\left(\frac{S}{N}\right)_D = \frac{S_D}{S_R}\frac{S_R}{N_D} = \frac{A_R^2 m^2 \overline{x^2(t)}}{\frac{A_R^2}{2}\left[1 + m^2\,\overline{x^2(t)}\right]} \cdot \frac{S_R}{2\eta W}$$

$$= \frac{m^2\,\overline{x^2(t)}}{\left(1 + m^2\,\overline{x^2(t)}\right)} \cdot \frac{S_R}{\eta W} = \left[\frac{m^2\,\overline{x^2(t)}}{\left[1 + m^2\,\overline{x^2(t)}\right]}\right] \cdot \gamma$$

∴
$$\left(\frac{S}{N}\right)_{D\ \substack{\text{AM}\\ \text{Sync-det}}} = \left[\frac{m^2\,\overline{x^2(t)}}{\left(1 + m^2\,\overline{x^2(t)}\right)}\right]\gamma \tag{7.34}$$

7.6.2 AM System with Envelope Detector

An envelope detector ideally extracts the envelope of the signal given to it as input. If there were to be no channel noise the input signal to the detector block in Fig. 7.2 would be

$$y(t) = Kx_c(t) = KA_c[1 + mx(t)]\cos \omega_c t$$

and the output of the detector would be its envelope.

i.e., $z(t) = A_R[1 + mx(t)]$, where $A_R \triangleq KA_c$.

However, with the channel noise, the detector input is

$$y(t) = KA_c[1 + mx(t)]\cos \omega_c t + n(t) \tag{7.35}$$

where $n(t)$ is band pass noise centered on f_c and having a bandwidth of $2W$. This noise changes the envelope. To see how it affects the envelope of the AM signal, let us replace $n(t)$ by its inphase and quadrature components.

∴
$$y(t) = A_R[1 + mx(t)]\cos \omega_c t + n_i(t)\cos \omega_c t - n_q(t)\sin \omega_c t$$
$$= \{A_R[1 + mx(t)] + n_i(t)\}\cos \omega_c t - n_q(t)\sin \omega_c t \tag{7.36}$$

∴
$$R_y(t) = \text{Envelope of } y(t)$$
$$= [\{A_R[1 + mx(t)] + n_i(t)\}^2 + n_q^2(t)]^{1/2} \tag{7.37}$$

and the phase angle $\theta_y(t)$ is

$$\theta_y(t) = \tan^{-1}\left[\frac{n_q(t)}{A_R[1+mx(t)]+n_i(t)}\right] \qquad (7.38)$$

Since an envelope detector is totally insensitive to the phase variations of its input signal, we can totally ignore $\theta_y(t)$.

Generally, for satisfactory intelligibility of the message signal output from an envelope detector, the signal-to-noise ratio at the input to the detector must be at least around 8 to 10 dB. So, we can safely assume that the carrier-to-noise power ratio is quite high at the input to the envelope detector. So, we assume that

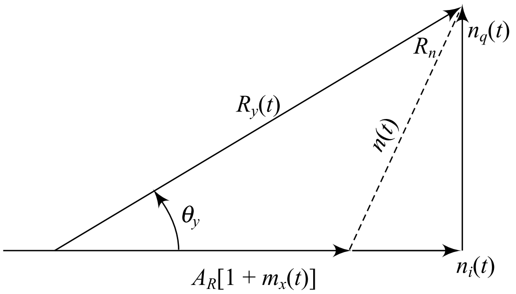


Fig. 7.4 *Phasor diagram of the components of y(t), the input to the detector.* $A_R^2 >> \overline{n^2}$ *is assumed*

$$A_R^2 >> \overline{n^2(t)} \qquad (7.39)$$

as this will enable us to write the output of the detector, viz., the envelope of $y(t)$ as the sum of a signal component and a noise component. This will allow us to write down the expression for the destination SNR immediately.

From Fig. 7.4, in view of the assumption of Eq. (7.39), we may say that

$P[A_R\{1+mx(t)\} >> n_q(t)]$ is almost equal to unity.

$$\therefore \qquad R_y(t) \approx A_R[1+mx(t)]+n_i(t) \qquad (7.40)$$

The dc component A_R in this envelope will be blocked by the coupling capacitor at the output of the detector. (Note that A_R is the mean because both $x(t)$ and $n_i(t)$ are zero-mean processes.) Hence, the signal at the output of the receiver is

$$w(t) = A_R mx(t) + n_i(t) \qquad (7.41)$$

Since $A_R mx(t)$ represents the signal component and $n_i(t)$, the noise component of this output signal, we have

$$\left(\frac{S}{N}\right)_D = \frac{A_R m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} \qquad (7.42)$$

But

$$\overline{n_i^2(t)} = \overline{n^2(t)} = \eta B_T = 2\eta W \qquad (7.43)$$

(Since $n(t)$ is of zero mean, the variances of $n_i(t)$ and $n(t)$ will be the same.)

Hence, we may write Eq. (7.42) as

$$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} = \frac{A_R^2 m^2 \overline{x^2(t)}}{S_R} \cdot \frac{S_R}{\eta B_T}$$

But $$S_R = \frac{1}{2}A_R^2\left[1+m^2\overline{x^2(t)}\right] \qquad \text{(see Eq. (7.27))}$$

$$\therefore \qquad \left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\frac{1}{2}A_R^2\left[1+m^2\overline{x^2(t)}\right]} \cdot \left(\frac{S_R}{\eta B_T}\right)$$

$$= \frac{m^2\overline{x^2(t)}}{\left[1+m^2\overline{x^2(t)}\right]}\left(\frac{S_R}{\eta W}\right) \text{ since } B_T = 2W$$

$$\therefore \quad \left(\frac{S}{N}\right)_{\substack{D \\ \text{AM-Env.det}}} = \left[\frac{m^2\overline{x^2(t)}}{\left[1+m^2\overline{x^2(t)}\right]}\right]\cdot\gamma \tag{7.44a}$$

Comparing Eqs. (7.34) and (7.44a), we find they are exactly the same. *However, it must be noted that Eq. (7.44) gives the destination SNR for AM with an envelope detector only if the carrier-to-noise ratio at the input to the detector is large and provided m, the modulation index, is not more than one. It must also be noted that there are no such conditions in the case of AM with coherent or synchronous detector, for Eq. (7.34) to be valid.*

The figure of merit for AM is therefore obtained from Eqs. (7.7) and (7.44a).

$$\underset{\text{(AM)}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{m^2\overline{x^2}\gamma}{1+m^2\overline{x^2}}\cdot\frac{1}{\gamma}$$

$$= \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} \tag{7.44b}$$

Example 7.2 Find the figure of merit of an AM system when the depth of modulation is (a) 100 %, (b) 50 %, and (c) 30 %. **(Bangalore University, April, 1997)**

Solution Figure of merit (FOM) of an AM system = $\dfrac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$ (From Eq. (7.42a))

(a) $m = 1$, *i.e., 100% modulation*

FOM = $\dfrac{\overline{x^2}}{1+\overline{x^2}}$. Since nothing has been mentioned about the average power of the modulating signal, if a single tone is assumed, $\overline{x^2} = 1/2$

$$\underset{m=1}{\text{FOM}} = \frac{1/2}{1+1/2} = \frac{1}{2}\times\frac{2}{3} = \frac{1}{3}$$

(b) $m = 0.5$, *i.e.,* 50%

$$\text{FOM} = \frac{(0.5)^2\overline{x^2}}{1+(0.5)^2\overline{x^2}} = \frac{0.25\overline{x^2}}{1+0.25\overline{x^2}}$$

For an $x(t)$ which is a single tone, $\overline{x^2} = 1/2$

$$\therefore \quad \underset{m=0.5}{\text{FOM}} = \frac{0.25\times 0.5}{1+0.25\times 0.5} = \frac{0.125}{1+0.125} = 0.111$$

(c) $m = 0.3$, *i.e., 30% modulation*

$$\text{FOM} = \frac{0.09\overline{x^2}}{1+0.09\overline{x^2}}$$

For an $x(t)$ which is a single tone,

$$\underset{m=0.3}{\text{FOM}} = \frac{0.09\times 0.5}{1+0.09\times 0.5} = \frac{0.045}{1+0.045} = 0.04306$$

Example 7.3 Prove that the figure of merit of an AM system for single-tone modulation with 100% modulation is 1/3.

Solution $$\underset{\text{of an AM system}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}\gamma \cdot \frac{1}{\gamma} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$$

Since $\overline{x^2}$ represents the mean-squared value of the normalized message signal, normalized such that $|x(t)| \le 1$, for the case of a single-tone message signal (i.e., sinusoidal message signal), it means that its peak value is 1. Hence, its RMS value is $\frac{1}{\sqrt{2}}$ and the mean-squared value $\overline{x^2}$ is 1/2. Further, for 100% modulation, $m = 1$

$$\therefore \quad \left.\begin{matrix}\text{Figure of merit for AM}\\ \text{with } m=1 \text{ and single-tone}\\ \text{modulating signal}\end{matrix}\right\} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} = \frac{1/2}{3/2} = 1/3$$

7.6.3 Threshold Effect for AM with Envelope Detector

In case the Signal-to-Noise Ratio (SNR) at the input to the envelope detector becomes very much less than unity, noise completely dominates over the signal and the behavior of the envelope detector would be entirely different. It can be shown that in such a situation, there will be no separate term in the output of the detector, which can be identified as the message signal; *the message signal and noise become intermingled.*

Referring to Eq. (7.37), we may write the expression for the envelope of the detector input as

$$\begin{aligned} R_y(t) &= \sqrt{\{A_R[1+mx(t)]+n_i(t)\}^2 + n_q^2(t)} \\ &= \sqrt{A_R^2(1+mx(t))^2 + n_i^2(t) + n_q^2(t) + 2A_R n_i(t)[1+mx(t)]} \end{aligned}$$

Since the SNR at the input to the detector is much smaller than 1, $A_R^2(1+mx(t))^2$ can be neglected in comparison with the rest of the terms under the square-root sign in the above expression. Hence, $R_y(t)$ may be written as

$$R_y(t) = \sqrt{\left[n_i^2(t)+n_q^2(t)\right]\left[1+\frac{2A_R n_i(t)}{\left[n_i^2(t)+n_q^2(t)\right]}[1+mx(t)]\right]} \tag{7.45}$$

Under the assumption of a small SNR at the input to the detector, the following will be true:

$$\left\{\frac{2A_R n_i(t)}{\left[n_i^2(t)+n_q^2(t)\right]}[1+mx(t)]\right\} << 1$$

If we represent the above expression by $\in$, then in Eq. (7.45), we may make use of the approximation that when $\in << 1$

$$\sqrt{1+\in} \approx \left(1+\frac{\in}{2}\right)$$

$$\therefore \quad R_y(t) = \sqrt{n_i^2(t)+n_q^2(t)}\left[1+\frac{A_R n_i(t)[1+mx(t)]}{n_i^2(t)+n_q^2(t)}\right] \tag{7.46}$$

(Compare this with the $Ry(t)$ given by Eq. (7.40) for the case $A_R^2 >> \overline{n^2(t)}$.)

Thus, at the output of the envelope detector, the message signal term $mx(t)$ gets multiplied by the noise terms and cannot therefore be distinguished from noise. This is called the 'threshold effect' in envelope detection of AM.

Example 7.4 An AM receiver, operating with a sinusoidal modulating wave and 80% modulation has an output signal-to-noise ratio of 30 dB. What is the corresponding carrier-to-noise ratio?

(**VTU, March 2001**)

Solution For an AM system with modulation index, m, the output SNR is given by

$$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} \qquad \text{(See Eq. (7.32))}$$

This is given to be 30 dB = 10^3; $m = 0.8$ and $\overline{x^2(t)} = \frac{1}{2}$ (single tone)

$$\therefore \qquad \frac{A_R^2 \times 0.64 \times 1/2}{\overline{n_i^2(t)}} = 1000. \qquad \therefore \frac{A_R^2}{\overline{n_i^2(t)}} = \frac{1000}{0.32}.$$

But we know that $\overline{n_i^2(t)} = \overline{n^2(t)}$ and that the carrier-to-noise ratio (CNR) is defined as

$$CNR = \frac{A_R^2/2}{\overline{n^2(t)}} \qquad \therefore CNR = \frac{1000}{0.64} = 1562.5$$

or

$$(CNR)_{\text{dB}} = 10 \log_{10} 1562.5 = 31.9 \text{ dB}$$

Example 7.5 A message signal $x(t)$ of 5 kHz bandwidth and having an amplitude probability density as shown in Fig. 7.5, amplitude modulates a carrier to a depth of 80%. The AM signal so obtained transmitted over a channel with additive noise power spectral density of $\eta = 2 \times 10^{-12}\ W/\text{Hz}$ (one sided). The received signal is demodulated using an envelope detector.

(a) If a $\left(\frac{S}{N}\right)_D \geq 40$ dB is desired, what should be the minimum value of A_c, the peak amplitude of the carrier?

(b) Assuming $(S/N)_{\text{th}}$ for envelope detection to be 10 dB, determine the threshold value of A_c?

Solution

(a) For AM systems, the destination SNR is given by

$$\left(\frac{S}{N}\right)_D = \left(\frac{m^2 \overline{x^2}}{1 + m^2 \overline{x^2}}\right) \cdot \gamma \qquad \text{Here, } m = 0.8$$

So let us first find $\overline{x^2}$, the average power of the message, using the given amplitude probability density function of $x(t)$.

We know $\overline{x^2} = \int_{-\infty}^{\infty} x^2 f_X(x) dx$.

$$\therefore \qquad \overline{x^2} = 2\int_0^1 x^2(1-x)dx = \frac{1}{6}$$

$$\therefore \qquad \left(\frac{m^2 \overline{x^2}}{1 + m^2 \overline{x^2}}\right) \cdot \gamma \geq 40 \text{ dB}$$

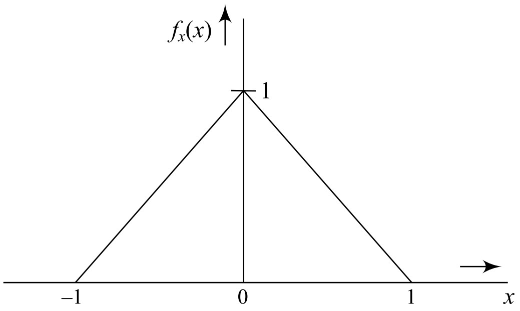


Fig. 7.5 *PDF of signal for Example 7.5*

Since 40 dB = a ratio of 10^4, substituting for m and $\overline{x^2}$, we get

$$\left(\frac{(0.8)^2\left(\frac{1}{6}\right)}{1+(0.8)^2\left(\frac{1}{6}\right)}\right)\cdot\gamma \geq 10^4 \qquad \therefore \gamma \geq \frac{10^4\times(1+0.1066)}{0.1066}$$

$\therefore$ $\gamma \geq 103808.63$ $\therefore$ minimum value of $\gamma = 103808.63$

$\therefore$ if the AM signal is $x_c(t) = A_c[1+mx(t)]\cos\omega_c t$

We know that S_R = Received signal power

$$= \overline{x_c^2(t)} = \frac{A_c^2}{2}[1+m^2\overline{x^2(t)}]$$

Here, we have used the fact that $\overline{x(t)}$ = Average value of $x(t)$ = 0. This is because

$$\overline{x(t)} = \int_{-1}^{1} xf_X(x)dx = \int_{-1}^{0} x(x+1)dx + \int_{0}^{1} x(1-x)dx = 0$$

$\therefore$

$$S_R = \frac{A_c^2}{2}\left[1+0.64\times\left(\frac{1}{6}\right)\right] = \frac{A_c^2}{2}(1.1066)$$

Since

$$\gamma_{\min} = 103808.63 = \frac{S_R}{\eta W} = \frac{A_c^2\times 1.1066}{2\times 2\times 10^{-12}\times 5\times 10^3}$$

$$A_{c_{\min}}^2 = \frac{103808.63\times 20\times 10^{-9}}{1.1066} = 187617.26\times 10^{-8}$$

$\therefore$

$$A_{c_{\min}} = \sqrt{187617.26\times 10^{-8}} = 433\times 10^{-4} = 43.3 \text{ mV}$$

$\therefore$

$$A_{c_{\min}} = 43.3 \text{ mV}$$

(b) We are given that $\left(\frac{S}{N}\right)_{i\text{th}} = 10 \text{ dB} = 10 \text{ (ratio)}$

For AM,

$$\left(\frac{S}{N}\right)_i = \frac{S_R}{N_R} = \frac{\frac{A_c^2}{2}(1+m^2\overline{x^2})}{(\eta/2B_T)} = \frac{A_c^2(1+m^2\overline{x^2})}{2\eta W}$$

$$\frac{A_c^2\left[1+0.64\times\left(\frac{1}{6}\right)\right]}{2\times 2\times 10^{-12}\times 5\times 10^3} = \frac{1.1066A_c^2}{2\times 10^{-8}}$$

$\therefore$ at threshold,

$$\frac{S_R}{N_R} = 10$$

$\therefore$

$$\frac{1.1066A_c^2}{2\times 10^{-8}} = 10 \quad \text{or,} \quad A_c^2 = \frac{2\times 10^{-7}}{1.1066} = 0.18073\times 10^{-8}$$

$\therefore$

$$A_{c_{\text{th}}} = 0.425\times 10^{-4} \text{ V} \quad \text{or} \quad 0.0425 \text{ MV}$$

Example 7.6 An AM system employing an envelope detector in the receiver, is operating at threshold. Determine the increase in transmitter power (in dB) needed if an $\left(\frac{S}{N}\right)_D$ of 40 dB is desired. Assume $m = 1$ and tone modulation.

Solution In the first case, when the system is operating at threshold, let the received average signal power be S_{R_1}. If N_R is the average noise power that has entered the receiver, we have

$$\frac{S_{R_1}}{\eta B_T} = \frac{S_{R_1}}{2\eta W} = 10 \quad \therefore \frac{S_{R_1}}{\eta W} = 20; \quad \text{or} \quad S_{R_1} = 20\,\eta \text{W}$$

where W is the bandwidth of the modulating signal, i.e., the frequency of the modulating signal, since it is given as tone modulation. Let S_{R_2} be the received signal power required to obtain $(S/N)_D$ of 40 dB.

Then, $$\underset{\text{AM}}{\left(\frac{S}{N}\right)_D} = \left(\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}\right)\cdot\left(\frac{S_{R_2}}{\eta W}\right) = 10^4 = 10{,}000 \qquad \text{(since 40 dB} = 10^4\text{)}$$

Since $m = 1$ and it is tone modulation,

$$\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} = \left(\frac{1\cdot(1/2)}{1+1\cdot(1/2)}\right) = \frac{1}{3}$$

$$\therefore \quad \underset{\text{AM}}{\left(\frac{S}{N}\right)_D} = \frac{1}{3}\cdot\left(\frac{S_{R_2}}{\eta W}\right) = 10{,}000 \qquad S_{R_2} = 30{,}000\,\eta \text{W}$$

$$\therefore \quad \left(\frac{S_{R_2}}{S_{R_1}}\right) = \left(\frac{30000\,\eta W}{20\,\eta W}\right) = 1500$$

$$\therefore \quad \left.\begin{array}{l}\text{Increase in transmitter power (in decibels)}\\ \text{required to have a destination SNR of 40 dB}\end{array}\right\} = 10\log_{10}1500 = 31.76\text{ dB}$$

7.6.4 Comparison of Noise Performance of AM, DSB-SC and SSB-SC

We find that SSB-SC and DSB-SC have the same destination signal-to-noise ratio, as is evident from Eqs. (7.14) and (7.23), and that it is γ. In this connection, it must be noted that this is on the basis of the average power *in each sideband* being the same in the two cases. This is because, S_R, the received signal power is $\left(\frac{1}{4}\right)A_R^2\overline{x^2(t)}$ for SSB-SC and $\left(\frac{1}{2}\right)A_R^2\overline{x^2(t)}$ for DSB-SC. So, if the same message signal, $x(t)$ is considered in the two cases, DSB-SC gives twice as much signal power at the input to the detector as compared to SSB-SC. However, the transmission bandwidth, B_T, of DSB-SC being twice that of SSB-SC, the noise power that it brings in at the detector input is also twice as much when compared to SSB-SC. That is why their noise performances are the same.

From Eq. (7.27) it is clear that the total sideband power in the AM is $\left(\frac{1}{2}\right)A_R^2m^2\overline{x^2(t)}$, where m is the modulation index. This means that if we make $m = 1$, we will be able to compare the noise performance of AM with the noise performance of DSB-SC and SSB-SC *on the basis of equal average signal power per sideband*, i.e., the same basis on which we compared the noise performance of DSB-SC and SSB-SC. So, assuming that $m = 1$, the destination SNR for AM is

$$\underset{\substack{\text{AM}\\ m=1}}{(S/N)_D} = \left(\frac{\overline{x^2(t)}}{1+\overline{x^2(t)}}\right)\cdot\gamma \tag{7.47}$$

Note $\overline{x^2(t)}$ *is always less than or equal to 1 since x(t) is the normalized signal, normalized so that* $|x(t)| \leq 1$.

Since $\overline{x^2(t)}$ has got to be non-negative, this means that whatever may be the message signal $x(t)$, the destination SNR for AM is always less than γ, i.e., it is always inferior to DSB-SC and SSB-SC. This, of course, can be attributed to the fact that in the case of AM, the carrier power is rejected after demodulation and does not contribute to signal power at the destination.

The RHS of Eq. (7.47) makes it clear that the value of $\overline{x^2(t)}$ determines how small the value of the destination SNR with $m = 1$ would be as compared to γ.

1. **For tone modulation:** $\overline{x^2(t)} = 1/2$

$$\underset{\substack{\text{AM}\\ m=1}}{(S/N)_D} = \left(\frac{1/2}{1+1/2}\right)\cdot\gamma = \gamma/3$$

So, in the case of tone modulation, even with m = 1, the performance of AM is about 5 dB poorer compared to DSB-SC or SSB-SC.

2. **If** $\overline{x^2(t)}$ **takes its maximum possible value of 1** (as it would, for example, when $x(t)$ is a square wave), and with $m = 1$

$$\underset{\substack{\text{AM}\\ m^2\overline{x^2(t)}=1}}{(S/N)_D} = \left(\frac{1}{1+1}\right)\cdot\gamma = \gamma/2$$

So, even in this case, AM is still 3 dB poorer compared to DSB-SC and SSB-SC.

In the above two cases, we have assumed that $m = 1$ and $\overline{x^2(t)}$ was 0.5 in the case of tone modulation and 1 in the other case. But in actual practice, we have speech signal as the message signal. For this signal, m can hardly reach a value of 0.2 for most of the time since a speech signal has occasional large peaks and a very small amplitude in between. Further, this makes $\overline{x^2(t)}$ also very small. Because of these reasons, with speech as the modulating signal, the destination SNR of AM will be very much smaller than γ making its performance poorer than that of DSB-SC or SSB-SC by as much as 10 dB. However, peak limiting and volume compression of the audio, used in all broadcast transmitters will ensure a fairly good value of m for most of the time and this will help in improving the noise performance of AM to some extent.

Example 7.7 A message signal has a bandwidth of 15 kHz. This signal is to be transmitted over a channel whose attenuation is 80 dB and the two-sided noise PSD is 10^{-12} W/Hz. If it is desired to have a destination signal-to-noise ratio of 40 dB, what will be transmitter power (average) needed and what will be the transmission bandwidth, if the modulation is (a) SSB-SC, and (b) DSB-SC.

Solution

(a) $\gamma = \frac{S_R}{\eta W} = \frac{S_R}{2\times 10^{-12}\times 15\times 10^3} = \frac{10^8 S_R}{3}$

Channel attenuation = 80 dB $\quad\therefore\quad 10\log_{10}(S_T/S_R) = 80$

where S_T is the average transmitted power and S_R is the average received power.

$\therefore\quad S_T = 10^8 S_R \quad$ or $\quad S_R = 10^{-8} S_T$.

$$\therefore\quad \gamma = \frac{10^8\times 10^{-8}\times S_T}{3} = \frac{S_T}{3} = 40\text{ dB} = 10^4$$

$\therefore\quad S_T = 3\times 10^4 = 30$ kW. Since it is SSB-SC, $B_T = W = 15$ kHz.

(b) For DSB-SC, $(S/N)_D = \gamma = \dfrac{10^8 S_R}{3}$ and $S_R = 10^{-8} S_T$

$\therefore \dfrac{S_T}{3} = 10^4$ or $S_T = 30$ kW

Since both the sidebands are transmitted in DSB-SC, the bandwidth B_T required is $2W = 30$ kHz.

Example 7.8 A message signal with maximum amplitude of ±5 V is uniformly distributed and has a bandwidth of 15 kHz. Using AM with a modulation index of 0.6, it is transmitted over a channel whose attenuation is 60 dB and whose noise power spectral density (two sided) is 10^{-11} *W*/Hz. Determine the average power of the transmitter and the transmission bandwidth required, if a post-detection signal-to-noise ratio of 40 dB is desired.

Solution $\gamma = \dfrac{S_R}{\eta W} = \dfrac{S_R}{2 \times 10^{-11} \times 15 \times 10^3} = \dfrac{S_R \times 10^7}{3}$

Average power in the message signal (before normalization) = $\int_{-5}^{5} x^2 \dfrac{1}{10} dx = \dfrac{25}{3}$

$\overline{x^2(t)}$ = Average power in the message signal after normalization so that $|x(t)| \le 1$

$$= \frac{25}{3} \times \frac{1}{5^2} = \frac{1}{3}$$

$$\therefore \quad m^2 \overline{x^2(t)} = 0.36 \times \frac{1}{3} = 0.12 \qquad \therefore \frac{m^2 \overline{x^2(t)}}{1 + m^2 \overline{x^2(t)}} = \frac{0.12}{1.12}$$

Since channel attenuation = 60 dB = $10 \log_{10}(S_T/S_R)$

$$\therefore \quad \log_{10}\left(\frac{S_T}{S_R}\right) = 6 \qquad \therefore S_R = 10^{-6} S_T$$

$$\therefore \quad \left[\frac{m^2 \overline{x^2(t)}}{1 + m^2 \overline{x^2(t)}}\right] \cdot \gamma = \left(\frac{S}{N}\right)_D = \frac{10^{-6} S_T \cdot 10^7}{3} \cdot \frac{0.12}{1.12} = \frac{0.04 S_T}{1.12}$$

But $\underset{\text{AM}}{\left(\dfrac{S}{N}\right)_D}$ has to be 40 dB (= 10^4)

$$\therefore \quad \frac{0.04 S_T}{1.12} = 10^4 \quad \text{or} \quad S_T = \frac{10^4 \times 1.12}{0.04} = 280 \text{ kW}; \quad B_T = 2 \times 15 \text{ kHz} = 30 \text{ kHz}.$$

Example 7.9 A transmitter, transmitting an unmodulated carrier power of 20 kW is amplitude modulated to a depth of 0.8, by a message signal *x*(*t*) of 15 kHz bandwidth, which has an average power of 0.78 *W* when normalized so that $|x(t)| \le 1$. The modulated signal is transmitted over a channel whose attenuation is 60 dB and has an additive white noise with two-sided PSD of 10^{-12} *W*/Hz. Determine the pre-detection and post-detection SNRs at the receiver.

Solution P_c = unmodulated carrier power = 20×10^3 W

$$m = 0.8$$

$$\overline{x^2(t)} = 0.78\, W \quad \text{when} \quad |x(t)| \le 1$$

$\therefore$ average total power of the modulated signal = P_T

where $P_T = P_c[1 + m^2 \overline{x^2}] = 20 \times 10^3[1 + 0.64 \times 0.78] = 30 \times 10^3\, W$

$\therefore$ S_T = transmitted power = $30 \times 10^3\ W$

Attenuation of the channel = 60 dB (= a ratio of 10^{-6})

$$\therefore \quad \frac{S_T}{S_R} = 10^6 = \frac{30 \times 10^3}{S_R} \qquad \therefore S_R = 30 \times 10^{-3}\ W$$

$$\eta B_T = (2 \times 10^{-12}) \times (2 \times 15 \times 10^3) = 6 \times 10^{-8}\ W$$

$\therefore$ noise power at the input to the detector = $N_R = 6 \times 10^{-8}\ W$

$$\therefore \text{ pre-detection SNR} = \left(\frac{S}{N}\right)_R = \frac{S_R}{\eta B_T} = \frac{30 \times 10^{-3}}{6 \times 10^{-8}} = 5 \times 10^5 = \frac{1}{2} \times 10^6$$

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{\eta(B_T/2)} = 10 \times 10^5$$

$$\frac{m^2\,\overline{x^2(t)}}{1 + m^2\,\overline{x^2(t)}} = \frac{0.64 \times 0.78}{1 + 0.64 \times 0.78} = \frac{0.5}{1 + 0.5} = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$$

$$\therefore \text{ post-detection SNR} = \left[\frac{m^2\,\overline{x^2(t)}}{1 + m^2\,\overline{x^2(t)}}\right]\gamma = \frac{1}{3} \times 10^6$$

Example 7.10 An AM transmitter is to transmit a message signal having a bandwidth of 20 kHz and an average power (when normalized such that $|x(t)| \le 1$) of 1, over a transmission channel characterized by an additive white noise of two-sided PSD of 0.5×10^{-15} W/Hz and a total transmission loss of 100 dB. If the modulation index $m = 1$, determine the average transmitted power if destination SNR is to be 10^4.

Solution

For AM: $$(S/N)_D = \left(\frac{m^2\,\overline{x^2}}{1 + m^2\,\overline{x^2}}\right)\gamma = \frac{1 \cdot \gamma}{1+1} = 0.5\gamma$$

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{10^{-15} \times 20 \times 10^3} = \frac{S_R}{2 \times 10^{-11}}$$

$$\therefore \quad \left(\frac{S}{N}\right)_D = \frac{0.5 S_R}{2 \times 10^{-11}} = 0.25 S_R \times 10^{11} = 10^4$$

$$\therefore \quad S_R = 4 \times 10^{-7}\ W. \quad L = 100\text{ dB} = 10^{10}$$

$$\therefore \quad S_T = 4 \times 10^{-7} \times 10^{10} = 4\text{ kW}$$

Example 7.11 A message signal of bandwidth 5 kHz is to be transmitted using SSB-SC over a transmission channel characterized by an additive white noise of two-sided PSD $\eta/2 = 0.5 \times 10^{-15}$ W/Hz and a transmission loss of 100 dB. If a destination SNR of 40 dB is required, determine the average transmitter power required.

Solution $$\left(\frac{S}{N}\right)_{D\atop \text{SSB}} = \gamma = \frac{S_R}{\eta W} = 10^4$$

$$\therefore \quad S_R = 10^{-15} \times 5 \times 10^3 \times 10^4 = 10^{-8}\ W. \text{ Also, 100 dB = a ratio of } 10^{10}$$

$$\therefore \quad S_R = 10^{-8}\ W. \text{ But } \frac{S_T}{S_R} = 10^{10} \quad \therefore S_T = 10^{-8} \times 10^{10} = 10^2 = 100\ W$$

7.7 NOISE PERFORMANCE OF FREQUENCY MODULATED SYSTEMS

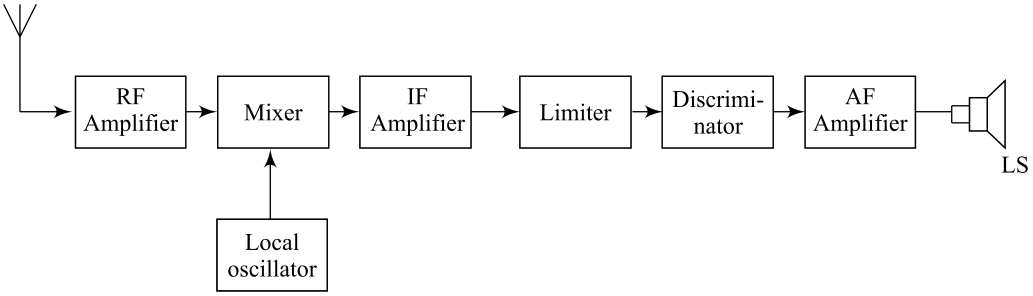


Fig. 7.6 *Block diagram of a FM broadcast superheterodyne receiver*

The block diagram of an FM broadcast superheterodyne receiver is shown in Fig. 7.6. For the purpose of noise performance evaluation, we model the receiver as shown in Fig. 7.7.

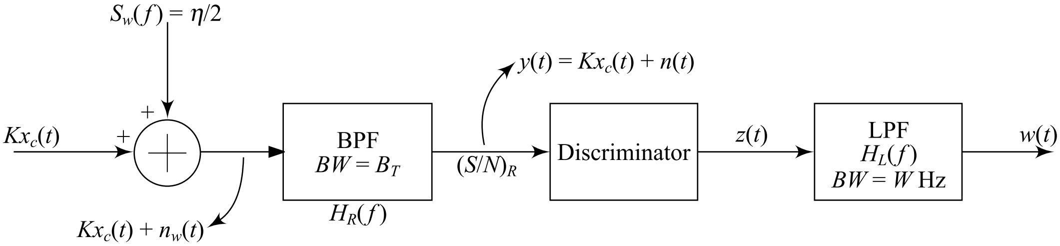


Fig. 7.7 *Receiver model for noise performance evaluation*

Additive noise of the channel is modeled as zero-mean white Gaussian noise of a two-sided power spectral density η/2 · *K* represents the channel attenuation. The modulated signal in this case, is given by

$$x_c(t) = A_c \cos[\omega_c t + \phi(t)], \quad f_c = f_{i \cdot f} \text{ of the receiver} \tag{7.48}$$

where

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha) d\alpha \tag{7.49}$$

Note *In Eq. (7.49), k_f is the frequency deviation constant if the message signal x(t), is not normalized. If x(t) is normalized, k_f denotes the peak frequency deviation.*

The ideal BPF whose response characteristic is shown in Fig. 7.8, is used to represent the combined effect of the RF and IF amplifiers.

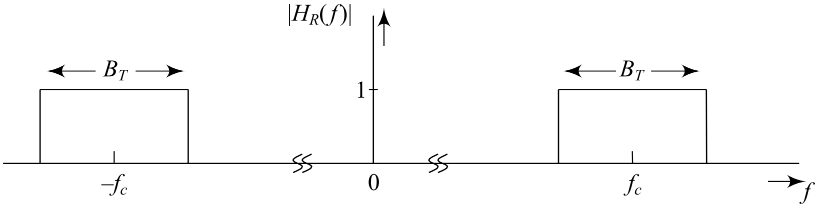


Fig. 7.8 *Response characteristic of the BPF*

The bandwidth of this BPF is the transmission bandwidth B_T of the modulated signal, $x_c(t)$, and is also the bandwidth of the front-end of the receiver. The signal at the input to the filter is $Kx_c(t) + n_w(t)$, i.e., the modulated signal and the additive white noise. Its output, however is $Kx_c(t) + n(t)$, where $n(t)$ is band pass noise centered on f_c and obtained by filtering the white noise using the BPF of bandwidth B_T, with center frequency $f_c = f_{i \cdot f}$, the intermediate frequency of the superheterodyne receiver. The FM detector, called the discriminator produces an output voltage which at any instant, is proportional to the deviation of the instantaneous frequency of the input signal from the carrier (i.e., in this case the IF) frequency.

The input signal for the discriminator is $Kx_c(t) + n(t)$ where $x_c(t)$ is the FM signal and $n(t)$ is band pass noise centered on f_c. In the case of amplitude modulation, the additive noise would simply add to the amplitude modulated signal $x_c(t)$ and thus change its envelope which the envelope detector would extract. So, in the case of AM, the additive noise *directly* affects that parameter of the input signal (envelope) which the detector tries to extract. So the effect of the additive noise is considerable in the case of AM. But in the case of FM, the discriminator extracts the frequency deviation of the carrier of the input signal each instant, and produces an output voltage proportional to the instantaneous frequency deviation. The additive noise does not *directly* affect the frequency deviation of the incoming FM signal. It affects it only indirectly, as we will be see in the following paragraphs. Thus, in a qualitative way, we may say that FM will not be affected by the channel noise to the same extent as AM.

Since the bandwidth of the BPF is B_T and the two-sided PSD of the additive white noise in the channel is $\eta/2$, the noise power entering the receiver is

$$\overline{n^2(t)} = \frac{\eta}{2} \times 2B_T = \eta B_T \triangleq N_R \tag{7.50}$$

The received signal power is equal to the average power of the component $Kx_c(t)$ of $y(t)$, the input to the discriminator. This is denoted by S_R and is given by

$$S_R = \frac{(KA_c)^2}{2} = \frac{A_R^2}{2} \tag{7.51}$$

∴ the pre-detection SNR is given by

$$\left(\frac{S}{N}\right)_R = \frac{S_R}{N_R} = \frac{A_R^2}{2} \cdot \frac{1}{\eta B_T} = \frac{A_R^2}{2\eta B_T} \tag{7.52}$$

As mentioned earlier, $n(t)$ is band pass noise centered on f_c and we may represent it by its inphase and quadrature components as

$$n(t) = n_i(t)\cos \omega_c t - n_q(t)\sin \omega_c t \tag{7.53}$$

Alternatively, we may use the envelope and phase angle representation (see Section 2.8 Eq. (2.164)) and write as

$$n(t) = R_n(t)\cos[\omega_c t + \phi_n(t)] \tag{7.54}$$

where $R_n(t)$, the envelope is related to $n_i(t)$ and $n_q(t)$ by

$$R_n(t) = \sqrt{n_i^2(t) + n_q^2(t)} \tag{7.55}$$

and is Rayleigh distributed. The phase angle, $\phi_n(t)$ is given by

$$\phi_n(t) = \tan^{-1}\left[\frac{n_q(t)}{n_i(t)}\right] \tag{7.56}$$

As it is more convenient in the present analysis to use the envelope and phase representation, we shall write $y(t)$, the input to the discriminator as

$$\begin{aligned} y(t) &= A_R \cos[\omega_c t + \phi(t)] + n(t) \\ &= A_R \cos[\omega_c t + \phi(t)] + R_n(t)\cos[\omega_c t + \phi_n(t)] \end{aligned} \tag{7.57}$$

We shall make use of Eq. (7.57) to examine how the noise term $n(t)$ affects the angle $\phi(t)$ of the FM signal and thus changes its frequency deviation. However, this is going to be quite involved. So, we shall proceed by making the simplifying and reasonable assumption that the SNR at the input to the discriminator is high.

i.e., $$\left(\frac{S}{N}\right)_R >> 1 \tag{7.58}$$

If under this assumption, we draw the phasor diagram for Eq. (7.57), it will appear as shown in Fig. 7.9.

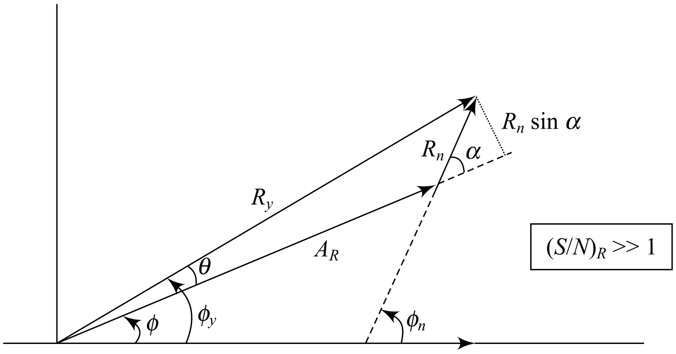


Fig. 7.9 *Phasor diagram of Eq. (7.54) when (S/N)$_R$ >> 1*

For the band pass signal $y(t)$, if $R_y(t)$ is the envelope and $\phi_y(t)$, the phase angle, we may write

$$y(t) = R_y(t)\cos[\omega_c t + \phi_y(t)] \tag{7.59}$$

Since $y(t)$ is the input to the discriminator, what the discriminator does is, it produces an output $z(t)$ which at any instant, is proportional to the instantaneous frequency deviation given by

$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}[\phi_y(t)] \tag{7.60}$$

So let

$$z(t) = \frac{1}{2\pi}\frac{d}{dt}[\phi_y(t)] \tag{7.61}$$

The phasor diagram of Fig. 7.9 shows how the additive noise component $n(t)$ affects the phase angle ϕ and thereby the frequency deviation, of the incoming FM signal. $\phi(t)$ is the phase angle of the received FM signal, $\phi_n(t)$ is the phase angle of the band pass noise component $n(t)$. The sum of the phasors A_R and R_n gives R_y, the envelope of $y(t)$, the phase angle of which is $\phi_y(t)$. Note that because of our assumption that the pre-detection SNR is very much greater than 1,

$$P[R_n(t) << A_R] \text{ is almost equal to } 1. \tag{7.62}$$

But from Fig. 7.7

$$\sin\theta(t) = [R_n(t)\sin\alpha(t)]/R_y(t) \tag{7.63}$$

However, from Eq. (7.62), it follows that the following small-angle approximation can be made so that

$$\sin\theta(t) \cong \theta(t)$$

and hence, Eq. (7.63) may be rewritten as

$$\theta(t) = \frac{R_n(t)\sin\alpha(t)}{R_y(t)} \tag{7.64}$$

Thus, since

$$\phi_y(t) = \phi(t) + \theta(t) \tag{7.65}$$

we have

$$\phi_y(t) = \phi(t) + \frac{R_n(t)\sin\alpha(t)}{R_y(t)} \tag{7.66}$$

But because of Eq. (7.62), we may make the following approximation:

$$R_y(t) \cong A_R \tag{7.67}$$

Hence, from Eqs. (7.61) and (7.65), we have

$$z(t) = \text{discriminator output signal}$$

$$= \frac{1}{2\pi}\frac{d}{dt}[\phi_y(t)] = \frac{1}{2\pi}\frac{d}{dt}\phi(t) + \frac{1}{2\pi}\frac{d}{dt}\theta(t) = k_f x(t) + n_d(t) \tag{7.68}$$

Since $\phi(t)$ is the phase angle caused due to frequency modulating the carrier by the message signal $x(t)$, from Eqs. (7.48) and (7.49), the first term in Eq. (7.68) clearly represents the message signal component in the output of the discriminator. Since $\theta(t)$ is the additional phase caused by noise, the second term of Eq. (7.68) represents the noise term in the output of the discriminator and is denoted by $n_d(t)$.

To see how much of this noise goes past the low pass filter and reaches the destination, we have to examine the spectrum of the noise term in Eq. (7.68). For this purpose, let us rewrite it as follows:

$$\frac{1}{2\pi}\frac{d}{dt}\theta(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\frac{R_n(t)\sin\alpha(t)}{A_R}\right] \qquad \text{(From Eqs. (7.64) and (7.67))} \qquad (7.69)$$

From the phasor diagram of Fig. 7.7, we find that

$$\alpha(t) = \phi_n(t) - \phi(t) \qquad (7.70)$$

This seems to indicate that the post-detection noise, $n_d(t)$, is dependent on the modulation angle $\phi(t)$. Now, $\phi_n(t)$ is the phase angle of the band pass noise in its envelope – phase angle representation. But, we know that in such a representation, the envelope is *Rayleigh distributed* while the phase angle $\phi_n(t)$ is *uniformly distributed* over $-\pi$ to $+\pi$ (see Example 7.13). If we can assume that $\alpha(t)$, which is $[\phi_n(t) - \phi(t)]$, is itself uniformly distributed over $-\pi$ to $+\pi$, then this coupling between the post-detection noise and the modulation angle will be removed and $n_d(t)$ will be independent of modulation. Rice has shown that such an assumption is justified provided the carrier-to-noise ratio is large. In that case, we may, for a moment, assume that there is no modulation and that only an unmodulated carrier is transmitted. In such a case, the phasor diagram, will appear as shown in Fig. 7.10 (since $\phi(t) = 0$ when there is no modulation).

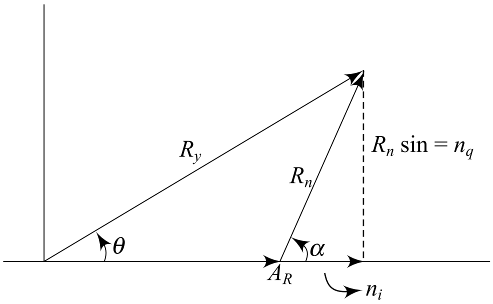


Fig. 7.10 *Phasor diagram with no modulation $(S/N)_R >> 1$*

Since $\phi(t) = 0$, $\alpha(t) = \phi_n(t)$ and so

$$R_n(t)\sin\alpha t = R_n(t)\sin\phi_n(t) = n_q(t) \qquad (7.71)$$

Hence, Eq. (7.69) may be rewritten as

$$\frac{1}{2\pi}\frac{d}{dt}\theta(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\frac{R_n(t)\cdot\sin\alpha t}{A_R}\right]$$

$$\therefore \qquad n_d(t) = \frac{1}{A_R}\frac{1}{2\pi}\frac{d}{dt}[n_q(t)] \qquad (7.72)$$

Therefore, to determine how much power of this post-detection noise goes past the low pass filter with a cut-off frequency of W Hz, we have to determine the power spectrum of $n_d(t)$. To do this, we first note that $n_q(t)$ is the low pass equivalent of the band pass noise, $n(t)$, that has entered the receiver. Since the BPF at the front-end of the receiver has a transfer function of $H_R(f)$, its output, $n(t)$, will have a power spectrum of

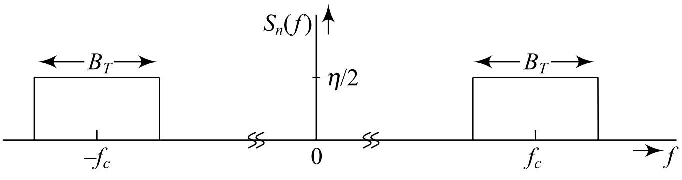


Fig. 7.11 *Power spectral density of n(t)*

$$S_n(f) = S_{n_w}(f)|H_R(f)|^2 = \frac{\eta}{2}|H_R(f)|^2 \qquad (7.73)$$

The PSD of the band pass noise, $n(t)$, is shown in Fig. 7.11, and that of its low pass equivalent, $n_q(t)$, is shown in Fig. 7.12.

$$\therefore \qquad S_{nq}(f) = \eta\Pi(f/B_T) \tag{7.74}$$

The power spectrum of $\frac{1}{A_R}\cdot\frac{1}{2\pi}n_q(t)$ is then given by

$$\frac{1}{(2\pi)^2}\cdot\frac{1}{A_R^2}\cdot S_{nq}(f) \tag{7.75}$$

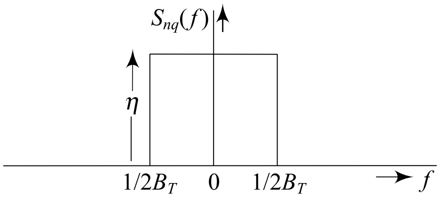


Fig. 7.12 *Power spectral density of $n_q(t)$, low pass equivalent of $n(t)$*

To find the power spectrum of the post-detection noise $n_d(t)$, in view of Eq. (7.72), we proceed as in Fig. 7.13.

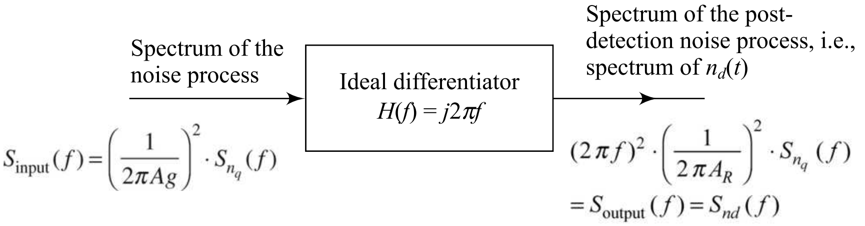


Fig. 7.13 *Deriving the spectrum of post-detection noise*

Substituting for $S_{nq}(f)$ in the expression for the $S_{nd}(f)$ and simplifying, we get

$$S_{n_d}(f) = \left(\frac{\eta f^2}{A_R^2}\right)\Pi\left(\frac{f}{B_T}\right)$$

$$= \left(\frac{\eta f^2}{2S_R}\right)\Pi\left(\frac{f}{B_T}\right) \tag{7.76}$$

A sketch of the post-detection noise spectrum is given in Fig. 7.14. While the message has a bandwidth of only W Hz, this noise process has a bandwidth of $B_T/2$, which is much greater than W. Hence, there is considerable noise outside the message bandwidth. This out-of-band noise has to be removed using a low pass filter having a cut-off frequency of W Hz.

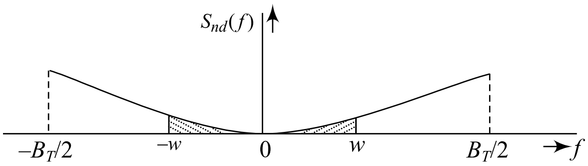


Fig. 7.14 *Power spectral density of post-detection noise*

The average power of the noise at the output of the low pass filter = N_D = Destination noise power = Area of the shaded region

$$= \int_{-W}^{W}\left(\frac{\eta f^2}{2S_R}\right)\Pi\left(\frac{f}{B_T}\right)df$$

$$= \int_{-W}^{W}\frac{\eta f^2}{2S_R}df = \frac{\eta W^3}{3S_R}$$

$$\therefore \qquad N_D = \frac{\eta W^3}{3S_R} \tag{7.77}$$

The message signal component at the output of the discriminator has been found (refer to Eq. (7.68)) to be $k_f\,x(t)$. Since this has a bandwidth of W, all of it passes through the low pass filter. Hence, the destination signal power is given by

$$S_D = k_f^2\,\overline{x^2(t)} \tag{7.78}$$

∴ destination signal-to-noise ratio is given by

$$\left(\frac{S}{N}\right)_D = \left(\frac{S_D}{N_D}\right) = \frac{k_f^2\,\overline{x^2(t)}}{(\eta W^3/3S_R)}$$

$$= 3\left(\frac{k_f}{W}\right)^2 \overline{x^2(t)}\left(\frac{S_R}{\eta W}\right)$$

But, we know that when $x(t)$ is the normalized message signal, k_f denotes the peak frequency deviation (refer to the note under Eq. (7.49)). Since we have k_f over W as a factor in the above expression for the destination SNR, let us replace that factor by the deviation ratio denoted by D.

∴

$$\boxed{\underset{\text{FM}}{\left(\frac{S}{N}\right)_D} = 3D^2\,\overline{x^2(t)}\,\gamma} \tag{7.79a}$$

Hence, from Eqs. (7.79a) and (7.7), the figure of merit for FM systems may be written as

$$\boxed{\underset{\text{(FM)}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{3D^2\,\overline{x^2}\gamma}{\gamma} = 3D^2\,\overline{x^2}} \tag{7.79b}$$

Example 7.12 For tone modulation, show that the Figure of Merit of an FM system is given by $\frac{3}{2}\beta_f^2$, where β_f is the modulation index.

Solution D = Deviation ratio = $\left(\frac{k_f A_m}{W}\right)$, where A_m is the peak amplitude and $W = f_m$, is the frequency of the single-tone modulating signal. Because $|x(t)| \le 1$, $A_m = 1$ and so $\overline{x^2} = 1/2$, substitution in Eq. (7.79b) gives

$$\text{Figure of Merit} = \frac{3}{2}\left(\frac{k_f A_m}{f_m}\right) = \frac{3}{2}\beta_f^2 \qquad \text{(see Eq. (5.18))}$$

Remark

(i) *Although derived under certain assumptions, the result represented by Eq. (7.79) is indeed a very significant one. This is because, it says that as long as the assumptions under which it is derived are not violated, the destination signal-to-noise ratio can be increased just by increasing the deviation ratio without having to increase the average transmitted power. When the deviation ratio D is increased, we know that the transmission bandwidth, B_T, increases, because $B_T = 2(D+1)W$. So, Eq. (7.79) tells us that the destination SNR can be increased by increasing the transmission bandwidth without increasing the transmitter power. This means there is a 'power-bandwidth' trade-off possible in the case of FM. This is something which is not possible in the case of AM, where the bandwidth is fixed and does not depend on the value of the modulation index, m.*

(ii) *This 'power-bandwidth' trade-off is, however, not without a limit. We must realize that as the transmission bandwidth B_T is increased to get better destination signal-to-noise ratio, the average noise power entering the receiver also increases, since it is equal to ηB_T; but the received signal power does not, because it is equal to $\frac{A_R^2}{2}$. Thus, along with the bandwidth the received noise power increases, making $(S/N)_R$ smaller and smaller. Hence, a situation will arise at some value of the B_T, in which the assumption that $(S/N)_R$ is large, which we made use of while deriving Eq. (7.79), will no longer be valid.*

(iii) The relative immunity that it enjoys with regard to the additive noise on the channel, its ability to handle message bandwidths up to even 15 to 20 kHz (with very little increase in transmission bandwidth) which makes it extremely useful for transmission of high quality music, and the flexibility that it offers through the 'power-bandwidth' trade-off, make FM a really attractive proposition.

Example 7.13 A single-tone modulating signal $f(t) = E_m \cos \omega_m t$ phase modulates a carrier signal $A_c \cos \omega_c t$. Show that the figure of merit for PM = $\frac{1}{2} m_f^2$ where m_f is the modulation index for FM.

Solution We know that in the case of phase modulation,

$$\phi(t) = k_p x(t) = k_p E_m \cos \omega_m t$$

∴ peak frequency deviation produced by this phase modulation

$$= \left[\left| \frac{d}{dt} \phi(t) \right| \right]_{\max} = k_p E_m \omega_m = \text{say } (\Delta\omega)_{\text{PM}}$$

If m_f is the modulation index for FM,

$$m_f = \frac{(\Delta f)}{f_m} = \frac{\Delta\omega}{\omega_m} = k_p E_m$$

In the case of PM, the figure of merit is given by

$$(\text{FOM})_{\text{PM}} = k_p^2 \overline{x^2(t)} = k_p^2 \cdot \frac{E_m^2}{2}; \quad \text{But } k_p E_m = m_f$$

∴ $$(\text{FOM})_{\text{PM}} = \frac{1}{2} m_f^2$$

Example 7.14 Show that narrowband FM does not offer any better destination signal-to-noise ratio than AM.

Solution

For AM: $\left(\frac{S}{N}\right)_D = \left(\frac{m^2 \overline{x^2(t)}}{1 + m^2 \overline{x^2(t)}}\right) \cdot \gamma$

The maximum value of this occurs when $m^2 \overline{x^2(t)} = 1$, i.e., $m = 1$ and $\overline{x^2(t)} = 1$.

Then $$\left(\frac{S}{N}\right)_D = \frac{1}{2} \cdot \gamma$$

For FM: $$\left(\frac{S}{N}\right)_D = 3\beta_f^2 \overline{x^2(t)} \gamma$$

Assuming $\overline{x^2(t)} = 1$ for this case also,

$$\left(\frac{S}{N}\right)_D = 3\beta_f^2 \gamma$$

∴ if this is to be better than the $\left(\frac{S}{N}\right)_D$ for AM,

$$3\beta_f^2 \gamma = \frac{1}{2} \gamma \quad \text{or} \quad \beta_f^2 = \frac{1}{6} \quad \therefore \beta_f > 0.408$$

i.e., for FM to be better than AM, β_f should be greater than 0.408. But for NBFM, $\beta_f < 0.2$.

∴ NBFM is no better than AM.

7.8 PRE-EMPHASIS AND DE-EMPHASIS

In the last section, we found that the power spectral density of the post-detection noise varies as the square of the frequency. This means that within the message bandwidth $-W \le f \le W$, the high frequency components of the message signal will, after detection, encounter a much higher noise power than the low frequency components. This tends to make the destination signal-to-noise ratio poor for the high frequency components of the message. Unfortunately, there exists another factor, associated with the power spectral density of the message itself, which too tends to make the destination SNR worse for the high frequency components of the message. Audio message signals in general, and speech message signals in particular, generally have a power spectral density that tends to fall rather sharply beyond about 800 Hz to 1 kHz. Thus, compared to the low frequency components, the high frequency components are much weaker and produce much smaller frequency deviation. Hence, at the output of the discriminator in the receiver, the high frequency message signal components at these frequencies will be quite weak; but the noise frequency components at these frequencies will be quite strong. Thus, the SNR for high frequency components of the message will be poor. This will reduce the overall destination SNR of the receiver.

'Pre-emphasis and de-emphasis' is a technique quite often used in all FM systems in order to overcome the problem stated above, and improve the destination SNR. The 'pre-emphasis' part of the process, performed at the transmitting end, consists of boosting-up of the high frequency components of the message signal before using it for modulation, so as to make the PSD of the message more uniform within its bandwidth of $-W \le f \le +W$. Because of pre-emphasis, the signal at the output of the discriminator will be a distorted version of the original message. Hence, the output of the discriminator (signal plus noise) is subjected to the de-emphasis process so as to restore the original relative amplitude values of the various frequency components of the message signal. The de-emphasis process consists of appropriately attenuating the high frequency components of the output of the discriminator, to compensate for the 'boosting-up' or pre-emphasis done at the transmitting end. In this process of de-emphasis therefore, while the message spectrum is restored to its original form, amplitudes of the *high frequency components of the noise* at the output of the discriminator are also reduced, thereby improving the SNR at the destination. *This method is effective because the boosting-up of the high frequency components is done at the transmitter* ***before*** *channel noise enters and attenuating of the high frequency components is done in the receiver at the output of the discriminator so that high frequency components of both the message signal and the post-detection noise, are attenuated.* For introducing pre-emphasis and de-emphasis, a pre-emphasis filter $H_{pe}(f)$ is included in the transmitter and a de-emphasis filter $H_{de}(f)$ is included in the receiver after the discriminator stage, as shown in Fig. 7.15.

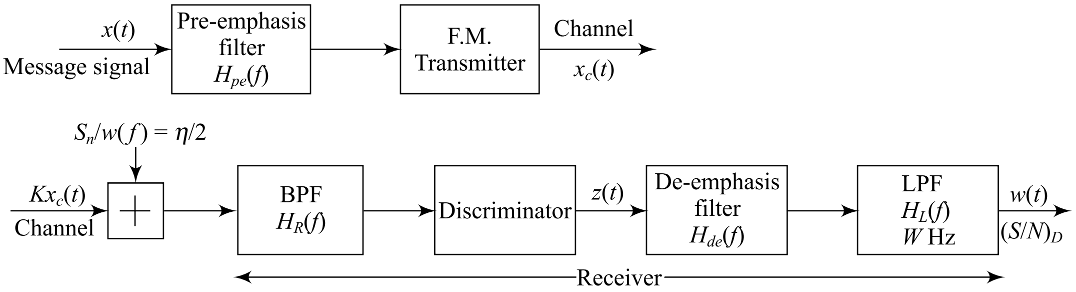


Fig. 7.15 *Pre-emphasis and de-emphasis in an FM system*

The de-emphasis filter should come after the discriminator stage and may be placed either before, or after the LPF. This is because, both the LPF and $H_{de}(f)$ being linear time-invariant systems, the order in which they are placed is immaterial.

Ideally, the transfer functions of the pre-emphasis and de-emphasis filters should be inverses of each other, at least over the message bandwidth, W.

$$H_{de}(f) = \frac{1}{H_{pe}(f)}; \qquad |f| \le W \tag{7.80}$$

Since the pre-emphasis filter should boost up the high frequency components of the message signal and leave the low frequency components practically unaffected, the following simple transfer function is generally used for it:

$$H_{pe}(f) = \left(1 + j\frac{f}{f_0}\right) \tag{7.81}$$

In this, f_0 is a fixed frequency. The corresponding $H_{de}(f)$ is

$$H_{de}(f) = \frac{1}{\left(1 + j\dfrac{f}{f_0}\right)} \tag{7.82}$$

Figures 7.16 and 7.17 show the typical magnitude responses of the pre-emphasis and the de-emphasis filters. As long as $f << f_0$, the magnitude responses of both the filters remain practically constant. At $f = f_0$, the response of the pre-emphasis filters is +3 dB, while that of the de-emphasis filter is –3 dB.

$H_{pe}(f)$, which is *essentially* the response of a differentiator, can be closely realized by a simple RC filter shown in Fig. 7.18(a) and $H_{de}(f)$, which is *essentially* the response of an integrator, can be closely realized by the simple RC filter shown in Fig. 7.18(b).

For commercial FM broadcasting, for which W = 15 kHz, the value of the time constant rc is set equal to 75 μs so that f_0, the 3-dB frequency is equal to 2122 Hz.

Since the response of the pre-emphasis filter is almost constant for low message frequencies, and that of a differentiator for high message frequencies, and noting that the message signal passes through this filter before being used for frequency modulating the carrier, we may say that the pre-emphasis filter makes the low frequency compo-

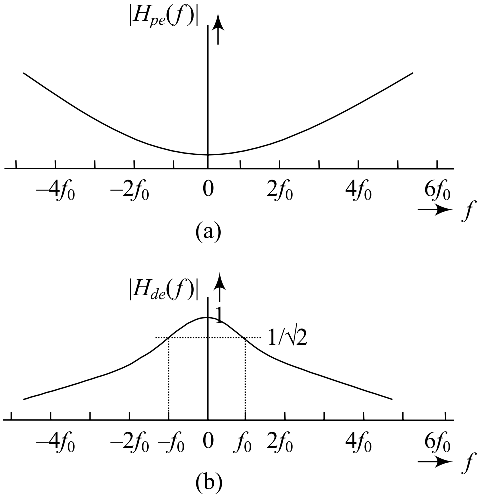


Fig. 7.16 *Magnitude responses (a) Pre-emphasis and (b) de-emphasis filters (linear frequency scale)*

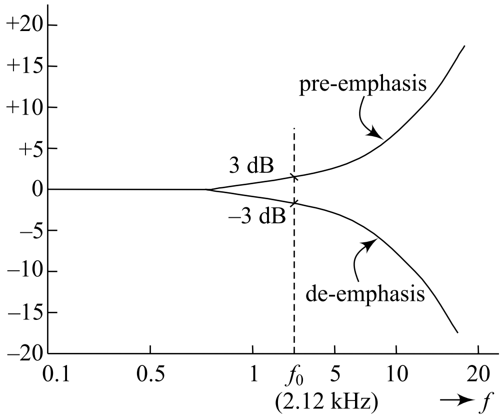
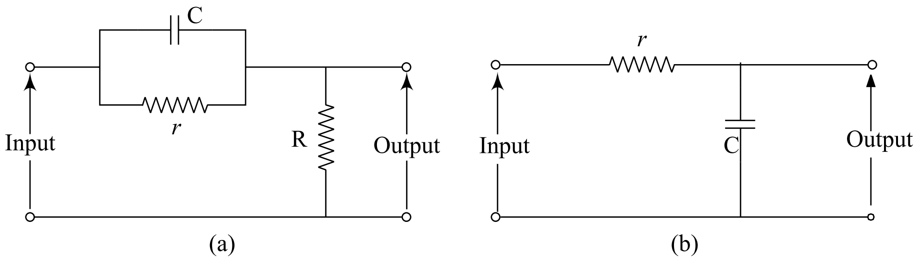


Fig. 7.17 *Magnitude responses (in decibels) of pre-emphasis and de-emphasis. Filters plotted using logarithmic scale for frequency*



(a) (b)

Fig. 7.18 *(a) Pre-emphasis filter, (b) De-emphasis filter* nents of the message to frequency modulate the carrier while making the high frequency components of the message signal to phase modulate it. Similarly, the discriminator together with the de-emphasis filter may be considered to be working as a frequency demodulator for low message frequencies and as a phase demodulator for high message frequencies.

7.8.1 Improvement in Destination SNR due to Pre-emphasis and De-emphasis

To make a quantitative evaluation of the improvement in $(S/N)_D$ caused by the 'pre-emphasis, de-emphasis' technique, we first note that S_D, the signal power at the destination is unaffected by the presence or absence of the pre-emphasis filter at the transmitter and de-emphasis filter in the receiver. It is only the destination noise power that is getting reduced and this reduction is caused only by the de-emphasis filter in the receiver. Hence, a good quantitative measure of the destination SNR improvement due to the use of pre-emphasis and de-emphasis is given by the following ratio:

$$I \underline{\Delta} \left(\frac{\text{Noise power output at the destination without de-emphasis}}{\text{Noise power output at the destination with de-emphasis}} \right) \tag{7.83}$$

Earlier, we found (see Eq. (7.76)) that the PSD of the noise at the output of the discriminator is given by

$$S_{nd}(f) = \left(\frac{\eta f^2}{2S_R} \right) \Pi \left(\frac{f}{B_T} \right)$$

So, this is the PSD of the noise at the input to the 'de-emphasis filter – baseband low pass filter' combination.

To find how the combination of these two filters will modify this post-detection noise spectrum, let us say $H_c(f)$ is the overall transfer function of the cascade connection of these two filters.

Then

$$H_c(f) = H_{de}(f) \cdot H_L(f) \tag{7.84}$$

and

$$H_L(f) = \Pi(f/2W) \tag{7.85}$$

since the baseband filter has been modeled as an ideal LPF with a cut-off frequency of W Hz.

Hence, noise PSD at the destination is given by

$$S_D(f) = S_{nd}(f) \cdot |H(f)|^2 \tag{7.86}$$

∴ average noise power at the destination, i.e., at the output of the receiver *with* pre-emphasis and de-emphasis is given by

$$N_D = \int_{-W}^{W} S_D(f) df = \int_{-W}^{W} S_{nd}(f) |H_c(f)|^2 df$$

$$= \int_{-W}^{W} \frac{\eta f^2}{2S_R} \cdot \Pi(f/B_T) \Pi(f/2W) |H_{de}(f)|^2 df$$

$$= \int_{-W}^{W} \frac{\eta f^2}{2S_R} |H_{de}(f)|^2 df = \frac{\eta}{A_R^2} \int_{-W}^{W} f^2 |H_{de}(f)|^2 df \tag{7.87}$$

Average noise power at the output of the receiver *without* the pre-emphasis and de-emphasis, is given by

$$N_0 = \int_{-W}^{W} S_{nd}(f) df = \int_{-W}^{W} \frac{\eta f^2}{2S_R} df$$

$$= \frac{\eta}{A_R^2} \int_{-W}^{W} f^2 df = \frac{2\eta W^3}{3A_R^2} \tag{7.88}$$

Note

The limits for the integral are –W and +W because of the baseband filter $H_L(f)$. Substituting these in Eq. (7.83), we have

$$\boxed{I = \frac{N_0}{N_D} = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2\, df}} \tag{7.89}$$

For the type of de-emphasis filter used in FM broadcast receivers, the destination SNR improvement due to the use of pre-emphasis and de-emphasis works out to about 13 dB, which represents a substantial improvement.

Example 7.15 Show that the improvement in $(S/N)_D$ due to the use of de-emphasis filter in a broadcast FM receiver is of the order of 13 dB.

Solution From Eq. (7.82), we have

$$|H_{de}(f)|^2 = \left(\frac{f_0^2}{f_0^2 + f^2}\right)$$

∴ from Eq. (7.89), we get

$$\text{Improvement in } \left(\frac{S}{N}\right)_D = I = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2\, df}$$

$$= \frac{2W^3}{3\int_{-W}^{W} f^2 \left[\frac{f_0^2}{f_0^2 + f^2}\right] df} = \frac{1}{3}\left[\frac{(W/f_0)^3}{(W/f_0) - \tan^{-1}(W/f_0)}\right]$$

In a commercial FM broadcast receiver, $W = 15$ kHz and $f_0 = 2122$ Hz. Substituting these values in the above

$I \approx 22$ ∴ improvements of $(S/N)_D$ in dB = $10 \log_{10} 22 \approx 13$ dB.

Example 7.16 The ratio of $(S/N)_D$ to γ is referred to as the figure of merit of a system. Assuming that the normalized message signal has a bandwidth of W Hz and an average power of 0.5, determine the Figures of Merit for an FM system (without pre-emphasis, de-emphasis) and an AM system.

Solution

For FM: $\left(\frac{S}{N}\right)_D = 3\beta^2\, \overline{x^2(t)}\,\gamma$

with $\overline{x^2(t)} = 0.5$, the figure of merit for an FM system is

$$F = [(S/N)_D/\gamma] = \frac{3}{2}\beta^2$$

For AM: $\left(\frac{S}{N}\right)_D = \left[\frac{m^2\,\overline{x^2}}{1 + m^2\,\overline{x^2}}\right]\gamma$

∴ $$F = \frac{(m^2/2)}{1 + \frac{1}{2}m^2} = \left[\frac{m^2}{2 + m^2}\right]$$

7.9 THRESHOLD EFFECT IN FM

For wideband FM, we have seen that (refer to Eq. (7.79)) the destination SNR is given by

$$\underset{\text{WBFM}}{\left(\frac{S}{N}\right)_D} = 3D^2\,\overline{x^2}\,\gamma \tag{7.90}$$

As has already been observed in point ii. of Remarks below Example 7.12, it then follows that just by increasing the deviation ratio, D, it is possible to increase the destination signal-to-noise ratio without increasing the transmitted power. In other words, it means that it is possible to exchange transmitter power for the bandwidth of the transmitted signal.

However, as explained in that remark, this exchange is not without a limit. As we increase the bandwidth, B_T, by increasing the deviation ratio D at the transmitter while keeping the average transmitted power $(A_c^2/2)$ constant, the destination signal-to-noise ratio at the output of the receiver will, of course, increase initially. But, as B_T is increased, the average noise power entering the receiver, given by ηB_T, will also be increasing. However, since $(A_c^2/2)$, the average transmitted power is held constant, the signal power entering the receiver, given by $(K^2A_c^2/2) = (A_R^2/2)$, is also constant, while the average noise power entering the receiver increases. So, the receiver input signal-to-noise ratio, $(S_R/N_R) = (S_R/\eta B_T)$ goes on decreasing. Thus, as we go on increasing D to get better $(S/N)_D$, a stage will be reached at some value of D and B_T, at which, the input SNR for the receiver will be so low that the basis on which we had derived Eq. (7.79), viz., that the $(S/N)_R$ is quite high, will be violated, making the application of Eq. (7.79) no longer appropriate.

This fact, that Eq. (7.79) is not applicable below a certain value of input SNR, is clearly brought out by a plot of $(S/N)_D$ vs. (S_R/N_R) for a fixed D. For the purpose of plotting this curve, let us assume single-tone modulation so that

$$x(t) = \cos \omega_m t \tag{7.91}$$

since $x(t)$ has been normalized in such a way that $|x(t)| \le 1$.

∴ from Eq. (7.91), we have

$$\overline{x^2(t)} = \frac{1}{2} \tag{7.92}$$

Further, since $x(t)$ is single tone, and has been normalized,

$$D = \beta \tag{7.93}$$

From the above, Eq. (7.90) may therefore be written as

$$\underset{\text{WBFM}}{\left(\frac{S}{N}\right)_D} = \frac{3}{2}\beta^2\gamma \tag{7.94}$$

or

$$\left(\frac{S}{N}\right)_{D;\text{dB}} = 10\log_{10}\left(\frac{3}{2}\beta^2\right) + 10\log_{10}\gamma$$

$$\therefore \quad \boxed{\left(\frac{S}{N}\right)_{D;\text{dB}} = 10\log_{10}\left(\frac{3}{2}\beta^2\right) + \gamma_{\text{dB}}} \tag{7.95}$$

Note

Relationship between ρ and γ

$$\rho \,\underline{\Delta}\, \frac{S_R}{\eta B_T} \quad \textit{whereas} \quad \gamma \,\underline{\Delta}\, \frac{S_R}{\eta W}$$

$$\therefore \quad \gamma = \rho\left(\frac{B_T}{W}\right) \text{ but } B_T = 2W(1+\beta)$$

$$\therefore \quad \gamma = 2\rho(1+\beta)$$

$$\therefore \quad 10\log_{10}\gamma = 10\log_{10}2 + 10\log_{10}\rho + 10\log_{10}(1+\beta)$$

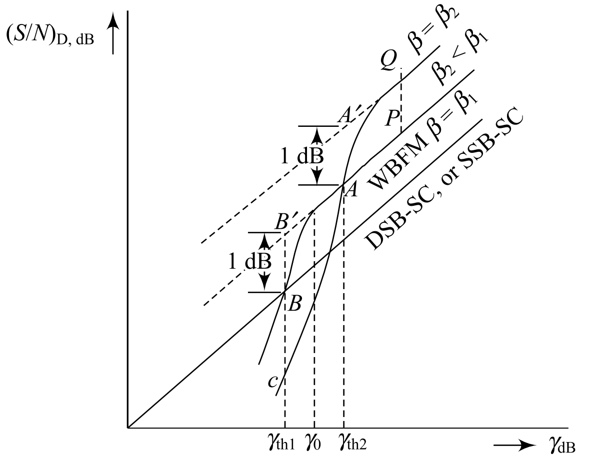


Fig. 7.19 *Plots of output SNR to input SNR for WBFM and DSB-SC or SSB-SC*

For linear modulation schemes DSB-SC and SSB-SC, we had seen that output SNR is equal to γ. The output SNR for baseband transmission. So for these modulation schemes, the output SNR vs. input SNR plots give a straight line passing through the origin as shown in Fig. 7.19. For WBFM too, as per Eq. (7.95), the plot of $(S/N)_{D;\text{dB}}$ vs. γ_{dB} for a given β yields a straight line. So, for large values of γ_{dB} for which Eq. (7.79) is valid, for WBFM also we get straight lines; but they will not pass through the origin. For different values of β like β_1 and β_2, etc., we get parallel lines as shown provided γ_{dB} values are in the range for which Eq. (7.79) and therefore Eq. (7.95) will be valid. As γ_{dB} is reduced, we find that $(S/N)_{D;\,\text{dB}}$ comes down rapidly below a certain value of γ_{dB}, thus exhibiting the phenomenon of *threshold* in WBFM. For larger values of β, we find that the threshold input (SNR) value is also higher (i.e., $\gamma_{th_2} > \gamma_{th_1}$ if $\beta_2 > \beta_1$). *The threshold input SNR for any given value of β, is arbitrarily defined as that input SNR for which the $(S/N)_D$ falls by 1 dB with respect to the straight line portion or its extension.* For $\beta = \beta_2$, as shown in Fig. 7.19, this happens at point *A* on the characteristic, since at *A*, the output SNR has fallen by 1 dB with respect to the value it would have had for $(S/N)_D$ at that input (S/N) corresponding to the point *A*, if it had not deviated from the straight line characteristic. Similarly, it happens at point *B* on the characteristic corresponding to a value of $\beta = \beta_1$. The corresponding input (SNR) values at *A* and *B* are the threshold values for $\beta = \beta_2$ and $\beta = \beta_1$, respectively. One interesting point which should be observed is that if we are operating at point *P* on the straight line portion (i.e., above the threshold) of the $\beta = \beta_1$ curve and if we increase the modulation index β to a higher value β_2, the output SNR increases. On the other hand, if we are operating below the threshold, an increase in β value actually produces a deterioration in the output SNR, as may be seen from points *B* and *C*. Another interesting observation is that γ_{th} depends on β. It is approximately 13 dB for most of the FM receivers, since ρ_{th} is about 10 dB (see the note in the box below Eq. (7.95)).

One may wonder why the output SNR falls steeply when the input SNR is reduced below some value. This leads us to a discussion on the physical phenomenon that causes this.

7.9.1 Causes for Threshold Effect

In an FM receiver, the noise at the output, as heard through the loudspeaker, appears '*soft*' and '*smooth*' when it is operated above the threshold and '*spiky*' and coming out like '*bursts*', when the receiver is operated below the threshold. That is, the nature of the output noise changes as we go below the threshold value of the input SNR.

For a discussion on the mechanism responsible for this change in the nature of the output noise, let us, for the sake of simplicity, assume without loss of generality, that there is no modulation and that only the carrier is being received along with the channel noise. Let the noise entering the receiver be represented as in Eq. (7.54).

$\therefore$
$$n(t) = R_n(t)\cos[\omega_c t + \phi_n(t)] \tag{7.96}$$

where
$$\phi_n = \tan^{-1}\left(\frac{n_q(t)}{n_i(t)}\right) \tag{7.97}$$

∴ the input to the discriminator is (see Fig. 7.5)

$$y(t) = K \cdot A_c \cos \omega_c t + n(t) = A_R \cos \omega_c t + n(t) \tag{7.98}$$

Combining Eqs. (7.96) and (7.98), we may write $y(t)$ as

$$y(t) = R_y(t) \cos [\omega_c t + \xi(t)] \tag{7.99}$$

When the receiver is operated well above the threshold, the input SNR is high and so $R_n << A_R$. Hence, under this condition the phasor diagram will be as shown in Fig. 7.20.

Since $\left(\frac{S}{N}\right)_R >> 1$, $R_n << A_R$ with a high probability and so

$$\xi \approx \tan^{-1}\left(\frac{n_q}{A_R}\right) \tag{7.100}$$

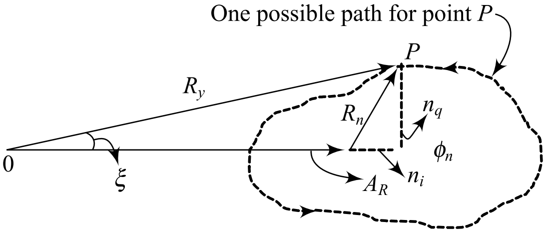


Fig. 7.20 *Phasor diagram when $(S/N)_R >> 1$*

Further, since $\left(\frac{n_q}{A_R}\right) << 1$ for most of the time, we may write

$$\xi \approx \left(\frac{n_q}{A_R}\right) \tag{7.101}$$

Now, $R_n(t)$ and $\xi(t)$ vary randomly with time, with R_n having Rayleigh density and ξ having uniform distribution. Since the $\left(\frac{S}{N}\right)_R << 1$, $R_n << A_R$ for most of the time. Further, because $\phi_n(t)$ is also randomly varying, point P in the phasor diagram moves randomly around the tip of the phasor A_R and may take paths such as the one shown by the dotted line. However, since R_n is quite small compared to A_R, for most of the time, point P while moving along such random paths will be closed to the tip of the phasor A_R. But, of course, occasionally, R_n may take large values, i.e., values larger than A_R and the random path traversed by point P may enclose point 'O' as shown in Fig. 7.19. Whenever such a thing happens, ξ changes by 2π radians. However, we know that the discriminator produces an output proportional to the rate of change of the phase angle of its input signal, $y(t)$. So, when $\xi(t)$ suddenly changes by 2π radians as shown in Fig. 7.22(a) the discriminator output $z(t)$ which is given by

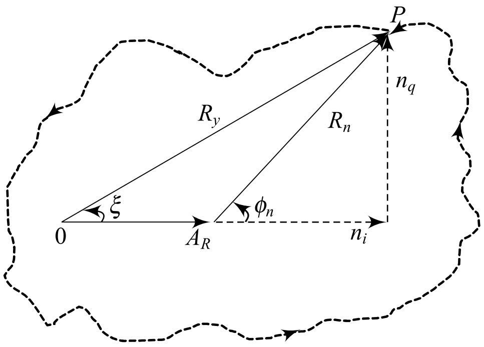


Fig. 7.21 *Phasor diagram when $(S/N)_R << 1$ showing one possible path traversed by P*

$$z(t) = \frac{d}{dt}\xi(t) \tag{7.102}$$

suddenly takes a large value causing a spike in the voltage $z(t)$ and a loud click to be produced by the loudspeaker (refer to Fig. 7.22(b)). However, since $A_R >> R_n$ for most of the time, this phenomenon occurs very rarely. But when the receiver is operated at a low input $(S/N)_R$, the probability of R_n becoming larger than A_R will be high and so, the occurrence of spikes, at the output of the discriminator, will become more frequent. Since a large amount of energy is associated with each spike, the average noise power at the output of the receiver increases considerably with the onset of the occurrence of spikes and so the output signal-to-noise ratio falls rather steeply, causing a 'threshold phenomenon' insofar as the input SNR is concerned, in the case of a FM receiver.

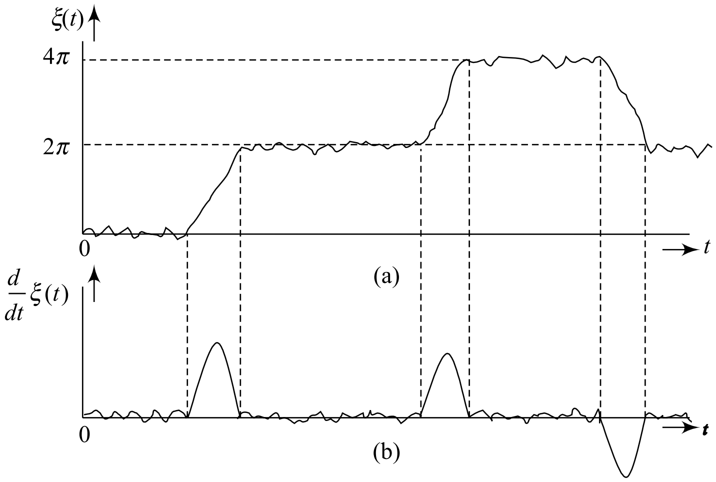


Fig. 7.22 *(a) $\xi(t)$ vs. t, (b) $\dot{\xi}(t)$ vs t showing spikes in the discriminator output, z(t)*

Example 7.17 It is required to transmit, using WBFM, a normalized message signal with $\overline{x^2} = 1$ and W = 15 kHz, over a channel whose bandwidth is 200 kHz. Additive white noise on the channel has $\eta = 10^{-8}$ W/Hz. The destination signal-to-noise ratio should be at least 40 dB. If the signal attenuation during its passage through the channel is 40 dB, find the minimum transmitter power required.

Solution As stated in Section 7.9, the value of β to be used may be restricted either by power considerations, or bandwidth considerations. We shall first examine this.

$$\gamma = \frac{S_R}{\eta W} \text{ and } \rho = \frac{S_R}{\eta B_T} \qquad \therefore \gamma = \rho \cdot \left(\frac{B_T}{W}\right) = \rho \frac{2W(\beta+1)}{W} = 2\rho(\beta+1)$$

$$\therefore \qquad \gamma_{\text{th}} = 2\rho_{\text{th}}(\beta+1)$$

But

$$\rho_{\text{th}} = 10 \text{ dB} = 10 \qquad \therefore \gamma_{\text{th}} = 20(\beta+1)$$

But

$$\left(\frac{S}{N}\right)_D = 3\beta^2\,\overline{x^2}\gamma \qquad \therefore \left(\frac{S}{N}\right)_{D,\text{th}} = 3\beta^2\,\overline{x^2}\gamma_{\text{th}}$$

$$= 3\beta^2\,\overline{x^2}\,20(\beta+1)$$

$$\therefore \qquad \left(\frac{S}{N}\right)_{D,\text{th}} = 60\beta^2(1+\beta)\overline{x^2}$$

$$\therefore \qquad 10^4 = 60\beta^2(1+\beta)\cdot\frac{1}{2} \qquad \therefore \beta^2(1+\beta) = \frac{10^4}{30} = 333$$

$$\therefore \qquad \beta^3 + \beta^2 - 333 = 0 \qquad \therefore \beta \approx 6.6 \text{ from power considerations}$$

Now, if we look at it from bandwidth point of view:

$W = 15$ kHz, $B_T = 2(\beta+1)W = 30 \times 10^3 \times (\beta+1) = 200$ kHz

$\therefore\ \beta + 1 = 6.6$ and $\therefore \beta = 5.6$ from bandwidth point of view.

We find that the maximum value of β is restricted by the channel bandwidth and not by power. We shall therefore choose

$$\beta = 5.6$$

With this β, and an $(S/N)_D$ of 10^4, the value of γ is

$$\gamma = \frac{10^4}{3 \times (5.6)^2 \times \frac{1}{2}} = \frac{10^4}{47} = 212.76$$

But $$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{10^{-8} \times 10^3 \times 15} = \frac{S_R}{15 \times 10^{-5}} = 212.76$$

$\therefore$ $$S_R = 3191.5 \times 10^{-5} = 0.031915 \quad \text{But} \quad \left(\frac{S_T}{S_R}\right) = 10^4$$

$\therefore$ $$S_T = 3191.5 \times 10^{-5} \times 10^4 = 3191.5 \times 10^{-1}$$
$$= 317.15 \text{ W}$$

Example 7.18 A message signal normalized so that $|x(t)| \leq 1$ and having an average power of 1 W and a bandwidth of 15 kHz, is to be transmitted using WBFM with $\beta = 5$, over a channel with additive noise of two-sided PSD = $\eta/2 = 0.5 \times 10^{-13}\ W/\text{Hz}$ and a total transmission loss of 100 dB. If a destination SNR of 40 dB is required, what should be the average transmitted power? Check whether the system is above the threshold.

Solution $$\left(\frac{S}{N}\right)_D = 3\beta^2 \overline{x^2} \gamma = 3 \times 25 \times 1 \times \frac{S_R}{15 \times 10^3 \times 10^{-13}} = 10^4$$

$\therefore$ $$S_R = 2 \times 10^{-7}\ W$$

But $$\left(\frac{S_T}{S_R}\right) = 10^{10} \quad \therefore S_T = 2 \times 10^{-7} \times 10^{10} = 2 \text{ kW}$$

$\therefore$ $$\frac{S_R}{\eta W} = \frac{10^{-7} \times 2}{10^{-13} \times 15 \times 10^3} = \frac{400}{3} = 21.3 \text{ dB} = \gamma$$

$$\left(\frac{S}{N}\right)_{D,\text{th}} = 60\beta^2(\beta + 1)\overline{x^2} = 60 \times 25(5+1) \times 1 = 60 \times 150 = 9000$$
$$= 3 \times 25 \times \gamma_{\text{th}} \times 1$$

$\therefore$ $$\gamma_{\text{th}} = \frac{9000}{3 \times 25} = \frac{3000}{25} = 120 \quad \text{and} \quad 10 \log_{10} 120 = 20.8 \text{ dB}$$

$\therefore$ the system is operating above the threshold since $\gamma > \gamma_{\text{th}}$.

7.9.2 Threshold Extension

As we have stated earlier, for most WBFM receivers, γ_{th} is about 13 dB. This corresponds to a value of 10 dB for ρ_{th}, the actual input SNR to the discriminator. So, for satisfactory operation of the receiver, we have to ensure that the input SNR is always kept above 10 dB. While this may not be a problem in the case of FM broadcast systems, in the case of wideband satellite communications and space communications, such a large value of threshold does pose problems. The reason for this is easy to see. Since

$$\rho = \frac{S_R}{\eta B_T}$$

if we desire to operate above ρ_{th}, we have to either increase the transmitter power, or decrease the transmission bandwidth. But both these options are not feasible in the case of satellite to earth or space communications, where power is at a premium and wide bandwidth is a must.

This underscores the need to have some methods for reduction of the threshold ρ_{th} below the 10-dB value. These methods are called 'threshold extension techniques', and they permit the receiver to operate satisfactorily even when the input SNR is very low.

7.9.3 Threshold Extension Techniques

Basically there are two threshold extension techniques that are available. These are as follows:

1. Frequency Modulation Feedback (FMFB) technique
2. Phase Lock Loop (PLL) technique

Actually these two techniques work on similar lines and are equally effective in lowering the threshold. However, the PLL method is simpler and is therefore generally preferred. In practice, they reduce the threshold ρ_{th} by about 5 to 7 dB, i.e., when either of these techniques is used, the value of the threshold ρ_{th} effectively has a value of 3 to 5 dB.

FMFB technique As we have already discussed, the onset of threshold conditions occurs when the input signal-to-noise ratio (S_R/N_R), of the discriminator falls below some critical value. We know that

$$S_R = \frac{A_R^2}{2}$$

and $$N_R = \eta B$$

where B = Bandwidth of the noise at the input of the discriminator.

In a normal FM broadcast superheterodyne receiver, B is equal to the bandwidth of the incoming FM signal, which is being denoted by us as B_T, the transmission bandwidth, and the IF amplifier bandwidth is designed to be equal to B_T. So, in a normal FM broadcast receiver, by not employing any threshold extension techniques, the bandwidth of the noise at the input to the receiver's discriminator is B_T.

Hence, by keeping S_R, the input signal power the same, if we can reduce N_R by reducing the noise bandwidth below B_T, we can improve the SNR at the input to the discriminator and thus achieve threshold extension. Basically, this is precisely what the FMFB technique for threshold extension tries to do.

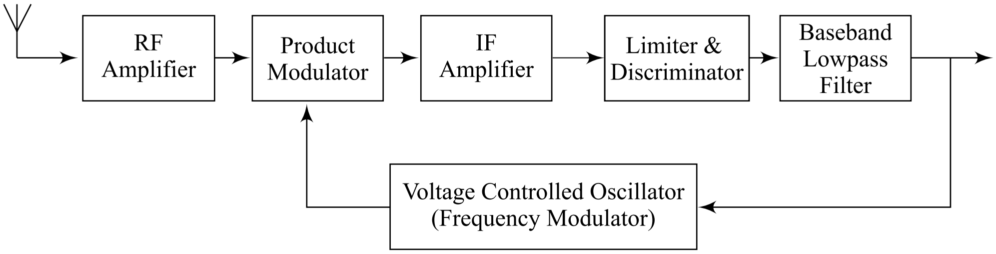


Fig. 7.23 *FMFB method for threshold extension*

Referring to Fig. 7.23, the normal local oscillator of the receiver is replaced by a Voltage Controlled Oscillator (VCO), which may as well be considered as a frequency modulator. The VCO is adjusted, in the absence of the control voltage, to oscillate at a frequency f_0 which is f_{if} hertz below the carrier frequency, f_c, to which the receiver is tuned. The control voltage applied to it is the output audio signal of the receiver, which is an approximation to the message signal $x(t)$ of the FM signal being received. The output of the VCO is thus a frequency modulated signal with a carrier frequency f_0 and $x(t)$ as the modulating signal. The product modulator multiplies the incoming FM signal having a carrier frequency f_c with the output of the VCO. In the output of the product modulator, only the difference frequency component is passed on to the IF amplifier. The input to the IF amplifier is therefore an FM signal with $(f_c - f_0) = f_{if}$ as the carrier frequency and $x(t)$ as the modulating signal. However, its peak deviation will be less than that of the incoming FM signal, since it is the difference frequency component coming out of the product modulator. Because of the smaller deviation, its bandwidth will be less than B_T, the transmission bandwidth of the incoming FM signal. Since the IF stage bandwidth is much less than B_T, say B, the noise power at the input to the discriminator is only ηB instead of ηB_T. Thus, the input (S/N) ratio for the discriminator is increased and consequently the onset of threshold is made to occur at a much smaller value of $(S/N)_R$ than the value at which it would have occurred in the absence of the feedback.

PLL technique Earlier, in Section 5.8, we had seen how a PLL could be used as an FM demodulator.

For convenience, Fig. 5.36 showing the linearized equivalent circuit of the PLL has been reproduced here as Fig. 7.24. This circuit was analyzed in Section 5.8. It was shown there that

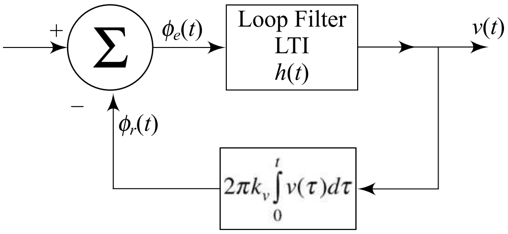


Fig. 7.24 *Linearized equivalent circuit of the PLL*

$$\Phi_e(f) = \frac{\Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)} \tag{7.103}$$

and that

$$V(f) = \frac{H(f) \cdot \Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)} \tag{7.104}$$

As pointed out there, if the gain of the loop filter is high enough, so that

$$\left(\frac{k_v}{jf}\right)H(f) >> 1 \quad \text{for} \quad |f| < W$$

then $v(t)$, the output of the PLL is given by

$$v(t) = \left(\frac{k_f}{k_v}\right)x(t)$$

where $x(t)$ is the modulating signal of the incoming FM wave. Hence, for good tracking, i.e., for $\phi_c(f)$ to be very small, the loop filter's gain must be adjusted to be high.

Now, since $v(t)$ is proportional to the modulating signal, $x(t)$, if $x(t)$ is band limited to W Hz, i.e., $X(f) = 0$ for $|f| \geq W$, then $V(f)$ will also be zero for $|f| \geq W$. Since $v(t)$ is the output of the loop filter, and since it is band limited to W Hz, we need to provide a bandwidth of only W Hz to the loop filter. As in the case of the FMFB, this implies that the threshold is lowered and that the receiver can operate satisfactorily with even smaller values of input SNRs.

Generally a second-order filter of the 'proportional plus integral type', shown in Fig. 7.25, is used as the loop filter.

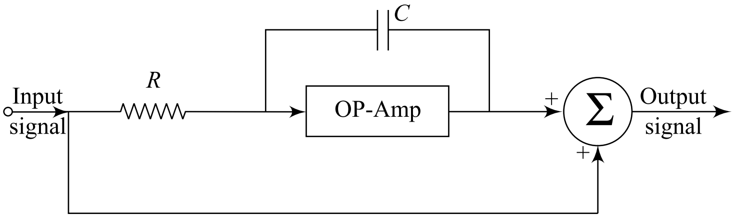


Fig. 7.25 *Loop filter for a second-order PLL*

7.9.4 Comparison between AM and FM

Table 7.1

S. No.	Amplitude Modulation	Frequency Modulation
1.	It is the amplitude parameter of the carrier which is varied.	It is the frequency parameter of the carrier which is varied.
2.	Average power of the modulated signal changes with the depth of modulation.	Average power of the modulated signal does not change with modulation index.
3.	Depth of modulation depends only on the amplitude of the modulating signal.	Modulation index, β, depends both on the amplitude as well as the frequency of the modulating signal.
4.	For a single-tone modulating signal, the modulated signal has only two side frequencies besides the carrier, $B_T = 2f_m$.	Even for a single-tone modulating signal, the modulated signal theoretically contains an infinite number of side-frequencies besides the carrier. Theoretically, B_T is infinite.
5.	The carrier component in the modulated signal has fixed amplitude and it does not change with the modulation index.	The carrier component in the modulated signal varies with the modulation index and it becomes zero for some values of the modulation index.
6.	Bandwidth is constant and equal to $2W$ irrespective of the depth of modulation. $B_T = 10$ kHz for commercial AM broadcasting.	Effective bandwidth changes with modulation index β. $B_T = 2W(\beta + 1) \approx 180$ kHz for commercial FM broadcasting.
7.	Bandwidth increases in direct proportion to the frequency of the modulating signal.	Effective bandwidth increases only slightly with the frequency of the modulating signal.
8.	The maximum audio frequency handled by an AM broadcast transmitter is generally limited to 5 kHZ.	The maximum audio frequency handled by an FM broadcast transmitter is generally 15 kHZ.
9.	Additive noise on the channel directly affects an amplitude modulated signal.	Additive noise on the channel can affect the FM signal only indirectly by producing a change in its phase. Thus, compared to AM, FM enjoys some immunity against channel noise.
10.	AM systems do not permit any trade-off between transmission bandwidth and the average transmitted power.	Trade-off is possible between transmission bandwidth and the average transmitted power.
11.	When the channel includes devices like TWT amplifier which generally has a non-linear input-output relation, an AM signal gets terribly distorted (*see Section 5.7*).	Input-output non-linearity of the channel does not cause any distortion. It only changes the amplitude of the FM signal.
12.	Even weak interfering signals close to the frequency of desired signal can cause some interference.	Interfering signals which are weak compared to the desired signal do not cause interference due to capture effect.

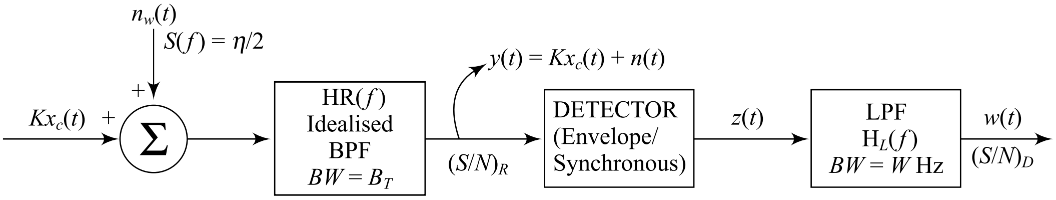
Summary

- For a baseband transmission system, i.e., when the baseband or message signal is transmitted without any modulation,

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} \Delta \gamma \text{, where } S_R = \text{Received signal power}$$

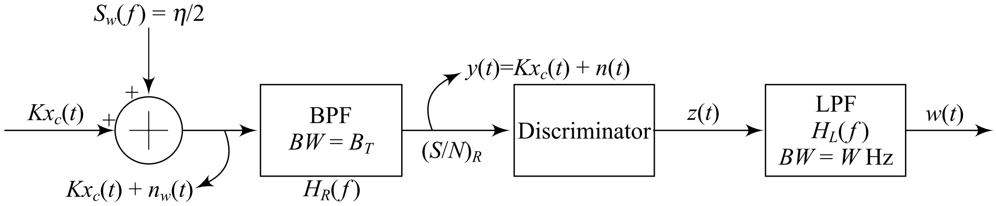
$\eta/2$ = PSD of white noise on the channel
W = Bandwidth of baseband signal

- Model used for linear modulation systems:



- Pre-detection $SNR = \left(\frac{S}{N}\right)_R = \gamma\left(\frac{W}{B_T}\right)$, where B_T is the bandwidth of the transmitted signal.
- (a) $\underset{\text{SSB-SC}}{\left(\frac{S}{N}\right)_D} = \gamma = \underset{\text{DSB-SC}}{\left(\frac{S}{N}\right)_D}$

  (b) $\underset{\text{AM}}{\left(\frac{S}{N}\right)_D} = \left[\frac{m^2\overline{x^2(t)}}{1+m^2\overline{x^2(t)}}\right]\gamma; \underset{\substack{\text{AM}, m=1\\ \text{tone mod}}}{\left(\frac{S}{N}\right)_D} = \left(\frac{\gamma}{3}\right)$, i.e., 5 dB less than $\left(\frac{S}{N}\right)_D$ of SSB-SC and DSB-SC.
- There is a threshold effect for AM with an envelope detector. That is, when the SNR at the input to the envelope detector is small compared to unity, the message signal and noise become intermingled at the output of the detector.
- *Model used for FM systems:*



- The PSD of the noise at the output of the discriminator in the FM receiver varies as the square of the frequency.
- $\underset{\text{FM}}{\left(\frac{S}{N}\right)_D} = 3D^2\overline{x^2(t)}\gamma$ provided $\left(\frac{S}{N}\right)_R >> 1.$
- There is a power-bandwidth trade-off possible in WBFM as shown by $\left(\frac{S}{N}\right)_D$ for FM.
- (a) Pre-emphasis consists of boosting up the higher message frequencies *before modulation* at the FM transmitter.

  (b) De-emphasis consists of de-emphasizing the higher message frequencies back to their original level *after the discriminator stage in an FM receiver*.

  (c) Pre-emphasis, de-emphasis technique is used to improve the destination SNR in an FM system.
- There is a threshold effect in FM reception, i.e., if the input SNR for an FM receiver falls below a certain threshold value, the output of the receiver will be only noise.
- For WBFM receivers, the threshold value of the input SNR is approximately 10 dB.
- Threshold extension technique like FMFB and PLL methods reduce the threshold input SNR *to* about 3 to 5 Db, i.e., they reduce the threshold by 5 to 7 dB.
- Figure of Merit (FOM) of a communication system is defined as

$$\text{FOM} \underline{\Delta} \frac{\text{(SNR at the destination)}}{\text{(SNR at the input to the detector but with the noise considered only over message bandwidth)}}$$

$$= \frac{(S/N)_D}{(S_R/\eta W)} = \frac{1}{\gamma}\left[\left(\frac{S}{N}\right)_D\right] \quad \text{since} \quad \gamma = \left(\frac{S_R}{\eta W}\right)$$

- FOM Values: SSB-SC : 1; DSB-SC : 1, AM : $\left[\dfrac{m^2\,\overline{x^2(t)}}{1+m^2\,\overline{x^2(t)}}\right]$
  WBFM: $3D^2\,\overline{x^2(t)}$ and $\underset{\text{(tone mod)}}{WBFM}:\dfrac{3}{2}\beta^2$

References and Suggested Reading

1. Proakis, John G., and Masoud Salehi, *Communication Systems Engineering*, 2nd Edition (Eastern Economy Edition), Prentice-Hall, New Delhi, 2006.
2. Haykin, Simon, *Communication Systems*, 3rd Edition, John Wiley & Sons, 1995.
3. Carlson, A. Bruce, P. B. Crilly, and Janet C. Rutledge, *Communication Systems*, International (4th Edition) Edition, McGraw-Hill, 2002.
4. Blake, Roy, *Electronic Communication Systems*, 2nd Edition, Thomson-Delmar, 2002.
5. Couch, Leon W. II, *Digital and Analog Communication Systems*, 6th Edition, Pearson Education (Asia), 2001.
6. Taub, Herbert, and Donald L. Schilling, *Principles of Communication Systems*, International Student Edition, McGraw-Hill Kogakusha,1971.

Review Questions

1. Draw the block diagram of the model used for the channel and the receiver to study the noise performance of various modulation systems.
2. Derive an expression for the destination SNR of a baseband system. How is the receiver modeled for this case?
3. What is the model used for a synchronous detector?
4. Derive an expression for the destination SNR in the case of an AM system employing synchronous detection.
5. What is the model used for an envelope detector?
6. Derive an expression for the destination SNR of a DSB-SC system in terms of that of a baseband system.
7. Show that an SSB-SC system gives the same destination SNR as a baseband system.
8. Critically compare the noise performance of AM, DSB-SC, and SSB-SC systems.
9. Discuss the effect of channel noise on the phase angle and frequency of an FM signal.
10. Derive an expression for the PSD of noise at the output of the discriminator of an FM receiver.
11. Explain the meaning of the following statement:
    ‘FM systems permit a trade-off between bandwidth and power’.
12. Explain the need for pre-emphasis and de-emphasis in the case of FM systems. How is it implemented?
13. Draw the circuit diagram of the filters used for pre-emphasis and de-emphasis. Write down the expressions for their transfer functions and sketch their frequency response.
14. Derive an expression for the improvement in the destination SNR obtained by the use of pre-emphasis and de-emphasis in an FM system.
15. What is meant by the ‘threshold effect’ in FM receivers?
16. Clearly explain the physical processes that lead to the occurrence of threshold in a FM receiver.
17. Clearly explain the basic principle of extension of threshold using the FMFB technique.
18. How can a PLL be used for threshold extension?

Problems

1. An AM transmitter is used to send a message signal with $\overline{x^2}=0.5$ and a bandwidth of 5 MHz over a channel which introduces additive white noise with a power spectral density of 10^{-12} *W*/Hz. The modulation index is equal to 1. If the channel introduces a loss of 100 dB, and if the average transmitted power is 200 *W*, find the destination signal-to-noise ratio that can be obtained.
2. Determine the post-detection SNR to pre-detection SNR ratio for the following types of communication systems:
   (a) AM with a modulation index of *m* (b) SSB-SC
   (c) DSB-SC (d) FM with modulation index β_f

3. A DSB-SC signal is transmitted over a channel with additive white noise of two-sided PSD of $(\eta/2) = 0.5 \times 10^{-12}\ W/\text{Hz}$. If the received signal power is $S_R = 20 \times 10^{-9}$ W and the message bandwidth $W = 5 \times 10^6$ Hz, find the destination SNR.

4. It is proposed to transmit a message signal whose amplitude is uniformly distributed over [–1, 1] and whose bandwidth is 1.5 MHz over a channel with an additive white noise two-sided PSD of 0.5×10^{-13} *W*/Hz and introducing a loss of 80 dB between the transmitter and receiver. If destination SNR of 40 dB is desired, for each of the following cases, determine the transmitter power that will be required:

- (a) SSB-SC modulation
- (b) AM with a modulation index of $m = 0.6$
- (c) DSB-SC modulation

5. A message signal, band limited to 2 kHz, is uniformly distributed in the interval [–1, 1]. It is used for amplitude modulating (AM) a sinusoidal carrier of peak amplitude 5 V and frequency f_c Hz, the modulation index being 0.5. The modulated signal is transmitted over a channel with additive white noise of PSD (two sided) 0.5×10^{-12} *W*/Hz and the channel introduces an attenuation of 80 dB. The received signal is first filtered using a BPF centered on f_c and having a transfer function *H*(*f*) as shown in Fig. P7.5.

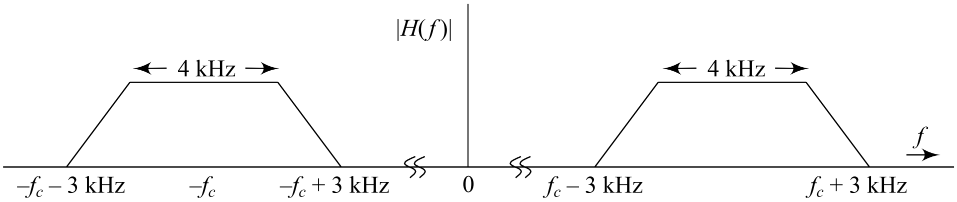


Fig. P7.5 *Transfer function of the front-end BPF*

Then it is demodulated using a synchronous detector consisting of a product device (to which the locally generated carrier and the filtered received signal are applied) followed by an ideal LPF with a cut-off frequency of 2 kHz. Determine the pre-detection and destination SNRs.

6. A transmitter is producing an average transmitted power of 20 kW. The channel with an additive white noise of PSD (two sided) 0.5×10^{-10} *W*/Hz introduces an attenuation of 70 dB. The message signal has a bandwidth of 10 kHz and a normalized average power of 0.2 W.

- (a) Find the pre-detection SNR
- (b) Find the destination SNR if
  - (i) the modulation is AM with a modulation index of $m = 0.8$
  - (ii) the modulation is DSB-SC
  - (iii) the modulation is SSB-SC

7. While deriving the destination SNR for a WBFM system, we had assumed that the baseband filter in the receiver is an ideal LPF with a cut-off frequency of *W* Hz. Derive the expression for the $(S/N)_D$ assuming that the baseband filter is a Butterworth filter of order *n* with a 3-dB cut-off frequency of *W* Hz.

8. A message signal, *x*(*t*), normalized so that $|x(t)| \leq 1$, has a bandwidth of 4 kHz and an average power of 0.2 W. It is used for modulating a carrier and the modulated signal is transmitted over a channel of bandwidth 100 kHz. Find the ratio of the destination SNRs obtained for the following two cases:

- (a) The message frequency modulates the carrier and the modulated signal fully utilizes the full bandwidth of the channel.
- (b) The message amplitude modulates (AM) the carrier to a depth of 0.5.

9. A message signal, *x*(*t*), with a bandwidth of 500 kHz and an average power of 0.33 W, frequency modulates a carrier having a peak amplitude of 22.36 V, producing a peak frequency deviation of 2 MHz. This modulated signal is transmitted over a channel with additive white noise of two-sided PSD equal to 0.5×10^{-15} *W*/Hz and a transmission loss of 80 dB. If the receiver uses a post-detection de-emphasis filter having a transfer function,

$$H_{de}(f) = \frac{1}{\sqrt{1 + (f/B)^2}} \quad \text{where } B = 5 \text{ kHz}$$

followed by an ideal LPF of 500 kHz cut-off frequency, determine the destination SNR.

10. A message signal with a bandwidth of 6 kHz and an average power of 0.5 W, is transmitted using FM, over a channel characterized by a bandwidth of 60 kHz and additive white noise of two-sided PSD equal to 10^{-11} *W*/Hz.

(a) If a destination SNR of 60 dB is desired without pre-emphasis and de-emphasis, what should be the transmitted power?

(b) If a destination (SNR) of 60 dB is desired using pre-emphasis and de-emphasis filters of time constant 75 micro-sec., what should be the transmitted power?

11. A communication system makes use of a message signal with an average power of 0.5 W and a bandwidth $W =$ 10 kHz. The modulated signal is transmitted over a channel with additive white noise having a two-sided PSD of 0.5×10^{-14} *W*/Hz, and a transmission loss of 80 dB. A destination SNR of 40 dB is needed. Determine the transmitter power required if

(a) AM with $m = 0.5$ is used
(b) SSB-SC is used
(c) WBFM with $D = 5$ is used (No pre-emphasis and de-emphasis)
(d) WBFM with $D = 5$ is used and pre-emphasis and de-emphasis filters of 75 µs time constant are used.

12. An FM receiver employing FMFB for threshold extension is shown in Fig. P7.12.

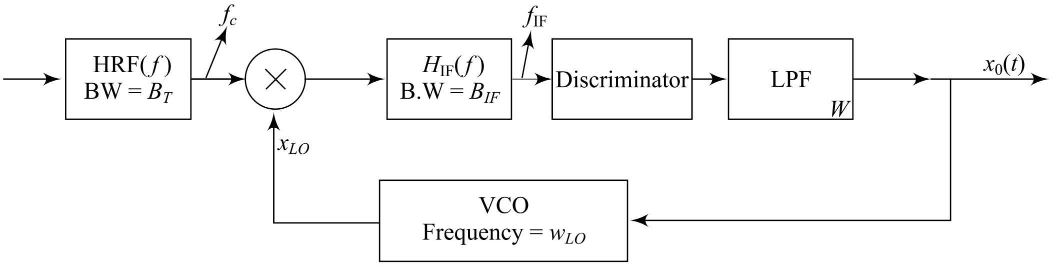


Fig. P7.12

Assume that the received FM signal is noise-free and that it has a carrier frequency of f_c. The VCO produces a signal given by

$$x_{LO} = 2\cos\left[(\omega_c - \omega_{IF})t + K2\pi\int_0^t x_0(\tau)d\tau\right]$$

Show that the deviation ratio of the FM signal with f_{if} as the carrier, is $D/(1 + K)$, where D is the deviation ratio of the received FM signal with carrier frequency f_c. How is B_{IF} related to B_T?

Multiple-Choice Questions

1. The channel noise has a two-sided PSD of η/2 *W*/Hz and the incoming FM signal has a bandwidth of B_T Hz. The peak amplitude of the FM signal at the input to the discriminator is A_R volts. The pre-detection SNR is

(a) $\dfrac{A_R^2}{\eta B_T}$ (b) $\dfrac{1}{2}\left(\dfrac{A_R^2}{\eta B_T}\right)$ (c) $\dfrac{2A_R^2}{\eta B_T}$ (d) $\dfrac{2\eta B_T}{A_R^2}$

2. In the receiver model used for discussing the noise performance of different modulation schemes, the pre-detection and post-detection stages of the receiver are modeled respectively as

(a) band pass filter and low pass filter (b) high pass filter and low pass filter
(c) low pass filter and low pass filter (d) band pass filter and high pass filter

3. If γ denotes the destination SNR for a baseband transmission system, that of a DSB-SC system with carrier peak amplitude of A_R is given by

(a) $\gamma/2$ (b) γ (c) 2γ (d) $\gamma/4$

4. For AM, the destination SNR is given by

(a) $\left(\dfrac{m^2\overline{x^2}}{1+m^2\overline{x^2}}\right)\gamma$ (b) $\left(\dfrac{m\overline{x^2}}{1+m\overline{x^2}}\right)\gamma$ (c) $\left(\dfrac{m^2\overline{x}}{1+m^2\overline{x}}\right)\gamma$ (d) $\left(\dfrac{m\overline{x}}{1+m\overline{x}}\right)\gamma$

5. In an AM system transmitting a single tone message at 100% modulation, the destination SNR is given by

(a) γ (b) (1/2) γ (c) (1/3) γ (d) 2γ

6. At the output of the discriminator in an FM receiver, the PSD of the noise

(a) increases linearly with frequency (b) decreases as the square of the frequency
(c) increases as the square of the frequency (d) decreases linearly with frequency

7. The average signal power at the input to the detector in the case of an AM system is given by

(a) $\left(1+m\overline{x^2}\right)$ (b) $\frac{A_R^2}{2}\left[1+m\overline{x^2}\right]$ (c) $\left(1+m^2\overline{x^2}\right)$ (d) $\frac{A_R^2}{2}\left(1+m^2\overline{x^2}\right)$

8. For the same average power transmitted and with tone modulation,

(a) $(S/N)_D$ will be the same for AM and SSB-SC
(b) $(S/N)_D$ for AM is greater than $(S/N)_D$ for SSB-SC by 5 dB
(c) $(S/N)_D$ for AM is less than $(S/N)_D$ for SSB-SC by 5 dB
(d) $(S/N)_D$ for AM is less than $(S/N)_D$ for SSB-SC by 10 dB

9. Pre-emphasis is

(a) boosting up of the high frequency components of the message signal after detection in the receiver
(b) boosting up of the high frequency components of the message signal at the transmitter before modulation
(c) boosting up of the low frequency components of the message signal after detection in the receiver
(d) boosting up of the low frequency components of the message signal at the transmitter before modulation

10. In standard FM broadcasting systems, the time constants of the pre-emphasis and de-emphasis filters are respectively

(a) 75 μs and 100 μs (b) 75 μs and 75 μs (c) 100 μs and 75 μs (d) 100 μs and 100 μs

11. For standard FM broadcast receivers, the threshold input SNR, i.e., ρ_{th} is approximately

(a) 10 dB (b) 13 dB (c) 5 dB (d) 7 dB

12. Use of some type of threshold extension technique is absolutely necessary in the case of

(a) FM broadcast receivers
(b) wideband FM communication from the earth station to a satellite
(c) wideband FM communication from a satellite to the earth station
(d) None of the above

Key to Multiple-Choice Questions

1. (b)	2. (a)	3. (b)	4. (a)	5. (c)	6. (c)	7. (d)	8. (c)
9. (b)	10. (b)	11. (a)	12. (c)				

SAMPLING AND ANALOG PULSE MODULATION

8

> *"Everyone thinks of changing the world, but no one thinks of changing himself."*
>
> **Leo Nikolaevich Tolstoy (1828–1910)**
> *Russian writer*

Learning Objectives

After going through this chapter, students will be able to

- *clearly understand the meaning of terms like: 'Band-limited Signals', Nyquist rate, Aliasing, etc., derive the low pass sampling theorem and explain its implication,*
- *appreciate the usefulness and limitations of various methods of sampling, and explain the way a band-limited low pass signal may be reconstructed from its samples,*
- *explain the basic concept of time-division-multiplexing,*
- *understand the way the amplitude of each sample of a continuous-time band-limited signal, is represented in PAM, PDM and PPM,*
- *understand that bandwidth deficiency of the channel causes cross-talk in PAM, PDM and PPM signals, and*
- *mathematically analyze the noise performance of PAM, PDM and PPM systems and compare their noise performance.*

8.1 INTRODUCTION

In this chapter, we will be first discussing the low pass sampling theorem. In essence, this theorem tells us that a low pass signal $x(t)$, band limited to W Hz, i.e., one which does not have any frequency components at or above W Hz, can be completely recovered *for all time* from its samples taken at regular intervals T_s, provided $T_s \le \frac{1}{2W}$ sec. As we are going to see, the process of reconstructing, or recovering, $x(t)$ from its samples, is extremely simple. All that we need to do is to pass the samples through a low pass filter having an appropriate cut-off frequency.

In all the continuous-wave modulation techniques – AM, FM or PM, information about the message signal is transmitted continuously in terms of corresponding variations of the amplitude, frequency, or the phase of the carrier wave as the case may be. In this context, what the low pass sampling theorem states, has tremendous practical implication. It makes it clear that if a message is band limited, it is not necessary to transmit it continuously; it is enough if we transmit its samples, since the receiver can reconstruct the message from these samples.

There are different methods that one can adopt for representing the sample values and transmit them to the receiver. These different methods of representing the sample values give rise to the different pulse modulation schemes. Some of them like pulse amplitude modulation (PAM), pulse duration modulation (PDM or PWM) and pulse position modulation (PPM) are analog pulse modulation techniques, while pulse code modulation (PCM), etc., are digital pulse modulation techniques. We will, of course, confine our discussion to only analog pulse modulation systems in this chapter.

Information about the sample value at a sampling instant is carried by the amplitude of a pulse occurring at that instant in the case of PAM, by the width of the pulse occurring at that sampling instant in the case of PDM and by the shift in its position with respect the sampling instant, in the case of PPM. Since the width of the pulse, τ, is very small compared to the interval between consecutive pulses, the average power in a pulse modulated signal is very low compared to that in a continuous wave modulation system. Of course, in the course of this chapter, we will be discussing this and other advantages and disadvantages of pulse communication systems, but for the present, we will simply list them as follows.

1. The average transmitted power is very low. This is especially useful when the energy to be radiated is obtained from devices like magnetron or a laser, which can give large pulsed powers but only a very small average power.
2. It is possible to have Time Division Multiplexing (TDM) for transmission of several message signals simultaneously over the same physical channel by making the pulses pertaining to different message signals to share the available time T_s between two consecutive samples of the same message signal.
3. Pulse modulation has the disadvantage of requiring large transmission bandwidths.

Since the pulses contain considerable dc content and low frequency components in addition to the high frequency components, they cannot be radiated directly. So, when transmission over long distances is desired, these pulses must be made to modulate a high frequency carrier. For short distances, however, they can be transmitted over a cable, or a pair of wires.

Pulse modulation systems are mostly used for time division multiplexing of several message signals as in the case of data telemetry and in instrumentation systems.

8.2 SAMPLING OF BAND LIMITED LOW PASS SIGNALS

If $x(t)$ is an analog signal, the process of sampling it should result in the set of samples, $\{x(nT)\}$, where T is the sampling interval and $x(nT)$ is the value of $x(t)$ at $t = nT$, the nth sampling instant.

An easy way of visualizing the sampling process, and perhaps a simple way of implementing it may be through a switch, as shown in Fig. 8.1. Although a mechanical switch is shown in Fig. 8.1, in actual practice, an electronic switch, making use of a diode bridge clamper, a diode bridge linear gate or a shunt transistor gate, may be used.

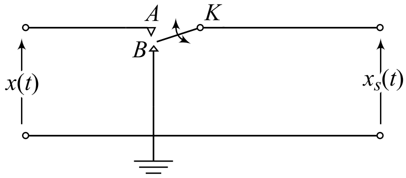


Fig. 8.1 *A switch used for sampling*

Let the switch make contact with A once every T sec. Then $x_s(t)$ consists of samples of $x(t)$ taken every T sec, provided the switch makes contact with A instantaneously. However, in practice, the contact will be made for a finite amount of time, say, τ sec.

Then the sampled version is as shown in Fig. 8.2. This consists of strips of $x(t)$ of width τ occurring at regular intervals of T sec; and may be visualized as the waveform that results when $x(t)$ is multiplied by a 'sampling function' shown in Fig. 8.3.

This sampling function may be expanded using Fourier series, as it is a periodic function with period T.

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi f_0 nt};\ T = \text{sampling period} = \frac{1}{f_s} \tag{8.1}$$

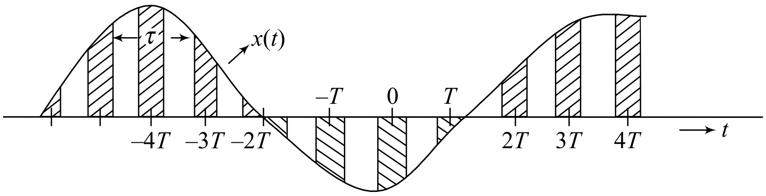


Fig. 8.2 *Sampling waveform $x_s(t)$*

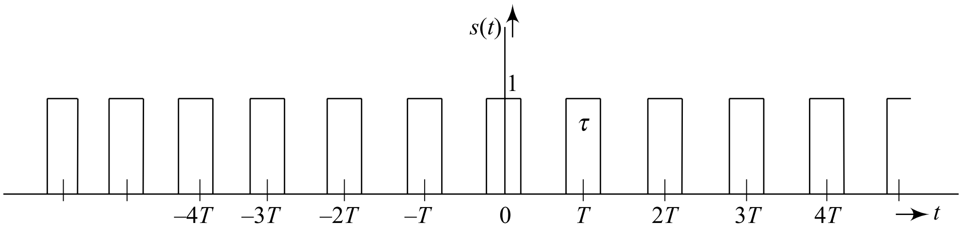


Fig. 8.3 *Sampling function*

where

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} s(t)\, e^{-j2\pi n f_0 t}\, dt \tag{8.2}$$

However, over the interval $-T/2$ to $T/2$,

$$s(t) = \Pi(t/\tau), \text{ since } \tau << T$$

Hence, we may write Eq. (8.2) as

$$c_n = f_s \int_{-\tau/2}^{\tau/2} \Pi(t/\tau) e^{-j2\pi n f_0 t}\, dt \tag{8.3}$$

As $\Pi(t/\tau) = 0$ outside the limits of integration, we may write

$$c_n = f_s \int_{-\infty}^{\infty} \Pi(t/\tau) e^{-j2\pi (nf_s) t}\, dt = f_s[\mathcal{F}\{\Pi(t/\tau)\}]\Big|_{f=nf_s}$$

since $\mathcal{F}[\Pi(t/\tau)] = \tau \operatorname{sinc} f\tau$

$$c_n = f_s \tau \operatorname{sinc} n f_s \tau \tag{8.4}$$

$$\therefore \quad s(t) = \sum_{n=-\infty}^{\infty} f_s \tau \operatorname{sinc} n f_s \tau\, e^{+j2\pi n f_s t} = \left[2 f_s \tau \sum_{n=1}^{\infty} \operatorname{sinc}(n f_s \tau) \cos 2\pi n f_s t\right] + f_s \tau \tag{8.5}$$

Since $f_s = \frac{1}{T}$ and $\frac{\tau}{T}$ = duty ratio of the sample function = α,

$$s(t) = \alpha + 2\alpha \operatorname{sinc}(\alpha)\cos 2\pi f_s t + 2\alpha \operatorname{sinc}(2\alpha)\cos 2\pi 2 f_s t + 2\alpha \operatorname{sinc}(3\alpha)\cos 2\pi 3 f_s t + \ldots \tag{8.6}$$

$$\therefore \quad x_s(t) = x(t)\cdot s(t) = \alpha x(t) + 2\alpha[x(t) \operatorname{sinc} \alpha \cdot \cos \omega_s t + x(t) \operatorname{sinc}(2\alpha)\cdot \cos \omega_s t + \ldots] \tag{8.7}$$

Taking Fourier transform on both sides, we get

$$X_s(f) = \alpha X(f) + c_1[X(f - f_s) + X(f + f_s)] + c_2[X(f - 2f_s) + X(f + 2f_s)] + \ldots \tag{8.8}$$

where $c_k = \alpha \operatorname{sinc} k\alpha$ (8.9)

If the signal $x(t)$ has a spectrum as shown in Fig. 8.4, the spectrum of $X_s(f)$, the sampled version of $x(t)$, will be as shown in Fig. 8.5.

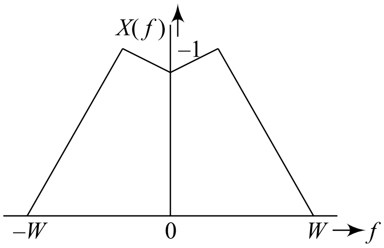


Fig. 8.4 *Spectrum of x(t)*

This figure showing $X_s(f)$ has been drawn assuming that $(f_s - W) > W$, or $f_s > 2W$. It is interesting to note from this figure that the spectrum of $x(t)$, viz., $X(f)$ appears in it without any distortion. It is only scaled by the factor α, the duty cycle of the sampling function. If we can, by some means, separate out this part of the spectrum from $X_s(f)$, say, by using a low pass filter with a cut-off frequency of B Hz, where B is such that $W < B < (f_s - W)$ and whose gain is constant at least up to W Hz, then, in time domain, it means that we are able to get back our $x(t)$ without any distortion, from its sampled version. If $f_s = 2W$, then $W = (f_s - W)$ and so there will not be any guard band. So, to recover $x(t)$ from $x_s(t)$, one has to use an ideal LPF with a cut-off frequency equal to W.

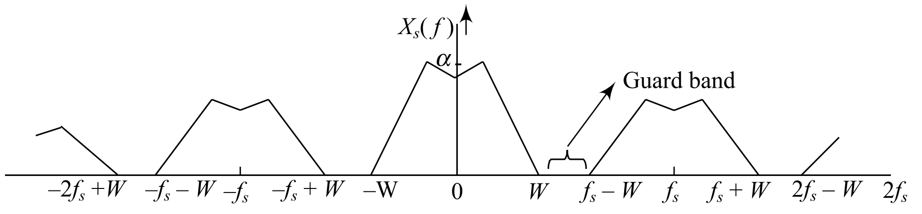


Fig. 8.5 *Spectrum of $x_s(t)$ when $f_s > 2W$*

In case f_s is less than $2W$, the spectrum $X_s(f)$, of the sampled version of $x(t)$, viz., $x_s(t)$, will be as shown in Fig. 8.6. In this case, we find that there is no guard band.

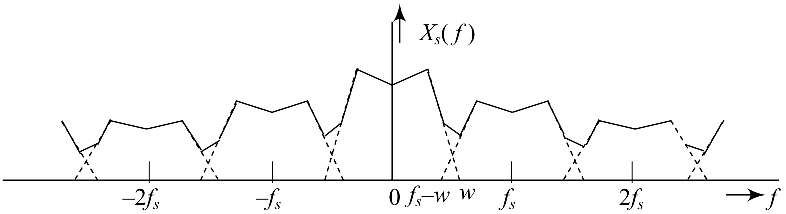


Fig. 8.6 *Spectrum of $x_s(t)$ when $f_s < 2W$*

In fact, the spectra overlap and it is impossible to retrieve $x(t)$ from $x_s(t)$ without distortion. Thus, we find that, in general, there are two basic conditions to be satisfied if $x(t)$ is to be recovered from its samples. These are:

1. $x(t)$ should be band limited to some frequency, W.
2. The sampling frequency should be at least twice the band limiting frequency.

If W is the band limiting frequency, $f_s - 2W$ is called the Nyquist rate of sampling and represents the *theoretical* minimum sampling frequency that can be used if the signal is to be recovered without any distortion from its samples. It is the 'theoretical minimum' because when the Nyquist rate of sampling is used, only an ideal LPF can be used to extract $X(f)$ from $X_s(f)$, i.e., to recover $x(t)$ from $x_s(t)$. However, if $f_s > 2W$, any practical LPF with constant gain over the frequency range $-W$ to W and a phase shift that is proportional to the frequency, will be able to recover $x(t)$, without any distortion from $x_s(t)$.

With the above background, we shall now proceed to the low pass sampling theorem – an extremely important theorem that forms the basis for all modern digital communications. It summarizes the results obtained in the foregoing and guarantees that it is possible to recover the continuous-time signal, $x(t)$, for all time, from its samples taken at regular intervals, if the signal $x(t)$ is band limited and if the sampling is done at or above the Nyquist rate.

8.3 LOW PASS SAMPLING THEOREM

Statement Let $x(t)$ be a band-limited low pass signal, band limited to W Hz, i.e., $X(f) = 0$ for $|f| \geq W$. Then it is possible to recover $x(t)$ completely, without any distortion whatsoever from its samples, if the sampling interval, T_s, is such that $T_s \leq 1/2W$. Specifically, $x(t)$ can be expressed in terms of its samples, $x(kT_s)$ as follows:

$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s)\,\text{sinc}\,2B(t - kT_s) \tag{8.10}$$

where B is any frequency such that $W \leq B \leq (f_s - W)$

Proof Let $x(t)$ have a spectrum $X(f)$ as shown in Fig. 8.7.

Consider $\tilde{X}(f)$ shown in Fig. 8.8, which is a periodic repetition of $X(f)$ at regular intervals of frequency equal to f_s, where $f_s > 2W$.

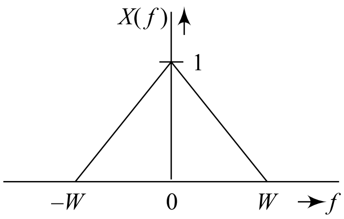


Fig. 8.7 *Spectrum of x(t)*

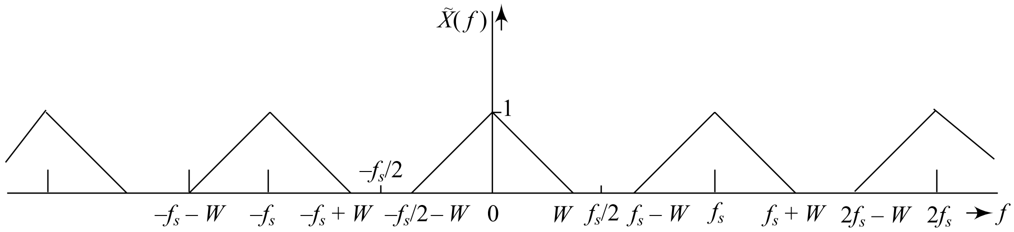


Fig. 8.8 *Spectrum of $x_s(t)$*

Since $f_s > 2W$, dividing by 2 on both sides, $f_s/2 > W$.
Hence, $f_s/2 - W > 0$. Now, adding $f_s/2$ on both sides, we get $f_s - W > f_s/2$. Hence, we have

$$W < \frac{f_s}{2} < (f_s - W) \tag{8.11}$$

i.e., $f_s/2$ lies between W and $(f_s - W)$.

Since $\tilde{X}(f)$ is periodic in frequency with a period of f_s, we can expand it as a Fourier series. Let us say

$$\tilde{X}(f) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi nT_s f}; \quad T_s = \frac{1}{f_s} \tag{8.12}$$

where

$$c_n = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} \tilde{X}(f) e^{-j2\pi nfT_s} df \tag{8.13}$$

Since

$$\tilde{X}(f) = X(f) \ \text{ for } |f| \leq \frac{f_s}{2} \quad \text{and} \quad X(f) = 0 \ \text{ for } |f| \geq W$$

Equation (8.13) may be written as

$$c_n = T_s \int_{-\infty}^{\infty} X(f) e^{-j2\pi nfT_s} df \tag{8.14}$$

But

$$\int_{-\infty}^{\infty} X(f) e^{+j2\pi nfT_s} df = \left\{\mathcal{F}^{-1}[X(f)]\right\}\Big|_{t=-nT_s}$$

Hence, Eq. (8.14) may be written as

$$c_n = T_s x(t) \Bigg|_{t=-nT_s} = T_s x(-nT_s)$$

Substituting this in Eq. (8.12), we have

$$\tilde{X}(f)=\sum_{n=-\infty}^{\infty} T_s x(-nT_s)e^{j2\pi nfT_s}$$

If we put $k=-n$

$$\tilde{X}(f)=T_s\sum_{k=-\infty}^{\infty} x(kT_s)e^{-j2\pi fkT_s} \tag{8.15}$$

If we now define a gate pulse $W_{2B}(f)$ of width $2B$ Hz in the frequency domain, i.e., if

$$W_{2B}(f)\triangleq\Pi\left(\frac{f}{2B}\right) \tag{8.16}$$

then

$$W_{2B}\tilde{X}(f)=X(f) \tag{8.17}$$

$\therefore$

$$X(f)=T_s\sum_{k=-\infty}^{\infty} x(kT_s)e^{-j2\pi fkT_s}\cdot W_{2B}(f) \tag{8.18}$$

But

$$x(t)=\mathcal{F}^{-1}[X(f)]=T_s\,\mathcal{F}^{-1}\left[\sum_{k=-\infty}^{\infty} x(kT_s)e^{-j2\pi fkT_s}\cdot W_{2B}(f)\right]$$

$$=T_s\sum_{k=-\infty}^{\infty} x(kT_s)[\mathcal{F}^{-1}\{e^{-j2\pi fkT_s}\cdot W_{2B}(f)\}]$$

$$=T_s\sum_{k=-\infty}^{\infty} x(kT_s)\{\mathcal{F}^{-1}[e^{-j2\pi fkT_s}]\}*\{\mathcal{F}^{-1}[W_{2B}(f)]\} \tag{8.19}$$

In Eg. 8.19 we made use of the convolution theorem of Fourier transform now, noting that

$$\mathcal{F}^{-1}[e^{-j2\pi fkT_s}]=\delta(t-kT_s) \tag{8.20}$$

and

$$\mathcal{F}^{-1}[W_{2B}(f)]=\mathcal{F}^{-1}[\Pi(f/2B)]=2B\,\text{sinc}\,2Bt, \tag{8.21}$$

we have

$$x(t)=2BT_s\sum_{k=-\infty}^{\infty} x(kT_s)[\delta(t-kT_s)*\text{sinc}\,2Bt]$$

Using the replication property of an impluse, we have,

$\therefore$

$$\boxed{x(t)=2BT_s\sum_{k=-\infty}^{\infty} x(kT_s)\,\text{sinc}\,2B(t-kT_s)} \tag{8.22}$$

Equation (8.22) tells us how we may reconstruct the signal $x(t)$ from its samples, $x(kT_s)$. It says that $x(t)$ is the weighted sum of an infinite number of the interpolating functions sinc $2B(t-kT_s)$ with $x(kT_s)$ as the weightage given to the sinc function delayed by an amount of time kT_s. Since $2B$ sinc $2Bt$ is the impulse response of an ideal low pass filter whose cut-off frequency is B Hz and whose pass band gain is 1, Eq. (8.22), in fact, gives us the clue as to how we may reconstruct $x(t)$ from its samples – obtain a sequence of impulses at regular intervals of T_s, with the impulse at $t=kT_s$ having a strength equal to $x(kT_s)$, the value of the k^{th} sample of $x(t)$, and then give this sequence of impulses as input to an ideal LPF whose cutoff frequency is B Hz. The output of the ideal LPF will then be proportional to $x(t)$.

To get a better appreciation of the foregoing, let us first consider what is generally called 'Ideal Sampling', 'Impulse Sampling' or 'Instantaneous Sampling'.

8.4 IDEAL OR IMPULSE SAMPLING

Earlier, in Section 8.2, we had considered sampling of a continuous-time waveform using periodic rectangular pulses of width τ.

Ideally, sampling should be done instantaneously so that the k^{th} element of the sequence obtained by sampling represents the value of $x(t)$ at $t=kT_s$. However, for obtaining this instantaneous sampling, if we try to reduce τ, the pulse width in the sampling function, to zero, the duty ratio α will be zero and hence, $x_s(t)$ will be zero, as may be seen from Eq. (8.7).

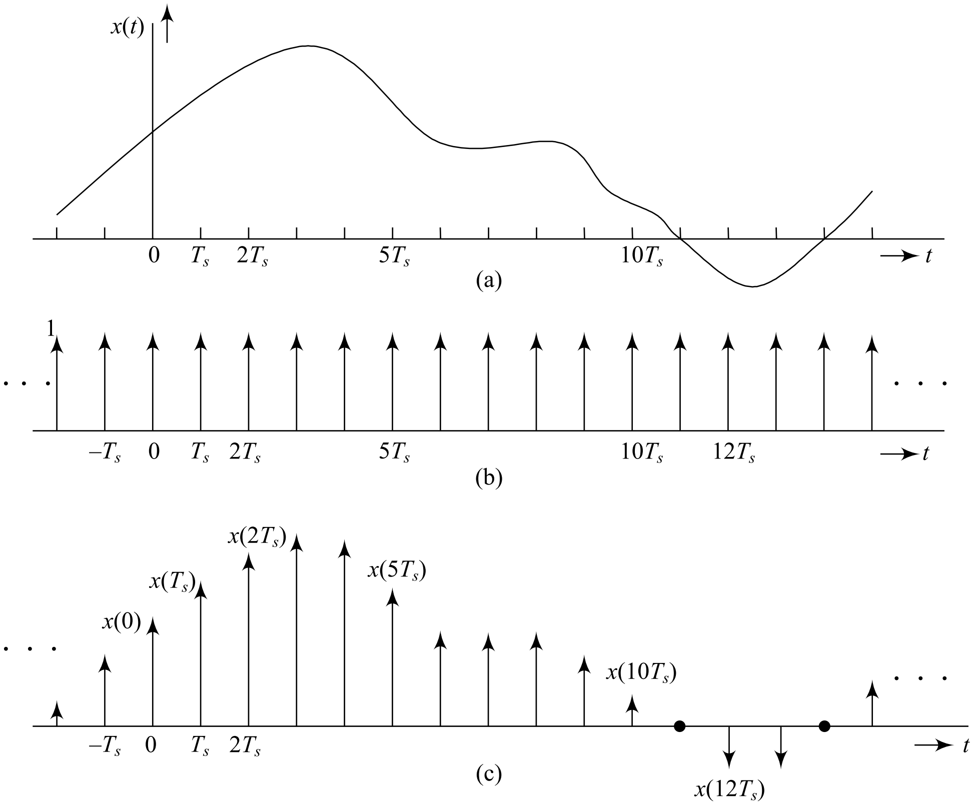


Fig. 8.9 *(a) x(t), (b) s(t), (c) $x_\delta(t)$*

To overcome this difficulty, we will consider a sampling function that is a sequence of unit impulses as shown in Fig. 8.9(b) instead of a sequence of unit amplitude pulses of zero width. This $s(t)$ may be expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{8.23}$$

If we again model the sampling process as multiplication of $x(t)$ by the sampling function $s(t)$, we have the sampled version $x_s(t)$, or, in this case, $x_\delta(t)$, given by

$$x_\delta(t) = x(t) \cdot s(t) \tag{8.24}$$

or

$$X_\delta(f) = X(f) * S(f) \tag{8.25}$$

To find $S(f)$, let us make use of the fact that $s(t)$ is a periodic function with a period of T_s. Hence, we may write its Fourier series expansion as

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t}; \quad f_s = \frac{1}{T_s}; \quad -\infty < t < \infty \tag{8.26}$$

where

$$c_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} s(t) e^{-j2\pi n f_s t} dt \tag{8.27}$$

However, for $-\frac{T_s}{2} \le t \le \frac{T_s}{2}$, $s(t) = \delta(t)$, as may be seen from Fig. 8.9(b).

Hence,
$$c_n = \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} \delta(t)\, e^{-j2\pi n f_s t}\, dt = \frac{1}{T_s} e^{-j2\pi n f_s t}\Big|_{t=0} = f_s \cdot 1 = f_s \tag{8.28}$$

Since $c_n = f_s$ for all values of n, substituting this in Eq. (8.26), we get

$$s(t) = f_s \sum_{n=-\infty}^{\infty} e^{j2\pi n f_s t}$$

Taking Fourier Transform on both sides,

$$S(f) = f_s \mathcal{F}\left[\sum_{n=-\infty}^{\infty} e^{j2\pi n f_s t}\right] = f_s \sum_{n=-\infty}^{\infty} \mathcal{F}[e^{j2\pi n f_s t}]$$

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

$$\therefore \quad S(f) = f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \tag{8.29}$$

Substituting this in Eq. (8.25) and realizing that

$$X(f) * \delta(f - nf_s) = X(f - nf_s)$$

and invoking the linearity theorem of FT, we have

$$\boxed{X_\delta(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)} \tag{8.30}$$

Equation (8.30) tells us that the spectrum of the ideally sampled version of $x(t)$, viz., $X_\delta(f)$ is nothing but a periodic repetition of $X(f)$, the spectrum of $x(t)$, with a period of repetition f_s and is scaled by the factor f_s. Hence, if $x(t)$ is a low pass signal band limited to W, with a spectrum as shown in Fig. 8.10(a), then $X_\delta(f)$ would be as shown in Fig. 8.10(b).

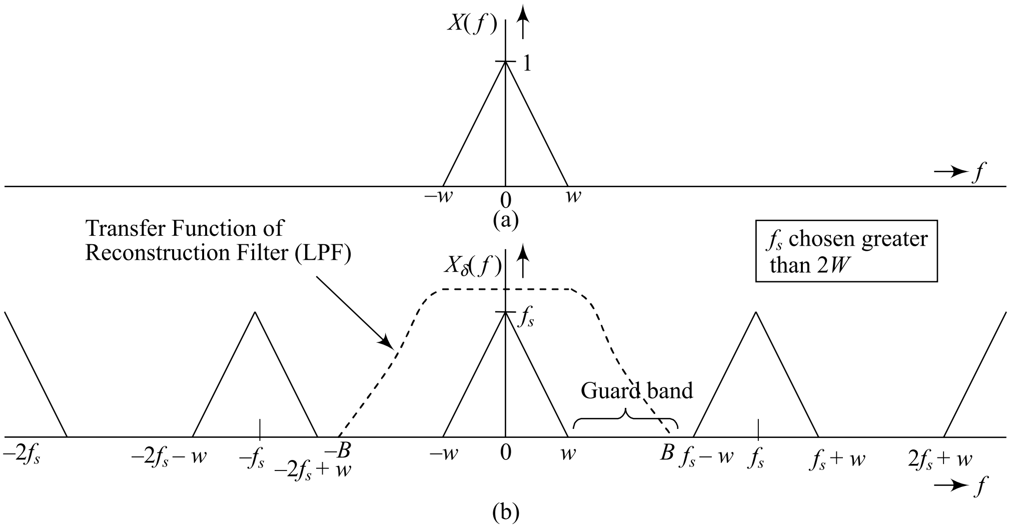


Fig. 8.10 *(a) shows X(f) and (b) shows $X_\delta(f)$*

For the sake of drawing Fig. 8.10(b), it ha*s been assumed that* $f_s > 2W$*, i.e., the sampling is done above the Nyquist rate. Hence* $f_s - W > W$ and a guard band appears in the spectrum of $x_\delta(t)$. Because of the presence of the guard band, as shown in the figure, it is possible to recover $X(f)$ from $X_\delta(f)$, i.e., $x(t)$ from $x_\delta(t)$ without any distortion using any practical low pass filter whose pass band gain is constant over the range of frequencies 0 Hz to W Hz within which all the frequency components present in the signal $x(t)$ are contained.

If f_s is equal to $2W$, i.e., if the sampling is done exactly at the Nyquist rate, $f_s - W = W$ and therefore the spectrum of $x_\delta(t)$ would appear as shown in Fig. 8.11.

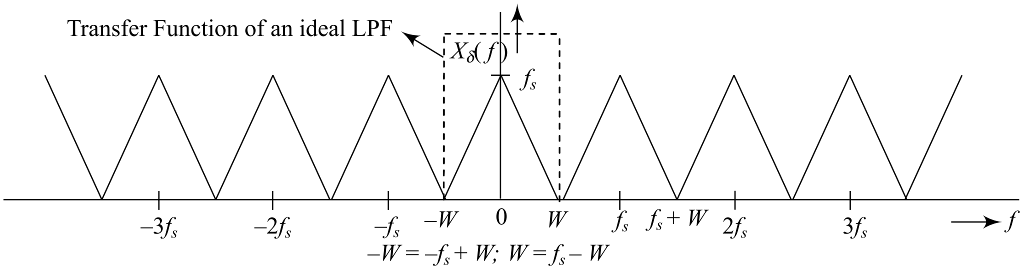


Fig. 8.11 *Spectrum of* $x_\delta(t)$ *when* $f_s = 2W$

As before, a low pass filter may be employed to separate out $X(f)$ from the rest of the spectrum of $x_\delta(t)$. However, as the pass band gain of this filter has to be constant at least from $-W$ to $+W$ for obtaining $x(t)$ without any distortion and as there is no guard band in the present case, only an ideal low pass filter with a cut-off frequency of W will have to be used, as shown in dotted lines in Fig. 8.11.

If the sampling is done at less than the Nyquist rate, i.e., if $f_s < 2W$, then $f_s - W < W$ and therefore the spectrum of $x_\delta(t)$ would appear as shown in Fig. 8.12.

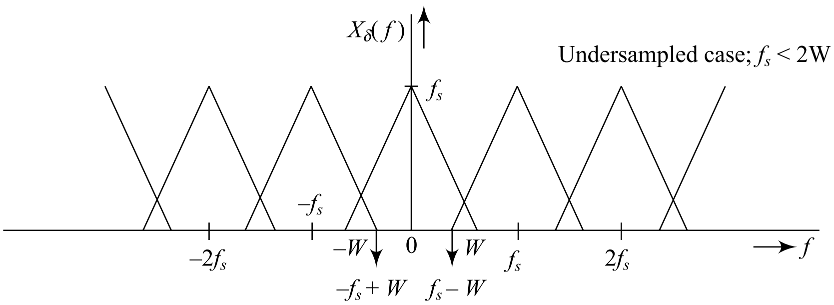


Fig. 8.12 *Spectrum of* $x_\delta(t)$ *when* $f_s < 2W$

In this case, we find that the spectra overlap and hence it is not possible to separate $X(f)$ from $X_\delta(f)$, i.e., it is not possible to recover $x(t)$ from $x_\delta(t)$ even if we were to use an ideal LPF. As may be seen from Fig. 8.12, because of this overlapping, the high frequency components of $x(t)$ reappear as low frequency components. This phenomenon is therefore appropriately referred to as '*Aliasing*'. It is also called '*Frequency Folding Effect*'.

We may summarize the foregoing discussion on the effect of sampling rate as follows.

1. $X_\delta(f)$ is a repetitive version of $X(f)$, with $X(f)$ repeating itself at regular intervals of f_s, the sampling frequency.
2. If $f_s > 2W$, then there is a guard band and it is easy to separate out $X(f)$ from $X_\delta(f)$, i.e., easy to recover $x(t)$ from $x_\delta(t)$ using a practical LP filter.

3. If $f_s = 2W$, i.e., Nyquist rate, no guard band exists and an ideal LPF is needed to recover $x(t)$ from $x_\delta(t)$.
4. If $f_s < 2W$, aliasing takes place and it is not possible to recover $x(t)$ from $x_\delta(t)$ without distortion.
5. To avoid aliasing, it should be ensured that
   (a) $x(t)$ is strictly band limited
   (b) f_s is greater than $2W$

We have all along been assuming that $x(t)$ is band limited to W. But, it must be realized that in practice, signals are time limited, i.e., no practical signal exists from –∞ of time to +∞ of time. This means that no signal will, in practice, be strictly band limited. For example, if the spectrum of a signal $x(t)$ is as shown in Fig. 8.13, it is necessary to first band limit $x(t)$ to some appropriate frequency W such that most part of the energy is retained.

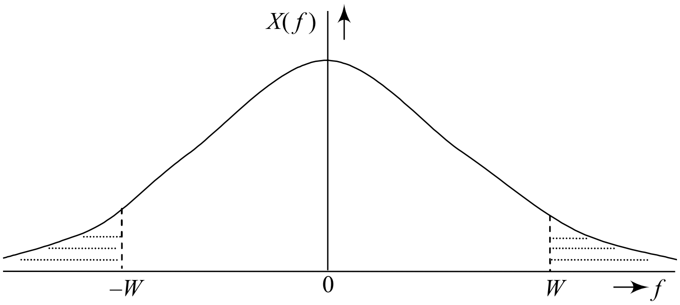


Fig. 8.13 *Spectrum of an x(t)*

We then choose a sampling frequency f_s such that it is more than $2W$. The choice of W depends on the application. For example, speech signals can have frequencies up to even 15 kHz if it is a female voice. But, for digital telephony it is band limited to 3.4 kHz and sampled at 8 kHz. This is because, for this application, intelligibility is the criterion governing the choice of W. The minimum possible value of W is chosen for the sake of reducing the required bit rate, consistent with the requirement that the speech should be intelligible at the destination. A value of W equal to 3.4 kHz has been found to satisfy the requirement. This filter, an LPF, used for band limiting a signal before sampling, is generally referred to as an *anti-aliasing filter* since it is used primarily for preventing aliasing. Incidentally, this anti-aliasing filter helps in cutting off the out-of-band noise, if any, present along with the signal. This noise would otherwise alias into the useful band 0 Hz to W Hz after sampling. Similarly, for high fidelity music, a minimum bandwidth (W) of 20 kHz is needed. That is why, in CD music systems, a sampling frequency of 44.1 kHz, which is slightly more than the Nyquist rate, is used.

Example 8.1 The signal $x(t) = 10 \cos 150\,\pi t$ is ideally sampled at a frequency f_s = 200 samples per second. Sketch the spectrum of $x_\delta(t)$.

Solution

$$X(f) = \mathcal{F}[x(t)] = \mathcal{F}[10 \cos 150\,\pi t]$$
$$= 5[\delta(f-75) + \delta(f+75)]$$

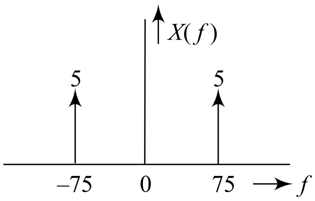


Fig. 8.14 *Sketch of X(f)*

Since the spectrum $X_\delta(f)$ of $x_\delta(t)$ is given by

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

the sketch of it is as follows:

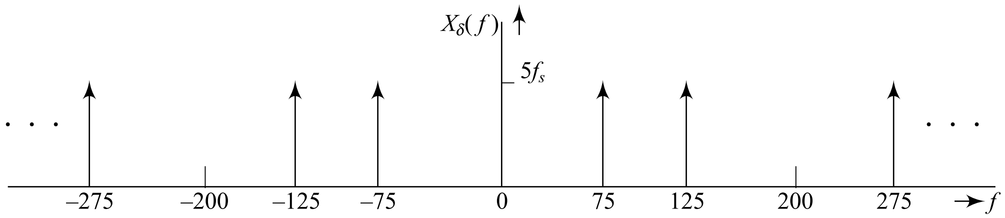


Fig. 8.15 *Sketch of $X_\delta(f)$ with f_s = 200 samples per second*

Example 8.2 For the $x(t)$ of Example 8.1, sketch $x_\delta(t)$, the spectrum of $x_\delta(t)$, the ideally sampled version of $x(t)$, if the sampling is done at a frequency $f_s = 100$ sps.

Solution

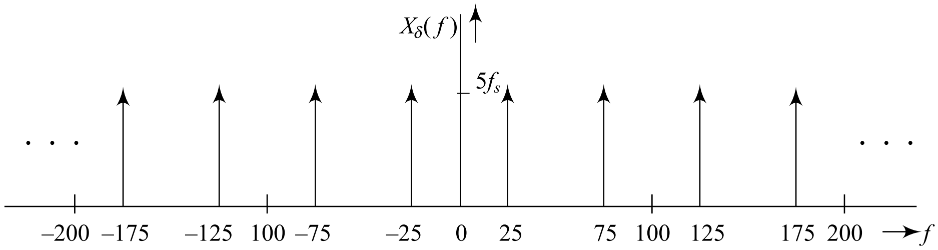


Fig. 8.16 *Sketch of $X_\delta(f)$ of Example 8.2 with f_s = 100 samples per second*

Note the presence of the 25 Hz component in the spectrum of $x_\delta(t)$ even though $x(t)$ contains only the 75 Hz component. This is because the sampling frequency in this example is 100 samples per second while the frequency of the signal is 75 Hz. Thus, the sampling rate of 100 samples per second is less than the Nyquist rate of sampling which is equal to 150 Hz. Hence, aliasing takes place and we should recognize the fact that the 75 Hz component of $x(t)$ is itself reappearing as a low frequency component at 25 Hz because of aliasing.

Example 8.3 What is the minimum number of samples required to exactly describe the following signal?

$$x(t) = 10\cos(6\pi t) + 4\sin(8\pi t)$$

Solution If $x(t)$ is periodic then it can be described exactly by a finite number of samples – corresponding to those in one period of $x(t)$. So, let us first check whether $x(t)$ is periodic

$$T_1 = \text{period of } 10\cos 6\pi t = \frac{2\pi}{6\pi} = \frac{1}{3}$$

$$T_2 = \text{period of } 4\sin 8\pi t = \frac{2\pi}{8\pi} = \frac{1}{4}$$

$\therefore \frac{T_1}{T_2} = \frac{1}{3} \times \frac{4}{1} = \frac{4}{3}$, which is a rational number.

Hence, $x(t)$ is periodic. Now, to determine its period T,

$$T = \text{LCM}\left(\frac{1}{3}, \frac{1}{4}\right) = 1 \qquad \therefore\ T = 3T_1 = 4T_2$$

The maximum frequency present in $x(t)$ is 4 Hz, which is the frequency of the $\sin(8\pi t)$ component.

$\therefore$ the minimum sampling frequency required = 8 samples per second.

$\therefore$ the number of samples in one period of $x(t)$ is equal to 8 since $T = 1$ sec and the sampling frequency is 8 samples per second.

Example 8.4 Determine the minimum sampling frequency to be used to sample the signal

$$x(t) = 100\,\text{sinc}^2\,100t$$

if the signal $x(t)$ is to be recovered from the samples without any distortion.

Solution $x(t) = 100\,\text{sinc}^2\,100t = (10\,\text{sinc}\,100t)\cdot(10\,\text{sinc}\,100t)$

We know that $10\,\text{sinc}\,100t \xleftrightarrow{\text{FT}} 0.1\,\Pi(f/100)$

$$\therefore \qquad \mathcal{F}[100\,\text{sinc}^2\,100t] = \left[0.1\,\Pi(f/100)\right] * 0.1\,\Pi(f/100)$$

Referring to Section 2.5, of Chapter 2, we find that convolution of two identical rectangular pulses results in a triangular pulse whose base width is twice that of each rectangular pulse.

$$\mathcal{F}[100\,\text{sinc}^2\,100t] = [0.1\,\Pi(f/100)] * [0.1\,\Pi\,(f/100)]$$

$$\therefore \qquad = 0.01 \times 100\Lambda\,(f/200)$$

where $\Lambda(f/200)$ denotes a triangular pulse as shown.

Thus, the signal $x(t) = 100\,\text{sinc}^2\,100t$ is a low pass signal band limited to 100 Hz.

Hence the Nyquist rate for it is 200 samples per second.

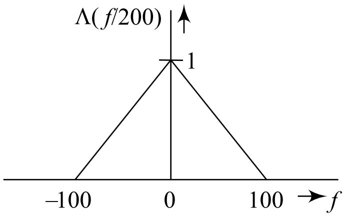


Fig. 8.17 *Triangular spectrum of x(t) = 100 sinc² 100t*

8.5 RECONSTRUCTION

As already mentioned earlier, 'recovering $X(f)$ from $X_\delta(f)$' and 'reconstructing $x(t)$ from $x_\delta(t)$' are one and the same; the only difference being that in the former case, it is looked upon as a frequency-domain operation, while in the latter, it is looked upon as a time-domain operation.

We shall now briefly analyze and see how the signal $x(t)$ is recovered in each case. First we shall consider the frequency-domain operation.

Then, as shown in Fig. 8.11, let us assume $f_s = 2W$ and that an ideal LPF is used to recover $X(f)$ form $X_\delta(f)$. Let the ideal LPF have a gain of T_s in the pass band and let it introduce τ sec time delay. Then, we can write down its transfer function $H(f)$ as

$$H(f) = T_s \Pi(f/f_s) e^{-j\omega\tau} \tag{8.31}$$

Hence, the spectrum of the output of the filter is

$$Y(f) = X_\delta(f) \cdot H(f) = T_s f_s X(f) e^{-j\omega\tau}$$

$$\therefore \qquad Y(f) = X(f) e^{-j\omega\tau} \tag{8.32}$$

or taking the inverse Fourier transform on both sides,

$$y(t) = x(t-\tau) \tag{8.33}$$

Thus, the output of the LPF is a time-shifted version of the signal $x(t)$.

We now consider the reconstruction operation in the time domain.

$x_\delta(t)$ is a sequence of weighted impulses given by

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t-nT_s) \tag{8.34}$$

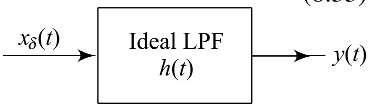


Fig. 8.18 *Recovering x(t) from x_δ(t), the sampled version*

This weighted sequence, when given as the input to the ideal LPF with impulse response $h(t)$, gives an output signal $y(t)$ given by

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT_s)h(t-nT_s) \tag{8.35}$$

where $h(t)$, the impulse response of the ideal LPF is given by

$$h(t) = \mathcal{F}^{-1}[H(f)] = \mathcal{F}^{-1}[T_s \Pi(f/2W) e^{-j\omega\tau}]$$

$$= 2BT_s\,\text{sinc}\,2B(t-\tau) \tag{8.36}$$

In our case, the cut-off frequency B of the LPF = $W = f_s/2$.

$$\therefore \qquad h(t) = 2\frac{f_s}{2} T_s\,\text{sinc}\,2B(t-\tau)$$

$$= \text{sinc}\,2B(t-\tau) \tag{8.37}$$

Taking the time delay τ introduced by the LPF equal to zero, and substituting for $h(t)$ in Eq. (8.35) using Eq. (8.37), we have

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{sinc}\,2B(t - nT_s) \tag{8.38}$$

But from Eq. (8.22), we realize that RHS of Eq. (8.38) is nothing but $x(t)$, since $B = \dfrac{f_s}{2} = \dfrac{1}{2T_s}$ in this case.

$$\therefore \qquad x(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\mathrm{sinc}\,2B(t - nT_s) \tag{8.39}$$

As explained earlier in Section 8.3 in connection with the low pass sampling theorem and Eq. (8.22), when $x_\delta(t)$, a sequence of weighted impulses, is given as input to an ideal LPF, the output will be a sequence of weighted sinc pulses (since sinc pulse is the impulse response of an LPF) as shown in Fig. 8.19. When these are all added, together with their precursors and post-cursors, Eq. (8.39) tells us that we get $x(t)$. Hence, when the sampled version $x_\delta(t)$ is fed as input to the LPF, $x(t)$ appears at the output. Since the LPF reconstructs the original signal $x(t)$ from its sampled version, it is generally referred to as the 'Reconstruction Filter'.

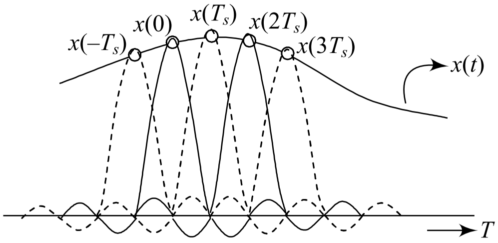


Fig. 8.19 *Reconstruction of x(t) from its samples (A sketch of RHS of Eq. (8.39))*

8.6 SAMPLING USING A SEQUENCE OF PULSES – NATURAL SAMPLING

Instead of a sequence of unit impulses as the sampling function $s(t)$, one may use a sequence of pulses $p(t)$ of width τ along the time axis occurring at regular intervals of $T_s = 1/f_s$ such that $\tau << T_s$. The actual shape of the pulse $p(t)$ is not important, although for the sake of illustration it is shown as a rectangular pulse in Fig. 8.20(b). Again, modeling the sampling process as multiplication of $x(t)$ by $s(t)$, we have

$$x_s(t) = x(t)\cdot s(t) \tag{8.40}$$

where

$$s(t) = \sum_{k=-\infty}^{+\infty} p(t - kT_s) \tag{8.41}$$

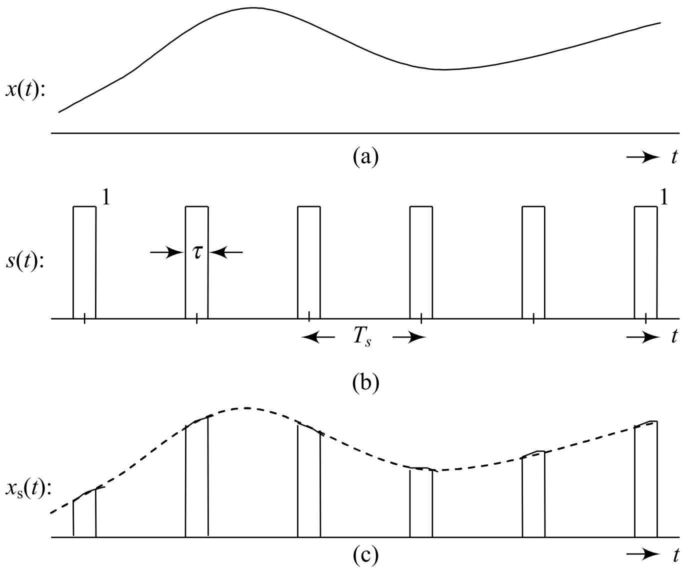


Fig. 8.20 *(a) Signal x(t), (b) Sampling function s(t), (c) Sampled version of x(t), i.e., $x_s(t)$*

As $s(t)$ is a periodic pulse train, let us write its Fourier series expansion

$$s(t) = \sum_{k=-\infty}^{+\infty} p(t - kT_s) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t} \tag{8.42}$$

where

$$c_n = \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} s(t)e^{-j2\pi nf_st}dt$$

Since τ, the width of $p(t)$ is very much less than T_s and $p(t) = 0$ for $|t| \geq \tau/2$, we may write

$$c_n = \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} s(t)e^{-j2\pi nf_st}dt = \frac{1}{T_s}\int_{-\infty}^{\infty} p(t)e^{-j2\pi nf_st}dt$$

$$c_n = f_sP(nf_s) \tag{8.43}$$

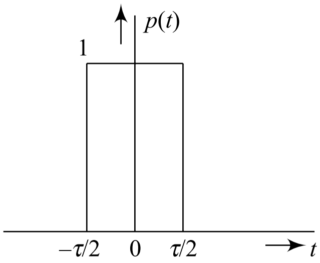


Fig. 8.21 *Pulse p(t)*

where

$$P(nf_s) = \mathcal{F}[p(t)]\Big|_{f=nf_s} \tag{8.44}$$

$$\therefore \quad s(t) = f_s\sum_{n=-\infty}^{\infty} P(nf_s)e^{j2\pi nf_st} \tag{8.45}$$

and

$$X_s(f) = \mathcal{F}[x_s(t)] = \mathcal{F}\left[f_s\sum_{n=-\infty}^{\infty} P(nf_s)x(t)e^{j2\pi nf_st}\right]$$

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

Since $\mathcal{F}[e^{j2\pi nf_st}] = \delta(f - nf_s)$

Hence,

$$X_s(f) = f_s\sum_{n=-\infty}^{\infty} P(nf_s)X(f - nf_s) \tag{8.46}$$

If $x(t)$ has a spectrum as shown in Fig. 8.10(a), $X_s(f)$, the spectrum of the sampled version of $x(t)$ will appear as shown in Fig. 8.22.

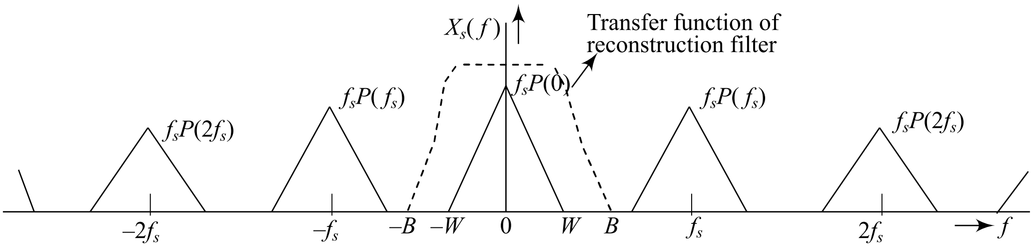


Fig. 8.22 *Spectrum of the sampled version of x(t)*

From the figure, it is clear that $X(f)$ can be recovered from $X_s(f)$, i.e., $x(t)$ can be recovered from $x_s(t)$, if $f_s > 2W$ by using a low pass filter whose pass band gain is constant at least up to W Hz and whose cut-off frequency B is such that $W < B < f_s - W$, as shown. This is true whatever may be the pulse shape, as mentioned earlier.

Example 8.5 The signal $x(t) = 2\cos 200\pi t + 6\cos 180\pi t$ is ideally sampled at a frequency of 150 samples per sec. The sampled version $x_\delta(t)$ is passed through a unit gain ideal LPF with a cut-off frequency of 110 Hz. What frequency components will be present in the output of the LPF. Write down an expression for its output signal.

Solution $x(t) = 2\cos 200\pi t + 6\cos 180\pi t$

$$= 2\cos 2\pi(100)t + 6\cos 2\pi(90)t$$

Hence, taking FT on both sides,

$$X(f) = [\delta(f+100) + \delta(f-100)] + 3[\delta(f+90) + \delta(f-90)]$$

This is depicted in Fig. 8.23.

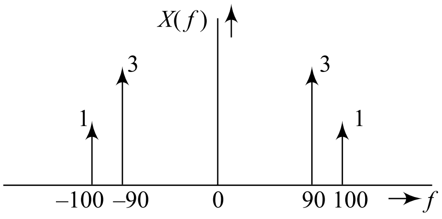


Fig. 8.23 *Spectrum of x(t)*

The spectrum of $x_\delta(t)$, the ideally sampled version of $x(t)$ is a periodic repetition of $X(f)$ at regular intervals of f_s, i.e., 150 Hz; and will be as shown in Fig. 8.24.

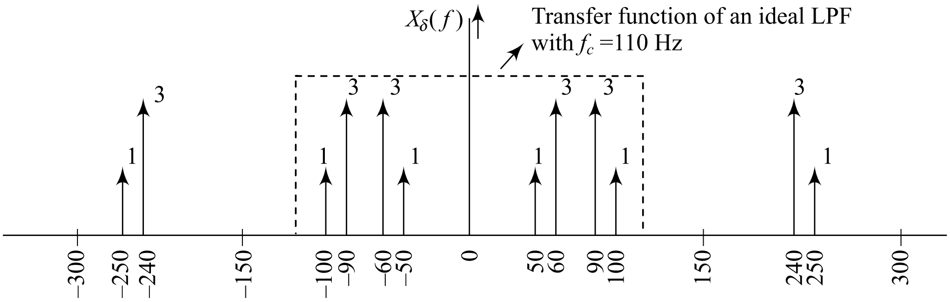


Fig. 8.24 *Spectrum of sampled version of x(t)*

From Fig. 8.24, it is clear that the output of the LPF contains frequency components at 50 Hz, 60 Hz, 90 Hz and 100 Hz, although the original analog signal contains only 90 Hz and 100 Hz components. As $x(t)$ is under sampled, aliasing is taking place. 50 Hz component is the alias of the 100 Hz component and 60 Hz component is the alias of the 90 Hz component.

The expression for the output of the LPF is given by

$$x(t) = 2[\cos 2(50)\pi t + \cos 2\pi(100)t] + 6[\cos 2\pi(60)t + \cos 2\pi(90)t]$$

Example 8.6 The signal $x(t) = 12\cos(800\pi t)\cos^2(1800\pi t)$ is ideally sampled at 4600 samples per second. What is the minimum allowable sampling frequency? What is the range of permissible cut-off frequencies for the ideal low pass filter to be used for reconstructing the signal?

Solution $x(t) = 12\cos(800\,\pi t)\left[\frac{1}{2}\{1+\cos 3600\,\pi t\}\right]$

$$= 6\cos 800\,\pi t + 6\cos 800\,\pi t \cdot \cos 3600\,\pi t$$

$$= 6\cos 800\,\pi t + 3\cos 4400\,\pi t + 3\cos 2800\,\pi t$$

Hence, the maximum frequency component present in $x(t)$ has a frequency of 2200 Hz. So, the minimum allowable sampling frequency, i.e., the Nyquist rate is 4400 samples per second.

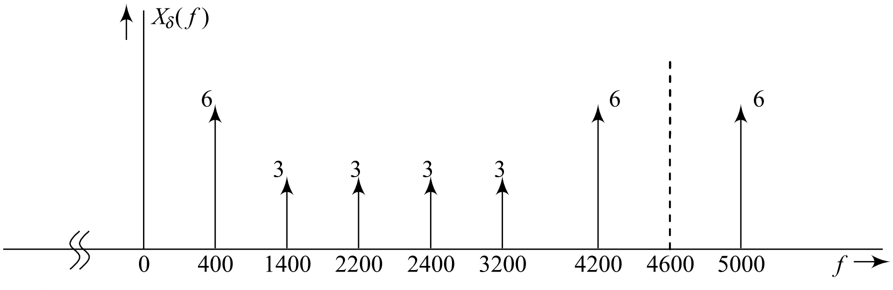


Fig. 8.25 *Spectrum of ideally sampled version of x(t) (only one-sided spectrum drawn)*

From Fig. 8.25, it is clear that in order to recover the three frequency components at 400 Hz, 1400 Hz and 2200 Hz which are present in $x(t)$ and avoid other frequencies, the cut-off frequency of the ideal LPF should be above 2200 Hz but less than 2400 Hz.

8.7 PRACTICAL SAMPLING

In practice, sampling is done using what is generally referred to as the 'Sample – and – Hold' circuit which produces 'flat – top sampling' unlike in the previous case wherein the sampled version consisted of pulses whose top followed the contour of $x(t)$. The schematic of a 'Sample – and – Hold' (S/H) circuit is shown in Fig. 8.26 and a typical output waveform from a S/H circuit is shown in Fig. 8.27.

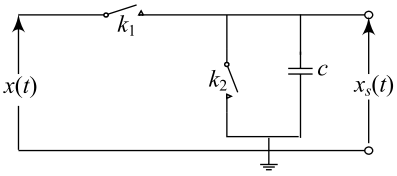


Fig. 8.26 *Schematic of an S/H circuit*

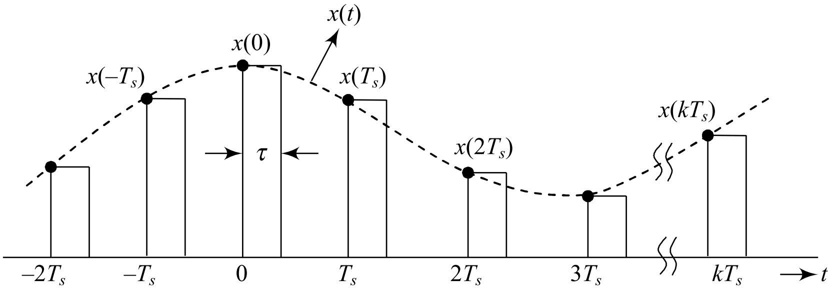


Fig. 8.27 *Signal x(t) and output of S/H circuit*

The S/H circuit essentially consists of two switches k_1 and k_2 and a capacitor C, connected as shown in Fig. 8.26. With k_2 open, k_1 is closed for a very brief period at each sampling instant. The capacitor C then gets charged to a voltage equal to the value of the input signal $x(t)$ at the sampling instant and holds it for a period τ at the end of which, k_2 is closed to allow the capacitor to discharge. This sequence of operations is repeated at the next and all subsequent sampling instants. The switches k_1 and k_2 are generally FET switches and are operated by giving appropriate pulses to their gates. An actual S/H circuit uses one or two op–amps also. The voltage across C appears as $x_s(t)$ and is sketched in Fig. 8.27.

It may be observed that the amplitude of each pulse in Fig. 8.27 is equal to the amplitude of the sample of $x(t)$ at that sampling instant. Hence, the output of S/H circuit is a Pulse Amplitude Modulated or a PAM signal. In Fig. 8.27, the pulses shown are all of single polarity because the $x(t)$ is positive throughout. In case $x(t)$ takes negative values also, the PAM signal obtained will be of double polarity.

From the figure, it is obvious that the sampled version, $x_s(t)$ consists of a sequence of rectangular pulses, the leading edge of the k^{th} pulse being at $t = kT_s$ and the amplitude of the pulse being the value of $x(t)$ at $t = kT_s$, i.e., $x(kT_s)$. Hence, we may write

$$x_s(t) = \sum_{k=-\infty}^{\infty} x(kT_s)p(t - kT_s) \tag{8.47}$$

where $$p(t) \triangleq \Pi\left(\frac{t - \tau/2}{\tau}\right) \tag{8.48}$$

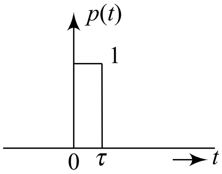


Fig. 8.28 *Pulse p(t)*

and is as shown in Fig. 8.28.

Since $$p(t - kT_s) = p(t) * \delta(t - kT_s),$$

we may write Eq. (8.47) as

$$x_s(t) = p(t) * \sum_{k=-\infty}^{\infty} x(kT_s)\delta(t - kT_s) \tag{8.49}$$

Now, taking Fourier transform on both sides,

$$X_s(f) = \mathcal{F}[x_s(t)] = \mathcal{F}[p(t)] \cdot \mathcal{F}\left[\sum_{k=-\infty}^{\infty} x(kT_s)\delta(t - kT_s)\right]$$

$$\therefore \qquad X_s(f) = p(f) \cdot X_\delta(f) \tag{8.50}$$

where $X_\delta(f)$ is the Fourier transform of $x_\delta(t)$ the ideally sampled version of $x(t)$.

Note that $$P(f) = \mathcal{F}[p(t)]$$

$$= (\tau \operatorname{sinc} f\tau)e^{-j2\pi f\tau/2} = \tau\left(\frac{\sin \pi f\tau}{\pi f\tau}\right)e^{-j\pi f\tau} \tag{8.51}$$

We shall now assume that $x(t)$ has a spectrum as shown in Fig. 8.29 (This shape of $X(f)$ is deliberately chosen for this illustration, as it helps in clearly bringing out the 'Aperture Effect', to be discussed later).

Since $p(t)$ is a rectangular pulse of width τ, its Fourier transform $P(f)$, which is a sinc function, will have an 'inverted bowl' shape as shown in Fig. 8.30(a) and will have its first zero values only at $f = -1/\tau$ and $+1/\tau$. Since $\tau << T_s$, these zero values of $|P(f)|$ which occur at $\pm 1/\tau$, will be far away from f_s and $-f_s$. Since $X_s(f) = P(f) \cdot X_\delta(f)$, its plot will be as shown in Fig. 8.30(b).

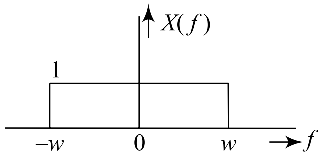


Fig. 8.29 *Assumed shape of X(f)*

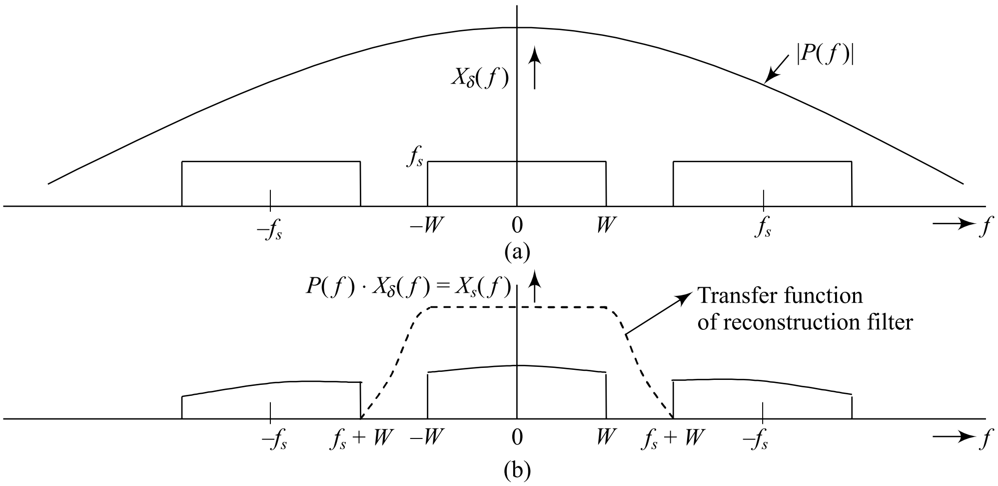


Fig. 8.30 *(a) Plot of P(f) and X_δ(f); (b) Plot of X_s(f) = P(f) · X_δ(f)*

As before, if we pass the sampled version $x_s(t)$ through the reconstruction filter (an LPF), what we get at the output of the filter will not be exactly $x(t)$. It will be a distorted version of $x(t)$ – distorted because, the magnitudes of the high frequency components are relatively reduced, as compared to the magnitudes of the low frequency components, as can be seen in Fig. 8.30(b), because of the multiplication of $X_\delta(f)$ by $P(f)$. *This distortion of x(t), wherein the amplitudes of the high frequency components are reduced relative to the amplitudes of the low frequency components, in the reconstructed signal x(t) obtained from the flat–top sampled version of the signal, is referred to as the 'aperture effect'.*

$P(f)$, which is in the form of an invented bowl as shown in Fig. 8.30(a), will have a relatively flat shape in the message frequency band $-W$ to W if it reaches its zero value at a frequency far greater than W, i.e., if $1/\tau >> W$. This will reduce the 'aperture effect'. Hence, the time τ for which the 'sample – and – hold' circuit holds the sample value, should be made as small as possible, in order to reduce the aperture effect. But this makes the average power in $x_s(t)$ and hence in the reconstructed message, very low. So, we keep the pulse width τ reasonably large and try to reduce the distortion within the message frequency band arising out of the aperture effect by using an equalizer with transfer function $H_e(f)$ in cascade with the reconstruction filter and adjusting $H_e(f)$ so that

$$H_e(f) = \frac{1}{P(f)}; \qquad |f| \le W \tag{8.52}$$

Example 8.7 Figure 8.31(a) shows the spectrum of a particular message signal $x(t)$. If this $x(t)$ is sampled at a rate of 1 kHz using flat-top pulses, each of 0.5 m/sec duration and unit amplitude, determine and sketch the spectrum of the PAM signal that results.

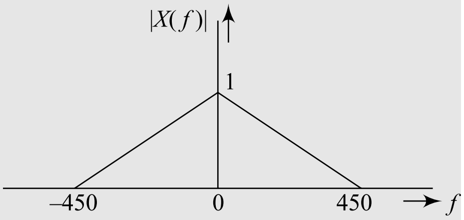


Fig. 8.31(a)

Solution From Eq. (8.50), we know that the spectrum of the flat-top sampled version of $x(t)$, viz., $X_s(f)$ is given by

$$X_s(f) = P(f)X_\delta(f)$$

where $X_\delta(f)$ is the spectrum of the ideally sampled version of $x(t)$, which, we know, is a periodic repetition of $X(f)$ at regular frequency intervals of f_s. $P(f)$ is the spectrum of the sampling pulse $p(t)$ and is given by

$$P(f) = \tau \text{sinc } f\tau e^{-j\pi f\tau} = \tau\left(\frac{\sin \pi f\tau}{\pi f\tau}\right)e^{-j\pi f\tau} \qquad \text{(See Eq. (8.51))}$$

Here, τ is the width of the pulse and is given to be 10^{-4} sec.

Since it is the attenuation of the high frequency components of $x(t)$ relative to the low frequency components that causes the distortion, the constant factor τ and the phase factor $e^{-j\pi f\tau}$ can be ignored and attention can be focused only on $[(\sin \pi f\tau)/(\pi f\tau)]$ to see how it varies with f, the frequency, over the frequency range of interest, i.e., from 0 Hz to 450 Hz.

We know that $\left(\frac{\sin \pi f\tau}{\pi f\tau}\right) = 1$ for $f = 0$ Hz.

f $\lvert X(f)\rvert$	**0** **1**	**100** **0.7777**	**200** **0.5555**	**300** **0.3333**	**400** **0.1111**	**450** **0**
$\left(\frac{\sin \pi f\tau}{\pi f\tau}\right) = 1$	1	0.96639	0.9836	0.9629	0.93547	0.91878
$\lvert X(f)\rvert P(f)\rvert = \lvert X_s(f)\rvert$	1	0.7749	0.54638	0.3209	0.10393	0

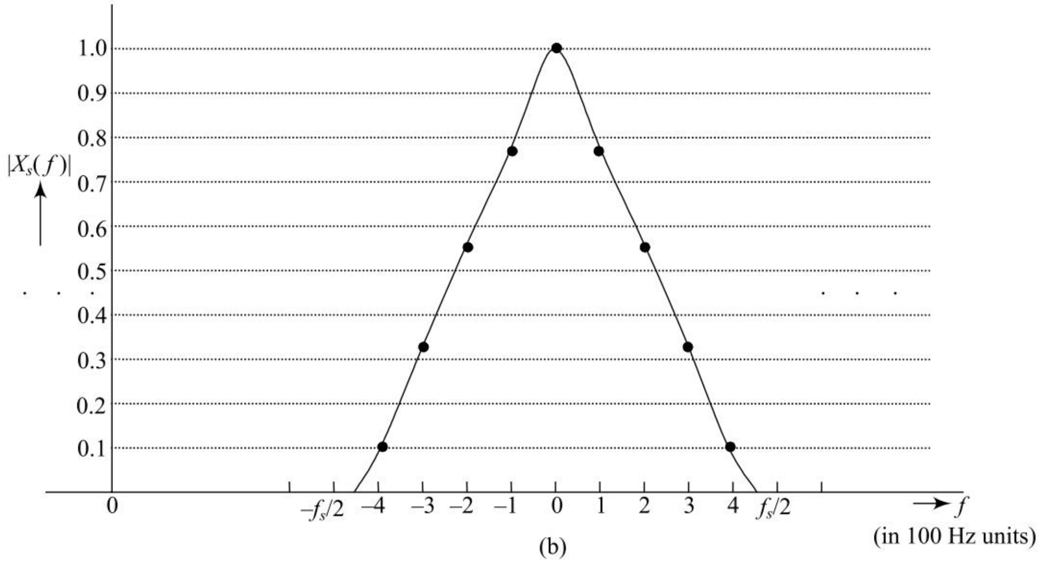


Fig. 8.31(b) *Spectrum of the PAM signal*

8.8 ANTI – ALIASING AND RECONSTRUCTION FILTERS

From the low pass sampling theorem, we know that an analog signal $x(t)$ can be recovered without any distortion from its uniformly sampled version, provided the sampling frequency, f_s, is, at least, twice the highest frequency component present in $x(t)$. If f_s is less than twice the highest frequency component, $x(t)$ cannot be recovered from the sampled version because of the distortion caused by aliasing.

However, in practice, no signal will be strictly band limited, as every practical signal has to be time limited. Hence, prior to sampling, we have to band limit the signal to some frequency W, keeping in view the frequency band of interest in the spectrum of the signal. For this purpose, we use, what is generally called an ‘anti–aliasing filter’ just before the sampler. Such a filter will be helpful in removing ‘out of band frequency components’, or out of band noise, if any, in the original analog signal $x(t)$.

An ideal LPF with a brick wall type of transfer function and a cut-off frequency W, less than $f_s/2$ would be best suited for use as an anti-aliasing filter. However, since such a filter cannot be realized in practice, and since practical filters will have a transition frequency band, the attenuation of the filter should slowly increase from zero at the pass band edge f_p to some desired value at the stop band edge f_{st}, where

$$f_p < f_{st} \le \frac{f_s}{2} \tag{8.53}$$

If the signal were to be band limited to W, we will obviously choose $f_p = W$. In this case, since the gain of the filter is designed to remain almost constant within the pass band, the filter will not distort the signal much, especially if it has a linear phase response too. However, since the signal is not going to be strictly band limited, we have to choose an appropriate portion of the spectrum of $x(t)$, keeping in view the application, and fix the pass band edge, f_p, accordingly. But, since the filter response of the non-ideal filter also is slowly decreasing from f_p onwards, the spectrum of the output of the filter may extend even beyond $f_s/2$. When this happens, the sampling of the output of the filter will create severe aliasing problems. In this connection, it must be realized that it is the frequency components in the band $(f_s - f_p)$ to f_s that alias into the pass band 0 to f_p Hz, which is the useful part of the baseband, as shown in Fig. 8.32.

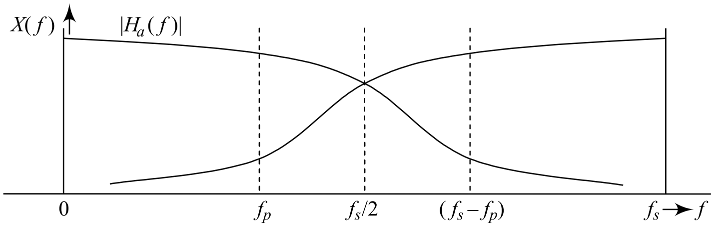


Fig. 8.32 *Anti-aliasing filter response and frequency components aliasing into the baseband*

To keep the amplitude of these aliased components low, the anti-aliasing filter must be so designed that its response falls adequately for frequencies beyond $(f_s - f_p)$. So, if all these aliased components are to be, at least, say, 60 dB below the corresponding ones in the pass band, the filter has to be so designed that it has 60 dB attenuation at a frequency of $(f_s - f_p)$. Noting that the frequency $(f_s - f_p)$ aliases and appears as a frequency component at a frequency of f_p in the pass band, the other frequency components between $(f_s - f_p)$ and f_s which reappear in the pass band between f_p and 0 Hz, will suffer more than 60 dB attenuation.

Butterworth, Chebyshev, Elliptic or Bessel type of analog low pass filters of appropriate order, may be used as anti-aliasing filters. Butterworth filters give reasonably good magnitude as well as phase response. However, if linear phase response is more important, one may go in for Bessel filters – but they give slightly poorer magnitude response. If better magnitude response rather than linear phase response is important, then elliptic or Chebyshev filters may be used.

In applications where distortion due to aliasing has to be kept very low, f_s, the sampling frequency is chosen to be high compared to f_p, the pass band edge, typically about four times. But where it is not critical, f_s is chosen to be a little more than the Nyquist rate, as in the case of digital telephony for which f_p is chosen as 3.6 KHz while f_s is chosen as 8.0 KHz.

As the reader must have realized by now, achieving low aliasing distortion with an f_s that is not much greater than the Nyquist rate, would necessitate the use of a very sharp cut-off low pass filter. So it will be an analog filter of high order and will be quite complex. Sometimes in such cases, to ease the stringent roll-off requirements of the anti-aliasing filter, deliberately an extremely high value of f_s is used for the analog signal, and decimation circuits are used to bring down the sampling frequency of the digital signal at a later stage. Such a deliberate over sampling and a down sampling at a later stage are resorted to in the case of VLSI realization of Digital Signal Processing of analog signals. In compact disk encoding of audio signals, sampling frequencies as high as 3175.2 KHz are used.

8.8.1 Reconstruction Filter

Reconstruction filter is a system that is used to reconstruct the analog signal $x(t)$ from its samples. That is, if the sampled version of $x(t)$ is given as input to the system, ideally it should give $x(t)$ as the output. In frequency domain terms, it means that the transfer function of the reconstruction filter should, as shown in Fig. 8.33, separate out the baseband, i.e., the spectrum of $x(t)$, from the spectrum of $x_s(t)$, which, as we know, consists of periodic repetitions of $X(f)$ at regular intervals of f_s.

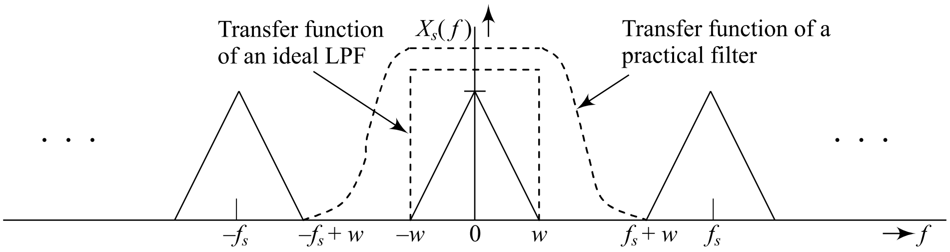


Fig. 8.33 *Action of the reconstruction filter in the frequency domain*

In principle, an ideal LPF with a cut-off frequency of W as shown in Fig. 8.33 would be best suited for being used as a reconstruction filter. However, an ideal LPF is not physically realizable, as its impulse response function, which is the inverse Fourier transform of its transfer function, is a sinc function that extends from minus infinity to plus infinity of time. Hence, any practical low pass filter with a flat amplitude response up to W Hz and whose gain reduces to zero before $(f_s - W)$ may be used.

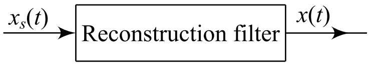


Fig. 8.34 *Action of reconstruction filter in the time domain*

The action of the reconstruction filter when viewed in the time domain is shown in Fig. 8.34. Since the input to the filter is the sequence of samples of $x(t)$, the job of the reconstruction filter is one of interpolating between successive samples. The best interpolator is the ideal LPF. However, in practice, we invariably employ a zero – order – hold (ZOH) for this purpose. Figure 8.35(a) shows the block schematic of a ZOH while Fig. 8.35(b) shows its interpolating action.

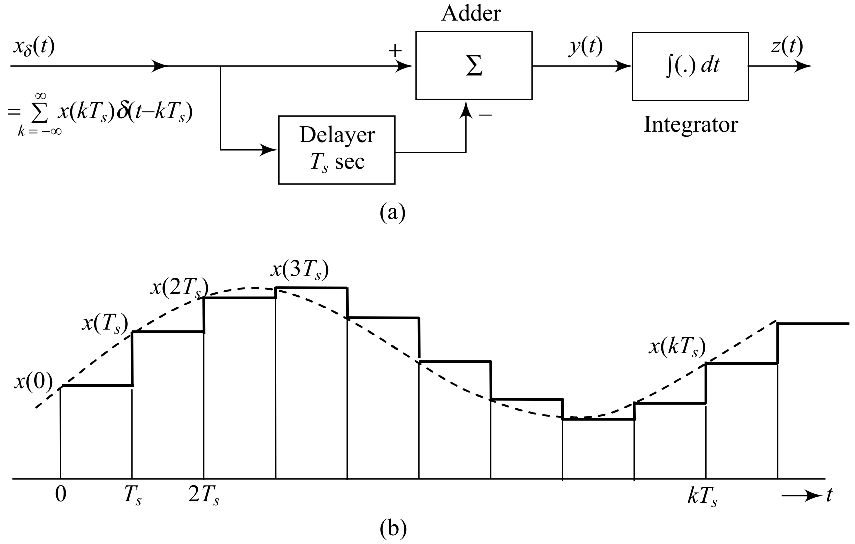


Fig. 8.35 *(a) A Zero-Order-Hold Circuit (ZOH) (b) Interpolation using a ZOH*

It is easy to find the impulse response of a zero–order–Hold as can be seen from the following example:

Example 8.8 Determine the impulse response $h(t)$ and the transfer function $H(f)$, for a ZOH.

Solution If a unit impulse, $\delta(t)$, is given as input to the system,

$$y(t) = \delta(t) - \delta(t - T_s)$$

$$\therefore \quad z(t) = \int y(t)dt = u(t) - u(t - T_s) = p(t) \tag{8.54}$$

where $p(t)$ = Impulse response $h(t)$ and is as shown in Fig. 8.36.

The transfer function $H(f)$ is therefore given by

$$H(f) = \mathcal{F}[h(t)] = T_s \operatorname{sinc}(fT_s)e^{-j\pi fT_s} \tag{8.55}$$

Hence, the output of the ZOH for an input of

$$x_\delta(t) = \sum_{k=-\infty}^{\infty} x(kT_s)\delta(t - kT_s)$$

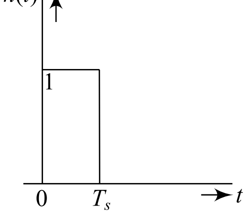


Fig. 8.36 *Impulse response of ZOH*

is a staircase waveform as shown in Fig. 8.35(b). This contains several high frequency components outside the baseband. Hence, the ZOH is generally followed by an LPF with a cut-off frequency of W. To compensate for the aperture effect, an amplitude equalizer of appropriate transfer function, as discussed earlier, will also be added in tandem with the ZOH and LPF.

From Eq. (8.55) which gives the transfer function of the ZOH, two things are quite clear:

1. ZOH gives a linear phase shift corresponding to a time delay of $T_s/2$.
2. Since the spectrum of the reconstructed signal is equal to $X_\delta(f) \cdot H(f)$ for $|f| \le W$, and since $H(f)$ is a sinc function while $X_\delta(f) = X(f)$ for $|f| \le W$ when $f_s \ge 2W$, it follows that the reconstructed signal is a distorted version of $x(t)$. As mentioned earlier, we make use of an amplitude equalizer to reduce this distortion.

Example 8.9 An L-section RC low pass filter with a 30 dB cut-off frequency f_c is used for band-limiting a signal which is to be sampled at a frequency f_s, what is the minimum value of f_s if the response to the aliased component at the edge of the pass band, i.e., at f_c is to be at least 30 dB below the response at f_c?

Solution For L-section RC low pass filter, the transfer function is

$$H(f) = \frac{1}{1 + j(f/f_c)}$$

where f_c = 3 dB cut-off frequency = $\dfrac{1}{2\pi RC}$

$$\text{(Response at } f_c) = \frac{1}{1+j} \qquad \therefore |H(f)|\Big|_{f=f_c} = \frac{1}{\sqrt{2}}$$

Now, referring to Fig. 8.32, in which f_p is now f_c, response at $(f_s - f_c)$ which appears as a frequency f_c because of aliasing, is given by

$$|H(f)|\Big|_{f=(f_s-f_c)} = \frac{1}{1 + j\left(\dfrac{f_s - f_c}{f_c}\right)} = \frac{1}{1 + j(x-1)}$$

where $x \underline{\Delta} \left(\dfrac{f_s}{f_c}\right)$

$$\therefore \qquad \frac{|H(f)|\Big|_{f=f_c}}{|H(f)|\Big|_{f=(f_s-f_c)}} = \left(\frac{\dfrac{1}{\sqrt{2}}}{\left(\dfrac{1}{\sqrt{1+(x-1)^2}}\right)}\right)$$

$$\therefore \qquad 10\log_{10}\left(\frac{\dfrac{1}{2}}{\left(\dfrac{1}{1+(x-1)^2}\right)}\right) \ge 30 \quad \text{or} \quad \frac{1+(x-1)^2}{2} \ge 10^3$$

i.e., $x \ge [\sqrt{2000-1} + 1] = 45.2$

$\therefore$ $f_s \ge 45.2 f_c$

8.8.2 Sampling of Band Pass Signals

When we discussed the low pass sampling theorem, we had seen that a signal band limited to W Hz has to be sampled, at least, at $2W$ samples per second if the analog signal is to be reconstructed, without distortion, from its samples, i.e., if no aliasing is to occur. This however, is true only for low pass signals, i.e., signals for which $X(f) = 0$ for all frequencies such that $|f| \geq W$, where W is some finite frequency. There are signals for which this rule does not apply, for instance, in the case of band pass signals, i.e., signals for which $X(f) = 0$ for all frequencies outside the range $f_1 \leq f \leq f_2$ where $f_1 \neq 0$. For this class of signals, while there will be no aliasing if $f_s > 2f_2$, there might be no aliasing even if $f_s < 2f_2$, provided f_s satisfies certain conditions. These conditions are well spelt out in what is called the 'Band Pass Sampling Theorem', which we state below.

Band pass sampling theorem Let a band pass signal, $x(t)$, have a spectrum as shown in Fig. 8.37.

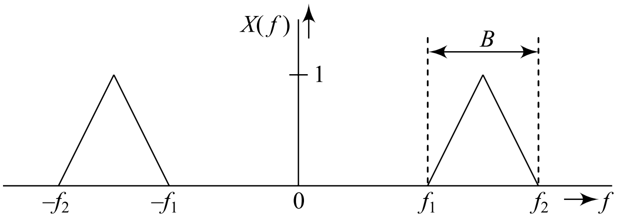


Fig. 8.37 *Spectrum of the band pass signal x(t)*

Then $x(t)$ can be recovered without any error whatsoever from its samples $x(kT_s)$ taken at regular intervals of T_s if the sampling rate f_s is such that

$$f_s = \frac{1}{T_s} = \frac{2f_2}{m} \tag{8.56}$$

where m is the largest integer not exceeding f_2/B.

It must be noted that sampling frequencies higher than what is given in Eq. (8.56) may not always permit recovery of $x(t)$ without distortion (i.e., they may not be able to avoid aliasing) unless $f_s > 2f_2$. If $x(t)$ is ideally sampled, i.e., using impulses, the signal $x(t)$ can be recovered from its samples by an ideal band pass filter with transfer function $H(f)$ given by

$$H(f) = \begin{cases} 1 & \text{for} \quad f_1 < |f| < f_2 \\ 0 & \text{otherwise} \end{cases} \tag{8.57}$$

In fact, as stated in the sampling theorem, the required sampling rate for a band pass signal depends upon 'm', i.e., on (f_2/B). This relationship is depicted in Fig. 8.38.

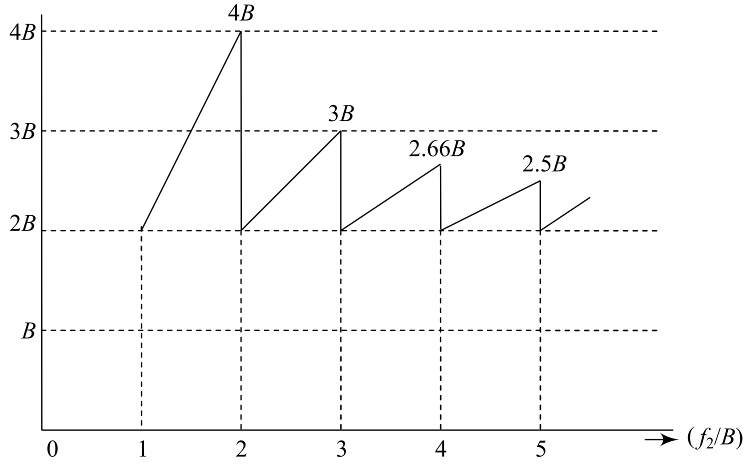


Fig. 8.38 *Relationship between (f_2/B) and the minimum sampling frequency*

As (f_2/B) increases indefinitely, the minimum sampling frequency approaches $2B$. Further it can be shown that if $f_s > 2f_2$, there will not be any aliasing and perfect reconstruction is possible. Also, if $f_2 = kB$ where k is an integer, a sampling rate $f_s = 2B$ would suffice and will not produce any aliasing.

Example 8.10 A band pass signal has a spectrum as depicted in Fig. 8.39. What is the minimum sampling frequency that can be used? By sketching the $X_\delta(f)$, show that no aliasing takes place when this sampling frequency is used.

Solution Here, f_2, the highest frequency component is 25 kHz. If we use the low pass sampling theorem, a minimum sampling frequency of 50 kHz would be needed.

However, it is not necessary to use such a high sampling frequency. Since it is a band pass signal, a sampling frequency equal to f_s would suffice, where

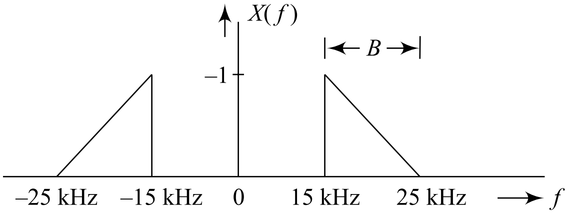


Fig. 8.39 *Spectrum of the band pass signal x(t)*

$$f_s = \frac{2 \times f_2}{m}$$

where m is the largest integer less than $25/10 = 2.5$

Thus, $m = 2$
$$f_s = \frac{2 \times 25 \text{ kHz}}{2} = 25 \text{ kHz}$$

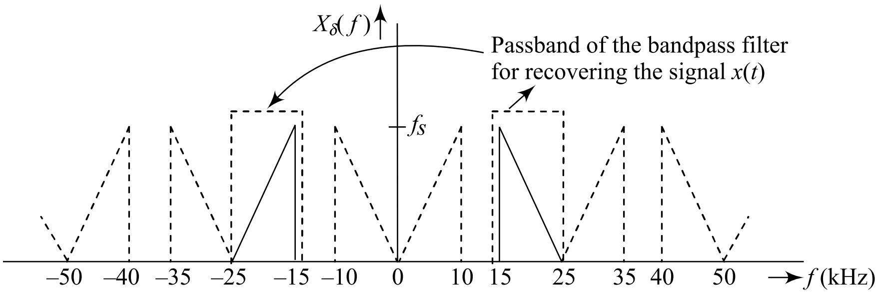


Fig. 8.40 *Spectrum of the sampled version of x(t) with f_s = 25 kHz*

8.8.3 Quadrature Sampling of Band Pass Signals

In the previous section, we had seen that a band pass signal centered on f_c for which $X(f) = 0$ for all frequencies outside the range $f_1 \le f \le f_2$ can be directly sampled at a frequency $f_s > 2f_2$ without aliasing, although aliasing may not occur for $f_s < 2f_2$ also provided f_s satisfies certain conditions. These conditions were, of course, spelt out by the band pass sampling theorem.

Recall that a band pass signal $x(t)$ centered on f_c may be represented by the 'Inphase' and 'Quadrature' component representation (see Section 3.4.5) as follows.

$$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \tag{8.58}$$

It was shown in Section 3.4.5 that if the band pass signal $x(t)$ has a bandwidth of $2W$ centered on f_c, the inphase component $x_I(t)$ and the quadrature component $x_Q(t)$ are low pass signals with a bandwidth of W for each. Since they are low pass signals band limited to W Hz, we may sample each of these without aliasing, by using a sampling frequency $f_s > 2W$. For this purpose, we shall first see how we can obtain $x_I(t)$ and $x_Q(t)$ from a given band pass signal, $x(t)$. Consider multiplying $x(t)$ by $\cos\omega_c t$. From Eq. (8.56), we get

$$x(t)\cdot\cos\omega_c t = x_I(t)\cos^2\omega_c t - x_Q(t)\sin\omega_c t\cdot\cos\omega_c t$$
$$= \frac{1}{2}x_I(t) + \frac{1}{2}x_I(t)\cos 2\omega_c t - \frac{1}{2}\sin 2\omega_c t$$

Hence, low pass filtering of $x(t)\cdot\cos\omega_c t$ will yield $\frac{1}{2}x_I(t)$. Similarly, low pass filtering of $x(t)\cdot\sin\omega_c t$ will yield $\frac{1}{2}x_Q(t)$. Thus, the inphase and quadrature components can be obtained, except for a scaling factor, by the arrangement shown in Fig. 8.41. Sampling these low pass signals at an $f_s \ge 2W$ will yield the inphase and quadrature component samples and there will be no aliasing.

This method of obtaining the inphase and quadrature component samples is referred to as 'quadrature sampling' The band pass signal $x(t)$, may be reconstructed from these samples as shown in Fig. 8.42.

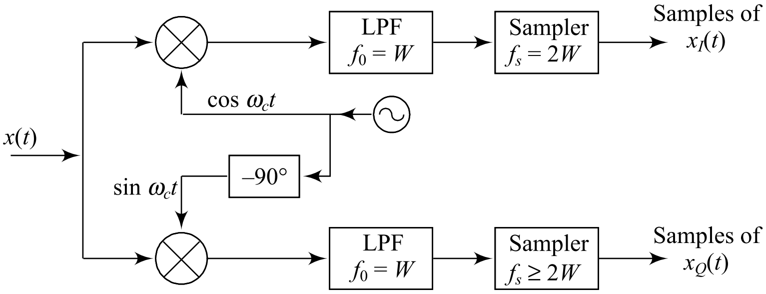


Fig. 8.41 *Quadrature sampling*

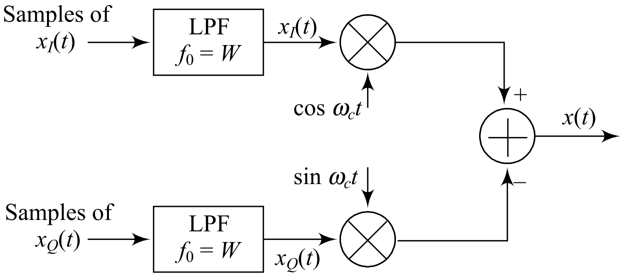


Fig. 8.42 *Reconstruction of a band pass signal from the inphase and quadrature samples*

8.9 PAM AND TIME DIVISION MULTIPLEXING

PAM signals may be generated straight away by flat-top sampling discussed in Section 8.7 (see Figs. 8.26 and 8.27).

The PAM signal of Fig. 8.43 is unipolar because the continuous-time signal $x(t)$, from which it is derived by flat-top sampling, is positive throughout. If that was not the case, there would have been zero amplitude pulses, or missing pulses in the PAM signal. Missing pulses cause synchronization problems in time division multiplexing and so have to be avoided. Since PAM is invariably used only for time division multiplexing, we shall consider only unipolar PAM.

A unipolar flat-top PAM signal may be analytically represented as

$$\underset{\text{PAM}}{x_s(t)} = \sum_k A_0[1 + mx(kT_s)]\,p(t - kT_s) \tag{8.59}$$

where $p(t)$ is a unit-amplitude flat-top pulse of width $\tau << T_s$ and having its leading edge at $t = 0$ as shown in Fig. 8.44; m is the modulation index and is such that $0 < m < 1$, $|x(t)| \le 1$ and A_0 is the unmodulated pulse amplitude. From this, it is clear that

$$1 + mx(kT_s) > 0 \tag{8.60}$$

for all k and that therefore it is ensured that $x_s(t)$ is a unipolar PAM signal.

If $x(t)$ has a spectrum as shown in Fig. 8.29, the spectrum of the unipolar PAM signal of Eq. (8.59) will be similar to what has been shown in Fig. 8.30 except that there will be impulses in the spectrum at $f = 0$,

$\pm f_s$, $\pm 2f_s$, …. As shown in Fig. 8.30 we may use a low pass filter for recovering $x(t)$; but now, to block the dc component (represented by the impulse at $f = 0$ in the spectrum) we have to use a blocking condenser too, and also an equalizer to reduce the aperture effect.

8.9.1 Time Division Multiplexing (TDM)

The low pass sampling theorem forms the basis for TDM. This theorem tells us that a band limited continuous-time signal can be completely recovered without any distortion, from its samples taken at regular intervals provided the sampling frequency is, at least, equal to the Nyquist rate. This means that we need not transmit the band limited continuous-time signal which engages the transmission channel all the time. Instead, we can transmit only the samples and reconstruct the continuous-time signal from the received samples. In this case, as the samples are separated in time by the sampling interval, the transmission channel is not engaged all the time; it is engaged only whenever a sample occurs. It is this fact that gives scope for the use of TDM. The interval between two successive samples of one message signal during which time the transmission channel is free, may be utilized to transmit the samples of each of the other message signals, i.e., we may interleave the samples of various message signals as shown, so that samples of different messages occupy different non-overlapping time slots.

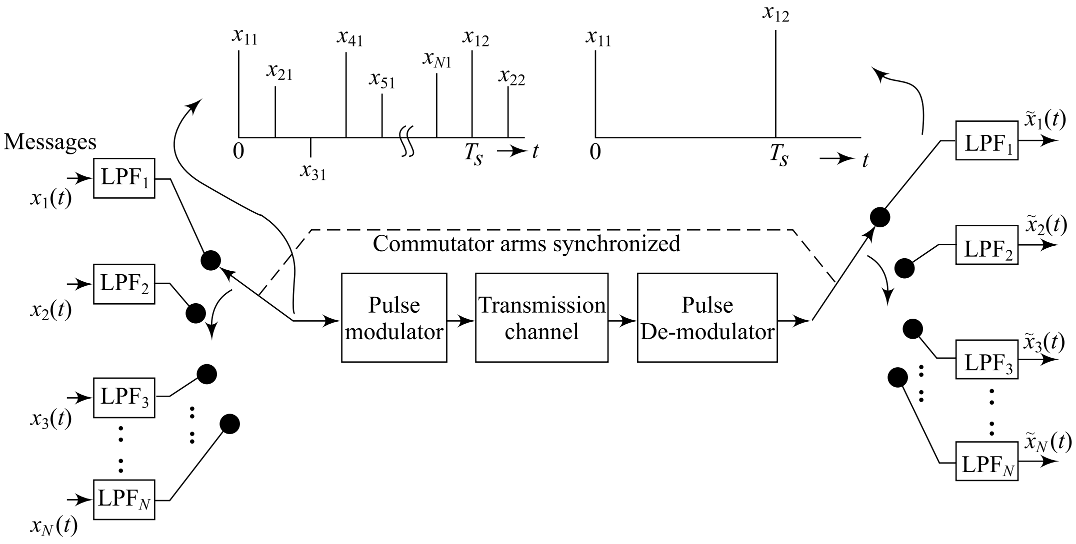


Fig. 8.43 *A PAM/TDM system*

messages $x_1(t)$, $x_2(t)$, …, $x_N(t)$ which are all to be time division multiplexed, are first band-limited using low pass filters. These band-limited signals are then sequentially sampled by the arm of the commutator at the sending-end. This commutator arm therefore carries samples of messages as shown, where, x_{11} is the first sample of the first message, x_{21} is the first sample of the second message, and so on. x_{12} is the second sample of the first message. These samples are fed to a pulse modulator and then transmitted over the channel. If the arms of the sending-end and receiving-end commutators are synchronized (neglecting the propagation delay caused by the transmission over the channel) x_{i1}, x_{i2}, x_{i3}, … which are all samples of $x_i(t)$, are fed to LPF$_i$ at the receiving-end which reconstructs the continuous-time signal and gives $\tilde{x}_i(t)$, an approximation to $x_i(t)$.

8.9.2 Bandwidth of TDM-ed Baseband PAM Signals

Assumed that messages $x_1(t), x_2(t), \ldots, x_N(t)$, each bandlimited to W Hz, have been flat-top sampled by narrow pulses (which can be approximated by impulses). Assume that this base-band TDM-ed signal is transmitted over a channel with finite bandwidth. For convenience, let us model the channel by an ideal LPF with a cut-off frequency of f_c Hz, where $f_c > W$.

When an impulse of strength I is fed as input at $t = 0$ to the channel, its output is a sinc pulse extending from $t = -\infty$ to $t = +\infty$, but having its peak at $t = 0$. Now the baseband TDM signal is a sequence of impulses regularly spaced at intervals of T_s/N and having strengths proportional to the sample values at the respective sampling instants.

Therefore, if $(T_s/N) = (1/2\,f_c)$ and if the arm of the de-commutator samples the successive sinc pulses exactly at the time instants marked as A, B, C, etc., in Fig. 8.44, then each sample so collected is directly proportional to a sample value of only one of the messages and so there will not be any cross-talk.

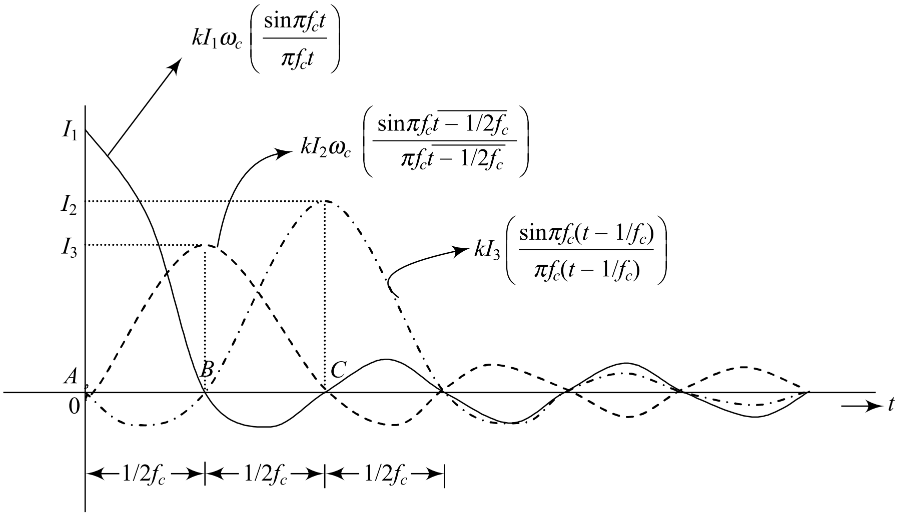


Fig. 8.44 *Response of the channel (ideal LPF) to successive samples (impulses) fed to it at t = 0, 1/2f$_c$, 1/f$_c$, etc.*

$$\therefore \qquad \frac{T_s}{N} = \frac{1}{2f_c}$$

Hence,

$$f_c = \frac{N}{2T_s}$$

Since $T_s \le \frac{1}{2W}$, we have

$$f_c \ge \frac{N}{\left(2 \cdot \frac{1}{2W}\right)} = NW$$

$$\therefore \qquad f_c \ge NW \tag{8.61}$$

Remark

(i) Therefore, the minimum bandwidth required for the baseband signal consisting of the TDM-ed samples of N identically band-limited messages is NW, where W is the bandwidth of each message.

(ii) Note that we would have got the same minimum bandwidth even if we had multiplexed these N message signals using Frequency Division Multiplexing (FDM), by making use of SSB subcarrier modulation (with no guard bands).

Example 8.11 24 different message signals, each band limited to 4 kHz are to be multiplexed and transmitted. What is the minimum bandwidth required for each of the following methods of multiplexing and modulation?

(a) FDM with SSB modulation, and (b) TDM with pulse amplitude modulation.

Solution

(a) *FDM with SSB modulation:* With SSB, each message channel occupies 4 kHz and the 24 messages can be accommodated in 24 non-overlapping frequency slots, each of width 4 kHz. Hence, total bandwidth required for the Frequency Division Multiplexed signal, is 24 × 4 = 96 kHz. It is assumed here that no guard bands have been provided. Since we are required to find the minimum bandwidth.

(b) *TDM with pulse amplitude modulation:* Equation (8.61) tells us that for TDM-PAM of N different messages, each of W Hz bandwidth, the minimum bandwidth required is NW Hz = 96 kHz.

Example 8.12 Signals $x_1(t)$, $x_2(t)$ and $x_3(t)$ are to be TDM-ed. $x_1(t)$ and $x_2(t)$ have a bandwidth of 10 kHz and $x_3(t)$ has a bandwidth of 15 kHz. Determine a commutator switching system so that all the three signals are sampled at their respective Nyquist rates.

Solution Since $x_1(t)$ and $x_2(t)$ have bandwidths of 10 kHz each, the Nyquist rate of sampling for them is 20 kilo samples per second. The Nyquist rate of sampling for $x_3(t)$ with 15 kHz bandwidth is 30 kilo samples/sec. So the commutator arrangement shown satisfies the requirement.

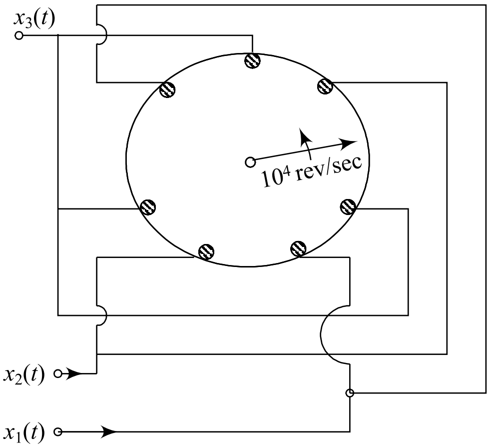


Fig. 8.45

8.9.3 Cross-Talk in PAM

As shown in Fig. 8.43, when we send a number of messages using PAM/TDM, we interleave the samples of the various messages. In such a situation, cross-talk can take place unless the communication circuit is carefully designed. We say that cross-talk is taking place if a sample of one message signal, say, $x_i(t)$ can influence the received sample value of a sample pertaining to some other message signal, say, $x_j(t)$, where, $j \neq i$. Cross-talk should be avoided, since it results in distortion of the message reconstructed from received samples.

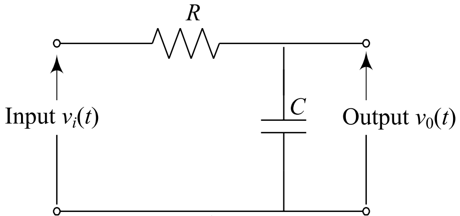
Cross-talk can occur due to the following reasons:

1. High-frequency limitation of the channel
2. Low-frequency limitation of the channel

8.9.4 Cross-Talk in PAM/TDM due to Frequency Limitations of the Channel

As mentioned earlier, in PAM/TDM, the samples of various messages are interleaved. A sample is represented by a narrow pulse whose width τ is very small compared to the sampling interval T_s and whose amplitude is proportional to the value of the sample. Samples of various channels (or messages) occur in non-overlapping time slots. Actually a time slot is an interval of time that can accommodate a pulse of width τ and also a guard time τ_g. Thus, each pulse of width τ is separated from its preceding pulse as well as the next pulse by guard times of τ_g on each side.

When a pulse is transmitted over a channel, it is affected in three ways. First, it is attenuated. Second, it is corrupted by noise. And third, it suffers some distortion because of the high-frequency and low-frequency deficiencies of the channel. Insofar as our interest is on cross-talk, the first two are of no consequence. We will, of course, be discussing the effect of noise separately later. The low and high frequency deficiencies of the channel cause a pulse to get distorted and also spill out into the guard time and some times even into the adjacent time slot. When that happens, it affects the value of the sample in the next time slot and thus causes cross-talk. We shall now see how these bandwidth deficiencies of the channel can result in cross-talk.



$$f_c = \frac{1}{2\pi RC}$$

= 3dB cutoff frequency

$RC = \tau_c << \tau$, the pulse width

Fig. 8.46 *A lowpass RC filter*

Cross-talk due to high frequency deficiency To study this, let us model our channel as a lowpass *R-C* filter with a time constant $RC = \tau_c << \tau$.

Figure 8.47(a) shows the waveform of the input pulse and Fig. 8.47(b) shows the waveform at the output of the channel. Since $\tau_c << \tau$, the pulse rises almost to the full value V (attenuation caused by the channel is ignored) within the time τ and from $t = \tau$, begins to fall exponentially towards zero as shown, again with a time constant of τ_c. It is obvious from Fig. 8.47(b) that cross-talk would be considerably reduced if τ_c is very small compared to even τ_g. From this figure, it is clear that a sample of message-1, transmitted in time slot-1 will, at the receiving-end, appear partly in time slot-2 also, thus causing cross-talk. The degree of cross-talk is generally specified by a 'cross-talk factor', denoted by K and defined as

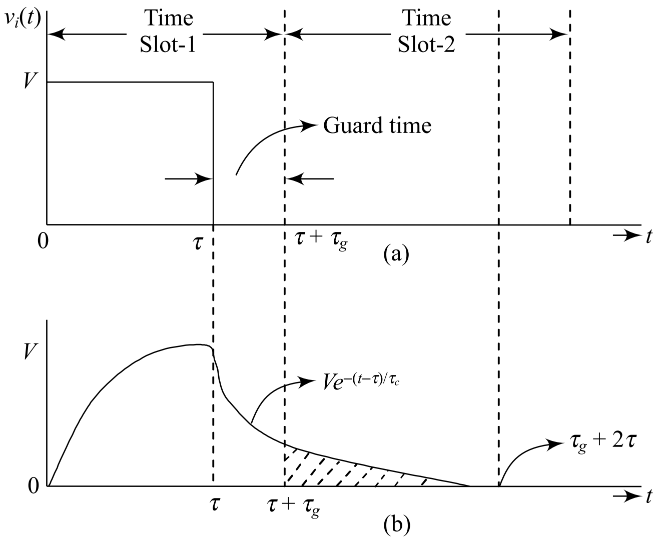


Fig. 8.47 *(a) Waveform of the transmitted pulse, (b) Waveform of the received pulse*

$$K \triangleq \frac{A_{12}}{A_2} \tag{8.62}$$

where A_{12} = shaded area in Fig. 8.47(b).

and A_2 = Area under the pulse transmitted in time slot-2.

Assuming that the sample values in time slots-1 and 2 are equal,

$$K = \frac{A_{12}}{A_1} \tag{8.63}$$

But

$$A_1 = V\tau$$

and

$$A_{12} = \int_{t=\tau+\tau_g}^{2\tau+\tau_g} Ve^{-(t-\tau)/\tau_c} dt$$

Putting $(t-\tau) \triangleq \lambda$, $dt = d\lambda$

and when $t = \tau + \tau_g,\ \lambda = \tau_g.$ Also, when $t = 2\tau + \tau_g,\ \lambda = \tau + \tau_g$

$$\therefore \quad A_{12} = \int_{\lambda=\tau_g}^{\tau+\tau_g} Ve^{-\lambda/\tau_c} d\lambda = V\tau_c e^{-\tau_g/\tau_c}[1 - e^{-\tau/\tau_c}]$$

However, since $\tau >> \tau_c,\ e^{-\tau/\tau_c} \approx 0$ and so we have

$$A_{12} \approx V\tau_c e^{-\tau_g/\tau_c} \tag{8.64}$$

Substituting this for A_{12} in Eq. (8.63), we get

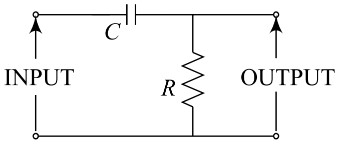
$$K = \frac{A_{12}}{A_1} = \frac{V\tau_c e^{-\tau_g/\tau_c}}{V\tau} = \left(\frac{\tau_c}{\tau}\right) e^{-\tau_g/\tau_c}$$

$$\therefore \quad \boxed{K = \left(\frac{\tau_c}{\tau}\right) e^{-\tau_g/\tau_c}} \tag{8.65}$$

From this equation, we find that

1. When $\tau_g >> \tau_c,\ K \cong 0$ and there is no cross-talk.
2. If it is specified that the cross-talk factor K should not exceed some particular value, we can determine (τ_g/τ_c) for given values of τ and either τ_c or τ_g.

Cross-talk due to low frequency deficiency We shall now investigate how the low-frequency limitation of the channel can cause cross-talk. For this purpose, to simplify matters, we shall once again model the channel as an *RC* filter, but of the high pass type, as shown in Fig. 8.48.



RC highpass Filter with
Time constant RC = $\tau_c << \tau$,
3-dB cutoff frequency = f_1
Where $f_1 = 1/(2\pi RC) = 1/(2\pi\tau)$

Fig. 8.48 *RC high pass filter used for modeling the channel*

As in the previous case, here too

$$A_1 \approx V\tau \qquad (\because \Delta << V) \tag{8.66}$$

The low frequency deficiency of the channel causes a 'tilt', or 'droop' denoted here by Δ and the received pulse waveform will be as shown in Fig. 8.49(b). Because the time constant τ_c is quite large compared to τ, the undershoot dies down rather slowly. Because of this, we may consider the shaded region to be a rectangle of area $\Delta\tau$.

Now,

$$\Delta = V - [Ve^{-t/\tau_c}]\Big|_{t=\tau} = V - Ve^{-\tau/\tau_c}$$

$$= V - V\left[1 - \frac{\tau}{\tau_c} + \frac{\tau^2}{\tau_c^2} - \ldots\right] \cong \frac{V\tau}{\tau_c}$$

(since $\tau << \tau_c$)

$$\therefore \quad A_{12} \cong \Delta\tau = \frac{V\tau^2}{\tau_c}$$

Hence,

$$K = \frac{A_{12}}{A_1} = \frac{V\tau^2}{\tau_c} \cdot \frac{1}{V\tau} = \frac{\tau}{\tau_c}$$

$$\therefore \quad \boxed{K = \frac{\tau}{\tau_c}} \qquad (8.67)$$

It may be noted that the cross-talk resulting from high frequency limitation of the channel might at the most affect the immediately adjacent channel only, because of the low time constant of the channel. But, because of the large time constant, the cross-talk arising from low-frequency deficiency will affect not just the immediately adjacent channel but even up to a *few* adjacent channels.

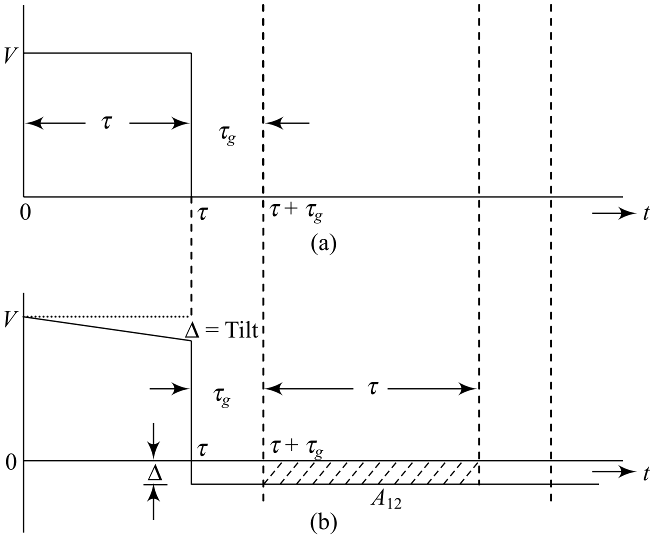


Fig. 8.49 *(a) Waveform of the transmitted pulse, (b) Waveform of the received pulse*

Example 8.13 12 speech signals, each band limited to 3.5 kHz, and sampled at a rate of 8 kHz, are to be transmitted as PAM signals over a certain channel using time division multiplexing. Assuming a guard time of half the pulse width, calculate the minimum bandwidth of the channel if the cross-talk factor (arising from high frequency limitation of the channel) between adjacent channels is less than 10^{-3}.

Solution f_s = sampling rate = 8000 samples/sec

$$\therefore \quad T_s\text{, the sampling period} = \frac{1}{f_s} = \frac{1}{8000} = 125\ \mu\text{s}$$

Since there are 12 message signals to be TDM-ed, the duration of the time slot for each, i.e., τ is given by

$$\tau = \frac{125}{12}\text{ micro-seconds} = 10.41\text{ micro-sec}$$

Since time slot includes pulse width and guard time, and since guard time is given to be half of the pulse width,

$$\text{Pulse width} = \tau = 10.41 \times \frac{2}{3} = 6.94 \text{ micro-sec}$$

$$\therefore \quad \text{guard time} = \tau_g = 10.41 - 6.94 = 3.47 \text{ micro-sec}$$

We know that $$K = \left(\frac{\tau_c}{\tau}\right) e^{-\tau_g/\tau_c}$$

In the above, we know K, τ and τ_g.

$\therefore$ solving for τ_c, we get τ_c = 0.75 micro-sec = RC

$$\therefore \text{ upper 3-dB cut-off frequency of the channel} = \frac{1}{2\pi RC} = \frac{10^6}{2\pi \times 0.75} = 212.314 \text{ kHz}$$

$\therefore$ the minimum bandwidth of the channel = 212.314 kHz.

8.10 PULSE TIME MODULATION SYSTEMS

We had seen that in pulse amplitude modulation, information regarding the sample value at any particular sampling instant is carried by the amplitude of a flat-top pulse located at that sampling instant. In the case of pulse time modulation, the information regarding the sample value at any particular sampling instant is carried not by the amplitude, but an 'interval of time' associated with a flat-top pulse. In the case of Pulse Width Modulation (PWM), the 'interval of time', is the width of the flat-top pulse located at that sampling instant; and in the case of Pulse Position Modulation (PPM), the 'time interval' is the displacement in time, given to the position of the flat-top pulse, relative to the sampling instant under consideration. Since PWM and PPM are closely related, they are generally clubbed together under the common name, 'Pulse Time Modulation'.

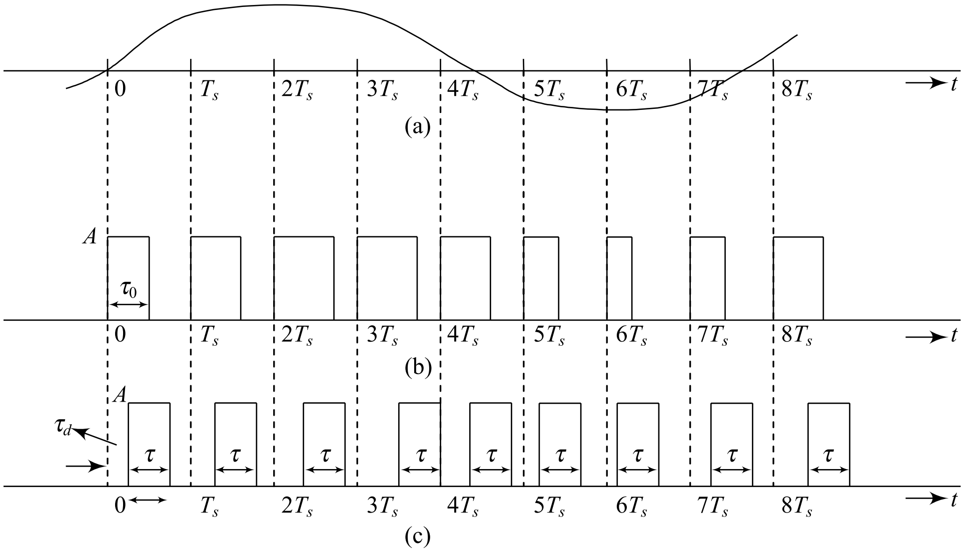


Fig. 8.50 *(a) Message signal x(t), (b) PDM signal, (c) PPM signal*

Assuming that $x(t)$ has been normalized so that $|x(t)| \le 1$, the width of the pulse at k^{th} sampling instant, i.e., at $t = kT_s$ in the PDM waveform shown in Fig. 8.50(b), is given by

$$\tau_k = \tau_0[1 + mx(kT_s)] \tag{8.68}$$

where τ_0 is the unmodulated pulse width, m is the modulation index and is such that $0 < m < 1$. Again in this case too,

$$1 + mx(kT_s) > 0$$

which ensures that there will not be any missing pulses. Of course, τ_0 must be so chosen that τ_k is always less that T_s. To ensure this, we choose

$$\tau_0 < \frac{T_s}{2} \tag{8.69}$$

Under certain simplifying assumptions, it can be shown that a PDM signal contains a dc component, the message signal, $x(t)$, and groups of phase modulated waves with the sampling frequency f_s and its harmonics as the carrier frequencies, and that as long as τ_0 is chosen as indicated in Eq. (8.69), the side frequencies of these phase modulated waves do not overlap much in the message signal bandwidth especially if $f_s >> W$, so that $x(t)$ can be recovered without much distortion from the PDM signal with a low pass filter followed by a blocking capacitor to reject the dc component.

Just as the pulse width in the case of PDM, as indicated in Eq. (8.68), contains a dc or constant pulse width plus a pulse width component that is directly proportional to the pertinent sample value $x(kT_s)$, in the case of PPM also, the delay in the occurrence of the pulse relative to the sampling instant also has two components, the dc component shown as τ_d in Fig. 8.50 and another component directly proportional to the pertinent sample value. Hence, we may write the expression for the instant τ_k at which the leading edge of the pulse appears, as

$$\tau_k = kT_s + \tau_d + t_0 x(kT_s) \tag{8.70}$$

where t_0 is a proportionality constant having units of seconds per volt.

8.10.1 Generation of PTM Signals

We shall now discuss briefly, a few methods for the generation of PTM signals.

One way of generating PDM and PPM signals by first generating PAM signal, is illustrated in Fig. 8.51. In this method, the PAM signal and an inverse ramp signal are generated synchronously, as shown. These two are then added and fed to a comparator whose triggering level is so adjusted that it is in the sloping portion of the sum waveform. The second crossing of the comparator trigger level with the sum waveform coincides with the trailing edge of the PDM wave and the leading edge of the PPM pulse. All these PPM pulses are of constant width.

For the generation of PDM and PPM signal, it is not necessary that one should first produce PAM, although the above method is based on such a procedure. We now give two more methods of generation of PDM and PPM and these methods do not need the generation of PAM first – they generate PDM and PPM directly from the message signal.

In the first of these two methods, as illustrated in Fig. 8.52, a periodic inverse ramp signal with a period T_s is added to the message signal and the sum signal is fed to a comparator whose triggering level is set to fall in the ramp portion of the sum signal. The leading edge of the PDM signal coincides with the first intersection of the comparator trigger and the vertical side of the inverse ramp. The trailing edge occurs at the instant at which the second intersection occurs. The leading edge of each PPM pulse coincides with the trailing edge of the corresponding PDM pulse, and these PPM pulses will be of the same amplitude and width.

The circuit diagram given in Fig. 8.53 gives yet another direct method of generation of PDM and PPM signals.

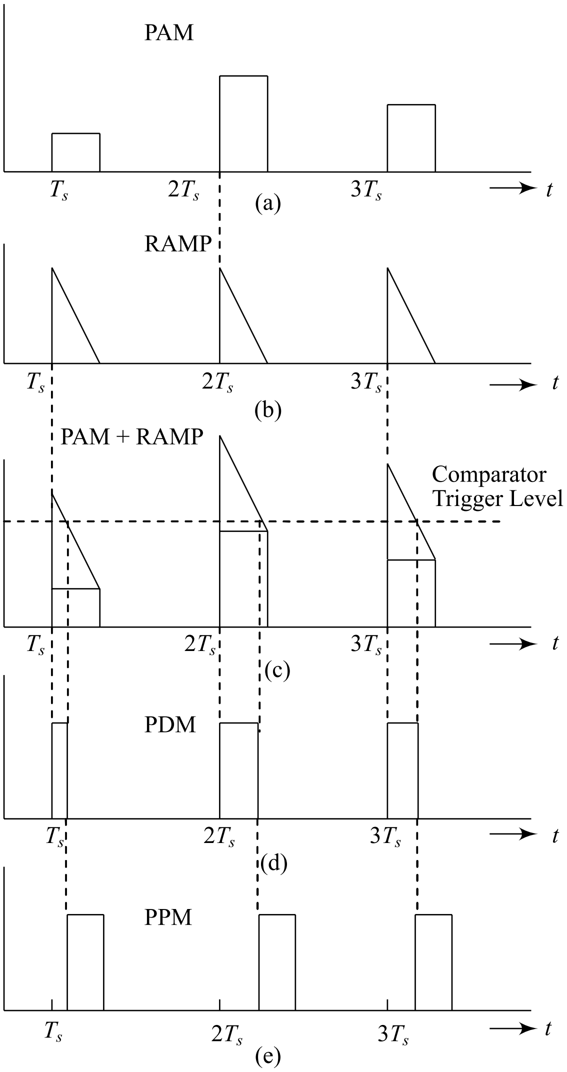


Fig. 8.51 *(a) PPM, (b) RAMP, (c) PPM + RAMP, (d) PDM, (e) PPM*

This circuit is an emitter-coupled one-shot or mono-stable multivibrator. In its stable state, transistor T_1 is in cut-off condition and T_2 is in conducting state. However, when a trigger pulse of sufficient amplitude is applied to its base, T_1 suddenly goes into conduction and T_2 is temporarily cut off. We know that the duration τ of the pulse that results at the collector of T_2 is linearly related to the bias applied to the base of T_1. This bias is, as may be seen from the figure, the sum of a fixed dc component, and the message signal. Thus, τ is linearly related to the amplitude of the message signal at the instant at which the trigger pulse is applied and since the trigger pulses are applied at regular intervals of T_s, we get PDM/PWM signal at the collector of T_2. By differentiating this PDM signal, and using the negative trigger pulses occurring at the location of the variable edge (of the PDM signal) for triggering another mono-stable multivibrator with a fixed bias for its normally cut-off transistor, one can obtain a PPM signal.

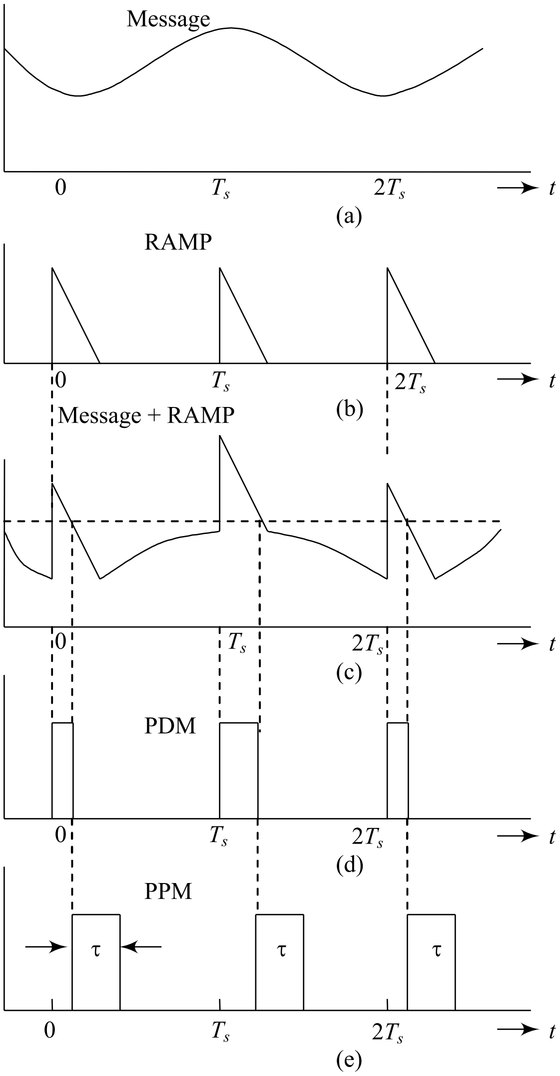


Fig. 8.52 *(a) Message signal, (b) Inverse ramp sequence, (c) x(t) + ramp, (d) PDM, (e) PPM*

8.10.2 Detection of PTM Signals

Earlier, while discussing the frequency components that make up a PDM signal, we have pointed out that recovery of the message signal from the PDM signal by directly low pass filtering is possible but it results in some distortion.

Apart from direct low pass filtering, there is another approach possible for recovery of the message signal from a PDM signal. This approach consists of first converting the PDM signal into a PAM signal from which the message signal may be recovered with very little distortion by low pass filtering and equalization. Actually, if the pulse width in this PAM signal is quite small, equalization may not be necessary at all; simple low pass filtering will suffice. This approach is applicable to the detection of PPM signals also. This method is illustrated in Fig. 8.54.

The PDM signal is first integrated and the value of the output of the integrator at the end of each PDM pulse is held till the next sampling instant, at which time the capacitor of the integrator is discharged suddenly and the integrator of the next pulse is allowed to start. We get the waveform shown in Fig. 8.54(b). To this we add locally generated constant amplitude pulse sequence having a period of T_s in such a way that the pulses sit over the pedestal portion. This waveform is then subjected to clipping with the clipper level so adjusted that it is above the level of the highest pedestal. The clipper output, shown in Fig. 8.54(d) is a PAM representation of the original PDM signal. For converting a PPM signal into a PAM signal we may first convert it into a PDM signal by generating pulses with their leading edges at the sampling instants and trailing edges at

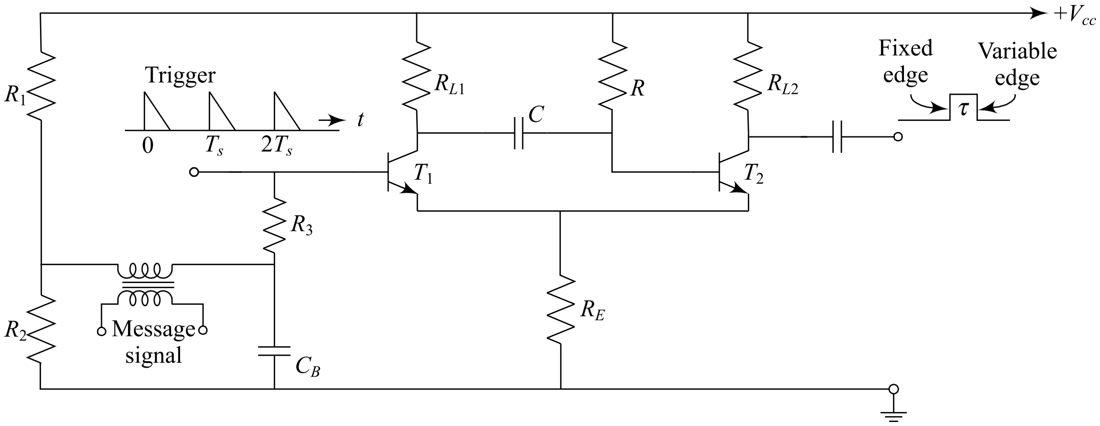


Fig. 8.53 *A circuit for generating PDM (PWM) signals*

the leading edges of the PPM pulses. Once the PAM signal is obtained, it can be low pass filtered to recover the original signal. If the locally generated pulse sequence has very narrow pulses, the resulting PAM signal also will have only narrow pulses. In that case, the distortion due to aperture effect will be negligible and no equalizer need be used after low pass filtering of the PAM signal.

8.10.3 Cross-Talk in PTM

Cross-talk can occur in PTM time division multiplexed systems too, just as it occurs in TDM-ed PAM systems because of the low frequency and high frequency deficiencies of the channel. However, one basic difference between the two should always be borne in mind. Pulse transmitted in one time slot extends at the receiving end into the following time slot (or time slots) because of the low frequency, or high frequency deficiency of the channel in both the cases (PAM and PTM). In PAM, cross-talk results from such an extension into the following time slot because of its effect on the *amplitude* of the pulse in that time slot. But in the case of PTM, such an extension causes cross-talk by influencing the *width* of the pulse in the following time slot in the case of PDM and the *position* of the pulse in the following time slot in the case of PPM.

Cross-talk due to high frequency limitation of the channel In Fig. 8.55, the transmitted pulses in the N^{th} and $(N + 1)^{th}$ time slots are shown by the solid line. t_1 represents the instant at which the *trailing edge* of the *unmodulated* transmitted pulse occurs, if it is PDM. If it is PPM, it represents the instant at which the *leading edge* of the *unmodulated* pulse occurs. t_2 represents the instant at which the trailing edge of the transmitted pulse occurs when pulse duration modulation is present. If the modulation is PPM, t_2 represents the position of the leading edge of the transmitted pulse with modulation. Figure 8.55(b) represents the received pulses in the two time slots when there is no modulation. Note that owing to the high frequency deficiency of the channel, there is distortion. However, there is no cross-talk. Figure 8.55(c) shows the two received pulses when modulation is present in channel-1. It is, of course, assumed for the purpose of drawing this figure, that there is no modulation in channel-2. Note that the cross-talk has caused a timing error Δt in the time slot-2. Since time translates into amplitude at the time of detection of PDM and PPM signals, this timing error creates distortion of the received signal in channel-2.

Cross-talk due to low frequency deficiency In Fig. 8.56(b), we find that if the high frequency response of the channel is very good, the low frequency deficiency of the channel does not lead to any timing error because the rising and falling edges of the received pulses will also be absolutely vertical. Hence, in such a case, the low frequency deficiency does not cause any cross-talk.

However, if the high frequency response of the channel is not very good, its low frequency deficiency can cause timing error and cross-talk. This is due to both the tilt in the top of the received pulse owing to the low frequency deficiency and also the finite rise time and fall time of the received pulse owing to the 'not-so-good' high frequency response.

8.10.4 Synchronization in TDM-ed PAM and PTM

In a TDM system, arrangements must be made for proper synchronization of the commutator at the transmitting end and the de-commutator at the receiving-end. During each time slot, at the transmitting-end, the channel must be connected to the particular message channel the sample value of which must be transmitted during that time slot. This is accomplished by the use of a clock signal (at the transmitter) from which the necessary gate signals are derived. Similarly, at the receiving-end, the channel must be connected to the baseband recovery circuits of the various message channels in a sequential manner. Again, this also is accomplished with the help of a clock signal generated at the receiving-end. For proper functioning of the TDM system, it is necessary that these two clocks, one at the transmitting-end and the other at the receiving-end, work in synchronism.

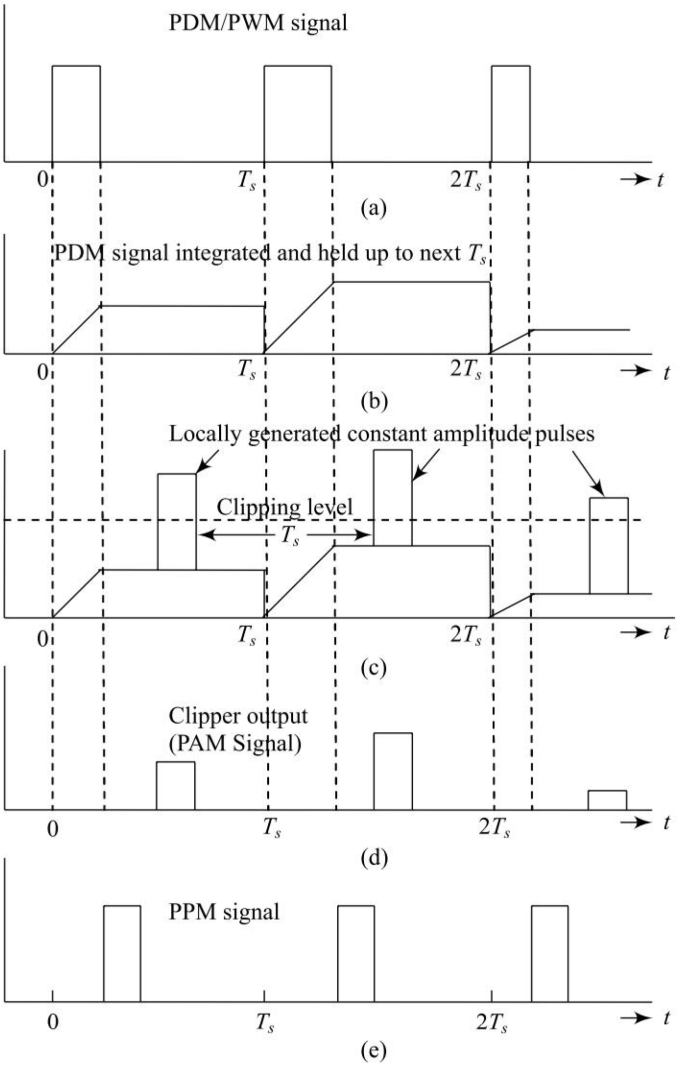


Fig. 8.54 *(a) PDM/PWM signal, (b) PDM signal integrated and held up to next sampling instant, (c) Locally generated constant amplitude pulses added on the pedestal, (d) Output of the clipper, (e) PPM signal*

For this purpose, a special pulse called synchronization pulse distinguishable from the normal signal pulses is transmitted along with the signal pulses, but in a separate time slot, at regular intervals. Hence, if there are N message signals TDM-ed, in addition to the N time slots required per frame, an extra time slot is provided to accommodate the synchronization pulse. A frame is said to be completed when one sample of each one of all the messages is sent in a sequential manner. In case the frame time is too long because the number of messages to be TDM-ed is very large, more than one synchronization pulse will have to be included per frame, in order to ensure that the receiving-end clock does not go out of synchronization.

In the case of PAM, the synchronization pulse is made to have much larger amplitude than any of the signal pulses. Since the time of arrival of the synch pulse is important, rather than its amplitude, the instant at which the received synch pulse crosses the set level of a comparator, is used for synchronizing.

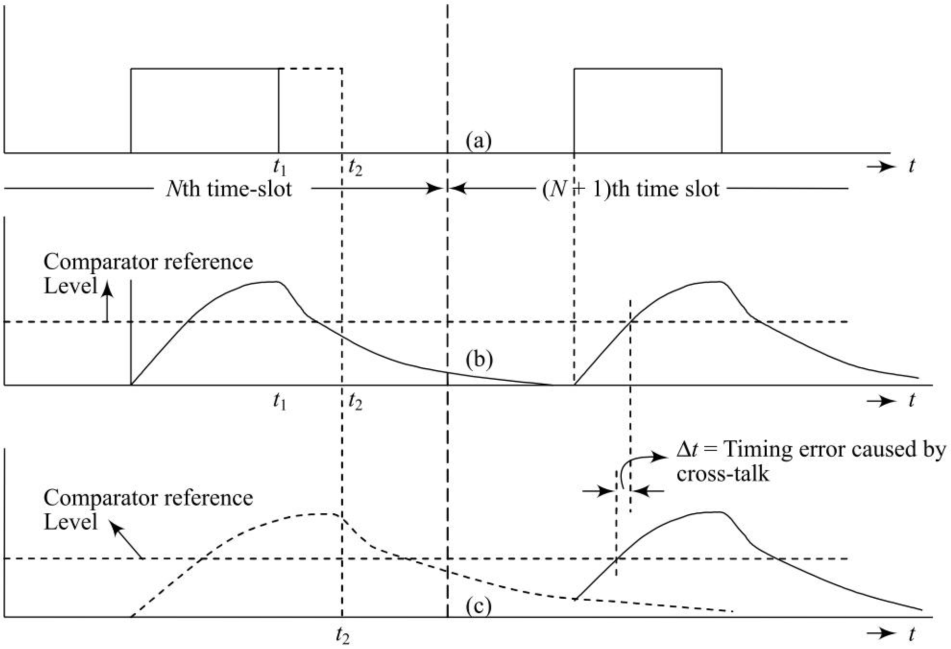


Fig. 8.55 *(a) Transmitted PTM signal without modulation, (b) Received signal without modulation, (c) Received signal when modulation is present*

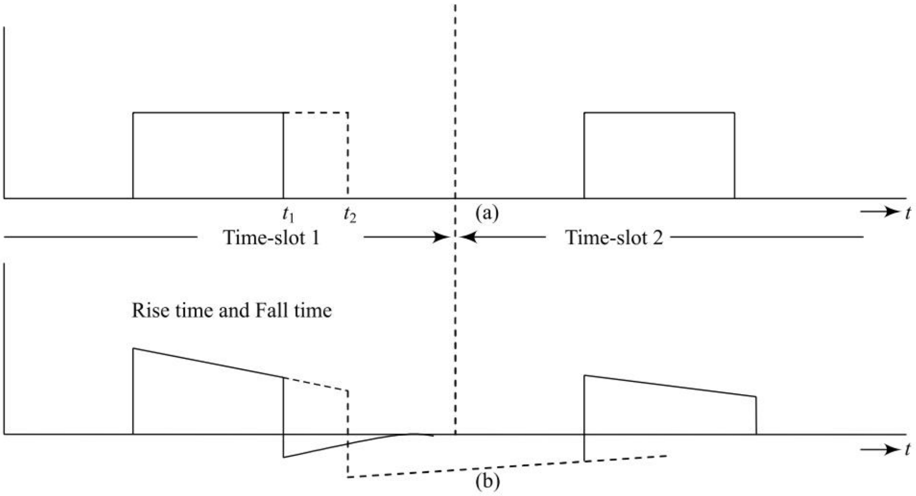


Fig. 8.56 *(a) Transmitted pulses (dashed line shows change due to modulation), (b) Received pulses (assuming excellent high frequency response for channel)*

For PDM, the synch pulse is made to have a much larger width than any of the signal pulses.

The pulses shown in Fig. 8.57(a) are inverted and as shown in Fig. 8.58, applied to the base of a transistor which acts as a switch. In the absence of any external input, the transistor conducts heavily passing saturation current. Hence, the capacitor *C* connected across it will have no voltage across it. But when these pulses are inverted and applied, for each pulse duration, the transistor goes into the cut-off state and so the capacitor

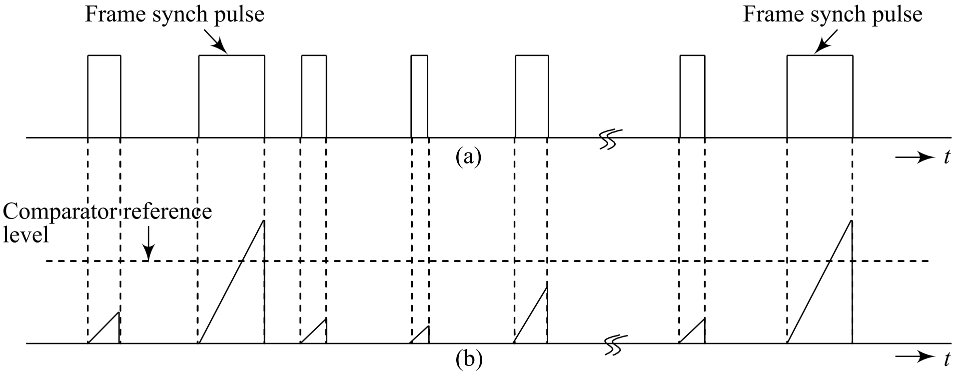


Fig. 8.57 *(a) TDM-ed PDM pulses along with one synch pulse per frame, (b) Output of an integrator circuit*

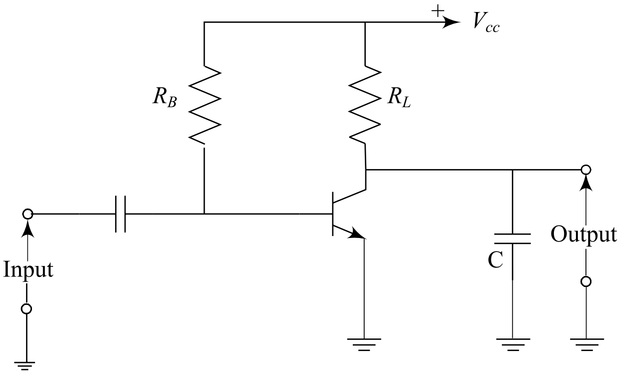


Fig. 8.58 *Saw-tooth waveform generating circuit*

charges from V_{cc} through the collector load resistance R_L. When a pulse ceases to exist, the transistor conducts heavily and the capacitor discharges quickly. Thus, we get a number of saw-tooth pulses, as shown in Fig. 8.57(b). The synch pulses have a very large width and so the corresponding saw-tooth pulses will have very large amplitudes compared to the amplitudes of the saw-tooth pulses produced by the signal pulses. A comparator, whose reference level is adjusted to be far above the smaller saw-tooth pulses, produces an output trigger pulse whenever the large saw-tooth pulse produced by a frame synchronization pulse crosses the reference level. This trigger is used for synchronization of the clock that controls the operations at the de-commutator.

8.10.5 Comparison between TDM and FDM

1. TDM hardware is much simpler than that required for FDM, as there is no need for subcarrier modulators, band pass filters, etc.
2. In FDM cross-talk occurs mainly due to non-linear cross-modulation and imperfect band pass filtering. In TDM, cross-talk is mainly due to inadequate transmission bandwidth of the channel.
3. It is much easier to time division multiplex baseband signals having widely different bandwidths, whereas it is not that easy in the case of FDM.
4. Short-term fading of the transmission channel affects *all* the message channels in the case of FDM. However, in the case of TDM, *only a few* sample pulses transmitted during the occurrence of the fading will be affected, causing slight distortion only in the few affected channels.

Example 8.14 Five low pass message signals, each of bandwidth 2 kHz are to be sampled at 5 kHz and PAM/TDM-ed using pulses of width 20 μs. What is the guard time available?

Solution Since f_s = Sampling frequency = 5×10^3 samples/sec,

$$T_s = \frac{1}{f_s} = \frac{1}{5 \times 10^3} \text{ sec} = 0.2 \text{ m.sec}$$

Hence, the interval between successive samples of a particular message signal is 0.2 m-sec. If five such message signals are to be TDM-ed, it means that five pulses, each of width 20 µs (as specified) are to be interleaved in the interval between two successive samples of any one message signal, as shown in Fig. 8.59.

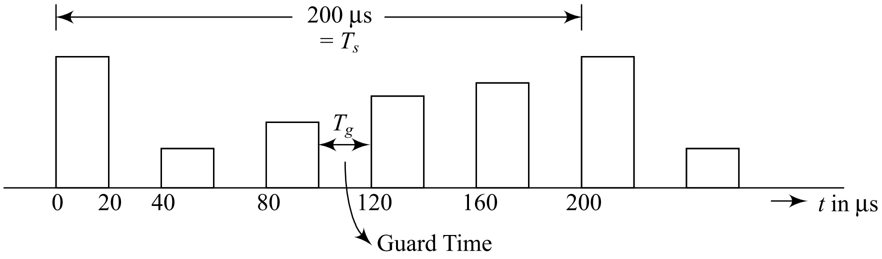


Fig. 8.59

Therefore, as can be seen from the figure, the guard time between adjacent pulses in the PAM/TDM-ed signal is 20 µs.

Example 8.15 A TDM signal is shown in Fig. 8.60. Show that it is possible to detect it using a time-averaging low pass filter.

Solution

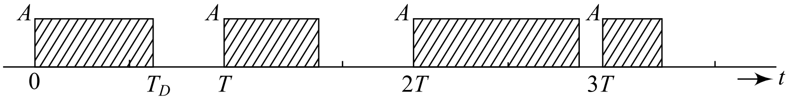


Fig. 8.60

A time-averaging filter takes the average value over each period T.

$\therefore$ output of the filter at $t = T$ is given by

$$V_0(t) = \frac{1}{T}\int_0^{T_D} A\,dt + \frac{1}{T}\int_{T_D}^{T} 0\,dt = \frac{AT_D}{T}$$

$\therefore$ $V_0(T)$ is proportional to T_D, the width of the pulse. Hence, a time-averaging filter can be employed to detect a TDM signal.

8.11 NOISE PERFORMANCE OF ANALOG PULSE MODULATION SYSTEMS

Before we proceed to a study of the noise performance of PAM and PTM, it would be proper to have a brief discussion on certain aspects of baseband pulse transmission.

We know that a rectangular pulse of width τ seconds will, in general, have a spectrum extending from dc up to very high frequencies. Smaller the value of τ, the width of the pulse, more will be the high frequency content in the spectrum. Therefore, one question that immediately arises in one's mind, is: 'How much transmission bandwidth is to be provided for pulse transmission'? The answer to this question depends upon what our requirement is. If we would like the pulse to be reproduced at the receiving-end of the channel with very little distortion, i.e., if our requirement is to preserve the pulse shape; the channel should produce a phase shift that is proportional to frequency and should have a bandwidth, B_T, which is very large. In this case

$$B_T >> \frac{1}{\tau} \tag{8.71}$$

On the other hand, if our interest is only to detect the presence of a pulse, or measure the amplitude of the received pulse, a bandwidth B_T given by

$$B_T \geq \frac{1}{2\tau_{\min}}, \tag{8.72}$$

would be sufficient, where $\tau_{\min}$ is the smallest *output* pulse duration.

Yet another type of scenario in which we will be interested is one wherein two closely spaced rectangular pulses which have been transmitted over a channel may have to be resolved, or identified as two separate pulses when they arrive at the receiving-end. We will be interested in knowing what minimum bandwidth the channel should have for a given separation $\tau_{\min}$ between the two pulses, each of which is of width τ sec. It has been found that the minimum spacing is to be at least equal to τ and that the bandwidth required with that spacing is

$$B_T = \frac{1}{2\tau} \tag{8.73}$$

For this bandwidth, if the spacing is reduced below τ, or for the spacing of τ if the bandwidth of the channel is less than the value specified by Eq. (8.73), there will be considerable overlap between the two output pulses and it will be difficult to recognize them as two separate pulses.

In case we are interested in measuring the time of occurrence (i.e., the position) of an output pulse relative to some reference instant, the rise time and/or fall time of the output pulse become important. We then fall back on the well-known relationship between the rise time of the output pulse and the channel bandwidth, and write

$$B_T \geq \frac{1}{2\tau_{r_{\min}}} \tag{8.74}$$

when $\tau_{r_{\min}}$ is the minimum rise time of the output pulse.

We are now ready to study the noise performance of PAM and PTM. In connection with this study, the following remarks are very pertinent, as they put the derivations in the proper perspective, and so are to be borne in mind.

1. The pulse-modulated signals (PAM and PTM) that we consider are *baseband signals and have no high frequency carrier*.
2. Because there is no carrier modulation, the *noise entering the receiving system is low pass noise and not band pass noise* as was the case when we considered the noise performance of continuous-wave modulations like AM and FM.
3. Whereas in CW modulation systems we were interested in receiving the transmitted message waveform without much distortion, in the pulse modulation case, our interest is limited to measuring the amplitude, or the time of arrival, of the received pulse rather than ensuring that the received pulses are replicas of the corresponding transmitted pulses.
4. We may, at the receiving-end, know the shape of the transmitted pulse in advance.

A continuous-time signal $x(t)$, band limited to W Hz is the modulating signal which has been sampled at regular intervals of $T_s = \frac{1}{2W}$ and the sample values are represented by a PAM, PDM, or a PPM signal and this baseband pulse-modulated signal is transmitted to the receiver through a channel characterized ideally by additive white noise of two-sided PSD equal to $\eta/2$ W/Hz. So, whatever may be the actual method adopted by the receiver for de-modulation, the demodulation process may ideally be visualized as one of converting back the pulse modulated signal (PAM, PDM or PPM) plus the additive noise into a sequence of weighted impulses (corresponding to ideal sampling). The original message plus noise will be obtained when this impulse train is passed through an ideal LPF which acts as the reconstruction filter. Thus, we shall use the model shown in Fig. 8.61 for studying the noise performance of analog pulse modulated systems.

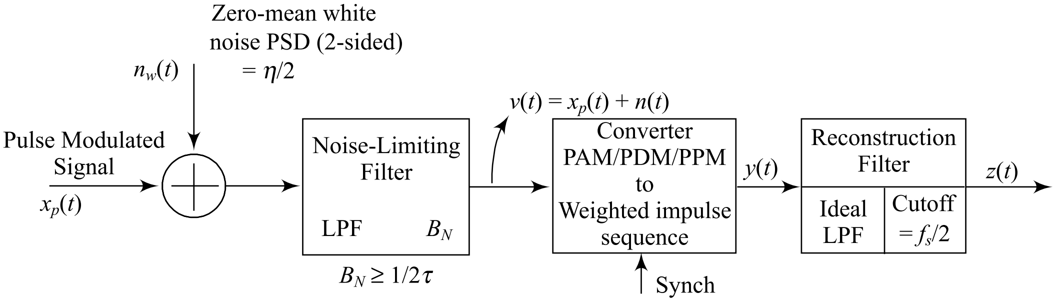


Fig. 8.61 *Model for an analog pulse modulation receiver*

We shall now derive the expression for destination signal-to-noise ratios for PAM, PDM and PPM systems by making use of the above model.

8.11.1 Pulse Amplitude Modulation

Pulse amplitude modulated signal plus white noise tries to enter the receiver. The noise limiting low pass filter has a cut-off frequency $B_N \geq \frac{1}{2\tau}$, where τ is the pulse width of the PAM signal. Its output therefore is

$$v(t) = x_p(t) + n(t) \tag{8.75}$$

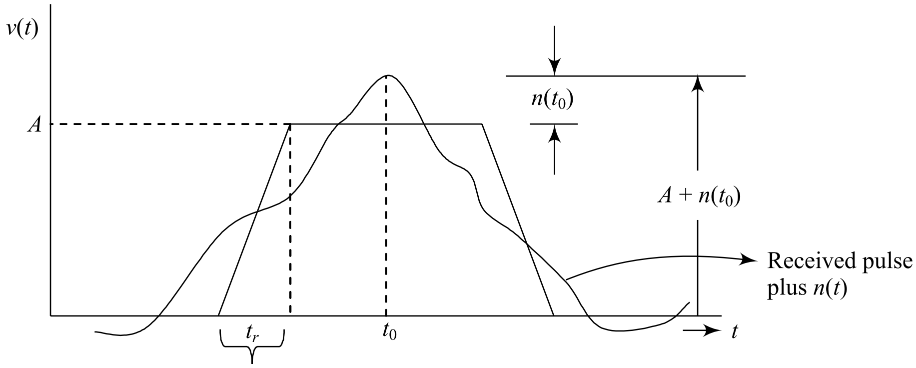


Fig. 8.62 *Received pulse plus low pass filtered noise pulse*

Because of the finite rise time and fall time, the pulse amplitude is generally measured near the middle of the time slot at some instant such as t_0. So the measured value is

$$v(t)\big|_{t=t_0} = v(t_0) = A + n(t_0) = A + \in_0 \tag{8.76}$$

where $\in_0$ represents the amplitude error. This error has a variance equal to the average power in $n(t)$, the filtered (white) noise. Therefore, it is given by

$$\sigma_0^2 = \overline{n^2} = \eta B_N \tag{8.77}$$

The output of the converter, $y(t)$, which is a train of weighted impulses spaced T_s sec apart may therefore be written as

$$y(t) = \sum_k [A_c m x(kT_s) + \in_k]\delta(t - kT_s) \tag{8.78}$$

where m is the modulation index and $\in_k$ is the error in the measurement of the amplitude of the k^{th} received pulse because of noise. The reconstruction filter, assumed to be an ideal LPF with a cut-off frequency of $f_s/2$, a pass band gain of T_s (this is purely arbitrarily chosen, just for convenience) and zero delay, will given an output $z(t)$ which may be written as

$$\begin{aligned} z(t) &= \sum_k [A_c mx(kT_s) + \in_k]\text{sinc}[f_s(t-kT_s)] \\ &= \sum_k A_c mx(kT_s)\text{sinc}[f_s(t-kT_s)] + \sum_k \in_k \text{sinc}[f_s(t-kT_s)] \end{aligned} \tag{8.79}$$

The first term in the RHS of the above equation gives the output signal component and the second term gives the output noise component. Hence, we may write

$$z(t) = A_0 mx(t) + n_0(t) \tag{8.80}$$

As shown in Fig. 8.55, $B_N \geq \frac{1}{2\tau}$. Since $\tau << T_s$, it follows that

$$B_N > \frac{1}{T_s} \tag{8.81}$$

Hence, the values of the error, i.e., $\in_k$'s can be considered to be uncorrelated. Further, since the channel noise has been assumed to be zero mean and since the noise limiting filter is an LTI system, $n(t)$ is also zero mean. Hence, $\in_k$s have a zero mean and are uncorrelated. Thus, the average noise power at the destination, viz., N_D is given by

$$N_D = \overline{n_D^2(t)} = \overline{\in_k^2} \tag{8.82}$$

But, we have already shown that the variance of the measurement error (see Eq. (8.77)) is equal to the average noise power at the output of the noise limiting filter and that this is given by ηB_N.

$\therefore$
$$N_D = \eta B_N \tag{8.83}$$

Now, to determine the average signal power at destination, we proceed as follows.

$$\left.\begin{array}{l}\text{Average energy per pulse in} \\ \text{the PAM signal}\end{array}\right\} = \tau A_0^2 \overline{[1+mx(kT_s)]^2} = E_p \tag{8.84}$$

Number of pulses per second in the PAM signal $= f_s$ (8.85)

$\therefore$ received average signal power $= S_R = \tau A_0^2 \overline{[1+mx(kT_s)]^2} \cdot f_s$ (8.86)

But, $\overline{[1+mx(kT_s)]^2} = 1 + m^2 \overline{x^2(t)}$ (8.87)

since $x(t)$ is assumed to be of zero mean so that

$$m\overline{x(kT_s)} = 0 \tag{8.88}$$

$\therefore$ we may rewrite Eq. (8.86) as follows:

$\therefore$
$$S_R = A_0^2 f_s \tau \left[1 + m^2 \overline{x^2}\right] \tag{8.89}$$

From Eq. (8.80), average signal power at the destination is given by

$$S_D = A_0^2 m^2 \overline{x^2(t)} \tag{8.90}$$

$\therefore$ using Eqs. (8.83) and (8.90), we may write

$$\left(\frac{S}{N}\right)_D = \frac{m^2 \overline{x^2(t)} \cdot A_0^2}{\eta B_N}$$

But $B_N \geq \frac{1}{2\tau}$

$\therefore$ the minimum value of $B_N = \frac{1}{2\tau}$ and this gives the maximum destination SNR for a given modulation index, m.

$$\left(\frac{S}{N}\right)_{D;\max} = \frac{2A_0^2 m^2 \overline{x^2(t)}\cdot\tau}{\eta} \tag{8.91}$$

But, as per Eq. (8.89), τ is given by

$$\tau = \frac{S_R}{A_0^2 f_s\left[1+m^2\overline{x^2(t)}\right]}$$

Substituting this for τ in Eq. (8.91),

$$\left(\frac{S}{N}\right)_{\substack{D;\max\\ \text{PAM}}} = \frac{WA_0^2 m^2\overline{x^2(t)}\cdot 2(S_R/\eta W)}{A_0^2 f_s\left[1+m^2\overline{x^2(t)}\right]} = \left[\frac{m^2\overline{x^2(t)}}{1+m^2\overline{x^2(t)}}\right]\left(\frac{2W}{f_s}\right)\gamma \tag{8.92}$$

This can be further maximized by choosing $m = 1$. Then

$$\left(\frac{S}{N}\right)_{\substack{D;\max\\ \text{PAM}}} = \left[\frac{\overline{x^2}}{1+\overline{x^2}}\right]\left(\frac{2W}{f_s}\right)\gamma \tag{8.93}$$

Since $\overline{x^2}$ can at the most be 1 (since $|x(t)| \le 1$) and since $f_s \ge 2W$, it follows that $(S/N)_D$ is less than or equal to $(\gamma/2)$. It is therefore at least 3 dB inferior to baseband transmission. However, this has not much significance, since PAM, when used, is not for its good noise performance but only for its simplicity and for time division multiplexing.

8.11.2 Noise Performance of PDM and PPM

In the case of PDM/PWM, information regarding the k^{th} sample value of the message signal is incorporated into the width τ_k of the k^{th} pulse.

$$\tau_k = \tau_0[1 + mx(kT_s)]$$

Here, τ_0 is the width of the unmodulated pulse. The amplitude of the pulse is constant and equal to A.

In the case of PPM, information about the value of the k^{th} sample is incorporated into the delay t_k, in the arrival of the leading edge of the k^{th} pulse.

$$t_k = kT_s + \tau_d + t_0 x(kT_s)$$

where τ_d represents the delay when the sample value kT_s is zero. So, a PDM receiver has to measure the pulse duration time while a PPM receiver has to measure the pulse arrival time. Since the leading and trailing edges of the received pulse will be having finite slopes and are superimposed by additive noise, the exact instant at which the pulse begins or ends will not be easy to identify. Hence, the instant t_0 at which the pulse attains say 50% of its final value, i.e., a value of $A/2$, is generally identified. An error $\in$ is caused by the noise in this measurement, as shown in Fig. 8.63. Triangles PQR and $P'Q'R'$ are similar

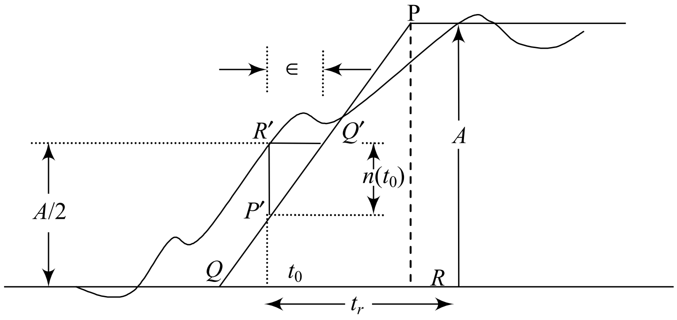


Fig. 8.63 *Received pulse and position error ∈*

$$\therefore \qquad (\in/n(t_0)) = (t_r/A) \qquad \therefore \in = \left(\frac{t_r}{A}\right)\cdot n(t_0) \tag{8.94}$$

and $$\sigma^2 = \overline{\in^2} = \left(\frac{t_r}{A}\right)^2 \overline{n^2} = \left(\frac{t_r}{A}\right)^2 \cdot \eta B_N \tag{8.95}$$

But $$t_r \cong \frac{1}{2B_N} \text{ and } A^2 = \frac{E_p}{\tau_0} \text{ for PDM and } \frac{E_p}{\tau^2} \text{ for PPM}$$

where $$Ep = \text{Average energy per pulse} = \begin{cases} A^2\tau_0 & \text{for PDM} \\ A^2\tau & \text{for PPM} \end{cases}$$

and τ_0 = unmodulated pulse width for PDM
τ = Pulse width for PPM

$\therefore$ $$\sigma^2 = \frac{\eta}{4B_N A^2} = \begin{cases} \dfrac{\eta\tau_0}{4B_N E_p} & \text{for PDM} \\ \dfrac{\eta\tau}{4B_N E_p} & \text{for PPM} \end{cases} \tag{8.96}$$

$$S_R = E_p \cdot f_s = \begin{cases} A^2\tau_0 f_s & \text{for PDM} \\ A^2\tau\ f_s & \text{for PPM} \end{cases} \tag{8.97}$$

$$S_D = \begin{cases} m^2\tau_0^2\overline{x^2} & \text{for PDM; where } \tau_0 \text{ is unmodulated pulse width} \\ m^2 t_0^2 \overline{x^2} & \text{for PPM; where } t_0 \text{ is the proportionality constant for conversion} \\ & \qquad\qquad \text{form amplitude to time in sec/volt} \end{cases}$$

$N_D = \sigma^2$ and is as given by Eq. (8.96)

For PDM:

$$\frac{S_D}{N_D} = \left(\frac{S}{N}\right)_D = \frac{m^2\tau_0^2\overline{x^2}}{\sigma^2} = \frac{m^2\tau_0^2\overline{x^2}4B_N A^2}{\eta}$$

But $$A^2\tau_0 f_s = S_R \quad \text{and} \quad \frac{S_R}{\eta W} = \gamma \quad \text{and} \quad B_N \approx B_T$$

$\therefore$ $$\left(\frac{S}{N}\right)_D = 4m^2\tau_0 B_T \overline{x^2}\left(\frac{W}{f_s}\right)\gamma \tag{8.98}$$

To maximize $\left(\frac{S}{N}\right)_D$, we note that

$$f_{s\min} = 2W; \quad \tau_{0\max} = \frac{T_{s\max}}{2} = \frac{1}{2f_{s\min}} = \frac{1}{4W} \quad \text{and} \quad m_{\max} = 1$$

$\therefore$ $$\left(\frac{S}{N}\right)_D \le \frac{1}{2}\left(\frac{B_T}{W}\right)\overline{x^2}\gamma \tag{8.99}$$

For PPM:

$$\left(\frac{S}{N}\right)_D = \frac{S_D}{N_D} = \frac{m^2 t_0^2\overline{x^2}}{\sigma^2} = \frac{m^2 t_0^2\overline{x^2}4B_N A^2}{\eta}$$

But $$A^2\tau f_s = S_R \quad \text{and} \quad \frac{S_R}{\eta W} = \gamma. \quad \text{Also,} \quad B_N \approx B_T$$

$$\therefore \qquad \left(\frac{S}{N}\right)_D = \frac{4m^2\overline{x^2}B_T t_0^2 S_R}{\eta\tau f_s} \tag{8.100}$$

Note that $f_{s\min} = 2W; \quad t_0 \le \frac{T_s}{2}$ so that $t_{0\max} = \frac{T_s}{2} = \frac{1}{4W}$

Also, the pulse width, $\tau \ge 2t_r = 1/B_T \quad \therefore \tau_{\min} = \frac{1}{B_T}$

Substituting the above in Eq. (8.100) in order to maximize $(S/N)_D$, we get

$$\left(\frac{S}{N}\right)_{D;\,\max} = \frac{1}{8}m^2\overline{x^2}\left(\frac{B_T}{W}\right)^2\gamma \tag{8.101}$$

i.e.,

$$\left(\frac{S}{N}\right)_D \le \frac{1}{8}m^2\overline{x^2}\left(\frac{B_T}{W}\right)^2\gamma \tag{8.102}$$

Thus, just like in CW wideband FM, for PPM also the destination SNR varies as the square of the transmission bandwidth. *In practice, however, the $(S/N)_D$ for PPM will be less than the maximum value given by Eq. (8.102) by about 10 dB*. Nevertheless, PPM has the advantage of low average power requirement for the transmitter and so is used in situations where average transmitter power is at a premium.

It may be noted that as suggested by Eqs. (8.99) and (8.102), both PDM and PPM offer a trade-off between transmission bandwidth and average transmitter power. However, since the destination SNR of PPM varies as the square of B_T, while that of PDM varies proportional to B_T only, the PPM offers a better trade-off than PDM.

Example 8.16 A message signal has $\overline{x^2} = 0.1$ and is band limited to 100 Hz. It is sampled at a rate of 250 samples/sec and converted into a PDM signal with $m = 0.2$ and an unmodulated pulse width of 80 μs, which is then transmitted over a channel of bandwidth 3 kHz. If the two-sided PSD of the additive noise on the channel is 0.5×10^{-12} W/Hz, find the value of the received average signal power, S_R given that the $(S/N)_D$ is to be at least 40 dB.

Solution From Eq. (8.99), we have

$$\left(\frac{S}{N}\right)_D = 4m^2\tau_0 B_T\overline{x^2}\left(\frac{W}{f_s}\right)\gamma = 4\times(0.2)^2\times 80\times 10^{-6}\times 0.1\left(\frac{100}{250}\right)\gamma$$

$$\therefore \qquad \gamma = \left(\frac{S}{N}\right)_D \cdot \frac{250\times 10^6}{4\times 0.04\times 80\times 0.1\times 100} = \left(\frac{S}{N}\right)_D \cdot \frac{250\times 10^6}{128} = 1953125\left(\frac{S}{N}\right)_D.$$

But $\left(\frac{S}{N}\right)_D$ should be at least 40 dB, i.e., 10^4. $\quad \therefore \left(\frac{S}{N}\right)_D \ge 10^4$

$$\therefore \qquad \gamma \ge 1953125\times 10^4.\ \text{But } \gamma = \frac{S_R}{\eta W} \quad \therefore S_R = \gamma\,\eta W$$

$$\therefore \qquad S_R \ge 1953125\times 10^4 \times 10^{-12}\times 100 = 1953125\times 10^{-6}\ W$$

$$\therefore \qquad S_R \ge 1.953125W$$

8.11.3 Comparison between FDM and TDM

Both FDM and TDM achieve the same objective – that of transmitting several message signals simultaneously over the same physical channel; only, the techniques used are different and so each one has its own advantages. However, TDM has a definite edge over FDM because of its simplicity.

Table 8.1

S. No	FDM	TDM
1	Individual message channels are allocated different non-overlapping frequency slots.	Individual message channels are allocated distinct, non-overlapping time slots.
2	Requires subcarrier modulator, bandpass filter and de-modulator for each message channel.	Uses inexpensive digital VLSI circuitry for switching operations at the commutator and the de-commutator.
3	Synchronization required for the carrier generated at the receiving-end, in the case of SSB-SC modulation.	Synchronization of the commutator and the de-commutator is essential and is more elaborate.
4	Short-term fading of the channel affects all the message channels.	Short-term fading affects at the most only a few channels
5	Slow, narrow-band fading of the channel may affect at the most one or two FDM channels only.	Slow, narrow-band fading of the channel affects all the message channels of TDM.
6	Multiplexing message channels of widely different bandwidths is not easy.	Multiplexing message channels of widely different pulse rates (bandwidths) is not difficult.
7	Cross-talk in FDM is caused by non-linear cross-modulation and imperfect bandpass filtering.	Cross-talk in TDM is caused by high frequency and low-frequency deficiencies and dispersion, if any, in the channel.

MATLAB Example 8.1 *Sampling of bandlimited signal and its recovery from the samples*

The three samples are $x_1(t)$ = 1/3 Sin 2π5t, $x_2(t)$ =1/3 * Sin 2π10t, $x_3(t)$ = 1/3*Sin 2π15t. Form a signal x(t) bandlimited to 15 Hz as follows:

$$x(t) = 1 + [x_1(t) + x_2(t) + x_3(t)]$$

MATLAB Program

```
% This program calls the following functions
% 1. fftseq
%
% Natural sampling using rectangular pulses
%
clc
clear
% Displaying one rectangular pulse of unit amplitude and 5 mS duration
df = 0.1;
T = 10*10^-3 % Sampling period
ts = 1/10000;
fs = 1/ts;
tn = 0:ts:T
N = size(tn)
s = zeros(N)
s(1:50) = ones(size(s(1:50)))
figure (1)
subplot(2,1,1)
plot(tn,s,'r')
xlabel('Time in Seconds')
ylabel('Amplitude');
title('Rectangular Pulse');
ylim([0 1.2]);
xlim([0 0.015])
%
% Spectrum of the rectangular pulse
```

```
%
subplot(2,1,2)
K = 256;
s1 = abs(fft(s,K))
s1 = fftshift(s1);
f = [-K/2:K/2-1]/K;
plot(100*f,s1),
grid on
xlabel('frequency / f s')
ylabel('Amplitude');
title('Spectrum of Rectangular Pulse');
%
% Combination of three sinusoidal signals
%
ts = 1/fs;
df = 0.1
tn = 0:ts:0.6
x= 1+1/3*(sin(2*pi*5*tn)+sin(2*pi*10*tn)+sin(2*pi*15*tn));
figure (2)
subplot(2,1,1)
plot(tn,x);
xlabel('time')
ylabel('Amplitude')
title('Orginal signal')
grid on
[X,x1,df] = fftseq(x,ts,df);
X = X/fs;
n = length(fftshift(abs(X)))
f = fs*[-n/2:n/2-1]/n;
subplot(2,1,2)
plot(f,fftshift(abs(X)));
xlabel('Frequency');
ylabel('Amplitude')
title('Frequency Spectrum of Orginal Signal')
grid on
xlim([-20 20]);
%
% Making 100 pulses
%
N = length(tn)
s(1:N) = 0;
 n = 0
   for i = 1:60:N-1
     s(i+n:i+n+30) = 1;
     n = n+1;
  end
%
% Plotting of the pulses
%
```

```
figure (3)
subplot(2,1,1)
plot(tn,s(1:length(tn)))
ylim([0 1.2])
grid on
xlabel('Time');
ylabel('Amplitude')
title('Plot of 100 pulses')
%
% Sampled analog signal
%
length(x)
y = x.*s(1:length(tn));
subplot(2,1,2)
plot(tn,y)
grid on
xlabel('Time');
ylabel('Amplitude')
title('Sampled analog signal')
%
% Spectrum of the sampled signal
%
[X,x1,df] = fftseq(y,ts,df);
X = X/fs;
n = length(fftshift(abs(X)))
f = fs*[-n/2:n/2-1]/n;
figure(4)
subplot(2,1,1)
plot(f,fftshift(abs(X)));
xlabel('Frequency');
ylabel('Amplitude')
title('Spectrum of Sampled Signal');
grid on
xlim([-20 20]);
%
% Lowpass filter implementation
%
y1 = lpf(50,0.1,y);
subplot(2,1,2)
plot(y1)
xlabel('Frequency');
ylabel('Amplitude')
title('Lowpass filtered output of sampled signal')
%
% Getting the envelop of the sampled analog signal
%
N = length(y1)
  for i = 1:N-1
    if y1(i) <=0
      y1(i) = 0;
```

```
    end
  end
  n = 1
  yi(1) = y1(1)
  for i = 15:60:N-1
    n = n+1;
      yi(n) = max(y1(i-14:i+15));
    end
    newn = n;
subplot(2,1,1)
figure (4)
t = 0:n-1;
t = t*0.006
plot(t,yi)
disp('the value of n is')
xlabel('time');
ylabel('Amplitude')
title('Recovered sIgnal');
grid on
%
% Frequency spectrum of the filtered output
%
[X,x1,df] = fftseq(yi,ts/60,df);
fs = fs/60;
X = X/fs;
n = length(fftshift(abs(X)))
f = fs*[-n/2:n/2-1]/n;
subplot(2,1,2)
plot(f,fftshift(abs(X)));
xlabel('Frequency');
 ylabel('Amplitude')
 title('Recovered Signal Spectrum')
grid on
xlim([-20 20]);
figure (5)
subplot(2,1,1)
plot(tn,x);
xlabel('time');
ylabel('Amplitude');
title('Original signal');
t = 0:newn-1;
t = t*0.006;
subplot(2,1,2)
plot(t,yi);
xlabel('time');
ylabel('Amplitude');
title('Recovered sIgnal');
```

function [M,m,df] = fftseq(m,ts,df)

```
% [M,m,df] = fftseq(m,ts,df)
% [M,m,df] = fftseq(m,ts,df)
% FFTSEQ generates M, the FFT of the sequence m.
% The sequence is zero padded to meet the required frequency resolution df.
% ts is the sampling interval. The output df is the final frequency.
% Resolution. Output m is the zero padded version of input m. M is the FFT.
%
fs = 1/ts;
if nargin == 2
n1 = 0;
else
n1 = fs/df;
end
n2 = length(m)
n = 2^(max(nextpow2(n1),nextpow2(n2)));
M = fft(m,n);
m = [m,zeros(1,n-n2)];
df = fs/n;
return
```

Results

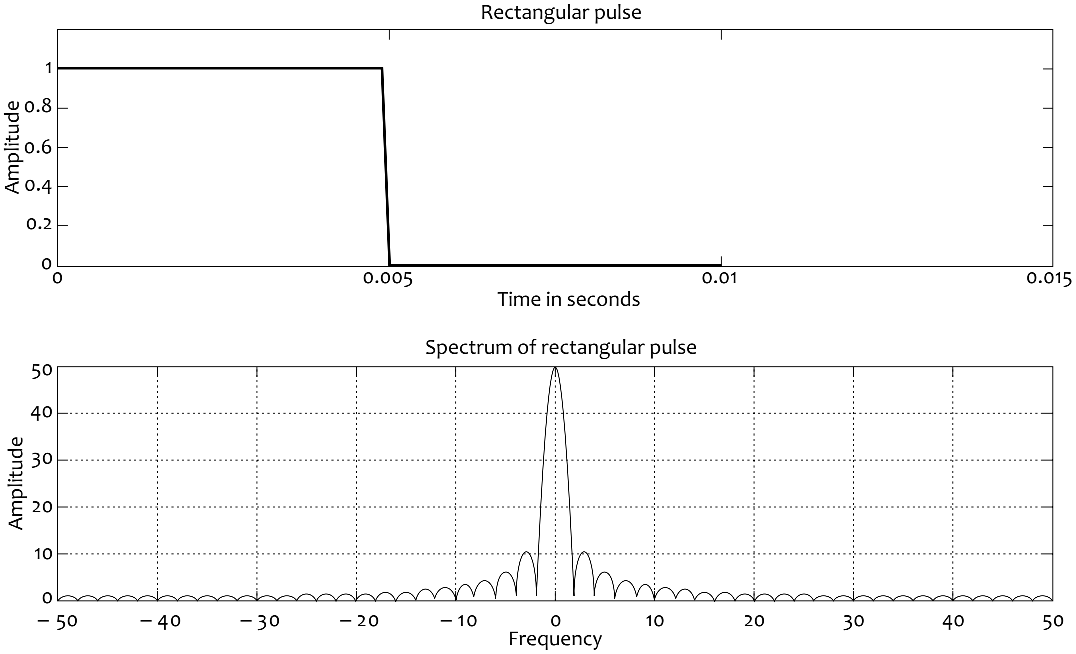


Fig. 8.64 *(Contd.)*

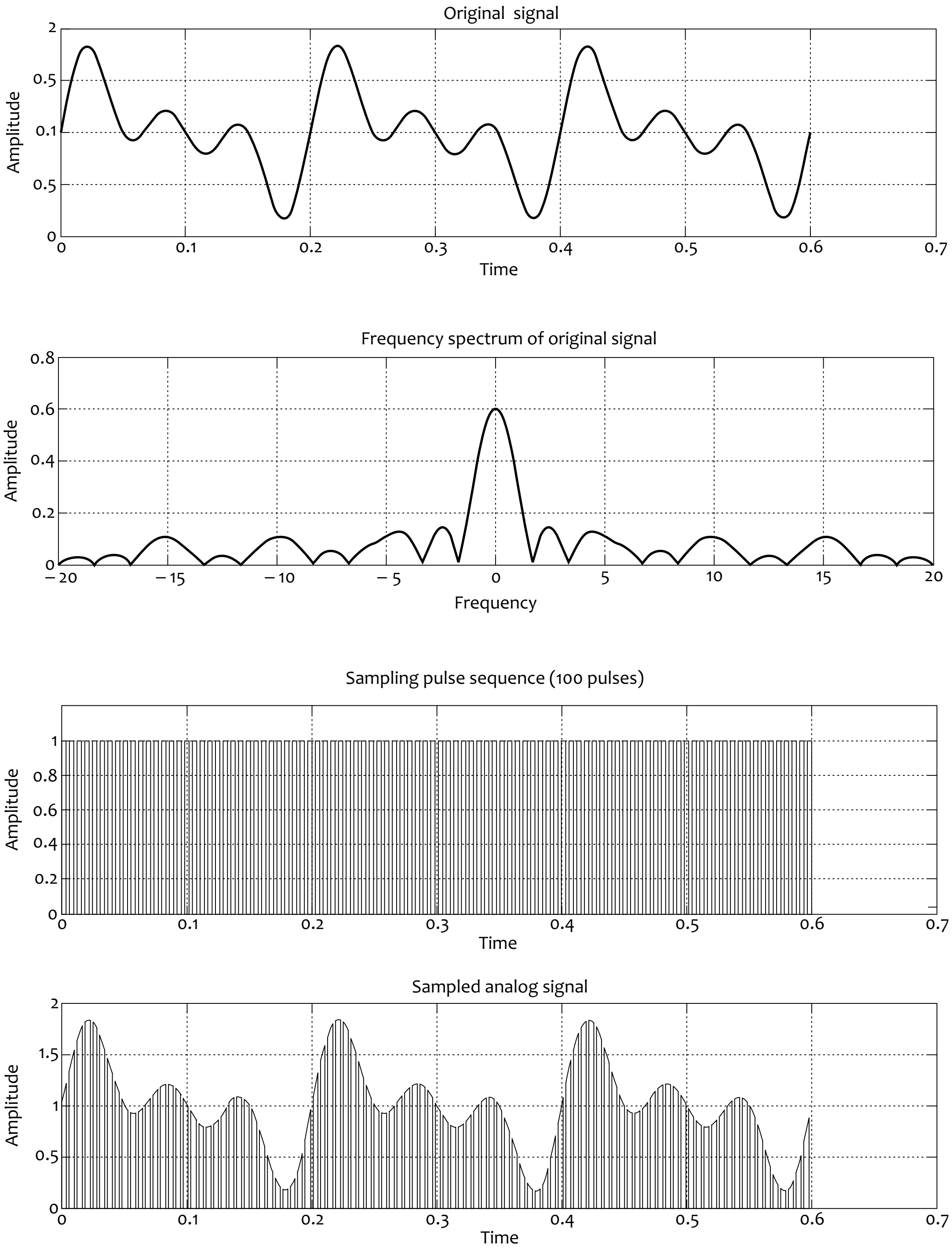


Fig. 8.64 *(Contd.)*

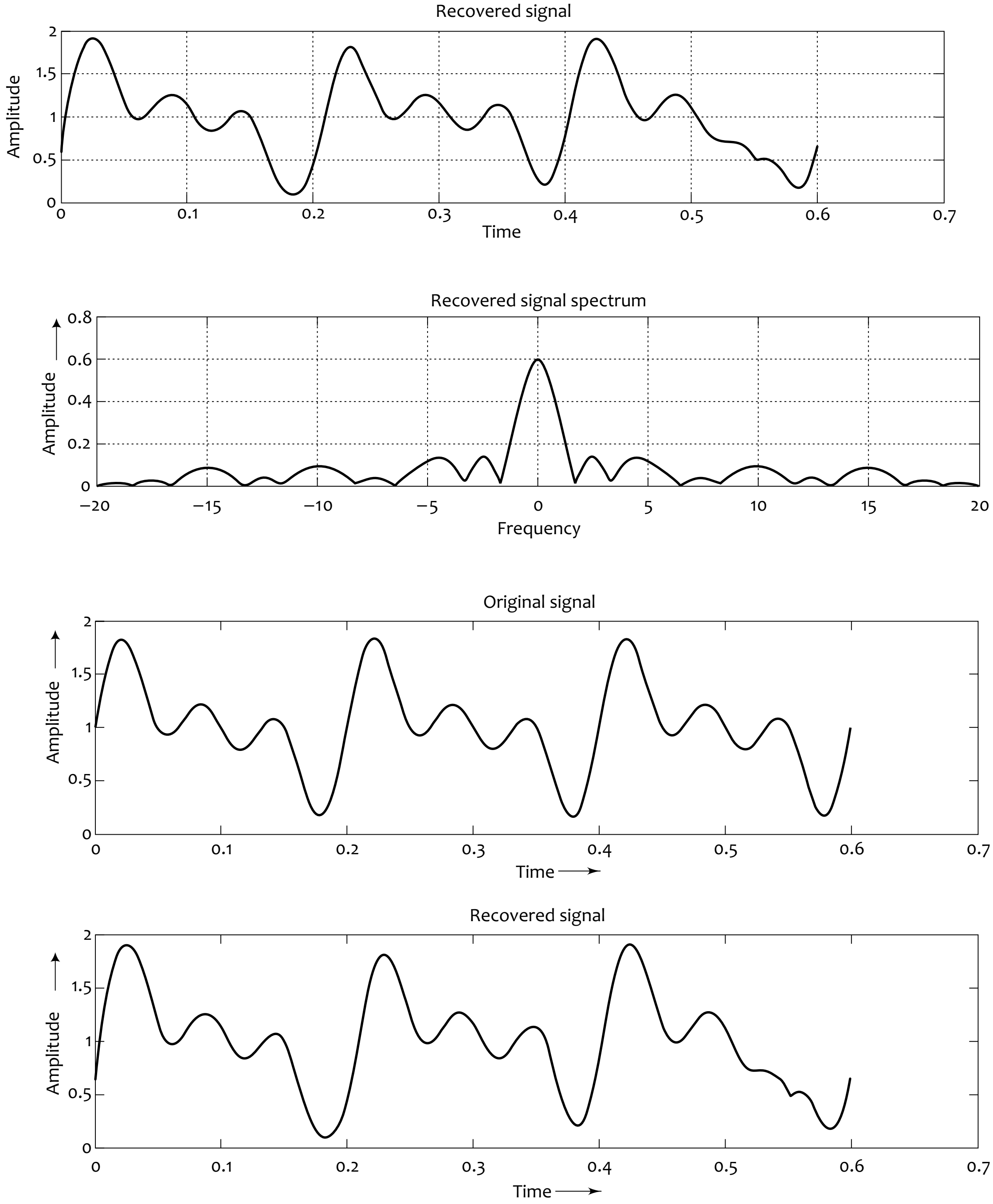

Recovered signal
Amplitude
2
1.5
1
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Time
Recovered signal spectrum
Amplitude
0.8
0.6
0.4
0.2
0
−20
−15
−10
−5
0
5
10
15
20
Frequency
Original signal
Amplitude
2
1.5
1
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Time
Recovered signal
Amplitude
2
1.5
1
0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Time

Fig. 8.64

MATLAB Example 8.2 In this example, we compare linear and cubic interpolation strategies. We compare the interpolations for a sine wave. Such interpolations are often used in practice in applications where sinusoids of different frequencies need to be generated using one cycle of a sinusoid that is stored.

```
t = pi * [ 0 : 10] * 0.4 ; x = sin (t) ; %generate the basic signal
ti = pi * [0 : 40] * 0.1 ; xi = sin (ti) ; %generates closely spaced samples of
                                             the actual signal
xic = interp1 (t, x, ti. ‘spline’) ; xil = interpl (t, x, ti, ‘linear’);
%generates the cubic spline and the linear interpolated versions
Y = [xil.’ ,xic.’ ,xi.’];
h = stem (ti (11 : 30) , y (11 : 30, :), ‘k’);
set (h(1), ‘MarkerFaceColor’ , ‘black’)
set (h(2), ‘MarkerFaceColor’ , ‘black’ , ‘Marker’ , ‘square’)
set (h(3), ‘MarkerFaceColor’ , ‘black’, ‘Marker’ , ‘diamond’)
title(‘Linear and Cubic Interpolations’);
ylabel (‘value’); xlabel (‘sample index’);
legend(‘Linear’, ‘Cubic’ , ‘Actual’)
%It is clear from the figure that while linear interpolation leads to
%large errors, cubic interolations will work better.
```

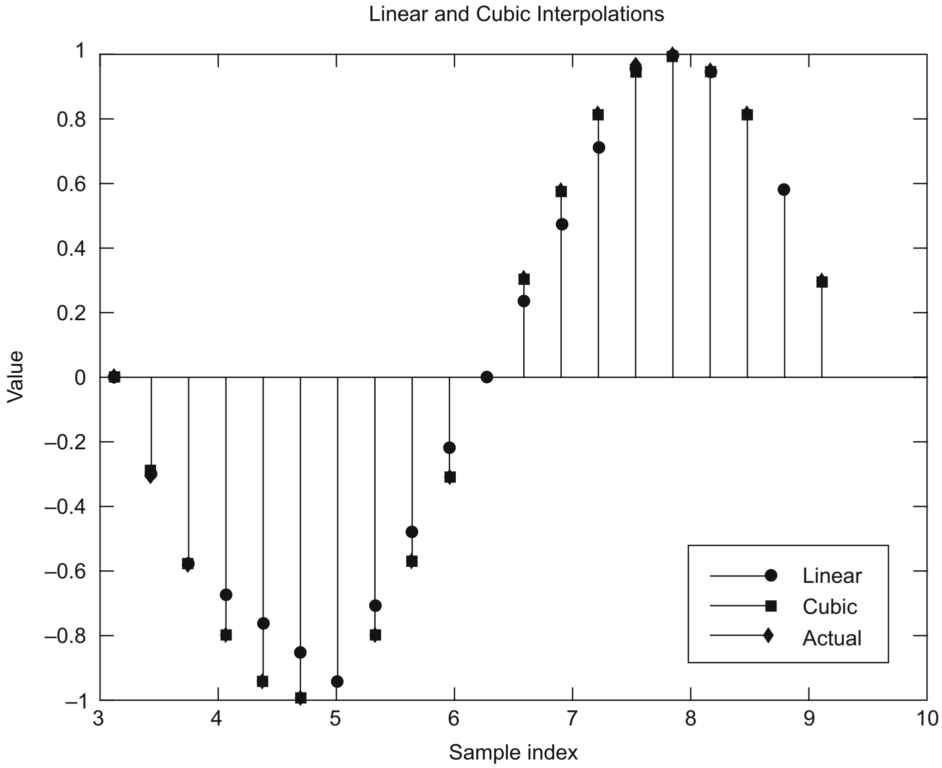


Fig. 8.65 *The linear and the cubic spline interpolated version of a sinusoid*

MATLAB Example 8.3 (Pulse Width Modulation and Pulse Position Modulation)

Modulating (message) signal is a sinusoidal signal with f = 100, and sampling frequency is 4000 samples/s. Pulse carrier will be as shown in Fig. 8.66:

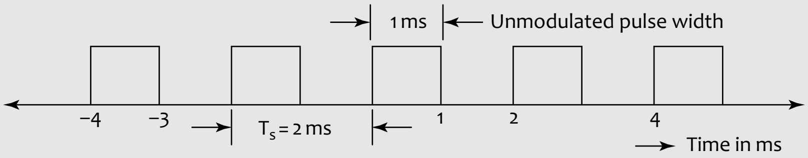


Fig. 8.66

At the peak of the modulating sinusoidal signal, the pulse carrier width should increase from the unmodulated value. Generate the Pulse Width Modulated (PWM) signal. Then do the following:

(a) Display the PWM signal obtained (for I full cycle of modulating signal).
(b) Display its spectrum.
(c) Demonstrate the PWM signal and display two cycles of the recovered message signal.
(d) Replace PWM in your program by PPM and repeat the above three steps.

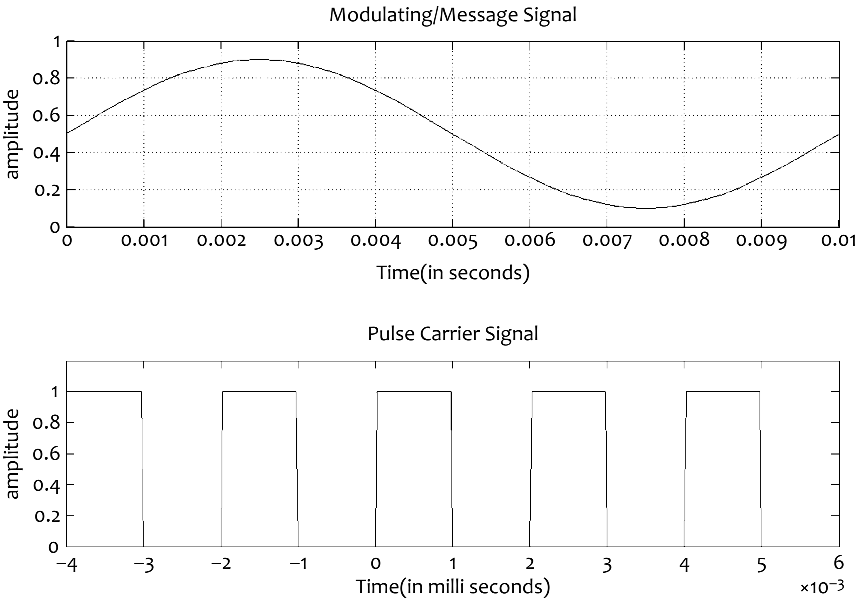
MATLAB Program

```
clc
Fc = 100;  % modulating signal frequency
Fs = 4000; % sampling frequency
ts = 1/Fs  % sampling time 0.0025 (0.25 milli seconds)
t = [0:ts:1/Fc];    % a total 4500 samples for 9 milli seconds (500 samples per
millisecond)
size(t)
%
% Generation of message wave
%
figure (1)
subplot (2,1,1)
x = 0.5+0.4*sin(2*pi*Fc*t); % message or modulating signal
plot(t,x)
grid on
ylabel (‘amplitude’);
xlabel (‘time(in secs)’)
title (‘modulating/message signal’);
%
% Generation of pulse carrier
%
tt = (-4/1000:(ts/10):6/1000)
m = zeros(size(tt));
m(1:40) = ones(size(m(1:40)));
m(81:120) = ones(size(m(81:120)));
m(161:200) = ones(size(m(161:200)));
m(241:280) = ones(size(m(241:280)));
m(321:360) = ones(size(m(321:360)));
subplot (2,1,2)
```

```
plot(tt,m)
ylim ([0 1.2])
ylabel ('amplitude');
xlabel ('time(in milli secs)')
title ('Pulse carrier signal');
%
% Generation of PULSE WIDTH MODULATION
%
y = modulate(x,Fc,Fs,'pwm','centered');
k = 1:1:length(y)
k = k/(length(y)*100);
figure (2)
subplot (2,2,2)
plot(k,y)
title('Modulated Signal');
ylim ([0 1.2])
subplot (2,2,1)
plot(t,x)
grid on
ylabel ('amplitude');
xlabel ('time(in secs)')
title ('modulating/message signal');
% Demodulated signal of modulated signal
%
m1 = demod(y,Fc,Fs,'pwm','centered');
subplot(2,2,3)
plot (t,m1);
title ('PWM Demodulated Signal');
grid on
%
Sx = fftshift(abs(fft(y))) % power spectrum of PWM
f = -length(Sx)/2:1:(length(Sx-1)/2)-1
subplot (2,2,4)
plot (f,Sx);
grid;
title('Magnitude Spectrum of x(n)');
xlabel('Frequency, Hz');
ylabel('Magnitude, dB');
%
% Generation of PULSE POSITION MODULATION
%
k = 1:1:length(Sx)
k = k/(length(Sx)*100);
y1=modulate(x,Fc,Fs,'ppm');
figure (3)
subplot (2,2,2)
plot (k,y1)
title('Pulse position Modulated Signal');
ylim ([0 1.2])
subplot (2,2,1)
```

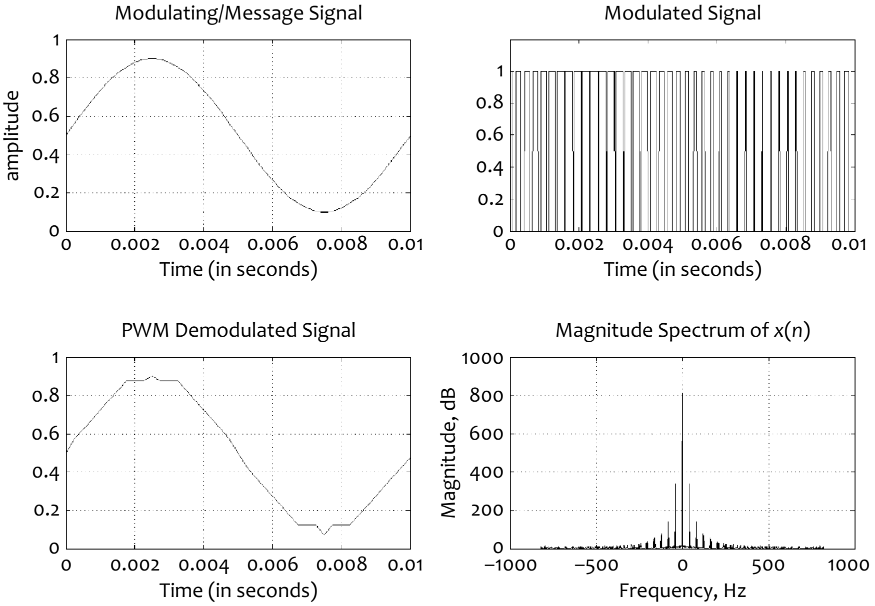
```
plot(t,x)
grid on
ylabel (‘amplitude’);
xlabel (‘time(in secs)’)
title (‘modulating/message signal’);
%
% Demodulated signal of PPM
%
m2 = demod(y1,Fc,Fs,‘ppm’);
subplot (2,2,3)
plot (t,m2);
grid on
ylabel (‘amplitude’);
xlabel (‘time(in secs)’)
title (‘PPM Demodulated signal’);%
% Spectrum of the pulse position modulated signal
%
Sx1 = fftshift(abs(fft(y1))) % power spectrum of PPM;
f = -length(Sx1)/2:1:(length(Sx1-1)/2)-1
subplot (2,2,4)
plot (f,Sx1)
grid on;
title(‘Magnitude Spectrum of x(n)’);
xlabel(‘Frequency, Hz’);
ylabel(‘Magnitude, dB’);
```

Results



Pulse Width Modulation

Fig. 8.67 *(Contd.)*



Pulse Position Modulation

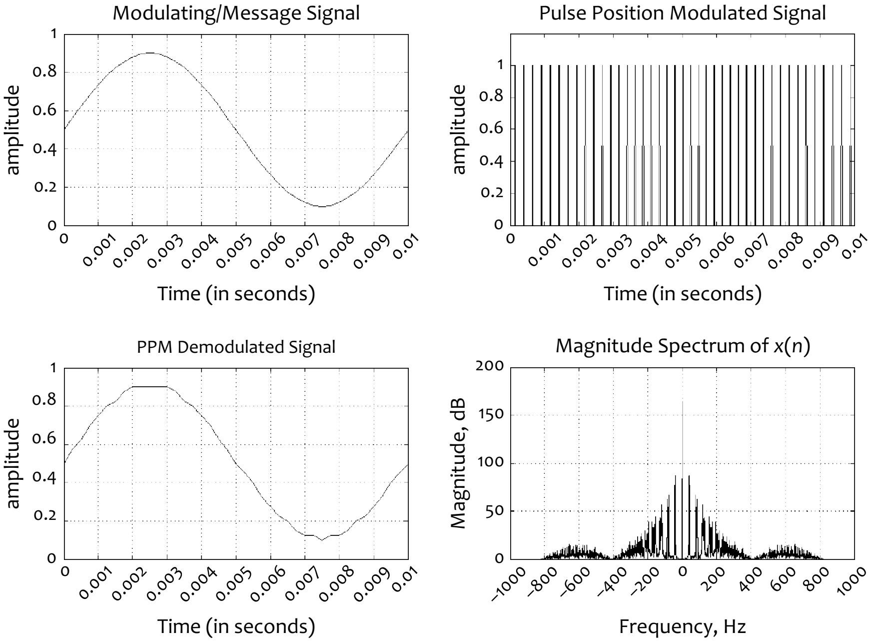


Fig. 8.67

Summary

- Statement of low pass sampling theorem: if $x(t)$ is a low pass signal, band limited to W Hz, i.e., if $X(f) = 0$ for all $|f| \geq W$, it is possible to recover $x(t)$ completely, without any distortion whatsoever, from its samples taken at intervals $T_s \leq 1/2W$. $x(t)$ can be expressed in terms of its samples as

$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s) \operatorname{sinc} 2B(t - kT_s)$$

- The low pass sampling theorem provides the basis for all analog pulse modulation systems as well as all digital communication systems.
- $f_s = 2W$ represents the minimum sampling rate that can be used, for sampling a low pass signal band limited to W Hz, if the signal is to be recovered from its samples. This minimum sampling rate is called 'Nyquist Rate'.
- There are basically three types of sampling – impulse or ideal sampling, natural sampling using pulses of finite width, and flat-top sampling using finite-width pulses.
- It is possible to recover $x(t)$, without any distortion, from its samples in the case of ideal sampling as well as natural sampling.
- In the case of flat-top sampling, which is the most commonly used sampling method in practice, it is not possible to recover $x(t)$ without any distortion because of 'aperture effect'. However, this distortion can be reduced using an amplitude equalizer with an appropriate transfer function.
- 'Aliasing', or 'folding-over effect' occurs because of under-sampling, i.e., sampling below the 'Nyquist rate', and manifests itself as some of the high frequency components of $x(t)$ reappearing as low frequency components in the spectrum of the sampled signal.
- Aperture effect is a distortion that appears in the message signal recovered from its samples taken using flat-top sampling. Because of this effect, high frequency components of the recovered message signal $x(t)$, suffer relatively higher attenuation compared to its low frequency components.
- A zero-order-hold (ZOH) may be used to reconstruct the message from its samples – it gives a staircase approximation of the message.
- PAM, PDM and PPM are, strictly speaking, not modulation techniques at all, as there is no frequency translation and these signals cannot be radiated directly. They are actually signal processing methods – methods used for representing a sample value in terms of the amplitude of a pulse in the case of PAM the width of a pulse in the case of PDM/PWM and the shift/delay in the position of a pulse in the case of PPM.
- A PPM signal may be obtained by directly flat-top sampling an analog signal. It can be detected by making use of a low pass filter, followed by, if necessary, an equalizer.
- A PDM signal may be generated either from a PAM signal, using a ramp signal and a comparator, or directly from the analog signal by using a mono-stable multivibrator.
- The spectrum of a PDM signal consists of a dc component, the message signal and groups of phase-modulated waves with sampling frequency f_s and its harmonics as the carrier frequencies.
- A PDM signal may be detected either by first converting into a PAM signal and low pass filtering this PAM signal, or by directly low pass filtering the PDM signal itself.
- A PPM signal may be generated by first generating a PAM signal and converting it into a PDM and then using a trigger pulse produced by the trailing edge of the PDM when it is differentiated, to generate a rectangular pulse from a pulse generator so that the leading edge of the resultant PPM pulse coincides with the trailing edge of the PDM pulse.
- Cross-talk occurs in PAM, PDM as well as PPM transmission if the channel bandwidth is inadequate. In PAM, it causes an amplitude error while in PDM and PPM it causes a timing error.
- In TDM-ed transmission of PAM, PPM, or PDM, synchronization of the commutators at the two ends is necessary.
- In comparison with FDM, the TDM hardware is much simpler. It has several other advantages over FDM, such as its ability to easily handle baseband signals having widely different bandwidths and its relative robustness with regard to short-term fading.
- As far as noise performance is concerned, PAM is at least 3 dB inferior direct baseband analog message signal transmission.

- The destination SNR of PDM is proportional to (B_T/W) while that of PPM is proportional to the square of (B_T/W). Hence, power to bandwidth trade-off is possible in both, with PPM offering a better trade-off. However PPM is inferior to WBFM by about 10 dB.

References and Suggested Readings

1. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 4th Edition, International Edition, McGraw-Hill, 2002.
2. Couch II, Leon W., *Digital and Analog Communication Systems*, 6th Edition, Pearson Education (Asia), 2001.
3. Haykin, Simon, *Communication Systems*, 3rd Edition, John Wiley & Sons, 1995.
4. Rao, P. Ramakrishna, *Signal and Systems*, Tata McGraw-Hill, New Delhi, 2008.
5. Taub, Herbert, and Donald L. Schilling, *Principles of Communication Systems*, International Student Edition, McGraw-Hill Kogakusha, Tokyo, 1971.

Review Questions

1. What is aliasing? How can it be reduced or avoided?
2. What is meant by aperture effect? How can it be reduced?
3. What is a Zero-Order-Hold? How can it be used as a reconstruction filter?
4. State the low pass sampling theorem and briefly explain its significance.
5. Discuss the advantages and disadvantages of analog pulse modulation as compared to continuous-wave modulation.
6. Explain how a PAM signal may be generated. How can it be demodulated?
7. Describe with the help of neat sketches of waveforms, any two methods of generation of PDM/PWM and PPM.
8. How do you de-modulate a PDM signal?
9. Explain how a PPM signal may be converted into a PAM signal.
10. What is Time Division Multiplexing?
11. If N voice signals, each of bandwidth W Hz, are TDM-ed, show that the TDM-ed signal needs a minimum transmission bandwidth of NW Hertz.
12. What is meant by cross-talk with reference to TDM-ed signals?
13. Explain how the low frequency deficiency of a channel causes cross-talk consider PAM signal and model the channel as a high pass RC filter.
14. By considering a PAM signal and using a low pass RC filter as the model for the channel, show how high frequency deficiency of a channel can cause cross-talk.
15. Derive an expression for the destination signal-to-noise ratio of a PAM system, and show that it cannot exceed $\gamma/2$.
16. With the help of a neat sketch, show how additive noise on the channel can cause an error in the measurement of the arrival time of a pulse.
17. Show that in the case of PPM, the $(S/N)_D$ takes a maximum value which is proportional to the square of the ratio of transmission bandwidth B_T to the message bandwidth W.
18. Critically compare FDM and TDM.

Problems

1. Determine the Nyquist rate of sampling for the following signals:
   (a) $x(t) = 10\ \text{sinc}\ 100\ t$ (b) $x(t) = 10\cos^2(100\ \pi t)$
   (c) $x(t) = 10\ \text{sinc}^2(100\ t)$
2. For each of the signals listed below, identify the minimum sampling frequency needed to ensure that no aliasing takes place,
   (a) $x(t) = 5\,\text{sinc}(10t)\cos(100\ \pi t)$ (b) $x(t) = 10\cos^2 100\ \pi t$
   (c) $x(t) = 4\Pi(t/10^{-2})\cos(10^2\pi t)$

3. A unipolar rectangular wave of unit amplitude, 0.3 duty cycle and a period of *T* sec is used as the sampling function for sampling a signal $x(t)$ with a maximum frequency component 1 kHz. What is the largest value of *T* for which reconstruction of $x(t)$ from the samples would be possible? Determine a suitable system for reconstruction of $x(t)$ from the samples.
4. To completely describe a periodic band-limited signal, it is enough if we have the samples from one period. How many samples are needed to exactly describe the following band-limited periodic signals?
   (a) $x(t) = 5\cos(300\,\pi t) + 15\sin(200\,\pi t)$ (b) $x(t) = 16\cos(5\pi t) + 6\sin(8\pi t)$
5. The signal $x(t) = 12\cos 40\,\pi t$ is ideally sampled at $f_s = 50$ samples/sec. Plot the spectrum of the sampled version up to a frequency of ±180 Hz.
6. The schematic diagram of a bipolar chopper is shown in Fig. P8.6(a).
   (a) Sketch the waveform of the sampling function $s(t)$, assuming switch *k* starts at *A* at $t = 0$ and makes contact alternately at *A* and *B* staying at each stud for $T_s/2$ sec.
   (b) If the spectrum of $x(t)$ is as shown in Fig. P8.6(b).
      (i) Sketch the spectrum of the sampling function $s(t)$
      (ii) Sketch the spectrum of the sampled signal $x_s(t)$ assuming $f_s > 2W$.
   (c) Comment on the filter to be used for reconstruction of $x(t)$ from $x_s(t)$.

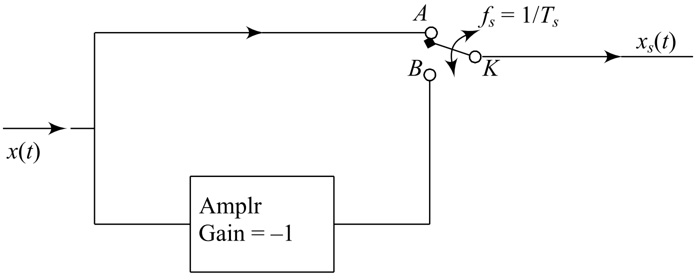


Fig. P8.6(a)

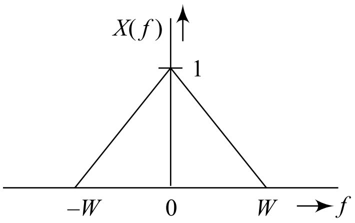


Fig. P8.6(b)

7. For the low pass sampling theorem of Section 8.2, there is a dual. It says that if $x(t)$ is time-limited, i.e., if $x(t) = 0$ for $|t| \geq T$, then the frequency-domain representation of $x(t)$, namely, $X(f)$, can be determined without any error from its samples taken at regular frequency intervals of $f_0 \leq 1/2T$. Prove this.
8. A PAM is represented by
$$x_p(t) = \sum_k A_0[1 + mx(kT_s)]p(t - kT_s)$$
   (a) Show that its spectrum is given by
$$X_p(f) = A_0 f_s P(f)\left[\sum_k \{\delta(f - nf_s) + mx(f - nf_s)\}\right]$$
   (b) Sketch $X_p(f)$ when $p(t)$ is a rectangular pulse of amplitude 1 and base width equal to half the sampling period; $m = 1$ and $x(t) = \cos 2\pi \times 200t$ when $f_s = 500$ Hz. Take $A_0 = 1$.
9. What is the transmission bandwidth needed for a PDM signal for which the sampling frequency is 8 kHz, $m = 0.8$ and $|x(t)| \leq 1$ and unmodulated pulse width $\tau_0 = T_s/5$. It is desired that the rise time t_r should not be greater than a quarter of the minimum pulse width in the PDM signal.
10. 15 voice signals, each band limited to 4 kHz, are sampled at a rate that allows us to provide a guard band of 1.5 kHz to facilitate reconstruction. The samples are transmitted using PAM with AM of a continuous wave, i.e., PAM/AM, the duty cycle being 0.25. Calculate the required transmission bandwidth.
11. 10 message signals, each band limited to 2 kHz are sampled at a frequency f_s that permits a 1 kHz guard band. The multiplexed samples are transmitted by (a) PAM/AM with 25% duty cycle, and (b) PAM/FM with baseband filtering and a peak frequency deviation of ±75 kHz.

Multiple-Choice Questions

1. A band limited low pass signal is sampled at twice its Nyquist rate with $f_s = 2000$ samples/sec. The signal is band limited to

(a) 250 Hz (b) 1000 Hz (c) 500 Hz (d) 2000 Hz

2. A certain low pass signal $x(t)$ is sampled and the spectrum of the sampled version has guard band from 1500 Hz to 1900 Hz. The sampling frequency used is

(a) 1500 samples/sec (b) 1900 samples/sec (c) 1700 samples/sec (d) 3400 samples/sec

3. A low pass signal band limited to 1200 Hz was sampled and it was found that the 1000 Hz frequency component was reappearing in the recovered signal, because of aliasing, as 400 Hz component. The sampling frequency used is

(a) 1400 samples/sec (b) 1600 samples/sec (c) 2200 samples/sec (d) 800 samples/sec

4. $x(t) = 3 \cos^2 250\, \pi t$. This signal is sampled at regular intervals of T sec. The maximum value of T for which $x(t)$ may be recovered from the sampled version without any distortion is equal to

(a) 1 m.sec (b) 2 m.sec (c) 4 m.sec (d) 0.5 m.sec

5. A cosinusoidal signal $x(t) = 5 \cos 240\, \pi t$ was sampled at a frequency f_s. The signal recovered from the samples was, however, found to be $3 \cos 110\, \pi t$. The sampling frequency f_s is equal to

(a) 175 samples/sec (b) 350 samples/sec (c) 130 samples/sec (d) 65 samples/sec

6. Aperture effect

(a) amplifies the high frequency components (b) attenuates the low frequency components
(c) amplifies the low frequency components (d) attenuates the high frequency components

7. A continuous-time signal $x(t)$ is ideally sampled using a unit impulse train with a sampling interval of T sec. The sampled version is

(a) a sequence of samples of $x(t)$, the k^{th} sample being equal to $x(kT)$ and located at $t = kT$
(b) a periodic version of $x(t)$ with period of T sec.
(c) a sequence of impulses, the k^{th} impulse having a strength of $x(kT)$ and located at $t = kT$
(d) None of the above

8. The most commonly used sampling method is

(a) ideal or impulse sampling (b) natural sampling using rectangular pulses
(c) sample-and-hold method (d) None of the above

9. The distortion in the signal arising from aperture effect, can be reduced by

(a) reducing the width of the pulses used for flat-top sampling
(b) reducing the sampling frequency
(c) properly band limiting the signal before sampling it
(d) using flat-top sampling

10. The impulse response function, $h(t)$, of a zero-order-hold circuit is

(a) an impulse (b) a rectangular pulse (c) a triangular pulse (d) None of these

11. A PAM signal may be generated using

(a) impulse sampling (b) a sample-and-hold circuit
(c) natural sampling (d) a clipper circuit

12. A PAM signal may be demodulated using

(a) a low pass filter (b) a differentiator followed by a low pass filter
(c) an integrator (d) a low pass filter followed by an equalizer

13. Cross-talk occurs in PAM/TDM-ed system because of

(a) only low frequency deficiency of the channel
(b) only high frequency deficiency of the channel
(c) either low frequency deficiency or high frequency deficiency, or both
(d) non-linear cross-modulation

14. In general, cross-talk decreases with increasing bandwidth.
   (a) It reduces more rapidly in PPM than in PAM
   (b) It reduces more rapidly in PAM than in PPM
   (c) It reduces at the same rate in PAM and PPM
   (d) None of the above

15. Noise performance of PAM is
   (a) better than that of direct base-band transmission
   (b) better than CW amplitude modulation
   (c) poorer than that of direct base-band transmission
   (d) better than that of PDM

16. $(S/N)_D$ of PDM is
   (a) proportional to the transmission bandwidth
   (b) proportional to the square of the transmission bandwidth
   (c) proportional to the square-root of the transmission bandwidth
   (d) independent of the transmission bandwidth

17. $(S/N)_D$ of PPM is
   (a) proportional to the transmission bandwidth
   (b) proportional to the square of the transmission bandwidth
   (c) proportional to the square-root of the transmission bandwidth
   (d) independent of the transmission bandwidth

18. Short-term fading of the channel
   (a) affects only a few message channels of an FDM system
   (b) affects all the message channels of a TDM system
   (c) affects all the message channels of an FDM system
   (d) does not have much effect on both TDM and FDM systems

Key to Multiple-Choice Questions

1. (c)	2. (d)	3. (a)	4. (b)	5. (a)	6. (d)	7. (c)	8. (c)
9. (a)	10. (b)	11. (b)	12. (d)	13. (c)	14. (a)	15. (c)	16. (a)
17. (b)	18. (c)						

DIGITAL CODING OF ANALOG SIGNALS (PCM AND DM)

9

> *"If your actions inspire others to dream more, learn more, do more and become more, you are a leader."*
>
> **John Quincy Adams (1767–1848)**
> *Sixth President of the United States (1825–1829)*

Learning Objectives

After going through this chapter, students will be able to

- *understand the need for and the effect of quantization, different types of quantizers, and the need for companding of speech signals in PCM systems,*
- *explain the relationship between Q, the number of quantization levels; n, the bits per codeword; r, the bit-rate and B_T, the transmission bandwidth of a pulse-code modulated signal,*
- *draw the block diagrams of the transmitter and receiver of a baseband PCM system and explain the function of each block,*
- *determine the signal-to-quantization noise ratio as well as signal-to-noise ratio for PCM systems,*
- *explain the SNR-BW trade-off in PCM systems,*
- *explain the operation of DM, ADM, DPCM, ADPCM systems using the block diagrams of their transmitters and receivers,*
- *explain the different methods used for speech compression and the principle of linear predictive vocoders, and also get familiar with the speech compression standards, and*
- *understand the principle of TDM, and become familiar with various details like bit and word interleaving, frame synchronization methods, T-carrier system formats and TDM hierarchy.*

9.1 INTRODUCTION

There are three steps in the process of digital coding of analog signals. These are sampling, quantization and coding. Different methods of sampling, issues involved in each of them and the reconstruction of a band limited low pass analog signal from its samples, have already been discussed in detail in Chapter 8. A sampled signal is only a discrete-time signal and not a digital signal. Quantization is the process of discretizing the sampled version in its amplitude also. We will show that quantization of the samples results in quantization noise and study its dependence on the number of quantization levels. Signals like the speech signal, which have low amplitudes for most part of the time, but have large peaks occurring occasionally, will have poor SNR if uniform quantization is used, i.e., if the quantization levels are equally spaced over the dynamic range of the signal, unless the levels are very closely spaced. As close spacing of the levels increases the bandwidth required for the signal, we resort to non-uniform quantization. When these quantized sample values are encoded using a suitable code, like the binary code, it results in a digital signal.

The most basic form of digital communication is by the use of the Pulse Code Modulation (PCM), which results when a binary code is used in encoding the quantized samples. Channel noise, which causes severe problems in long-haul analog communication using a number of repeaters, can be made irrelevant in properly designed long-haul PCM transmission systems using regenerative repeaters.

Variants of PCM like delta modulation, adaptive delta modulation, and differential pulse code modulation, their principles, advantages and disadvantages, and fields of application, are discussed in detail, along with Time Division Multiplexing (TDM), its principle, signaling and synchronization details, and TDM hierarchy.

The different time-domain and frequency-domain techniques used for compression of speech, and the present-day standards for compression of these signals are also discussed in this chapter.

9.2 QUANTIZATION

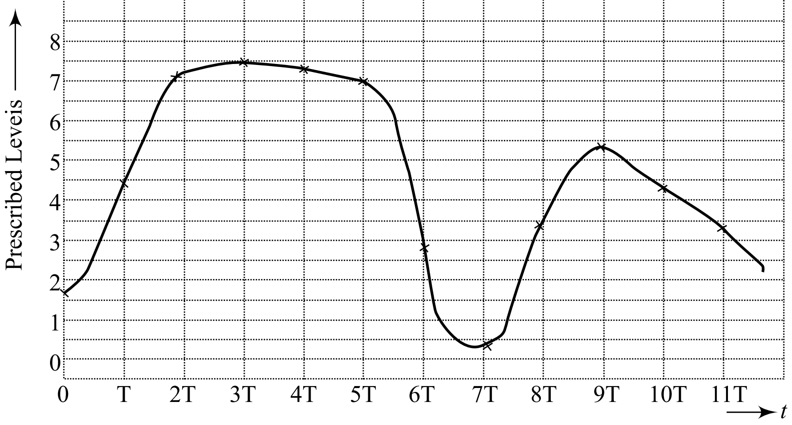
Sampling of an analog signal merely converts it into a discrete-time signal, i.e., a signal which is discretized in time but not in its amplitude. A quantizer discretizes the discrete-time signal in its amplitude also. There are two types of quantization – uniform and non-uniform. Both of these produce what is called ‘quantization noise’. As we are going to see, from the ‘signal-to-quantization noise ratio’ point of view, uniform quantization is suitable for signals whose amplitude probability distribution is uniform over their dynamic range and the non-uniform quantization is more suitable for signals with low amplitude but occasional large peaks, such as the speech signals.

9.2.1 Uniform Quantization

A continuous-time signal is defined for all values of time and its amplitude can have any value within the limits set by its dynamic range. When we sample it, what we get is only a discrete-time signal and not a digital signal. Sampling, as we have seen, discretizes the continuous-time signal only in time but not in amplitude. The samples obtained by the sampling process can have a continuum of values – they are not restricted to any finite set of prescribed values. So, the next step in the digitization of an analog signal is the discretization of the amplitudes of these samples obtained through the sampling process. For this, let us say we divide the dynamic range of the analog signal into a certain finite number of equal segments, as shown in Fig. 9.1. We round-off a sample value falling within a particular segment to the value represented by the ‘prescribed level’ passing through the middle of that segment. This process of rounding-off is called quantization.

Quantization certainly introduces errors into the values of the samples. This in turn may be viewed as deliberately distorting the original message signal. However, we now have samples whose values are restricted to a certain finite set of prescribed values. This means that we have discretized the amplitudes of the samples. Thus, the original message signal has been discretized in time by the sampling process and in amplitude by the quantization process. Since the permitted, or prescribed levels are equidistant, this quantization is called ‘uniform quantization’. It must be noted that the quantization process which we are adopting is such that it is instantaneous or memory-less in the sense that the way the rounding-off of a sample is done is solely dependent upon its actual value only and is not in any way influenced by the values of the other samples preceding it or succeeding it. Before proceeding to the quantizers, their types and characteristics, let us define ‘quantization’.

Definition of quantization It is the process of assigning to each one of the sample values of the message signal, a discrete value from a prescribed set of a finite number of such discrete values, called the ‘quantized values’.



Sample Number	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}
Actual values of the samples	1.7	4.4	7.2	7.4	7.3	7.0	2.4	0.4	3.4	5.3	4.3	3.2
Rounded-off values	2	4	7	7	7	7	2	0	3	5	4	3
Binary code for rounded value	010	100	111	111	111	111	010	000	011	101	100	011

Fig. 9.1 *Illustrating the quantization process*

We note that for any given dynamic range of the analog signal, if the number of the 'prescribed levels', or the 'quantized levels', is increased, the interval between two successive levels, called the 'step size', becomes smaller and so the error due to quantization, which can be at the most ±0.5 (step size), also becomes smaller. This will certainly give a better approximation to the original analog message. However, in order to get a better approximation by increasing the quantization levels, we have to pay a price in the form of increased bandwidth, as we will see later.

9.2.2 Quantizers

Quantization can be performed by feeding the samples of the analog signal to a 'quantizer', which transforms each of the samples fed to it into a 'quantized sample' having an amplitude corresponding to the 'prescribed level' used for representing any sample value falling within the pertinent interval (step) into which that analog signal sample falls. This is shown diagrammatically in Figs. 9.2(a) and (b).

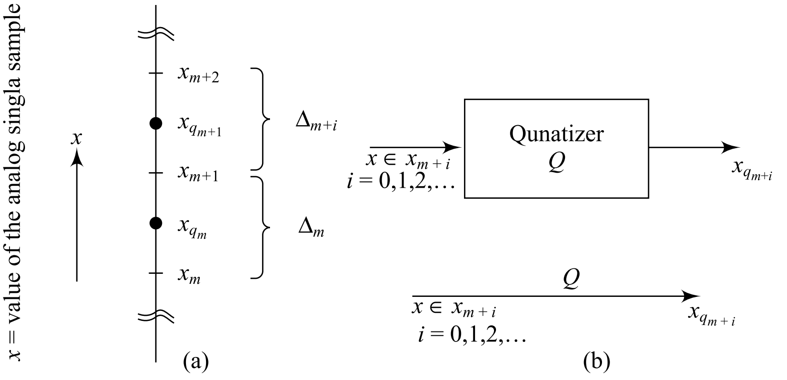


Fig. 9.2 *(a) and (b) Action of a quantizer*

As shown in the above figures, any sample value x falling in the interval x_m to x_{m+1}, i.e., in the interval (or step) Δ_m, will be mapped by the quantizer into the prescribed value x_{q_m} corresponding to that interval.

As stated earlier, a quantization process in which the quantization levels are uniformly spaced is called ‘uniform quantization’, and a quantizer which performs uniform quantization is called a ‘uniform quantizer’. A quantization process in which the quantization levels are not uniformly spaced, is called a non-uniform quantization and the quantizer performing the non-uniform quantization, is called a ‘non-uniform quantizer’. We shall presently discuss only the uniform quantizers.

Uniform quantizers These are of two types depending on the shape of the input-output characteristic:

1. Mid-tread type
2. Mid-rise type

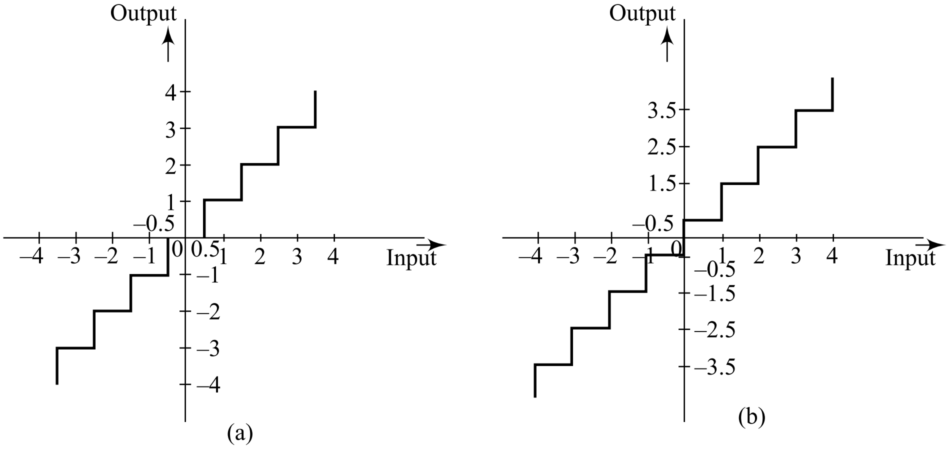


Fig. 9.3 *Quantization characteristic: (a) Mid-tread quantizer, (b) Mid-rise quantizer*

As shown in these figures, the quantizer characteristics have a staircase shape irrespective of whether it is a mid-tread type or a mid-rise type. The only difference is that in the mid-tread characteristic, the origin is in the ‘tread’ portion and that is why it is called mid-tread; whereas for the other one, the origin is in the ‘rise’ portion of the staircase type of characteristic and so it is known as the mid-rise type of quantizer. Note that in the mid-tread type, any input value between –0.5 and +0.5 is mapped to an output value of ‘zero’; any input value between 0.5 and 1.5 is mapped to an output value of 1, and so on. In the mid-rise quantizer, on the other hand, any input value between 0 and 1 is mapped to an output value of 0.5; any input value between 1 and 2 is mapped to an output value of 1.5, and so on.

Example 9.1 The temperature at a particular place varies between 14°C and 34°C. For the purpose of transmitting the temperature record of that place using PCM, the record is sampled at an appropriate sampling rate and the samples are quantized. If the error in representation of the samples due to quantization is not to exceed ±1% of the dynamic range, what is the minimum number of quantization levels that can be used?

Solution If Q is the number of levels and Δ is the step size,

$$\Delta = \frac{(34-14)}{Q} \quad \text{Quantization error } = \pm\frac{\Delta}{2} \text{ in the representation of any sample}$$

$$\therefore \qquad 100 \times \left(\frac{\Delta/2}{20}\right) = 1 \quad \text{or} \quad \frac{\Delta}{2} = \frac{1}{5} \qquad \therefore \Delta = \frac{2}{5}\,°\text{C}$$

$$\therefore \qquad Q = \frac{20}{\Delta} = \frac{20 \times 5}{2} = 50$$

∴ the minimum number of prescribed levels, or quantization levels that can be used is 50.

9.2.3 Quantization Noise

Let the message signal $x(t)$ be a zero-mean random process. Then its sample, x, is a zero-mean random variable. If this sample, x, is fed to a quantizer, and if x_q is the prescribed quantization level to which x is approximated, the quantization error e_q is given by

$$e_q = x - x_q$$

Now, since in a quantizer, the levels x_qs are symmetrically located on either side of zero level, and since x is a zero-mean random variable, the random variable e_q must also be zero mean. Since x_q is at the center of the interval of sample amplitudes into which, the sample, x, has fallen, we also know that

$$-\frac{\Delta}{2} < e_q < \frac{\Delta}{2} \tag{9.1}$$

where Δ is the 'interval' of amplitudes, or the 'step size'. That is, the error may take any value from $-\frac{\Delta}{2}$ to $+\frac{\Delta}{2}$. However, the way the error random variable, e_q, is distributed over this range of values is not known. But, if the step size, Δ, is small, i.e., if the number of prescribed quantization levels, Q is large, it is quite reasonable to assume that the random variable, e_q, is uniformly distributed over the interval $-\frac{\Delta}{2}$ to $+\frac{\Delta}{2}$. So, we shall assume that e_q has a PDF as shown in Fig. 9.4.

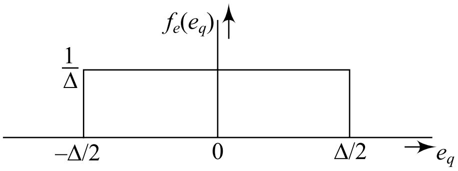


Fig. 9.4 *PDF of the quantization error random variable*

These random errors caused by quantization in the successive samples, appear as noise, called the *quantization noise*. The mean-squared value of this noise can be evaluated, using the probabilistic model we have developed for the quantization-error random variable, e_q.

Mean-squared value of the error = Average power in the quantization noise

$$\overline{e_q^2} = \int_{-\infty}^{+\infty} e_q^2 f_e(e_q)\,de_q$$

But $f_e(e_q)$, the probability density function of e_q, is

$$f_e(e_q) = \begin{cases} 1/\Delta; & -\frac{\Delta}{2} \le e_q \le \frac{\Delta}{2} \\ 0; & \text{otherwise} \end{cases}$$

$$\therefore \qquad \overline{e_q^2} = \int_{-\infty}^{\infty} e_q^2 f_e(e_q)\,de_q = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} e_q^2\,de_q = \frac{1}{\Delta}\left[\left.\frac{e_q^3}{3}\right|_{-\Delta/2}^{\Delta/2}\right]$$

$$\therefore \qquad \boxed{\overline{e_q^2} = \frac{\Delta^2}{12}} \tag{9.2}$$

The step size, of course, depends upon the dynamic range of the message signal $x(t)$ and the number of prescribed quantization levels Q. So, let us normalize the message signal $x(t)$ so that $|x(t)| \le 1$. Then the dynamic range of values of $x(t)$ is –1 to +1. So, we may write

$$\boxed{\Delta = \frac{2}{Q} \qquad\qquad (\text{since } Q >> 1)} \tag{9.3}$$

Substituting this for Δ in the expression for $\overline{e_q^2}$, we get

$$\boxed{\underset{(|x(t)| \le 1)}{\text{Average power in quantization noise}} = \overline{e_q^2} = \frac{1}{3Q^2}} \tag{9.4}$$

9.2.4 Non-Uniform Quantization (Companding)

Consider a uniform quantization with characteristic as shown in Fig. 9.3 for which an input of, say 8 V peak to peak constitutes full load. Let us first consider the case where it is given full 8 V peak-to-peak sinusoidal signal as input. Then the signal power is

$$S = \left(\frac{4^2}{2}\right) = 8 \text{ W}$$

and the noise power is

$$N_q = \frac{\Delta^2}{12} = \frac{1}{12} \text{ W} \qquad \text{(See Eq. (9.2))}$$

Hence, the $(SNR)_q = \left(\frac{S}{N_q}\right) = 96 = 19.8 \text{ dB}$

But suppose, to the same quantizer whose full load input is 8 V peak to peak, we give a 2 V peak-to-peak sinusoidal signal as input. Then the signal power is 0.5 W but quantization noise power, which, as per Eq. (9.2) depends only on Δ, the step size, remains at the same value of (1/12) watt. Thus, the $(SNR)_q$ now is 6, which is equal to 7.78 dB.

Thus, we find that a uniform quantizer gives good $(SNR)_q$ for large amplitude signals but a poor $(SNR)_q$ for low amplitude signals. This is not desirable. A quantizer which gives almost the same $(SNR)_q$ for signals whose powers vary over a wide range, is called a *robust quantizer*. As we will be seeing presently, employing a non-uniform quantization is the way to realize robust quantizer.

Thus, uniform quantization is quite suitable for message signals which have an amplitude probability distribution which is uniform over their dynamic range. But, unfortunately, in the case of speech, which is perhaps the most important message signal in communication engineering, the amplitude probability distribution is not at all uniform. Since the ratio of intensities corresponding to a loud shout and a whisper can be as high as 1000:1, the dynamic range of a speech signal is quite large. However, a typical speech signal is characterized by low amplitude for most part of the time, corresponding to *normal speech*, with occasional peaks of large amplitude but short duration, corresponding to shouts. Thus, while its dynamic range is large, its average power is quite small. So, for speech signals, if we wish to use only uniform quantization, we are left with two alternatives, both of which are unattractive:

1. Keep the quantization error and quantization noise low by using a large number of quantization levels, i.e., have a large Q. As mentioned earlier, and as would be seen later when we discuss PCM in detail, if Q is large, the transmission bandwidth would be correspondingly large.
2. Keep the bandwidth requirement low by using a small number of prescribed quantization levels, i.e., a low Q, in which case, the quantization noise will be high.

Non-uniform quantization provides a good solution for the above problem, as it allows us to keep the quantization noise low even when Q is small, i.e., even when we use a small number of prescribed quantization levels.

As noted earlier, the mean-squared value of quantization noise is proportional to the square of the step size. Further, for speech signals, the signal amplitude is low for most part of the time and large peaks occur only very rarely and they are of very short duration. Hence, in the case of speech signals, in order to obtain a good signal-to-noise ratio, it makes sense to provide smaller step size in the low amplitude region, and a large step size in the large amplitude region of the dynamic range of the signal. This means that we need to employ non-uniform quantization, wherein the prescribed quantization levels are *not* uniformly spaced.

Non-uniform quantization is equivalent to first subjecting the samples of the message speech signal to amplitude compression by passing them through a compressor and then applying uniform quantization to these compressed samples. Compression of the input samples is accomplished according to a specific law governing the relationship between amplitudes of the input and output samples. There are two different compression laws in vogue and they are:

1. μ-law
2. A-law

Both these laws provide a near linear relation between the output amplitude and the input amplitude for small amplitudes of the input samples, and a somewhat logarithmic relation for the larger amplitude input samples.

1. μ-law: If x and y are the *normalized* input and output values, respectively, this compression law is defined by the following input-output relation:

$$\boxed{|y| = \frac{\log[1+\mu|x|]}{\log[1+\mu]}} \tag{9.5}$$

where μ is a positive constant. The typical input-output characteristics of a μ-law compressor are given in Fig. 9.5 for three different values of the constant μ. It may be noted that when $\mu = 0$, it corresponds to uniform quantization and that compression of the larger sample values is higher for larger values of μ.

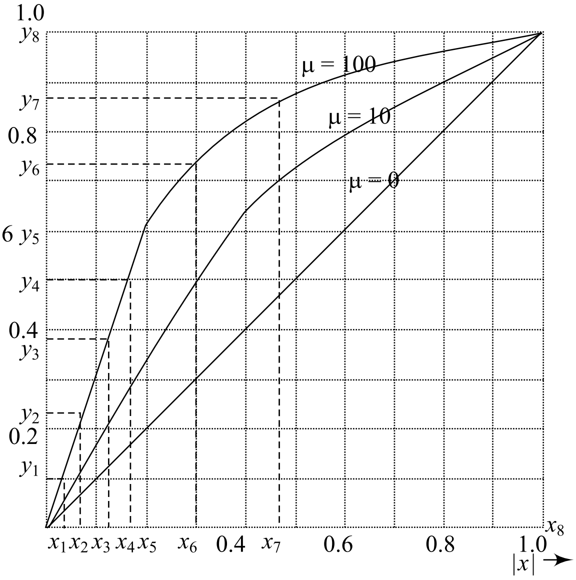


Fig. 9.5 *μ-law characteristic*

2. A-law: The other compression law in vogue is the A-law which is defined by the following input-output relation:

$$|y| = \begin{cases} \dfrac{A|x|}{1+\log A}; & 0 \le |x| \le \dfrac{1}{A} \\[2ex] \dfrac{1+\log[A|x|]}{1+\log A}; & \dfrac{1}{A} \le |x| \le 1 \end{cases} \tag{9.6}$$

where A is a constant. Figure 9.6 shows the input-output relation for an A-law compressor. Clearly, $A = 1$ corresponds to uniform quantization. Larger the value of the constant A, more is the compression.

In both μ-law as well as in A-law, as μ or A increases, the compression for the larger amplitudes also increases and so the dynamic range increases. Again, in both the cases, the signal-to-quantization noise ratio becomes poorer and poorer for larger amplitudes (because of larger step size) as the μ or A increases. Hence, the choice of the value of μ or A is a matter of compromise and in practice $\mu = 255$ and $A = 87.6$ are generally used.

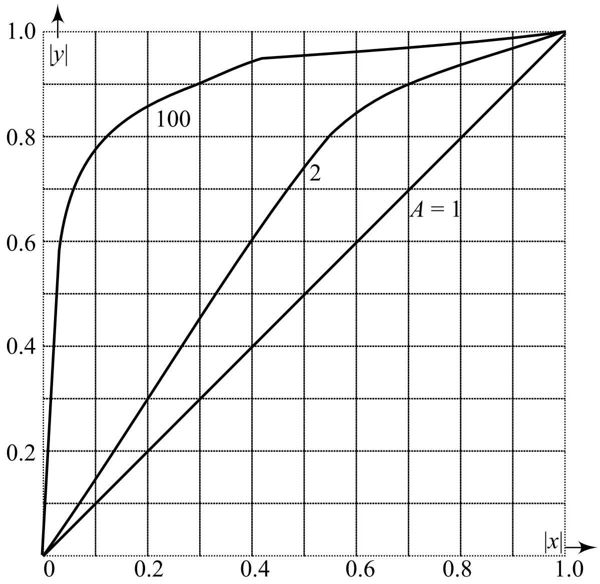


Fig. 9.6 *A-law characteristics*

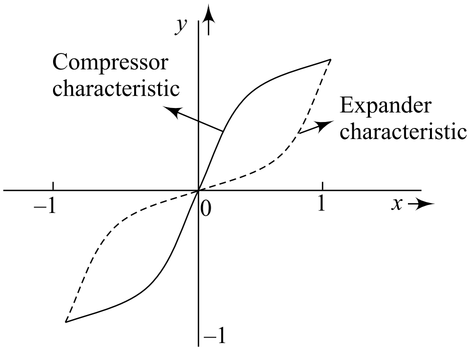


Fig. 9.7 *Compressor and expander characteristics for both positive and negative values of input*

Since the message signal is compressed at the transmitter, in the receiver it has to be expanded to remove the distortion. For this purpose, an *expander* having an input-output characteristic which is the exact inverse of that of the compressor at the transmitter, will have to be used. *This process of compressing message signals like speech, at the transmitter and expanding them at the receiver is called '**companding**'*. Companding is a word coined by combining the words, 'compressing' and 'expanding'. Figure 9.7 shows the compressor and expander characteristics together, for both positive as well as negative values of input samples.

Note *It must be noted that in μ-law as well as A-law, the input represented by x is the normalized input. Correspondingly the output represented by y, is also the normalized output.*

Earlier we had stated that non-uniform quantization of the message x may be considered equivalent to compressing x and then subjecting the compressed signal, y, to uniform quantization. This can easily be understood from Fig. 9.5 which shows the μ-law compressor characteristic. x is the message input to the compressor and y is the compressor output. In that figure, the compressor output is subjected to uniform quantization, the dotted lines passing through $y_1, y_2, \dots, y_8$ being the uniformly spaced quantization levels. Insofar as the input, x, is concerned, these quantization levels passing through $y_1, y_2, \dots, y_8$ are equivalent to input side quantization levels passing through $x_1, x_2, \dots, x_8$. Clearly, these prescribed quantization levels of the input are closer for smaller values and farther apart for larger input values, making the step size smaller for smaller values of the input and large for larger values of input.

Example 9.2 A message signal $A_m \cos \omega_m t$, sampled and applied to a quantizer with Q permitted quantization levels, fully loads it. Find the signal-to-quantization noise ratio at the output of the quantizer.

Solution Since the signal fully loads the quantizer, it means that the Q levels cover the full range of $-A_m$ to $+A_m$ of the signal.

Hence, step size $\Delta = \frac{2A_m}{Q}$

$\therefore$ quantization noise mean-squared value = $\frac{\Delta^2}{12} = N_q$

$$= \frac{4A_m^2}{12Q^2} = \frac{A_m^2}{3Q^2}$$

S = Average signal power = $\frac{A_m^2}{2}$

$$\therefore \qquad \left(\frac{S}{N_q}\right) = \frac{A_m^2}{2} \times \frac{3Q^2}{A_m^2} = \frac{3}{2}Q^2$$

Example 9.3 A message signal with a dynamic range of –8 V to +8 V is non-uniformly quantized using a μ-law compressor with $\mu = 255$ and 64 quantization levels. Determine the smallest and largest step sizes obtained. Assume that the quantizer is fully loaded.

Solution We know that non-uniform quantization of the message signal is obtained by giving the message signal as input to the compressor and subjecting the output of the compressor to uniform quantization (see Fig. 9.5). Since the compressor output is subjected to uniform quantization, and since 64 levels will give 63 steps, for the output of the compressor, the step size is

$$\Delta_0 = \frac{16}{63}$$

We know that on the input side of the compressor, the smallest step size occurs near the origin and the largest step size occurs farthest from the origin. For obtaining the smallest step size, we put $y = y_1$ (see Fig. 9.5)

$$y_1 = \frac{16}{63} = \frac{\log[1 + \mu |x_1|]}{\log(1 + \mu)} = \frac{\log[1 + 255|x_1|]}{\log 256} \quad \therefore \quad x_1 = \frac{3.088}{255}$$

$\therefore$ The smallest step size = $8 \times x_1 = \frac{8 \times 3.088}{255} = 0.0969$ V

(Since x_1 used in the μ-law is the normalized value.)

The largest step-size is obtained by taking the farthest step on the y side.

$$\left(1 - \frac{16}{63}\right) = \frac{\log[1 + \mu |x_l|]}{\log(1 + \mu)}$$ Putting $\mu = 255$ and solving for x_1, we get

$x_l = 0.2415$ V. Hence on the input side, the last step (normalized) = $(1 - x_l)$

$\therefore$ the largest step size = $8 \times (1 - x_l) = 6.068$ volts

9.3 PULSE CODE MODULATION (PCM)

In PCM, the message signal, $x(t)$, is first sampled, using a sampling frequency f_s that is greater than twice the maximum frequency present in $x(t)$, i.e., at $f_s = 2W$. Then the samples are quantized using an appropriate number of quantization levels, Q. This Q is generally chosen to be an integer power of 2, i.e., $Q = 2^n$ where n is a positive integer, and such that this Q is equal to or greater than the number of quantization levels needed for a specified minimum accuracy in the representation of the sample values after quantization (see Example 8). We know that in the process of quantization, every sample value is rounded off to the nearest prescribed quantization level. Hence *any* quantized sample will have a value that corresponds to one of the Q quantization levels. Therefore, transmitting a quantized sample value is equivalent to transmitting the number of the corresponding quantization level. Since there are Q quantization levels, where

$$Q = 2^n \tag{9.7}$$

we know that any quantization level may be uniquely represented by an n-digit binary number where

$$n = \log_2 Q \tag{9.8}$$

Hence, corresponding to each sample value, the PCM transmitter transmits just a sequence of n binary digits, i.e., a binary number that represents the number of the quantization level to which the quantized value of the sample corresponds (see Fig. 9.1).

At this stage, in order not to lose sight of the logic behind all these processes, let us recall that the lowpass sampling theorem tells us that a continuous-time signal, band limited to W Hz can be completely recovered without any error, for all instants of time, from the samples of it taken at regular intervals not exceeding $(1/2W)$ sec. This implies that we need not transmit a band-limited continuous-time (CT) signal at all instants of time – enough if we send its samples so that the receiver reconstructs the CT message. Since the samples have a continuum of values, in order to discretize the signal in amplitude also, we adopted quantization. Now, instead of transmitting the actual sample values, we have to transmit the quantized sample values. However, since any quantized sample value corresponds to one of the Q quantization levels, we do not transmit even the quantized sample value – instead, we transmit the number of the quantization level to which it corresponds. Since $Q = 2^n$ and any quantization level can be uniquely represented by an n-digit binary number, we transmit that binary number each time a sample is taken, i.e., once in T_s sec, we transmit an n-digit binary number called the *code word*, where T_s is the sampling interval.

Although we have, in the above discussion, assumed a binary code for representing the quantization level numbers, any suitable code like the ternary code may be used that gives a unique representation of the level numbers. *However, binary code is almost universally used because of its simplicity, ease of regeneration and ability of binary symbols to withstand high levels of noise*. A PCM system using a binary code for representing the quantization level numbers is referred to as *binary PCM*.

We know that an n-digit binary code word consists of a string of n binary digits, each one of which is either a 'zero' or a 'one'. There are a variety of ways of electrically representing this string of zeros and ones. These are called '*line codes*' and we shall discuss these in detail, later in Chapter 10. For the time being, we shall represent a binary 'one' by a *positive* rectangular pulse and a binary 'zero' by either a *negative* rectangular pulse of the same duration and amplitude, or a zero voltage for the same duration, as shown in the figure.

If in a PCM system, an n-bit binary code word is used to represent each quantized sample value, then it is referred to as an *n-bit binary PCM system*. Since each sample value must be transmitted before the next sample is taken and since the sampling interval is $T_s = \frac{1}{f_s}$, an n-bit code word must be accommodated in T_s sec., or the maximum time available for each bit is (T_s/n) sec. We always make full use of the available time T_s for transmitting the codeword because, if we use less time, the pulse width will be smaller and the transmission bandwidth will be more. The time interval available for each bit (1 or 0), is called one '*time-slot*', and is of T sec. duration where

$$\text{Time slot} = T = (T_s/n) \text{ sec.} \tag{9.9}$$

These binary PCM signals of the type shown in Fig. 9.8(a) or (b), consisting of a string of 1s and 0s, may either be transmitted over a suitable channel directly, or may be used for modulating a high frequency carrier signal before transmission. In the former case, it

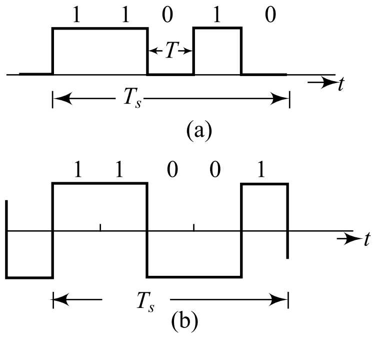


Fig. 9.8 *Representation of binary digits using electrical voltage: (a) Unipolar non-return to zero (NRZ) binary PCM signal, (b) Polar binary NRZ PCM signal*

is called *baseband PCM system* and in the latter case where a high frequency carrier is used, it is called a *bandpass PCM system*.

Earlier, while discussing quantization and quantization noise, we had shown that if Δ is the step size, the mean-squared value of the quantization noise is given by

$$e_q^2 = \frac{\Delta^2}{12}$$

Also, by assuming a normalized message signal, $x(t)$, normalized so that $|x(t)| \le 1$, we approximated the step size to

$$\Delta = \frac{2}{Q}$$

We then obtained, by combining the above two equations

$$e_q^2 = \frac{\Delta^2}{12} = \frac{4}{Q^2} \cdot \frac{1}{12} = \frac{1}{3Q^2} \tag{9.10}$$

∴ if S is the average power of the message signal,

$$(SNR)_q = S \Big/ \left(\frac{1}{3Q^2}\right) = 3Q^2 S$$

But $Q = 2^n$ for an n-bit binary PCM.

∴ for an n-bit binary PCM, the signal-to-quantization noise ratio is given by

$$\boxed{\underset{\text{Binary PCM}}{(SNR)_q} = 3S \cdot Q^2 = 3S \cdot (2^n)^2 = 3S(2^{2n}); \quad \text{if } |x(t)| \le 1} \tag{9.11}$$

where S is the average message signal power.

Hence, for a sinusoidal message signal with peak amplitude A_m, the $(SNR)_q$ of an n-bit binary PCM system is

$$\underset{\text{Binary PCM}}{(SNR)_q} = \frac{S}{e_q^2}$$

where $$\overline{e_q^2} = \frac{\Delta^2}{12} = \left(\frac{2A_m}{Q}\right)^2 \cdot \frac{1}{12} = \frac{A_m^2}{3Q^2}$$

but $Q = 2^n$ for binary b-bit PCM.

∴ $$(SNR)_q = \frac{S}{(A_m^2/3Q^2)} = \frac{(A_m^2/2)}{A_m^2 \cdot 3Q^2} = \frac{3}{2}Q^2 = \frac{3}{2}2^{2n}$$

(It has been assumed here that the quantization has not changed the message signal's average power.)

∴ for an n-bit binary PCM with a sinusoidal message signal, the signal-to-quantization noise is given by

$$\boxed{(SNR)_q = \frac{3}{2}2^{2n}} \tag{9.12}$$

If this $(SNR)_q$ is to be expressed in decibels,

$$(SNR)_q \text{ in dB} = 10 \log_{10} \frac{3}{2} 2^{2n} = 10 \log_{10} 1.5 + 2n \log_{10} 2$$

∴ $$\boxed{(SNR)_q \text{ in dB} = 1.8 + 6n} \tag{9.13}$$

Equation (9.13) shows that for each additional bit in the codeword transmitted by a binary PCM system, the output $(\text{SNR})_q$ increases by 6 dB when the message signal is sinusoidal.

Table 9.1 shows how the signal-to-quantization noise ratio (expressed in dB) varies with n, the number of bits/sample, for an n-bit binary PCM with a sinusoidal message signal.

Table 9.1

N (bits/sample)	$Q = 2^n$ Number of Quantization levels	$(SNR)_q$ in decibels
4	16	25.8
5	32	31.8
6	64	37.8
7	128	43.8
8	256	49.8

Example 9.4 A PCM system uses a uniform quantizer followed by a 7-bit binary encoder. The bit rate of the system is equal to 50×10^6 bits/sec.

(a) What is the maximum message signal bandwidth for which the system operates satisfactorily?

(b) Calculate the output signal-to-quantization noise ratio when a full-load sinusoidal modulating wave of frequency 1 MHz is applied to the input (**UP Tech., 2002–03**)

Solution

(a) 7-bit binary encoder is used $\therefore\ n = 7$

i.e., each sample is represented by a 7-bit code word.

Now 'bit rate' of a PCM system = (Number of samples/sec.) × (Number of bits used to represent each sample)

$$\therefore \quad r = f_s \times n = \text{Bit rate of the PCM system}$$

This is given to be 50×10^6 bits/sec

$$\therefore \quad f_s \times n = 7 f_s = 50 \times 10^6$$

$$\therefore \quad f_s = \frac{50}{7} \times 10^6 \text{ samples/sec.}$$

We know that in order to avoid aliasing, f_s has to be at least $2W$, where W is the message bandwidth.

$$\therefore \quad f_s \geq 2W \quad \therefore \quad W \leq \frac{f_s}{2} = \frac{1}{2}\left[\frac{50}{7} \times 10^6\right] = 3.57 \text{ MHz}$$

$$\therefore \quad \left.\begin{array}{l}\text{Maximum message signal bandwidth}\\ \text{for which the system operates satisfactorily}\end{array}\right\} = 3.57 \text{ MHz}$$

(b) We know from Eq. (9.13) that for a uniform quantizer with a full-load sinusoidal modulating/message signal, the output signal-to-quantization noise ratio (in dB) is given by

$$(SNR)_q = (1.8 + 6n)\text{dB}$$

Since $n = 7$, $(SNR)_q = (1.8 + 6 \times 7)\text{dB} = 43.8 \text{ dB}$

Example 9.5 A signal having a bandwidth equal to 3.5 kHz is sampled, quantized and coded by a PCM system. The coded signal is then transmitted over a transmission channel that supports a transmission rate of 50×10^3 bits/sec. Determine the maximum signal-to-noise ratio that can be obtained by this system. The input signal has a peak-to-peak value of 4 V and an RMS value of 0.2 V.

(**Pune University, 1998**)

Solution Since the message signal has a peak value of 2 V and an RMS value of only 0.2 V, it is not a sinusoidal signal. So Eq. (9.13), which gives the $(SNR)_q$ when the message signal is sinusoidal, cannot be used. We will use the more general expression of Eq. (9.11) which is applicable to any message signal. This says that

$$\underset{\text{Binary PCM}}{(SNR)_q} = 3S(2^{2n}); \quad |x(t)| \le 1$$

where S is the average power of the normalized message signal and n is the number of bits used for representing each sample value. In the above expression for SNR, S, the average power of the normalized signal is a constant and has been given to be having an RMS value of 0.2, i.e., an average power of 0.04 W. Since our message signal is not normalized so that $|x(t)| \le 1$, its average power after normalization will be

$$S = \frac{\{\text{Average power of the given } x(t)\} = 0.04}{\{\text{Square of the peak value of } x(t)\} = (2)^2} = \frac{0.04}{4} = 0.01$$

$\therefore$ since S in the expression for $(SNR)_q$ is constant, the maximum value of $(SNR)_q$ is obtained corresponding to the largest possible value of n that can be used.

$$\underset{\max}{r} = (f_s \cdot n)_{\max} = 50 \times 10^3$$

To maximize n (in order to obtain maximum SNR), let us take the minimum value of $f_s = 7 \times 10^3$ Hz, which is the Nyquist rate for the 3.5 kHz bandwidth signal.

$$\therefore \quad 7 \times 10^3 \times n_{\max} = r_{\max} = 50 \times 10^3$$

$$\therefore \quad n_{\max} = \frac{50 \times 10^3}{7 \times 10^3} = \frac{50}{7} = 7.14$$

Note

Since $n_{max} = 7.14$ and n has to be an integer, we have to take n = 7 and not 8. If we take 8, we will have a transmission rate $r = 8 \times 7 \times 10^3 = 50 \times 10^3$ bits/sec which the channel will not support.

Hence, choosing n = 7, and substituting this in the expression for $(SNR)_q$, we get the max value of $(SNR)_q$.

$$[(SNR)_q]_{max} = 3 \times 0.01 \times 2^{2\times 7} = 491.52$$

$$\therefore \quad [(SNR)_q]_{max} \text{ in dB} = 10 \log_{10} 49.52 = 26.9154$$

Example 9.6 A message signal, band limited to 4 kHz is to be transmitted using a PCM system. If the quantization error of any sample is to be at the most ±1% of the dynamic range of the message signal, determine the minimum value of n, the minimum sampling rate and the corresponding bit rate of transmission.

Solution Let the dynamic range of the message signal be from –V to +V volts. Let Q be the number of quantization levels. We know that step size

$$\Delta = \frac{\text{Dynamic range}}{\text{No. of quantization levels}} = \frac{2V}{Q}\text{V}$$

Also, we know that quantization error in any sample is

$$(e_q)_{\max} = \pm\frac{\Delta}{2} = \pm\frac{V}{Q}\text{V}$$

This maximum error is given to be 1% of the dynamic range

$$\therefore \quad \frac{V}{Q} \le 0.01 \times 2V \quad \therefore \quad Q \ge 50$$

$\therefore$ Minimum value of Q to be used = 64, as Q has to be a power of 2.

$\therefore$ Let $$2^n \geq 64$$

$\therefore$ $n \geq 6$, or the min value of $n = 6$.

$\therefore$ $$n_{\min} = 6$$

Since W = Band-limiting frequency of the message signals is

$$W = 4 \times 10^3 \text{ Hz}, \quad f_{s_{\min}} = \text{Min. sampling rate} = 8 \times 10^3 \text{ samples per second}$$

$\therefore$ The corresponding transmission rate = $r_{\min}$

where $$r_{\min} = n_{\min} \times f_{s_{\min}} = 6 \times 8 \times 10^3 \text{ bits/sec.}$$

$\therefore$ $$r = 48 \text{ kilo bits/sec.}$$

Till now, we have discussed in detail, how the message signal is sampled, quantized and encoded so that the continuous-time message signal is converted into a digital signal. Because of the binary encoding employed, during each time slot, the digital signal can have one of two possible values, a 'one' or, a 'zero'. A binary 'one' is represented by a rectangular pulse of say, V volts in the time slot and a binary zero by, say, the absence of any pulse during the time slot as shown in Fig. 9.8(a). Alternatively, it may also be represented by a rectangular pulse of $-V$ volts in the time slot as shown in Fig. 9.8(b). For the present, for convenience, we shall assume that the encoded signal is as shown in Fig. 9.8(a). Let this PCM signal be transmitted without any further high frequency carrier modulation, through an appropriate channel.

During transmission through the channel, the PCM signal will be

1. Attenuated
2. Distorted due to the finite bandwidth of the channel
3. Corrupted by the additive noise introduced in the channel

Hence, the signal received at the output of the channel, which is the signal given as input to the receiver will appear somewhat as shown in Fig. 9.9, assuming that the code word shown in Fig. 9.8(a) is transmitted.

What the receiver does, on receiving the signal, is to make a considered decision, *during each time slot*, as to whether what has been received during that time slot is a '1' or a '0', i.e., a 'pulse', or 'no pulse'. If there were to be no noise, it would be quite easy to make this decision without committing an error. A simple though not an elegant way of taking this decision may be to take a sample of the received signal during each time slot, preferably at the center of the time slot, compare the sample amplitude with a fixed pre-set threshold and declare it as a '1' if it exceeds the threshold and declare it as a '0' during that time slot, if it is less than the threshold. The threshold may be set at 50% of the maximum amplitude attained by the received pulse. More sophisticated and practically used methods of decision making and choosing an appropriate threshold value in order to minimize the probability of committing an error, called optimum detection techniques are discussed in Chapter 11.

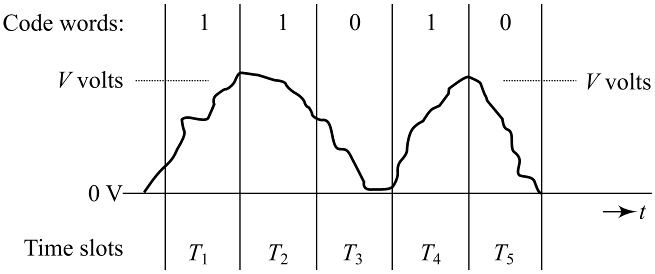


Fig. 9.9 *Received signal*

Once it is decided that it is a 1 in a time slot, a local pulse generator in the receiver is triggered and it gives a clean, noise-free rectangular pulse. Thus, although the received pulses were distorted and corrupted by additive noise, in their place, through this process of detection of the presence of a pulse and generating a clean pulse locally in the receiver, a process which is known as '*regeneration*', we are able to get rid of the additive noise of the channel completely, provided of course, there are no errors in the decision making process. The possibility of errors being committed in the decision making process will be negligibly small, provided the received signal (pulse) amplitude is large enough compared to noise.

This sequence of clean 'pulses' and no-pulses', obtained through the process of regeneration, are then fed to a decoder which converts each codeword into the corresponding quantized sample value. This quantized sample value will be the correct one (i.e., the same quantized sample value that was encoded and transmitted) provided no errors have been committed in all the decisions pertaining to the *n* digits in that code word. These quantized samples obtained at the output of the decoder (i.e., the D/A converter), are then passed through a low pass reconstruction filter, which reconstructs an analog signal from these samples. This analog signal will be an approximation to the original analog message signal that was sought to be transmitted. The difference between the two is the result of the errors introduced, deliberately, in the sample values due to the quantization process at the transmitting end and can be reduced by employing a larger *Q*, i.e., larger number of bits/sample. Of course, in all this, we have assumed that sufficient care has been taken to ensure that there are no decision-making errors. Such errors, if caused in any codeword, will result in large deviation in the quantized sample values (obtained at the output of the decoder) from the correct values and cause severe distortion of the reconstructed analog signal at the output of the receiver.

From the foregoing, it is clear that unlike analog modulation systems, the PCM system, because of the digital modulation, can be made almost immune to channel noise, if it is ensured that signal power input to the receiver is sufficiently large so as not to cause any decision-making errors. Only quantization noise will be present at the output but that can be made small by using a larger value of *n*.

9.3.1 PCM Transmitter and Receiver

Since we have discussed in some detail, the basic principle of a PCM system, we will now get into a few more details of its working through a study of the block diagrams of the transmitter and receiver of a PCM system.

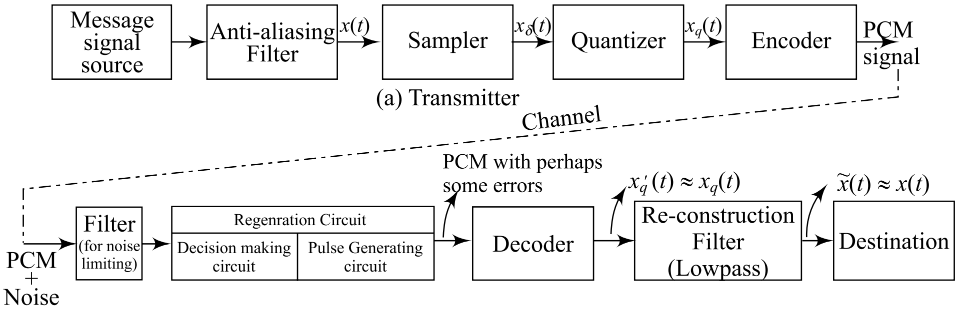


Fig. 9.10 *Transmitter and receiver of a baseband PCM system*

The low pass sampling theorem tells us that for the message signal to be reconstructed from its samples, the sampling frequency, f_s, must be greater than or equal to the Nyquist rate which is equal to $2W$, where W is the bandwidth of the message signal. However, message signals, in practice, can never be strictly band limited. They may have very little energy beyond some frequency say W Hz. Hence, in order to ensure that aliasing does not take place, we have to ensure that they are almost strictly band limited to some known frequency, W Hz, so that we can choose an f_s that is convenient and is greater than $2W$. For this purpose, we use an analog low pass filter of high order with a cut-off frequency of W Hz and pass the original message signal through this filter before it is sampled. Since, this filter is meant to avoid aliasing, it is known as '*anti-aliasing filter*'. The samples are then quantized using a convenient number of uniformly spaced quantization levels, Q, keeping in view the transmission bandwidth constraint and the output *SNR* desired. The quantized samples are then encoded using a binary encoder to obtain a binary PCM signal, which is given as input to the channel over which it is to be transmitted.

Note *In case the message signal is, say, a speech signal, and 'companding' has to be used in order to get a good signal-to-quantization noise ratio without unduly increasing the required transmission bandwidth, we place a compressor between the sampler and the quantizer.*

As shown in Fig. 9.10(b), the output of the channel, which is the PCM signal plus noise, is fed to the 'regeneration circuit', which consists of a decision-making circuit followed by a pulse generator. The output of this regeneration circuit will be a PCM signal devoid of any additive noise, but may have some occasional errors (some of the 1s in the transmitted PCM signal may appear in this as 0s and vice versa). The decoder of the receiver (in fact a D/A converter), converts this digital (PCM) signal into a sequence of samples, i.e., quantized samples. But the output of this decoder, viz; $x'_q(t)$, the sequence of quantized samples, may not be an exact replica of $x_q(t)$, the sequence of quantized samples at the output of the quantizer in the transmitter. The output of the reconstruction filter will not be an exact replica of the band limited message signal, $x(t)$ because of quantization performed at the transmitter and also due to occasional decision-making errors if any, at the receiver.

Note *In case a compressor is used in the transmitter, we place an expander, whose input-output characteristic is the inverse of the input-output characteristic of the compressor used at the transmitting end, between the decoder and the reconstruction filter of the receiver.*

9.3.2 Bandwidth of a PCM Signal

We know that when a sequence of pulses is transmitted through a channel, adequate time-domain resolution of the pulses at the output of the channel would require that the channel bandwidth B_T be such that

$$B_T \geq \frac{1}{2\tau} \tag{9.14}$$

where τ is the pulse width.

For an n-bit binary PCM, if f_s is the sampling frequency employed, n time slots are to be provided in one sampling interval T_s. Thus, in our case

$$\tau = \frac{T_s}{n} = \frac{1}{nf_s} \tag{9.15}$$

$\therefore$ substituting for τ in Eq. (9.14) using Eq. (9.15)

$$B_T \geq \frac{1}{2}(nf_s)$$

But

$$\text{Signaling rate } r = nf_s \quad \text{(in bits/sec.)} \tag{9.16}$$

$\therefore$

$$B_T \geq \frac{r}{2} \tag{9.17}$$

Hence, the minimum bandwidth of an n-bit PCM signal is

$$B_T = \frac{1}{2}r = \frac{1}{2}nf_s\Big|_{f_s=2W} = nW \tag{9.18}$$

where W is the message signal bandwidth.

Example 9.7 Using n-bit binary PCM, a message signal which is uniformly distributed between $-x_{\max}$ and $+x_{\max}$, is transmitted. Show that the signal-to-quantization ratio that is obtained is $6n$ dB.

Solution The mean-squared value of a random variable X which is uniformly distributed between $-x_{\max}$ and $+x_{\max}$ is

$$\overline{X^2} = \int_{-x_{\max}}^{+x_{\max}} x^2 f_X(x)\,dx$$

Hence for the process $x(t)$, i.e., the message signal, which is given to be uniformly distributed between $-x_{\max}$ and $+x_{\max}$

$$\overline{x^2(t)} = \int_{-x_{\max}}^{x_{\max}} x^2 \cdot \left(\frac{1}{2x_{\max}}\right) dx = \frac{x_{\max}^2}{3}$$

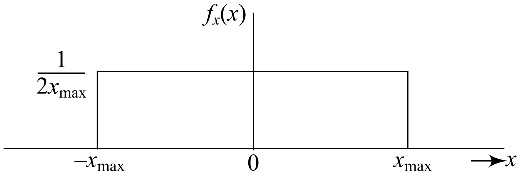


Fig. 9.11 *Message signal for Example 9.7*

If $x(t)$ is normalized so that $|x(t)| \le 1$, then this normalized $x(t)$ will have an average power S given by

$$S = \frac{x_{\max}^2}{3} \cdot \frac{1}{(x_{\max})^2} = \frac{1}{3}$$

But Eq. (9.11)

$$(SNR)_q = 3SQ^2; \quad \text{If } |x(t)| \le 1$$

where $S \triangleq \overline{x^2(t)}$

∴ Substituting for S in Eq. (9.11), we get

$$(SNR)_q = 3 \cdot \left(\frac{1}{3}\right) \cdot Q^2 = Q^2$$

But $Q = 2^n$ $\quad \therefore \quad (SNR)_q = 2^{2n}$ (9.19)

∴ $(SNR)_q$ in dB $= 10 \log_{10} 2^{2n} = 6n$ dB

Example 9.8 If a TV signal of 4.5 MHz bandwidth is transmitted using 8-bit binary PCM, determine

(a) the maximum signal-to-quantization noise ratio
(b) the minimum bit rate
(c) the minimum transmission bandwidth needed.

Solution

(a) We know that $(SNR)_q = 3SQ^2$ (See Eq. (9.11))
where

$$S = \overline{x^2(t)} \quad \text{and} \quad |x(t)| \le 1$$

In this case, we do not know how the $x(t)$, viz. the TV signal, is distributed. So it is not possible to determine $\overline{x^2(t)}$. Since we have been asked to determine only the *maximum* $(SNR)_q$, and since as per Eq. (9.11) $(SNR)_q$ is proportional to S for a fixed Q (or n), let us assume $S = 1$, the maximum possible value it can have under the constraint, $|x(t)| \le 1$.

Then $(SNR)_{q_{\max}} = 3 \times 1 \times Q^2 = 3 \times 1 \times (2^n)^2 = 3 \times 2^{2n}$

∴ $(SNR)_{q_{\max}}$ in dB $= 10 \log_{10} [3 \times 2^{2n}] = (4.8 + 6n)$dB

In our case, n is given to be 8.

$$\therefore \quad (SNR)_{q_{\max}} \text{ in dB} = (4.8 + 6 \times 8) = 52.8 \text{ dB}$$

(b) We know that bit-rate $r = nf_s$

where f_s is the sampling frequency

$\therefore$ minimum bit-rate = $r_{\min} = nf_{s,\min} = 2nW = 2 \times 8 \times 4.5$

$$\therefore \quad r_{\min} = 72.0 \text{ Mbps}$$

(c) $B_T = \frac{1}{2}r \quad \therefore B_{T\min} = \frac{1}{2}r_{\min} = 36.0 \text{ MHz}$

9.3.3 Noise in PCM Systems

As has already been discussed, there are two distinct and *independent* sources of noise in a PCM system. These are:

1. The quantization noise introduced deliberately at the transmitter by rounding off the actual values of the samples of the message signal to the nearest quantization levels so as to discretize the message signal in amplitude also. This noise which originates at the quantizer stage of the transmitter, travels along with the signal all the way through the channel and the receiver right up to the destination. It is message-dependent for its existence and ceases to exist once the message signal is switched off.
2. The channel noise, which is not message dependent, and is always present. It is partly contributed by the electronics of the transmitter and receiver and partly by noises from external agencies entering into the channel.

These two types of noise exist simultaneously in a PCM system. However, the way they affect the performance of the PCM system is different.

Quantization noise, originating in the quantizer of the transmitter in the form of quantization errors, finally manifests as noise at the output of the reconstruction filter of the receiver. Additive random noise of the channel, on the other hand, can cause error in the decision making process at the receiver. Noise peaks occurring during a particular time slot, can force the decision-making circuit to wrongly interpret a 1 as a 0 and a 0 as a 1 in that time slot. Either way, when such a thing happens, an 'error' is said to have occurred in that time-slot and the bit transmitted during that time slot, is received erroneously. In that sense, a *bit-error* is said to have occurred in the code word to which that particular bit belongs. Since peaks of channel noise of sufficient magnitude that can cause decoding errors occur randomly in time, bit-errors also occur randomly in the bit sequence that is transmitted. When we consider a very long sequence of transmitted bits, the average rate at which bit errors occur, is referred to as the '*bit-error rate*'. A typical bit-error rate may be 1 in 10^8 bits. Correspondingly, the probability of a bit being in error, called the *bit-error probability*, P_e, is 10^{-8}.

When *one or more bits* of a code word are erroneously received, a *decoding error* occurs in the sense that the particular code word, when decoded by the decoder, will result in a totally different quantization level than what was actually intended to be conveyed by the original code word. Depending upon which bits of the code word were erroneously received, there can be a drastic change in the decoded value of the quantized sample. When the analog signal is reconstructed from these decoded quantized samples, the decoding error manifests as noise in the output of the reconstruction filter. Since P_e is kept quite small in any practical system and is of the order of 10^{-8}, i.e., on the average, only one in a hundred million bits is affected by noise, and since n, the number of bits in a code word, is only of the order of 6 to 14, it is only very rarely that one of the bits in any code word is affected.

Now consider an n-bit code word. The bit-error probability being P_e, the probability of error for each one of the bits is P_e and so the probability that *any one* of the n bits of the code word is erroneous, is nP_e. Since this in itself is quite small, the probability that two or more bits of that same code word are erroneous would be extremely small. So such an event is almost an impossible event and can be ignored.

Let us now estimate the mean-squared value of the noise generated by one of the bits in a code word being erroneous. The change or error caused in the amplitude of the quantized sample that results from decoding this erroneous codeword, would depend upon the position of the affected bit in the code word. This is because, each digit in a binary number, has a weight that depends on its position, as shown below, where b_ks are either 0 or 1 for k =1 to n

Weight	2^{n-1}	2^{n-2}	$\cdots$	2^1	2^0
Binary number	b_n	b_{n-1}	$\cdots$	b_2	b_1

As the above binary number represents an n-bit codeword, let us consider it as one of the code words transmitted. *Assuming* that this codeword is affected by the channel noise and that one of its bits is changed from 0 to 1 or 1 to 0 in the decision-making process, if the bit b_{n-1} is the one which is erroneously interpreted, the error caused in the value of the codeword is 2^{n-2}. Since all the n-bits of the affected codeword have equal probability of being affected, this probability is $(1/n)$. Hence, the mean-squared value of the error caused in the *value of the codeword* is

$$\overline{\in^2} = \frac{1}{n}\cdot\left(2^{n-1}\right)^2 + \frac{1}{n}\cdot\left(2^{n-2}\right)^2 + \ldots + \frac{1}{n}\left(2^1\right)^2 + \frac{1}{n}\left(2^0\right)^2$$

$$= \frac{1}{n}\left[\left(2^{n-1}\right)^2 + \left(2^{n-1}\right)^2 + \ldots + \left(2^1\right)^2 + \left(2^0\right)^2\right] \tag{9.20}$$

Now, we know that each code word represents one of the Q quantization levels. If $x(t)$, the message signal, has been normalized so that $|x(t)| \le 1$, as we have always been assuming, then the dynamic range of the normalized message signal is 2. Hence, with Q quantization levels, the step size $\Delta = 2/Q$. thus, since the codeword transmitted represents a quantization level number, when the codeword $(b_n b_{n-1} \ldots b_2 b_1)$ is transmitted, it means that the quantized sample value corresponds to the $(b_n, b_{n-1}, \ldots b_2, b_1)^{\text{th}}$ quantization level, and this quantization level corresponds to a *sample value* of $(b_n b_{n-1} \ldots b_2 b_1) \times \left(\frac{2}{Q}\right)$ for the normalized message signal. Hence, when a codeword is affected by noise, the *sample value* represented by that code word has a mean-squared error given by

$$\overline{\in^2} = \frac{1}{n}\left[\left(2^{n-1}\cdot\frac{2}{Q}\right)^2 + \frac{1}{n}\cdot\left(2^{n-2}\cdot\frac{2}{Q}\right)^2 + \ldots + \frac{1}{n}\left(2^1\cdot\frac{2}{Q}\right)^2 + \frac{1}{n}\left(2^0\cdot\frac{2}{Q}\right)^2\right]$$

$$= \frac{4}{nQ^2}\sum_{k=0}^{n-1} 2^{2k} = \frac{4}{nQ^2}\left[\frac{2^{2n}-1}{(2^2-1)}\right] = \frac{4}{nQ^2}\frac{(Q^2-1)}{3} \approx \frac{4}{3n} \quad \text{since generally } Q >> 1.$$

(Sum of n terms of a geometric progression with common ratio r equal to 2^2.)

$$\therefore \qquad \overline{\in_n^2} = \left(\frac{4}{3n}\right)$$

This is the mean-squared error in the sample value represented by a codeword, *if that codeword is erroneous*, i.e., affected by channel noise.

But as we have already stated, the probability that *any one* of the bits of an n-bit codeword would be affected, i.e., that the codeword would be erroneous in nP_e, where P_e is the bit-error probability.

Hence mean-squared value of the noise caused by the random noise of the channel is given by

$$\overline{\in_n^2} = \left(\frac{4}{3n}\right)nP_e = \frac{4}{3}P_e \tag{9.21}$$

Since the mean-squared value of the quantization noise for Q levels and $|x(t)| \le 1$, is given by

$$e_q^2 = \frac{1}{3Q^2} \qquad \text{(Refer to Eq. (9.9))}$$

and since the two noise process are independent, the mean-squared value of the *total noise* is given by the sum of their individual mean-squared values.

$\therefore$ N_D = mean-squared value of the total noise at destination

$$= \overline{e_q^2} + \overline{\epsilon_n^2} = \frac{1}{3Q^2} + \frac{4}{3}P_e$$

$$\therefore \quad N_D = \frac{1 + 4Q^2 P_e}{3Q^2} \tag{9.22}$$

Hence, if $\overline{x^2}$ represents the mean-squared value of the normalized message signal

$$\boxed{\underset{\text{Binary PCM}}{\left(\frac{S}{N}\right)_D} = \text{Destination } SNR = \frac{3Q^2\overline{x^2}}{1 + 4Q^2 P_e}} \tag{9.23}$$

Note

The above expression for (S/N)$_D$ takes into account both quantization noise as well as channel noise.

Remarks

From Eq. (9.23), it follows that

(i) When $4Q^2 P_e << 1$, i.e., when $P_e << \frac{1}{4Q^2}$, the (SNR)$_D$ is given by

$$\left(\frac{S}{N}\right)_D = 3Q^2\overline{x^2} \tag{9.24}$$

This means that when $P_e << \frac{1}{4Q^2}$, the quantization noise dominates and (S/N)$_D$ is decided only by this noise.

(ii) When $4Q^2 P_e >> 1$, i.e., when $P_e >> \frac{1}{4Q^2}$, the (S/N)$_D$ is given by

$$\left(\frac{S}{N}\right)_D = \frac{3\overline{x^2}}{4P_e} \tag{9.25}$$

This means that when $P_e >> \frac{1}{4Q^2}$, the channel noise dominates and the (S/N)$_D$ is decided only by this noise.

In the first case, when the quantization noise dominates as per Eq. (9.24) we find that $(S/N)_D$ is independent of the probability of the error P_e. However, as we will be deriving later, P_e is related to the input SNR, i.e., $(S/N)_R$ of the receiver and is given by

$$P_e = \begin{cases} Q\left[\sqrt{(S/N)_R}\right] & \text{for bipolar signals} \quad (9.26a) \\ Q\left[1/2\sqrt{(S/N)_R}\right] & \text{for unipolar signals} \quad (9.26b) \end{cases}$$

assuming the channel noise to be Gaussian and white.

Note *Q[x] denotes Q-function with argument x. (See Appendix B)*
(We will be deriving the result of Eq. (9.26) in Chapter 10.)

Further, when $P_e >> \left(\frac{1}{4Q^2}\right)$ and the channel noise dominates, from Eq. (9.25) we find that $(S/N)_D$ decreases as P_e increases. However, from Eq. (9.26), we find that P_e increases when $(S/N)_R$ decreases, since the *Q-function* $Q[x]$ is a monotonically decreasing function of its argument, x. Using these relationships, we shall now plot the $(S/N)_D$ vs. $(S/N)_R$ for a binary PCM system employing bipolar signaling, assuming that the message signal $x(t)$ is a sinusoidal signal normalized so that $|x(t)| \le 1$ (This means that $\overline{x^2} = 1/2$). The channel noise is assumed to be white Gaussian. The curves have been plotted for four values of Q. We find that there is a sharp decline in the destination *SNR* as the input *SNR* is decreased, indicating a threshold effect. The *threshold effect* manifests in the form of increased decoding errors as the input (*SNR*) is decreased below a certain value called the '*threshold*', whose value depends upon the value of Q. When the probability of error, P_e becomes very much larger than $1/4Q^2$ — and this happens when $(S/N)_R$ is lower than the threshold value, decoding errors occur very frequently making the reconstructed message signal at the output of the receiver totally different from the transmitted message signal.

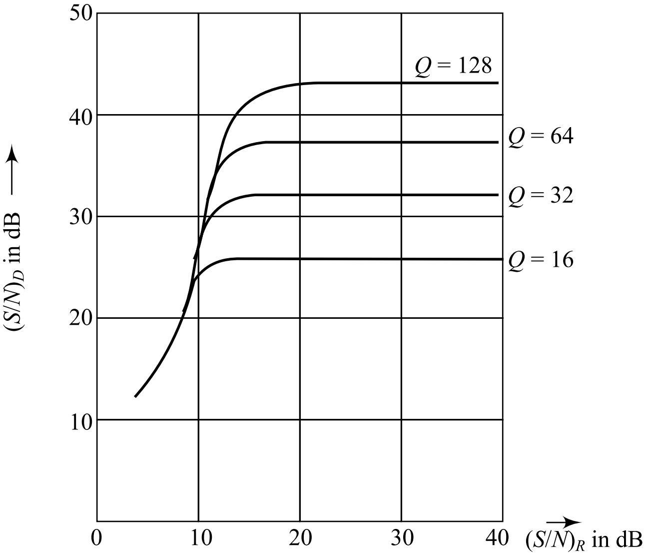


Fig. 9.12 *Noise performance of binary PCM*

There is no unique way of defining 'threshold'. Since the onset of threshold is accompanied by a rapid increase in the probability of error, P_e and a consequent rapid fall in the $(S/N)_D$, either of these two parameters may be used to define threshold. *In both the cases, the definition is rather arbitrary*. The following are the two ways it is defined:

Definition 1 The threshold value of the $(S/N)_R$ is that value of $(S/N)_R$ at which the decoding noise reduces the $(S/N)_D$ by 1 dB compared to the value it has for large values of $(S/N)_R$, when quantization noise dominates.

Definition 2 The threshold is that value of $(S/N)_R$ for which the bit-error probability P_e is greater than 10^{-5}.

The first definition is more widely used, but it is difficult to use it in analysis. The second one is easier to use in mathematical analysis.

Example 9.9 A binary PCM system, employing bipolar signaling, uses 256 quantization levels. Calculate the bit-error probability, P_e at the threshold for this system. Use Definition 1. Also find the corresponding $(S/N)_R$ at threshold.

Solution From Eq. (9.23), we have

$$\left(\frac{S}{N}\right)_D = \frac{3Q^2\overline{x^2}}{1+4P_eQ^2}, \text{ where } Q \text{ is the number of levels used.}$$

As per definition 1, the threshold is reached when $(S/N)_D$ drops by 1 dB from the constant value it has for high values of $(S/N)_R$. We know that the value of $(S/N)_D$ when the quantization noise dominates, is given by

$$\left(\frac{S}{N}\right)_D = 3Q^2\overline{x^2}$$

If threshold condition is reached when $(S/N)_D$ falls by 1 dB relation to $3Q^2\overline{x^2}$, , it means that if we take the expression for $(S/N)_D$ given by Eq. (9.23) which takes both the noises into account, the dominator $(1 + 4P_eQ^2)$ becomes equal to 1 dB at threshold.

$\therefore$ $$10\log_{10}(1+4P_eQ^2) = 1 \text{ dB at threshold,}$$

$\therefore$ if $P_{e\text{ th}}$ is the bit-error probability at threshold

$$10\log_{10}(1+4P_{e\text{th}}Q^2) = 1 \quad \text{or} \quad (1+4P_{eth}Q^2) = 10^{0.1} = 1.2589$$

Since $Q^2 = (256)^2$, we have $1 + 4(256)^2 P_{e\text{th}} = 1.2589$

$\therefore$ $$P_{e\text{th}} = \frac{0.2589}{4\times(256)^2} = 9.876\times10^{-7}$$

To find the corresponding $(S/N)_{R\text{ th}}$, we use Eq. (9.26a)

$\therefore$ $$P_{e\text{th}} = Q\left[\sqrt{(S/N)_{R\text{th}}}\right] = 9.876\times10^{-7}$$

$\therefore$ $$\left(\frac{S}{N}\right)_{R\text{th}} = 24 \qquad \text{or} \qquad 13 \text{ dB (i.e., } 10\log_{10} 24)$$

Example 9.10 A message signal of bandwidth 4 kHz is sampled at twice the Nyquist rate and the samples are transmitted by unipolar n-bit binary PCM. If an output *SNR* of 47 dB is required, determine (a) n, the number of digits per codeword, (b) the minimum value of (S_R/η) required to maintain the operation of the system above the threshold. Assume that the message signal is normalized so that $|x(t)| \le 1$ and that it is uniformly distributed.

Solution

(a) Since the system will be operated well above the threshold so that only the quantization noise needs to be considered

$$\left(\frac{S}{N}\right)_D = 3Q^2\overline{x^2} = \frac{3}{3}Q^2 = 47 \text{ dB} = 501187 \text{ (ratio)}$$

(Note that $\overline{x^2}$ is taken as 1/3 since $x(t)$ is uniformly distributed and $|x(t)| \le 1$.)

$\therefore$ $$Q = 707.9$$

Since Q has to be an integer and a power of two in the case of binary PCM, let us choose

$$Q = 1024 = 2^{10}$$

$\therefore$ the required value of n is 10. This will infact give better than 47 dB output *SNR*.

(b) At threshold, $10\log_{10}(1+4P_{e\text{th}}Q^2) = 1$ dB

$\therefore$ $$(1+4P_e\,Q^2) = 10^{0.1} = 1.2589$$

$\therefore$ $$P_{e\text{th}} = \frac{0.2589}{4\times(1024)^2} = 6.1726\times10^{-8}$$

But $$P_{e\text{th}} = Q\left[\sqrt{1/2(S/N)_{R\text{th}}}\right] \qquad \text{(See Eq. (9.26b))}$$

From the Q-function graph, for a $P_{e\text{th}} = 6.1726\times10^{-8}$ the value of $Q[\sqrt{1/2(S/N)_{R\text{th}}}] = 5.3$

$$\therefore \qquad \left(\frac{S}{N}\right)_{R,\text{th}} = 56.18$$

But $N_R = \frac{\eta}{2} \times 2B_T = \eta B_T$, where B_T = transmission bandwidth

But $B_T = nW$ $\quad\therefore\ N_R = \eta nW = \eta \times 10 \times 4 \times 10^3$

$$\therefore \qquad \left(\frac{S}{N}\right)_{R,\text{th}} = \left(\frac{S_R}{\eta \times 40 \times 10^3}\right) = 56.18$$

$$\therefore \qquad \left(\frac{S_R}{\eta}\right) = 2.2472 \times 10^6$$

9.3.4 Bandwidth-Power Trade-Off in PCM Systems

We have seen (refer to Eq. (9.18)) that the minimum bandwidth required for an n-bit *baseband* PCM signal is given by

$$B_T = nW \tag{9.27}$$

where W is the bandwidth of the message signal. Hence, the transmission bandwidth is directly proportional to n, the number of bits used per codeword.

The destination *SNR*, viz. $(S/N)_D$, for binary PCM, *assuming operation of the system well above the threshold, so that we can ignore decoding or channel noise*, is given by

$$\left(\frac{S}{N}\right)_D = 3SQ^2 = 3S2^{2n}$$

where S is the average power of the normalized message signal, normalized so that $|x(t)| \le 1.$ If this signal is uniformly distributed, we know that $S = 1/3$ (refer to Example 9.7). Therefore, in that case

$$\left(\frac{S}{N}\right)_D = 2^{2n} = 2^{2\left(\frac{B_T}{W}\right)} = 2^{2\mathsf{B}} \tag{9.28}$$

where $\mathsf{B} \triangleq (B_T/W)$

More than the exact expression, what is important to us right now, is the fact that the output SNR of PCM is exponentially related to n.

From Eq. (9.28) it is clear that as n increases, $(S/N)_D$ increases rapidly (*exponentially*). But at the same time, the required transmission bandwidth B_T also increases, *but only linearly with n*. Thus, without increasing the transmitter power (in order to increase $(S/N)_D$), we can just increase n and get an improved destination *SNR* – but at a price. The price is the consequent increase in required transmission bandwidth as n is increased. Thus, we can save power at the cost of bandwidth and vice versa or in other words, *there is a power-bandwidth trade-off possible in PCM*.

This trade-off, however, is not without a limit. It is not possible to maintain a certain destination SNR if we go on reducing the transmitter power and try to compensate for it by increasing n, and thereby the bandwidth. This is because, as the bandwidth increases, the channel noise power entering the receiver, which is equal to

$$N_R = \left(\frac{\eta}{2}\right)2B_T$$

also goes on increasing and soon a stage will be reached when the bit-error probability P_e becomes quite high (due to increased decoding errors caused by large N_R) and our assumption that only quantization noise need be considered and channel noise can be ignored, used in deriving Eq. (9.28), is violated. Then the trade-off is not possible.

However, as long as we are operating the system well above the threshold, the trade-off is possible. It may be noted in this context, that the terms of the trade-off are much better in the case of PCM as compared to either wideband FM or PPM. This is because $(S/N)_D$ increases *exponentially with bandwidth* in the case of PCM while it increases approximately as only the square of the bandwidth in the case of WBFM and PPM.

9.3.5 Regenerative Repeaters and Long-Haul Transmission of PCM Signals

The term, 'long-haul' transmission, refers to a situation wherein the transmitter and the final destination, i.e., the receiver, are separated by very long distance. In such a situation, the signal has to be amplified and retransmitted, by what are called repeaters, at a number of intermediate points so as to take care of the attenuation that the signal suffers. In such a long-haul transmission, PCM, by virtue of its being a digital modulation which permits the use of regenerative repeaters, has a definite edge over any analog modulation. In what follows, we shall discuss this aspect in some detail.

First, let us examine as to what happens when we use analog modulation for long-haul transmission. The repeater receives the signal from the transmitter, or may be the preceding repeater, amplifies the received signal and re-transmits it. When the received signal is amplified, the signal as well as the noise which has contaminated it during the course of its travel through the channel are both amplified to the same extent. In fact, the amplifier adds some more noise – its internally generated noise, with the result that the *SNR* at the output of any repeater will always be poorer than the *SNR* at its input. So, when there are a large number of repeaters, as required in a long-haul transmission, the *SNR* at the destination will be so low that the signal is likely to be completely drowned in noise. This is because of the fact that noise goes on accumulating and increases at each repeater.

On the other hand, suppose we use pulse-code modulation. If, in the design of the system, it is ensured that the input *SNR* at each repeater is large enough and if regenerative repeaters are used, the effect of channel noise can be made negligibly small even when a large number of repeaters are used. A regenerative repeater receives the PCM signal from the preceding repeater and decodes it. Depending upon the decision made by the decision-making section of the decoder during each time slot, whenever the decision is that a 1 has been received, the pulse generator section of the decoder produces a *clean* rectangular pulse. If there are no decision-making errors in the decoding at a repeater, its output pulse sequence will be an exact replica of the pulse sequence transmitted by the transmitter. Thus, the effect of channel noise may be almost completely eliminated. The problem then boils down to one of ensuring that there are no decoding errors at the repeaters.

Since the Q-function, $Q(x)$, is a monotonically decreasing function of its argument, x, from Eqs. (9.26a) and (9.26b), we find that it is possible to make the bit-error probability, P_e, at the output of each regenerative repeater extremely small by making the input *SNR* for the decoder of a repeater adequately high. A high input SNR for a repeater can be ensured either by increasing the transmitted power of the previous repeater, or by decreasing the distance between consecutive repeaters, or by a combination of both.

Thus, we find that the use of PCM together with regenerative repeaters helps us to almost eliminate the effect of channel noise in long-haul circuits.

9.3.6 Power-Bandwidth Trade-Off and a Comparison of PCM and WBFM

From Eq. (9.28), we found that a power-bandwidth trade-off, or exchange, is possible in the case of PCM. In analog modulation too, in the case of WBFM we found that such an exchange is possible. So, a comparison of PCM with these analog modulation schemes which also offer this power-bandwidth exchange will be interesting.

In the case of WBFM, the $(S/N)_D$ increases as the square of the modulation index. Since the effective bandwidth of WBFM is *almost* directly proportional to the modulation index, β_f (refer to Carson's rule), it means that the destination *SNR* of WBFM increases as the square of the bandwidth. However, in the case of PCM, the destination *SNR* increases exponentially with the bandwidth, i.e., $(SNR)_D$ of PCM increases at a much faster rate with bandwidth, than the (*SNR*) of WBFM. Thus, PCM offers a much better power-bandwidth exchange, than WBFM.

Advantages of PCM

1. It offers a very efficient power-bandwidth exchange.
2. It is very robust since it is almost immune to channel noise and interference.
3. Because of the possibility of the use of regenerative repeaters, it is extremely useful in long-haul communication.
4. It makes it possible to integrate baseband signals of different types, like audio, video, etc., into a common format for easy multiplexing using TDM.
5. Because of its digital nature, coding techniques are available for PCM signals for efficient compression, encryption and error-correction.
6. In a TDM system, it is relatively easy to either *add* or *drop* any message signal and PCM signals can easily be TDM*ed*.

Disadvantages of PCM

1. PCM signal generation and reception involve complex processes and require the use of somewhat complex systems. This, however, does not constitute a real problem nowadays because of the availability of VLSI chips for performing these various operations.
2. For the same message bandwidth, PCM requires a much larger transmission bandwidth than some of the analog modulation schemes like AM. For instance, if message bandwidth, *W* is 3.4 kHz as in the case telephone quality voice channel, AM requires a bandwidth of only 6.8 kHz; whereas PCM with standard 8 kHz sampling frequency and 8-bits per codeword, requires 64 kbps bit rate, or a 32 kHz transmission bandwidth. This is nearly five times the bandwidth required for AM.

However, this is not a serious problem with the availability of very efficient data compression techniques as well as wideband satellite channels and fiber-optic channels which can support high data rates.

9.4 DELTA MODULATION (DM)

Whenever we sample a voice, or video signal a little above the Nyquist rate, the adjacent samples are correlated, indicating that there is redundancy. Transmitting these samples using conventional PCM as discussed in the previous section, may not therefore be the most efficient way of transmitting information. Suppose we use a much higher sampling frequency than the Nyquist rate, i.e., we over-sample the message signal. Adjacent samples will now be highly correlated. Thus, the sequence representing the *difference between adjacent samples* will have a much smaller variance than the variance of the original sequence of samples. Therefore, it has a much smaller dynamic range and needs a much smaller number of bits for a reasonably accurate representation of its sample values. This, in general, is the basic principle of differential pulse-code modulation, or DPCM. As an extreme case, if we use a single-bit representation for the sample values in the difference sequence, it is called Delta Modulation (DM).

A delta modulator simply consists of a comparator, a single-bit quantizer (which is nothing but a hard limiter) and an accumulator, connected together as shown in Fig. 9.14. Thus, in delta modulation, we compare the present sample $x(n)$ of the message signal $x(t)$ with an approximation to the previous sample, viz. $\tilde{x}(n-1) = x_q(n-1)$, and the difference between $x(n)$ and $x_q(n-1)$, i.e., $e(n)$ is applied to the single-bit quantizer, or the hard limiter. So, if the output of the comparator is positive, whatever may be its actual magnitude, the hard limiter gives an output of $+\Delta$ and if the comparator output is negative, irrespective of its actual amplitude, the hard-limiter output will be $-\Delta$. This output of the hard limiter which we shall call as $e_q(n)$ goes to the encoder and through it, to the channel. $e_q(n)$ is also used for generating $x_q(n-1)$, i.e., the

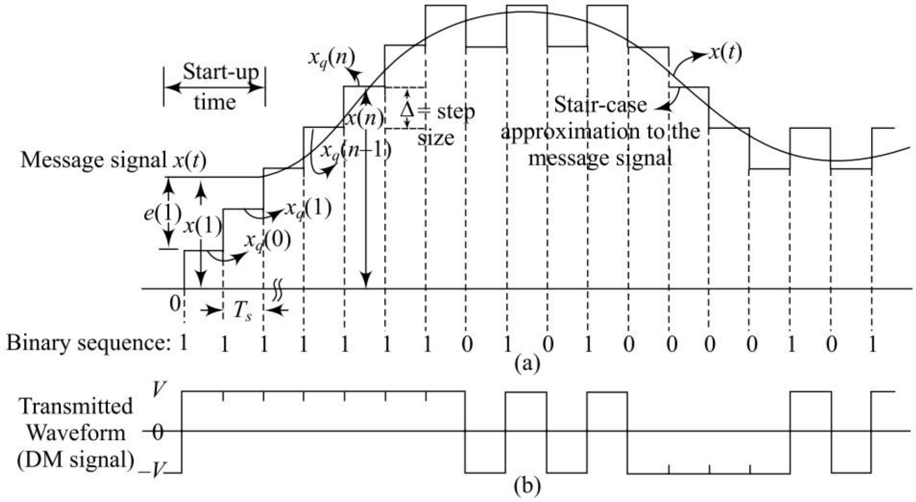
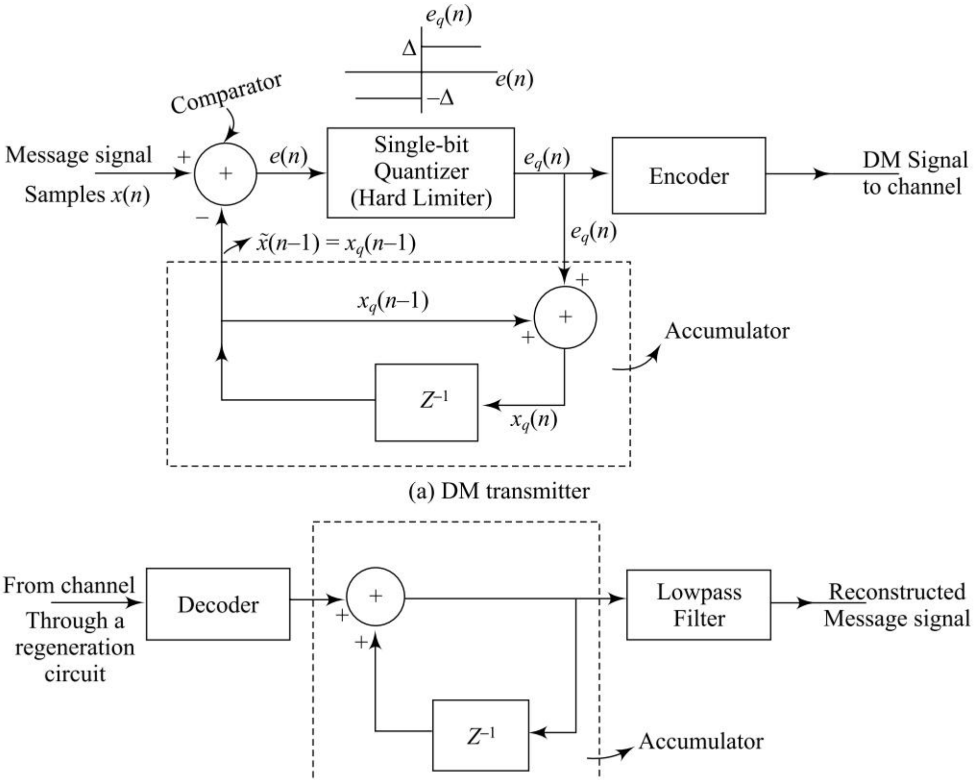


Fig. 9.13 *Illustrating the working of a Delta modulator*



(b) DM Receiver

Fig. 9.14 *Delta modulation system*

approximation to the previous sample $x(n-1)$ of $x(t)$, by giving it as input to the accumulator, as shown in Fig. 9.14.

As shown in Fig. 9.13(a), initially the '*approximate signal'* starts building up from zero to catch-up with the message signal. As shown there, $x(T_s) = x(1)$ is greater than $x_q(0) = \Delta$. So, $e(1) = \lfloor x(1) - x_q(0) \rfloor$ and $e_q(1) = +\Delta$. In the accumulator, this adds to $x_q(0)$ and gives $x_q(1)$ which is equal to $x_q(0) + e_q(1) = x_q(0) + \Delta$, as shown in the figure. Thus, the Delta modulator produces a staircase approximation to the message signal by trying to track the message signal as per the following defining equations:

$$\left.\begin{aligned} e(n) &= x(n) - x_q(n-1) \\ e_q(n) &= \Delta \cdot \mathrm{sgn}[e(n)] \\ x_q(n) &= x_q(n-1) + e_q(n) \end{aligned}\right\} \tag{9.29}$$

where sgn(z) is called the signum function and is defined by

$$\mathrm{sgn}(z) = \begin{cases} 1 & \text{if } z > 0 \\ -1 & \text{if } z < 0 \end{cases}$$

When $e_q(n) = +\Delta$, it is a binary 'one' and whenever $e_q(n)$ is $-\Delta$, it is a binary 'zero'. This sequence of $e_q(n)$'s is encoded by the encoder whose output waveform is shown in Fig. 9.13(b). This is the binary waveform transmitted over the channel.

When this waveform travels through the channel, it will be distorted and also corrupted by additive noise. So, in the receiver, a regeneration circuit, consisting of a decision-making part and a pulse generator part, will first decide, during each time slot, whether a positive pulse, or a negative pulse is received during that time slot and accordingly produces clean rectangular positive or negative pulse during that time slot. These are then decoded by the decoder which gives at its output a $+\Delta$ volts or a $-\Delta$ volts at each sampling instant. This sequence is then fed to an accumulator in the receiver. Recall that $e_q(n)$s, i.e., a sequence of $e_q(n)$ or $-e_q(n)$ fed to the accumulator in the transmitter resulted in the generation of $x_q(n)$s by the addition of $x_q(n-1)$ to $e_q(n)$. Similarly, the receiver accumulator output gives the staircase approximation signal. The low pass filter, having a cut-off frequency of W Hz (frequency to which message signal, $x(t)$ is band limited), removes the high frequency out of band noise components present in the staircase waveform and gives at its output, a waveform that closely approximates $x(t)$.

9.4.1 Slope Overload Noise and Step Size

As can be seen from Fig. 9.13(a), when the delta modulator tracks the message signal, the $x_q(t)$ waveform (i.e., the approximation waveform) increases or decreases linearly with time since the step size is constant. For this reason, this type of delta modulator using a fixed step size, is sometimes referred to as a *linear delta modulator*.

Although it has been shown in that figure that the delta modulator is perfectly tracking the message signal, $x(t)$, it may not always do so. The average rate of increase or decrease of $x_q(t)$, the staircase approximation, is given by (Δ/T_s). If this is smaller than the maximum rate of change of $x(t)$, the message, the Linear Delta Modulator (LDM) will not be able to track the $x(t)$ properly and the approximation, i.e., the staircase waveform will be very much different from the message signal, $x(t)$, as shown in Fig. 9.15. This inability of the LDM to correctly track the message signal, $x(t)$ when $x(t)$ has steep changes, is referred to as 'slope overload' condition. This leads to severe distortion of the reconstructed message signal and appears as noise at the destination. Therefore it is to be avoided. Thus, 'slope overload' in an LDM system can be avoided by ensuring that

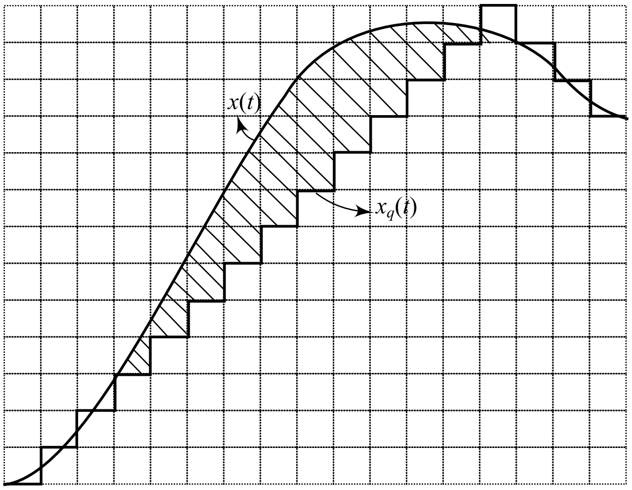


Fig. 9.15 *Illustrating the phenomenon of slope overload in linear Delta modulation*

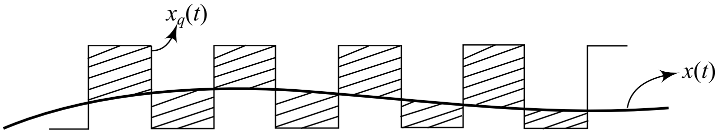


Fig. 9.16 *Granular noise*

$$\frac{\Delta}{T_s} > \left|\frac{dx(t)}{dt}\right|_{\max} \tag{9.30}$$

where Δ is the fixed step size and $T_s = 1/f_s$ is the sampling interval. Therefore, with a fixed sampling frequency, the step size has to be large in order to avoid 'slope overload'.

But, consider the situation when the message signal is relatively flat. As may be seen from Fig. 9.16, the granular noise, arising from the 'hunting' that takes place when the signal is not changing much, will increase as the stepsize Δ is increased. This granular noise is similar to the quantization noise of PCM.

Thus, we find two conflicting requirements – a large step size in order to avoid slope-overload noise, and a small step size to reduce the granular noise. A solution for this problem would be to use an adaptive system, in which the step size automatically varies with the rate of change of $x(t)$, the message signal, giving a large step size when $\left|\frac{dx(t)}{dt}\right|$ is large and a small step size when it is small. But, before we discuss the adaptive delta-modulation system, let us examine the noise performance of a linear delta modulation system.

Destination SNR for linear Delta modulation (only granular noise)

We shall now determine the destination signal-to-granular noise ratio for a linear delta modulation system, assuming a sinusoidal message signal. It is assumed that for the step size and sampling frequency of the system the peak-amplitude and frequency of the assumed sinusoidal message signal are such that there is no slope-overload noise.

Referring to Fig. 9.17(b), it is clear that the value of quantization error is limited to ± Δ, where Δ is the step size. But we do not know how this error is distributed within these limits. If the sampling frequency is high, as is always the case with delta modulation, it is quite reasonable to assume that the error, e_q is uniformly distributed between –Δ and +Δ. Further, it has been experimentally established that the two-sided power spectrum of this periodic error waveform (it will be periodic if T, the period of the message signal is an integer multiple of the sampling interval T_s) is white (i.e., constant) and extends from $-f_s$ to $+f_s$.

Since e_q, the error is uniformly distributed between – Δ and + Δ, its mean-square value is given by

$$\overline{e_q^2} = \int_{-\Delta}^{\Delta} e_q^2 f_{eq}(e_q)\,de_q = \int_{-\Delta}^{\Delta} e_q^2\left(\frac{1}{2\Delta}\right)de_q = \frac{\Delta^2}{3} \tag{9.31}$$

Since the power spectral density of this quantization noise is white and from $-f_s$ to $+f_s$ and since the total area under any PSD curve of a signal must be equal to the average power of the signal, the PSD is given by

$$PSD = \frac{\text{total power}}{2f_s} = \left(\frac{\Delta^2/3}{2f_s}\right) \text{since PSD curve is a rectangle}$$

$$\therefore \quad P(f) = \text{PSD} = \frac{\Delta^2}{6f_s}; \quad -f_s \le f \le f_s \tag{9.32}$$

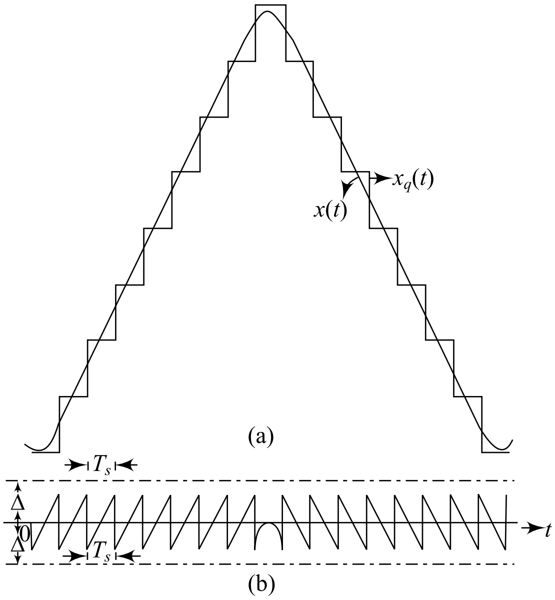


Fig. 9.17 *(a) Staircase approximation for an assumed sinusoidal message signal, (b) Waveform granular or quantization noise, $e_q(t)$*

As there is a low pass reconstruction filter as the last stage of the DM receiver, and since its cut-off frequency is W Hz, the bandwidth of the message signal, and since $f_s >> W$, the full quantization noise power equal to $\overline{e_q^2}$ which has been shown to be given by ($\Delta^2/3$), does not pass through this low pass filter and reach the destination

$$\therefore \qquad N_D = \text{Destination noise power} = \left(\frac{\Delta^2}{3}\right)\cdot\left(\frac{2W}{2f_s}\right) = \frac{\Delta^2}{3}\cdot\left(\frac{W}{f_s}\right) \tag{9.33}$$

Now, let us determine the destination signal power. Let the sinusoidal message signal be represented by

$$x(t) = A \sin \omega_m t$$

The maximum rate of change of this signal is

$$\left|\frac{d}{dt}x(t)\right|_{\max} = |A\omega_m \cos \omega_m t|_{\max} = A\omega_m \tag{9.34}$$

Since we have assumed that there is no slope overload noise, as per Eq. (9.30)

$$A\omega_m < \frac{\Delta}{T_s} \tag{9.35}$$

Since the maximum rate of change of the sinusoidal message signal depends on the product of its peak amplitude and its frequency, and since this product has got to be less than (Δ/T_s), let us choose the worst case scenario by assuming $f_m = W$, the highest frequency in the passband of the baseband low pass filter of the receiver, and then determine A, the peak-amplitude of the sinusoidal message signal as the maximum permissible peak amplitude for that sinusoid of frequency W Hz under the 'no slope overload' constraint.

∴ Putting $f_m = W$ in Eq. (9.35) and evaluating A, we get

$$A < \left(\frac{\Delta f_s}{2\pi W}\right) \tag{9.36}$$

Therefore we shall assume that A is chosen so as to have the highest possible value. In that case,

$$x(t) = \left(\frac{\Delta f_s}{2\pi W}\right)\sin 2\pi Wt \tag{9.37}$$

The signal power at the destination is therefore given by

$$S_D = \frac{A^2}{2} \le \frac{\Delta^2 f_s^2}{2W^2 4\pi^2} \tag{9.38}$$

The destination SNR is therefore equal to

$$\frac{S_D}{N_D} = \left(\frac{S}{N}\right)_D \le \frac{\Delta^2 f_s^2}{2\times 4\pi^2 W^2}\,\frac{3f_s}{\Delta^2 W}$$

∴

$$\left(\frac{S}{N}\right)_D \le \left(\frac{3}{8\pi^2}\right)\left(\frac{f_s}{W}\right)^3 \tag{9.39}$$

Example 9.11 A DM transmitter with a fixed step of 0.5 V, is given a sinusoidal message signal. If the sampling frequency is twenty times the Nyquist rate, determine (a) the maximum permissible amplitude of the message signal, if slope overload is to be avoided, and (b) the maximum destination SNR under the above condition.

Solution

(a) Let A be the maximum permissible peak amplitude of the sinusoidal message signal for avoiding slope overload. Then

$$A < \frac{\Delta}{2\pi}\left(\frac{f_s}{W}\right) \qquad \text{(See Eq. (9.36))}$$

Since Nyquist rate is $2W$ samples per second for a signal of frequency W Hz, and since the sampling frequency f_s is given to be 20 times the Nyquist rate, f_s is given by

$$f_s = 20\times 2W = 40W \quad \therefore \left(\frac{f_s}{W}\right) = 40$$

∴ the maximum permissible value of A is

$$A_{\max} = \frac{\Delta}{2\pi}\times 40 = \frac{0.5\times 40}{2\pi} = \frac{10}{\pi} = 3.18 \text{ volts}$$

(b) From Eq. (9.39), we have

$$\left(\frac{S}{N}\right)_D \le \left(\frac{3}{8\pi^2}\right)\left(\frac{f_s}{W}\right)^3$$

Note *The 'equal to' sign holds when A is chosen as the maximum value according to Eq. (9.36).*

∴ the maximum destination SNR is given by

$$\left(\frac{S}{N}\right)_{D,\max} = \left(\frac{3}{8\pi^2}\right)(40)^3 = \frac{3\times 64000}{8\pi^2} = 2432$$

$$\therefore \qquad \left(\frac{S}{N}\right)_{D,\max} \text{ in dB} = 10\, log_{10}\, 2432 = 33.85 \text{ dB}$$

Example 9.12 The pulse rate in a DM system is 56,000 per sec. The input signal is $5\cos(2\pi\, 1000t) + 2\cos(2\pi\, 2000t)$ volts, with t in seconds. Find the minimum value of step-size which will avoid slope overload distortion. What would be the disadvantages of choosing a value larger than the minimum? **(GATE Examination, 1998)**

Solution We have to find the maximum slope, *not individually of each component of the message signal* $x(t)$, but that of $x(t)$ itself and then find the minimum value of step size that can avoid slope overload even when this maximum slope of $x(t)$ is encountered.

$$x(t) = 5\cos 2000\pi t + 2\cos 4000\pi t$$

Putting $\omega_1 = 2000\pi,$ we have

$$x(t) = 5\cos \omega_1 \pi t + 2\cos 2\omega_1 \pi t$$

$$\left|\frac{d}{dt}x(t)\right|_{\max} = \left|-5\omega_1 \sin \omega_1 t - 2\times 2\omega_1 \sin 2\omega_1 t\right|_{\max}$$

$$= \left|5\omega_1 \sin \omega_1 t + 4\omega_1 \sin 2\omega_1 t\right|_{\max}$$

To find the maximum value of $5\omega_1 \sin\omega_1 t + 4\omega_1 \sin 2\omega_1 t \triangleq y(t)$, differentiate it with respect to t and equate the derivative to zero.

$$\frac{dy(t)}{dt} = 5\omega_1^2 \cos\omega_1 t + 8\omega_1^2 \cos 2\omega_1 t = 0$$

Putting this $\omega_1 t$ as θ for convenience

$$5\omega_1^2 \cos\theta + 8\omega_1^2 \cos 2\theta = 0 \quad \therefore \frac{\cos 2\theta}{\cos\theta} = -\frac{5}{8} = -0.625$$

Solving the above transcendental equation for θ, we get $\theta = 55.4°$

$\therefore$ $x(t)$ has maximum absolute value of slope when $\omega_1 t = 55.4°$; and this maximum value is given by

$$\left|\frac{d}{dt}x(t)\right|_{\max} = 5\times 2000\pi \sin 55.4° + 2\times 4000\pi \sin 110.8° = 26080.689$$

If Δ is the minimum step size required to avoid slope overload noise, we know from Eq. (9.30) that

$$\Delta_{\min} = \left|\frac{dx(t)}{dt}\right|_{\max} \div f_s = \frac{26080.689}{56000} = 0.4657 \text{ V}$$

Granular noise will increase unnecessarily, if we use a step size larger than this required minimum value.

Advantages and disadvantages of DM

1. DM transmitter and receiver require very simple and inexpensive hardware. Although a higher sampling rate is used in DM, since only one bit is used to represent the error which is actually transmitted, the bit rate is, however, not too high; but will generally be higher than the bit rate of PCM.
2. If the input message waveform has steep gradients, severe slope overload distortion results, since the step size is fixed.
3. $e(n)$, which is given as input to the single-bit quantizer, or hard limiter, in the DM transmitter, is the difference between the present sample and an estimate of the previous sample. This operation of taking the difference amounts to discrete-time differentiation of the input. Because of this, transmission channel noise can cause accumulation of errors in the receiver.

9.4.2 Delta-Sigma Modulation System

The last drawback of DM mentioned earlier, can be overcome by the use of a Sigma–Delta modulation system, which is more commonly known as the Delta–Sigma modulation system although in fact, the 'sigma operation' or 'integration' precedes the Delta modulator in this system.

Integration of the message signal prior to its Delta modulation gives the following advantages:

1. Integration de-emphasizes sudden changes, i.e., high frequency content and emphasizes the low frequency content contributed by slow changes. Since the variation from one sample to the next is thus reduced, the error variance is reduced. This permits the Delta modulator to perform well even with slightly lower sampling frequencies than those used in the LDM system.
2. The receiver becomes extremely simple – just an LPF.

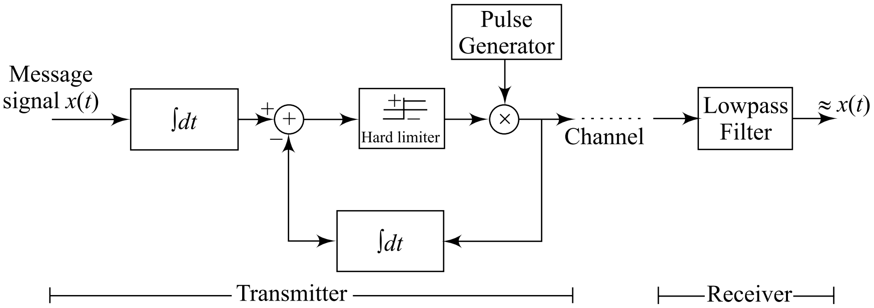


Fig. 9.18 *A Delta–Sigma modulation system*

Figure 9.18 shows the block diagram of a Delta–Sigma modulation system. As the message is subjected to integration before Delta modulation (at the transmitter), an inverse operation, viz. differentiation needs to be done in the receiver. As the DM receiver anyhow has an integrator (accumulator), these two cancel and so the receiver will be just only the baseband filter, which is an LPF.

The transmitter part of the Delta–Sigma modulation system of Fig. 9.18 can be further simplified. As integration is a linear operation, we can combine the two integrators as shown in Fig. 9.19.

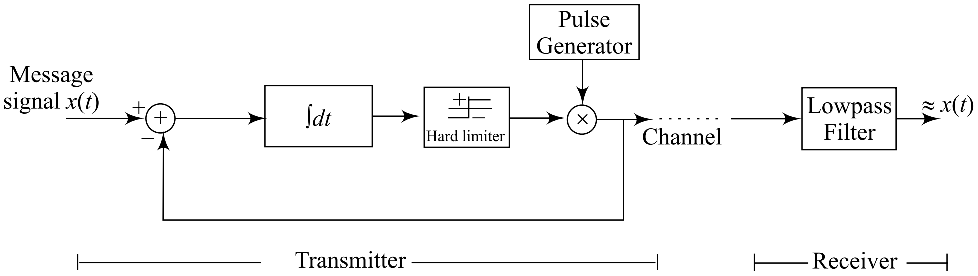


Fig. 9.19 *Simplified form of the Delta–Sigma modulation system*

Applications of Delta–Sigma modulation system A Sigma–Delta modulator is also known as a Sigma–Delta A/D converter, or an over-sampling A/D converter in DSP literature. The design of the analog low pass anti-aliasing filter that precedes the sampler in any A/D converter becomes difficult and its implementation becomes expensive because of the extremely sharp cut-off characteristics needed if the sampling frequency is close to the Nyquist frequency. This is because it needs to be of high order, needs precision analog components for implementation and VLSI technology cannot be used. So, to simplify the design and implementation of this anti-aliasing filter, the signal is over-sampled using an f_s that is many times more than the Nyquist rate and then converted using a Sigma–Delta A/D converter. The output of this is later decimated to bring down the sampling rate nearer to the Nyquist rate.

Because high over-sampling is employed in Sigma–Delta A/D converters, they are most useful in low frequency applications such as digital telephony, digital audio (compact-disc encoders) and in digital spectrum analyzers.

9.4.3 Adaptive Delta Modulation (ADM)

While discussing ‘slope overload’, we had stated that while a large value of step size is needed to avoid slope overload when the signal is changing steeply, the large step size would produce too much of granular noise when the signal is either constant, or is changing rather slowly. Thus, the overall noise can be reduced and the destination *SNR* can be improved considerably, if we make the step size to vary depending upon the way the signal itself changes with time – having a large value when the signal is changing steeply with time and a small value when the signal is relatively constant or changing very slowly. This is what is done in an adaptive delta modulator.

In an adaptive delta modulator, the step size may be made to vary with steepness of variation of the message signal either continuously, or in a discrete manner. In both the types, the sensing of the steepness of the message signal is done in the same way. As shown in Fig. 9.13, except during the stat-up time, whenever the signal is changing steeply, the binary output from the modulator continues to be the same – a series of 1s if signal is steeply rising and a series of 0s if the signal is steeply decreasing. On the other hand, when the signal is changing very slowly, or is constant the binary output from the Delta modulator is alternate 1s and 0s. This fact is utilized to change the step size. Figure 9.20 shows an adaptive Delta modulator in which the step size changes in a continuous manner (CVSDM).

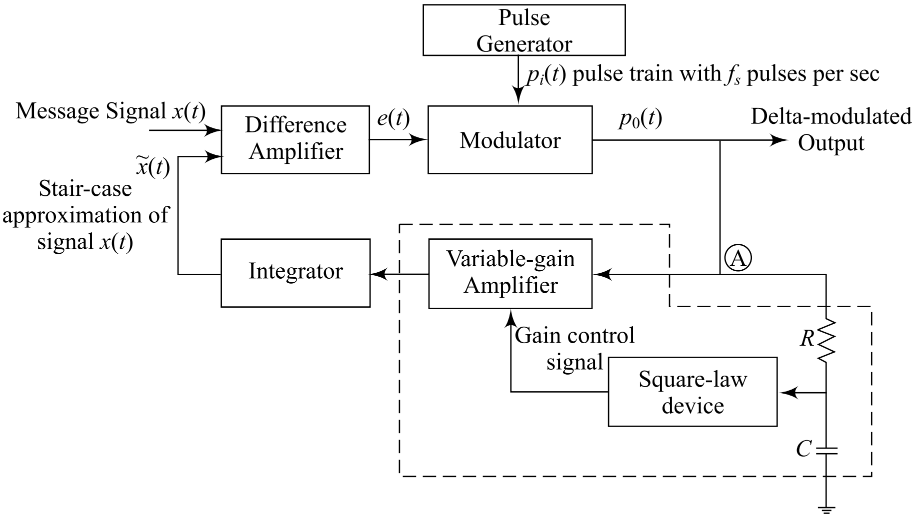


Fig. 9.20 *An adaptive Delta modulator with continuously variable step size*

The pulse generator produces narrow pulses of fixed amplitude at a rate equal to the desired sampling rate. The modulator consists of a hard limiter followed by a product device, or a multiplier. Whatever may be the actual value of $e(t)$, the hard limiter output will be +1 if $e(t)$ is positive and –1 if $e(t)$ is negative. So the polarity of the pulse $p_0(t)$ depends on the sign of $e(t)$. The subsystems within the dotted-line box are for ‘adaptation’. For a moment, assume this part is not there and point marked (A) is directly connected to the input of the integrator. Let us approximate the narrow pulses in the pulse train $p_0(t)$ by impulses. Since integration of an impulse gives a step, integration of the train of impulses occurring at regular intervals of $T_s = 1/f_s$ will result in a staircase signal approximation of the message signal $x(t)$. The step size Δ in this staircase approximation depends only on the amplitude of the pulses in $p_0(t)$ and the gain of the integrator. So we get a staircase approximation with a fixed step size.

Now, assume that the 'adaptation' circuit shown in the dotted-line box is connected. The pulses in the pulse train $p_0(t)$ try to charge the capacitor C through the resistance R. However, if for a short segment of time, the pulses are alternatively positive and negative – and this happens when the message signal is either not changing at all, or is changing very slowly, there will not be any charge accumulation on the capacitor and the voltage across it will be zero, or negligible. So the gain control voltage is almost zero or is zero and there will not be any change in the amplitude of the pulses at the output of the variable-gain amplifier. As the gain of this amplifier is adjusted initially to be low when the gain-control voltage level is zero, the amplitude of the pulses fed as input to the integrator and so, the step size of $\tilde{x}(t)$, the staircase approximation, will be small. We have thus ensured that the step size is small when $x(t)$ is almost constant, or is changing very slowly. Now, if the $x(t)$ is steeply rising or falling for some time, the consecutive pulses in the pulse train $p_0(t)$ will be either all positive or all negative over that segment of time. So the capacitor will be charged. Irrespective of whether it is charged positively or negatively, the square law device output which is the gain-control voltage, will be positive and its value will depend upon the length of time for which the polarity of the gain of the amplifier and consequently the step size, will go on increasing till the rate of change of $x(t)$ becomes less. Once rate of change of $x(t)$ becomes less, the gain of the amplifier and the step size will reduce automatically to suit the new conditions.

In the absence of any adaptation, the receiver just consists of a decoder (a decision device followed by a pulse generator) and an integrator followed by a low pass filter with cut-off frequency, W Hz, the band limiting frequency of $x(t)$. In the absence of any decoding errors due to channel noise, the output pulse train from the pulse generator part of the decoder will be an exact replica of the transmitted pulse train $p_0(t)$. These impulse-like narrow pulses, when fed to the integrator, produce a staircase approximation of $x(t)$ and the LPF, the last stage, removes the out-of-band frequency components from this staircase approximation to give an estimate of $x(t)$. When we consider an ADM system, the receiver will have a structure as shown in Fig. 9.21.

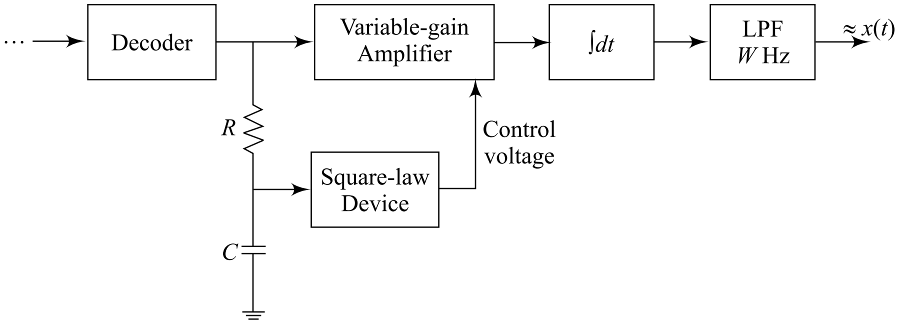


Fig. 9.21 *Receiver for an ADM (CVSDM) system with continuously variable step size*

Just as in the transmitter, here too, the gain of the variable gain amplifier is controlled by the voltage developed across the capacitor. This will be large when the polarity of output pulses from the decoder continues to be the same – corresponding to steeply rising portion of $x(t)$, and will be almost zero when decoder output pulses are alternatively positive and negative – corresponding to a flat segment of $x(t)$. When the gain changes the step size also changes correspondingly, thus giving a variable step size. The LPF output will be devoid of the out-of-band frequency components of the variable step size staircase approximation produced by the integrator.

9.4.4 ADM System with Discrete Set of Values for Step Size

ADM system in which the step size can take a discrete set of values make use of a logic circuit in the transmitter as well as the receiver to control the step size. The logic circuit in the transmitter senses the steepness of variations in $x(t)$ by observing the binary output-sequence from the modulator and accordingly either increases, or reduces the step size values in discrete steps. The step size, Δ, is permitted to vary in discrete steps within a certain range of values: $\Delta_{\min}$ to $\Delta_{\max}$. Figure 9.22 shows the block diagram of such an adaptive delta modulation system.

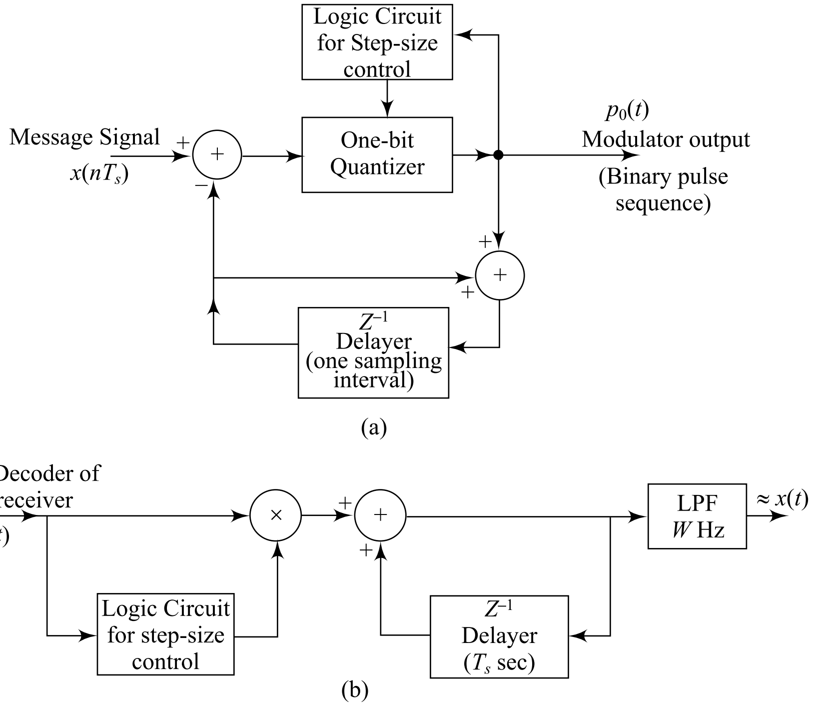


Fig. 9.22 *An ADM system with a discrete set of values for the step size: (a) Transmitter, (b) Receiver*

The step size will initially be $\Delta_{\min}$. If $\Delta(nT_s)$ is the step size at the n^{th} sampling instant, it is so arranged that

$$\Delta(nT_s) = \begin{cases} K\Delta(nT_s - T_s) \text{ if } b(nT_s) = b(nT_s - T_s) \\ (1/K)\ \Delta(nT_s - T_s) \text{ if } b(nT_s) \neq b(nT_s - T_s) \end{cases} \tag{9.40}$$

where $b(nT_s)$ is the binary pulse at $t = nT_s$. Hence, if two consecutive binary pulses in the output binary pulse sequence $p_0(t)$ are alike, which indicates that $x(t)$ is steeply changing, the step size is increased by a factor K compared to its previous value. K is generally taken as 1.5 for speech and image signals. On the other hand, if two consecutive binary pulses of $p_0(t)$ are *not* alike, which indicates that $x(t)$ is varying slowly, the step size is decreased by the factor K. It has been reported by Jayant and Knoll (See Reference) that for a wide range of bit-rate values like 20 kbps to 60 kbps, a value of $K = 1.5$ is quite satisfactory and that this type of ADM system with $K = 1.5$ gives about 10 dB better $(SNR)_D$ as compared to an LDM system for which the step size is fixed. Figure 9.23 shows the waveforms of the approximation signal $\hat{x}(t)$, obtained in the case of ADM and LDM.

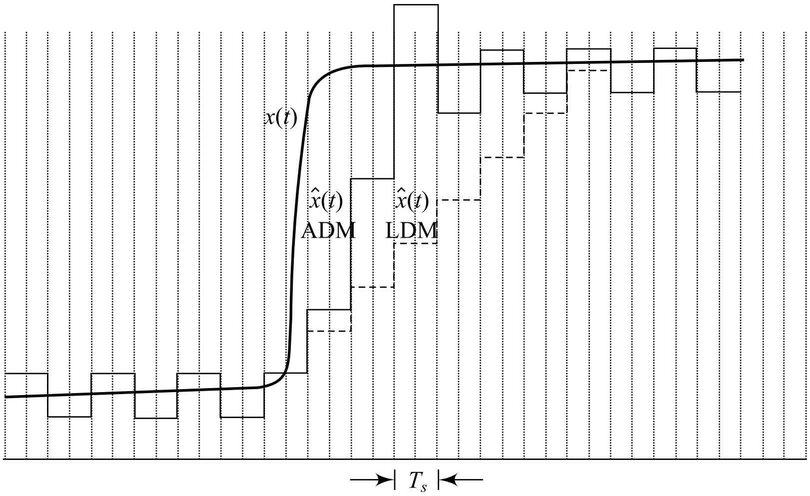


Fig. 9.23 *Waveforms of $\hat{x}(t)$ for LDM and ADM*

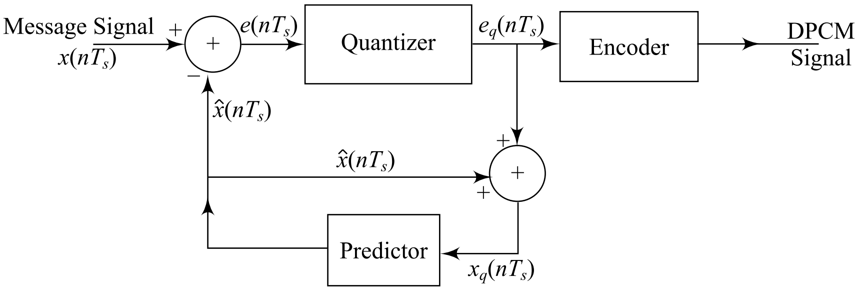
9.4.5 Comparison of PCM and DM

In PCM, we sample the message signal slightly above the Nyquist rate, quantize each sample and represent each quantized sample value using *n* bits. Since one sample is produced at each sampling instant, and *n* bits are used for representing each sample, the bit-rate is nf_s bits/sec. for an *n*-bit binary PCM with a sampling frequency of f_s. In delta modulation, on the other hand, we use a much higher sampling frequency in order to reduce the difference between adjacent samples so that the difference could be represented with a reasonably good accuracy by just one bit. If the sampling frequency for delta modulation is denoted by $f_{s\Delta}$, the bit rate of the binary stream produced by delta modulation is given by $1 \times f_{s\Delta} = f_{s\Delta}$ bits/sec. Although the bit rate is generally higher for DM, it would depend on the nature of the message signal as well as the quality specifications for any specific application.

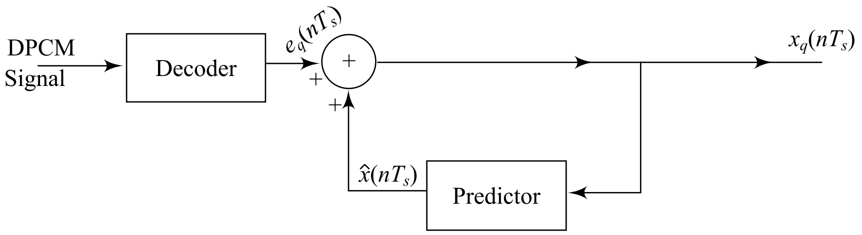
For example, for a speech signal with 3.5 kHz as the band-limiting frequency, as will be the case in telephone signal transmission, with the standard 8 kHz sampling rate and 8-bit binary PCM, the transmission bit-rate is 64 kbps. To achieve the same quality of received signal with a DM system, we need to use a sampling frequency of approximately 100 kHz. This implies that DM needs a bit rate of 100 kbps. But if some compromise on the quality front is permitted, Delta modulation will score over PCM. Further, DM system is quite simple and inexpensive while a PCM system is generally more complex and expensive.

9.4.6 Differential Pulse Code Modulator (DPCM)

When an audio or a video message signal is sampled slightly above the Nyquist rate, as is generally done in PCM, adjacent samples have a good degree of correlation. This implies that by directly encoding these sample values (of course, after quantization), we are permitting a good degree of redundancy in the PCM signal, i.e., we are using a higher bit rate than what is actually needed. In fact, correlation between adjacent samples suggests that it should be possible to predict, with a fair degree of accuracy, the present value of the message signal from a knowledge of its immediate past behavior. The DPCM system, in fact, employs a predictor, which predicts the present sample value making use of a few immediate past sample values by taking the linear combination of those past samples. How exactly the coefficients or the weights of this linear combination are chosen, we will be discussing in detail later. What is done in the DPCM system is that this predicted value of the present sample is compared to the actual value of the present sample and the difference between the two is pulse-code modulated. The advantage in this lies in the fact that if the prediction is reasonably good, the difference between the actual value and the predicted value, called the error, will have a much smaller dynamic range than the original message itself and therefore needs far fewer bits per each error sample than what would have been needed for the original samples themselves.



(a) Transmitter



(b) Receiver

Fig. 9.24 *A differential pulse code modulation system*

Figure 9.24 shows a DPCM system. In the transmitter, the current sample, $x(nT_s)$ is compared to the predicted value $\hat{x}(nT_s)$ and the difference is quantized using an appropriate number of quantization levels, and then encoded and transmitted in the form of a stream of binary pulses. If the predictor utilizes 'p' previous samples to predict the present sample, it is said to be a p^{th}-order predictor. As the order of the predictor is increased, initially, the prediction error, given by

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s)$$

decreases but after a certain value of predictor order, it just changes very little and there is no point in increasing the order any further since it only increases the complexity of the predictor, without in any way increasing the prediction accuracy.

The quantized error sample $e_q(nT_s)$ goes into the feedback loop, gets added to $\hat{x}(nT_s)$ to give $x_q(nT_s)$ which is then used by the predictor along with $x_q[(n-1)T_s]$ and a few more, i.e., $(p-1)$ past estimates of the quantized sample values, to produce at its output, the predicted value of the next sample, i.e., $\hat{x}[(n+1)T_s]$, which is compared with the next message sample $x[(n+1)T_s]$ to produce $e[(n+1)T_s]$. This is then quantized to give $e_q[(n+1)T_s]$ and this again goes to the encoder as well as the adder in the transmitter feedback loop. The process continues like that.

In the receiver, the decoder decodes the received bit stream. Neglecting the errors that might be caused in the decoding process owing to channel noise, the decoder output will be $e_q(nT_s)$ since the decoding operation is the inverse of the encoding operation. Just like in the transmitter feedback loop, here too, $e_q(nT_s)$ adds to $\hat{x}(nT_s)$, the predictor output, to give $x_q(nT_s)$ sequence which is the output sequence of the receiver. In case an analog output is desired, this $x_q(nT_s)$ sequence can be low pass filtered by an LPF with a cut-off frequency of

W Hz (the band-limiting frequency of the original message), to get $\tilde{x}(t)$ (which is an approximation to $x(t)$), together with some inband quantization noise.

9.4.7 Linear Prediction and Predictors

When adjacent samples of a message have good correlation, as in the case of audio and video message samples encoded using PCM, it is possible to predict the value of a future sample by making use of the present and some previous samples. Suppose, we want to predict $x(nT_s)$, the n^{th} sample. For this purpose, we may make use of 'p' previous samples, $x(\overline{n-1}\,T_s), x(\overline{n-2}\,T_s), \ldots, x(\overline{n-p}\,T_s)$. One way of using them is, to take their linear combination $\hat{x}(nT_s)$ to give an estimate of $x(nT_s)$. Let us say

$$\hat{x}(nT_s) = h_1 x(\overline{n-1}\,T_s) + h_2 x(\overline{n-2}\,\,T_s) + \ldots + h_p x(\overline{n-p}\,\,T_s) \tag{9.41}$$

where $h_1, h_2, \ldots, h_p$ are some real numbers, called the weights used in the linear combination and $\hat{x}(nT_s)$ is called the predicted value. Since a linear combination of the previous sample values is used for obtaining the predicted value, the prediction process is called '*Linear Prediction*'; and since p previous sample values have been used for prediction, the predictor is said to be of p^{th} order. How these weights, or coefficients of the linear combination, are to be chosen, or determined, we shall discuss a little later. Since only a simple linear combination of the 'p' previous samples is needed, we can implement the predictor as a simple FIR digital filter, generally called a *transversal filter* as shown in Fig. 9.25.

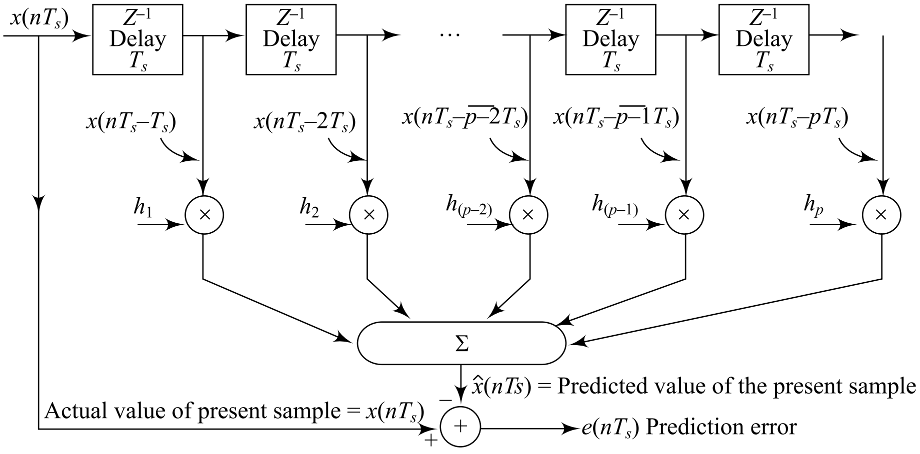


Fig. 9.25 *A p^{th} order prediction filter and prediction error*

In the receiver, the predictor is used for the reverse operation, i.e., obtaining $x(nT_s)$ from the $e(nT_s)$ which is given as the output of the decoder. Here, the predictor feedback loop is configured as shown in Fig. 9.26.

The predictor weights or coefficients must be so chosen that the 'prediction error' is minimized in some sense. Usually the error is minimized in the 'mean-square' sense, i.e., the mean-squared value of the error is minimized by an appropriate choice of the 'p' weights, or coefficients of the linear combination. For the purpose of obtaining an optimum set of weights in this sense, let us assume that our message signal $x(t)$ is a zero-mean stationary random process. (Strictly speaking, speech is not a stationary process.) Now,

$$\text{Error} = e(nT_s) = x(nT_s) - \hat{x}(nT_s) \tag{9.42}$$

We have to minimize

$$E[e^2(nT_s)] = E\left[\{x(nT_s) - \hat{x}(nT_s)\}^2\right],$$

where the symbol $E[\cdot]$ is used for representing the '*ensemble average*'.

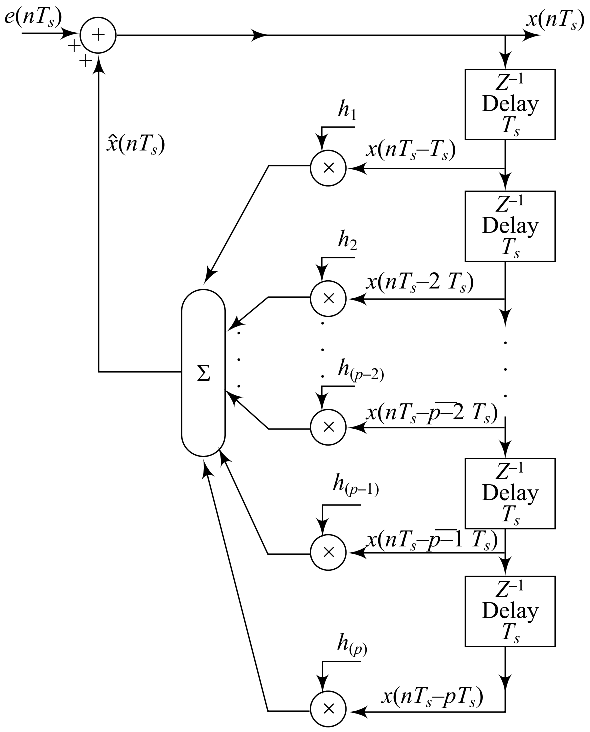


Fig. 9.26 *Prediction feedback loop of the receiver*

But $\hat{x}(nT_s)$ = Output of the FIR filter of order 'p'

$$= \sum_{k=1}^{p} h(kT_s)x(nT_s - kT_s) \tag{9.43}$$

For the sake of convenience, let us drop T_s from the above. Then, we have to minimize J, where

$$J \underline{\Delta} E[e^2(n)] = E\left[x(n) - \sum_{k=1}^{p} h(k)x(n-k)\right]^2 \tag{9.44}$$

With respect to the weights $h(k)$s, $k = 1$ to p.

$$\therefore \qquad J = E\left[x^2(n) - 2\sum_{k=1}^{p} h(k)x(n)x(n-k) + \sum_{l=1}^{p}\sum_{k=1}^{p} h(l)h(k)x(n-l)x(n-k)\right] \tag{9.45}$$

Taking expectation on the RHS term by term and noting that

$$E[x^2(n)] = R_X(0); \qquad -2\sum_{k=1}^{p} h(k)E\{x(n)x(n-k)\} = -2\sum_{k=1}^{p} h(k)R_X(k);$$

and

$$\sum_{l=1}^{p}\sum_{k=1}^{p} h(l)h(k)E\{x(n-l)x(n-k)\} = \sum_{l=1}^{p}\sum_{k=1}^{p} h(l)h(k)R_X(k-l)$$

we get

$$J = R_X(0) - 2\sum_{k=1}^{p} h(k)R_X(k) + \sum_{l=1}^{p}\sum_{k=1}^{p} h(l)h(k)R_X(k-l) \tag{9.46}$$

To minimize J with respect to the coefficients or weights, let us take the partial derivatives of J with respect to each weight and equate it to zero.

$$\therefore \qquad \frac{\partial J}{\partial h(k)} = -2R_X(k) + 2\sum_{l=1}^{p} h(l)R_X(k-l) = 0 \tag{9.47}$$

From Eq. (9.47), we therefore have the following p equations:

$$R_X(k) = R_X(-k) = 2\sum_{l=1}^{p} h(l)R_X(k-l); \quad k = 1, 2, 3, \dots, p \tag{9.48}$$

Writing this set of p equations in expanded form:

$$\begin{aligned}
R_X(1) &= R_X(-1) = h(1)R_X(0) + h(2)R_X(-1) + h(3)R_X(-2) + \dots + h(p)R_X(1-p)\\
R_X(2) &= R_X(-2) = h(1)R_X(1) + h(2)R_X(0) + h(3)R_X(-1) + \dots + h(p)R_X(2-p)\\
R_X(3) &= R_X(-3) = h(1)R_X(2) + h(2)R_X(1) + h(3)R_X(0) + \dots + h(p)R_X(3-p)\\
&\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots\\
R_X(p) &= R_X(-p) = h(1)R_X(p-1) + h(2)R_X(p-2) + h(3)R_X(p-3) + \dots + h(p)R_X(0)
\end{aligned}$$

This set of equations may be written in matrix form as

$$\begin{bmatrix} R_X(1)\\ R_X(2)\\ R_X(3)\\ .\\ .\\ .\\ R_X(p) \end{bmatrix} = \begin{bmatrix} R_X(0) & R_X(-1) & R_X(2) & . & . & . & R_X(p-1)\\ R_X(1) & R_X(0) & R_X(-1) & . & . & . & R_X(p-2)\\ R_X(2) & R_X(1) & R_X(0) & . & . & . & R_X(p-3)\\ . & . & . & & & & .\\ . & . & . & & & & .\\ . & . & . & & & & .\\ R_X(p-1) & R_X(p-2) & R_X(p-3) & . & . & . & R_X(0) \end{bmatrix} \begin{bmatrix} h(1)\\ h(2)\\ h(3)\\ .\\ .\\ .\\ h(p) \end{bmatrix} \tag{9.49}$$

Equation (9.49) may be written in a compact form as

$$[R_X]\,\boldsymbol{h} = r_X \tag{9.50}$$

where $[R_X]$ is the $p \times p$ auto-correlation matrix figuring in Eq. (9.49), $\boldsymbol{h}$ is the column vector of weights defined by

$$\boldsymbol{h} = [\, h_1\ h_2\ h_3 \dots h_p]^T,$$

and $\boldsymbol{r}_{\mathbf{X}}$ is the column vector of auto-correlation for different lags and defined by

$$\boldsymbol{r}_{\boldsymbol{X}} = [\, R_X(1)\ R_X(2)\ R_X(3) \dots R_X(p)]^T$$

The set of equations defined by Eqs. (9.48), (9.49), and (9.50) are called '*Wiener–Hopf Equations*' for linear prediction.

It may be noted that the $p \times p$ auto-correlation matrix $[R_X]$ is a *Toeplitz matrix* since all the elements of not only the main diagonal, but even the other diagonals parallel to the main one, are also equal.

From Eq. (9.50), we have the weights vector given by

$$\boldsymbol{h} = R_X^{-1}\boldsymbol{r}_{\boldsymbol{X}} \tag{9.51}$$

$[R_X]$, being the auto-correlation matrix whose main diagonal elements are equal to $R_X(0)$, is non-singular. So, the optimum weights that minimize the error in the mean-squared sense can be obtained by solving the matrix Eq. (9.51). The auto-correlation matrix $[R_X]$ is known uniquely once the auto-correlation values $R_X(0) \dots R_X(p-1)$ are known and the auto-correlation vector $\boldsymbol{r}_{\boldsymbol{X}}$ is known once the auto-correlation values $R_X(1) \dots R_X(p)$ are determined. Further, very efficient algorithms are available for the inversion of a Toeplitz matrix. Hence, the optimum weights can be determined uniquely, once the message auto-correlation values are known for lags 0, 1, 2, … , p when a p^{th}-order predictor is used.

Since $R_X(0) = \sigma_X^2 =$ average power (variance) of the zero-mean message process, the minimum mean square prediction error is given by

$$J_{\min} = \sigma_X^2 - \boldsymbol{r}_X^T [R_X]^{-1} \boldsymbol{r}_X \tag{9.52}$$

Further, since the quadratic term

$$\boldsymbol{r}_x^{\boldsymbol{T}} [R_X]^{-1} \boldsymbol{r}_X \tag{9.53}$$

is always positive, Eq. (9.52) tells us that the minimum mean-square error is always less than the variance of the message signal samples $x(nT_s)$.

Remarks

From the above discussion on DPCM and a comparison of Fig. 9.24 which shows a DPCM system, with Fig. 9.13 which shows a DM system, it is clear that basically there is considerable similarity between the two systems. However, there are two conspicuous differences between them. These are:

(i) Whereas a DPCM employs a predictor to produce an approximation to the present message sample $x(nT_s)$, a DM system, on the other hand, employs just a unit delayer whose output is the quantized version of the previous sample, and uses it as an approximation for $x(nT_s)$.

(ii) Whereas the DPCM employs an n-bit quantizer for pulse code modulating the error samples, the DM employs a one-bit quantizer to encode the error samples.

From point (ii)., we may view that a DM system as a 'one-bit version' of a DPCM system. Further, just like the DM, the DPCM system also has the quantization noise, as well as the slope-over load problems.

9.4.8 Processing Gain of DPCM and Comparison with PCM

In PCM, the samples are directly quantized, encoded and transmitted. In DPCM, however, we quantize and encode *not* the original message samples, but only the *error samples* obtained by subtracting the *predicted* value of each message sample from the *actual* value of that message sample. Does this give us any benefit in terms of an improvement of the output $(SNR)_Q$? If it does, to what extent is the $(SNR)_Q$ improved? What determines the extent of improvement, if there is any? We shall now examine these aspects.

For a DPCM system, we define the output signal-to-quantization noise ratio in the usual way, as

$$\left(\frac{S}{N}\right)_{D,Q} = \frac{\text{variance of message} = \sigma_X^2}{\text{variance of quantization error} = \sigma_Q^2} \tag{9.54}$$

As stated earlier, we are assuming the message to be a zero-mean stationary process.

Now, let σ_p^2 denote the variance of the prediction error in the DPCM system. Then, we shall rewrite Eq. (9.54) as

$$\left(\frac{S}{N}\right)_{D,Q} = \frac{\sigma_X^2}{\sigma_Q^2} = \left(\frac{\sigma_X^2}{\sigma_p^2}\right)\left(\frac{\sigma_p^2}{\sigma_Q^2}\right) \tag{9.55}$$

Since σ_p^2 is the variance of the prediction error and since the prediction error is the one which is PCM-ed in a DPCM system, the ratio (σ_p^2/σ_Q^2) may be taken as the usual signal-to-quantization noise ratio of PCM. If we now define a factor, called the processing gain, denoted by G_p, as

$$G_p \underline{\Delta} \left(\frac{\sigma_X^2}{\sigma_p^2}\right) \tag{9.56}$$

then this processing gain is obtained because of the 'differential quantization', i.e., because we are quantizing the prediction error.

Now, the quantity G_p, the processing gain, may be greater than one, or less than one, depending upon how good the prediction is, which in turn depends upon our selection of the weights of the predictor. If we minimize the prediction error (in the mean-square sense) as detailed earlier by an optimum choice of the weights, then the variance of prediction error, σ_p^2, is minimized. Since the message variance σ_X^2 is fixed for a given message, it follows from Eq. (9.56) that the processing gain G_p will be maximized. Thus, for a given order of the predictor, optimum selection of weights maximizes G_p, the processing gain, by minimizing the prediction error *for that order of the predictor*. So, further improvement in G_p can be obtained by increasing the order of the prediction filter, and using optimum weights for each value of the predictor order. However, as stated earlier, as the predictor order is increased from an initial value of zero (which implies no prediction) to first order, second order, and so on, it is only an increase from zero order to first order that produces about 4 to 5 dB of processing gain and further increases of order results in only small increments in G_p. With a prediction filter order of 5, it is found that DPCM gives about 11 dB improvement in the $(SNR)_{D,Q}$ as compared to PCM.

Till now, we have seen DPCM and PCM from the point of view of their output signal-to-quantization noise ratios. We may look at them for a specified $(SNR)_{D,Q}$. For a sampling rate of 8 kHz, DPCM may give a saving in bit rate to the extent of 1 to 2 bits/sample, i.e., about 8 to 16 kbps, as compared to PCM.

9.4.9 Adaptive Differential Pulse Code Modulation (ADPCM)

As we have already seen, determination of the optimum coefficient set for the predictor requires a knowledge of the auto-correlation coefficients of the message signal for lags of 0, T_s, $2T_s$, … , pT_s. The problem is these auto-correlation coefficients are not usually known. Further, the message signal may not be a stationary process, in which case, its auto-correlation values and all other statistical parameters will be varying with time. Both the above situations underscore the need for adaptively adjusting the predictor coefficients.

In addition, there is also one more problem. As was stated earlier, just like DM, the DPCM also has the slope-overload problem and the quantization noise, or granular noise problem. As in the case of DM, here too, in the case of speech communication, changes in the levels of speech signal are quite common and these again call for adaptively changing the step size in order to have good signal-to-quantization noise ratio at the destination besides avoiding slope-overload noise.

Thus, in DPCM, prediction as well as quantization need to be adaptively controlled, depending respectively upon the spectrum and the changing levels of the message signal. Insofar as the step size is concerned, it may be controlled adaptively just in the same manner as was done in the case of DM. For obtaining adaptive prediction, an iterative algorithm, called the 'Least Mean Square', or LMS algorithm, may be used in order to adjust the predictor weights to their optimum values so as to obtain error in the mean-square sense. This arrangement is illustrated in Fig. 9.27.

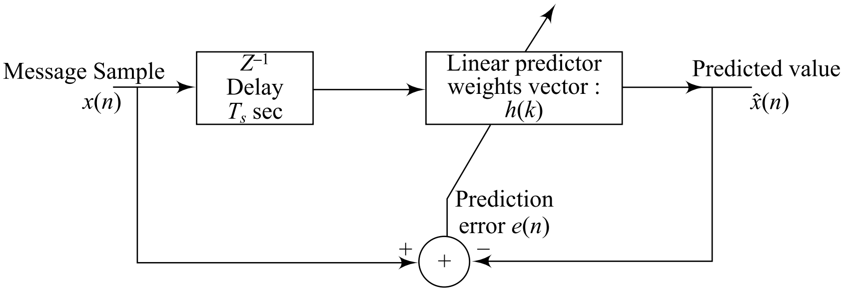


Fig. 9.27 *Illustrating adaptive linear prediction*

Example 9.13 A first-order predictor gives the predicted sample value as a constant times the previous sample values, i.e.,

$$\hat{x}(nT_s) = wx(nT_s - T_s)\ )$$

where $x(nT_s)$ is a sample of a zero-mean stationary process, T_s is the sampling period and ω is the weight. Determine the following:

(a) σ_E^2, the variance of the prediction error.

(b) Optimum value of the weight w for obtaining minimum mean-square error.

(c) The minimum value of the prediction-error variance.

Solution

(a) Prediction error = $e(nT_s) = [x(nT_s) - \hat{x}(nT_s)]$

$\therefore$ mean-squared value of the prediction error $= E\left[e^2(nT_s)\right]$

$$= E\left[\{x(nT_s) - \hat{x}(nT_s)\}^2\right]$$

$$= E\left[x^2(nT_s) - 2x(nT_s)\hat{x}(nT_s) + \hat{x}^2(nT_s)\right]$$

Substituting $wx(nT_s - T_s)$ for $\hat{x}(nT_s)$ and taking the expectation of the quantity in the rectangular brackets term by term, we get

$$J = E\left[e^2(nT_s)\right] = E[x^2(nT_s)\Big] - 2wE\left[x(nT_s)x(nT_s - T_s)\right] + w^2E\left[x^2(nT_s - T_s)\right]$$

Now, since $x(t)$ is given to be a zero-mean stationary process, we may write

$E\left[e^2(n)\right] = \sigma_E^2$ variance of the predictor error

$E\left[x^2(nT_s)\right] = \sigma_X^2$ = variance of the message signal

= Average power of the message signal

$E\left[x^2(nT_s - T_s)\right] = \sigma_X^2$ since $x(t)$ is a stationary process

$$E\left[x(nT_s)x(nT_s - T_s)\right] = R_X(T_s)$$

$$\therefore \qquad J = \sigma_E^2 = \sigma_X^2 - 2wR_X(T_s) + w^2\sigma_X^2$$

$$\therefore \qquad \sigma_E^2 = \sigma_X^2\left[1 + w^2 - 2w\frac{R_X(T_s)}{\sigma_X^2}\right]$$

(b) Now, $$\frac{\partial J}{\partial \omega} = 0 - 2R_X(T_s) + 2w\sigma_X^2$$

$\therefore$ equating the above to zero for minimizing J, we have

$$w_{\text{opt}} = \left[\frac{R_X(T_s)}{\sigma_X^2}\right]$$

(c) The minimum value of the variance of the predictor error is obtained by substituting the value of w_{opt} for w in the expression for σ_E^2

Thus, $$\sigma_{E\,\min}^2 = \sigma_X^2 + \sigma_X^2\frac{R_X^2(T_s)}{\sigma_X^4} - 2\frac{R_X(T_s)}{\sigma_X^2}\frac{R_X(T_s)}{\sigma_X^2}\cdot\sigma_X^2$$

$$\therefore \qquad \min \sigma_E^2 = \sigma_X^2 - \left[\frac{R_X^2(T_s)}{\sigma_X^2}\right]$$

Example 9.14 A zero-mean stationary process with a variance of $\overline{x^2}$ is given as the message signal to a DPCM system using a second-order predictor. The auto-correlations of the message signal for lags of T_s and $2T_s$ are 0.8 $\overline{x^2}$ and 0.5 $\overline{x^2}$ respectively. Determine

(a) the optimum predictor weights which will give the minimum mean square prediction error.
(b) the minimum variance of the prediction error.
(c) the processing gain of the system.

Solution

We can conveniently put $\overline{x^2} = 1$ and $T_s = 1$ sec.

From the given data: $R_X(1) = 0.8\overline{x^2} = 0.8$

$R_X(2) = 0.5\overline{x^2} = 0.5$

Further, since $R_X(0) = \overline{x^2}$, $R_X(0) = 1$

$$\therefore \qquad [R_X] = \begin{bmatrix} \overline{x^2} & 0.8\overline{x^2} \\ 0.8\overline{x^2} & \overline{x^2} \end{bmatrix} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

(a) $\therefore$ from Eq. (9.49), we have $\begin{bmatrix} R_X(1) \\ R_X(2) \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \begin{bmatrix} h(1) \\ h(2) \end{bmatrix}$

$$\therefore \quad \begin{bmatrix} h(1) \\ h(2) \end{bmatrix} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 2.77^{\cdot} & -2.22^{\cdot} \\ -2.22^{\cdot} & 2.27^{\cdot} \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1.111^{\cdot} \\ -3.388^{\cdot} \end{bmatrix}$$

$\therefore$ $\boxed{h(1) = 1.111^{\cdot} \text{ and } h(2) = -0.388^{\cdot}}$. These are the optimum filter coefficients.

(b) Processing gain = $G \underline{\Delta} \dfrac{\sigma_X^2}{\sigma_p^2}$ (Refer to Eq. (9.56))

where σ_X^2 is the variance of the message signal = $R_X(0)$ since it is zero mean, and σ_p^2 is the variance of the prediction error when optimum filter weights are used.

But we know that $\sigma_p^2 = J_{\min} = \sigma_X^2 - \boldsymbol{r}_X^T [R_X]^{-1} \boldsymbol{r}_X$ (Refer to Eq. (9.52))

$$\therefore \qquad \sigma_p^2 = 1 - \begin{bmatrix} 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} 2.77^{\cdot} & -2.22^{\cdot} \\ -2.22^{\cdot} & 2.77^{\cdot} \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.5 \end{bmatrix}$$

$$1 - \begin{bmatrix} 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} 1.111 \\ -0.388 \end{bmatrix} = 1 - 0.694 = 0.306 \quad \therefore \boxed{\sigma_p^2 = 0.306}$$

(c) $\boxed{\therefore \text{Processing gain} = \dfrac{1}{0.306} = 3.26}$

9.4.10 Speech Compression/Low Bit Rate Coding of Speech

Normally speech signal, band-limited to 3.2 kHz is used for obtaining 'telephone quality' of speech. It is PCM encoded by sampling at 8000 samples per second and using an 8-bit representation for each sample. The bit rate therefore works out to 64 kbps. This bit rate does not, of course, pose any problem in normal telephone channel transmission. However, in the case of certain wireless channels used for secure communication, which can support only narrow bandwidths, or, low bit rates, speech transmission at 64 kbps does pose severe problems. This makes it necessary to resort to low bit-rate coding of speech. Systems used for voice coding at low bit rates are called voice coders, or VOCODERS.

Ideally, we would like to achieve the low bit-rate coding, or compression of speech in such a manner that no information is lost in the process, by removing only just the redundancy. This is called 'lossless compression'. However, almost any method of speech coding involves some amount of loss of information. Hence, what is sought is a low bit-rate coding which maintains certain acceptable standards with regard to the quality of reproduced speech. There are three distinct approaches to low bit-rate coding of speech. These are:

1. Use of time-domain techniques
2. Use of frequency/transform domain techniques
3. Use of certain model-based techniques

Now, we shall discuss each of these approaches in some detail.

1. Time-domain techniques:

(a) *DPCM* As we have already seen, DPCM achieves some compression compared to PCM, as it pulse-code modulates not the original samples of the message, but only the difference between the original samples and their predicted values. The dynamic range of this difference being much smaller than that of the original samples, less number of bits are needed to represent each difference sample and therefore the bit rate required is less. For the same quality of reproduced speech, as given by PCM, an ADPCM system cuts down the bit rate by a factor of 2, i.e., at 32 kbps, ADPCM gives the same speech quality as PCM at 64 kbps.

(b) *Sub-band coding* In this, the speech signal is divided into a number of non-overlapping sub-bands – say, 400 Hz to 800 Hz, 800 Hz to 1600 Hz, and 1600 Hz to 3200 Hz, by using a bank of what are called, Quadrature Mirror Filters (QMF). The output signal from each of these filters, representing the time-domain signal having frequency components in a particular band, is then encoded in the time-domain using PCM or DPCM. Signals representing the low frequency bands like 0 to 400 Hz and 400 to 800 Hz are alloted more number of bits/sample, compared to the number of bits per sample assigned to the high frequency bands. This is done because of the following reasons:

(i) Most of the energy of a speech signal segment lies in the low frequency sub-bands and therefore samples of the signals representing these sub-bands need to be represented more accurately.
(ii) Quantization noise is more easily sensed by the human ear in the low frequency sub-bands.

Actually, the coding is made highly adaptive and dynamic. The pitch frequency, estimated from the speech samples, is used to improve prediction. Further, the formant frequencies, i.e., the resonant frequencies of the vocal tract of the speaker are also estimated and use is made of the 'noise-masking phenomenon' to assign very few bits to samples near these formant frequencies.

Thus, by giving very few bits to the high frequency sub-bands and by adopting adaptive assignment of bits to the samples, as described above, we are able to save considerably in terms of the overall bit rate. Adaptively sub-band coded speech at 16 kbps sounds as good as the 64 kbps PCM encoded speech, in its quality.

2. Frequency/Transform domain techniques: The speech samples stream is divided into frames of convenient length say, N samples. These frames are then transformed into spectral coefficients using a suitable discrete transform. The spectral coefficients, so obtained, are then quantized and encoded, assigning larger number of bits for the significant spectral coefficients and very few bits to represent each of the rest, by using an appropriate adaptive algorithm. It is these encoded spectral coefficients that are transmitted. At the receiver, these are first decoded, and then the resulting spectral coefficients are transformed back into time-domain samples, using an inverse discrete transform. These time-domain samples are then passed through a reconstruction filter, which is an LPF with a cut-off frequency of W Hz, the band-limiting frequency of the original message signal, $x(t)$. A time-domain analog signal $\tilde{x}(t)$ closely approximating $x(t)$, is thus obtained at the output of the receiver.

A number of discrete transforms are available for this purpose. Of all these, the Karhunen–Loeve Transform (KLT) is the best among them, theoretically so to speak. This is because it gives spectral coefficients that are totally uncorrelated. Thus, the redundancy existing in the time domain is absent in the spectral domain, if KLT is used. Hence, from the data compression point of view, it is the ideal transform and gives optimum results. However, since the KLT uses the N-length eigenvectors of the $N \times N$ co-variance matrix of the time-domain samples, it is a data (message) dependent transform and is computationally time consuming and expensive. Hence, it is used only as a benchmark for comparing the compression efficacy of the other discrete transforms. Generally, in practice, the Discrete Cosine Transform (DCT) which is the next best, or the Discrete Fourier Transform (DFT), are used although they give only sub-optimal results. Both these transforms have fast algorithms, the FCT and the FFT. Using adaptive coding with DFT or DCT, speech can be compressed to a bit rate of 9.6 kbps which can still be of the same quality as the usual 64 kbps telephone quality speech. Adaptive Transform coding, described above, is made use of in speech encoders of MP3 standard.

3. Model-based compression – Linear Predictive coding (VOCODER): Linear Predictive Coding (LPC) is a model-based method for low bit rate encoding of speech. It uses speech analysis and synthesis techniques and can give speech at very low bit rates like 2.4 kbps to 4.8 kbps. However, when encoded using this method, the reproduced speech sounds a little synthetic.

Speech sounds are categorized as ‘voiced’ and ‘unvoiced’. Voiced speech is produced when the air in the lungs is pushed out through the trachea, the pharynx cavity and finally through the mouth and the nasal cavities with the vocal cords tensed. These vocal cords vibrate thereby modulating the air into discrete pulses. The vocal cords are tensed when voiced sounds like vowels are spoken. Thus, voiced speech is quasi-periodic and repeats at the ‘pitch frequency’. On the other hand, in the case of sounds like ‘*s*’, ‘*c*’, and ‘*f*’, etc., air gushes out from the lungs through a constriction created by the lips or the tongue, with the vocal cords playing no role at all. They produce a sort of ‘hissing sound’ and are called unvoiced sounds.

Vocal tract is that part of human body extending from the glottis to the lips, and it plays a key role in speech production. Speech is not a stationary process, since the shape and size of this vocal tract goes on changing as we speak. Since it takes approximately 20–30 ms for the vocal tract to change its shape and size, speech segments of 20–30 ms duration can be considered to be stationary. In speech analysis, there exist techniques for detecting voiced/ unvoiced speech and algorithms for determination of pitch period of voiced speech.

The vocal tract is generally modeled as a linear, time-varying auto-regressive (i.e., all-pole) discrete-time filter of some order ‘p’ (generally about 10), with a system function $H(z)$ given by

$$H(z) = \frac{G}{\left[1 - \sum_{k=1}^{p} a_k z^{-k}\right]} \tag{9.57}$$

where G represents the gain of the all-pole filter. Hence, speech may be considered as the response of this filter to an input. The sound produced at its output will, of course, depend not only on the type of excitation given to it, but also on the values of the filter coefficients, a_ks, $k = 1$ to ‘p’. Since voiced speech is produced due to periodic pressure variations at the pitch frequency caused by the vibrating vocal cords, we try to simulate voiced speech by the response of this all-pole filter when the excitation (input) to it is a train of impulses occurring at a regular interval equal to the pitch period. Similarly, the unvoiced speech can be simulated by the response of this filter to a ‘white-noise’ input, as shown in Fig. 9.28 which gives the block diagram of a speech synthesizer.

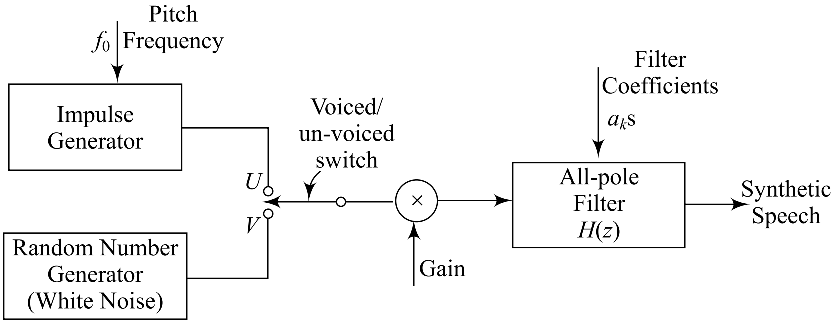


Fig. 9.28 *A speech synthesizer*

The unit in the speech analyzer which detects whether the speech is voiced or unvoiced, actuates the voiced/ unvoiced switch in the synthesizer. The pitch frequency which is estimated in the analyzer is fed to the impulse generator so as to control the period of the impulse sequence generated by it. The speech analyzer estimates the optimum filter coefficients which will minimize the difference between the actual speech and the synthesized speech in the minimum mean-square sense. These filter coefficients and gain, so estimated, are fed to the all-pole filter and the multiplier unit (respectively) of the synthesizer.

The LPC transmitter therefore need not transmit the speech or its samples. Instead, it simply transmits (i) the binary encoded version of $e(n)$, samples of the difference between the actual speech and the synthesized version of it, (ii) information about whether the speech is voiced/unvoiced, (iii) the gain parameter, G, (iv) the 'p' coefficients of the p^{th} order all-pole filter, and (v) the pitch frequency/period. Information pertaining to (ii), (iii), (iv) and (v) above, is updated every 20–30 milliseconds, as speech is not a stationary process and its statistics change whenever the vocal tract size and or shape change. Figure 9.29 shows the arrangement used at the transmitter.

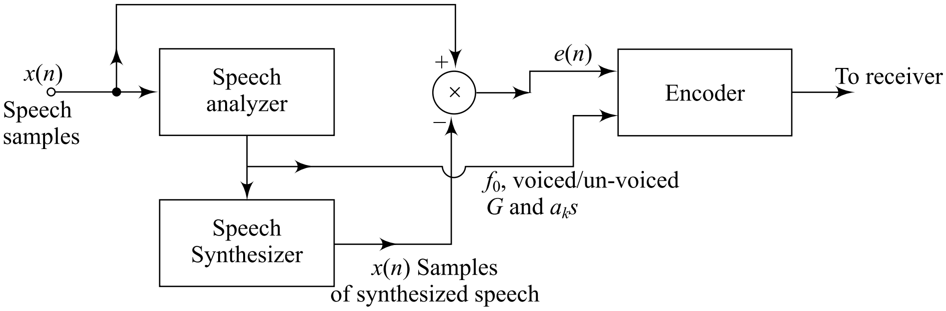


Fig. 9.29 *LPC transmitter*

The all-pole digital filter used in the synthesizer is generally of 10th order. *The filter coefficients must be transmitted to a high degree of accuracy, since even small inaccuracies can sometimes make the receiver all-pole filter unstable*. So, often, about 8 to 10 bits are used to transmit the value of each filter coefficient. About five bits are used for the gain parameter, six for pitch period and one bit for information regarding voiced/unvoiced speech. Since these are updated every 20 ms to 30 ms, the bit rate required works out to approximately 5 to 6 kbps.

The receiver reconstructs the speech signal from the information transmitted by the LPC transmitter, by using a speech synthesizer in an arrangement as shown in Fig. 9.30.

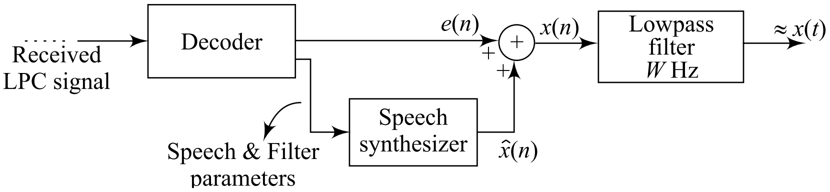


Fig. 9.30 *LPC receiver*

There are several LPC-based Vocoders more sophisticated than the one described above. These are (i) the Residual Excited LPC (RELP), (ii) the Multi-Pulse LPC (MPLP), (iii) the Code-Excited LPC (CELP), and (iv) Vector-Sum Excited LPC (VSELP). Among these, the last two, viz. the CELP and VSELP give toll grade quality speech at 4.8 kbps and are used in cellular mobile communications.

Comparison of compression techniques

Table 9.2

1.	PCM without companding and 8 bits/sample and 8 kHz sampling rate	64 kbps
2.	PCM with companding, 7–8 bits/sample and 8 kHz sampling rate	56–64 kbps
3.	Delta Modulation (DM)	32–64 kbps
4.	Differential PCM with 4–6 bits/error sample	≈32 kbps
5.	Adaptive DPCM with 3–4 bits/error sample	24–32 kbps
6.	Adaptive Delta Modulation (ADM)	16–32 kbps
7.	Sub-band Coding	≈16 kbps
8.	Adaptive Transform-Domain Techniques	≈9.6 kbps
9.	Linear Predictive Coding (LPC)	2.4–4.8 kbps

9.5 DIGITAL MULTIPLEXING

9.5.1 Time Division Multiplexing (TDM) Principle

Definition Time Division Multiplexing or TDM is the time interleaving of samples from several sources in order to transmit information from all these sources serially over a single communication channel.

The low pass sampling theorem forms the basis for TDM. It tells us that we need not transmit a band-limited continuous-time signal which engages the transmission channel continuously as long as the signal is transmitted. Instead, we can transmit only its samples taken at regular intervals at a rate above the Nyquist rate so that the receiver can reconstruct the analog signal from the received samples. As the samples are separated in time by the sampling interval, the transmission channel is not engaged all the time; it is engaged only whenever a sample occurs. Therefore, the interval between two successive samples of one message signal during which the transmission channel is free, may be utilized to transmit the samples of each of the other message signals. That is, we may interleave the samples from various message signals. A typical PAM/TDM system is shown in Fig. 9.31.

Suppose we now extend this concept of time division multiplexing to N number of n-bit binary PCM-ed message channels which are identically band limited. Then, since the sampling interval has to accommodate N codewords of n-bits each, the pulse width of the PCM/TDM signal is $(T_s/N \cdot n)$. Just as we interleaved the samples in the PAM/TDM system, here let us interleave the codewords as shown in Fig. 9.32.

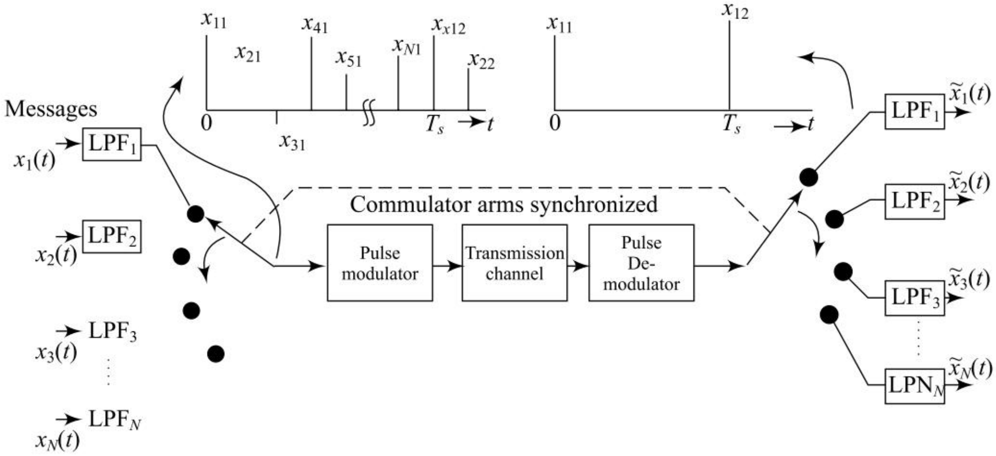


Fig. 9.31 *Time division multiplexing of analog signals*

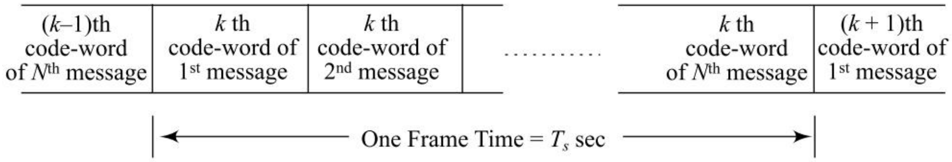


Fig. 9.32 *Illustration of the concept of a frame*

when one codeword, say the k^{th} codeword, from each one of the N-message channels is transmitted sequentially, we say one TDM 'frame' is transmitted; and the time taken for transmitting one frame is called the 'frame time'. In this case, the frame time is equal to the sampling interval of any one message channel and so is equal to T_s.

Although we have considered in the above example N message channels all of which are analog and are identically band limited, we need not have to restrict ourselves like that. Once an analog message signal is sampled, quantized and encoded, i.e., once it is digitized, it can be regarded as data which is to be multiplexed with other data coming from other sources like computers which are digital data sources. Since any digital signal is just a sequence of symbols, we may regard multiplexing of digital signals simply as interleaving of these symbols. We thus free ourselves of the rather rigid constraints like periodic sampling and waveform preservation associated with TDM. Thus, multiplexing of digital signals, or '*digital multiplexing*' as we will be referring to it hereafter, is similar to TDM except that it is more flexible because of the freedom from the constraints stated above. Digital multiplexing is therefore based on 'interleaving of symbols' from two or more digital signals.

9.5.2 Functions Performed by a Binary Digital Multiplexer (MUX)

The interleaving of the symbols, or the digital multiplexing, is done by a selector switch which sequentially takes symbols from each incoming line and delivers them to a common high-speed line to form what is known as the *multiplexed signal*. This multiplexed signal consists of binary digits from various input sources interleaved either bit by bit, or word by word. At the receiving-end, a de-multiplexer separates the low-speed components (from the different individual sources) and delivers them to their respective destinations. This concept of a 'multiplexing' and 'de-multiplexing' is illustrated in Fig. 9.33.

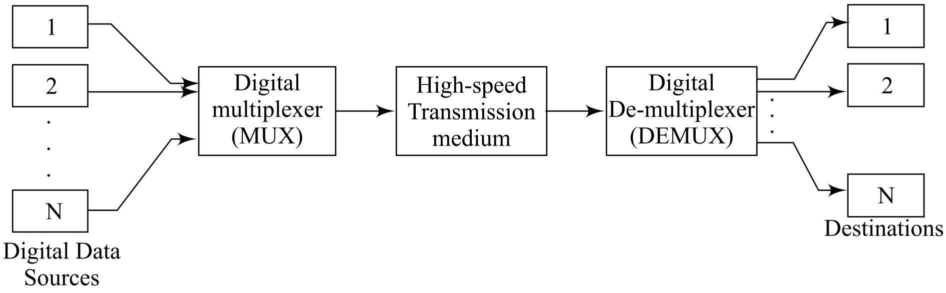


Fig. 9.33 *Illustration of the use of multiplexers and de-multiplexers*

To enable proper de-multiplexing at the receiving end, it is necessary that the multiplexed signal be constructed with a constant bit rate. This necessitates the use of signaling (for frame identification, etc.) and synchronization information also to be embedded into the multiplexed signal. The broad functional responsibilities of digital multiplexers may therefore be identified as

1. constructing a 'frame' as the smallest unit that contains at least one bit from each input.
2. assigning within each frame an appropriate number of time slots for each input.
3. inserting control bits for frame identification and synchronization.
4. providing for any small variations in the bit rates of the data from the various inputs.

Besides slight variations about a specified nominal value, which always occur when the input sources are independent and are not centrally controlled, some sources may be active only occasionally and not continuously. In that case, the time-slots allotted to them in each frame will not be having any bits when these sources are not active. Even when a source is active, it may not be giving data at a constant bit rate at its output, as happens when the data is obtained from a computer keyboard. Because of all these factors, different types of digital multiplexers are available, to cater to different types of situations. Multiplexers may therefore be broadly classified into three different categories:

1. Synchronous multiplexers: In this case, all the sources are controlled by a central master clock so as to eliminate the possibility of bit-rate variations. If, for some time any of the sources do not output any data, it is taken care of by 'stuffing' dummy pulses in the time slots allotted to those sources. Synchronous multiplexing gives maximum throughput efficiency. But it needs very elaborate arrangements for distributing the master clock signal.

2. Asynchronous multiplexers: These are used when the data sources are such that they produce bursts of characters with variable interval between the bursts, as in the case of keyboard terminals. In this type of scenario, the timing is precise only for the bits within a character or word. Start–Stop signaling is used and each character consists of a 'start bit' that starts the receiver clock and ends with one or two 'stop bits' which terminate the receiver clock. Hence synchronizing the receiver clock with a master clock is not necessary. The frequency of the receiver clock is accurate enough to ensure correct bit timing within each character or word. Because of the nature of these sources (operating in bursts of characters), only characters are interleaved and not bits. It is, of course, possible to merge these sources into a synchronous multiplexed bit stream by using buffers and character interleaving.

3. Quasi-synchronous multiplexers: When the bit rates of the input sources have the same nominal value but vary around that value within certain specified limits, quasi-synchronous multiplexers are used. As the individual clocks of the various input sources are not exactly synchronized in frequency, the data from different sources are likely to have slightly different bit rates. Hence, when a new input bit is not available at the multiplexer clocking time, 'stuff bits', which are dummy bits, are inserted to fill such time slots. These multiplexers are used for combining different sources into a high speed digital TDM signal, and when arranged in a hierarchy of increasing bit rates, provide the common carriers such as the T-1, T-2, T-3 digital lines which are the basic building blocks of interconnected digital communications systems.

9.5.3 Advantages of Digital Multiplexing

Digital multiplexing and transmission has several advantages over analog multiplexing and transmission. First, the hardware cost is low because of the use of digital ICs. Second, we can use regenerative repeaters and keep the bit-error rate low even in long-haul transmission. Third, it has much greater flexibility and can be used for multiplexing digital signals originating from a variety of sources such as voice signals, TV, digital data, videophone, etc., as shown in Fig. 9.34.

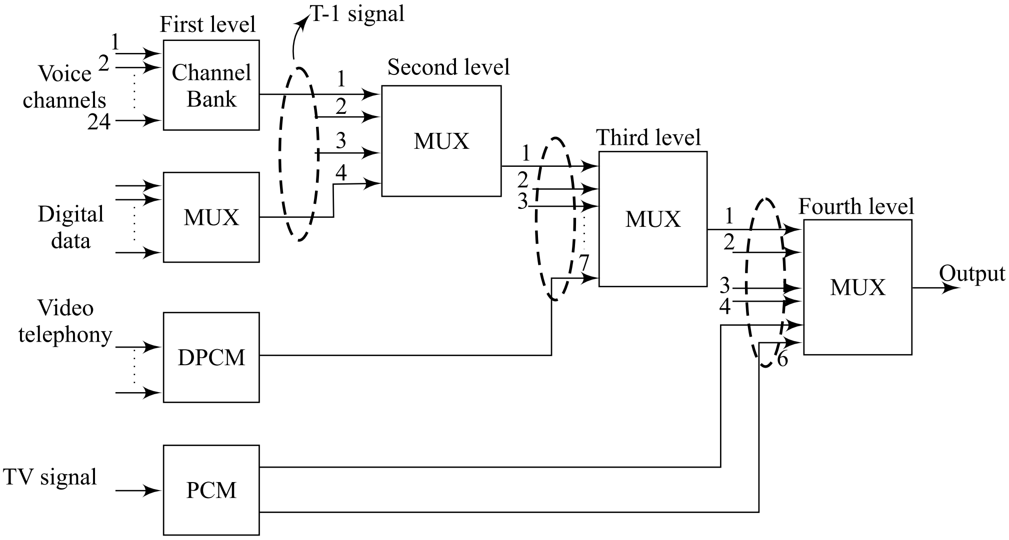


Fig. 9.34 *Multiplexing different types of signals*

In the above figure, what is shown as a channel bank consists of the following. Each voice channel has an anti-aliasing LPF followed by a sample-and-hold unit. The outputs of all these S/H units are added up in a summer. Since the S/H units are made to take the samples of the various analog sources in a sequential manner, the output of the adder will be in the form of sequence of frames, where the first frame will have the first samples of all the analog sources in the sequential order, the second frame will have the second samples of all the sources in the sequential order and so on. Each of these samples is quantized and encoded by the encoder to which these samples from the adder are fed, as shown in Fig. 9.35. The output of the encoder is a TDM-ed PCM signal with appropriate line coding. Finally, the processor adds or appends the control and signaling bits and produces a T-1 digital signal at its output.

The anti-aliasing filters band limit the analog voice signals to 3.2 kHz and the sampling is done at the rate of 8000 samples/sec. An 8-bit PCM encoding is done so that each sample is represented by eight binary bits. Since there are 24 message channels and each frame has one sample of each of these message channels, we may calculate the transmission speed of T-1 carrier as follows:

$$\frac{24 \text{ channels}}{\text{frame}} \times \frac{8 \text{ bits}}{\text{channel}} = 192 \text{ bits/frame}$$

and

$$\frac{192 \text{ bits}}{\text{frame}} \times \frac{8000 \text{ frames}}{\text{second}} = 1.536 \text{ Mbps}$$

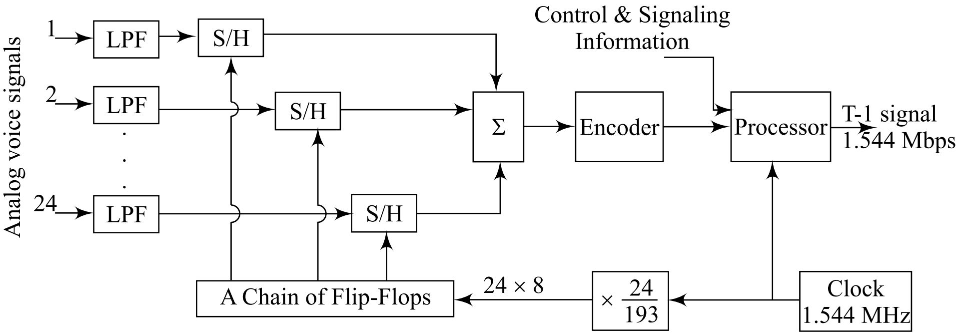
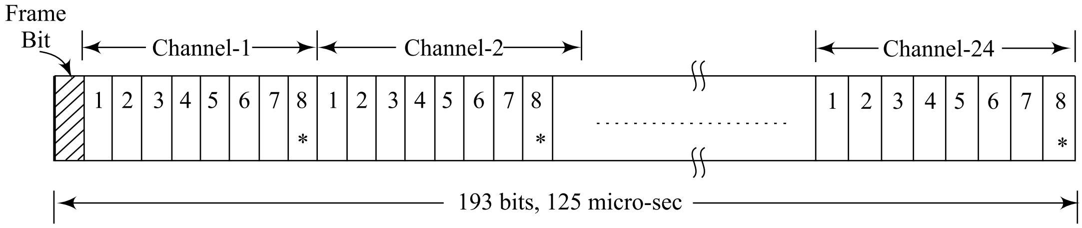


Fig. 9.35 *Details of the channel bank shown in Fig. 9.34*

In practice, a 'framing bit' is added in each frame. This is used in the receiver to maintain frame and sample synchronization between the transmitter and receiver. Hence, instead of 192, there are actually 193 bits per frame and so the transmission rate becomes:

$$\text{Transmission rate} = \frac{193 \text{ bits}}{\text{frame}} \times \frac{8000 \text{ frames}}{\text{second}} = 1.544 \text{ Mbps}$$

Channel banks In the initial stages, channel banks for T-1 carrier used only a 7-bit PCM code which represented only the sample magnitude. These channel banks, called the D-1 channel banks used analog companding with $\mu = 100$. Over the years, these channel banks have undergone several modifications and versions D-2, D-3, D-4, D-5 and D-6 have appeared. Of these, the last three, namely D-4, D-5 and D-6 use 8-bit sign-magnitude digitally compressed PCM codes with $\mu = 255$. In the earlier versions, out of the 8-bit sign-magnitude PCM codes one bit used to be allotted for signaling so that effectively the magnitude used to be represented by only six bits. In addition to reducing the resolution for magnitude, this was found to be creating another problem. One signaling bit per channel produced an 8 kbps signaling rate and this was excessive on the standard telephone voice circuits. Hence at present, a signaling bit is inserted in the position of the least significant bit (of the 8-bit PCM code), that too not in every frame, but only once in every six frames. This method of inserting signaling bits is called 'bit-robbing'. The signaling rate on each channel now reduces to 1.333 kbps and the average number of bits per sample increases to $7\frac{5}{6}$. Thus, the frame structure is as shown in Fig. 9.36. Thus, in T-1 carrier system, instead of using a separate channel for signaling purposes, the LSB slots normally used for voice information, are themselves used once in six frames, for the purpose of signaling. This arrangement is therefore referred to as 'channel associated signaling'.



* In this LSB slot, a signaling bit is inserted in every sixth frame only.

Fig. 9.36 *Frame structure showing frame bit and signaling bits*

9.5.4 Frame Synchronization

As already noted earlier, synchronization between the multiplexer and the de-multiplexer is essential for proper working of a digital communication system. Only when there is proper synchronization between the two, will the de-multiplexer be able to send the received data from each channel correctly to the corresponding channel at its output. Since data from various channels is interleaved sequentially, identifying the first channel data will enable the de-multiplexer to properly separate the data of all the channels of the received frame. So, the de-multiplexer must somehow be able to correctly identify the commencement of each frame. For this purpose, as shown in Fig. 9.37, one frame bit is included at the beginning of every frame. The pattern formed by 12 such frame bits occurring in 12 successive frames gives a 12-bit code called the *frame sync word*, which is known a priori to the receiver, and used by it for synchronization. This 12-bit code is repeatedly transmitted once every 12 frames.

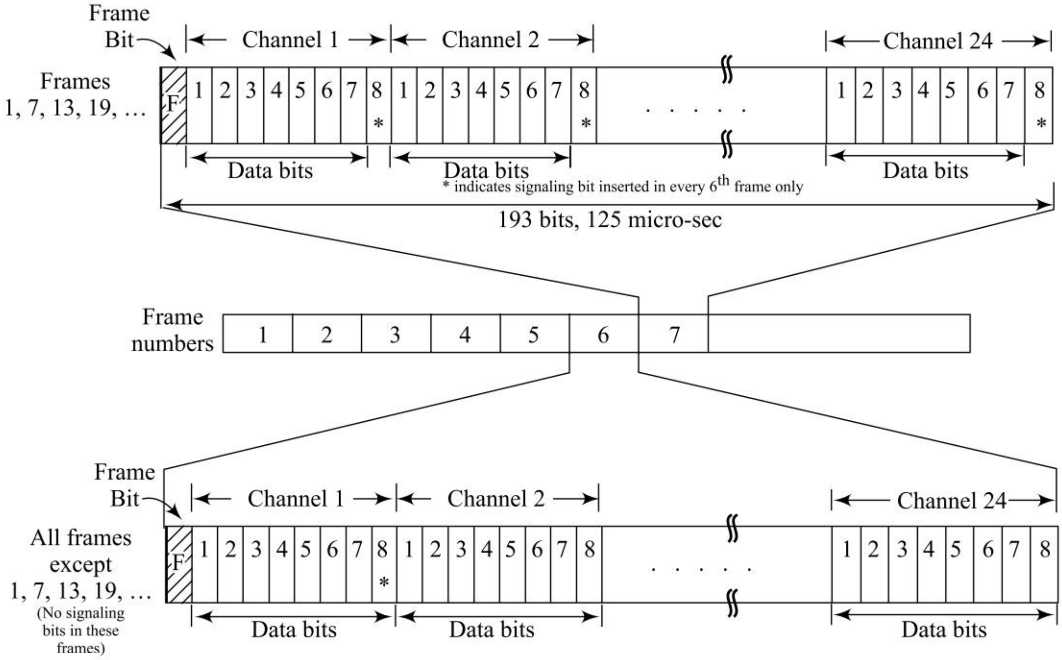


Fig. 9.37 *Frame details of a T1 carrier*

In the receiver, the received data stream is passed through a shift register of length equal to the length of the sync word, and an associated logic circuit decodes the sync word. If this sync word so decoded, exactly matches with the stored sync word, a synchronization pulse is generated and this is used by the receiver for frame synchronization. In an arrangement like this, there is a possibility that data bits, one from each successive frame and located exactly a frame length apart from each other, may by chance, exactly coincide with the frame synchronization word. This phenomenon is referred to as 'false synchronization' and results in the receiver identifying the frame boundary at a wrong location. The probability of occurrence of false synchronization is obviously

$$P_{FS} = (0.5)^N \tag{9.58}$$

where N is the number of bits in the frame sync word.

Further, since any bit in a frame may be affected by channel noise, there exists a possibility of one or more of the frame synchronization bits being decoded wrongly. When this happens, frame synchronization fails, as the receiver fails to decode the sync word. If the bit error probability is P_e, the probability of all the N bits of the sync code word being received correctly is $(1–P_e)^N$. Since there will be failure of detection of the sync code word if one or more of its bits are wrongly decoded in the receiver, the probability of frame sync code word detection failure is given by

$$P_{DF} = 1 - (1 - P_e)^N \cong NP_e \tag{9.59}$$

Example 9.15 What is the bit duration and what is the minimum bandwidth required for the T-1 carrier?

Solution We had already found that the bit-rate for a T-1 carrier is 1.544 Mbps.

Hence, minimum bandwidth required = $\frac{1}{2}$ (Bit-rate) = 772 kHz

The frame duration is (1/8000) sec = 125 μs since the sampling frequency is 8000 sps. Since 193 bits are included in each frame,

$$\text{Bit duration} = \frac{125 \times 10^{-6}}{193} = 0.64766\ \mu\text{s}$$

9.5.5 Digital Multiplexing Hierarchy

The very basic digital signal is the PCM-ed voice signal. Since the voice signal is sampled at the rate of 8000 sps and since each sample is represented by an 8-bit codeword, the bit-rate of the standard PCM-ed voice signal is 8 × 8000 = 64 kbps. This is generally referred to as the DS-0 signal or Digital Signal of level zero. Several such DS-0 signals are multiplexed together to form a higher level digital signal and this process is continued to get a very high bit-rate digital signal that represents the multiplexed version of several thousands of the basic PCM-ed voice signal. The several stages of multiplexing, using which, the high speed digital signal is formed, is what is generally referred to as the 'multiplexing hierarchy'.

There are two different hierarchies in vogue. One is the 'North American Hierarchy', or the AT&T hierarchy, which is followed by the North American countries and with some minor changes, by Japan. The other is the CCITT hierarchy, recommended by the ITU, which is mostly followed in the European countries.

1. North American hierarchy: The bit rates and the capacities (in terms of number of voice frequency channels) of these DS lines, also called the T-carriers are given in Table 9.3. The T-1 carriers carry binary digital signals which are represented using a Bipolar RZ line code (see line codes discussed in Section 10.2).

Table 9.3 *Bit-rates and capacities of T-carriers*

DS number and T-carrier No.		Bit rate in Mbps	No. of 64 kbps PCM VF Voice Channels	Transmission Media Used
DS-0		0.064	1	Twisted wire pairs
DS-1	T-1	1.544	24	Twisted wire pairs
DS-2	T-2	6.312	96	Twisted wire pairs / Fiber
DS-3	T-3	44.736	672	Coaxial cable / Radio / Fiber
DS-4	T-4	274.176	4032	Coaxial cable / Fiber
DS-5	T-5	560.160	8064	Coaxial cable / Fiber

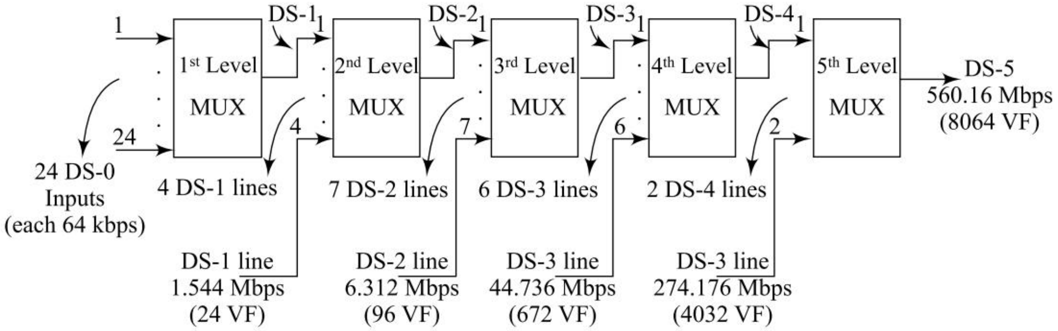


Fig. 9.38 *North American hierarchy*

2. CCITT digital multiplexing hierarchy: In this hierarchy, the first level of multiplexing involves 30 numbers of 64 kbps PCM-ed voice channels. This gives a 2.048 Mbps digital signal. Four such signals are multiplexed in the second level multiplexing to obtain a 8.448 Mbps digital signal. The third, also involves only four inputs to give a 34.368 Mbps multiplexed signal. Four such signals are multiplexed in the fourth-level multiplexer to obtain a 139.264 Mbps digital signal. Again four such signals are multiplexed in the 5th level to get a 565.148 Mbps signal.

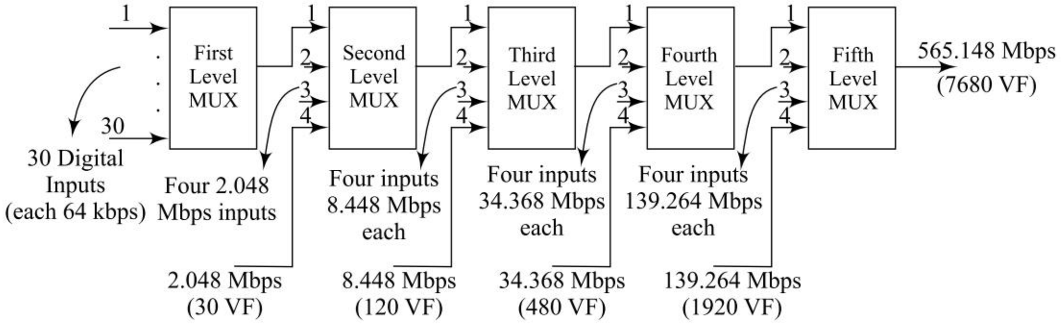


Fig. 9.39 *CCITT digital multiplexing hierarchy*

Bandwidth efficiency of a multiplexing system The bandwidth efficiency of a multiplexing system is defined as the ratio of the total bandwidth of all the message signals that have been multiplexed to the bandwidth of the multiplexed signal.

$$\therefore \quad \text{Bandwidth efficiency} = \frac{\text{No. of voice signals} \times \text{BW of each voice signal}}{\text{Bandwidth of the multiplexed signal}}$$

Example 9.16 Determine the minimum transmission bandwidth needed for a DS-5 signal in North American (or AT&T) hierarchy. Also, determine the bandwidth efficiency of this level of multiplexing.

Solution From Fig 9.38, the bit rate of the DS-5 signal is equal to 560.16 Mbps.

$$\text{The minimum transmission bandwidth} = \frac{1}{2}(\text{bit-rate}) = \frac{560.16}{2} \times 10^6$$

The DS-5 signal carries 8064 voice frequency signals, each of say 4 kHz.

$$\therefore \quad \text{bandwidth efficiency} = \frac{8064 \times (4 \times 10^3)}{280.08 \times 10^6} = 0.11516$$

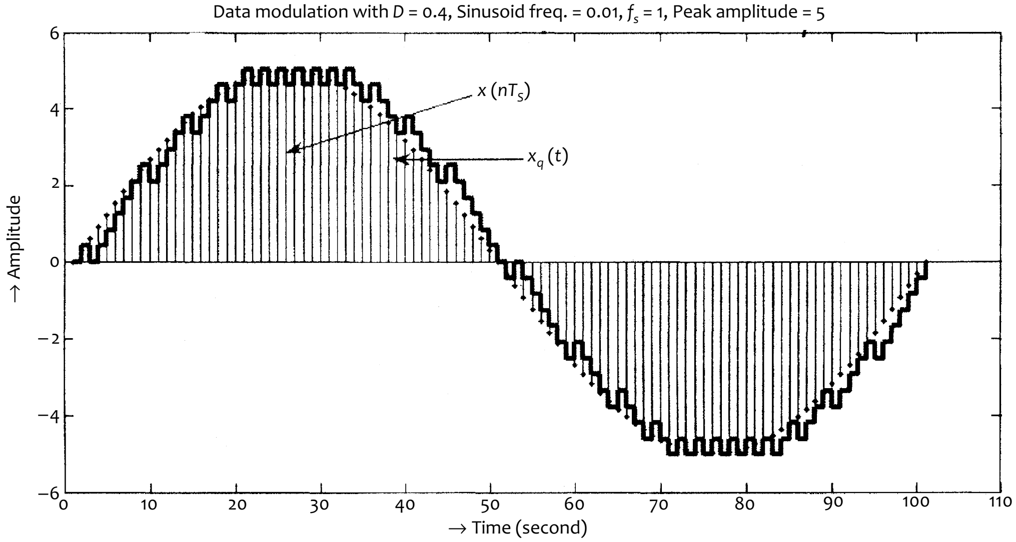
$$\therefore \quad \%\ \text{bandwidth efficiency} = 0.11516 \times 100 = 11.516\%$$

MATLAB Example 9.1 In this problem, we wish to demonstrate the effect of step size, Δ, on the output of a delta modulation encoder. For this purpose, we take a sinusoidal signal of peak amplitude $A = 5$ and frequency $f = 0.01$ Hz as the signal to be encoded. We sample this signal at the rate of 1 sample per second; i.e., 100 samples per cycle of the sinusoidal signal. We study the effect of the step size by encoding the sinusoidal signal at three different values of Δ.

MATLAB Program

```
function [y] = Delta_Modulation (del, A)
5 del = step size
% A = amplitude of signal
% y = output binary sequence
t = 0:2* pi/100:2* pi;
x = A* sin (t);
stem (x)
hold on
y = [0];
xr = 0;
for i = 1: length (x)-1
if xr(i) < = x(i)
d = 1;
xr (i + 1) = xr (i) + del;
else
d = 0;
xr (i + 1) = xr (i)-del;
end
y = [y];
end
stairs (xr)
hold off
end
```

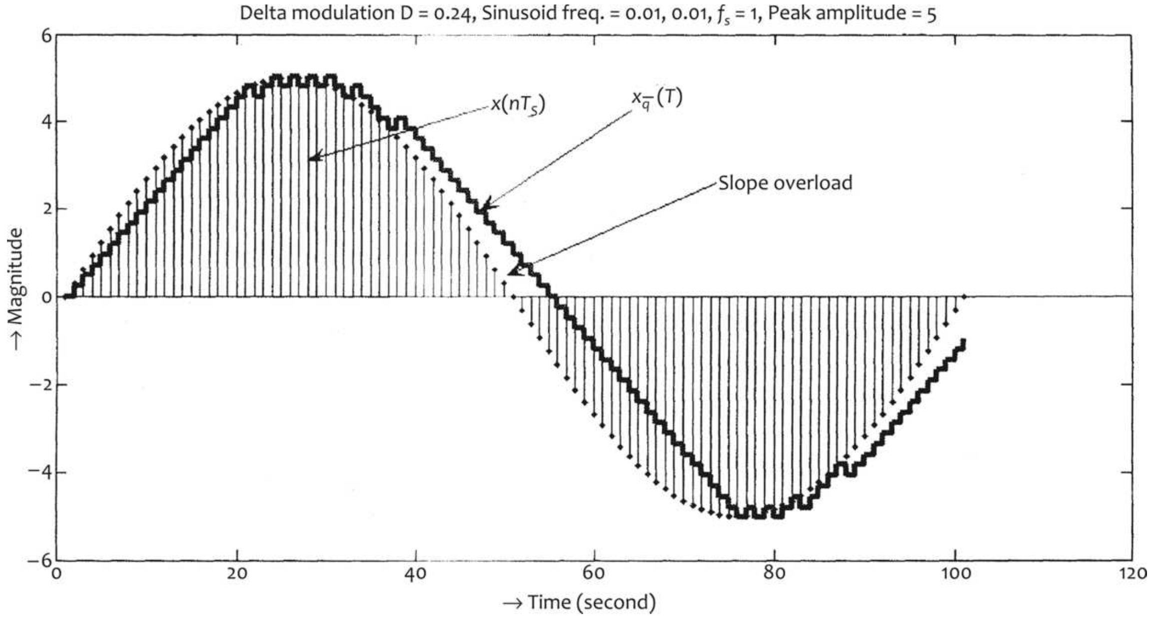
Results



Staircase approximation with Δ = 0.4

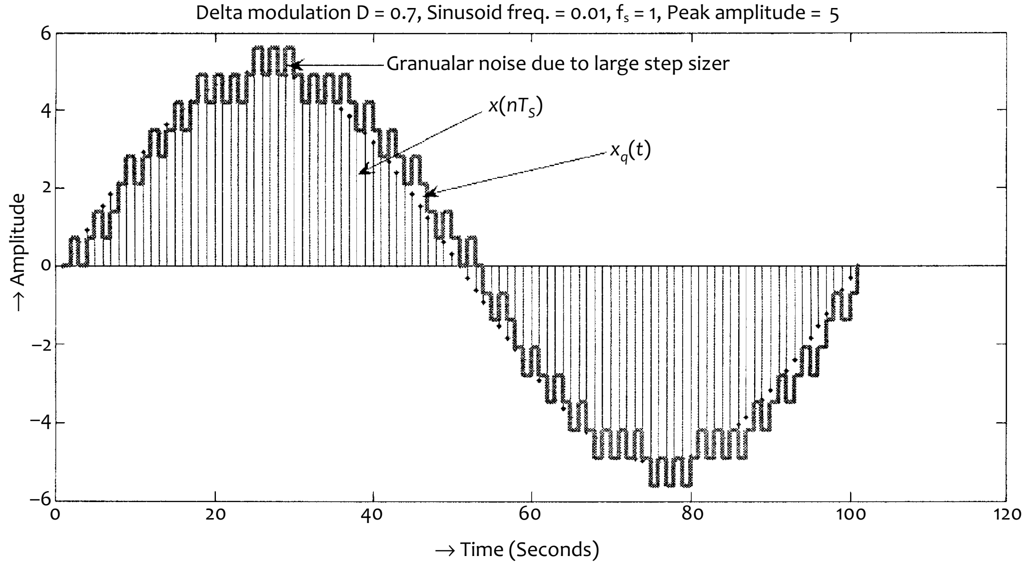
Fig. 9.39 (*Contd.*)

At Δ = 0.4, the staircase approximation generated by the encoder tracks the sinusoid fairly well without slope overload and with very little granular noise as can be seen in the figure



Staircase approximation with Δ = 0.24

At Δ = 0.24, there is considerable slope overload and negligible granular noise.



Staircase approximation with Δ = 0.7

At Δ = 0.7, there is no slope overload but there is considerable granular noise near the peaks of the sinusoid.

Fig. 9.39

Summary

- The steps involved in converting an analog $x(t)$ signal into a digital signal are (a) Band limiting $x(t)$ to W Hz, (b) Sampling at $f_s > 2W$, (c) Quantizing the samples, and (d) Encoding the quantized samples.
- In uniform quantization, the quantization levels are uniformly spaced.
- For signals like speech signal which have a large dynamic range but very low average power, compression is done for the samples before quantization at the transmitter and the decoded samples are subjected to expansion before being fed to the reconstruction filter in the receiver. This improves the *SNR*.
- If uniform quantization is used and Δ is the step size, $\overline{e_q^2}$ is the mean-square value of quantization error, then by assuming that e_q is uniformly distributed between $-\Delta/2$ and $+\Delta/2$, it can be shown that

$$\overline{e_q^2} = \frac{\Delta^2}{12} \quad \text{and} \quad (SNR)_q = 3\overline{x^2}\,2^{2n} \quad \text{if} \quad |x(t)| \le 1$$

- For n-bit binary PCM *with sinusoidal message signal,*

$$(SNR)_q \text{ in dB} = 1.8 + 6n$$

- For an n-bit binary PCM signal, the minimum transmission bandwidth is $B_T = nW$, where W Hz is the message bandwidth.
- $$\underset{\text{Binary PCM}}{(SNR)_D} = \underset{\text{(channel noise \& Quantization noise)}}{\text{Destination } SNR} = \frac{3Q^2\overline{x^2}}{1+4Q^2P_e}$$
- PCM with regenerative repeaters is best suited for long-haul signal transmission using a number of repeater stations.
- $(SNR)_q$ of PCM increases exponentially with the bandwidth provided the input SNR for the receiver is maintained above the threshold.
- In DM, the message sample, $x(n)$ is compared with $x_q(n-1)$, the quantized value of the previous message sample, and the difference between these two is subjected to single-bit quantization and binary encoding before being transmitted.
- DM suffers from quantization noise as well as slope overload noise. Slope overload noise can be eliminated by making $\frac{\Delta}{T_s} > \left[\frac{dx(t)}{dt}\right]$.
- Segments of $x(t)$ with steep gradients require a large value of step size Δ in order to avoid slope overload distortion. But segments of $x(t)$ in which rate of change is low require a small step size in order to have low granular noise.
- The destination *SNR* for DM is given by $\left(\frac{S}{N}\right)_D \le \left(\frac{3}{8\pi^2}\right)\left(\frac{f_s}{W}\right)^3$.
- DM requires high sampling rate and consequently the transmission bit-rate and bandwidth are also high. But its implementation is simple and inexpensive.
- In adaptive DM, the step size Δ is varied automatically depending upon the rate of change of $x(t)$. The Δ may be varied either continuously, or it may take a discrete set of values.
- In DPCM, a predictor is used to predict the present sample making use of a few past samples. The difference between the actual sample value and the predicted value is PCM-ed and transmitted.
- The linear predictor used in DPCM takes the linear combination of a few past samples to predict the present sample. The predictor coefficients are selected so to minimize the prediction error in the mean-square sense.
- $\boldsymbol{h}_{opt} = [R_X]^{-1}\boldsymbol{r}_X$, where $\boldsymbol{h} = [h_1 h_2 \ldots h_p]^T =$ weights vector, and $\boldsymbol{r}_X \underline{\Delta} [R_X(1) R_X(2) \ldots R_X(p)]^T$, where $R_X(k)$ is the auto-correlation of $x(t)$ for a lag of k.
- Processing gain of DPCM $= G_p \underline{\Delta} \left(\frac{\sigma_X^2}{\sigma_p^2}\right)$, where σ_X^2 and σ_p^2 are the variances of $x(t)$ and the prediction error, respectively.

- Just like DM, DPCM too has the granular noise and slope overload problems. That is why in ADPCM, both the step size Δ as well as the prediction coefficients are adaptively adjusted.
- Systems used for voice coding at low bit-rates are called VOCODERS.
- With a sampling rate of 8000 samples per second and 8-bit PCM, telephone quality speech needs a bit-rate of 64 kbps.
- There are three main approaches to compression of speech signals.
  - (a) *Time-domain techniques* like DPCM, ADPCM, Sub-band coding. DPCM and ADPCM give telephone quality speech at 24 kbps to 32 kbps. Sub-band coding gives telephone quality speech at 16 kbps.
  - (b) *Adaptive transform-domain techniques* can give telephone quality speech at 9.6 kbps.
  - (c) *Model-based techniques*: Linear predictive coding, a model-based technique gives telephone quality speech at as low a rate as 2.4 kbps to 4.8 kbps.
- Digital multiplexing is similar to TDM except that it is more flexible because of the freedom from constraints like periodic sampling and waveform preservation.
- A 'frame' is the smallest unit in the digital stream that contains at least one bit from each input.
- Digital multiplexers are basically of three types:
  - (a) Synchronous multiplexers
  - (b) Asynchronous multiplexers
  - (c) Quasi-synchronous multiplexers
- Advantages of digital multiplexing over analog multiplexing are:
  - (a) Low cost because of use of digital ICs for hardware
  - (b) Regenerative repeaters can be used for long-haul transmission
  - (c) It offers greater flexibility
- In the North American (or AT&T) digital multiplexing hierarchy, 24 PCM-ed voice channels, each of 64 kbps bit rate, are multiplexed into a DS-1 signal (or T-1) of 1.544 Mbps in the first level multiplexing. In the second level MUX, four DS-1 signals are multiplexed into a 6.312 Mbps DS-2. Then 7 DS-2 signals are multiplexed into a 44.736 Mbps DS-3 signal. Then 6 such DS-3 signals are multiplexed into a 274.176 Mbps DS-4 signal.
- In the CCITT hierarchy, the first stage has 30 inputs, each of 64 kbps. The second stage has 4 inputs, each of 2.048 Mbps; the third stage has 4 inputs, each of 8.448 Mbps; the fourth stage has 4 inputs each of 34.368 Mbps; and the fifth stage has 4 inputs each of 139.264 Mbps to give output multiplexed signal of 565.148 Mbps; which carries 7680 VF channels.
- In both AT&T as well as CCITT hierarchies, word interleaving only is used. Each frame of AT&T system has 193 bits including the 'frame synchronization bit'.
- Bandwidth efficiency of a digital multiplexing system is defined as:

$$\text{Bandwidth efficiency} = \frac{\text{No. of voice signals} \times \text{BW of each voice signal}}{\text{Bandwidth of the multiplexed signal}}$$

References and Suggested Reading

1. Abate, J.E., Linear and Adaptive Delta Modulation, *Proceedings of IEEE*, Vol. 55, pp. 298–309, 1967.
2. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems—An Introduction to Signals and Noise in Electrical Communication*, (McGraw-Hill International Edition), 4th Edition, McGraw-Hill, 2002.
3. Flanagan, J.L., M.R. Schroeder, B.S. Atal, R.E. Crochiere, N.S. Jayant, and J.M. Tribolet, Speech Coding, *IEEE Transactions on Communications*, Vol. com-27, pp. 710–737, April 1979.
4. Haykin, Simon, *Communication Systems*, 4th Edition, Wiley Student Edition, Wiley & Sons, New Delhi, 2001.
5. Jayant, N.S., Coding Speech at Low Bit Rates, *IEEE spectrum*, Vol. 23, No.8, pp, 58–63, 1986.
6. Jayant, N.S., Digital Encoding of Speech Waveforms, *Proceedings of IEEE*, May, 1974, pp. 611–632.
7. Jayant, N.S., and P. Knoll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, New Jersey, 1984.
8. R. Rabiner, Lawrence, and Bernard Gold, *Theory and Applications of Digital Signal Processing*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1975.

9. Reeves, A.H., The Past, Present and Future of PCM, *IEEE spectrum*, Vol.12, No. 5, pp. 89–63, 1975.
10. Sklar, B., *Digital Communications*, Prentice Hall, Englewood Cliffs, New Jersey, 1998.
11. Spanias, A.S., Speech Coding: A Tutorial Review, *Proceedings of IEEE*, Vol. 82, No. 10, October, 1994, pp. 1541–1582.
12. Taub, Herbert, Donald Schilling, and Gautam Saha, *Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill, New Delhi, 2008.
13. Tomasi, Wayne, *Advanced Electronic Communication Systems*, 6th Edition, (Low Price Edition), Pearson Education, 2004.
14. W. Couch, Leon II, *Digital and Analog Communication Systems*, 6th Edition, Pearson Education (Asia), 2001.
15. Zelinski, P., and Knoll, P., Adaptive Transform Coding of Speech Signals, *IEEE Transactions on ASSP*, Vol. 25, pp. 299–309, 1977.

Review Questions

1. What is meant by quantization and quantization noise?

2. Sketch the quantization characteristics of (a) a mid-tread quantizer, and (b) a mid-rise quantizer.

3. What is companding? Sketch the characteristics (input–output) of a compressor and an expander.

4. Briefly explain the principle of pulse-code modulation.

5. A PCM system uses a step size of Δ. Assuming that the quantization error is uniformly distributed, determine the mean-square value of the quantization error.

6. With the help of block schematic diagrams of the transmitter and the receiver, explain the working of binary PCM system.

7. In case companding is used, how will you modify the above block diagrams of the PCM transmitter and receiver?

8. If the message bandwidth is W Hz, show that the minimum possible bandwidth required for an n-bit PCM system is nW.

9. Define 'bit-error probability'.

10. When quantization noise as well as channel noise are considered, derive an expression for the destination (SNR) of a binary PCM system.

11. Sketch a typical input $(SNR)_{\text{dB}}$ vs. output $(SNR)_{\text{dB}}$ for a binary PCM system for $n = 4$, 6 and 8.

12. With reference to a PCM system, what is 'threshold effect'? How do you define the threshold $(SNR)_{\text{input}}$?

13. Explain the terms of the 'power-bandwidth trade-off' possible in a binary PCM system. Compare it with the trade-off possible in WBFM.

14. Discuss why PCM with regenerative repeaters, is considered the best option for long-haul signal transmission requiring the use of a large number of repeaters.

15. Discuss the advantages and disadvantages of PCM.

16. By drawing the block schematic diagrams of the transmitter and receiver and with the help of relevant waveforms, explain the working of a DM system.

17. What is meant by slope-overload distortion in a DM system? How can it be avoided?

18. Assuming a sinusoidal message signal and no slope overload, derive an expression for the maximum value of the destination signal-to-quantization noise ratio.

19. What are the advantages and disadvantages of DM?

20. What is a Delta–Sigma modulator? Where is it used?

21. With the help of relevant block diagrams explain the working of a continuously variable step-size adaptive Delta modulation system.

22. Compare PCM and DM.

23. With the help of block diagrams of the transmitter and receiver, explain the working of an ADM system with a discrete set of values for the step size.

24. Explain the basic principle of DPCM.

25. What is a linear predictor? On what basis are the predictor coefficients determined?

26. What is the need for adaptive DPCM?

27. State the different approaches for low bit-rate coding of speech.
28. Briefly explain how sub-band coding achieves low bit rate coding of speech.
29. Describe the transform domain approach for compression of speech.
30. List the various features/parameters of speech that are determined through speech analysis for the purpose of being used in speech synthesis?
31. Compare the efficacy of the various time-domain, frequency/transform-domain and model-based methods available for speech compression.
32. What is multiplexing?
33. What is TDM and what is the difference between analog TDM and digital multiplexing?
34. Describe the working of a typical channel bank.
35. Describe the North American digital multiplexing hierarchy.
36. What is the need for 'frame synchronization'? How is it achieved?
37. Describe the CCITT digital multiplexing hierarchy.
38. What is a frame? With the help of a neat sketch describe the frame structure in the AT&T system.
39. What is meant by bandwidth efficiency of a digital multiplexing system?
40. What is channel associated signaling?
41. What is meant by bit-robbing?

Problems

1. Consider a non-uniform quantizer with Δ_k as the k^{th} step size. If the probability of the message signal amplitude falling in the k^{th} interval is p_k, and if Δ_k is small compared to the peak-to-peak amplitude of the message signal, show that the mean-square value of the quantization error is approximately given by

$$\overline{e_q^2} = \frac{1}{12}\sum_k \Delta_k^2 p_k$$

2. A μ-law compressor is used to compress a message signal having a dynamic range of –40 V to +40 V, employing 256 quantization levels. Assuming $\mu = 255$, determine
   (a) the interval between two consecutive levels if no compression is used.
   (b) the minimum interval and the maximum interval between consecutive levels, if compression is used.
3. Repeat the above problem for A-law compressor for $A = 81.6$ and for the same dynamic range of the message signal as in the previous problem. The number of quantization levels are also the same, i.e., 256.
4. A message signal, $x(t)$ has an amplitude probability density function as shown in Fig. P9.4. Determine the locations of the eight quantization levels such that the probability of the message signal $x(t)$ taking a value between any two consecutive levels, or between the extreme levels and the edge voltages +1 V and –1 V is the same.

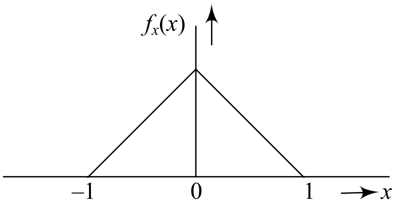


Fig. P9.4

5. An analog waveform with an amplitude range from –10 V to +10 V and a bandwidth of 200 Hz is to be PCM-ed and transmitted with an accuracy of ± 0.2% of the dynamic range of the signal. Determine the following:
   (a) The minimum sampling rate needed.
   (b) The number of bits/codeword.
   (c) The minimum bit-rate needed.
   (d) The minimum transmission bandwidth needed.
6. Channel noise is causing a bit-error rate of 1 in 10^4 in a certain binary PCM system employing polar NRZ signaling. If the system is to give a destination signal-to-noise ratio of at least 30 dB, determine (a) the number of bits/codeword and (b) the minimum transmission bandwidth needed if the analog message signal has a bandwidth of 3.2 kHz.
7. A sinusoidal message signal of 20 V peak to peak, is to be transmitted using n-bit binary PCM with $n = 10$ bits.

Determine the values of Q, Δ and its destination $(\text{SNR})_q$.

8. Determine the minimum memory size required to store sampled and quantized voice for 15 minutes, if the signal-to-quantization noise is to be at least 30 dB and the sampling rate is 8000 samples/sec.

9. A PCM system uses a uniform quantizer followed by an 8-bit binary encoder. If the bit rate of the system is 56×10^6 bits/sec., determine

(a) the maximum message bandwidth for which the system operates satisfactorily.

(b) the output $(S/N)_q$ when a full load sinusoidal message signal of frequency 2 MHz is applied to the quantizer of the system.

10. When the bit-error probability in a PCM system is $P_e \cong \left(\frac{1}{15}\right)2^{-2n}$, the $(S/N)_D$ decreases by 1 dB, i.e., threshold conditions are reached. Determine the $\gamma_{\text{th}} = \left(\frac{S_R}{\eta W}\right)_{\text{th}}$ for values of Q = 16,256 and 1024. Assume $S_R = 1$ and plot $(S/N)_D$ in dB against γ_{th} in dB.

11. In a binary PCM system, the output signal-to-quantizing noise ratio is to be held to a minimum value of 40 dB. Determine the number of required levels and find the corresponding output signal-to-quantizing noise ratio.

(GATE Examinations, 1997)

12. Consider an audio signal with spectral components limited to the frequency band of 300 Hz to 3300 Hz. A PCM signal is generated with a sampling rate of 8000 samples/sec. The required output signal-to-quantizing noise ratio is 30 dB.

(a) What is the minimum number of uniform quantizing levels needed, and what is the minimum number of bits per sample needed?

(b) Calculate the minimum system bandwidth needed?

(c) Repeat parts (a) and (b) when a μ-law compander is used with $\mu = 255$.

(Karnataka University, 1997)

13. A DPCM system has a processing gain of 6 dB. Show that a codeword of this DPCM system needs one bit less than that required for a binary PCM system, all other factors remaining the same.

14. A long-haul binary PCM communication system has nine regenerative repeaters. So, including the final decision at the receiver, there are a total of 10 sequential decisions taken on the transmitted binary PCM signal. If a binary 1 or 0 transmitted through the system has an independent probability p of being inverted in any repeater, determine the probability that a binary symbol finally received by the receiver is erroneous.

15. A DM system can handle message signals of bandwidth up to 5 kHz and has a sampling rate of 50 kHz. A sinusoidal signal of peak amplitude 1.5 V and frequency 2 kHz is applied to the system. Determine

(a) the step-size Δ required to avoid slope overload.

(b) the $(S/N)_q$ from the system for the given sinusoidal signal.

16. An LDM system operates with a sampling frequency of 30 kHz. If a sinusoidal signal $x(t)$, normalized so that $|x(t)| \leq 1$ whose frequency is 3 kHz, is applied, what value of Δ minimizes the slope overload?

17. Determine the processing gain of a DPCM system with a first order predictor, if the message signal has a normalized auto-correlation function of 0.8 for a lag of one sampling period, assuming that the predictor is designed to minimize the mean-square value of the prediction error.

18. A message signal $x(t)$ has a variance of $\overline{x^2}$ and a bandwidth of 3.2 kHz. It is known that $R_X(1) = 0.82R_X(0)$ and $R_X(2) = 0.56R_X(0)$. A DPCM system employing a second-order linear predictor is to be designed for transmission of the signal, $x(t)$, so that its destination $(SNR)_q$ is at least 30 dB. What will be the bit rate of this DPCM system?

19. A sinusoidal signal of frequency f_0 Hz is sampled at 15 f_0 Hz and the samples are given to a first-order linear predictor. Determine the weight w_1 of this predictor that would minimize the prediction error variance. Also determine the minimum prediction error.

20. A certain stationary random process has auto-correlation coefficients given by

$$R_X(k) = 1 - 0.2|k|; \quad |k| \leq 3$$

Determine the weights of an optimum linear predictor of order three and calculate the variance of its prediction error when these weights are used.

21. A PCM-TDM system multiplexes 24 voice channels, each of 0 Hz to 4 kHz bandwidth. If 7-bit PCM is used and a framing bit is added to each frame, what is the minimum line speed in bits/sec. and the corresponding minimum bandwidth needed?

22. A number of high fidelity audio channels, each band limited to 15 kHz, are to be transmitted using 12-bit binary PCM. Calculate how many of these PCM signals can be accommodated by the first level multiplexer of the AT&T multiplexing hierarchy. Also calculate the corresponding bandwidth efficiency.

Multiple-Choice Questions

1. Quantization is done in order to
  (a) improve the quality of the signal
  (b) improve the *SNR* at high frequencies
  (c) discretize the signal in the amplitude domain also
  (d) improve the *SNR* at low frequencies

2. Companding is used in the case of signals
  (a) having uniform amplitude distribution
  (b) having large peak-to-peak amplitude but small average power
  (c) having a large peak at high amplitude levels in their amplitude probability density function
  (d) with large average powers

3. Companding is used in PCM in order to
  (a) keep the quantization noise low for low-amplitude segments of a signal
  (b) avoid quantization noise
  (c) reduce the effect of impulse, or channel noise
  (d) reduce the complexity of the PCM system

4. In the mid-tread type of quantizer, any input value lying between –0.5 to +0.5 is mapped into an output value of
  (a) 0.5 (b) 1 (c) –0.5 (d) 0

5. In the mid-rise type of quantizer, any input value lying between 0 to 0.5 units is mapped into an output of
  (a) 0 (b) 0.25 (c) 1 (d) 0.5

6. In uniform quantization, as the step size is decreased the mean-square value of the quantization error will
  (a) decrease (b) increase (c) not change (d) None of these

7. A sinusoidal message signal is being transmitted by an 8-bit binary PCM. If the bits/codeword is reduced by a factor of 2, the output signal-to-quantization noise ratio will
  (a) reduce by 3 dB (b) reduce by 12 dB (c) reduce by 6 dB (d) reduce by 24 dB

8. A message signal with its amplitude uniformly distributed between –2 V and +2 V is transmitted by a 4-bit binary PCM system. The $(SNR)_q$ is equal to
  (a) 256 (b) 1024 (c) 512 (d) 768

9. A binary PCM system with 256 quantizing levels has a sampling frequency of 7 kHz. The bit rate of the system is
  (a) 56 kbps (b) 28 kbps (c) 1792 kbps (d) 896 kbps

10. Channel noise dominates the $(S/N)_D$ is decided only by this noise when
  (a) $P_e << \frac{1}{4Q^2}$ (b) $P_e < \frac{1}{Q^2}$ (c) $P_e >> \frac{1}{4Q^2}$ (d) $P_e >> 4Q^2$

11. In binary PCM, the $(SNR)_D$
  (a) increases linearly with the number of bits/codeword
  (b) increases as the square of the number of bits/codeword
  (c) does not depend upon the number of bits/codeword
  (d) increases exponentially with the number of bits/codeword

12. In a linear DM system
  (a) only granular noise will be present

(b) only slope overload noise will be present
(c) both granular noise as well as slope overload noise can be eliminated
(d) granular noise will be present but slope overload noise can be avoided by proper design

13. In an LDM system, a large step size will
(a) increase the slope overload noise, and reduce the granular noise
(b) reduce the slope overload noise but will increases the granular noise
(c) reduce the slope overload noise as well as the granular noise
(d) increase both the overload noise and the granular noise

14. For an LDM system with a sinusoidal signal, a peak amplitude A and frequency f_0 have been found to just avoid slope overload condition. Now if the frequency is doubled, slope overload can be just avoided by
(a) halving the peak amplitude of the sinusoid (b) doubling the peak amplitude of the sinusoid
(c) making the peak amplitude to be $(A/4)$ (d) making the peak amplitude to be $4A$

15. An LDM system with a sinusoidal message signal of peak amplitude A and frequency f_0, employs a sampling frequency which is $20f_0$. It is found that a peak amplitude of A_1 and frequency f_0 of the message signal just avoid slope overload condition. If now the peak amplitude is made $2A_1$, to just avoid slope overload,
(a) the sampling should be done at $10f_0$ (b) the sampling should be done at $5f_0$
(c) the sampling should be done at $40f_0$ (d) the sampling should be done at $80f_0$

16. The greatest advantage of LDM is
(a) it requires very low sampling rates
(b) it does not produce quantization noise
(c) it uses only one-bit representation of each error sample and so the bit rate is quite low
(d) its transmitter and receiver require very simple and inexpensive hardware

17. The linear predictor in a DP|CM system is generally implemented as
(a) a transversal digital filter (b) a recursive digital filter
(c) an analog *R*-*C* filter (d) None of the above

18. Generally the linear predictor's weights are so chosen that
(a) the absolute value of the prediction error is minimized
(b) the mean-square value of the prediction error is minimized
(c) the maximum value of the prediction error is minimized
(d) None of the above

19. The number of bits per frame and the bit-rate of a T-1 carrier signal are respectively
(a) 192, 1.544 Mbps (b) 193, 1.536 Mbps (c) 192, 1.536 Mbps (d) 193, 1.544 Mbps

20. The number of input lines for the first, second, third, fourth and fifth level digital multiplexers in the North American hierarchy are respectively
(a) 24, 4, 6, 7, 2 (b) 24, 4, 7, 6, 2 (c) 24, 6, 4, 7, 2 (d) 30, 4, 4, 4, 4

21. The bandwidth efficiency of the first level multiplexing in the CCITT hierarchy, assuming 4 kHz bandwidth for each audio channel is
(a) 11.72% (b) 12.46% (c) 10.23% (d) 1

Answers for Multiple-Choice Questions

1.	(c)	2.	(b)	3.	(a)	4.	(d)	5.	(d)	6.	(a)	7.	(d)	8.	(b)
9.	(a)	10.	(c)	11.	(d)	12.	(d)	13.	(b)	14.	(a)	15.	(c)	16.	(d)
17.	(a)	18.	(b)	19.	(d)	20.	(b)	21.	(a)						

DIGITAL DATA TRANSMISSION TECHNIQUES—BASEBAND AND BAND PASS

10

> *"The best and safest thing is to keep a balance in your life, acknowledge the great powers around us and in us. If you can do that, and live that way, you are really a wise man."*
>
> **Euripides (c. 480–406 BC)**
> *Greek writer*

Learning Objectives

After going through this chapter, students will be able to

- *understand the difference between 'binary signaling' and 'multilevel signaling', and their relative merits,*
- *become familiar with NRZ, RZ, Polar, Bipolar and Manchester line codes, their properties and areas of application,*
- *understand the causes for the occurrence of ' inter-symbol interference' and how techniques such as 'pulse shaping' and correlative coding' are helpful,*
- *understand the need for equalization and the techniques to be adapted for equalization of a dispersive digital base band channel,*
- *understand the usefulness of an 'eye pattern' and draw interferences on the performance of a data transmission system from its eye pattern,*
- *understand the different basic digital band pass modulation schemes like binary ASK, FSK and PSK; the quadrature modulation schemes such as QPSK, OQPSK, π/4-shifted QPSK and QAM; and the M-ary modulation schemes like M-ary QAM, M-ary FSK and M-ary PSK,*
- *determine the power spectra and the bandwidth of ASK, FSK, QPSK, and MSK signals, draw the signal space diagrams and sketch the signal constellations, and*
- *understand the need and the techniques adopted for symbol-level and word-level synchronization in all digital communication systems and the need and the techniques adopted for carrier recovery in the case of coherent bandpass digital communication systems.*

10.1 INTRODUCTION

In Chapter 9, we had discussed the various waveform coding methods like PCM, DM and DPCM for generation of digital signals. We had also discussed methods for multiplexing digital data from different sources.

Having studied the methods of generation and multiplexing of digital signals, in the present chapter, we will be looking at the problem of transmission of digital data. Transmission of digital data without sinusoidal carrier modulation is known as **baseband transmission**. These baseband signals possess considerable low frequency content and therefore cannot be transmitted over a radio link channel. Hence, they are transmitted over a pair of wires, a coaxial cable, or an optical fiber. If the digital data modulates a sinusoidal carrier and the modulated band pass signal is transmitted, it is called **band pass transmission**. These modulated digital signals can be transmitted either on terrestrial radio links, or on satellite radio links. In this chapter, therefore, we will discuss both baseband transmission as well as band pass transmission.

Generally, the digital baseband channels are dispersive and the pulses are not confined to their respective time slots when they arrive at the receiver. Each pulse is influenced by its adjacent pulses, causing what is referred to as 'intersymbol interference', or, ISI. This can lead to wrong decisions by the decoder of the receiver which has to decide during each time slot whether what was transmitted during that time slot was a 1 or a 0. Hence, we will be examining certain methods like 'pulse shaping', duo-binary signaling and equalization techniques to combat the effect of ISI.

There are three basic types of modulation schemes, viz., Amplitude-Shift Keying (ASK), Frequency-Shift Keying (FSK), and Phase-Shift Keying (PSK). These are somewhat analogous to their counter part analog modulation schemes—Amplitude Modulation (AM), Frequency Modulation (FM), and Phase Modulation (PM), respectively. As in the case of FM and PM, their counterparts FSK and PSK also have the constant envelope feature, which makes them best suited for transmission over non-linear band pass channels such as those encountered in microwave terrestrial and satellite links in which the traveling-wave tube (TWT) amplifiers work near their maximum power handling capacity by going beyond the linear region of their operation. There are some digital modulation schemes which may be viewed as hybrid type. A typical example of such a hybrid modulation scheme is the Quadriphase Amplitude Modulation. As we will be seeing later, this has some very attractive features.

Baseband transmission as well as band pass transmission can either be binary or M-ary. Baseband transmission in which a symbol, a bit, can take one of two possible values, like A and 0 (unipolar), or A and $-A$ (polar), is called binary baseband transmission. In M-ary baseband transmission, on the other hand, a symbol can take M possible values. In M-ary band pass signaling, in a symbol period of τ seconds which is equal to nT seconds where T is the bit interval, any one of a set of $M = 2^n$ possible signals, $s_1(t)$, $s_2(t)_,$ … $s_M(t)$ may be transmitted. In M-ary baseband signaling, these M signals, $s_1(t)$, $s_2(t)_,$ … $s_M(t)$, will be rectangular pulses, but of M distinct amplitudes. In M-ary ASK, these M possible signals will have M distinct carrier levels. In M-ary FSK, they will all have the same amplitude but M different carrier frequencies. In M-ary PSK, they will have the same amplitude and frequency but M different phase angles. In Chapter 2, we made use of the analogy between signals and vectors and developed the concept of a signal space. **Signal constellation** is the set of message points in a signal space corresponding to the set of all the transmitted signals.

In the study of digital band pass signaling, it is useful to examine the power spectra of the modulated signals, as such spectra provide an insight into the bandwidth of the modulated signal and also the possibility of interference with adjacent channels when multiplexing is used. However, we do not derive the spectra of the modulated signals. Instead, we make use of the fact that the spectrum of a band pass signal is known, once the spectrum of its complex envelope, or its low pass equivalent, is known. Hence, it is enough if we examine the PSD of the baseband signal.

Irrespective of whether it is baseband signaling or band pass signaling and whether it is binary or M-ary, one more aspect of digital transmission that deserves our attention and which we have discussed to some extent in connection with baseband signaling, is the 'bandwidth efficiency', or 'spectral efficiency'. This is defined as the ratio (R/B_T), i.e., the ratio of 'signaling rate' to the 'bandwidth required'. The bandwidth efficiency is important because it tells us how fast we can signal for a given bandwidth, be it baseband or band pass signaling.

In the case of baseband systems, it depends on pulse-shaping and on whether it is binary or M-ary signaling. As we have already seen, M-ary signaling gives better bandwidth efficiency. With a pulse shape corresponding to a ", it depends on the roll-off factor, ρ, and decreases with increasing value of ρ (between 0 and 1). In band pass systems, the bandwidth efficiency depends on the type of modulation and for a given modulation scheme, on whether it is binary or M-ary signaling. For a specified rate of transmission, as M increases, the required bandwidth decreases and so the spectral efficiency improves.

However, whether it is baseband signaling or band pass signaling, the value of M cannot be increased indefinitely, as the power requirement goes up if a specified probability of error is to be maintained.

10.2 LINE CODES FOR BINARY SIGNALS

Earlier, while discussing PCM, we had seen that each of the 2^n quantization levels could be uniquely identified by an n-digit binary number. This made it possible to state the value of a quantized sample by the binary code number associated with the corresponding quantization level. Then we stated that these binary code words could be transmitted to the receiver, where, by passing the quantized samples through a reconstruction filter, the message signal could be reconstructed.

To transmit the code words over the channel, we must devise a method to electrically represent the binary digits '0' and '1'. While introducing PCM, for simplicity, we had stated that we could conveniently represent a binary 1 by a positive pulse of V volts amplitude and a binary '0' by a 'no-pulse', i.e., a zero volts pulse or alternatively a –V volts pulse. In addition to these two, there are several other ways of representing the binary digits electrically. *These electrical representations of binary codes are called 'line codes*'. We shall now briefly discuss five popular line codes including the two stated earlier. Any binary data stream is a random sequence of the binary digits '0' and '1'. Hence, the power spectral density (PSD) of such a random binary sequence would depend upon the line code used to represent the binary digits. The shapes of the PSDs of the line codes are useful because they give the spectral content of a bit stream for any particular line code that is used. Figure 10.1 gives the waveform representing the codeword 01100101 when each of the line codes is used, and Fig. 10.2 gives the sketches of their PSDs under the following assumptions: (i) that binary 0s and 1s are equally probable; (ii) that frequency is normalized with respect to $1/T$, where T is the bit-slot time. (iii) that the average powers of the electrical waveforms obtained with different line codes is normalized to a value of 1. We shall first describe these five line codes illustrated in Fig. 10.1.

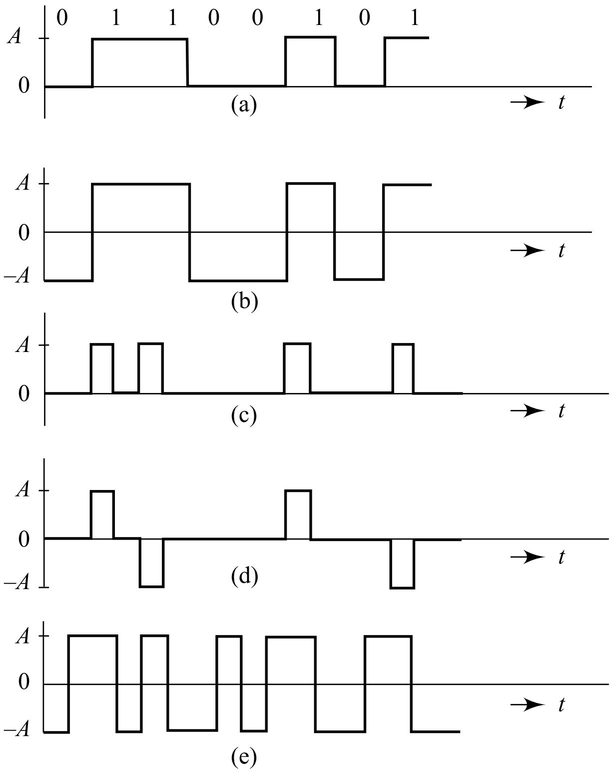


Fig. 10.1 *Line codes for binary data: (a) Unipolar NRZ signaling, (b) Polar NRZ signaling, (c) Unipolar RZ signaling, (d) Bipolar RZ signaling (also called AMI signaling), (e) Split-phase, or Manchester code*

10.2.1 Unipolar Non-Return to Zero (NRZ) Code

In this line code, a binary 1 is represented by a positive pulse of some amplitude, say, A, for the duration of the time slot; and a '0' is represented by a 'no-pulse', i.e., absence of any pulse for the full duration of its time slot. This is shown in Fig. 10.1(a). This code suffers from the disadvantage of a large dc component which results in wastage of transmitted power. This dc component appears as a delta function of strength $A/2$ in its PSD, as shown in Fig. 10.2(a). Further, the spectrum has a fairly large value near and at zero frequency, indicating large low-frequency content.

10.2.2 Polar Non-Return-to-Zero (NRZ) Code

In this code, a binary '1' is represented by a positive pulse occupying the full duration of the a time slot, and a binary '0' is represented by a negative rectangular pulse of equal amplitude, for the full duration of the time slot, T, as shown in Fig.10.1(b). Although easy to generate, this code suffers from the disadvantage of its PSD having a very large value at and near zero frequency, indicating a very large low frequency content. But, of course, it does not have a dc component.

10.2.3 Unipolar Return-to-Zero Code

In this code, a binary '1' is represented by a rectangular positive pulse of amplitude say, A, but of duration equal to only half the duration of the time slot; and a binary '0' is represented by a no-pulse for the entire duration of the time slot T, as shown in Fig.10.1(c). As may be seen from Fig. 10.2(c), its PSD has delta functions at $f = 0$, as well as $\pm(1/T)$. This delta function at $f = \pm 1/T$ means that there is a frequency component at that frequency in the data transmitted using this line code. This can conveniently be used for clock (bit timing) recovery at the receiving-end. However, it has two disadvantages: (i) it has a D.C. component and a fairly large low frequency content; and (ii) it requires 3 dB more power compared to polar return-to-zero signaling for the same probability of error.

10.2.4 Bipolar Return-to-Zero (BRZ) Code

Half slot-width duration positive and negative rectangular pulses of amplitude A are used for representing alternate binary 1s and a no-pulse' is used for representing the binary '0'. It is also known as *Alternate Mark Inversion* (AMI) *code*. The main advantage of this code is that the PSD of the bit stream will not have any dc component and further, it has negligible low-frequency content, when binary 1s and 0s occur with equal probability, as assumed earlier.

10.2.5 Split Phase or Manchester Code

In this code, as shown in Fig. 10.1(e), binary 1 is represented by a positive rectangular pulse of amplitude A followed by a negative rectangular pulse of the same amplitude, both these pulses having widths of half the time slot T. Binary '0' is represented again by two half-width pulses of the same amplitude and opposite signs, but occurring in the reverse order compared to the way they occur in the representation of a binary '1'. That is, for binary '0', the negative half-width pulse comes first and is followed by the positive half-width pulse. When this line code is used, irrespective of whether a '1' and a '0' have the same probability of occurrence, or not, the PSD of the bit stream will have no dc component.

Note

(i) Frequency has been normalized with respect to (1/T).

(ii) Average power has been normalized to unity (That is why different values have been given to A).

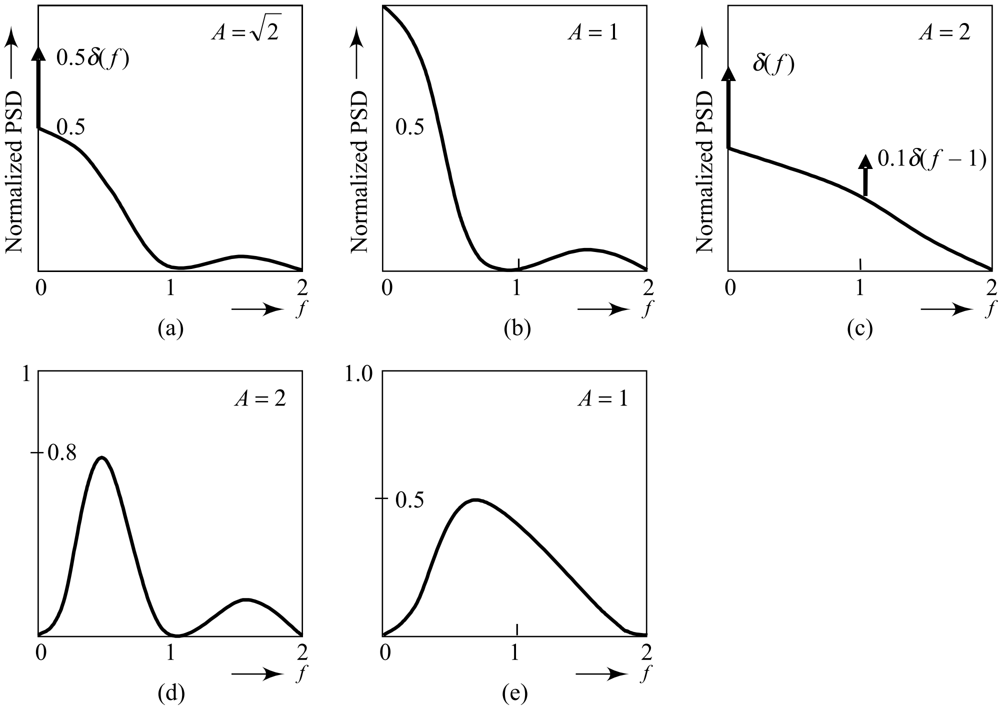


Fig. 10.2 *Power spectra of various line codes: (a) Unipolar NRZ, (b) Polar NRZ, (c) Unipolar RZ, (d) Bipolar RZ (AMI), (e) Split-phase (Manchester) code*

Some desirable properties of line codes

1. It would be possible to have an ac coupling to the channel if the spectrum of the line-code has negligible low-frequency components and no dc component.
2. If the line code has in-built timing information (as in the case of unipolar RZ code), it would be easy for the receiver to extract the clock signal.
3. The line code should have a sufficiently small bandwidth so that the channel bandwidth is adequate for transmission of digital data (using this line code) without ISI causing any problem.
4. Channel encoding for incorporating error-detecting capabilities should be easy.

10.3 ISI AND NYQUIST CRITERION

In a baseband digital transmission system, the two main sources of error are the ISI and the additive channel noise. We will discuss ISI and the various ways of combating its effect in this chapter. Methods of tackling the effect of channel noise will be discussed in Chapter 11 in which the detection of baseband as well as band pass digital signals is proposed to be discussed in detail.

A digital data signal, as we have seen, consists of a sequence of pulses, where each pulse is confined to its own time slot. However, when this signal is transmitted over a channel that is dispersive, the pulses are no longer confined to their respective time slots by the time the signal arrives at the receiving-end. Instead, the pulses spill over to the adjacent time slots, influencing the amplitudes of the pulses in those adjacent time slots. This is referred to as ‘inter-symbol interference’, or ISI. This can lead to the decision-making device making wrong decisions with regard to the symbols that were transmitted in those affected time slots, thus causing errors.

Since inter-symbol interference anyhow does take place in a dispersive channel, we shall adopt the following approach to overcome the ISI problem. We would like to examine whether there is any particular class of received pulse shapes which would allow us to correctly determine the amplitude of the received pulse pertaining only to a particular time slot even in the presence of ISI, so that we can avoid the effects of ISI. Once we identify any such *desirable* received pulse shape, we can use appropriately designed filters at the transmitting and receiving ends of the channel so that after passage through the transmit filter, the channel and the receive filter, the received pulse at the input to the sampler in the decision circuit has that desired pulse shape. This, of course, assumes that the transfer function of the channel is known *a priori*.

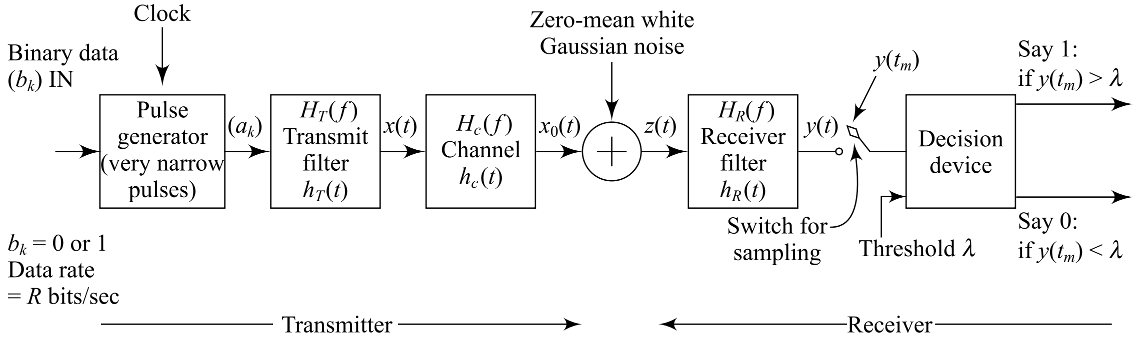


Fig. 10.3 *Binary baseband system*

A binary baseband system is shown in Fig. 10.3. Binary sequence $\{b_k\}$, $b_k = 0$ or 1, is fed to a narrow-width pulse generator which produces a sequence of very narrow-width pulses of amplitude a_k, where,

$$a_k = \begin{cases} 1 & \text{if } b_k = 1 \\ -1 & \text{if } b_k = 0 \end{cases} \tag{10.1}$$

These pulses are so narrow that they can be approximated as impulses. The transmit filter which is excited by these impulses of strength a_k, produces a sequence of weighted impulse responses. Let the impulse response be represented for convenience, as $p_T(t)$, i.e., $p_T(t) = h_T(t)$ Hence, we may write

$$x(t) = \sum_{k=-\infty}^{\infty} a_k p_T(t - k/r) \tag{10.2}$$

Let $p_T(0)$ be assumed to have been normalized to a value 1. If the bit rate of the data is r bits/sec and if we define

$$T \underset{-}{\Delta} \frac{1}{R_b} \tag{10.3}$$

Then,

$$x(t) = \sum_{k=-\infty}^{\infty} a_k p_T(t - kT);\ p_T(0) = 1 \tag{10.4}$$

When this signal $x(t)$ passes through the channel and the receive filter,

1. the pulses $p_T(t)$s will get distorted because of the finite and inadequate bandwidth.
2. Noise will get added to the signal. This noise is assumed to be additive zero-mean white Gaussian noise.
3. A certain time-delay t_d will be introduced.

We shall assume, without loss of generality, that $t_d = 0$. If the pulse at the output of the receive filter is designated as $p(t)$, (a distorted version of $p_T(t)$), then the signal $y(t)$ at the output of the receive filter may be written down as

$$y(t) = \left[\sum_{k=-\infty}^{\infty} A_k p(t - k/R_b) \right] + n(t) \tag{10.5}$$

where

A_k = attenuated version of a_k

$n(t)$ = noise component in $y(t)$. This noise will be Gaussian and zero mean, but will not be white.

$p(\cdot)$ = distorted form of the pulse $p_T(t)$. It is assumed that this pulse is normalized so that $p(0) = 1$.

The sampler, shown as a switch in Fig.10.3, samples $y(t)$ at some pre-set optimum instant t_m during each time slot of duration T sec. The sample of $y(t)$ at $t = t_m$, viz, $y(t_m)$, is a random variable, and it forms the 'observed variable', based on the value of which the next section, i.e., the decision device decides (by comparing $y(t_m)$ with the threshold, say λ), whether what was transmitted during that time slot, was a binary 1 or a binary 0. In fact, if

$y(t_m) > \lambda$, the receiver decides it was a binary 1

and if $y(t_m) < \lambda$, the receiver decides it was a binary 0

When we sample $y(t)$ at $t = t_m = m/R_b$, Eq. (10.5) may be written as

$$y(t_m) = \left[\sum_{k=-\infty}^{\infty} A_k p\left(\frac{m-k}{r}\right) \right] + n(t_m)$$

$$= \left[\sum_{k=-\infty}^{\infty} A_k p(mT - kT) \right] + n(t_m)$$

$$\therefore \quad y(t_m) = A_m + \sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} A_k p(mT - kT) + n(t_m) \tag{10.6}$$

In the above equation, the first term A_m represents the correct, or the desired output from the sampler. Ideally, we would have got only this term in the absence of ISI and noise. The second term, viz.,

$$\sum_{\substack{k=-\infty \\ k \neq m}}^{\infty} A_k p(mT - kT) \tag{10.7}$$

is the ISI term because it represents the output of the sampler contributed by all the past and future digits – all except the correct one, i.e., the digit corresponding to the time slot in which the sample has been taken. The last term represents the noise component in the sample.

The pulse $A_k p(t)$ occurring in Eq. (10.5) is obtained, as shown in Fig. 10.4 by feeding the impulse response of the transmit filter to the channel and then feeding the channel output to the input of the receive filter.

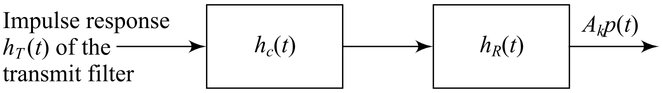


Fig. 10.4 *Generation of the pulse p(t)*

$$\therefore \quad A_k p(t) = h_T(t) * h_C(t) * h_R(t) \tag{10.8}$$

By taking the Fourier transform on both sides of the above equation, we get

$$A_K P(f) = H_T(f) \cdot H_C(f) \cdot H_R(f) \tag{10.9}$$

Since both ISI and noise cause the decoding errors, the method adopted in the design of $H_T(f)$ and $H_R(f)$ for a given, or known $H_C(f)$, is one that is aimed at reducing the effect of both these on the probability of error. Now, we shall first examine how the shape of the pulse $p(t)$ might be selected so as to eliminate ISI, and then how the filters might be designed, for obtaining the particular shape of $p(t)$ at the output of the receive filter.

Nyquist criterion for distortionless baseband transmission As stated earlier, it is the random variable $y(t_m)$, the sample of the receive filter output corresponding to the m^{th} time slot, that is used by the decision device for making decision about that time slot. So, Eq. (10.6) says that if $y(t_m)$ is to be simply equal to A_m so that in the absence of noise, there is no error, the effect of the ISI term should be equal to zero. This means, we should ensure that

$$p(mT - kT) = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{if } k \neq m \end{cases} \tag{10.10}$$

So, the shape of the pulse $p(t)$ at the receive filter output should be such that the above conditions are satisfied. When these conditions are satisfied, insofar as the observed variable $y(t_m)$ is concerned, ISI does not have any component in it and so it cannot play any role in the decision making. Thus, the *effect* of ISI is completely eliminated even though ISI itself is not.

Since both m and k in Eq. (10.10) are integers, the conditions for zero ISI implied in that equation may be restated as

$$\underset{l\text{ is an integer}}{p(lT)} = \begin{cases} 1 & \text{if } l = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \tag{10.11}$$

Equation (10.11) states that $p(0)$ should be equal to 1 and that $p(t)$ should be equal to 0 for $t = lT$ where $l = \pm1, \pm2, \pm3, \ldots$ for zero ISI. Equation (10.11) thus defines a class of signals that make it possible to have zero ISI. The reader would have already guessed that a sinc function satisfies the two conditions of Eq. (10.11) and that it therefore belongs to that class of signals.

Equation (10.11) states in the time domain, the two conditions that the received pulse shape $p(t)$ should satisfy for zero ISI. A frequency domain version of this equation throws some more light on this class of signals. For this purpose, imagine that $p(t)$ is ideally sampled at regular intervals of T, by taking the product of $p(t)$ with a unit-strength impulse train with impulses located at $t = 0, \pm T, \pm 2T, \ldots$. Since $p(t)$ is such that $p(0) = 1$ and $p(lT) = 0$ for $l = \pm1, \pm2, \ldots$, it follows that the product of $p(t)$ with the impulse train yields $p(0)\,\delta(t) = \delta(t)$ only, since $\delta(t - kT_s)$ for $k = \pm1, \pm2, \ldots$ face zero values of $p(t)$. Hence,

$$p_\delta(t) = \delta(t) \tag{10.12}$$

Taking the Fourier transform of $p_\delta(t)$, we have

$$\mathcal{F}[p_\delta(t)] = P_\delta(f) = \frac{1}{T}\left[\sum_{n=-\infty}^{\infty} p(f - nf_s)\right] = \mathcal{F}[\delta(t)] = 1 \quad \left(\text{where } f_s \Delta \frac{1}{T}\right)$$

$$\therefore \qquad \frac{1}{T}\sum_{n=-\infty}^{\infty} P(f - nf_s) = 1$$

or

$$\sum_{n=-\infty}^{\infty} P(f - nf_s) = T \tag{10.13}$$

Hence, the pulse $p(t)$ should have a spectrum $P(f)$ that satisfies Eq. (10.13) if it is to have zero ISI, since we have made use of both the conditions implied in Eq. (10.11) in order to arrive at Eq. (10.13). Figure 10.5 illustrates the implication of Eq. (10.13). This equation is referred to as ***Nyquist criterion for distortionless baseband transmission of digital data in the absence of noise***. We may summarize this by saying that the time function $p(t)$ that satisfies the conditions laid down in Eq. (10.11), or the frequency function $P(f)$ satisfying Eq. (10.13) eliminates ISI for samples taken at regular intervals of T. It may be noted that $P(f)$ is dependent upon $H_T(f)$, $H_C(f)$ and $H_R(f)$, as stipulated in Eq. (10.9).

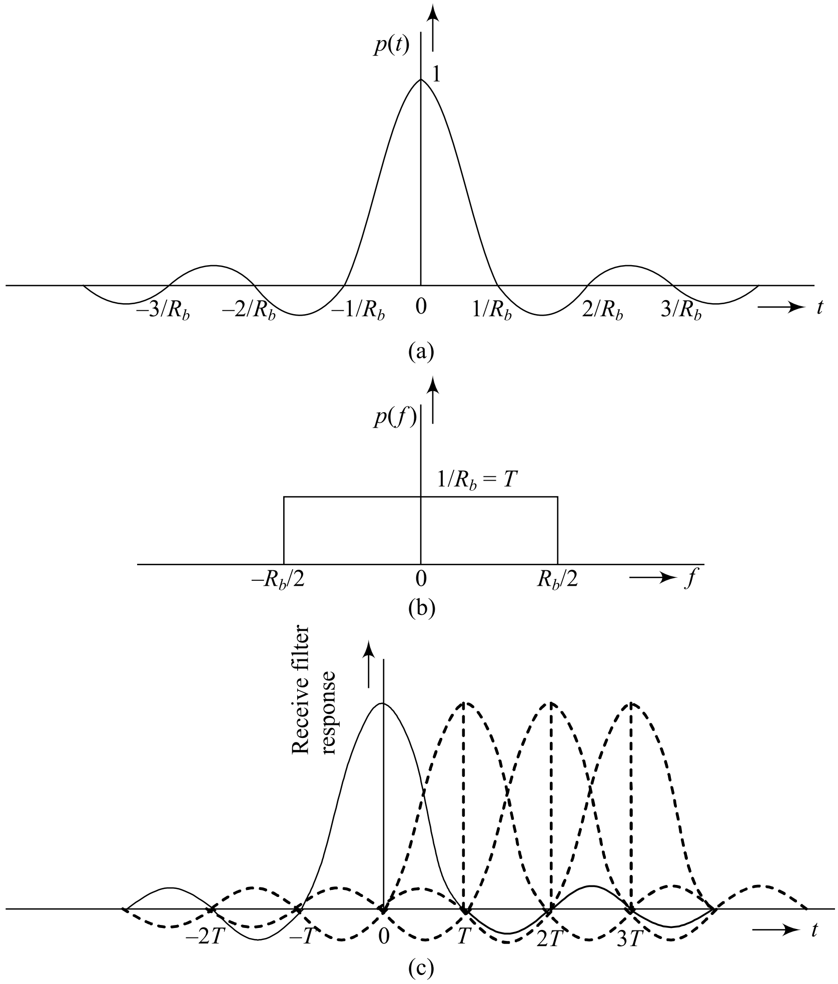


Fig. 10.5 *(a) A sinc pulse that satisfies Nyquist criterion. Note that p(0) = 1 and p(t) = 0 for t = ±T, ±2T, etc., (b) Spectrum of the pulse p(t). Note that it satisfies Eq. (10.13), (c) A sequence of sinc pulses at regular intervals of T sec (All assumed positive). Note that sampling at t = 0, T, 2T, etc., avoids ISI*

Ideal Nyquist channel The ideal Nyquist channel is the one in which the output pulses $p(t)$s from the receive filter have the shape of a sinc function. Correspondingly, these output pulses have a spectrum that has the shape of the magnitude response of an ideal LPF, as shown in Fig. 10.5(b). As can be seen from Fig. 10.5(a), (b) and (c), these pulses satisfy the conditions for zero ISI. Analytically, we may describe $p(t)$ and $P(f)$ of an ideal Nyquist channel as follows:

$$P(f) = \begin{cases} \dfrac{1}{2W}; & -W < fW \\ 0; & |f| > W \end{cases} \tag{10.14}$$

i.e.,

$$P(f) = \frac{1}{2W}\Pi(f/2W) \tag{10.15}$$

where the overall system bandwidth W is defined by

$$W = \frac{R_b}{2} = \frac{1}{2T} \tag{10.16}$$

Also,

$$p(t) = \text{sinc}(2Wt) = \frac{(\sin 2\pi Wt)}{(2\pi Wt)} \tag{10.17}$$

As shown in Fig.10.5(c), the sinc pulses are coming out from the receive filter at the rate of one pulse for every T sec, i.e., the bit rate is $R = 1/T = 2W$. But, the bandwidth needed for the channel is simply the bandwidth occupied by the sinc pulse, i.e., the bandwidth of $P(f)$, which is equal to $R_b/2 = W$. ***Hence, in the ideal Nyquist channel, it is (theoretically) possible to transmit at a bit rate of R = 2W even with a channel of bandwidth W. This bit rate of R_b = 2W is called the Nyquist rate and W, the minimum bandwidth required for this rate without ISI is called the Nyquist bandwidth. Also, the channel itself is called the Nyquist channel***.

Disadvantages Even though it appears that it is possible to transmit without ISI at a rate that is twice the minimum bandwidth by using the Nyquist channel, there are several difficulties involved. These are:

1. The magnitude response of the Nyquist channel, which is equal to $|P(f)|$, is as shown in Fig. 10.5(b), which, we know, is not physically realizable because of the sudden transitions.
2. As shown in Fig.10.5(c), ISI can be avoided in this channel only if the output sinc pulses from the receive filter are sampled ***exactly*** at $t = 0, T, 2T$, etc., and even a slight sampling jitter will introduce the ISI.
3. The rate of decrease of the sinc pulse $p(t)$ is only $1/|t|$ even at large values of t. So the side-lobe amplitudes do not decrease fast. This results in considerable ISI even for small errors in sampling time.

Raised cosine spectrum There are shapes of the received pulses, other than the sinc pulse used in the Nyquist channel, which can be practically realized and which also offer the possibility of zero ISI. Let $p(t)$ be the pulse and $P(f)$ its spectrum. Consider

$$P(f) = \begin{cases} \dfrac{1}{2W}; & |f| < f_1 \\ \dfrac{1}{4W}\left\{1 + \cos\left[\dfrac{\pi(|f| - f_1)}{2W - 2f_1}\right]\right\}; & f_1 < |f| < 2W - f_1 \\ 0; & |f| > 2W - f_1 \end{cases} \tag{10.18}$$

A plot of $P(f)$ is given in Fig.10.6(a).

Note

(i) From Eq. (10.18), we find that f_1 is the maximum frequency up to which the spectrum $P(f)$ has a value $(1/2W)$, i.e., up to which $2W\,P(f)$ is equal to 1.

(ii) We define a roll-off factor ρ as

$$\rho = \left[1 - \frac{f_1}{W}\right] \tag{10.19}$$

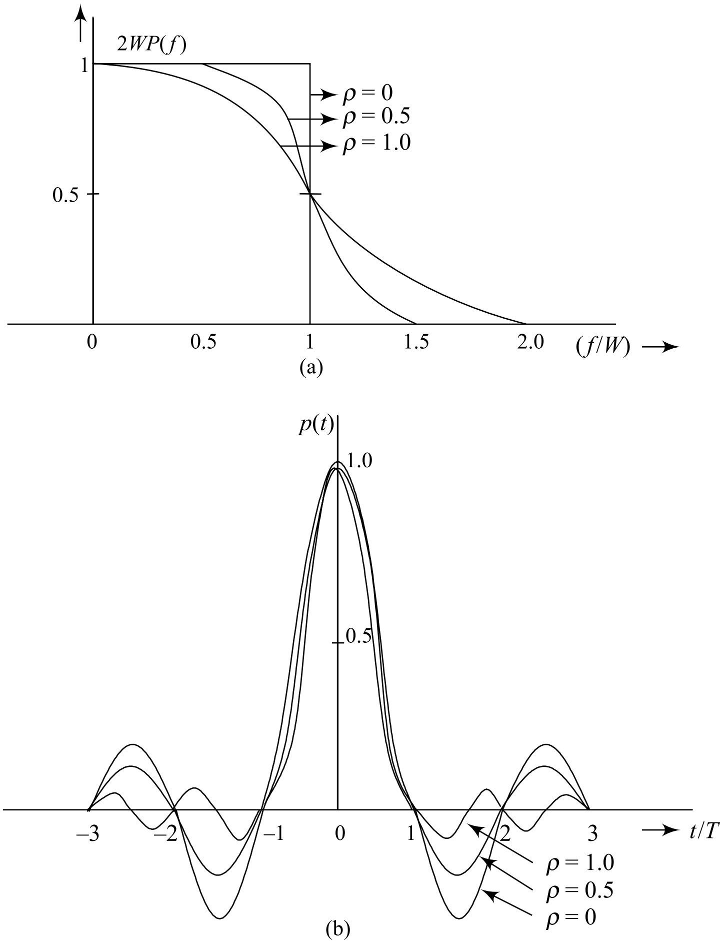
Roll-off factor is, in general, defined as the ratio of the excess bandwidth to the theoretical minimum bandwidth. It can be seen that the way ρ has been defined in Eq. (10.19) is in conformity with this.

(iii) From Fig. 10.6(a), it is clear that the transmission bandwidth B_T is given by

$$B_T = 2W - f_1 \tag{10.20}$$

(iv) f_1 depends on the roll-off factor and

$$\left.\begin{aligned} f_1 &= W \text{ for } \rho = 0 \\ \text{and} \quad f_1 &= 0 \text{ for } \rho = 1 \end{aligned}\right\} \tag{10.21}$$



(a)

(b)

Fig. 10.6 *(a) Spectrum P(f) of the pulse, (b) Pulse p(t) for different roll-off factors*

The inverse Fourier transform of $P(f)$ given in Eq. (10.18) is

$$p(t) = \text{sinc}(2Wt)\left[\frac{\cos 2\pi\rho Wt}{1 - 16\rho^2 W^2 t^2}\right] \tag{10.22}$$

1. For $\rho = 0$, $p(t) = \text{sinc}(2Wt)$, the pulse that we had for the Nyquist channel.
2. For $\rho = 1$, $p(t)$ simplifies to: $\quad p(t) = \dfrac{\text{sinc}(4Wt)}{1 - 16W^2 t^2}$ (10.23)

Remark

(i) The pulse function p(t) given in Eq. (10.22) has two distinct factors; the first factor sinc 2Wt, associated with ideal Nyquist channel, and the second factor is shown inside the rectangular brackets, which decreases as $1/|t|^2$ for large values of $|t|$. While the first one, being a sinc function, ensures that there are zero-crossings at regular intervals, so that they can be used for sampling (to avoid ISI), the second factor ensures that the tails of the pulse decay faster than what we had for the pure sinc pulse used in the ideal Nyquist channel.

(ii) *Whereas the ideal Nyquist filter is not realizable at all, this filter called the '**raised cosine filter**', with a value of $\rho = 1$, can easily be approximated by giving an adequate amount of time-delay.*

(iii) *Since the tails of p(t) reduce very fast when $\rho = 1$, timing jitter of the sampler in the receiver does not produce much of inter symbol interference (ISI).*

(iv) *However, to avoid ISI with this raised cosine channel ($\rho = 1$), the minimum transmission bandwidth required is 2W, i.e., twice that required for the ideal Nyquist channel, for the same pulse rate.*

(v) (a) *the pulse p(t) corresponding to $\rho = 1$, i.e., p(t) for the raised cosine filter takes a value of 0.5 at $t = \pm T/2 = \pm 1/4\,W$. This means that the pulse width, measured at half the maximum amplitude (of 1), is exactly equal to the time-slot duration T.*

(b) *For the pulse p(t) of the raised cosine filter with $\rho = 1$, there are two zero-crossing points at $t = \pm 1.5T$ and $\pm 2.5T$ in addition to the usual zero-crossings at $t = \pm 1, \pm 2, \pm 3, \ldots$*

Note *Properties 5(a) and 5(b) are very useful since they enable the receiver to extract the timing signals required for synchronization.*

Example 10.1 Binary data is transmitted at the rate of 56 kbps using a baseband binary PAM system designed to have a raised cosine spectrum. What is the transmission bandwidth required if the roll-off factor $\rho = 0.25, 0.5$?

Solution $$\rho = \left[1 - \frac{f_1}{W}\right] \qquad \text{(Refer to Eq. (10.19))}$$

$$W = \frac{R_b}{2} = \frac{56\text{ kbps}}{2} = 28\text{ kbps} \qquad \text{(Refer to Eq. (10.16))}$$

(a) when $\rho = 0.25$

$$\rho = \left[1 - \frac{f_1}{W}\right] = 0.25 = \left[1 - \frac{f_1}{28 \times 10^3}\right] \quad \therefore f_1 = 21 \times 10^3\text{ Hz}$$

From Eq. (10.20), we have

$$B_T = 2W - f_1 = (56 \times 10^3 - 21 \times 10^3) = 35 \times 10^3\text{ Hz}$$

(b) When $\rho = 0.\ 5$

$$\rho = \left[1 - \frac{f_1}{W}\right] = 0.5 = \left[1 - \frac{f_1}{28 \times 10^3}\right] \quad \therefore f_1 = 14 \times 10^3\text{ Hz}$$

But, $B_T = \text{Transmission bandwidth} = 2W - f_1 = (56 \times 10^3 - 14 \times 10^3)\text{Hz}$

$\therefore$ $B_T = 42\text{ kHz}$

10.3.1 Correlative Coding

Till now, we have been looking upon inter-symbol interference (ISI) as nuisance, or something which is undesirable and is to be avoided as it produces decoding errors and thereby degrades the system performance. We have seen that it is possible to transmit binary data at a speed of 2*W* even on a channel of bandwidth *W*, avoiding ISI. In this section, we will discuss some schemes like the correlative level or partial response signaling schemes, in which ISI is deliberately introduced in a controlled manner in order to achieve a data speed of 2*W* on a channel of bandwidth *W*. The underlying principle behind all these schemes is that since ISI is introduced in a controlled manner and is known, the receiver can take care of it.

Duo-binary signaling This is a specific case that belongs to the general class of correlative-level coding schemes. The general arrangement for duo-binary signaling scheme is shown in Fig. 10.7.

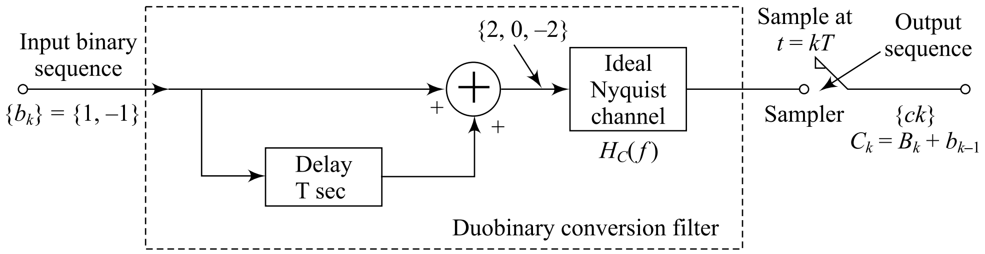


Fig. 10.7 *Duo-binary signaling scheme*

Since an ideal delay element producing a delay of T sec has a transfer function of $\exp[-j2\pi fT]$, the delay line filter has a transfer function of $[1 + \exp(-j2\pi fT)]$. This delay line filter is in cascade with the ideal Nyquist channel, whose transfer function is denoted by $H_C(f)$. Thus, the overall transfer function of the duo-binary conversion filter is given by

$$H(f) = H_C(f)[1 + \exp(-j2\pi fT)] \tag{10.24}$$

$$= H_C(f)[e^{j\pi fT} + e^{-j\pi fT}]e^{-j\pi fT}$$

$$\therefore \qquad H(f) = [2H_C(f)\cos \pi fT]e^{-j\pi fT} \tag{10.25}$$

We know that an ideal Nyquist channel is an ideal LPF with a bandwidth of W Hz and that

$$W = \frac{1}{2T} \qquad \text{(Refer to Eq. (10.16))}$$

Hence,

$$H_C(f) = \begin{cases} 1; & |f| \le \dfrac{1}{2T} \\ 0; & \text{otherwise} \end{cases}$$

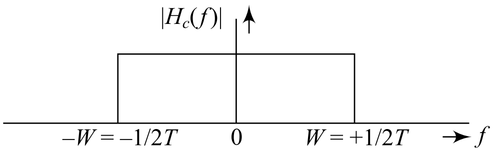


Fig. 10.8 *Transfer function of an ideal Nyquist channel*

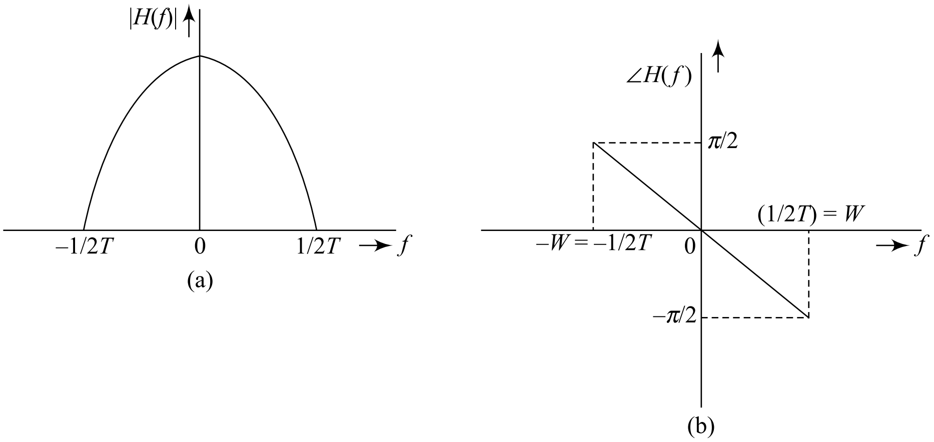
Thus, the overall frequency response is

$$H(f) \cdot H_C(f) = H(f) = \begin{cases} 2\cos \pi fT \exp(-j\pi fT); & |f| \le \dfrac{1}{2T} \\ 0; & \text{otherwise} \end{cases} \tag{10.26}$$

$$\therefore \qquad |H(f)| = \begin{cases} 2\cos \pi fT & ; \ |f| \le \dfrac{1}{2T} \\ 0 & ; \ \text{otherwise} \end{cases} \tag{10.27}$$

Making use of Eq. (10.26), the magnitude and phase responses of the duo-binary signaling scheme are drawn in Fig. 10.9.

This response of the duo-binary signaling scheme can easily be approximated as it is a smooth cosine pulse. This can also be seen from the impulse response of the duo-binary filter plotted in Fig. 10.10. The impulse response is not zero for negative values of time. But it can easily be approximated by truncating it for some negative value and then introducing enough time delay.



(a)

(b)

Fig. 10.9 *Transfer function of the duo-binary conversion filter as well as the overall frequency response: (a) Magnitude response, (b) Phase response*

Impulse response of the duo-binary filter From Eq. (10.24), by taking its inverse Fourier transform on both sides, we get the impulse response $h(t)$ of the duo-binary conversion filter as

$$\begin{aligned} h(t) &= \frac{\sin(\pi t/T)}{(\pi t/T)} + \frac{\sin[\pi(t-T)/T]}{[\pi(t-T)/T]} \\ &= \frac{\sin(\pi t/T)}{(\pi t/T)} + \frac{\sin(\pi t/T)}{\pi(t-T)/T} \\ &= \frac{T^2 \sin(\pi t/T)}{\pi t(T-t)} \end{aligned} \tag{10.28}$$

This impulse response is shown in Fig. 10.10.

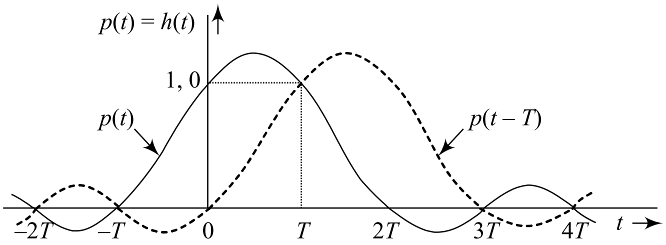


Fig. 10.10 *Impulse response of the duo-binary conversion filter*

If the binary symbols 1 and 0 are represented at the input to the system shown in Fig.10.7 by a narrow positive pulse of amplitude 1 for $b_k = 1$ and a narrow negative pulse of amplitude one for $b_k = -1$, the impulse response $h(t)$ shown in Fig. 10.10 actually will be the pulse $p(t)$ received at the output of the channel.

If we now consider $p(nT)$, the pulse received during the n^{th} time slot, then referring to Fig. 10.10,

$$p(nT) = \begin{cases} 1 \text{ for } n = 0 \text{ and } 1 \\ 0 \text{ for any other value of } n \end{cases} \tag{10.29}$$

Since we are using polar signaling, corresponding to $b_k = 1$, a pulse $p(t)$ and corresponding to $b_k = -1$, a pulse $-p(t)$, will be received at the input to the sampler. The sampler samples the received pulses at $t = nT$. As may be seen from Fig. 10.10, the samples will all be zero except the ones taken at $t = 0$ (i.e., $n = 0$) and $t = T$ (i.e., $n = 1$). Thus pulse $p(t)$ will produce ISI only with the next pulse marked as $p(t - T)$ in the figure and not with any other pulse. Although both $p(t)$ and $p(t - T)$ have been shown for convenience to be positive, it is possible that they may have opposite polarities. If they are both of the same polarity, the sample taken at $t = T$ will be +2 or –2. But if these adjacent pulses are of opposite sign, the sample at $t = T$ will be zero. Thus, even though the input pulses are having only two levels, i.e., +1 and –1, the samples of the received pulses can have any of the three possible values, viz., +2, –2 and 0. Hence, the following decision strategy may be adopted.

1. If the sample taken in the present time-slot is positive and the previous bit was identified as a 1, then in this time-slot is also 1.
2. If the sample taken in the present time-slot is negative and the previous bit was identified as a –1(i.e., $b_k = 0$), then the present bit in this time-slot is also a – 1.
3. If the sample value in the present time-slot is zero, then the present bit is the complement of the previous bit.

This scheme of detection is illustrated in the following table.

Table 10.1 *Duo-binary scheme of detection*

Transmitted sequence $\{b_k\}$. Starting bit = 1	1	1	0	0	0	1	1	0	1	1	1	0	1
Samples of the received signal $\{c_k\}$	2	2	0	–2	–2	0	2	0	0	2	2	0	0
Detected binary sequence $\left\{\hat{b}_k\right\}$	1	1	0	0	0	1	1	0	1	1	1	0	1

From the above discussion we find that the pulses are transmitted at a rate of $(1/T)$. Also, from Fig. 10.9(a), we find that the bandwidth of the duo-binary filter is $(1/2T)$. Hence, we are able to achieve a transmission rate that is twice the bandwidth even while avoiding the effects of ISI.

Decision feedback and error propagation It is clear from the above detection strategy that decision on the bit during the present-time slot is based on the knowledge of the previous detected bit which has been stored. This technique is referred to as 'decision feedback'. A serious drawback with decision feedback is that if an error is made in the decision making in any time slot, the error tends to propagate. A technique used for avoiding the error propagation is called 'pre-coding'. We shall now see how pre-coding is done and how it avoids error propagation.

Pre-coding for avoiding error propagation A pre-coded duo-binary scheme is shown in Fig. 10.11. The input binary sequence $\{b_k\}$ is now given not directly to the duo-binary encoder, but through a '*pre-coder*' and a level shifter. The pre-coder produces a new binary sequence $\{d_k\}$ through the following operation:

$$d_k = b_k \oplus d_{k-1} \tag{10.30}$$

where ⊕ denotes modulo-2 addition of the binary digits. Since modulo-2 addition is nothing but 'Exclusive-OR' operation,

$$d_k = \begin{cases} 1 & \text{if either } b_k \text{ or } d_{k-1} \text{ is a 1, but not both} \\ 0 & \text{otherwise} \end{cases} \tag{10.31}$$

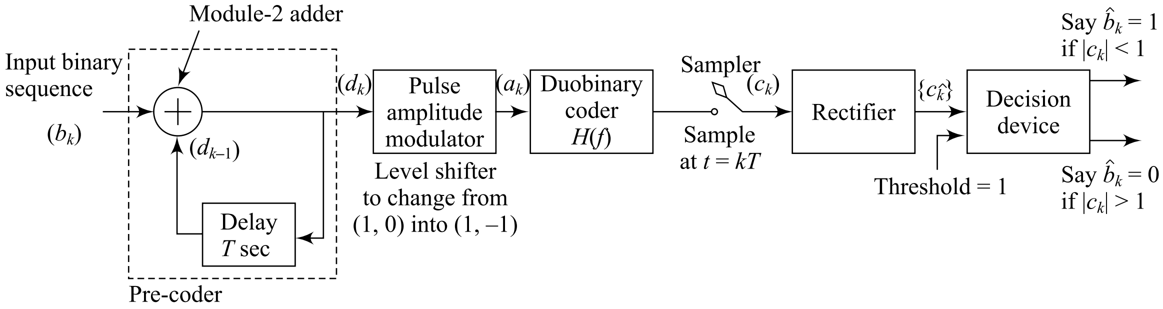


Fig. 10.11 *A pre-coded duo-binary scheme*

The pulse amplitude modulator, or level-shifter converts this sequence $\{d_k\}$ of 1s and 0s into a new sequence $\{a_k\}$, where, $a_k = +1$ or -1. This sequence is now fed to the duo-binary encoder. So, referring to Fig. 10.7, a_ks now play the role of b_ks shown in the figure. At the output of the sampler, therefore, we get a sequence $\{c_k\}$, where

$$c_k = a_k + a_{k-1} \tag{10.32}$$

Hence, from Eqs. (10.31) and (10.32), we have

$$c_k = \begin{cases} 0 & \text{if } b_k = 1 \\ \pm 2 & \text{if } b_k = 0 \end{cases} \tag{10.33}$$

From Eq. (10.33), it is clear that

$$\left.\begin{aligned} &\hat{b}_k = 1 \text{ if } |c_k| < 1 \\ \text{and} \qquad &\hat{b}_k = 0 \text{ if } |c_k| > 1 \end{aligned}\right\} \tag{10.34}$$

Hence, the decision device is given a reference, or threshold voltage of 1 volt as shown in Fig. 10.11. The probability of $|c_k|$ being exactly equal to 1 is zero and so we have ignored that possibility.

The scheme of detection when a pre-coder is used is illustrated in Table 10.2.

Table 10.2 *Scheme of detection with a pre-coder*

Binary Sequence $\{b_k\}$	1	1	0	0	0	1	1	0	1	1	1	0	1
The sequence $\{d_k\}$ start-up bit =1	0	1	1	1	1	0	1	1	0	1	0	0	1
The sequence $\{a_k\}$ start-up bit =1	–1	1	1	1	1	–1	1	1	–1	1	–1	–1	1
The sequence $\{c_k\}$	0	0	2	2	2	0	0	2	0	0	0	–2	0
Detected binary sequence $\hat{b}_k = 1$ if $\lvert c_k \rvert < 1$; $\hat{b}_k = 0$ if $\lvert c_k \rvert > 1$	1	1	0	0	0	1	1	0	1	1	1	0	1

10.3.2 Modified Duo-Binary Signaling

The magnitude response $|H(f)|$ shown in Fig. 10.9(a) clearly indicates that the PSD of the transmitted pulse $p(t)$ is not zero at $f = 0$ since $|H(f)|$ is non-zero at $f = 0$. This is not desirable in certain applications that make use of communication channels which cannot carry a dc component. This deficiency of the duo-binary signaling is overcome by incorporating some modifications in it which make $|H(f)| = 0$ at $f = 0$. This modified duo-binary signaling scheme is shown in Fig. 10.12.

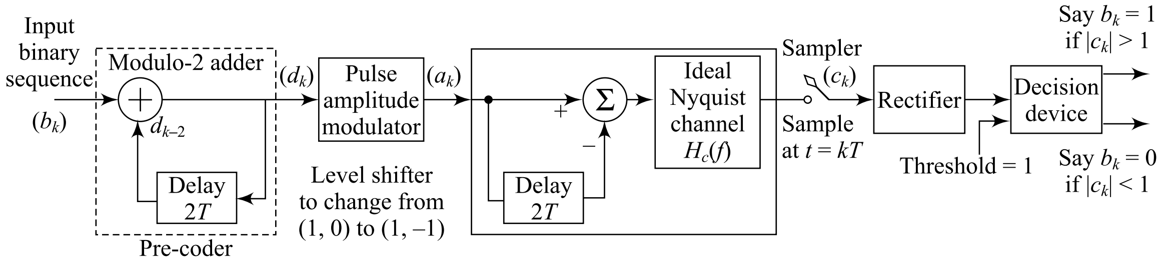


Fig. 10.12 *Modified duo-binary signaling scheme*

As may be seen from Fig. 10.12, we now have the following relations:

$$d_k = b_k \oplus d_{k-2}$$

$$d_k = \begin{cases} 1 & \text{if either } b_k \text{ or } d_{k-2} \text{ is a 1 but not both} \\ 0 & \text{otherwise} \end{cases}$$

$$a_k = 1 \text{ or } -1$$

$$c_k = a_k - a_{k-2}$$

$$c_k = +2, 0, \text{ or } -2$$

The transfer function $H_M(f)$ of the modified duo-binary filter may be written down in the same way as Eq. (10.24), except for the difference that in the present case, the delay is $2T$ and not T. Hence,

$$\begin{aligned} H_M(f) &= H_C(f)[1-\exp(-j4\pi fT)] \\ &= 2jH_C(f)\sin 2\pi fT \exp(-j2\pi fT) \end{aligned} \tag{10.35}$$

where $H_C(f)$ is the transfer function of the ideal Nyquist filter. Since

$$H_c(f) = 1\cdot\Pi(f/2W) \tag{10.36}$$

as shown in Fig. 10.8 and $W = 1/2T$, it follows that the overall transfer function of the modified duo-binary filter as

$$H_M(f) = \begin{cases} 2j\sin 2\pi fT\exp(-j2\pi fT) & \text{for } |f| \le (1/2T) \\ 0 & \text{otherwise} \end{cases} \tag{10.37}$$

From Eq. (10.35), since $H_M(f)$ is

$$H_M(f) = H_C(f)[1-\exp(-j4\pi fT)]$$

and since $H_C(f)$ is a rectangular function as shown in Fig. 10.8, when we take the inverse Fourier transform of $H_M(f)$, we get two sinc pulses that are displaced in time by $2T$ sec. Thus, the impulse response $h_M(t)$ of this modified duo-binary filter is

$$\begin{aligned} h_M(t) &= \frac{\sin(\pi t/T)}{(\pi t/T)} - \frac{\sin[\pi(t-2T)/T]}{[\pi(t-2T)/T]} \\ &= \frac{\sin(\pi t/T)}{(\pi t/T)} - \frac{\sin(\pi t/T)}{\pi(t-2T)/T} \\ &= \frac{2T^2\sin(\pi t/T)}{\pi t(2T-t)} \end{aligned} \tag{10.38}$$

From Eq. (10.37), the magnitude response and the phase response of the modified duo-binary filter are plotted in Figs. 10.13(a) and (b), respectively.

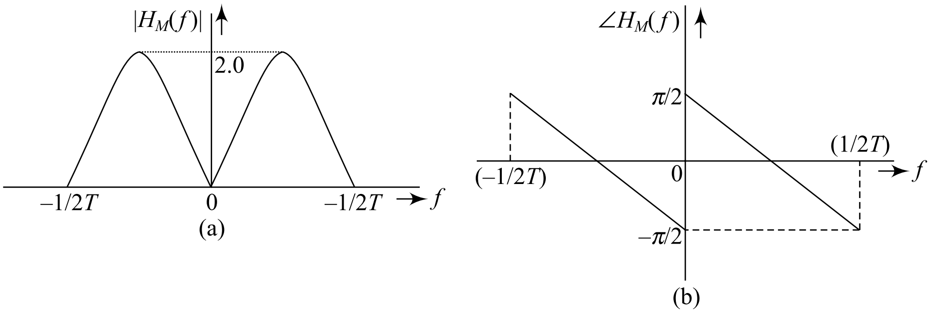


Fig. 10.13 *Transfer function of the modified duo-binary filter: (a) Magnitude response, (b) Phase response*

As can be seen from Fig. 10.13(a), the magnitude of $H_M(f)$ is zero at $f = 0$ indicating that the modified duo-binary coder has no dc component at its output – a feature we desired to have.

The denominator of the expression for $h_M(t)$, given in Eq. (10.38) clearly shows that the impulse response, and hence the pulse $p(t)$ has tails that decay as $1/|t|^2$. Further, the plot of $h_M(t)$, given in Fig. 10.14 shows that there are three distinct levels at the sampling instants, these levels being 1, 0 and –1.

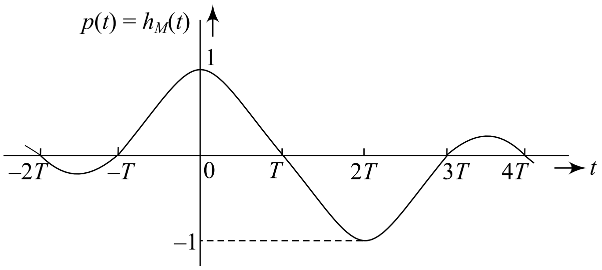


Fig. 10.14 *Impulse response function of the modified duo-binary conversion filter*

Detection strategy From Fig. 10.12, we have

$$d_k = b_k \oplus d_{k-2} \tag{10.39}$$

$$= \begin{cases} \text{binary 1} & \text{if either } b_k \text{ or } d_{k-2} \text{ (but not both) is a 1} \\ \text{binary 0} & \text{otherwise} \end{cases}$$

Since the level-shifter converts the unipolar {1, 0} sequence $\{d_k\}$ into a polar {1, –1} sequence $\{a_k\}$, the output digit c_k can take any of the three values –2, 0, or 2. This can easily be seen by following b_k, d_k and a_k.

Case 1: Let $b_k = 1$ and d_{k-2} be 0. Then d_k is 1. Hence $a_k = 1$. Since d_{k-2} has been assumed to be zero, the corresponding a_{k-2} would have been –1. Hence c_k will be +2.

Case 2: Let $b_k = 1$ and d_{k-2} be 1. Then d_k is 0. Hence $a_k = -1$. Since d_{k-2} has been assumed to be 1, the corresponding $a_{k-2} = 1 \therefore c_k = -2$ in this case.

Case 3: Let $b_k = 0$ and d_{k-2} be 0. Then $d_k = 0$, and $a_k = -1$. Since d_{k-2} has been assumed to be 0, the corresponding $a_{k-2} = -1$. Hence $c_k = 0$ in this case.

Case 4: Let $b_k = 0$ and d_{k-2} be 1. Then $d_k = 1$. $\therefore a_k = 1$. Since d_{k-2} has been assumed to be 1, correspondingly $a_{k-2} = 1 \therefore c_k = 0$.

Thus, the detection strategy adopted is

1. If $|c_k| > 1,$ ask the receiver to say $b_k = 1$
2. If $|c_k| < 1,$ ask the receiver to say $b_k = 0$ (10.40)

Note

(i) The magnitude response of the modified duo-binary filter in Fig. 10.13(a) clearly shows that the bandwidth is $W = 1/2T = R_b/2$. Hence, signaling at a rate that is twice the bandwidth is possible.

(ii) There is no dc component in the communication channel.

Generalized correlative coding (Partial-response signaling) Referring to Fig. 10.7, we find that the duo-binary conversion filter is given a two-level input sequence $\{1, -1\}$ and that it correlates the input signal with a delayed signal, delayed by T sec. The modified duo-binary conversion filter of Fig. 10.12 too has a two-level input sequence $\{1, -1\}$ but it correlates the input signal with a delayed signal, delayed by $2T$ sec. Both of them enable binary baseband data transmission at a rate that is twice the channel bandwidth. But, the modified duo-binary conversion filter has better frequency response (no dc component)

This idea used in the modified duo-binary conversion filter can easily be generalized so as to have a filter in which the input signal is correlated by a number of delayed signals delayed by 0 sec, T sec, $2T$ sec … , $(N - 1)T$ sec, with appropriate weights given to each delayed signal. Signaling using such filters is referred to as 'correlative-level coding', or 'partial response signaling'. An arrangement for such a generalized correlative coding scheme, is shown in Fig. 10.15.

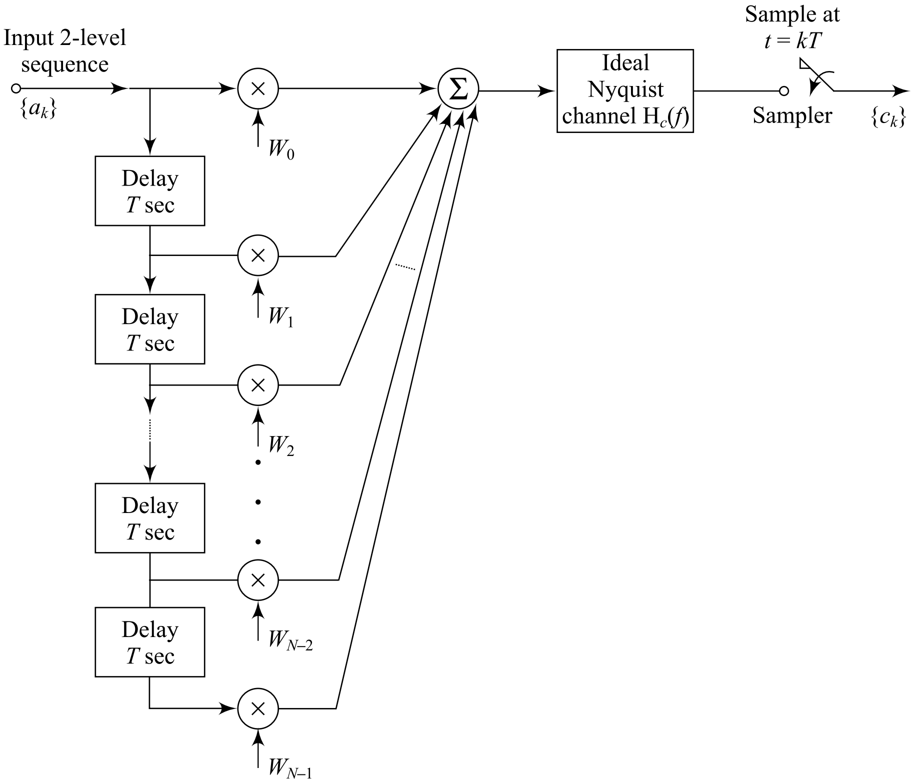


Fig. 10.15 *A generalized correlative coding scheme*

By using different tap weights w_i, $i = 0, 1, \ldots, (N-1)$, it is possible to create a variety of spectral shapes for the overall channel. As in the case of the duo-binary conversion filter and the modified duo-binary conversion filter, with the generalized correlative coding scheme too data rates close to the Nyquist rate can be achieved while avoiding ISI.

In the presence of channel noise, however, P_e may be a little higher.

Example 10.2 A binary PAM wave with bit duration of 10 µs is to be transmitted over a channel with a maximum bandwidth of 75 kHz. Determine a suitable raised cosine spectrum for this purpose.

Solution From Eq. (10.18), we know that a raised cosine spectrum is completely known once the two parameters W and f_1 are known.

Since T is given to be 10 µs = 10^{-5} sec, the transmission rate

$$R_b = \frac{1}{T} = 10^5 \text{ bits/sec} = 100 \text{ kbps}$$

$\therefore$ W = Minimum bandwidth required when ideal Nyquist filter is used, is given by

$$W = \frac{R_b}{2} = \frac{100 \times 10^3}{2} = 50 \text{ kHz}$$

B_T = Transmission bandwidth of the channel is given to be 75 kHz.

But, when a raised cosine filter is used,

$$B_T = 2W - f_1 \qquad \therefore 75 \times 10^3 = 2 \times 50 \times 10^3 - f_1$$

$$\therefore \qquad f_1 = 100 \times 10^3 - 75 \times 10^3 = 25 \text{ kHz}$$

$$\therefore \qquad \text{Roll-off factor } \rho = 1 - \frac{f_1}{W} = 1 - \frac{25 \times 10^3}{50 \times 10^3} = 0.5$$

Determining the raised cosine filter means specifying transfer function. This can be obtained by substituting the value of W and f_1 in Eq. (10.18).

Example 10.3 The binary data stream 001 101 101 is applied to the input of a duo-binary system. If no pre-coder is used, determine the duo-binary coder output and the resulting receiver output.

Solution We know that in a duo-binary system, the coder output in any particular time slot will be +2 if the pulses in the present-time slot and the previous-time slot are both positive; –2 if both are negative and 0 if they are of opposite polarity. Since the previous pulse is needed, we take a starting bit $b_k = 1$.

$\{b_k\}$ starting bit =1	0	0	1	1	0	1	1	0	1
Samples of the received signal (duo-binary coder output)	0	–2	0	2	0	0	2	0	0
Received output $\{\hat{b}_k\}$	0	0	1	1	0	1	1	0	1

The receiver output sequence $\{\hat{b}_k\}$ is estimated using the detection strategy outlined just before Table 10.1.

Example 10.4 Repeat Example 10.3 when a pre-coder is used.

Solution

Binary Sequence $\{b_k\}$	0	0	1	1	0	1	1	0	1
Sequence $\{d_k\}$ start-up bit =1	1	1	0	1	1	0	1	1	0
Sequence $\{a_k\}$; start-up bit = 1 a_k = 1 if d_k = 1; a_k = –1 if d_k = 0	1	1	–1	1	1	–1	1	1	–1
Sequence $\{c_k\}$	2	2	0	0	2	0	0	2	0
Detected binary sequence $\hat{b}_k = 1$ if $\lvert c_k \rvert < 1; \hat{b}_k = 0$ if $\lvert c_k \rvert > 1$	0	0	1	1	0	1	1	0	1

Note

$$d_k = \begin{cases} 1 & \text{if either } b_k \text{ or } d_{k-1} \text{ is a 1, but not both} \\ 0 & \text{otherwise} \end{cases}$$

$$c_k = a_k + a_{k-1}$$

10.4 M-ARY BASEBAND SIGNALING

Till now we have discussed binary baseband transmission, in which a symbol could take one of two possible values – like A and 0 (unipolar) or A and $-A$ (polar). We have been representing the symbol duration by T sec. Since a binary choice was involved during each T sec period, with the two values A and 0 or A and $-A$ equi-probable, each symbol carried 1 bit of information and the information rate could be expressed as $(1/T) = R$ bits/sec.

In M-ary baseband transmission, on the other hand, a symbol can take M possible values. If we assume that all the M values are equally probable, during each symbol duration of say, τ sec, an information of $\log_2 M$ bits is transmitted and therefore the information rate is

$$R_M = \frac{1}{\tau} \log_2 M \text{ bits/sec} \tag{10.41}$$

A simple example of M-ary baseband signaling is provided by the 'quarternary coding' of binary data. We will use this to illustrate M-ary baseband signaling. Let us say we would like to transmit the following binary data of eight binary digits in a total time of 8 milli-seconds.

$$\{b_k\} = 0\,0\,0\,1\,1\,0\,1\,1$$

In quartenary coding, we take two successive binary digits, group them and give one level to it. Let us say we encode the four possible groups as follows:

$$\text{symbols} \left\{ \begin{array}{cc} \text{Group} & \text{Level} \\ 01 & -3 \\ 00 & -1 \\ 10 & 1 \\ 11 & 3 \end{array} \right.$$

Then the given binary digit sequence may be represented by the following waveform, using this quarternary coding:

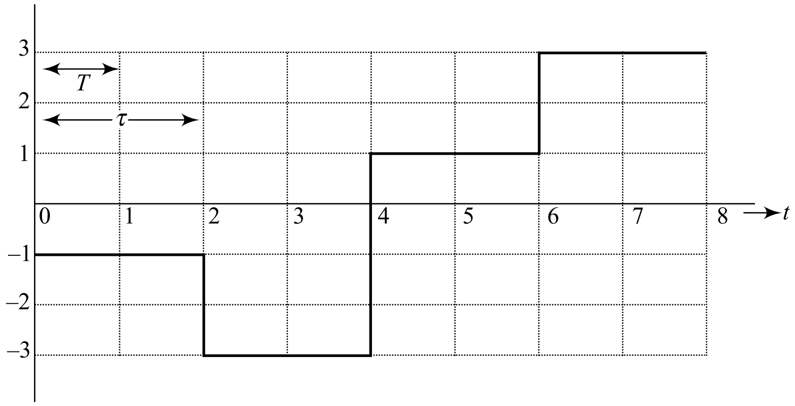


Fig. 10.16 *Waveform of the quarternary coded binary sequence*

Here, two binary symbols are grouped together and since there are four such distinct groups possible, the number of levels is also four; i.e., $M = 4$. In general, if n binary digits are grouped together, there will be 2^n such groups and so

$$\left.\begin{aligned} M &= 2^n \\ \text{or,} \quad n &= \log_2 M \end{aligned}\right\} \tag{10.42}$$

If T is the time duration of each binary digit, since a symbol consists of n binary digits, the symbol duration is

$$\tau = nT \tag{10.43}$$

In Fig. 10.16, since $M = 4$, $n = 2$ and $\tau = 2T$. Since one binary digit is transmitted in T sec. in baseband binary transmission, the transmission rate is R_b (suffix b is added to indicate binary transmission) where,

$$R_b = \frac{1}{T} \text{ bits/sec} \tag{10.44}$$

For baseband M-ary transmission, each symbol has a duration of $\tau = nT$ sec. Hence, the rate of *symbol transmission* is

$$R_M = \frac{1}{nT} = \frac{R_b}{\log_2 M} \tag{10.45}$$

10.4.1 Baud and Bit Rate

'Baud' is a term used to indicate symbol rate. As such, it is not necessary to say '*baud rate*', as the idea of rate is already there in 'baud' itself.

Bit rate, on the other hand, indicates the rate of transmission of information '*bits*'. If we are talking of transmission of binary symbols, 'baud' and bit rate will be the same, as each binary symbol carries one bit of information. But, if we are talking of transmission of some other symbols, since each such symbol may carry several bits of information, the symbol rate, baud and the bit rate, will be different. For example, in the transmission of quarternary coded symbols which we considered, the symbol rate is

$$R_M = \frac{1}{nT} = \frac{R_b}{\log_2 M} \text{ baud} \tag{10.46}$$

But since each of these symbols carries $\log_2 M$ bits of information, the bit rate is

$$R_M \cdot \log_2 M = R_b \text{ bits/sec.} \tag{10.47}$$

Since the binary symbol carries the least *possible information, viz., 1 bit per symbol, in general, for any M-ary baseband transmission, (*$M > 2$), the symbol rate will be less than the bit-rate.

10.4.2 Bandwidth for M-ary Baseband Transmission

For a given *bit-rate* of R_b bits/sec, using binary baseband transmission, we know that the absolute minimum transmission bandwidth is the Nyquist bandwidth which is given by

$$W_b = \frac{R_b}{2} \text{ Hz} \tag{10.48}$$

The same *bit-rate* can, however, be achieved at a transmission bandwidth that is lower than the ideal Nyquist bandwidth if we use M-ary baseband signaling. This is because the bandwidth required depends only on the pulse rate (the shape of the received pulse having been fixed as a sinc function) and in the case of M-ary baseband signaling, the pulses are wider by a factor of $\log_2 M$ compared to the binary case, making the *pulse rate* in M-ary baseband signaling smaller by that factor. However, the *bit-rate* will be the same since each M-ary pulse carries $\log_2 M$ bits whereas each binary pulse carries only one bit. Thus, the absolute minimum transmission bandwidth required for achieving a bit-rate of R_b bits/sec using M-ary baseband transmission, is

$$W_m = \frac{R_b}{2} \cdot \frac{1}{\log_2 M} = W_b / (\log_2 M) \tag{10.49}$$

It should be noted, however, that there is a price to be paid to achieve this while maintaining the same probability of error as in the binary case. This is because the probability of error primarily depends upon the spacing between the amplitude levels, irrespective of whether it is binary or M-ary signaling. When the same spacing (as in the binary case) between adjacent amplitude levels is maintained, the M-ary signal will have much larger power. Thus, binary signaling which has just two amplitude levels only, gives the best noise immunity for a given SNR. M-ary signaling on the other hand, needs more *SNR* but less bandwidth for the same bit-rate. ***M-ary signaling is therefore better suited for digital voice channels which have limited bandwidth but a large signal-to-noise ratio***.

Example 10.5 A computer is generating binary words, each consisting of 16 bits, at the rate of 15,00 words per second. (a) Find the bandwidth required to transmit its output as a binary PAM signal. (b) Find M so that the output could be transmitted as an M-ary signal on a channel whose bandwidth is limited to 30 kHz.

Solution Bit rate required = $16 \times 15000 = 240{,}000$ bits/sec = R_b

(a) If we use ideal Nyquist channel, the bandwidth required = $\frac{R_b}{2} = 120 \text{ kHz} = W$

On the other hand, if we use a channel with raised cosine spectrum and a roll-off factor $\rho = 1$, the transmission bandwidth required will be

$$B_T = 2W = 240 \text{ kHz}$$

(b) When we use M-ary signaling, as per Eq. (10.49),

$$W_M = \frac{R_b}{2} \cdot \frac{1}{\log_2 M} = \frac{120 \times 10^3}{2} \cdot \frac{1}{\log_2 M} = 30 \times 10^3$$

$$\therefore \qquad \log_2 M = \frac{120 \times 10^3}{2} \cdot \frac{1}{30 \times 10^3} = 2$$

$$\therefore \qquad M = 2^2 = 4$$

10.4.3 Equalization

An equalizer, as we already know, is a system which is designed to have a frequency response that is the inverse of the frequency response of the channel, so that when it is kept in cascade with the channel, the overall frequency response will be flat and the signal distortion caused by the channel is eliminated. Unlike in the case of analog communication, where signal waveform preservation is of paramount importance, in the case of digital communications, what is important is that during each time slot, the receiver should be able to decide correctly whether what was transmitted during that time-slot was a binary 1 or a binary 0.

Zero-forcing equalizer In this approach to the optimization of the receiver, the channel noise is totally ignored and the receiver is viewed as a 'zero-forcing equalizer' followed by the decision-making device. The equalizer, in this approach, is used to force the inter-symbol interference to be zero at all sampling instants $t = kT$ at which the channel output taken through the equalizer is sampled, except for $k = 0$ for which the distorted, but desired pulse occurs. A transversal (tapped delay) equalizer with $(2N + 1)$ taps is used and the samples of the equalizer output $p_{eq}(t)$ are taken at regular intervals of T. The tap gains, or the weights of the equalizer, c_i, $i = -N$ to $+N$, are so chosen that

$$p_{eq}(t_k) = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k = \pm 1, \pm 2, \pm 3, \ldots, \pm N \end{cases} \tag{10.50}$$

This means that we are forcing N-zero values to exist on each side of the peak of $p_{eq}(t)$, where $p_{eq}(t)$ is the output pulse from the equalizer and $p_{eq}(t_k)$ is the kth sample of $p_{eq}(t)$, taken at $t = kT$. It should be noted that it is required that $p_{eq}(t)$, the output pulse satisfies the Nyquist criterion or, the controlled ISI criterion as the case may be.

If, as shown in Fig. 10.17, $p_r(t)$ is the output pulse from the channel which is given as input to the zero-forcing equalizer, then, since the equalizer output pulse is the sum of all the delayed pulses from the outputs of the various delay elements of the equalizer, we have

$$p_{eq}(t) = \sum_{n=-N}^{N} c_n p_r(t - nT - NT) \tag{10.51}$$

The $-NT$ term in the argument of $p_r(\cdot)$ in Eq. (10.51), represents a constant time delay which we may ignore during the analysis and re-introduce at the end.

$$\therefore \quad p_{eq}(t) = \sum_{n=-N}^{N} c_n p_r(t - nT) \tag{10.52}$$

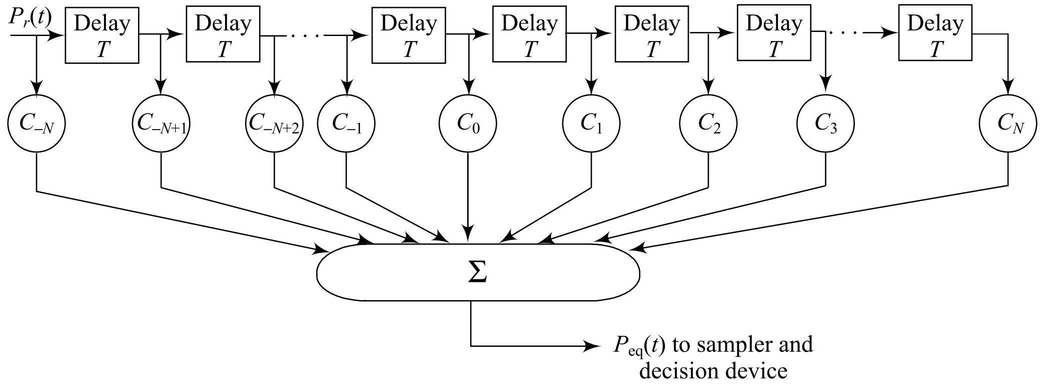


Fig. 10.17 *Zero-forcing equalizer*

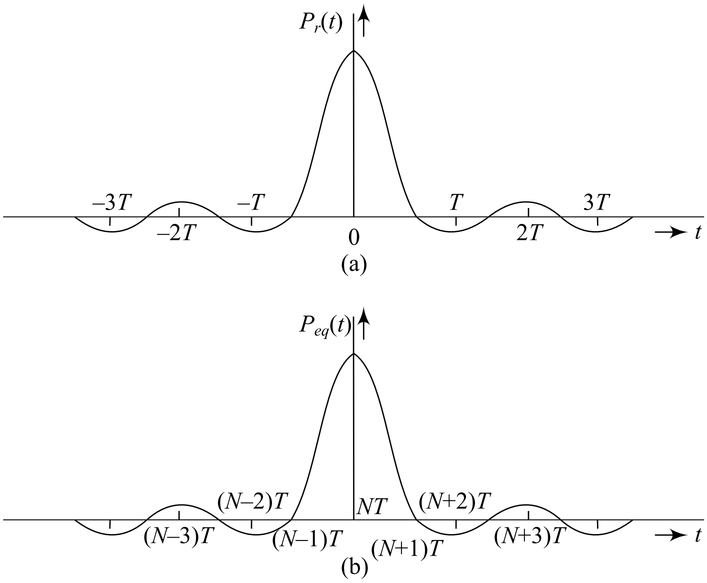


Fig. 10.18 *The input pulse $p_r(t)$ and the output pulse $p_{eq}(t)$*

Now, this $p_{eq}(t)$ is sampled at $t = kT$, $k = 0, \pm1, \pm2, \ldots$. Hence, at $t = kT$, we have

$$p_{eq}(kT) = \sum_{n=-N}^{N} c_n p_r(kT - nT); \quad k = 0, \pm1, \pm2, \ldots \tag{10.53}$$

For convenience in notation, let us drop T for the time being, so that

$$p_{eq}(k) = \sum_{n=-N}^{N} c_n p_r(k - n); \quad k = 0, \pm1, \pm2, \ldots \tag{10.54}$$

Equation (10.54) represents a set of an infinite number of simultaneous equations with only $(2N + 1)$ variables, viz., c_ns, $n = -N$ to N, which we have to determine with the constraint that

$$p_{eq}(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k = \pm1, \pm2, \pm3, \ldots, \pm N \end{cases} \tag{10.55}$$

It is not possible to solve this set of an infinite number of equations. However, since $p_{eq}(t)$ satisfies the Nyquist criterion, or the controlled ISI criterion, the pre-cursor and post-cursor amplitudes of $p_{eq}(t)$ decay rather rapidly. So we can as well modify the constraint (Eq. (10.55)) and limit the number of samples on either side of $k = 0$, to N values only. Thus, we are implicitly assuming that the ISI is limited only to N preceding and N succeeding values – an assumption that is quite justifiable if $N > 2$. Thus, we rewrite the constraint equation in Eq. (10.55) as

$$p_{eq}(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k = \pm1, \pm2, \ldots, \pm N \end{cases} \tag{10.56}$$

Once this modified constraint condition is imposed on the samples of $p_{eq}(t)$, as shown in Eq. (10.54), we get only $(2N + 1)$ simultaneous linear equations involving $(2N + 1)$ variables, viz., the c_ns. We can write these $(2N + 1)$ equations in matrix form as follows.

$$\begin{bmatrix} p_r(0) & p_r(-1) & \cdot & \cdot & \cdot & p_r(-2N) \\ p_r(1) & p_r(0) & \cdot & \cdot & \cdot & p_r(-2N+1) \\ \cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots & \cdots & \cdots\cdots \\ p_r(N-1) & p_r(N-2) & \cdot & \cdot & \cdot & p_r(-N-1) \\ p_r(N) & p_r(N-1) & \cdot & \cdot & \cdot & p_r(-N) \\ p_r(N+1) & p_r(N) & \cdot & \cdot & \cdot & p_r(-N+1) \\ \cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots & \cdots & \cdots\cdots \\ p_r(2N-1) & p_r(2N-2) & \cdot & \cdot & \cdot & p_r(1) \\ p_r(2N) & p_r(2N-1) & \cdot & \cdot & \cdot & p_r(0) \end{bmatrix} \begin{bmatrix} c_{-N} \\ c_{-N+1} \\ \cdots \\ c_{-1} \\ c_0 \\ c_1 \\ \cdots \\ c_{N-1} \\ c_N \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \\ 0 \\ \cdots \\ 0 \\ 0 \end{bmatrix} \tag{10.57}$$

The zero-forcing equalizer's tap gains, or weights are obtained by solving this matrix equation for c_ns.

Remarks

(i) The constant time delay, NT, in Eq. (10.51) which was ignored in the rest of the analysis has been re-introduced for drawing Fig.10.18(b).

(ii) As stated earlier, the zero-forcing approach totally ignores the channel noise. This, in fact, leads to an overall performance degradation owing to 'noise enhancement'.

(iii) Determination of zero-forcing tap-gains, i.e., c_ns using Eq. (10.57) requires that we know the $p_r(\cdot)$ values in the $(2N + 1) \times (2N + 1)$ square matrix. In the case of changes in channel characteristics, as happens in switched telephone links and slowly changing radio links, an a priori knowledge of $p_r(\cdot)$ values is not possible. In such cases, the tap gains are directly adjusted on-line by using a training sequence that is transmitted before the actual message sequence is transmitted.

Example 10.6 A zero-forcing equalizer is to be designed using 3 taps. Assume that the input pulse $p_r(t)$ to the equalizer is as shown in Fig. 10.18(a) in which $p_r(-2T) = 0.08$, $p_r(-T) = -0.25$, $p_r(0) = 1$, $p_r(T) = -0.20$ and $p_r(2T) = 0.10$

Solution Here, $2N + 1 = 3 \therefore N = 1$. Substituting the given values of $p_r(kT)$ in Eq. (10.57), we get

$$\begin{bmatrix} p_r(0) & p_r(-1) & p_r(-2) \\ p_r(1) & p_r(0) & p_r(-1) \\ p_r(2) & p_r(1) & p_r(0) \end{bmatrix} = \begin{bmatrix} 1 & -0.25 & 0.08 \\ -0.20 & 1 & -0.25 \\ 0.10 & -0.20 & 1 \end{bmatrix}$$

Hence, for 'zero forcing', the tap-gains c_ns must satisfy the following equation (Eq. (10.57)):

$$\begin{bmatrix} 1 & -0.25 & 0.08 \\ -0.20 & 1 & -0.25 \\ 0.10 & -0.20 & 1 \end{bmatrix} \begin{bmatrix} c_{-1} \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Solving the above set of three linear equations for c_{-1}, c_0 and c_1, we get

$$c_{-1} = 0.90825; \quad c_0 = 3.4386; \quad c_1 = -0.6075$$

Adaptive equalizer In all our discussions on ISI, till now, we have been assuming the channel characteristics (i.e., $h_c(t)$, its impulse response, or $H_c(f)$, its transfer function) to be known to us *a priori*. In practice, however, these channel characteristics are not only not known, in addition, they change with time. As already stated in point iii. of remarks under Eq. (10.57), a fixed tapped-delay equalizer will not therefore be able to give satisfactory performance. The tap gains of the equalizer will have to be automatically adjusted, taking into account the changes in the channel characteristics. Such an equalizer which automatically adapts itself to the channel is called an 'adaptive equalizer'.

The general arrangement in an adaptive equalizer is shown in Fig. 10.19. As shown there, an error sequence is generated by comparing the actual output of the equalizer with the 'desired output sequence' and this error sequence is used to appropriately change the tap-gains of the equalizer so as to minimize the error in some sense – usually, in the mean-square sense.

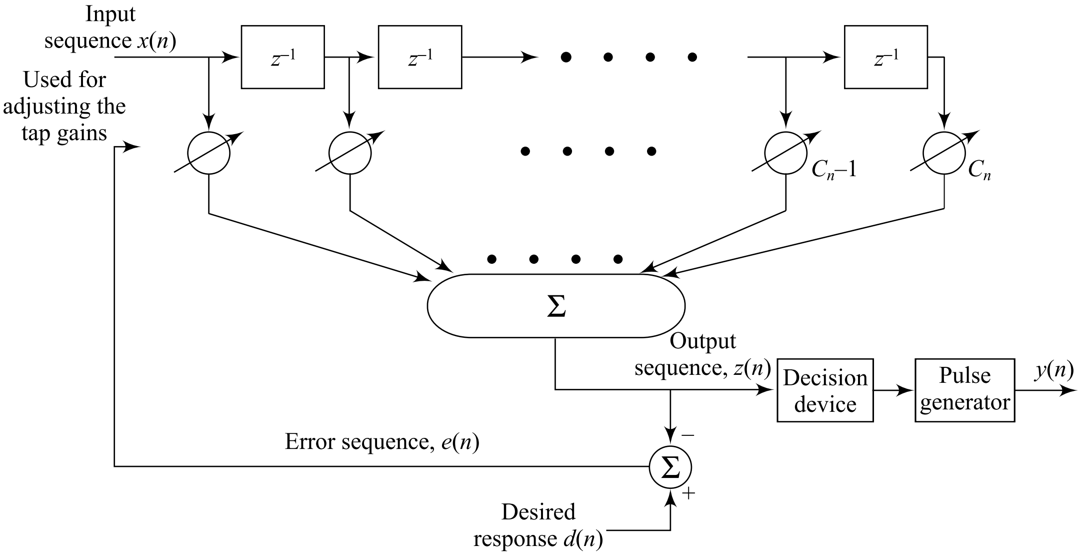


Fig. 10.19 *Adaptive equalizer block diagram*

Ideally, the desired output sequence from a receiver is nothing but the transmitted sequence itself. Hence, the question naturally arises: 'How can the desired output sequence (i.e., the correct transmitted sequence) be available at the receiving-end'? As has been mentioned earlier, what is done in practice is, a certain known sequence, called the 'training sequence', is transmitted from the transmitting-end before commencing the transmission of the actual data. A replica of this training sequence, which is already stored at the receiving-end, is applied as the desired signal (see Fig.10.19) after a time delay equal to the estimated / measured time delay of the channel. The error sequence now created is used to appropriately adjust the tap gains of the equalizer. This time used for initial adjustment of the tap gains using the training sequence, is called the '*training period*'.

Since the equalizer tap gains have converged to their optimum values by the end of the training period, the sequence at the output of the decision device and the pulse generator that follows it, will be sufficiently reliable, so that it may be used for continuing the tap-gain adjustment process. Hence, after the training period, the training sequence is switched off and normal data transmission is started at the transmitter. At the receiving-end too, the desired sequence will be supplied by the output of the pulse generator that follows the decision device, instead of the stored version of the training sequence, which is switched off after the transmitted training sequence is completely received. Since the output of the receiver decision device is used as the reference, or desired signal for adoption purpose, such a strategy is called '***decision-directed adaptation***'.

10.4.4 Scrambling

Scrambling is basically a process of randomizing the binary message bit stream at the transmitting end. It has the following beneficial effects:

1. It eliminates long strings of zeros or ones which might affect the receiver synchronization.
2. It makes it difficult to have unauthorized access to the data being transmitted.

It is, of course, necessary to undo, at the receiving end, the scrambling done at the transmitter to restore the original bit sequence of the bit stream as it existed before the scrambling.

As shown in Fig.10.20, the scrambler consists of a feedback shift register while the unscrambler consists of a feed-forward shift register.

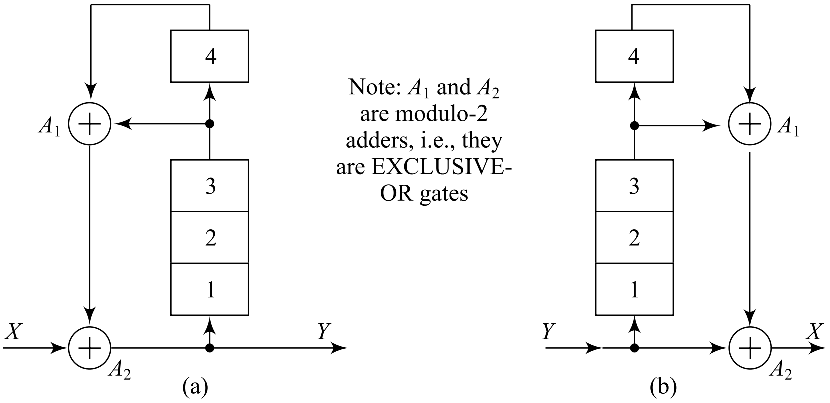


Fig. 10.20 *(a) Scrambler, (b) Unscrambler*

We know that as a bit passes through one stage of the shift register, there will be a unit delay, D, which is equal to the clock period. At the tick of the clock, the bit gets shifted from one stage of the shift register to the next. Hence, for the scrambler, we may write the following equation that relates its input sequence, X, with its output sequence, Y:

$$Y = D^3Y \oplus D^4Y \oplus X \tag{10.58}$$

Performing modulo-2 addition of $(D^3Y \oplus D^4Y)$ to both sides of the above equation, and noting that modulo-2 addition of any binary sequence to itself yields a zero sequence, we get

$$X = (1 \oplus D^3 \oplus D^4)Y \tag{10.59}$$

Since the unscrambler has to retrieve X from Y, it simply implements the above equation, as may be readily seen from Fig. 10.20(b).

There is, however, one serious problem with this arrangement – error propagation at the unscrambler. Any erroneous bit in the input Y gives to the unscrambler will cause several output bits to be in error. The error propagation will stop only when all the bits in the unscrambler shift register are correct.

Example 10.7 Assuming the initial contents of all the shift registers of the scrambler of Fig.10.20(a) to be zeros, find the output sequence Y for an input sequence X given by $X =$ **1 0 1 0 1 0 1 1 1 1 1 1**.

Solution As the shift register contents are all zero initially, when we apply X, the output Y will be initially equal to X. Then as X enters the shift registers, it returns as $(D^3 \oplus D^4)X$ where D is the unit delay operator.

If we represent $(D^3 \oplus D^4)$ by W, then the output sequence Y is given by

$$Y = X \oplus WX \oplus W^2X \oplus W^3X \ldots$$

But, $W \triangleq (D^3 \oplus D^4) \quad \therefore W^2 = D^6 \oplus D^8; \quad \therefore W^3 = D^9 \oplus D^{10} \oplus D^{11} \oplus D^{12}$

Substituting for W in the expression for Y and simplifying, we get

$$Y = X \oplus D^3X \oplus D^4X \oplus D^6X \oplus D^8X \oplus D^9X \oplus D^{10}X \oplus D^{11}X \oplus D^{12}X \oplus \ldots$$

We now write all these terms (up to only $D^{12}X$ only since the rest of the terms involving higher powers of the delay operator D will not affect the first 12 digits, as our input sequence X has been given only up to the first 12 digits) and perform the modulo-2 addition to get Y.

$$\begin{aligned}
X &= 1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^3X &= 0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^4X &= 0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^6X &= 0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^8X &= 0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^9X &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^{10}X &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^{11}X &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
D^{12}X &= 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,1\,1\,1\,1\,1\\
\therefore \quad Y &= \mathbf{1\,0\,1\,1\,0\,1\,1\,0\,0\,1\,0\,1}
\end{aligned}$$

Note that even though X, the input to the scrambler contains some periodicities and a long string of ones, the output Y of the scrambler is devoid of these undesirable patterns in the input bit stream.

10.4.5 Eye Pattern

The eye pattern is an experimental tool that can be used in the field for evaluating the combined effect of ISI and channel noise in the case of digital transmission systems. The only equipment needed to observe the eye pattern is an oscilloscope with a high persistence screen. To generate the eye diagram, the received baseband signal at the output of the channel which is given by

$$y(t) = \left[\sum_k A_k P(t - k/R_b\right] + n(t) \tag{10.60}$$

is applied the input to the Y-plates of the oscilloscope. The time base is triggered at the pulse rate (1/T) so that a pattern of duration T seconds is produced. This pattern, displayed on the screen of the oscilloscope, is actually a synchronized superposition of the pulses in successive time slots (or symbol intervals if it is M-ary baseband signaling). It is called an eye pattern, or an eye diagram because it resembles the human eye for a binary baseband signal applied to the scope. Figure 10.21a shows the sketch of the output of the channel when polar NRZ pulse train is transmitted. Figure 10.21b illustrates the formation of an eye pattern or the screen of the oscilloscope.

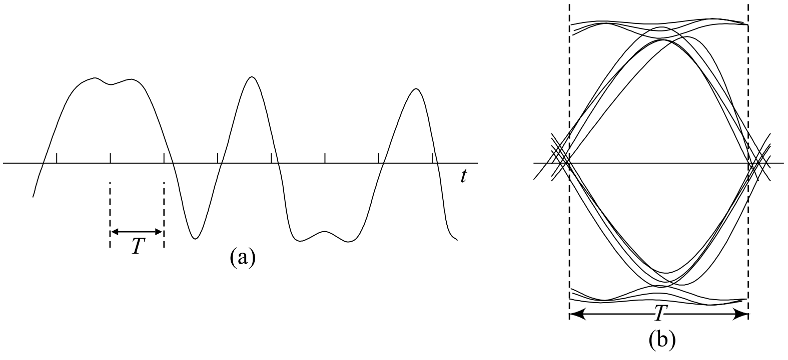


Fig. 10.21 *(a) Polar NRZ binary signal distorted by transmission through the channel (Noise is not shown), (b) Corresponding eye pattern*

Figure 10.22(a) shows a computer generated (using MATLAB) eye diagram of a polar NRZ signaling system using a raised cosine pulse with a roll-off factor of 0.5 without channel noise. Figure 10.22(b) shows the eye diagram for the same transmitted signal but with channel noise (AWGN) added so as to produce an *SNR* of 20 dB. Note the decrease in the eye opening and the increase in the zero-crossing jitter.

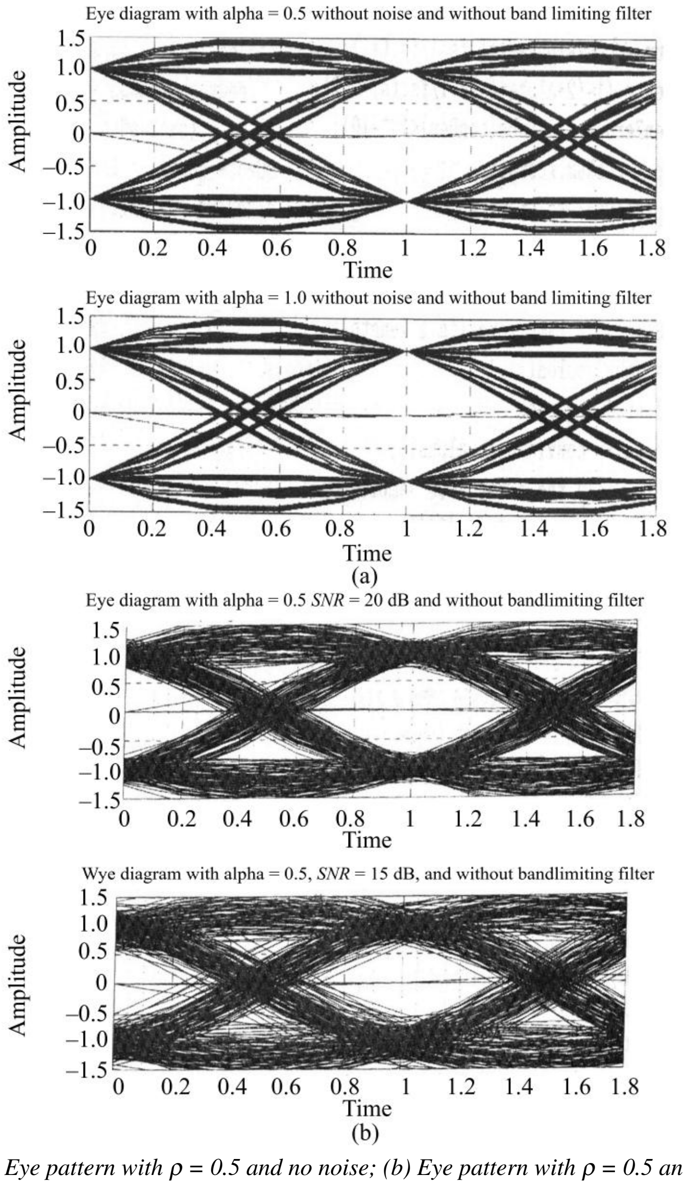


Fig. 10.22 *(a) Eye pattern with ρ = 0.5 and no noise; (b) Eye pattern with ρ = 0.5 and SNR = 20 dB*

Quite a lot of useful information is provided by the eye pattern as to how the baseband transmission system is performing. One can, by observing the pattern, draw inferences regarding the extent of ISI, the extent of zero-crossing jitter, the noise margin available, etc. The interpretation of the various features of an eye pattern are all marked on the generalized binary eye pattern shown in Fig. 10.23 and these will be helpful in drawing inferences on the performance of a system by observing the eye pattern.

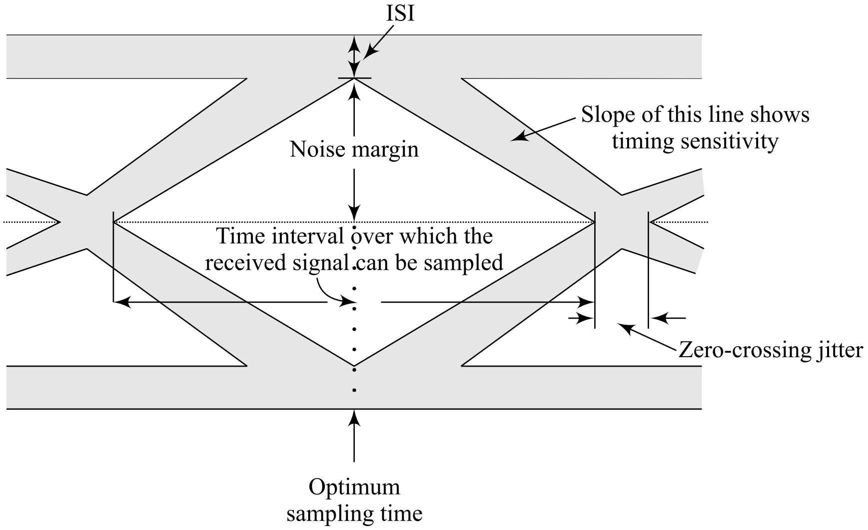


Fig. 10.23 *Generalized binary eye pattern and its interpretations*

Remarks

(i) *The optimum sampling time corresponds to maximum eye opening as this gives maximum noise margin. Sampling at any other instant results in the noise margin getting reduced – that is tolerance to noise is reduced. In such a situation, if a pulse encounters a noise peak of opposite polarity with an amplitude greater than this reduced noise margin, then the polarity of the pulse will change and an error will creep in.*

(ii) *ISI at the sampling instant partially closes the eye and thus reduces the noise margin. The noise margin is maximum at the optimum sampling time.*

(iii) *The width of the eye opening shows the time interval over which the received signal can be sampled.*

(iv) *For M-ary baseband signaling, the pattern that we get on the screen of the oscilloscope would be (M – 1) 'eyes' stacked vertically.*

10.5 BAND PASS DIGITAL TRANSMISSION

In this chapter, till now we have considered transmission of digital baseband signals. As we have seen, these signals have considerable low frequency content. So they cannot be directly radiated using antennas of practicable size. So, in order to transmit these digital baseband signals over long distances using terrestrial radio links or satellite links, we have to shift them to a higher frequency by making them to modulate a high frequency carrier. Recall that this is exactly what we did with **analog** baseband signals like speech. In the case of digital baseband signals too, modulating a high frequency carrier not only makes it possible to radiate them using practical antennas, but also makes it possible to transmit several signals simultaneously over the same physical channel, which is termed as 'multiplexing'. For one way communication, we need a modulator at the transmitting-end, and a demodulator at the receiving end. For two-way or duplex communication, we need a modulator as well as demodulator at both the ends. So, these two are usually packaged into a single unit and this unit goes by the name '**modem**'

Binary digital modulation schemes There are basically three types of digital modulation schemes. These are:

1. Amplitude Shift Keying (ASK) or, On-off Keying (OOK)
2. Frequency Shift Keying (FSK)
3. Phase Shift Keying (PSK)

These are analogous respectively to Amplitude Modulation (AM), Frequency Modulation (FM) and Phase Modulation (PM) in analog continuous-wave modulation schemes. The following table gives details of these digital modulation schemes.

Table 10.3 *Basic digital modulation schemes*

S. No.	**Type of modulation**	**Baseband binary digit**	
		$b_k = 0$	$b_k = 1$
		$s_1(t)$	$s_2(t)$
		$0 \leq t \leq T$	$0 \leq t \leq T$
1	Amplitude Shift Keying (ASK) or ON-OFF Keying (OOK)	0	$A \cos \omega_c t$ or $A \sin \omega_c t$
2	Frequency Shift Keying (FSK)	$A \cos(\omega_c - \omega_d)$ Or $A \sin (\omega_c - \omega_d)t$	$A \cos(\omega_c + \omega_d)t$ or $A \sin(\omega_c + \omega_d)t$
3	Phase Shift Keying (PSK)	$-A \cos \omega_c t$ or $-A \sin \omega_c t$	$A \cos \omega_c t$ or $A \sin \omega_c t$
Note *f_c and f_d are suitably chosen in relation to T, as explained later.*			

For each type of modulation, corresponding to binary digit $b_k = 0$ a signal $s_1(t)$ of duration T seconds and corresponding to binary digit $b_k = 1$, a signal $s_2(t)$ of duration T seconds, are produced as shown in Table 10.3. The waveforms of $s_1(t)$ and $s_2(t)$ for the three basic digital modulation schemes are shown in Fig. 10.24.

Note *To ensure that each bit interval of T sec contains an integer number of cycles of the carrier wave, the carrier frequency f_c is chosen to be n/T, where n is a suitable fixed integer.*

10.5.1 Methods of Generation

Amplitude shift keying (ASK) As Fig. 10.24(b) suggests, binary ASK signal is generated by simply switching on the carrier oscillator for $b_k = 1$ and switching it off for $b_k = 0$. This may be implemented by multiplying the carrier signal with a unipolar NRZ coded bit stream, making use of a product device as shown in Fig. 10.25.

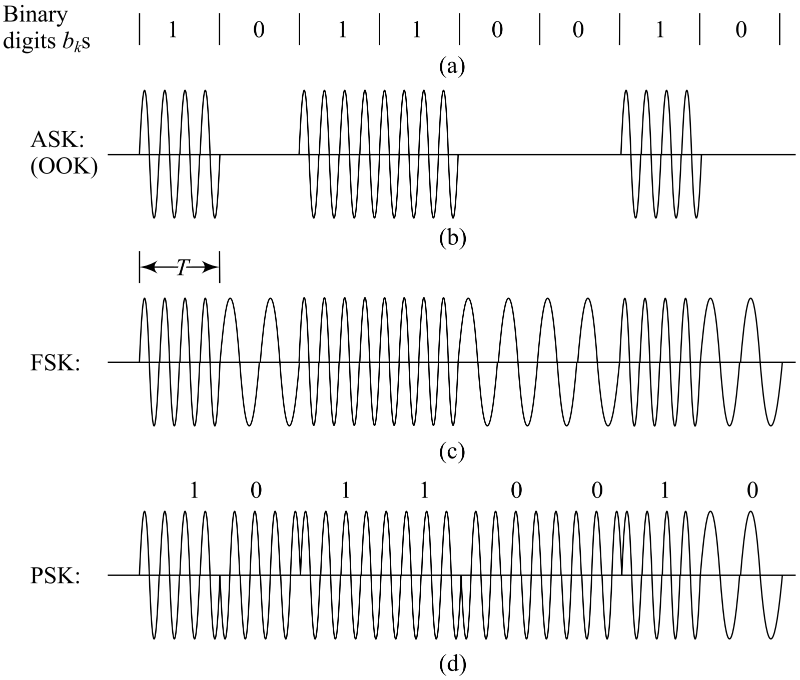


Fig. 10.24 *Waveforms of digitally modulated binary signals: (a) Binary digits, (b) ASK, (c) FSK, (d) PSK*

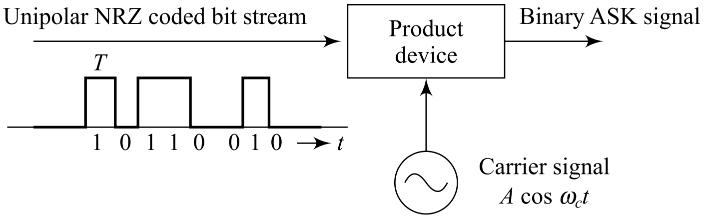


Fig. 10.25 *Generation of binary ASK signal*

Frequency shift keying (FSK) One easy method of generation of a BFSK signal is to have two oscillators, one of frequency $(f_c - f_d)$ and the other of frequency $(f_c + f_d)$ and switching in the former whenever $b_k = 0$ and the latter whenever $b_k = 1$. This can be achieved by making the switch to be operated by the binary baseband signal.

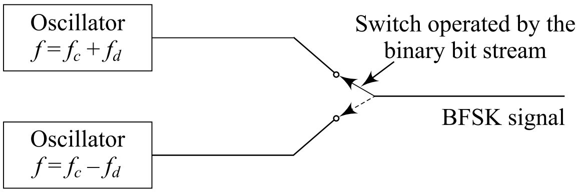


Fig. 10.26 *Generation of BFSK (discontinuous phase) signal*

Though simple to implement, this method of generation of a BFSK signal has one serious disadvantage. Because of the switching, the modulated signal is discontinuous at each switching instant. The spectrum of the resultant BFSK signal will contain large side-lobes which do not carry any information and only waste the bandwidth. To avoid these discontinuities, we may go in for what is referred to as the ‘continuous-phase FSK’ (CPFSK) which can be obtained by having only one oscillator and frequency modulating it using the binary baseband bit stream as the modulating signal.

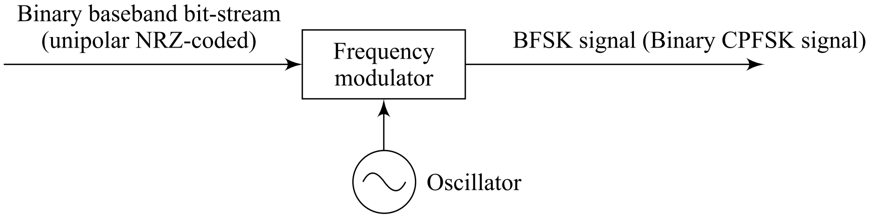


Fig. 10.27 *Generation of BFSK (continuous phase) signal*

Binary phase shift keying (PSK) Since $s_1(t)$ and $s_2(t)$, the two signals produced corresponding respectively to $b_k = 0$ and $b_k = 1$ are such that $s_1(t) = -s_2(t)$ in the case of BPSK, a simple arrangement as shown in Fig. 10.28 may be used for generating a BPSK signal.

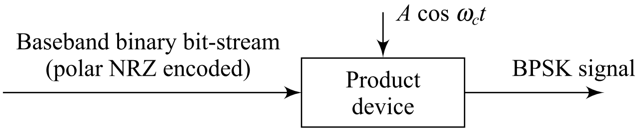


Fig. 10.28 *Generation of BPSK signals*

10.5.2 Power Spectra of Binary ASK, FSK and PSK Signals

Binary ASK From Fig. 10.24, we find that a binary ASK signal is generated by multiplying a unipolar NRZ-coded binary baseband bit stream with the carrier signal of frequency f_c. Hence, the power spectrum of binary ASK signal is simply the frequency-shifted version of the power spectrum of a unipolar NRZ line code given earlier in Fig. 10.25(a). Hence, the PSD of a binary ASK signal is as shown below in Fig. 10.29.

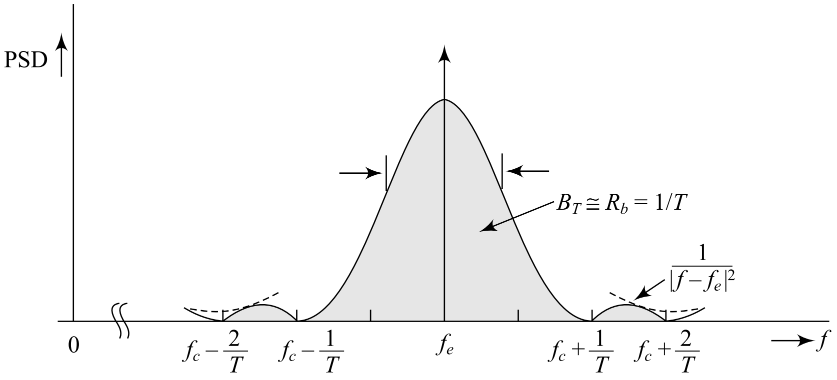


Fig. 10.29 *PSD of a binary ASK signal*

This PSD, being a sinc^2 function, has a roll-off proportional to $|f - f_c|^2$ as we move away from the carrier frequency f_c. Since $(1/T)$ equals R, the bit rate, the null-to-null bandwidth is $2R$ as shown. Also, the transmission bandwidth of the binary ASK signal is $B_T = 2B$, where B is the baseband transmission bandwidth, since ASK is nothing but an AM type of modulation. Since the minimum baseband transmission bandwidth = $R_b/2$ the minimum transmission bandwidth B_T for ASK is R, where R_b is the binary bit rate = $1/T$.

Binary FSK As shown in Table 10.3 and Fig. 10.24(c), in binary FSK, the two signals $s_1(t)$ and $s_2(t)$ transmitted corresponding to $b_k = 0$ and $b_k = 1$ differ only in frequency. Of the different types of BFSK signals, as stated in the discussion on the method of generation of BFSK, we shall consider here the continuous phase type called Sunde's FSK. For BFSK, we know

$$b_k = 0 : s_1(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos 2\pi(f_c - f_d)t; & 0 \le t \le T \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{10.61}$$

and
$$b_k = 1 : s_2(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos 2\pi(f_c + f_d)t; & 0 \le t \le T \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{10.62}$$

Here, E_b is the bit energy and T is the bit duration. The peak amplitude of the cosine function in $s_1(t)$ and $s_2(t)$ is taken as $\sqrt{2E_b/T}$ so as to normalize and ensure that each signal has an energy of E_b. The frequency offset is f_d. This f_d is taken as $(1/2T)$ in Sunde's FSK. Hence, we shall write

$$s_1(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos 2\pi(f_c - 1/2T)t; & 0 \le t \le T \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{10.63}$$

and
$$s_2(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos 2\pi(f_c + 1/2T)t; & 0 \le t \le T \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{10.64}$$

We may conveniently club these two equations and write

$$s(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos\left(2\pi f_c \mp \frac{\pi}{T}\right)t; & 0 \le t \le T \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{10.65}$$

where $s(t) = s_1(t)$ when the minus sign is taken and $s\,(t) = s_2(t)$ when the plus sign is taken.

Now, for finding the PSD of Sunde's BFSK, we need to concentrate on $s(t)$ over the interval $0 \le t \le T$.

$$s(t) = \sqrt{\frac{2E_b}{T}} \cos\left(2\pi f_c \mp \frac{\pi}{T}\right)t; \quad 0 \le t \le T \tag{10.66}$$

$$= \sqrt{\frac{2E_b}{T}} \cos\left(\mp\frac{\pi t}{T}\right)\cos 2\pi f_c t - \sqrt{\frac{2E_b}{T}} \sin\left(\mp\frac{\pi t}{T}\right)\sin 2\pi f_c t$$

$$= \sqrt{\frac{2E_b}{T}} \cos\left(\frac{\pi t}{T}\right)\cos 2\pi f_c t \pm \sqrt{\frac{2E_b}{T}} \sin\left(\frac{\pi t}{T}\right)\sin 2\pi f_c t \tag{10.67}$$

In the RHS of the above equation, the first term is the inphase component and the second term is the quadrature component. The plus sign corresponds to $b_k = 0$ and the minus sign corresponds to $b_k = 1$. As per our practice, we shall assume that the 0s and 1s occur with equal probability and that they are emitted by the source independently.

We find that the first term, i.e., the inphase component of $s(t)$ is independent of which symbol is transmitted while the second term, i.e., quadrature component is dependent on whether $b_k = 0$ or 1. Thus, the two components are not only orthogonal, they are statistically independent. So we can find the PSD of each of these components and add their PSDs to get the PSD of $s(t)$.

The inphase component is independent of b_k and so is deterministic and its value is $\sqrt{2E_b/T}\cos(\pi t/T)$ for all time. Hence its PSD works out to

$$\text{PSD of inphase component } = \frac{E_b}{2T}\left[\delta\left(f-\frac{1}{2T}\right)+\delta\left(f+\frac{1}{2T}\right)\right] \tag{10.68}$$

i.e., it consists of two delta functions each of strength $E_b/2T$ and located at $f = -\frac{1}{2T}$.

The quadrature component value, on the other hand, is random and is given by

$$\begin{cases} \sqrt{\frac{2E_b}{T}}\sin\left(\frac{\pi t}{T}\right); & 0 \le t \le T, \quad \text{if } b_k = 1 \\ -\sqrt{\frac{2E_b}{T}}\sin\left(\frac{\pi t}{T}\right); & 0 \le t \le T, \quad \text{if } b_k = 0 \\ 0; & \text{outside the interval } 0 \le t \le T \end{cases}$$

Being of finite duration, it is an energy signal and its power spectral density is given by its energy spectral density divided by its duration, T.

$$\text{PSD of quadrature component } = \frac{8E_b\cos^2(\pi fT)}{\pi^2(4f^2T^2-1)^2} \tag{10.69}$$

Hence $S_B(f)$, the baseband PSD of Sunde's BFSK signal works out to be the sum of the PSDs of the inphase and quadrature components (which are independent)

$$\therefore \qquad S_B(f) = \frac{E_b}{2T}\left[\delta\left(f-\frac{1}{2T}\right)+\delta\left(f+\frac{1}{2T}\right)\right] = \frac{8E_b\cos^2(\pi fT)}{\pi^2(4f^2T^2-1)^2} \tag{10.70}$$

Noting that carrier modulation simply translates this baseband power spectrum along the frequency axis by f_c, the PSD of Sunde's BFSK signal is plotted in Fig. 10.30.

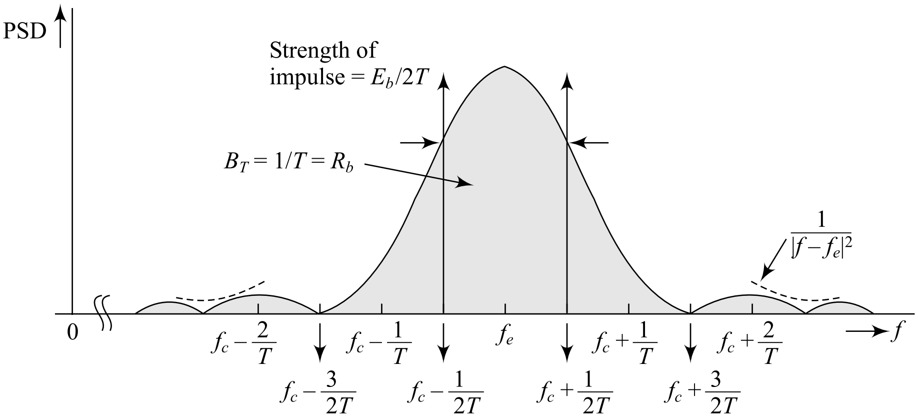


Fig. 10.30 *Power spectrum of Sunde's BFSK signal*

Remark

(i) Since $f_d = \frac{1}{2T}$ *in Sunde's FSK, from Fig. 10.30, we find that the impulses located at* $\left(f_c - \frac{1}{2T}\right)$ *and* $\left(f_c + \frac{1}{2T}\right)$ *correspond to the keyed frequencies* $(f_c - f_d)$ *and* $(f_c + f_d)$.

(ii) The spectrum has a fourth-order roll off.

(iii) Because of the rapid roll off, the spectrum has very small values beyond $|f - f_c| > \frac{1}{T}$, *i.e., for* $|f - f_c| > R_b$. *Hence, the transmission bandwidth* B_T *is generally taken as* $B_T = R_b$.

(iv) The two frequency impulses located at $\left(f_c - \frac{1}{2T}\right)$ *and* $\left(f_c + \frac{1}{2T}\right)$ *are helpful in arranging synchronization of the receiver with the transmitter.*

Binary PSK From Fig. 10.28, we find that we may write $s_1(t)$ and $s_2(t)$ in the case of BPSK as:

$$s_1(t) = -A\cos\omega_c t;\quad 0 \le t \le T$$

and

$$s_2(t) = A\cos\omega_c t;\quad 0 \le t \le T$$

Since we are representing the bit energy as E_b, we will replace A by $\sqrt{2E_b/T}$ and write $s_1(t)$ and $s_2(t)$ as

$$\left.\begin{aligned} s_1(t) &= -\sqrt{\frac{2E_b}{T}}\cos\omega_c t;\quad 0 \le t \le T \\ \text{and}\qquad s_2(t) &= \sqrt{\frac{2E_b}{T}}\cos\omega_c t;\quad 0 \le t \le T \end{aligned}\right\} \tag{10.71}$$

As usual, we shall assume that a binary 0 and a binary 1 are equally likely and that the symbols in adjacent time slots are statistically independent. The BPSK modulated signal will therefore have a corresponding baseband binary sequence which will be a random polar binary wave with amplitudes of either $\sqrt{2E_b/T}$ or $-\sqrt{2E_b/T}$ with equal probability.

Since the signal

$$\left.\begin{matrix} \sqrt{\dfrac{2E_b}{T}}; & 0 \le t \le T \\ 0; & \text{otherwise} \end{matrix}\right\} \tag{10.72}$$

is Fourier transformable, we may find the energy spectral density of the above random signal and divide it by the symbol duration T to get the baseband PSD for a BPSK signal. The energy spectral density itself is the squared magnitude of the Fourier transform of the signal in Eq. (10.72). Thus, the baseband PSD of BPSK signal is obtained as

$$S_B(f) = \frac{2E_b \sin^2(\pi fT)}{(\pi fT)^2} = 2E_b\,\text{sinc}^2(fT) \tag{10.73}$$

Since the carrier simply shifts the baseband spectrum by a frequency of f_c, we obtain the PSD of a BPSK signal as shown in Fig. 10.31.

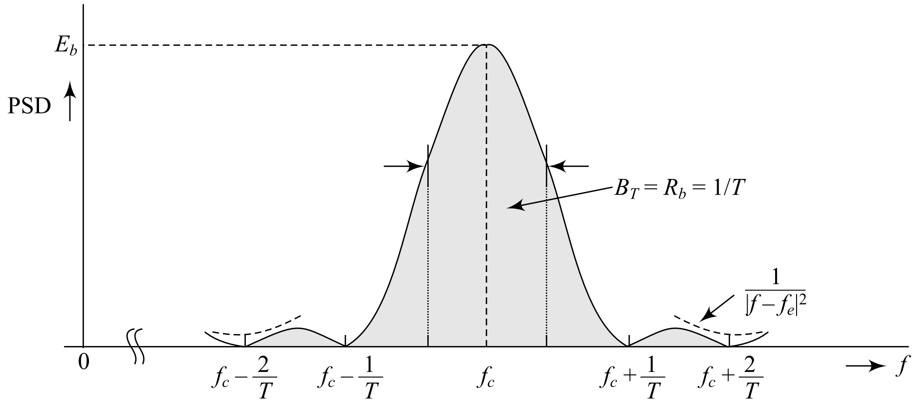


Fig. 10.31 *PSD of a BPSK signal*

Remarks

(i) The spectrum has a second-order roll off.

(ii) The transmission bandwidth $B_T \cong \frac{1}{T} = R_b$.

10.5.3 Detection of Binary ASK, FSK and PSK Signals

As mentioned earlier in the introduction to this chapter, band pass digital signals may be detected coherently or non-coherently, with the exception of PSK which can be detected only coherently. Although coherent reception requires elaborate carrier recovery arrangements, it gives optimum performance which the non-coherent receivers will not be able to give. We shall now discuss coherent detection of binary ASK, FSK and PSK and non-coherent detection of binary ASK and FSK. Coherent reception implies the usage of a locally generated carrier signal by the receiver. This locally generated carrier must be in frequency and phase synchronization with the carrier signal at the transmitter. This, in turn, requires the usage of carrier recovery arrangements in the receiver, which recover the carrier signal from the received band pass digital signal itself. In addition, irrespective of whether it is coherent, or non-coherent reception, synchronization of the receiver with the transmitter is absolutely necessary for proper operation of the system, as it is necessary for the receiver to know precisely when the new symbol starts, so that the sampling is done at the correct instant during each symbol period. This necessitates symbol-timing recovery, or clock recovery. Methods of carrier recovery and clock recovery are discussed in Section 10.9 of this chapter.

At this stage, we propose to discuss only the principle of detection of each of the modulation schemes without bothering about the effect of channel noise. A detailed analysis of the noise performance of each of these modulation schemes is reserved for the next chapter.

Coherent detection of binary ASK signals If we denote the modulated signal by $x_c(t)$, for a binary ASK signal, we have

$$x_c(t) = \begin{cases} s_2(t) = A\cos\omega_c t & \text{for } 0 \le t \le T \text{ if } b_k = 1 \\ s_1(t) = 0 & \text{for } 0 \le t \le T \text{ if } b_k = 0 \end{cases} \tag{10.74}$$

whereas discussed earlier, T denotes the binary symbol period and b_k denotes the binary symbol.

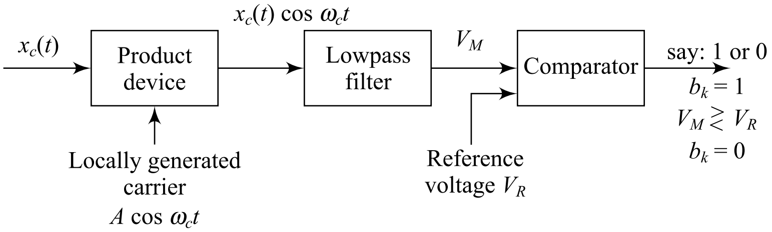


Fig. 10.32 *Coherent detection of an ASK signal*

$$Ax_c(t)\cos\omega_c t = \begin{cases} A^2\cos^2\omega_c t = \frac{A^2}{2}[1+\cos 2\omega_c t] & \text{if } b_k = 1 \\ 0 & \text{if } b_k = 0 \end{cases}$$

Low pass filtering removes the high frequency component, viz., $(A^2/2)\cos 2\omega_c t$. Hence, V_M equals $(A^2/2)$ if $b_k = 1$ and 0 if $b_k = 0$. In the absence of channel noise, V_R, the reference voltage, can as well as set at zero volts. As will be discussed in the next chapter, in the presence of channel noise, V_R is set at an optimum value, which of course, will not be zero.

Non-coherent detection of binary ASK signals As in the case of AM, an ASK signal can be detected non-coherently also. From the waveform of an ASK signal shown in Fig. 10.24(a), it is clear that as in the case of AM, for ASK too we may make use of envelope detection. Since the envelope detector extracts the envelope of the input signal, the detector output will be almost equal to the peak value of the received sinusoidal signal for binary baseband symbol '0'. Hence, an arrangement shown in Fig. 10.33 can be used.

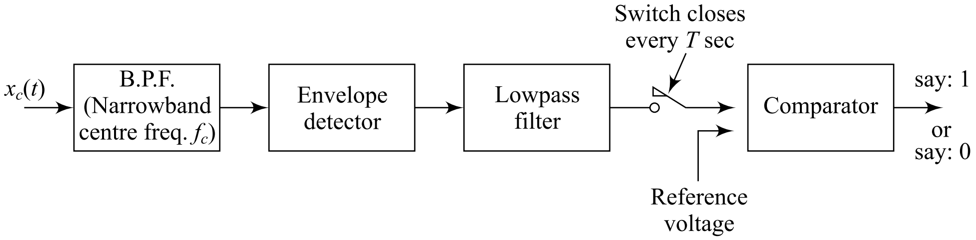


Fig. 10.33 *Non-coherent detection of a binary ASK signal*

Coherent detection requires a complex arrangement while non-coherent detection of ASK needs only a simple and relatively inexpensive arrangement. The received signal at the output of the channel will not be just $x_c(t)$, the ASK signal plus additive white noise. A part of the noise, falling within the pass band of the BPF reaches the input of the envelope detector. Hence, for reasonably high values of received *SNR*s, non-coherent detection is used for ASK. In fact, if one is prepared to go in for coherent detection, PSK is a better option than the ASK, the only attractive feature of which is its simplicity. That is why coherent detection of ASK is seldom used.

Coherent detection of binary FSK signals For binary FSK, during each time slots, one of the two signals $s_1(t)$ and $s_2(t)$ will be transmitted, where

$$s_1(t) = A\cos(\omega_c - \omega_d)t; \quad 0 \le t \le T \quad \text{for} \quad b_k = 0$$

and

$$s_2(t) = A\cos(\omega_c + \omega_d)t; \quad 0 \le t \le T \quad \text{for} \quad b_k = 1$$

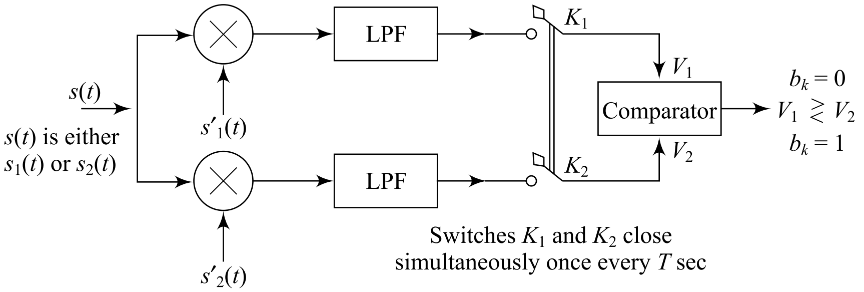


Fig. 10.34 *Coherent detection of binary FSK*

Signals $s_1'(t)$ and $s_2'(t)$ which are in frequency and phase synchronism with $s(t)$, are generated in the receiver and used for being correlated with the $s(t)$ to be detected. If $s(t)$ is equal to $s_1(t)$ during a given time slot, $s_1'(t)$ correlates well with it while $s_2'(t)$ does not. So, during that time slot, when the switches k_1 and k_2 are closed, $V_1 > V_2$ and the $s(t)$ being detected is declared to be $s_1(t)$ during that time-slot. On the other hand, if $s(t)$ were to be $s_2(t)$ during a time slot, $V_2 > V_1$ and this fact is used to declare that $s(t) = s_2(t)$ during the time slot.

Non-coherent detection of binary FSK Since $s_1(t)$ and $s_2(t)$ differ only in frequency, we may make use of the following arrangement shown in Fig. 10.35 for non-coherent detection of binary FSK.

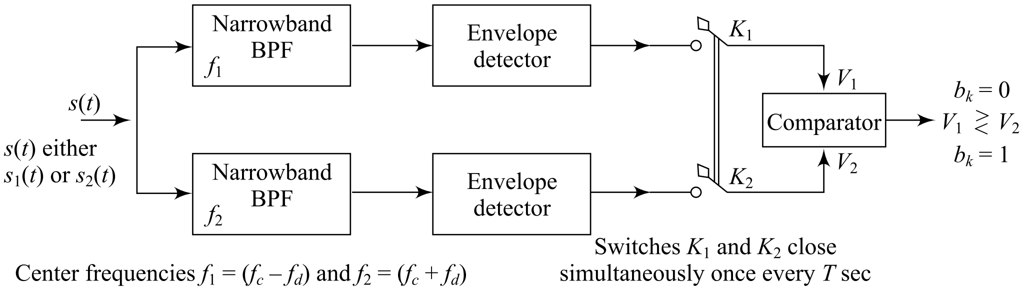


Fig. 10.35 *Non-coherent detection of binary FSK*

Suppose $s(t) = s_1(t)$ during a particular time slot of T seconds. Then ideally the output of the top narrowband filter with center frequency at f_1 will be equal to $s_1(t)$ and that of the other narrowband filter will be zero. Since an envelope detector extracts the envelope of the signal given as input to it, the output of the top envelope detector would be almost equal to the peak value of the signal $s_1(t)$ while the output of the lower arm envelope detector would be zero. Hence, $V_1 > V_2$ and the decision would be that $b_k = 0$ during that time slot. A similar thing happens if $s(t)$ were to be equal to $s_2(t)$ during a given time slot. Then $V_2 > V_1$ and the receiver's decision would be that $b_k = 1$ during that time slot.

Detection of binary PSK Since the information regarding whether $b_k = 0$ or 1 during a given time slot is contained in the phase of the binary PSK signal, coherent detection alone is possible in this case. Hence, the following arrangement, shown in Fig. 10.36, may be used for detection. For PSK,

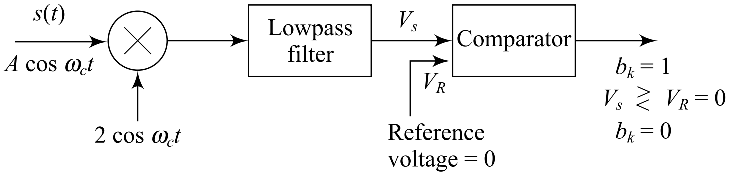


Fig. 10.36 *Detection of binary PSK signals*

If $b_k = 1$:

$$s(t) 2 \cos \omega_c t = s_2(t) 2 \cos \omega_c t = 2A \cos^2 \omega_c t = A(1 + \cos 2\omega_c t)$$

The LPF removes the $\cos 2\omega_c t$ term making $V_s = A$.
If $b_k = 0$:

$$s(t) 2 \cos \omega_c t = s_1(t) 2 \cos \omega_c t = -2A \cos^2 \omega_c t = -A(1 + \cos 2\omega_c t)$$

Since the LPF rejects the $\cos 2\omega_c t$ term, $V_s = -A$.
Thus, if $V_s > 0$, the decision would be that $b_k = 1$ and if $V_s < 0$, it would be that $b_k = 0$.

10.6 DIFFERENTIAL PHASE SHIFT KEYING (DPSK)

In ordinary binary phase shift keying, a signal, $s_2(t) = A \cos \omega_c t$ is transmitted in a given time-slot, if the corresponding baseband binary digit in that time slot is a 1 and a signal $s_1(t) = -A \cos \omega_c t = A \cos(\omega_c t + \pi)$ is transmitted if the binary baseband digit is a 0.

However, BPSK requires coherent detection, which, at the receiving-end, involves the use of a carrier signal that is in frequency and phase synchronism with the carrier at the transmitter. A technique, known as 'Differential Phase Shift Keying', or DPSK, obviates the need for a coherent detector, but is sub-optimum in its performance. In this method, the waveform transmitted during a given time slot is $A\cos\omega_c t$ if the baseband binary digits, in that time slot and the preceding one, are alike, i.e., both are either 1s or 0s; and the waveform transmitted is $-A\cos\omega_c t$ if they are not alike. Thus, at the transmitter, the two major operations are differentially encoding the binary baseband data stream (which is to be transmitted) and then phase shift keying of the encoded bit stream.

The receiver therefore makes use of the signal received in the preceding time slot itself for the reference phase. This, it does, by correlating the signal received during the present time slot with that received in the previous time slot. The receiver, therefore, must have storage facility. If the two signals are correlated, it decides that the message bit in the present time slot is a 1, and if they are anti-correlated, it decides that the present message bit is a 0. Instead of using a clean carrier waveform for phase reference, this method utilizes the noisy received signal of the preceding time slot, so it is a sub-optimum method.

10.6.1 Differential Encoding and Phase Shift Keying

Since differential encoding entails comparison of the present bit with the previous bit of the input bit stream, there is a need to have a 'reference', or 'start-up digit'. This is always taken as 1. The encoding process is illustrated in Table 10.4.

Table 10.4 *Differential encoding and decoding*

Message sequence	1	1	0	0	1	1	0	0
Encoded sequence; Reference digit 1	1	1	0	1	1	1	0	1
Transmitted phase 0	0	0	π	0	0	0	π	0
Decoded sequence	1	1	0	0	1	1	0	0

Figure 10.37 shows the block diagram of a DPSK transmitter wherein the differential encoding and phase shift keying are done.

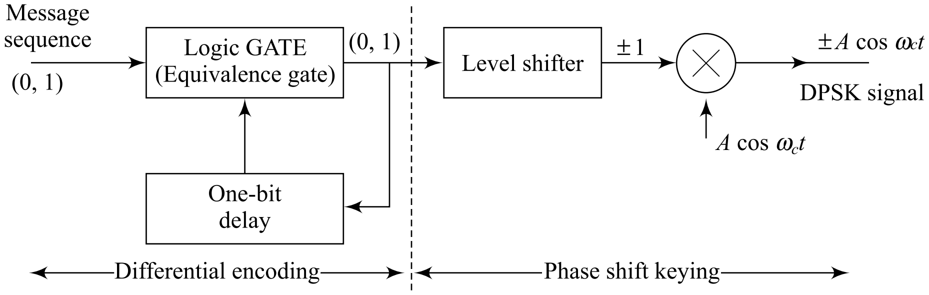


Fig. 10.37 *Block diagram of a DPSK transmitter*

The logic gate, called the equivalence gate, has to output a 1 if the present digit and the immediate previous digit of the message bit-stream are alike; and a 0 as output if they are not alike. Hence, the truth table of the equivalence gate is as shown in Table 10.5.

Table 10.5 *Truth table of the equivalence gate*

X_1	X_2	Y
0	0	1
0	1	0
1	0	0
1	1	1

A look at the truth table of the equivalence gate reveals that the operation it performs is just the complement of that of an Exclusive – OR gate. The sequence of 0s and 1s at the output of the equivalence gate is converted by the level shifter into a sequence of –1s and 1s, i.e., the level shifter converts the non-polar sequence into a polar NRZ sequence. This polar sequence is now used to modulate the carrier signal $A \cos \omega_c t$ by a process of multiplication, in order to get the DPSK signal.

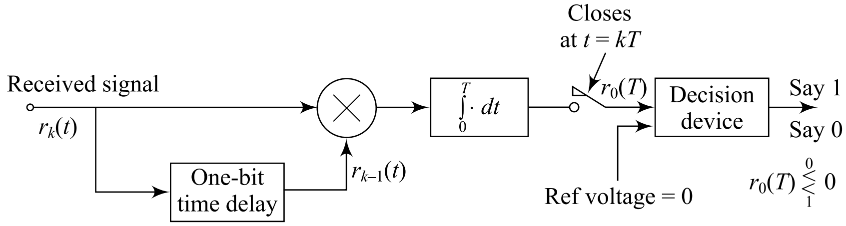


Fig. 10.38 *Block diagram of a DPSK receiver*

Let the received signal during the *k*th time slot be

$$r_k(t) = A \cos \omega_c t; \ (k-1)T \le t \le kT \tag{10.75}$$

The received signal in the previous time slot may be either $A \cos \omega_c t$ or $-A \cos \omega_c t$. Taking it as $A \cos \omega_c t$, we have

$$r_{k-1}(t) = A \cos \omega_c t; \ (k-2)T \le t \le (k-1)T \tag{10.76}$$

Then,

$$r_0(T) = \int_0^T A^2 \cos^2 \omega_c t \, dt = \frac{1}{2} A^2 T, \quad \text{since } T = \frac{n}{2f_c} \tag{10.77}$$

But, if

$$r_{k-1}(t) = -A \cos \omega_c t; \ (k-2)T \le t \le (k-1)T \tag{10.78}$$

Then,

$$r_0(T) = -\int_0^T A^2 \cos^2 \omega_c t \, dt = -\frac{1}{2} A^2 T \tag{10.79}$$

Similarly, if $r_{k-1}(t) = -A \cos \omega_c t$ and $r_k(t)$ is also $-A \cos \omega_c t$, then also $r_0(T)$ will be equal to $+\frac{1}{2}A^2T$, while it will be equal to $-\frac{1}{2}A^2T$ if $r_{k-1}(t)$ and $r_k(t)$ are of opposite sign.

Thus, if $r_0(T) > 0$, the receiver decides that the present bit is 1 and if $r_0(T) < 0$, the receiver decides that the present bit is a 0.

10.6.2 Differentially Encoded Phase Shift Keying (DEPSK)

DEPSK differs from DPSK only in the manner in which the receiver operates. There is no difference insofar as the transmitters are concerned. But the receiver configurations are different.

DPSK receiver makes use of a delay of *T at the carrier frequency*. Of course the DPSK transmitter also requires a delay of *T* sec, but that delay is at baseband level, not at carrier frequency. The DPSK system obviates the use of a *T* sec delay element at the carrier frequency by *coherent detection* of the DPSK signal to recover the encoded sequence of digits. From the encoded sequence, the message sequence is then recovered by making use of a baseband decoder shown in Fig.10.39.

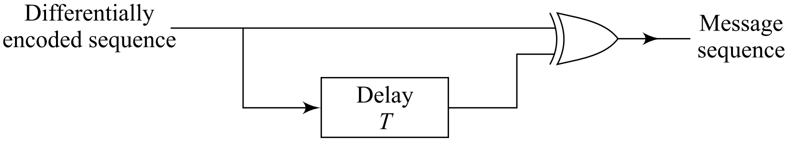


Fig. 10.39 *A baseband decoder to recover message bit stream from the differentially encoded bit stream*

Advantage It avoids the use of delay at carrier frequency.

Disadvantage Needs coherent detection to recover the differentially encoded bit stream.

10.6.3 Coherent Quadriphase-Shift Keying (QPSK)

Although it is a special case of M-ary PSK discussed in Section 10.8.8, we are discussing it separately here because of its importance. Quadriphase-shift keying (QPSK), is a bandwidth efficient band pass digital modulation scheme that makes use of quadrature multiplexing. Just like in BPSK, in QPSK too, the message information is carried in the phase of the transmitted signal. In QPSK, any one of the four possible signals, which have four equally spaced values, $\pi/4$, $3\pi/4$, $5\pi/4$ and $7\pi/4$ for the carrier phase, is transmitted during the given symbol period of τ seconds. The QPSK signal set may be represented as

$$s_i(t) = \begin{cases} \sqrt{\dfrac{2E}{\tau}} \cos\left[2\pi f_c t + (2i-1)\dfrac{\pi}{4}\right]; & 0 \le t \le \tau \\ 0 \qquad ; & \text{otherwise} \end{cases} \tag{10.80}$$

Hence, $i = 1, 2, 3, 4$ and E is the signal energy in τ sec, the symbol duration. As usual, the carrier frequency f_c is so chosen that an integer number of carrier cycles are completed in the symbol period τ by making $f_c = \dfrac{n}{\tau}$, where n is a fixed integer.

In the baseband binary bit-stream consisting of 1s and 0s, two adjacent digits will have one of the four possible ways of ordering of the digits. Each such pair of bits is called a 'dibit'. These four 'dibits' may be ordered as per the 'Gray Code' in which two consecutive dibits will differ in one binary digit only, as shown in Table 10.6. The four phase values of the QPSK signal are represented by these Gray-encoded dibits.

Table 10.6 *Dibits, QPSK phase, and signal coordinates*

Signal	Gray Encoded Dibits	Phase of Qpsk Signal in Radians	Signal Coordinates along $\phi_1(T)$ and $\phi_2(T)$
$s_1(t)$	1 1	$\pi/4$	$\sqrt{E/2}, \sqrt{E/2}$
$s_1(t)$	0 1	$3\pi/4$	$-\sqrt{E/2}, \sqrt{E/2}$
$s_1(t)$	0 0	$5\pi/4$	$-\sqrt{E/2}, -\sqrt{E/2}$
$s_1(t)$	1 0	$7\pi/4$	$\sqrt{E/2}, -\sqrt{E/2}$

10.6.4 Signal Space and QPSK Signal Constellation

The signal set of Eq. (10.80) may conveniently be rewritten as

$$\underset{i=1,2,3,4}{s_i(t)} = \begin{cases} \sqrt{\dfrac{2E}{\tau}} \cos 2\pi f_c t \cos(2i-1)\dfrac{\pi}{4} - \sqrt{\dfrac{2E}{\tau}} \sin 2\pi f_c t \sin(2i-1)\dfrac{\pi}{4}; & 0 \le t \le \tau \\ 0 & ; \text{ otherwise} \end{cases} \tag{10.81}$$

For the signal space containing this signal set, we may conveniently choose the following as ***basis signals***. Being cosine and *sine* functions, they are obviously orthogonal and therefore form an orthogonal basis set. The corresponding orthonormal basis signals are

$$\left.\begin{aligned} \phi_1(t) &= \sqrt{\frac{2}{\tau}} \cos \omega_c t \\ \text{and} \qquad \phi_2(t) &= -\sqrt{\frac{2}{\tau}} \sin \omega_c t \end{aligned}\right\} \tag{10.82}$$

Now,

$$s_1(t) = \sqrt{\frac{2E}{\tau}} \cos 2\pi f_c t \cos\frac{\pi}{4} - \sqrt{\frac{2E}{\tau}} \sin 2\pi f_c t \sin\frac{\pi}{4}$$

The coordinates of this $s_1(t)$ along $\phi_1(t)$ and $\phi_2(t)$ are:

$$(s_1(t), \phi_1(t)) = \int_0^\tau s_1(t)\phi_1(t)dt = \sqrt{\frac{E}{2}} \text{ along } \phi_1(t)$$

and

$$(s_1(t), \phi_2(t)) = \int_0^\tau s_1(t)\phi_2(t)dt = \sqrt{\frac{E}{2}} \text{ along } \phi_2(t)$$

The coordinates of $s_2(t)$, $s_3(t)$ and $s_4(t)$ along $\phi_1(t)$ and $\phi_2(t)$ can also be found in a similar way. These are all listed in the last column of Table 10.6. Figure 10.40 shows the location of these signals $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$ in the signal space spanned by the basis signals $\phi_1(t)$ and $\phi_2(t)$.

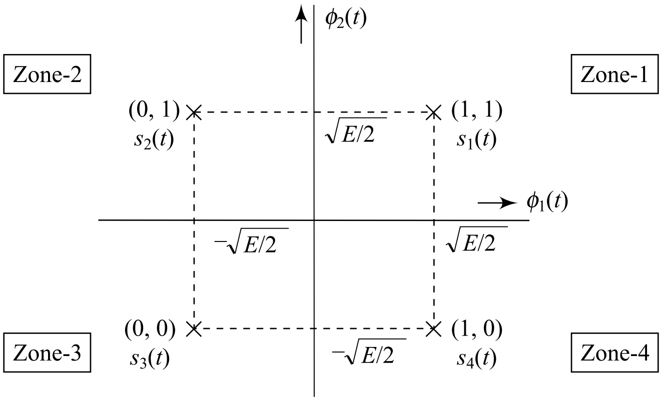


Fig. 10.40 *Signal space and signal constellation for QPSK*

We find that the points representing the signals $s_1(t)$ in the signal space is lying in the first quadrant, the one representing $s_2(t)$ is in the second quadrant, the one representing $s_3(t)$ is in the third quadrant and the one representing $s_4(t)$ is in the fourth quadrant. Thus, our signal space is two dimensional and the decision boundaries are the quadrant boundaries themselves. We therefore mark the quadrant in which signal $s_i(t)$ is laying as zone $-i$, $i = 1, 2, 3, 4$.

10.6.5 Generation and Detection of QPSK Signals

Methods used for generation and detection of QPSK signals are directly based on the equivalence of a QPSK system to two BPSK systems – the inphase channel and the quadrature channel, operating in parallel, as discussed earlier.

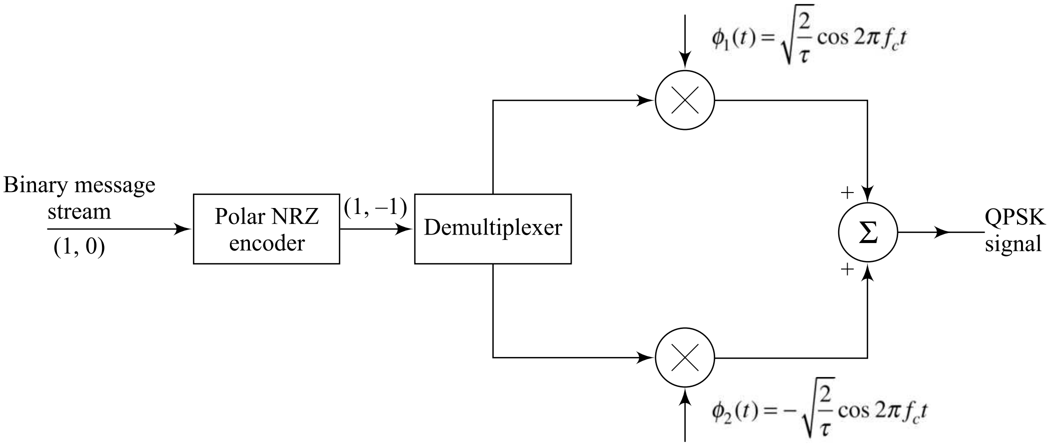


Fig. 10.41 *QPSK transmitter*

The message binary bit stream of 1s and 0s is encoded into a polar non-return-to-zero bit stream of 1s and –1s. The demultiplexer segregates the odd and even indexed polar binary digits, routing all odd-indexed bits through the upper inphase channel where they modulate the inphase carrier $\phi_1(t)$. The demultiplexer routes all even-indexed polar binary digits through the lower quadrature channel where, by a process of multiplication, they modulate the inphase carrier $\phi_2(t)$. The inphase carrier BPSK signal produced in the upper arm and the quadrature carrier BPSK signal produced in the lower arm are fed to an adder and we get the QPSK signal from its output.

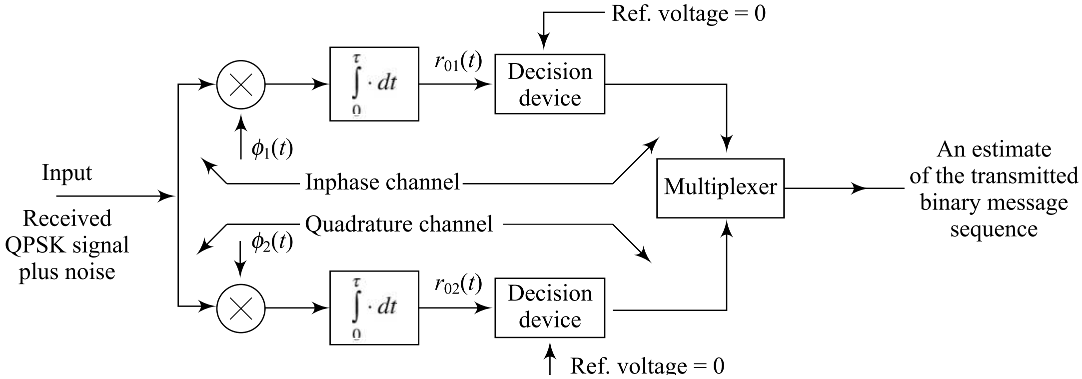


Fig. 10.42 *A coherent QPSK receiver*

The inphase and quadrature channels of the QPSK coherent receiver are typical BPSK coherent receivers. Locally generated carrier signals $\phi_1(t)$ and $\phi_2(t)$ are supplied to the correlators of the inphase and quadrature channels respectively. Although the received QPSK signal is applied to both the channels of the receiver, the inphase channel correlator produces zero output for all bits with quadrature carrier modulation. Similarly, the quadrature channel correlator produces zero output for all bits with inphase carrier modulation. $r_{01}(\tau)$ is the output of the inphase channel correlator for an inphase carrier modulated bit. Similarly, $r_{02}(\tau)$ is the output of the quadrature channel correlator for a quadrature carrier-modulated bit. If $r_{01}(\tau) > 0$ the inphase channel decision device decides it is a 1 and if $r_{01}(\tau) < 0$, it decides, it is a binary 0 and it outputs a binary digit accordingly. Similarly, if $r_{02}(\tau) > 0$ the quadrature channel's decision device decides that it is a binary 1 and if $r_{02}(\tau) < 0$, it decides that it is a binary 0 and outputs a binary 0. These binary digits from the outputs of the decision devices in the two channels are then multiplexed to obtain an estimate of the transmitted binary message sequence with minimum possible probability of error (for AWGN) since it is correlation reception.

10.6.6 Power Spectrum of QPSK Signal

We shall derive the PSD of a QPSK signal under the following assumptions:

1. In the message binary sequence, the binary digits 0 and 1 are equally probable.
2. Adjacent transmitted symbols are statistically independent.

Referring to Eq. (10.81), since

$$\cos\left[(2i-1)\frac{\pi}{4}\right] = \pm 1/\sqrt{2} \quad \text{for } i = 1, 2, 3, 4$$

the equivalent low pass (or baseband) signal corresponding to either the inphase component, or the quadrature component of the QPSK signal will be a random binary sequence

$$g(t) = \begin{cases} \pm\sqrt{\dfrac{E}{\tau}}; & 0 \le t \le \tau \\ 0 \quad ; & \text{otherwise} \end{cases} \tag{10.83}$$

where E is the symbol energy and τ is the symbol duration. As usual, we shall find the PSD of the low pass signal $g(t)$ and use it for sketching the PSD of the band pass QPSK signal by appropriate frequency translations. First we note that

$$g(t) = \pm\sqrt{\frac{E}{\tau}}\,\Pi\left(\frac{t-\tau/2}{\tau}\right)$$

The PSD of this random binary sequence may be written as

$$G(f) = \frac{E[F\cdot T\cdot \text{of } g(t)]^2}{\tau} \tag{10.84}$$

where E stands for 'expectation'.

$$\therefore \qquad G(f) = \frac{\left[\sqrt{\dfrac{E}{\tau}}\cdot\tau\cdot\text{sinc}f\tau\right]^2}{\tau}$$

$$= E\,\text{sinc}^2 f\tau \tag{10.85}$$

This is the PSD of the low pass equivalent of either the inphase component, or, the quadrature component of the QPSK signal. However, since the inphase and quadrature components are statistically independent, the power spectrum $P(f)$ of the low pass equivalent of the QPSK signal will be the sum of the PSDs of the low pass equivalents of the two components. Thus,

$$P(f) = 2E\,\text{sinc}^2(f\tau)$$

$$= 4E_b\,\text{sinc}^2(2fT) \tag{10.86}$$

where E_b is the bit energy and T is the bit duration.

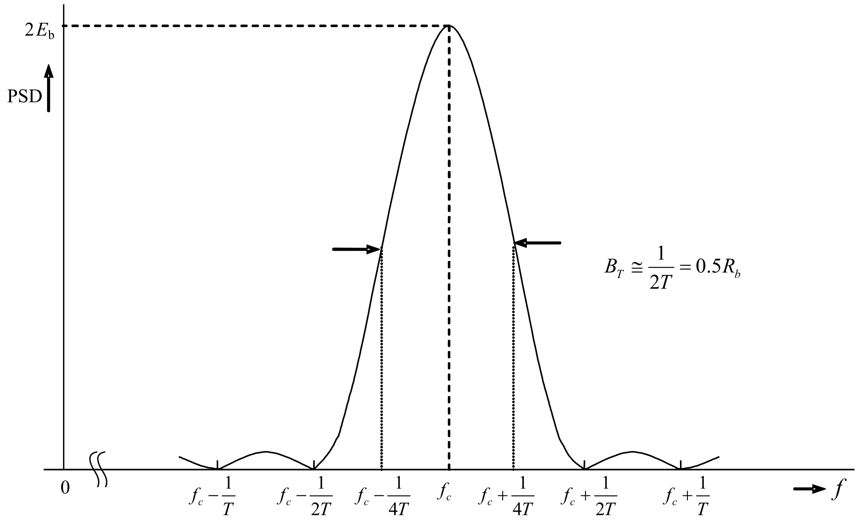


Fig. 10.43 *Power spectrum of a QPSK signal*

10.6.7 Non-coherent QPSK

A non-coherent QPSK is possible that makes use of differential encoding at the transmitter and phase-comparison method of detection at the receiver, just like in DPSK. However, a differentially encoded QPSK (DQPSK) requires about 2.3 dB more signal energy than the coherent QPSK for any specified probability of error.

10.6.8 Offset QPSK (OQPSK)

This is a variant of QPSK in which the bit-stream for the quadrature component is delayed by half a symbol duration with respect to the inphase component bit-stream. Thus the two orthonormal basis signals for the OQPSK are:

$$\phi_1(t) = \sqrt{\frac{2}{\tau}} \cos 2\pi f_c t;\ 0 \le t \le \tau$$

$$\phi_2(t) = \sqrt{\frac{2}{\tau}} \sin 2\pi f_c t;\ \tau/2 \le t \le 3\tau/2 \tag{10.87}$$

The need for this $T/2$ offset arises from the fact that in the QPSK, sign changes of the inphase and quadrature components cause sudden carrier phase changes. These phase changes of the carrier may be by ±90°, or sometimes even by ±180°, depending upon whether the sign change occurs only for one of the two components, or for both. As can be seen from the signal constellation of QPSK, whenever adjacent dibits in the binary message sequence differ only in one of the digits (like 1, 1 and 0, 1), the transition from the message point corresponding to the first dibit to the message point corresponding to the next dibit of the message sequence involves a change of carrier phase by only ±90°. On the other hand, if the adjacent dibits of the message sequence differ in both the binary digits (like 0, 0 and 1, 1 or 1 0 and 0 1) the transition from the message point corresponding to the first dibit to the message point corresponding to the next dibit of the message sequence will involve a change of carrier phase by ±180°.

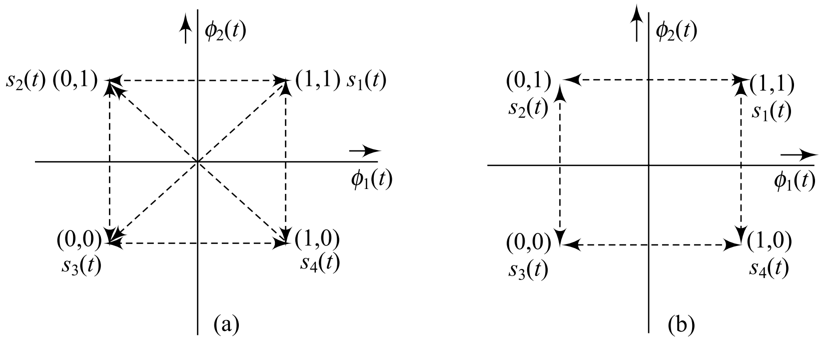


Fig. 10.44 *(a) Possible transitions in QPSK, (b) Possible transitions in offset QPSK*

Such sudden phase changes in the carrier can result in reduction of the amplitude of the QPSK signal when it is filtered. So, if the QPSK signal, during the course of its transmission, is passed through a filter before the signal is detected, the resulting amplitude reduction of the signal can lead to errors in detection. Carrier phase changes by ±180° in particular, cause considerable reduction in the envelope amplitude and so are to be avoided.

In OQPSK, because of the offset, carrier phase changes are confined to ±90° only and so the extent of amplitude changes is reduced, thus reducing the probability of occurrence of symbol errors in the detection process.

Theoretically, the average probability of symbol error is exactly the same for QPSK and OQPSK for coherent detection.

10.6.9 $\pi/4$-Shifted QPSK

As pointed out in our discussion on offset QPSK, the sudden carrier phase changes of ±90° and ±180° that can take place in ordinary QPSK lead to sudden reduction in the amplitude of the envelope of the carrier and these can cause decoding errors. Offset QPSK, as we have seen already, tries to reduce the occurrence of such decoding errors by not allowing ±180° phase changes of the carrier due to transition from one message point to another. However, ±90° phase changes do occur in the offset QPSK.

$\pi/4$-shifted QPSK is yet another variant of QPSK and it goes one step further compared to offset QPSK by not allowing even ±90° carrier phase changes to occur. The phase transition from one message point to the next in the case of $\pi/4$ shifted QPSK are restricted only to $\pm\pi/4$ and $\pm3\pi/4$ radians. This it does by the following method:

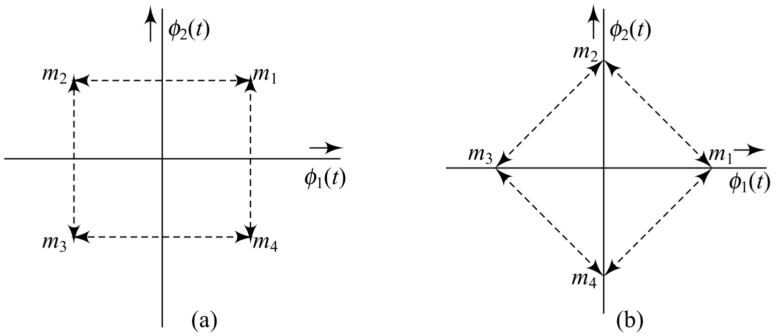


Fig. 10.45 *Two QPSK constellations*

Two commonly used QPSK constellations are shown in (a) and (b) of Fig. 10.45. The carrier phase transmitted corresponding to the first dibit of the binary baseband message sequence is say as at message point m_1 in the first constellation. For the next dibit of the message sequence, the carrier phase transmitted will be as per the second constellation – it may correspond to m_1, m_2, m_3 or m_4 in that constellation, and depends on the message dibit. The third dibit will be transmitted as per the first constellation like this, the carrier phase is chosen alternatively from one constellation and then the other. Thus in $\pi/4$ QPSK, there will be eight possible phase-states, as shown in Fig. 10.46.

1. Note that from each state there are only four possible transitions, and all of them involve a phase change of $\pm\pi/4$, or $\pm 3\pi/4$ radians and not $\pm\pi/2$ or $\pm\pi$.
2. $\pi/4$-shifted QPSK signals, unlike offset QPSK signals, lend themselves to non-coherent detection. Also, they can be differentially encoded, like the QPSK signals.

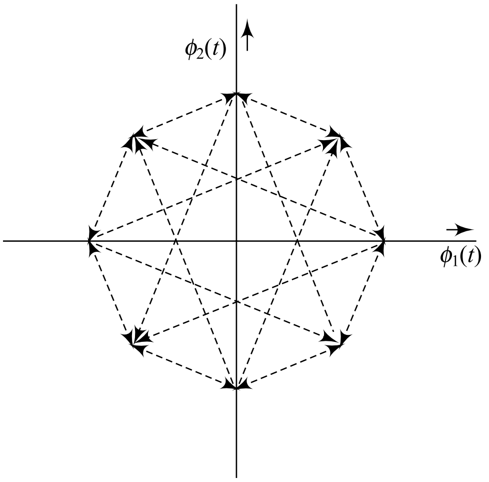


Fig. 10.46 *Possible phase-states and transitions for $\pi/4$-shifted QPSK*

Example 10.8 Given the input binary sequence 1100100010, sketch the waveforms of the inphase and quadrature components of the modulated wave as well as the modulated wave obtained by QPSK, based on the signal set shown in Fig. 10.40.

Solution

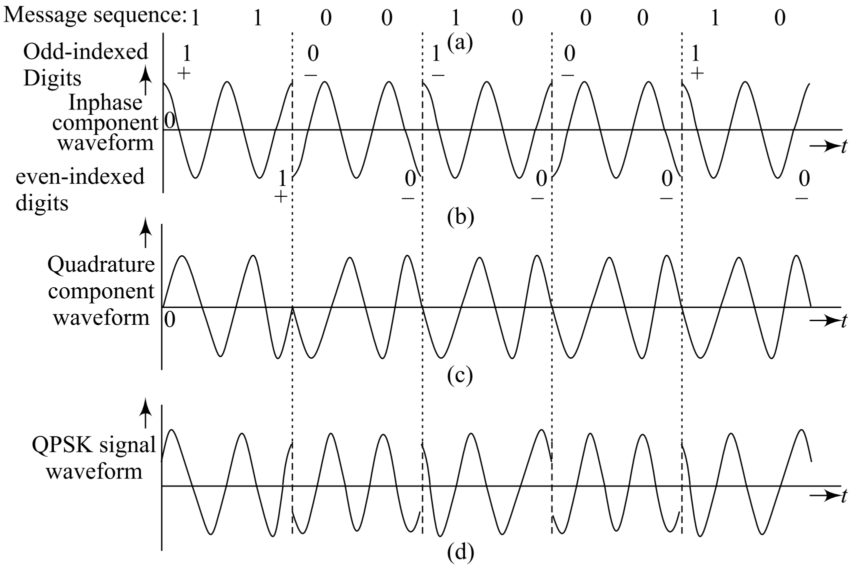


Fig. 10.47 *(a) Given sequence, (b) Inphase component waveform, (c) Quadrature component waveform, (d) QPSK signal*

10.7 COHERENT MINIMUM SHIFT KEYING (MSK)

Minimum shift keying (MSK) is a type of continuous phase FSK that produces orthogonal signaling and is bandwidth efficient. It gives a probability of error that is as good as that of binary PSK and QPSK. A variant of this, called Gaussian Minimum Shift Keying (GMSK) is extensively used in GSM cellular mobile communications.

Coherent BFSK receivers use the phase information contained in the received signal only for synchronization but not for improved detection. In MSK, the phase information of the received signal is made use of in the detection process to improve the noise performance. In Section 10.5.1, under the methods of generation of BFSK signals, it was pointed out that the method making use of switching in of oscillators of frequencies $(f_c + f_d)$ and $(f_c - f_d)$ corresponding to baseband message digits 1 and 0 respectively, will result in a BFSK signal with discontinuities at the switching instants. To avoid these discontinuities and the resultant wastage of bandwidth, it was suggested that frequency modulating a single carrier oscillator using the baseband polar binary data stream could be resorted to. Let us now examine the effect of such a frequency modulation on the phase of the signal. For this purpose, let us represent the frequency modulated signal as:

$$s(t) = \sqrt{\frac{2E_b}{T}} \cos[2\pi f_c t + \phi(t) + \phi(0)] \tag{10.88}$$

where, E_b is the energy of the signal in one bit-duration T, $\phi(0)$ is the initial phase of the signal at $t = 0$ and $\phi(t)$ is the phase contributed by the modulation and is given by:

$$\phi(t) = 2\pi k_f \int_{\lambda=0}^{t} x(\lambda) d\lambda \tag{10.89}$$

In the above equation, $x(t)$ is the modulating signal. In our case the modulating signal is the baseband binary (polar) sequence in which

$$x(t) = \begin{cases} +A \text{ if } b_k = 1 \\ -A \text{ if } b_k = 0 \end{cases} \quad \text{for } 0 \le t \le T \tag{10.90}$$

$$\therefore \qquad \phi(t) = 2\pi k_f \int_{\lambda=0}^{t} \pm A d\lambda = \pm 2\pi k_f At; 0 \le t \le T \tag{10.91}$$

where, k_f is the frequency deviation constant.

$\therefore$ $\phi(t)$ increases (or decreases) linearly with time t during the bit duration, commencing from a value of zero.

The instantaneous frequency $f_i(t)$ at the instant t, is given, as usual by

$$\begin{aligned} f_i(t) &= \frac{1}{2\pi}\frac{d}{dt}[2\pi f_c t + \phi(t) + \phi(0)] \\ &= f_c \pm k_f \cdot A; 0 \le t \le T \end{aligned} \tag{10.92}$$

(using Eq. 10.91 and noting that $\phi(0)$ is a constant)

Hence $\qquad f_1 = f_c - k_f A; b_k = 0$ i.e., $s_1(t)$ is transmitted

and $\qquad f_2 = f_c + k_f A; b_k = 1$ i.e; $s_2(t)$ is transmitted $\qquad$ (10.93)

$$\therefore \qquad (f_2 - f_1) = 2k_f \cdot A \tag{10.94}$$

Let us for convenience introduce a parameter h which we shall define later, and which is such that

$$(f_2 - f_1) = 2k_f \cdot A = \frac{h}{T} \tag{10.95}$$

Since $2k_f A$ represents frequency deviation,

$$h = 2k_f \cdot AT = (2k_f \cdot A) / R_b = (f_2 - f_1)T = (f_2 - f_1) / R_b \tag{10.96}$$

Represents the normalized frequency deviation, normalized with respect to the bit-rate, and is therefore called the deviation ratio. ***So, the dimensionless parameter, h, represents ($f_2 - f_1$) normalized with respect to the bit-rate.***

If $s_1(t)$ and $s_2(t)$ are the signals transmitted corresponding to binary baseband symbols 0 and 1 respectively, then

$$s_1(t) = \sqrt{\frac{2E_b}{T}} \cos[2\pi f_1 t + \phi(0)];\ 0 \le t \le T$$

and
$$s_2(t) = \sqrt{\frac{2E_b}{T}} \cos[2\pi f_2 t + \phi(0)];\ 0 \le t \le T \tag{10.97}$$

where,

$$f_1 = f_c - k_f \cdot A = f_c - \frac{h}{2T} \quad \text{(from Eq.10.95)}$$

and
$$f_2 = f_c + k_f \cdot A = f_c + \frac{h}{2T} \tag{10.98}$$

Now, in Eq. (10.88), if we put

$$\theta(t) = \phi(t) + \phi(0) \tag{10.99}$$

Then using Eqs. (10.91) and (10.95) we may write

$$\theta(t) - \phi(0) = \phi(t) = \pm 2\pi k_f A t$$

$$= \pm \pi t \cdot \frac{h}{T};\quad \left.\begin{matrix} +\text{ sign for } b_k = 1 \\ -\text{ sign for } b_k = 0 \end{matrix}\right\} \tag{10.100}$$

10.7.1 Phase Tree or Phase Trellis

From Eq. (10.100), at $t = T$, we may write:

$$\theta(T) - \phi(0) = \begin{cases} +\pi h & \text{if } b_k = 1 \\ -\pi h & \text{if } b_k = 0 \end{cases} \tag{10.101}$$

Equation (10.101) tells us that the phase of a CPFSK signal advances by πh radians when a binary 1 is transmitted and retards by πh radians when a binary 0 is transmitted. Thus, $[\theta(t) - \phi(0)]$ when plotted against t for a given baseband binary sequence, will trace out a particular path made up of a sequence of straight lines with positive and negative slopes. When we plot all such possible paths (corresponding to different binary baseband sequences), we get a tree-like structure as shown in Fig. 10.48. This is generally referred to as the ***phase tree*** or '***phase trellis***'. Fig. 10.48 shows a phase trellis.

Earlier, while discussing BFSK, we had talked of one type of continuous phase BFSK called Sunde's BFSK, in which f_1 and f_2 were given by (refer to Eqs. (10.63) and (10.64))

$$\left.\begin{aligned} f_1 &= f_c - \frac{1}{2T} \\ \text{and}\quad f_2 &= f_c + \frac{1}{2T} \end{aligned}\right\} \tag{10.102}$$

And in which the two signals, $s_1(t)$ with frequency f_1 and $s_2(t)$ with frequency f_2, are orthogonal over $[0, T]$.

Comparing RHS of Eq. (10.98) with the RHS of Eq. (10.102), we find that Sunde's BFSK signals are indeed only a special case of CPBFSK signals that we are discussing and that $h = 1$ for them. But, as one can see from Eq. (10.101), when the frequency deviation ratio, h, is equal to unity, the phase change

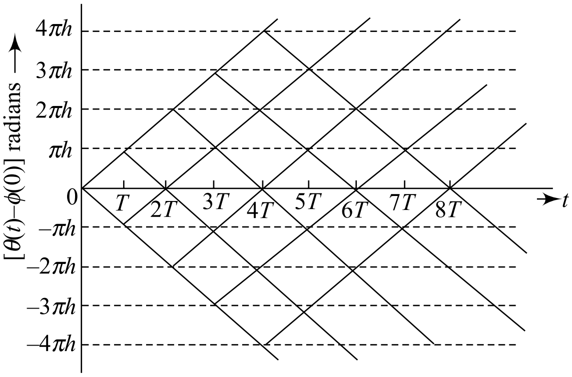


Fig. 10.48 *A phase trellis*

at the end of each bit duration T will be either $+\pi$ or $-\pi$ depending on whether $b_k = 1$ or 0, respectively. Since there is no difference between $+\pi$ and $-\pi$, it amounts to there being no memory for Sunde's FSK in the sense that a knowledge of what phase change has occurred in the previous bit interval will not be of any use in the present bit interval. However, if we take $h = 1/2$, corresponding to

$$\left.\begin{aligned} f_1 &= f_c - \frac{1}{4T} \\ \text{and} \qquad f_2 &= f_c + \frac{1}{4T} \end{aligned}\right\} \qquad (10.103)$$

then not only will the two CPBFSK signal $s_1(t)$ and $s_2(t)$ be orthogonal, the phase change in each bit-interval will also be either $+\pi/2$ or $-\pi/2$ (refer to Eq. (10.101)), which are clearly distinguishable. ***Further, h = 1/2 represents the minimum frequency deviation ratio for which the two signals, $s_1(t)$ and $s_2(t)$ will be orthogonal. It is for this reason that the CPBFSK with h = 1/2 is called the Minimum Shift Keying' or, MSK***.

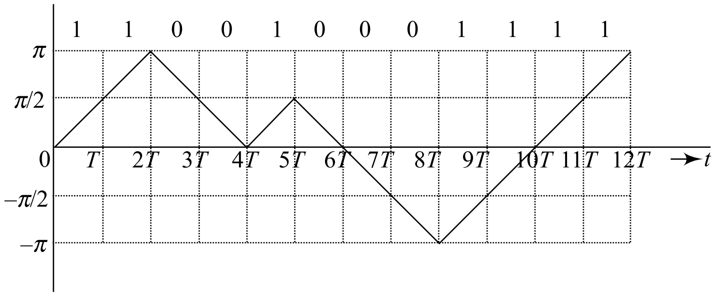


Fig. 10.49 *Phase path of the sequence 110010001111 with h = 1/2*

10.7.2 Signal Space of MSK

We have noted that with $h = 1/2$, the phase of the signal $s(t)$, the MSK signal, changes by $\pm\pi/2$ during each period T in a linear fashion depending on whether the baseband digit transmitted during that bit interval T is a 1 or a 0. ***Since the phase changes by ±π/2 for each T, and since there is no difference between +π and −π, the phase can be either 0 or π radians for even multiples of T and ±π/2 for odd multiples of T***.

Denoting $[\phi(t) + \phi(0)]$ by $\theta(t)$ as indicated in Eq. (10.99), we may rewrite Eq. (10.88) as:

$$\begin{aligned} s(t) &= \sqrt{\frac{2E_b}{T}} \cos[2\pi f_c t + \phi(t) + \phi(0)] = \sqrt{\frac{2E_b}{T}} \cos[2\pi f_c t + \theta(t)] \\ &= \sqrt{\frac{2E_b}{T}} [\cos\theta(t) \cdot \cos 2\pi f_c t - \sin\theta(t) \cdot \sin 2\pi f_c t] \end{aligned} \qquad (10.104)$$

The first term in the RHS of the above equation is the inphase term and the second one is the quadrature term. First, let us examine the inphase component given by

$$s_I(t) = \sqrt{\frac{2E_b}{T}} \cos\theta(t) \qquad (10.105)$$

In this,

$$\cos\theta(t) = \cos[\phi(t) + \phi(0)]$$

But from Eq. (10.100)

$$\begin{aligned} \theta(t) &= \phi(0) \pm \left(\frac{\pi h}{T}\right)t; 0 \le t \le T \quad \text{with} \begin{cases} + \text{ sign for } b_k = 1 \\ - \text{ sign for } b_k = 0 \end{cases} \\ &= \phi(0) \pm \left(\frac{\pi}{2T}\right)t; 0 \le t \le T \text{ for MSK} \end{aligned} \qquad (10.106)$$

Now, it can easily be shown that throughout the time interval $-T \le t \le T$, the *polarity* of $\cos\theta(t)$ depends only on the value of $\phi(0)$ and is independent of the binary sequence transmitted both prior to $t = 0$ and after $t = 0$. This may be demonstrated as follows.

Since $\phi(0) = 0$ or π, let us consider both the cases.

1. **$\phi(0) = 0$**: $\phi(t)$ varies from 0 to $\pm\pi/2$ in a linear way in the interval 0 to T. Throughout this interval, $\theta(t)$, which is $\phi(t) + \phi(0)$, varies between 0 and $\pm\pi/2$ (positive sign for $b_k = 1$ and negative sign for $b_k = 0$). But, when θ varies between 0 to $+\pi/2$ or 0 to $-\pi/2$, $\cos\theta(t)$ will be non-negative throughout.
2. **$\phi(0) = \pi$**: Since $\phi(t)$ varies from 0 to $\pm\pi/2$, and since $\theta(t) = \phi(t) + \phi(0)$, $\theta(t)$ will vary from π to $3\pi/2$ or π to $\pi/2$. In both the cases, $\cos\theta(t)$ will be non-positive throughout.

Thus, the sign of $\cos\theta(t)$ in the interval 0 to T depends only on the value of $\phi(0)$ but not on whether what was transmitted in that bit-interval was a binary 1 or a binary 0. By arguing out in a similar way, it can be shown that the sign of $\cos\theta(t)$ in the interval $-T$ to 0 also, is independent of the binary digit transmitted in that bit-interval and that it depends only on the value of $\phi(0)$.

From Eq. (10.106), we have

$$\cos\theta(t) = \cos\left[\phi(0) \pm \left(\frac{\pi}{2T}t\right)\right]$$

$$= \cos\phi(0)\cos\left(\frac{\pi}{2T}\right)t \mp \sin\phi(0)\sin\left(\frac{\pi}{2T}\right)t$$

Since $\phi(0) = 0$ or π, $\quad \sin\phi(0) = 0$

$$\therefore \quad \cos\theta(t) = \cos\phi(0)\cos\left(\frac{\pi}{2T}\right)t = \pm\cos\left(\frac{\pi}{2T}\right)t$$

$\therefore$ the inphase component, $S_I(t)$ is given by (refer Eq. (10.105))

$$s_I(t) = \sqrt{\frac{2E_b}{T}}\cos\phi(0)\cos\left(\frac{\pi}{2T}\right)t$$

$$\therefore \quad s_I(t) = \pm\sqrt{\frac{2E_b}{T}}\cos\left(\frac{\pi}{2T}\right)t; \quad -T \le t \le T \tag{10.107}$$

The above equation is valid over the interval $-T \le t \le T$ because, as we had discussed earlier, over the entire interval $-T \le t \le T$, $\cos\theta(t)$ depends only on the value of $\phi(0)$. In *Eq. (10.107), the plus sign corresponds to the 0 value for $\phi(0)$ and the minus sign corresponds to the value π for $\phi(0)$*. Also, we find that the signal $s_I(t)$ is just a *single cosine pulse*, i.e., a half-cycle cosine pulse.

In a similar manner, we can show that the quadrature component of $s(t)$, viz., $s_Q(t)$ is given by

$$s_Q(t) = \sqrt{\frac{2E_b}{T}}\sin\phi(0)\sin\left(\frac{\pi t}{2T}\right)$$

$$= \pm\sqrt{\frac{2E_b}{T}}\sin\left(\frac{\pi}{2T}\right)t; \quad 0 \le t \le 2T \tag{10.108}$$

where the plus sign has to be taken if $\phi(T) = \pi/2$ and the minus sign if $\phi(T) = -\pi/2$. Note that the quadrature component of $s(t)$ is a *single sinusoidal pulse*, i.e., a half-cycle sinusoidal pulse. Also, just as the sign of $\cos\theta(t)$, over the interval $-T \le t \le T$ is dependent solely on the value of $\phi(0)$, the value of $\sin\theta(t)$ over the interval $0 \le t \le 2T$, is dependent solely on the value of $\phi(T)$. Thus, the following four possibilities arise:

1. $\phi(0) = 0$ and $\phi(T) = \pi/2$ (corresponding to transmission of $b_k = 1$ in the interval $0 \le t \le T$
2. $\phi(0) = \pi$ and $\phi(T) = \pi/2$, which corresponds to transmission of $b_k = 0$ in the interval $0 \le t \le T$, as the phase got reduced from π at $t = 0$ to $\pi/2$ at $t = T$.

3. $\phi(0) = \pi$ and $\phi(T) = -\pi/2,$ which corresponds to transmission of $b_k = 1$ in the interval $0 \le t \le T,$ since $-\frac{\pi}{2} = \frac{3\pi}{2}$, i.e., phase has increased from π at $t = 0$ to $\frac{3\pi}{2}$ at $t = T$.
4. $\phi(0) = 0$ and $\phi(T) = -\pi/2,$ which corresponds to the transmission of binary 0 over $0 \le t \le T$.

Thus, $s(t)$, the MSK signal itself can take any one of four possible forms corresponding to the four possible phase state combinations. This, again, means that the signal space of the MSK signal is going to be two dimensional. We may, for the sake of drawing the signal constellation, take the following two orthonormal signals as the basis set for this two-dimensional signal space.

$$g_1(t) = \sqrt{\frac{2}{T}} \cos\left(\frac{\pi}{2T}\right) t \cdot \cos 2\pi f_c t, 0 \le t \le T \tag{10.109}$$

$$g_2(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{\pi}{2T}\right) t \cdot \sin 2\pi f_c t, 0 \le t \le T \tag{10.110}$$

Note that the interval $0 \le t \le T$ is the common interval over which the expression given by Eq. (10.107) for $s_I(t)$ and Eq. (10.108) for $s_Q(t)$ are both valid.

We can now express $s(t)$, the MSK signal, as a linear combination of the two orthonormal basis signals $g_1(t)$ and $g_2(t)$ and write

$$s(t) = s_1 g_1(t) + s_2 g_2(t) \tag{10.111}$$

where s_1 and s_2 are constants representing the coordinates of $s(t)$ along $g_1(t)$ and $g_2(t)$ respectively. Hence, s_1 and s_2 can be found by taking the inner products of $s(t)$ with $\phi_1(t)$ and $\phi_2(t)$ respectively.

$$\therefore \quad s_1 = (s(t), g_1(t)) = \int_{-T}^{T} s(t) \cdot g_1(t) dt = \int_{-T}^{T} s_I(t) \cdot g_1(t) dt$$

$$= \int_{-T}^{T} \sqrt{\frac{2E_b}{T}} \cos \phi(0) \cdot \cos\left(\frac{\pi}{2T} t\right) \cdot \sqrt{\frac{2}{T}} \cos\left(\frac{\pi}{2T} t\right) dt$$

$$\therefore \quad s_1 = \sqrt{E_b} \cos \phi(0); -T \le t \le T \tag{10.112}$$

Similarly,

$$s_2 = (s(t), g_2(t)) = \int_{0}^{2T} s(t) \cdot g_2(t) dt$$

$$= -\sqrt{E_b} \sin \phi(T); 0 \le t \le 2T \tag{10.113}$$

The signal constellation of the MSK signal is shown in Fig. 10.50.

Although there is a similarity between the signal constellation of the QPSK system and the MSK system, one important difference exists and it must be taken note of. This is that in QPSK, each message point corresponded to one of the four dibits and hence, one of the four messages. ***In MSK, there are only two possible messages. However, two message points are shown for each of the possible messages. But only one of these two message points is used to represent the transmitted symbol or transmitted message at any one time, depending on the value of $\phi(0)$.***

The labeling $b_k = 0$ or $b_k = 1$ in each of the quadrants refers to the binary symbol transmitted during the bit interval $0 \le t \le T,$ corresponding to the message point in that quadrant. Message point m_1 corresponds to the phase states $\phi(0) = 0$ and $\phi(T) = -\pi/2$. Since the phase has reduced by $\pi/2$ radians, the transmitted bit in that interval is a 0. Similarly, message point m_2 corresponds to the phase states $\phi(0) = \pi$ and $\phi(T) = -\pi/2$. Since $-\pi/2$ is the same as $+\frac{3\pi}{2}$, it means the phase has increased from π to $\frac{3\pi}{2}$ over the interval 0 to T. Hence the transmitted bit in $0 \le t \le T$ is a 1.

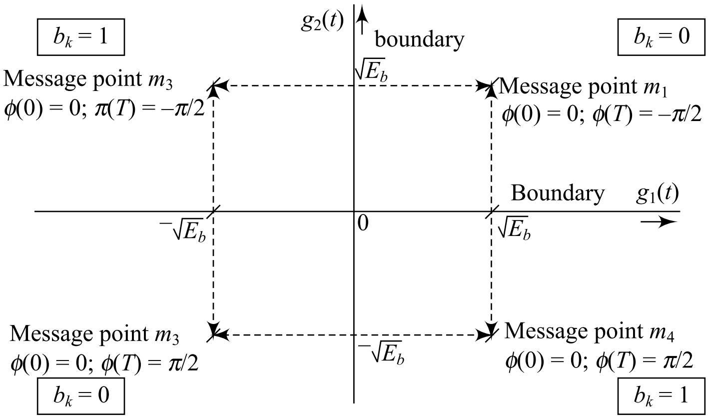


Fig. 10.50 *Signal space diagram of MSK showing the signal constellation*

10.7.3 Waveforms of the Inphase and Quadrature Components

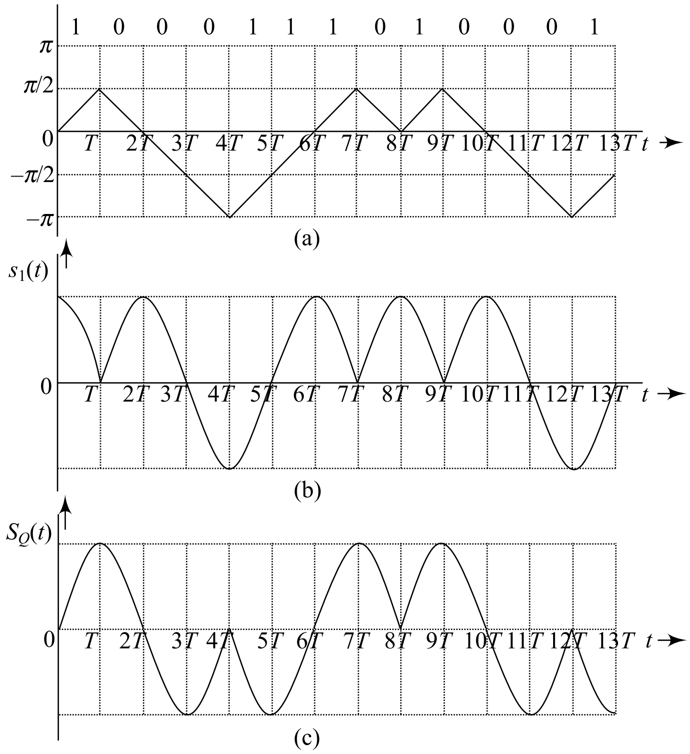


Fig. 10.51 *(a) Phase path for the message sequence 1000111010001, (b) Inphase component $s_I(t)$, (c) Quadrature component $s_Q(t)$*

10.7.4 Generation and Detection of MSK Signals

A method of generating MSK signals is shown in Fig. 10.52(a). It is directly based on Eq. (10.111). As shown in the figure, $g_1(t)$ and $g_2(t)$, the orthogonal basis signals given by Eqs. (10.109) and (10.110) are generated as follows:

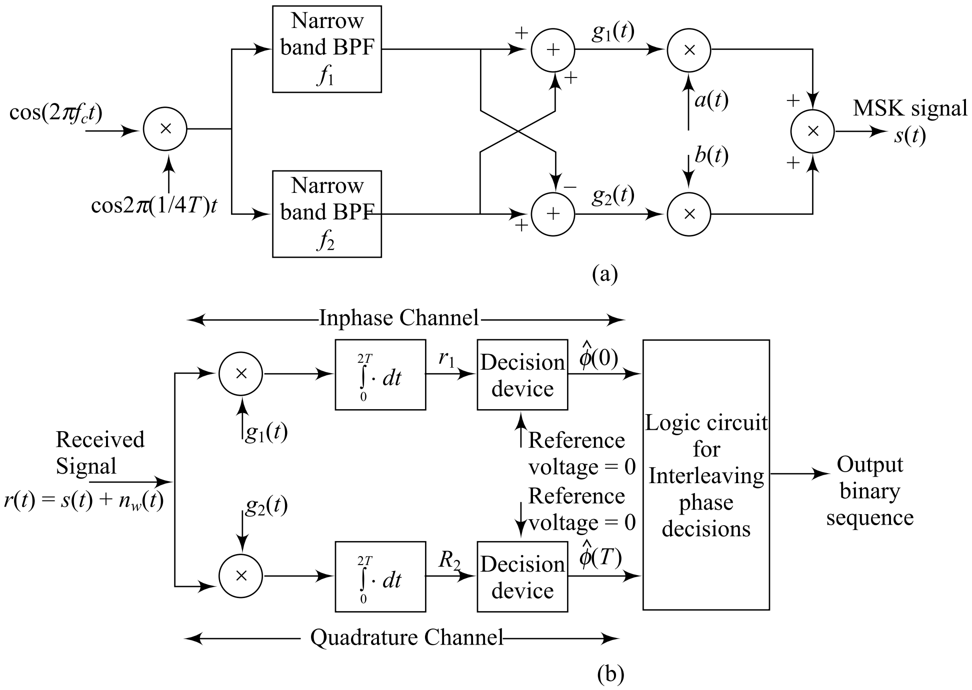


Fig. 10.52 *(a) MSK transmitter, (b) MSK receiver*

First, we multiply two cosinusoidal signals, one of frequency f_c and the other of frequency (1/4T) where T is the bit duration.

This multiplication gives rise to two phase coherent signals, one with the sum frequency, i.e., $\cos 2\pi(f_c + 1/4T)t$, and the other with the difference frequency, i.e., $\cos 2\pi(f_c - 1/4T)t$. These two signals are then separated by applying the output of the product device simultaneously to two very narrow band band pass filters, one centered on the difference frequency which is equal to f_1 (see Eq.10.102), and the other centered on the sum frequency which is equal to f_2 (see Eq. (10.102)). Then these two signals, one at a frequency f_1 and the other at a frequency f_2, are linearly combined as shown in the figure, to obtain $g_1(t)$ and $g_2(t)$. This is because,

$$\frac{1}{2}\left[\cos 2\pi\left(f_c - \frac{1}{4T}\right)t + \cos 2\pi\left(f_c + \frac{1}{4T}\right)t\right] = \cos 2\pi f_c t \cdot \cos 2\pi\left(\frac{1}{4T}\right)t$$
$$= g_1(t)$$

and
$$\frac{1}{2}\left[\cos 2\pi\left(f_c - \frac{1}{4T}\right)t - \cos 2\pi\left(f_c + \frac{1}{4T}\right)t\right] = \sin 2\pi f_c t \cdot \sin 2\pi\left(\frac{1}{4T}\right)t$$
$$= g_2(t)$$

Now, as per Eq. (10.111), we must multiply $g_1(t)$ by s_1 and $g_2(t)$ by s_2 where s_1 and s_2 are as given in Eqs. (10.112) and (10.113). Note that

$$\therefore \qquad s_1 = \sqrt{E_b}\cos\phi(0); \quad -T \le t \le T$$

where $\phi(0)$ is either 0 or π and correspondingly $\cos\phi(0)$ is either 1 or –1 depending on the phase state at $t = 0, 2T, 4T, \ldots$, which, as shown in Fig. 10.51(a), in turn, depends on the message sequence. So, to generate the sequence of values that s_1 takes at $t = 0, 2T, 4T, \ldots$, the phase values $\phi(0)$, $\phi(2T)$, $\phi(4T)$, … are extracted from the given message sequence. Since these phase values can only be either 0 or π radians, correspondingly s_1 will be either $+\sqrt{E_b}$ or $-\sqrt{E_b}$. Thus, the sequence of s_1 values will be in the form of a polar NRZ binary wave denoted, by $a(t)$ in Fig. 10.52(a), with a bit rate of $(1/2T)$. Also, $b(t)$ denotes the sequence of s_2 values. As per Eq. (10.111), the MSK signal $s(t)$ is then obtained by adding $s_1g_1(t)$ and $s_2g_2(t)$.

The block diagram of an MSK receiver is shown in Fig. 10.52(b). The orthonormal basis signals $g_1(t)$ and $g_2(t)$ are locally generated and are used for correlating the received signal with $g_1(t)$ in the inphase channel and with $g_2(t)$ in the quadrature channel to obtain r_1 and r_2 the observed random variables in the inphase and quadrature channels respectively. These are compared with the reference voltages, which are zero volts in the two channels, to obtain the phase estimates $\hat{\phi}(0)$ and $\hat{\phi}(T)$ which are then used by the logic circuit to produce the output binary baseband sequence with minimum P_e.

10.7.5 Power Spectra of MSK Signals

For determining the power spectrum, we shall assume, as we have always been doing, that the message band binary sequence is a random binary wave with equal probability of occurrence for 1s and 0s and that each symbol is statistically independent of those preceding it.

As we have already seen, the MSK signal is given by

$$s(t) = s_1g_1(t) + s_2g_2(t)$$

The low-frequency equivalent of the inphase component $s_1g_1(t)$ is given by

$$\begin{cases} +\sqrt{E_b}\cdot\sqrt{\dfrac{2}{T}}\cos\left(\dfrac{\pi}{2T}\right)t; & -T \le t \le T \text{ if } \phi(0) = 0 \\ 0; & \text{otherwise} \end{cases}$$

and

$$\begin{cases} -\sqrt{E_b}\cdot\sqrt{\dfrac{2}{T}}\cos\left(\dfrac{\pi}{2T}\right)t; & -T \le t \le T \text{ if } \phi(0) = \pi \\ 0; & \text{otherwise} \end{cases} \tag{10.114}$$

Similarly, the low frequency equivalent of the quadrature component is given by

$$\begin{cases} \sqrt{E_b}\cdot\sqrt{\dfrac{2}{T}}\sin\left(\dfrac{\pi}{2T}\right)t; & 0 \le t \le 2T, \text{ if } \phi(T) = -\pi/2 \\ 0; & \text{otherwise} \end{cases}$$

and

$$\begin{cases} -\sqrt{E_b}\cdot\sqrt{\dfrac{2}{T}}\sin\left(\dfrac{\pi}{2T}\right)t; & 0 \le t \le 2T \text{ if } \phi(T) = +\pi/2 \\ 0; & \text{otherwise} \end{cases} \tag{10.115}$$

Since the PSD of the inphase component (low frequency equivalent) is equal to the energy spectral density over a period of $2T$ divided by $2T$, it is given by

$$P_I(f) = \frac{\left[F\left(\left\{\sqrt{\dfrac{2E_b}{T}}\cos\left(\dfrac{\pi}{2T}\right)t\right\}\{\Pi(t/2T)\}\right)\right]^2}{2T}$$

$$= \frac{16E_b}{\pi^2}\left[\frac{\cos(2\pi fT)}{(4fT)^2 - 1}\right]^2 \tag{10.116}$$

The PSD of the low-frequency equivalent of the quadrature component also is the same as that of the inphase component. But since the inphase and quadrature components are statistically independent, the PSD of their sum is equal to the sum of their individual PSDs. Thus, the baseband PSD of the MSK signal is given by

$$\underset{\text{MSK}}{P(f)} = 2P_I(f) = \frac{32E_b}{\pi^2}\left[\frac{\cos(2\pi fT)}{16f^2T^2 - 1}\right]^2 \tag{10.117}$$

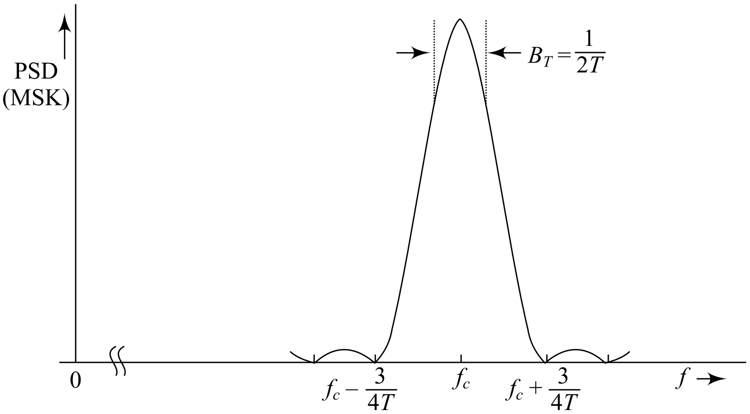


Fig. 10.53 *Power spectrum of an MSK signal*

As can be seen from the expression for the power spectrum, the PSD of MSK signal falls off as the inverse fourth power of frequency. It may be recalled that in the case of QPSK, the PSD is proportional to *sinc*^{2}(*fT*) and therefore falls of as the inverse square of the frequency. *Thus, the interference caused outside the signal band of interest is less in the case of MSK as compared to QPSK.*

10.7.6 Gaussian MSK (GMSK)

Although the power spectral density of MSK falls off as the fourth power of the frequency, it is still not good enough for multiuser wireless communication applications, which require very stringent standards to be maintained with regard to adjacent channel interference.

To meet those stringent requirements, the PSD of an MSK signal is modified by passing the NRZ binary data stream through a pulse-shaping low pass filter before carrier modulation. While having a narrow bandwidth together with sharp cut-off characteristic, this LPF has to have negligible overshoot in its impulse response. Further, it should permit the final modulated signal in that it has a carrier phase of 0 or π radians at even multiples of T and $+\pi/2$ or $-\pi/2$ radians at odd multiples of T.

All the above stated requirements are met by an LPF whose frequency response is having the shape of a Gaussian function. Because of the use of such a Gaussian filter, the modified MSK generated by using this filter, is called as 'Gaussian-filtered MSK', or Gaussian MSK (GMSK).

As mentioned earlier, the Gaussian filter operates at baseband level. So, if we take the 3-dB baseband bandwidth of this filter to be W Hz, its transfer function is given by

$$H(f) = \exp\left[-\frac{\log 2}{2}\left(\frac{f}{W}\right)^2\right] \tag{10.118}$$

and the corresponding impulse response, $h(t)$ is given by

$$h(t) = \sqrt{\frac{2\pi}{\log 2}}\, W \exp\left[-\frac{2\pi^2}{\log 2}W^2t^2\right] \tag{10.119}$$

Since this pulse-shaping Gaussian filter acts on the input binary data which has been encoded in the form of polar NRZ waveforms, let us consider its response to a unit amplitude rectangular pulse of duration *T* sec centered on the origin. This response works out to

$$g(t) = \frac{1}{2}\left[erfc\left\{\pi\sqrt{\frac{2}{\log 2}}WT\left(\frac{t}{T}-\frac{1}{2}\right)\right\} - erfc\left\{\pi\sqrt{\frac{2}{\log 2}}WT\left(\frac{t}{T}+\frac{1}{2}\right)\right\}\right] \tag{10.120}$$

Thus, the time-bandwidth product *WT* plays an important role in the shape of the response of the pulse-shaping filter. Since the output of this filter to the rectangular polar NRZ pulses, is used for frequency-shift keying, the $g(t)$ given above is the 'frequency-shaping pulse'. Its shape for different values of *WT*, the time-bandwidth product, is plotted in Fig. 10.54.

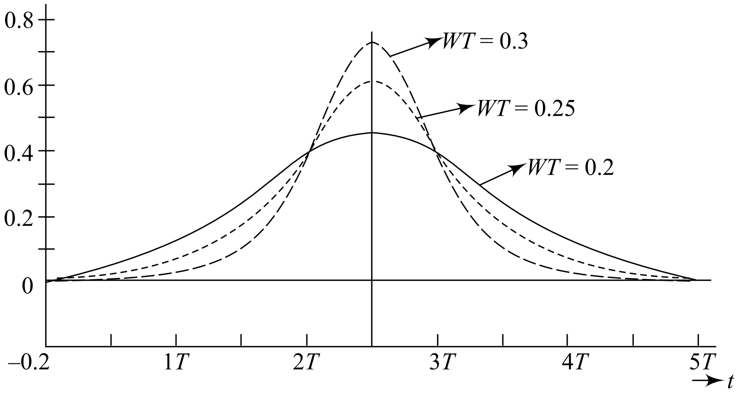


Fig. 10.54 *Causal approximation $\hat{g}(t)$, for three different values of WT*

As $g(t)$ is given in Eq. (10.120) is non-causal, it has to be truncated for some negative value of time and then shifted in time to the right (delayed) so as to get $\hat{g}(t)$ which is causal and is as good an approximation as possible with tolerable delay.

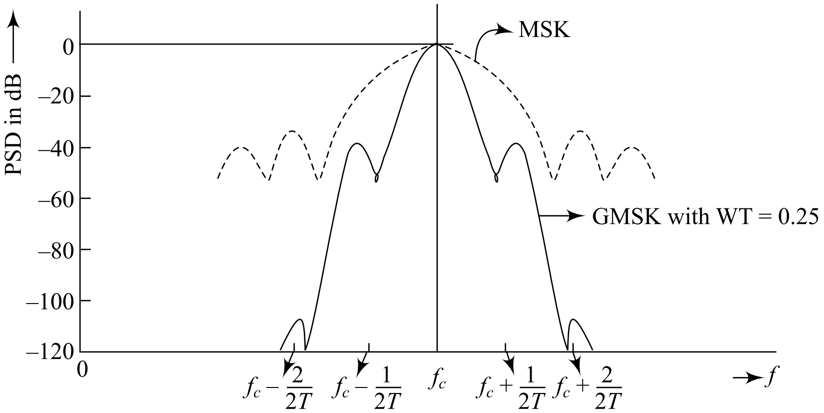


Fig. 10.55 *Power spectra of MSK and GMSK with WT = 0.25*

The PSD of a GMSK signal with *WT* = 0.25 is shown along with PSD of an MSK signal, in Fig. 10.55. With a small value of *WT* like 0.25, we find that most of the signal energy is packed much closer to the carrier frequency in the case of GMSK.

The probability of error for GMSK is approximately given by the empirical formula

$$P_e = \frac{1}{2} erfc\left[\sqrt{\frac{\alpha E_b}{2\eta}}\right] \tag{10.121}$$

where E_b is the signal energy per bit and α is a constant that depends on *WT*.

10.7.7 Advantages and Disadvantages of GMSK

Advantages GMSK has the advantage of good spectral compactness and so needs less bandwidth than MSK (see Fig. 10.55)

Disadvantages

1. Filtering of the message binary NRZ bit stream introduces some ISI. This increases as *WT* is decreased, as can be seen from Fig. 10.54.
2. As a value of $WT \approx 0.25$ to 0.3 is generally used (because of 1 above), and as the complementary error function is a monotonically decreasing function, from Eq. (10.121), we find that there will be a degradation of performance as compared to MSK.

Because of its spectral compactness, and preservation of the essential good features of MSK, Gaussian-filtered MSK is widely used in GSM cellular mobile communications.

Example 10.9 Given the input binary sequence 100010110, sketch the waveforms of the inphase and quadrature components of an MSK signal. Sketch the MSK signal also.

Solution

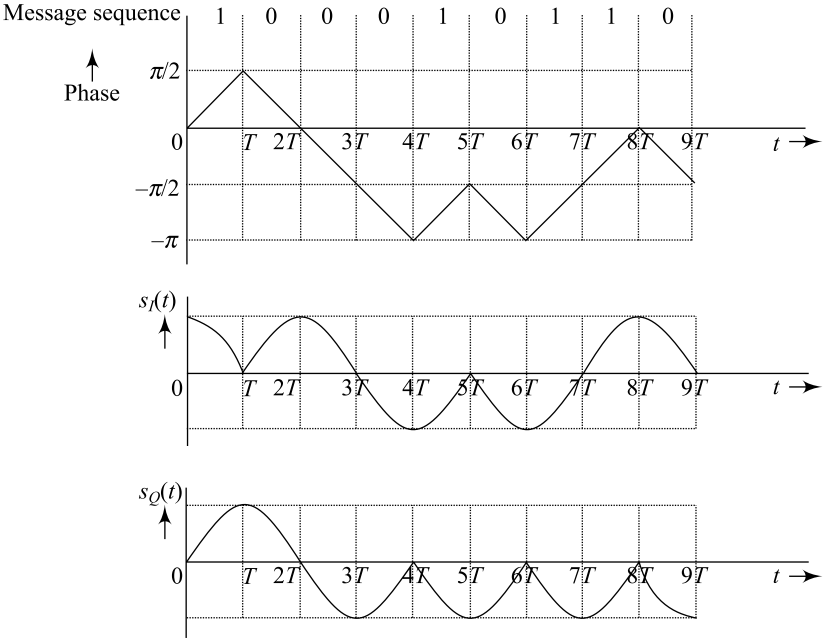


Fig. 10.56 *Phase, $s_I(t)$ and $s_Q(t)$ for Example 10.9*

1. Phase: The first message digit being 1, the phase increases from zero at $t = 0$ to $\pi/2$ at $t = T$. The next 3 digits being 0s, the phase decreases $\pi/2$ radians during each bit interval and is $-\pi$ at $4T$. Since the fifth digit is a 1 and the sixth is a 0, the phase increases from $-\pi$ to $-\pi/2$ (at $5T$) but again decreases to $-\pi$ at $6T$. Since the seventh and eighth digits are 1s, the phase increases from $-\pi$ at $6T$ to 0 at $8T$. But since the 9th digit is a 0, it again decreases to $-\pi/2$ at $9T$.

2. $s_I(t)$: If at any even multiple of T, including $0T$, if the phase is 0, then $s_I(t)$ will be a single half cosine pulse of width $2T$ centered on that even multiple of T (including zero). But if the phase is π or $-\pi$ at any even multiple of T, $s_I(t)$ will be a negative half cosine pulse of $2T$ and centered on that even multiple of T. At $t =$ 0, $2T$ and $8T$ the phase is 0. Hence, we get half-cosine pulses centered on 0, $2T$ and $8T$. At $t = 4T$ and $6T$, however, the phase is $-\pi$. So we get negative half-cosine pulses at these points.

3. $s_Q(t)$: At any odd multiple of T if the phase is $\pi/2$, $s_Q(t)$ will be a half-sine pulse of width $2T$ centered on that odd multiple of T. If the phase is $-\pi/2$, $s_Q(t)$ will be a negative half sine pulse of width $2T$ centered on that odd multiple of T. At $t = T$ the phase is $+\pi/2$ while at $t = 3T$, $5T$, $7T$ and $9T$, it is $-\pi/2$. So the $s_Q(t)$ is drawn accordingly.

10.8 M-ARY BAND PASS SIGNALING

In M-ary signaling, any one of M possible signals is transmitted during each time slot, where $M = 2^n$ and n is an integer. In baseband signaling, it was the M different pulse amplitudes that distinguished the M signals. In band pass M-ary signaling, the M signals can differ in their amplitudes, as in the case of M-ary ASK, or in their frequencies, as in the case of M-ary FSK or in their phases, as in the case of M-ary PSK. It is also possible to have some hybrid forms of M-ary signaling, notable among them being M-ary QAM, which has some useful features.

The discussion and analysis pertaining to baseband M-ary signaling, i.e., M-ary PAM given in section 10.4, apply equally well to band pass M-ary ASK too. As pointed there, if we wish to maintain some specified minimum probability of error, the difference between adjacent levels will have to be maintained the same even while increasing M in order to achieve a higher rate of transmission in bits/sec. Thus, the average transmitted power will increase with M if the probability of error is to remain the same. This means that in M-ary PAM, the average transmitted power sets a limit to the transmission rate that can be achieved while maintaining a specified probability of error. This applies equally well to M-ary ASK also.

In BPSK, QPSK and M-ary PSK, all the signals transmitted during a time slot have the same amplitude. They differ only in their phase and so all the message points pertaining to each of these systems will lie on the circumference of a circle in the signal space, with the origin as the center and a radius whose value is dependent on the symbol energy. As detection error in the presence of noise depends on the distance between any two adjacent message points, it is obviously advantageous to have not only phase difference but also amplitude difference between the M possible signals of an M-ary system. This leads to a *hybrid* type of M-ary band pass modulation scheme, known as M-ary QAM (also called M-ary QASK), i.e., M-ary quadrature amplitude modulation, which is a hybrid amplitude and phase modulation.

10.8.1 Basic QAM Signal

In QAM, two carrier signals of the same frequency but in phase quadrature, are independently amplitude modulated by discrete amplitudes, a_ks and b_ks so that the k^{th} transmitted signal may be written as

$$s_k(t) = \sqrt{\frac{2E_0}{\tau}} a_k \cos 2\pi f_c t - \sqrt{\frac{2E_0}{\tau}} b_k \sin 2\pi f_c t \qquad k = 0, \pm 1, \pm 2, \ldots \tag{10.122}$$

where E_0 is the energy of the signal with the smallest amplitude among all the M different signals that can possibly be transmitted. We will now choose two orthogonal basis signals $\phi_1(t)$ and $\phi_2(t)$ for the signal space in which the $s_k(t)$s lie, and define them as

$$\left.\begin{aligned} &\phi_1(t) = \sqrt{\frac{2}{\tau}} \cos 2\pi f_c t; \quad 0 \le t \le \tau \\ \text{and} \quad &\phi_2(t) = \sqrt{\frac{2}{\tau}} \sin 2\pi f_c t; \quad 0 \le t \le \tau \end{aligned}\right\} \tag{10.123}$$

10.8.2 Signal Constellation of QAM

Then $s_k(t)$ of Eq. (10.122) may be expressed as a linear combination of $\phi_1(t)$ and $\phi_2(t)$ as follows:

$$s_k(t) = a_k\sqrt{E_0}\phi_1(t) - b_k\sqrt{E_0}\phi_2(t) \tag{10.124}$$

Hence, the coordinates of $s_k(t)$ along $\phi_1(t)$ and $\phi_2(t)$ are respectively $a_k\sqrt{E_0}$ and $-b_k\sqrt{E_0}$ respectively. Thus, a_k^2 and b_k^2 represent the normalized energies of the inphase and quadrature components of $s_k(t)$ normalized with respect to E_0.

Now, let

$$\left.\begin{aligned} &\sqrt{M} = L \\ &\text{where } L \text{ is a positive integer} \end{aligned}\right\} \tag{10.125}$$

Then we may view the M message points in the signal constellation of QAM as having been generated by the Cartesian product of the two coordinate sets of the message points of an L-ary ASK with $\phi_1(t)$ as the carrier and another L-ary ASK with $\phi_2(t)$ as the carrier. The Cartesian product of two sets of coordinates each one representing coordinates in a one-dimensional space like the one in Fig. 10.57, is the set of all possible ordered pairs of coordinates with the first coordinate of each ordered pair drawn from the first set and the second coordinate (of the ordered pair) drawn from the second set. For instance, if $M = 16$ so that $L = 4$, the one-dimensional signal constellation of the 4-ary ASK with $\phi_1(t)$ as the carrier will be as shown in Fig. 10.57.

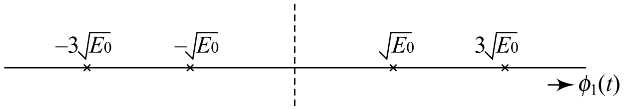


Fig. 10.57 *One-dimensional signal constellation of the 4-ary ASK with $\phi_1(t)$ as the carrier*

An exactly identical one-dimensional signal constellation of the 4-ary ASK with $\phi_2(t)$ as the carrier, may also be drawn. Then the two-dimensional constellation of the 4-ary QAM with $M = 16$ will have a signal constellation comprising 16 message points with their coordinates obtained by taking the Cartesian product of the coordinate set pertaining to the one-dimensional space $\phi_1(t)$ and the coordinate set pertaining to the one-dimensional space $\phi_2(t)$, in that order. The signal constellation of the 16-ary QAM is shown in Fig. 10.57. In this, for convenience, $\sqrt{E_0}$, is represented by 'a'.

10.8.3 M-ary QAM Generation and Detection

The generation is based directly on Eq. (10.122) and is shown in Fig. 10.59. Let $p(t)$ be an appropriately shaped baseband pulse. Then two baseband binary digital signals, $m_1(t)$ and $m_2(t)$ are produced such that $m_1(t) = a_k\, p(t)$ and $m_2(t) = b_k\, p(t)$. These are then used to modulate the two carriers which are in phase quadrature as shown in Fig. 10.59.

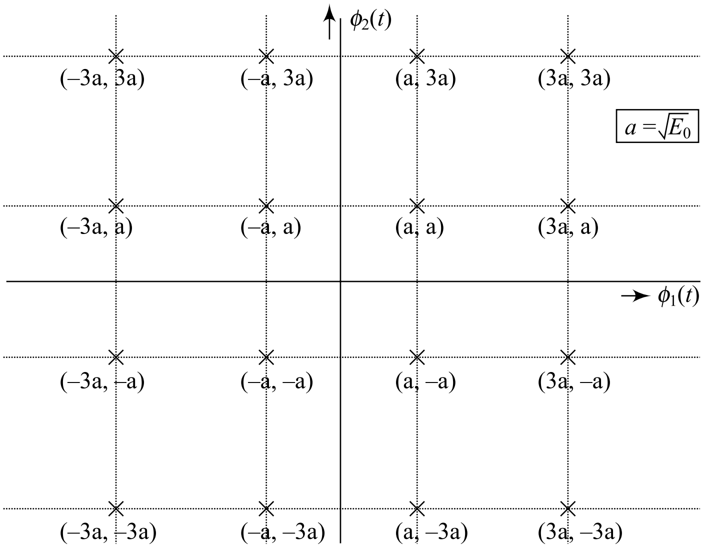


Fig. 10.58 *Signal constellation of 16-ary QAM*

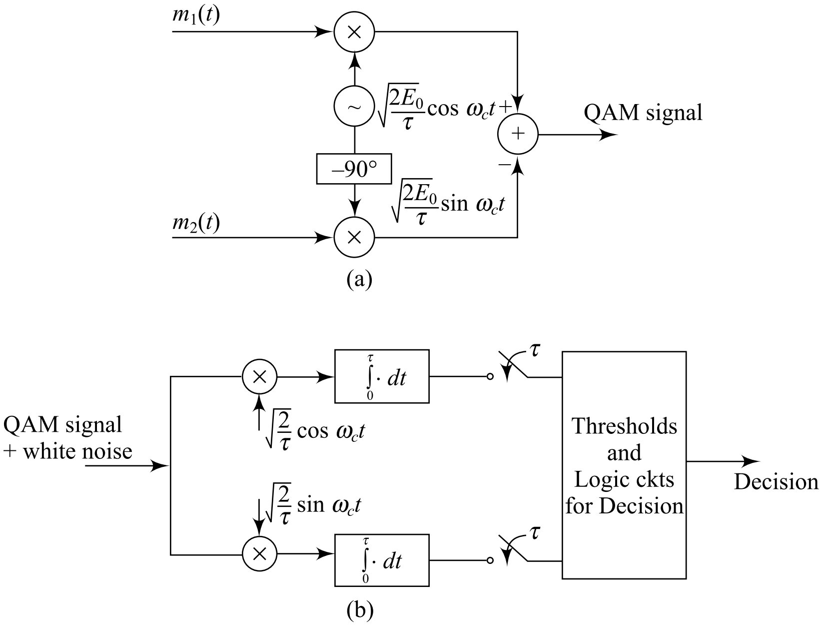


Fig. 10.59 *(a) M-ary QAM signal generation, (b) M-ary QAM receiver structure*

The receiver has two correlator units — one for the inphase carrier and the other for the quadrature carrier. The correlator outputs are sampled every τ seconds and fed to the threshold devices and logic circuits for taking decisions on which a_k and b_k have been used at the transmitter and then decide on which of the M possible signals has been transmitted during that symbol interval τ.

10.8.4 Bandwidth Efficiency of QAM

We know from the principle of quadrature carrier multiplexing that quadrature carrier ASK (binary) will have a bandwidth efficiency that is twice that of ordinary binary ASK. Since ordinary binary ASK has a bandwidth efficiency of 1, quadrature carrier ASK will have a bandwidth efficiency of 2. Since M-ary QAM is composed of two L-ary ASK systems on quadrature carriers, and since an L-ary ASK system will have a bandwidth efficiency that is $\log_2 L$ times that of a binary ASK, it follows that the bandwidth efficiency of M-ary QAM is $2\log_2 L = 2\log_2\sqrt{M} = \log_2 M$

$$\therefore \qquad \left(\frac{R_b}{B_T}\right)\text{for M-ary QAM} = \log_2 M$$

M-ary FSK In M-ary FSK, during each interval of τ seconds, one of a set of M signals having different frequencies, will be transmitted. The individual signal frequencies are separated by $(1/2\tau)$ hertz so that the M signal from an orthogonal set. That the signals will be orthogonal over τ sec when their frequencies are separated by $(1/2\tau)$, has already been demonstrated when we discussed BFSK. In addition, as usual, the carrier frequency f_c is so chosen that

$$f_c = \frac{n}{2\tau};\ n,\text{ a fixed integer} \tag{10.126}$$

Such a set of M signals may be represented mathematically as

$$s_k(t) = \sqrt{\frac{2E}{t}}\cos\left[\frac{\pi}{\tau}(n+k)t\right];\ 0 \le t \le \tau \qquad k = 1, 2, 3, \dots M \tag{10.127}$$

where E is the energy of each signal during a time-slot and the amplitude $\sqrt{2E/T}$ is chosen so as to make

$$\int_0^\tau s_k(t)\,s_m(t)\,dt = \begin{cases} E & \text{if } m = k \\ 0 & \text{otherwise}\end{cases} \tag{10.128}$$

Since E is the norm square of each signal, to normalize the set of signals we can divide each signal by its norm, i.e., $\sqrt{E}$ and obtain an orthonormal set of basis signals given by

$$\phi_k(t) = \frac{1}{\sqrt{E}}s_k(t);\ 0 \le t \le \tau \qquad k = 1, 2, \dots, M \tag{10.129}$$

Since the signals, $s_k(t)$s, are M orthogonal signals, and hence are linearly independent, the signal space of $s_k(t)$s is an M-dimensional one and we can use $\phi_k(t)$s, which are orthonormal and M in number, as the basis functions.

10.8.5 Signal Space and Signal Constellation

Since it is not possible to diagrammatically represent an M-dimensional space for $M > 3$, in Fig. 10.60, we have shown the signal space and constellation for $M = 3$. Obviously, it is just an extension of the signal space diagram for BFSK.

As is clear from the signal space diagram, the distance between any two signals is $\sqrt{2E}$.

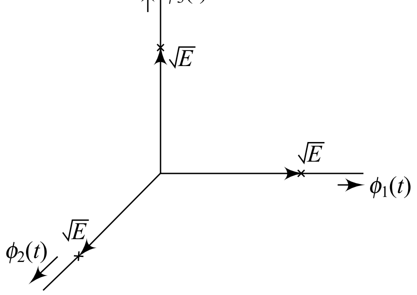


Fig. 10.60 *Signal space and constellation for M-ary FSK with M = 3*

10.8.6 Receiver Configuration

A coherent receiver for M-ary FSK simply consists of a bank of M correlators or matched filters. To the k^{th} correlator, $1 \le k \le M$, the signal $\phi_k(t)$ is supplied. At the end of each τ sec the receiver compares the outputs of all the M correlators or matched filters and selects the largest among them as per the maximum likelihood decision making strategy and decides on the baseband symbol accordingly.

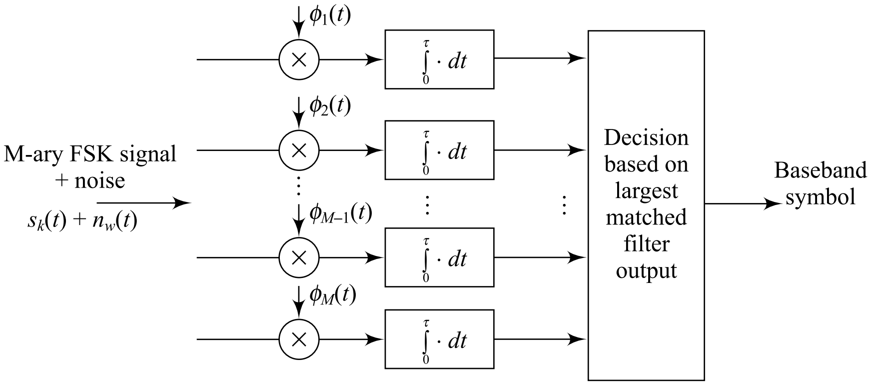


Fig. 10.61 *Coherent reception of M-ary FSK*

An exact expression for probability of symbol error is rather involved and difficult to derive. However, an upper-bound for the symbol error is given by

$$\underset{\text{(M-ary FSK)}}{P_e} \le \frac{1}{2}(M-1)\, erfc\left(\sqrt{\frac{E}{2\eta}}\right) \tag{10.130}$$

This upper bound is approached as (E/η) is increased and is almost reached when (E/η) is large enough to make $P_e \le 10^{-3}$. In the case $M = 2$, i.e., for BFSK (coherent), the equality sign holds and RHS of Eq. (10.130) reduces exactly to that obtained by us for coherent BFSK.

10.8.7 Bandwidth and Bandwidth Efficiency

Determination of the bandwidth by deriving the spectra of M-ary FSK is rather complicated and so we will try to get an approximate expression by using the fact that there are M possible signals separated from each other in frequency by $1/2T$ hertz. Hence, the transmission bandwidth B_T may be taken as

$$B_T \ge M \times \frac{1}{2T} = W\frac{R}{2} = \frac{MR_b/2}{\log_2 M}$$

where R is the symbol rate and R_b is the bit rate. The bandwidth efficiency, which is the ratio of the bit rate to the bandwidth, is given by

$$\rho = \text{Bandwidth efficiency of M-ary FSK}$$
$$= \left(\frac{2\log_2 M}{M}\right) \tag{10.131}$$

Remarks

(i) *Equation (10.131) clearly indicates that as M is increased the bandwidth efficiency of M-ary FSK decreases. In fact, while it is equal to 1 for M = 2, it is only 0.5 for M = 16. This is just the opposite of what happens in the case of M-ary PSK, for which ρ increases with M, as we will be seeing later.*

(ii) *Further, M-ary FSK requires considerably more bandwidth in comparison with M-ary PSK.*

10.8.8 M-ary PSK

Signal In M-ary PSK, the phase of the carrier can take one of the M possible values

$$\theta(k) = 2\pi(k-1)/M \; ; k = 1, 2, \ldots M \tag{10.132}$$

Hence, the M possible signals, one of which is transmitted during each signaling interval of T sec, are given by

$$s_k(t) = \sqrt{\frac{2E}{\tau}} \cos\left[2\pi f_c t + \frac{2\pi}{M}(k-1)\right] \qquad k = 1, 2, 3, \ldots M \tag{10.133}$$

Signal space, signal constellation and P_e We may expand the RHS of the above equation and write

$$s_k(t) = \sqrt{\frac{2E}{\tau}} \cos 2\pi f_c t \cdot \cos\frac{2\pi}{M}(k-1) - \sqrt{\frac{2E}{\tau}} \cdot \sin\frac{2\pi}{M}(k-1) \cdot \sin 2\pi f_c t \qquad k = 1, 2, 3, \ldots M \tag{10.134}$$

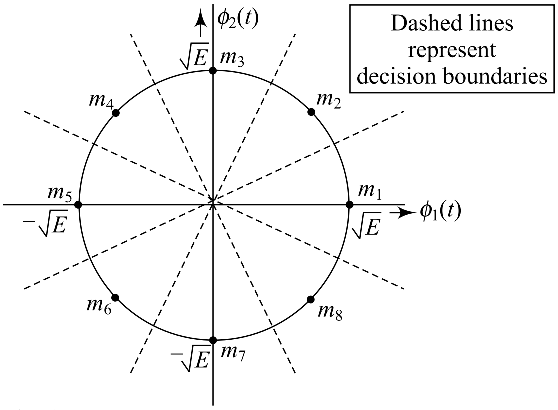
Since the first and second terms of the RHS of the above equation are orthogonal to each other (one being the inphase component and the other being the quadrature component), just as in the case of BPSK, here too, we may choose the following two orthogonal signals as the basis signals for the signal space of M-ary PSK.

$$\left.\begin{aligned} \phi_1(t) &= \sqrt{\frac{2}{\tau}} \cos 2\pi f_c t; \quad 0 \le t \le \tau \\ \text{and} \qquad \phi_2(t) &= \sqrt{\frac{2}{\tau}} \sin 2\pi f_c t; \quad 0 \le t \le \tau \end{aligned}\right\} \tag{10.135}$$

Thus, *the signal space of M-ary PSK is two dimensional* and for any k, $1 \le k \le M$, $s_k(t)$ can be expressed as a linear combination of these two basis signals. $\phi_1(t)$ and $\phi_2(t)$. Since all the M signals of M-ary PSK have the same amplitude, their energies over the symbol interval, τ sec, is the same, viz., E. Hence, all the message signal points lie on the circumference of a circle of radius $\sqrt{E}$ which is the coordinate of $s_k(t)$ along $\phi_1(t)$ for k =1. The signal space diagram, signal constellation and the decision boundaries for detection, are shown in Fig. 10.62 for M = 8.

Suppose the signal $s_1(t)$, corresponding to k = 1 and represented in the signal space as message point m_1, is transmitted. Then, unless the ratio E/η is too small, the probability of signal $s_1(t)$ being mistaken by the receiver for any $s_k(t)$ other than $s_2(t)$ or $s_8(t)$, the nearest neighbors of m_1, is very little. Thus, for reasonably large values of E/η, we have to consider only the probability of m_1 being mistaken for either m_2 or m_8 because of the presence of noise. Proceeding on these lines, it can be shown for $M \ge 4$ that

$$\text{M-ary PSK:} \quad \underset{\text{(symbol error)}}{P_e} \approx erfc\left[\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{M}\right)\right] \tag{10.136}$$



Dashed lines represent decision boundaries

Fig. 10.62 *Signal space and signal constellation of 8-ary PSK*

Remark *When M = 4, P_e as given by Eq. (10.136) reduces exactly to the expression for QPSK derived in Chapter 11.*

10.8.9 Bandwidth and Bandwidth Efficiency of M-ary PSK

The method to be adopted for determining the PSD of M-ary PSK signals is similar to the one we had employed for the PSK and QPSK signals. The PSD of the low frequency equivalent of an M-ary PSK signal can be shown to be given in terms of symbol energy E, as

$$P(f) = 2E\,\text{sinc}^2(f\tau) \tag{10.137}$$

But, the symbol duration τ is given by

$$\tau = T\log_2 M \tag{10.138}$$

$$\therefore \quad P(f) = 2E\,\text{sinc}^2[fT\log_2 M]$$

$$= 2E_b\log_2 M\,\text{sinc}^2[fT\log_2 M] \tag{10.139}$$

where E_b is the bit energy.

A plot of $P(f)$, normalized with respect to $2E_b$, is shown in Fig. 10.63.

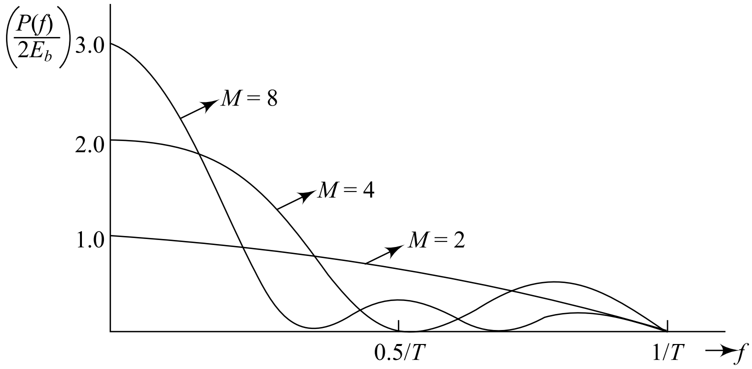


Fig. 10.63 *Power spectra of M-ary PSK for M = 8, 4 and 2*

The null-to-null, or the main-lobe bandwidth of M-ary PSK signals may be taken to be $2/\tau$. But the 3-dB bandwidth is

$$B_T = \frac{1}{\tau} \tag{10.140}$$

where τ is the symbol duration, which is related to the bit duration T_b through the relation given in Eq. (10.138).

$$B_T = \frac{1}{T\log_2 M} = \frac{R_b}{\log_2 M} \tag{10.141}$$

where R_b is the bit rate.

Since bandwidth efficiency, ρ is the ratio of the bit rate to the bandwidth,

$$\rho = \frac{R_b}{B_T} = \log_2 M \tag{10.142}$$

Equation (10.142) tells us that as M increases, the bandwidth efficiency of an M-ary PSK system improves. However, as M increases, the distance between message signal points on the circumference of the circle, will decrease and so the probability of error increases. This can be countered, of course, by increasing the radius of the circle, i.e., by increasing the transmitted power, since the radius is equal to $\sqrt{E}$.

Example 10.10 An M-ary PSK system is to operate with 2^n symbols over a 100 kHz channel. The bit rate is required to be at least 750 kilobits/sec. What minimum CNR is required if the bit-error probability should be equal to or better than $P_b = 10^{-6}$? Assume ISI free conditions.

Solution The maximum symbol rate under ISI-free conditions is

$$R_s = \frac{1}{\tau} = \text{Bandwidth } B_T \qquad \text{(Eq. (10.140))}$$

Since channel Bandwidth is limited to 100 kHz,

$$R_s \leq 100 \times 10^3 \text{ symbols/sec}$$

Minimum value of R_b required = 750 × 10^3 bits/symbol

Minimum number of bits/symbol = $\dfrac{750 \times 10^3}{100 \times 10^3} = 7.5$ bits/symbol

∴ for the M-ary PSK, $M \geq 2^{7.5}$

Since M must be an integer power of 2, let us take $M = 2^8$

∴ $M = 2^8 = 256$

When Gray coding of bits is used obtain PSK symbols,

Probability of symbol error = $P_e = P_b \cdot \log_2 M = 10^{-6} \cdot \log_2 256$

$$\therefore \quad \underset{\text{(symbol error)}}{P_e} = 8 \times 10^{-6}$$

But from Eq. (10.136), we know that for M-ary PSK with $M \geq 4$,

$$\underset{\text{(symbol error)}}{P_e} \approx erfc\left[\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{M}\right)\right] \text{ where } E \text{ is the symbol energy}$$

∴ $\text{erfc}\left[\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{M}\right)\right] = 8 \times 10^{-6}$. From error-function tables, we get

$$\left[\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{8}\right)\right] = 3.2$$

$$\therefore \quad \frac{E}{\eta} = 69.924$$

Now, $R_s = \dfrac{R_b}{\log_2 M} = \dfrac{750 \times 10^3}{8} = 93.75 \times 10^3$ symbols/sec.

∴ E = Symbol duration × Average carrier power

$$\therefore \quad \frac{E}{\eta} \cdot \frac{1}{T} = \frac{\text{Average carrier power}}{\eta} = \frac{E}{\eta} \cdot R_s$$

$$= 69.924 \times 93.75 \times 10^3 = 6555.375 \times 10^3$$

$$\therefore \quad (CNR)_{\min} = 10 \log_{10}(6555375) = 68.166 \text{ dB}$$

10.9 SYNCHRONIZATION

In both baseband signaling as well as band pass signaling, it was noted that the receiver has to observe the received signal (corrupted by channel noise) and must decide whether what was transmitted during the time slot under question, was a binary 1 or a binary 0 in the case of binary signaling. In the case of M-ary signaling, the receiver has to decide which signal, out of the M possible signals, was transmitted during each symbol duration. For proper operation of the system it is therefore necessary for the receiver to know exactly, when a bit is starting and when it is ending. Similarly, it must know the time of commencement and time of ending of a word (or a frame). If it is band pass signaling and a coherent system is used, then the receiver must have, available to it, a carrier signal that is having exactly the same frequency and phase as the carrier signal at the transmitter. All these requirements underscore the need for what is called 'synchronization' of the transmitter and receiver. Synchronization is the name given to the process of ensuring that the two clocks, one at the transmitter, and the other at the receiver, tick together, giving of course due allowance to the propagation/transmission delay.

From the foregoing it is clear that all digital communication systems, irrespective of whether they are baseband or band pass systems, need the following two levels of synchronization.

1. Bit synchronization
2. Word synchronization

In addition to the above two, in the case of band pass systems of the coherent type, there is a need for

3. Carrier synchronization, or carrier recovery.

Techniques adopted for each of the above may be broadly classified as

(a) Data-aided synchronization
(b) Non-data-aided synchronization

10.9.1 Data-Aided Synchronization Systems

In these systems, a preamble containing information about the carrier and symbol/bit timing is transmitted periodically along with the data carrying signal in a time-multiplexed manner. The receiver extracts the information contained in the preamble and utilizes it for synchronization.

Advantage Time required for synchronization is small.

Disadvantages

1. Data throughput is somewhat reduced as a part of each frame is used for sending synchronization information.
2. Power efficiency is also somewhat reduced as a part of the power is diverted for transmission of synchronization information.

10.9.2 Nondata-Aided Synchronization Systems

These are also called self-synchronizing systems. In these systems, no preamble is transmitted, instead, the receiver has to extract the required bit, word and carrier synchronization information from the transmitted data-bearing signal itself.

Advantages

1. Better data throughput
2. Better power efficiency

10.9.3 Bit Synchronization

The complexity of the methods for bit synchronization depends on the line code used at the transmitter for representing the binary sequence. If it uses unipolar RZ code, the bit-synchronization problem becomes almost trivial since the PSD of this line code has a delta function at the bit rate as shown in Fig. 10.2(a). Therefore, narrowband filtering of the received unipolar RZ signal using a filter tuned to a frequency equal to the bit rate will yield the necessary clock signal. Alternatively, a PPL locked to the frequency $1/T_b$ can be used to extract the synchronizing signal. On the other hand, if a polar NRZ line code is used, the noise-corrupted received NRZ waveform is first filtered and then fed to a square-law device to convert into a unipolar RZ signal. Since this unipolar RZ signal will have a Delta function in its PSD, a PLL or a narrowband filter can be used to extract the clock signal. The problem with this method is that since it relies heavily on the zero crossings of the received signal (baseband) it works well when the zero crossings are spaced at integer multiples of the bit duration. Also, if there is, at any time, a long string of either 1s or 0s, since there will be no zero crossing during that period, synchronism may be lost. To overcome this problem, data scramblers are used at the transmitter and descramblers are used at the receiver.

A totally different approach that does not depend on zero crossing is what is called the '*early-late bit synchronization technique*'. This technique makes use of the fact (as may be seen from the eye pattern of a polar NRZ data stream) that a properly filtered digital signal has peaks at the ideal or optimum sampling instants, and is fairly symmetric on either side of the peak. Hence, if $\Delta < T_b/2$ and if t_0 is synchronized

$$|y(t_0 - \Delta)| \approx |y(t_0 + \Delta)| < y(t_0) \tag{10.143}$$

A delayed (or late) synchronization signal will result in making

$$|y(t_0 - \Delta)| > |y(t_0 + \Delta)|$$

whereas an early synchronization will result in making

$$|y(t_0 - \Delta)| < |y(t_0 + \Delta)|$$

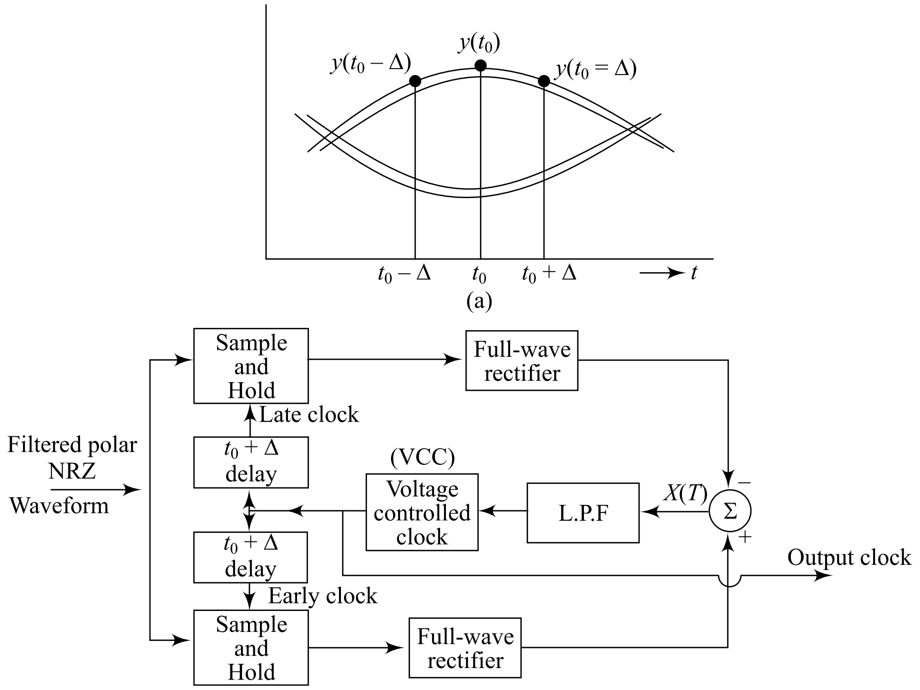


Fig. 10.64 *(a) Eye-diagram wave form showing optimum sampling time, (b) Block diagram of early-late synchronizer for bit synchronization*

The 'early-late synchronizer' shown in Fig. 10.63(b) makes use of the above-mentioned features and produces a control voltage for the VCC. In case the control signal, $x(t)$, which is equal to $|y(t_0 - \Delta)| - |y(t_0 + \Delta)|$ is positive, it makes the clock to run faster. On the other hand, if $x(t)$ is negative it makes the voltage controlled clock to run slower. In this way it tries to keep Δ very small.

10.9.4 Carrier Synchronization

As word frame synchronization has already been discussed in Chapter 9, we will proceed now with carrier synchronization.

If the spectrum of the modulated signal has a carrier component present in it, achieving carrier coherence is a very simple matter. One has to simply lock on to the carrier component using a PLL. In case a carrier component is not present in the modulated signal spectrum, the same techniques that are used for carrier recovery in the coherent reception of DSB-SC signals, may be employed. These are the costas loop and the squaring loop. These have been discussed in detail earlier.

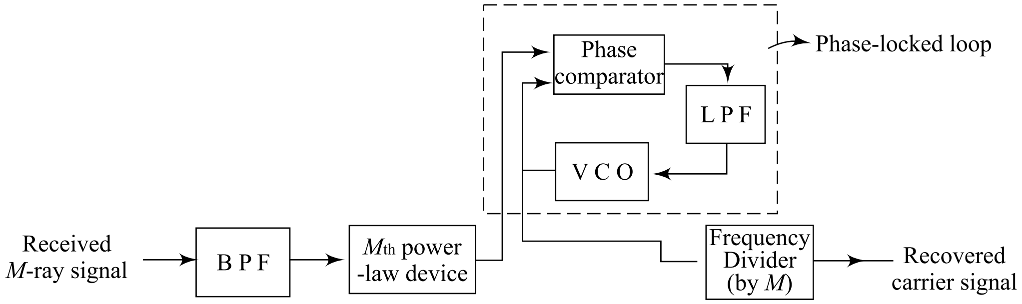


Fig. 10.65 *M-th power loop for carrier recovery*

For carrier recovery in the case of M-ary PSK, etc., what is generally referred to as the 'M-th power loop' consisting of an M-th power law device, a BPF, a PLL and a frequency divider by M, all connected in cascade, may be used, as shown in Fig. 10.65.

Matlab Example 10.1 In this problem, we will study the effect of channel noise and bandwidth limitation of the channel on the eye diagram for a binary rectangular polar baseband pulse transmission system using raised cosine pulse shaping with roll-off factors $\alpha = 0.5$ and $\alpha = 1.0$ and a constant Nyquist bandwidth, W, of 0.5 that is equal to $\frac{1}{2T}$, where T is the bit duration. First we plot the impulse response of the raised cosine filter for these two values of α. Then we plot the eye diagrams obtained under the following four conditions, using the above parameters.

1. Impulse response of the raised cosine filter
2. Eye diagram without channel noise and without a bandlimiting filter
3. Eye diagram with channel noise but without the bandlimiting filter
4. Eye diagram without channel noise but with the bandlimiting filter
5. Eye diagram with channel noise as well as the bandlimiting filter

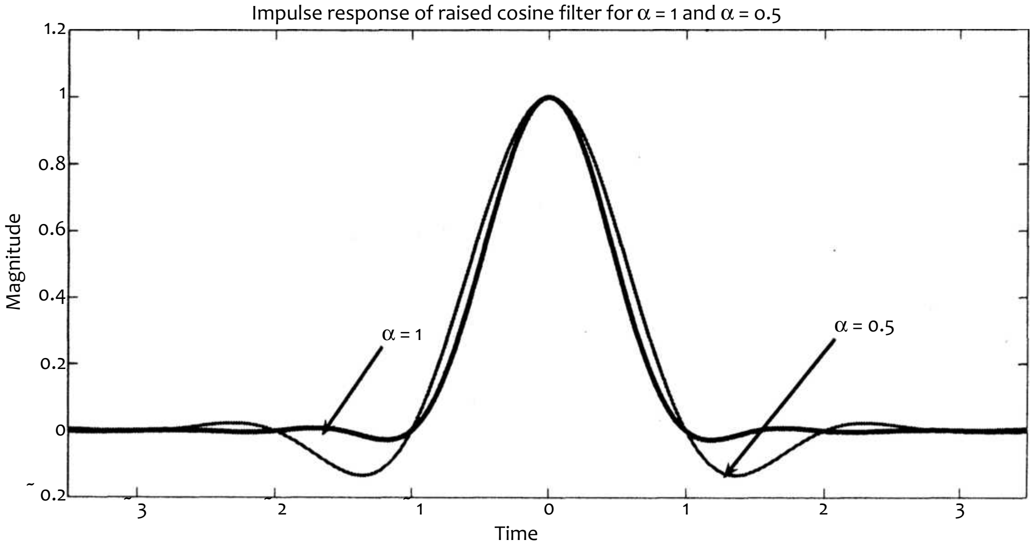
1. Impulse Response of the Raised Cosine Filter

MATLAB Program

```
fs= 5; % sampling frequency in Hz
% defining the sinc filter
```

```
sincNum = sin(pi*[-fs:0.01:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:0.01:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0

% raised cosine filter
alpha = 0.5;
cosNum = cos(alpha*pi*[-fs:0.01:fs]);
cosDen = (1-(2*alpha*[-fs:0.01:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha5 = sincOp.*cosOp;
fs1=-fs:0.01:fs;
% defining the sinc filter
sincNum = sin(pi*[-fs:0.01:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:0.01:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0
% raised cosine filter
alpha = 1.0;
cosNum = cos(alpha*pi*[-fs:0.01:fs]);
cosDen = (1-(2*alpha*[-fs:0.01:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha1 = sincOp.*cosOp;
plot(fs1,gt_alpha1,fs1,gt_alpha5)
```



Note: The bandlimiting filter is an RC lowpass filter with RC = 1.67 seconds

Fig. 10.66

2. Without Channel Noise and Without a Bandlimiting Filter

MATLAB Program

```
 clear
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
fs= 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0

% raised cosine filter
alpha = 0.5;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha5 = sincOp.*cosOp; %bandlimiting filter
% upsampling the transmit sequence
amUpSampled = [am;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha5 = conv(amU,gt_alpha5);
% taking only the first 4000 samples
st_alpha5 = st_alpha5([1:4000]);
st_alpha5_reshape = reshape(st_alpha5,fs*2,N*fs/10).';
close all
subplot(211);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=0.5');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
N  = 800; % number of symbols
 randn('state',0);
 am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
 fs = 5; % sampling frequency in Hz
 % defining the sinc filter
 sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
 sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
 sincDenZero = find(abs(sincDen) < 10^-10);
 sincOp = sincNum./sincDen;
```

```
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0
% raised cosine filter
alpha = 1.0;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha1 = sincOp.*cosOp; %bandlimiting filter
% upsampling the transmit sequence
amUpSampled = [am;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha1 = conv(amU,gt_alpha1);
% taking only the first 4000 samples
st_alpha1 = st_alpha1([1:4000]);
st_alpha1_reshape = reshape(st_alpha1,fs*2,N*fs/10).';
subplot(212);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=1.0');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
```

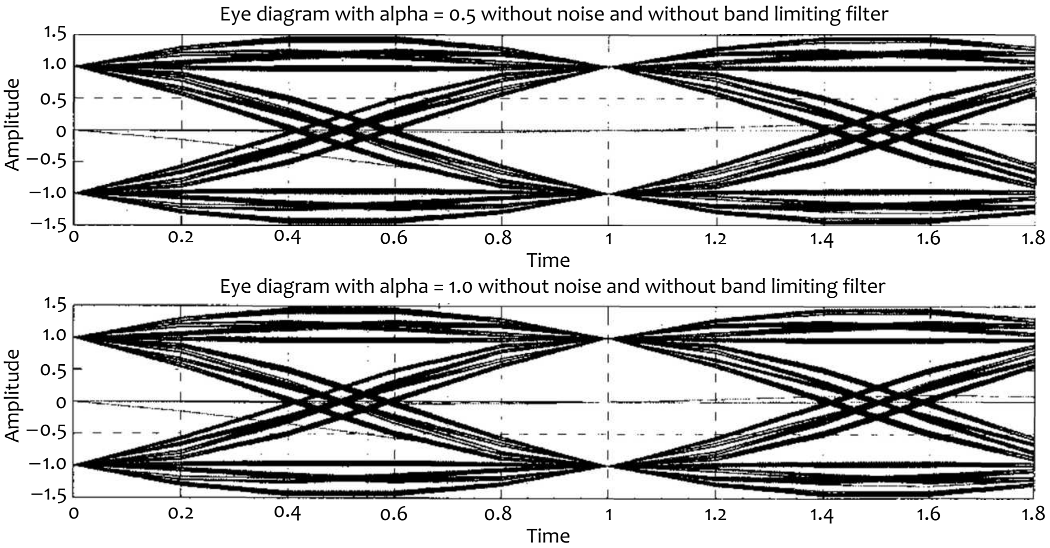


Fig. 10.67

In this case as there is no noise and no bandlimiting filter, it can be seen from the above eye diagrams that the eye is fully open both in the horizontal and vertical directions.

3. With Channel Noise but Without Bandlimiting Filter

MATLAB Program

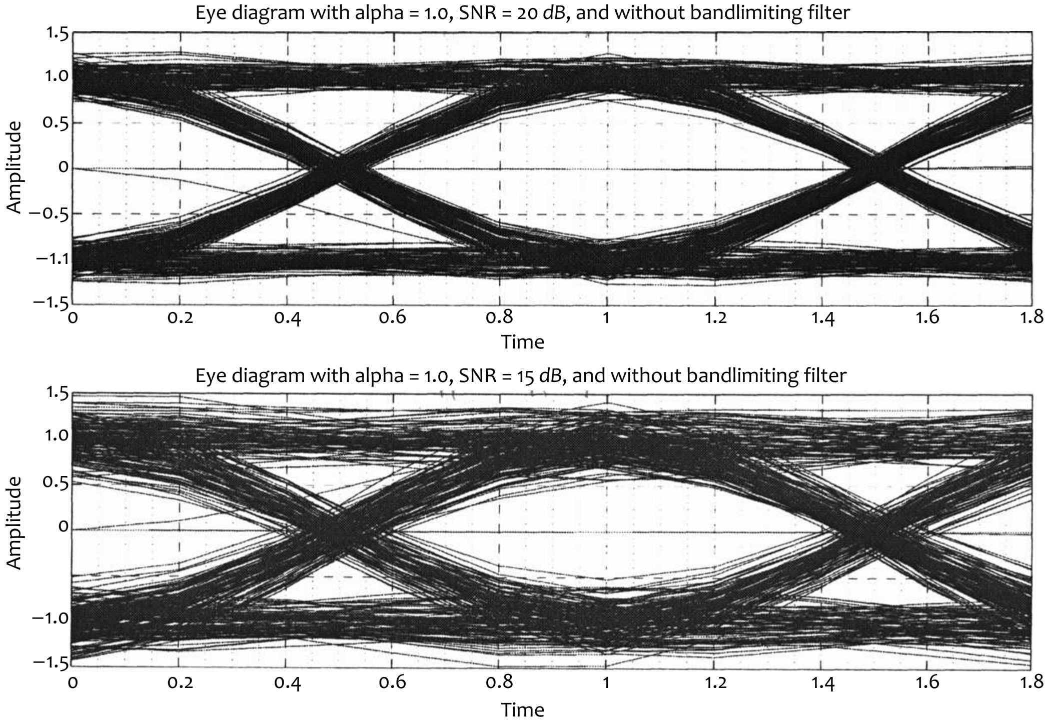
```
 clear
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
am1=awgn(am,20);%Noise-added sequence
fs = 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0

% raised cosine filter
alpha = 1.0;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha1 = sincOp.*cosOp ;%Band-limiting RC lowpass filter
% upsampling the transmit sequence
amUpSampled = [am1;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha1 = conv(amU,gt_alpha1);
% taking only the first 4000 samples
st_alpha1 = st_alpha1([1:4000]);
st_alpha1_reshape = reshape(st_alpha1,fs*2,N*fs/10).';
close all
subplot(211);
plot([0:1/fs:1.99],real(st_alpha1_reshape).','b');
title('eye diagram with alpha=1.0');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
am1=awgn(am,15);%Noise-added sequence
fs = 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
```

```
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0
% raised cosine filter
alpha = 1.0;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha1 = sincOp.*cosOp ;%Band-limiting RC lowpass filter
% upsampling the transmit sequence
amUpSampled = [am1;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha1 = conv(amU,gt_alpha1);
% taking only the first 4000 samples
st_alpha1 = st_alpha1([1:4000]);
st_alpha1_reshape = reshape(st_alpha1,fs*2,N*fs/10).';
subplot(212);
plot([0:1/fs:1.99],real(st_alpha1_reshape).','b');
title('eye diagram with alpha=1.0');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
```


Eye diagram with alpha = 1.0, SNR = 20 *dB*, and without bandlimiting filter

Eye diagram with alpha = 1.0, SNR = 15 *dB*, and without bandlimiting filter

Fig. 10.68

Note that by changing the value of α to 0.5 in the above program, we obtain the following results.

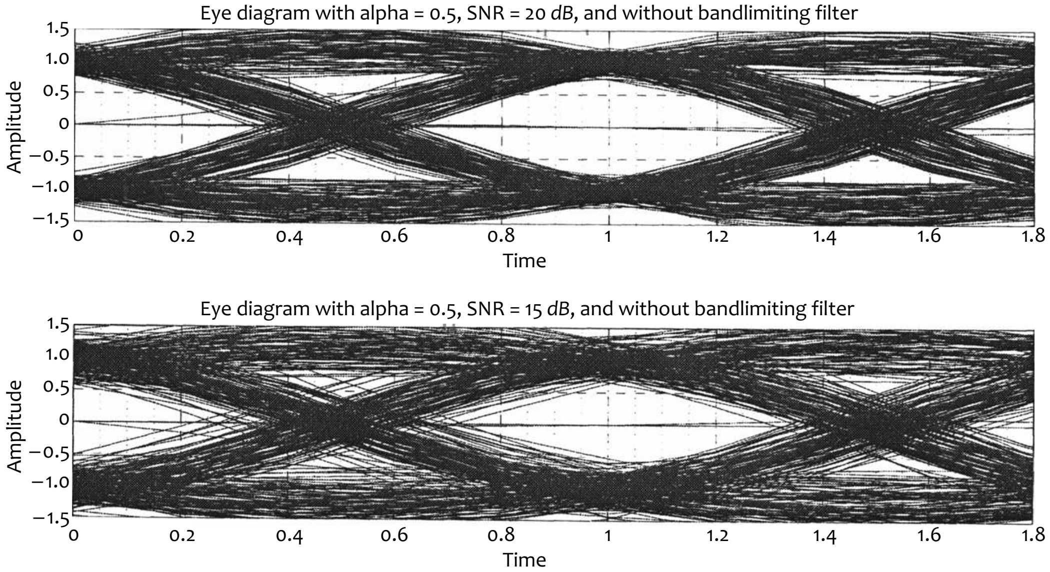


Fig. 10.69

In this case because of the channel noise, the eye opening in the vertical direction is reduced while the opening in the horizontal direction is unaltered. Note the reduction in the vertical eye opening when the SNR is reduced to 15 *dB*.

4. Without Channel Noise but With Bandlimiting Filter

MATLAB Program

```
 clear
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
fs= 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0

% raised cosine filter
alpha = 0.5;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha5 = sincOp.*cosOp.*exp(-0.67.*[-fs:1/fs:fs]); %bandlimiting filter
% upsampling the transmit sequence
amUpSampled = [am;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
```

```
% filtered sequence
st_alpha5 = conv(amU,gt_alpha5);
% taking only the first 4000 samples
st_alpha5 = st_alpha5([1:4000]);
st_alpha5_reshape = reshape(st_alpha5,fs*2,N*fs/10).';
close all
subplot(211);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=0.5');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
fs = 5; % sampling frequency in Hz
% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0
% raised cosine filter
alpha = 1.0;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha1 = sincOp.*cosOp.*exp(-0.67.*[-fs:1/fs:fs]); %bandlimiting filter
% upsampling the transmit sequence
amUpSampled = [am;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha1 = conv(amU,gt_alpha1);
% taking only the first 4000 samples
st_alpha1 = st_alpha1([1:4000]);
st_alpha1_reshape = reshape(st_alpha1,fs*2,N*fs/10).';
subplot(212);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=1.0');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
```

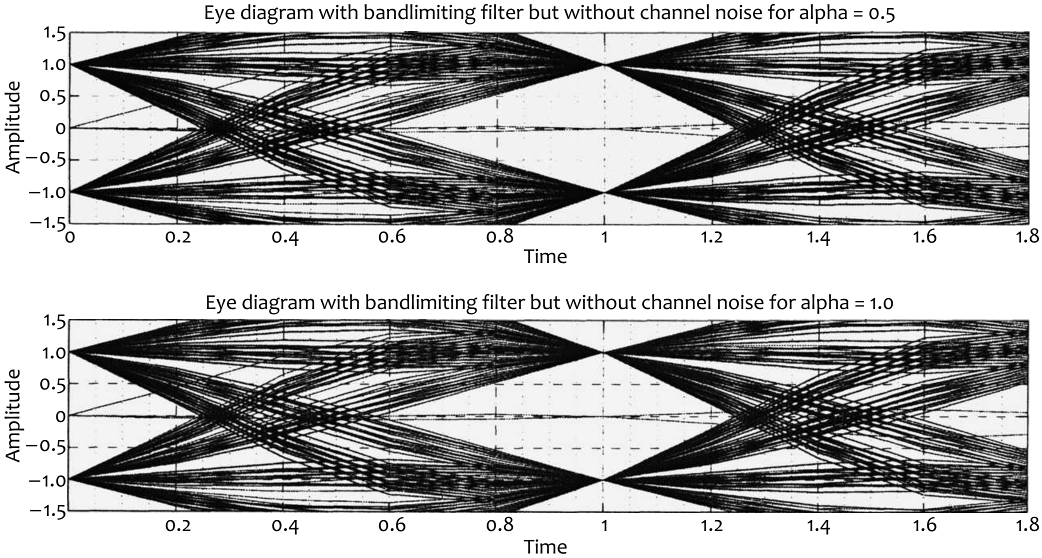


Fig. 10.70

In this case owing to the absence of the channel noise, the vertical opening of the eye is unaltered while the opening in the horizontal direction is reduced due to the band limiting filter.

5. With Channel Noise and also the Bandlimiting Filter

MATLAB Program

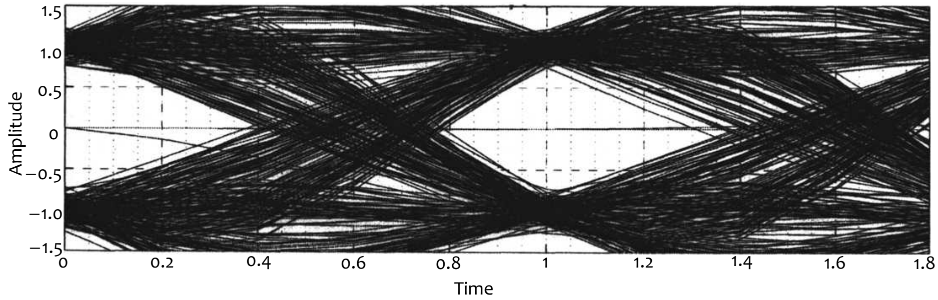
```
 clear
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
am1=awgn(am,20);%Noise-added sequence
fs = 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0

% raised cosine filter
alpha = 0.5;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha5 = sincOp.*cosOp.*exp(-0.67*[-fs:1/fs:fs]);%Bandlimiting RC lowpass
% upsampling the transmit sequence
amUpSampled = [am1;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
```

```
% filtered sequence
st_alpha5 = conv(amU,gt_alpha5);
% taking only the first 4000 samples
st_alpha5 = st_alpha5([1:4000]);
st_alpha5_reshape = reshape(st_alpha5,fs*2,N*fs/10).';
close all
subplot(211);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=0.5');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
N  = 800; % number of symbols
randn('state',0);
am = 2*(rand(1,N)>0.5)-1 ;% generating random binary sequence
am1=awgn(am,15);%Noise-added sequence
fs = 5; % sampling frequency in Hz

% defining the sinc filter
sincNum = sin(pi*[-fs:1/fs:fs]); % numerator of the sinc function
sincDen = (pi*[-fs:1/fs:fs]); % denominator of the sinc function
sincDenZero = find(abs(sincDen) < 10^-10);
sincOp = sincNum./sincDen;
sincOp(sincDenZero) = 1; % sin(pix/(pix) =1 for x =0
% raised cosine filter
alpha = 0.5;
cosNum = cos(alpha*pi*[-fs:1/fs:fs]);
cosDen = (1-(2*alpha*[-fs:1/fs:fs]).^2);
cosDenZero = find(abs(cosDen)<10^-10);
cosOp = cosNum./cosDen;
cosOp(cosDenZero) = pi/4;
gt_alpha5 = sincOp.*cosOp.*exp(-0.67*[-fs:1/fs:fs]);%Band-limiting RC lowpass
% upsampling the transmit sequence
amUpSampled = [am1;zeros(fs-1,length(am))];
amU = amUpSampled(:).';
% filtered sequence
st_alpha5 = conv(amU,gt_alpha5);
% taking only the first 4000 samples
st_alpha5 = st_alpha5([1:4000]);
st_alpha5_reshape = reshape(st_alpha5,fs*2,N*fs/10).';
subplot(212);
plot([0:1/fs:1.99],real(st_alpha5_reshape).','b');
title('eye diagram with alpha=0.5');
xlabel('time')
ylabel('amplitude')
axis([0 2 -1.5 1.5])
grid on
```


Eye diagram with alpha = 0.5, SNR = 20 *dB*, and bandlimiting filter

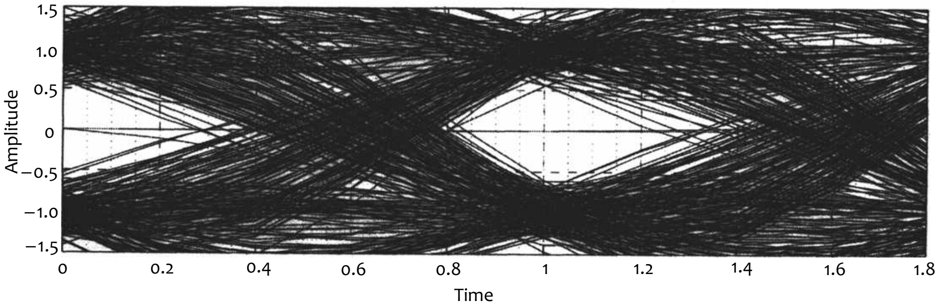
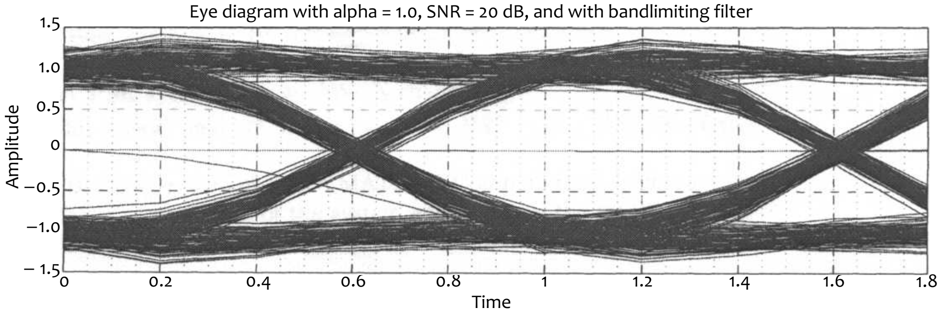

Eye diagram with alpha = 0.5, SNR = 15 *dB*, and bandlimiting filter

Fig. 10.71

Note that by changing the value of α from 0.5 to 1.0 in the above program, we get the following results.


Eye diagram with alpha = 1.0, SNR = 20 dB, and with bandlimiting filter

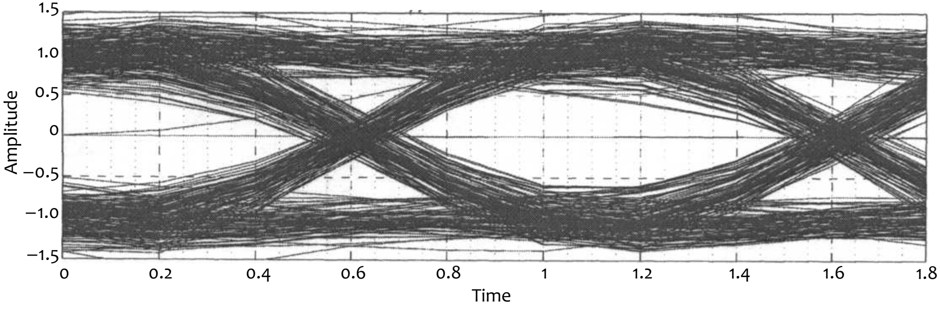

Eye diagram with alpha = 1.0, SNR = 15 dB, and with bandlimiting filter

Fig. 10.72

In this case since channel noise as well as the bandlimiting filter are present, the eye opening is reduced in both the directions. It may also be noted that for SNR of 15 dB the eye opening in the vertical direction is very much reduced compared to SNR of 20 dB.

Summary

- *Line codes* These are used for electrical representation of a binary data stream. Popular line codes are: (a) Unipolar NRZ, (b) Polar NRZ, (C) Unipolar RZ, (d) bipolar RZ, and (e) Split-phase, or Manchester code.
- Desirable properties of line codes
  (a) Absence/suppression of dc component
  (b) Possibility of timing recovery
  (c) Possibility of error detection.
- *Probability of error* It is the average fractional number of erroneously received symbols when a very large number of received symbols are considered.
- *ISI* When a digital signal is transmitted through a dispersive channel, the pulses are no longer confined to their respective time slots when they arrive at the receiver. They spill over into adjacent time slots causing inter-symbol interference (ISI).
- *Nyquist Criterion* $\sum_{n=-\infty}^{\infty} P(f - nf_z) = T$
- This equation is referred to as *Nyquist criterion* for distortionless baseband transmission of digital data.
- A pulse of sinc shape satisfies Nyquist criterion.
- *Ideal Nyquist channel* An ideal Nyquist channel is one in which the output pulses from the receive filter have the shape of a sinc function.
- Bandwidth and pulse rate of ideal Nyquist channel
  R = bit rate = $1/T$; W = Bandwidth of Nyquist channel = $R/2 = 1/2T$
  $\therefore$ $R = 2W$ bits/sec.
- Disadvantages of the ideal Nyquist channel
  (a) Not physically realizable
  (b) Even slight sampling jitter causes ISI.
  (c) Rate of decrease of sinc pulse is only $1/|t|$.
- *Raised cosine spectrum* A pulse $p(t)$ which has a raised cosine spectrum also satisfies Nyquist criterion and gives zero ISI.

$$P(f) = \begin{cases} \dfrac{1}{2W}; & |f| < f_1 \\ \dfrac{1}{4W}\left\{1 + \cos\left[\dfrac{\pi(|f| - f_1)}{2W - 2f_1}\right]\right\}; & f_1 < |f| < 2W - f_1 \\ 0; & |f| > 2W - f_1 \end{cases}$$

and
$$p(t) = \text{sinc } 2Wt\left[\frac{\cos 2\pi\rho Wt}{1 - 16\rho^2 W^2 t^2}\right]$$

where $\rho \triangleq$ roll-off factor and $0 \le \rho \le 1$.

- *Correlative coding* It is technique by which a transmission speed of $2W$ is achieved on a channel of bandwidth W by introducing controlled ISI.
- *Duo-binary signaling* It is a particular form of correlative coding. It gives Nyquist speed of transmission but suffers from the following disadvantages:
  (a) Non-zero PSD at $f = 0$, (b) Error propagation
- *Pre-coding* It is a technique used for avoiding error propagation in a duo-binary signaling system.
- *Modified duo-binary system* It is a signaling system that gives the Nyquist speed of signaling but without the disadvantages of the duo-binary scheme. It uses a pre-coder with a delay of $2T$ instead of T along with a modified

version of the duo-binary filter which too uses a delay of $2T$ instead of T. There is no error propagation, PSD = 0 at $f = 0$ and further, the tails decay at the rate of $1/|t|^2$.

- *M-ary baseband signaling* The baseband pulse will have M-levels (where M is a power of 2) instead of only 2 levels as in binary case.

$$R_M = Tr.\ \text{Rate with M-ary signaling} = \frac{1}{nT} = \frac{R_b}{\log_2 M}$$

where n = (symbol duration / time-slot duration).

(a) Bandwidth for M-ary: $W_m = (W_b/\log_2 M)$

(b) Probability of error P_e: $\underset{\substack{\text{(Polar M-ary}\\ \text{Baseband)}}}{P_e} = 2\left(1 - \frac{1}{M}\right)Q\left(\frac{A}{2\sigma}\right)$

- *Eye pattern* The eye pattern is an experimental tool that can be used for evaluating the combined effect of channel noise and ISI on a baseband data transmission system under operational conditions. It gives information on zero-crossing jitter, noise margin, timing sensitivity and time interval available for sampling.
- The three basic binary digital communication schemes are:
  (a) Amplitude Shift Keying (ASK) or ON-OFF Keying (OOK)
  (b) Frequency Shift Keying (FSK)
  (c) Phase Shift Keying
- *ASK* $s_1(t) = 0$ for $b_k = 0$ and $s_2(t) = A\cos\omega_c t$ for $b_k = 1$

  FSK $s_1(t) = A\cos(\omega_c - \omega_d)t$ for $b_k = 0$; $s_2(t) = A\cos(\omega_c + \omega_d)t$ for $b_k = 1$

  PSK $s_1(t) = -A\cos\omega_c t$ for $b_k = 0$; $s_2(t) = A\cos\omega_c t$ for $b_k = 1$

- *Spectra ASK* Main lobe width = $2/T$ and $B_T \cong \frac{1}{T} = R_b$

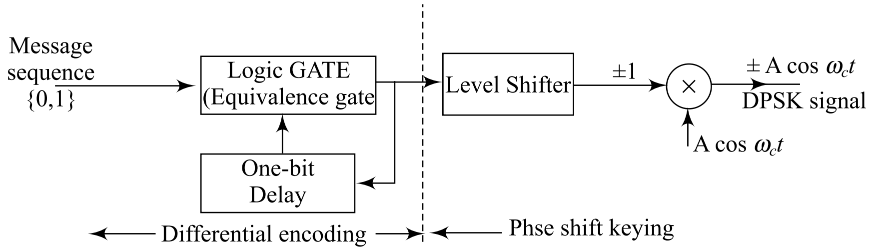
  Roll off $1/(f - f_c)^2$

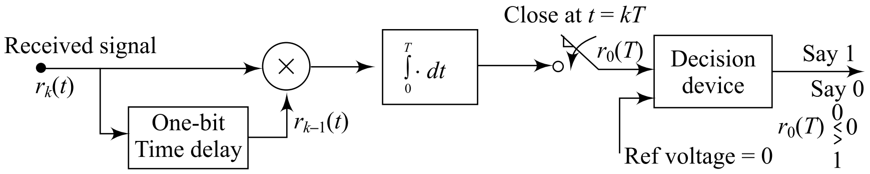
  FSK Main lobe width = $3/T$; Impulses at $f = f_c \pm \frac{1}{2T}$

  $B_T = \frac{1}{T} = R_b$; Roll off $1/(f - f_c)^4$

  PSK Main lobe width = $2/T$; $B_T \cong \frac{1}{T} = R_b$; Roll off $1/(f - f_c)^2$

- *DPSK principle* In DPSK, the waveform transmitted in a given time slot is $A\cos\omega_c t$ if the baseband binary digits in that time slot and the preceding one are alike; and the waveform transmitted is $-A\cos\omega_c t$ if they are not alike.





- *DEPSK (Differentially Encoded PSK)*
  - (a) Transmitters of DPSK and DEPSK are the same
  - (b) In DEPSK receiver, the one-bit delay element is used at the baseband level rather than at the carrier frequency level.
- *QPSK* QPSK is a bandwidth efficient band pass digital modulation scheme that makes use of quadrature multiplexing.

$$s_i(t) = \begin{cases} \sqrt{\dfrac{2E}{\tau}}\cos[2\pi f_c t + (2i-1)\pi/4]; & 0 \le t \le \tau; i = 1, 2, 3, 4, \\ 0; & \text{otherwise} \end{cases}$$

- *QPSK signal space* It is two dimensional. The signal constellation is shown in Fig. 10.38. $\phi_1(t) = \sqrt{2/\tau}\cos\omega_c t$, $\phi_2(t) = -\sqrt{2/\tau}\sin\omega_c t$
- Bandwidth and bandwidth efficiency of QPSK

  3-dB bandwidth of QPSK = $B_T = \dfrac{1}{2T}$ $\quad\therefore$ bandwidth efficiency = 2
- *Offset QPSK* This is a variant of QPSK in which the bit-stream for the quadrature component is delayed by half a symbol duration w.r.t, the inphase component bit stream. In this the carrier phase changes are confined to ±90°.
- *π/4 shifted QPSK* In this, the carrier phase changes are restricted to ±45° only. Thus, sudden changes in the amplitude of the envelope will have small magnitude.
- *MSK (Minimum Shift Keying)* It is a type of CPFSK that produces orthogonal signaling and is bandwidth efficient

$$\underset{\text{(MSK)}}{s(t)} = \sqrt{\frac{2E_b}{T}}[\cos\theta(t)\cos 2\pi f_c t - \sin\theta(t)\sin 2\pi f_c t]$$

  where $\quad \theta(t) = \phi(0) \pm \left(\dfrac{\pi}{2T}\right)t; 0 \le t \le T; + \text{sign for } b_k = 1, -\text{sign for } b_k = 0$

  $\phi(0)$ being the initial phase of the carrier.
- *Bandwidth efficiency and P_e of MSK*
  - (a) $\underset{\text{(3-dB)}}{B_T} = \dfrac{1}{2T} = \dfrac{R_b}{2}$ $\quad\therefore$ bandwidth efficiency, $\dfrac{R_b}{B_T} = 2$
  - (b) MSK signal's PSD falls off as the fourth power of $|(f - f_c)|$
  - (c) $\underset{\text{(MSK)}}{P_e} = \dfrac{1}{2}erfc\left[\sqrt{\dfrac{E_b}{\eta}}\right]$, same as that of QPSK and BPSK.
- *Gaussian MSK (GMSK)* This makes the power spectrum of MSK more compact and thus reduces the B_T. For this purpose, the NRZ binary data stream is passed through a pulse-shaping LPF whose frequency response is a Gaussian function of frequency. It is extensively used in GSM cellular mobile communications.
- *M-ary band pass signaling*
  - (a) If the bit energy is E_b and symbol energy is E, if R_b is binary bit rate and R_M the M-ary symbol rate, and if $n = \log_2 M$, then

$$R_M = \frac{R_b}{n}; E_b = \frac{E}{n}$$

- *QAM (Quadrature Amplitude Modulation)* In this, two carriers in phase quadrature are independently amplitude modulated by discrete amplitudes a_ks and b_ks.

$$\underset{\text{(M-ary QAM)}}{P_e} = 2\left(1 - \frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{3E_{\text{av}}}{2(M-1)\eta}}\right)$$

  Bandwidth efficiency: $\dfrac{R_b}{B_T} = \log_2 M$

  Signal space is two dimensional.

- *M-ary FSK* $P_e \le \frac{1}{2}(M-1)\,erfc\left(\sqrt{\frac{E}{2\eta}}\right)$

  Bandwidth efficiency = $\left(\frac{2\log_2 M}{M}\right)$
- *M-ary PSK* $s_k(t) = \sqrt{\frac{2E}{\tau}}\cos\left[\omega_c t + \frac{2\pi}{M}(k-1)\right]$

  $$k = 1, 2, 3, \ldots, M$$
- Signal space is 2-dimensional and all the M message signal points lie on the circumference of a circle with center at the origin and radius equal to $\sqrt{E}$ where E is symbol energy.
- $\underset{\text{(M-ary PSK)}}{P_e} \cong erfc\left(\sqrt{\frac{E}{\eta}}\sin\left(\frac{\pi}{n}\right)\right)$; ρ = Bandwidth efficiency = $\log_2 M$
- *Synchronization* For coherent band pass signaling systems, three levels of synchronization are needed – (a) Bit synchronization, (b) Word synchronization, and (c) Carrier synchronization. For non-coherent systems, the last one is not necessary.
- (i) An early-late synchronizer may be used for bit synchronization.
  (ii) A preamble is sent along with the information sequence in a time-multiplexed way periodically, for word/frame synchronization.
  (iii) An M-th power loop may be used for carrier recovery.

References and Suggested Reading

1. Carlson, A. Bruce, Paul B. Crilly and Janet C. Rutledge, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 4th Edition (International Edition), McGraw-Hill, 2002.
2. Couch II, Leon W., *Digital and Analog Communication Systems*, 6th Edition, Pearson Education Asia, 2001.
3. Gronemeyer, S. A., and A.L. Mc Bride, MSK and Offset QPSK Modulation, *IEEE Transactions on Communications*, Vol. 24, pp. 809–820, Aug. 1976.
4. Haykin, Simon, *Communication Systems*, 4th Edition, Wiley India, 2001.
5. Kobal, P., and Pasupathy, S., Partial Response Signaling, *IEEE Transactions on Communications*, Vol.23, pp. 921–934, 1975.
6. Lafrance, Pierre, *Fundamental Concepts of Communication*, Prentice Hall of India, New Delhi, 1992.
7. Lathi, B.P, *Modern Digital and Analog Communication Systems*, 3rd Edition, Oxford University Press, 1998.
8. Lee, Edward A., and David G. Messerschmitt, *Digital Communication*, 2nd Edition, Allied Publications Limited in Association with Kluwer Academic Publishers, 1994.
9. Pasupathy, S., Correlative Coding – A Bandwidth Efficient Signaling Scheme, *IEEE Communication Magazine*, Vol.15, No.4, pp. 4–11, 1977.
10. Pasupathy, S., Minimum Shift Keying – A Spectral Efficient Modulation, *IEEE Communication Magazine*, Vol.17, No. 4, pp. 14–22, 1979.
11. Pasupathy, S., Nyquist's Third Criterion, *Proceedings of IEEE*, Vol.62, pp. 860–861, 1974.
12. Proakis, John G., *Digital Communications*, 4th Edition (McGraw-Hill International Edition), McGraw-Hill, 2001.
13. Proakis, John G.,and Masoud Salehi, *Fundamentals of Communication Systems*, Low Priced Edition, Pearson Education, 2005.
14. Qureshi, S., Adaptive Equalization, *IEEE Communication Magazine*, Vol.20, No.2, pp.9–16, 1982.
15. Qureshi, S., Adaptive Equalization, *Proceedings of the IEEE*, Vol.73, pp.1349–1387, 1985.
16. Taub, Herbert, Donald Schilling, and Gautam Saha, *Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill, New Delhi, 2008.
17. Wozencraft, John M., and Irwin Mark Jacobs, *Principles of Communication Engineering*, John Wiley & Sons Inc, 1965.
18. Ziemer, R.E., and W.H. Tranter, *Principles of Communications – Systems, Modulation and Noise*, 3rd Edition, Jaico Publishing House, 2002.

Review Questions

1. What are 'line codes'? Name some popular line codes.
2. State the desirable properties of line codes.
3. Sketch the power spectra of (a) Polar NRZ, and (b) bipolar RZ signals.
4. Compare the various line codes, listing the merits and demerits of each.
5. With the help of a block schematic diagram, explain the basic structure of a binary baseband receiver.
6. Explain the meaning of the term: Probability of error.
7. Draw the block diagram of a coherent binary baseband receiver.
8. What is ISI and what is it that causes ISI?
9. Explain how even though ISI is present, its effect on P_e is avoided.
10. Write down the expression for, and sketch the spectrum and the waveform of the received pulse in the case of an ideal Nyquist channel.
11. What are the drawbacks of the ideal Nyquist channel?
12. What is a 'raised cosine spectrum'? How does it help us to avoid ISI?
13. Discuss the advantages and disadvantages of the ideal Nyquist channel and the channel with a raised cosine spectrum for the received pulse?
14. What is a correlative level coding? Draw the block schematic diagram of a duo-binary signaling scheme.
15. Draw sketches of (a) the magnitude response, (b) the phase response, and (c) the impulse response of a duo-binary conversion filter.
16. With reference to duo-binary signaling, explain how error propagation occurs.
17. What is 'pre-coding' and how does it help in avoiding error propagation that occurs in duo-binary signaling?
18. What is 'modified duo-binary signaling scheme'? Draw the block diagram of this signaling scheme. What are its advantages and disadvantages, as compared to the ordinary duo-binary signaling scheme without pre-coding?
19. Draw the sketches of (a) the magnitude response, (b) the phase response, and (c) the impulse response of a modified duo-binary conversion filter.
20. Distinguish between 'baud' and bit rate.
21. What are the advantages and disadvantages of M-ary signaling over binary signaling insofar as baseband data transmission is concerned?
22. What is a zero-forcing equalizer? Explain the basic principle of it.
23. What is the need for adaptive equalization?
24. What is the LMS algorithm? What are the various steps in using the LMS algorithm for adaptation of an equalizer?
25. What is meant by 'decision-directed adaptation'?
26. What is an eye pattern?
27. Draw a generalized eye pattern and label the various interpretations possible.
28. Name the three basic digital modulation schemes. Explain the way the carrier is changed in each case.
29. Given unipolar NRZ binary data stream, how do you generate ASK signals?
30. What is continuous phase FSK? How is a CPFSK signal generated? What is the advantage of CPFSK?
31. Distinguish between coherent and non-coherent reception.
32. With the help of a block schematic diagram, explain the operation of a non-coherent BFSK receiver.
33. With the help of a block schematic diagram, explain the operation of a coherent BPSK receiver.
34. What is Sunde's FSK? Show that Sunde's BFSK signals are orthogonal.
35. Explain the principle of DPSK encoding.
36. With the of neat block schematic diagram, explain the working of a DPSK transmitter and receiver.
37. Explain the difference between DPSK and DEPSK.
38. What is QPSK? Write down an expression for the signal set. Draw the signal space diagram and show the signal constellation.
39. Explain how a QPSK signal can be generated.
40. What is MSK? What are its advantages over ordinary binary FSK?
41. Draw the signal space diagram and show the signal constellation for an MSK signal.

42. Distinguish between signal constellations of QPSK and MSK systems.
43. What OQPSK? What are its advantages and disadvantages? Compare it with MSK.
44. What is $\pi/4$ – shifted QPSK? How is it an improvement over ordinary QPSK?
45. With reference to MSK, what does the dimensionless parameter, h, represent? For MSK, what is its value? What is the reason for choosing that particular value for h?
46. Draw the block diagram of an MSK transmitter and explain its operation.
47. Draw the block diagram of an MSK receiver and explain its operation.
48. Derive an expression for the power spectrum of an MSK signal, clearly stating all the assumptions made in the derivation.
49. What is GMSK? Sketch and compare the power spectra of an MSK signal and a GMSK signal.
50. What are the advantages and disadvantages of M-ary band pass signaling as compared to binary band pass signaling?
51. Write down the expression for the transmitted signal of an M-ary QAM signal. What is the dimension of its signal space? Choose some suitable orthonormal basis signals for this signal space.
52. Show the signal space diagram of 8-ary PSK. What is its dimension? Sketch the signal constellation.
53. Explain the principle of working of an ‘early-late bit synchronizer’.
54. Explain any method of ‘carrier recovery’ for a coherent band pass signaling system.

Problems

1. For the 9-bit binary bit stream 111001001, draw the signal waveforms, if the line-code used is (a) unipolar NRZ, (b) Polar NRZ, (c) bipolar RZ, and (d) Manchester coding.
2. Binary words, each consisting of 16 bits are being generated at the rate of 15,000 words/second. Find the bandwidth required to transmit the data so generated as (a) a binary PAM signal, and (b) as an M-ary signal with $M = 4$.
3. A computer outputs binary data at the rate of 64 kbps. The binary digits of its output are encoded into 4-level PAM signal by coding each set of two successive bits into one of four possible levels. The resulting signal is transmitted using a 4-level PAM system having a raised cosine spectrum with a roll-off factor $\rho = 0.5$. Determine the bandwidth required.
4. An analog signal band-limited to 6 kHz, is sampled at a rate of 20×10^3 samples/sec. The samples are then quantized into 256 levels and coded into M-ary amplitude pulses that satisfy Nyquist’s criterion with a roll-off factor $\rho = 0.2$. If these multi-amplitude pulses are to be transmitted over an available channel that has a bandwidth of 32 kHz, determine the minimum acceptable value of M.
5. The binary data stream 110010110 is applied to the input of a duo-binary system. Determine the duo-binary coder output and the corresponding receiver output.
6. Using Nyquist criterion pulses, binary data at a rate of 8 kbps is to be transmitted over a channel of bandwidth 6 kHz. What is the maximum value of the roll-off factor ρ that can be used?
7. Binary data is transmitted over a channel of bandwidth 5 kHz. Determine the maximum possible rate of transmission if (a) Polar signal with rectangular full-width pulses are used. (b) Polar signal using Nyquist criterion pulses with $\rho = 0.2$ are used.
8. The binary data stream 100011011 is applied to the input of a modified duo-binary system. Determine the output of the modified duo-binary coder and the corresponding receiver output (a) without a pre-coder, and (b) with a pre-coder.
9. In a binary data transmission using duo-binary conversion filter without a pre-coder, the received sample values $\{c_k\}$ were found to be:

2 2 0 –2 –2 0 2 –2 0 2 2 0

The starting b_k was 1.

(a) Do you feel there is an error in these values?
(b) If there is an error, can you guess the correct $\{b_k\}$ sequence? Is this unique?
(c) If the obtained $\{b_k\}$ sequence is not unique, write down all the possible correct $\{b_k\}$ sequences.

Note *More than one error in $\{c_k\}$ sequence is extremely unlikely.*

10. Design a 3-tap zero-forcing equalizer if the received pulse $p_r(t)$, in a particular binary communication system using Nyquist criterion pulses, has the following values at the sampling instants:

$$p_r(0)=1;\ p_r(T)=0.1;\ p_r(2T)=-0.03$$

$$p_r(-T)=0.25;\ p_r(-2T)=-0.06$$

11. Binary non-coherent FSK with a frequency shift of 170 Hz is commonly used in the amateur radio bands. What is the maximum data rate such systems can support?
12. The discontinuous phase binary FSK signal obtained by switching between two unsynchronized carrier oscillators one of frequency f_1 and the other of frequency f_2, can be considered to be the one obtained by interleaving two binary ASK signals. Using this approach, find the PSD of discontinuous phase BFSK signal.
13. In a BPSK system for a band limited channel with $B_T = 4$ kHz, the spectral envelope must be at least 30 dB below the maximum at frequencies outside the channel bandwidth. What is the maximum data rate R_b up to which this condition is satisfied?
14. For the input binary sequence 1100110011, sketch the waveforms of the inphase and quadrature components of a QPSK signal based on the signal set of Fig.10.40. Also sketch the QPSK signal waveform itself.
15. A random binary sequence of 1s and 0s occurring with equal probability and represented by polar NRZ modulates a carrier using offset QPSK modulation. Assuming that symbols in different time slots are statistically independent and identically distributed, determine the PSD of the modulated wave.
16. The M^{th} power loop shown in the Fig.10.65 has a phase ambiguity problem. Show that it exhibits M phase ambiguities in the interval $[0, 2\pi]$. How does the ambiguity problem arises and how can we overcome it?

Multiple-Choice Questions

1. The line code which suppresses the dc component is
   (a) bipolar return-to-zero code (b) Manchester code
   (c) polar non-return-to-zero code (d) unipolar non-return-to-zero code
2. The line code which has an inbuilt error-detecting capability against sign inversion of transmitted pulse due to channel noise, is
   (a) bipolar return-to-zero code (b) Manchester code
   (c) polar non-return-to-zero code (d) unipolar return-to-zero code
3. The line code which permits bit-timing recovery at the receiver, is
   (a) bipolar return-to-zero code (b) Manchester code
   (c) polar non-return-to-zero code (d) unipolar return-to-zero code
4. Using an ideal Nyquist channel, with an overall system bandwidth of W, we can transmit without ISI affecting the P_e, at a maximum rate of
   (a) $\frac{1}{2}W$ (b) W (c) $2W$ (d) $\frac{3}{2}W$
5. The rate of decrease of a sinc pulse $p(t)$ is
   (a) $\frac{1}{t}$ (b) $\frac{1}{|t|}$ (c) $\frac{1}{|t|^2}$ (d) None of these
6. A raised cosine filter with a roll-off factor $\rho = 1$ gives a transmission bandwidth B_T equal to
   (a) $2W$ (b) $0.5\ W$ (c) W (d) $1.5W$
7. A roll-off factor value for the raised cosine spectrum that enables the receiver to extract the timing signals required for synchronization, is
   (a) 0 (b) 1.5 (c) 0.5 (d) 1
8. For the same transmission bandwidth W, the maximum transmission speeds attainable by the ideal Nyquist channel and the duo-binary conversion filter, are respectively
   (a) $(2W, 2W)$ (b) $(2W, W)$ (c) $(W, 2W)$ (d) (W, W)

9. In a duo-binary system without pre-coder, the sample of the received signal in a particular time slot, was $c_k = 0$. Then the binary digit in that time slot is

(a) zero
(b) 'zero', provided the detected bit in the previous time slot was a 'one'
(c) 'zero', provided the detected bit in the previous time slot was also a 'zero'
(d) 'one', provided the detected bit in the previous time slot was also a 'one'

10. A duo-binary system with pre-coding has which of the following drawback?

(a) For a given bandwidth W, it can support a maximum transmission speed of W bits/sec.
(b) Error propagation
(c) The PSD of the transmitted pulse $p(t)$ is not zero at $f = 0$.
(d) None of the above

11. In quaternary baseband transmission, the ratio of transmission speeds in bits/sec and bauds is

(a) 2 (b) 4 (c) 0.5 (d) 0.25

12. The extent of maximum eye opening in the vertical direction indicates

(a) ISI (b) Timing sensitivity (c) Zero-crossing jitter (d) Noise margin

13. Which one of the following band pass digital modulation schemes is not suitable for transmission over non-linear band pass channels?

(a) FSK (b) ASK (c) PSK (d) QFSK

14. Which one of the following digital band pass modulation schemes cannot be detected non-coherently?

(a) FSK (b) ASK (c) PSK (d) Sunde's BFSK

15. Which one of the following digital band pass modulation schemes is a hybrid type?

(a) QAM (b) QPSK (c) MSK (d) Sunde's BFSK

16. Sunde's BFSK has a power spectrum with a rolloff factor inversely proportional to

(a) $(f - f_c)$ (b) $(f - f_c)^3$ (c) $(f - f_c)^2$ (d) $(f - f_c)^4$

17. In Sunde's BFSK, the difference in frequency between the two possible signal, is equal to

(a) $\frac{1}{T}$ (b) $\frac{2}{T}$ (c) $\frac{1}{2T}$ (d) $\frac{1}{3T}$

18. The envelope of band pass noise is

(a) uniformly distributed (b) Rayleigh distributed
(c) Ricean distributed (d) Gaussian distributed

19. The envelope of a sinusoid pulse band pass noise has

(a) normal distribution (b) uniform distribution
(c) Ricean distribution (d) Rayleigh distribution

20. In a 16-ary QAM system, the energy of the signal with the smallest amplitude is E_0. The distance between any two adjacent message points in the signal space, is

(a) $2\sqrt{E_0}$ (b) $\sqrt{E_0}$ (c) $2E_0$ (d) E_0

21. For any 4-ary FSK, the signal set is given by

$$s_k(t) = \sqrt{\frac{2E}{T}} \cos\left[\frac{\pi}{4}(n+k)t\right]; 0 \le t \le T$$

$$k = 1, 2, 3, 4$$

the dimension of its signal space is

(a) 1 (b) 2 (c) 3 (d) 4

22. For the signal set of MCQ 24, the distance between any two message signal points in the signal space is

(a) $2E$ (b) $2\sqrt{E}$ (c) $\sqrt{2E}$ (d) There is no specific answer

23. If ρ_1 and ρ_2 represent the bandwidth efficiencies of M-ary PSK and M-ary FSK respectively, as M is increased, ρ_1 and ρ_2 will (in that order)

(a) (increase, increase) (b) (increase, decrease)
(c) (decrease, increase) (d) (decrease, decrease)

24. The signal set for a 4-ary PSK system is given by

$$s_k(t) = \sqrt{\frac{2E}{T}} \cos\left[\omega_c t + \frac{\pi}{2}(k-1)\right]; 0 \le t \le T$$
$$k = 1, 2, 3, 4$$

The distance between adjacent message signal points in its signal space is given by

(a) $2E$ (b) $2\sqrt{E}$ (c) $\sqrt{E}$ (d) $\sqrt{2E}$

25. Which one of the following digital modulation scheme is not preferred when the channel is non-linear?

(a) QAM (b) BFSK (c) BPSK (d) MSK

Key to Multiple-Choice Questions

1. (b)	2. (a)	3. (d)	4. (c)	5. (b)	6. (a)	7. (d)	8. (a)
9. (b)	10. (c)	11. (a)	12. (d)	13. (b)	14. (c)	15. (a)	16. (d)
17. (a)	18. (b)	19. (c)	20. (a)	21. (d)	22. (c)	23. (b)	24. (d)
25. (a)							

NOISE PERFORMANCE OF DIGITAL COMMUNICATION SYSTEMS

11

> *"In learning to know other things, and other minds, we become more intimately acquainted with ourselves, and are to ourselves better worth knowing."*
>
> **Philip Gilbert Hamilton**
> *American author*

Learning Objectives

After going through this chapter, students will be able to

- *understand that probability of error, P_e, is an appropriate index for the noise performance of digital communication systems,*
- *apply statistical hypothesis testing techniques for determining probability of error of a digital communication system in the presence of Additive White Gaussian Noise (AWGN),*
- *realize that matched filter is the optimum filter for detection of known signals in the presence of AWGN,*
- *determine the matched filters for some simple signals,*
- *understand the equivalence of matched filtering and correlation, and can draw the correlation receiver structure for detection of a known signal in the presence of noise,*
- *draw the signal-space diagrams, sketch the signal constellations, identify the decision boundaries and can calculate the probabilities of error for various digital band pass modulation schemes like ASK, FSK, PSK, QAM, QPSK and MSK, and*
- *compare the various digital band pass modulation schemes with reference to their probabilities of error and their bandwidth efficiencies.*

11.1 INTRODUCTION

In this chapter, our focus will be on evaluation of the performance of digital communication systems. Naturally, any performance evaluation must be based on certain appropriate performance indices. Earlier in Section 1.8 of Chapter 1, we had mentioned that fidelity and 'bandwidth efficiency' constitute the two important performance indices of a communication system, whether it is an analog or a digital communication system. Fidelity, in the case of an analog communication system refers to the extent to which the baseband analog modulating signal waveform is faithfully reproduced at the output of the demodulator. As was stated earlier, the destination signal-to-noise ratio (*SNR*) serves as a good parameter that can be used for quantifying 'fidelity'. In the case of digital communication systems, the degree of faithful reproduction, at the decoder output, of the baseband digital sequence used at the transmitter, is dependent on how correctly the decoder is able to decide, during each time slot, whether what was transmitted during that time slot was a binary 1 or 0. Hence, in this case, the probability of error, P_e, which gives the average rate of occurrence of decoding errors, is the appropriate parameter for quantifying the '**fidelity**' of a system. Of course, P_e again depends upon the SNR at the input of the decoder. For a given channel, since the signal power at the input to the decoder depends upon the average transmitted power, the system that gives the smallest probability of error for a given average transmitted power is utilizing the communication resource, power, more efficiently, and so it is to be preferred. This again prompts us to examine the question of optimizing the receiver structure so as to minimize the probability of error.

Bandwidth efficiency, which indicates how efficiently the communication system is utilizing the other communication resource – the bandwidth, is defined in the case of digital communication systems as follows.

$$\text{Bandwidth efficiency } \underline{\Delta} \frac{R_b}{B_T}$$

where R_b denotes the transmission rate in bits per second and B_T is the transmission bandwidth utilized, in Hertz.

Hence, our study of the performance of digital communication systems will involve the analysis of baseband as well as various digital modulation systems with a view to determine their error probabilities and bandwidth efficiencies. In short, we would like to compare the systems on the basis of their probabilities of error for a fixed average transmitted power and bit rate, and their bandwidth efficiencies for a given bit rate and at a fixed probability of error, P_e.

11.2 BASEBAND RECEPTION AND PROBABILITY OF ERROR

Earlier in Chapter 9, we had studied the structure of a baseband PCM receiver (see Fig. 9.10). Now that we are concerned with the baseband reception of digital data in general, we do not need the D/A conversion and the reconstruction filter, which were required there for the recovery of the continuous wave message signal. The basic objective in baseband reception of digital data is one of recovering the original digital data (that was transmitted) with minimum possible errors although the transmitted digital signal suffers distortion due to bandwidth limitations of the channel, and gets corrupted by the noise added to it. Distortion of the digital signal leads to the phenomenon of inter-symbol interference (ISI) that causes P_e to increase. Various techniques for combating ISI have already been discussed in detail in Chapter 10. Hence, for the present, we shall consider only the additive noise of the channel and its effect on the portability of error. Hence, the basic structure of the baseband receiver for digital data will be as shown in Fig. 11.1.

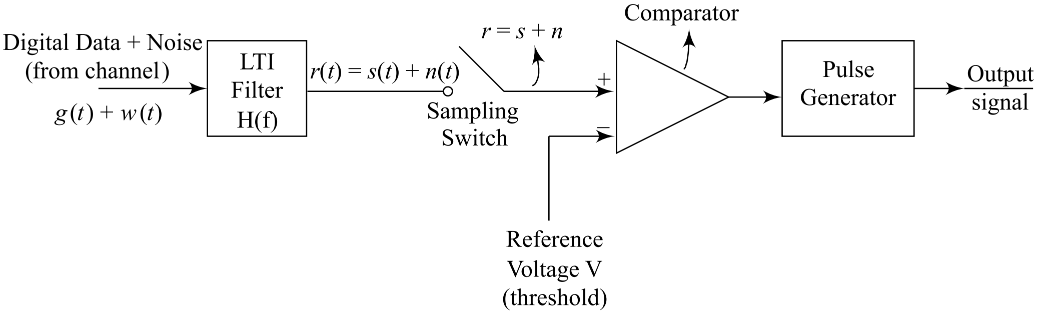


Fig. 11.1 *Basic structure of a baseband receiver*

Let us assume that the digital data is being transmitted using unipolar NRZ line code. Then during each time slot of say T seconds, $g(t)$ in the above figure takes a value of either $+A$ V or 0 V. Let $w(t)$, which represents channel noise, be a zero-mean white Gaussian noise. The sampling switch shown in the figure takes a sample of $r(t)$, the output of the filter, during each time slot. It is assumed that some arrangement has been made to ensure that the value of $s(t)$ in any time slot is not in any way affected by its value in the adjacent time slots. The sample of $r(t)$, namely r is just a voltage with a signal component and a noise component. The comparator compares this voltage r with the reference voltage V. If $r > V$, the comparator outputs a 1 that triggers the pulse generator to produce a clean positive rectangular pulse of amplitude A occupying a width T equal to the time-slot duration. If $r < V$, then the comparator outputs a 'zero' and the pulse generator does not produce any pulse. The case of r being exactly equals to V will have zero probability of occurrence and so need not be considered.

Thus, the '*observed variable*' for decision making is the random variable r, a real number. The total range of values that the observed variable can possibly take, constitutes what is called the '*Observation space*'. In this case, it is *one dimensional*, and is the real line, R. The observation space has thus been divided into two regions: (i) Region R_0 defined by $r < V$ and (ii) Region R_1 defined by $r > V$. In any time-slot, if the observed variable r falls in the region R_0, the receiver decides that what was transmitted during that time slot was a binary '0', and if the observed random variable falls in the region R_1, it decides that what was transmitted during that time slot was a binary '1'. The choice of V, the reference voltage, or the 'threshold', thus has a considerable bearing on the decision-making process. The question therefore arises – how can we choose V so that the wrong decisions are minimized? Further, since the decision is solely based on the value of the 'observed variable', r, and since it has a signal component, 's', and a noise component, 'n', is there anything we can do to make the decision more reliable by enhancing the signal component and reducing the noise component in r? The analysis that follows will provide answers for all these questions.

Since unipolar NRZ line code is assumed to have been used for transmission of the binary data, and since it is assumed that distortion of the signal during transmission through the channel will not be considered now, we have

$$r(t) = s(t) + n(t); \quad 0 \le t \le T = \text{Duration of time-slot} \tag{11.1}$$

where

$$s(t) = \begin{cases} A & \text{If Hypothesis } H_1 \text{ is true} \\ 0 & \text{if Hypothesis } H_0 \text{ is true} \end{cases} \tag{11.2}$$

Remarks

(i) Hypothesis H_1 assumes that a binary '1' was transmitted and hypothesis H_0 assumes that a binary '0' was transmitted during that time slot $0 \le t \le T$.

(ii) Since the channel noise w(t) has been assumed to be zero-mean white Gaussian noise and the filter is an LTI system, n(t) is also zero mean and Gaussian but not white.

(iii) A represents the amplitude of the pulse at the output of the filter. The filter, which is intended to reject the out-of-band noise, is assumed to have sufficient bandwidth to pass the pulse without distortion.

During each time slot, the sampler takes a sample of $r(t)$. Let it take the sample at $t = T_0$ in the time slot under consideration, i.e., $0 \le t \le T$. Then,

$$r(t)\big|_{t=T_0} = r(T_0) = s(T_0) + n(T_0); \quad 0 \le T_0 \le T \tag{11.3}$$

Since $n(t)$ is a zero-mean Gaussian random process, its sample $n(T_0)$ is a zero-mean Gaussian random variable. Let us represent its variance by $\sigma^2_{n(T_0)}$. Then the probability density function $p_{n(T_0)}(n(T_0))$ is given by

$$p_{n(T_0)}(n(T_0)) = \frac{1}{\sqrt{2\pi\sigma^2_{n(T_0)}}} e^{-n^2(T_0)/2\sigma^2_{n(T_0)}} \tag{11.4}$$

Now, we may rewrite Eq. (11.3) as follows

$$r(T_0)=\begin{cases} A+n(T_0) & \text{if } H_1 \text{ is true} \\ 0+n(T_0) & \text{if } H_0 \text{ is true} \end{cases} \tag{11.5}$$

Since $n(T_0)$ is a zero-mean Gaussian random variable with variance $\sigma^2_{n(T_0)}$ and A is a constant, $r(T_0)$ is a Gaussian random variable with the same variance as $\sigma^2_{n(T_0)}$ but of mean A if H_1 is true and of mean zero if H_0 is true. Thus, the density function of $r(T_0)$ conditioned on H_1 being true is

$$p_{r|H_1}(r(T_0)|H_1)=\frac{1}{\sqrt{2\pi\sigma^2_{n(T_0)}}}e^{-(r(T_0)-A)^2/2\sigma^2_{n(T_0)}} \tag{11.6}$$

Similarly, the density function of $r(T_0)$ conditioned on H_0 being true is

$$p_{r|H_0}(r(T_0)|H_0)=\frac{1}{\sqrt{2\pi\sigma^2_{n(T_0)}}}e^{-r^2(T_0)/2\sigma^2_{n(T_0)}} \tag{11.7}$$

A plot of the density function of $n(T_0)$ given by Eq. (11.4) is shown in Fig. 11.2 and the conditional density functions given by Eqs. (11.6) and (11.7) are shown in Fig. 11.3.

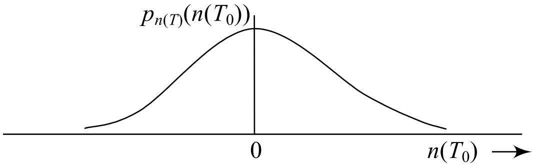


Fig. 11.2 *PDF of the random variable $n(T_0)$*

As stated earlier, we have arranged matters in such a way that when $r(T_0)$, the observed random variable, is greater than V, the receiver decides that a binary '1' was transmitted and when $r(T_0)$ is less than V, the receiver decides that a binary '0' was transmitted. However, the shaded area P_{e_0} under $p_{r|H_0}(r(T_0)\,|\,H_0)$ indicates that there is a non-zero probability (equal to P_{e_0}) that the observed variable $r(T_0)$ might take a value greater than V even when H_0 is true, i.e., there is a probability P_{e_0} of the receiver saying that a binary 1 was transmitted, even-though, in fact, a binary 0 was transmitted. Similarly, the area marked P_{e_1} shows that there is a probability P_{e_1} of the receiver saying that a binary 0 was transmitted, even though, in fact, a binary 1 was transmitted.

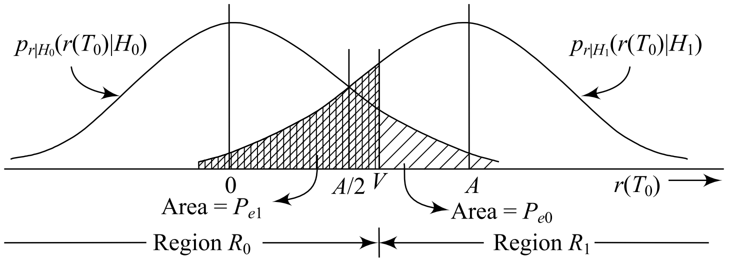


Fig. 11.3 *Conditional density functions, threshold V and the decision regions R_0 and R_1*

If P_1 denotes the probability of transmission of a binary 1 and P_0 the probability of transmission of a binary 0, the average probability of error, P_e may be written as

P_e = (probability of transmission of a 1) × (probability of the receiver misinterpreting the 1 as 0) + (probability of transmission of a 0) × (probability of the receiver misinterpreting the 0 as 1).

$$\therefore \qquad P_e = P_1P_{e_1} + P_0P_{e_0} \tag{11.8}$$

Referring to Fig. 11.3, we may write the expressions for P_{e_0} and P_{e_1} as

$$P_{e_0} = P[r(T_0) > V|H_0] = \int_V^{\infty} p_{r|H_0}(r(T_0)|H_0)\,dr \tag{11.9}$$

and

$$P_{e_1} = P[r(T_0) < V|H_1] = \int_{-\infty}^{V} p_{r|H_1}(r(T_0)|H_1)\,dr \tag{11.10}$$

From Fig. 11.3, it is clear that if V, the reference voltage, or the threshold voltage, is decreased, P_{e_0} increases while P_{e_1} decreases and when V is increased, P_{e_0} decreases while P_{e_1} increases. That is, as V is changed, while one type of errors decrease, the other type of errors increase. Since we are interested in minimizing the *average probability of error* P_e by an appropriate choice of the reference voltage, V, let V_{opt} be the value of V for which P_e takes a minimum value.

i.e.,

$$\frac{dP_e}{dV} = 0 \text{ when } V = V_{\text{opt}} \quad \text{i.e., when } r(T_0) = V_{\text{opt}} \tag{11.11}$$

∴ differentiating both sides of Eq. (11.8) with respect to V, we get

$$\frac{dP_e}{dV} = P_0\frac{dP_{e_0}}{dV} + P_1\frac{dP_{e_1}}{dV}$$

$$= P_0\frac{d}{dV}\left[\int_V^{\infty} p_{r|H_0}(r(T_0)|H_0)\,dr\right] + P_1\frac{d}{dV}\left[\int_{-\infty}^{V} p_{r|H_1}(r(T_0)|H_1)\,dr\right]$$

Hence, from Eq. (11.11)

$$\left.\frac{dP_e}{dV}\right|_{V=V_{\text{opt}}} = [-P_0 p_{r|H_0}(r(T_0)\,|\,H_0) + P_1 p_{r|H_1}(r(T_0)\,|\,H_1)]\,|_{r(T_0)=V_{\text{opt}}} = 0$$

i.e.,

$$P_0 p_{r|H_0}(r(T_0)|H_0)|_{r(T_0)=V_{\text{opt}}} = P_1 p_{r|H_1}(r(T_0)|H_1)|_{r(T_0)=V_{\text{opt}}}$$

i.e.,

$$\frac{P_1}{P_0} = \left.\frac{p_{r|H_0}(r(T_0)|H_0)}{p_{r|H_1}(r(T_0)|H_1)}\right|_{r(T_0)=V_{\text{opt}}} \tag{11.12}$$

If $P_1 = P_0 = 0.5$, as is generally the case in a communication scenario, Eq. (11.12) tells us that V_{opt} is that value of $r(T_0)$ for which $p_{r|H_0}(r(T_0)|H_0) = p_{r|H_1}(r(T_0)|H_1)$. That is, if $P_1 = P_0 = 0.5$, V_{opt} is that value of $r(T_0)$ which corresponds to the intersection of the two conditional density curves. From the symmetry of the two curves in Fig. 11.3, it means that

$$V_{\text{opt}} = A/2 \text{ for unipolar NRZ case} \tag{11.13}$$

When $P_1 \neq P_0$, we have to use Eq. (11.12) to find V_{opt}

Now,

$$P_{e_0} = \int_V^{\infty} p_{n(T_0)}(n(T_0))\,dn(T_0) = Q\left[\frac{V}{\sigma_{n(T_0)}}\right] \tag{11.14}$$

(where $Q(\cdot)$ is the Q-function. Refer to Appendix E)

and

$$P_{e_1} = \int_V^{\infty} p_{n(T_0)}(n(T_0) - A)\,dn(T_0) = Q\left[\frac{A - V}{\sigma_{n(T_0)}}\right] \tag{11.15}$$

Thus, when $P_0 = P_1$, since $V_{\text{opt}} = A/2$ and $P_{e_0} = P_{e_1}$,

$$P_e = 0.5P_{e_0} + 0.5P_{e_1} = P_{e_0} = P_{e_1} = Q\left[\frac{A}{2\sigma_{n(T_0)}}\right] \tag{11.16}$$

With $P_0 = P_1 = 0.5$, for a unipolar case, the average signal power at the input to the decision-making circuit

$$= S_R = P_0 \cdot (0)^2 + P_1 A^2 = 0.5A^2$$

$$\left.\begin{array}{l}\text{The average noise power at the}\\ \text{input to the decision-making circuit}\end{array}\right\} = N_R = \sigma^2_{n(T_0)} \quad \text{...(since } n(t) \text{ is of zero mean)}$$

$$\therefore \qquad \left(\frac{S}{N}\right)_R = \frac{A^2}{2\sigma^2_{n(T_0)}} \tag{11.17}$$

$\therefore$ from Eqs. (11.16) and (11.17), we may write

$$\underset{\text{(unipolar)}}{P_e} = \begin{array}{l}\text{Average probability of}\\ \text{error for unipolar NRZ case}\end{array} = Q\left[\frac{1}{2}\frac{A}{\sigma_{n(T)}}\right] = Q\left[\frac{1}{\sqrt{2}}\sqrt{\left(\frac{S}{N}\right)_R}\right] \tag{11.18}$$

Proceeding on similar lines, for the polar signaling case, it can be shown that

$$\underset{\text{(polar)}}{P_e} = Q\left[\sqrt{\left(\frac{S}{N}\right)_R}\right] \tag{11.19}$$

Since $Q(x)$ is a monotonically decreasing function of x, the value of $Q(x)$ decreases as x increases and vice versa. Thus we conclude that

1. For polar as well as unipolar signaling, the probability of error, P_e, decreases as the (S/N) at the input to the decision-making circuit (i.e., at the output of the filter) increases.
2. Other things remaining the same, polar signaling gives a smaller value of P_e than unipolar signaling.
3. If time-slot duration is T, the signaling rate $r = (1/T)$. But we know that to support a signaling rate of r pulses/sec, the low pass noise-limiting filter must have a cut-off bandwidth $B \geq r/2$. This means, N_R which is equal to $(\eta/2)(2B) = \eta B$, is given by

$$N_R = \eta B \geq \eta(r/2) \tag{11.20}$$

This shows that rapid signaling needs more signal power to get any specified P_e.

11.2.1 ML and MAP Detection Strategies

When the *a priori* probabilities P_0 and P_1 are equal, the value of the observed random variable $r(T_0)$ corresponding to the point of intersection of the two conditional density functions $p_r(r(T_0)|H_1)$ and $p_r(r(T_0)|H_0)$, i.e., $r(T_0) = A/2$ in Fig.11.3, is taken as the optimum threshold. Then whenever the value r_1 of the observed random variable $r(T_0)$ is greater than $A/2$,

$$p_r(r_1|H_1) > p_r(r_1|H_0)$$

and so in such a case, we ask the receiver to decide in favor of H_1. So, whenever $r_1 < A/2$, we have

$$p_r(r_1|H_0) > p_r(r_1|H_1)$$

So, in this case, we ask the receiver to decide in favor of H_0. The logic behind this decision-making strategy is that whenever $r_1 > A/2$, the observed value of the random variable $r(T_0)$, it is more likely that a binary 1 was transmitted (i.e., H_1 being true). Detection based on this decision-making strategy is referred to as the '***Maximum Likelihood Detection***' or ML detection. If r_1 is the value of the observed random variable $r(T_0)$, this ML detection strategy is

$$\left.\begin{aligned} &\text{Say } H_1 \text{ if}: \quad p_r(r_1|H_1) > p_r(r_1|H_0) \\ \text{and} \quad &\text{Say } H_0 \text{ if}: \quad p_r(r_1|H_0) > p_r(r_1|H_1) \end{aligned}\right\} \tag{11.21}$$

But suppose the *a priori* probabilities P_0 and P_1 are known and are unequal. In that case, as we have already seen, the value of $r(T_0)$ for which the two conditional probabilities are equal, i.e., the value of $r(T_0)$ corresponding to the point of intersection of the conditional probability curves, will not be the optimum threshold value that gives the least probability of error. It is some other value V_{opt} as defined by Eq. (11.12). So, now, if the value r_1 of the observed random variable $r(T_0)$ is greater than V_{opt}, the receiver decides in favor of H_1 and if r_1 is less than V_{opt}, the receiver decides in favor of H_0. When r_1 is greater than V_{opt},

$$P_1 p_r(r_1|H_1) > P_0 p_r(r_1|H_0)$$

This inequality tells us that when $r_1 > V_{\text{opt}}$, the probability of this being caused by a binary 1 having been transmitted is more than the probability of its being caused by a zero having been transmitted. Hence, in this case, the decision in favor of H_1 or H_0 is based on the probabilities of the binary symbols 1 and 0. Hence, detection based on this decision-making strategy, is called '***Maximum a posteriori detection***', or simply, MAP detection. In MAP detection therefore, the decision rule is

$$\left.\begin{aligned} &\text{Say } H_1 \text{ if}: \quad \frac{p_r(r_1|H_1)}{p_r(r_1|H_0)} > \frac{P_0}{P_1} \\ &\text{Say } H_0 \text{ if}: \quad \frac{p_r(r_1|H_0)}{p_r(r_1|H_1)} > \frac{P_1}{P_0} \end{aligned}\right\} \tag{11.22}$$

where r_1 is the value of the observed random variable $r(T_0)$. ***It is obvious from Eq. (11.22) that when P_0 equals P_1, MAP detection reduces to ML detection***.

11.2.2 ML Detection with Zero-Mean AWGN

If A_0 and A_1 are transmitted corresponding to H_0 and H_1 respectively

$$p_r(r(T_0)|H_0) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{\{r(T_0) - A_0\}^2}{2\sigma_n^2}\right]$$

and

$$p_r(r(T_0)|H_1) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{\{r(T_0) - A_1\}^2}{2\sigma_n^2}\right]$$

$\therefore$ if r_1 is the value of the observed random variable $r(T_0)$, and if, $p_r(r_1|H_1) > p_r(r_1|H_0)$, we say it is H_1 in ML detection.

i.e., if $\dfrac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\dfrac{\{r_1 - A_1\}^2}{2\sigma_n^2}\right] > \dfrac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\dfrac{\{r_1 - A_0\}^2}{2\sigma_n^2}\right]$: say H_1

when simplified, this reduces to the following decision rule:

Say H_1 if: $r_1 > \dfrac{A_0 + A_1}{2}$

and similarly, if $p_r(r_1|H_0) > p_r(r_1|H_1)$, receiver says H_0. When the Gaussian conditional densities are substituted, on simplification, this leads to the decision rule for ML detection in zero-mean AWGN as

Say H_0 if: $r_1 < \dfrac{A_0 + A_1}{2}$,

Note *We can take the logarithms on both sides and write like this because: (i) the density functions are always non-negative, and (ii) the logarithm is a monotonically increasing function of its argument.*

Thus, for ML detection, with zero-mean additive white Gaussian channel noise and A_0 and A_1 being the values received corresponding to binary '0' and binary '1', the decision rule of Eq. (11.21) reduces to

$$\left.\begin{aligned} &\text{if} \;:\; r_1 > \frac{(A_0 + A_1)}{2}, \text{ say } H_1 \\ \text{and} \quad &\text{if} \;:\; r_1 < \frac{(A_0 + A_1)}{2}, \text{ say } H_0 \end{aligned}\right\} \tag{11.23}$$

where r_1 is the value of the observed random variable.

Similarly in the case of MAP detection, if r_1 is the value of the observed random variable, the decision rule of Eq. (11.22) reduces to

$$\left.\begin{aligned} &\text{if}: \; r_1 > \frac{\sigma_{n_0}^2}{(A_1 - A_0)}\left[\log_e\left\{\frac{P_0}{P_1}\right\} + \frac{A_1^2 - A_0^2}{2\sigma_{n_0}^2}\right], \text{ say } H_1 \\ \text{and} \quad &\text{if}: \; r_1 < \frac{\sigma_{n_0}^2}{(A_1 - A_0)}\left[\log_e\left\{\frac{P_0}{P_1}\right\} + \frac{A_1^2 - A_0^2}{2\sigma_{n_0}^2}\right], \text{ say } H_0 \end{aligned}\right\} \tag{11.24}$$

Example 11.1 A polar NRZ waveform, taking the values +1 V for binary 1 and –1V for binary '0', corrupted by additive zero-mean white Gaussian noise whose variance is 0.2 V^2, is received by an MAP receiver. Determine the optimum threshold voltage for the receiver for each of the following *a priori* probabilities of transmission of a binary 1. (a) $P_1 = 0.5$, and (b) $P_1 = 0.3$.

Solution

(a) When $P_1 = P_0 = 0.5$, the MAP detector reduces to an ML detector and so

$$V_{\text{opt}} = \text{Threshold voltage} = \frac{1 + (-1)}{2} = 0 \text{ V}$$

(b) When $P_1 = 0.3$, $P_0 = 0.7$. Since, the threshold voltage for an MAP detector is given by (see Eq. (11.24))

$$V_{\text{opt}} = \frac{\sigma_{n_0}^2}{(A_1 - A_0)}\left[\log_e\left\{\frac{P_0}{P_1}\right\} + \frac{A_1^2 - A_0^2}{2\sigma_{n_0}^2}\right],$$

Substituting the given values in this

$$V_{\text{opt}} = \frac{0.2}{2}\left[\log_e\left\{\frac{0.7}{0.3}\right\} + 0\right] = 0.1 \times 0.8473 = 0.08473 \text{ V}$$

Example 11.2 A received binary baseband polar NRZ signal is either +1 V or –1 V. If $P(1) = 0.75$, assuming that zero-mean white Gaussian noise of variance 0.2 V^2 is corrupting the signal. (a) Find the threshold voltage to keep P_e to a minimum value. (b) Find the corresponding minimum P_e.

Solution

(a) *To find the threshold voltage, let us apply Eq. (11.24)*

$$V_{\text{opt}} = \frac{\sigma_{n_0}^2}{(A_1 - A_0)}\left[\log_e\left\{\frac{P_0}{P_1}\right\} + \frac{A_1^2 - A_0^2}{2\sigma_{n_0}^2}\right]$$

$$= \frac{0.2}{2}\left[\log_e\left\{\frac{0.25}{0.75}\right\} + 0\right] = -(0.1 \times 0.477) = -0.0477 \text{ V}$$

(b) *To find the corresponding minimum P_e*

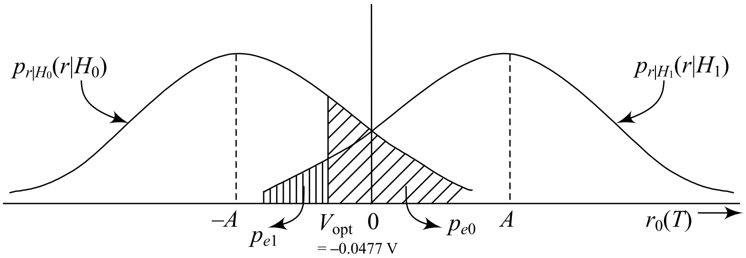


Fig. 11.4 *Conditional density functions for Example 11.2*

$$P_{e_{\min}} = P_0 \cdot P_{e_0} + P_1 \cdot P_{e_1}$$

$$= 0.25\left[\int_{V_{\text{opt}}}^{\infty} p_{r|H_0}(r|H_0)\right]dr + 0.75\left[\int_{-\infty}^{V_{\text{opt}}} p_{r|H_1}(r|H_1)\right]dr$$

$$= 0.25Q\left[\frac{1}{\sqrt{0.2}}\right] + 0.75Q\left[\frac{1.0477}{\sqrt{0.2}}\right]$$

$$= 0.25Q[2.236] + 0.75Q[2.342]$$

$$= \frac{0.25}{2}erfc\left[\frac{2.236}{\sqrt{2}}\right] + \frac{0.75}{2}erfc\left[\frac{2.342}{\sqrt{2}}\right]$$

$$= 0.125\left[1 - erf\left\{\frac{2.236}{1.414}\right\}\right] + 0.375\left[1 - erf\left\{\frac{2.342}{1.414}\right\}\right]$$

By referring to the error function tables, and simplifying:

$$P_{e_{\min}} = 0.01 = 10^{-2}$$

11.2.3 Optimum Filter

In our discussion on the basic structure of a baseband receiver, it was shown that (refer to Eqs. (11.18) and (11.19)) the probability of error decreases monotonically as the signal-to-noise ratio at the input to the comparator (decision device) is increased. So, we now turn our attention to the problem of increasing the (*S*/*N*) at the input to the decision device, i.e., at the output of the filter. Although we had originally introduced this filter for the purpose of not allowing the out-of-band noise from entering the receiver, in the light of our above requirement, we would like to look at this filter from the point of view of obtaining the maximum possible (*S*/*N*) at its output so that the decision device takes decisions under the best possible condition, i.e., when the signal-to-noise ratio at its input is maximum. In order to make our analysis a little more general, we shall assume that the channel noise is only zero-mean Gaussian *and not* zero-mean white Gaussian noise.

A filter, whose transfer function is such that it maximizes the signal-to-noise ratio at its output under these conditions, is referred to as an '*optimum filter*'. Further, in order to make the analysis applicable to both binary baseband reception as well as binary band pass reception we shall assume that corresponding to a binary 1, a signal $s_2(t)$ will be transmitted and that corresponding to a binary '0', a signal $s_1(t)$ will be transmitted. So, when baseband signaling is being considered,

$$s_2(t) = +A$$

and

$$s_1(t) = \begin{cases} 0 & \text{for unipolar signaling} \\ -A & \text{for polar signaling} \end{cases}$$

On the other hand, when band pass signaling is being considered we interpret $s_2(t)$ and $s_1(t)$ appropriately depending upon whether it is ASK, FSK or PSK.

Now, we shall determine the transfer function, $H_{\text{opt}}(f)$, of the optimum filter for binary signaling. But, before we proceed with that, we need to discuss Schwarz's inequality, an inequality that we will be using in the derivation of $H_{\text{opt}}(f)$.

11.2.4 Transfer Function of Optimum Filter

For the purpose of determining the transfer function of an optimum filter, let us slightly modify Fig. 11.1 and consider a receiver structure shown below in Fig. 11.5.

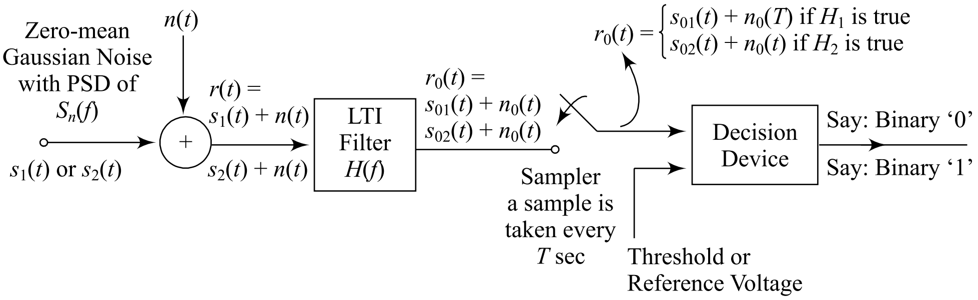


Fig. 11.5 *Receiver structure considered for determination of $H_{opt}(f)$*

The input to the filter is either $s_1(t) + n(t)$ or $s_2(t) + n(t)$ depending on whether hypothesis H_1 is true or hypothesis H_2 is true in the time slot being considered. Since the filter is LTI, at its output also the signal and noise components will be separate and the output is

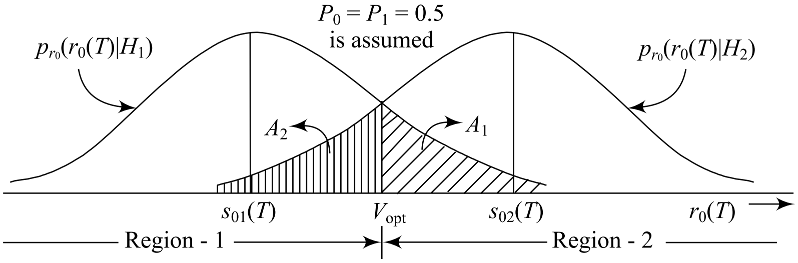
$$r_0(t) = \begin{cases} s_{01}(t) + n_0(t) & \text{if } H_1 \text{ is true; } 0 \le t \le T \\ s_{02}(t) + n_0(t) & \text{if } H_2 \text{ is true; } 0 \le t \le T \end{cases} \tag{11.25}$$

The sampler samples the filter output, taking one sample every T seconds, where T is the time-slot duration, i.e., the duration of a binary '0' or a binary '1'. The sampler output, which is the input to the decision device, may therefore be written as

$$r_0(T) = \begin{cases} s_{01}(T) + n_0(T) & \text{if } H_1 \text{ is true} \\ s_{02}(T) + n_0(T) & \text{if } H_2 \text{ is true} \end{cases} \tag{11.26}$$

Since $n(t)$, the channel noise is zero-mean Gaussian and since the filter is LTI, the noise component at the filter output, viz., $n_0(t)$ is also a zero-mean Gaussian process. Further, $s_{01}(t)$ is the filter output corresponding to $s_1(t)$, the signal component at its input if H_1 is true; and $s_{02}(t)$ is the filter output corresponding to $s_2(t)$, the signal component at its input if H_2 is true. $s_{01}(T)$ and $s_{02}(T)$ are constants (whose values depend on the filter's

H(*f*), the sampling instant *T* and on $s_1(t)$ and $s_2(t)$ respectively) and $n_0(T)$ is a zero-mean Gaussian random variable with variance $\sigma_{n_0}^2$, the same as that of $n_0(T)$. A sketch of the conditional densities $p_{r_0}(r_0(T)|H_1)$ and $p_{r_0}(r_0(T)|H_2)$ is shown in Fig.11.6. It is assumed that $s_{02}(T) > s_{01}(T)$.



If $r_0(T) > V_{\text{opt}}$, receiver says $s_2(t)$ was transmitted

$r_0(T) < V_{\text{opt}}$, receiver says $s_1(t)$ was transmitted

Fig. 11.6 *Conditional densities* $p_{r_0}(r_0(T)|H_1)$ *and* $p_{r_0}(r_0(T)|H_2)$

Since it has been assumed that $P_0 = P_1 = 0.5$, as pointed out earlier, the value of $r_0(T)$ for which we get the intersection of the two conditional densities, gives us the optimum value of the threshold or reference voltage. If H_1 is true,

$$r_0(T) = s_{01}(T) + n_0(T) \tag{11.27}$$

So, an error will occur if $n_0(T)$ is large enough to make

$$r_0(T) > V_{\text{opt}} \tag{11.28}$$

But
$$V_{\text{opt}} = \frac{1}{2}[s_{01}(T) + s_{02}(T)] \tag{11.29}$$

$\therefore$ an error will occur, if

$$r_0(T) = [s_{01}(T) + n_0(T)] > V_{\text{opt}} = \frac{1}{2}[s_{01}(T) + s_{02}(T)]$$

i.e., if
$$n_0(T) > \frac{s_{02}(T) - s_{01}(T)}{2} \tag{11.30}$$

The probability of occurrence of this = shaded area A_1.

$\therefore$ P_{e0} = Probability of the receiver misinterpreting a transmitted zero as a 1.
= Shaded area A_1

Similarly, therefore, P_{e1} = Probability of the receiver misinterpreting a transmitted 1 as a 0.
= Shaded area A_2

$\therefore$ average probability of error = $P_e = P_0 \cdot A_1 + P_1 \cdot A_2$ = Area A_1

(Since $P_0 = P_1 = 0.5$ and also $A_1 = A_{2.}$)

$\therefore$
$$P_e = \text{Area } A_1 = \int_{V_{\text{opt}} = \left[\frac{s_{01}(T) + s_{02}(T)}{2}\right]}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{n_0}^2}} e^{-[r_0(T) - s_{01}(T)]^2/2\sigma_{n_0}^2} dr_0(T) \tag{11.31}$$

Putting $z = \dfrac{r_0(T) - s_{01}(T)}{\sigma_{n_0}}$, we have $dr_0(T) = \sigma_{n_0} \cdot dz$, and $r_0(T) = z\sigma_{n_0} + s_{01}(T)$

when $r_0(T) = V_{\text{opt}} = \dfrac{s_{01}(T) + s_{02}(T)}{2}$, $\quad z = \dfrac{s_{02}(T) - s_{01}(T)}{2\sigma_{n_0}}$

Hence, Eq. (11.31) may be rewritten as

$$\therefore \qquad P_e = \frac{1}{\sqrt{2\pi}} \int_{\frac{s_{02}(T)-s_{01}(T)}{2\sigma_{n_0}}}^{\infty} e^{-z^2/2} dz = Q\left[\frac{s_{02}(T) - s_{01}(T)}{2\sigma_{n_0}}\right] = Q\left[\frac{d}{2}\right] \tag{11.32}$$

$$\text{where} \qquad d \,\underline{\Delta} \left[\frac{s_{02}(T) - s_{01}(T)}{\sigma_{n_0}}\right]$$

For the purpose of drawing Fig. 11.6 and for this derivation, we have assumed that $s_{02}(T) > s_{01}(T)$, since the values of these constants depend, as mentioned earlier, on the nature of the signals $s_{01}(t)$ and $s_{02}(t)$, and the value of *T*. So, in general

$$d = \left[\frac{|s_{02}(T) - s_{01}(T)|}{\sigma_{n_0}}\right] \tag{11.33}$$

Since $P_e = Q\left[\dfrac{d}{2}\right]$ from Eq. (11.32), and since the *Q*-function is a monotonically decreasing function of its argument, it follows that

$$P_{e\min} = Q\left[\frac{d_{\max}}{2}\right] \tag{11.34}$$

i.e., to minimize the probability of error, we have to maximize '*d*'. Hence, we have to choose *H*(*f*) in such a way that it maximizes *d* or d^2.

Now, suppose we define a signal *p*(*t*) as:

$$p(t) \,\underline{\Delta} \begin{cases} [s_2(t) - s_1(t)]; & 0 \le t \le T \\ 0 & ; \ \text{otherwise} \end{cases} \tag{11.35}$$

If *h*(*t*) is the impulse response of the filter, we have

$$p_0(T) = [s_{02}(T) - s_{01}(T)] = \int_0^T [s_2(\tau) - s_1(\tau)] h(T - \tau) d\tau \tag{11.36}$$

Since *p*(*t*) = 0 for *t* < 0 as well as for *t* > *T*, and since $h(T - \tau) = 0$ for $\tau > T$ in the case of a physically realizable (causal) filter, we may change the limits of integration in the RHS of Eq. (11.36) and write

$$p_0(T) = [s_{02}(T) - s_{01}(T)] = \int_{-\infty}^{\infty} p(\tau) h(T - \tau) d\tau \tag{11.37}$$

Further, if $P(f) \,\underline{\Delta}\, \mathcal{F}[p(t)]$, then

$$P_0(f) = P(f)H(f)$$

$$\therefore \qquad p_0(T) = \{\mathcal{F}^{-1}[P_0(f)]\}|_{t=T} = \int_{-\infty}^{\infty} \{P(f)H(f)\} e^{j2\pi fT} df \tag{11.38}$$

$$\therefore \qquad d^2 \,\underline{\Delta}\, \frac{|s_{02}(T) - s_{01}(T)|^2}{\sigma_{n_0}^2} \tag{11.39}$$

But

$$\begin{aligned}\sigma_{n_0}^2 &= \text{Noise variance at the output of the filter}\\ &= \text{Noise power at the output of the filter}\\ &\quad \text{(since the noise is zero-mean)}\\ &= \int_{-\infty}^{\infty} \{\text{PSD of the noise at the output of the filter}\}\, df\\ &= \int_{-\infty}^{\infty} S_n(f)\,|H(f)|^2 df \end{aligned} \tag{11.40}$$

$$\therefore \qquad d^2 = \frac{\left|\int_{-\infty}^{\infty} \{P(f)H(f)\}e^{j2\pi fT} df\right|^2}{\int_{-\infty}^{\infty} S_n(f)|H(f)|^2 df} = \left[\frac{p_0^2(T)}{\sigma_{n_0}^2}\right] \tag{11.41}$$

In order to minimize P_e, we have to maximize d^2. For this, we shall make use of 'Schwarz's Inequality' which can be stated for frequency domain functions as

$$\left|\int_{-\infty}^{\infty} U(f)V^*(f)df\right|^2 \le \left[\int_{-\infty}^{\infty} |U(f)|^2 df\right]\left[\int_{-\infty}^{\infty} |V(f)|^2 df\right]$$

or alternatively as

$$\frac{\left|\int_{-\infty}^{\infty} U(f)V^*(f)df\right|^2}{\int_{-\infty}^{\infty} |U(f)|^2 df} \le \int_{-\infty}^{\infty} |V(f)|^2 df \tag{11.42}$$

where the equality sign holds if and only if

$$U(f) = k \cdot V(f)\text{, where } k \text{ is a constant} \tag{11.43}$$

To apply Schwarz's inequality to Eq. (11.41), let us put

$$U(f) = k \cdot H(f) \cdot \sqrt{S_n(f)} \tag{11.44}$$

and

$$V(f) = k\frac{P^*(f)e^{-j2\pi fT}}{\sqrt{S_n(f)}} \tag{11.45}$$

where k is a constant.

Then the maximum value of d^2 is obtained (corresponding to the condition for the equality sign to hold good, as stated in Eq. (11.43) when

$$H(f)\sqrt{S_n(f)} = k\frac{P^*(f)e^{-j2\pi fT}}{\sqrt{S_n(f)}}$$

i.e., when

$$\boxed{H_{\text{opt}}(f) = k\left[\frac{P^*(f)e^{-j2\pi fT}}{S_n(f)}\right]} \tag{11.46}$$

where k is a constant.

Note that $|H_{\text{opt}}(f)|$ is directly proportional to $|P(f)|$ and inversely proportional to $S_n(f)$, the noise PSD. ***Thus, the optimum filter emphasizes frequencies where the signal spectrum is large and de-emphasizes those frequencies where the noise spectrum has large amplitudes. Thus, it improves the signal-to-noise ratio***.

Now, to find the probability of error, P_e when the optimum filter is used, we note that d^2 takes the maximum value when $H(f) = H_{\text{opt}}(f)$ in Eq. (11.41). Hence, substituting $H_{\text{opt}}(f)$ for $H(f)$ in Eq. (11.41) and using Eq. (11.46) for $H_{\text{opt}}(f)$, we get

$$\left[\frac{p_0^2(T)}{\sigma_{n_0}^2}\right]_{\max} = d_{\max}^2 = \frac{\left|\int_{-\infty}^{\infty}\frac{|P(f)|^2}{S_n(f)}df\right|^2}{\int_{-\infty}^{\infty}\frac{|P(f)|^2}{S_n(f)}df} = \int_{-\infty}^{\infty}\frac{|P(f)|^2}{S_n(f)}df \tag{11.47}$$

$\therefore \quad P_{e_{\min}} = Q\left|\frac{d_{\max}}{2}\right|$, where $d_{\max}$ is as given in Eq. (11.47)

$$P_{e_{\min}} = \frac{1}{2}erfc\left[\frac{d_{\max}}{2\sqrt{2}}\right] \tag{11.48}$$

11.2.5 Matched Filter

We have derived the transfer function $H_{\text{opt}}(f)$ of the optimum filter assuming the channel noise to have a PSD equal to $S_n(f)$, i.e., we have derived it for a general case. However, in case the channel noise is white and has a PSD (two-sided) of say $(\eta/2)$, as is generally assumed, then the $H_{\text{opt}}(f)$ is called a '*matched filter*'. Hence, from Eq. (11.46), we have

$$\boxed{H(f) = k\left[\frac{P^*(f)e^{-j2\pi fT}}{(\eta/2)}\right] = \alpha P^*(f)e^{-j2\pi fT}} \tag{11.49}$$

Since $\eta/2$ is a constant and the constant of proportionality, k, is arbitrary, we combine them and represent by a scaling factor α.

Thus, the transfer function $H(f)$ of the matched filter depends only on the signal $p(t) = [s_2(t) - s_1(t)]$ and that is why it is called a 'matched filter', indicating that the filter is matched to the signal to give minimum probability of error. Taking the inverse Fourier transform on both sides of Eq. (11.49), we get the impulse response $h(t)$ of the matched filter as:

$$h(t) = \alpha p(T-t)\,;\, 0 \le t \le T \tag{11.50}$$

$$= \alpha[s_2(T-t) - s_1(T-t)] \tag{11.51}$$

Equation (11.50) shows that the impulse response, $h(t)$, of the matched filter is obtained by reversing $p(t)$ in time and then shifting it by T sec to the right along the time axis. For example, if $p(t)$ is as shown in Fig. 11.7(a), the waveform of the impulse response of its matched filter is as shown in Fig. 11.7(c) waveform of time reversed $p(t)$ is shown in Fig. 11.7(b).

Probability of error with matched filter Substituting in Eq. (11.47) $\eta/2$ for $S_n(f)$, the two-sided power spectral density of the channel noise, we get

$$d_{\max}^2 = \frac{2}{\eta}\int_{-\infty}^{\infty}|P(f)|^2 df$$

where $\quad P(f) = \mathcal{F}[p(t)]$ and $p(t) = [s_2(t) - s_1(t)]$

We know from Parseval's theorem (also known as Rayleigh's theorem), that E_p the energy of the finite energy signal $p(t)$ is given by

$$E_p = \int_{-\infty}^{\infty} |p(t)|^2 dt = \int_{-\infty}^{\infty} |P(f)|^2 df$$

$$\therefore \quad d^2_{\max} = \frac{2}{\eta} E_p \tag{11.52}$$

1. **For the unipolar case:** Binary 1 represented by a positive rectangular pulse of A V over 0 to T sec and binary 0 represented by a no pulse, i.e., 0 V over 0 to T sec. That is $s_2(t) = A \; ; 0 \le t \le T$ and $s_1(t) = 0 \; ; 0 \le t \le T$.

$$\therefore \quad p(t) = (A - 0) \quad ; \quad 0 \le t \le T$$

and $\quad E_p = A^2 T$

$$\therefore \quad d^2_{\max} = \frac{2}{\eta} E_p = \frac{2A^2T}{\eta}$$

$$\therefore \quad d_{\max} = \sqrt{\frac{2A^2T}{\eta}}$$

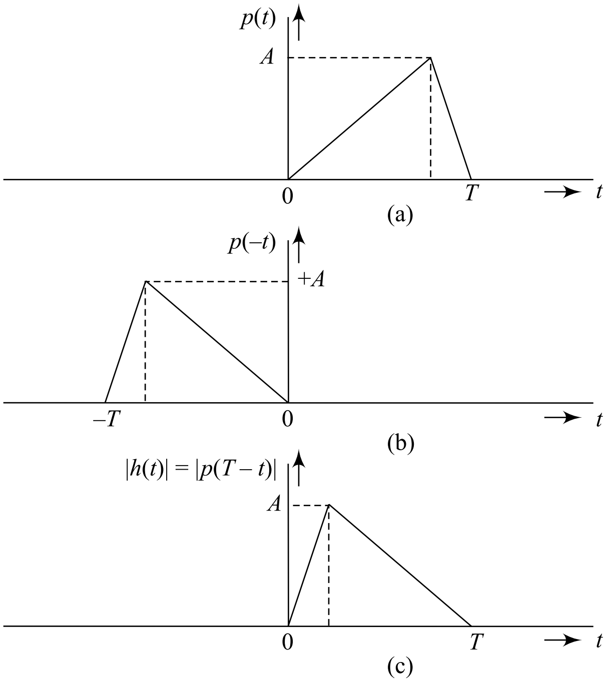


Fig. 11.7 *(a) Signal p(t), (b) Signal p(–t), (c) Signal p(T – t)*

and $$\frac{d_{\max}}{2} = \sqrt{\frac{A^2T}{2}\frac{1}{\eta}} \tag{11.53}$$

Assuming that the *a priori* probabilities of a 1 and a 0 are equal, since the pulses of amplitude A occur only for 50% of time, the average energy per time slot is $\left(\frac{A^2T}{2}\right)$. Representing the average energy per time slot by E_b, we have

$$\boxed{\underset{\substack{\text{matched filter}\\ \text{unipolar}}}{P_e} = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{E_b}{\eta}}\right]} \tag{11.54}$$

where $E_b = \frac{A^2T}{2}$ as stated earlier.

2. **For the polar case:** $s_2(t) = A$ and $s_1(t) = -A$

$\therefore$ In this case, $p(t) = s_2(t) - s_1(t) = 2A$

$$\therefore \qquad E_p = \int_0^T p^2(t)\,dt = 4A^2T$$

Hence, $$d^2_{\max} = \frac{2}{\eta} E_p = \frac{8A^2T}{\eta} \quad \text{and} \quad d_{\max} = 2\sqrt{\frac{2A^2T}{\eta}}$$

But in the polar case, E_b, the average energy per time slot is given by A^2T since a pulse of amplitude either $+A$ or $-A$ exists in every time slot.

$\therefore$ E_b = Average energy/time-slot = A^2T

$$\therefore \qquad d_{\max} = 2\sqrt{\frac{2E_b}{\eta}}$$

Hence,
$$P_e = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{2E_b}{\eta}}\right]$$

$$\therefore \qquad \boxed{\underset{\substack{\text{matched filter}\\ \text{polar case}}}{P_e} = Q\left[\sqrt{\frac{2E_b}{\eta}}\right]} \tag{11.55}$$

where E_b = Average energy per time slot = A^2T.

Matched filter for a rectangular pulse signal We will show that the matched filter for a rectangular pulse signal is what is called the 'integrate and dump' circuit. The integrate-and-dump receiver assumes importance because, for unipolar as well as polar signaling, $p(t) = [s_2(t) - s_1(t)]$ will be a *rectangular pulse* of width T, the duration of a time slot.

So, let $p(t)$ be a rectangular pulse of duration T sec, i.e.,

$$p(t) = \begin{cases} A; & 0 \le t \le T \\ 0\,; & \text{elsewhere} \end{cases} \tag{11.56}$$

Then, we know, from Eq. (11.46) that the impulse response of the matched filter for this signal, is given by

$$h(t) = \alpha\, p(T - t) = \alpha\, p(t) = \begin{cases} A; & 0 \le t \le T \\ 0\,; & \text{elsewhere} \end{cases} \tag{11.57}$$

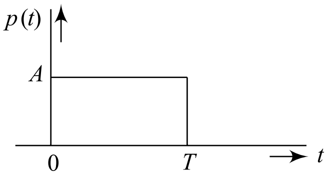


Fig. 11.8 *Rectangular pulse signal*

This is because, when a rectangular pulse of duration T sec extending from $t = 0$ to T is reversed in time and then translated in time to the right by T sec, we again get only the same rectangular pulse we started with. Thus, waveform of $h(t)$ is the same as $p(t)$, shown in Fig. 11.8.

Now, consider an ideal integrator. If we give a unit impulse as input to it, we know that we get $u(t)$.

So, if we allow the $\delta(t)$ to be integrated by the ideal integrator from $t = 0$ to T and suddenly bring the output of the integrator to zero, we get a rectangular pulse of duration T sec, extending from $t = 0$ to $t = T$. Thus, the impulse response of the matched filter for a 'rectangular pulse' also has exactly the same waveform as the output waveform of an 'integrate-and-dump' circuit.

Hence, we may conclude that the matched filter for a rectangular pulse extending from 0 to T sec is just an integrator operating from 0 to T sec whose output is dumped at $t = T$ sec.

Since the integration is only from 0 to T sec and the integrator should start afresh in the next time slot, we have to ensure that the energy storage devices in the integrator are all discharged at the end of each T sec duration, i.e., at the end of each time slot. That is why this matched filter configuration is called 'integrate-and-dump' circuit. A binary baseband receiver using an 'integrate-and-dump' implementation of the matched filter is shown in Fig. 11.9.

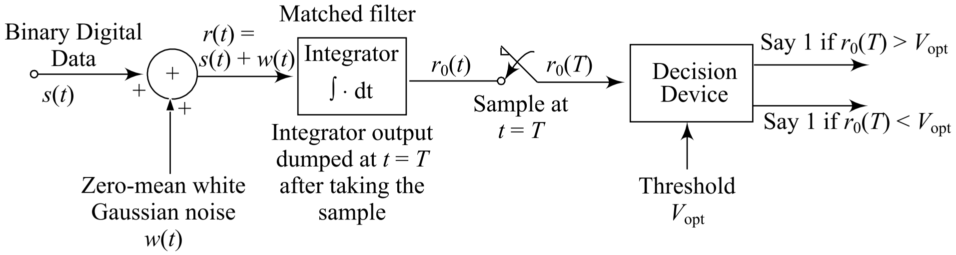


Fig. 11.9 *An 'integrate-and-dump' type of binary baseband receiver*

Before we conclude our discussion on the integrate-and-dump type of receiver, it is instructive to examine how this matched filter maximizes the signal-to-noise ratio of the observed random variable $r_0(T)$. Since $r_0(t) = s_0(t) + n_0(t)$,

$$\begin{aligned} r_0(T) &= s_0(T) + n_0(T) \\ &= \int_0^T s(t)dt + \int_0^T w(t)dt \\ &= \begin{cases} s_{01}(T) + n_0(T); & \text{if } s_1(t) \text{ is transmitted} \\ s_{02}(T) + n_0(T); & \text{if } s_2(t) \text{ is transmitted} \end{cases} \end{aligned} \tag{11.58}$$

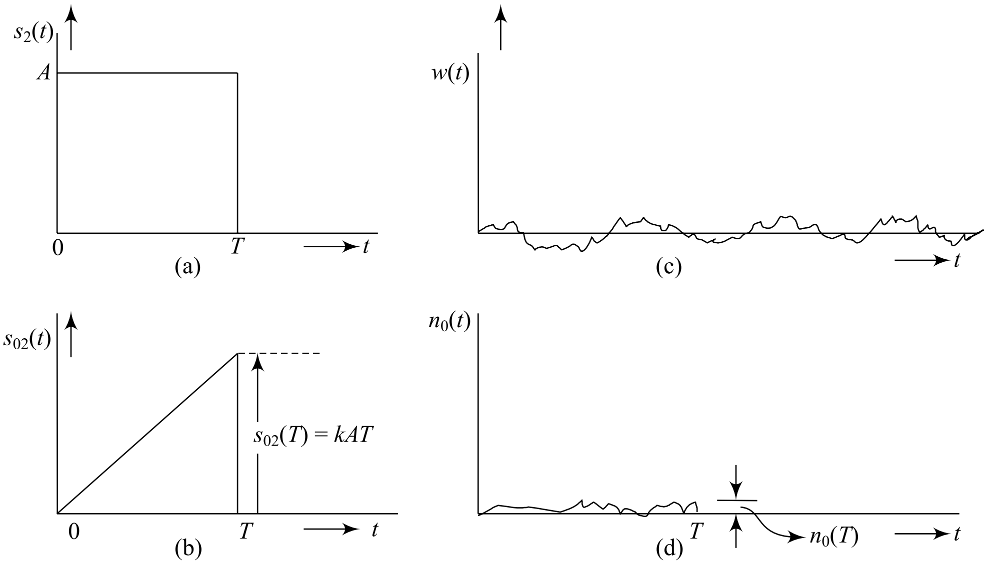


Fig. 11.10 *Illustrating the effect of integration of the rectangular signal and the zero-mean white Gaussian noise: (a) Rectangular signal, (b) Triangular pulse with peak value at t = T, (c) Zero-mean noise, (d) Noise after integration*

As can be seen from Fig. 11.10(b), the signal component at the output of the matched filter (here, an integrator) goes on increasing with time up to $t = T$, the instant at which the input pulse ceases to exist. If the integration process is allowed to continue beyond $t = T$ and if $s(t) = 0$ from $t = T$ to $2T$ (i.e., the next time slot) the integrator output would remain constant at the value kAT which it has attained at $t = T$ (as shown by the dotted line). But, we dump the output when the next time slot starts. Thus, the signal component of $r_0(t)$ is maximized at $t = T$.

The noise, being random, and of zero mean, when integrated over 0 to T, would however, be almost zero at the output of the filter. As the time of integration, T, cannot be extremely large, $n_0(T)$ will not be zero. It may have some small value. In fact, what is important is, that when the noise is integrated, the noise $n_0(t)$ at the output of the integrator will have a variance that goes on decreasing with the time for which the integration is carried out (see Example 11.3). Thus, while the power of the signal component $s_0(T)$ increases with time of integration, T, the variance of the noise component $n_0(T)$ goes on decreasing. Thus the signal-to-noise ratio of the random variable $r_0(T)$, by observing which, the decision device takes its decision, is maximized.

Example 11.3 Zero-mean white-noise of two-sided power spectral density $\eta/2$ is integrated by an ideal integrator. A sample of the output noise of the integrator is taken at $t = T$. Show that the variance of the noise sample so taken, is inversely proportional to T.

Solution Let the zero-mean white noise be represented by $w(t)$ and let the noise at the output of the integrator be represented by $n_0(t)$. So, the sample of $n_0(t)$ at $t = T$ is

$$n_0(t)|_{t=T} = n_0(T)$$

Then $n_0(T)$ is a zero-mean random variable since $w(t)$ is zero-mean and integration is a linear operation.

$$\text{Variance of } n_0(T) = E[n_0^2(T)] - [E\{n_0(T)\}]^2$$

$$= E[n_0^2(T)] \text{ as } E\{n_0(T)\} = 0 \quad \text{since } n_0(t) \text{ is a zero-mean noise}$$

But $$n_0(T) = \left\{\frac{1}{T}\int_0^T w(t)\,dt\right\}$$

$\therefore$ $$E[n_0^2(T)] = E\left[\left\{\frac{1}{T}\int_0^T w(t)\,dt\right\}\left\{\frac{1}{T}\int_0^T w(u)\,du\right\}\right]$$

$$= \frac{1}{T^2}\int_0^T\int_0^T E\{w(t)w(u)\}\,dt\,du$$

But $$E\{w(t)w(u)\} = R_w(t-u) = \frac{\eta}{2}\delta(t-u)$$

$\therefore$ $$E[n_0^2(T)] = \frac{1}{T^2}\int_0^T\int_0^T \left(\frac{\eta}{2}\right)\delta(t-u)\,dt\,du$$

But $$\int_0^T \left(\frac{\eta}{2}\right)\delta(t-u)\,dt = \frac{\eta}{2} \text{ from the defining equation of an impulse}$$

$\therefore$ $$\sigma^2_{n_0(T)} = E[n_0^2(T)] = \frac{1}{T^2}\cdot\frac{\eta}{2}\int_0^T du = \frac{\eta}{2T}$$

Thus, the variance of $n_0(T)$ is inversely proportional to T.

Example 11.4 A baseband binary system transmits the signal $s_1(t)$ for binary 1 and the signal $s_2(t)$ for binary 0, where $s_1(t)$ and $s_2(t)$ are given by

$$s_1(t) = \begin{cases} A & \text{for } 0 \le t \le T/2 \\ A/2 & \text{for } T/2 < t \le T \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad s_2(t) = \begin{cases} A/2 & \text{for } 0 \le t \le T/2 \\ -A/2 & \text{for } T/2 < t \le T \\ 0 & \text{elsewhere} \end{cases}$$

The channel may be assumed to be AWGN with noise PSD of $N_0/2$ and the symbols are equi-probable. Find the energy of the two transmitted signals $s_1(t)$ and $s_2(t)$ and hence find the average energy per bit, E_b. Also find the probability of bit error, P_e, in terms of E_b/N_0. **(University Examination Question)**

Solution Energy of the signal $s_1(t) = E_1 = \int_{-\infty}^{\infty} s_1^2(t)dt$

$$= \int_0^{T/2} A^2 dt + \int_{T/2}^{T} (A/2)^2 dt = \frac{A^2T}{2} + \frac{A^2T}{8} = \frac{5A^2}{8}$$

Energy of the signal $s_2(t) = E_2 = \int_{-\infty}^{\infty} s_2^2(t)dt$

$$= \int_0^{T/2} (A/2)^2 dt + \int_{T/2}^{T} (-A/2)^2 dt = \frac{A^2T}{8} + \frac{A^2T}{8} = \frac{A^2T}{4}$$

$\therefore$ bit energy (average) = $E_b = P_1E_1 + P_2E_2$

where P_1 = Probability of $s_1(t)$ and P_2 = Probability of $s_2(t)$

But it is given that the binary symbols 1 and 0 are equi-probable.

$\therefore$ $P_1 = P_2 = 0.5$

and $E_b = (E_1 + E_2)0.5 = \frac{7}{16}A^2T$

Now, from Eq. (11.52), $d_{\max}^2 = \frac{2}{N_0}E_p$ $\quad\therefore\quad \frac{d_{\max}}{2} = \sqrt{\frac{E_p}{2N_0}}$

Now E_p = Energy in $[s_1(t) - s_2(t)]$ = Energy in $p(t)$

From the given $s_1(t)$ and $s_2(t)$, we have

$$(s_1(t) - s_2(t)) = p(t) = \begin{cases} A/2 & \text{for } 0 \le t \le T/2 \\ A & \text{for } T/2 < t \le T \\ 0 & \text{elsewhere} \end{cases}$$

$$\therefore \quad \int_{-\infty}^{\infty} p^2(t) = E_p = \int_0^{T/2} A^2/4dt + \int_{T/2}^{T} A^2 dt = \frac{A^2T}{8} + \frac{A^2T}{2} = \frac{5}{8}A^2T$$

Since $E_b = \frac{7}{16}A^2T$ and $E_p = \frac{5}{8}A^2T$, we have $E_p = \frac{10}{7}E_b$

We know that $P_e = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{E_p}{2N_0}}\right]$

$$= Q\left[\sqrt{\frac{(10/7)E_b}{2N_0}}\right] = Q\left[\sqrt{\frac{5E_b}{7N_0}}\right]$$

$$\therefore \quad P_e = Q\left[\sqrt{\frac{5E_b}{7N_0}}\right]$$

11.2.6 Properties of a Matched Filter

Property 1: For signals corrupted by additive white Gaussian noise, a filter matched to the signal, i.e., a matched filter, maximizes the signal-to-noise ratio at its output.

Recall that as per Eq. (11.41), d^2 represents the signal-to-noise ratio at the output of the filter. By applying 'Schwarz's inequality' and imposing the conditions required for the equality sign to hold good, we have indeed maximized d^2, i.e., we have derived the $H(f)$ of the filter that maximizes the output signal-to-noise ratio. Thus, this property is indeed the very basis on which the matched filter transfer function was determined.

Property 2: The signal and the matched filter impulse response are mirror images of each other.

A look at Figs. 11.7(a) and (c) shows that this is indeed true. Figure 11.7(a) shows the signal and Fig. 11.7(c) shows the magnitude of the impulse response of the matched filter matched to the signal in Fig. 11.7(a). Except for a lateral inversion, they are exactly the same – a basic property of a mirror image.

Property 3: The impulse response $h_{\text{opt}}(t)$ of the matched filter matched to a signal $p(t)$, has the same magnitude spectrum as the signal itself, except for a scaling factor.

As given in Eq. (11.49), the transfer function, $H_{\text{opt}}(f)$ of a matched filter, matched to the signal $p(t)$ is

$$H_{\text{opt}}(f) = \alpha P^*(f)e^{-j2\pi fT}$$

So, if $h_{\text{opt}}(t)$ is the impulse response of this matched filter,

$$\mathcal{F}[h_{\text{opt}}(t)] = \alpha P^*(f)e^{-j2\pi fT}$$

$$\therefore \quad |\mathcal{F}[h_{\text{opt}}(t)]| = \text{Magnitude spectrum of } h_{\text{opt}}(t) = |\alpha P^*(f)e^{-j2\pi fT}| = |\alpha P(f)| = |\alpha||P(f)|$$

where $P(f)$ is the magnitude spectrum of the signal.

Thus, except for a scaling factor, $h_{\text{opt}}(t)$ and $p(t)$ have the same magnitude spectrum.

Property 4: If the signal is of duration T sec, the filter matched to it has to have a delay of at least T sec in order to be physically realizable.

An LTI filter is physically realizable *iff* it is causal; and the impulse response of any causal LTI system must be identically equal to zero for all negative values of time, i.e., $h(t) = 0$ for $t < 0$.

From Fig. 11.7(b), it is clear that unless there is a delay of T sec, which is the duration of the signal, the impulse response of the matched filter will have non-zero values for its impulse response for negative values of time.

Thus, a minimum delay of T sec. is needed in order to make the matched filter physically realizable.

Example 11.5 If a signal $x(t) = \Lambda(t-1)$, determine the transfer function of the filter matched to this signal.

Solution Since $x(t) = \Lambda(t-1)$, we have

$$X(f) = \text{sinc}^2(f)e^{-j2\pi f}$$

From Eq. (11.47) we then have the matched filter transfer function given by

$$H(f) = X^*(f)e^{-j\omega T}$$

From the given $x(t)$, we know that $T = 2$.

$$\therefore \quad H(f) = \text{sinc}^2(f)e^{+j2\pi f}e^{-j4\pi f}$$

$$= \text{sinc}^2(f)e^{-j2\pi f}$$

Example 11.6 $x(t)$ is a triangular pulse of width 1 m-sec and height 10^{-2} V. Assuming the channel noise to be white with a PSD of $\eta = 10^{-8}$ *W*/Hz, determine the signal-to-noise ratio at the output of the matched filter.

Solution The energy E_b of the triangular pulses is

$$E_b = 2\int_0^{T/2} x^2(t)\,dt = 2\int_0^{0.5\times10^{-3}} (20t)^2\,dt$$

$$= 0.33\times10^{-7}\ \text{volt}^2\cdot\text{sec}$$

From Eq. (11.55) we then find the *SNR* at the output of the matched filter is

$$\frac{S}{N} = \frac{2E_b}{\eta} = \frac{2\times0.33\times10^{-7}}{10^{-8}} = 6.67$$

$$\therefore \qquad \left(\frac{S}{N}\right)_{\text{dB}} = 10\log_{10} 6.67 = 8.2\ \text{dB}$$

Example 11.7 In Fig.11.13, show that in the absence of any channel noise,

$$[r_{01}(T) - r_{02}(T)] = E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1}\cdot E_{s_2}} = \int_0^T [s_1(t) - s_2(t)]^2\,dt$$

where $r_{01}(T)$ and $r_{02}(T)$ are the inputs to the decision device when $s_1(t)$ alone, and $s_2(t)$ alone are respectively given as inputs to the correlation receiver, ρ is defined by

$$\rho \,\Delta\, \frac{1}{\sqrt{E_{s_1}\cdot E_{s_2}}}\int_0^T s_1(t)\cdot s_2(t)\,dt$$

and E_{s_1} and E_{s_2} are respectively the energies in the two signals $s_1(t)$ and $s_2(t)$, each of which is zero outside the slot duration 0 to T sec.

Solution From Fig. 11.13, we find that

$$r_{01}(T) = \int_0^T s_1(t)[s_1(t) - s_2(t)]\,dt = \int_0^T s_1^2(t)\,dt - \int_0^T s_1(t)s_2(t)\,dt$$

$$= E_{s_1} - \rho\sqrt{E_{s_1}\cdot E_{s_2}}$$

Similarly,

$$r_{02}(T) = \int_0^T s_2(t)[s_1(t) - s_2(t)]\,dt = -\int_0^T s_2^2(t)\,dt + \int_0^T s_1(t)s_2(t)\,dt$$

$$= -E_{s_2} + \rho\sqrt{E_{s_1}\cdot E_{s_2}}$$

$$\therefore \qquad r_{01}(T) - r_{02}(T) = E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1}\cdot E_{s_2}}$$

Now consider

$$\int_0^T [s_1(t) - s_2(t)]^2\,dt = \int_0^T s_1^2(t)\,dt + \int_0^T s_2^2(t)\,dt - 2\int_0^T s_1(t)s_2(t)\,dt$$

$$= E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1}\cdot E_{s_2}}$$

Hence,

$$[r_{01}(T) - r_{02}(T)] = E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1} \cdot E_{s_2}} = \int_0^T [s_1(t) - s_2(t)]^2 dt$$

Example 11.8 During each time slot of 1 m-sec, a baseband binary communication system transmits either a rectangular pulse of amplitude V volts, or a triangular pulse of amplitude $-V$ volts. If $V = 10$ mV and the white noise on the channel has a PSD (two-sided) of 5×10^{-9} W/Hz, find the bit-error probability, assuming matched filter reception.

Solution

$$E_{s_1} = V^2T = (10 \times 10^{-3})^2 \times 1 \times 10^{-3} = 10^{-7}\ \text{V}^2$$

$$E_{s_2} = \text{Energy of the triangular pulse}$$

$$= 2\int_0^{T/2} (-20t)^2 \cdot dt = 2 \times 400 \frac{t^3}{3}\Bigg|_{t=0}^{t=0.5\times10^{-3}}$$

$$= \frac{800}{3} \times 0.125 \times 10^{-9} = 0.333 \times 10^{-7}\ \text{V}^2\ \text{sec}$$

$$\therefore \quad \sqrt{E_{s_1} \cdot E_{s_2}} = \sqrt{1 \times 10^{-7} \times 0.33 \times 10^{-7}}$$

$$= 0.57445624 \times 10^{-7}\ \text{V}^2\ \text{sec}$$

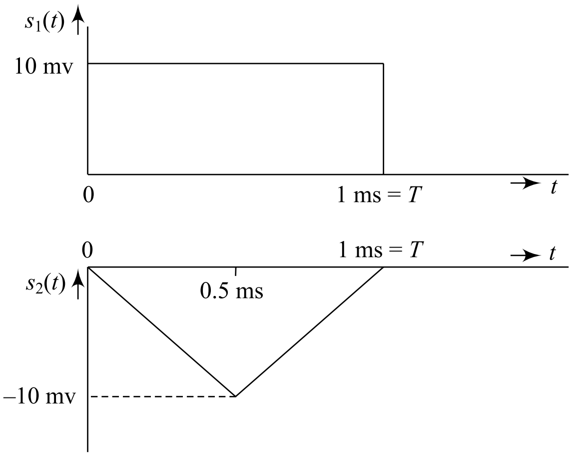


Fig. 11.11 *$s_1(t)$ and $s_2(t)$ of Example 11.8*

$$\rho = \frac{1}{\sqrt{E_{s_1} \cdot E_{s_2}}} \int_0^T s_1(t) \cdot s_2(t) dt$$

$$= \left(\frac{10^7}{0.57445624}\right) 2 \times \int_0^{0.5\times10^{-3}} 10 \times 10^{-3} \times (-20t)dt = \frac{-5 \times 10^{-8} \times 10^7}{0.57445624}$$

$$= \frac{-0.5}{0.57445624} = -0.87$$

Since the receive filter is matched to $p(t) = [s_1(t) - s_2(t)]$, we know from Eq. (11.48) that

$$P_{e\min} = \frac{1}{2} erfc\left[\frac{d_{\max}}{2\sqrt{2}}\right]$$

where

$$d^2_{\max} = \frac{2}{\eta} E_p$$

and

$$E_p = \int_0^T p^2(t) dt = \int_0^T [s_1(t) - s_2(t)]^2 dt = r_{01}(T) - r_{02}(T).$$

$$= E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1} \cdot E_{s_2}}$$

$$\therefore \qquad d_{\max}^2 = \frac{2}{\eta}[E_{s_1} + E_{s_2} - 2\rho\sqrt{E_{s_1} \cdot E_{s_2}}]$$

$$= \frac{2}{10 \times 10^{-9}}[10^{-7} + 10^{-7} \times 0.33 - 2(-0.87) \times 0.57445624 \times 10^{-7}]$$

$$= 46.5$$

$$\therefore \qquad d_{\max} = \sqrt{46.5} = 6.82 \quad \text{and} \quad \frac{d_{\max}}{2\sqrt{2}} = 2.41$$

$$\therefore \qquad P_{e\min} = \frac{1}{2} erfc\left[\frac{d_{\max}}{2\sqrt{2}}\right] = \frac{1}{2} erfc(2.41)$$

$$= Q[\sqrt{2} \cdot 2.41] = Q[3.40774] \cong 3 \times 10^{-4}$$

Example 11.9 An optimum receiver receives polar NRZ equi-probable binary baseband data. The received signal takes the values +6 mv and –6 mv corresponding respectively to binary 1 and binary 0. The channel noise is white with a two-sided PSD of 10^{-9} W/Hz. Optimum decision threshold is used. If the data rate is 9600 bits/sec, find (a) P_e, (b) P_e when the data rate is 12 kbps, and (c) if P_e is to remain the same at 12 kbps data rate as it was at 9600 bits/sec data rate, what should be the voltage levels corresponding to binary 1 and binary 0?

Solution

(a) From Eq. (11.55) we have $P_e = Q\left[\sqrt{\frac{2E_b}{\eta}}\right] = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

where $E_b = A^2T$ where T is the duration of each pulse = Time-slot duration
A = 6 mv and $(\eta/2) = 10^{-9}$ W/Hz
To determine T: It is the inverse of the data rate

$$\therefore \qquad T = \frac{10^3}{9600} \text{ms} = 0.104 \text{ ms}$$

$$E_b = A^2T = (6 \times 10^{-3}) \times 0.104 \times 10^{-3} W$$

$$= 36 \times 0.104 \times 10^{-9} = 3.744 \times 10^{-9} W$$

$$\therefore \qquad \sqrt{\frac{E_b}{\eta}} = \sqrt{\frac{3.744}{2}} = 1.3682$$

$\therefore$ from error-function table, $erf(1.3862) \cong 0.95$

$$\therefore \qquad erfc(1.3862) \cong 0.05 \qquad \therefore\ P_e = \frac{1}{2} erfc(1.3862) = 25 \times 10^{-3}$$

(b) When the data rate is 12 kbps, $T = \frac{1}{12 \times 10^3} = 0.083$ m.s

$$\therefore \qquad E_b = A^2T = 36 \times 10^{-6} \times 0.083 \times 10^{-3} = 3 \times 10^{-9} W$$

$$\therefore \qquad \sqrt{\frac{E_b}{\eta}} = \sqrt{\frac{3}{2}} = 1.2247$$

$$\therefore \qquad erf(1.2247) \cong 0.917 \quad \therefore \quad \frac{1}{2} erfc(1.2247) = P_e = 41.5 \times 10^{-3}$$

(c) If P_e is to remain the same at the higher bit rate of transmission of the signal, E_b must be maintained the same.

$$\therefore \qquad E_b = 3.744 \times 10^{-9} = A^2 \times 0.083 \times 10^{-3}$$

$$\therefore \qquad A = \sqrt{(3.744 \times 10^{-9})/(0.083 \times 10^{-3})}$$

$$= \sqrt{45.1 \times 10^{-6}} = 6.716 \text{ mv}$$

Example 11.10 For an NRZ polar binary data, the received signal is either +3 V or –3 V during a time slot. The signal is corrupted by white noise of PSD η equal to 10^{-4} V^2/Hz. If an integrate-and-dump type of receiver is used, what should be the minimum duration of the time slot, if P_e is not to exceed 10^{-5}?

Solution Polar data is received. Hence, $P_e = \frac{1}{2}\text{erfc}\left[\sqrt{\frac{E_b}{\eta}}\right]$. Let the time-slot duration be T sec. Then

$$E_b = A^2 T = 9T \text{ V}^2\text{-sec.}$$

It is given that P_e can have a maximum value of 10^{-5} corresponding value of $\sqrt{\frac{E_b}{\eta}}$ should be (by referring to the error-function tables, or curves) 3.

$$\therefore \qquad \sqrt{\frac{E_b}{\eta}} = 3 \quad \text{or} \quad \frac{E_b}{\eta} = 9 = \frac{9T}{10^{-4}} = 9 \times 10^4 T$$

$\therefore$ $T = 10^{-4}$ sec = 0.1 m-sec. This is the maximum time slot duration.

Example 11.11 Repeat Example 11.10 assuming NRZ unipolar signal with binary 1 and binary 0 represented by +3 V and 0 V respectively.

Solution Now, $P_e = Q\left[\sqrt{\frac{E_b}{\eta}}\right] = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

where $E_b = A^2 T/2$ (see Eq. (11.50))

If P_e is to be 10^{-5}, as we have seen in the last example, the error function should have an argument of 3.

$$\therefore \qquad \sqrt{\frac{E_b}{\eta}} = 3 \quad \text{or} \quad E_b = 9 \times 10^{-4}$$

But $E_b = A^2 T/2$ in this case.

$$\therefore \qquad A^2 T = 2 \times 9 \times 10^{-4} \quad \text{or} \quad T = \frac{2 \times 9 \times 10^{-4}}{A^2} = \frac{2 \times 9 \times 10^{-4}}{9} = 2 \times 10^{-4}$$

Hence the minimum time slot duration $= 2 \times 10^{-4}$ sec.

Example 11.12 An integrate-and-dump type of receiver is used to receive a binary baseband polar NRZ signal which takes the values of +A or –A during any time slot (of duration T sec) with probabilities of P_1 and P_0 respectively. The signal is corrupted by zero-mean white Gaussian noise of two-sided power spectral density (η/2) watts/Hz. (a) Find the optimum threshold for the receiver. (b) Write down an expression for P_e with the threshold as determined in (a).

Solution Refer to Fig. 11.9 of a baseband binary receiver of the integrate-and-dump type. Now, the observation variable is $r_0(T)$ and not $r(T)$. We find from Eq. (11.58) that $r_0(T)$ is given by

$$r_0(T) = \begin{cases} s_{01}(T) + n_0(T) & \text{if } H_1 \text{ is true i.e., if a 1 is sent} \\ s_{00}(T) + n_0(T) & \text{if } H_0 \text{ is true i.e., if a 0 is sent} \end{cases}$$

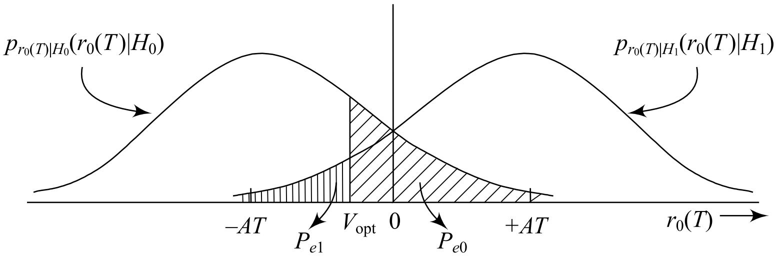


Fig. 11.12 *Conditional densities and threshold for Example 11.12*

Since the integrator is an LTI system and the input noise to it is Gaussian zero-mean white noise, the output noise component $n_0(t)$, and hence $n_0(T)$ are also Gaussian and zero mean.

$$s_{01}(T) = \int_0^T s_1(t)\,dt = \int_0^T A\,dt = AT$$

$$s_{00}(T) = \int_0^T s_0(t)\,dt = \int_0^T -A\,dt = -AT$$

Since $n_0(T)$ is a zero-mean Gaussian random variable, we may write the conditional densities of $r_0(T)$ as

$$H_1 : s_{01}(T) + n_0(T) = AT + n_0(T) \text{ and } p_{r_0|H_1}(r_0(T)|H_1) = \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}}\exp\left[\frac{-(r_{0T} - AT)^2}{2\sigma^2_{n_0(T)}}\right]$$

and $$H_0 : s_{00}(T) + n_0(T) = -AT + n_0(T) \text{ and } p_{r_0|H_0}(r_0(T)|H_0) = \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}}\exp\left[\frac{-(r_{0T} + AT)^2}{2\sigma^2_{n_0(T)}}\right]$$

If P_1 and P_0 are the *a priori* probabilities of occurrence of a binary '1' and a binary '0' respectively, we may make use of Eq. (11.24) to determine the optimum threshold for this case. We note in this connection that $\sigma^2_{n_0}$, A_1 and A_0 of Eq. (11.24) are now

1. $\sigma^2_{n_0} \to \sigma^2_{n_0(T)} = (\eta/2T)$ from Example 11.3
2. $A_1 \to AT$
3. $A_0 \to -AT$

$\therefore$ the optimum threshold V_{opt} for this case is given by

$$V_{opt} = \frac{\sigma^2_{n_0(T)}}{[AT - (-AT)]}\left[\log_e\left\{\frac{P_0}{P_1}\right\} + 0\right] = \left[\frac{\sigma^2_{n_0(T)}}{2AT}\right]\log_e\left[\frac{P_0}{P_e}\right]$$

If P_e is the probability of error with this optimum threshold value, with P_{e_0} and P_{e_1} as shown in Fig. 11.11, we have

$$P_e = P_1 P_{e_1} + P_0 P_{e_0} \qquad \text{(Refer to Eq. (11.8))}$$

But

$$P_{e_0} = \int_{V_{\text{opt}}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}} e^{-[r_0(T)+AT]^2/2\sigma^2_{n_0(T)}} \, dr_0(T) = Q[V_{\text{opt}}] = \frac{1}{2} erfc\left[\frac{V_{\text{opt}}}{\sqrt{2}}\right]$$

and

$$P_{e_1} = \int_{-\infty}^{V_{\text{opt}}} \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}} e^{-[r_0(T)-AT]^2/2\sigma^2_{n_0(T)}} \, dr_0(T) = \frac{1}{2} erfc\left[\frac{(AT - V_{\text{opt}})}{\sqrt{2}\sigma_{n_0(T)}}\right]$$

$$= \frac{1}{2}\text{erfc}\left[\frac{(AT - V_{\text{opt}})\sqrt{T}}{\sqrt{\eta}}\right] \text{ by substituting } \sqrt{\eta/2T} \text{ for } \sigma_{n_0(T)}$$

∴

$$P_e = P_1 P_{e_1} + P_0 P_{e_0} = \frac{P_1}{2} erfc\left[\frac{(AT - V_{\text{opt}})\sqrt{T}}{\sqrt{\eta}}\right] + \frac{P_0}{2} erfc\left[\frac{V_{\text{opt}}}{\sqrt{2}}\right]$$

where

$$V_{\text{opt}} = \left[\frac{\eta}{4AT^2}\right]\log_e\left[\frac{P_0}{P_e}\right]$$

11.2.7 Correlation Receivers

We have seen that matched filtering provides a way of achieving minimum probability of error for the reception of a given pair of signals $s_1(t)$ and $s_2(t)$ in the presence of additive zero-mean white Gaussian noise. For a signal $p(t) = [s_2(t) - s_1(t)], 0 \le t \le T,$ it was shown that the transfer function of the matched filter is

$$H_{\text{opt}}(f) = \alpha P^*(f) e^{-j2\pi fT}$$

and that the corresponding impulse response is

$$h_{\text{opt}}(t) = \alpha p(T - t); 0 \le t \le T$$

Hence, one way of *implementing* an optimum receiver for receiving binary baseband signals in the presence of additive zero-mean white Gaussian noise is to synthesize a filter with a transfer function of $H_{\text{opt}}(f)$ and use the receiver configuration shown in Fig. 11.5 with $H(f)$ equal to $H_{\text{opt}}(f)$ and $S_n(f)$, the two-sided PSD of the noise corrupting the signal taken as (η/2) since the noise is white. So, in this method of implementation of an optimum receiver, the received signal $r(t) = s(t) + n_w(t),$ where $s(t)$, which is either $s_2(t)$ or $s_1(t)$ (depending on whether a binary 1 or a binary 0 was transmitted), is convolved with the impulse response, $H_{\text{opt}}(f)$ of the matched filter to give $r_0(t)$.

There is, however, another interesting way of *implementing* an optimum receiver when the signal is corrupted by zero-mean additive *white* Gaussian noise. It is called a *correlation receiver* since it uses *correlatio*n instead of convolution. It correlates the received signal $r(t)$ with $p(t) = [s_2(t) - s_1(t)]$ in order to generate the observed variable which is used for decision making.

In fact, ***matched filtering and correlation can be shown to be equivalent operations***, *and* ***performance-wise there is no difference between the two***. We shall first establish the equivalence between the two.

Equivalence of matched filtering and correlation The received signal,

$$r(t) = \begin{cases} s_1(t) + n_w(t) & \text{if } H_0 \text{ is true} \\ s_2(t) + n_w(t) & \text{if } H_1 \text{ is true} \end{cases}$$

Since the two possible signals, $s_1(t)$ and $s_2(t)$ are known *a priori* to the receiver, it stores a copy of $p(t) = [s_2(t) - s_1(t)]$ and *correlates* the received signal $r(t)$ with $p(t)$.

i.e., it finds $r_0(t) \underline{\Delta} \int_0^t r(\lambda) p(\lambda) d\lambda$ (11.59)

Matched filter, on the other hand, convolves the received signal, $r(t)$ with its own impulse response function, $h_{\text{opt}}(t)$ and produces an output $z(t)$ given by

$$z(t) = \text{Matched filter output} = \int_0^t r(\lambda) h_{\text{opt}}(T - \lambda) d\lambda$$

But $h_{\text{opt}}(t) = p(T - t); 0 \le t \le T$

$$h_{\text{opt}}(T - \lambda) = p(\lambda); 0 \le \lambda \le T$$

$$\therefore \quad z(t) = \int_0^t r(\lambda) h_{\text{opt}}(T - \lambda) d\lambda = \int_0^t r(\lambda) p(\lambda) d\lambda$$

$$= \text{correlation between } p(t) \text{ and } r(t) \underline{\Delta} r_0(t) \text{ as per Eq. (11.59)}$$

Thus, matched filtering of the received signal $r(t)$, i.e., convolving it with the impulse response of the matched filter, and correlating $r(t)$ with $p(t)$, the known signal, lead to the same result and are thus *equivalent*.

Implementation of the correlation receiver We may implement the correlation receiver as shown in Fig. 11.13.

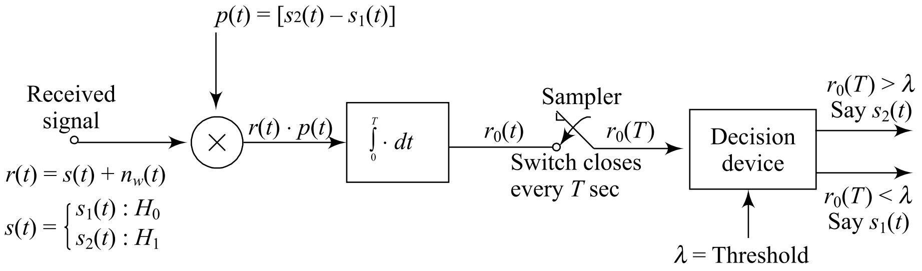


Fig. 11.13 *A correlation receiver*

Note that the above implementation is the same as the implementation shown in Fig. 11.14.

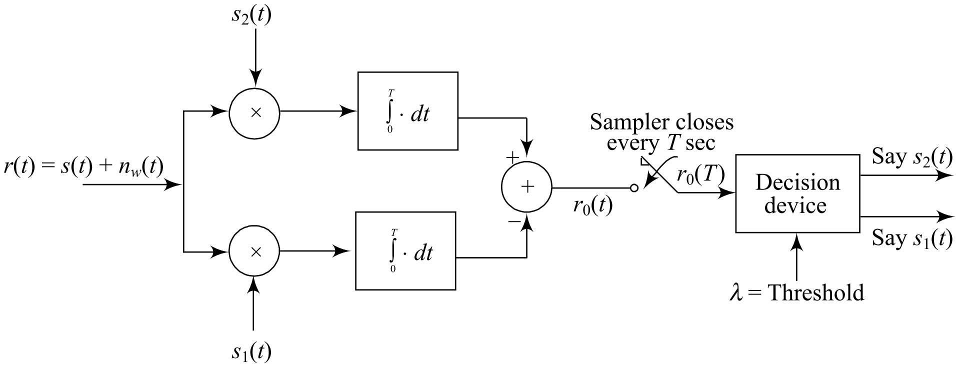


Fig. 11.14 *An equivalent configuration for the correlation receiver*

Note *As already stated earlier, performance-wise there is no difference between a matched filter receiver and a correlation receiver, and their probabilities of error will be the same.*

Probability of error in a PCM receiver For a polar signal, with an optimum receiver, the probability of error P_e is given by

$$P_e = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right] \tag{11.60}$$

We know that the complementary error function has an upper bound given by

$$erfc(x) < \frac{e^{-x^2}}{\sqrt{\pi} x} \tag{11.61}$$

Using Eq. (11.58), we may write the upper bound for P_e as

$$P_e < \frac{e^{-E_b/\eta}}{2\sqrt{\pi E_b/\eta}} \tag{11.62}$$

We thus find that in a PCM receiver, the probability of error decreases exponentially with increase of the ratio (E_b/η). This is shown in Fig. 11.15.

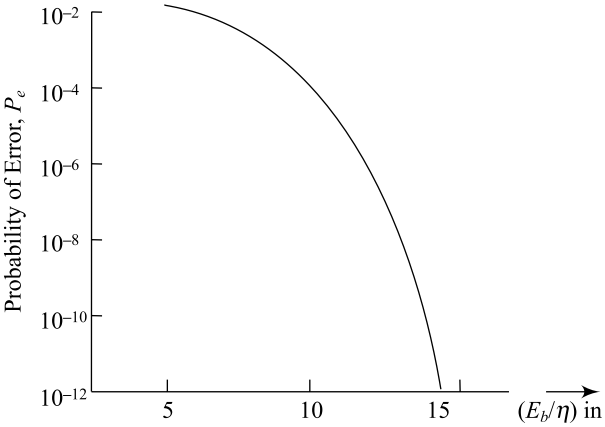


Fig. 11.15 *Probability of error vs.* E_b/η

Example 11.13 A binary baseband data transmission system transmits waveforms $s_1(t) = 0; 0 \le t \le T$ corresponding to binary zero and waveform $s_2(t) = \begin{cases} 2t/T; & 0 \le t \le T \\ \dfrac{2}{T}(T-t); & 0 \le t \le T \end{cases}$ corresponding to binary 1. Assume that $T = 20$ ms and that the additive zero-mean white Gaussian has a PSD (two sided) $(\eta/2) = 10^{-7}$ *W*/Hz. Find P_e for the optimum receiver assuming *a priori* probabilities of occurrence of a binary 1 and 0 to be 0.75 and 0.25 respectively.

Solution The waveform of $s_2(t)$ is as shown in Fig. 11.16.

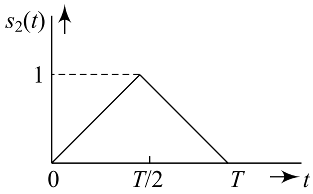


Fig. 11.16 *Waveform of* $s_2(t)$

$$p(t) = [s_2(t) - s_1(t)] = s_1(t)$$

$\therefore$ when $s_2(t)$ is received, the output at $t = T$ of the correlator will be $s_{02}(t)$ given by

$$s_{02}(T) = \int_0^T s_2(t)[s_2(t) - s_1(t)]\,dt = \int_0^T s_2^2(t)\,dt = 2\int_0^{T/2} \frac{4t^2}{T^2}\,dt$$

$$= \frac{8}{T^2}\int_0^{T/2} t^2\,dt = \frac{T}{3}$$

$$\therefore \qquad s_{02}(T) = \frac{T}{3}$$

Since $s_1(t) = 0, s_{01}(T) = 0.$

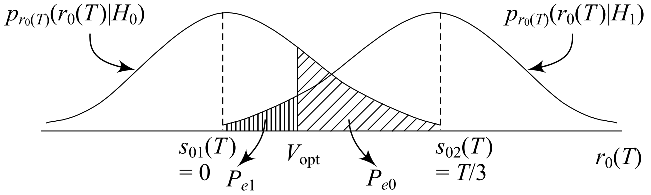


Fig. 11.17 *Plot of conditional density functions for Example 11.13*

From Eq. (11.24), the optimum threshold is given by

$$V_{\text{opt}} = \frac{\sigma^2_{n_0(T)}}{T/3}\left[\log_e \frac{P_0}{P_1} + \frac{T^2/9}{2\sigma^2_{n_0(T)}}\right]$$

But $$\sigma^2_{n_0(T)} = \frac{\eta}{2T} \qquad \text{(From Example 11.3)}$$

$$= \frac{10^{-7}}{T} = \frac{10^{-7}}{2\times 10^{-2}} = \frac{10^{-5}}{2}W$$

and $$\log_e\left(\frac{0.25}{0.75}\right) = -0.477$$

∴ substituting these in the expression for V_{opt}, we get

$$V_{\text{opt}} = \frac{10^{-5}}{2}\times\frac{3}{20\times 10^{-3}}\left[-0.477 + \frac{(20\times 10^{-3})\times 2}{18\times 10^{-5}}\right] = 2.975\times 10^{-3}\text{ V}$$

and $$s_{02}(T) = \frac{T}{3} = \frac{20\times 10^{-3}}{2} = 6.6\text{ mv}$$

$$P_e = 0.5P_{e_0} + 0.75P_{e_1}$$

But $$P_{e_0} = \int_{V_{\text{opt}}}^{\infty} p_{r_0(T)}(r_0(T)|H_0)\,dr_0(T) = Q[V_{\text{opt}}] = \frac{1}{2}erfc\left[\frac{V_{\text{opt}}}{\sqrt{2}}\right] = \frac{1}{2}erfc\left[\frac{2.973\times 10^{-3}}{\sqrt{2}}\right]$$

$$P_{e_1} = \int_{-\infty}^{V_{\text{opt}}} p_{r_0(T)}(r_0(T)|H_1)\,dr_0(T) = Q[s_{02}(T) - V_{\text{opt}}] = \frac{1}{2}erfc\left[\frac{1}{\sqrt{2}}\cdot 3.691\times 10^{-3}\right]$$

∴ $$P_e = 0.125erfc[0.00210] + 0.375erfc[0.00261]$$

$$\cong 3.9\times 10^{-2}$$

11.2.8 M-ary Baseband Signaling and Probability of Error

In Section 10.1, we had discussed the basic principles of baseband M-ary signaling while in Section 10.4.2, we had shown that

$$W_b = \frac{R_b}{2}\text{Hz} \qquad (11.63)$$

is the Nyquist bandwidth for binary baseband transmission at a rate of R_b bits per second, then the absolute minimum transmission bandwidth required for achieving a bit rate of R_b bits per second using M-ary baseband transmission is

$$W_b = \frac{R_b}{2} \cdot \frac{1}{\log_2 M} = \frac{W_b}{\log_2 M} \tag{11.64}$$

The above result clearly shows that for a given fixed transmission rate in bits per second, M-ary baseband transmission cuts down the transmission bandwidth requirements by a factor of $\log_2 M$ as compared to the bandwidth requirement of baseband binary transmission. Thus, M-ary baseband transmission saves bandwidth. But it must be noted that this saving in bandwidth is obtained at a cost, since the multilevel pulse amplitude requires more average transmitter power to maintain a specified probability of error, P_e, as compared to binary transmission, as will be evident from the following analysis that gives us the probability of error for M-ary baseband signaling.

For calculating the P_e for M-ary baseband signaling, we shall assume that

1. polar M-ary signaling is used, where M is even.
2. the noise at the input to the decision device is zero-mean Gaussian.
3. the signaling levels a_ks are equispaced.
4. all the levels are equally likely to occur.

Hence, let the signaling levels, viz., a_ks be

$$a_k = \pm A/2, \pm 3A/2, \pm 5A/2, \ldots, \pm (M-1)A/2 \tag{11.65}$$

For the purpose of drawing the conditional density functions, let us take $M = 4$. Then these conditional PDFs will be as shown in Fig. 11.18.

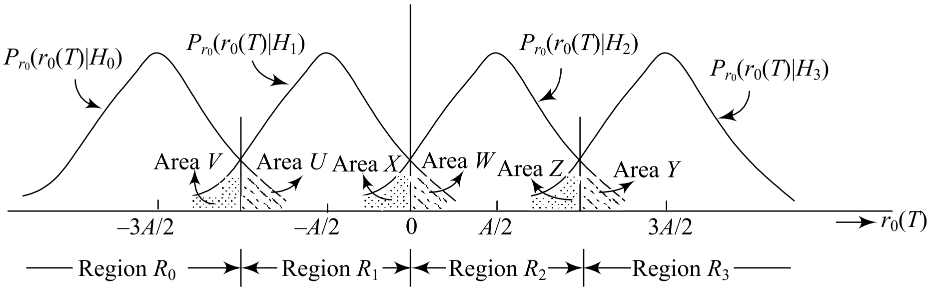


Fig. 11.18 *Conditional PDFs with M = 4 and Gaussian noise*

For reasonably large *SNR*, i.e., (A^2/σ^2), where σ^2 is the noise variance, it is clear from Fig. 11.18 that the decision-making device may mistake one level from its immediately adjacent level but not one which is farther away. For example, it may wrongly detect level $A/2$ as either $-A/2$ or $3A/2$ but not as $-3A/2$. Further, since all the levels are equally likely, the values $-A$, 0 and $+A$, etc., which are the values of the observation variable $r_0(T)$ corresponding to the intersection points, are the optimum threshold values that define the decision boundaries of the various decision region like R_0, R_1, R_2 and R_3 shown in Fig. 11.18.

From the figure, we find that

1. Probability of level $-3A/2$ being mistaken for level $-A/2$ is area *u*.
2. Probability of level $-A/2$ being mistaken for level $-3A/2$ is area *V*.
3. Probability of level $-A/2$ being mistaken for level $A/2$ is area *W*.
4. Probability of level $A/2$ being mistaken for level $-A/2$ is area *X*.
5. Probability of level $A/2$ being mistaken for level $3A/2$ is area *Y*.
6. Probability of level $3A/2$ being mistaken for level $A/2$ is area *Z*.

Each of these areas is given by $Q(A/2\sigma)$. Since all the levels have been assumed to be equally probable, the *a priori* probability of each of the M levels, is equal to $1/M$. We find from the above discussion that while the extreme two levels, the Oth level, i.e., the $-(M-1)A/2$ level and the $(M-1)$th level, i.e., the $+(M-1)A/2$ level, contribute only one error area (probability) the remaining $(M-2)$ contribute two error areas (probabilities).

Hence, the average error probability, P_e, is given by

P_e = (Probability of transmission of level $-(M-1)A/2$) × (Probability of level $-(M-1)A/2$ being mistaken as $-(M-2)A/2$ level) + (Probability of transmission of level $-(M-2)A/2$) × (Probability of level $-(M-2)A/2$ being mistaken as either level $-(M-2)A/2$ or $(M-3)A/2$) + … + (Probability of transmission of level $(M-2)A/2$) × (Probability of level $(M-2)A/2$ being mistaken as either level $(M-3)A/2$ or level $(M-1)A/2$) + (Probability of transmission of level $(M-1)A/2$) × (Probability of level $(M-1)A/2$ being mistaken as level $(M-2)A/2$).

$$= \frac{1}{M}[2 \times Q(A/2\sigma) + (M-2) \times 2Q(A/2\sigma)]$$

$$= \left(\frac{2M-2}{M}\right)Q(A/2\sigma) = 2\left(1-\frac{1}{M}\right)Q(A/2\sigma)$$

$$\therefore \qquad \boxed{\underset{(\text{polar M-ary baseband})}{P_e} = 2\left(1-\frac{1}{M}\right)Q(A/2\sigma)} \qquad (11.66)$$

Remarks

(i) For binary baseband polar transmission with the two levels at A/2 and –A/2, we know $P_e = Q(A/2\sigma)$ *when* $P_0 = P_1 = 0.5$ *and the optimum threshold is used. Eq. (11.66) also reduces to* $Q(A/2\sigma)$ *when* $M = 2$.

(ii) When M is very large so that 1/M << 1, Eq. (11.66) reduces to

$$P_e = 2Q(A/2\sigma)$$

(iii) We find that $P_e = \begin{cases} Q(A/2\sigma) & \text{when } M = 2 \\ 2Q(A/2\sigma) & \text{when } M \text{ is very large} \end{cases}$

For constant pulse amplitude A, we find that as M is increased, P_e also increases, becoming two times its value for binary signaling, when M is made very large. In other words, as M is increased, A too has to be increased in order to keep P_e fixed at a specified value. (This is because the Q-function is a monotonically decreasing function of its argument). But increasing the value of A amounts to increasing the average transmitted power. Hence, for a specified P_e M-ary baseband signaling needs more transmitter power than binary baseband signaling.

In Eq. (11.66), the probability of error is given in terms of a Q-function whose argument is $(A/2\sigma)$. In certain situations, it will be useful if it is in terms of a Q-function whose argument is related directly to the signal-to-noise ratio or signal-to-noise spectral density ratio. Hence, we now relate $(A/2\sigma)$ to the signal power and noise density, $\eta/2$.

$$\text{The average energy of an M-ary digit} = E_M = \overline{a_k^2}\ \tau,$$

where τ is the width of the M-ary pulse as shown in Fig. 10.16.

If the M amplitude levels are equally likely, as has been assumed by us earlier, then

$$\overline{a_k^2} = 2 \times \frac{1}{M}\sum_{i=1}^{M/2}(2i-1)^2(A/2)^2 = \frac{(M^2-1)}{12}\cdot A^2 \qquad (11.67)$$

If R_M is the M-ary pulse rate,

$$S_R = R_M \cdot E_M = \left(\frac{M^2-1}{12}\right) \cdot A^2 R_M \cdot \tau = \left(\frac{M^2-1}{12}\right) A^2$$

$$\therefore \qquad A^2 = \frac{S_R}{(M^2-1)} \cdot 12 = S_R\left(\frac{12}{(M^2-1)}\right) \tag{11.68}$$

and $\sigma^2 = N_R$

$$\therefore \qquad \left(\frac{A^2}{4\sigma^2}\right) = \left(\frac{S_R}{4N_R}\right)\frac{12}{(M^2-1)} = \frac{3}{(M^2-1)} \cdot \left(\frac{S_R}{N_R}\right) \tag{11.69}$$

But $N_R = \left(\frac{\eta}{2}\right) 2B_T = \eta B_T$. But $B_T = \frac{R_M}{2}$

$$\therefore \qquad \left(\frac{A^2}{4\sigma^2}\right) = \frac{6}{(M^2-1)}\left(\frac{S_R}{\eta}\right)\frac{1}{R_M} \tag{11.70}$$

Equation (11.69) expresses $(A/2\sigma)$ in terms of (S_R/N_R) while Eq. (11.70) expresses $(A/2\sigma)$ in terms of (S_R/η).

Example 11.14 A binary baseband transmission system is to have a bit rate of R_b = 500 kbps and a probability of error not exceeding 10^{-4}. The channel noise is zero-mean white Gaussian with a two-sided PSD of $\eta/2 = 0.5 \times 10^{-17}$ *W*/Hz. Find the minimum value of S_R, the received signal power, when (a) $M = 2$ and (b) $M = 8$, assuming that Gray coding is used.

Solution

(a) ***When $M = 2$, $R_M = R_b = 5 \times 10^5$ bits/sec.***

From the *Q*-function tables, we find that when $P_e = 10^{-4}$, $(A/2\sigma)$ is approximately equal to 3.8, as $P_e = Q(A/2\sigma)$.

$$\therefore \qquad \left(\frac{A}{2\sigma}\right)^2 \ge (3.8)^2 = 14.44$$

$$\text{But} \qquad \left(\frac{A}{2\sigma}\right)^2 = \left(\frac{3}{2^2-1}\right)\left(\frac{S_R}{N_R}\right) = \left(\frac{6}{M^2-1}\right)\left(\frac{S_R}{\eta}\right)\frac{1}{R_M}$$

$$\therefore \qquad S_R = \left(\frac{A}{2\sigma}\right)^2 \cdot R_M \cdot \eta \cdot \frac{(M^2-1)}{6} = \frac{14.44 \times 5 \times 10^5 \times 10^{-17} \times 3}{6}$$

$$\therefore \qquad S_R \ge 36.1 \times 10^{-12}\, W$$

(b) *When **M** = 8 and Gray code is used to fix the levels*

$$\text{Again} \qquad \left(\frac{A}{2\sigma}\right) \ge 3.8 \qquad \therefore \quad \left(\frac{A}{2\sigma}\right)^2 \ge 14.44$$

$$\text{But} \qquad \left(\frac{A}{2\sigma}\right)^2 = \frac{6}{M^2-1} \cdot \frac{S_R}{\eta} \cdot \frac{1}{R_M}$$

$$\therefore \qquad S_R = \left(\frac{A}{2\sigma}\right)^2 \frac{(M^2-1)\eta \cdot R_M}{6} \geq 14.44 \frac{(64-1)}{6} \times 10^{-17} \times \frac{5\times 10^5}{\log_2 8}$$

Since in this case $R_M = \dfrac{R_b}{\log_2 M}$

$$\therefore \qquad S_R \geq 14.44 \frac{63}{6} \times 10^{-17} \times \frac{5\times 10^5}{3} = 252.7 \times 10^{-17} \text{ W}$$

$$\therefore \qquad S_R \geq 2.527 \times 10^{-15} \text{watt}$$

Example 11.15 In a baseband M-ary PAM system using M equally likely amplitude levels, the average probability of error P_e is to be less than 10^{-6}. Show that the minimum value of received signal-to-noise ratio for the system is approximately equal to

$$\left(\frac{S}{N}\right)_{R,\min} \cong 7.8(M^2-1)$$

Solution From Eq. (11.66), we have the probability of error given by

$$P_e = 2\left(1-\frac{1}{M}\right)Q(A/2\sigma) \leq 10^{-6}$$

$$\therefore \qquad Q(A/2\sigma) \leq \frac{10^{-6}}{2\left(1-\dfrac{1}{M}\right)}$$

When $M >> 1$, $\left(1-\dfrac{1}{M}\right) \approx 1 \qquad \therefore\ Q(A/2\sigma) \leq 0.5 \times 10^{-6}$

By referring to the Q-function tables or graphs,

$$\frac{A}{2\sigma} \geq 4.83 \text{ to keep } P_e \text{ less than } 10^{-6}$$

$$\frac{A^2}{4\sigma^2} \geq (4.83)^2 \quad \text{or} \quad A^2 \geq 93.6\sigma^2$$

But from Eq. (11.69), we have

$$A^2 \frac{12\sigma^2}{(M^2-1)} \cdot \left(\frac{S_R}{N_R}\right) \quad \therefore \left(\frac{S_R}{N_R}\right) \geq \frac{93.6\sigma^2}{12\sigma^2}(M^2-1)$$

or

$$\left(\frac{S_R}{N_R}\right) \geq 7.8(M^2-1) \text{ approximately}$$

Example 11.16 A symbol probability of error of 10^{-4} is to be maintained by a 4-ary polar baseband signaling system using NRZ rectangular pulses. If the attenuation in the channel is 20 dB and the noise power at the 50Ω input of the detector is 10^{-5} W, determine the average signal power which must be transmitted.

Solution From Eq. (11.66), we have

$$P_e = 2\left(1-\frac{1}{M}\right)Q(A/2\sigma) = 2\left(1-\frac{1}{4}\right)Q(A/2\sigma) = 10^{-4}$$

$$\therefore \qquad Q(A/2\sigma) = \frac{2}{3}\times 10^{-4} = \frac{1}{2}erfc\left(\frac{A}{2\sqrt{2}\sigma}\right)$$

Hence, from the Q-function tables, or the error function tables, we get

$$\left(\frac{A}{2\sigma}\right) = 3.8$$

The noise power across 50Ω input resistance of the detector is given to be 10^{-5} W.

Since the noise is given to be of zero mean, its variance is equal to its average power and so its standard deviation σ is equal to its r.m.s. value.

If r.m.s. value of noise voltage at the input to the detector = σ, then

$$\frac{\sigma^2}{R} = \text{average power} = 10^{-5}\text{ W}$$

$$\therefore \qquad \sigma = \sqrt{10^{-5}\times 50} = 2.236\times 10^{-2}\text{ V}$$

Since $(A/2\sigma)$ has already been determined to be 3.8,

$$A = 2\times 2.236\times 10^{-2}\times 3.8 = 16.9936\times 10^{-2} = 0.17\text{ V}$$

and

$$\frac{A}{2} = \frac{0.17}{2} = 0.0855\text{ V} = 85.5\text{ mV}$$

As per Eq. (11.65), the signaling levels, viz., a_ks are

i.e., $a_k = \pm A/2$ and $\pm 3A/2$ for 4-ary signaling.

∴ the signaling levels are ± 85.5 mV and ± 256 mV

Then, since the four amplitude levels are equally likely, from Eq. (11.67), the average received signal power = $\overline{a_k^2}$, where

$$\overline{a_k^2} = \left(\frac{M^2-1}{12}\right)A^2$$

$$S_R = \overline{a_k^2} = \left(\frac{4^2-1}{12}\right)(0.17)^2 = \frac{5}{4}\times 0.0289 = 0.36125\text{ W}$$

Since the channel attenuation = 20 dB = a ratio of 100, the transmitted power S_T is given by

$$S_T = S_R\times 100 = 0.036125\times 100 = 3.6125\text{ W}$$

11.3 COMBINED EFFECT OF CHANNEL NOISE AND ISI

In Section 11.2.5, we had discussed about the optimum receiver and arrived at the matched filter receiver taking into account only the channel noise and completely ignoring the ISI caused by the dispersive nature of the channel.

Then, in Section 10.3, we discussed about the receiver only from the point of view of eliminating ISI and found that appropriately shaping the received pulse provided the solution, but here, the effect of channel noise was ignored.

In reality, however, the channel noise and ISI act together to cause errors. Hence, any attempt to optimize the baseband receiver must take their combined effect into account. There are two approaches adopted for this purpose. One is to use what is called a '*zero-forcing equalizer*' before the sampler of the decision device. The other, and a more refined approach, uses the '*mean-square error criterion*' for minimizing the effects of both the channel noise and also the ISI on the probability of error. In the following sections, we will discuss these two approaches very briefly. We shall take up the mean square error criterion approach first. The equalizer required for this, as well as the zero-forcing type of equalizer will be taken up in the next section which deals with equalization.

Optimum receiver using minimum mean-square error criterion Referring to the binary baseband transmission system shown in Fig. 10.3 we note that the output of the receive filter, $H_R(f)$ is obtained by convolving the channel output with the impulse response, $h_R(t)$, of the filter. Let $q(t)$ be the convolution of the impulse response, $h_T(t)$ of the transmit filter with the impulse response, $h_C(t)$ of the channel. Then the channel output is

$$x(t) = \sum_k a_k q(t - kT) + w(t)$$

where a_k is the symbol transmitted at time $t = kT$, and $w(t)$ is the channel noise. Since $x(t)$ consists of a signal component and a noise component, and since the receive filter is a linear one, which convolves its own impulse response with this channel output, the receive filter output also will have a signal component and a noise component. When this is sampled at $t = iT$, the ith sampling instant, let us say we get $y(iT)$. In fact, ideally, we should get only a_i as this sample. But what we get as $y(iT)$ is a random variable which is the sum of a signal component (which too, will not be equal to a_i because of ISI) and a noise component. The difference between $y(iT)$ and a_i is the error and is given by

$$e_i = y(iT) - a_i$$

The mean-square error is, say J, given by

$$J = E[e_i^2]$$

where E denotes the 'expectation'.

In order to get a receive filter that is optimum in the 'minimum mean-square error' sense, we differentiate J with respect to $h_R(t)$, the impulse response of the receive filter and equate the result to zero. The optimum $h_R(t)$ so obtained, is called the 'minimum mean-square equalizer' and is given by

$$\underset{\text{(MMSE)}}{H_{\text{opt}}(f)} = \frac{Q^*(f)}{S_q(f) + (\eta/2)} \tag{11.71}$$

where $Q(f) = \mathcal{F}[q(t)]$, $S_q(f)$ is the PSD of the pulse $q(t)$ and $\eta/2$ is the two-sided PSD of the zero-mean white noise on the channel.

The MMSE optimum filter, whose transfer function is given in Eq. (11.70), can be realized as the cascade connection of a matched filter with impulse response $q(-t)$, where $q(t)$ is equal to the convolution of $h_T(t)$ and $h_C(t)$, and a transversal or tapped delay-line equalizer whose transfer function is the inverse of $\left(S_q(f) + \frac{\eta}{2}\right)$. This equalizer, which is theoretically of infinite length, is generally approximated by a finite length transversal filter with $2N$ weights, whose delay elements produce exactly T sec delay each. A block diagram of the MMSE linear receiver is shown in Fig. 11.19.

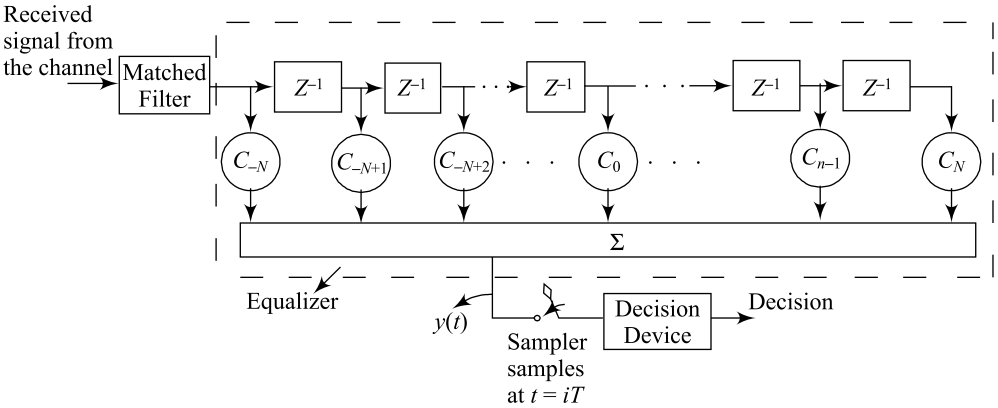


Fig. 11.19 *Optimum (MMSE) linear receiver*

11.4 DETECTION OF BAND PASS DIGITAL SIGNALS AND PROBABILITY OF ERROR

Recall our discussion of optimum baseband binary detection presented in Sections 11.2.5 and 11.3. There we adopted a notation that makes the analysis applicable to baseband as well as band pass binary reception by an appropriate interpretation of $s_1(t)$ and $s_2(t)$. In the case of baseband binary reception $s_1(t)$ and $s_2(t)$ represented two different levels –0 and 1 or, –1 and 1. But, in the case of band pass binary reception, $s_1(t)$ and $s_2(t)$ will be used to represent two different waveforms, as listed out in Table 10.1. Hence, in what follows, we will be making use of the results obtained in Section 11.2.3 with this interpretation.

11.4.1 Coherent Binary ASK Reception

Here
$$s_1(t) = 0 \text{ and } s_2(t) = A\cos\omega_c t \tag{11.72}$$

∴
$$s_{01}(t) = \int_0^T s_1(t)h(T-t)\,dt = \int_0^T 0\cdot dt = 0 \tag{11.73}$$

and ∴
$$s_{02}(t) = \int_0^T s_2(t)h(T-t)\,dt = \int_0^T s_2(t)p(t)\,dt = \int_0^T s_2(t)[s_2(t)-s_1(t)]\,dt$$
$$= \int_0^T s_2^2(t)\,dt = \int_0^T A^2\cos\omega_c t dt = \frac{A^2}{2}\int_0^T (1+\cos\omega_c t)\,dt$$
$$= \frac{A^2}{2}\int_0^T 1\cdot dt + \frac{A^2}{2}\int_0^T \cos 2\omega_c t\,dt$$

But as stated earlier, T contains an integer number of the carrier wave cycles. Hence, the second integral is zero.

∴
$$s_{02}(T) = \frac{A^2 T}{2} \tag{11.74}$$

Further, the optimum threshold (with $P_0 = P_1$ as we have always been assuming), is given by

$$V_{\text{opt}} = \frac{s_{01}(T) + s_{02}(T)}{2} = \frac{A^2 T}{4} \tag{11.75}$$

Also, $$d_{\max}^2 = \int_{-\infty}^{\infty} \frac{|P(f)|^2}{(\eta/2)}\,df = \frac{2}{\eta}\int_0^T p^2(t)\,dt \qquad \text{(Parseval's theorem)}$$

$$\therefore \quad d_{\max}^2 = \frac{2}{\eta}\int_0^T A^2\cos^2\omega_c t\,dt = \frac{A^2T}{\eta} \tag{11.76}$$

$$\therefore \quad P_{e\min} = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{A^2T}{4\eta}}\right] = \frac{1}{2}erfc\left[\sqrt{\frac{A^2T}{8\eta}}\right] \tag{11.77}$$

Since bit energy for $b_k = 1$ is $\int_0^T s_2^2(t)\,dt = \frac{A^2T}{2}$ and bit energy for $b_k = 0$ is equal to zero for ASK, with $P_0 = P_1$, the average bit energy is

$$E_{\text{av}} = \frac{A^2T}{4} \tag{11.78}$$

Substituting this in Eq. (11.77), we get

$$P_{e\min} = Q\left[\sqrt{\frac{E_{\text{av}}}{\eta}}\right] = \frac{1}{2}erfc\left[\sqrt{\frac{E_{\text{av}}}{2\eta}}\right]$$

$$\therefore \quad \boxed{\underset{\text{coherent ASK}}{P_{e\min}} = \frac{1}{2}erfc\left[\sqrt{\frac{E_{\text{av}}}{2\eta}}\right]} \tag{11.79}$$

11.4.2 Non-Coherent Binary ASK Reception

In binary ASK, for $b_k = 1$, the signal $A\cos\omega_c t$ is transmitted and for $b_k = 0$, nothing is transmitted. Hence, if the channel noise is assumed to be zero-mean AWGN, the received signal $r(t)$, under the two conditions, is given by

$$b_k = 0\ ;\ H_1:\ r(t) = 0 + n_W(t);\ 0 \le t \le T$$

$$b_k = 1\ ;\ H_2:\ r(t) = A\cos\omega_c t + n_W(t);\ 0 \le t \le T \tag{11.80}$$

where $n_W(t)$ represents zero-mean white Gaussian noise.

The detection process consists of simply determining during each bit period of T sec, whether $b_k = 0$ or 1, i.e., whether hypothesis H_1 is true or hypothesis H_2 is true. From the nature of the received signal as given above under the two conditions, the following receiver structure may be used:

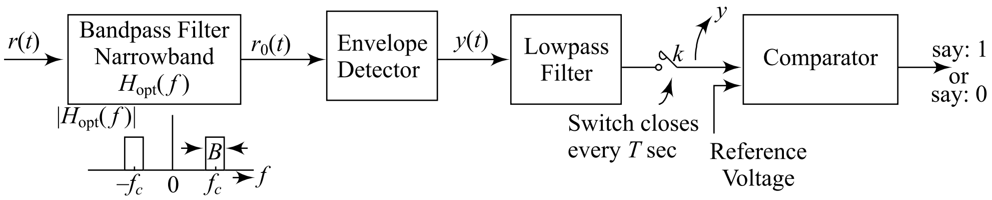


Fig. 11.20 *Non-coherent Binary ASK receiver*

The band pass filter centered on f_c is the matched filter for the signal $Ap(t)\cos\omega_c t$ where $p(t)$ is a rectangular pulse of amplitude 1 and duration T sec. The output of the BPF is $r_0(t)$. Then

$$H_1 : r_0(t) = 0 + n(t); \quad 0 \le t \le T$$

$$H_2 : r_0(t) = A \cos \omega_c t + n(t); \quad 0 \le t \le T \tag{11.81}$$

where $n(t)$ is zero-mean Gaussian band pass noise centered on f_c.

We know that the envelope detector extracts the envelope of the input given to it, which is $r_0(t)$ in the present case. Under hypothesis H_1, $r_0(t)$ is simply zero mean Gaussian band pass noise $n(t)$, whose variance is $\sigma^2 = \eta B$ and mean is zero, where $\eta/2$ is the two-sided PSD of the white noise on the channel and B is the bandwidth of the band pass filter. This band pass noise $n(t)$ centered on f_c may be represented by its inphase and quadrature components as:

$$n(t) = n_i(t) \cos \omega_c t - n_q(t) \sin \omega_c t \tag{11.82}$$

$\therefore$ envelope $R_n(t)$ of this band pass noise is given by

$$R_n(t) = \sqrt{n_i^2(t) + n_q^2(t)} \tag{11.83}$$

and

$$\theta(t) = \tan^{-1}\left[\frac{n_q(t)}{n_i(t)}\right] \tag{11.84}$$

Since $n(t)$ is zero-mean Gaussian process, $n_i(t)$ and $n_q(t)$ are also zero-mean Gaussian processes which are statistically independent and they have the same variance since

$$\overline{n_i^2(t)} = \overline{n_q^2(t)} = \overline{n^2(t)} = \sigma^2 \tag{11.85}$$

Hence, it follows that $R_n(t)$, i.e., the envelope of $n(t)$ will have Rayleigh density. Thus, under hypothesis H_1, the sample y of this envelope will be a random variable with Rayleigh density function. Thus,

$$H_1 : f_Y(y) = \frac{y}{\sigma^2} e^{-y^2/2\sigma^2}; \quad y \ge 0 \tag{11.86}$$

where σ^2 is the variance of the band pass noise.

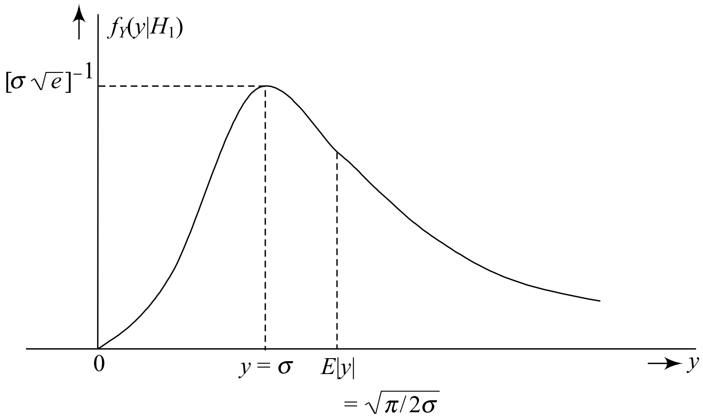


Fig. 11.21 *Rayleigh density function*

Under hypothesis H_2, i.e., for $b_k = 1$, the output of the band pass filter is

$$H_2 : r_0(t) = A \cos \omega_c t + n(t) \quad ; \quad 0 \le t \le T$$

Hence, in this case, the output of the envelope detector will be the envelope of a cosine signal with frequency f_c plus band pass noise (zero-mean Gaussian) centered on f_c. We know (see Section 6.4.2) that this envelope $y(t)$ will have Ricean distribution. Its sample y (under hypothesis H_2) will be a Ricean random variable with density function given by

$$f_Y(y|H_2) \cong \sqrt{\frac{y}{2\pi A\sigma^2}}\, e^{-\frac{(y-A)^2}{2\sigma^2}} \;;\; y \geq 0 \qquad (11.87)$$

Thus, a plot of the two conditional density functions will appear as shown in Fig. 11.23.

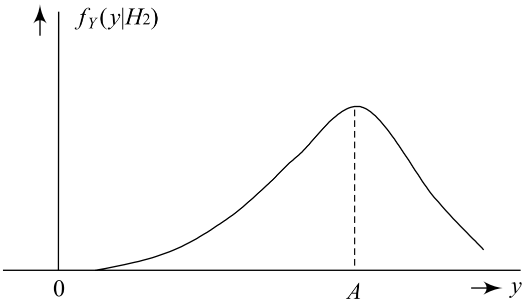


Fig. 11.22 *Ricean density function*

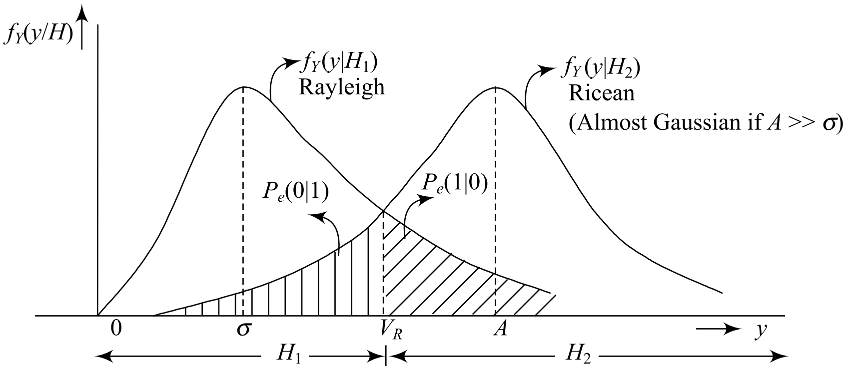


Fig. 11.23 *Conditional density functions and threshold V_R for $P_0 = P_1$*

Assuming the probabilities of transmission of a baseband 1 and 0 to be equal, i.e., $P_0 = P_1$, the optimum threshold is the value of y corresponding to the intersection of the two conditional density functions. This is marked as V_R. Substituting V_R for y in the two density functions, equating them and solving for V_R, we get

$$V_R \approx \frac{A}{2}\sqrt{1 + \frac{8\sigma^2}{A^2}} \qquad (11.88)$$

when the signal-to-noise ratio is large, i.e., when $(A^2/\sigma^2) >> 1$, the term $(8\sigma^2/A^2)$ can be neglected in comparison with 1 so that

$$V_R \approx \frac{A}{2} \qquad (11.89)$$

Let $P_e(1|0)$ = Probability of the receiver saying 1 even though a 0 was transmitted.
and $P_e(0|1)$ = Probability of the receiver saying 0 even though a 1 was transmitted.
These two probabilities are given by the areas marked in Fig. 11.23.
The overall average probability of error, P_e, is then given by

$$\begin{aligned} P_e &= P_0 P_e(1|0) + P_1 P_e(0|1) \\ &= 0.5P_e(1|0) + 0.5P_e(0|1) \end{aligned} \qquad (11.90)$$

From Fig. 11.23, we may write $P_e(1|0)$ and $P_e(1|0)$ as

$$P_e(1|0) = \int_{y=V_R=A/2}^{\infty} \frac{y}{\sigma^2}\exp\left[-\frac{y^2}{2\sigma^2}\right]dy = \exp\left[-\frac{A^2}{8\sigma^2}\right] \qquad (11.91)$$

and
$$P_e(0|1) \cong \int_{-\infty}^{y=A/2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-A)^2}{2\sigma^2}\right] dy$$

Note *Ricean density has been approximated by Gaussian density.*

$$= Q\left[\frac{A}{2\sigma}\right] \tag{11.92}$$

If we now use the approximation

$$Q[x] \approx (e^{-x^2/2})/x\sqrt{2\pi} \tag{11.93}$$

which is quite valid for large values of the argument, x, we may write $P_e(0|1)$ as

$$P_e(0|1) \cong \sqrt{\frac{4\sigma^2}{2\pi A^2}} \exp\left[-\frac{A^2}{2\sigma^2}\right] \tag{11.94}$$

$\therefore$ substituting for $P_e(1|0)$ and $P_e(0|1)$ in Eq. (11.90)

$$P_e \approx \frac{1}{2}\left[1 + \sqrt{\frac{4\sigma^2}{2\pi A^2}}\right] e^{-A^2/8\sigma^2} \approx \frac{1}{2} e^{-A^2/8\sigma^2} \tag{11.95}$$

But $\sigma^2 = \eta B_T$, where B_T is transmission bandwidth. However, $B_T = 1/T$

$\therefore$ $$\sigma^2 = \eta/T \qquad \therefore P_e = \frac{1}{2} e^{-A^2T/8\eta} \qquad \text{(From Eq. (11.95))}$$

But $\frac{A^2T}{4} = E_{b\text{av}} = E_b$

$\therefore$ $$\boxed{\underset{\substack{\text{ASK}\\(\text{non-coherent})}}{P_e} \cong \frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]} \tag{11.96}$$

Example 11.17 Determine the transmitted power needed to transmit binary data at a rate of 1 Mbps over a channel with zero-mean AWGN of two-sided PSD equal to 10^{-12} *W*/Hz and a total transmission loss of 40 dB if the system used is (a) non-coherent ASK, and (b) coherent ASK. In all the cases, P_e should not exceed 10^{-4}.

Solution

(a) *Non-coherent ASK*

$$P_e = \frac{1}{2} e^{-\frac{E_b}{2\eta}} = 10^{-4}$$

$\therefore$ $(E_b/2\eta) = 8.517$ since $\eta/2 = 10^{-12}$, $\eta = 2 \times 10^{-12}$ W/Hz

Hence, $E_b = 8.517 \times 4 \times 10^{-12}$. watt-sec = Average bit energy

Since r_b, the transmission rate = 10^6 bits/sec, average transmitted power is given by

$$P_{\text{av}} = E_b \cdot r_b = 8.517 \times 4 \times 10^{-12} \times 10^6 = 34.068 \times 10^{-6} W$$

Since the transmission loss L, of the channel is 40 dB,

$$10 \log_{10} L = 40 \quad \therefore \; L = 10^4$$

$\therefore$ transmitted average power = $34.068 \times 10^{-6} \times 10^4 = 0.34068$ W

(b) *Coherent ASK*

For coherent ASK, probability of error, $P_e = \frac{1}{2} erfc\left[\sqrt{\frac{E_{\text{av}}}{2\eta}}\right]$

$$\therefore \quad \frac{1}{2} erfc\sqrt{\frac{E_{\text{av}}}{2\eta}} = 10^{-4} \qquad \therefore \; \sqrt{\left(\frac{E_{\text{av}}}{2\eta}\right)} = \left(\frac{3.7}{\sqrt{2}}\right) \qquad \therefore \; E_{\text{av}} = 13.69 \times 2 \times 10^{-12} \text{ Watt-sec}$$

$\therefore$ average received power = $E_{\text{av}} \times R_b$ (where R_b is transmission rate in bits/sec)

$$\therefore \quad P_{\text{av}}(\text{received}) = 13.69 \times 2 \times 10^{-12} \times 10^6 = 27.38 \times 10^{-6} \text{Watt}$$

But transmission loss = 10^4 (i.e., 40 dB)

$\therefore$ average transmitted power = $27.38 \times 10^{-6} \times 10^4 = 0.2738\,W$

Example 11.18 A binary transmission system with a transmitted power of 200 mW uses a channel with zero-mean AWGN of two-sided PSD equal to 10^{-15} W/Hz and a total transmission loss of 90 dB. If the probability of error, P_e is not to exceed 10^{-4}, what is the maximum allowable bit rate using (a) non-coherent ASK, and (b) coherent ASK?

Solution Transmitted average power = 200 mW
Transmission loss = 90 dB = a ratio of 10^9.

$\therefore$ received average power = $200 \times 10^{-3} \times 10^{-9} = 2 \times 10^{-10}$ W

(a) *Non-coherent ASK*

$$P_e = 10^{-4} = \frac{1}{2} e^{-\frac{E_b}{2\eta}} \quad \therefore \; \frac{E_b}{2\eta} = 8.517$$

Since $\eta/2 = 10^{-15}$ W/Hz, $\eta = 2 \times 10^{-15}$ W/Hz

$$\therefore \quad E_b = 8.517 \times 2\eta = 8.517 \times 4 \times 10^{-15} = 34.068 \times 10^{-15} \text{Watt-sec}$$
$$= \text{Average bit energy}$$

$\therefore$ if r_b is the transmission rate in bits/sec, the average power is

$$P_{\text{av}} = E_b \cdot r_b = 34.068 \times 10^{-15} \times r_b$$
$$= \text{Average received power } = 2 \times 10^{-10} \text{W}$$

$$\therefore \quad r_b = \frac{2 \times 10^{-10}}{34.068 \times 10^{-15}} = \frac{200 \times 10^3}{34.068} \cong 6 \text{ kbps}$$

(b) *Coherent ASK*

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_{\text{av}}}{2\eta}} \qquad \therefore \; E_{\text{av}} = 27.38 \times 10^{-15} \text{ Watt-sec.}$$

$\therefore$ if R_b is the transmission rate in bits/sec, the average received power is

$$P_{av} = E_{av} \cdot R_b = 27.38 \times 10^{-15} \times R_b = 2 \times 10^{-10}$$

$$\therefore \qquad R_b = \frac{2 \times 10^{-10}}{27.38 \times 10^{-15}} = \frac{2 \times 10^5}{27.38} = 7.3046 \text{ kbps}$$

11.4.3 Coherent Reception of Sunde's Binary FSK

For binary FSK, the two signals $s_1(t)$ and $s_2(t)$ that are transmitted, are given by

$$s_1(t) = A\cos(\omega_c - \omega_d)t;\ 0 \le t \le T; \text{corresponding to } b_k = 0$$

and

$$s_2(t) = A\cos(\omega_c + \omega_d)t;\ 0 \le t \le T; \text{corresponding to } b_k = 1$$

As we are considering Sunde's continuous phase binary FSK, as stated earlier in our discussion on power spectrum of BFSK, we have

$$f_d = \frac{1}{2T}$$

$$\therefore \qquad \left.\begin{aligned} s_1(t) &= A\cos(\omega_c - \pi/T)t;\ 0 \le t \le T \text{ for } b_k = 0 \\ s_2(t) &= A\cos(\omega_c + \pi/T)t;\ 0 \le t \le T \text{ for } b_k = 1 \end{aligned}\right\} \qquad (11.97)$$

Refer to the derivation of the transfer function of an optimum filter in Section 11.2.4, which we made very general so as to be applicable to baseband as well as band pass signaling.

For convenience, we reproduce here Fig. 11.24 of a correlation receiver.

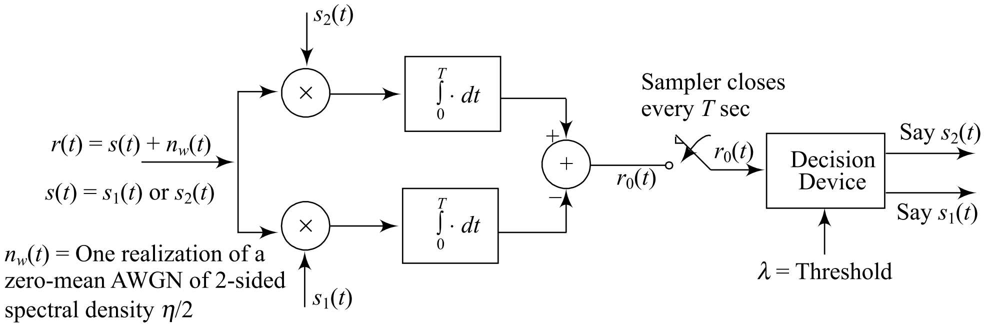


Fig. 11.24 *A correlation receiver for binary signals*

Let $r_0(T)$ = Output of the sampler at $t = T$ be denoted by $s_{01}(T)$ when $r(t) = s_1(t)$ alone.

Then

$$s_{01}(T) = -\int_0^T s_1(t) \cdot s_1(t)dt = -\int_0^T s_1^2(t)\,dt$$

$$= -\int_0^T A^2\cos^2(\omega_c - \pi/T)t\,dt = -A^2\int_0^T \left[\frac{1 + \cos 2(\omega_c - \pi/T)t}{2}\right]dt$$

$$= -\frac{A^2T}{2} - \frac{A^2}{2}\int_0^T \cos 2\pi(2f_c - 1/T)t\,dt$$

But, as already stated earlier, matters are always so arranged that there will be an integer number of carrier cycles say, n, in the bit duration T.

$$\therefore \qquad T = \frac{n}{f_c}$$

So $\cos 2\pi(2f_c - 1/T)t$ is a cosine signal which will complete $(2n - 1)$ full cycles of the carrier frequency f_c in the integration period T. Thus, the integral is zero.

$$\therefore \qquad s_{01}(T) = -\frac{A^2T}{2} \tag{11.98}$$

Similarly, if $s_{02}(T)$ denotes the value of $r_0(T)$, the output of the sampler at $t = T$ when $r(t) = s_2(t)$ alone,

$$s_{02}(T) = \int_0^T s_2^2(t)\,dt = \frac{A^2T}{2} \tag{11.99}$$

Assuming, as usual, that the probabilities of transmission of $s_1(t)$ and $s_2(t)$ are equal, the optimum threshold, i.e., λ_{opt} will be given by

$$\lambda_{\text{opt}} = V_{\text{opt}} = \frac{s_{01}(T) + s_{02}(T)}{2} = 0 \tag{11.100}$$

Further, from Eq. (11.100),

$$d_{\max}^2 = \int_{-\infty}^{\infty} [|P(f)|^2/S_n(f)]\,df \tag{11.101}$$

where in this case, $P(f) = \mathcal{F}[s_2(t) - s_1(t)]$ and $S_n(f) = \eta/2$.

$$\therefore \qquad d_{\max}^2 = \frac{2}{\eta}\int_{-\infty}^{\infty} |P(f)|^2 df = \frac{2}{\eta}\int_0^T p^2(t)\,dt \quad \text{(From Parseval's theorem of FT)} \tag{11.102}$$

where $\qquad p(t) = [s_2(t) - s_1(t)]$

$$\therefore \qquad d_{\max}^2 = \frac{2}{\eta}\int_0^T [s_2(t) - s_1(t)]^2 dt$$

$$= \frac{2}{\eta}(\text{Energy of } p(t)) \tag{11.103}$$

Now, before finding the energy of $p(t)$, we will show that signals $s_1(t)$ and $s_2(t)$ are orthogonal over the interval 0 to T.

i.e., $\qquad \int_0^T s_1(t)s_2(t)\,dt = 0$

To prove this, let us substitute for $s_1(t)$ and $s_2(t)$ using Eq. (11.97)

$\therefore$ $(s_1(t), s_2(t))$ = Inner product of $s_1(t)$ and $s_2(t)$

$$= A^2\int_0^T \{\cos(\omega_c - \pi/T)t \cdot \cos(\omega_c + \pi/T)t\}\,dt$$

$$= \frac{A^2}{2}\int_0^T \{\cos 2\omega_c t + \cos(2\pi t/T)\}\,dt$$

Since there will be $2n$ full cycles of the cosine signal $\cos 2\omega_c t$ in the integration period, the first integral is zero.

$$\therefore \qquad (s_1(t), s_2(t)) = \frac{A^2}{2}\int_0^T \cos\left(\frac{2\pi}{T}\right) t dt = \frac{A^2}{2}\frac{T}{2\pi}\sin\left(\frac{2\pi}{T}\right)t\Bigg|_{t=0}^{T}$$

$$= \frac{A^2T}{4\pi}(\sin 2\pi - \sin 0) = 0 \qquad (11.104)$$

Hence, the two signals $s_1(t)$ and $s_2(t)$ of Eq. (11.97) are orthogonal to each other.

Now, reverting to Eq. (11.103), we have

$$d^2_{\max} = \frac{2}{\eta}\text{ [Energy of }\{s_2(t) - s_1(t)\}] = \frac{2}{\eta}\int_0^T [s_2(t) - s_1(t)]^2 dt$$

$$= \frac{2}{\eta}\int_0^T s_2^2(t)\,dt + \frac{2}{\eta}\int_0^T s_1^2(t)\,dt - \frac{4}{\eta}\int_0^T [s_1(t)s_2(t)]\,dt$$

since $s_1(t)$ and $s_2(t)$ are orthogonal, the last integral is zero.

Hence,

$$d^2_{\max} = \frac{2}{\eta}\left[\frac{A^2T}{2} + \frac{A^2T}{2}\right] = \frac{2A^2T}{\eta}, \qquad (11.105)$$

since energy of each of the signals is equal to $A^2T/2$.

Then, from Eq. (11.48), we know that P_e is given by

$$P_{e_{\min}} = Q\left[\frac{d_{\max}}{2}\right]$$

If the bit energy is denoted by E_b,

$$E_b = \frac{A^2T}{2}$$

$$\therefore \qquad d^2_{\max} = \frac{2\times 2E_b}{\eta} \qquad \therefore\ d_{\max} = 2\sqrt{\frac{E_b}{\eta}}$$

$$\therefore \qquad P_{e_{\min}} = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{E_b}{\eta}}\right]$$

$$= \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$$

$$\boxed{\underset{\substack{\text{(CPBFSK)}\\ \text{coherent}}}{P_{e_{\min}}} = Q\left[\sqrt{\frac{E_b}{\eta}}\right] = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]} \qquad (11.106)$$

In obtaining Eq. (11.106), we have made use of the results obtained during the course of the derivation of the probability of error for an optimum filter. However, the same result may be obtained making use of signal space approach, identifying the boundary of the decision regions and then calculating the probability of error. This latter approach, though somewhat similar to the previous approach, gives a better insight into the problem and so will be outlined in what follows.

11.4.4 Probability of Error for Sunde's Continuous-Phase Binary FSK Using Signal-Space Concepts

Let the two waveforms transmitted be $s_1(t)$ and $s_2(t)$ given by

$$b_k = 0: \quad s_1(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T}} \cos 2\pi(f_c - 1/2T)t; & 0 \le t \le T \\ 0 & ; \text{ elsewhere} \end{cases} \tag{11.107}$$

and

$$b_k = 1: \quad s_2(t) = \begin{cases} \sqrt{\dfrac{2E_b}{T}} \cos 2\pi(f_c + 1/2T)t; & 0 \le t \le T \\ 0 & ; \text{ elsewhere} \end{cases} \tag{11.108}$$

Let the channel noise be zero-mean additive white Gaussian noise with a two-sided PSD of η/2, and let the a priori probability of transmission of $s_1(t)$ be the same as that of $s_2(t)$. We have already proved that Sunde's BFSK signals represented by Eqs. (11.107) and (11.108) are orthogonal over the time interval [0, *T*] (see Eq. (11.104)). They are not normalized since

$$\|s_1(t)\| = \|s_2(t)\| = \sqrt{E_b} \tag{11.109}$$

Dividing $s_1(t)$ and $s_2(t)$ of Eqs. (11.107) and (11.108) by the norm of each, viz., $\sqrt{E_b}$, we get the normalized signal pair $\phi_1(t)$ and $\phi_2(t)$ which are orthonormal.

$$\phi_1(t) = \begin{cases} \sqrt{\dfrac{2}{T}} \cos 2\pi(f_c - 1/2T)t; & 0 \le t \le T \\ 0 & ; \text{ elsewhere} \end{cases} \tag{11.110}$$

and

$$\phi_2(t) = \begin{cases} \sqrt{\dfrac{2}{T}} \cos 2\pi(f_c + 1/2T)t; & 0 \le t \le T \\ 0 & ; \text{ elsewhere} \end{cases} \tag{11.111}$$

Since these two are orthonormal, they can be conveniently used as the basis set for the two-dimensional signal space in which $s_1(t)$ and $s_2(t)$ are located.

The coordinate of $s_1(t)$ along $\phi_1(t)$ basis signal is given by their inner product:

$$\begin{aligned} (s_1(t), \phi_1(t)) &= \int_0^T \sqrt{\frac{2E_b}{T}} \cos 2\pi(f_c - 1/2T)t \cdot \sqrt{\frac{2}{T}} \cos 2\pi(f_c - 1/2T)t dt \\ &= \frac{1}{2}\frac{2}{T}\sqrt{E_b} \int_0^T [1 + \cos 4\pi(f_c - 1/2T)t] dt \\ &= \sqrt{E_b} \end{aligned} \tag{11.112}$$

Since $s_1(t)$ is orthogonal to $\phi_2(t)$, its components along $\phi_2(t) = 0$. Hence, in the signal space spanned by $\phi_1(t)$ and $\phi_2(t)$, the signal $s_1(t)$ is represented by the vector

$$s_1 = \begin{bmatrix} \sqrt{E_b} \\ 0 \end{bmatrix} \tag{11.113}$$

Similarly, $s_2(t)$ has a coordinate of 0 along $\phi_1(t)$ since it is orthogonal to $\phi_1(t)$ but has a coordinate of $\sqrt{E_b}$ along $\phi_2(t)$. Hence, we may represent the signal $s_2(t)$ by the vector

$$s_2 = \begin{bmatrix} 0 \\ \sqrt{E_b} \end{bmatrix} \tag{11.114}$$

Figure 11.25 shows the signal space spanned by $\phi_1(t)$ and $\phi_2(t)$ and the representation of signals $s_1(t)$ and $s_2(t)$ in this space as points m_1 and m_2 respectively.

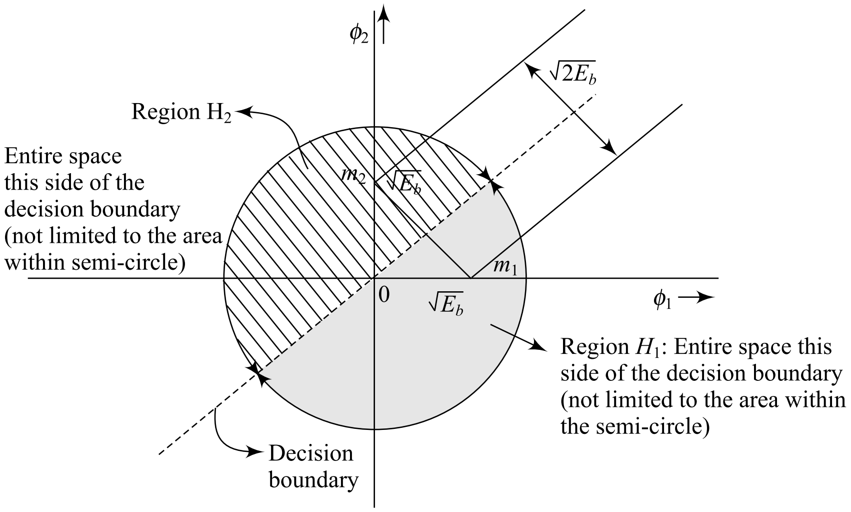


Fig. 11.25 *Signal space of CPBFSK system. Point m_1 represents $s_1(t)$ and point m_2 represents $s_2(t)$. Decision regions are on either side of the decision boundary (only a limited portion of each region is shown shaded).*

If the received signal is $r(t)$, we know that

$$r(t) = s(t) + n_w(t) \tag{11.115}$$

where $s(t)$ may be either $s_1(t)$ or $s_2(t)$ and $n_w(t)$ is one realization of a zero-mean white noise process with two-sided PSD of $\eta/2$. The problem is to observe $r(t)$ and correctly decide whether $s(t)$ equals $s_1(t)$ or $s_2(t)$. As shown in the diagram, point m_1 with coordinates $\left(\sqrt{E_b}, 0\right)$ represents signal $s_1(t)$ and point m_2 with coordinates $\left(0, \sqrt{E_b}\right)$ represents signal $s_2(t)$ in the signal space. The observation space is now divided into two regions H_1 and H_2 as shown, by the decision boundary running through the origin and the mid-point of the line joining m_1 and m_2.

Since $r(t)$, the received signal has a noise component, it will not entirely belong to this signal space spanned by $\phi_1(t)$ and $\phi_2(t)$, i.e., it may have a component orthogonal to this signal space S, spanned by $\phi_1(t)$ and $\phi_2(t)$. So, to determine whether $r(t)$ is closer to $s_1(t)$ or $s_2(t)$ we take its coordinates along $\phi_1(t)$ and $\phi_2(t)$. Let these coordinates be r_1 and r_2 respectively. Then we know that

r_1 = coordinate of $r(t)$ along $\phi_1(t)$

$$(r(t), \phi_1(t)) = \int_0^T r(t)\,\phi_1(t)dt \tag{11.116}$$

Similarly,

$$r_2 = (r(t), \phi_2(t)) = \int_0^T r(t)\,\phi_2(t)dt \tag{11.117}$$

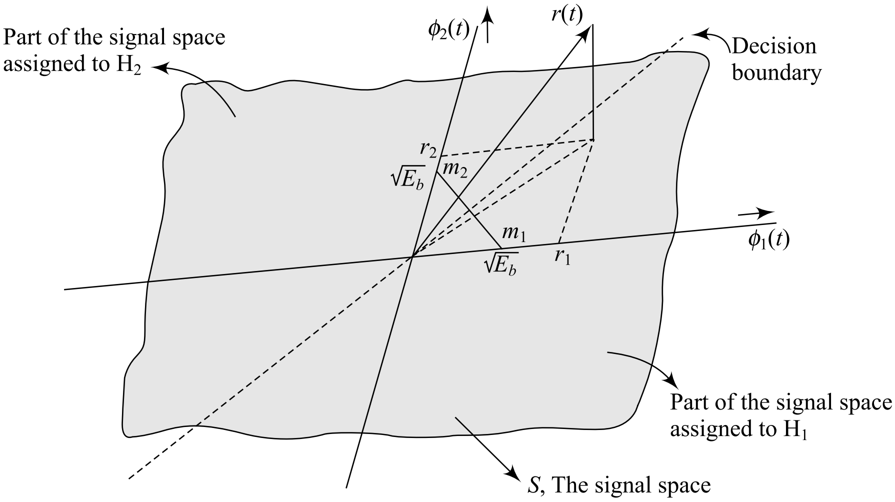


Fig. 11.26 *Signal space S, decision regions and the received signal r(t)*

Now, the decision-making strategy is

if $z \underline{\Delta} (r_1 - r_2)$ is greater than zero, receiver should say $b_k = 0$

if $(r_1 - r_2)$ is less than zero, receiver should say $b_k = 1$

Note that being the sum of a deterministic signal $s(t)$ (which is either $s_1(t)$, or $s_2(t)$) and a realization of white noise process, $r(t)$ is also a random process and so r_1 and r_2 are the sample values of random variables R_1 and R_2.

Now,

$$E[R_1|H_1] = E\left[\int_0^T \{r(t)|H_1\}\,\phi_1(t)dt\right] = \int_0^T E\{r(t)|H_1\}\,\phi_1(t)dt = \int_0^T s_1(t)\phi_1(t)dt = \sqrt{E_b}$$

Similarly, $$E[R_2|H_1] = \int_0^T E\{r(t)|H_1\}\,\phi_2(t)dt = \int_0^T s_1(t)\phi_2(t)dt = 0$$

$\therefore$ $$E[z|H_1] = \sqrt{E_b} \tag{11.118}$$

$$\begin{aligned}E[z|H_2] &= E[r_1|H_2] - E[r_2|H_2] \\ &= E\left[\int_0^T \{r(t)|H_2\}\phi_1(t)dt\right] - E\left[\int_0^T \{r(t)|H_2\}\phi_1(t)dt\right] \\ &= \int_0^T E\{s_2(t) + n_w(t)\}\phi_1(t)dt - \int_0^T E\{s_2(t) + n_w(t)\}\phi_2(t)dt \\ &= \int_0^T E\{s_2(t)\phi_1(t)\}dt + \int_0^T E\{n_w(t)\phi_1(t)\}dt - \int_0^T E\{s_2(t)\phi_2(t)\}dt - \int_0^T E\{n_w(t)\phi_2(t)\}dt\end{aligned}$$

The first integral is zero since $s_2(t)$ and $\phi_1(t)$ are orthogonal. The second integral is zero since $\phi_1(t)$ is deterministic and $n_w(t)$ is of zero-mean. The third integral equals $\sqrt{E_b}$. The last integral is zero since $n_w(t)$ is of zero mean.

Thus,

$$E[z|H_2] = -\sqrt{E_b} \tag{11.119}$$

Since $E[z|H_1] = \sqrt{E_b}$ while $E[z|H_2] = -\sqrt{E_b}$, the variance of z is not dependent on whether H_1 is true or H_2 is true. Also, it can be shown that the random variables R_1 and R_2 are statistically independent and that each has a variance of $\eta/2$. Hence, the variance of z equals the sum of the variances of R_1 and R_2, which is η.

$$\therefore \qquad \text{var}\,[z] = \eta \tag{11.120}$$

Since z is the observed random variable and since its mean but not its variance depends on whether H_1 or H_2 is true, let us examine the conditional density function

$$f(z|H_1) = \frac{1}{\sqrt{2\pi\eta}} \exp\left[-\frac{(z+\sqrt{E_b})^2}{2\eta}\right] \tag{11.121}$$

We have asked the receiver to say H_1 if $z > 0$.
Also

$$f(z|H_2) = \frac{1}{\sqrt{2\pi\eta}} \exp\left[-\frac{(z-\sqrt{E_b})^2}{2\eta}\right] \tag{11.122}$$

and we have asked the receiver to say H_2 if $z < 0$.

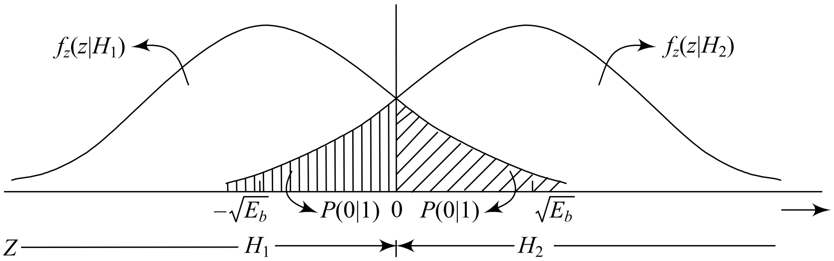


Fig. 11.27 *Conditional density function of z*

$P(1 \mid 0)$ = Probability of the receiver saying 1 even though a 0 was sent

$$= P(z > 0 \mid 0 \text{ was sent})$$

$$= \int_0^\infty f_z(z|H_1)\,dz = \int_0^\infty \frac{1}{\sqrt{2\pi\eta}} \exp\left[-\frac{(z+\sqrt{E_b})^2}{2\eta}\right] dz$$

On simplification, this gives

$$P(1 \mid 0) = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{2\eta}}\right] \tag{11.123}$$

Similarly,

$$P(0 \mid 1) = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{2\eta}}\right] \tag{11.124}$$

$$\therefore \qquad P_e = .5P(1 \mid 0) + 0.5P(0 \mid 1)$$

$$\therefore \qquad \boxed{\underset{\substack{\text{CPBFSK}\\ \text{coherent}}}{P_e} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]} \qquad (11.125)$$

Example 11.19 A BFSK system operating with an unmodulated carrier frequency of f_c, transmits rectangular pulses of cosinusoids of frequency $f_1 = f_c + (f_d/2)$ and $f_2 = f_c - (f_d/2)$ corresponding to binary symbols 1 and 0 respectively, where $f_d = f_1 - f_2 = 1/T$ where T is the duration of a binary 0 or 1. At the input to the receiver, the transmitter can create a maximum of 100 mW of power. Noise (one-sided) PSD at the input to the receiver is 10^{-10} W/Hz. What is the maximum bit rate which the system can support if the bit-error probability is to be 10^{-6}? What is the nominal bandwidth of the BFSK signal, if $f_2 = 85$ MHz?

Solution The BFSK system of this example is clearly a Sunde's continuous phase BFSK system for which as per Eq. (11.125),

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_b}{2\eta}} = 10^{-6} \text{ (Given)}$$

From the error-function tables, we find that the above equation gives

$$\frac{E_b}{\eta} = 22.6 \qquad \therefore\ E_b = 22.6 \times \eta = 22.6 \times 10^{-10} = 2.26 \times 10^{-9}\text{ J}$$

Since E_b is the bit energy, it is equal to the average received signal power multiplied by the bit duration T.

$$\therefore \qquad E_b = 2.26 \times 10^{-9} = 100 \times 10^{-3} \times T$$

$$\therefore \qquad T = \frac{2.26 \times 10^{-9}}{100 \times 10^{-3}} = 2.26 \times 10^{-8}\text{ sec}$$

The bit rate
$$R_b = \frac{1}{T} = \frac{1}{2.26 \times 10^{-8}} = 0.4424 \times 10^8 = 44.24\text{ Mbps}$$

$\therefore$ the maximum bit rate that can be supported = 44.24 Mbps

For Sunde's BFSK, the transmission bandwidth is equal to the bit rate since

$$\text{B.W.} = (f_1 - f_2) = (f_c + f_d/2) - (f_c - f_d/2) = f_d = \frac{1}{T}$$

$$\therefore \text{ if } f_2 = 85\text{ MHz}, \quad f_1 = f_2 + \frac{1}{T} = (85 + 44.24)\text{ MHz}$$
$$= 129.24\text{ MHz}$$

The transmission bandwidth, when $f_2 = 85$ MHz, is given by

$$B_T = 85\text{ MHz to }129.24\text{ MHz.}$$

11.4.5 Probability of Error for Non-Coherent Reception of Binary FSK

In binary FSK, the signal transmitted during any time slot will be either $s_1(t)$ with a frequency of $(f_c - f_d)$, or $s_2(t)$ with a frequency of $(f_c + f_d)$. Hence, a non-coherent FSK receiver will have a structure as shown in Fig. 11.28.

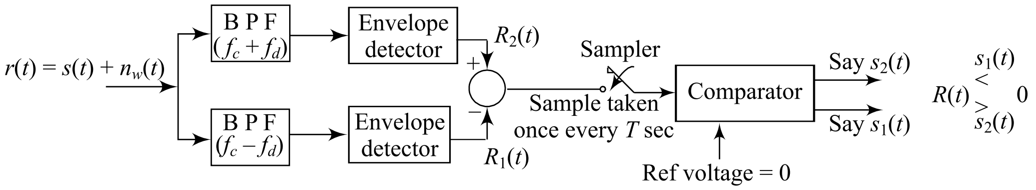


Fig. 11.28 *Non-coherent BFSK receiver structure*

Let the channel noise be zero-mean additive white Gaussian noise. As usual, let $P_1 = P_0 = 0.5$.

$$r(t) = \text{received signal} = s(t) + n_w(t), \tag{11.126}$$

where

$$s(t) = \begin{cases} s_1(t) = \sqrt{\dfrac{2E_b}{T}} \cos 2\pi(f_c - f_d)t & \text{if } b_k = 0 \\ s_2(t) = \sqrt{\dfrac{2E_b}{T}} \cos 2\pi(f_c + f_d)t & \text{if } b_k = 1 \end{cases} \tag{11.127}$$

For convenience let $A = \sqrt{\dfrac{2E_b}{T}}$

Let us suppose $s_1(t)$ has been transmitted. Then $R_2(t)$ will be the envelope of only band pass noise and so will have Rayleigh density, while $R_1(t)$ will be the envelope of a sinusoid (i.e., $s_1(t)$) plus band pass noise and so will have a Ricean density.

∴
$$f_{R_1(t)}(R_1(t)|H_1) = \frac{R_1}{\sigma^2} I_0\left[\frac{AR_1}{\sigma^2}\right] e^{-\left(\frac{R_1^2+A^2}{2\sigma^2}\right)}; R_1 \ge 0 \tag{11.128}$$

where A is the peak amplitude of $s_1(t)$ and

$$A = \sqrt{\frac{2E_b}{T}} \tag{11.129}$$

and

$$\sigma^2 = \eta B_T$$

where B_T is the bandwidth of the BPF = Transmission bandwidth (which is approximately equal to $2/T$).

Since $s_1(t)$ has been transmitted, the BPF in the top branch responds only to noise and so its envelope will have Rayleigh density function. This is given by

$$f_{R_2(t)}(R_2(t)|H_1) = \left(\frac{R_2}{\sigma^2}\right) e^{\frac{-R_2^2}{2\sigma^2}}; R_2 \ge 0 \tag{11.130}$$

Also, since $s_1(t)$ has been transmitted, $R_1(T)$ should be larger than $R_2(T)$, i.e., $R(t) < 0$. Hence, when $s_1(t)$ has been transmitted, if $R(T)$ turns out to be positive, we say that an error has occurred. From the symmetry of the problem, it is clear that the probability of a transmitted $s_1(t)$ being mistaken by the receiver as $s_2(t)$, is the same as the probability of a transmitted $s_2(t)$ being mistaken as an $s_1(t)$.

i.e.,
$$P_e(1|0) = P_e(0|1) \tag{11.131}$$

Also,
$$P_1 = P_0 = 0.5 \tag{11.132}$$

and
$$\begin{aligned} P_e &= P_0 P_e(1|0) + P_1 P_e(0|1) = 0.5 P_e(1|0) + 0.5 P_e(0|1) \\ &= P_e(1|0) = P_e(0|1) \end{aligned} \tag{11.133}$$

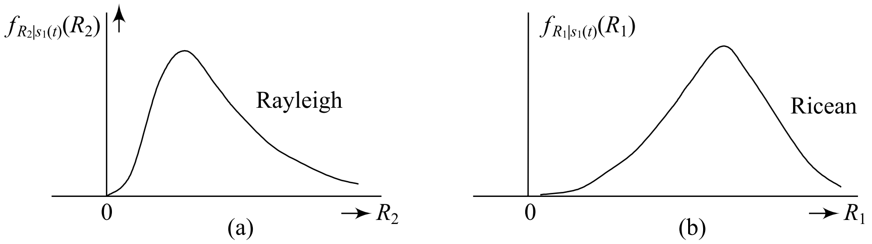


Fig. 11.29 *(a) Rayleigh density function, (b) Ricean density function*

But
$$P_e(1\,|\,0) = P_e[s_2(t)|s_1(t)] = P[R_2(T) > R_1(T)] \tag{11.134}$$

and
$$P[R_2(T) > R_1(T)] = \int_0^{\infty} f_{R_1|s_1(t)}(R_1)\left[\int_{R_1}^{\infty} f_{R_2|s_1(t)}(R_2)\,dR_2\right]dR_1 \tag{11.135}$$

Let
$$\int_{R_1}^{\infty} f_{R_2|s_1(t)}(R_2)\,dR_2 = \int_{R_1}^{\infty} \frac{R_2}{\sigma^2}e^{-R_2^2/2\sigma^2}dR_2 \,\underline{\Delta}\, I \tag{11.136}$$

Put
$$-\frac{R_2^2}{2\sigma^2} = z \quad \therefore \quad dz = -\frac{2R_2dR_2}{2\sigma^2} = -\frac{R_2}{\sigma^2}dR_2$$

Also, when $R_2 = R_1,\ z = -\dfrac{R_1^2}{2\sigma^2}$

$\therefore$
$$I = \int_{-\infty}^{-R_1^2/2\sigma^2} e^z dz = [e^z]_{-\infty}^{-R_1^2/2\sigma^2} = e^{-R_1^2/2\sigma^2}$$

$\therefore$
$$P[R_2 > R_1] = P_e = \int_0^{\infty}\frac{R_1}{\sigma^2}I_0\left(\frac{AR_1}{\sigma^2}\right)e^{-\frac{A^2+R_1^2}{2\sigma^2}}\cdot e^{-R_1^2/2\sigma^2}dR_1$$
$$= \int_0^{\infty}\frac{R_1}{\sigma^2}I_0\left(\frac{AR_1}{\sigma^2}\right)e^{-\left(\frac{2R_1^2+A^2}{2\sigma^2}\right)}dR_1$$

Putting $x = \sqrt{2}R_1,\ dR_1 = \dfrac{dx}{\sqrt{2}}$

$\therefore$
$$P_e = \left[\int_0^{\infty}\frac{x}{\sigma^2}I_0\left(\frac{Ax}{\sqrt{2}\sigma^2}\right)e^{-\left(\frac{x^2+(A/\sqrt{2})^2}{2\sigma^2}\right)}dx\right]\left[\frac{1}{2}e^{-A^2/4\sigma^2}\right] \tag{11.137}$$

But $\displaystyle\int_0^{\infty}\frac{x}{\sigma^2}I_0\left(\frac{Ax}{\sqrt{2}\sigma^2}\right)e^{-\left(\frac{x^2+(A/\sqrt{2})^2}{2\sigma^2}\right)}dx = 1$ (Since the integrand is a Ricean density function)

$\therefore$
$$\underset{\substack{\text{BFSK}\\\text{(non-coherent)}}}{P_e} = \frac{1}{2}e^{-A^2/4\sigma^2} = \frac{1}{2}e^{\frac{-1}{2}\left(\frac{E_b}{\sigma^2T}\right)}$$

But $\sigma^2 = \eta \cdot B_T = \eta/T$ since B_T (for FSK) $= \dfrac{1}{T}$ $\therefore$ $\sigma^2 T = \eta$

$$\therefore \qquad \boxed{\underset{\substack{\text{BFSK}\\\text{(Non-coherent)}}}{P_e} = \frac{1}{2} e^{\frac{-E_b}{2\eta}}} \tag{11.138}$$

11.4.6 Probability of Error for Binary Phase-Shift Keying

In the case of binary phase-shift keying, the two signals, $s_1(t)$ corresponding to $b_k = 0$ and $s_2(t)$ corresponding to $b_k = 1$, are given by

$$\left.\begin{aligned} H_1 : b_k = 0 : s_1(t) &= -A\cos\omega_c t; 0 \le t \le T \\ H_2 : b_k = 1 : s_2(t) &= A\cos\omega_c t; 0 \le t \le T \end{aligned}\right\} \tag{11.139}$$

where $A = \sqrt{2E_b/T}$ and $T = n/f_c$ so that there are an integer number of carrier cycles in a time slot T. We find that, except for a 180° phase shift there is no other difference between the two possible transmitted signals. Such signals are referred to as 'Antipodal Signals'. Since the only thing that distinguishes the two signals is the phase, only coherent detection is possible, as non-coherent detection ignores phase information.

As usual we shall assume channel noise to be AWGN with zero mean and a two-sided power spectral density of $\eta/2$. We shall also assume equiprobable symbols, i.e., $P_0 = P_1 = 0.5$. The received signal may therefore be written as

$$\left.\begin{aligned} H_1 : r(t) &= -A\cos\omega_c t + n_w(t); 0 \le t \le T \\ H_2 : r(t) &= A\cos\omega_c t + n_w(t); 0 \le t \le T \end{aligned}\right\} \tag{11.140}$$

If $s_{01}(T)$ and $s_{02}(T)$ are respectively the outputs of the integrator when $s_1(t)$ alone, and $s_2(t)$ alone are fed to the receiver, then

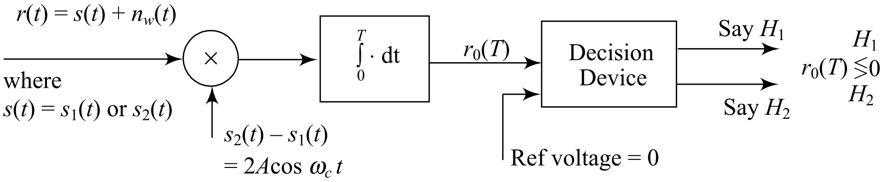


Fig. 11.30 *Receiver structure for (coherent) PSK*

$$\begin{aligned} s_{01}(T) &= \int_0^T s_1(t)[s_2(t) - s_1(t)]\,dt \\ &= \int_0^T (-A\cos\omega_c t)(2A\cos\omega_c t)\,dt = -A^2 T \end{aligned} \tag{11.141}$$

$$\begin{aligned} s_{02}(T) &= \int_0^T s_2(t)[s_2(t) - s_1(t)]\,dt \\ &= \int_0^T (A\cos\omega_c t)(2A\cos\omega_c t)\,dt = A^2 T \end{aligned} \tag{11.142}$$

$$\therefore \qquad \text{threshold} = V_{\text{opt}} = \frac{s_{01}(T) + s_{02}(T)}{2} = 0 \tag{11.143}$$

Further, from Eq. (11.47), we have

$$d_{\max}^2 = \int_{-\infty}^{\infty} \frac{|P(f)|^2}{S_n(f)} df = \frac{2}{\eta}\int_{-\infty}^{\infty} |P(f)|^2 df = \frac{2}{\eta}\int_{-0}^{T} p^2(t)\,dt$$

$$= \frac{2}{\eta}\int_0^T (2A\cos\omega_c t)^2\,dt = \frac{2}{\eta}\times 2A^2T = \frac{4A^2T}{\eta}$$

∴ from Eq. (11.48) we get

$$P_{e\min} = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{\frac{A^2T}{\eta}}\right] = \frac{1}{2}erfc\left[\sqrt{\frac{A^2T}{2\eta}}\right]$$

However,

$$\frac{A^2T}{2} = E_b = \text{bit energy}$$

∴

$$\boxed{\underset{\text{(PSK)}}{P_{e\min}} = Q\left[\sqrt{\frac{2E_b}{\eta}}\right] = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\right]} \tag{11.144}$$

11.4.7 $P_{e\min}$ for BPSK from Signal-Space Concepts

Equation (11.144), which gives the probability of error for PSK, has been derived using the results obtained earlier for optimum receivers. We shall now derive the expression for probability of error for BPSK using the signal space concepts by determining the signal constellation and identifying the decision boundaries, just the same way as we did in the case of binary FSK.

We shall assume that the two antipodal signals of the binary PSK are represented by $s_1(t)$ for binary 0 and $s_2(t)$ for binary 1 and are given by

$$\left.\begin{aligned} H_1 : b_k = 0 : s_1(t) = -\sqrt{\frac{2E_b}{T}}\cos\omega_c t;\ 0 \le t \le T \\ H_2 : b_k = 1 : s_2(t) = +\sqrt{\frac{2E_b}{T}}\cos\omega_c t;\ 0 \le t \le T \end{aligned}\right\} \tag{11.145}$$

whereas usual, E_b represents the bit energy. As in the previous case, we shall assume that the channel noise is zero-mean additive white Gaussian with a two-sided PSD of $\eta/2$. In order to have an integer number of cycles of the carrier in one time slot of duration T sec, f_c, the carrier frequency and T are so chosen that $T = (n/f_c)$, where n is an integer. Further, we assume $P_0 = P_1 = 0.5$.

From Eq. (11.145), we find that the two possible signals to be transmitted are just 180° apart. So the signal space is one–dimensional in this case. We may conveniently take, as the basis function, the following signal:

$$\phi(t) = \sqrt{\frac{2}{T}}\cos\omega_c t;\ 0 \le t \le T \tag{11.146}$$

Figure 11.31 shows the signal space, which is a straight line as it is one-dimensional, and the signal constellation comprising the two message points m_1 and m_2 used for representing the two antipodal signals $s_1(t)$ and $s_2(t)$ in that signal space.

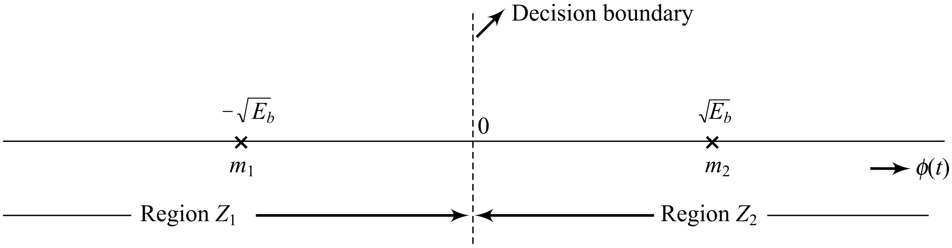


Fig. 11.31 *Signal space and signal constellation for binary PSK*

Coordinate of $s_1(t)$ along $\phi(t) = (s_1(t), \phi(t))$

$$= \int_0^T -\sqrt{\frac{2E_b}{T}} \cos \omega_c t \cdot \sqrt{\frac{2}{T}} \cos \omega_c t \, dt = -\sqrt{E_b} \tag{11.147}$$

Coordinate of $s_2(t)$ along $\phi(t) = (s_2(t), \phi(t))$

$$= \int_0^T \sqrt{\frac{2E_b}{T}} \cos \omega_c t \cdot \sqrt{\frac{2}{T}} \cos \omega_c t \, dt = \sqrt{E_b} \tag{11.148}$$

Hence, message point m_1 representing signal $s_1(t)$ is located at $-\sqrt{E_b}$ on the line representing the basis signal $\phi(t)$, and the message point m_2 representing the signal $s_2(t)$ is located at $+\sqrt{E_b}$ on the same line. Thus m_1 and m_2 are located at equal distances but on opposite sides on the line $\phi(t)$. That is why $s_1(t)$ and $s_2(t)$ are called antipodal signals. So, we now fix the boundary by drawing a line perpendicular to $\phi(t)$ line and passing through 0. The portion of the signal space, i.e., the line $\phi(t)$, which is to the left of this decision boundary constitutes the region Z_1 and the portion to the right constitutes the region Z_2.

Since the received signal $r(t)$ is given by

$$r(t) = s(t) + n_w(t) \quad ; \quad 0 \le t \le T$$

where $s(t)$ may be either $s_1(t)$ or $s_2(t)$ and $n_w(t)$ is zero-mean white Gaussian noise, $r(t)$ may not completely be within the signal space of $s_1(t)$ and $s_2(t)$. So, let us consider r, its coordinate along the basis signal $\phi(t)$. This is given by

$$r = \int_0^T r(t)\phi(t)dt \tag{11.149}$$

Now, if this observed random variable r falls in the region Z_1, we will ask the receiver to say that $s_1(t)$ has been transmitted and if it falls in the region Z_2, we will ask the receiver to say that $s_2(t)$ has been transmitted during that time slot.

If $s_1(t)$ corresponding to $b_k = 0$ has been transmitted, but due to the influence of the channel noise, the coordinate of the received signal along $\phi(t)$, viz., r falls in the region Z_2, the receiver will say that an $s_2(t)$, corresponding to $b_k = 1$ has been transmitted; and an error occurs. Let the probability of occurrence of such an error be denoted by $P(1|0)$, i.e., the probability of the receiver saying 1 even though a 0 was transmitted. Let $P(0|1)$ be the probability of the receiver saying 0 when in fact, a 1 was transmitted. From the symmetry of the signal constellation with respect to the decision boundary, however, it is clear that $P(1|0) = P(0|1)$. To determine these probabilities of the two types of errors, let us consider the conditional density functions of r conditioned on $s_1(t)$ and $s_2(t)$. These two conditional density functions, $f_{r|s_1(t)}(r|s_1(t))$ and $f_{r|s_2(t)}(r|s_2(t))$ have the same variance but have different means.

$$f_{r|s_1(t)}(r|s_1(t)) = \frac{1}{\sqrt{\eta\pi}}\exp\left[-\frac{(r+\sqrt{E_b})^2}{\eta}\right] \tag{11.150}$$

and

$$f_{r|s_2(t)}(r|s_2(t)) = \frac{1}{\sqrt{\eta\pi}}\exp\left[-\frac{(r-\sqrt{E_b})^2}{\eta}\right] \tag{11.151}$$

Note

It is left to the reader as an exercise, to show that $E[r|s_1(t)] = -\sqrt{E_b}$, $E[r|s_2(t)] = +\sqrt{E_b}$ *and that*

$$Var[r|s_1(t)] = Var[r|s_2(t)] = \eta/2$$

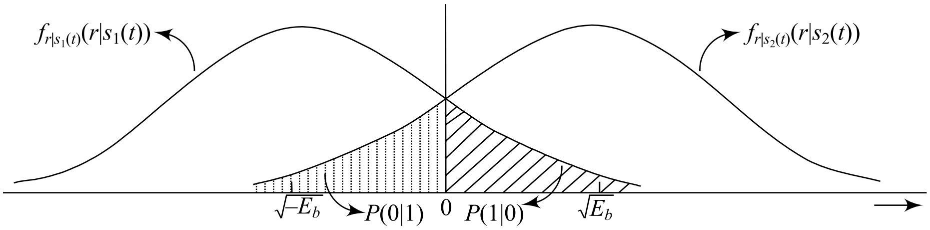


Fig. 11.32 *Conditional densities of the observed random variable*

From Fig. 11.32, we have

$$P(1|0) = \int_0^\infty f_{r|s_1(t)}(r|s_1(t))\,dr = \frac{1}{\sqrt{\eta\pi}}\int_0^\infty e^{-\frac{(r+\sqrt{E_b})^2}{\eta}}\,dr \tag{11.152}$$

Putting $\frac{1}{\sqrt{\eta}}(r+\sqrt{E_b})\,\underline{\Delta}\,z$, we get

$$P(1\,|\,0) = \frac{1}{\sqrt{\pi}}\int_{\sqrt{E_b/\eta}}^{\infty} e^{-z^2}\,dz = \frac{1}{2}\text{erfc}\left[\sqrt{\frac{E_b}{\eta}}\right] \tag{11.153}$$

Also,

$$\begin{aligned} P_e &= P_0 P(1\,|\,0) + P_1 P(0\,|\,1) \\ &= 0.5P(1\,|\,0) + 0.5P(0\,|\,1) \end{aligned} \tag{11.154}$$

However, from the symmetry of the conditional density functions (refer to Fig. 11.31), we find that

$$P(1\,|\,0) = P(0\,|\,1) \tag{11.155}$$

$\therefore$ from Eqs. (11.154) and (11.155), we have

$$P_e = P(1\,|\,0)$$

$\therefore$ from Eq. (11.153)

$$\boxed{\underset{\text{Binary PSK}}{P_e} = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\right]} \tag{11.156}$$

11.4.8 Comparison of the Performance of Binary ASK, FSK and PSK

1. ASK and FSK lend themselves to coherent as well as non-coherent detection; but PSK lends itself only to coherent reception. There is no non-coherent PSK system, although a sub-optimal PSK scheme called DPSK, or differential PSK can be detected non-coherently.
2. Since FSK and PSK signals have a constant envelope, they are immune to amplitude non-linearities which arise in microwave and radio channels. Hence, FSK and PSK signals are preferred to ASK in band pass data transmission over these channels.
3. For a given bit energy, PSK gives the lowest probability of error. This can be guessed from the fact that the PSK signals, being antipodal, have maximum separation for a given bit energy. Larger signal separation means less likelihood of the noise making the receiver commit a mistake in identifying which signal has been transmitted.

Table 11.1 *Comparison of binary ASK, FSK and PSK*

Binary digital modulation	Type of detection	Bit rate/bandwidth	Probability of error
Binary ASK/OOK	Non-coherent	1	$\frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]^*$
	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]^*$
BFSK	Non-coherent	1	$\frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$
	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$
BPSK	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

* $E_b = E_{av} = A^2T/4$.

Note *(bit rate/B_T) will be equal to 0.5 if B_T is taken as the full main lobe width in the power spectrum of the modulated signal.*

4. From the above table, it is evident that PSK, coherent or non-coherent ASK and FSK, all have the same bandwidth efficiency, but their bit-error rates are different.
5. The bit-error rates of the various basic binary digital modulation schemes are plotted against (E_b/η) in Fig. 11.32. From the figure, it is clear that
   (a) Coherent ASK and coherent FSK have the same P_e.
   (b) From probability of error point of view, starting from the best, viz., the PSK, they may be ranked in the following order:
      (i) PSK
      (ii) Coherent FSK and coherent ASK
      (iii) Non-coherent FSK
      (iv) Non-coherent ASK

6. Non-coherent ASK and non-coherent FSK give essentially the same performance. However, in practice, FSK is preferred over ASK because it has a fixed optimum threshold while ASK has a threshold that depends on the signal-to-noise ratio and is therefore susceptible to signal fading (refer to Eq. (11.88)).
7. Comparison of P_e of BPSK with that of coherent ASK or coherent FSK reveals that for achieving the same probability of error, ASK requires twice the pulse energy as compared to PSK, i.e., ASK requires 3 dB more power than PSK. Hence, if coherent detection is to be used, ASK and FSK are not considered.

However, for non-coherent reception, FSK is useful because of its simplicity and its superior performance over ASK for reasons stated in point 6 above.

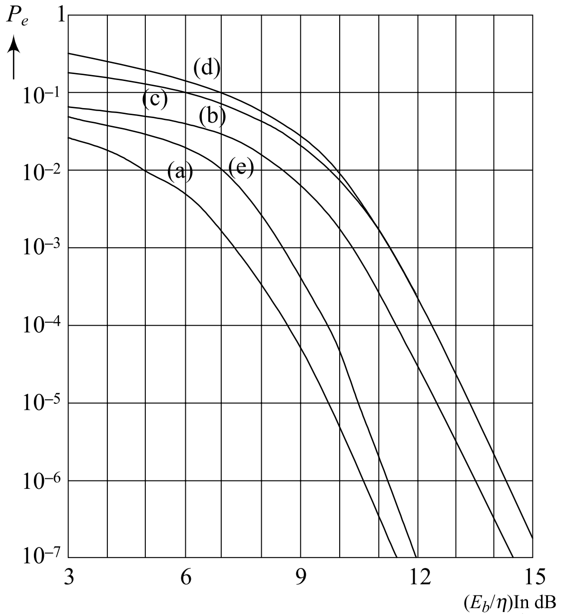


Fig. 11.33 *P_e vs. (E_b/η) for: (a) Binary PSK, (b) Coherent BFSK and ASK, (c) Non-coherent FSK, (d) Non-coherent ASK, (e) DPSK*

Example 11.20 (a) Draw the signal constellation for a binary phase-shift keyed signal set if the two signals are

$$s_1(t) = \sqrt{\frac{2E_b}{T}} \cos(\omega_c t - \pi/4)\,;\, 0 \le t \le T$$

and $$s_2(t) = \sqrt{\frac{2E_b}{T}} \cos(\omega_c t + \pi/4)\,;\, 0 \le t \le T$$

(b) Also determine and sketch the low pass equivalent spectrum of these PSK signals.

Solution

(a) $$s_1(t) = \sqrt{\frac{2E_b}{T}} \cos(\omega_c t - \pi/4)$$

$$= A\left[\cos\omega_c t \cdot \cos\frac{\pi}{4} + \sin\omega_c t \cdot \sin\frac{\pi}{4}\right] \text{ where } A = \sqrt{\frac{2E_b}{T}}$$

$$= \frac{A}{\sqrt{2}}[\cos\omega_c t + \sin\omega_c t]$$

In a similar way $$s_2(t) = \frac{A}{\sqrt{2}}[\cos\omega_c t - \sin\omega_c t]$$

Let $\phi_1 = \cos\omega_c t$ and $\phi_2 = \sin\omega_c t$. These two orthogonal signals can be used as the basis signal set for the signal space to which $s_1(t)$ and $s_2(t)$ belong.

$$\therefore \qquad s_1(t) = \frac{A}{\sqrt{2}}\phi_1(t) + \frac{A}{\sqrt{2}}\phi_2(t)$$

and $s_2(t) = \frac{A}{\sqrt{2}}\phi_1(t) - \frac{A}{\sqrt{2}}\phi_2(t)$

$$\therefore \quad s_1(t) = \left(\frac{A}{\sqrt{2}}, \frac{A}{\sqrt{2}}\right) \quad \text{and} \quad s_2(t) = \left(\frac{A}{\sqrt{2}}, -\frac{A}{\sqrt{2}}\right)$$

$s_1(t)$ is represented by point m_1 and $s_2(t)$ is represented by point m_2, in the signal space spanned by ϕ_1 and ϕ_2.

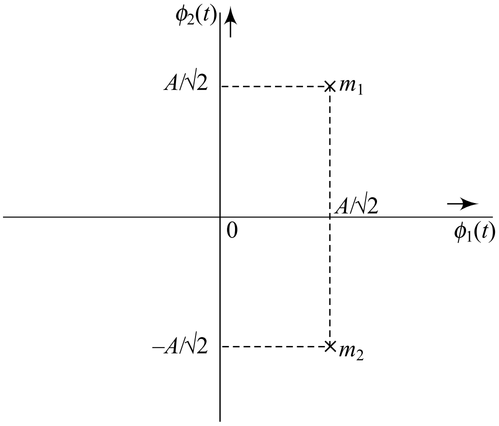


Fig. 11.34 *Signal constellation*

(b) The low pass equivalent of $s_1(t)$ or $s_2(t)$ is given by

$$p(t) = \sqrt{\frac{2E_b}{T}}\Pi(t/T)$$

$$P(f) = F[p(t)] = \sqrt{\frac{2E_b}{T}}T \text{ sinc}(fT)$$

$$\therefore \text{ power spectral density } = \frac{|P(f)|^2}{T} = 2E_b\text{sinc}^2(fT)$$

This PSD has already been sketched in Fig. 10.31 in Chapter 10.

Example 11.21 A microwave link is used for transmitting binary data at the rate of 1 Mbps. Assuming the PSD (two sided) of the noise at the input of the receiver to be 10^{-10} *W*/Hz, find the transmission bandwidth and the average carrier power required to be maintained if the probability of error, P_e is not to exceed 10^{-4}, when (a) coherent BPSK, and (b) coherent BFSK are used.

Solution

(a) *For coherent BPSK*

$$P_e = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\right] = 10^{-4}$$

$$\therefore \qquad erfc\left[\sqrt{\frac{E_b}{\eta}}\right] = 2\times 10^{-4} = 2Q\left[\sqrt{\frac{2E_b}{\eta}}\right]$$

From the Q-function graph, we find that $\left[\sqrt{\frac{2E_b}{\eta}}\right] = 3.7$

$$\sqrt{\frac{E_b}{\eta}} = \frac{3.7}{\sqrt{2}} \quad \text{and} \quad E_b = 2\times 10^{-10}\times\frac{(3.7)^2}{2} = 13.69\times 10^{-10}\text{W-sec.}$$

Since bit rate (= $1/T$) is given to be 10^6 Mbps

$$\text{Average power} = 13.69\times 10^{-10}\times \text{bit rate}$$
$$= 13.69\times 10^{-10}\times 10^6 = 1.369 \text{ mW}$$

(b) *Coherent BFSK*

$$P_e = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right] = 10^{-4}$$

Proceeding as in the previous case, from the Q-function graph,

$$\sqrt{\frac{E_b}{2\eta}} = \frac{3.7}{\sqrt{2}} \quad \therefore\ E_b = 2 \times 13.69 \times 10^{-10}\,\text{W-sec.}$$

$\therefore\ P_{\text{av}}$ = Average power = $E_b \times$ bit rate = $2 \times 13.69 \times 10^{-10} \times 10^6$

$\therefore\ P_{\text{av}} = 2.738$ mW

In both the cases, the transmission bandwidth $B_T = 1$ MHz.

Example 11.22 In a BPSK system, the correlator in the receiver, to which the received PSK signal is applied, supplied with a carrier signal whose phase is θ radians away from the exact carrier phase. Determine the effect of this phase error θ on the probability of error of the system.

Solution For convenience, the structure of a correlation receiver for BPSK signals is shown in Fig. 11.35 with the local carrier θ radians away from the correct phase.

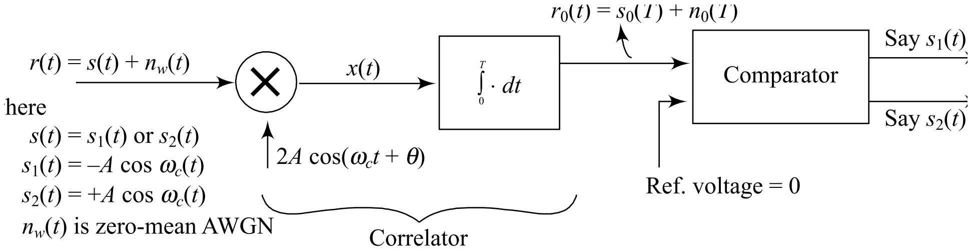


Fig. 11.35 *Correlation receiver for BPSK signals in which the local carrier is θ radians away from correct phase*

$$x(t) = [s(t) + n_W(t)]\,[2A \cos(\omega_c t + \theta)]$$
$$= 2As(t)\cos(\omega_c t + \theta) + 2An_W(t)\cos(\omega_c t + \theta)$$

$\therefore$ if
$$s_0(T) = \begin{cases} s_{01}(T) \text{ when } s(t) = s_1(t) \text{ and} \\ s_{02}(T) \text{ when } s(t) = s_2(t), \end{cases}$$

then

$$s_{01}(T) = \int_0^T [x(t)|s_1(t)]\,dt = -2A^2 \int_0^T \cos \omega_c t \cdot \cos(\omega_c t + \theta)\,dt$$
$$= \frac{-1}{2} 2A^2 \int_0^T \{\cos(2\omega_c t + \theta) + \cos\theta\}\,dt$$

But $\int_0^T \cos(2\omega_c t + \theta)\,dt = 0$. Since the carrier cosine function is arranged to have an integer number of full cycles in the period 0 to T sec.

$\therefore$
$$s_{01}(T) = -A^2 \int_0^T \cos\theta\, dt = -A^2 T \cos\theta$$

Similarly,
$$s_{02}(T) = A^2 T \cos\theta$$

$n_0(T)$ = noise component of $r_0(T)$

$$= \int_0^T 2An_W(t)\cos(\omega_c t + \theta)\,dt$$

$\therefore$ $n_0(T)$ is independent of which signal has been transmitted.
Now,

$$E[n_0(T)] = \int_0^T 2AE[n_W(t)]\cos(\omega_c t + \theta)\,dt = 0$$

Since $$E[n_W(t)] = 0$$

$\therefore$ $$\sigma^2_{n_0(T)} = E[n_0^2(T)] - \{E[n_0(T)]\}^2 = E[n_0^2(T)]$$

$$= 4A^2 \int_0^T\int_0^T E\{n_W(\alpha)n_W(\beta)\}\cos(\omega_c\alpha + \theta)\cos(\omega_c\beta + \theta)\,d\alpha d\beta$$

Now,

$$E\{n_W(\alpha)n_W(\beta)\} = \begin{cases} \eta/2, & \text{if } \beta = \alpha \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$ When $\beta = \alpha$,

$$\sigma^2_{n_0(T)} = 4A^2\int_0^T \left(\frac{\eta}{2}\right)\cos^2(\omega_c\alpha + \theta)\,d\alpha = \frac{2A^2\eta}{2}\int_0^T \left[1 + \cos 2(\omega_c\alpha + \theta)\right]d\alpha$$

$\therefore$ $$\sigma^2_{n_0(T)} = \eta A^2 T \tag{11.157}$$

Since $$r_0(T) = s_0(T) + n_0(T),$$

where $n_0(T)$ is a zero-mean Gaussian random variable, $r_0(T)$ is also a Gaussian random variable with variance $\sigma^2_{n_0(T)}$ and a mean that is equal to $-A^2T\cos\theta$ when $s_1(t)$ is transmitted and $A^2T\cos\theta$ when $s_2(t)$ is transmitted. The conditional densities of $r_0(T)$ may be written as

$$f_{r_0(T)}(r_0(T)|s_1(t)) = \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}}\exp\left[-\frac{(r_0(T) + A^2T\cos\theta)^2}{2\sigma^2_{n_0(T)}}\right]$$

and $$f_{r_0(T)}(r_0(T)|s_2(t)) = \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}}\exp\left[-\frac{(r_0(T) - A^2T\cos\theta)^2}{2\sigma^2_{n_0(T)}}\right]$$

Assuming the *a priori* probabilities of transmission of $s_1(t)$ and $s_2(t)$ to be equal, i.e., $P_0 = P_1 = 0.5$, the optimum threshold or reference voltage is zero, as shown in Fig. 11.36.

$$P_e' = 0.5P(0\,|\,1) + 0.5P(1\,|\,0) = P(1\,|\,0) \text{ since } P(0\,|\,1) = P(1\,|\,0) \text{ (From the symmetry)}$$

Now,

$$P_e' = P(1\,|\,0) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2_{n_0(T)}}}\exp\left[-\frac{(r_0(T) + A^2T\cos\theta)^2}{2\sigma^2_{n_0(T)}}\right]dr_0(T)$$

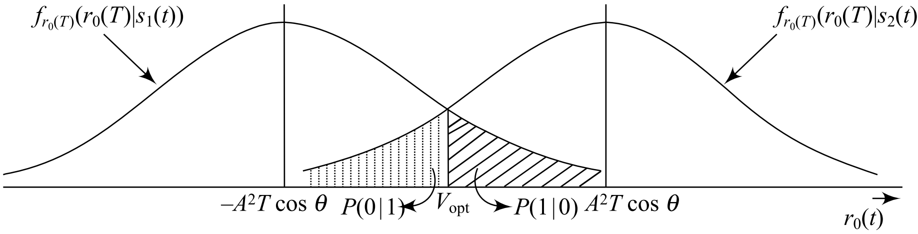


Fig. 11.36 *Conditional densities of the observed random variable $r_0(T)$*

Put $$z = \left[\frac{r_0(T) + A^2T\cos\theta}{\sigma_{n_0(T)}}\right] \qquad \therefore \text{ when } r_0(T) = 0, z = \frac{A^2T\cos\theta}{\sigma_{n_0(T)}}$$

and $$dz = \left(\frac{1}{\sigma_{n_0(T)}}\right) dr_0(T)$$

$\therefore$ substituting the above in the expression for P_e', we get

$$P_e' = \int_{\frac{A^2T\cos\theta}{\sigma_{n_0(T)}}}^{\infty} \frac{\sigma_{n_0(T)}}{\sqrt{2\pi\sigma^2_{n_0(T)}}} e^{-z^2/2} dz = Q\left[\frac{A^2T\cos\theta}{\sigma_{n_0(T)}}\right] = Q\left[\frac{A^2T\cos\theta}{A\sqrt{\eta T}}\right] \qquad \text{(From Eq. (11.156))}$$

$$= Q\left[A\sqrt{\frac{T}{\eta}}\cos\theta\right] = Q\left[\sqrt{\frac{A^2T}{\eta}}\cos\theta\right] = Q\left[\sqrt{\frac{2E_b}{\eta}}\cos\theta\right]$$

Since $$E_b = \left(\frac{A^2}{2}\cdot T\right)$$

$\therefore$ $$P_e' = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\cos\theta\right]$$

Without the phase error for the local carrier, ideally

$$P_e = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$$

While with a phase offset of θ radians, it is

$$\boxed{\underset{\substack{\text{(BPSK)}\\ \text{Phase offset } \theta}}{P_e'} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\cos\theta\right]} \qquad (11.158)$$

Since complementary error function is a monotonically decreasing function of its argument, and since cos $\theta \le 1$, for small but non-zero value of θ, $P_e' > P_e$. Hence the effect of phase offset is to increase the probability of error.

Example 11.23 If the frequency offset f_d in a binary FSK system satisfies the relation $W_dT = n\pi$ show that $s_1(t)$ and $s_2(t)$ are orthogonal.

Solution $W_dT = n\pi \quad \therefore 2\pi f_dT = n\pi$ or $f_d = \frac{n}{2T}$

$$\therefore \qquad s_1(t) = A\cos\left(\omega_c - \frac{\pi n}{T}\right)t;\ 0 \le t \le T$$

and
$$s_2(t) = A\cos\left(\omega_c + \frac{n\pi}{T}\right);\ 0 \le t \le T$$

To show that they are orthogonal over the interval [0, T], we have to show that their inner product, defined as

$$(s_1(t), s_2(t)) \underline{\Delta} \int_0^T s_1(t)s_2(t)\,dt = 0$$

Substituting for $s_1(t)$ and $s_2(t)$ in the above inner product, we get

$$(s_1(t), s_2(t)) = A^2\int_0^T \cos\left(\omega_c - \frac{n\pi}{T}\right)t \cdot \cos\left(\omega_c + \frac{n\pi}{T}\right)t\,dt$$

$$= \frac{A^2}{2}\int_0^T \cos 2\omega_c t\,dt + \frac{A^2}{2}\int_0^T \cos\left(\frac{2n\pi}{T}\right)t\,dt$$

In the above, the first integral is zero since f_c is so chosen that there will be an integer number of full cycles of the carrier wave in one T.

$$\therefore \qquad (s_1(t), s_2(t)) = \frac{A^2}{2}\int_0^T \cos\left(\frac{n2\pi}{T}\right)t \cdot dt = \frac{A^2}{2}\frac{T}{2n\pi}\sin\left(\frac{2n\pi}{T}\right)t\Big|_{t=0}^{T} = 0$$

Since their inner product is zero, the two signals $s_1(t)$ and $s_2(t)$ are orthogonal to each other.

Example 11.24 We have derived an expression for P_e applicable to continuous phase binary FSK. Derive an expression for discontinuous type binary FSK.

Solution $s_1(t) = A\cos(\omega_c - \omega_d)t$

and $s_2(t) = A\cos(\omega_c + \omega_d)t$

The correlation receiver structure is as shown in Fig. 11.37.

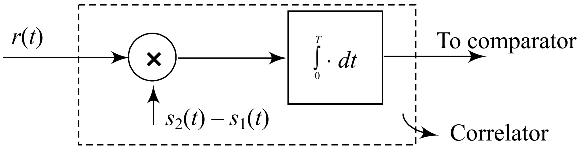


Fig. 11.37 *Correlation receiver of Example 11.24*

$s_{01}(T)$ = output of the correlator at $t = T$ when $s_1(t)$ alone is given as input to it.

$$= \int_0^T s_1(t)[s_2(t) - s_1(t)]\,dt$$

$$= \int_0^T [A\cos(\omega_c - \omega_d)t][A\cos(\omega_c + \omega_d)t]\,dt - A^2\int_0^T \cos^2(\omega_c - \omega_d)t\,dt$$

$$= \frac{A^2}{2}\int_0^T (\cos 2\omega_c t + \cos 2\omega_d t)\,dt - \frac{A^2}{2}\int_0^T [1 + \cos 2(\omega_c - \omega_d)t]\,dt$$

If $T >> \frac{1}{2(f_c - f_d)}$,

$$s_{01}(T) = \frac{A^2}{2}\left[\frac{1}{2\omega_d}\sin 2\omega_d t\right]\Bigg|_0^T - \frac{A^2T}{2} = -\frac{A^2T}{2} + \frac{A^2T}{2}\left(\frac{\sin 2\omega_d T}{2\omega_d T}\right)$$

Similarly,

$s_{02}(T)$ = correlator output when $s_2(t)$ alone is given as input

$$= \frac{A^2T}{2} + \frac{A^2T}{2}\left(\frac{\sin 2\omega_d T}{2\omega_d T}\right)$$

$$\therefore \qquad V_{\text{opt}} = \text{Threshold} = \frac{s_{01}(T) + s_{02}(T)}{2} = A^2T\left(\frac{\sin 2\omega_d T}{2\omega_d T}\right)$$

$$d^2_{\max} = \frac{2}{\eta}\int_0^T [s_2(t) - s_1(t)]^2 dt \qquad \text{(see Eq.11.52)}$$

$$= \frac{2A^2}{\eta}\int_0^T [\cos(\omega_c + \omega_d)t - \cos(\omega_c - \omega_d)t]^2 dt$$

$$= \frac{2A^2}{\eta}\left[\int_0^T \cos^2(\omega_c + \omega_d)t\,dt + \int_0^T \cos^2(\omega_c - \omega_d)t\,dt - 2\int_0^T \cos(\omega_c + \omega_d)t \cdot \cos(\omega_c - \omega_d)t\,dt\right]$$

$$= \frac{2A^2}{\eta}\left[T - \frac{\sin 2\omega_d T}{2\omega_d T}\right] = \frac{2A^2T}{\eta}\left[1 - \frac{\sin 2\omega_d T}{2\omega_d T}\right]$$

This quantity, $d^2_{\max}$ attains its largest value when $\omega_d T \cong \frac{3\pi}{4}$, i.e., when

$$\omega_d \approx \frac{3\pi}{4T}$$

Therefore, if the offset frequency, f_d, is so chosen that $\omega_d = \frac{3\pi}{4T}$, then $d^2_{\max}$ will be given by

$$d^2_{\max} = \frac{2A^2T}{\eta}\left[1 - \left(\frac{\sin 2\omega_d T}{2\omega_d T}\right)\Bigg|_{\omega_d = 3\pi/4T}\right]$$

$$= \frac{2A^2T}{\eta}[1 + 0.21] = \frac{2.42A^2T}{\eta}$$

But

$$P_{e\min} = Q\left[\frac{d_{\max}}{2}\right] = Q\left[\sqrt{0.605\frac{A^2T}{\eta}}\right]$$

$$= \frac{1}{2}erfc\left[\sqrt{0.6\frac{A^2T}{2\eta}}\right]$$

But
$$\frac{A^2T}{2} = \text{bit energy} = E_b$$

∴
$$\boxed{\underset{\substack{\text{BFSK-coherent}\\ \text{(discontinuous phase)}}}{P_{e\,\min}} = \frac{1}{2}erfc\left[\sqrt{0.6\frac{E_b}{\eta}}\right]} \quad (11.159)$$

Example 11.25 A binary band pass system transmits binary data at the rate of 2.5 × 10^6 bits/sec. During the course of transmission, zero-mean AWGN of two-sided PSD equal to 10^{-14} *W*/Hz is added to the signal. In the absence of noise, the amplitude of the received sinusoidal wave for digit 1 or 0 is 1 mV. Find the average probability of symbol error, for the following systems: (a) Coherent BFSK, (b) Non-coherent BFSK, and (c) BPSK (coherent).

Solution

(a) *Coherent BFSK*

For coherent BFSK, $$P_e = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$$

where $$E_b = \frac{A^2T}{2}$$

Now, $$T = \frac{1}{r_b} = \frac{1}{2.5 \times 10^6} = 0.4 \times 10^{-6}\text{ sec}$$

$A = 1$ mV $$\therefore\ E_b = \frac{A^2T}{2} = \frac{1 \times 10^{-6} \times 0.4 \times 10^{-6}}{2} = 0.2 \times 10^{-12}$$

$$\sqrt{\frac{E_b}{2\eta}} = \sqrt{\frac{0.2 \times 10^{-12}}{2 \times 2 \times 10^{-14}}} = 2.236$$

From error-function tables,

$$P_e = \frac{1}{2}\text{erfc}\,2.236 = 1.27 \times 10^{-3}$$

(b) *Non-coherent BFSK*

$$P_e = \frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$$

∴ $$P_e = \frac{1}{2}\exp\left[-\frac{0.2 \times 10^{-12}}{2 \times 2 \times 10^{-14}}\right] = \frac{1}{2}e^{-5} = 0.003369 = 3.369 \times 10^{-3}$$

(c) *BPSK*

$$P_e = \frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}} = \frac{1}{2}erfc\sqrt{\frac{0.2 \times 10^{-12}}{2 \times 10^{-14}}}$$

$$= \frac{1}{2}erfc\sqrt{10} = \frac{1}{2}erfc\,(3.1622) = 0.0000055$$

∴ $$P_e = 5.5 \times 10^{-6}$$

11.4.9 Probability of Error for DPSK

In Section 10.6.1 of Chapter 10, we had discussed the basic principle of operation and the methods of generation and reception of DPSK signals. We shall now derive an expression for the probability of error of a DPSK system.

As can be seen from Table 10.4, differential encoding basically consists of the encoder giving an output of 1 if the present and the preceding message bits are alike and giving an output of 0 in case they are not alike. A similar thing is done in the decoding of the received signals at the receiver. Whenever the signal in the present time slot and the preceding one have the same phase (either both 0 or π), we decode the present bit as 1 and if they do not have the same phase, we decode the present bit as 0.

Let $s_1(t)$ be used to represent the signal transmitted ***over two consecutive time slots when they both have 0 phase***.

$\therefore$

$$s_1(t) \underline{\Delta} \begin{cases} A\cos 2\pi f_c t; & 0 \le t \le T \\ A\cos 2\pi f_c t; & T \le t \le 2T \end{cases} \tag{11.160}$$

Similarly, let

$\therefore$

$$s_2(t) \underline{\Delta} \begin{cases} A\cos 2\pi f_c t; & 0 \le t \le T \\ A\cos(2\pi f_c t + \pi); & T \le t \le 2T \end{cases} \tag{11.161}$$

It is easy to show that the inner product of $s_1(t)$ and $s_2(t)$ is equal to zero.

i.e.,

$$(s_1(t), s_2(t)) \underline{\Delta} \int_0^{2T} s_1(t)s_2(t)\,dt = 0 \tag{11.162}$$

Hence, $s_1(t)$ and $s_2(t)$ are orthogonal over the internal $0 \le t \le 2T$. ***Thus a DSPK signal may be regarded as a special case of non-coherent orthogonal modulation, but with a time-slot of 2T instead of T***. Hence, if $s_1(t)$ and $s_2(t)$ are normalized so that

$$A = \sqrt{\frac{E_b}{2T}}$$

Then the energy of $s_1(t)$ or $s_2(t)$ over the interval 0 to $2T$ is equal to $2E_b$.

We have already discussed non-coherent binary FSK , where the signals were of continuous phase (Sunde's Continuous phase FSK) type and we had shown that the transmitted signals corresponding to a message bit 0 and a message bit 1 are orthogonal. ***So non-coherent binary FSK of Sunde's type, is also a case of non–coherent binary orthogonal modulation***. The probability of error was derived for it and it was shown to be

$$\underset{\substack{\text{(Sunde's non-coherent} \\ \text{binary FSK)}}}{P_e} = \frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$$

DPSK, as a special case of binary non-coherent orthogonal modulation (Orthogonality of the signals being over a period 0 to $2T$ instead of 0 to T and the energy being $2E_b$ over that period), will have a probability of error given by

$$\boxed{\underset{\text{(DPSK)}}{P_e} = \frac{1}{2}\exp\left[-\frac{E_b}{\eta}\right]} \tag{11.163}$$

A plot of P_e vs. (E_b/η) for DPSK is shown in Fig.11.33(e).

For BPSK, we had (refer to Eq. (11.156))

$$\underset{\text{(BPSK)}}{P_e} = \frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}}$$

But we know that for large values of the argument x, erfc x can be approximated by

$$\underset{(\text{for large } x)}{erfc\, x} \approx [e^{-x^2}]/(\sqrt{\pi}\, x) \tag{11.164}$$

Hence, for large values of SNR, P_e for BPSK can be written as

$$\underset{(\text{BPSK; large SNR})}{P_e} \approx \frac{\sqrt{\eta}}{2} \frac{e^{-E_b/\eta}}{\sqrt{\pi}\sqrt{E_b}} \tag{11.165}$$

For large *SNR*s, i.e., for $P_e \le 10^{-4}$, the difference in value between the *SNR* for DPSK and *SNR* for BPSK, is only of the order of 1 dB, that of DPSK being lower. Thus, DPSK can be viewed as an attractive alternative to PSK. However, it suffers from the following two disadvantages:

1. Since DPSK needs one–bit delay units in its transmitter as well as receiver, the transmission rate for the system is fixed.
2. Since the previous decoded bit is used for decoding the present bit, errors tend to occur in pairs.

Remark *The bit energy for non-coherent FSK (CP type) is E_b over a period T. In the case of DPSK, it is $2E_b$ but over a period of 2T. So bit energy over slot time is same. Hence, comparison of Eqs. (11.138) and (11.163) reveals that* ***DPSK is 3dB superior to non-coherent FSK****.*

Example 11.26 A binary DPSK system is to have an average probability of error, $P_e \le 10^{-4}$. If the average transmitted power is 150 mW, the channel attenuation is 80 dB and the additive zero-mean white Gaussian noise on the channel is having a two-sided PSD of $\eta/2 = 0.5(10^{-15})$ *W*/Hz, find the maximum allowable bit rate for transmission.

Solution $$\underset{(\text{DPSK})}{P_e} = \frac{1}{2} e^{-E_b/\eta} \le 10^{-4}$$

$$\therefore \quad e^{E_b/\eta} \ge \frac{1}{2} \times 10^{-4} = 5000. \qquad \text{Thus, } \frac{E_b}{\eta} \ge 8.517$$

$$\therefore \quad E_b \ge 8.517 \times 10^{-15} \text{ W-sec}$$

Now, average power of the received signal $= E_b \cdot R_b$

where R_b is the bit rate of transmission of the data.

$\therefore$ average received power = $S_R = 8.517 \times 10^{-15} \times R_b$

Since attenuation in the channel = 80 dB = 10^8 (ratio)

$$\frac{S_T}{S_R} = 10^8 = \frac{\text{Average transmitted power}}{\text{Average received power}} = \frac{150 \times 10^{-3}}{8.517 \times 10^{-15} \times R_b}$$

$$\therefore \quad R_b = \frac{150 \times 10^{-3} \times 10^{15}}{8.517 \times 10^8} = \frac{150 \times 10^4}{8.517} = 17.61 \times 10^4 \text{ bits/sec}$$

or $$R_b = \text{bit-rate} = 17.61 \times 10^4 \text{ bits/sec}$$

Example 11.27 A binary transmission system with phase modulation is to have $P_e \le 10^{-4}$. If it is a BPSK system with a phase offset of θ, find the maximum value of θ which will still make the BPSK to require less signal energy than DPSK.

Solution P_e for BPSK with phase offset θ is $\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\cos\theta\right]$ (Refer to Example 11.22)

$$\underset{\text{(DPSK)}}{P_e} = \frac{1}{2}e^{-E_b/\eta} \qquad \text{(Refer to Eq. (11.163))}$$

$$\therefore \qquad \frac{1}{2}e^{-E_b/\eta} = \frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\cos\theta\right] = 10^4$$

$$\frac{1}{2}e^{-E_b/\eta} = 10^{-4}\text{ gives } E_b/\eta = \log_e 5000 = 8.517$$

$$\therefore \qquad \sqrt{\frac{E_b}{\eta}} = \sqrt{8.517} = 2.918 \quad \text{let} \quad \sqrt{\frac{E_b}{\eta}}\cdot\cos\theta \,\underline{\Delta}\, y$$

Now, $\frac{1}{2}erfc\ y = 10^{-4} \Rightarrow 2Q[\sqrt{2}y] = 2\times 10^{-4}$

$\therefore$ from the Q-function graph, we get

$$y = 2.65$$

Since $\therefore$ $\sqrt{\frac{E_b}{\eta}} = 2.918$ and $\sqrt{\frac{E_b}{\eta}}\cdot\cos\theta = y = 2.65,$

$$\cos\theta = \frac{2.65}{2.918} = 0.908 \qquad \therefore\ \theta = \cos^{-1}0.908$$

Thus, $\theta = 24.76°$

This is the maximum value of the phase offset in the BPSK system which will still make BPSK to require less signal energy than what is required for DPSK while keeping $P_e \le 10^{-4}$.

11.5 PROBABILITY OF ERROR FOR QPSK

The basic principle, signal space and signal constellation as well as the methods of generation and detection of QPSK signals were all discussed in detail in Section 10.6.3. As mentioned there, in QPSK, any one of the four possible signals which have four equally spaced values $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$ for the carrier phase, is transmitted during each symbol period of τ seconds.

Let $r(t)$ be the received signal. This is one of the four QPSK signals, $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$, of course, corrupted by zero-mean additive white Gaussian noise having a two-sided power spectral density of $\eta/2$. The task before the receiver is to correctly identify which one it is

$$r(t) = s_i(t) + n_w(t) \quad ; \quad 0 \le t \le \tau \quad \text{and} \quad i = 1, 2, 3, 4 \qquad (11.166)$$

Referring to Fig. 10.40, because of the noise component, $r(t)$ will not lie entirely in the signal space spanned by $\phi_1(t)$ and $\phi_2(t)$. So, we would like to know in which quadrant of our signal space the orthogonal projection of $r(t)$ falls. We ask the receiver to say that the signal transmitted during that symbol time was $s_1(t)$ if the projection of $r(t)$ falls in zone-1, or that it was $s_2(t)$ if the projection of $r(t)$ falls in zone-2, and so on.

So, to determine the zone in which the orthogonal projection of $r(t)$ falls, we determine the coordinates of $r(t)$ along $\phi_1(t)$ and $\phi_2(t)$, the basis signals of our signal space.

$$r_{01}(\tau) = \text{coordinate of } r(t) \text{ along } \phi_1(t) = (r(t), \phi_1(t)) = \int_0^\tau r(t)\phi_1(t)\,dt$$

$$= \int_0^\tau [s_i(t) + n_w(t)]\phi_1(t)\,dt$$

$$\sqrt{E}\cos\left[(2i-1)\frac{\pi}{4}\right]+w_1$$

$$=\pm\sqrt{\frac{E}{2}}+w_1 \tag{11.167}$$

Similarly coordinates of $r(t)$ along $\phi_2(t)$ $=(r(t),\phi_2(t))=r_{02}(\tau)$

$$=\int_0^\tau r(t)\phi_2(t)\,dt=\int_0^\tau [s_i(t)+n_w(t)]\phi_2(t)\,dt$$

$$=\mp\sqrt{\frac{E}{2}}+w_2 \tag{11.168}$$

In Eqs. (11.167) and (11.168), w_1 and w_2 are zero-mean orthogonal Gaussian random variables and so they are uncorrelated and independent. The mean values of random variables $r_{01}(\tau)$ and $r_{02}(\tau)$ are therefore $\pm\sqrt{\frac{E}{2}}$ and $\mp\sqrt{\frac{E}{2}}$ respectively and they have the same variance, viz. $\eta/2$. The ± and ∓ signs used in the two equations should be interpreted according to the coordinate values given for each $s_i(t)$ along $\phi_1(t)$ and $\phi_2(t)$ respectively in the last column of Table 10.6.

Since the receiver is asked to say $s_i(t)$ if the projection of $r(t)$ falls in zone-1, an error will occur if the transmitted signal is $s_i(t)$ with i = say 1, but the projection of $r(t)$ with coordinates as given by Eqs. (11.167) and (11.168) falls in some other zone, other than zone-1 (where it would have fallen but for the noise).

A look at the QPSK signal constellation given in Fig. 10.40 and a comparison of it with the BPSK signal constellation given in Fig. 11.31 suggests that the QPSK system can be considered to be equivalent to two BPSK systems operating in parallel and having carrier signals which are of the same frequency but inphase quadrature. Just as $\sqrt{E_b}$ and $-\sqrt{E_b}$, where E_b is the bit energy of the BPSK system (of Fig. 11.31) were the coordinates of the two antipodal signals $s_2(t)$ and $s_1(t)$ of that system along the carrier of that system, now, in Fig.10.40, the $\sqrt{E_b/2}$ and $-\sqrt{E_b/2}$ along $\phi_1(t)$ axis, can be considered to be the coordinates along $\phi_1(t)$ of the antipodal BPSK signal pair generated with the inphase carrier. Similarly, the $\sqrt{E_b/2}$ and $-\sqrt{E_b/2}$ along $\phi_2(t)$ in Fig. 10.40 may be regarded as the coordinates along $\phi_2(t)$ of the antipodal BPSK signal pair generated with the quadrature carrier.

Thus, with reference to the QPSK signals represented uniquely by the four dibits, the first digit in each dibit is related to $\phi_1(t)$ and hence the inphase carrier BPSK system, while the second digit in each dibit is related to $\phi_2(t)$, and hence the quadrature carrier BPSK system (see Fig. 10.40).

From the foregoing, it is clear that the inphase carrier BPSK system as well as the quadrature carrier BPSK system have (E/2) as the signal energy *per bit*. The noise PSD, of course, is the same for each of these BPSK systems as it is for the QPSK system and it is $\eta/2$. Hence, using the expression for P_e for a BPSK system (Eq. (11.156)), the average probability of bit error in each of the two BPSK systems (the inphase and quadrature channels of the QPSK system) is

$$P'=\frac{1}{2}erfc\left[\sqrt{\frac{E/2}{\eta}}\right]=\frac{1}{2}erfc\left[\frac{E}{2\eta}\right] \tag{11.169}$$

Since the first digit in any dibit pertains to the inphase channel and the second to the quadrature channel, decoding errors in the two digits are independent. A dibit is therefore correctly decoded only when both the inphase and quadrature channels of the receiver decode their respective digits of the dibit correctly. Since the decoding error in the two channels are statistically independent, the average probability of a dibit being correctly decoded is

$$P_C = (1 - P')^2$$
$$= \left[1 - \frac{1}{2} erfc\left[\sqrt{\frac{E}{2\eta}}\right]\right]^2$$
$$= 1 - erfc\left(\sqrt{\frac{E}{2\eta}}\right) + \frac{1}{4} erfc^2\left(\sqrt{\frac{E}{2\eta}}\right)$$

Thus, the average probability of error in a QPSK system is

$$P_e = (1 - P_C)$$
$$= erfc\left(\sqrt{\frac{E}{2\eta}}\right) - \frac{1}{4} erfc^2\left(\sqrt{\frac{E}{2\eta}}\right)$$

For usual values of $\sqrt{\frac{E}{2\eta}}, \frac{1}{2} erfc\left(\sqrt{\frac{E}{2\eta}}\right)$ itself is very small and so its second power can be ignored to get a reasonable approximation for P_e of a QPSK system as

$$\underset{\substack{\text{QPSK}\\\text{(symbol error rate)}}}{P_e} \approx erfc\left(\sqrt{\frac{E}{2\eta}}\right) \tag{11.170}$$

In the above equation, E denotes the symbol energy, which is twice the bit energy E_b. Hence, we may write

$$\boxed{\underset{\substack{\text{QPSK}\\\text{(symbol error rate)}}}{P_e} = erfc\left(\sqrt{\frac{E_b}{\eta}}\right)} \tag{11.171}$$

or

$$\boxed{\underset{\substack{\text{QPSK}\\\text{(Bit-error rate)}}}{P_e} = \frac{1}{2} erfc\left(\sqrt{\frac{E_b}{\eta}}\right)} \tag{11.172}$$

Equation (11.172) for QPSK bit-error rate is exactly the same as that for the bit-error rate of BPSK. So, QPSK gives the same bit-error rate (for the same value of E_b/η and same bit rate) as BPSK, but requires only half the bandwidth required for BPSK since the symbol duration for QPSK is twice the bit duration of BPSK. That is why it is a bandwidth efficient system compared to BPSK. Put in a different way, we can say that for the same (E_b/η) and the same bandwidth as BPSK, we can transmit at twice the bit rate using QPSK. It is this aspect that has made it so popular and a preferred option over BPSK.

11.6 PROBABILITY OF ERROR FOR MSK

As usual, we shall assume the channel noise to be zero-mean additive white Gaussian with a two-sided power spectral density of $\eta/2$. The received signal will be the transmitted MSK signal $s(t)$ plus this noise. Let the received signal be $r(t)$. Then

$$r(t) = s(t) + n_w(t)$$

where $n_w(t)$ is one realization of the white noise on the channel.

Now, referring to Fig. 10.50 which shows the signal space diagram of MSK and its signal constellation, what the receiver does is, it finds the inphase and quadrature coordinates of this $r(t)$. For this purpose, we take the inner product of $r(t)$ with $g_1(t)$ and $g_2(t)$, the inphase and quadrature basis signals used for drawing the signal constellation.

$$\therefore \qquad r_1 \underline{\Delta} (r(t), g_1(t)) = \int_{-T}^{T} r(t) g_1(t)\, dt$$

$$= \int_{-T}^{T} s(t) g_1(t)\, dt + \int_{-T}^{T} n_w(t) g_1(t) dt$$

$$= s_1 + n_1; \quad -T \le t \le T \tag{11.173}$$

where s_1 is the inner product of $g_1(t)$ with $s(t)$ and as already shown in Eq. (10.112), it is given by

$$s_1 = \sqrt{E_b} \cos \phi(0); \quad -T \le t \le T \tag{11.174}$$

In Eq. (11.173), n, is a zero-mean Gaussian random variable with a variance which is the same as that of $n_w(t)$, i.e., η/2.

In a similar way, the quadrature coordinate of $r(t)$ is

$$r_2 \underline{\Delta} (r(t), g_2(t)) = \int_{0}^{2T} r(t) g_2(t) dt$$

$$= \int_{0}^{2T} s(t) g_2(t) dt + \int_{0}^{2T} n_w(t) g_2(t) dt$$

$$= s_2 + n_2; 0 \le t \le 2T \tag{11.175}$$

where s_2 is given by Eq. (11.172) as $-\sqrt{E_b} \sin \phi(T); 0 \le t \le 2T$, and n_2 is a zero-mean Gaussian random variable with variance η/2 and is independent of n_1.

If we look at the signal space and the signal constellation of MSK depicted in Fig. 10.49, we find that message points m_1 and m_4 are both having one feature in common, i.e., for both of them $\phi(0) = 0$. Similarly message points m_2 and m_3 have a common feature and that is that $\phi(0) = \pi$ for both of them. This suggests that if the projection of $r(t)$ on the signal space falls in the right half of the two-dimensional signal space of the MSK signal, then it means that the receiver chooses the estimate $\hat{\phi}(0) = 0$ and if it falls in the left half it chooses $\hat{\phi}(0) = \pi$. Hence, the receiver's decision on $\hat{\phi}(0)$ is as follows

$$r_1 \underset{\hat{\phi}(0)=\pi}{\overset{\hat{\phi}(0)=0}{\gtrless}} 0 \tag{11.176}$$

In a similar way, the Fig. 10.50 tells us that $\phi(T) = -\pi/2$ is the common feature of message points m_1 and m_2; and that $\phi(T) = +\pi/2$ is the common feature of message points m_3 and m_4. So, if the projection of the received signal $r(t)$ on the signal space falls in the upper half of the signal space, i.e., if $r_2 > 0$ then the receiver chooses the estimate $\hat{\phi}(T) = -\pi/2$. On the other hand, if the projection of $r(t)$ onto the signal space falls in the lower half of the signal space, i.e., if $r_2 < 0$, then the receiver decides in favor of the estimate $\hat{\phi}(T) = \pi/2$.

$$\therefore \qquad r_2 \underset{\hat{\phi}(T)=\pi/2}{\overset{\hat{\phi}(T)=-\pi/2}{\gtrless}} 0 \tag{11.177}$$

Hence, the decision regarding the baseband binary digit that is transmitted during $0 \le t \le T$ is made on the following basis:

(i) If the receiver chooses the estimates $\hat{\phi}(0) = 0$ and $\hat{\phi}(T) = -\pi/2$, or alternatively chooses the estimates $\hat{\phi}(0) = \pi$ and $\hat{\phi}(T) = \pi/2$, then it decides that a binary digit 0 was transmitted in the time interval $0 \le t \le T$.

(ii) If the receiver chooses the estimates $\hat{\phi}(0) = \pi$ and $\hat{\phi}(T) = -\pi/2$, or alternatively chooses the estimates $\hat{\phi}(0) = 0$ and $\hat{\phi}(T) = \pi/2$, then it decides that a binary 1 was transmitted during $0 \le t \le T$.

From the above statements and the signal-space diagram of Fig. 10.49, it follows that the receiver has to decide between the message points m_1 and m_3 for symbol 1. As shown in the receiver block diagram of Fig. 10.51, decision regarding the phase estimate $\hat{\phi}(0)$, i.e., whether it is 0 radians or π radians, is made in the I-channel (i.e., the inphase channel), while the decision regarding $\hat{\phi}(T)$, i.e., whether it is $+\pi/2$ or $-\pi/2$, is made in the *Q*-channel (i.e., the quadrature channel); and these decisions are made alternatively in the two channels. An error will be committed by the receiver if it commits a mistake in the decision making either in the *I*-channel or in the *Q*-channel.

From Eq. (11.173), we have

$$r_1 = s_1 + n_1; \quad -T \le t \le T$$

where $s_1 = \sqrt{E_b}\cos\phi(0)$ and n_1 is a zero-mean Gaussian random variable with a variance of $\eta/2$. Hence, r_1 is a Gaussian random variable with mean = s_1 and variance $\eta/2$. Now, the value of s_1 depends on whether $\phi(0) = 0$ or π. If $\phi(0) = 0$, s_1 equals $\sqrt{E_b}$

and if $\phi(0) = \pi$, s_1 equals $-\sqrt{E_b}$.

$$\therefore \qquad p(r_1 | \phi(0) = \pi) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x+\sqrt{E_b})^2/2\sigma^2}$$

$$p(r_1 | \phi(0) = 0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\sqrt{E_b})^2/2\sigma^2}$$

where $\sigma^2 = \eta/2$. These conditional probability density functions are shown in Fig. 11.38.

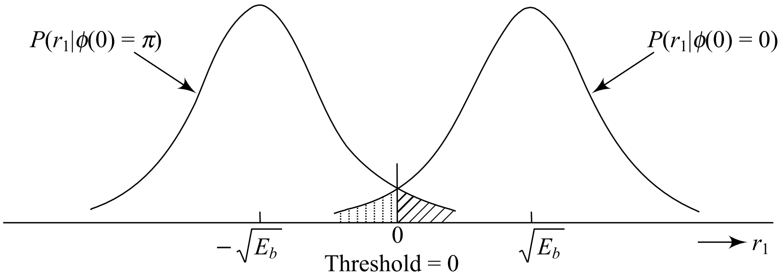


Fig. 11.38 *Conditional PDFs of the random variable r_1*

P_{e_1} = Probability of receiver deciding in favor of $\phi(0) = \pi$ even though $r_1 > 0$.

P_{e_2} = Probability of receiver deciding in favor of $\phi(0) = 0$ even though $r_1 < 0$.

But from symmetry, $P_{e_1} = P_{e_2}$

$$P_{e_1} = \frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty e^{-(r_1+\sqrt{E_b})^2/2\sigma^2}\, dr_1.$$

Putting $\dfrac{r_1+\sqrt{E_b}}{\sqrt{2}\,\sigma}=y, dy=\dfrac{1}{\sqrt{2}\sigma}dr_1$

$$\therefore \qquad P_{e_1}=\frac{1}{\sqrt{\pi}}\int_{\sqrt{E_b}/\sqrt{2}\sigma}^{\infty} e^{-y^2}dy=\frac{1}{2}erfc\sqrt{\frac{E_b}{2\sigma^2}}=\frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}}$$

Assuming $P[\phi(0)=0]=P[\phi(0)=\pi]=0.5$,

P_e' = Probability of erroneous decision in the Q-channel = $0.5P_{e1}+0.5P_{e2}=P_{e1}$

$$=\frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}}$$

Proceeding similarly, the probability of an erroneous decision in the Q-channel can also be shown to be the same. Now, decision in the two channels are independent. So, probability of an error in either of the channels is

$$\underset{\text{(symbol)}}{P_e}=erfc\sqrt{\frac{E_b}{\eta}}$$

Hence bit-error probability of MSK is

$$\boxed{P_e=\frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}}} \qquad (11.178)$$

This is the same as what we got for QPSK and binary PSK.

11.7 PROBABILITY OF ERROR FOR M-ARY BAND PASS SYSTEMS

11.7.1 P_e for M-ary Band Pass Signaling Schemes

P_e for M-ary QAM In Sections 10.8.1 and 10.8.2, we had discussed the basic structure of a QAM signal. There, we had observed that it is made up of two carrier signals of the same carrier frequency, f_c, but in phase quadrature, which are independently amplitude modulated by discrete amplitudes a_ks and $\underline{b_k}$s so that the kth QAM signal, $s_k(t)$, could be represented as

$$s_k(t)=a_k\sqrt{E_0}\phi_1(t)-b_k\sqrt{E_0}\phi_2(t);\ k=0,\pm1,\pm2,\ldots \qquad (11.179)$$

where

$$\phi_1(t)=\sqrt{\frac{2}{\tau}}\cos 2\pi f_c t;\quad 0\le t\le\tau$$

$$\phi_2(t)=\sqrt{\frac{2}{\tau}}\sin 2\pi f_c t;\quad 0\le t\le\tau$$

and E_0 is the energy of the signal with the smallest amplitude among all the different signals that can possibly be transmitted.

We know that the M signals appear as M distinct points in the signal space and that these M points constitute the signal constellation. For $M = 16$, this signal constellation is as shown in Fig. 10.58. If $M=L^2$, where L is a positive integer, this constellation may be viewed as having been generated by taking the Cartesian product of the constellation points of the two L-ary ASK signals, one with $\Phi_1(t)$ as the carrier and the other with $\Phi_2(t)$ as the carrier. We will now use this concept that the M-ary QAM system may be viewed as being made up of two L-ary ASK systems and derive the probability of error of the M-ary QAM system by making use of the expression for P_e of an L-ary ASK which is the same as that of an L-ary baseband system, and is given by Eq. (11.66).

Let $P_L(e)$ be the probability of error of an L-ary ASK system. Then, $P_L(e)$ is given by Eq. (11.66) as

$$P_L(e) = 2\left(1 - \frac{1}{L}\right)Q\left(\frac{A}{2\sigma}\right)$$

where A is the difference between adjacent levels. But in our case, this is equal to $2\sqrt{E_0}$ $\therefore$ $A/2 = \sqrt{E_0}$

Since $Q(x) = \frac{1}{2}erfc\left(\frac{x}{\sqrt{2}}\right)$, we may write

$$P_L(e) = 2\left(1 - \frac{1}{L}\right)erfc\left(\frac{A}{2\sqrt{2}\sigma}\right) = \left(1 - \frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{E_0}{\eta}}\right) \tag{11.180}$$

since $$\frac{A}{2\sqrt{2}\sigma} = \frac{\sqrt{E_0}}{\sqrt{2}\sigma} = \sqrt{\frac{E_0}{2\sigma^2}} = \sqrt{\frac{E}{\eta}}$$

We may now write the probability of correct detection in an M-ary QAM as

$$P_c = [1 - P_L(e)]^2 \tag{11.181}$$

The probability of symbol error for an M-ary QAM is then given by

$$P_e = (1 - P_c) = 1 - [1 - P_L(e)]^2 \tag{11.182}$$

As $P_L(e)$ itself will be small compared to 1, $P_L^2(e)$ can comfortably be ignored in comparison with 1. So, we may write

$$P_e = 2P_L(e) = 2\left(1 - \frac{1}{\sqrt{M}}\right)erfc\left(\sqrt{\frac{E_0}{\eta}}\right) \tag{11.183}$$

All the M signals of an M-ary QAM will not have the same amplitude and so their energies will also be different. So, let us find their average energy, E_{av}, in terms of E_0 which we have defined as the energy of the QAM signal with the smallest amplitude. This will enable us to conveniently express P_e, the probability of error of QAM in terms of this average energy, E_{av}. Now, the L amplitudes of the L-ary ASK system with $\phi_1(t)$ as the carrier, may be written as

$$a_k\sqrt{E_0} = \pm 1 \cdot \sqrt{E_0}, \pm 3 \cdot \sqrt{E_0}, \ldots, \pm (L-1) \cdot \sqrt{E_0} \tag{11.184}$$

Assuming that these L levels are equally likely,

$\overline{a_k^2}E_0$ = Average energy of the L signals of the L-ary ASK system with $\phi_1(t)$ as the carrier

$$= 2\left[\frac{1}{L}\sum_{i=1}^{L/2}(2i-1)^2(\sqrt{E_0})^2\right] = \frac{2E_0}{L}\sum_{i=1}^{L/2}(2i-1)^2 \tag{11.185}$$

Similarly,

$\overline{b_k^2}E_0$ = Average energy of the L signals of the L-ary ASK system with $\phi_2(t)$ as the carrier

$$= \frac{2E_0}{L}\sum_{i=1}^{L/2}(2i-1)^2 \tag{11.186}$$

$$\therefore \qquad E_{av} = 2\left[\frac{2E_0}{L}\sum_{i=1}^{L/2}(2i-1)^2\right] = \frac{2(L^2-1)E_0}{3}$$

$$\therefore \qquad E_{av} = \frac{2(M-1)E_0}{3} \tag{11.187}$$

Using Eq. (11.187) and substituting for E_0 in terms of E_{av} the expression for P_e given in Eq. (11.183),

$$\boxed{\underset{(\text{M-ary QAM})}{P_e} = 2\left(1 - \frac{1}{\sqrt{M}}\right) erfc\left(\sqrt{\frac{3E_{av}}{2(M-1)\eta}}\right)} \tag{11.188}$$

Remark *For M = 4, Eq. (11.188) reduces to the same expression that we had obtained for QPSK symbol error rate given in Eq. (11.170).*

P_e for M-ary FSK M-ary FSK was discussed in detail in Section 10.8.4 of Chapter 10 and it was stated there that the upper bound for the symbol error is given by

$$\underset{\substack{(\text{symbol error}\\ \text{M-ary FSK})}}{P_e} \le \frac{1}{2}(M-1) erfc\sqrt{\frac{E}{2\eta}} \tag{11.189}$$

where E is the energy of each one of the M-possible signals any one of which may be transmitted during a symbol period of τ seconds and $(\eta/2)$ is the two-sided PSD of the white noise on the channel.

As has been stated there, this upper bound is approached as (E/η) is increased and is almost reached when (E/η) is large enough to make $P_e \le 10^{-3}$. In the case of $M = 2$, that is for coherent BFSK, the equality sign holds and the RHS of the above equation reduces exactly to that obtained by us for coherent BFSK.

P_e for M-ary PSK M-ary PSK was discussed in detail earlier in Section 10.8.8 of Chapter 10 and as stated there, the probability of symbol error for M-ary PSK with $M \ge 4$ is approximately given by

$$\underset{\substack{(\text{symbol error}\\ \text{M-ary FSK})}}{P_e} \approx erfc\left[\sqrt{\frac{E}{2\eta}}\sin\left(\frac{\pi}{M}\right)\right] \tag{11.190}$$

where $M = 4$, the symbol error given by the above expression reduces exactly to the expression for symbol error for QPSK.

Note *E = Symbol Energy = 2E_b where E_b is the bit energy, in the case of QPSK.*

Example 11.28 An M-ary PSK system is to operate with 2^n symbols over a 100 kHz channel. The bit rate is required to be at least 750 kilobits/sec. What minimum CNR is required if the bit-error probability should be equal to or better than $P_b = 10^{-6}$? Assume ISI free conditions.

Solution The maximum symbol rate under ISI-free conditions, is

$$R_M = \frac{1}{\tau} \text{ Bandwidth } B_T$$

Since channel bandwidth is limited to 100 kHz,

$$R_M \le 100 \times 10^3 \text{ symbols/sec}$$

Minimum value of R_b required = 750×10^3 bits/symbol

Minimum number of bits/symbol = $\dfrac{750 \times 10^3}{100 \times 10^3} = 7.5$ bits/symbol

$\therefore$ for the M-ary PSK, $M \ge 2^{7.5}$

Since M must be an integer power of 2, let us take $M = 2^8$

$$\therefore \qquad M = 2^8 = 256$$

When Gray coding of bits is used obtain PSK symbols,

Probability of symbol error = $P_e = P_b \cdot \log_2 M = 10^{-6} \cdot \log_2 256$

$$\therefore \qquad \underset{(\text{symbol error})}{P_e} = 8 \times 10^{-6}$$

But, from Eq. (10.136), we know that for M-ary PSK with $M \geq 4$,

$$\underset{(\text{symbol error})}{P_e} \approx erfc\left[\sqrt{\frac{E}{\eta}}\sin\left(\frac{\pi}{M}\right)\right] \text{ where } E \text{ is the symbol energy}$$

$\therefore\ erfc\left[\sqrt{\frac{E}{\eta}}\sin\left(\frac{\pi}{M}\right)\right] = 8 \times 10^{-6}$. From error-function tables, we get

$$\left[\sqrt{\frac{E}{\eta}}\sin\left(\frac{\pi}{8}\right)\right] = 3.2$$

$$\therefore \qquad \frac{E}{\eta} = 69.924$$

Now, $R_M = \frac{R_b}{\log_2 M} = \frac{750 \times 10^3}{8} = 93.75 \times 10^3 \text{symbols/sec}$

$\therefore\ E$ = Symbol duration × Average carrier power

$$\therefore \qquad \frac{E}{\eta} \cdot \frac{1}{T} = \frac{\text{Average carrier power}}{\eta} = \frac{E}{\eta} \cdot R_s$$

$$= 69.924 \times 93.75 \times 10^3 = 6555.375 \times 10^3$$

$$\therefore \qquad (CNR)_{\min} = 10 \log_{10}(6555375) = 68.166 \text{ dB}$$

11.8 COMPARISON OF M-ARY DIGITAL MODULATION SCHEMES

Let the information come from a binary source at a bit rate of R_b bits/sec. Let the M-ary transmission rate be R_M and the *symbol* energy be E. Let the bit energy of the signal be E_b. Also let the bits per M-ary symbol be represented by n. Then we know that

$$n = \log_2 M \qquad (11.191)$$

$$R_M = \frac{R_b}{n} = \frac{R_b}{\log_2 M} \qquad (11.192)$$

$$E_b = \frac{E}{n} = \frac{E}{\log_2 M} \qquad (11.193)$$

Further, if the bit energy to white noise one-sided PSD be

$$\gamma_b \Delta E_b/\eta \qquad (11.194)$$

then

$$\gamma_b = \frac{E}{\eta n} \qquad (11.195)$$

In Table 11.2, we compare the various M-ary digital modulation schemes in terms of bandwidth efficiency R_b/B_T and the E_b/η for a specified P_e, viz., $P_e = 10^{-4}$ which is generally taken as the standard for comparison of various digital modulation schemes. Binary modulation schemes are also included therein to enable comparison of M-ary and binary systems.

Table 11.2 *Comparison of various digital modulation schemes*

Bit-error rate P_e is fixed at $P_e = 10^{-4}$

Type of modulation	Type of detection	R_b/B_T	γ_b in dB
BASK, or BFSK (with $f_d = R_b/2$ for BFSK)	Envelope Detection	1	12.3
Binary DPSK	Phase comparison with previous bit	1	9.3
BPSK	Coherent detection	1	8.4
MSK, QAM ($M = 4$), QPSK	Coherent quadrature detection	2	8.4
M-ary PSK with $M = 8$	Coherent quadrature detection	3	11.8
M-ary PSK with $M = 16$	Coherent quadrature detection	4	16.2
M-ary QAM with $M = 16$	Coherent quadrature detection	4	12.2

Remarks

(i) *In general, quadrature carrier systems like MSK and QPSK as well as all the M-ary band pass systems increase the bit rate of transmission and the bandwidth efficiency, but they do so at the expense of bit-error probability if transmitter power is fixed, or at the expense of transmitter power if P_e is fixed.*

(ii) *From Table 11.6, we find that QPSK, MSK and QAM (M = 4) offer the best trade-off between bandwidth efficiency and power.*

(iii) *Among the M-ary PSK systems, from the point of view of best trade-off between bandwidth efficiency and power, QPSK appears to be the best. It is for this reason that it is so popular.*

(iv) *M-ary PSK systems with M > 8 require excessively large average powers and so are not generally used.*

(v) *Insofar as bandwidth efficiency is concerned, M-ary PSK and M-ary QAM have similar performance. However, for M > 4, M-ary QAM requires less average power than M-ary PSK for the specified P_e. The reason for this is that for M > 4, they have different signal constellations. While the M-ary PSK has a circular constellation, M-ary QAM has a rectangular constellation. As M increases, for a circular constellation, the adjacent message signal points become nearer to each other for a given radius of the circle (which is equal to $\sqrt{E}$); but in the case of M-ary QAM, for the same average symbol energy they need not be that close because of the rectangular shape of the constellation. This is an important advantage of M-ary QAM. However, M-ary QAM, just like ASK, cannot generally be used in channels with non-linearities.*

MATLAB Example 11.1 **QAM Simulation** Perform a Monte Carlo simulation of an $M = 16$-QAM communication system. Assume a rectangular signal constellation. Plot the symbol error probability vs E_b/N_o(in dB).

MATLAB Program

```
  %
  % Assumed 16-QAM signal constellation for Monte Carlo simulation is as given
  % (-3*d 3d) (-d 3*d) (d 3*d) (3*d 3*d)
  % (-3*d d) (-d d) (d d) (3*d d)
  % (-3*d -d) (-d -d) (d -d) (3*d -d)
  % (-3*d - 3*d) (-d -3*d) (d -3*d) (3*d -3*d)
  % Monte Carlo simulations are done for transmission of 10,000 symbols at different
values of SNR
  % parameter Eb/No where Eb = Es/4 is the bit energy.
  % The program calls two functions named P = SM(snr) and Q
  % Assumed SNR range 0-20 in steps of 1 and 0.1

  SNR1 = 0:1:20
  SNR2 = 0:0.1:20;
  M = 16;
  N = 10000
  k = log2(M);
  d = 1;
  % 16-QAM signal constellation for Monte Carlo simulation
  mpg = [-3*d 3*d;
              - d 3*d;
              d 3*d;
              3*d 3*d;
              -3*d d;
              -d d;
              d d;
              3*d d;
              -3*d -d;
              -d -d;
              d -d;
              3*d -d;
              -3*d -3*d;
              -d -3*d;
              d -3*d;
              3*d -3*d];
  %
  % Calculation of simulated error rate
  %
  d = 1; % minimum distance between the symbols
  for i = 1:length(SNR1)
    sm_err_prb(i) = SM4(SNR1(i),N,d,M,mpg);
  end;
```

```
%
% signal to noise ratio
%
for i = 1: length(SNR2);
  snr = exp(SNR2(i)*log(10)/10);
  % Calculation of theoretical error rate
  theory_error_prb(i) = 4*qfunc(sqrt(3*k*snr/(M-1)));
end
%
semilogy(SNR1, sm_err_prb,'*',SNR2,theory_error_prb','b');
h = legend('Simulation error','Theory error',3)
xlabel('S/N in dB');
ylabel('Pe Error');
title('Pe Performance of M=16-QAM system ( Monte Carlo Simulation)');
xlim([0 15]);
ylim([10^-6 1]);
```

function [p] = SM(Snr,N,d,M,mpg)

```
% [p] = SM(Snr, N, d)
% This function finds the probability of error for a given value of snr in
% dB. Other arguments of the function are N number of symbols, d minimum
% distance between the symbols; % M is QAM Communication system; % energy % per symbol
%
% This function calls another function—gengauss
%
Eng_per_syb = 10*d^2;
% signal to noise per bit
snpb = 10^(Snr/10);
% noise variance
sigma = sqrt(Eng_per_syb/(8*snpb));
%
% Generation of data source
%
for i = 1:N;
  temp = rand;               % a uniform RV between 0 and 1
  source(i) = 1+floor(M*temp); % a number between 1 and 16 uniform
end;
mapping = mpg;
for i = 1:N
  qam_sig(i,:) = mapping(source(i),:);
end
% received signal
for i = 1:N
  [n(1) n(2)]= gengauss(sigma);
    r(i,:) = qam_sig(i,:)+n;
end
% Error probalility calculation and detection
Noerrs = 0;
```

```
for i=1:N
  for j = 1:M
    metrics(j) = (r(i,1) -mapping(j,1))^2+ (r(i,2)-mapping(j,2))^2;
  end;
  [min_metric decis] = min(metrics);
  if (decis~= source(i))
    Noerrs = Noerrs+1;
  end
end
p = Noerrs/(N);
```

function [gv1 gv2] = gengauss(m,sgma)

```
%    [gv1,grv2] = gengauss(m,sgma)
%    [gv1,grv2] = gengauss(sgma)
%    [v1,gsrv2] = gengauss

% 'gengauss' generates two independent Gaussian random variables with
% mean 'm' and standard deviation 'sgma'. If one of the input arguments is
% missing it takes the mean as 0. If neither the mean nor the variance is given, it %
generates two standard Gaussian random variables.

if nargin == 0
  m = 0; sgma = 1;
elseif nargin == 1
  sgma =m; m=0;
end
u = rand;                 % uniform random varioable in (0,1)
z = sgma*(sqrt(2*log(1/(1-u)))); % a Rayleigh distributed random variable
u = rand;
gv1 = m+z*cos(2*pi*u);
gv2 = m+z*sin(2*pi*u);
```

Results

```
SNR1 = 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

N =10000
```

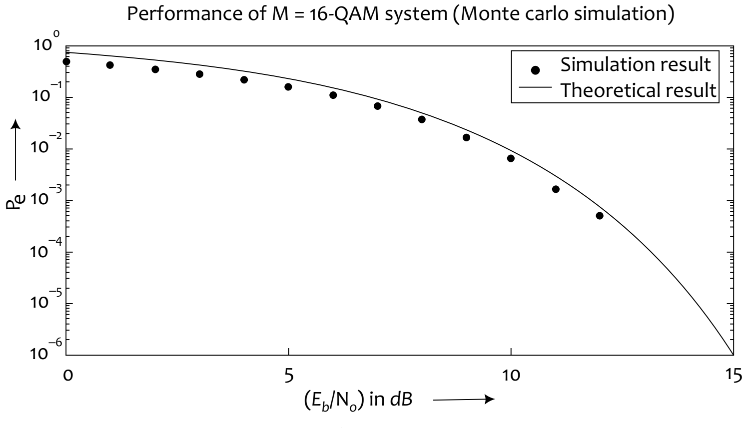


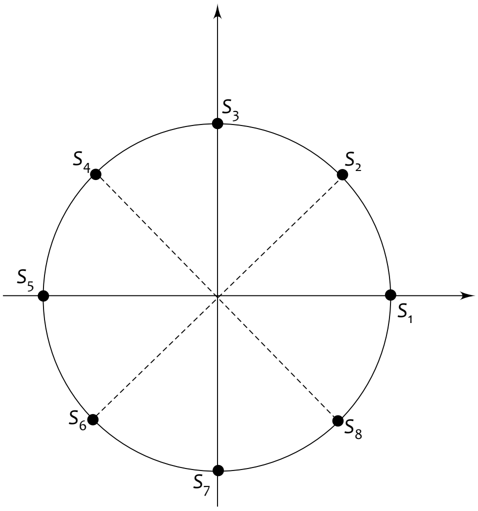
Fig. 11.39

MATLAB Example 11.2 Simulation of 8-ary PSK system

In this problem, we will study the variation of bit-error rate (BER) with E_b/N_o for an 8-ary PSK system with an AWGN channel.

E_b is the bit energy and $N_o/2$ is the two-sided PSD of the AWGN of the channel.

Using the BER tool provided in the MATLAB Communication toolbox, we perform Monte Carlo simulation of the system and compare the result with the theoretically calculated one by plotting the two together.



Signal constellation for 8-ary PSK

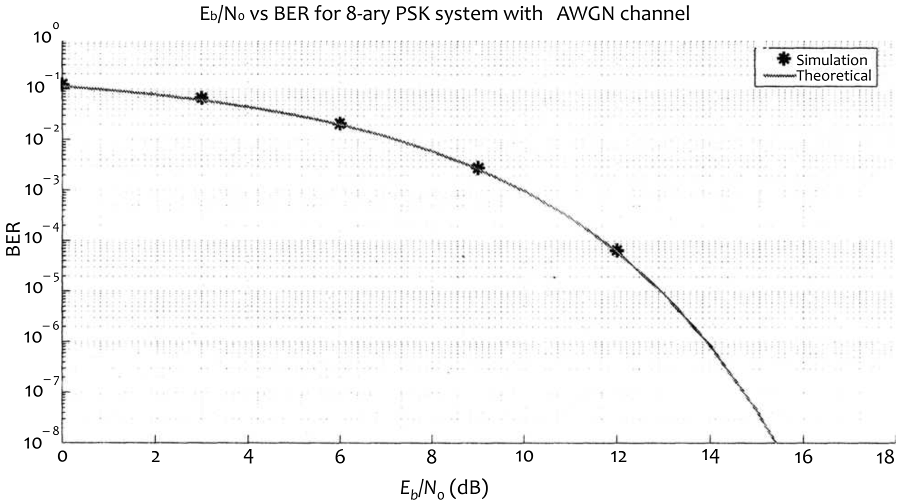


Fig. 11.40

Summary

- *Probability of error:* It is the average fractional number of erroneously received symbols when a very large number of received symbols are considered.
- *Matched filter:* It is a filter which is matched to a known signal of duration T sec, and which, when fed with this signal corrupted by additive white Gaussian noise, maximizes the output SNR at time $t = T$.
- $H(f)$ and $h(t)$ of a matched filter:

$$\underset{(\text{matched filter})}{H(f)} = kP^*(f)\exp(-j2\pi fT)$$

$$\underset{(\text{matched filter})}{h(t)} = kP(T-t)$$

- *Integrate-and-dump circuit:* Integrate-and-dump type of circuit is the matched filter for a rectangular pulse.

  (a) *For Unipolar NRZ* $P_e = Q\left[\frac{A}{2\sigma}\right]$

  (b) *For polar NRZ* $P_e = Q\left[\frac{A}{\sigma}\right]$

- *ML (Maximum likelihood) detection rule:*

  (a) *Decide* H_1 if $p_r(r_0|H_1) > p_r(r_0|H_0)$

  (b) *Decide* H_0 if $p_r(r_0|H_0) > p_r(r_0|H_1)$

  where r_0 is the observed random variable.

- *MAP (Maximum a posteriori probability) detection rule:*

  (a) *Decide in favor of* H_1 if $\frac{p_r(r_1|H_1)}{p_r(r_1|H_0)} > \frac{P_0}{P_1}$

  (b) *Decide in favor of* H_0 if $\frac{p_r(r_1|H_0)}{p_r(r_1|H_1)} > \frac{P_1}{P_0}$

  where r_1 is the value of the observed random variable $r(T_0)$.

- *Schwarz's inequality:*

$$\left|\int_{-\infty}^{\infty} s_1(t)s_2^*(t)\,dt\right|^2 \le \left[\int_{-\infty}^{\infty} |s_1(t)|^2 dt\right]\left[\int_{-\infty}^{\infty} |s_2(t)|^2 dt\right]$$

  where the equality sign holds *if and only if* $s_2(t) = cs_1^*(t)$

- *Transfer function of the optimum filter:*

$$H_{\text{opt}}(f) = k\left[\frac{P^*(f)e^{-j2\pi fT}}{S_n(f)}\right]$$

- *Probability of error with matched filter:*

  (a) *Unipolar signal* $P_e = Q\left[\sqrt{\frac{E_b}{\eta}}\right]$; $E_b = \frac{A^2T}{2}$

  (b) *Polar signal* $P_e = Q\left[\sqrt{\frac{2E_b}{\eta}}\right]$; $E_b = A^2T$

- *Matched filter reception and correlation are equivalent:* Performance-wise, there is no difference between the two.

- *M-ary baseband signaling:* The baseband pulse will have M-levels (where M is a power of 2) instead of only two levels as in binary case.

$$R_M = \text{Transmission rate with M-ary signaling} = \frac{1}{nT} = \frac{R_b}{\log_2 M}$$

where n = (symbol duration/time-slot duration).

(a) *Bandwidth for M-ary* $W_m = (W_b / \log_2 M)$

(b) *Probability of error* P_e $\underset{\substack{\text{(Polar M-ary}\\ \text{Baseband)}}}{P_e} = 2\left(1 - \frac{1}{M}\right)Q\left(\frac{A}{2\sigma}\right)$

- *Optimum receiver using MMSE criterion:* This is a receiver optimized by using the minimum mean-square error criterion and takes care of both channel noise as well as ISI. If $q(t)$ is the pulse at the input to the receiver, $Q(f)$ its Fourier transform and if $S_q(f)$ is the PSD of $q(t)$, then the MMSE optimum filter is given by

$$\underset{\text{(MMSE)}}{H_{\text{opt}}(f)} = \frac{Q^*(f)}{S_q(f) + (\eta/2)}$$

- *Noise performance of ASK, FSK and PSK:*

(a) *ASK* $\underset{\text{(Coherent ASK)}}{P_{e\min}} = \frac{1}{2} erfc\left[\sqrt{\frac{E_{\text{av}}}{2\eta}}\right]$; $\underset{\text{Non-Coherent}}{P_e} \cong \frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$

(b) *FSK* $\underset{\text{(Coherent CPBFSK)}}{P_{e\min}} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$; $\underset{\text{(Non-coherent)}}{P_e} = \frac{1}{2}e^{-A^2/4\sigma^2} = \frac{1}{2}e^{-\frac{1}{2}\left(\frac{E_b}{\sigma^2 T}\right)} = \frac{1}{2}e^{-\frac{E_b}{2\eta}}$

(c) *PSK* $\underset{\text{(PSK)}}{P_{e\min}} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

- *Comparison of binary ASK, FSK and PSK:*

Binary digital modulation	**Type of detection**	R_b/B_T	$P_b(e)$
ASK or OOK	Non-coherent	1	$\frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$*
	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$*
BFSK	Non-coherent	1	$\frac{1}{2}\exp\left[-\frac{E_b}{2\eta}\right]$
	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{2\eta}}\right]$
BPSK	Coherent	1	$\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

*For these two, $E_b = E_{\text{av}} = A^2 T/4$

(a) For a given bit energy, PSK gives the lowest probability of error.

(b) Non-coherent ASK and non-coherent FSK give essentially the same performance.

(c) For achieving a specified probability of error, ASK requires twice the pulse energy as compared to PSK, i.e., it requires 3 dB more power than PSK.

- *Phase offset in a BPSK system:*

  $$\underset{\text{(BPSK)}}{P_e} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$$ when there is no phase offset.

  $$\underset{\text{(BPSK)}}{P_e} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}} \cos\theta\right]$$ when a phase offset of θ is present.

- *Discontinuous phase BFSK:*

  $$\underset{\substack{\text{(BFSK-coherent)}\\ \text{(discontinuous phase)}}}{P_{e\,\min}} = \frac{1}{2} erfc\left[\sqrt{0.6\frac{E_b}{\eta}}\right]$$

- *P_e for DPSK:* $\underset{\text{(DPSK)}}{P_e} = \frac{1}{2}\exp\left[-\frac{E_b}{\eta}\right]$

- *P_e for QPSK:* $\underset{\substack{\text{(QPSK)}\\ \text{(bit-error rate)}}}{P_e} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$

- *Bandwidth and bandwidth efficiency of QPSK:*

  3-dB bandwidth of QPSK = $B_T = \frac{1}{2T}$ $\quad\therefore$ bandwidth efficiency = 2

- *Bandwidth efficiency and P_e of MSK:*

  (a) $\underset{\text{(3-dB)}}{B_T} = \frac{1}{2T} = \frac{R_b}{2}$ $\quad\therefore$ bandwidth efficiency, $\frac{R_b}{B_T} = 2$

  (b) MSK signal's PSD falls off as the fourth power of $|(f - f_c)|$

  (c) $\underset{\text{(MSK)}}{P_e} = \frac{1}{2} erfc\left[\sqrt{\frac{E_b}{\eta}}\right]$, same as that of QPSK and BPSK.

- *QAM (Quadrature Amplitude Modulation):* In this, two carriers in phase quadrature are independently amplitude modulated by discrete amplitudes a_ks and b_ks.

  $$\underset{\text{(M-ary QAM)}}{P_e} = 2\left(1 - \frac{1}{\sqrt{M}}\right) erfc\left(\sqrt{\frac{3E_{\text{av}}}{2(M-1)\eta}}\right)$$

  Bandwidth efficiency $\frac{R_b}{B_T} = \log_2 M$

  Signal space is two-dimensional.

- *M-ary FSK:* $P_e \le \frac{1}{2}(M-1)\, erfc\left(\sqrt{\frac{E}{2\eta}}\right)$

  Bandwidth efficiency = $\left(\frac{2\log_2 M}{M}\right)$

- *M-ary PSK:* $s_k(t) = \sqrt{\frac{2E}{\tau}} \cos\left[\omega_c t + \frac{2\pi}{M}(k-1)\right]$

  $$k = 1, 2, 3, \ldots, M$$

  Signal space is two-dimensional and all the M message signal points lie on the circumference of a circle with center at the origin and radius equal to $\sqrt{E}$ where E is symbol energy.

  $$\underset{\text{(M-ary PSK)}}{P_e} \cong erfc\left(\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{n}\right)\right); \rho = \text{Bandwidth efficiency} = \log_2 M$$

- *Comparison of digital band pass signaling schemes:*

Type of modulation	Type of detection	R_b/B_T	γ_b in dB
BASK, or BFSK (with $f_d = R_b/2$ for BFSK)	Envelope detection	1	12.3
Binary DPSK	Phase comparison with previous bit	1	9.3
BPSK	Coherent detection	1	8.4
MSK, QAM ($M = 4$), QPSK	Coherent quadrature detection	2	8.4
M-ary PSK with $M = 8$	Coherent quadrature detection	3	11.8
M-ary PSK with M = 16	Coherent quadrature detection	4	16.2
M-ary QAM with $M = 16$	Coherent quadrature detection	4	12.2

References and Suggested Reading

1. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 4th Edition (International Edition), McGraw-Hill Education, 2002.
2. Couch II, Leon W., *Digital and Analog Communication Systems*, 6th Edition, Pearson Education Asia, 2001.
3. Gronemeyer, S.A., and A.L. McBride, MSK and Offset QPSK Modulation, *IEEE Transactions on Communications*, Vol. 24, pp. 809–820, Aug. 1976.
4. Haykin, Simon, *Communication Systems*, 4th Edition, John Wiley & Sons India, 2001.
5. Haykin, Simon, *Digital Communications*, John Wiley & Sons, 1988.
6. Lathi, B.P., *Modern Digital and Analog Communication Systems*, 3rd Edition, Oxford University Press, 1998.
7. Lee, Edward A., and David G. Messerschmitt, *Digital Communication*, 2nd Edition, Allied Publications Limited in Association with Kluwer Academic Publishers, 1994.
8. Pasupathy, S., Minimum Shift Keying – A Spectral Efficient Modulation, *IEEE Communication Magazine*, Vol. 17, No. 4, pp 14–22, 1979.
9. Proakis, John G., *Digital Communications*, 4th Edition, McGraw-Hill International Edition, McGraw-Hill, 2001.
10. Proakis, John G., and Masoud Salehi, *Fundamentals of Communication Systems*, Low Priced Edition, Pearson Education, 2005.
11. Taub, Herbert, Donald Schilling, and Gautam Saha, *Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill, New Delhi, 2008.
12. Turin, G.L., An Introduction to Matched Filters, *IRE transactions on Information Theory*, Vol. 6, pp. 311–329, 1960.
13. Wozencraft, John M., and Irwin Mark Jacobs, *Principles of Communication Engineering*, John Wiley & Sons, 1965.

Review Questions

1. What is a matched filter?
2. Write down the expressions for the transfer function and the impulse response, of a matched filter for the signal $p(t)$ which is of duration T sec.
3. What is an 'integrate-and-dump' circuit? Show that it is the matched filter for rectangular pulse of duration T sec.
4. Show that a matched filter receiver and a correlation receiver are equivalents of each other.
5. The received signal (for binary baseband receiver) is $r(t) = s(t) + n(t)$

   where $s(t) = \begin{cases} A; 0 \le t \le T \text{ if } H_1 \text{ is true} \\ -A; 0 \le t \le T \text{ if } H_0 \text{ is true} \end{cases}$

   and $n(t)$ is zero-mean Gaussian noise of variance σ_n^2

   (a) Write down expressions for $p_{r|H_1}(r|H_1)$ and $p_{r|H_0}(r|H_0)$ where r is the sample of $r(t)$.

   (b) Sketch the two conditional density functions and show the optimum threshold value assuming $P_1 = P_0$.

6. With reference to baseband binary transmission what is an 'Optimum Linear Receiver'?
7. Draw the block diagram of an optimum linear receiver for binary baseband signaling?
8. Compare the noise performance of BASK, BFSK and BPSK for fixed average transmitted power and identical noise environment.
9. For a fixed bit-error probability, P_e, comment on the bandwidth efficiencies and the average transmitted power requirements of (a) BPSK, (b) QPSK, (c) 8-ary PSK, and (d) 8-ary QAM.
10. Qualitatively explain why M-ary QAM gives better noise performance than M-ary PSK for $M > 8$.

Problems

1. A binary data stream, with polar NRZ line code uses +V volts for binary 1 and –V volts for binary 0. Assuming that $P(1) = 0.6$ and $P(0) = 0.4$, and that channel noise is AWGN with two-sided spectral density of $\eta/2$, determine **(a)** the optimum threshold voltage, and **(b)** P_e, if an integrate-and-dump type of receiver is used.
2. A signal which takes the values $+A$, 0 and $-A$ volts for T seconds with equal probability, is transmitted over a channel with additive white Gaussian noise of two-sided PSD equal to $\eta/2$. An integrate-and-dump type of receiver is used. What threshold voltages should be used if the probability of the receiver committing an error is to be independent of which signal is transmitted?
3. An integrate-and-dump type of receiver, using an RC-integrator with a 3-dB cut-off frequency of f_c, is to detect the received signal $r(t) = s(t) + n_w(t)$, where $s(t) = \pm A$ for an interval of T sec and $n_w(t)$ is additive white Gaussian noise of two-sided PSD equal to $\eta/2$. What should be the value of f_c if at the sampling instant, the *SNR* is to be maximum? How does this *SNR* compare with what would have been obtained if an ideal integrator was sued?
4. Determine the matched filter for the following signal:

$$s(t) = \begin{cases} +A/2; & 0 \le t \le T/2 \\ -A/2; & T/2 \le t \le T \end{cases}$$

   Sketch the matched filter output as a function of time.

5. A baseband binary communication system uses two signals $s_1(t)$ and $s_2(t)$ corresponding to binary 1 and binary 0 respectively. $s_1(t)$ is same as the $s(t)$ of problem 5 above and $s_2(t)$ is given by

$$s_2(t) = \begin{cases} A; & 0 \le t \le T \\ 0; & \text{otherwise} \end{cases}$$

   Assuming that $s_1(t)$ and $s_2(t)$ are equally likely, and that the signals are corrupted by AWGN during transmission and further that the receiver uses two matched filters, one matched to $s_1(t)$ and the other matched to $s_2(t)$, determine the probability of error of this communication system.

6. A baseband binary transmission system uses Manchester code for representing binary symbols 1 and 0. If a 1 is represented by $s(t)$ and a 0 by $-s(t)$ where

$$s(t) = \begin{cases} 0; & 0 \le t \le T/2 \\ -1; & T/2 \le t \le T \\ 0; & \text{otherwise} \end{cases}$$

   and if $P(1) = P(0) = 0.5$, find the probability of error if a maximum likelihood receiver is used. Assume that an AWGN channel is employed.

7. A binary baseband long-haul transmission system using polar NRZ signals has 20 repeaters. If the input *SNR* at each repeater is 22 dB, find the probability of error assuming the repeaters to be regenerative. What will be the P_e if the repeaters are non-regenerative?
8. A pulse $p(t) = A\Pi\left(\frac{t - T/2}{T}\right)$ is to be detected in the presence of AWGN. Instead of using a matched filter for maximizing the peak signal-to-noise ratio, if we use an ideal LPF of bandwidth B, what is the optimum value of B for which the LPF is the best approximation to the matched filter insofar as maximizing the peak *SNR*? By how many decibels is this approximation inferior to the matched filter?

9. A PCM system with NRZ polar signaling and operating above the threshold, has a probability of error, $P_e = 10^{-6}$. If the signaling rate is doubled, what will be new P_e?
10. A baseband binary transmission system transmits the symbols $s_1(t)$ and $s_2(t)$ at a rate of 10 kbps. If

$$s_1(t) = \begin{cases} A & \text{for } 0 \le t \le T/2 \\ 0 & \text{for } T/2 \le t \le T \end{cases}$$

and $s_2(t) = 0 \quad ; \quad 0 \le t \le T$

Given that the channel noise at the input to the receiver has an r.m.s. value of 1 mV and that the channel attenuation is 35 dB, determine the minimum average signal power that must be transmitted to ensure that $P_e \le 10^{-6}$.

11. FSK need not necessarily use orthogonal signals. Find the frequency shift f_d between the two BFSK signal that will minimize the bit-error rate for a coherent BFSK system.
12. An optimum receiver for Sunde's BFSK is implemented in the form of two parallel matched filters, one matched to $s_1(t)$ and the other to $s_2(t)$. A sample of the difference between the outputs of these two matched filters, taken at the end of each time slot, is used for decision making. Sketch the amplitude response characteristics of the two filters.
13. A channel has 80-dB transmission loss and white noise with two-sided PSD of 0.5×10^{-10} W/Hz. Binary data is to be transmitted over this channel at a bit rate of 10^5 bits/sec. The bit-error rate is not to exceed 10^{-4}. Find the transmitted power needed for each of the following types of modulation:
    (a) Non-coherent FSK
    (b) DPSK
    (c) Coherent BPSK
14. Over a radio channel having a bandwidth of 200 kHz, binary data is to be transmitted at a bit-rate of 600 kbps.
    (a) Which modulation method needs minimum signal energy?
    (b) For that modulation method, calculate (E_b/η) for obtaining a bit-error rate of 10^{-5}.
15. By what factor must the symbol energy be changed to keep the probability of error unchanged for the following cases.
    (a) For converting a 16-ary QAM into DPSK.
    (b) For converting a 16-ary PSK system into a 16-ary QAM system.
16. Determine the reduction in the transmission bandwidth and the average signal energy of 64 QAM and 16 QAM for the same probability of error to be obtained for both.

Multiple-Choice Questions

1. The transfer function of the matched filter for a triangular pulse that is symmetrical about $t = T/2$ is
    (a) a sinc function
    (b) a sine function
    (c) a sinc square function
    (d) a rectangular function
2. For the same white noise corrupting the two signals shown in Fig. 11.M2, if the maximum *SNR*s at the outputs of the respective matched filters of $p(t)$ and $q(t)$ are to be equal, K should be equal to

    (a) $K = 2A$ (b) $K = \sqrt{3}A$ (c) $K = 1.5A$ (d) $K = \sqrt{2}A$

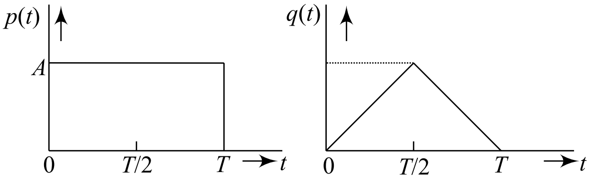


Fig. M11.2

3. For the same signaling speed and with the same white noise corrupting both the signals and with $P(0) = P(1)$, if a polar NRZ signal with amplitude A_1 and a unipolar NRZ signal with amplitude A_2 are to give the same P_e under matched filter conditions, A_1 and A_2 should be related as

(a) $A_2 = \sqrt{2}A_1$ (b) $A_2 = 2A_1$ (c) $A_2 = A_1$ (d) $A_1 = \sqrt{2}A_2$

4. Consider polar quarternary baseband transmission with adjacent levels separated by A V and binary unipolar NRZ transmission with a binary 1 represented by a pulse of amplitude A and a binary 0 by zero volts. If the probability of error in the former case is P_{eM} and in the latter case is P_{eB}, under the same noise conditions for both, the ratio (P_{eM}/P_{eB}) is equal to

(a) 3/4 (b) 3/2 (c) $3/2\sqrt{2}$ (d) 3/8

5. A certain non-coherent BFSK system is giving a bit error probability P_{e1}. When the average transmitted power is increased by n dB, keeping all other things unaltered, the bit-error probability is P_{e2}. If $(P_{e2}/P_{e1}) = 2\,P_{e1}$, n equal

(a) 2 dB (b) 3 dB (c) 4 dB (d) 6 dB

6. For a specified average transmitted power, the system that gives the lowest probability of error among the following, is

(a) Non-coherent FSK system (b) Coherent FSK system
(c) PSK system (d) Coherent ASK system

7. The probability of error for a BPSK system is

(a) $\frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}}$ (b) $\frac{1}{2}erfc\sqrt{\frac{2E_b}{\eta}}$ (c) $erfc\sqrt{\frac{E_b}{2\eta}}$ (d) $\frac{1}{2}erfc\sqrt{\frac{E_b}{2\eta}}$

where E_b is the bit energy and $\eta/2$ is the two-sided PSD of the white noise on the channel.

8. If there is a phase error of θ in a coherent BPSK system, the P_e is

(a) $\frac{1}{2}erfc\left[\sqrt{\frac{E_b}{\eta}\cos\theta}\right]$ (b) $\frac{1}{2}erfc\sqrt{\frac{2E_b}{\eta}}\cos\theta$ (c) $erfc\sqrt{\frac{E_b}{2\eta}}\cos\theta$ (d) $erfc\sqrt{\frac{E_b}{2\eta}\cos\theta}$

9. P_e for a DPSK system is

(a) $\exp\left[-\frac{E_b}{\eta}\right]$ (b) $\exp\left[-\sqrt{\frac{E_b}{\eta}}\right]$ (c) $\frac{1}{2}\exp\left[-\frac{E_b}{\eta}\right]$ (d) $\frac{1}{2}\exp\left[\frac{E_b}{\eta}\right]$

10. The signal space of a QPSK system has a dimension of

(a) 1 (b) 3 (c) 4 (d) 2

11. For the same bit-error and channel noise, the P_e of QPSK is the same as that of

(a) BPSK (b) BFSK Coherent (c) BFSK Non-coherent(d) DPSK

12. In a QPSK signal, the carrier phase can sometimes change by as much as

(a) $\pm\frac{\pi}{4}$ (b) $\pm\frac{\pi}{2}$ (c) $\pm\pi$ (d) $\pm\frac{3\pi}{2}$

13. In MSK, the frequency difference between the two signals that can possible be transmitted is

(a) $\frac{1}{4T}$ (b) $\frac{1}{2T}$ (c) $\frac{1}{T}$ (d) $\frac{2}{T}$

14. In MSK, the phase change in each bit interval is

(a) $+\frac{\pi}{2}$ or $-\frac{\pi}{2}$ (b) 0 or π (c) $+\frac{\pi}{4}$ or $-\frac{\pi}{4}$ (d) $+\pi$ or $-\pi$

15. The dimension of the signal space of MSK signal is

(a) 1 (b) 2 (c) M (d) 3

16. In MSK, during the interval $-T \le t \le T$, the inphase component is

(a) a full-cycle cosine pulse (b) a half-cycle sinusoidal pulse
(c) a half-cycle cosinusoidal pulse (d) a full-cycle sinusoidal pulse

17. The symbol-error probability of MSK is

(a) $\frac{1}{2} erfc \sqrt{\frac{E_b}{\eta}}$ (b) $\frac{1}{2} erfc \sqrt{\frac{E_b}{\eta}}$ (c) $\frac{1}{2} erfc \sqrt{\frac{2E_b}{\eta}}$ (d) $erfc \sqrt{\frac{E_b}{\eta}}$

18. The PSD of MSK falls off as the n^{th} power of $|f - f_c|$ where n is

(a) 1 (b) 2 (c) 3 (d) 4

19. In a 16-ary PSK, the symbol rate is 10 kbps. The bit rate is

(a) 160 kbps (b) 40 kbps (c) 2.5 kbps (d) (10/16) kbps

20. For M-ary PSK systems the best trade-off between bandwidth efficiency and transmitted power is given for a value of M equal to

(a) 2 (b) 4 (c) 8 (d) 16

Key to Multiple-Choice Questions

1. (c)	2. (b)	3. (a)	4. (b)	5. (b)	6. (c)	7. (a)	8. (b)
9. (c)	10. (d)	11. (a)	12. (c)	13. (b)	14. (a)	15. (b)	16. (c)
17. (d)	18. (d)	19. (b)	20. (b)				

INFORMATION THEORY AND SOURCE CODING

12

> *"You see things; and you say "Why?" But I dream things that never were; and I say "Why not?"*
>
> **George Bernard Shaw (1856–1950)**
> *Irish playwright*

Communication systems are basically meant to transfer *information* from one location to another. The broad objective of this chapter is therefore to look at communication systems from this perspective.

Learning Objectives

After going through this chapter, students will be able to

- *view information as removal of uncertainty, is familiar with the 'measure' of information and can determine the average rate at which a Discrete Memory Source (DMS) is giving information,*
- *understand the need for 'source coding' and can encode the output from a discrete memoryless source using Fano coding, Huffman coding or Lempel Ziv coding,*
- *understand the importance and the implications of Shannon's source coding theorem,*
- *understand the need for channel coding and the importance and implications of Shannon's channel coding theorem, and*
- *relate the concept of 'Mutual information' of a channel to information transfer through the channel and understands that Shannon's Information capacity theorem sets a fundamental limit on the rate at which error-free transmission can be achieved over power-limited, band-limited Gaussian channels.*

12.1 INTRODUCTION

The ultimate goal of any communication system is to transmit, over a channel, the information originating from a source in one location, to the destination in another location; and to do this as speedily and as reliably as possible. The transmitter connects the source to the channel and the receiver connects the channel to the destination. Any given channel has an inherent limitation with regard to the speed of transmission of information through it. Further, it introduces noise, which tends to limit the reliability of the transmitted information.

When we ponder over the problem of achieving the ultimate goal of a communication system, as stated earlier, several questions of a fundamental nature arise, like what is information? Can we measure it? To what extent can we compress it so that it can still be retrieved with very little, or no loss? Are speed and reliability of transmission inter-related? If so, what is the maximum rate at which information can be transmitted reliably?

Although some work in this direction was done earlier by Hartley, Nyquist, etc., it was the pioneering work of Claude E. Shannon of Bell Laboratories in 1948 that could bring all the concepts concerning information, its representation and transmission within a rigorous mathematical framework and provide satisfactory answers to all questions such as those posed earlier. In short, it revealed the fundamental limits of communication. His work, and that of a host of researchers all over the world who were spurred by his famous papers published in the Bell System Technical Journal (BSTJ) in 1948, ushered in a new branch of science, called the 'Information Theory' that provided satisfactory answers to all the questions raised earlier. Information theory has found interesting applications in such diverse fields as linguistics, computer science, genetics, statistical physics, and communication engineering. Shannon's famous source coding theorem provides the answer for the question concerning the limit on compression of information while his equally famous noisy channel coding theorem provides the answer for the question on the maximum rate at which information can be transmitted reliably.

Information theory deals with communication sources and systems in an abstract way using their mathematical models developed specifically from the point of information generation and its reliable transmission rather than in terms of systems, subsystems and their working principles, etc. It has developed a measure of information based on probability and models for sources which depict them as generating discrete random processes with finite alphabet and having some well-defined statistical properties.

12.2 INFORMATION MEASURE AND ENTROPY

In this section, we will first briefly discuss what we mean by information and then proceed to define a measure for the amount of information. This, we will do in a heuristic manner but our definition of an 'information measure' will satisfy all the properties that we intuitively associate with information. Building upon these ideas, we will determine the entropy, or the 'average information' of a message or a source. But, before proceeding further, it is appropriate and useful to distinguish between 'the amount' of information and the 'value', or 'usefulness' of that information. The 'amount' of information is determined by measurements carried out according to some well-defined scientific principles whereas, the 'usefulness' of that information is purely subjective and varies from person to person and is, therefore, out of our purview.

Information theory proceeds on the premise that removing uncertainty on any matter is equivalent to giving information on that matter. It therefore measures the 'amount of information' given by the occurrence of an event in terms of the 'amount of uncertainty' removed by its occurrence. Everyone knows that the sun rises in the East and there is absolutely no uncertainty about. So, if a message 'The sun will rise in the East tomorrow' is received, the message has not removed any uncertainty since there was not any, and hence the amount of information obtained from it is zero. Now, the 'amount of uncertainty' regarding the occurrence of an event is related to the 'probability' of its occurrence and is in fact, inversely proportional to the probability. Smaller the probability of occurrence of an event, larger is the uncertainty associated with its occurrence and therefore, larger is the amount of information associated with the occurrence of that event. Also, if the probability of occurrence is 1, i.e., if it is a sure, or certain event, like the sun rising in the East, such an event has zero information associated with it.

Thus, we find that the probability of occurrence of an event can be used to measure the information associated with the occurrence of that event. But then, we are faced with the question: 'What function of probability'? Can we use $(1/p)$ itself, where p is the probability of that event?' Let us examine and see whether it satisfies the properties that we *intuitively* associate with information.

Suppose we make two independent tossings of a coin and both the times the 'head' shows up. If p is the probability of a 'head' showing up in the tossing, the probability of getting head in both the tossings is p^2. The information associated with a 'head' showing up, is $(1/p)$. Since the two tossings are independent, we intuitively feel that the total information from the two tossings must be equal to the sum of the amount of information obtained from each tossing and so it should be $(2/p)$. But since the result of the combined experiment has a probability of p^2, according to our measure of information, the amount of information given by the combined experiment is $(1/p^2)$. But then $2/p$ is not in general equal to $(1/p^2)$. Thus, $(1/p)$ cannot be used as a measure of information.

What we used in the above argument is not the only property that we intuitively associate with information. Let p be the probability of occurrence of an event 'a' and $I(a)$ be a function of p that is used as a measure of information. As stated earlier, since the value or the usefulness of the information is not our concern, the information measure $I(a)$ should depend only on p and not on the nature or usefulness of the event whose probability of occurrence, is p. $I(a)$ must satisfy the following conditions also:

1. $I(a)$ must be a continuous function of p.
2. $I(a)$ must be a decreasing function of p.
3. If $p = p_1 p_2$ where p_1 is the probability of occurrence of an event a_1 and p_2 is the probability of occurrence of the event a_2 and p is the probability of occurrence of a_1 and a_2 (i.e., a_1 and a_2 are statistically independent events), then $I(a)$ must be equal to $[I(a_1) + I(a_2)]$.
4. $I(a)$ must be non-negative for $0 \le p \le 1$.

It can be shown that a logarithmic function, log $(1/p)$ is a suitable function to be used as a measure of information. This can seen from the fact that since p can take only positive values, the logarithm of $(1/p)$ is a continuous function of p and obviously a decreasing function of p. The third condition is also satisfied since

$$\log\left(\frac{1}{p_1}\cdot\frac{1}{p_2}\right) = \log\left(\frac{1}{p_1}\right) + \log\left(\frac{1}{p_2}\right)$$

The fourth condition, that $I(a)$ must be non-negative is also satisfied by this logarithmic measure, since $\log(1/p) \ge 0$ for $0 \le p \le 1$.

So, we shall hereafter associate an amount of information equal to log $(1/p)$ with the occurrence of an event whose probability of occurrence is p.

Whenever we measure something, there has to be some unit for measurement – like inches, or centimeters for length. The unit in which the amount of information is measured, depends on the base used for the logarithm. If the base is 2, the units are called 'bits' – a contraction of binary digits, if the base is e, then the units are 'nits', and if the base is 10, the units are called Hartley's. As a matter of rule, we will always use only a base of 2, i.e., units of bits.

Let us now see what '1 bit' of information represents. Consider the tossing of a fair coin. The probability of 'heads' is equal to the probability of 'tails' and each is 0.5. The information associated with the event of 'H' showing up, i.e., the information associated with a single tossing of a fair coin is

$$I(H) = \log_2\left(\frac{1}{0.5}\right) = 1 \text{ bit}$$

In general, '*one bit*' of information is associated with any binary decision like the above where the probability of either result is the same. ***So, one bit of information is given whenever a choice is made between two equiprobable events***.

Thus, the information obtained from the occurrence of an event a_k with probability of occurrence of p_k, is given by

$$I(a_k) = \log_2\left(\frac{1}{p_k}\right) \text{ bits} = -\log_2(p_k) \text{ bits} \tag{12.1}$$

Example 12.1 Of the two units of information – bit and nit, which is bigger? How are they related?

Solution If p is the probability of occurrence of an event, information associated with the occurrence of the event is

$$-\log_2 p \text{ bits} \quad \text{or} \quad -\log_e p \text{ nits}$$

But $\log_e p = \log_2 p / \log_2 e$

$\therefore$ $[-\log_2 p]\text{bits} = [-\log_2 p]/\log_2 e$ nits

Since $\log_2 e > 1$, obviously the nit is a bigger unit and they are related as follows:

$$1 \text{ nit} = \log_2 e \text{ bits}$$

Example 12.2 Find the information associated with the throwing of a fair die once.

Solution The die has six faces marked 1, 2, …, 5, 6. For a fair die, the probability $p_i = 1/6$ for $i = 1, 2, \ldots 6$.
$\therefore$ The information associated with the single throw of a die is

$$I(a_i) = -\log_2(p_i) = +\log_2\left(\frac{1}{1/6}\right) = \log_2 6$$

$$= (\log_{10} 6)/(\log_{10} 2) = \frac{0.77815}{0.3010} = 2.5852 \text{ bits}$$

Example 12.3 A book contains 400 pages with 450 words per page. Each word contains on the average, 6 symbols chosen at random from an alphabet of size 37 (26 letters, 10 digits from 0 to 9 and a blank space). Estimate the storage space in bits needed to store the information contained in the book on a compact disk. Assume that there is no statistical correlation between the symbols (including letters) and that all the symbols occur with equal probability.

Solution The choice of any one symbol from the alphabet set, when all the symbols occur with equal probability gives an amount of information

$$I(a_k) = \log_2\left(\frac{1}{(1/37)}\right) = \log_2 37 = \frac{\log_{10} 37}{\log_{10} 2} = \frac{1.5682}{0.3010} = 5.21 \text{ bits}$$

The total number of symbols chosen = $400 \times 450 \times 6$ = 1080 k bits
$\therefore$ the total amount of information contained in the book = $1080 \times 10^3 \times 5.21 = 5.6 \times 10^6$ bits
Hence, an amount of storage space needed = 5.6×10^6 bits

12.2.1 Sources

Before attempting to define the term entropy and finding an expression for the entropy of a given source, let us first see how a source may be conveniently represented or modeled. A source produces signals and as signals may be basically continuous time, or discrete time in nature, so are the sources too. The continuous-time signals like speech signals or video signals are, however, *essentially* band-limited though not *exactly* band limited. This is because the spectra of speech signals have very little power beyond about 4 kHz and similarly the spectra of video signals are mostly confined to about 6 MHz. Thus, these signals are, for all practical purposes, band limited and so can be recovered from their samples taken at the Nyquist rate, or above that. These information-bearing signals, being band limited, can as well be modeled by the samples of band limited random processes, i.e., as discrete random variables. Since the samples may have a continuum of values, we shall restrict our sources to emit discrete random variables X_i which take only a discrete set of values, are statistically independent and are identically distributed. Such sources are generally referred to as *Discrete Memoryless Sources* (DMS), in the sense that they emit discrete-time random processes that take only discrete amplitudes, in which all X_is are generated independently and with the same distribution. Hence, we shall hereafter consider our information sources to be Discrete Memoryless Sources (DMS), unless otherwise stated.

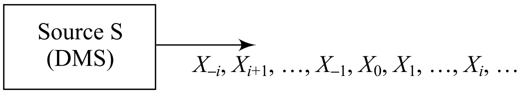


Fig. 12.1 *An information source S*

12.2.2 Entropy

Let us consider an arbitrary signaling interval. Let the source emit the symbol (discrete value) x_k with a probability p_k , $k = 0, 1, 2, \dots , (M - 1)$, where

$$\sum_{k=0}^{M-1} p_k = 1 \tag{12.2}$$

Thus, $I(x_k)$, the information produced by the source during that signaling interval, is itself a random variable which can take on the finite set of values $I(x_0), I(x_1), \dots, I(x_{M-1})$ with probabilities $p_0, p_1, p_2, \dots, p_{M-1}$. Since there are M discrete values or symbols in the alphabet S of the sources, the average information given by the source per signaling interval, or per symbol is

$$H(S) = E[I(x_k)]$$

$$= \sum_{k=0}^{M-1} p_k I(x_k) = -\sum_{k=0}^{M-1} p_k \log_2 p_k$$

$$\therefore \qquad \boxed{H(S) = -\sum_{k=0}^{M-1} p_k \log_2 p_k} \tag{12.3}$$

This quantity, $H(S)$, which represents the *average information per symbol* emitted by the DMS with a source alphabet S of size M, is called the '*entropy*' of the source.

Note *H(S) as used here is just a matter of notation and it does not represent a function of S. It should be read as 'entropy of the source S'.*

Example 12.4 A binary memoryless source produces the binary symbols 0 and 1 with probabilities p and $(1 - p)$ respectively. Determine the entropy of this source and sketch the variation of the entropy with the value of p.

Solution $H(S) = -\sum_{k=0}^{1} p_k \log_2 p_k = -p_0 \log_2 p_0 - p_1 \log_2 p_1$

where $p_0 = P(X = 0)$ and $p_1 = P(X = 1)$

since $P(X = 0) = p$, $p_0 = p$ and since $P(X = 1) = (1 - p)$, $p_1 = (1 - p)$

$$\therefore \qquad H(S) = -p_0 \log_2 p_0 - p_1 \log_2 p_1 = -p \log_2 p - (1 - p)\log_2(1 - p)$$

$$\therefore \qquad H(S) = -[p \log_2 p + (1 - p)\log_2(1 - p)] \text{ bits per symbol} \tag{12.3a}$$

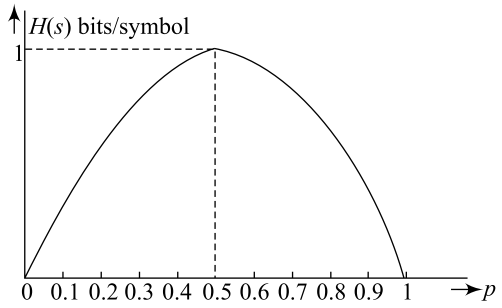


Fig. 12.2 *H(S) vs. p for a binary memoryless source*

Example 12.5 A source emits one of four symbols with probabilities $P_0 = 0.4$, $P_1 = 0.3$, $P_2 = 0.2$ and $P_3 = 0.1$. Find the amount of information gained by observing the source emitting each of these symbols.

(JNTU-K, Nov., 2010)

Solution Let the four symbols be x_0 with probability P_0, x_1 with probability P_1, x_2 with probability P_2 and x_3 with probability P_3. Then

Information gained by observing the source emitting x_0 =

$$I(x_0) = -\log_2 P_0 = -\log_2 0.4 = \log_2 2.5 = \frac{\log_{10} 2.5}{\log_{10} 2} = 1.32205 \text{ bits}$$

Information gained by observing the source emitting x_1 =

$$I(x_1) = -\log_2 0.3 = \log_2\left(\frac{1}{0.3}\right) = 1.73713 \text{ bits}$$

Similarly,

$$I(x_2) = -\log_2 0.2 = \log_2\left(\frac{1}{0.2}\right) = \log_2 5 = 2.32215 \text{ bits}$$

and

$$I(x_3) = -\log_2 0.1 = \log_2 10 = 3.32226 \text{ bits}$$

Assuming the symbols to be statistically independent, the average information gained per symbol

$$= 0.4 \times 1.32205 + 0.3 \times 1.73713 + 0.2 \times 2.32215 + 0.1 \times 3.32226$$

$\therefore$ $H(S) = 1.84661$ bits/symbol

Properties of entropy The entropy $H(S)$ of a discrete memoryless source is bounded as follows:

$$0 \le H(S) \le \log_2 M$$

where M is the size of the alphabet set.

Proof

1. The lower bound says that $H(S)$ is non-negative.

This follows immediately from the fact that

$$H(S) = -\sum_{k=0}^{M-1} p_k \log_2 p_k = \sum_{k=0}^{M-1} p_k \log_2(1/p_k)$$

Since p_k is a probability, it is always non-negative, and since p_k is less than 1, $(1/p_k)$ is greater than 1 and therefore $\log_2(1/p_k)$ will be non-negative. Since both p_k as well as $\log_2(1/p_k)$ are non-negative for all k, it follows that

$$H(S) = \sum_{k=0}^{M-1} p_k \log_2(1/p_k) \ge 0 \tag{12.4}$$

where the equality sign obviously holds good if and only if either $p_k = 0$ or 1. If $p_k = 1$ for some value of the index k, then $p_k = 0$ for all other values of k.

2. $H(S) \le \log_2 M$ implies that the highest value of the entropy of a DMS with an alphabet of size M is equal to $\log_2 M$ bits/symbol.

We shall prove this and also find out under what condition $H(S)$ takes this maximum value.

Let us try to find out for what values of p_ks the entropy $H(S)$ takes a maximum value. This is a constrained optimization problem, since $H(S)$ has to be maximized under the constraint that

$$\sum_{k=0}^{M-1} p_k = 1 \tag{12.5}$$

This constrained optimization can be carried out using Lagrange multiplier method that maximizes the expression (Refer to Appendix F)

$$J=\sum_{k=0}^{M-1} p_k\log_2 p_k-\lambda\left[\left\{\sum_{k=0}^{M-1} p_k\right\}-1\right] \tag{12.6}$$

where λ is the Lagrange multiplier, an undetermined constant.

Differentiating J with respect to p_j and equating the result to zero, we get the following M equations:

$$\frac{\partial J}{\partial p_j}=-\log p_j-(1/\ln 2)-\lambda=0 \tag{12.7}$$

$$j=0,1,2,\ldots,(M-1)$$

Equation (12.7) implies that $-\log p_j$ and hence p_j is a constant for all values of j, i.e., $j = 0, 1, 2, \ldots, (M-1)$. If $p_j = c$ for all j, then from Eq. (12.5), the constraint equation, it follows that

$$p_j=\frac{1}{M}\text{ for } j=0,1,2,\ldots,(M-1) \tag{12.8}$$

Thus, the maximum value of $H(S)$ is obtained when all the symbols are equally probable. Obviously, this maximum value is

$$\underset{\max}{H(S)}=\left[-\sum_{j=0}^{M-1} p_j\log_2 p_j\right]\Bigg|_{p_j=1/M}=-\sum_{j=0}^{M-1}(1/M)\log_2(1/M)$$

$$=+\sum_{j=0}^{M-1}\left(\frac{1}{M}\right)\log_2 M=\log_2 M$$

$\therefore$ the maximum value of $H(S)$ is $\log_2 M$ bits/symbol and this occurs when all the M symbols are equally probable.

$\therefore$
$$0\le H(S)\le\log_2 M \tag{12.9}$$

12.2.3 Extended Sources

Suppose a DMS, S, has an alphabet of size M. Instead of individual symbols given out by the source, suppose we consider blocks of such symbols, each block consisting of n symbols. Then, we may consider that a new source, called the extended source, is emitting such blocks as its symbols. Obviously, the alphabet size for the extended source is M^n since there will be that many *distinct blocks* that can be formed from the alphabet of the original source. Since the original source is a discrete memoryless source, its symbols (M in number) are statistically independent. If we denote the extended source by S^n, each of its symbols, consisting of n symbols of the original source, will have a probability that is the product of the probabilities of the n symbols of the original source constituting it. The entropy of the source S^n can therefore be expressed to be n times the entropy of S.

i.e.,
$$H(S^n)=nH(S) \tag{12.10}$$

Example 12.6 A discrete memoryless source S, has an alphabet $\{s_0, s_1\}$ with probabilities $P[s_0] = p_0 = 1/4$ and $P[s_1] = p_1 = ¾$. Find the entropies of the source S and the extended source S^3.

Solution

(a) Entropy of the source S

$$H(S)=-[p_0\log_2 p_0-p_1\log_2 p_1]=-\left[\frac{1}{4}\log_2(1/4)-\frac{3}{4}\log_2(3/4)\right]$$

$$=\frac{1}{4}\log_2 4+\frac{3}{4}\log_2(4/3)=\frac{1}{4}\times 2+\frac{3}{4}\left[\frac{\log_{10}(4/3)}{\log_{10} 2}\right]$$

$$=\frac{1}{2}+\frac{3}{4}\left[\frac{0.12492}{0.3010}\right]=\frac{1}{2}+\frac{3}{4}[0.415042]=0.8112 \text{ bits/symbol}$$

(b) S^3 will have $2^3 = 8$ distinct symbols. If we denote these 8 symbols as $x_0, x_1, x_2, x_3, \dots, x_7$ then

Symbol S^3	**Composition**	**Probability $p[x_i]$**	**Symbol S^3**	**Composition**	**Probability $p[x_i]$**
x_0	s_0, s_0, s_0	1/64	x_4	s_1, s_0, s_0	3/64
x_1	s_0, s_0, s_1	3/64	x_5	s_1, s_1, s_0	9/64
x_2	s_0, s_1, s_1	9/64	x_6	s_0, s_1, s_0	3/64
x_3	s_1, s_1, s_1	27/64	x_7	s_1, s_0, s_1	9/64

$\therefore$ entropy of the extended source $S^3 = -\sum_{i=0}^{7} P[x_i]\log_2 P[x_i]$

$$=-\left[\frac{1}{64}\log_2\left(\frac{1}{64}\right)+\frac{3}{64}\log_2\left(\frac{3}{64}\right)+\frac{9}{64}\log_2\left(\frac{9}{64}\right)+\frac{27}{64}\log_2\left(\frac{27}{64}\right)\right.$$

$$\left.+\frac{3}{64}\log_2\left(\frac{3}{64}\right)+\frac{9}{64}\log_2\left(\frac{9}{64}\right)+\frac{3}{64}\log_2\left(\frac{3}{64}\right)+\frac{9}{64}\log_2\left(\frac{9}{64}\right)\right]$$

$$= 2.433 \text{ bits/symbol}$$

This is exactly three times the entropy of S.

12.3 JOINT AND CONDITIONAL ENTROPIES

1. Joint entropy: In the last section, we had modeled a discrete memoryless source as one that emits an independent discrete random variable $\boldsymbol{X}$ during each signaling interval, where each of these random variables can take a finite set of discrete values $x_0, x_1, x_2, \dots, x_{M-1}$ with probabilities $p_0, p_1, p_2, \dots, p_{M-1}$, respectively. We then defined the entropy of such a source as the average information given by each such discrete random variable, $\boldsymbol{X}$ on the basis of the uncertainty associated with it, since $\boldsymbol{X}$ may take any one of the values $x_0, x_1, x_2, \dots, x_{M-1}$ with the specified probability p_i for each x_i, $i = 0, 1, \dots, M-1$.

We shall now extend this concept of entropy to a situation wherein we have more than one random variable. Let us consider two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, where, $\boldsymbol{X}$ can take any one of the M possible values $x_0, x_1, x_2, \dots, x_{M-1}$ with probabilities p_i, $i = 0, 1, \dots, M-1$ respectively and $\boldsymbol{Y}$ can take any one of the L possible values $y_0, y_1, y_2, \dots, y_{L-1}$, with probabilities p_j, $j = 0, 1, 2, \dots, L-1$, respectively. $\boldsymbol{X}$ and $\boldsymbol{Y}$ may be visualized as being produced by two separate sources, which are not necessarily independent. For example, $\boldsymbol{X}$ may be the random variable at the input of a noisy channel, while $\boldsymbol{Y}$ may be that at the output of the same channel. Now, let

$$p(x_i, y_j) \triangleq P\lfloor \boldsymbol{X} = x_i \text{ and } \boldsymbol{Y} = y_j \rfloor = p_{i,j} \tag{12.11}$$

with $i = 0, 1, 2, \dots, (M-1)$ and $j = 0, 1, 2, \dots, (L-1)$

Thus, with $\boldsymbol{X}$ taking any one of M possible values and $\boldsymbol{Y}$ taking any one of L possible values, our random experiment, with the pairs of observed values of $\boldsymbol{X}$ and $\boldsymbol{Y}$ as the outcomes will have a total of ML possible outcomes. Since the probability of the joint occurrence of the outcomes x_i and y_j is $p(x_i, y_j)$, we should have

$$\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j) = 1 \tag{12.12}$$

Further, the amount of information associated with this joint event of $\boldsymbol{X}$ taking the value x_i and $\boldsymbol{Y}$ taking the value y_j, is

$$I(\boldsymbol{X} = x_i, \boldsymbol{Y} = y_j) = -p(x_i, y_j)\log_2 p(x_i, y_j) \tag{12.13}$$

Therefore, we may define the *joint entropy* of these two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$H(\boldsymbol{X},\boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i, y_j) \tag{12.14}$$

Although only two random variables, ***X*** and ***Y*** have been considered in the above discussion, we may generalize Eq. (12.14) to the case of *N* random variables, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ and say that *their joint entropy* is

$$H(\boldsymbol{X}_1, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_N) = -\sum_{x_1, x_2, \ldots, x_N}^{M-1} p(x_1, x_2, \ldots, x_N)\log_2 p(x_1, x_2, \ldots, x_N) \tag{12.15a}$$

From Eq. (12.14), it is clear that $H(\boldsymbol{X},\boldsymbol{Y}) = H(\boldsymbol{Y},\boldsymbol{X})$ (12.15b)

2. Conditional entropy: With the random variables ***X*** and ***Y*** used in our discussion above on joint entropy, we shall now define the conditional uncertainty of ***Y*** given that $\boldsymbol{X} = x_i$ as:

$$H(\boldsymbol{Y}/\boldsymbol{X} = x_i) = -\sum_{j=0}^{L-1} p(y_j|x_i)\log_2 p(y_j|x_i) \tag{12.16}$$

where $p(y_j|x_i)$ is the conditional probability of ***Y*** taking the value y_j given that ***X*** has taken the value x_i. For example, suppose ***X*** is a random variable at the input of a *noisy channel* and can take either of the values $x_0 = 0$ or $x_1 = 1$; and ***Y*** is the random variable that corresponds to the output of the channel and can either of the values $y_0 = 0$ or $y_1 = 1$. In such a situation, we will be interested in knowing $P[\boldsymbol{Y} = 1 \text{ when } \boldsymbol{X} = x_0 = 0]$, i.e., the probability of channel output being a 1 even though its input was a 0, i.e., $p(y_1|x_0)$. Of course, the conditional probability $p(y_j|x_i)$ must satisfy the condition

$$\sum_{j=0}^{L-1} p(y_j|x_i) = 1 \tag{12.17}$$

The average of $H(\boldsymbol{Y}|\boldsymbol{X} = x_i)$ given in Eq. (12.16), will therefore give us the conditional entropy of ***Y*** given ***X***, i.e.,

$$H(\boldsymbol{Y}/\boldsymbol{X}) \underline{\Delta} \sum_{i=0}^{M-1} p(x_i) H(\boldsymbol{Y}/\boldsymbol{X} = x_i)$$

$$-\sum_{i=0}^{M-1} p(x_i)\left[\sum_{j=0}^{L-1} p(y_j|x_i)\log_2 p(y_j|x_i)\right] \tag{12.18}$$

Making use of the result that the joint probability of x_i and y_j is

$$p(x_i, y_j) = p(x_i)\, p(y_j|x_i) \tag{12.19}$$

we get

$$H(\boldsymbol{Y}/\boldsymbol{X}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(y_j|x_i) \tag{12.20}$$

where $H(\boldsymbol{Y}/\boldsymbol{X})$ represents the conditional entropy of ***Y*** given ***X***.

We are now in a position to derive an important and useful result that relates the joint entropy of two random variables ***X*** and ***Y*** with their individual entropies and conditional entropies. This result says

$$H(\boldsymbol{Y},\boldsymbol{X}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}/\boldsymbol{X})$$

$$= H(\boldsymbol{Y}) + H(\boldsymbol{X}/\boldsymbol{Y})$$

Proof $H(\boldsymbol{X},\boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i, y_j)$

$$= -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)[\log p(x_i)\, p(y_j|x_i)]$$

$$= -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log p(x_i) - \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log p(y_j|x_i)$$

$$= -\sum_{i=0}^{M-1} p(x_i)\log p(x_i) + H(\boldsymbol{Y}|\boldsymbol{X})$$

where $H(\boldsymbol{Y}/\boldsymbol{X})$ is used to represent the conditional entropy and is given by

$$H(\boldsymbol{Y}|\boldsymbol{X}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log p(y_j|x_i) \tag{12.21}$$

$\therefore$ $$H(\boldsymbol{X},\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}/\boldsymbol{X})$$

Similarly
$$\left.\begin{aligned} H(\boldsymbol{X},\boldsymbol{Y}) &= H(\boldsymbol{X}) + H(\boldsymbol{Y}/\boldsymbol{X}) \\ H(\boldsymbol{X},\boldsymbol{Y}) &= H(\boldsymbol{Y}) + H(\boldsymbol{X}/\boldsymbol{Y}) \end{aligned}\right\} \tag{12.22}$$

If ***X*** and ***Y*** are statistically independent random variables the above equations reduce to

$$H(\boldsymbol{X},\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}) \tag{12.23}$$

The above equation is fully in tune with what we feel intuitively, viz., that when two sources are totally independent, the joint entropy of the two must be equal to the sum of their marginal (or individual) entropies.

Example 12.7 Determine the average information content in bits associated with the tossing of a pair dice. Assume that a given pair of numbers is regarded as a distinct symbol regardless of which die shows up which number.

Solution As there are 6 faces for each die, there are 36 pairs possible altogether. Each of these can occur in two ways if we do not bother about which die has shown up which number.

$\therefore$ probability of any given pair of numbers = $2 \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{18}$

$\therefore$ information obtained whenever any pair of numbers shows up = $-\log_2 \frac{1}{18} = \log_2 18$

Average information = $18 \times \frac{1}{18}\log_2 18$ bits $= 4.1703$ bits/ pair of numbers

Example 12.8 A source produces three symbols, A, B and C with the following marginal and conditional probabilities:

i	*p(i)*
A	1/4
B	1/4
C	1/2

		j		
	p(*j* | *i*)	A	B	C
i	A	1/8	1/4	5/8
	B	1/2	1/8	3/8
	C	3/8	5/8	0

(a) Assuming that there is no inter-symbol influence, calculate the entropy of the source.
(b) If index *i* refers to ***X*** and index *j* refers to ***Y***, determine the conditional entropy $H(\boldsymbol{Y}/\boldsymbol{X})$

Solution

(a) When the symbols are emitted independent of each other, i.e., when there is no inter-symbol influence,

$$H(\boldsymbol{X}) = \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 4 + \frac{1}{2}\log_2 2 = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = 1.5 \text{ bits/symbol}$$

(b) When inter-symbol influence exists with the conditional probabilities as given in the table on the right-hand side:

We know from Eq. (12.21) that the conditional entropy is

$$H(\boldsymbol{Y}/\boldsymbol{X}) = -\sum_{i=0}^{2}\sum_{j=0}^{2} p(x_i, y_j)\log_2 p(y_j|x_i)$$

using the marginal probabilities given in the table on the left-hand side and the conditional probabilities in the second table, we have to first calculate the joint probabilities $p(x_i, y_j)$s for $i = 0, 1, 2,$ and $j = 0, 1, 2$ using the relation:

$$p(x_i, y_j) = p(y_j|x_i)p(x_i)$$

$$p(x_0, y_0) = \frac{1}{32}; \qquad p(x_0, y_1) = \frac{1}{16}; \qquad p(x_0, y_2) = \frac{5}{32}$$

$$p(x_1, y_0) = \frac{1}{8}; \qquad p(x_1, y_1) = \frac{1}{32}; \qquad p(x_1, y_2) = \frac{3}{32}$$

$$p(x_2, y_0) = \frac{3}{16}; \qquad p(x_2, y_1) = \frac{5}{16}; \qquad p(x_2, y_2) = 0$$

$$\therefore \qquad H(\boldsymbol{Y}|\boldsymbol{X}) = \frac{1}{32}\log_2 8 + \frac{1}{8}\log_2 2 + \frac{3}{16}\log_2 8/3 + \frac{1}{16}\log_2 4 + \frac{1}{32}\log_2 8 + \frac{5}{16}\log_2 8/5 + \frac{5}{32}\log_2 8/5 + \frac{3}{32}\log_2 8/3 + 0$$

$$\therefore \qquad H(\boldsymbol{Y}|\boldsymbol{X}) = 1.1534 \text{ bits/symbol}$$

Alternatively, one may calculate $H(\boldsymbol{X}, \boldsymbol{Y})$ using the joint probabilities and then subtract $H(\boldsymbol{X})$ from it so as to get $H(\boldsymbol{Y}/\boldsymbol{X})$.

$$\therefore \qquad H(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{32}\log_2 32 + \frac{1}{16}\log_2 16 + \frac{5}{32}\log_2 32/5 + \frac{1}{8}\log_2 8 + \frac{1}{32}\log_2 32 + \frac{3}{32}\log_2 32/3 + \frac{3}{16}\log_2 16/3 + \frac{5}{16}\log_2 16/5 + 0$$

$\therefore$ $H(\boldsymbol{X}, \boldsymbol{Y}) = 2.6538$. Hence $H(\boldsymbol{Y}|\boldsymbol{X}) = H(\boldsymbol{X}, \boldsymbol{Y}) - H(\boldsymbol{X}) = 1.1533$ bits/symbol

12.4 SOURCE CODING AND SHANNON'S THEOREM

A discrete source produces symbols. These symbols may have to be represented in such a way that it would be possible to transmit them over a given channel. For example, if the channel is a binary channel, i.e., one that accepts binary symbols 0 and 1, the source output, which is in the form of a sequence of source symbols, must be converted into a sequence of what are called the '*code elements*', which in this case are the binary symbols. This process is called '*encoding*' and the device, or system which performs this encoding is called an encoder. The encoder assigns a unique sequence of code elements, called a '*codeword*', for representing each source symbol.

The objective of source coding is to remove or reduce the redundancy in the source output so as to give an efficient representation of the message information given by the source by using less number of bits.

The encoder must do its job efficiently; otherwise, it will be wasting the precious communication resources – bandwidth and power. For it to be efficient, it must have prior knowledge of the probability of occurrence of each source symbol. It must give the shortest (in terms of the number of code elements used) codeword for

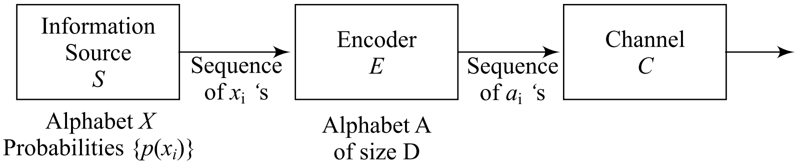


Fig. 12.3 *Encoding process*

the most frequently occurring source symbol and the longest codeword for the least frequent source symbol. Even in the early 1830s, long before Shannon's information theory came into existence, the genius of Samuel Morse, the inventor of telegraphy, realized the importance of this feature of source coding and had, in his famous telegraph code, allotted the shortest symbol, a mere 'dot' for the most frequent letter, E, of the English alphabet, and the longest symbol, 'dash dash dot dot' for the least frequent letter Q. The encoder must also ensure that the encoded version is uniquely decipherable so that it will be possible to get back the original version (source symbol sequence) unambiguously. Thus, the two basic requirements to be met by any source encoder are:

1. Minimum average length of a codeword for a given set of source alphabet $\{\boldsymbol{X}\}$ and the source symbol probability set $\{p(x_i)\}$.
2. Unique decipherability of the encoded sequence.

Definition Codes having the two properties stated above, are called '*Optimal codes*'.

Before proceeding further we shall examine the meaning of the two requirements stated above in a little more detail.

Average length of a code For fast transmission of any message from the source to the destination, it is necessary that the average length of a code is as small as possible. The average length $\bar{n}$ of a code is defined as

$$\bar{n} \underline{\Delta} \sum_{i=0}^{M-1} n_i p(x_i) \tag{12.24}$$

where n_i is the length of the codeword corresponding to the symbol x_i, which occurs with a probability $p(x_i)$.

12.4.1 Unique Decipherability of a Coded Sequence

Even though each codeword may be distinct, sequences of codewords coming out from the encoder may not have this property. As an example, let us examine the two codes given in Table 12.1.

Table 12.1

Source Symbols	Code *A*	Code *B*	Code *C*	Code *D*
x_0	0 0	0	0	0
x_1	0 1	0 1	0 1	1 0
x_2	1 0	0 1 0	0 1 1	1 1 0
x_3	1 1	1 1 0	0 1 1 1	1 1 1

Codes *A* and *B* both have distinct codewords for the four source symbols. Code *A* is a simple binary coding with all the codewords having equal length. When this code is used, unique decipherability of a coded sequence is guaranteed since the codeword boundaries are fixed and the codewords themselves are distinct. However, if the probabilities of occurrence of the four source symbols are *not equal*, then it may be preferable to use a code with codewords of unequal length, the shortest codeword being assigned to the most probable source symbol and the longest codeword to the least probable source symbol. Code *B* is such a code. However, it does not lead to unique decipherability since a code sequence like **0 1 0** can be decoded either as x_2 or as $x_1\, x_0$.

It is not difficult to find the reason for the ambiguity arising in the decoding of the encoded sequence. The problem is caused because the codeword 0 1 for the source symbol x_1 happens to be a '*prefix*' for the codeword 0 1 0 assigned to the source symbol x_2. So, we should avoid a situation wherein one codeword is a prefix to another codeword. It may be noted, however, that this is not a necessary condition, as can be seen from the fact that code *C* of the table is uniquely decipherable, although many of the codewords are prefixes of other codewords. Unique deciphering of the encoded sequences (using this code) is possible by subdividing the sequence of 0s and 1s to the left of every 0. Thus, the first element of every codeword, viz., '0', acts as a sort of a 'comma' between one codeword and another. Such codes are referred to as '*comma codes*'.

Definition Codes in which no codeword is a prefix to another codeword, are called '*instantaneous codes*', or '*prefix-free codes*'.

Code *D* of the table is an example of an instantaneous or prefix-free code.

Since we are interested in codes with minimum average codeword length and having unique decipherability property, i.e., in optimal codes, we now give, without proof, a useful theorem by McMillan and Karush.

Theorem 12.1 If for a given source *S*, a code is optimal among the instantaneous codes, then it is optimal among all uniquely decipherable codes.

From the above theorem, it is clear that we can limit our search for optimal codes only to the set of instantaneous codes.

12.4.2 Kraft's Inequality (also known as Kraft–McMillan Inequality)

For a code with codewords of unequal lengths, the requirement of unique decipherability places certain constraints on its structure, i.e., on its codeword lengths and the number of codewords. Kraft's inequality spells out this constraint. It states that a necessary and sufficient condition for the existence of an instantaneous code having word lengths of $n_0, n_1, n_2, \ldots, n_{M-1}$ is given by

$$\sum_{i=0}^{M-1} D^{-n_i} \leq 1 \tag{12.25}$$

where *D* is the size of the encoder's alphabet, i.e., the number of symbols comprising the code alphabet (usually 2 since binary symbols 0 and 1 are generally used as code elements).

Conversely, if integers $n_0, n_1, n_2, \ldots, n_{M-1}$ satisfy the condition in Eq. (12.25), a prefix-free or instantaneous code can be found, whose word lengths are given by $n_0, n_1, n_2, \ldots, n_{M-1}$.

We shall now discuss two theorems which fix the lower and upper bounds for the length of the codeword of a code in terms of the entropy of the source and the size of the code alphabet.

Theorem 12.2 Given a source with alphabet $x_0, x_1, x_2, \ldots, x_{M-1}$ with probabilities $p(x_0), p(x_1), \ldots, p(x_{M-1})$, the average length *of a uniquely decipherable code* with alphabet size D satisfies the inequality:

$$\bar{n} \geq \frac{H(S)}{\log D} \tag{12.26}$$

where *H*(*S*) is the entropy of the source (information/symbol) *D* is the size of the code alphabet (generally 2) and the base of the logarithm is arbitrary, but generally taken as 2 since information units used are generally bits. If *D* = 2 and the base for the logarithm is taken as 2, $\log_2 D = 1$.

Proof $H(S) - \bar{n} \log D$

$$= -\sum_{i=0}^{M-1} p(x_i) \log p(x_i) - \sum_{i=0}^{M-1} p(x_i) n_i \log D$$

$$= \sum_{i=0}^{M-1} p(x_i) \log \left[\frac{D^{-n_i}}{p(x_i)} \right] \le \sum_{i=0}^{M-1} p(x_i) \left[\frac{D^{-n_i}}{p(x_i)} - 1 \right] \tag{12.27}$$

(since $\log y \le (y-1)$)

$$\therefore \qquad H(S) - \bar{n} \log D \le \left\{ \left[\sum_{i=0}^{M-1} D^{-n_i} \right] - \sum_{i=0}^{M-1} p(x_i) \right\} = \left[\sum_{i=0}^{M-1} D^{-n_i} \right] - 1 \tag{12.28}$$

Since the code is uniquely decipherable, Kraft–McMillan inequality must be satisfied. That is

$$\sum_{i=0}^{M-1} D^{-n_i} \le 1$$

$\therefore$ from Eq. (12.28), we have

$$H(S) - \bar{n} \log D \le 0 \tag{12.29}$$

This implies that

$$\bar{n} \ge \frac{H(S)}{\log D} \tag{12.30}$$

Remarks

(i) *Since $\log y = (y-1)$ if and only if $y = 1$, from Eq. (12.27), it follows that in Eq. (12.30), the equality $\bar{n} = H(S)/\log D$ holds good, if and only if, $p(x_i) = D^{-n_i}$ for all i, i.e., iff $-\log_D p(x_i)$ is an integer for all i.*

(ii) *Theorem 12.2 tells us that one can have*

$$\bar{n} \log D = H(S) = \text{Entropy of the source} \tag{12.30a}$$

iff

$$p(x_i) = D^{-n_i} \text{ for } i = 0, 1, 2, \ldots, (M-1) \tag{12.30b}$$

Suppose, for convenience, we choose $D = 2$, i.e., our prefix-free code uses binary alphabet 0 and 1. Also, since we always use 'bits' as units of information, let the base for the logarithm be 2. Then $\log_2 D = 1$ and so we get from Eq. (12.30a) that

$$\left.\begin{array}{l}\text{Average number of binary}\\ \text{digits in a codeword}\end{array}\right\} = \bar{n} = H(S) = \left\{\begin{array}{l}\text{Entropy of the source}\\ \text{in bits per symbol}\end{array}\right.$$

This implies perfect matching between the source and the prefix-free code. Thus, Theorem 12.2 says that it is possible to construct a prefix-free code that perfectly matches with the source if and only if $p(x_i) = D^{-n_i}$ for all i. Noting that $p(x_i)$ depends on the nature of the source, the question arises: If an arbitrary source is given for which $-\log_2 p(x_i)$ is not an integer for all i, (i.e., if condition (12.30b) is not satisfied), how do we match the prefix-free code to such a source?

As we shall see Theorem 12.3, provides an answer to this question.

Theorem 12.3 For a source with alphabet $\{x_0, x_1, x_2, \ldots, x_{M-1}\}$ having probabilities $\{p(x_0), p(x_1), \ldots, p(x_{M-1})\}$, it is possible to construct an instantaneous code using a code alphabet of size D in such a way that

$$\bar{n} < \frac{H(S)}{\log D} + 1 \tag{12.31}$$

Proof For the source symbol x_i, let us choose a codeword of length n_i given by

$$n_i = \lceil -\log_D p(x_i) \rceil \tag{12.32}$$

where $\lceil z \rceil$ denotes the smallest integer greater than or equal to z.

But $$n_i = \lceil -\log_D p(x_i) \rceil \text{ implies that}$$

$$n_i \geq -\log_D p(x_i) \Rightarrow D^{-n_i} \leq p(x_i)$$

$\therefore$ $$\sum_{i=0}^{M-1} D^{-n_i} \leq \sum_{i=0}^{M-1} p(x_i) = 1 \tag{12.33}$$

Equation (12.33) means that the code under consideration satisfies Kraft–McMillan's inequality (see Eq. (12.25)). Hence a prefix-free, or instantaneous code can be found with its codewords having lengths n_0, n_1, n_2, …, n_{M-1}.

Further,

$$n_i = \lceil -\log_D p(x_i) \rceil \Rightarrow n_i < -\log_D p(x_i) + 1$$

$\therefore$ $$\sum_{i=0}^{M-1} p(x_i) n_i < -\sum_{i=0}^{M-1} p(x_i) \log_D p(x_i) + \sum_{i=0}^{M-1} p(x_i)$$

i.e.,

$$\overline{n} < \frac{H(S)}{\log D} + 1 \tag{12.34}$$

This specifies the upper bound for $\overline{n}$, the average length of the codeword while Eq. (12.30) specifies the lower bound. Let us now see how we may approach this lower bound for $\overline{n}$.

Till now, we have been considering that each individual symbol of the source *S* is encoded separately. But suppose that we encode each block of *N* symbols generated by an N^{th} order extension of the original source *S*. Then this extended source will have an entropy of *NH*(*S*) where *H*(*S*) is the entropy of the original source *S* (see Eq. (12.10)). Further, the average codeword length will now be $N\overline{n}$. Therefore, applying Theorems 12.2 and 12.3 to this N^{th} order extension of the original source,

$$\frac{NH(S)}{\log D} \leq N\,\overline{n} < \frac{NH(S)}{\log D} + 1 \tag{12.35}$$

i.e., $$\frac{H(S)}{\log D} \leq \overline{n} < \frac{H(S)}{\log D} + \frac{1}{N} \tag{12.36}$$

Thus, by increasing *N*, the block length, i.e., the order of extension, $\overline{n} \log D$ can be made arbitrarily close to *H*(*s*), the entropy of the source.

Remarks

(i) By making N larger and larger, we are not changing the lower bound on $\overline{n}$. We are only approaching it more and more closely.

(ii) Increasing N increases the encoder complexity as well as the size of the buffer required for storing the x_is.

(iii) Since D represents the number of code symbols in the code alphabet, $\log_2 D$ represents the number of bits per code symbol, and $\overline{n} \log_2 D$ represents the number of bits of information in a codeword of length $\overline{n}$, the average length of a codeword.

(iv) If the size of code alphabet, i.e., D = 2 (as in the case of binary codes) and if the base of the logarithm, which is arbitrarily, chosen to be 2, then $\log_2 D = 1$ and so we may write Eq. (12.36) as

$$H(S) \leq \overline{n} < H(S) + \in$$

where, $\in = 1/N$, can be made arbitrarily small by choosing N very large.

Hence, Shannon's source coding theorem is generally stated as follows:

12.4.3 Shannon's Source Coding Theorem

For a discrete memoryless source with entropy $H(S)$, the minimum value of $\overline{n}$, the average length of a codeword of a perfectly decipherable code, is bounded by

$$H(S) \leq \overline{n} < H(S) + \in \tag{12.37}$$

where $\in$ can be made arbitrarily small by appropriate coding.

In the introduction to source coding, we had stated that the aim of source coding is to represent the information from the source in the most economical manner possible, i.e., by using the smallest number of *code elements* per source symbol. Shannon's source coding theorem states that the maximum extent to which the source data can be compressed using a code, is limited by the fact that the *average length (in bits) of a codeword cannot be less than the average information in bits per source symbol*, i.e., the entropy $H(s)$ of the source. The theorem merely states this bound, but does not tell us how we may design a code that permits us to attain that bound. To what extent a given code is able to reach the bound, will therefore tell us how efficient the code is.

The efficiency of a source code is therefore defined by

$$\text{Coding efficiency} \underline{\Delta} \frac{H(S) = \overline{n}_{\text{min}}}{\text{Actual } \overline{n} \text{ of the code}} \tag{12.38}$$

12.5 SOURCE CODING FOR DISCRETE MEMORYLESS SOURCES

The output of a physical source, in its original form, generally contains lot of redundancy and directly transmitting it as it is, will result in wastage if time, bandwidth and power. The objective of source coding is to remove that redundancy and make the data more compact. The ultimate limit for data compaction without loss of any information is set, as we have already seen from Shannon's source coding theorems, by the entropy of the source.

In the following sections, we will discuss a few prefix-free coding schemes for discrete memoryless sources, called compact coding schemes, which are suboptimum, in the sense that, they give, on the average, longer codewords than the optimum value which is $H(S)$, the entropy of the source. So these codes will be uniquely decipherable, but *not necessarily* optimal.

12.5.1 Shannon–Fano Coding Scheme

This code may be constructed as per the following algorithm:

1. Write down the message or source symbols in the order of decreasing probabilities.
2. Draw a line after say the k^{th} symbol such that the total probability of the symbols above the line and below the line are approximately equal, i.e., divide the source symbols into two groups of almost equal probability. To each symbol above the line, assign a '0' and to each symbol below the line, assign a '1'.
3. Apply step 2 to each of the groups formed and continue the process till all the subgroups have only one symbol. When that stage is reached, the coding is complete.

We now illustrate the Shannon-Fano coding scheme by a few examples.

Example 12.9 A source is producing sequences of independent symbols A, B, C and D with the following probabilities: $A = 0.5$, $B = 0.25$, $C = 0.125$ and $D = 0.125$.

(a) Devise an unambiguous binary code for the output of this source.
(b) Compute the coding efficiency of your code.

Solution

(a) Devising an unambiguous code.

Symbol x_i	Probability $p(x_i)$	Steps			Symbol code
		1	2	3	
A	0.5	**0**			0
B	0.25	1	**0**		1 0
C	0.125	1	1	**0**	1 1 0
D	0.125	1	1	1	1 1 1

(b) Source entropy $= H(S) = 0.5 \log_2 2 + 0.25 \log_2 4 + 0.125 \log_2 8 + 0.125 \log_2 8$

$\therefore \quad H(S) = 1.750$ bits/symbol

Average value of codeword length $= \bar{n} = \sum_{i=0}^{3} p(x_i) n_i$

$$= 0.5 \times 1 + 0.25 \times 2 + 0.125 \times 3 + 0.125 \times 3 = 1.750 \text{ bits/symbol}$$

$$\therefore \quad \text{coding efficiency} = \frac{1.750}{1.750} \times 100 = 100\%$$

Remark

One may wonder how this suboptimum coding scheme could give a code with a coding efficiency of 100%. The given p(x$_i$*)s are such that* $p(x_i) = D^{-n_i}$ *for each i. Hence, as stated in point i. Remarks under Eq. (12.30), the equality sign holds in Eq. (12.30) and* $\bar{n} = \dfrac{H(s)}{log\ D} = H(s)$ *for binary code.*

Example 12.10 A source is producing sequences of independent symbols A, B, C, D and E, with the following probabilities:

$A = 1/2$, $B = 1/6$, $C = 1/12$, $D = 1/6$, $E = 1/12$

(a) Devise an unambiguous binary code for these symbols.

(b) Compute the coding efficiency of your code.

Solution

(a)

Symbol x_i	Probability $p(x_i)$	Steps				Symbol code
		1	2	3	4	
A	1/2	**0**				0
B	1/6	1	**0**			1 0
D	1/6	1	1	**0**		1 1 0
C	1/12	1	1	1	**0**	1 1 1 0
E	1/12	1	1	1	1	1 1 1 1

(b) $$H(S) = -\left[\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{6}\log_2 \frac{1}{6} + \frac{1}{12}\log_2 \frac{1}{12} + \frac{1}{6}\log_2 \frac{1}{6} + \frac{1}{12}\log_2 \frac{1}{12}\right]$$

$$= 1.9591 \text{ bits/symbol}$$

$$\bar{n} = \text{Average length of codeword} = \sum_{i=0}^{M-1} p(x_i)n_i$$

$$= \frac{1}{2}\times 1 + \frac{1}{6}\times 2 + \frac{1}{6}\times 3 + \frac{1}{12}\times 4 + \frac{1}{12}\times 4$$

$$= 1.999 \text{ bits/symbol}$$

$$\therefore \quad \text{coding efficiency} = \frac{H(S)}{\bar{n}} = \frac{1.9591}{1.999} = 0.9800 = 98.00\%$$

12.5.2 Huffman Coding Scheme

This is yet another compact coding scheme that is suboptimal. The following steps describe this coding algorithm:

Step 1: Rearrange the source symbols in the order of decreasing probability.

Step 2: Assign 0 and 1 to the two symbols of lowest probability. This forms stage 1.

Step 3: Combine the last two symbols into one new symbol with probability equal to the sum of the probabilities of the two original symbols. List the probabilities of the original symbols (except the last two) and the new symbol in decreasing order.

Step 4: This process adopted in step 3 is to be repeated till we are left with only two symbol statistics (i.e., probabilities) to which a 0 and a 1 are assigned. This forms the last stage.

The code for each original source symbol is then obtained by tracing out the sequence of 0s and 1s which we have to go through when we work backwards to arrive at the original source symbol.

The following examples clearly illustrate the procedure that is to be followed.

Example 12.11 A discrete memoryless source produces symbols x_i, $i = 0$ to 5 with the following probabilities: $p(x_0) = 0.1$; $p(x_1) = 0.2$; $p(x_2) = 0.15$; $p(x_3) = 0.09$; $p(x_4) = 0.20$; $p(x_5) = 0.26$. Design a Huffman code for the above source. Find the coding efficiency of your code.

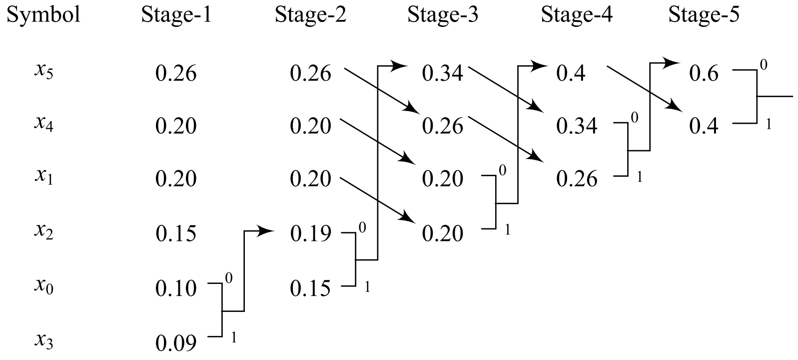
Solution

(a) We first arrange the source symbols in decreasing order of probability as shown in the following table.

(b) We club the probabilities of the last two symbols and put their sum 0.19 at the appropriate level in the decreasing order of $p_i s$ in stage 2. Assign 0 to probability 0.1 and 1 to the probability 0.09 in stage 1.

(c) The two lowest probabilities 0.19 and 0.15 are assigned 0 and 1 respectively and their sum, viz., 0.34 is taken to the top in stage 3, where it is the highest probability. The rest of the probabilities of stage 2 are arranged at appropriate places in stage 3.

This process of assigning 0 and 1 to the lowest two probabilities, clubbing them and taking their sum to an appropriate level in the next stage, is continued. When we come to stage 5, there are only two probabilities and their sum, as is to be expected, is 1. So, these two are assigned 0 and 1 as shown. The process ends there.



Now, to find the codeword for each source symbol, we illustrate for the source symbol x_3 which has a probability of 0.09. Starting at 0.09, trace the path from one stage to the next, till the last stage is reached, noting down the 0s and 1s in its path. It runs as 1 – 0 – 0 – 0. Now, reverse this to get the codeword for x_3 as 0 0 0 1. Proceeding in a similar way, we have

Source Symbol	Codeword
x_0	0 0 0 0
x_1	1 1
x_2	0 0 1
x_3	0 0 0 1
x_4	1 0
x_5	0 1

From the given probabilities for the source symbols, the entropy of the source is

$$H(S) = 0.1 \log_2\left(\frac{1}{0.1}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.15 \log_2\left(\frac{1}{0.15}\right) + 0.09 \log_2\left(\frac{1}{0.09}\right)$$

$$+ 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.26 \log_2\left(\frac{1}{0.26}\right)$$

$$= 0.1\left(\frac{\log_{10} 10}{\log_{10} 2}\right) + 0.2\left(\frac{\log_{10} 5}{\log_{10} 2}\right) + 0.15\left(\frac{\log_{10} 6.6}{\log_{10} 2}\right) + 0.09\left(\frac{\log_{10} 11.1}{\log_{10} 2}\right)$$

$$+ 0.2\left(\frac{\log_{10} 5}{\log_{10} 2}\right) + 0.26\left(\frac{\log_{10} 3.846}{\log_{10} 2}\right)$$

$$\therefore \quad H(S) = 2.4895 \text{ bits/symbol}$$

$$\bar{n} = \sum_{i=0}^{5} p(x_i)n_i = 0.1 \times 4 + 0.2 \times 2 + 0.15 \times 3 + 0.09 \times 4 + 0.2 \times 2 + 0.26 \times 2$$

$$\therefore \quad \bar{n} = 2.53 \text{ bits/symbol}$$

$$\therefore \quad \text{coding efficiency} = \frac{H(S)}{\bar{n}} \times 100 = \frac{2.4895}{2.53} \times 100 = 98.4\%$$

Example 12.12 A discrete memoryless source has the alphabet *A*, *B*, *C*, *D*, *E*, *F* and *G* with corresponding probabilities {0.08, 0.2, 0.12, 0.15, 0.03, 0.02, 0.4}.

(a) Devise a Huffman code for the above source and determine the average length of the codeword.
(b) Determine the coding efficiency of the Huffman code designed.

Solution

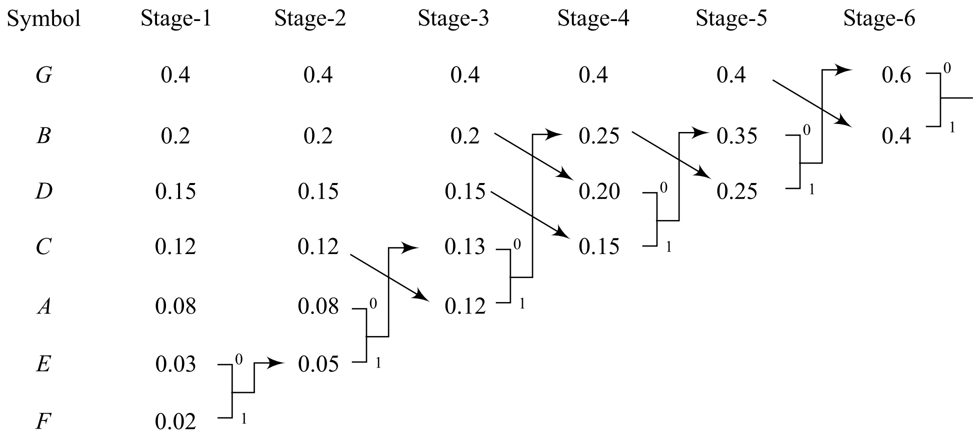
(a) The codewords are listed below:

A – 0 1 0 0; B – 0 0 0; C – 0 1 1; D – 0 0 1; E – 0 1 0 1 0; F – 0 1 0 1 1; G – 1

Hence, the average length of a codeword is

$$\bar{n} = \sum_{i=0}^{M-1} p(x_i)n_i = 0.08 \times 4 + 0.2 \times 3 + 0.12 \times 3 + 0.15 \times 3 + 0.03 \times 5 + 0.02 \times 5 + 0.4 \times 1 = 2.38$$

$$\therefore \quad \bar{n} = 2.38 \text{ bits/symbol}$$



(b) $H(S)$ = Entropy of the source $= -\sum_{i=0}^{M-1} p(x_i)\log_2 p(x_i)$

$$= 0.08\log_2\frac{1}{0.08} + 0.2\log_2\frac{1}{0.2} + 0.12\log_2\frac{1}{0.12} + 0.15\log_2\frac{1}{0.15}$$

$$+ 0.03\log_2\frac{1}{0.03} + 0.02\log_2\frac{1}{0.02} + 0.4\log_2\frac{1}{0.4}$$

$= 2.31$ bits/symbol

$\therefore$ coding efficiency $= \frac{2.31}{2.38} \times 100 = 97.135\%$

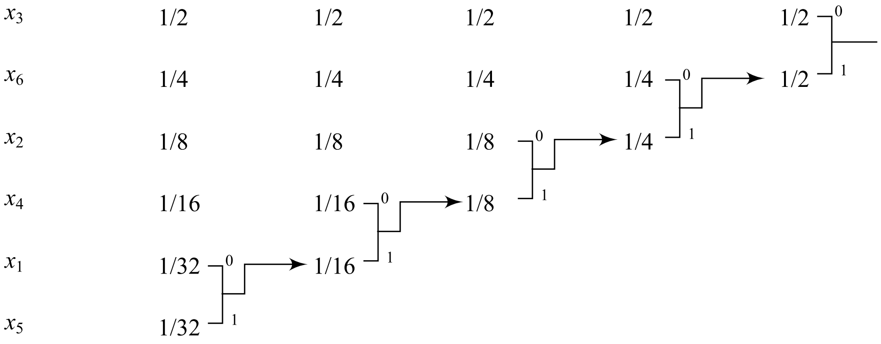
Example 12.13 A discrete memoryless source is described by the alphabet $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ with probabilities {1/32, 1/8, 1/2, 1/16, 1/32, 1/4} respectively.

(a) Design a Huffman code for the above source and find the average length of the codeword.

(b) Can you improve the Huffman code designed in part (a) by encoding the second-order extension of the source? Give reason(s) for your answer.

Solution

(a)



The codewords are listed below:

x_1: 1 1 1 1 0; x_2: 1 1 0; x_3: 0; x_4: 1 1 1 0; x_5: 1 1 1 1 1; x_6: 1 0

Average length of the codeword $= \sum_{i=0}^{6} p(x_i)n_i$

$$= 5 \times \frac{1}{32} + 3 \times \frac{4}{32} + 1 \times \frac{16}{32} + 4 \times \frac{2}{32} + 5 \times \frac{1}{32} + 2 \times \frac{8}{32}$$

$\therefore \quad \bar{n} = 1.9375$ bits/symbol

(b) From the given probability values for the various symbols, we find that $p(x_i) = 2^{-n_i}, i = 1, 2, \ldots, 6$ where n_is are all integers. From Eq. (12.30), we find that $\bar{n} = H(S) = \bar{n}_{\text{min}}$.

As per Shannon's theorem, it is not possible to reduce the value of $\bar{n}$ below the value of the entropy of the source. Hence, encoding the second order source, or for that matter, any other method too, cannot reduce $\bar{n}$ any further.

Example 12.14 A discrete memoryless source *S*, produces the symbols *A*, *B* and *C* with probabilities 0.4, 0.25 and 0.35, respectively.

(a) Can the output of this source be compressed so that the average codeword length is 2 bits? Give reason(s) for your answer. Devise a Huffman code for this source and determine its coding efficiency.

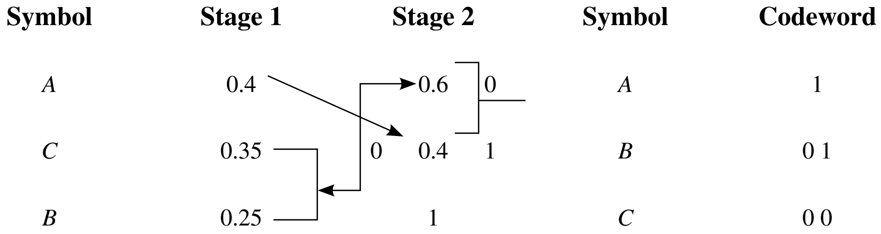
(b) Devise a Huffman code for the second-order extension of this source and find the average length of the codeword for this code. What is the coding efficiency?

Solution Entropy of the source $S = H(S) = -\sum_{i=0}^{M-1} p(x_i) \log_2 p(x_i)$

$$= 0.4 \log_2\left(\frac{1}{0.4}\right) + 0.25 \log_2\left(\frac{1}{0.25}\right) + 0.35 \log_2\left(\frac{1}{0.35}\right) = 1.55$$

$\therefore \quad H(S) = 1.55$ bits/symbol

(a) Since $H(S) = 1.55$ bits/symbol the average number of bits used for representing a symbol cannot be less than 1.55. So, it is possible to have $\bar{n} < 2$ bits; but it cannot be less than 1.55.



$\therefore$ average length of the codeword $= 1 \times 0.4 + 2 \times 0.35 + 2 \times 0.25$

$\therefore \quad \bar{n} = 1.6$ bits/codeword

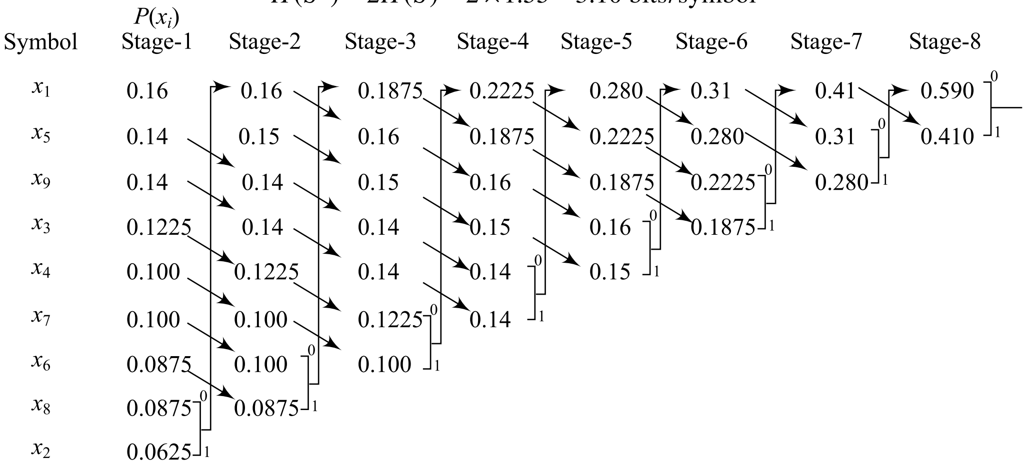
$$\text{Coding efficiency} = \frac{1.55}{1.60} = 96.875\%$$

(b) The second-order extension source S^2, will have the following alphabet with probabilities as shown below.

Symbol	Probability	Symbol	Probability
$AA = x_1$	0.16	$AC = x_5$	0.140
$BB = x_2$	0.0625	$BC = x_6$	0.0825
$CC = x_3$	0.1225	$BA = x_7$	0.100
$AB = x_4$	0.100	$CB = x_8$	0.0875
		$CA = x_9$	0.140

Entropy of the second-order source = 2 × Entropy of original source

$$\therefore \qquad H(S^2) = 2H(S) = 2 \times 1.55 = 3.10 \text{ bits/symbol}$$



Symbol (x_i)	Codeword	Length (n_i)
x_1	0 0 0	3
x_2	0 0 1 1	4
x_3	1 0 0	3
x_4	1 0 1	3
x_5	0 1 0	3
x_6	1 1 1	3
x_7	1 1 0	3
x_8	0 0 1 0	4
x_9	0 1 1	3

$$\text{Average length of the codeword} = \sum_{i=1}^{9} p(x_i)n_i = \overline{n}$$

$$= 3 \times 0.16 + 4 \times 0.0625 + 3 \times 0.1225 + 3 \times 0.100 + 3 \times 0.14$$

$$+ 3 \times 0.0875 + 3 \times 0.100 + 4 \times 0.0875 + 3 \times 0.14$$

$$= 3.15$$

$$\overline{n} = 3.15 \text{ bits/codeword (for new symbols)}$$

Since each new symbol is composed of two symbols of the original source, the average length of the new codeword is actually 3.15/2 = 1.57 bits ≈ $H(s)$ per original symbol. Thus, by encoding a second-order source, we are almost able to reach the *Shannon bound*.

$$\text{Coding efficiency} = \frac{3.10}{3.15} = 0.984 = 98.4\%$$

Although Huffman code is an optimal code in the sense that it gives the minimum average codeword length for a discrete memoryless source with a given source statistics, it suffers from the following two disadvantages in practical applications:

1. As it depends heavily on source statistics, an *a priori* knowledge of the probabilities of occurrence of the source symbols is a must.
2. As most of the sources that we come across in practice are not memoryless, (i.e., the probability of occurrence of a symbol is not independent of which symbols have preceded it) and since Huffman coding takes into account only the individual symbol probabilities, use of this coding in practical applications does not yield good compression.

12.5.3 Lempel–Ziv Source Coding Algorithm

This belongs to a class of coding schemes called the ‘Universal Coding Schemes’, which are not dependent upon the source statistics and so is quite popular in practical applications like coding a text. As we are aware, in English language there is considerable interdependence of the letters. For example, the conditional probability of the letter ‘u’ being the next letter, is very high given that the letter ‘q’ has occurred. Even though determination of the probability of occurrence of the letters of English alphabet can be done on the basis of relative frequency of their occurrence in a long string of words, determination of their interdependence is not that easy. It is in this context that universal coding schemes come in handy.

The Lempel–Ziv source coding algorithm proposed in the 1970s by two Israeli scientists, Abraham Lempel and Jacob Ziv, is, unlike the Huffman and Fano codes, a variable-to-fixed length code and is quite simple and widely used in practice for compressing computer files. In this scheme, a sequence of symbols from a source are passed into what are called ‘*unique phrases*’ which are of unequal length and these are then represented by codewords of fixed length. The encoding proceeds as follows.

The parser maintains a codebook. A 0 and 1 are initially stored, in that order, in this codebook. The parser then observes the output sequence from the source, parses it into segments which are shortest possible ‘*subsequences*’ that have not been observed earlier. These subsequences are stored in the codebook after the initially stored 0 and 1, in the order of their occurrence, and a numerical position is assigned to each of these. For instance, if the source sequence is given as

0 1 1 0 0 0 1 0 0 1 1 1 1 1 0 1 0…

Then the codebook entries and the progress made in parsing may be shown as

$$\left.\begin{array}{l}\text{subsequences stored in}\\ \text{the codebook}\end{array}\right\} 0, 1$$

Data yet to be parsed: 0 1 1 0 0 0 1 0 0 1 1 1 1 1 0 1 0

Now the parser encounters a ‘0’ in the data. As a ‘0’ is already stored in the codebook, the parser goes to the next digit in the data and finds that the subsequence 0 1 which it has come across is not available in the codebook and so enters it there. At this stage, the codebook entries and the data yet to be parsed may be shown as

$$\left.\begin{array}{l}\text{subsequences stored in}\\ \text{the codebook}\end{array}\right\} 0, 1, 01$$

Data yet to be parsed: 1 0 0 0 1 0 0 1 1 1 1 1 0 1 0 0 1 . . .

The parser continues this process till the source sequence is completely parsed. As mentioned earlier, for each entry in the codebook, a distinct numerical position is assigned. The codebook entries other than the initially stored 0 and 1 are now given compact numerical representation in terms of the numerical position assigned to the previously entered subsequences in the following manner:

Numerical position	1	2	3	4	5	6	7	8	9
subsequences stored in the codebook	0	1	01	10	00	100	11	111	010

Numeriacl representation of the codebook entries	12 21 11 41 22 72 31
Binary encoding of the numerical representation	0011 0100 0010 1000 0101 1111 0110

The numerical representation of the subsequence 0 1 is 12 because 0 1 is a concatenation of codebook entry 0 with numerical position 1 and codebook entry 1 with numerical position 2. Similarly 1 0 is the concatenation of codebook entry 1 with numerical position 2 and codebook entry 0 with numerical position 1 and so subsequence 1 0 is given the numerical representation of 21. Subsequence with three digits, like say 1 0 0 are always taken as concatenation of a subsequence with 2 digits (here 1 0) and a subsequence with single digit number (here 0). Since the numerical position of subsequence 1 0 is 4 and that of the subsequence 0 is 1, the numerical representation of the subsequence 1 0 0 is 41. Similarly, if a subsequence in the codebook is having four digits, it is considered as a concatenation of 3-digit subsequence and a single digit subsequence. Thus, the last digit of any subsequence has a special significance in that, if we append it to any subsequence other than the initially stored 0 and 1, the resulting subsequence is different from any of the previous subsequences. It is therefore given a special name, '*innovation symbol*'. These numerical representations of the various codebook entries are then binary encoded in a particular way. Since the first digit of the numerical representation is going up to 7 in our illustration, we use a three bit binary coding to get a unique representation for the first digit of the numerical representation. The second digit of the numerical representation is always a 1 or a 2 corresponding respectively to the initially stored 0 or 1. Thus, for instance, the numerical representation 41 is binary encoded as 1 0 0:0; while 72 is binary encoded as 111 : 1. Thus, in this illustration in which the final encoded blocks are of four-digit length, the last digit, is the innovation bit while the rest, called the pointer, represent what is generally referred to as the '*root subsequence*'. With Lempel–Ziv coding, the decoding process is extremely simple. First, the pointer of each block of the encoded sequence is used for identifying the root subsequence. Then the innovation bit, i.e., the last bit of each code block, is appended to it.

Advantages of Lempel–Ziv coding

1. It uses fixed length coding and is therefore quite well suited for synchronous transmission.
2. Unlike Huffman coding, it takes care of even inter-character redundancy and therefore can give better compaction of given data. For instance, it has been reported that for a text in English language, the level of compaction that can be achieved with Lempel–Ziv algorithm is as much as 55% while with Huffman coding it is only about 43%.

Because of these advantages, Lempel–Ziv coding is almost invariably being used nowadays in the place of Huffman coding for data compaction.

Example 12.15 Encode the following binary sequence using Lempel–Ziv coding scheme.

1 1 1 0 1 0 0 1 1 0 0 0 1 0 1 …

Assume that binary symbols 0 and 1 are already there in the codebook.

Solution

Numerical Position	1	2	3	4	5	6	7	8
Subsequences in the codebook	0	1	11	10	100	110	00	101
Numerical representation of codebook entries			22	21	41	31	11	42
Binary encoding of the numerical representation			0101	0100	1000	0110	0010	1001

12.6 DISCRETE MEMORYLESS CHANNELS (DMCs)

In the earlier sections, we had discussed about discrete memoryless sources (DMS) and the methods used for source coding. There, we had observed that source coding was only for an efficient/economical representation of the source output in order to save time, bandwidth and power required in transmitting it over a channel. In this section, we will focus our attention on another important aspect of digital data transmission, viz., reliability. For this purpose, we shall consider a simple, yet very useful statistical model of a channel, called the Discrete Memoryless Channel (DMC). A DMC is one which has discrete random variables, $\boldsymbol{X}$ and $\boldsymbol{Y}$ as its input and output respectively. Random variable $\boldsymbol{X}$ may take any one of the M possible values x_i, $i = 0, 1, \ldots, (M-1)$ and the random variable $\boldsymbol{Y}$ may take any one of the L possible values, y_j, $j = 0, 1, 2, \ldots,(L-1)$. In other words, the input may be any symbol from an alphabet $\{x_0, x_1, \ldots, x_{M-1}\}$ and the output may be any symbol from an alphabet $\{y_0, y_1, \ldots, y_{L-1}\}$. It is called a *Discrete* Memoryless Channel because the input and output are *discrete* random variables with a finite set of alphabet for each, and it is memoryless because its *present output* symbol depends only on the *present input* symbol but not on any of the previous input symbols. A diagrammatic representation of a discrete memoryless channel is shown in Fig. 12.4. When the channel is given a certain input x_i, it emits an output, say y_j, with a certain probability $p(y_j|x_i)$. Hence, a complete description of a discrete memoryless channel comprises specification of its input alphabet $\{x_0, x_1, \ldots, x_{M-1}\}$, its output alphabet $\{y_0, y_1, \ldots, y_{L-1}\}$ and a set of what are called the transitional probabilities $p(y_j|x_i)$, $j = 0, 1, \ldots, (L\text{-}1)$ and $i = 0, 1, \ldots, (M-1)$. These transitional probabilities, ML in number, are generally given in the form of an M by L matrix as shown. *Note that M and L need not be equal.* If channel coding is done then $M < L$, but if two input symbols lead to the same output symbol (see binary erasure channel) then $M > L$. If the channel were to be ideal, $M = L$ and $y_j = x_j$ for every j.

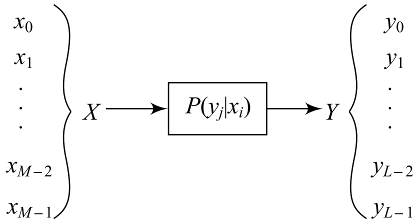


Fig. 12.4 *Diagrammatic representation of a discrete memoryless channel*

		Y							
	$p(y_j\|x_i)$	y_0	y_1	y_2	.	.	.	y_{L-2}	y_{L-1}
	x_0	$p(y_0\|x_0)$	$p(y_1\|x_0)$	$p(y_2\|x_0)$	.	.	.	$p(y_{L-2}\|x_0)$	$p(y_{L-1}\|x_0)$
	x_1	$p(y_0\|x_1)$	$p(y_1\|x_1)$	$p(y_2\|x_1)$	.	.	.	$p(y_{L-2}\|x_1)$	$p(y_{L-1}\|x_1)$
X	x_2	$p(y_0\|x_2)$	$p(y_1\|x_2)$	$p(y_2\|x_2)$	.	.	.	$p(y_{L-2}\|x_2)$	$p(y_{L-1}\|x_2)$
	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
	x_{M-2}	$p(y_0\|x_{M-2})$	$p(y_1\|x_{M-2})$	$p(y_2\|x_{M-2})$	.	.	.	$p(y_{L-2}\|x_{M-2})$	$p(y_{L-1}\|x_{M-2})$
	x_{M-1}	$p(y_0\|x_{M-1})$	$p(y_1\|x_{M-1})$	$p(y_2\|x_{M-1})$	.	.	.	$p(y_{L-2}\|x_{M-1})$	$p(y_{L-1}\|x_{M-1})$

Fig. 12.5 *Matrix of transitional probabilities*

In the matrix of transitional probabilities, $p(y_j|x_i)$ represents the probability of the random variable $\boldsymbol{Y}$ taking the value y_j given that random variable $\boldsymbol{X}$ has taken the value x_i, i.e., the probability of receiving symbol y_j when the symbol x_i is transmitted (because of noise in the channel).

i.e.,
$$p(y_j|x_i) = P\lfloor \boldsymbol{Y} = y_j | \boldsymbol{X} = x_i \rfloor ;\; i = 0, 1, \ldots, (M-1);\;\; j = 0, 1, \ldots, (L-1) \tag{12.39}$$

12.6.1 Properties of Transition Matrix/Channel Matrix

1. Each row of the transition matrix corresponds to a particular fixed input symbol to the channel.
2. Each column of the transition matrix corresponds to a certain fixed channel output symbol.
3. The sum of the probabilities along any particular row of the transition matrix is equal to one.

i.e.,
$$\sum_{j=0}^{L-1} p(y_j|x_i) = 1 \text{ for all } i \tag{12.40}$$

If the input symbol probabilities are known, i.e., if $p(x_i)$, $i = 0, 1, \ldots, (M - 1)$ are known, the joint distribution $p(x_i, y_j)$ $i = 0, 1, \ldots, (M - 1)$ and $j = 0, 1, \ldots, (L - 1)$, as well as the marginal distribution $p(y_j)$, $j = 0, 1, \ldots, (L - 1)$ can be determined as follows:

$$p(x_i, y_j) = p(y_j|x_i)p(x_i) \tag{12.41}$$

and
$$p(y_j) = \sum_{i=0}^{M-1} p(y_j|x_i)p(x_i);\ j = 0, 1, \ldots, (L-1) \tag{12.42}$$

12.6.2 Binary Symmetric Channel (BMC)

The input as well as the output alphabet size for a binary symmetric channel is two. Hence, $M = L = 2$. It is usually represented as shown in Fig. 12.6, when p is used to represent the transition probability. Since the channel is symmetric, the probability of a transmitted 0 being received as a 1 and a transmitted 1 being received as a 0 are equal.

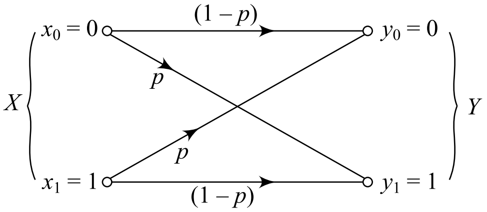


Fig. 12.6 *A binary symmetric channel*

The transition or channel matrix of a binary symmetric channel (BSC) may be written down as

$$\text{Transition matrix } \boldsymbol{P} = \begin{bmatrix} (1-p) & p \\ p & (1-p) \end{bmatrix} \tag{12.43}$$

12.6.3 Binary Erasure Channel (BEC)

The binary erasure channel has an input alphabet size of 2 and an output alphabet size of 3. Hence, for this channel, $M = 2$ and $L = 3$. Sometimes, due to noise, it may not be possible to identify the output symbol as one or the other of the input symbols. In that case, it is erased, i.e., ignored and a request is sent to the transmitter to retransmit. That is why it is called a binary erasure channel. It is generally represented as shown in Fig. 12.7.

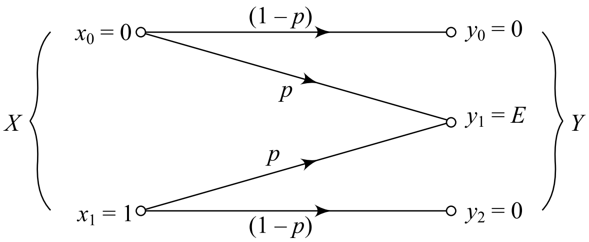


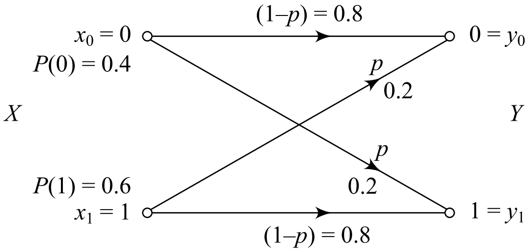
Fig. 12.7 *A binary erasure channel*

The transition, or channel matrix for a BEC is readily seen to be

$$\boldsymbol{P} = \begin{bmatrix} (1-p) & p & 0 \\ 0 & p & (1-p) \end{bmatrix} \tag{12.44}$$

Example 12.16 A binary symmetric channel has an error probability $p = 0.2$. The *a priori* probabilities of a 0 and 1 at the input are 0.4 and 0.6, respectively. What is the probability of receiving a 1 at the receiving end?

Solution



From Eq. (12.42), we have

$$p(y_1) = p(\mathbf{Y}=1) = \left[\sum_{i=0}^{1} p(y_j|x_i)p(x_i)\right]\Bigg|_{j=1}$$

$$= p(y_1|x_0)p(x_0) + p(y_1|x_1)p(x_1) \qquad (12.45)$$

$$\therefore \qquad p_Y(1) = 0.2 \times 0.4 + 0.8 \times 0.6 = 0.08 + 0.48 = 0.56$$

In fact, we could have written down this straight away. A 1 may be observed at the output either when a 1 is transmitted, or even when a 0 is transmitted.

$\therefore$ $p_Y(1)$ = (Probability of a 1 at the input) × (Probability of this 1 being received as a 1) + (Probability of a 0 at the input) × (Probability of a 0 being received as a 1)

Note

Equation (12.45) is precisely this only.

$$P_Y(1) = P(1) \times P(1|1) + P(0) \times P(1|0)$$
$$= 0.6 \times 0.8 + 0.4 \times 0.2 = 0.56$$

12.7 MUTUAL INFORMATION AND CHANNEL CAPACITY

As usual, let the source have an alphabet of size M, i.e., the random variable $\mathbf{X}$ can take one of the M values $\{x_0, x_1, \ldots, x_{M-1}\}$. Also, let the channel output have an alphabet size L, i.e., the random variable $\mathbf{Y}$ can take one of the L possible values $\{y_0, y_1, \ldots, y_{L-1}\}$. Of course, if the channel were to be ideal, L would be equal to M and y_j would be equal to x_j for all j.

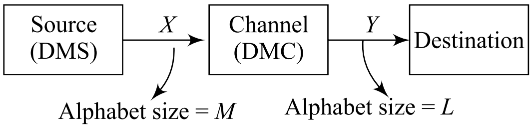


Fig. 12.8 *Information transfer from source to destination*

When the source symbol x_i is transmitted, let us say the channel output is the symbol y_j. Then there are two probabilities with which we are concerned: (i) The *a priori* probability $p(x_i)$ of x_i being transmitted, and (ii) The *a posteriori* probability $p(x_i|y_j)$. This is the probability of x_i having been transmitted, given that y_j has been received. $p(x_i)$ is related to our 'state of knowledge' at the destination, before x_i is transmitted, regarding which symbol from the source alphabet would be transmitted. However, once x_i is transmitted and y_j is received at the destination, our 'state of knowledge' at the destination regarding which symbol from the alphabet of the source has been transmitted, is represented by the *a posteriori* probability $p(x_i|y_j)$ of x_i having been transmitted given that y_j has been received. If $p(x_i|y_j)$ is equal to 1 for a particular x_i and zero for all the other x_i, as it should be under ideal conditions, once we observe y_j, the uncertainty at the destination regarding which source symbol has been transmitted, would be completely removed. But unfortunately, owing to the presence of noise and other channel imperfections, $p(x_i|y_j)$ will not be one for one x_i and zero for the rest. Instead, it will be large (but not equal to one) for some x_i and small but not exactly zero for the rest. Because of this, the uncertainty at the destination regarding which x_i was actually transmitted will not be completely removed even after the channel output y_j is observed. So, the situation may be described by saying: There was some uncertainty at the destination regarding which source symbol would be transmitted before the symbol y_j was received; but even after receiving it, there is still some uncertainty regarding which symbol was transmitted. Of course the uncertainty persisting after receiving y_j would be less compared to what it was before it was received, indicating that some amount of information has been transferred from the source to the destination.

Now, the amount of information at the destination after receiving the output of the channel, is given by

$$I_2 = \log\left[\frac{1}{p(x_i|y_j)}\right] \tag{12.46}$$

and the amount of information at the destination before receiving the symbol y_j is given by

$$I_1 = \log\left[\frac{1}{p(x_i)}\right]$$

Therefore the amount of information transferred when y_j is observed is given by

$$I_1 - I_2 = \log\left[\frac{1}{p(x_i)}\right] - \log\left[\frac{1}{p(x_i|y_j)}\right]$$

or

$$I(x_i, y_j) = \log\left[\frac{p(x_i|y_j)}{p(x_i)}\right] \tag{12.47}$$

This $I(x_i, y_j)$ represents the difference in the amount of information at the destination after and before the reception of the symbol y_j consequent to the transmission of the symbol x_i, and it is called 'Mutual Information' of the channel, between the transmitted symbol x_i and the received symbol y_j.

But since the transmitted symbol can be any x_i and the received symbol also can be any y_j, it is more appropriate to talk about the average mutual information, which is the difference between the average information given by the source, i.e., its entropy $H(\boldsymbol{X})$ and the average information at the destination after receiving the output. The latter, is given by $H(\boldsymbol{X} \mid \boldsymbol{Y})$ using Eq. (12.20) we may write

$$H(\boldsymbol{X}|\boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j) \log p(x_i|y_j)$$

$\therefore$

$$H(\boldsymbol{X}|\boldsymbol{Y}) = -\sum_{j=0}^{L-1}\sum_{i=0}^{M-1} p(y_j)\{p(x_i|y_j)\log p(x_i|y_j)\} \tag{12.48}$$

Note that the RHS of Eq. (12.48) is the average value of I_2 (of Eq. (12.46)) when all x_is and y_js are considered.
$\therefore$ average mutual information of the channel is given by

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) \text{ in bits/symbol} \tag{12.49}$$

$I(\boldsymbol{X}; \boldsymbol{Y})$, the mutual information of the channel represents the average amount of information transferred through the channel in bits/symbol.

Similar to Eq. (12.49), we may write

$$I(\boldsymbol{Y};\boldsymbol{X}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) \tag{12.50}$$

where $H(\boldsymbol{Y})$ is the entropy of the channel output and $H(\boldsymbol{Y}|\boldsymbol{X})$ is the conditional entropy of the channel output given the channel input.

12.7.1 Properties of Mutual Information

Property 12.1 The mutual information of a channel is symmetric, i.e.,

$$I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) \tag{12.51a}$$

Proof From Eq. (12.22), we have

$$H(\boldsymbol{X}, \boldsymbol{Y}) = H(\boldsymbol{Y}) + H(\boldsymbol{X}|\boldsymbol{Y})$$
$$= H(\boldsymbol{X}) + H(\boldsymbol{Y}|\boldsymbol{X})$$

From the above, it follows that

$$[H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y})] - [H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})] = 0$$

i.e., $H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})$

From Eqs. (12.49) and (12.50), it then follows that

$$I(\boldsymbol{X}; \boldsymbol{Y}) = I(\boldsymbol{Y}; \boldsymbol{X})$$

Property 12.2 The mutual information is non-negative, i.e., $I(\boldsymbol{X}, \boldsymbol{Y}) \geq 0$

Proof We know $I(\boldsymbol{X}; \boldsymbol{Y}) = H(\boldsymbol{X}) - H(\boldsymbol{X} \mid \boldsymbol{Y})$ (from Eq. (12.49))

But $$H(\boldsymbol{X}) = -\sum_{i=0}^{M-1} p(x_i)\log_2 p(x_i)$$

and $$H(\boldsymbol{X}/\boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i|y_j)$$

$\therefore$ $$I(\boldsymbol{X}; \boldsymbol{Y}) = -\sum_{i=0}^{M-1} p(x_i)\log_2 p(x_i) + \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i|y_j)$$
$$= \left[-\sum_{i=0}^{M-1} p(x_i)\log_2 p(x_i)\sum_{j=0}^{L-1} p(y_j|x_i)\right] + \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i|y_j)$$

That $\sum_{j=0}^{L-1} p(y_j|x_i) = 1$ has been made use of in the above expression.

$\therefore$ $$-I(\boldsymbol{X}; \boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i)\,p(y_j|x_i)\log_2 p(x_i) - \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i|y_j)$$

Combining the two double summations using the fact that $p(x_i)p(y_j|x_i) = p(x_i, y_j)$, we get

$$-I(\boldsymbol{X}; \boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)}{p(x_i|y_j)}\right] \tag{12.51b}$$
$$= \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i|y_j)p(y_j)}\right]$$
$$= \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right] \tag{12.52}$$

$\therefore$ $$-I(\boldsymbol{X}; \boldsymbol{Y}) = \frac{1}{\log_e 2}\left\{\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_e\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right]\right\}$$

But we know that $\log_e x \leq (x - 1)$

$\therefore$ $$-I(\boldsymbol{X}; \boldsymbol{Y}) \leq \frac{1}{\log_e 2}\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)} - 1\right]$$

or $$-I(\boldsymbol{X};\boldsymbol{Y}) \le \frac{1}{\log_e 2}\left[\sum_{i=0}^{M-1}\sum_{j=0}^{L-1}\{p(x_i)p(y_j) - p(x_i, y_j)\}\right]$$

But $$\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i)p(y_j) = 1 = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)$$

∴ $$-I(\boldsymbol{X};\boldsymbol{Y}) \le 0$$

or $$\therefore\ I(\boldsymbol{X};\boldsymbol{Y}) \ge 0 \tag{12.53}$$

This implies that even on a noisy channel, by observing the output of the channel, on the average we cannot lose any information. At the most, the mutual information may be zero, i.e., we do not gain any information by observing the output, and this happens when the input and output symbols of the channel are statistically independent.

Property 12.3 The mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ of a channel is related to the marginal entropies $H(\boldsymbol{X})$ and $H(\boldsymbol{Y})$ of the input and output and their joint entropy $H(\boldsymbol{X}, \boldsymbol{Y})$ as per the following relationship:

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{X}, \boldsymbol{Y})$$

Proof We know that

$$H(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(x_i, y_j)}\right]$$

$$= \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right] + \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(x_i)p(y_j)}\right]$$

But from Eq. (12.52), the first term is $-I(\boldsymbol{X};\boldsymbol{Y})$

∴ $$H(\boldsymbol{X},\boldsymbol{Y}) = -I(\boldsymbol{X};\boldsymbol{Y}) + \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(x_i)p(y_j)}\right] \tag{12.54}$$

The second term on the RHS may be rewritten in the following manner:

$$\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(x_i)p(y_j)}\right] = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(x_i)}\right] + \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{1}{p(y_j)}\right]$$

$$= \left\{\sum_{i=0}^{M-1}\log_2\left[\frac{1}{p(x_i)}\right]\sum_{j=0}^{L-1} p(x_i, y_j)\right\} + \left\{\sum_{i=0}^{M-1}\log_2\left[\frac{1}{p(y_j)}\right]\sum_{j=0}^{L-1} p(x_i, y_j)\right\}$$

But $$\sum_{j=0}^{L-1} p(x_i, y_j) = p(x_i) \text{ and } \sum_{i=0}^{M-1} p(x_i, y_j) = p(y_j)$$

∴ the second term on the RHS of Eq. (12.54) is equal to

$$\sum_{i=0}^{M-1} p(x_i)\log_2\left[\frac{1}{p(x_i)}\right] + \sum_{j=0}^{L-1} p(y_j)\log_2\left[\frac{1}{p(y_j)}\right] = H(\boldsymbol{X}) + H(\boldsymbol{Y})$$

∴ $$H(\boldsymbol{X},\boldsymbol{Y}) = -I(\boldsymbol{X};\boldsymbol{Y}) + H(\boldsymbol{X}) + H(\boldsymbol{Y})$$

or $$\boxed{I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{X},\boldsymbol{Y})} \tag{12.55}$$

The relationship between the various channel entropies is diagrammatically depicted in Fig. 12.9. Entropy of the channel input, $H(\boldsymbol{X})$, is shown as the circle on the left, channel output entropy, $H(\boldsymbol{Y})$ is shown as the circle on the right. The common area is $I(\boldsymbol{X};\boldsymbol{Y})$, the mutual information.

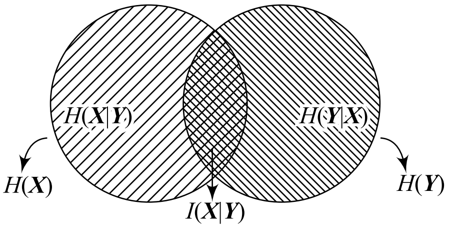


Fig. 12.9 *Relationship between H(X), H(Y) and I(X;Y)*

Example 12.17 Find the mutual information of a binary symmetric channel with a transition probability of p and an *a priori* probability of occurrence of a binary '0' equal to α.

Solution

From Eq. (12.52), we have

$$-I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right]$$

$$= p(x_0, y_0)\log_2\left[\frac{p(x_0)p(y_0)}{p(x_0, y_0)}\right] + p(x_0, y_1)\log_2\left[\frac{p(x_0)p(y_1)}{p(x_0, y_1)}\right]$$

$$+ p(x_1, y_1)\log_2\left[\frac{p(x_1)p(y_1)}{p(x_1, y_1)}\right] + p(x_1, y_0)\log_2\left[\frac{p(x_1)p(y_0)}{p(x_1, y_0)}\right] \quad (12.56)$$

Now, from Fig. 12.10, we may determine the various probabilities involved in Eq. (12.56) as follows:

$$p(x_0, y_0) = p(y_0|x_0)p(x_0) = (1-p)\alpha; \quad p(x_1, y_1) = p(y_1|x_1)p(x_1) = (1-p)(1-\alpha)$$

$$p(x_0, y_1) = p(y_1|x_0)p(x_0) = p\alpha; \quad p(x_1, y_0) = p(y_0|x_1)p(x_1) = p(1-\alpha)$$

Further,

$$p(y_0) = p(x_0)p(y_0|x_0) + p(x_1)p(y_0|x_1) = \alpha(1-p) + (1-\alpha)p$$

$$p(y_1) = p(x_0)p(y_1|x_0) + p(x_1)p(y_1|x_1) = \alpha p + (1-\alpha)(1-p)$$

Then, $-I(\boldsymbol{X};\boldsymbol{Y})$ is obtained by substituting all the above probabilities into Eq. (12.56).

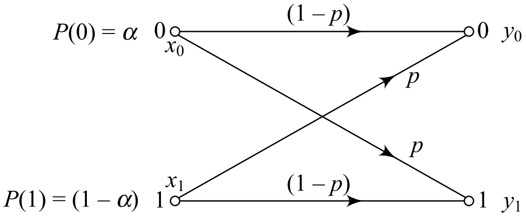


Fig. 12.10

Example 12.18 Given a channel as shown in Fig. 12.11, determine

(a) Channel matrix
(b) $P(y_1)$ and $P(y_2)$
(c) $P(x_1, y_2)$ and $P(x_2, y_1)$

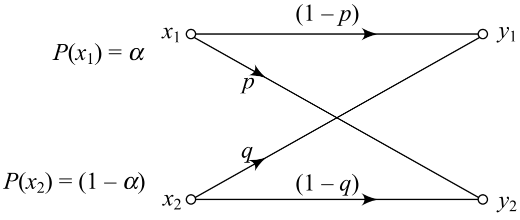


Fig. 12.11

Solution

(a) The transition matrix may be written down as

		Y	
	$P(y_j \mid x_i)$	y_1	y_2
X	x_1	$(1-p)$	p
	x_2	q	$(1-q)$

$$\therefore \quad [P] = \text{Channel matrix} = \begin{bmatrix} (1-p) & p \\ q & (1-q) \end{bmatrix}$$

(b)
$$P(y_1) = p(x_1)p(y_1|x_1) + p(x_2)p(y_1|x_2)$$
$$= [\alpha(1-p) + (1-\alpha)q]$$
$$P(y_2) = p(x_1)p(y_2|x_1) + p(x_2)p(y_2|x_2)$$
$$= [\alpha p + (1-\alpha)(1-q)]$$

(c)
$$P(x_1, y_2) = p(y_2|x_1)p(x_1) = p\alpha$$
$$P(x_2, y_1) = p(y_1|x_2)p(x_2) = q(1-\alpha)$$

Example 12.19 Show that the conditional entropy $H(\boldsymbol{X} \mid \boldsymbol{Y}) = 0$ for a lossless channel.

Note *A lossless channel is one which has only one non-zero element in each column of the channel matrix.*

Solution From Eq. (12.48), we know that

$$H(\boldsymbol{X}|\boldsymbol{Y}) = -\sum_{j=0}^{L-1}\sum_{i=0}^{M-1} p(y_j)p(x_i|y_j)\log_2 p(x_i|y_j)$$

From the properties of a channel matrix, we know that the sum of all the elements in each row must be equal to one. Also, since there will be only one non-zero element in each column of a lossless channel, and since each column in a channel matrix corresponds to one particular output symbol, it follows that each input symbol will be giving one *distinct* output symbol. Thus, each of the conditional probabilities appearing in the expression for $H(\boldsymbol{X} \mid \boldsymbol{Y})$ will be either 1 or 0.

This means that $H(\boldsymbol{X} \mid \boldsymbol{Y})$ will be zero for a lossless channel.

Example 12.20 Find the overall channel matrix if two channels are connected in cascade. Assume each channel to be the one given in Example 12.18.

Solution

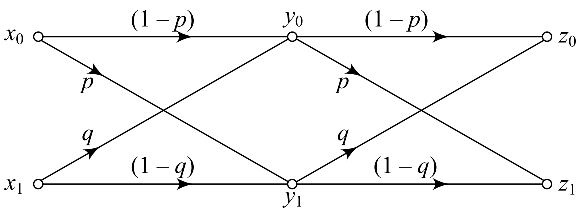


Fig. 12.12

$$\left.\begin{array}{l}\text{The transition matrix, or the channel}\\ \text{matrix of the single channel}\end{array}\right\} = [P] = [P(\boldsymbol{Y}|\boldsymbol{X})] = \begin{bmatrix}(1-p) & p\\ q & (1-q)\end{bmatrix}$$

$\therefore$
$$[P(\boldsymbol{Y})] = [p(y_0) \quad p(y_1)] = [p(x_0) \quad p(x_1)]\begin{bmatrix}(1-p) & p\\ q & (1-q)\end{bmatrix}$$

But
$$[P(\boldsymbol{Z})] = [p(z_0) \quad p(z_1)] = [p(y_0) \quad p(y_1)]\begin{bmatrix}(1-p) & p\\ q & (1-q)\end{bmatrix}$$

Combining the two equations, we get

$$[p(z_0) \quad p(z_1)] = [p(x_0) \quad p(x_1)]\begin{bmatrix}(1-p) & p\\ q & (1-q)\end{bmatrix}\begin{bmatrix}(1-p) & p\\ q & (1-q)\end{bmatrix}$$

$\therefore$
$$[p(z_0) \quad p(z_1)] = [p(x_0) \quad p(x_1)]\begin{bmatrix}(1-p)^2 + pq & p(1-p) + p(1-q)\\ q(1-p) + q(1-q) & pq + (1-q)^2\end{bmatrix}$$

Hence, $[P(\boldsymbol{Z} \mid \boldsymbol{X})]$ = Transition matrix of the overall channel

$$= \begin{bmatrix}(1+p^2-2p+pq) & (2p-p^2-pq)\\ (2q-q^2-pq) & (1+q^2-2q+pq)\end{bmatrix}$$

12.7.2 Channel Capacity

From Eq. (12.51a), we have

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i|y_j)}{p(x_i)}\right]$$

But
$$p(x_i, y_j) = p(y_j|x_i)p(x_i) = p(x_i|y_j)p(y_j)$$

$\therefore$
$$\frac{p(x_i|y_j)}{p(x_i)} = \frac{p(y_j|x_i)}{p(y_j)}$$

$\therefore$
$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(y_j|x_i)}{p(y_j)}\right]$$

But
$$p(x_i, y_j) = p(y_j|x_i)p(x_i)$$

and
$$p(y_j) = \sum_{i=0}^{M-1} p(x_i, y_j) = \sum_{i=0}^{M-1} p(y_j|x_i)p(x_i)$$

$\therefore$
$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(y_j|x_i)p(x_i)\log_2\left[\frac{p(y_j|x_i)}{\sum_{i=0}^{M-1} p(y_j|x_i)p(x_i)}\right]$$

From the above equation, it is clear that in order to determine the mutual information, one has to know the transition probabilities $p(y_j|x_i)$ for all i and j and also the $p(x_i)$s for all i – the probability distribution of the input symbols. For a given channel, the transition probabilities are all fixed. Hence, for a given channel, if we want to maximize the mutual information, i.e., maximize the information per symbol, on the average, transferred from the input to output of the channel, we need to maximize $I(\boldsymbol{X}; \boldsymbol{Y})$ with respect to the probability distribution for the input symbols. The maximum value so obtained, of the mutual information of a given channel is called the capacity of the channel. Thus, channel capacity is defined as

$$\text{Channel Capacity} = C_s = \max_{\{p(x_i)\}} I(\boldsymbol{X};\boldsymbol{Y}) \text{ in bits/symbol} \tag{12.57}$$

and it represents the maximum average information that can be transferred per symbol over the channel.

Although the mutual information, $I(\boldsymbol{X};\boldsymbol{Y})$ depends on the input symbol probability distribution, i.e., $p(x_i)$s, as well as the transition probabilities which define the channel, the channel capacity C_s, however, depends only on the channel (i.e., the transition probabilities) and is independent of the probability distribution of the input symbols.

Since the maximization is with respect to all possible input symbol probability distributions, and since $p(x_i)$s must always satisfy the following two conditions whatever may be the distribution, determination of the channel capacity involves a constrained optimization, the two constraints being

$$p(x_i) \geq 0 \text{ for all } i$$

and

$$\sum_i p(x_i) = 1$$

Thus, except in certain very simple cases, determination of the channel capacity of a given channel is not an easy task.

Note *Channel Capacity in bits/second =* $C = R_s \cdot \max_{\{p(x_i)\}} I(\boldsymbol{X};\boldsymbol{Y})$ *(12.58)*

where R_s *is the transmission rate in symbols/sec.*

Example 12.21 Determine $I(\boldsymbol{X};\boldsymbol{Y})$, the mutual information of a binary symmetric channel assuming $p(x_0) = p(x_1) = 0.5$.

Solution In a binary symmetric channel, $p = q = 0.5$ as shown in Fig. 12.13.

Refer to Example 12.17. Substituting $p = q = 0.5$ and $\alpha = 0.5$ in the results obtained therein, we have

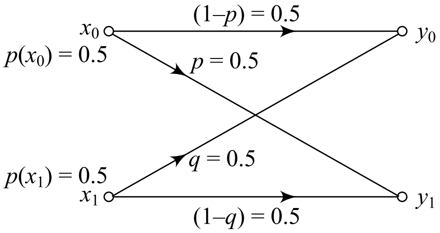


Fig. 12.13 *A binary symmetric channel*

$$p(x_0, y_0) = p(x_0, y_1) = p(x_1, y_0) = p(x_1, y_1) = 0.25$$

Also, $$p(y_0) = p(y_1) = 0.5$$

$$-I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right]$$

$$= p(x_0, y_0)\log_2\left[\frac{p(x_0)p(y_0)}{p(x_0, y_0)}\right] + p(x_0, y_1)\log_2\left[\frac{p(x_0)p(y_1)}{p(x_0, y_1)}\right]$$

$$+ p(x_1, y_1)\log_2\left[\frac{p(x_1)p(y_1)}{p(x_1, y_1)}\right] + p(x_1, y_0)\log_2\left[\frac{p(x_1)p(y_0)}{p(x_1, y_0)}\right]$$

Substituting for $p(x_i)$, $p(y_j)$ and $p(x_i, y_j)$ in the above, we find that

$$-I(\boldsymbol{X};\boldsymbol{Y}) = 0.25 \log_2 1 + 0.25 \log_2 1 + 0.25 \log_2 1 + 0.25 \log_2 1 = 0$$

$$\therefore \quad I(\boldsymbol{X};\boldsymbol{Y}) = 0$$

This result is to be expected since in this case,

$$p(y_0|x_0) = p(y_1|x_0) = 0.5 \text{ and } p(y_0|x_1) = p(y_1|x_1) = 0.5$$

This means that irrespective of which symbol is transmitted, the probability of receiving y_0 and y_1 are same, each equal to 0.5.

In such a situation, transmission of symbols through the channel is absolutely of no use, since the output symbol can as well be determined each time by flipping a coin.

Example 12.22 For a binary symmetric channel, show that

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) + p\log_2 p + (1-p)\log_2(1-p)$$

Solution We know that $I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})$ so it would suffice if we show that in the case of a BSC,

$$H(\boldsymbol{Y}|\boldsymbol{X}) = -[p\log_2 p + (1-p)\log_2(1-p)]$$

We know that

$$-H(\boldsymbol{X}|\boldsymbol{Y}) = \sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2[p(y_j|x_i)]$$

Setting $M = L = 2$ and expanding, we get

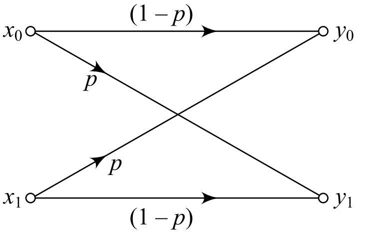


Fig. 12.14 *A binary symmetric channel*

$$\begin{aligned}-H(\boldsymbol{Y}|\boldsymbol{X}) &= p(y_0, x_0)\log_2 p(y_0|x_0) + p(y_0, x_1)\log_2 p(y_0|x_1)\\ &\quad + p(y_1, x_0)\log_2 p(y_1|x_0) + p(y_1, x_1)\log_2 p(y_1|x_1)\\ &= p(x_0)p(y_0|x_0)\log_2 p(y_0|x_0) + p(x_1)p(y_0|x_1)\log_2 p(y_0|x_1)\\ &\quad + p(x_0)p(y_1|x_0)\log_2 p(y_1|x_0) + p(x_1)p(y_1|x_1)\log_2 p(y_1|x_1)\end{aligned}$$

From Fig. 12.14, we find that

$$p(y_0|x_0) = (1-p);\quad p(y_0|x_1) = p;\quad p(y_1|x_0) = p;\quad p(y_1|x_1) = (1-p)$$

Substituting these values and simplifying,

$$-H(\boldsymbol{Y}|\boldsymbol{X}) = [p(x_0) + p(x_1)][(1-p)\log_2(1-p)] + [p(x_0) + p(x_1)][p\log_2 p]$$

But $p(x_0) + p(x_1) = 1$

$$\therefore \quad -H(\boldsymbol{Y}|\boldsymbol{X}) = (1-p)\log_2(1-p) + p\log_2 p$$

$$\therefore \quad I(\boldsymbol{X}|\boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) = H(\boldsymbol{Y}) + (1-p)\log_2(1-p) + p\log_2 p \tag{12.59}$$

Example 12.23 Find the channel capacity of a binary symmetric channel. Sketch C_s vs. p, the transition probability.

Solution In the previous example, we had shown that for a BSC, the mutual information is given by

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) + (1-p)\log_2(1-p) + p\log_2 p$$

To find the channel capacity C_s, we have to maximize $I(\boldsymbol{X};\boldsymbol{Y})$ with respect to the probability distribution of the input random variable $\boldsymbol{X}$ which can take any of the two values, x_0 and x_1. This amounts to saying that we have to maximize the $H(\boldsymbol{Y})$ on the RHS of that equation, since $[p\log_2 p + (1-p)\log_2(1-p)]$ is independent of the probability distribution of the input random variable. From the symmetry of a BSC, we know that $H(\boldsymbol{Y})$, like $H(\boldsymbol{X})$, attains the maximum value when x_0 and x_1 are equi-probable, i.e., when

$$p(x_0) = p(x_1) = 0.5$$

From Fig.12.15, we find that

$$\begin{aligned}p(y_0) &= p(x_0)p(y_0|x_0) + p(x_1)p(y_0|x_1)\\ &= p(x_0)(1-p) + p(x_1)p\end{aligned}$$

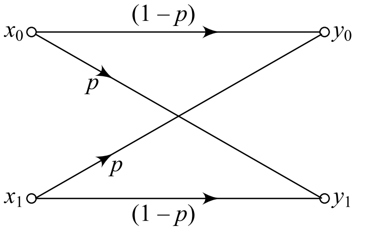


Fig. 12.15 *A binary symmetric channel*

$\therefore$ when $p(x_0) = p(x_1) = 0.5$,

$$p(y_0) = 0.5(1-p) + 0.5p = 0.5$$

Hence, $p(y_1) = 0.5$

But $$H(\boldsymbol{Y}) = -p(y_0)\log_2 p(y_0) - p(y_1)\log_2 p(y_1)$$

$\therefore$ when $p(y_0) = p(y_1) = 0.5$, $H(\boldsymbol{Y})$ takes a maximum value given by

$$\underset{(\max)}{H}(\boldsymbol{Y}) = 0.5\log_2 2 + 0.5\log_2 2 = 1$$

$$\therefore \quad \underset{(\max)}{I}(\boldsymbol{X};\boldsymbol{Y}) = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

$$\therefore \quad \boxed{\underset{(\text{BSC})}{C_s} = \underset{(\max)}{I}(\boldsymbol{X};\boldsymbol{Y}) = 1 + p\log_2 p + (1-p)\log_2(1-p)} \tag{12.60}$$

Recall that (refer Eq, (12.3a) of Example 12.4) the entropy, $H(S)$ of a binary DMS is given by

$$H(S) = -[\alpha\log_2\alpha + (1-\alpha)\log_2(1-\alpha)]$$

where α is the probability of one of the two possible source symbols.

So, we may write the channel capacity of a binary symmetric channel as

$$\underset{(\text{BSC})}{C_s} = \underset{(\max)}{I}(\boldsymbol{X};\boldsymbol{Y}) = [1 - H(p)] \text{ bits/symbol}$$

where $H(\cdot)$ is the familiar entropy function. In Fig.12.2, we had plotted variation of $H(S)$ with p. It takes a maximum value of 1 when $p = 0.5$ and is in the form of an inverted bowl, taking zero value at $p = 0$ as well as $p = 1$.

Hence, variation of C_s of a BSC will be as shown in Fig.12.16, with respect to the transition probability p of the BSC.

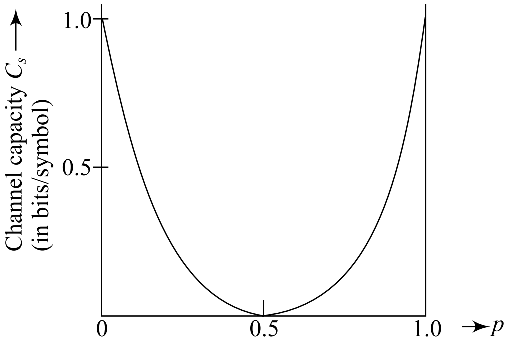


Fig. 12.16 *Channel capacity variation with transition probability, p*

Remarks

(i) If p = 0, corresponding to noise-free conditions on the channel, C_s = 1 bit per symbol, the maximum possible value.

(ii) If p = 1/2 owing to channel noise, the channel capacity takes the maximum value of zero bits per symbol. This corresponds to the case of $p(y_0|x_0) = p(y_1|x_0) = 0.5$ and $p(y_1|x_1) = p(y_0|x_1) = 0.5$. So, whenever a symbol is transmitted, it is likely to be received either as that symbol, or the other symbol with equal probability. Thus, receiving a symbol does not remove any uncertainty, rendering the transmission over the channel useless (refer to Example 12.21).

Example 12.24 Determine the channel capacity of the binary erasure channel shown in Fig. 12.17.

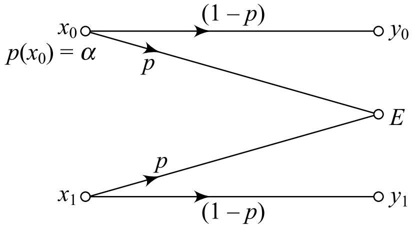


Fig. 12.17 *Binary erasure channel*

Solution We know that $I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})$

But

$$-H(\boldsymbol{Y}) = p(y_0)\log_2 p(y_0) + p(E)\log_2 p(E) + p(y_1)\log_2 p(y_1)$$

Also,
$$p(y_0) = p(y_0|x_0)p(x_0) = (1-p)\alpha$$
$$p(y_1) = p(y_1|x_1)p(x_1) = (1-p)(1-\alpha)$$
$$p(E) = p(E|x_0)p(x_0) + p(E|x_1)p(x_1) = p\alpha + p(1-\alpha) = p$$

Substituting these in the above expression for $-H(\boldsymbol{Y})$, we have
$$-H(\boldsymbol{Y}) = (1-p)\alpha\log_2[(1-p)\alpha] + p\log_2 p + (1-p)(1-\alpha)\log_2[(1-p)(1-\alpha)]$$

Also,
$$-H(\boldsymbol{Y}|\boldsymbol{X}) = p(y_0|x_0)p(x_0)\log_2 p(y_0|x_0) + p(E|x_0)p(x_0)\log_2 p(E|x_0)$$
$$+ p(E|x_1)p(x_1)\log_2 p(E|x_1) + p(y_1|x_1)p(x_1)\log_2 p(y_1|x_1)$$
$$-H(\boldsymbol{Y}|\boldsymbol{X}) = (1-p)\alpha\log_2(1-p) + p\alpha\log_2 p + p(1-\alpha)\log_2 p + (1-p)(1-\alpha)\log_2(1-p)$$
$$= (1+\alpha-p)\log_2(1-p) + p\log_2 p$$

To maximize $I(\boldsymbol{X};\boldsymbol{Y})$, put $\alpha = (1-\alpha) = 0.5$ because from the symmetry of the BEC, the output entropy will be maximized if the input entropy is maximized.

$$\therefore \quad \underset{(\alpha=0.5)}{-H(\boldsymbol{Y})} = 0.5(1-p)\log_2[(1-p)/2] + p\log_2 p + 0.5(1-p)\log_2[(1-p)/2]$$
$$= (1-p)\log_2(1-p) - (1-p)\log_2 2 + p\log_2 p$$

and
$$\underset{(\alpha=0.5)}{-H(\boldsymbol{Y}|\boldsymbol{X})} = 0.5(1-p)\log_2(1-p) + p\log_2 p + 0.5(1-p)\log_2(1-p)$$

$$\therefore \quad I(\boldsymbol{X};\boldsymbol{Y}) = [H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X})]_{\max} = (1-p)$$

12.8 CHANNEL CODING THEOREM

The average probability of error, P_e, of a digital communication system indicates how reliable the communication system is. Obviously higher the value of P_e, less reliable is the system. For a fixed average transmitted power, the value of P_e depends upon how noisy the channel is. What degree of reliability is needed, i.e., what maximum value of P_e is considered as acceptable, of course depends on the application. In practical systems, it may range anywhere from about 10^{-5} to 10^{-8}.

To achieve such small values of P_e even when the channel is noisy, it is necessary to make the data transmitted over the channel resistant to the tendency of the channel noise to cause decoding errors. Channel coding is the technique adopted for achieving this goal.

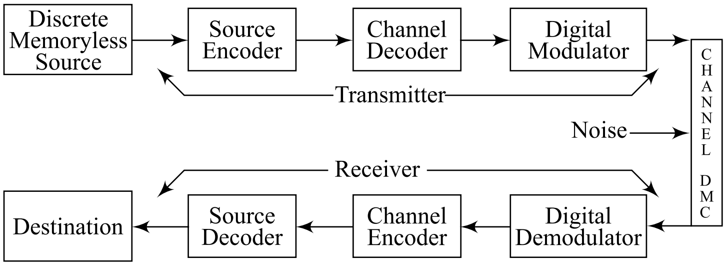


Fig. 12.18 *Block diagram of a digital communication system*

As shown in Fig.12.18, in the transmitter there is a channel encoder inserted between the source encoder and the carrier modulator. Correspondingly, there is a channel decoder in the receiver inserted between the carrier demodulator and the source decoder. The channel encoder introduces controlled redundancy into the data stream coming from the source encoder in order to make the data transmitted through the channel resistant to the effect of channel noise. The goal, of course, is to ensure that the data stream at the output of the channel decoder in the receiver is exactly identical to the one at the output of the source coder at the transmitting end. Thus, while the source coder attempts to remove the redundancy in the source output in order to improve the transmission *efficiency*, the channel encoder deliberately introduces some controlled redundancy in order to improve *reliability*.

Channel codes are broadly of two types – block codes and convolutional codes. Detailed discussion of channel coding is reserved for Chapter 7. As our objective here is only to briefly describe how channel coding introduces redundancy, block codes are most convenient for this purpose. When block codes are used, the channel encoder takes a block of k bits of the data stream at a time from the output of the source encoder and maps this block of k bits into a block of n bits ($n > k$) according to some pre-determined rule, thereby producing data at a higher rate. The $(n - k)$ additional bits, called the parity-check bits, constitute the redundancy introduced by the channel encoder in each block. At the source encoder output, if each bit occupies T_s seconds, then at the output of the channel encoder, each bit occupies only T_c seconds, where

$$T_c = \left(\frac{k}{n}\right)T_s \tag{12.61}$$

The ratio of the number of message bits to the total number of bits in a block, viz. k/n, is called the **code rate**.

$$\text{code rate} = r = \frac{k}{n} \tag{12.62}$$

Of course, r is less than one, and for a given k, it goes on decreasing as the redundancy per block is increased. In fact, smaller the value of P_e to be achieved, larger will be the value of n, and smaller will be the value of the code rate r. But, a small code rate implies inefficient transmission. So, the question naturally arises, 'Is it possible to have a code that enables one to achieve an arbitrarily small error rate which is prescribed, without making r, the code rate too low? Shannon's channel coding theorem, considered by many as his most valuable contribution, assures us that it is certainly possible to have such a code, but only under certain conditions.

If the channel has a capacity of C_s bits/symbol, we know from the way the channel capacity has been defined, that on the average, the maximum information that *each symbol* transmitted through the channel can carry, is C_s bits. If the symbols are transmitted through the channel at the rate of $(1/T_c)$, then the maximum rate of flow of information through the channel is (C_s/T_c) bits/second. Now, if the source encoder output has an entropy of $H(S)$ bits per output symbol of the source encoder, and if $(1/T_s)$ is the rate at which the source encoder is giving its output symbols, then the average rate at which information is being supplied by the source encoder is $H(S)/T_s$ bits per second.

12.8.1 Statement of Shannon's Channel Coding Theorem

Given a discrete memoryless source with an entropy of $H(S)$ bits per symbol emitting symbols at the rate of $(1/T_s)$ symbols per second, and given a discrete memoryless channel with a capacity of C_s bits per symbol and through which the symbols are transmitted at the rate of $(1/T_c)$ symbols per second, it is possible to construct a channel code which would make it possible to transmit the source symbols through the channel and be reconstructed with arbitrarily small probability of error, if and only if

$$\frac{H(s)}{T_s} \le \frac{C_s}{T_c} \tag{12.63}$$

Note that Shannon's theorem merely states that codes do exist that enable us to transmit information over a given channel with arbitrarily small probability of error as long as the rate of transmission (in bits/sec) is less than the channel capacity (in bits/sec) of the channel. *It does not tell us anything about how such codes may be constructed.*

12.8.2 Implications for a Binary Symmetric Channel

When Shannon's channel coding theorem is applied to a Binary Symmetric Channel (BSC), the condition for the existence of a channel code capable of achieving an arbitrarily low probability of error, reduces to the simple form

$$r \le C_s \tag{12.64}$$

where r is the code rate of the channel code (in bits/symbol) and C_s is the channel capacity of the BSC in bits per symbol. This may be shown as follows:

Suppose a DMS is emitting binary symbols, 0 and 1 with equal probability once every T_s seconds. Then the source entropy is 1 bit/symbol and the information rate equals $(1/T_s)$ bits/second. Let the channel encoder to which this source is connected, have a code rate of r and let it produce an encoded symbol every T_c seconds for transmission over a binary symmetric channel with a channel capacity of C_s bits/symbol. Since one encoded symbol passes through the channel every T_c seconds, the channel capacity may also be written as (C_s/T_c) bits per second. Now, applying Eq. (12.63) to this case, the condition for existence of a channel code that is capable of giving an arbitrarily small probability of error is

$$\frac{1}{T_s} \le \frac{C_s}{T_c} \quad \text{i.e.,} \quad \frac{T_c}{T_s} \le C_s$$

But $(T_c/T_s) = r$ = code rate

$$\therefore \qquad r \le C_s \tag{12.65}$$

12.9 CONTINUOUS SOURCES AND DIFFERENTIAL ENTROPY

Till now, the sources that we considered were producing ensembles of discrete random variables which could take any one of a finite number of discrete amplitudes. We had discussed the entropy of two discrete random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$. Finally, we developed the concept of 'Mutual Information', $I(\boldsymbol{X}; \boldsymbol{Y})$ of two discrete random variables. In this section, we shall try to extend all those concepts to the case of continuous sources, using appropriate modifications wherever necessary. Recall that we have defined the entropy of a discrete source producing a sequence of discrete random variables $\boldsymbol{X}_i$, where each $\boldsymbol{X}_i$ can take any of the possible discrete amplitudes x_k, $k = 0$ to $(M - 1)$ with probabilities $p(x_k)$, $k = 0$ to $(M - 1)$ respectively, as

$$H(X) \triangleq \sum_{k=0}^{M-1} p(x_k) \log_2 \left[\frac{1}{p(x_k)} \right] \tag{12.66}$$

A continuous source produces a set of continuous-time signals. The set of all such possible signals is assumed to be forming an ensemble of waveforms generated by an ergodic random process which is band limited, so that any realization $x(t)$ of this process is completely characterized by the samples taken at or above the Nyquist rate. Thus, at any sampling instant, the set of all possible sample values constitutes a continuous random variable $\boldsymbol{X}$ with a certain probability density function $f_X(x)$.

Consider now such a continuous random variable $\boldsymbol{X}$ with a probability density function $f_X(x)$.

From Fig. 12.19, it is clear that we may represent $p(x_k)$ of the discrete case by $f_X(x_k) \cdot \Delta x$ which represents the probability of the continuous random variable $\boldsymbol{X}$ taking any value between $\left(x_k - \frac{1}{2}\Delta x\right)$ to $\left(x_k + \frac{1}{2}\Delta x\right)$. In essence, we are only trying to represent a continuous random variable as the limiting case of a discrete random variable that takes discrete values x_ks separated by Δx, as Δx tends to zero. Thus, we may write the entropy of a continuous random variable $\boldsymbol{X}$ as

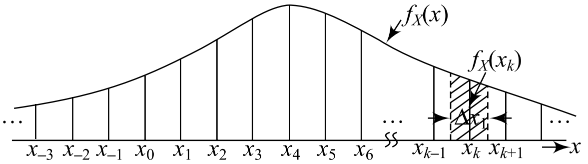


Fig. 12.19 *Approximating a continuous random variable by a discrete random variable*

$$H(\boldsymbol{X}) = \lim_{\Delta x \to 0} \sum_{k=-\infty}^{\infty} f_X(x_k)\Delta x \log_2\left[\frac{1}{f_X(x_k)\Delta x}\right]$$

$$= \lim_{\Delta x \to 0}\left[\sum_{k=-\infty}^{\infty} f_X(x_k)\log_2\left[\frac{1}{f_X(x_k)}\right]\Delta x - \log_2 \Delta x \sum_{k=-\infty}^{\infty} f_X(x_k)\Delta x\right]$$

$$= \int_{-\infty}^{\infty} f_X(x)\log_2\left[\frac{1}{f_X(x)}\right]dx - \lim_{\Delta x \to 0} \log_2 \Delta x \int_{-\infty}^{\infty} f_X(x)dx$$

But, $\int_{-\infty}^{\infty} f_X(x)dx = 1$

$$\therefore \quad H(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x)\log_2\left(\frac{1}{f_X(x)}\right)dx - \lim_{\Delta x \to 0} \log_2 \Delta x \tag{12.67}$$

In analogy with Eq. (12.66), if we define the entropy of a continuous random variable ***X***, as

$$h(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x)\log_2\left(\frac{1}{f_X(x)}\right)dx \tag{12.68}$$

Then we find that it differs from the value given in Eq. (12.67) by an amount equal to

$$\lim_{\Delta x \to 0} \log_2 \Delta x$$

This tends to infinity as $\Delta x \to 0$. Thus, the true or absolute entropy of a continuous random variable is infinitely large. Intuitively also this is what we expect since the continuous random variable takes a continuum of values and the probability of the event of its taking any particular value will be infinitesimally small, making the corresponding information of such an event infinitely large. Therefore, to avoid this problem, we call $h(\boldsymbol{X})$ as given by Eq. (12.68) as 'differential entropy'. This name is quite justified because the information transmitted over a channel is actually the difference between two absolute entropies. So, if we consider $H(\boldsymbol{X})$ as given by Eq. (12.67) as the absolute entropy of a continuous ransom variable, with the second term, viz., $\lim_{\Delta x \to 0} \log_2 \Delta x$, as a sort of common reference value, then the difference between two absolute entropies will be the same as the difference between the corresponding differential entropies. Hence, hereafter we shall use $h(\boldsymbol{X})$ as given by Eq. (12.68) as the *differential entropy* of the continuous random variable ***X***.

Example 12.25 A continuous random variable ***X*** is uniformly distributed over the interval –*a*/2 to +*a*/2 so that

$$f_X(x) = \begin{cases} \frac{1}{a}; & -\frac{a}{2} < x < \frac{a}{2} \\ 0; & \text{otherwise} \end{cases}$$

Find the differential entropy of the random variable ***X***.

Solution $h(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x)\log_2\left(\frac{1}{f_X(x)}\right)dx$

But,

$$f_X(x) = \begin{cases} \frac{1}{a} & \text{for } -\frac{a}{2} < x < \frac{a}{2} \\ 0 & \text{otherwise} \end{cases}$$

$\therefore$ $$h(\boldsymbol{X}) = \int_{-a/2}^{a/2} \frac{1}{a}\log_2\left(\frac{1}{(1/a)}\right)dx = \log_2 a$$

Remark *Since $\log_2 x$ will be negative for $x < 1$, this simple example shows that unlike in the case of $H(\boldsymbol{X})$ of discrete sources, $h(\boldsymbol{X})$ of a continuous source can be negative.*

Example 12.26 Show that the differential entropy of a continuous random variable having Gaussian distribution is given by

$$h(\boldsymbol{X}) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

where σ^2 is the variance of $\boldsymbol{X}$.

Solution Let $\boldsymbol{X}$ have a probability density function given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$$

Then,

$$h(\boldsymbol{X}) = -\int_{-\infty}^{\infty} f_X(x)[\log_2 f_X(x)]dx$$

But $$\log_2 f_X(x) = \log_2 e[\log_e f_X(x)]$$

$$= \log_2 e\left[\log_e\left\{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-a)^2/2\sigma^2}\right\}\right]$$

$$= \log_2 e\left[\log_e\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-a)^2}{2\sigma^2}\right] = -\log_2 e\left[\log_e\left(\sqrt{2\pi\sigma^2}\right) + \frac{(x-a)^2}{2\sigma^2}\right]$$

$\therefore$ $$h(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x)\cdot\log_2 e\cdot\log_e\sqrt{2\pi\sigma^2}\,dx + \int_{-\infty}^{\infty}\log_2 e\left\{\frac{(x-a)^2}{2\sigma^2}\cdot f_X(x)\right\}dx$$

$$= \log_2\sqrt{2\pi\sigma^2}\int_{-\infty}^{\infty} f_X(x)dx + \frac{1}{2\sigma^2}\log_2 e\int_{-\infty}^{\infty}(x-a)^2 f_X(x)dx$$

But $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and by a change of variable, $\int_{-\infty}^{\infty}(x-a)^2 f_X(x)dx$ can be shown to be equal to σ^2

$\therefore$ $$h(\boldsymbol{X}) = \log_2\sqrt{2\pi\sigma^2} + \frac{1}{2}\log_2 e$$

$$= \frac{1}{2}\log_2(2\pi\sigma^2) + \frac{1}{2}\log_2 e = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Remarks

(*i*) *$h(\boldsymbol{X})$ will be negative if $(2\pi e\sigma^2) < 1$.*

(*ii*) *The differential entropy of a Gaussian continuous random variable $\boldsymbol{X}$ is completely determined by its variance σ^2.*

(*iii*) *It can be shown (see Example 12.27) that of all random variables with the same variance, a Gaussian random variable has maximum differential entropy.*

Example 12.27 Show that the differential entropy $h(\boldsymbol{X})$ of a continuous random variable $\boldsymbol{X}$ is maximum when it has a Gaussian distribution and that this maximum value is $\frac{1}{2}\log_2[2\pi e\sigma^2]$

Solution Let σ^2 be the variance and A be the mean of the random variable $\boldsymbol{X}$. Then we know that

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{i}$$

and

$$\int_{-\infty}^{\infty} (x-A)^2 f_X(x)dx = \sigma^2 \tag{ii}$$

where $f_X(x)$ is the density function of the random variable $\boldsymbol{X}$.
From Eq. (12.68), we know that the differential entropy of $\boldsymbol{X}$ is

$$h(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x)\log_2\left(\frac{1}{f_X(x)}\right)dx \tag{iii}$$

we have to maximize this differential entropy with respect to the probability density function $f_X(x)$ under the two constraints given by Eqs. (i) and (ii) above. For this purpose, let us make use of 'Lagrange multipliers' method. As detailed in Annexure D, we proceed as follows.

$$g[f_X(x), \lambda_1, \lambda_2] = h(X) + \lambda_1\left[\int_{-\infty}^{\infty} f_X(x)dx - 1\right] + \lambda_2\left[\int_{-\infty}^{\infty} (x-A)^2 f_X(x)dx - \sigma^2\right]$$

where λ_1 and λ_2 are the Lagrange undetermined multipliers. For maximizing $g(\cdot)$, we take the partial derivation of $g(\cdot)$ with respect to $f_X(x)$ and equate it to zero after substituting for $h(\boldsymbol{X})$ using Eq. (iii).

i.e., $$\frac{\partial g[f_X(x), \lambda_1, \lambda_2]}{\partial f_X(x)} = -\log_2 f_X(x) - \log_2 e + \lambda_1 + \lambda_2(x-A)^2 = 0$$

Solving the above for $f_X(x)$, we get

$$f_X(x) = \exp\left[-1 + \frac{\lambda_1}{\log_2 e} + \frac{\lambda_2(x-A)^2}{\log_2 e}\right] \tag{iv}$$

In the above expression, unless λ_2 is negative, when we substitute the above expression for $f_X(x)$ in Eqs. (i) and (ii) the integrals will not converge.
If we represent

$$\exp\left[-1 + \frac{\lambda}{\log_2 e}\right] \text{ by } \alpha$$

and $$\frac{\lambda_2}{\log_2 e} \text{ by } -\beta^2,$$

Equation (iv) may be represented as

$$f_X(x) = \alpha e^{-\beta^2(x-A)^2} \tag{v}$$

Substituting for $f_X(x)$ in (i) using Eq. (v), we have

$$= \alpha\int_{-\infty}^{\infty} e^{-\beta^2(x-A)^2}dx = 1$$

Evaluating the left-hand side integral by putting $(x-A)\beta = y$,

$$\beta dx = dy \qquad \therefore dx = \frac{1}{\beta} dy$$

$$= \left(\frac{\alpha}{\beta}\right) \int_{-\infty}^{\infty} e^{-y^2} dx = 1 = \frac{\alpha}{\beta}\sqrt{\pi}$$

$$\therefore \qquad \left(\frac{\alpha}{\beta}\right) = \frac{1}{\sqrt{\pi}} \tag{vi}$$

Substituting for $f_X(x)$ in Eq. (ii) using Eq. (v), we get

$$\alpha \int_{-\infty}^{\infty} (x - A) e^{-\beta^2 (x-A)^2} dx = \sigma^2 \tag{vii}$$

or

$$\left(\frac{\alpha}{\beta^3}\right) = \frac{2\sigma^2}{\sqrt{\pi}}$$

From Eqs. (vi) and (vii), dividing one by the other

$$\beta^2 = \frac{1}{2\sigma^2} \quad \therefore \beta = \frac{1}{\sqrt{2}\sigma}$$

Substituting for β in Eq. (vi), we get $\alpha = \dfrac{1}{\sqrt{2\pi}\sigma}$

Now, substituting the values of α and β in Eq. (v),

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-A)^2/2\sigma^2}$$

This is a Gaussian density function. Hence, $h(\boldsymbol{X})$, the differential entropy of the random variable $\boldsymbol{X}$, has a maximum value when $\boldsymbol{X}$ is Gaussian.

It has already been shown in Example 12.26 that for a given variance σ^2 for $\boldsymbol{X}$ which is Gaussian, the value of $h(\boldsymbol{X})$ is equal to

$$h(X) = \frac{1}{2}\log_2 2\pi e \sigma^2$$

12.9.1 Mutual Information

Analogous to the way we had defined the mutual information between two discrete random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ (see Eq. (12.51a)), we now define the mutual information between a pair of continuous random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$I(\boldsymbol{X};\boldsymbol{Y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\log_2\left[\frac{f_X(x|y)}{f_X(x)}\right] dxdy \tag{12.69}$$

where $f_X(x|y)$ represents the conditional density function of $\boldsymbol{X}$ given that $\boldsymbol{Y} = y$ and $f_{X,Y}(x, y)$ is the joint probability density function. While the mutual information $I(\boldsymbol{X}; \boldsymbol{Y})$ in the case of a *discrete* memoryless channel represents the average information in bits transferred over the channel for each received *symbol*, in the case of *continuous* channels, it represents the average information in bits transferred over the channel per each *sample value* of the signal received at the channel output.

Now, analogous to the expression for the conditional entropy of $\boldsymbol{Y}$ given $\boldsymbol{X}$ (where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are discrete random variables) as given in Eq. (12.20), we write down the expression for the conditional differential entropy of a continuous random variable $\boldsymbol{Y}$ given the continuous random variable $\boldsymbol{X}$, as

$$H(\boldsymbol{Y}|\boldsymbol{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\log_2\left[\frac{1}{f_Y(\boldsymbol{Y}|\boldsymbol{X})}\right]dxdy \tag{12.70}$$

We now state the properties of the mutual information between continuous random variables ***X*** and ***Y***. They are exactly similar to the properties of mutual information between two discrete-time random variables ***X*** and ***Y***, which we had stated in Eqs. (12.51), (12.53) (12.49) and (12.50), respectively.

Properties

1. $I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X})$ (12.71)
2. $I(\boldsymbol{X};\boldsymbol{Y}) \geq 0$ (12.72)
3. $I(\boldsymbol{X};\boldsymbol{Y}) = h(\boldsymbol{X}) - h(\boldsymbol{X}|\boldsymbol{Y})$

   $= h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X})$ (12.73)

12.9.2 Shannon's Third Theorem — Information Capacity Theorem

Shannon's Information Capacity Theorem is also known as Shannon's Third Theorem, or sometimes as Shannon–Hartley Theorem. It deals with the channel capacity of a band limited and power limited continuous channel which is corrupted by additive white Gaussian noise of zero mean. It is one of the important results of information theory as it shows the relationship between the channel capacity and the three key parameters—channel bandwidth, average transmitted power and the power spectral density of the white noise on the channel. We will be discussing about the importance of this theorem and its implications after first deriving it.

The input to the channel is in the form of a sequence of the samples of a zero-mean Gaussian distributed stationary random process $x(t)$ band limited to B Hz. These samples are taken uniformly exactly at the Nyquist rate of $2B$ samples per second. If they are taken more frequently than the Nyquist rate, they do not contain any additional information. These samples enable us to reconstruct the input signal without any distortion in the absence of noise. It is assumed that the average signal power is constrained to be S. Since the channel is band limited to B Hz, the received signal as well as noise are also band limited to B Hz. If the input sample is a continuous zero-mean Gaussian random variable, ***X***, the corresponding output sample is a Gaussian random variable ***Y*** given by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N} \tag{12.74}$$

where ***N*** is a sample of one realization of a zero-mean Gaussian noise process which is bandlimited to B Hz. Since the signal and noise are statistically independent, their average powers add to give the average power of ***Y***.

$$\therefore \qquad E[\boldsymbol{Y}^2] = E[\boldsymbol{X}^2] + E[\boldsymbol{N}^2] \tag{12.75}$$

Further, ***Y*** is also a Gaussian random variable with a mean that is zero and a variance that is equal to the sum of the variances of ***X*** and ***N***, i.e., $E[\boldsymbol{Y}^2]$.

As per Eq. (12.73), the mutual information of the channel is given by

$$I(\boldsymbol{X};\boldsymbol{Y}) = h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}) \text{ bits/sample}$$

The channel capacity C is the maximum value of this mutual information under the constraint that the average signal power has to be equal to S, the maximization being with respect to the probability density function of ***X***.

$$\text{i.e.,} \qquad C = \underset{f_X(x)}{I_{\max}}[I(\boldsymbol{X};\boldsymbol{Y}): E(\boldsymbol{X}^2) = S] \text{ bits/sample} \tag{12.76}$$

Before discussing further about this constrained maximization, we shall first determine the conditional differential entropy, $h(\boldsymbol{Y} | \boldsymbol{X})$. By analogy with Eq. (12.48), we may write

$$h(\boldsymbol{Y}|\boldsymbol{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\log\left[\frac{1}{f_Y(y|x)}\right]dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x) f_Y(y|x)\log\left[\frac{1}{f_Y(y|x)}\right]dxdy$$

$$= \int_{-\infty}^{\infty} f_X(x)dx \int_{-\infty}^{\infty} f_Y(y|x)\log\left[\frac{1}{f_Y(y|x)}\right]dy$$

$$\therefore \qquad h(\boldsymbol{Y}|\boldsymbol{X}) = \int_{-\infty}^{\infty} f_Y(y|x)\log\left[\frac{1}{f_Y(y|x)}\right]dy, \quad \text{since} \quad \int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{12.77}$$

Now, from Eq. (12.74), we have

$$y = x + n$$

So, given $\boldsymbol{X} = x$, the output sample random variable $\boldsymbol{Y}$ will have a distribution that is the same as that of $\boldsymbol{N}$. If $f_N(n)$ denotes the probability density function of the noise sample $\boldsymbol{N}$, then we have:

$$f_Y(y|x) = f_N(y - x)$$

$$\therefore \qquad h(\boldsymbol{Y}|\boldsymbol{X}) = \int_{-\infty}^{\infty} f_N(y - x)\log\left[\frac{1}{f_N(y - x)}\right]dy \qquad \text{(From Eq.(12.77))}$$

If we denote $(y - x)$ by say z,

$$h(\boldsymbol{Y}|\boldsymbol{X}) = \int_{-\infty}^{\infty} f_N(z)\log\left[\frac{1}{f_N(z)}\right]dy = h(\boldsymbol{N}) \tag{12.78}$$

$$\therefore \qquad I(\boldsymbol{X};\boldsymbol{Y}) = h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}) = h(\boldsymbol{Y}) - h(\boldsymbol{N}) \text{ bits/sample}$$

$$\text{i.e.,} \qquad I(\boldsymbol{X};\boldsymbol{Y}) = [h(\boldsymbol{Y}) - h(\boldsymbol{N})] \text{ bits/sample} \tag{12.79}$$

Since $h(\boldsymbol{N})$, the differential entropy of the channel noise sample is independent of $f_X(x)$, maximization of the mutual information $I(\boldsymbol{X}; \boldsymbol{Y})$ with respect to $f_X(x)$, as required by Eq. (12.76), to determine the channel capacity of this Gaussian channel which is band limited and power limited, can be achieved by maximizing $h(\boldsymbol{Y})$ with respect to $f_X(x)$. We had observed in remark iii under Example 12.26 that $h(\boldsymbol{Y})$ will be maximum, for a given variance of $\boldsymbol{Y}$, only when $\boldsymbol{Y}$ is Gaussian distributed, But $\boldsymbol{Y}$ can be a Gaussian distributed random variable if and only if $\boldsymbol{X}$, the input random variable is Gaussian. That is why we have assumed in the beginning that the input signal $x(t)$ is a zero-mean Gaussian distributed stationary random process with an average power of S. So, $I(\boldsymbol{X}; \boldsymbol{Y})$ is maximized.

Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent, as stated in Eq. (12.75), the total average power of $\boldsymbol{Y}$ is

$$E[\boldsymbol{Y}^2] = E[\boldsymbol{X}^2] + E[\boldsymbol{N}^2] = S + N \tag{12.80}$$

where we have used $\boldsymbol{N}$ to denote $E[\boldsymbol{N}^2]$. Hence, the maximum value of the differential entropy $h(\boldsymbol{Y})$ of $\boldsymbol{Y}$ is

$$\underset{\max}{h(\boldsymbol{Y})} = \frac{1}{2}\log_2[2\pi e(S + N)] \qquad \text{(As shown in Example 12.26)}$$

$$\text{But} \qquad I_{\max}(\boldsymbol{X};\boldsymbol{Y}) = h_{\max}(\boldsymbol{Y}) - h(\boldsymbol{N})$$

(as per Eq. 12.79)

$$\therefore \qquad I_{\max}(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\log[2\pi e(S + N)] - h(\boldsymbol{N}) \tag{12.81}$$

But, since the channel noise is Gaussian, $h(\boldsymbol{N})$ is given by (see Example 12.26)

$$h(\boldsymbol{N}) = \frac{1}{2}\log 2\pi eN$$

$$\therefore \qquad I_{\max}(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2}\log_2 2\pi e(S+N) - \frac{1}{2}\log_2 2\pi eN$$

$$= \frac{1}{2}\log_2\left[\frac{2\pi e(S+N)}{2\pi eN}\right] = \frac{1}{2}\log_2\left(1+\frac{S}{N}\right) \text{bits/sample}$$

Since $2B$ samples are transmitted per second over the channel, *assuming* these samples of $x(t)$ taken at Nyquist rate to be statistically independent, we may write

$$C = 2BI_{\max}(\boldsymbol{X};\boldsymbol{Y}) = B\log_2\left(1+\frac{S}{N}\right) \text{bits/second}$$

Hence, the channel capacity of the band limited and power limited Gaussian channel with input power constrained to S, is given in bits per second as

$$\boxed{C = B\log_2(1+S/N) \text{ bits/second}} \qquad (12.82)$$

This is called Shannon–Hartley law, or Shannon's information capacity theorem.

Recall that we have assumed in the course of this derivation that the input samples are statistically independent. *This will be possible if and only if the power spectrum of the input process X(t) is constant, i.e., if and only if the input signal is not only Gaussian but is also white*. Then and only then, can the maximum rate of information transfer as given by Eq. (12.82) be possible.

Remarks

(i) In Eq. (12.82), N, the average noise power = 2B × η/2 = ηB.

(ii) While the channel capacity is linearly related to the bandwidth B, it is logarithmically related to the signal to noise ratio (S/N).

(iii) Because of point (ii) above, it is easier to increase the capacity of a given communication channel by increasing its bandwidth rather than by increasing the transmitted power.

(iv) Combining Shannon's channel coding theorem with Shannon–Hartley law, we find that for a given average transmitted power and bandwidth B, it would be possible to transmit information over the Gaussian channel at a rate of C bits/sec as given by Eq. (12.82) with arbitrarily small probability of error and that it is not possible to transmit at a rate greater than C given by this equation.

(v) Shannon's third theorem, or the information capacity theorem clearly sets a fundamental limit on the rate at which error-free transmission can be achieved using a power limited, band limited Gaussian channel.

(vi) As mentioned earlier, the channel capacity, as given by Eq. (12.82) can be approached only if the input signal is not only Gaussian, but also white.

Example 12.28 An analog signal having 4 kHz bandwidth is sampled at 1.25 times the Nyquist rate and each sample is quantized into one of 256 equally likely levels. Assume that the successive samples are statistically independent

(a) What is the information rate of this source?

(b) Can the output of this source be transmitted without error over an AWGN channel with a bandwidth of 10 kHz and an (*S*/*N*) ratio of 20 dB?

(c) Find the *S*/*N* ratio required for error-free transmission for part (a)

(d) Find the bandwidth required for an AWGN channel for error-free transmission of the output of this source if the *S*/*N* ratio is 20 dB? (**UPSC, IES Examination, 1999**)

Solution

(a) Nyquist rate = $2 \times f_m = 2 \times 4 \times 10^3$ = 8 kHz

$$\therefore \quad \text{sampling rate} = 1.25 \times 8 \times 10^3 = 10 \text{ ksps} = f_s$$

$$H(S) = -\sum_i p(x_i) \log_2 p(x_i) \text{ bits/symbol}$$

Since there are 256 equally likely levels, the probability of any particular level = (1/256) = $p(x_i)$.

$$-\log_2\left(\frac{1}{256}\right) = \log_2 256 = 8 \text{ bits/sample}$$

$$\therefore \quad \underset{\text{bits/sample}}{H(S)} = 256 \times \frac{1}{256} \times 8 = 8 \text{ bits/sample}$$

Since the sampling rate is 10 kilo samples/second, and the entropy is 8 bits/sample, the information rate from the source is given by

$$\underset{\text{bits/sec}}{H(S)} = f_s \times \underset{\text{bits/sample}}{H(S)} = 10^4 \times 8 = 80 \text{ kb/sec}$$

∴ information rate from the source = 80 k bits/sec.

(b) The output of this source can be transmitted over the given AWGN channel without error only if the source information rate, i.e., 80 k bits/sec is less than the channel capacity of the given channel. So first let us find the channel capacity C. For this, we use Shannon–Hartley law.

$$C = B \log_2\left(1 + \frac{S}{N}\right) \text{ bits/sec}$$

$B = 10^4$ and (S/N) = 20 dB = 100 (ratio)

$$\therefore \quad C = 10^4 \log_2(1+100) = 10^4 \log_2 101 = 10^4\left(\frac{\log_{10} 101}{\log_{10} 2}\right) = 6.6588 \times 10^4$$

$$\therefore \quad C = 66.6 \text{ k bits/sec}$$

∴ since the source rate is higher than the channel capacity, it is not possible to transmit the source output over this channel without error.

(c) For transmitting the source output over the channel, the channel capacity should be above 80 kb/sec.

$$\therefore \quad 80 \times 10^3 = 10^4 \log_2\left(1 + \left(\frac{S}{N}\right)_{\min}\right) \text{ bits/sec}$$

$$\therefore \quad 2^8 = 1 + \left(\frac{S}{N}\right)_{\min} \qquad \therefore \left(\frac{S}{N}\right)_{\min} = 256 - 1 = 255 \text{ (ratio)}$$

$$\therefore \quad \left(\frac{S}{N}\right) \geq 10 \log_2 255 \text{ dB} = 24.065 \text{ dB}$$

(d) $80 \times 10^3 = B \log_2\left(1 + \frac{S}{N}\right)$ but S/N = 20 dB = 100 (ratio)

$$B = \left[\frac{80 \times 10^3}{\log_2 101}\right] = \frac{80 \times 10^3}{6.659} = 12.01 \text{ kHz}$$

Example 12.29 A Gaussian channel has 1.5 MHz bandwidth. If the signal power-to-noise power spectral density ratio is $(S/\eta) = 10^4$ Hz, determine the channel capacity C.

Solution $C = B\log_2\left(1+\frac{S}{N}\right) = B\log_2\left(1+\frac{S}{\eta B}\right)$

$$= 1.5\times10^6 \log_2\left(1+\frac{10^4}{1.5\times10^6}\right) = 14.38 \text{ k bits/sec}$$

Example 12.30 Using 4 × 10^5 pixels per frame, with each pixel taking any one of the 10 different brightness levels with equal probability, a black and white TV transmitter transmits 25 frames per second. If the signal to noise ratio is 30 dB, determine the minimum bandwidth required to transmit the video signal.

Solution Number of pixels transmitted per second = $4\times10^5\times25 = 10^7$ pixels/sec.
There are 10 different brightness levels.

$$p(x_i) = (1/10) = 0.1 \qquad \text{(Since they are all equally likely)}$$

$$\therefore \quad \text{information per pixel} = -\log_2 0.1 = \log_2 10 = \frac{1}{0.3010} = 3.322 \text{ bits}$$

Assuming all the pixels to be statistically independent, the amount of information transmitted per second

$$I = 10^7\times3.322 \text{ bits}$$

Applying Shannon–Hartley law, and noting that 30 dB = 10^3 as a ratio

$$10^7\times3.322 = B\log_2(1+10^3) = B\times9.96822$$

$$\therefore \quad B = \frac{10^7\times3.322}{9.96822} = 3332591 \text{ kHz} = 3.332591 \text{ MHz}$$

12.9.3 Ideal System and Shannon Limit

We will now define an ideal system, the objective being to use it for assessing the performance of practical systems by determining how close they come to the ideal one.

Definition An ideal system is defined as one that transmits data at a bit rate that is equal to the channel capacity *C*, in bits per second.

Suppose the energy per bit is E_b. Then the average power transmitted is

$$S = E_b\cdot C \tag{12.83}$$

We know that if *B* is the bandwidth,

$$\frac{S}{N} = \frac{E_b\cdot C}{(\eta/2)(2B)} = \frac{E_b\cdot C}{\eta B} \tag{12.84}$$

where $\eta/2$ W/Hz is the two-sided PSD of the white noise on the channel. Substituting for (*S*/*N*) in Shannon–Hartley law using Eq. (12.84), we get

$$C = B\log_2\left(1+\frac{E_b}{\eta}\cdot\frac{C}{B}\right) \text{ bits/sec} \tag{12.85}$$

$$\therefore \quad \frac{E_b}{\eta} = \frac{2^{C/B}-1}{C/B} = \frac{B}{C}(2^{C/B}-1) \tag{12.86}$$

This expression is useful in establishing the performance of the ideal system defined earlier.

$$\text{When } B >> C,\ 2^{C/B} = e^{(C/B)\log_e 2} \tag{12.87}$$

Using the approximation $e^x \approx 1 + x$ when $x << 1$, we get

$$2^{C/B} = e^{(C/B)\ln 2} \cong 1 + \left(\frac{C}{B}\right)\ln 2 \tag{12.88}$$

Substituting for $2^{C/B}$ in Eq. (12.86) by making use of Eq. (12.88),

$$\frac{E_b}{\eta} \cong \frac{\left(1 + \frac{C}{B}\ln 2\right) - 1}{\left(\frac{C}{B}\right)} = \ln 2 = -1.6 \text{ dB} \tag{12.89}$$

This is called the Shannon Limit for an AWGN channel.

Using Eq. (12.86) with R_b replacing C, we may plot the relation between (E_b/η) and (R_b/B). This plot is shown in Fig.12.20. When $R_b = C$, we get the ideal system. Observe that the curve corresponding to $R_b = C$, (i.e., the ideal system) is the boundary between two distinct regions. The region, marked $R_b < C$ is the one in which it is possible to make the probability of error arbitrarily small and is the one which is of interest in practice. The other region, marked $R_b > C$ is the one in which it is not possible to achieve arbitrarily small probabilities of error.

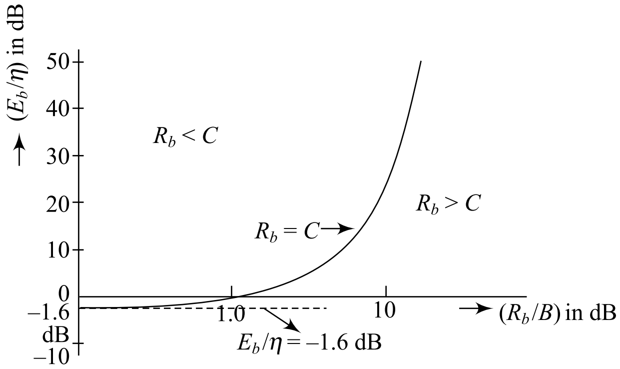


Fig. 12.20 *Relationship between (E_b/η) and (R_b/B) for an AWGN channel*

From the figure, it is evident that when (R_b/B) is large, a very large value of (E_b/η) is needed to keep $R_b < C$; i.e., a large average power is needed. But, suppose the bit-rate R_b is fixed and the channel bandwidth is increased, so that $B >> R_b$. In this case, operation in the region $R_b < C$ requires only that E_b/η be only slightly greater than the Shannon limit, i.e., –1.6 dB. The corresponding signal power required is

$$S \approx R_b \eta \log_e 2 \quad \text{W} \tag{12.90}$$

This represents the minimum signal power required for operation in the $R_b < C$ region. For *power-limited operation*, therefore, the bandwidth should be very large compared to the bit rate so that only minimum signal power represented in Eq. (12.90) is required.

But, suppose the system has to operate with $R_b >> B$. Then, from the figure, it is clear that (E_b/η) necessary for operation in the $R_b < C$ region, is quite large. Then the system is said to be operating in the *bandwidth-limited condition*.

12.9.4 Information Capacity Theorem and Some Practical Systems

Hartley–Shannon law, or the information capacity theorem tells us how trade-off is possible between *SNR* and bandwidth in an ideal system. It would therefore be interesting to see how some of the practical systems behave in this respect and compare them with the ideal system.

1. PCM: Let us first take up an M-ary PCM and compare its performance with that of the ideal system. With $M = 2$, we can deduce the result for binary PCM.

Let us say the message signal is quantized into L levels and each quantized sample is encoded into $\log_M L$ number of M-ary pulses. Assuming the message signal to be band limited to B Hz, we sample it at Nyquist rate of $2B$ samples/sec. Since $\log_M L$ M-ary pulses are used for representing each sample and $2B$ samples are produced per second, the number of M-ary pulses produced per second is

$$R_M = 2B \cdot \log_M L \tag{12.91}$$

Therefore the transmission bandwidth may be taken to be

$$B_T = \text{Half of } R_M = B \log_M L \tag{12.92}$$

In Eq. (4.120), we had shown that if the levels a_ks are as

$$a_k = \pm\frac{A}{2}, \pm 3\frac{A}{2}, \pm 5\frac{A}{2}, \ldots, \pm(M-1)\frac{A}{2}$$

with the basic level at $A/2$, and if the M-levels are equally likely, their mean-square value $\overline{a_k^2}$ is

$$\overline{a_k^2} = \frac{(M^2-1)}{12} \cdot A^2 = \frac{(M^2-1)}{3} \cdot \left(\frac{A}{2}\right)^2 \qquad \text{(From Eq. (4.120))}$$

If E_p denotes the energy of the pulse with basic level $A/2$, we know that

$$E_p = \left(\frac{A}{2}\right)^2 \cdot \frac{1}{R_M} = \left(\frac{A}{2}\right)^2 \cdot T_M \tag{12.93}$$

$$\therefore \qquad \text{average input power } \overline{a_k^2} = S_R = \frac{(M^2-1)}{3} \cdot E_p \cdot R_M \tag{12.94}$$

Also, input noise power N_R is given by

$$N_R = \left(\frac{\eta}{2}\right) 2B_T = \eta B_T$$

where B_T is the transmission bandwidth.

But from Eq. (12.92),

$$B_T = \frac{R_M}{2}$$

$$\therefore \qquad N_R = \eta \cdot \frac{R_M}{2} \tag{12.95}$$

As shown in Eq. (12.91), we are transmitting $2B \log_M L$ pulses (M-ary pulses) per second and each of these pulses is carrying an information of $\log_2 M$ bits. Hence, the rate of transmission of information is equal to R_b bits/sec where,

$$\begin{aligned} R_b &= (2B \log_M L)(\log_2 M) \\ &= 2B_T \log_2 M = B_T \log_2 M^2 \text{ bits/sec} \end{aligned} \tag{12.96}$$

Now, from Eq. (12.94), we have

$$M^2 = \left(1 + \frac{3S_R}{E_p \cdot R_M}\right),$$

and from Eq. (12.95), we have

$$N_R = \left(\frac{2N_R}{\eta}\right)$$

$\therefore$
$$M^2 = \left(1 + \frac{3S_R\eta}{2E_p \cdot N_R}\right)$$

Substituting this expression for M^2 in Eq. (12.96), we get

$$R_b = B_T \log_2\left[1 + \frac{3S_R\eta}{2E_pN_R}\right] \text{bits/sec} \tag{12.97}$$

Now, from Eq. (4.119)

$$P_e = 2\left(1 - \frac{1}{M}\right)Q\left(\frac{A}{2\sigma}\right) \approx 2Q\left(\frac{A}{2\sigma}\right) \quad \text{when } M >> 1. \tag{12.98}$$

But, from Eq.(4.122)

$$\left(\frac{A^2}{4\sigma^2}\right) \cong \left(\frac{6}{M^2}\right) \cdot \left(\frac{S_R}{\eta R_M}\right)$$

But, $\frac{\eta R_M}{2} = N_R$ (From Eq. (12.95))

$\therefore$
$$\left(\frac{A^2}{4\sigma^2}\right) \cong \frac{3}{M^2} \cdot \frac{S_R}{N_R} = \left(\frac{3}{M^2}\right)\left(\frac{M^2}{3} \cdot \frac{E_pR_M}{N_R}\right) \quad \text{(From Eq. (12.94))}$$

$$E_p \cdot \frac{R_M}{N_R} = \frac{E_p \cdot 2}{\eta} \quad \text{(From Eq. (12.95))}$$

$\therefore$ substituting this in Eq. (12.98), we get

$$P_e = 2Q\left(\sqrt{\frac{2E_p}{\eta}}\right) \text{when } M >> 1 \tag{12.99}$$

If $P_e \leq 10^{-6}$, we can consider the reception to be almost error-free.

But
$$P_e = 2Q\left(\sqrt{\frac{2E_p}{\eta}}\right) = 10^{-6} \Rightarrow \frac{2E_p}{\eta} \cong 24 \tag{12.100}$$

(from Q-function table)

$\therefore$ substituting 24 for $\frac{2E_p}{\eta}$ in Eq. (12.97), we get

$$R_b = B_T \log_2\left[1 + \frac{1}{8}\left(\frac{S_R}{N_R}\right)\right] \text{bits/sec} \tag{12.101}$$

Thus, we find that for $M >> 1$, an M-ary PCM can transmit information at the rate of R_b over a channel of bandwidth B_T. However, as per the information capacity theorem, over the same channel bandwidth of B_T and with the same (S_R/N_R), an ideal system can transmit at the rate of

$$C = B_T \log_2\left[1 + \frac{S_R}{N_R}\right] \text{bits/sec} \tag{12.102}$$

Equations (12.101) and (12.102) clearly indicate that while operating slightly above the threshold, M-ary PCM requires eight times as much power as the ideal system to provide the same rate of transmission of information. This means that M-ary PCM ($M >> 1$) even when operating slightly above the threshold, is 9 dB inferior to the ideal system. A plot of (R_b/B_T) vs (S_R/N_R) is shown in Fig. 12.21 for different values of M alongside the curve for an ideal system.

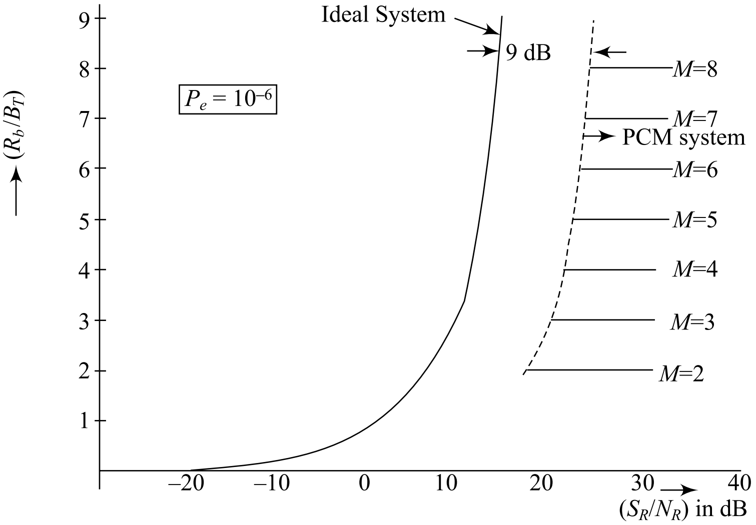


Fig. 12.21 *Comparison of performance of a PCM system with that of an ideal system*

Thus, in the case of M-ary PCM, as M is increased the bandwidth efficiency as well as the (S_R/N_R) required for operation above the threshold, also increase.

2. M-ary PSK: We now compare the bandwidth-power trade-off possible using M-ary PSK with that obtained for an ideal system. For this purpose, we consider a coherent M-ary PSK system that makes use of M non-orthogonal phase-shifted signals. We know that each of these M signals represents a particular symbol which contains an information of $\log_2 M$ bits. From Eq. (5.196), the bandwidth efficiency of M-ary PSK is given by

$$\rho = \frac{R_b}{B_T} = \log_2 M \tag{12.103}$$

Note *We defined the bandwidth as 3-dB width of the main lobe. If we define it as full main-lobe width, then $\rho = 0.5 \log_2 M$.*

The probability of symbol error for M-ary PSK is given by

$$\underset{\text{(symbol error)}}{P_e} \cong \text{erfc}\left[\sqrt{\frac{E}{\eta}} \sin\left(\frac{\pi}{M}\right)\right] \tag{12.104}$$

As in the previous case, we shall consider the transmission to be almost error-free when $P_e \le 10^{-6}$. For this P_e, the values of (R_b/B_T) are plotted against the values of (E_b/η) for M = 2, 4, 8, 16, 32 and 64 in Fig.12.22. The curve for the ideal system is also shown.

From Fig. 12.22, we observe that as M is increased, the bandwidth efficiency, (R_b/B_T) increases, but the (E_b/η) value required for error-free transmission moves away from the Shannon limit.

3. M-ary FSK: We will now consider a coherent M-ary FSK system employing orthogonal signals by arranging adjacent frequencies to differ by $1/2T$ Hz, where T is the symbol period. Again, each of these M-signals represents a particular symbol which contains an information of $\log_2 M$ bits.

We know that the bandwidth efficiency of M-ary FSK is given by

$$\rho = \frac{R_b}{B_T} = \frac{2 \log_2 M}{M} \quad \text{(see Eq. (10.131))} \tag{12.105}$$

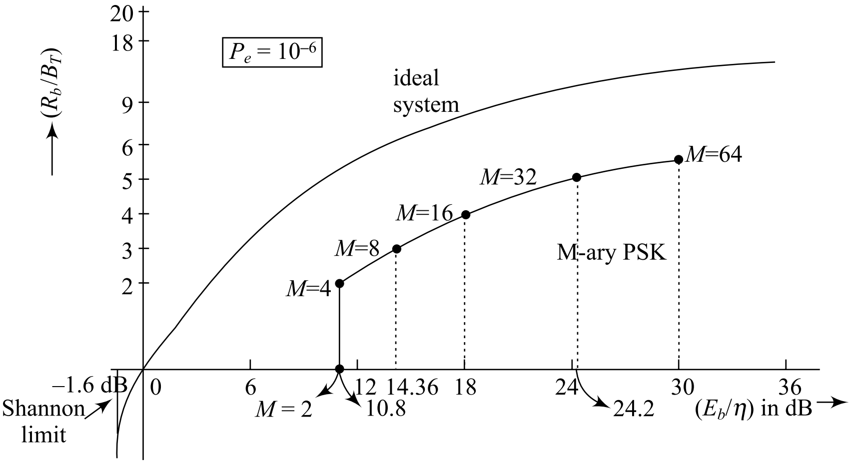


Fig. 12.22 *Comparison of M-ary PSK with the ideal system*

and $$P_e = \frac{1}{2}(M-1)\,erfc\sqrt{\frac{E}{2\eta}} \quad \text{(see Eq. (10.130))} \tag{12.106}$$

We also know that the following relations hold good:

T = Symbol period = $T_b \log_2 M$ where T_b is bit duration

and E = Symbol energy = $E_b \log_2 M$ where E_b is bit energy.

As in the case of M-ary PSK, here too, we shall assume the transmission to be totally error-free if $P_e \le 10^{-6}$. So, for this $P_e \le 10^{-6}$, the (E_b/η) required and the bandwidth efficiency (R_b/B_T) are determined for each value of M-ary FSK. A plot of the bandwidth efficiency against (E_b/η) required for $P_e \le 10^{-6}$, is shown in Fig. 12.23. The performance curve for the ideal system is also shown alongside for comparison.

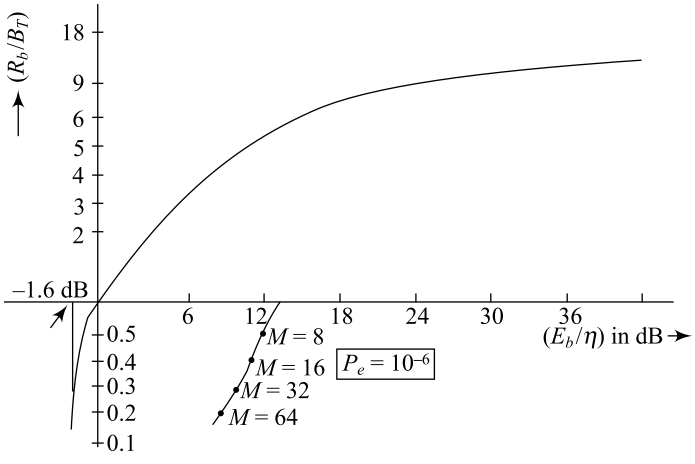


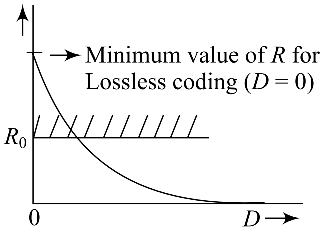
Fig. 12.23 *(E_b/η)Comparison of M-ary FSK with the ideal system*

As can be seen from the figure in the case of M-ary FSK, as M is increased, bandwidth efficiency decreases and the (E_b/η) value required for error-free transmission comes closer and closer to the Shannon limit.

12.10 RATE DISTORTION THEORY

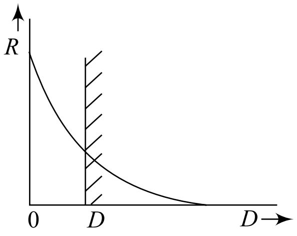
From Shannon's source coding theorem (see Section 12.4.3) we know that for a discrete memoryless source with entropy $H(s)$, the lower bound for the average length $\bar{n}$ of a perfectly decipherable code is $H(s)$ bits per source symbol. Hence, if we use a code with $\bar{n} = H(s)$, error-free restoration of the source symbols is possible and so there is no loss of information involved in such a compression of the source data. Such a compression is therefore called ***lossless compression***. However, since $\bar{n}$ bits are used on the average for representing each source symbol, the average code rate R in bits/codeword is equal to $\bar{n}$. In certain situations such as magnetic disk storage, usage of a code with $\bar{n} = H(s)$ itself may not be possible, as the disk-space requirement may become too large. In such cases, we may be forced to use an $\bar{n} = R$ that is even less than the source entropy $H(s)$. But then, error-free restoration of source symbols will not be possible. That is, at such low values of R, there will be loss of information and so such a compression is referred to as lossy compression and it always results in distortion of the source signal. A very familiar case of lossy coding arises in the encoding of continuous-time signals. Samples of these signals have a continuum of values. So, theoretically an infinite codeword length is needed for their exact or error-free representation. That is, an infinite value of code-rate R is needed. We overcome this difficulty by resorting to quantization of these samples; and quantization results in distortion.

Thus, in the two cases discussed above, we are forced to permit some amount of distortion so as to have some desired or specified value of $\bar{n}$ or R. The question that arises then is: For a given R, what is the minimum distortion D that we can possibly achieve? We may pose the question the other way about and ask: for a specified maximum distortion that can be tolerated, what is the minimum value of R that can be used? These two types of lossy compression problems are illustrated in Figs. 12.24(a) and (b).



For some specified code rate R_0 how can we minimize the distortion D?

Fig. 12.24(a)



For some specified acceptable level of distortion D^*, how can we minimize the code rate R

Fig. 12.24(b)

Rate distortion theory is a branch of information theory that deals with problems such as those posed above. In fact, it originated from Shannon's proposal that the information rate of a continuous source be measured by employing some specific measure for distortion.

Before we proceed any further, it is necessary to discuss 'distortion' and 'distortion measures'.

12.10.1 Distortion

Distortion is related to the fidelity of reproduction of a signal and it indicates how different or how far apart the reproduced signal is with respect to the original signal at the output of the source. This statement is equally applicable to continuous-time and discrete-time signal. For a continuous-time signal, $x(t)$, produced by a source, if $y(t)$ is the reproduced signal, we may have any of the following distortion measures:

1. Absolute value of the maximum difference between $y(t)$ and $x(t)$ for all t, i.e.,

$$d(x(t), y(t)) \underline{\Delta} \max_t [|x(t) - y(t)|]$$

2. Average value with respect to time of the difference between $x(t)$ and $y(t)$, i.e.,

$$d(x(t), y(t)) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [|x(t) - y(t)|]dt$$

3. The mean-squared value of the difference between $x(t)$ and $y(t)$, i.e.,

$$d(x(t), y(t)) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [x(t) - y(t)]^2 dt$$

However, from a practical point of view, any distortion measure should be mathematically tractable and also must closely approximate the human perception of what constitutes distortion and what does not, with reference to any particular physical signal. For instance, phase distortion in an audio signal need not be considered as a distortion since our ear is insensitive to phase. However, when dealing with video signals, the distortion measure has to certainly take phase distortions into account since our eyes are quite sensitive to phase changes.

In order to discuss in a little detail about distortion and distortion measures with reference to the coding of a discrete memoryless source (DMS) consider now Fig.12.4. In that figure, for our present discussion, let the DMS give a set X of statistically independent source symbols x_i, $i = 0, 1, \ldots, (M-1)$ with probabilities of occurrence p_i, $i = 0, 1, \ldots, (M-1)$. Assume that the output of this DMS is encoded and that R is the code rate, i.e., the average number $\bar{n}$ of bits per codeword. From Shannon's source coding theorem, we know that there will be perfect representation of the source symbols by the codewords and hence no distortion, if $R \geq H(s)$, where $H(s)$ is the entropy of our DMS. Let a decoder decode the codewords and let Y be the alphabet at the output of the decoder and let this consist of symbols y_js, $j = 0, 1, \ldots, (L-1)$ as shown in Fig. 12.4. We may now define what is called the 'per-symbol' distortion measure as follows.

1. Hamming measure: $d(x_i, y_j) \triangleq \begin{cases} 1 & \text{if } y_j \neq x_i \\ 0 & \text{otherwise} \end{cases}$ (12.107)

2. Squared-error distortion measure: $d(x_i, y_j) \triangleq (x_i - y_j)^2$ (12.108)

The statistical average of $d(x_i, y_j)$ taken over all values of i and j, is called the average distortion and is denoted by D. It is given by

$$D = E[d(x_i, y_j)] = \sum_{t=0}^{M-1} \sum_{j=0}^{L-1} [p(x_i)p(y_j|x_i)]d(x_i, y_j) \tag{12.109}$$

D is non-negative and is a continuous function of the transition probabilities $p(y_j|x_i)$, which of course, are dependent on the encoder and decoder.

Now recall our original problem: For a given maximum distortion that is permissible, what is the minimum value of the code-rate R that can be used? From Eq. (12.108) it is clear that if D is restricted to a value less than or equal to some D^*, the transition probabilities $p(y_j|x_i)$ that can be assigned, will be restricted to some set, say, P_{D^*} which is defined by

$$P_{D^*} = \{p(y_j|x_i); D \leq D^*\} \tag{12.110}$$

We will now use this set to define an important function, called the rate distortion function.

12.10.2 Rate Distortion Function $R(D^*)$

We know that in a situation like the one discussed above, the average mutual information between the symbols in X and those in Y, is given by

$$I(X;Y) = H(X) - H(X|Y)$$

$$= \sum_{i=0}^{M-1} \sum_{j=0}^{L-1} p(x_i, y_j) \log_2 \left[\frac{p(y_j|x_i)}{p(y_j)} \right] \tag{12.111}$$

We now define what is called the ***rate distortion function***, $R(D^*)$, which is the smallest possible coding rate R in bits for which the ***average*** distortion D is guaranteed not to exceed a specified maximum value D^*. This rate distortion function is given by

$$R(D^*) = \min_{D \le D^*} [H(X) - H(X|Y)\} \tag{12.112}$$

$$= \min_{p(y_j|x_i) \in P_{D^*}} [I(X;Y)$$

Since a given source symbol x_i will have to appear at the output of the decoder as one of the possible y_js, $j = 0, 1, \ldots, (L\text{-}1)$, in the above equation, although the transition probabilities have to be only from the set P_{D^*}, they have to satisfy the condition that

$$\sum_{j=0}^{L-1} p(y_j|x_i) = 1 \text{ for } i = 0, 1, \ldots, (M-1) \tag{12.113}$$

Obviously, as the maximum permissible distortion, D^*, is increased, it will be possible to use a smaller rate R. This means that $R(D^*)$ is a decreasing function with respect to D^*.

Recall that the basic question of the rate distortion theory is: For a specified maximum distortion D^*, what is the minimum value of R that can be used? We now find that the answer to this question is $R(D^*)$, as given by Eq. (12.108). Determining $R(D^*)$ for a given set of $\{x_i\}$ with specified probabilities, p_i, $i = 0, 1, \ldots, (M-1)$, and a specified per symbol distortion measure $d(x_i, y_j)$ boils down to finding the set P_{D^*} of 'D^* - admissible' transitional probabilities $p(y_j|x_i)$, , $j = 0, 1, \ldots, (L-1)$ and $i = 0, 1, \ldots, (M-1)$, subject to the constraint stated in Eq. (12.112). Thu, determination of $R(D^*)$ involves solving of a variational problem and this, in general, is not an easy matter.

Examples of rate distortion functions

1. Consider a binary memoryless source with $P(1) = p$ and $P(0) = 1 - p$ for which the Hamming distortion measure is used.

In this case, the rate distortion function $R(D^*)$ works out to

$$R(D^*) = \begin{cases} H(p) - H(D^*); & 0 \le D^* \le \min[p, 1-p] \\ 0; & \text{otherwise} \end{cases} \tag{12.114}$$

Note *Here, H(p) is the value of the entropy function H(s) of a binary memoryless source (see Fig. 12.2) at* $P(1) = p$. *Similarly,* $H(D^*)$ *is its value at* D^*.

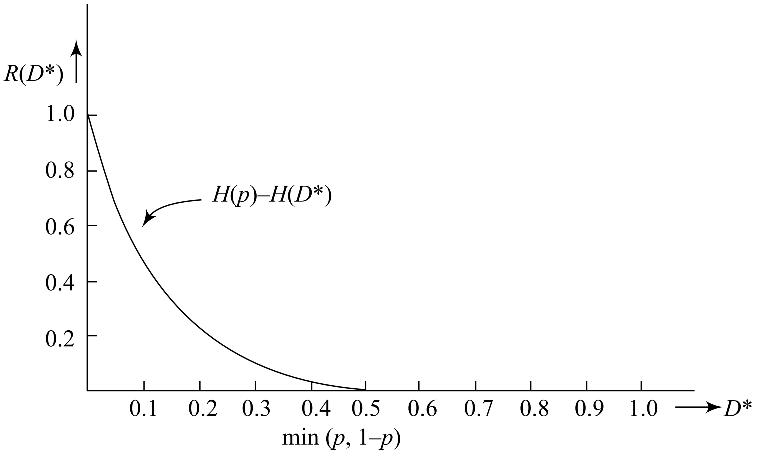


Fig. 12.25 *$R(D^*)$ for a binary memoryless source with Hamming distortion measure*

2. Rate distortion function for a Gaussian source: Consider a discrete memoryless zero-mean Gaussian source with a variance σ^2. A symbol of this will have a continuum of values and so has to be quantized to permit representation by a finite length code. Let the distortion measure be the squared-error measure, i.e.,

$$d(x, y) = (x - y)^2$$

The rate distortion function for this Gaussian source is then given by

$$R(D^*) = \begin{cases} \dfrac{1}{2}\log\left(\dfrac{\sigma^2}{D^*}\right); & 0 \le D^* \le \sigma^2 \\ 0; & D^* > \sigma^2 \end{cases} \tag{12.115}$$

We find that $R(D^*) \to \infty$ as $D^* > \sigma^2$ and $R(D^*) = 0$ for $D^* = \sigma^2$. Hence, a part of $R(D^*)$ vs. D^*/σ^2 is as given in Fig.12.26. Figure 12.27 gives a plot of $R(D^*)$ in bits vs. SNR in dB, where, *SNR* is defined as

$$SNR = 10\,\log_{10}(\sigma^2/D^*)\text{dB}$$

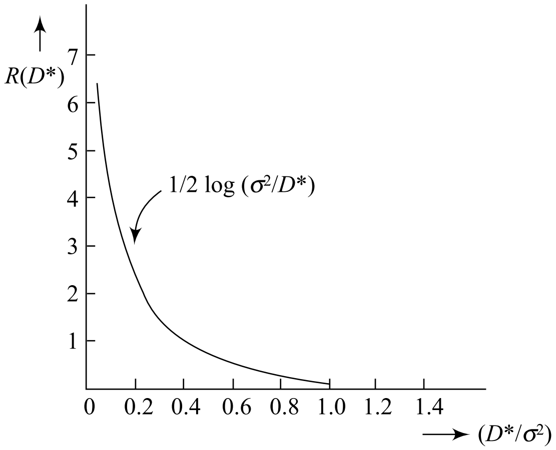


Fig. 12.26 *RDF for a zero-mean Gaussian source with variance σ^2*

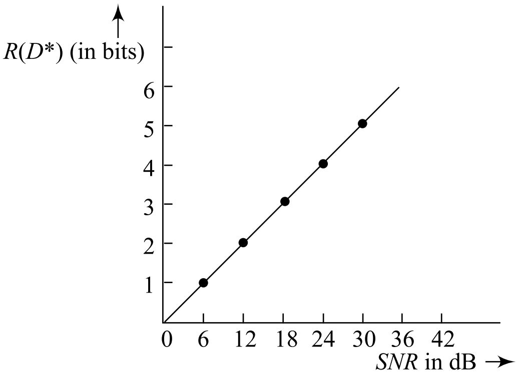


Fig. 12.27 *R(D*) in bits vs. SNR for the zero-mean Gaussian source with variance σ^2*

Summary

- The information obtained from the occurrence of an event a_k with a probability of occurrence of p_k, is $I(a_k) = \log_2(1/p_k)$ bits $\log_e(1/p_k)$ nits.
- 1 nit = $\log_2 e$ bits. Obviously, 'nit' is a bigger unit than a 'bit'.
- A discrete memoryless source (DMS), is one which emits discrete random variables, $\boldsymbol{X_i}$, which take only a discrete set of values and are statistically independent and identically distributed.
- The average information per symbol emitted by a source is called its entropy and is denoted by $H(S)$.
- For a DMS with alphabet S of size M, the entropy is given by

$$H(S) = -\sum_{k=0}^{M-1} p_k \log_2 p_k$$

- where p_k is the probability of the symbol $x(k)$, $k = 0, 1, \ldots, (M - 1)$
- For a DMS with alphabet $x(k)$, $k = 0, 1, \ldots, (M - 1)$ with probabilities p_k, the entropy of the source is maximum when $p_k = 1/M$ for all k, and this maximum value of entropy equals $H(S)_{\max} = \log_2 M$.
- $H(s)$ for a DMS with alphabet size M has the following properties:

- (a) $0 \le H(S) \le \log_2 M$, and (b) $H(S) \le \log_2 M$ and the equality sign holds good only if all the symbols are equi-probable.
- If ***X*** and ***Y*** are two random variables with X taking values x_i, $i = 0, 1, \ldots, (L-1)$ with probabilities $p(y_j)$:
  (a) the joint entropy of ***X*** and ***Y*** is

$$H(\boldsymbol{X}, \boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(x_i, y_j)$$

  (b) The conditional entropy $H(\boldsymbol{Y} \mid \boldsymbol{X} = x_i)$ is

$$H(\boldsymbol{Y}|\boldsymbol{X} = x_i) = -\sum_{j=0}^{L-1} p(y_j|x_i)\log_2 p(y_j|x_i)$$

  (c) The conditional entropy of ***Y*** given ***X*** is

$$H(\boldsymbol{Y}/\boldsymbol{X}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2 p(y_j|x_i)$$

  Further,

$$\begin{aligned} H(\boldsymbol{X}|\boldsymbol{Y}) &= H(\boldsymbol{X}) + H(\boldsymbol{Y}|\boldsymbol{X}) \\ &= H(\boldsymbol{Y}) + H(\boldsymbol{X}|\boldsymbol{Y}) \end{aligned}$$

- The two basic requirements of any source coder are:
  (a) Minimum average length of a codeword for a given set of source alphabet $\{\boldsymbol{X}\}$ and the source symbol probability set $\{p(x_i)\}$.
  (b) Unique decipherability of the encoded sequence.
- Codes having the above two properties are called 'optimal codes'.
- *Prefix-free or instantaneous codes:* Codes in which no codeword is a prefix to another codeword, are called prefix-free codes, or instantaneous codes.
- *McMillan and Karush theorem:* If for a given source, *S*, a code is optimal among the instantaneous codes, then it is optimal among all uniquely decipherable codes.
- *Kraft's inequality:* Kraft's inequality spells out the constraint on the structure of uniquely decipherable codes. It states that a necessary and sufficient condition for the existence of an instantaneous code having word lengths of $n_0, n_1, \ldots, n_{M-1}$ is given by

$$\sum_{i=0}^{M-1} D^{-n_i} \le 1$$

  where *D* is the size of the encoder's alphabet ($D = 2$ for binary codes).
- *Lower and upper bounds for the average length* of a codeword, of a code which is uniquely decipherable:

$$\frac{H(S)}{\log D} \le \bar{n} < \frac{H(S)}{\log D} + 1$$

  where *D* is the size of the encoder's alphabet and $H(s)$ is the entropy of the source.
- *Shannon's source coding theorem:* For a discrete memoryless source with entropy $H(s)$, the minimum value of $\bar{n}$, the average length of a codeword of a perfectly decipherable code is bounded as follows:

$$H(S) \le \bar{n} < H(S) + \in$$

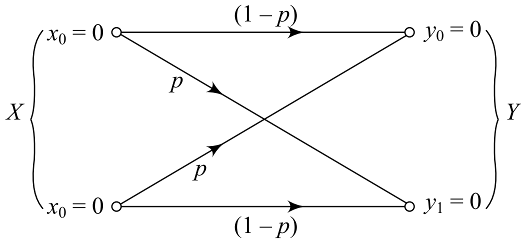
  where, $\in$ can be made arbitrary small by appropriate coding.
- *Coding efficiency:* coding efficiency $\underline{\Delta}\ \dfrac{H(S) = \bar{n}_{\text{min}}}{\text{Actual } \bar{n} \text{ of the code}}$
- Shannon–Fano coding and Huffman coding make use of the source statistics while Lempel–Ziv source coding does not depend on the source statistics and is therefore a 'universal code'. While the two former codes are unequal length codes, the latter one is an equal length code.

- *Discrete memoryless channel (DMC):* A channel with discrete random variables with finite alphabet size as its input and output whose present output symbol depends only on the present input symbol but not on any of the previous input symbol, is called a discrete memoryless channel.
- *Transition matrix/channel matrix:*

Y

	$p(y_j\|x_i)$	y_0	y_1	y_2	·	·	·	y_{L-2}	y_{L-1}
	x_0	$p(y_0\|x_0)$	$p(y_1\|x_0)$	$p(y_2\|x_0)$	·	·	·	$p(y_{L-2}\|x_0)$	$p(y_{L-1}\|x_0)$
	x_1	$p(y_0\|x_1)$	$p(y_1\|x_1)$	$p(y_2\|x_1)$	·	·	·	$p(y_{L-2}\|x_1)$	$p(y_{L-1}\|x_1)$
X	x_2	$p(y_0\|x_2)$	$p(y_1\|x_2)$	$p(y_2\|x_2)$	·	·	·	$p(y_{L-2}\|x_2)$	$p(y_{L-1}\|x_2)$
	·	·	·	·	·	·	·	·	·
	·	·	·	·	·	·	·	·	·
	·	·	·	·	·	·	·	·	·
	x_{M-2}	$p(y_0\|x_{M-2})$	$p(y_1\|x_{M-2})$	$p(y_2\|x_{M-2})$	·	·	·	$p(y_{L-2}\|x_{M-2})$	$p(y_{L-1}\|x_{M-2})$
	x_{M-1}	$p(y_0\|x_{M-1})$	$p(y_1\|x_{M-1})$	$p(y_2\|x_{M-1})$	·	·	·	$p(y_{L-2}\|x_{M-1})$	$p(y_{L-1}\|x_{M-1})$

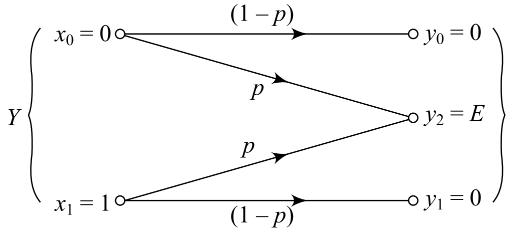
- *Binary symmetric channel (BSC):*



Transition matrix

$$P = \begin{bmatrix} (1-p) & p \\ p & (1-p) \end{bmatrix}$$

- *Binary erasure channel (BEC):*



Transition matrix

$$P = \begin{bmatrix} (1-p) & p & 0 \\ 0 & p & (1-P) \end{bmatrix}$$

- *Mutual information:* Mutual information, $I(\boldsymbol{X}\,;\boldsymbol{Y})$ of a channel represents the average amount of information transferred through the channel, in *bits per symbol.*

$$I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) = H(\boldsymbol{X}) - H(\boldsymbol{X}|\boldsymbol{Y}) \geq 0$$

- $I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{X}|\boldsymbol{Y})$

$$I(\boldsymbol{X};\boldsymbol{Y}) = -\sum_{i=0}^{M-1}\sum_{j=0}^{L-1} p(x_i, y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i, y_j)}\right]$$

- Channel capacity C:

$$C_s = \max_{\{P(x)\}} I(\boldsymbol{X};\boldsymbol{Y}) \text{ in bits/symbol}$$

If R is the transmission rate in symbols/second,

$$\underset{(\text{in bits/sec})}{C} = R \cdot C_s$$

- *Shannon's channel coding theorem:* Given a DMS with an entropy of $H(S)$ bits/symbol emitting symbols at the rate if $(1/T_s)$ symbols/sec, and given a DMC with a capacity of C_s bits/symbol through which the symbols are transmitted at the rate of $(1/T_c)$ symbols per second, it is possible to construct a channel code which would make it possible to transmit the source symbols through the channel and be reconstructed with arbitrarily small probability of error if and only if

$$\frac{H(S)}{T_s} \le \frac{C_s}{T_c}$$

- *Differential entropy:* The differential entropy of a continuous random variable $\boldsymbol{X}$ having a PDF $f_X(x)$, is

$$h(\boldsymbol{X}) = \int_{-\infty}^{\infty} f_X(x) \log_2 \left[\frac{1}{f_X(x)} \right] dx$$

- Properties of differential entropy:
  (a) unlike $H(S)$ of a discrete source, the $h(\boldsymbol{X})$ of a continuous source can be negative.
  (b) $h(\boldsymbol{X})$ is completely determined by the variance σ^2 of $\boldsymbol{X}$.
  (c) For a fixed variance σ^2, the random variable $\boldsymbol{X}$ with Gaussian density has the maximum differential entropy and it is given by $h(\boldsymbol{X}) = \frac{1}{2}\log_2(2\pi e \sigma^2)$.
- *Mutual information (continuous channel):*
  (a) $I(\boldsymbol{X};\boldsymbol{Y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,\, y) \log_2 \left[\frac{f_X(x|y)}{f_X(x)} \right] dxdy$

  (b) $I(\boldsymbol{X};\boldsymbol{Y}) = I(\boldsymbol{Y};\boldsymbol{X}) = h(\boldsymbol{X}) - h(\boldsymbol{X}|\boldsymbol{Y})$
  $= h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X}) \ge 0$
- Shannon—Hartley Law/Shannon's information capacity theorem:

$$\text{Channel Capacity} = C = B \log_2 \left[1 + \frac{S}{N} \right] \text{ bits/second}$$

- *Ideal system:* An ideal system is defined as one that transmits data at a bit rate that is equal to the channel capacity, C in bits per second.
- *Shannon limit for an AWGN channel:*

$$\frac{E_b}{\eta} \approx \log_e 2 = -1.6 \text{ dB}$$

  This represents the theoretical minimum value of E_b/η for an AWGN channel working in the $R_b < C$ region.
- In the case of M-ary PSK, as M is increased, the bandwidth efficiency (R_b / B_T) increases but (E_b / η) value required for error-free transmission moves away from the Shannon limit.
- In the case of M-ary FSK, as M is increased, the bandwidth efficiency decreases and the (E_b / η) value required for error-free transmission comes closer to the Shannon limit.

References and Suggested Reading

1. Abramson, Norman, *Information Theory and Coding*, McGraw-Hill, New York, 1963.
2. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems – An introduction to Signals and Noise in Electrical Communication*, 4th Edition (International Edition), McGraw-Hill, 2002.
3. Gallager, Robert G., *Information Theory and Reliable Communication*, John Wiley & Sons, New York, 1968.
4. Haykin, Simon, *Communication Systems*, 4th Edition, Wiley India, 2001.
5. Haykin, Simon, *Digital Communications*, John Wiley & Sons, Inc, 2002.

6. Lafrance, Pierre, *Fundamental Concepts in Communication*, Eastern Economy Edition, Prentice-Hall, New Delhi, 1992.
7. Lathi, B. P., and Zhi Ding, *Modern Digital and Analog Communication Systems*, International 4th Edition, Oxford University Press, 2010.
8. Lee, Edward A., and David G. Messerschmidt, *Digital Communications*, 2nd Edition, Allied Publishers Limited, 1994.
9. Proakis, John G., and Masoud Salehi, *Fundamentals of Communication Systems*, Low Price Edition, Pearson Education, 2005.
10. Taub, Herbert, Donald L. Schilling, and Gautam Saha, *Taub's Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill, New Delhi, 2008.
11. Wells, Richard B., *Applied Coding and Information Theory for Engineers*, Low Price Edition, Pearson Education, 2004.
12. Wozencraft, John M., and Irwin Mark Jacobs, *Principles of Communication Engineering*, Wiley & Sons, New York, 1965.
13. Ziemer, R.E., and W.H. Tranter, *Principles of Communications – Systems, Modulation and Noise*, 3rd Edition, Jaico Publishing House, 2002.

Review Questions

1. Justify the use of $\log(1/p)$ as the measure of information given by the occurrence of an event whose probability of occurrence is p.
2. Show that 1 nit of information = $\log_2 e$ bits of information.
3. What is the information (in bits) given when a single fair die is thrown?
4. A discrete memoryless source has the symbols A, B, C and D as its alphabet. What is the maximum information that can be associated with each symbol?
5. Define the term 'entropy' of a discrete memoryless source.
6. Explain what is meant by an extended source.
7. How are the joint entropy, marginal entropies and the conditional entropies of two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ related?
8. What is the need for source coding?
9. What are the basic requirements to be met by any source encoder?
10. What is an 'optimal code'?
11. What is meant by the average length of a code?
12. Explain what you mean by an 'instantaneous code'?
13. Sate Kraft's inequality and explain its significance.
14. What are the lower and upper bounds for the average length of a uniquely decipherable code?
15. What is the technique to be adopted for making $\bar{n}$, the average code length of a uniquely decipherable code approach its lower bound $[H(S)/\log D]$?
16. State Shannon's source coding theorem and explain briefly its implications.
17. Define coding efficiency of a source encoder. How can we make this approach 100%?
18. Why are Shannon–Fano codes and Huffman codes called as suboptimal codes?
19. A Shannon–Fano code has given 100% coding efficiency. What can you say about the source?
20. For a given source with a certain set of symbol probabilities, does Huffman coding lead to a unique set of codewords for the various symbols? Whether your answer is in the affirmative or not, justify it.
21. What are the main disadvantages of Huffman coding?
22. What are the advantages of Lempel–Ziv coding?
23. What is a 'transition', or 'channel' matrix? State its properties.
24. What is a binary symmetric channel? Write down its transition matrix in terms of p, the transition probability.

25. What is a binary erasure channel? Write down its transition matrix in terms of p, the probability of erasure of a symbol.
26. What is the physical meaning of the 'mutual information' of a channel? State the properties of $\boldsymbol{I}(\boldsymbol{X};\boldsymbol{Y})$, the mutual information of the discrete random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$.
27. Explain the meaning of the term 'channel capacity' of a channel.
28. When p, the probability of error of a binary symmetric channel is 0.5, the channel capacity is zero. Briefly explain why it is so, in a qualitative manner.
29. Explain, in your own words, the meaning and the importance of Shannon's channel coding theorem.
30. Write down the expression for the differential entropy of a continuous random variable $\boldsymbol{X}$ with a PDF of $f_X(x)$. Why is it called the differential entropy?
31. What are the properties of the differential entropy of a continuous random variable $\boldsymbol{X}$?
32. Write down Shannon–Hartley law and explain its implications.
33. What is 'Shannon Limit' with reference to an AWGN power limited Gaussian channel?
34. Define an ideal system. How is this concept useful?

Problems

1. p denotes the probability of an event A. Plot the amount of information gained by the occurrence of A for $0 \le p \le 1$.
2. A source produces five output symbols A, B, C, D and E with probabilities 0.35, 0.25, 0.20, 0.15 and 0.5 respectively. Assuming successive symbols to be statistically independent, determine (a) the information associated with each one of the symbols, and (b) the entropy of the source.
3. A DMS has symbols a, b and c as its alphabet. If these have probabilities 0.65, 0.20 and 0.15 respectively, calculate
   (a) the entropy of the source.
   (b) the entropy of the second-order extension of the source.
4. A channel has the following transition matrix:

$$P = \begin{bmatrix} 0.34 & 0.16 & 0.5 \\ 0.20 & 0.65 & 0.15 \\ 0.15 & 0.15 & 0.7 \end{bmatrix}$$

   (a) Sketch the channel diagram
   (b) If the input symbols are equally likely, calculate probabilities of the outputs.
   (c) Determine the input-output joint probability matrix.
5. A DMS with alphabet s_0, s_1 and s_2 produces them with probabilities of 0.7, 0.2 and 0.1 respectively. Using Shannon–Fano algorithm and devise an unambiguous code for the source output. Determine the coding efficiency of your code.
6. For the output of the source given in Problem 5, devise an unambiguous code using Huffman algorithm. Find the average codeword length of the code. Now apply Huffman algorithm to the second-order extension of this source and determine the average codeword length.
7. For a BSC, the input binary symbols 0 and 1 occur with probabilities 0.25 and 0.75. Find the probabilities of the binary symbols 0 and 1 appearing at the output.
8. A channel is described by the matrix

$$P = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

   Determine the channel capacity of this channel and find the probabilities of the source symbols that yield this capacity.
9. Two binary symmetric channels, one with a transition probability of 0.1 and the other with a transition probability of 0.2 are connected in cascade. Determine the equivalent channel.
10. A continuous r.v. $\boldsymbol{X}$ is constrained to a peak magnitude of M so that $-M < \boldsymbol{X} < M$. Show that the differential entropy $h(\boldsymbol{X})$ is maximum when $\boldsymbol{X}$ is uniformly distributed between $-M$ and $+M$.

11. A continuous bandpass channel can be modeled as shown in Fig. P12.1. Assuming a signal power of 15.0 W and a one-sided noise PSD of 10^{-4} W/Hz, plot the capacity of the channel as a function of the channel bandwidth.

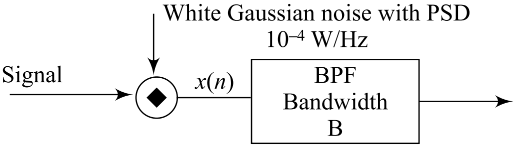


Fig. P12.11

12. A voice-grade channel of the telephone network has a bandwidth of 3.5 kHz

(a) Calculate the information capacity of this telephone channel for a signal-to-noise ratio of 40 dB.

(b) Calculate the minimum signal-to-noise ratio required if information is to be transmitted through the channel at the rate of 9.6 kbps.

Multiple-Choice Questions

1. A fair die is thrown and simultaneously a fair coin is also tossed. The die showed up 3 and the coin showed up 'head'. The information associated with these is

(a) 1.08 bits (b) 0.58 bits (c) 3.585 bits (d) 1.82 bits

2. If 1 nit = k bits, k is equal to

(a) 1.4428 (b) 0.69 (c) 1.359 (d) 0.735

3. The maximum entropy of a DMS with an alphabet size of 128 is

(a) 4.852 bits (b) 7 nits (c) 2.107 bits (d) 7 bits

4. The maximum entropies of two DMSs S_1 and S_2 are in the ratio of 3 : 2. The ratio of the size of their alphabet is

(a) 2 : 3 (b) 2.25 : 1 (c) 1.585 : 1 (d) 4 : 9

5. A discrete memoryless source produces four symbols whose probabilities are in the ratio of 0.25 : 0.5 : 0.75 : 1.0. The entropy of the source is

(a) 1.8466 bits (b) 1.213 bits (c) 0.7853 bits (d) 2.2468 bits

6. The probabilities of occurrence of the output symbols of a binary memoryless source are in the ratio of 1 : 3. For the 3rd order extension of the source, the ratio of the probabilities of occurrence of the most frequent to the least frequent symbols is

(a) 3 : 1 (b) 9 : 1 (c) 27 : 1 (d) 1 : 3

7. The transition matrix for a source producing three symbols A, B and C is as shown. The values of x, y and z are respectively

		j		
	$p(j \mid i)$	A	B	C
i	A	0	Y	1/5
	B	X	1/2	0
	C	1/2	2/5	Z

(a) $\left(\frac{1}{2}, \frac{1}{10}, \frac{4}{5}\right)$ (b) $\left(\frac{1}{2}, \frac{4}{5}, \frac{1}{10}\right)$ (c) $\left(\frac{4}{5}, \frac{1}{2}, \frac{1}{10}\right)$ (d) $\left(\frac{1}{10}, \frac{1}{2}, \frac{4}{5}\right)$

8. For given source alphabet and source statistics, an optimal code for the source is one that

(a) has minimum average code length

(b) has the property of unique decipherability and fixed code length

(c) has both minimum average code length as well as unique decipherability

(d) has 100% coding efficiency

9. A DMS is producing symbols x_i, $i = 0, 1, \ldots, (M - 1)$ with probabilities $p(x_i)$, $i = 0, 1, \ldots, (M - 1)$. A particular binary code is giving code lengths n_i for each x_i in such a way that $p(x_i) = 2^{-n_i}$. Then

(a) the coding efficiency is 100%

(b) the code gives minimum code length but not 100% coding efficiency

(c) the code is not uniquely decipherable
(d) None of the above

10. Huffman coding gives
(a) an equal length code which is unique (b) a variable length code which is unique
(c) an equal length code which is not unique (d) a variable length code which is not unique

11. One of the following is a universal coding scheme. Which one is it?
(a) Fano coding scheme (b) Huffman coding scheme
(c) Lempel–Ziv coding scheme (d) None of the above

12. A binary symmetric channel has a transition matrix as shown. The transition probability and the channel capacity of the channel are respectively

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

(a) (0.5, 1) (b) (0.5, 0) (c) (1, 0) (d) (0, 1)

13. Mutual information $I(\boldsymbol{X} ; \boldsymbol{Y})$ between two discrete random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by
(a) $H(\boldsymbol{X}) + H(\boldsymbol{Y}) - H(\boldsymbol{X}, \boldsymbol{Y})$ (b) $H(\boldsymbol{X}) - H(\boldsymbol{Y} \mid \boldsymbol{X})$
(c) $H(\boldsymbol{Y}) - H(\boldsymbol{X} \mid \boldsymbol{Y})$ (d) $H(\boldsymbol{X}) + H(\boldsymbol{Y}) + H(\boldsymbol{X}, \boldsymbol{Y})$

14. For a continuous channel, the maximum rate of transfer of information, as given by the Shannon–Hartley law, is possible if and only if the input signal is
(a) uniformly distributed (b) Gaussian distributed
(c) Gaussian distributed and has a flat spectrum (d) a flat spectrum

15. With $B >> R_b$ for operation to be in the $R_b < C$ region, the Shannon limit of –1.6 dB represents the theoretical minimum required value of
(a) the average transmitted power
(b) the signal-to-noise ratio
(c) (E_b/η) where E_b is the bit-energy and η is the one-sided PSD of the white noise on the AWGN channel
(d) None of the above

16. For M-ary PCM, as M is increased, ρ, the bandwidth efficiency and the input signal-to-noise ratio (S_R/N_R) required for operation above the threshold, will respectively
(a) increase, increase (b) increase, decrease (c) decrease, increase (d) decrease, decrease

17. For M-ary PSK, as M is increased, bandwidth efficiency ρ and (E_b/η) required for error-free transmission, will respectively
(a) increase, increase (b) increase, decrease (c) decrease, increase (d) decrease, decrease

18. For M-ary FSK, as M is increased, bandwidth efficiency ρ and (E_b/η) required for error-free transmission, will respectively
(a) increase, increase (b) increase, decrease (c) decrease, increase (d) decrease, decrease

Answers for Multiple-Choice Questions

1. (c)	2. (a)	3. (d)	4. (b)	5. (a)	6. (c)	7. (b)	8. (c)
9. (a)	10. (d)	11. (c)	12. (b)	13. (a)	14. (c)	15. (c)	16. (a)
17. (a)	18. (d)						

ERROR-CONTROL CODING

13

> *"Until you value yourself, you will not value your time. Until you value your time, you will not do anything with it."*
>
> **M. Scott Peck (1936 – 2005)**
> *American psychiatrist and author*

Learning Objectives

After going through this chapter, students will be able to

- *understand the different ARQ systems and analyze their performance,*
- *determine the output data rate and the error-correcting capability of a given linear block code,*
- *describe the structure of a systematic block code and determine all the codewords and the d_{min} of the code (including cyclic codes),*
- *use a knowledge of the block code employed, to decode a received word for all simple error patterns,*
- *describe the operation of a given convolutional encoder, draw its state diagram and trellis diagram, and*
- *apply Viterbi algorithm for maximum likelihood decoding of a given received word by making use of the trellis diagram of the convolutional encoder used.*

13.1 INTRODUCTION

A high transmission rate and good reliability, i.e., low probability of error are the two important desirable features that one looks for in any digital communication system. The two parameters available to us for achieving the above twin objectives are the average signal power and bandwidth. For a given modulation scheme, as we had seen in Chapter 11, it is (E_b/η), the ratio of the bit-energy to the noise spectral density that determines the bit-error probability, P_e, that can be achieved. Practical difficulties, however, may not permit one to increase E_b sufficiently to achieve the desired bit-error probability. In such situations, channel coding will be helpful. Channel coding, as stated in Chapter 1, is intended to introduce *controlled redundancy* in order to provide some amount of error-detecting and correcting capability to the data being transmitted. This controlled redundancy helps in detecting erroneously decoded bits and makes it possible to correct the errors before passing on the data to the source decoder. From this, it should not be construed that channel coding will undo the data compression achieved by the source coder. It is always better to remove the redundancy in the source output before introducing some *controlled redundancy* for achieving error correction capability. In some cases, channel coding may be used even for conserving transmitted power, for a given probability of error. Channel coding employing error-correcting codes can, in fact, be used as an alternative technique to approach Shannon limit. Channel coding may be used either for error-detection or error-correction, depending on the amount of redundancy introduced. Error-detecting codes are used in systems using Automatic Repeat

Request (ARR). Use of channel coding, however, increases the transmission bandwidth as the data rate is increased due to the redundancy introduced. It also increases the system complexity in the form of a channel encoder at the transmitter and a channel decoder at the receiver.

13.1.1 Types of Errors

Depending upon the nature of the noise, the bit stream passing through the channel is affected differently. The type of errors caused may be categorized into two types:

1. **Random errors:** Noise affects the transmitted symbols independently, for example, deep space and satellite communication channels.
2. **Burst errors:** Channel noise affects several consecutive bits and errors tend to occur in clusters, for example, HF links in which multipath produces severe fading lasting over several bits.

13.2 ERROR-CONTROL STRATEGIES

Different error-control strategies can be employed in the transmission of digital data. These are:

13.2.1 Forward Error-Correction

It consists of a channel encoder at the transmitter and a channel decoder at the receiver, as shown in Fig. 13.1, and depends upon error-correcting codes.

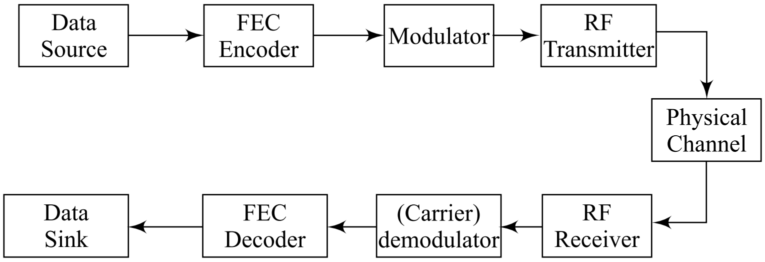


Fig. 13.1 *Block diagram of a system employing FEC*

The FEC encoder and modulator are shown as separate units in the transmitter and correspondingly the detector and FEC decoder are also shown as two separate units in the receiver. However, in certain cases, where bandwidth efficiency is of major concern, the functions of the FEC encoder and modulator at the transmitter and those of the FEC decoder and the demodulator at the receiver are combined.

The advantages and disadvantages in using FEC are as follows.

1. No return path, or feedback channel is needed as in the case of ARQ systems.
2. The ratio of the number of information, or message bits to the total number of bits transmitted, defined as the information throughput efficiency, is constant in FEC systems.
3. A constant overall delay is obtained.
4. On the other hand, the FEC systems need expensive input and output buffers for the encoders and decoders and sometimes buffer overflows cause problems.
5. Only a relatively moderate information throughput is obtained.
6. When very high reliability is needed, selection of an appropriate error-correcting code and implementing its decoding algorithm may be difficult.
7. Reliability of the received data is sensitive to channel degradations.

13.2.2 Automatic Repeat Request (ARQ)

It requires a return path or feedback path from the receiver to the transmitter. It makes use of error-detection at the receiver. Broadly, there are two types of ARQ. These are:

1. Stop-and-wait ARQ
2. Continuous ARQ

In the case of ARQ systems, the throughput efficiency is defined as the ratio of the average number of message bits accepted at the receiver per unit time to the number of message bits that would be accepted per unit time if the ARQ was not used.

1. Stop-and-wait ARQ: In the stop-and-wait ARQ, the transmitter transmits a codeword and then waits. On receiving the transmitted codeword, the receiver checks up whether there are any errors in it. If no errors are detected, the receiver sends an 'acknowledgement' (ACK) signal through the return or feedback path. On receipt of an acknowledgement (ACK) signal, the transmitter transmits the next codeword. In case, one or more errors are detected in the received codeword, the receiver sends a negative acknowledgement (NAK) to the transmitter, which, on receipt of the NAK, retransmits the same codeword that was sent earlier.

Disadvantage: A serious drawback of the stop-and-wait system is that the time interval between two successive transmissions is slightly greater than the round trip delay. So, in satellite channels, in which the round trip delay is quite large, use of stop-and-wait ARQ will very much degrade the transmission efficiency.

Advantage: These ARQ systems are very simple and so they are used on terrestrial microwave links as the round-trip delay is very small in these links.

2. Continuous ARQ: Continuous ARQ systems are of two types:

- (a) Go back-N ARQ systems
- (b) Selective repeat ARQ systems

(a) In a Go back-N ARQ system, the transmitter sends the message continuously without waiting for an ACK signal from the receiver. However, if the receiver detects an error in say the k^{th} message, a NAK signal is sent to the transmitter indicating that the k^{th} message is in error. The transmitter, on receiving the NAK signal, goes back to the k^{th} codeword and starts transmitting all the codewords from the k^{th} onwards.

The Go back-N ARQ is quite useful in satellite links in which the round trip delay is quite large. But buffering is generally its greatest drawback.

(b) In a selective repeat ARQ system, the transmitter goes on sending the messages one after the other without waiting for an ACK. In case the receiver detects an error in the k^{th} codeword, it informs the transmitter indicating that the k^{th} word is in error. The transmitter then immediately sends the k^{th} word and then resumes transmission of the messages in a sequential order starting from where it broke the sequence in order to send the k^{th} word.

From the throughput efficiency point of view, the selective ARQ is the best among all the ARQ systems; but its implementation is expensive.

13.2.3 Combination of FEC and ARQ

If FEC alone is used, the codes may become too long for achieving the desired level of reliability and the system may become too complex and expensive.

If ARQ alone is used, the throughput efficiency is reduced because of the retransmissions caused by error detection.

So, a hybrid system employing both FEC and ARQ, in which the FEC system is contained within an ARQ system, may be used. Such an arrangement is shown in Fig. 13.2.

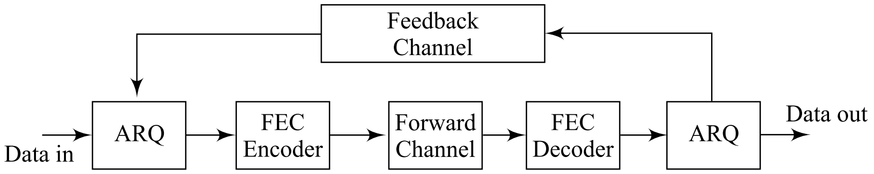


Fig. 13.2 *A hybrid FEC-ARQ system*

As we have seen, the various types of ARQ systems have one good feature in common, i.e., that they need only error-detection and not error-correction. This makes the decoders relatively simple and inexpensive. However, FEC does have the advantage that it does not need a feedback path and this makes it useful in many applications despite the fact that its decoder is quite complex since it has to not only detect errors, but also correct them. But with the advent of VLSI technology and availability of microprocessors, it has become possible to implement these decoders relatively easily.

For the rest of this chapter, we shall focus our attention on the study of different types of error-detecting and correcting codes, their properties, decoding algorithms, and the encoding and decoding techniques. In this context, the following comment is quite pertinent.

Shannon's channel coding theorem merely assures that codes do exist that enable one to transmit information over a noisy channel with a probability of error that can be made arbitrarily small, provided the rate of transmission of information over the channel is less than the rate corresponding to its 'channel capacity'. But, unfortunately, it does not provide any clue, whatsoever on the way such codes can be designed.

A code consists of a set of codewords. Each codeword is a finite length sequence of code elements. If these code elements are drawn from the binary number field, which consists of only two digits – 0 and 1, the code is said to be a binary code. In this book, we will be discussing only binary codes. It is therefore necessary to be familiar with modulo-2 arithmetic since all arithmetic operations like addition, subtraction, multiplication and division in a binary number field will have to be as per this arithmetic.

13.3 MODULO-2 ARITHMETIC

The binary number field, unlike say, the real number field, has only two digits – 0 and 1. So the basic arithmetic operations in this field proceed as follows:

1. Addition

$$0+0=0$$
$$0+1=1$$
$$1+0=1$$
$$1+1=0$$

Note

(i) Strictly speaking, we should use the notation ⊕ for addition operation in a binary field. But, we will be using simply +, as it is more convenient to do so.

(ii) Since $1+1=0$, it means that $-1=1$. It then follows that there is no difference between addition and subtraction operations in a binary field.

2. Multiplication:

$$0\times 0=0$$
$$0\times 1=0$$
$$1\times 0=0$$
$$1\times 1=1$$

3. Division: Division by 0 is not permitted

$$0\div 1=0$$
$$1\div 1=1$$

The reader might have noticed that modulo-2 addition is nothing but EXCLUSIVE-OR operation and that multiplication of binary digits follows AND logic.

13.4 ERROR-CORRECTING CODES

Over the years, several error-correcting codes have been found. However, all of them can be classified broadly as (i) Block codes, and (ii) Convolutional codes. Although both these types of codes introduce redundancy in order to provide error-correcting capability to the code, they differ in the way the redundancy is introduced and the presence or absence of memory in the encoder.

The encoder of a block code takes successive segments of length k binary digits from the message bit stream. Using each segment of *k message bits*, it produces at its output a *codeword* of n binary digits where, n is greater than k and is called the *block length*. The $(n - k)$ additional bits introduced in the encoding process are called *parity check bits* and are generated from the k message bits by taking $(n - k)$ different predetermined linear combinations of them. These $(n - k)$ different linear combinations define the mathematical structure of the code. The parity check bits, being related to the message bits, add redundancy to the message bit-stream and help in error correction. The ratio k/n is referred to as the code rate, denoted by r and is such that $0 < r < 1$.

The encoder of a convolutional code, on the other hand, operates on the message bit stream on a continuous basis by performing modulo-2 discrete convolution on the message sequence as it passes through the encoder's memory, the duration of which is equal to the duration of its own finite length impulse response.

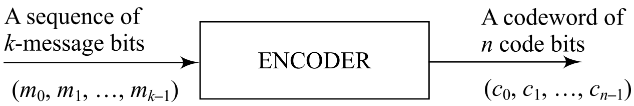


Fig. 13.3 *An encoder for a block code*

In the next few sections, we will be discussing in more detail, about these two types of codes, taking up the block codes first.

13.5 BLOCK CODES

Of all the block codes, we shall discuss only about what are called the 'Linear block codes'. The linearity property of any code may be explained in a simple way by saying that ***a code is said to be linear if the sum (modulo-2 sum of corresponding bits) of any two of its code vectors result again in a code vector***, i.e., the codewords of the code should obey the closure property with respect to modulo-2 addition of corresponding bits.

Recall that we had earlier stated (see Fig.13.3) that the encoder of a block code maps each k-bit segment of the message bit stream into a codeword of n bits. If this mapping is such that the n-bit codeword consists of k ***unaltered*** message bits $m_0, m_1, \ldots, m_{k-1}$ and the remaining $(n - k)$ are parity check bits $b_0, b_1, \ldots, b_{n-k-1}$ obtained by taking $(n - k)$ different linear combinations of the k message bits, we call the resulting code as a '***systematic code***', or simply that the code is in a '***systematic form***'. Block codes in the systematic form are preferred because their implementation becomes much simpler. Hence, the codeword structure for a systematic code is as follows:

$$\text{Codeword C} = (n - k) \text{ check bits} \vdots k \text{ message bits} \tag{13.1}$$

which means that if

$$c = (c_0, c_1, c_2, \ldots, c_{n-1}) \tag{13.2}$$

Then, for a systematic code,

$$c_i = \begin{cases} b_i; & i = 0, 1, \ldots, (n - k - 1) \\ m_{i-(n-k)}; & i = (n - k), (n - k + 1), \ldots, (n - 1) \end{cases} \tag{13.3}$$

A segment of k message digits (binary) is taken each time by the encoder to produce an n-length codeword. Since there can be 2^k distinct k-length binary sequences, it follows that there should correspondingly be 2^k distinct codewords. The full set of 2^k distinct codewords corresponding to the 2^k possible message sequences, is said to be constituting the (n, k) block code. In fact, these 2^k code vectors, each of length n, form a subspace of the n-dimensional vector space formed by the set of all possible 2^n vectors, each of length n having entries from the binary number field. Since there can be 2^n distinct n-length binary sequences, it means that out of these 2^n possible n-length sequences, only 2^k of them are legitimate codewords. The remaining $(2^n - 2^k)$ n-length sequences are not codewords. Since the transmitter transmits the output of the encoder (perhaps after carrier modulation), only legitimate codewords are transmitted. But channel noise might affect one or more of the bits in the transmitted codeword. So unless the channel noise has transformed the transmitted codeword into another legitimate codeword, the received n-length sequence will be one of those $(2^n - 2^k)$ sequences of length n which are not codewords. The receiver then knows that the received word is erroneous and the problem is only to find out which particular codeword was transmitted.

13.5.1 Generator Matrix

At this stage, it is necessary for us to adopt compact notation for representation of the message sequence, codewords, etc. so that the equations that we will be writing henceforth will not be too unwieldy. Hence, we shall use the following row-vector notation for these sequences:

$$\underset{(\text{length}\,k)}{\text{Message sequence}}: \boldsymbol{m} = (m_0, m_1, \ldots, m_{k-1});\ \text{a } (1 \times k) \text{ row vector}$$

$$\underset{(\text{length}\,n\text{–}k)}{\text{Parity bit sequence}}: \boldsymbol{b} = (b_0, b_1, \ldots, b_{n-k-1});\ \text{a } (1 \times n - k) \text{ row vector}$$

$$\underset{(\text{length}\,n)}{\text{Code word}}: \boldsymbol{C} = (c_0, c_1, \ldots, c_{n-1});\ \text{a } (1 \times n) \text{ row vector} \tag{13.4}$$

So, using Eq. (13.1), we may represent $\boldsymbol{C}$ as

$$\boldsymbol{C} = [\boldsymbol{b} \vdots \boldsymbol{m}] \tag{13.5}$$

Since each parity check bit b_i is a distinct linear combination of the message bits, let us write

$$b_i = i^{\text{th}} \text{ check bit} = p_{0,i} \cdot m_0 + p_{1,i} \cdot m_1 + \ldots + p_{k-1,i} \cdot m_{k-1}$$
$$i = 0, 1, 2, \ldots, (n - k - 1) \tag{13.6}$$

Hence, Eq. (13.6) represents a set of k linear equations all of which together may be conveniently represented as a single matrix equation as follows.

$$(m_0, m_1, \ldots, m_{k-1}) \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,n-k-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ p_{k-1,0} & p_{k-1,1} & \cdots & p_{k-1,n-k-1} \end{bmatrix}_{k \times (n-k)} = [b_0, b_1, \ldots, b_{n-k-1}] \tag{13.7}$$

or more compactly as

$$\boldsymbol{m}[P]_{k \times (n-k)} = \boldsymbol{b} \tag{13.8}$$

where

$$[P]_{k \times (n-k)} = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,n-k-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ p_{k-1,0} & p_{k-1,1} & \cdots & p_{k-1,n-k-1} \end{bmatrix} \tag{13.9}$$

Since as per Eq. (13.5),

$$\boldsymbol{C} = [\boldsymbol{b} \vdots \boldsymbol{m}]$$

∴ we may write

$$\boldsymbol{C} = \boldsymbol{m}[P \vdots I_k] = \boldsymbol{m}[\boldsymbol{G}] \tag{13.10a}$$

where

$$\boldsymbol{G} = [P \vdots I_k] \triangleq \text{ generator matrix} \tag{13.10b}$$

$\boldsymbol{C}$ is a $1 \times n$ row vector, $\boldsymbol{m}$ is a $1 \times k$ row vector, $[P]$ is the $k \times (n–k)$ matrix, i.e., the parity check matrix, and $[I_k]$ is a $k \times k$ identity matrix as given below

$$[I_k]_{k\times k} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}_{k\times k} \tag{13.11}$$

The $(k \times n)$ matrix $[P \vdots I_k]$ of Eq. (13.10) is called the generator matrix $[\boldsymbol{G}]$ of the code because when it is premultiplied by the $(1 \times k)$ message bit row vector, it gives the $(1 \times n)$ codeword corresponding to that message vector. So, by using each one of the 2^k possible $(1 \times k)$ message vectors, we can generate all the 2^k valid code vectors of the (n, k) code by using this generator matrix. Thus,

$$\boldsymbol{C} = \boldsymbol{m}[G]_{k\times n} = \boldsymbol{m}[P \vdots I_k]_{k\times n} \tag{13.12}$$

Note *$[G] \triangleq [P \vdots I_k]$ here because we have assumed that in the systematic (n, k) block code, each code vector is of the form $\boldsymbol{C} = [\boldsymbol{b} \vdots m]$ (see Eq. (13.5)). On the other hand, if we assume that $\boldsymbol{C} = [\mathbf{m} \vdots \mathbf{b}]$, then [**G**] of the systematic code will be of the form $G = [I_k \vdots P]$.*

The k rows of the generator matrix $\boldsymbol{G}$ are linearly independent in the sense that the linear combination of no two of its rows will result in any of the other rows. (In fact, these k linearly independent rows of $\boldsymbol{G}$ matrix form a basis set for the k-dimensional subspace formed by the 2^k code vectors. This in turn means that the rows of the $\boldsymbol{G}$ matrix are also codewords.) The matrix $\boldsymbol{G}$ is therefore said to be in the canonical form.

Example 13.1 A binary linear block code has a generator matrix

$$\boldsymbol{G} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Determine all its codewords.

Solution The generator matrix is a $k \times n$ matrix. Hence in this code, $k = 3$ and $n = 7$. Since it is a binary code and $k = 3$, there are 8 possible distinct message sequences of length 3 and these are:

$$\boldsymbol{m}_1 = (0,0,0) \quad \therefore \boldsymbol{C}_1 = \boldsymbol{m}_1[G] = [0,0,0][G] = [0,0,\ldots,0]_{1\times 7}$$
$$\boldsymbol{m}_2 = (0,0,1) \quad \therefore \boldsymbol{C}_2 = \boldsymbol{m}_2[G] = [0,0,1][G] = [0,1,1,1,0,0,1]$$
$$\boldsymbol{m}_3 = (0,1,0) \quad \therefore \boldsymbol{C}_3 = \boldsymbol{m}_3[G] = [0,1,0][G] = [1,1,1,0,0,1,0]$$
$$\boldsymbol{m}_4 = (0,1,1) \quad \therefore \boldsymbol{C}_4 = \boldsymbol{m}_4[G] = [0,1,1][G] = [1,0,0,1,0,1,1]$$
$$\boldsymbol{m}_5 = (1,0,0) \quad \therefore \boldsymbol{C}_5 = \boldsymbol{m}_5[G] = [1,0,0][G] = [1,1,0,1,1,0,0]$$
$$\boldsymbol{m}_6 = (1,0,1) \quad \therefore \boldsymbol{C}_6 = \boldsymbol{m}_6[G] = [1,0,1][G] = [1,0,1,0,1,1,1]$$
$$\boldsymbol{m}_7 = (1,1,0) \quad \therefore \boldsymbol{C}_7 = \boldsymbol{m}_7[G] = [1,1,0][G] = [0,0,1,1,1,1,0]$$
$$\boldsymbol{m}_8 = (1,1,1) \quad \therefore \boldsymbol{C}_8 = \boldsymbol{m}_8[G] = [1,1,1][G] = [0,1,0,0,1,1,1]$$

Note *As stated earlier, the k rows of the* **G** *matrix are indeed k of the 2^k code vectors. It may be checked that* C_8 *is row one of* **G**, C_3 *is row 2 of* **G**, *and* C_2 *is the* 3^{rd} *row of* **G**.

Example 13.2 An (n, k) block code has a generator matrix $\boldsymbol{G}$. Using $\boldsymbol{G}$, show that the sum of any two codewords results in another codeword.

Solution Let $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ be two arbitrary distinct message sequences. Then the corresponding codewords $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are given by

$$\boldsymbol{C}_1 = \boldsymbol{m}_1[G]$$

and

$$\boldsymbol{C}_2 = \boldsymbol{m}_2[G]$$

Hence,

$$\boldsymbol{C}_1 + \boldsymbol{C}_2 = (\boldsymbol{m}_1 + \boldsymbol{m}_2)[G] \tag{i}$$

Now, $\boldsymbol{m}_1$ is a binary message sequence of length k.
$\boldsymbol{m}_2$ is another binary message sequence also of length k.

$\therefore$ when $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are added using modulo-2 arithmetic, we get another binary sequence which is also of length k and which therefore must be one of the 2^k possible message sequences of length k.

Let

$$\boldsymbol{m}_1 + \boldsymbol{m}_2 = \boldsymbol{m}_i \text{ where } 1 \le i \le 2^k$$

Then from Eq. (i),

$$\boldsymbol{C}_1 + \boldsymbol{C}_2 = (\boldsymbol{m}_1 + \boldsymbol{m}_2)[G] = \boldsymbol{m}_i[G] = \boldsymbol{C}_i$$

where $\boldsymbol{C}_i$ is the codeword corresponding to the k-length message sequence $\boldsymbol{m}_i \in$ (the set of 2^k possible message sequences).

Hence, the sum of any two codewords gives another valid codeword.

13.5.2 Parity Check Matrix

We have already seen that the $k \times n$ generator matrix [**G**] completely characterizes a linear (n, k) block code in the sense that knowledge of [**G**] enables us to determine all the 2^k codewords of the code. This fact is clear from Eq. (13.12) and from Example 13.1.

Apart from [**G**] matrix, there is another matrix, [**H**] which also completely characterizes the code. This **H**-matrix is called the parity check matrix. As can be seen from Eq. (13.6), each one of the parity check bits is a linear combination of the message digits. Thus, in an (n, k) block code, the $(n–k)$ parity check digits can be determined for any arbitrary set of k message digits, provided we have $(n–k)$ parity-check equations, as clearly exemplified by Eq. (13.7). Thus, the parity-check equations give another way of characterizing a block code.

Let us consider a $(n–k) \times n$ matrix H defined as:

$$[H]_{(n-k)\times n} \triangleq \left[I_{n-k} \vdots P^T \right] \tag{13.13}$$

Then

$$[H][G]^T = \left[I_{n-k} \vdots P^T \right] \begin{bmatrix} P^T \\ \cdots \\ I_k \end{bmatrix} = I_{(n-k)} \cdot P^T + P^T \cdot I_k \tag{13.14}$$

Recalling that P^T is of size $(n–k) \times k$ we find that both the matrix multiplications are quite compatible. Hence, even though P^T is rectangular, it does not matter and each of the above matrix multiplications will yield P^T.

i.e.,

$$\boldsymbol{H} \cdot \boldsymbol{G}^T = I_{(n-k)} \cdot P^T + P^T \cdot I_k = P^T + P^T = [0]_{(n-k)\times k} \tag{13.15}$$

Because we have to add the corresponding entries of the two matrices using modulo-2 arithmetic and since the corresponding entries are either both zero, or both 1, all the entries of the resultant $(n-k) \times k$ matrix are zero, i.e., we get an $(n-k) \times k$ *null matrix.*

Now, taking the transposes of the matrices on the two sides of Eq. (13.15), we get

$$\boldsymbol{G} \cdot \boldsymbol{H}^T = [0]_{k\times(n-k)} \tag{13.16}$$

Equation (13.12) says that

$$\boldsymbol{C} = \boldsymbol{m}[G]$$

Post multiplying by $\boldsymbol{H}^T$ on both sides, we get

$$\boldsymbol{C} \cdot \boldsymbol{H}^T = \boldsymbol{m}[G]\boldsymbol{H}^T = \boldsymbol{m}[0]_{k\times(n-k)} = [0]_{1\times(n-k)}$$

$$\therefore \qquad \boxed{\boldsymbol{C} \cdot \boldsymbol{H}^T = \boldsymbol{0}} \tag{13.17}$$

The matrix $[H]_{(n-k)\times n}$ is called the parity-check matrix. The reason for calling it as parity check matrix may be seen from the following.

Equation (13.17) may also be written as

$$\boldsymbol{H} \cdot \boldsymbol{C}^T = \boldsymbol{0} \tag{13.17a}$$

Since

$$[H] = [I_{n-k} \vdots P^T] = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & p_{0,0} & p_{1,0} & \cdots & p_{k-1,0} \\ 0 & 1 & 0 & \cdots & 0 & p_{0,1} & p_{1,1} & \cdots & p_{k-1,1} \\ 0 & 0 & 1 & \cdots & 0 & p_{0,2} & p_{1,2} & \cdots & p_{k-1,2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & p_{0,n-k-1} & p_{1,n-k-1} & \cdots & p_{k-1,n-k-1} \end{bmatrix}_{(n-k)\times n} \tag{13.18}$$

$$\boldsymbol{C}^T = [c_0 \quad c_1 \ldots c_{n-k-1} \quad c_{n-k} \quad c_{n-k+1} \ldots c_{n-1}]^T \tag{13.19}$$

But from Eqs. (13.4) and (13.5), we may rewrite the above as

$$\boldsymbol{C}^T = [b_0 \quad b_1 \ldots b_{n-k-1} \quad m_0 \quad m_1 \ldots m_{n-1}]^T \tag{13.20}$$

Substituting in Eq. (13.17a) using Eq. (13.20) for $\boldsymbol{C}^T$ and Eq. (13.18) for $[H]$:

$$\boldsymbol{H} \cdot \boldsymbol{C}^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & p_{0,0} & p_{1,0} & \cdots & p_{k-1,0} \\ 0 & 1 & 0 & \cdots & 0 & p_{0,1} & p_{1,1} & \cdots & p_{k-1,1} \\ 0 & 0 & 1 & \cdots & 0 & p_{0,2} & p_{1,2} & \cdots & p_{k-1,2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & p_{0,n-k-1} & p_{1,n-k-1} & \cdots & p_{k-1,n-k-1} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ m_{k-1} \end{bmatrix} = [0] \tag{13.21}$$

This gives $(n-k)$ equations, which are actually parity-check equations. To see this, let us take the first equation:

$$b_0 + 0 + 0 + \ldots + b_{n-k-1}0 + p_{0,0} \cdot m_0 + p_{1,0} \cdot m_1 \ldots p_{k-1,0} \cdot m_{k-1} = 0$$

$$\Rightarrow \qquad b_0 = m_0 \cdot p_{0,0} + m_1 p_{1,0} + \ldots + m_{k-1} \cdot p_{k-1,0} \tag{13.22}$$

which is the parity check equation that gives the parity check bit b_0. Like this, these $(n-k)$ equations give the $(n-k)$ parity check bits $(b_0, b_1, b_2, \ldots, b_{n-k-1})$ of the code vector; the remaining elements of the code vector being the k unaltered message bits $(m_0, m_1, m_2, \ldots, m_{k-1})$ as given in Eq. (13.20).

13.5.3 Dual Code

Equation (13.16) says that $\boldsymbol{G} \cdot \boldsymbol{H}^T = \boldsymbol{0}$

$$\text{Transposing both sides,} \qquad \boldsymbol{G} \cdot \boldsymbol{H}^T = \boldsymbol{0} \tag{13.23}$$

From Eq. (13.23), it is clear that we can visualize a code for which the ***H***-matrix is a generator matrix and the ***G***-matrix is the parity-check matrix. This (n, $n - k$) linear block code is called the *dual* of the original (n, k) block code.

Note *While the **G**-matrix of a linear block code is useful in generating the code vectors (as output of the channel encoder at the transmitter), the H-matrix is useful at the decoder of the receiver. Since Eq. (13.17) is satisfied by **C** if and only if it is a legitimate code vector, the decoder of the receiver uses the received vector **r** in Eq. (13.17) in the place of **C** to check whether **r** satisfies that equation. If it does, then it is a valid code vector. If it does not, then the receiver decides that one or more bits of the received vector are erroneous.*

Example 13.3 Consider a (6, 3) generator matrix

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 1 & 0 & \vdots & 1 & 0 & 1 \\ 0 & 0 & 1 & \vdots & 1 & 1 & 0 \end{bmatrix}$$

Find (a) all the code vectors of this code, (b) the parity-check matrix of this code, and (c) the minimum weight of this code. **(JNTU, Nov., 2009)**

Solution As pointed out in the note under Eq. (13.12), here the structure of the given [***G***] matrix implies that the code vectors are of the form

$$\boldsymbol{C} = [m \vdots b]$$

(a) Since ***G*** is a $k \times n$ matrix for an (n, k) block code, from the given ***G***, we find that $k = 3$ and $n = 6$. Since $k = 3$, there will be $2^k = 8$ codewords corresponding to the $2^k = 8$ possible message sequences. These message sequences are (0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0) and (1, 1, 1). If a message sequence is (m_0, m_1, m_2), then we know that the code vector corresponding to this message sequence is given by

$$\boldsymbol{C} = (c_0, c_1, c_2, c_3, c_4, c_5) = [m_0, m_1, m_2] \begin{bmatrix} 1 & 0 & 0 & \vdots & 0 & 1 & 1 \\ 0 & 1 & 0 & \vdots & 1 & 0 & 1 \\ 0 & 0 & 1 & \vdots & 1 & 1 & 0 \end{bmatrix}$$

$$\therefore \quad \boldsymbol{C} = (c_0, c_1, c_2, c_3, c_4, c_5) = [m_0, m_1, m_2, m_1 + m_2, m_0 + m_2, m_0 + m_1]$$

For the message sequence $\boldsymbol{m}_0 = (0, 0, 0)$ $\therefore$ $\boldsymbol{C}_0 = [0, 0, 0, 0, 0, 0]$

For $\boldsymbol{m}_1$, which is (0, 0, 1), $m_0 = 0$, $m_1 = 0$ and $m_2 = 1$.

$$\therefore \quad \boldsymbol{C}_1 = [0, 0, 1, 1, 1, 0]$$

In a similar way, by finding values of m_0, m_1 and m_2 for each of the 8 message sequences and then substituting those values in the above expression for ***C***, we can get the remaining four code vectors, $\boldsymbol{C}_2$, $\boldsymbol{C}_3$, $\boldsymbol{C}_4$ and $\boldsymbol{C}_5$, $\boldsymbol{C}_6$, and $\boldsymbol{C}_7$

$$\boldsymbol{C}_2 = [0, 1, 0, 1, 0, 1]; \quad \boldsymbol{C}_3 = [0, 1, 1, 0, 1, 1]$$
$$\boldsymbol{C}_4 = [1, 0, 0, 0, 1, 1]; \quad \boldsymbol{C}_5 = [1, 0, 1, 1, 0, 1]$$
$$\boldsymbol{C}_6 = [1, 1, 0, 1, 1, 0]; \quad \boldsymbol{C}_7 = [1, 1, 1, 0, 0, 0]$$

(b) Since $\boldsymbol{G} = [I_k \vdots P]$ for the given matrix, correspondingly the parity-check matrix [***H***] is given by $[P^T \vdots I_{n-k}]$. From the given G matrix, we find that

$$P_{k\times n-k}=\begin{bmatrix}0&1&1\\1&0&1\\1&1&0\end{bmatrix}\qquad\therefore\qquad P^T=\begin{bmatrix}0&1&1\\1&0&1\\1&1&0\end{bmatrix}$$

$$\therefore\qquad [H]=\begin{bmatrix}0&1&1&\vdots&1&0&0\\1&0&1&\vdots&0&1&0\\1&1&0&\vdots&0&0&1\end{bmatrix}_{(n-k)\times n}$$

(c) The minimum weight of a code is equal to the minimum number of 1s in any code vector among all the non-zero code vectors of the code. So, excluding $\boldsymbol{C}_0$, among the other code vectors, we find that $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, $\boldsymbol{C}_4$ and $\boldsymbol{C}_7$ have three 1's. Hence, the minimum weight of the code is 3.

Example 13.4 A parity check encoder appends a single even parity bit b to each block of k message bits $(m_1, m_2, \ldots, m_k)$. If $k = 3$ (a) determine the 2^k possible codewords of this code, (b) determine its G and H matrices, (c) show that $\boldsymbol{C}\cdot\boldsymbol{H}^T = \boldsymbol{0}$, and (iv) show that for a single error, the received vector $\boldsymbol{r}$ is such that $\boldsymbol{r}\cdot\boldsymbol{H}^T = 1$.

Solution

(a) Since $k = 3$, there will be 8 possible message blocks and correspondingly 8 codewords. The 8 message bits, the corresponding check bit for each and the resulting code vectors are all listed in the following table:

Message bits	Check bit	Codeword
$m_0\ m_1\ m_2$	b_0	$c_0\ c_1\ c_2\ c_3$ $m_0\ m_1\ m_2\ b_0$
0 0 0	0	0 0 0 0
0 0 1	1	0 0 1 1
0 1 0	1	0 1 0 1
0 1 1	0	0 1 1 0
1 0 0	1	1 0 0 1
1 0 1	0	1 0 1 0
1 1 0	0	1 1 0 0
1 1 1	1	1 1 1 1

(b) Hence, for this code, $k = 3$ and $n = 4$. We know that the coefficient matrix, P is of size $k\times(n-k)$, i.e., 3×1 – a column vector with three entries. We know that

$$\boldsymbol{m}[P]_{k\times(n-k)}=\boldsymbol{b}$$

i.e.,

$$[m_0, m_1, m_2]\begin{bmatrix}p_0\\p_1\\p_2\end{bmatrix}=b_0\Rightarrow m_0p_0\oplus m_1p_1\oplus m_2p_2=\text{Parity bit}$$

Since even parity is needed, we should have

$$m_0p_0\oplus m_1p_1\oplus m_2p_2=\begin{cases}0 & \text{if } m_0\oplus m_1\oplus m_2=0\\1 & \text{if } m_0\oplus m_1\oplus m_2=1\end{cases}\qquad\text{(i)}$$

Since multiplication by 1 is not going to change the value of m_0, m_1, and m_2, it follows that

$$p_0=p_1=p_2=1$$

will satisfy the condition stipulated in Eq. (i),

$$\therefore \qquad [P]_{k\times(n-k)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad [G]_{k\times n} = [I_k \vdots P] = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

To find the ***H***-matrix, we know that

$$[H]_{(n-k)\times n} = [P^T \vdots I_{n-k}]$$

$$\therefore \qquad [H] = [1 \quad 1 \quad 1 \quad \vdots \quad 1] = [1 \quad 1 \quad 1 \quad 1]$$

(c) To show that $\boldsymbol{C}\cdot\boldsymbol{H}^T = \mathbf{0}$

$$\boldsymbol{CH}^T = [m_0 \quad m_1 \quad m_2 \quad b_0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = m_0 \oplus m_1 \oplus m_2 \oplus b_0$$

Because the check bit b_0 is appended to m_0, m_1, m_2 to get even parity, it means that $m_0 \oplus m_1 \oplus m_2 \oplus b_0 = 0$

$$\therefore \qquad \boldsymbol{CH}^T = \mathbf{0}$$

(d) To show that $\boldsymbol{r} \cdot \boldsymbol{H}^T = 1$ for a single error

A single error will upset the even parity and so

$$\boldsymbol{r} \cdot \boldsymbol{H}^T = m_0 \oplus m_1 \oplus m_2 \oplus b = 1$$

where $m' = m_0 + 1$ since m_0 is assumed to be the digit affected, i.e., has been changed from 0 to 1 or 1 to 0, which is equivalent to adding 1 to m_0.

Example 13.5 For a certain code, the generator matrix ***G*** is

$$\boldsymbol{G} = \begin{bmatrix} 1 & 1 & 0 & \vdots & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \vdots & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & \vdots & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & \vdots & 0 & 0 & 0 & 1 \end{bmatrix}$$

Find all the codewords of its dual code.

Solution For the given code, we find (from ***G***) that $k = 4$ and $n = 7$

$$\text{Also,} \qquad P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad \therefore \quad [H]_{(n-k)\times n} = \left[I_{n-k} \vdots P^T \right]$$

$$\therefore \qquad \boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} = \boldsymbol{G'}$$

where ***G′*** is the generator matrix of the dual code.

This is a (n', k') code where $n' = 7$ and $k' = 3$.

Since $k' = 3$, there are $2^3 = 8$ possible distinct message sequences each of length 3. These are (0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0) and (1, 1, 1). Now, using the relation $\boldsymbol{C} = \boldsymbol{m}[\boldsymbol{G'}]$, where $\boldsymbol{m}$ is a message sequence, we can find out all the 8 code vectors corresponding to the eight message sequences. For instance, if the code vector corresponding to the message sequence (0, 1, 1) is needed, we have to simply add the second and third rows of the $\boldsymbol{G}'$. Thus,

Message sequence	code vector
(0, 1, 1)	0, 1, 1, 1, 0, 0, 1

The other code vectors can be found out in a similar way.

Example 13.6 Given the $\boldsymbol{H}$ matrix as

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Determine all the codewords beginning with 1 1 1 …

Solution From the given $\boldsymbol{H}$ matrix, $n–k = 3$ and $n = 6$ $\therefore$ $k = 3$.

Also, since $[H] = [P^T \vdots I_{n-k}]$, we have $P^T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

$$\therefore \quad \boldsymbol{G} = [I_k \vdots P] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Since $\boldsymbol{C} = [m_0 \quad m_1 \quad m_2 \ \vdots\ b_0 \quad b_1 \quad b_2]$, for the codeword beginning with 111, the message sequence = (1, 1, 1) = $\boldsymbol{m}$

Also
$$\boldsymbol{C} = \boldsymbol{mG} = [1 \quad 1 \quad 1]\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

= Sum of all the three rows of the $\boldsymbol{G}$ matrix

$$\therefore \quad \boldsymbol{C} = [1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]$$

13.5.4 Repetition Code

The repetition codes are the simplest linear block codes used for error correction. The coding simply consists of transmitting each message digit n times. Of these, (n–1) digits are parity check digits used for error-correction and the remaining one is the message digit. Since the message digit can be either a 1 or a 0 (in binary transmission), it follows that in repetition codes, the codewords will be either all zero, or all one, type. Since $k = 1$, the repetition code which transmits the same message bit n times, constitutes an (n, 1) block code. The code rate for an (n, 1) repetition code is

$$r = \frac{1}{n} \text{ bits/symbol}$$

Decoding of a received sequence simply consists of deciding on the basis of majority. If the transmitted codeword is n-zeros corresponding to a message digit 'zero', it is extremely unlikely that the noise on the channel would convert a majority of the 0s into 1s. So, if the majority of digits of the received sequence are decoded as 1s, the decoder decides that the transmitted message digit was a 1 and if the majority of the digits are 0s, it decides that the message digit was a 0.

Repetition codes possess good error correction capacity provided the channel noise is random and if P_e is low. However, their code rate is generally very low.

Example 13.7 Determine (a) code rate, (b) the generator matrix ***G***, (c) the parity-check matrix ***H***, and (d) the coefficient matrix, P for a (3, 1) repetition code.

Solution Since it is a (3, 1) code, $n = 3$ and $k = 1$.

Hence, the code rate $= \dfrac{1}{n} = \dfrac{1}{3}$ bits/symbol

The vector $\boldsymbol{m} = 1$, i.e., a row vector of length 1.
The vector $\boldsymbol{b} = [\ 1\ 1\ ]$, a row vector of length 2.
From Eq. (13.7) it then follows that $[P] = [1\ 1]$ = coefficient matrix.

Now, $$[G]_{k\times n} = [P \vdots I_k]_{k\times n} = [111]_{k\times n=1\times 3}$$

and $$[H]_{(n-k)\times n} = \left[I_{n-k} \vdots P^T\right] = \begin{bmatrix} 1 & 0 & \vdots & 1 \\ 0 & 1 & \vdots & 1 \end{bmatrix}$$

Example 13.8 Data stream encoded using a (3, 1) repetition code, is transmitted over a binary symmetric channel having a transition probability of $P = 10^{-2}$. Determine the probability of error P_e of the symbols received at the destination.

Solution Channel encoder produces codewords (0, 0, 0) and (1, 1, 1) respectively for message binary digits 0 and 1, respectively. Irrespective of whether a (0, 0, 0) is transmitted, or a (1, 1, 1) is transmitted, since there are 3 digits in the codeword, the length-3 sequence at the output of the BSC will have to be one of the following eight length-3 sequences. Let us assume the message digit is 1 and that consequently the codeword given as input to the BSC is (1, 1, 1).

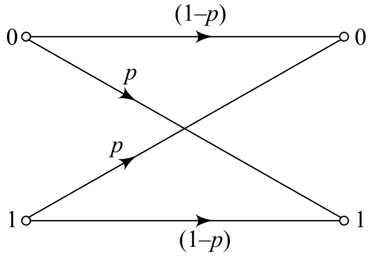


Fig. 13.4 *A binary symmetric channel*

Message bit	Encoded word		Received Word	Decoded word	Received digit
1	(1, 1, 1)		(0 0 0)	(0 0 0)	0
			(0 0 1)	(0 0 0)	0
			(0 1 0)	(0 0 0)	0
			(0 1 1)	(1 1 1)	1
			(1 0 0)	(0 0 0)	0
			(1 0 1)	(1 1 1)	1
			(1 1 0)	(1 1 1)	1
			(1 1 1)	(1 1 1)	1

Since the message bit is 1, an error is said to have taken place if the received bit is a 0. As can be seen from the above table, this can happen if the decoded word is (0 0 0). The decoded word would be (0 0 0) if the received word is either all zeros, or if two of the three digits are 0s. That is, when either all three digits that are transmitted are received erroneously, the probability of which is p^3, or two of the digits are received erroneously and the third is received correctly, the probability of which is $p^2(1-p)$. Since the latter case can arise if the received word is either (0 0 1) or (0 1 0) or (1 0 0), the probability of the received digit being erroneous is given by

Probability of a message bit 1 being received as a 0 after decoding = $p^3 + 3p^2(1-p)$.
From the symmetry of the problem, the problem of a message bit 0 being received as 1 is also the same.
$\therefore$ assuming that the message bit is a 1 or a 0 with equal probability, the average probability of error, P_e is given by

$$\begin{aligned} P_e &= (\text{Problem of message bit being 1}) \times (\text{Problem of receiving it as a 0}) \\ &\quad + (\text{Problem of message bit being 0}) \times (\text{Problem of receiving it as a 1}) \\ &= 0.5\lfloor p^3 + 3p^2(1-p)\rfloor + 0.5\lfloor p^3 + 3p^2(1-p)\rfloor \\ &= p^3 + 3p^2(1-p) \end{aligned}$$

Substituting $p = 10^{-2}$, we get

$$P_e = 2.98 \times 10^{-4}$$

13.5.5 Minimum Distance $d_{\min}$ of a Linear Block Code

Let $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ be two *distinct* binary sequences of the same length. Though they are of the same length, since they are two distinct sequences, their digits in corresponding locations will be different in one or more locations. We may then use the number of locations in which they differ as a measure of the distance between the two sequences or vectors. Accordingly, we define the Hamming distance as follows:

Hamming distance The Hamming distance $d(\boldsymbol{C}_1, \boldsymbol{C}_2)$ between two code vectors having the same number of elements is defined as the number of locations in which their respective elements differ.

Hamming weight w(c) of a code vector The Hamming weight of a code vector is defined as the number of non-zero elements in it.

Obviously, the Hamming weight of a code vector will be the same as the Hamming distance between that code vector and the all-zero code vector of the same length.

13.5.6 Minimum Distance $d_{\min}$ of a Code

The minimum distance $d_{\min}$ of a linear block code is the smallest Hamming distance between any two code vectors of the code

We know that when two binary code vectors of a linear block code are added (same as subtraction because of modulo-2 arithmetic) the resulting binary sequence also will be a code vector since the code is a linear code. We also know that this resultant code vector will have 1s only in those positions in which the elements differ. The number of positions in which they differ is the Hamming distance between them; but since the resultant code vector has 1s only at those locations, it is also equal to the Hamming weight of the resultant vector. *Thus, **we may state that the minimum distance $d_{\min}$ of a linear block code is the minimum value of the Hamming weight among all the non-zero code vectors of the code***.

The value of $d_{\min}$ of a linear block code depends on the structure of the code and it can be shown (see Example 13.9) that $d_{\min}$ of a linear block code is also equal to the minimum number of rows of the H^T matrix which, when added will result in the all-zero vector.

Example 13.9 Show that $d_{\min}$ of a linear block code is equal to the minimum number of the rows of the H^T matrix which will add up to zero.

Solution Consider a binary linear (n, k) block code. We know that its parity-check matrix H is of size $(n-k) \times n$. So, let us represent this matrix as follows:

$$[H]_{(n-k)\times n} = [\boldsymbol{h}_0 \quad \boldsymbol{h}_1 \quad \boldsymbol{h}_2 \ldots \boldsymbol{h}_{n-1}]$$

where $\boldsymbol{h}_j$, $j = 0$ to $(n - 1)$ are the n columns of the H matrix. So, we may write $[H^T]_{n\times(n-k)}$ matrix as follows:

$$[H^T]_{n\times(n-k)} = \begin{bmatrix} h_{0,0} & h_{0,1} & \cdots & h_{0,n-k-1} \\ h_{1,0} & h_{1,1} & \cdots & h_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ h_{n-1,0} & h_{n-1,1} & \cdots & h_{n-1,n-k-1} \end{bmatrix}$$

Let $\boldsymbol{C} = (c_0, c_1, \ldots, c_{n-1})$ be the non-zero code vector with minimum Hamming weight. We know that the elements c_0, c_1, etc., are either 0 or 1, and that the Hamming weight represents the number of 1s in $\boldsymbol{C}$. We know that when we consider

$$\boldsymbol{CH}^T = (c_0, c_1, \ldots, c_{n-1}) \begin{bmatrix} h_{0,0} & h_{0,1} & \cdots & h_{0,n-k-1} \\ h_{1,0} & h_{1,1} & \cdots & h_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ h_{n-1,0} & h_{n-1,1} & \cdots & h_{n-1,n-k-1} \end{bmatrix} = [0]_{1\times(n-k)}$$

the resulting $1 \times (n–k)$ vector is the sum of those rows of $\boldsymbol{H}^T$ matrix which correspond to the location of 1s in the code vector $\boldsymbol{C}$. For instance, if $\boldsymbol{C}$ has 1s only in the i^{th} and m^{th} positions, then the $1 \times (n–k)$ vector obtained by taking $\boldsymbol{CH}^T$ is the sum of the i^{th} and m^{th} rows of the $\boldsymbol{H}^T$ matrix. Since $d_{\min}$ corresponds to the minimum Hamming weight among all the non-zero code vectors, it follows that $d_{\min}$ equals the minimum number of rows of $\boldsymbol{H}^T$ matrix that would add up to zero vector.

13.5.7 Syndrome and Its Properties

The transmitter transmits only codewords. But owing to channel noise, what is received may or may not be a codeword and even if it is, it may not be the same codeword which was transmitted. This is because the noise can change one or more of the n binary digits of a transmitted codeword. We know that if we add a 1 to a binary digit, it is changed – a 0 to a 1 and a 1 to a 0. So, suppose an n-length codeword is affected in two positions. We can model this change as one of adding an n-length vector $\boldsymbol{e}$ to the transmitted code vector $\boldsymbol{C}$, where $\boldsymbol{e}$ has 1s only at those positions in which the code vector is affected and zeros everywhere else. So, if $\boldsymbol{r}$ is the resultant received vector.

$$\boldsymbol{r} = \boldsymbol{C} + \boldsymbol{e} \tag{13.24}$$

where $\boldsymbol{e} = (e_0, e_1, \ldots, _{n-1})$

with
$$e_i = \begin{cases} 1 & \text{1 if an error has occurred in the } i^{\text{th}} \text{ error-position} \\ 0 & \text{otherwise} \end{cases} \tag{13.25}$$

In Eq. (13.24), the vector $\boldsymbol{e}$ is called the error-pattern, since its structure depends on the error locations.

From Eq. (13.17), we know that $\boldsymbol{CH}^T = \boldsymbol{0}$ if and only if $\boldsymbol{C}$ is a code vector. So, if $\boldsymbol{C}$ is not a code vector, $\boldsymbol{CH}^T$ will not be an all-zero $(n–k)$ length row vector. Suppose we now consider

$$\boldsymbol{S} = \boldsymbol{rH}^T \tag{13.26}$$

where $\boldsymbol{r}$ is the n-length received vector, and the $(n–k)$ row vector $\boldsymbol{S}$ is what is called the '*Syndrome*'. The word syndrome actually means 'a group of symptoms that are characteristic of a specific disease'. The appropriateness of this name given to the vector $\boldsymbol{S}$ will be evident in the discussion that follows.

Note

It must be noted that while $\boldsymbol{S} \neq \boldsymbol{0}$ implies that $\boldsymbol{r}$ is not a code vector and that there are some errors, $\boldsymbol{S}$ = 0 does not automatically imply that there are no errors. It only implies that $\boldsymbol{r}$ is a codeword; but it may or may not be the codeword that was transmitted. The transmitted codeword might have been transformed into another codeword.

There are certain properties of this 'syndrome' which are extremely useful in understanding the way the error correction is done. These properties of the syndrome are:

Property 1: The syndrome is independent of the transmitted code vector. It depends only on the error pattern.

$$\begin{aligned} \boldsymbol{S} &= \boldsymbol{r}\boldsymbol{H}^T \\ &= (\boldsymbol{C}+\boldsymbol{e})\boldsymbol{H}^T = \boldsymbol{C}\boldsymbol{H}^T + \boldsymbol{e}\boldsymbol{H}^T \end{aligned}$$

But $\boldsymbol{C}\boldsymbol{H}^T = \boldsymbol{0}$ since $\boldsymbol{C}$ is a codeword.

$$\therefore \qquad \boldsymbol{S} = \boldsymbol{e}\boldsymbol{H}^T \tag{13.27}$$

which shows that the syndrome $\boldsymbol{S}$ is independent of which codeword has been transmitted and that it depends only on the error pattern $\boldsymbol{e}$.

Property 2 Error patterns differing by a codeword will have the same syndrome. Suppose $\boldsymbol{e}_1$ is an error pattern and $\boldsymbol{e}_2 = \boldsymbol{e}_1 + \boldsymbol{C}$ where $\boldsymbol{C}$ is any codeword. Then the syndrome corresponding to the error pattern $\boldsymbol{e}_1$ is

$$\boldsymbol{S}_1 = \boldsymbol{e}_1\boldsymbol{H}^T$$

The syndrome of the error pattern $\boldsymbol{e}_2$ is given by

$$\boldsymbol{S}_2 = \boldsymbol{e}_2\boldsymbol{H}^T = (\boldsymbol{e}_1+\boldsymbol{C})\boldsymbol{H}^T = \boldsymbol{e}_1\boldsymbol{H}^T + \boldsymbol{C}\boldsymbol{H}^T = \boldsymbol{e}_1 \cdot \boldsymbol{H}^T = \boldsymbol{S}_1$$

$$\therefore \qquad \boldsymbol{S}_2 = \boldsymbol{S}_1$$

Thus, error patterns differing by a codeword have the same syndrome.

Cosets Suppose $\boldsymbol{e}$ is some arbitrary error pattern.

$$\text{Let} \qquad \boldsymbol{e}_1 \underline{\Delta}\, \boldsymbol{e} + \boldsymbol{C}_i;\ i = 0, 1, 2, \ldots, (2^k-1) \tag{13.28}$$

(Since there are 2^k code vectors for an (n, k) code.)

Then, we know from property 2 of the syndromes that all the 2^k error patterns, $\boldsymbol{e}_i$s of Eq. (13.28) have the same syndrome. This set of 2^k error patterns defined by $\boldsymbol{e}_i$, $i = 0, 1, 2, \ldots, (2^k-1)$ in Eq. (13.28), having a common syndrome, is said to form a ***coset*** of the code.

Definition A coset of an (n, k) block code is a set of 2^k error patterns, characterized by a unique syndrome for all its elements.

We know that we have altogether 2^n distinct error patterns. If a set of 2^k error patterns form a coset with a common syndrome, then the number of such cosets that we can have for an (n, k) linear block code, is $2^n \div 2^k = 2^{n-k}$ cosets.

$$\therefore \qquad \text{number of cosets} = 2^{n-k} \tag{13.29}$$

Error pattern Although the syndrome is determined entirely by the error pattern and the structure of the code (i.e., $\boldsymbol{H}$), it is not possible to uniquely determine the error pattern from knowledge of the syndrome and the parity-check matrix, $\boldsymbol{H}$. This is because, the relation

$$\underset{(1\times n-k)}{\boldsymbol{S}} = \underset{(1\times n)}{\boldsymbol{e}}\ \underset{(n\times n-k)}{\boldsymbol{H}^T}$$

represents only a set of $(n-k)$ linear equations, each equation involving one of the $(n-k)$ syndrome elements on one side and a linear combinations of the error pattern elements and parity check bits on the other side.

Since we have 2^n unknowns – the error pattern elements, i.e., $e_0, e_1, \ldots, e_{2^n-1}$, and only 2^{n-k} linear equations, the set of equations is under-determined. So, it is not possible to solve the equations and get a unique solution for $\boldsymbol{e}$, the error pattern.

13.5.8 $d_{\min}$ and Error-Correcting Capability of the Code

If the error pattern can be exactly determined from the syndrome of the received vector, we can just add that error pattern to the received vector and get the code vector that was actually transmitted. But, as pointed out earlier, exact determination of the error pattern is not possible. So, some other decoding strategy which will enable us to pick the best possible codeword from knowledge of $\boldsymbol{r}$ has to be adopted. The '*minimum distance strategy*' gives such an approach.

If the received vector, $\boldsymbol{r}$, has a non-zero syndrome, certainly $\boldsymbol{r}$ is not a codeword and the transmitter would not have transmitted it. To find the codeword that is *most likely* to have been transmitted, in the minimum distance strategy, we pick that codeword $\boldsymbol{C}_i$ which is closest to the received vector $\boldsymbol{r}$. In other words, that $\boldsymbol{C}_i$ for which $d(\boldsymbol{C}_i, \boldsymbol{r})$ is the least. It can be shown that provided d_{min}, the minimum distance between any two codewords of the code is $(2t + 1)$, this approach enables us to detect and correct all error patterns whose Hamming weight $w(\boldsymbol{e})$ is less than or equal to t, i.e., any received vector $\boldsymbol{r}$ having at the most t errors can be correctly decoded provided d_{min} is at least $(2t + 1)$.

Although it is by no means a proof for the above assertion, the following geometrical interpretation provides a justification for it. A code vector, being an n-tuple, is represented as a point in an n-dimensional space. Let $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ be two code vectors. An error pattern with t errors is also an n-tuple and since it has only t number of 1s, its Hamming weight is t. Imagine that we have circles drawn with centers at $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ and radii equal to t. If $\boldsymbol{C}_i$ is the code vector that is transmitted and if the received vector $\boldsymbol{r}$ is having t or less number of errors, the point $\boldsymbol{r}$ corresponding to the received vector will be within the circle of radius t drawn with center at $\boldsymbol{C}_i$, as shown. Then $d(\boldsymbol{C}_i, \boldsymbol{r}) < d(\boldsymbol{C}_j, \boldsymbol{r})$ if the two circles are non-intersecting as shown in (a), i.e., if $d(\boldsymbol{C}_i, \boldsymbol{C}_j) > 2t$. In such a case, the received vector $\boldsymbol{r}$ will be identified by the decoder as $\boldsymbol{C}_i$. However, if the two circles intersect, as shown in Fig. 13.5(b), then there is no guarantee that $d(\boldsymbol{C}_i, \boldsymbol{r}) < d(\boldsymbol{C}_j, \boldsymbol{r})$ even if the error-pattern has a Hamming weight that is less than or equal to t. Hence, all error-patterns with at the most t error can be corrected *if and only if* $d(\boldsymbol{C}_i, \boldsymbol{C}_j) \geq (2t + 1)$ for all $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$. In other words, it means that an (n, k) linear block code with minimum distance d_{min} can correct up to t errors *if and only if*

(a)

(b)

Fig. 13.5 *Illustration of conditions for correct decoding*

$$t \leq \left\lfloor \frac{1}{2}(d_{\text{min}} - 1) \right\rfloor \tag{13.30}$$

where $\lfloor x \rfloor$ denotes the *largest integer* less than or equal to x.

13.5.9 Syndrome Decoding

Now, we are in a position to discuss a decoding method, called 'Syndrome decoding'.

Let us partition the 2^n possible received vectors into 2^k non-overlapping subsets as shown in the following array, called the '*standard array*'. Thus, there are 2^k columns in the array, each column being led by a code vector commencing with the all-zero vector at the left-most corner. Since there are 2^k columns, there will be $2^n \div 2^k = 2^{n-k}$ rows. Each of these rows forms a '***coset***' and the left-most element of each coset is called the '***coset leader***'. The first row comprises the 2^k possible zero-error received vectors. The coset leader for the second row is say an error pattern $\boldsymbol{e}_2$ which is most likely. The other elements of this row, i.e., the coset are $(\boldsymbol{C}_2 + \boldsymbol{e}_2), (\boldsymbol{C}_3 + \boldsymbol{e}_2), \ldots, (\boldsymbol{C}_{2^k} + \boldsymbol{e}_2)$. The coset leader for the third row, viz.; $\boldsymbol{e}_3$ is then selected and the other members of this coset, viz: $(\boldsymbol{C}_2 + \boldsymbol{e}_3), (\boldsymbol{C}_3 + \boldsymbol{e}_3), \ldots, (\boldsymbol{C}_{2^k} + \boldsymbol{e}_3)$ are filled up. Every time we pick a coset leader, we should make sure that it has not already appeared in the standard array. The coset leaders must be so chosen they are the most likely error patterns – those with *smallest* Hamming weight.

The decoding procedure consists of the following steps:

1. Determine the syndrome of the received vector $\boldsymbol{r}$:
$$\boldsymbol{S} = \boldsymbol{r} \cdot \boldsymbol{H}^T$$
2. Identify the coset with this syndrome and let its coset leader be an error pattern $\boldsymbol{e}$.
3. Decode the received vector $\boldsymbol{r}$ into the code vector $\boldsymbol{C} = \boldsymbol{r} + \boldsymbol{e}$.

Standard Array

$$\begin{array}{ccccccc}
\boldsymbol{C}_1 = \boldsymbol{0} & \boldsymbol{C}_2 & \boldsymbol{C}_3 & \cdots & \boldsymbol{C}_i & \cdots & \boldsymbol{C}_{2^k} \\
\boldsymbol{e}_2 & \boldsymbol{C}_2 + \boldsymbol{e}_2 & \boldsymbol{C}_3 + \boldsymbol{e}_2 & \cdots & \boldsymbol{C}_i + \boldsymbol{e}_2 & \cdots & \boldsymbol{C}_{2^k} + \boldsymbol{e}_2 \\
\boldsymbol{e}_3 & \boldsymbol{C}_2 + \boldsymbol{e}_3 & \boldsymbol{C}_3 + \boldsymbol{e}_3 & \cdots & \boldsymbol{C}_i + \boldsymbol{e}_3 & \cdots & \boldsymbol{C}_{2^k} + \boldsymbol{e}_3 \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\boldsymbol{e}_J & \boldsymbol{C}_2 + \boldsymbol{e}_J & \boldsymbol{C}_3 + \boldsymbol{e}_J & \cdots & \boldsymbol{C}_i + \boldsymbol{e}_J & \cdots & \boldsymbol{C}_{2^k} + \boldsymbol{e}_J \\
\vdots & \vdots & \vdots & & \vdots & & \vdots \\
\boldsymbol{e}_{2^{n-k}} & \boldsymbol{C}_2 + \boldsymbol{e}_{2^{n-k}} & \boldsymbol{C}_3 + \boldsymbol{e}_{2^{n-k}} & \cdots & \boldsymbol{C}_i + \boldsymbol{e}_{2^{n-k}} & \cdots & \boldsymbol{C}_{2^k} + \boldsymbol{e}_{2^{n-k}}
\end{array}$$

Fig. 13.6 *Standard array for an (n, k) linear block code*

The storage or memory space requirement for array decoding increases exponentially with the number of parity check bits used in the code. To store the 2^{n-k} coset leaders, each with n digits, we need $n.2^{n-k}$ digits storage. Further, to store the 2^{n-k} syndromes each of $(n–k)$ digits, we need $(n–k)\ 2^{n-k}$ digits storage. Thus, the total storage requirement is $(2n–k)\ 2^{n-k}$ bits.

Example 13.10 A linear (n, k) block code has a generator matrix:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

(a) Find all its codewords
(b) Find its $\boldsymbol{H}$ – matrix
(c) What is the minimum distance of the code and what is its error-correcting capacity.

Solution

(a) Since $k = 2$, the possible message sequences are (0, 0), (0, 1), (1, 0), (1, 1). The corresponding code vectors are obtained by premultiplying $\boldsymbol{G}$ by the row vector representing the message sequence.

$$\begin{aligned}
\boldsymbol{C}_0 &= [0 \quad 0][\boldsymbol{G}] = [0 \quad 0 \quad 0 \quad 0] \\
\boldsymbol{C}_1 &= [0 \quad 1][\boldsymbol{G}] = [0 \quad 1 \quad 1 \quad 0] \\
\boldsymbol{C}_2 &= [1 \quad 0][\boldsymbol{G}] = [1 \quad 0 \quad 1 \quad 1] \\
\boldsymbol{C}_3 &= [1 \quad 1][\boldsymbol{G}] = [1 \quad 1 \quad 0 \quad 1]
\end{aligned}$$

(b) To find the $\boldsymbol{H}$-matrix: $\boldsymbol{G} = [I_2 \vdots p]$

Hence, $\boldsymbol{H} = [P^T \vdots I_{n-k}] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

(c) $d_{\min}$ and t: We know that $d_{\min}$ of the code is given by the minimum Hamming weight of the non-zero code vectors. $\boldsymbol{C}_1$, $\boldsymbol{C}_2$, and $\boldsymbol{C}_3$ are the non-zero code vectors and of them, $\boldsymbol{C}_1$ has the minimum Hamming weight of 2.

Hence, $d_{\min} = 2$ for this code

If t errors are to be corrected, $d_{\min}$ should be equal to or greater than $(2t + 1)$. As no positive integer value of t satisfies the equation

$$d_{\min} = 2 \geq (2t + 1)$$

the value of $t = 0$.

Example 13.11 The generator matrix of a (6, 3) linear block code is given by

$$[\boldsymbol{G}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

(a) Find all its code vectors.

(b) What is its error correcting capacity?

Solution

(a) Since $k = 3$, there should be 8 codewords corresponding to the 8 possible distinct binary message sequences of length 3. These are:

Message sequence	Codeword		
0 0 0	0 0 0 0 0 0	=	All-zero code vector
0 0 1	0 0 1 1 1 0	=	r_3
0 1 0	0 1 0 1 1 1	=	r_2
0 1 1	0 1 1 0 0 1	=	$r_2 + r_3$
1 0 0	1 0 0 0 1 1	=	r_1
1 0 1	1 0 1 1 0 1	=	$r_1 + r_3$
1 1 0	1 1 0 1 0 0	=	$r_1 + r_2$
1 1 1	1 1 1 0 1 0	=	$r_1 + r_2 + r_3$

where r_1, r_2 and r_3 are the first, second and third rows of the generator matrix $\boldsymbol{G}$.

(b) The minimum Hamming weight of the non-zero code vector is 3.

$\therefore$ $d_{\min} = 3$

Hence, $t = 1$ since $d_{\min} = 2t + 1$

i.e., the given code has single-error correcting capacity.

Example 13.12 The parity-check equations for a (6, 3) systematic code are

$$b_0 = c_3 = m_0 + m_1 + m_2$$
$$b_1 = c_4 = m_0 + m_1$$
$$b_3 = c_5 = m_0 + m_2$$

(a) Determine the generator matrix of the code.

(b) Determine the parity-check matrix of the code.

(c) List out all the code vectors of the code.

(d) What is the error-correcting capability of the code?

(e) Prepare an appropriate decoding table.

(f) Decode the following received words: (i) 1 0 1 1 0 0, (ii) 0 1 0 0 1 1, and (iii) 0 0 1 1 0 0

Solution

(a) and (b) The given parity-check equations may be written as

$$1\cdot m_0 + 1\cdot m_1 + 1\cdot m_2 + 1\cdot b_0 + 0\cdot b_1 + 0\cdot b_2 = 0$$

$$1\cdot m_0 + 1\cdot m_1 + 0\cdot m_2 + 0\cdot b_0 + 1\cdot b_1 + 0\cdot b_2 = 0$$

$$1\cdot m_0 + 0\cdot m_1 + 1\cdot m_2 + 0\cdot b_0 + 0\cdot b_1 + 1\cdot b_2 = 0$$

$$\therefore\ \boldsymbol{H} = \begin{bmatrix} 1 \quad 1 \quad 1 & 1 \quad 0 \quad 0 \\ 1 \quad 1 \quad 0 & 0 \quad 1 \quad 0 \\ \underbrace{1 \quad 0 \quad 1}_{\boldsymbol{p}^T} & \underbrace{0 \quad 0 \quad 1}_{\boldsymbol{I}_{n-k}} \end{bmatrix} \qquad \therefore\ \boldsymbol{G} = \begin{bmatrix} 1 \quad 0 \quad 0 & 1 \quad 1 \quad 1 \\ 0 \quad 1 \quad 0 & 1 \quad 1 \quad 0 \\ \underbrace{0 \quad 0 \quad 1}_{\boldsymbol{I}_{n-k}} & \underbrace{1 \quad 0 \quad 1}_{\boldsymbol{P}} \end{bmatrix}$$

(c) We know that $\boldsymbol{C} = (c_0, c_1, c_2, c_3, c_4, c_5)$

$$= (m_0, m_1, m_2, b_0, b_1, b_2)$$

Since $k = 3$, there will be $2^3 = 8$ distinct message sequences. We shall first find out the code vectors corresponding to each of these 8 message sequences by finding b_0, b_1 and b_2 using the parity-check equations. We tabulate the result as follows.

Message $\boldsymbol{m}$	$b_0\ b_1\ b_2$	$c_0\ c_1\ c_2\ c_3\ c_4\ c_5$		$\boldsymbol{C}$
0 0 0	0 0 0	0 0 0 0 0 0	=	$\boldsymbol{C}_1$
0 0 1	1 0 1	0 0 1 1 0 1	=	$\boldsymbol{C}_2$
0 1 0	1 1 0	0 1 0 1 1 0	=	$\boldsymbol{C}_3$
0 1 1	0 1 1	0 1 1 0 1 1	=	$\boldsymbol{C}_4$
1 0 0	1 1 1	1 0 0 1 1 1	=	$\boldsymbol{C}_5$
1 0 1	0 1 0	1 0 1 0 1 0	=	$\boldsymbol{C}_6$
1 1 0	0 0 1	1 1 0 0 0 1	=	$\boldsymbol{C}_7$
1 1 1	1 0 0	1 1 1 1 0 0	=	$\boldsymbol{C}_8$

(d) From the above list of codewords, we find that $d_{\min} = 3$ = Minimum Hamming weight of any code vector.

But $d_{\min} = 2t + 1$ where t is the error-correcting capability

$\therefore\ t = 1$. $\quad\therefore$ It is a single-error correcting code

(e) Now, we will list the most likely error patterns and the corresponding syndromes. This will form the decoding table:

Error Patterns	$\boldsymbol{H}^T$	$\boldsymbol{S}$ = syndrome
$\boldsymbol{e}_0 = [0\,0\,0\,0\,0\,0]$	1 1 1	[0 0 0]
$\boldsymbol{e}_1 = [0\,0\,0\,0\,0\,1]$	1 1 0	[0 0 1]
$\boldsymbol{e}_2 = [0\,0\,0\,0\,1\,0]$	1 0 1	[0 1 0]
$\boldsymbol{e}_3 = [0\,0\,0\,1\,0\,0]$	1 0 0	[1 0 0]
$\boldsymbol{e}_4 = [0\,0\,1\,0\,0\,0]$	0 1 0	[1 0 1]
$\boldsymbol{e}_5 = [0\,1\,0\,0\,0\,0]$	0 0 1	[1 1 0]
$\boldsymbol{e}_6 = [1\,0\,0\,0\,0\,0]$		[1 1 1]

(f) (i) For the given received vector $\boldsymbol{r}$ = 1 0 1 1 0 0, the syndrome is

$$\boldsymbol{S} = r_1\boldsymbol{H}^T = [1 \quad 1 \quad 0]$$

From the above table in part (e), we find that $\boldsymbol{e} = \boldsymbol{e}_5 = [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]$

∴ the transmitted code vector = $\boldsymbol{C} = \boldsymbol{r} + \boldsymbol{e}_5$

∴ $$\boldsymbol{C} = [1 \;\; 1 \;\; 1 \;\; 1 \;\; 0 \;\; 0] = \boldsymbol{C}_7$$

(ii) For the given received vector $r_2 = [0 \;\; 1 \;\; 0 \;\; 0 \;\; 1 \;\; 1]$, the syndrome is

$$\boldsymbol{S} = r_2\boldsymbol{H}^T = [1 \;\; 0 \;\; 1]$$

For this syndrome, from the table given in part (e), we find that the error-pattern is

$$\boldsymbol{e} = \boldsymbol{e}_4 = [0 \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; 0]$$

∴ The transmitted code vector = $\boldsymbol{C} = r_2 + \boldsymbol{e}_4$

∴ $$\boldsymbol{C} = [0 \;\; 1 \;\; 0 \;\; 0 \;\; 1 \;\; 1] + [0 \;\; 0 \;\; 1 \;\; 0 \;\; 0 \;\; 0] = [0 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \;\; 1] = \boldsymbol{C}_4$$

(iii) For the given received vector $r_3 = [0 \;\; 0 \;\; 1 \;\; 1 \;\; 0 \;\; 0]$, the syndrome is

$$\boldsymbol{S} = r_3\boldsymbol{H}^T = [0 \;\; 0 \;\; 1]$$

From the table given in part (e), we find that for this syndrome, the corresponding error pattern is

$$\boldsymbol{e} = \boldsymbol{e}_1 = [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 1]$$

∴ $$\boldsymbol{C} = [0 \;\; 0 \;\; 1 \;\; 1 \;\; 0 \;\; 0] + [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 1] = [0 \;\; 0 \;\; 1 \;\; 1 \;\; 0 \;\; 1] = \boldsymbol{C}_1$$

∴ The transmitted code vector is = $\boldsymbol{C}_1 = [0 \;\; 0 \;\; 1 \;\; 1 \;\; 0 \;\; 1]$

13.5.10 Hadamard Codes

With k denoting the number of bits in the uncoded message sequence and $n = 2^k$ denoting the number of bits in a codeword, these codewords are the rows of a $n \times n$ Hadamard matrix. The smallest size Hadamard matrix is a 2 × 2 square matrix and corresponds to $k = 1$. In order to avoid confusion, instead of using H_2, we shall use $\boldsymbol{M}_2$ to denote a Hadamard matrix of order 2.

$$\boldsymbol{M}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{13.31}$$

In the Hadamard code, except the all-zero codeword, all the other codewords will have equal number of zeros and 1s. Hadamard matrices of larger sizes 4, 8, etc., can be constructed from the smallest sized ones using the following formula.

$$\boldsymbol{M}_4 = \begin{bmatrix} \boldsymbol{M}_2 & \boldsymbol{M}_2 \\ \boldsymbol{M}_2 & \overline{\boldsymbol{M}_2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{13.32}$$

where $\overline{\boldsymbol{M}_2}$ is the matrix obtained by replacing each element of $\boldsymbol{M}_2$ by its complement.

Another interesting feature of a Hadamard code is that the various codewords are orthogonal to each other. In an (n, k) Hadamard linear block code, since each codeword with $n = 2^k$ bits has k message bits, the number of parity check bits in a codeword = $(2^k – k)$. Obviously, as k increases, the number of parity bits in a codeword will increase very rapidly, bringing down the code rate.

$$\text{r = code rate} = \frac{k}{n} = \frac{k}{2^k} = k2^{-k} \tag{13.33}$$

Since the n bits of a codeword will have to occupy the same time as the original k message bits did, the transmission bandwidth of Hadamard encoded sequence will be very high. Hence, Hadamard codes are used only in situations where a large transmission bandwidth requirement is not a problem.

Since the codewords are orthogonal to each other, it means that for a codeword length of n bits, each codeword must differ from any other codeword in exactly $n/2$ places. This means that $d_{\min}$ for an (n, k) Hadamard block code is

$$d_{\min} = \frac{n}{2} = \frac{2^k}{2} = 2^{k-1} \tag{13.34}$$

with

$$t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor \quad , \qquad \text{(See Eq. (13.30))}$$

$$t = \left\lfloor \frac{1}{2}(2^{k-1} - 1) \right\rfloor = \left\lfloor 2^{k-2} - \frac{1}{2} \right\rfloor = (2^{k-2} - 1) \tag{13.35}$$

where $\lfloor x \rfloor$ denotes the largest integer less than x.

$$\therefore \qquad t = (2^{k-2} - 1) \tag{13.36}$$

From the above equation, it is clear that (i) k should be greater than 2 for this code to have any error correction capability, and (ii) As k is increased, although the code rate decreases as per Eq. (13.33), the error-correcting capability increases substantially.

Hamming code A Hamming code is a (n, k) linear block code with the following structure:

If $m \underline{\Delta} (n - k)$ = Number of parity check bits in a codeword, where n is the number of bits in a codeword and k is the number of message digits in a codeword, then

1. $n = (2^m - 1)$ where $m = (n - k) \geq 3$
2. k = no. of message bits = $(2^m - 1) - m$
3. Minimum distance $d_{\min} = 3$
4. Error-correcting capability = $t = 1$.

A systematic (n, k) Hamming code can be constructed by following the procedure indicated below:

As an illustration, let us construct a Hamming code with $m = 3$.

$\therefore$ $n = 2^3 - 1 = 7$. Also $k = (n - m) = (7 - 3) = 4$. Since $m = 3 = (n - k)$ and $k = 4$, construct a coefficient matrix, P of size $k \times (n - k) = (4 \times 3)$ size, by using all 3-bit words with 2 or more 1s, arranging them in any order.

$$\therefore \qquad P_{(k \times n-k)} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\therefore \text{ the generator matrix } \boldsymbol{G} = [P \vdots I_k]_{k \times n} = \begin{bmatrix} 1 & 0 & 1 & \vdots & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & \vdots & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & \vdots & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & \vdots & 0 & 0 & 0 & 1 \end{bmatrix}$$

Since $k = 4$, there will be $2^k = 16$ possible codewords. These can be determined by premultiplying the generator matrix G by each of the 16 4-bit binary message sequences.

Example 13.13 Determine all the codewords and their Hamming weights for the (7, 4) Hamming code whose generator matrix was determined in the above illustration as

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 1 & \vdots & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & \vdots & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & \vdots & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & \vdots & 0 & 0 & 0 & 1 \end{bmatrix}$$

Solution With $k = 4$, there will be $2^4 = 16$ possible *distinct* message sequences. The code vectors, determined using each of these message sequences, are tabulated below.

Codewords		Weight of codeword
Parity bits	Message bits	
0 0 0	0 0 0 0	0
0 1 1	0 0 0 1	3
1 1 0	0 0 1 0	3
1 0 1	0 0 1 1	4
1 1 1	0 1 0 0	4
1 0 0	0 1 0 1	3
0 0 1	0 1 1 0	3
0 1 0	0 1 1 1	4
1 0 1	1 0 0 0	3
1 1 0	1 0 0 1	4
0 1 1	1 0 1 0	4
0 0 0	1 0 1 1	3
0 1 0	1 1 0 0	3
0 0 1	1 1 0 1	4
1 0 0	1 1 1 0	4
1 1 1	1 1 1 1	7

Example 13.14 For the (7, 4) Hamming code of Example 13.13 determine the ***H***-matrix and the decoded codeword if the received codeword is 0 1 1 1 0 1 1

Solution Since

$$\boldsymbol{G} = \begin{bmatrix} \underbrace{\begin{matrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{matrix}}_{P} & \underbrace{\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}}_{I_k} \end{bmatrix}$$

The corresponding ***H***-matrix is

$$\boldsymbol{H} = \begin{bmatrix} \underbrace{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}}_{I_{n-k}} & \underbrace{\begin{matrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{matrix}}_{P^T} \end{bmatrix}$$

$$\boldsymbol{S} = \boldsymbol{r} \cdot \boldsymbol{H}^T = [0 \;\; 1 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \;\; 1] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = [0 \;\; 1 \;\; 1]$$

An error pattern [0 0 0 0 0 0 1] would have produced this syndrome, because the syndrome [0 1 0] is the seventh row of the $\boldsymbol{H}^T$ matrix. Hence, the 7th digit in the received vector is in error; or the error pattern is $\boldsymbol{e}$ = [0 0 0 0 0 0 1]. Hence, the decoded word is

$$\begin{aligned} \boldsymbol{C} = \boldsymbol{r} + \boldsymbol{e} &= [0 \;\; 1 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \;\; 1] + [0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 1] \\ &= [0 \;\; 1 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \;\; 0] \end{aligned}$$

13.5.11 Cyclic Codes

Although cyclic codes are a subclass of the linear block codes, they have some very good features which make them extremely useful. In fact, some of the very good block codes like the Hamming codes, BCH codes and Golay codes are in fact cyclic codes. The advantages of cyclic codes are:

1. They have an excellent mathematical structure which makes the design of error-correcting codes with multiple-error correction capability relatively easier.
2. The encoding and decoding circuits for cyclic codes can be easily implemented using shift registers.
3. Because of the availability of very efficient decoding methods that do not depend upon a look-up table, large memories are not needed for decoding. So powerful codes with $n >> 1$ can be used.
4. Cyclic codes can correct errors caused by bursts of noise that affect several successive bits.

It is because of these attractive features that almost all Forward Error Correcting (FEC) systems make use of cyclic codes.

Definition A linear block code is said to be a cyclic code if any cyclic shift of a codeword is also a codeword.

At this stage, a brief explanation of what we mean by a cyclic shift may be in order. Let $(c_0, c_1, \ldots, c_{n-1})$ be a codeword of a cyclic code.

Original codeword	$\boldsymbol{C} = (c_0, c_1, c_2, \ldots, c_{n-2}, c_{n-1})$	(13.37)
When $\boldsymbol{C}$ is given a cyclic shift to the right	$(c_{n-1}, c_0, c_1, \ldots, c_{n-3}, c_{n-2})$	
When $\boldsymbol{C}$ is given two cyclic shifts to the right	$(c_{n-2}, c_{n-1}, c_0, c_1, \ldots, c_{n-4}, c_{n-3})$	
$\vdots$		
When $\boldsymbol{C}$ is given $(n-1)$ cyclic shifts to the right	$(c_1, c_2, c_3, \ldots, c_{n-1}, c_0)$	

All the above n-tuples are also codewords.

We may also think of a left cyclic shift to a codeword. If the original codeword $\boldsymbol{C}$ is given one left cyclic shift, we get

When $\boldsymbol{C}$ is given one left cyclic shift $\quad (c_1, c_2, c_3, \ldots, c_0) \quad$ (13.38)

***Notice that the result of giving k cyclic shifts to the right is the same as giving (n–k) cyclic shifts to the left*.**

To explain the various algebraic properties and operations pertaining to cyclic codes, it is necessary to associate a codeword of the cyclic code with a polynomial $c(x)$ as shown below:

Codeword $\boldsymbol{C} = (c_0, c_1, c_2, \ldots, c_{n-1})$

$$\text{Code polynomial } c(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-1}x^{n-1} \tag{13.39}$$

Here, the coefficients $c_0, c_1, c_2, \ldots, c_{n-1}$ of the polynomial are the elements $c_0, c_1, c_2, \ldots, c_{n-1}$ of the codeword and so are binary digits 0 or 1.

Effect of multiplying c(x) by x^k $\quad c(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-k}x^{n-k} + \ldots + c_{n-1}x^{n-1}$

$$\therefore \quad x^k \cdot c(x) = c_0 x^k + c_1 x^{k+1} + c_2 x^{k+2} + \ldots + c_{n-k}x^n + \ldots + c_{n-1}x^{n+k-1} \tag{13.40}$$

Rearranging the terms, we may write Eq. (13.40), as

$$\left[x^k \cdot c(x) = c_{n-k}x^n + c_{n-k+1}x^{n+1} + \ldots + c_{n-1}x^{n+k-1} + c_0 x^k + c_1 x^{k+1} + \ldots + c_{n-k-1}x^{n-1}\right] \tag{13.41}$$

$$= \left[c_{n-k}(x^n + 1) + c_{n-k+1}x(x^n + 1) + \ldots + c_{n-1}x^{k-1}(x^n + 1)\right]$$

$$+ \left[c_{n-k} + c_{n-k+1}\, x + \ldots + c_{n-1}x^{k-1}\right] + \left[c_0 x^k + c_1 x^{k+1} + \ldots + c_{n-k-1}x^{n-1}\right] \tag{13.42}$$

Note

All the terms in the second rectangular bracket of the above equation are for canceling identical additional terms deliberately introduced in the first rectangular bracket. For instance

$$c_{n-k}(x^n+1)+c_{n-k}=c_{n-k}x^n+(c_{n-k}+c_{n-k})=c_{n-k}x^n,$$

since $c_{n-k,}$ etc., are binary digits and modulo-2 addition of a binary digit with itself yields zero.

Hence, rearranging the terms of Eq. (13.42), we have

$$x^k c(x)=[c_{n-k}+c_{n-k+1}x+\ldots+c_{n-1}x^{k-1}+c_0x^k+c_1x^{k+1}+\ldots+c_{n-k-1}x^{n-1}]$$
$$+[c_{n-k}(x^n+1)+c_{n-k+1}x(x^n+1)+\ldots+c_{n-1}x^{k-1}(x^n+1)] \quad (13.43)$$

Now we define

$$q(x)\,\underline{\Delta}\,(c_{n-k}+c_{n-k+1}x+\ldots+c_{n-1}x^{k-1}),$$

Then, the term inside the second rectangular bracket in Eq. (13.43), may be recognized as

$$q(x)[x^n+1]$$

Further, the quantity in the first rectangular brackets of Eq. (13.43), may be recognized as the polynomial corresponding to the following n-tuple:

$$\boldsymbol{C}^{(k)}=(c_{n-k},c_{n-k+1},\ldots,c_{n-1},c_0,c_1,\ldots,c_{n-k-1}) \quad (13.44)$$

In fact, $\boldsymbol{C}^{(k)}$ of Eq. (13.44) is the sequence obtained by giving k right cyclic shifts to the original code vector of Eq. (13.37). So we shall use the notation $c^{(k)}(x)$ to represent the polynomial corresponding to this k-shifted sequence.

$$\therefore \quad c^{(k)}(x)\,\underline{\Delta}\,c_{n-k}+c_{n-k+1}x+\ldots,c_{n-1}x^{k-1}+c_0x^k+c_1x^{k+1}+\ldots,c_{n-k-1}x^{n-1} \quad (13.45)$$

Thus, we may rewrite Eq. (13.43) as

$$x^k c(x)=q(x)(x^n+1)+c^{(k)}(x) \quad (13.46)$$

The Eq. (13.46) tells us that $c^{(k)}(x)$ is obtained as the remainder when we divide $x^k c(x)$ by $(x^n + 1)$. We state this as

$$c^{(k)}(x)=x^k c(x)\text{ module }(x^n+1) \quad (13.47)$$

and $c^{(k)}(x)$ also is a code polynomial since $c^{(k)}$ is also a code vector. Since k is an arbitrary integer, it follows that the cyclic code property reduces to saying that

If $c(x)$ is a code polynomial, then $c^{(k)}(x)$, which is a k-times right-shifted version of $c(x)$, is also a code polynomial.

It should be noted that since Eq. (13.47) states that $c^{(k)}(x)=x^k c(x)$ modulo (x^n+1), it means that $x^n = 1$ and so $x^n c(x)=c(x)$. Hence, when we go on giving cyclic shifts to $c(x)$, after the n^{th} shift the original polynomial is obtained.

13.5.12 Generator Polynomial

A factor g(x) of (x^n + 1) which is of degree (n – k) is called a generator polynomial of an (n, k) linear cyclic block code.

Note

(i) There may be several factors of degree (n – k) for the polynomial (x^n + 1). All such factors of degree (n – 1) are generator polynomials and can be used for constructing (n, k) linear cyclic codes. But all of these codes may not be equally good.

(ii) Suppose we write g(x) as follows:

$$g(x)=g_0+g_1x+g_2x^2+\ldots+g_{n-k}x^{n-k} \quad (13.48)$$

Then $g_0 = g_{n-k} = 1$. *This may be justified as follows:*
The coefficients g_0, g_1, g_2, …, g_{n-k} *are all binary digits and may be either 1 or 0. Suppose* $g_0 = 0$. *Then*

$$g(x) = x(g_1 + g_2x + \ldots + g_{n-k}x^{n-k}) = xg'(x)$$

Since g(x) is a factor of $(x^n + 1)$, *x must also be a factor of* $(x^n + 1)$ *but it cannot be. So* $g_0 \neq 0$ $\therefore$ $g_0 = 1$. *Similarly, if* $g_{n-k} = 0$, *the g(x) is not of degree (n – k). It is (n – k – 1) or less. So, it does not qualify to be a generator polynomial (as per our definition of a generator polynomial). g(x) is called a generator polynomial because it generates all the code polynomials of the (n, k) linear cyclic block code. If c(x) is a code polynomial, then*

$$c(x) = a(x)g(x) \tag{13.49}$$

Now, we shall see how, given a generator polynomial, we may find the code polynomial (in systematic form) corresponding to a given message sequence $(m_0, m_1, \ldots, m_{k-1})$. Since we want a systematic code, the structure of a codeword would be of the form:

$$\boldsymbol{C} = (\underbrace{b_0, b_1, \ldots, b_{n-k-1}}_{(n-k)\text{ parity bits}}, \underbrace{m_0, m_1, \ldots, m_{k-1}}_{k\text{ message bits}}) \tag{13.50}$$

So, let us first construct a message polynomial using the message sequence:

$$m(x) = m_0 + m_1x + \ldots + m_{k-1}x^{k-1} \tag{13.51}$$

We may also write a parity-bit polynomial $b(x)$ as

$$b(x) = b_0 + b_1x + \ldots + b_{n-k-1}x^{n-k-1} \tag{13.52}$$

So, using Eq. (13.50), we may write the code polynomial as

$$\begin{aligned} c(x) &= (b_0 + b_1x + b_2x^2 + \ldots + b_{n-k-1}x^{n-k-1}) + (m_0x^{n-k} + m_1x^{n-k+1} + \ldots + m_{k-1}x^{n-1}) \\ &= b(x) + x^{n-k}m(x) \end{aligned} \tag{13.53}$$

From Eq. (13.49), we have

$$c(x) = a(x)g(x) = b(x) + x^{n-k}m(x)$$

$$\therefore \quad x^{n-k}m(x) = a(x)g(x) + b(x) \quad \text{(modulo-2 arithmetic)}$$

$$\therefore \quad \frac{x^{n-k}m(x)}{g(x)} = a(x) + \frac{b(x)}{g(x)} \tag{13.54}$$

This equation indicates that when $x^{n-k}m(x)$ is divided by $g(x)$, the quotient is $a(x)$ and the remainder is $b(x)$. Hence, the steps involved in constructing an (n, k) linear cyclic code with systematic structure are:

1. From the message bits, form the message polynomial $m(x)$ and multiply it by x^{n-k}.
2. Divide $x^{n-k}m(x)$ by $g(x)$, the generator polynomial. Let the remainder be $b(x)$.
3. Add $b(x)$ to $x^{n-k}m(x)$ to obtain the code polynomial $c(x)$ corresponding to the message polynomial $m(x)$. $c(x)$ will be in systematic form.

Example 13.15 If $c(x)$ is a cyclic code polynomial, show that $xc(x)$ divided by $(x^n + 1)$ gives $c^{(1)}(x)$.

Solution Let $c(x) = c_0 + c_1x + c_2x^2 + \ldots + c_{n-1}x^{n-1}$

$$\therefore \quad xc(x) = c_0x + c_1x^2 + c_2x^3 + \ldots + c_{n-1}x^n$$

$$\begin{array}{r|l} & \quad c_{n-1} \\ \cline{2-2} (1 + x^n) & c_0x + c_1x^2 + \ldots + c_{n-1}x^n \\ & c_{n-1} \qquad\qquad\qquad + c_{n-1}x^n \\ \cline{2-2} & c_{n-1} + c_0x + c_1x^2 + \ldots + c_{n-2}x^{n-1} \ \text{ which is } c^{(1)}(x) \end{array}$$

∴ when $xc(x)$ is divided by $(x^n + 1)$, the remainder is $c^{(1)}(x)$ which is a polynomial obtained by cyclically shifting the code sequence corresponding to $c(x)$ to the right by one step.

Example 13.16 For a (7, 4) cyclic linear block code, show that there are two generator polynomials possible.

Solution Since it is a (7, 4) code, $n = 7$.

Hence, if $g(x)$ is a generator polynomial, then it must be a factor of $x^n + 1 = x^7 + 1$ and also, it must be of degree $(n - k) = 3$.

Now,
$$x^7 + 1 = (x + 1)(x^3 + x^2 + 1)(x^3 + x + 1)$$

Thus, both $(x^3 + x^2 + 1)$ and $(x^3 + x + 1)$ qualify to be generator polynomials as both of them are factors of $x^7 + 1$ and are of degree $(n - k) = 3$.

Example 13.17 Taking $x^3 + x + 1$ as the generator polynomial for the (7, 4) cyclic linear block code, determine the code vectors in systematic form for the following message sequences: (a) 1 0 1 1, (b) 1 1 1 1, and (c) 1 0 0 0.

Solution

(a) $\boldsymbol{m}$ = message sequence = 1 0 1 1

∴ message polynomial $m(x) = 1 + 0.x + 1.x^2 + 1.x^3$
$$= (1 + x^2 + x^3)$$

Further, since $n = 7$ and $k = 4$, $x^{n-k} = x^3$

∴ $x^{(n-k)}m(x) = x^3(1 + x^2 + x^3) = x^3 + x^5 + x^6$

Now, to find $b(x)$, we know from Eq. (13.54) that it is the remainder left when $x^{(n-k)}m(x)$ is divided by $g(x)$. $g(x)$ has been given to be $(x^3 + x + 1)$

$$\begin{array}{r|l}
 & x^3 + x^2 + x + 1 \\
\cline{2-2}
x^3 + x + 1 & x^6 + x^5 + x^3 \\
 & x^6 + x^4 + x^3 \\
\cline{2-2}
 & x^5 + x^4 \\
 & x^5 + x^3 + x^2 \\
\cline{2-2}
 & x^4 + x^3 + x^2 \\
 & x^4 + \quad + x^2 + x \\
\cline{2-2}
 & x^3 + x \\
 & x^3 + x + 1 \\
\cline{2-2}
 & \quad 1 = \text{remainder}
\end{array}$$

∴ $a(x) = 1 + x + x^2 + x^3$ and $b(x) = 1$

Now $c(x) = b(x) + x^{(n-k)}m(x) = 1 + x^3(1 + x^2 + x^3)$
$$= 1 + x^3 + x^5 + x^6 = \mathbf{1} + \mathbf{0}.x + \mathbf{0}.x^2 + \mathbf{1}.x^3 + \mathbf{0}.x^4 + \mathbf{1}.x^5 + \mathbf{1}.x^6$$

∴ code vector $\boldsymbol{C}$ = (1 0 0 1 0 1 1)

This is obviously in the systematic form.

(b) For $\boldsymbol{m} = 1\ 1\ 1\ 1,\ m(x) = 1 + x + x^2 + x^3$

$$x^{n-k} = x^3 \quad \therefore \quad x^{(n-k)}m(x) = x^3 + x^4 + x^5 + x^6$$

$g(x)$ is given to be $(x^3 + x + 1)$

To find $b(x)$, we divide $x^{(n-k)}m(x)$ by $g(x)$ and find the remainder

$$\begin{array}{r|l} & x^3 + x^2 + 1 \\ \hline x^3 + x + 1 & x^6 + x^5 + x^4 + x^3 \\ & \underline{x^6 \qquad + x^4 + x^3} \\ & x^5 \\ & \underline{x^5 + x^3 + x^2} \\ & x^3 + x^2 + 0 \\ & \underline{x^3 + x + 1} \end{array}$$

Remainder = $(x^2 + x + 1)$

$$\therefore \quad b(x) = (1 + x + x^2) = (1 + 1.x + 1.x^2)$$

$$\therefore \quad c(x) = b(x) + x^{(n-k)}m(x) = b(x) + x^3(1 + x + x^2 + x^3)$$

$$c(x) = \mathbf{1} + \mathbf{1}.x + \mathbf{1}.x^2 + \mathbf{1}.x^3 + \mathbf{1}.x^4 + \mathbf{1}.x^5 + \mathbf{1}.x^6$$

$$\boldsymbol{C} = (1, 1, 1, 1, 1, 1, 1)$$

(c) For $\boldsymbol{m} = 1\ 0\ 0\ 0,\ m(x) = 1$

$$\therefore \ x^{n-k} = x^3 \quad \text{and} \quad x^{(n-k)}m(x) = x^3$$

$$\begin{array}{r|l} & 1 \\ \hline x^3 + x + 1 & x^3 \\ & \underline{x^3 + x + 1} \\ & (x + 1) = \text{remainder} \end{array}$$

$$\therefore \ c(x) = \mathbf{1} + \mathbf{1}.x + \mathbf{0}.x^2 + \mathbf{1}.x^3 + \mathbf{0}.x^4 + \mathbf{0}.x^5 + \mathbf{0}.x^6$$

$$\therefore \ \boldsymbol{C} = (1\ 1\ 0\ 1\ 0\ 0\ 0)$$

which is in systematic form since the last four digits of $\boldsymbol{C}$ are the unaltered message digits.

13.5.13 Parity-Check Polynomial h(x)

Earlier, we had defined the generator polynomial of a cyclic (n, k) linear block code and stated that knowledge of this polynomial enables one to construct all the code polynomials of the code for all the 2^k possible message polynomials. Thus, the generator polynomial $g(x)$ is equivalent to the generator matrix, $\boldsymbol{G}$, of the block code.

Similarly, we can visualize a parity-check polynomial, $h(x)$ which is the equivalent of a parity-check matrix $\boldsymbol{H}$. For a linear block code, we know that its generator matrix, $\boldsymbol{G}$ and parity-check matrix $\boldsymbol{H}$ must satisfy the relation (Eq. (13.16))

$$\boldsymbol{G} \cdot \boldsymbol{H}^T = \mathbf{0}$$

Analogous to this, for the generator and parity-check polynomials, we may write the relation

$$g(x)h(x) \equiv 0 \quad \text{modulo} \quad (x^n + 1) \tag{13.55}$$

This relation implies that $s(x)$ and $h(x)$ are both factors of $(x^n + 1)$. Since $g(x)$ is of degree $(n - k)$, we may now define the parity-check polynomial of an (n, k) linear cyclic block code as:

Definition *A parity-check polynomial, h(x) of a cyclic (n, k) linear block code is a polynomial of degree k which is a factor of (x^n + 1).*

Although the coefficients of this polynomial $h(x)$

$$h(x) = h_0 + h_1 x + h_2 x^2 + \ldots + h_k x^k$$

are binary digits 0 and 1, just like in the case of $g(x)$, here too it is necessary that

$$h_0 = h_k = 1 \tag{13.56}$$

Example 13.18 Find $g(x)$ and $h(x)$, the generator polynomial and the parity-check polynomial of a systematic (7, 4) cyclic code. Determine ***G*** and ***H*** matrices of the code.

Solution Since $k = 4$ and $n = 7$, $g(x)$ must be a polynomial of degree 3 which is a factor of $(x^7 + 1)$. As we had already seen in Example 13.16,

$$x^7 + 1 = (x + 1)(x^3 + x^2 + 1)(x^3 + x + 1)$$

Both the polynomials $(x^3 + x^2 + 1)$ and $(x^3 + x + 1)$ qualify to be generator polynomials for this code. We shall choose

$$g(x) = x^3 + x + 1$$

In Example 13.17, we had determined the code polynomials in the systematic form for a few message sequences. Following the same procedure, we find out the code polynomials in systematic form, for the four message sequences of Hamming weight 1. The result is tabulated below.

Message sequence	Message polynomial	Code polynomial
0 0 0 1	1	$1 + x + x^3 = \mathbf{1} + \mathbf{1}.x + \mathbf{0}.x^2 + \mathbf{1}.x^3 + \mathbf{0}.x^4 + \mathbf{0}.x^5 + \mathbf{0}.x^6$
0 0 1 0	x^1	$x + x^2 + x^4 = \mathbf{0} + \mathbf{1}.x + \mathbf{1}.x^2 + \mathbf{0}.x^3 + \mathbf{1}.x^4 + \mathbf{0}.x^5 + \mathbf{0}.x^6$
0 1 0 0	x^2	$1 + x + x^2 + x^5 = \mathbf{1} + \mathbf{1}.x + \mathbf{1}.x^2 + \mathbf{0}.x^3 + \mathbf{0}.x^4 + \mathbf{1}.x^5 + \mathbf{0}.x^6$
1 0 0 0	x^3	$1 + x^2 + x^6 = \mathbf{1} + \mathbf{0}.x + \mathbf{1}.x^2 + \mathbf{0}.x^3 + \mathbf{0}.x^4 + \mathbf{0}.x^5 + \mathbf{1}.x^6$

The ***G***-matrix is of size $k \times n = 4 \times 7$ and we know that its rows are code vectors.

$$\therefore \qquad \boldsymbol{G} = \begin{bmatrix} 1 & 1 & 0 & \vdots & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & \vdots & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & \vdots & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & \vdots & 0 & 0 & 0 & 1 \end{bmatrix}_{k \times n}$$

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & \vdots & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & \vdots & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & \vdots & 0 & 1 & 1 & 1 \end{bmatrix}_{(n-k) \times n}$$

In fact, these are the same matrices that we had obtained as the ***G*** and ***H*** matrices of a (7, 4) Hamming code, except, of course, that the rows of the P-matrix are permuted. Thus, we find that Hamming codes are also cyclic codes. In fact, any cyclic code generated by a primitive polynomial is a Hamming code of minimum distance 3. ***An irreducible polynomial of degree m is said to be primitive if the smallest positive integer n for which the polynomial divides*** $x^n + 1$ ***is*** $2^m - 1$.

Both $(1 + x + x^3)$ and $(1 + x^2 + x^3)$ are irreducible polynomials which are primitive.

Now, to find $h(x)$, the parity check polynomial, we use the fact that

$$g(x)h(x) \quad \text{mod}\ (x^n+1)=0$$

$\therefore$ dividing (x^n+1) by the $g(x)$, which is x^3+x+1

$$\begin{array}{r|l}
 & x^4+x^2+x+1 \\ \hline
x^3+x+1 & x^7+1 \\
 & \underline{x^7+x^5+x^4} \\
 & x^5+x^4+1 \\
 & \underline{x^5+x^3+x^2} \\
 & x^4+x^3+x^2+1 \\
 & \underline{x^4+\quad+x^2+x} \\
 & x^3+x+1 \\
 & \underline{x^3+x+1} \\
 & 0
\end{array}$$

$\therefore$ $$h(x)=1+x+x^2+x^3$$

13.5.14 Calculation of Syndrome and Decoding of Cyclic Codes

Just as in the case of other block codes, for cyclic codes also syndrome calculation is the first step in decoding. If $c(x)$ is a code polynomial of the code and is transmitted and if the received polynomial is $r(x)$, then

$$r(x)=c(x)+e(x)$$

where $e(x)$ is the error polynomial corresponding to the error pattern created by the channel noise.

Let $$r(x)=r_0+r_1x+r_2x^2+\ldots+r_{n-1}x^{n-1} \qquad (13.57)$$

Divide $r(x)$ by the generator polynomial. Let $q(x)$ be the quotient and $s(x)$ be the remainder.

i.e., $$r(x)=q(x)\cdot g(x)+s(x) \qquad (13.58)$$

Since we have divided by $g(x)$ which is of degree $(n-k)$, the remainder, $s(x)$ will have to be of degree $(n-k-1)$ or less. $s(x)$ is called the ***syndrome polynomial***. It has the following interesting properties:

Property 1: If $s(x)$ is the syndrome of the received polynomial, then it is the syndrome of the error polynomial also.

Proof: Since $r(x)=c(x)+e(x)$

we may write $$e(x)=r(x)+c(x) \qquad (13.59)$$

we had seen that $$c(x)=a(x)g(x) \qquad \text{(See Eq. (13.49))}$$

Also, $$r(x)=q(x)\cdot g(x)+s(x) \qquad \text{(See Eq. (13.58))}$$

$\therefore$ substituting these in Eq. (13.59), we get

$$\begin{aligned} e(x)&=[a(x)+q(x)]g(x)+s(x) \\ &=u(x)\cdot g(x)+s(x) \qquad (13.60) \end{aligned}$$

where $$u(x)\ \underline{\Delta}\ a(x)+q(x)$$

Equation (13.60) shows the result of dividing $e(x)$ by $g(x)$ is a quotient of $a(x)$ and a remainder of $s(x)$. Hence, the syndrome is the same for $r(x)$ and $e(x)$.

Property 2: If $s(x)$ is the syndrome of $r(x)$, the received polynomial, then $x\,r(x)$ will have a syndrome of $x\,s(x)$.

Proof: Since $$r(x)=q(x)\cdot g(x)+s(x), \qquad \text{(See Eq. 13.58)}$$

$$xr(x)=xq(x)g(x)+xs(x) \qquad (13.61)$$

We know that multiplying $r(x)$ by x is equivalent to giving one right cyclic shift to the received word ***r***. Equation (13.61) implies that the remainder obtained after dividing $xr(x)$ by $g(x)$ is $xs(x)$.

Hence, the syndrome of $xr(x)$ is $xs(x)$. Instead of giving one right cyclic shift to $r(x)$, suppose we give i right cyclic shifts, then the resulting shifted received word is represented by $x^i r(x)$. So, we may generalize the above result and say that if $s(x)$ is the syndrome polynomial of the received polynomial $r(x)$, then $x^i s(x)$ will be the syndrome polynomial of $x^i r(x)$.

Property 3: If the errors occur only in the parity check bits of the transmitted codeword, the syndrome polynomial and the error pattern polynomial will be the same.

Proof: $r(x) = r_0 + r_1 x + r_2 x^2 + \ldots + r_{n-k-1} x^{n-k-1} + r_{n-k} x^{n-k} + r_{n-k+1} x^{n-k+1} + \ldots + r_{n-1} x^{n-1}$

In this $(r_0, r_1, r_2, \ldots, r_{n-k-1})$ are the received parity-check bits. By saying that the errors are confined only to the parity check bits, we are implying that in the error pattern polynomial:

$$e(x) = \left(e_0 + e_1 x + e_2 x^2 + \ldots + e_{n-k-1} x^{n-k-1}\right) + \left(e_{n-k} x^{n-k} + \ldots + e_{n-1} x^{n-1}\right)$$

coefficients e_{n-k} to e_{n-1} are all zero, i.e., $e(x)$ is of degree $(n-k-1)$ or less. Then from Eq. (13.60), which says that when $e(x)$ is divided by $g(x)$ which is of degree $(n-k)$, the quotient is $u(x)$ and the remainder is the syndrome $s(x)$. Thus, if $e(x)$ itself is of degree $(n-k-1)$ or less, since $g(x)$ is of degree $(n-k)$, $u(x)$ must be zero. Then from that equation, it follows that $s(x) = e(x)$.

The following example illustrates how errors may be corrected by calculating the syndrome and making use of some of the above properties of it.

Example 13.19 A channel encoder uses a (7, 4) linear cyclic block code in the systematic form, the generator polynomial being $(x^3 + x + 1)$. Determine the correct codeword transmitted, if the received word is (a) 1 0 1 1 0 1 1 (b) 1 1 0 1 1 1 1.

Solution In order to perform error correction using the syndrome, we shall first prepare a list of the syndromes for each of the seven possible single errors. In the first three error patterns listed, the errors are confined only to the parity check bits and so, using property 3 above, we have $s(x) = e(x)$. For the remaining four error patterns, the errors are not in the parity check bits – they are in the message bits. So, we use property 1, i.e., we divide $e(x)$ by $g(x)$ and take the remainder as the syndrome polynomial $s(x)$. We shall calculate these first.

e = error pattern = 0 0 0 1 0 0 0 $\quad\therefore\quad e(x) = 0.1 + 0.x + 0.x^2 + 1.x^3 + 0.x^4 + 0.x^5 + 0.x^6$

$\therefore\quad e(x) = x^3$

$$\begin{array}{r|l} & 1 \\ \hline x^3 + x + 1 & x^3 \\ & \underline{x^3 + x + 1} \\ & \text{Remainder} = (x+1) \end{array} \qquad \therefore s(x) = (1 + x)$$

e = error pattern = 0 0 0 0 1 0 0 $\quad\therefore\quad e(x) = 0.1 + 0.x + 0.x^2 + 0.x^3 + 1.x^4 + 0.x^5 + 0.x^6$

$$\begin{array}{r|l} & x \\ \hline x^3 + x + 1 & x^4 \\ & \underline{x^4 + x^2 + x} \\ & (x + x^2) = \text{Remainder} \end{array} \qquad \therefore s(x) = (x + x^2)$$

e = error pattern = 0 0 0 0 0 1 0 $\quad \therefore \quad e(x) = x^5$

$$\begin{array}{r|l} & x^2+1 \\ \cline{2-2} x^3+x+1 & x^5 \\ & \underline{x^5+x^3+x^2} \\ & x^3+x^2 \\ & \underline{x^3+x+1} \\ & (1+x+x^2) = \text{Remainder} \end{array} \qquad \therefore s(x) = (1+x+x^2)$$

e = error pattern = 0 0 0 0 0 0 1 $\quad \therefore \quad e(x) = x^6$

$$\begin{array}{r|l} & x^2+1 \\ \cline{2-2} x^3+x+1 & x^6 \\ & \underline{x^6+x^4+x^3} \\ & x^4+x^3 \\ & \underline{x^4+x^2+x} \\ & x^3+x^2+x \\ & \underline{x^3+x+1} \\ & x^2+1 = \text{Remainder} \end{array} \qquad \therefore s(x) = (1+0.x+x^2)$$

Error Pattern	**Error poly** ***e*(*x*)**	**Syndrome poly** ***s*(*x*)**	**Syndrome *S***
1 0 0 0 0 0 0	1	$1+0.x+0.x^2$	1 0 0
0 1 0 0 0 0 0	x	$0+1.x+0.x^2$	0 1 0
0 0 1 0 0 0 0	x^2	$0+0.x+1.x^2$	0 0 1
0 0 0 1 0 0 0	x^3	$1+1.x+0.x^2$	1 1 0
0 0 0 0 1 0 0	x^4	$0+1.x+1.x^2$	0 1 1
0 0 0 0 0 1 0	x^5	$1+1.x+1.x^2$	1 1 1
0 0 0 0 0 0 1	x^6	$1+0.x+1.x^2$	1 0 1

(a) Received word = 1 0 1 1 0 1 1 $\quad \therefore \quad r(x) = 1 + x^2 + x^3 + x^5 + x^6$

Syndrome polynomial is obtained as the remainder when we divide $r(x)$ by the generator polynomial $g(x)$, which is given to be $x^3 + x + 1$

$$\begin{array}{r|l} & x^3+x^2+x+1 \\ \cline{2-2} x^3+x+1 & x^6+x^5+x^3+x^2+1 \\ & \underline{x^6+x^4+x^3} \\ & x^5+x^4+x^2+1 \\ & \underline{x^5+x^3+x^2} \\ & x^4+x^3+1 \\ & \underline{x^4+x^2+x} \\ & x^3+x^2+x+1 \\ & \underline{x^3+x+1} \\ & \quad x^2 = \text{remainder} \end{array}$$

$\therefore \quad s(x) = x^2 = 0.1 + 0.x + 1.x^2$

$\therefore \quad S = (0, 0, 1)$

From the table, $e = (0, 0, 1)$ for $S = (0, 0, 1)$

$$\therefore \quad \boldsymbol{C} = r + e = \begin{array}{r} 1\,0\,1\,1\,0\,1\,1 \\ +\,0\,1\,1 \\ \hline 1\,0\,0\,1\,0\,1\,1 \end{array}$$

∴ Transmitted code vector ***C*** is

$\boldsymbol{C} = (1\ 0\ 0\ 1\ 0\ 1\ 1)$

(b) Received word = 1 1 0 1 1 1 1 $\quad \therefore \quad r(x) = 1 + x + x^3 + x^4 + x^5 + x^6$

$$\begin{array}{r|l} & x^3 + x^2 + 1 \\ \cline{2-2} x^3 + x + 1 & x^6 + x^5 + x^4 + x^3 + x + 1 \\ & \underline{x^6 + x^4 + x^3} \\ & x^5 + x + 1 \\ & \underline{x^5 + x^3 + x^2} \\ & x^3 + x^2 + x + 1 \\ & \underline{x^3 + x + 1} \\ & \quad x^2 = \text{remainder} \end{array}$$

$\therefore \quad s(x) = x^2 = 0.1 + 0.x + 1.x^2$

$\therefore \quad \boldsymbol{S} = (0, 0, 1)$

From the table, the corresponding e is

$\therefore \quad e = (0, 0, 1)$

$\therefore \quad \boldsymbol{C} = r + e = (1, 1, 1, 1, 1, 1, 1)$

13.5.15 Encoders and Decoders for Cyclic Codes

The function of the encoder is to give the codeword in systematic form when a message sequence is fed to it. For this, it needs to perform the operation implied in Eq. (13.54) in order to give out the parity bits. As we have already observed, multiplying a polynomial, say $m(x)$ by $x^{(n-k)}$ can be obtained by just subjecting $m(x)$ to $(n–k)$ cyclic shifts to the right. The polynomial division operation needed to be performed can also be implemented very easily and effectively by the shift register encoder shown in Fig.13.6.

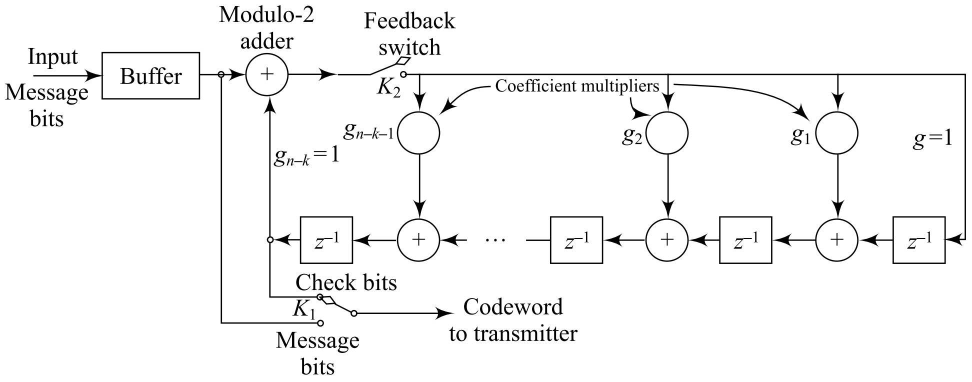


Fig. 13.6 *A shift register encoder for a cyclic (n, k) code*

As already stated earlier, the coefficients g_0 and g_{n-k} of the generator polynomial have to be equal to 1, only a solid connection is shown in both the cases. If any g_i is zero, only an open circuit will be shown.

The shift register contents are initialized to zero. Switch K_1 is kept in the message bits position and the feedback switch K_2 is kept closed. An external clock is used to shift the contents of the shift registers in the direction indicated by the arrowhead. The k message bits are first shifted into the shift register. Since K_1 switch is in the message bits position, the message bits are also passed on to the transmitter simultaneously. After the k shifts, the registers contain the $(n–k)$ parity check bits. At this time, the feedback switch K_2 is opened and the switch K_1 is thrown to the check bits position, making the check bits available to the transmitter. Since the parity bits are preceded by the k unaltered message bits, the arrangement shown in Fig. 13.6 gives the codewords in the systematic form.

For calculating the syndrome, we have to divide the received polynomial $r(x)$ by the generator polynomial $g(x)$. Syndrome is obtained as the remainder resulting from the division operation. So an arrangement as shown in Fig. 13.7 is used for implementing this division.

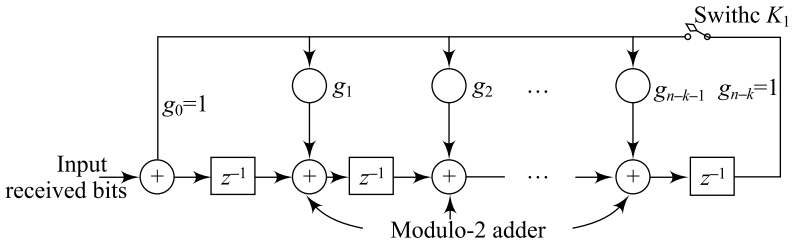


Fig. 13.7 *Syndrome calculator for an (n, k) cyclic code*

With switch K_1 closed, all the n bits of the received word are shifted into the registers. As soon as all the bits have been shifted, the syndrome is available as the contents of the shift registers.

Example 13.20 A systematic (7, 4) cyclic code is to be generated making use of $(x^3 + x + 1)$ as the generator polynomial. Draw the block schematic diagram of the encoder. By tracing the contents of the shift registers after shifting-in of each of the message bits, determine the parity check bits and the codeword in systematic form, assuming a message sequence $\boldsymbol{m}$ = (1, 1, 1, 1).

Solution Since the $g(x) = 1 + x + x^3$, we have $g_0 = 1$, $g_1 = 1$, $g_2 = 0$ and $g_3 = 1$. So, the encoder shown in Fig. 13.6 takes the following form as shown in Fig. 13.8.

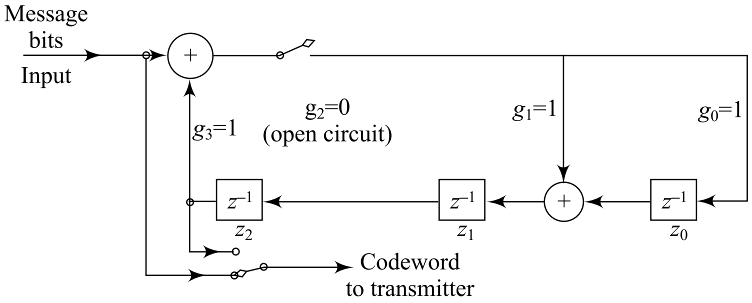


Fig. 13.8 *Encoder for the (7, 4) cyclic code with g(x) = 1 + x + x^3*

The following table shows the contents z_0, z_1 and z_2 of the three shift registers as the shifting of the input message bits progresses.

Input *m*	z_2	z_1	z_0	$z_2' = z_1$	$z_1' = z_0 \oplus z_2 \oplus m$	$z_0' = z_2 \oplus m$
$m_0 = 1$	0	0	0	0	1	1
$m_1 = 1$	0	1	1	1	0	1
$m_2 = 1$	1	0	1	0	1	0
$m_3 = 1$	0	1	0	1	1	1

$\leftarrow b \rightarrow \quad \leftarrow m \rightarrow$

Hence, the codeword is $\boldsymbol{C}$ = (1 1 1 1 1 1 1).

This is in agreement with the result that we got for part (b) of Example 13.13.

13.5.16 Bose–Chaudhary–Hocquenghem Codes

Popularly known as BCH codes, these form a subset of the larger class of cyclic codes. They were invented by Hocquenghem in 1959 and independently by Bose and Chaudhary in 1960. In fact, the single-error correcting Hamming codes can also be considered as BCH codes. The general class of BCH codes is indeed a remarkable generalization of the Hamming codes, for multiple error correction. In 1961 Gorenstein and Zeirler developed non-binary BCH codes as a generalization of the BCH codes.

A detailed study of the theory and the method of construction of BCH codes is beyond the scope of this book. We will, however, briefly describe one important subclass of the binary BCH codes, called the *primitive* BCH codes, which have the capability to detect and correct up to t random type of errors per codeword.

A t-error correcting binary BCH code exists with the following parameters for any pair of positive integers m and t, where $m \geq 3$ and $t < 2^{m-1}$.

Block length $n = 2^m - 1$

Number of parity check digits = $(n - k) \leq mt$

Minimum distance of the code: $d_{\min} \geq 2t + 1$

Besides their multiple error correction capability, an attractive feature of BCH codes is the flexibility in the choice of block length and code rate. For a given code rate and a given block length, BCH code are the best known codes. For decoding of BCH codes, Berlekamp's iterative algorithm and Chein's search algorithm are very efficient.

Reed-Solomon codes, a type of non-binary BCH codes, are being extensively used for forward error correction (FEC). They are being used in digital storage systems, modems, etc. For detailed discussion of BCH codes and Reed_Solomon codes, the interested readers may refer to the textbook by Shu Lin and Costello.

13.5.17 Burst-Error Detection and Correction

Till now, we have considered a few important types of error detecting and correcting block codes for random errors, i.e., errors that occur randomly, generally affecting one or, in some cases, a few bits at random locations in a codeword.

But errors can occur in clusters too, i.e., several successive bits of a codeword may be affected. Such errors may be caused in wireless transmission systems by lightning, or deep fading lasting for several bit intervals, and in the case of storage systems, by defects in the magnetic materials. For error detection and correction in such cases, ordinary random error codes are inefficient. A special class of codes, called burst error detecting and correcting codes have to be used.

In connection with codes for burst errors, there are two useful results.

1. For *detecting* burst errors of length p or less digits using a block code, it has been shown that it is necessary and sufficient for the block code to have p parity check digits.
2. For *correcting* burst errors of length p or less, a linear block code should have at least $2p$ parity-check digits.

An interesting thing about the two results quoted above, is that the number parity-check bits needed depends only on the burst length but not on the message sequence length, k, or the code length, n. Because of this, block codes for burst-error detection find application in packet switching where, the packet length may vary from packet to packet.

13.5.18 Coding Gain of Block Codes

As stated earlier, we resort to error-correction through channel coding to reduce the probability of error, P_e. But, as we have already seen, for an uncoded system with a given channel, P_e depends upon the average bit energy, which, for a given average power, depends upon the bit-rate. For the same channel, in the case of coded systems, it depends on the type of code too, in addition to the average power and bit rate.

Suppose we consider an (n, k) block code. k information bits are mapped into an n-bit codeword, where, $n > k$. If the bit rate of the uncoded system, which we may call as the information rate, is say, r, then the bit rate for the coded system will be $r.(n/k)$ is we assume that the time taken to transmit k information bits is the same in both the systems. Thus, the coded system has a higher bit-rate than the uncoded system and so it requires a bandwidth that is (n/k) times the bandwidth required for the uncoded system. Thus, the coded system scarifies bandwidth in order to give better reliability.

Now, for the comparison to be fair, let us say the energy utilized by the coded system to transmit k information bits (using n coded bits for this purpose), is the same as the energy utilized by the uncoded system to transmit the k information bits. Then, E_b, the bit energy of the coded system and so, it is less. Since P_e decreases with E_b, whatever may be the code, the question naturally arises: 'Can the coded system with a smaller bit energy give more reliability than the uncoded system?

The answer to the above question is: 'Yes, to some extent, if a certain condition is satisfied'. The condition to be satisfied is that if the code used is a t-error correcting code, the actual number of errors occurring per codeword should *not* be more that t. If this condition is not satisfied, the coded system can have a P_e that is worse than that of an uncoded system. As long as this condition is satisfied, there will be some reduction in the bit-error rate (BER). To what extent the BER will be reduced, will, of course, depend on the type of code used.

From the above discussion, two points emerge:

1. For a given channel, a coded system requires a smaller value of bit energy E_b than an uncoded system to give a specified P_e at the same information rate.
2. For a given channel and a given information rate, the extent of reduction possible in the bit energy when an error-correcting code is used, will depend on the type of code used.

The comparison between a coded and an uncoded system, as well as that between different codes, can be made in quantitative terms by making use of a parameter called the '**coding gain**', which may be defined as follows.

Definition Coding gain of a code is the ratio of the bit energies required by an uncoded system and a coded system for achieving a specified P_e, with the information rate being the same for both. (It is generally expressed in dB.)

Note *The term, 'information rate', used above, refer to the number of message bits, or information bits transmitted per unit time.*

It has been shown that the coding gain of an (n, k) block code is equal to the product of the code rate r $(= k/n)$ and the value of the minimum Hamming distance of the code (reference 5). Since r is always less than 1 and $d_{\min}$ is greater than or equal to one, the coding gain, as a ratio, can be either greater than, or less than 1. However, there are many block codes which provide good coding gains (i.e., above 1). For given values of n and k, it then follows that the (n, k) code with the largest value of its $d_{\min}$ will give the highest coding gain.

13.6 CONVOLUTIONAL CODES

The encoder for a block code, as we have seen, takes a k-bit message block, produces $(n-k)$ parity bits each of which is obtained according to some predetermined linear combination of the k message bits and produces an encoded sequence of n bits, where $n > k$, by appending the $(n-k)$ parity bits to the k message bits. Thus, the codewords are produced on a block-by-block basis. The encoder for a block code therefore needs a buffer for storing the k message bits. An encoder for a convolutional code, on the other hand, acts on the message bits coming serially. It generates the output codeword by modulo-2 discrete convolution of its own impulse response (finite duration) with the sequence formed by the present message bit and a few message bits preceding the present one. The encoder for convolutional encoding is quite simple – a tapped shift register. Further, convolutional codes offer much better performance by sophisticated decoding techniques. Most of the block codes have rates k/n above 0.95. On the other hand, most of the convolutional codes have rates below 0.90. *However, their low code rate is more than compensated by their very powerful error-correcting capabilities*. Convolutional codes are generally preferred in space and satellite communication systems that require simple encoders which achieve high performance and for use in very noisy channels.

13.6.1 Convolutional Encoders

A convolutional encoder in its simplest form is a finite state machine and consists of a tapped shift register with say $(L+1)$ stages whose outputs are connected to a modulo-2 adder through coefficient multipliers, as shown in Fig. 13.9. Since the message bits as well as the coefficients $g_0, g_1, \ldots, g_L$ are all binary numbers, the arrangement is just a binary FIR filter whose finite impulse response is given by the coefficients $g_0, g_1, \ldots, g_L$.

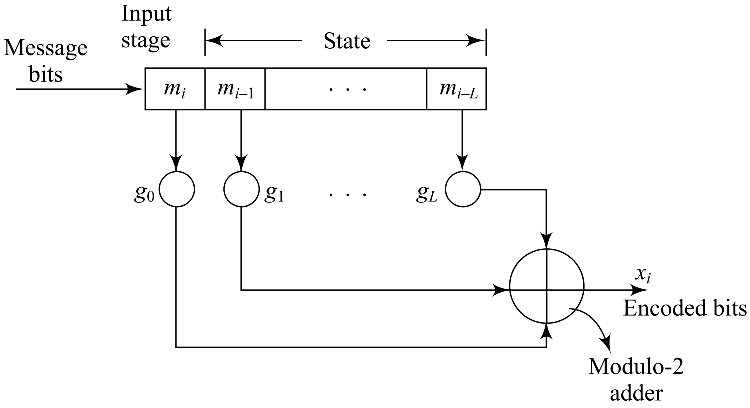


Fig. 13.9 *A shift register convolutional encoder*

Since $g_0, g_1, \ldots, g_L$ are either 0 or 1, if any g_i is a one, only a simple direct connection is shown, and if it is a 0, then the corresponding connection is not shown at all, as it is an open circuit. The first stage to which the input is given is referred to as the input stage and the remaining L stages of the shift register define the state of the encoder. With the contents of the shift register stages as marked in the figure, the output bit is given by

$$\begin{aligned} x_i &= m_i \cdot g_0 \oplus m_{i-1} \cdot g_1 \oplus \ldots + m_{i-L} \cdot g_L \\ &= \sum_{j=0}^{L} m_{i-j} g_j \quad \text{(modulo-2 addition)} \end{aligned} \tag{13.62}$$

The reader might have identified the RHS of the above equation as the familiar 'convolutional sum' (similar to the convolutional integral for continuous-time signals) that gives the i^{th} element of the output sequence in the case of a discrete-time filter. That is why these encoders are called *convolutional encoders* and the code, as a *convolutional code*.

Whenever a message bit is shifted into the input stage of the shift register, one output bit, as determined by Eq. (13.62) is given out. Thus, for each input message bit only one output bit is obtained from the encoder. Hence, there is no redundancy introduced and so no error-detecting or correcting capability exists for the code produced by this simple encoder. Before we proceed to describe a more practical form of a convolutional encoder, there is one important point that deserves mention. That is, that each message bit shifted into the input stage of the shift register has to go through all the $(L+1)$ stages: and as long as it stays in one or the other of these $(L+1)$ stages, it influences the output bit. *Thus, each message bit influences (L + 1) successive encoded bits coming out from the encoder*.

A more practical form of a convolutional encoder is shown in Fig. 13.10. As pointed out earlier, for the generated code to have error-correcting capability, it is necessary that each input bit should give rise to more than one bit of the encoded bit stream. This is achieved by using two or more modulo-2 summers and interleaving the output bits from them by using a commutator switch as shown in the figure.

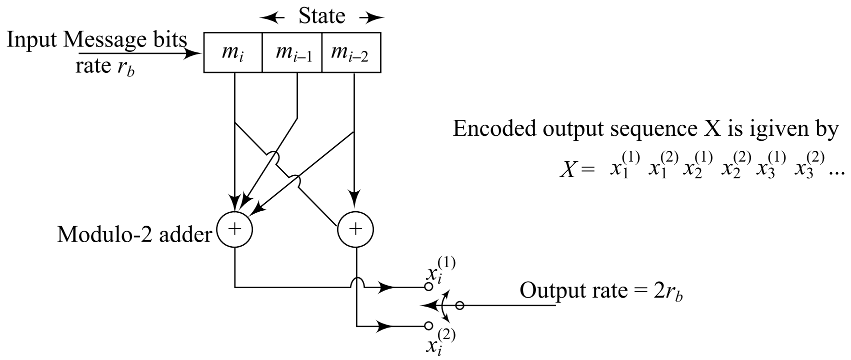


Fig. 13.10 *A convolutional encoder with n = 2, k = 1 and L = 2*

Since each message bit shifted into the input stage produces one encoded output bit at each of the two modulo-2 adders, and the output commutator switch collects them as one pair of ordered output bits, $k = 1$ and $n = 2$. Since there are 3 shift-register stages, $L + 1 = 3$ or $L = 2$. The code produced by a practical convolutional encoder is generally referred to as an (n, k, L) code. The encoder of Fig. 13.10 thus produces a (2, 1, 2) convolutional code. Whereas in the encoder of Fig. 13.9, each message bit was influencing $(L+1)$ successive encoded bits, in the encoder of Fig. 13.10, each message bit influences $n(L+1)$ encoded bits, which is 6 in this case. This quantity, $n(L+1)$ is referred to as the constraint length.

Definition The constraint length of a convolutional code is defined as the number of encoder output bits influenced by each message bit.

$$\text{Constraint length of } (n, k, L) \text{ code} = n(L+1) \text{ bits} \tag{13.63}$$

Note *The definition of constraint length is not consistent in the literature. Some authors define it as simply the number of memory elements in the shift register. In that case, it is $(L+1)$. So, one must be careful and check up the way an author has defined the term. In this book, we adhere to the above definition and use the expression given for it in Eq. (13.63).*

Since each time a message bit is shifted into the shift register, n-encoded output bits are produced, the code rate of the (n, k, L) code is said to be $1/n$. But, in fact, there is an oversimplification in the way we arrived at this value of the code rate. As we are going to see presently, this value of code rate is correct only when the length of the message sequence is very large compared to $(L+1)$, the number of shift register elements. Suppose we have a message sequence of length K bits. Then it is only after K shifts that the last message bit would enter into the input stage of the shift register. After that it will take another L shifts to make this last message bit to come out of the shift register. So, a message sequence of K bits length needs altogether $(K + L)$ shifts for the last message bit also to leave the shift register. We know that for an (n, k, L) code, the convolutional encoder has n number of modulo-2 adders, and that each time a shifting is done, n encoded bits are produced (one at each adder). Hence, the output sequence has a length of $n(K + L)$. *Thus, a K length message sequence produces* $n(K + L)$ *output encoded bits*. So, the code rate r is given by

$$\text{Code rate } r = \frac{K}{n(K+L)} \tag{13.64}$$

If $K >> L$, then the code rate is

$$r \cong \frac{1}{n} \quad (\text{if } K >> L) \tag{13.65}$$

Time-domain approach Equation (13.62), which gives the convolution summation, forms the basis for the time-domain analysis of a convolutional encoder. We shall now use this approach in the following example to demonstrate the two output sequences $x_i^{(1)}$ and $x_i^{(2)}$ of the encoder shown in Fig.13.10.

Example 13.21 Determine the output sequences $x_i^{(1)}$ and $x_i^{(2)}$ and the encoded output sequence X of the encoder of Fig.13.10, given that the message sequence is 1 0 1 0 0 1.

Solution $m_0 = 1, m_1 = 0, m_2 = 1, m_3 = 0, m_4 = 0, m_5 = 1$ and from Fig.13.10,

$$g_0^{(1)} = 1, \quad g_1^{(1)} = 1 \quad \text{and} \quad g_2^{(1)} = 1$$

From Eq. (13.62), we have: $x_i^{(1)} = \sum_{j=0}^{2} m_{i-j} g_i^{(1)}; \quad i = 0, 1, \ldots$

$\therefore$

$$x_0^{(1)} = m_{0-0} g_0^{(1)} \oplus m_{0-1} g_1^{(1)} = m_0 g_0^{(1)} = 1.1 = 1$$

$$x_1^{(1)} = m_1 g_0^{(1)} \oplus m_0 g_1^{(1)} = 0.1 \oplus 1.1 = 1$$

$$x_2^{(1)} = m_2 g_0^{(1)} \oplus m_1 g_1^{(1)} \oplus m_0 g_2^{(1)} = 1.1 \oplus 0.1 \oplus 1.1 = 1 \oplus 1 = 0$$

$$x_3^{(1)} = m_3 g_0^{(1)} \oplus m_2 g_1^{(1)} \oplus m_1 g_2^{(1)} = 0.1 \oplus 1.1 \oplus 0.1 = 1$$

$$x_4^{(1)} = m_4 g_0^{(1)} \oplus m_3 g_1^{(1)} \oplus m_2 g_2^{(1)} = 0.1 \oplus 0.1 \oplus 1.1 = 1$$

$$x_5^{(1)} = m_5 g_0^{(1)} \oplus m_4 g_1^{(1)} \oplus m_3 g_2^{(1)} = 1.1 \oplus 0.1 \oplus 0.1 = 1$$

$$x_6^{(1)} = m_5 g_1^{(1)} \oplus m_4 g_2^{(1)} = 1.1 \oplus 0.1 = 1$$

$$x_7^{(1)} = m_5 g_2^{(1)} = 1.1 = 1$$

$\therefore$ The sequence $x_i^{(1)}$ = **1 1 0 1 1 1 1 1**

To find the other sequence, i.e., $x_i^{(2)}$, we have

$g_0^{(2)} = 1 \quad g_1^{(2)} = 0 \quad g_2^{(2)} = 1$ and again the message digits are

$$m_0 = 1 \quad m_1 = 0 \quad m_2 = 1 \quad m_3 = 0 \quad m_4 = 0 \quad m_5 = 1$$

Again, making use of Eq. (13.61),

$$x_0^{(2)} = m_0 g_0^{(2)} = 1$$

$$x_1^{(2)} = m_1 g_0^{(2)} \oplus m_0 g_1^{(2)} = 0.1 \oplus 1.0 = 0$$

$$x_2^{(2)} = m_2 g_0^{(2)} \oplus m_1 g_1^{(2)} \oplus m_0 g_2^{(2)} = 1.1 \oplus 1.0 \oplus 1.1 = 1 \oplus 1 = 0$$

$$x_3^{(2)} = m_3 g_0^{(2)} \oplus m_1 g_2^{(2)} \oplus m_2 g_1^{(2)} = 0.1 \oplus 0.1 \oplus 1.0 = 0$$

$$x_4^{(2)} = m_4 g_0^{(2)} \oplus m_3 g_1^{(2)} \oplus m_2 g_2^{(2)} = 0.1 \oplus 0.0 \oplus 1.1 = 1$$

$$x_5^{(2)} = m_5 g_0^{(2)} \oplus m_4 g_1^{(2)} \oplus m_3 g_2^{(2)} = 1.1 \oplus 0.0 \oplus 0.1 = 1$$

$$x_6^{(2)} = m_5 g_1^{(2)} \oplus m_4 g_2^{(2)} = 1.0 \oplus 0.1 = 0$$

$$x_7^{(2)} = m_5 g_2^{(2)} = 1.1 = 1$$

Hence, $x_i^{(2)}$ = **1 0 0 0 1 1 0 1**

As shown in Fig.13.10, the encoded output sequence x is obtained by interleaving $x_i^{(1)}$ and $x_i^{(2)}$ sequences. Thus,

x = (**1 1, 1 0, 0 0, 1 0, 1 1, 1 1, 1 0, 1 1**)

It is clear from the above example that the time-domain approach is quite cumbersome and time consuming. We will now outline the transform domain approach for the analysis of convolutional encoders and then use that approach for solving the same problem.

Transform approach to convolutional encoder analysis As we have seen, a convolutional encoder performs a discrete linear convolution of two binary sequences – the input sequence and the encoder's own impulse response. Since a linear convolution of two sequences is much easier to determine in the transform domain rather than in the time domain, we now define the modulo-2 D-transform of a binary sequence m_k as

$$M(D) = \ldots \oplus m_{-1}D^{-1} \oplus m_0 D^0 \oplus m_1 D^1 \oplus m_2 D^2 \oplus \ldots \tag{13.66}$$

whereas usual, ⊕ denotes modulo-2 addition. As can easily be seen, it is just like the *Z*-transform, with the differences that *D* is used instead of Z^{-1} and that the additions are modulo-2 additions.

Hence, if two binary sequences, m_k and g_k are linearly convolved and if a binary sequence, *x*, results, i.e., if

$$x = m_k * g_k \tag{13.67}$$

then in the *D*-transform domain, we can write

$$X(D) = M(D)G(D) \tag{13.68}$$

We may now define the transfer function *h*(*D*) of a convolutional encoder as the *D*-transform of its impulse response sequence, h_k. *h*(*D*) is called the generator polynomial of the encoder. The notation used for this is not in tune with what we have been using. This is because, we have reserved the notation *H*(*D*) for the parity-check matrix of the code.

A convolutional encoder may be described making use of a generator matrix. The entries in this matrix are not 0s and 1s instead, they are polynomials in *D* with binary coefficients. They are, in fact, the transfer functions. The number of rows is equal to the number of inputs for the encoder and the number of columns is equal to the number of outputs. As in the case of block codes, here too, this generator matrix gives us the output if we know the inputs. We shall now illustrate the concept of a generator matrix for a convolutional coder by using the following examples.

Example 13.22 Find the generator matrix ***G***(*D*) for the (2, 1, 2) convolutional encoder of Fig. 13.10.

Solution This encoder has one input and two outputs. So ***G***(*D*) will be a 1 × 2 matrix. The first column entry in this will be the transfer function pertaining to the input and the first output, i.e., between m_k and $x_i^{(1)}$. This is equal to the *D*-transform of the impulse response, $h_k^{(1)}$, for the first output. But, we know that

$$h_k^{(1)} = \left(g_0^{(1)}\; g_1^{(1)}\; g_2^{(1)}\right) = (1\;\; 1\;\; 1)$$

∴ taking its *D*-transform, we have

$$h^{(1)}(D) = (1 \oplus D \oplus D^2)$$

Proceeding in a similar fashion, with respect to the second output,

$$h_k^{(2)} = \left(g_0^{(2)}\; g_1^{(2)}\; g_2^{(2)}\right) = (1\;\; 0\;\; 1)$$

$$\therefore \qquad h^{(2)}(D) = (1 \oplus D^2)$$

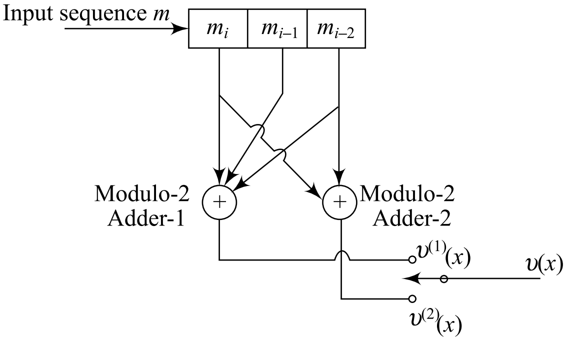


Fig. 13.11 *(a) A (2, 1, 2) encoder*

Hence, $$G(D) = [1 \oplus D \oplus D^2, \quad 1 \oplus D^2]$$

$\therefore$ if the D-transform of the outputs is represented by the vector $C(D)$, we have

$$\begin{aligned} C(D) &= M(D)[1 \oplus D \oplus D^2, \quad 1 \oplus D^2] \\ &= [M(D) \cdot (1 \oplus D \oplus D^2), \quad M(D) \cdot (1 \oplus D^2)] \end{aligned}$$

Example 13.23 Determine the output sequence $x_i^{(1)}$ and $x_i^{(2)}$ and the interleaved output x for the convolutional encoder of Fig. 13.10 assuming an input sequence m = 1 0 1 0 0 1. Use transform-domain approach. What is the code rate achieved?

Solution The time-domain convolutional operation described by Eq. (13.60) becomes a product operation in the *D*-transform domain.

$$X_i^{(1)}(D) = M(D) \cdot G^{(1)}(D)$$

where $X_i^{(1)}(D)$ is the D-transform of the output sequence at the first output, $M(D)$ is the transform of the input sequence, m and $G^{(1)}(D)$ is the transform of the impulse response $g^{(1)}$ pertaining to the input and first output of the encoder.

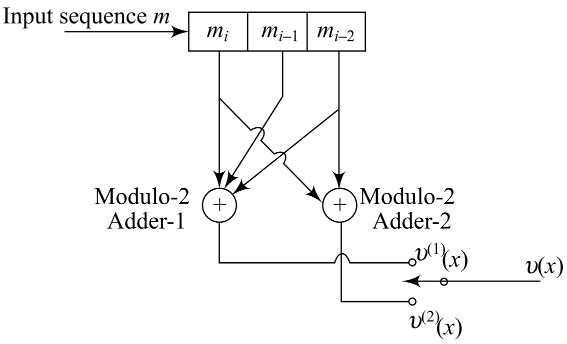


Fig. 13.11 *(b) A(2, 1, 2) encode*

$$M(D) = (1 \oplus D^2 \oplus D^5)$$

and $G^{(1)}(D) = D\text{-transform of } [1\ 1\ 1] = (1 \oplus D \oplus D^2)$

$$\begin{aligned} \therefore \quad X_i^{(1)}(D) &= (1 \oplus D^2 \oplus D^5)(1 \oplus D \oplus D^2) \\ &= 1 \oplus D^2 \oplus D^5 \oplus D \oplus D^3 \oplus D^6 \oplus D^2 \oplus D^4 \oplus D^7 \\ &= 1 \oplus D \oplus D^3 \oplus D^4 \oplus D^5 \oplus D^6 \oplus D^7 \end{aligned}$$

$\therefore \quad x_i^{(1)} = \mathbf{1\,1\,0\,1\,1\,1\,1\,1}$

Similarly, $X_i^{(2)}(D) = M(D) \cdot G^{(2)}(D)$

where $$G^{(2)}(D) = D\text{-transform of } \left[g_0^{(2)}\, g_1^{(2)}\, g_2^{(2)}\right]$$

$$= D\text{-transform of } [1\ 0\ 1] = (1 \oplus D^2)$$

$$\begin{aligned} \therefore \quad X_i^{(2)}(D) &= (1 \oplus D^2 \oplus D^5)(1 \oplus D^2) \\ &= 1 \oplus D^2 \oplus D^5 \oplus D^2 \oplus D^4 \oplus D^7 \end{aligned}$$

$$\therefore \quad X_i^{(2)}(D) = 1 \oplus D^4 \oplus D^5 \oplus D^7$$

Thus, $x_i^{(2)} = \mathbf{1\,0\,0\,0\,1\,1\,0\,1}$

Interleaving $x_i^{(1)}$ and $x_i^{(2)}$, we get the overall output as

$$x = (\mathbf{1\,1, 1\,0, 0\,0, 1\,0, 1\,1, 1\,1, 1\,0, 1\,1})$$

Note that exactly the same result was obtained earlier when we used the time-domain approach.

$$\text{Code rate achieved} = \frac{\text{No. of digits in } m}{\text{No. of digits in } x} = \frac{6}{16} = \frac{3}{8}.$$

Example 13.24 For the two-input, three-output convolutional encoder shown, determine the generator matrix $\boldsymbol{G}(D)$ and the parity-check matrix $\boldsymbol{H}(D)$. Show that $\boldsymbol{G}(D)\,\boldsymbol{H}'(D) = \mathbf{0}$.

Solution If $X^{(1)}(D), X^{(2)}(D)$ and $X^{(3)}(D)$ are the D-transforms of the three outputs $x^{(1)}, x^{(2)}$ and $x^{(3)}$, then from the figure, we find

$$X^{(1)}(D) = M^{(1)}(D).1$$

$$X^{(2)}(D) = M^{(2)}(D).1$$

and $$X^{(3)}(D) = (1 \oplus D)M^{(1)}(D) + DM^{(2)}(D)$$

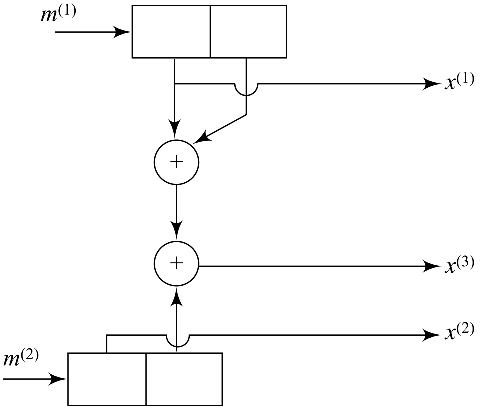


Fig. 13.12 *Encoder of Example 13.24*

We know that in the generator matrix, $\boldsymbol{G}(D)$, the rows correspond to the inputs and the columns to the outputs. We also know that each entry of $\boldsymbol{G}(D)$ is a transfer function between the input corresponding to the row in which it is located and the output corresponding to the column in which it is located. Hence, using the above three equations, we may write down

$$\boldsymbol{G}(D) = \begin{bmatrix} 1 & 0 & 1\oplus D \\ 0 & 1 & D \end{bmatrix}$$

This is in the form $\boldsymbol{G}(D) = [I_{2\times 2} \vdots P_{2\times 1}]$. Hence, by comparing this with Eq. (13.10a) which gives the structure of the generator matrix for a block code, we may write $\boldsymbol{H}(D)$, the parity-check matrix corresponding to the above $\boldsymbol{G}(D)$ as

$$\boldsymbol{H}(D) = [1\oplus D \quad D \vdots I_{1\times 1}]$$

$$= [1\oplus D \quad D \quad 1]$$

$$\boldsymbol{G}(D)\boldsymbol{H}^T(D) = \begin{bmatrix} 1 & 0 & 1\oplus D \\ 0 & 1 & D \end{bmatrix} \begin{bmatrix} 1\oplus D \\ D \\ 1 \end{bmatrix} = \begin{bmatrix} (1\oplus D)(1\oplus D) \\ D\oplus D \end{bmatrix} = \boldsymbol{0}$$

13.7 GRAPHICAL REPRESENTATION OF CONVOLUTIONAL CODES

Convolutional code structure is generally presented in graphical form using any one of the following three equivalent ways:

1. By drawing the code tree
2. By drawing the trellis for the code
3. By means of the state diagram

We will now discuss each one of these in detail, using the convolutional code generated by the encoder of Fig. 13.10 for illustration.

13.7.1 Code Tree

In the code tree, the dark black dots are nodes, and they represent the state of the encoder. From each node, there are two branches emanating – one upwards and the other downwards. When the encoder is in a particular state, from the node representing that state, we move upwards if the input (m_i) is 0 and downwards if the input (m_i) is a 1. Initially, the contents of all the shift-register elements are set to zero. As we go on giving the input bits serially, starting at the left-most node, the path takes the upper branch if the input bit m_i is zero. The corresponding output 0 0 is marked on that branch and the next state is represented by the node on which the branch terminates. When the next message digit is shifted into the input element of the shift register, it will move up or down from that node, depending on whether the input bit (m_i), now shifted into the shift register is a 0 or 1, respectively, and this process continues. The encoder has a memory of $(L + 1) = 3$ bits. Hence, when the fourth message bit is shifted into the register, the first message bit is shifted out of the

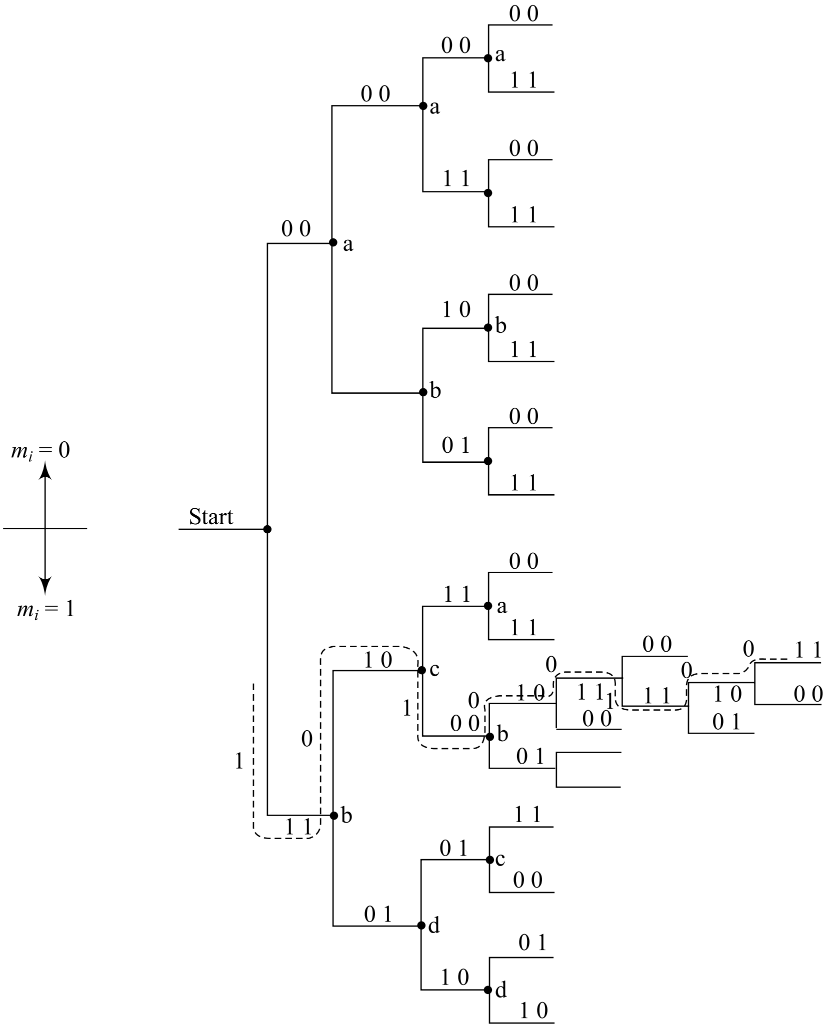


Fig. 13.13 *Code tree for the convolutional encoder of Fig. 13.10*

shift register (see Fig. 13.14). Therefore, after the third branch, the tree repeats. So, all the nodes marked **a** may be joined together. Same thing applies to nodes marked **b**, **c** and **d**. Joining like nodes together leads to the '*trellis*' representation.

The path traced out along the tree for the input sequence 1 0 1 0 0 1 considered in Example 13.21 is shown by the dotted line. The corresponding outputs along this path are: 1 1, 1 0, 0 0, 1 0, 1 1, 1 1, 1 0, 1 1 – exactly the same as what we got in Example 13.21. Note that eventhough there are only six digits in the input sequence, after the last input digit is shifted into the register, three more times we have to shift (with 0 input) in order to shift out the contents of the registers caused by the previous inputs. That is why the last three additional digits shifted into the shift register have been shown as 0s while tracing the path taken by the encoder for the given input. These last three digits (zeros) appended by us to the actual input message sequence, constitute what is called, the 'tail of the message'.

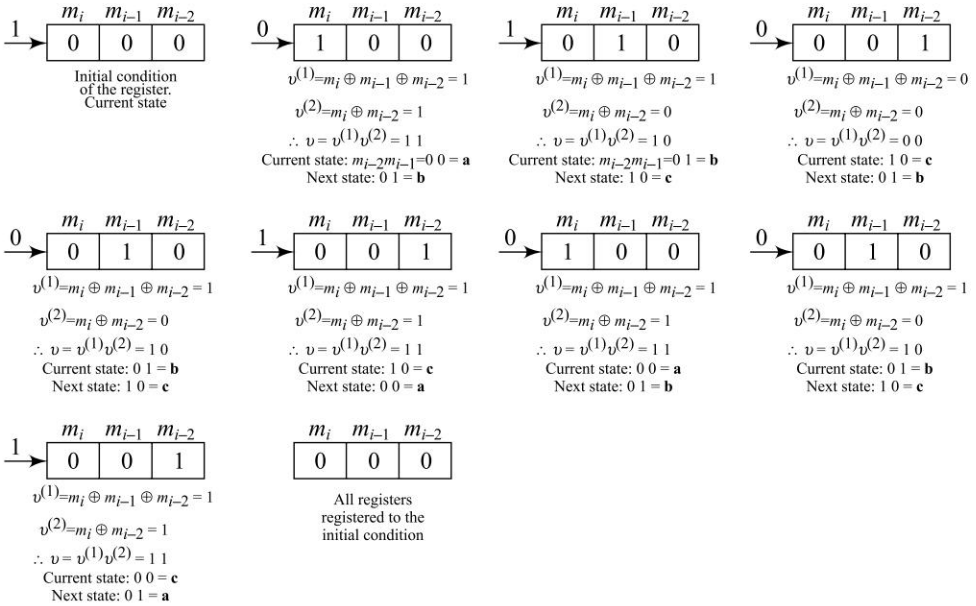


Fig. 13.14 *Pictorial display of the contents of the shift registers as the message digits 1 0 1 0 0 1 are shifted through the registers of the convolutional encoder of Fig. 13.10*

The sequence of changes in the shift register contents, the outputs $\upsilon^{(1)}$ and $\upsilon^{(2)}$, the state of the encoder each time a message digit is shifted into the shift register, is pictorially displayed in the following figure, i.e., Fig. 13.14.

13.7.2 Code Trellis

As we have observed during the discussion on the code tree, the tree is repetitive, the period of repetition, in terms of number of branches, depending upon the number of elements in the shift register used in the encoder. Since it is repetitive, a more compact description of a convolutional encoder may be obtained by joining together all the nodes in the code tree of Fig. 13.13 which are labeled using the same letter, say all nodes labeled '**a**', all nodes labeled '**b**' and so on. This results in an encoder representation, called the '*Code Trellis*'. In fact, the trellis is a more appropriate representation than the code tree for a convolutional encoder, which is a finite state machine. The code trellis for the (2, 1, 2) encoder of Fig. 13.10 is shown in Fig. 13.15. As shown in Fig. 13.14, the next state of the encoder is dependent on the current state and the bit m_i that has just been shifted into the shift register. So in the code trellis, we show the current state on the left side and the next state on the right side. As we have already seen, there are only four possible states – **a**, **b**, **c** and **d** for the (2, 1, 2) encoder since m_{i-2} and m_{i-1} can each be either a 0 or a 1 and so $m_{i-2}\, m_{i-1}$ which defines the state of the encoder, can take only the values 00, 01, 10, and 11, which we have represented by **a**, **b**, **c** and **d** respectively. A solid line has been used for the branch joining the current state with a next state if $m_i = 0$ and a broken line is used if $m_i = 1$. On each branch (either a solid line, or a broken line) the corresponding output υ is also marked.

13.7.3 State Diagram

Imagine that we fold the code trellis diagram vertically so as to make the **a**, **b**, **c** and **d** nodes in Fig. 13.15 representing current state coalesce with the corresponding nodes on the right representing the next states, we then get the state diagram, which is shown in Fig. 13.16. As shown in Fig. 13.15, it is only at nodes '**a**' and

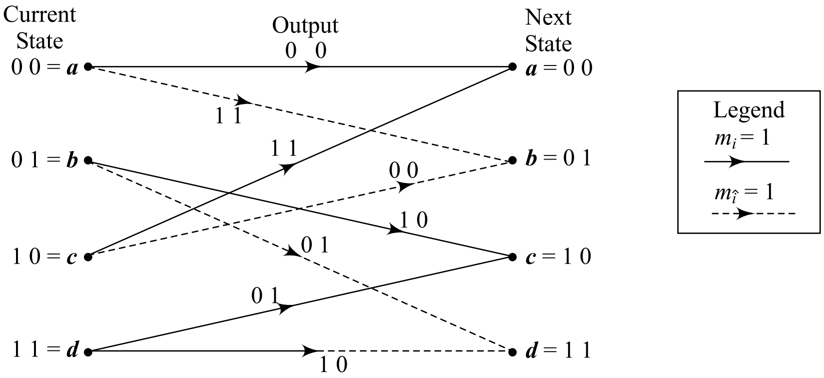


Fig. 13.15 *Code trellis for the (2, 1, 2) encoder of Fig. 13.10*

'**b**' that the current state and the next state can be the same – in the former case when m_i is 0 and in the latter, when m_i is a 1. So, these transitions appear as self-loops at those nodes a solid line loop at node '**a**' and a broken line loop at node '**d**'.

The code trellis, or the state diagram may be used for finding the resulting state sequence and output bits for a given message sequence.

13.7.4 The Trellis Diagram

The code trellis of Fig.13.15 shows only the steady-state transitions. A more useful trellis representation, referred to as the *trellis diagram*, is shown in Fig.13.17 and it can be drawn directly from the state diagram of Fig.13.16. The encoder is assumed to be initially in state '**a**', i.e., all the shift register contents are zeros. Looking at the state diagram of Fig.13.16, we find that an input of binary 0 to the encoder results in its remaining in the same state as indicated by the loop around state **a** in the state diagram and the horizontal line through the state **a** in the trellis diagram of Fig. 13.17. When the encoder is in state **a**, an input of binary 1, however, results in a transition to state **b** (see the state diagram) and this is indicated in the trellis diagram by a line from state **a** to state **b** and since it is the 'next state' from the initial state, it is shown against state **b** under the second node. As may be seen from the state diagram, any of the other three states can be reached from a given state by a sequence of two input bits. A third input results in transitions as shown. After the third input, the trellis becomes repetitive. In the trellis diagram, each trellis branch is labeled as shown, where the upper single digit represents the input and the lower double digit represents the resulting output.

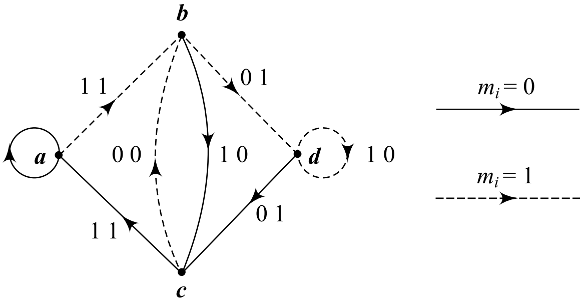


Fig. 13.16 *The state diagram for the (2, 1, 2) encoder of Fig. 13.10*

Example 13.25 Use the code trellis of Fig.13.14 and determine the output sequence and the sequence of states which the encoder goes through when an input sequence of 1 1 0 1 1 0 0 1 0 is given. Assume an initial state of 0 0, i.e., **a**.

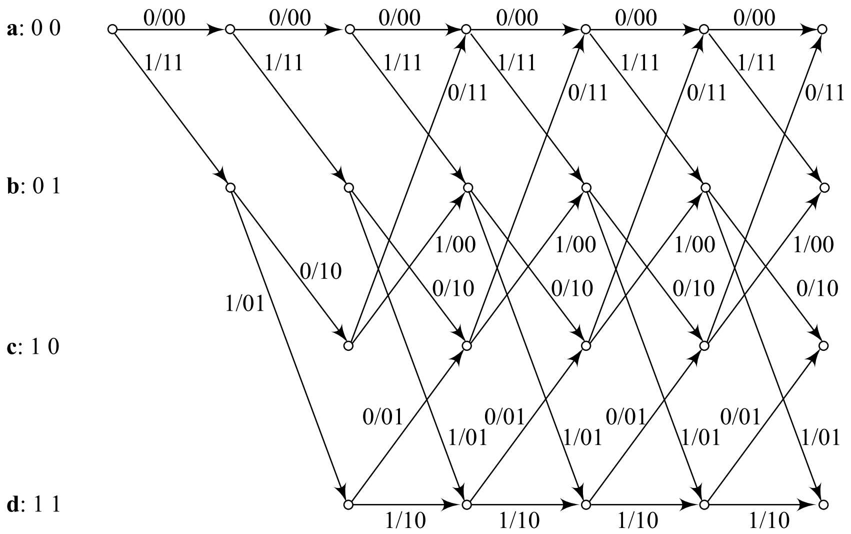


Fig. 13.17 *Trellis diagram for the encoder of Fig. 13.10*

Solution Using the code trellis of Fig. 13.14, we prepare the following table

Input		1		1		0		1		1		0		0		1		0	
State	**a**		**b**		**d**		**c**		**b**		**d**		**c**		**a**		**b**		**c**
Output		11		01		01		00		01		01		11		11		10	

Example 13.26 Sketch the state diagram for the convolutional encoder shown in Fig. 13.18. Code rate is 1/2 and constraint length $L = 2$. **(JNTU, NOV. 2009)**

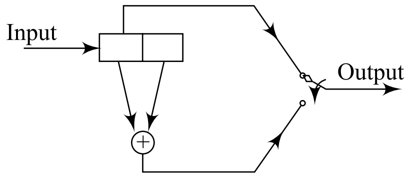


Fig. 13. 18

Solution From the diagram of the encoder, it is clear that the definition of constraint length followed is as mentioned in the note given under Eq. (13.62). So, what has been given as $L = 2$ in the problem actually refers to the total number of shift registers in the encoder. Hence, as per the notation we are following, this is an (n, k, L) encoder with $n = 2$, $k = 1$ and $L = 1$. Referring to Fig. 13.19.

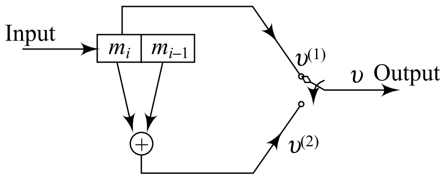


Fig. 13.19

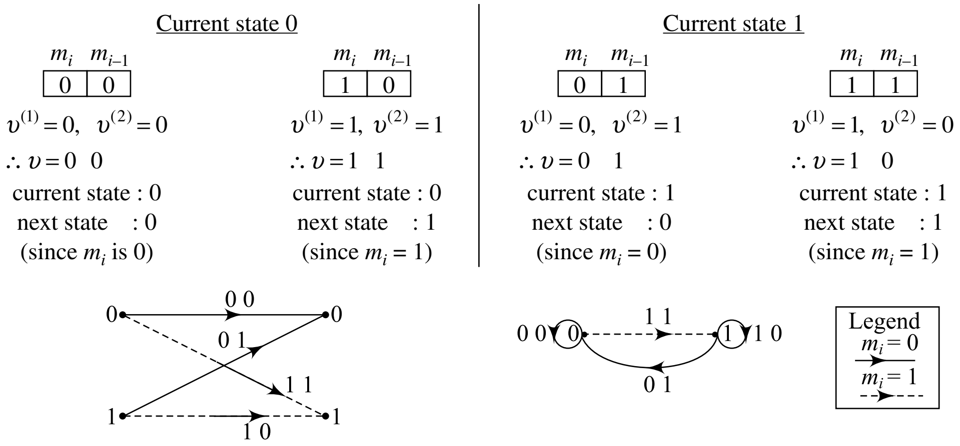
$$\upsilon^{(1)} = m_i$$

$$\therefore \qquad \upsilon^{(2)} = m_i \oplus m_{i-1}$$

The state = m_{i-1}

Since m_{i-1} can take only one of the two values, there are only two possible states.

Let the contents of the registers be 0, 0 initially. As shown below, there are only four possible values for $m_i\, m_{i-1}$



(a) Trellis **(b)** State diagram

Fig. 13.20

13.7.5 Decoding Methods

Convolutional codes may be decoded using tree-searching techniques. In Fig. 13.12, the path traced out along the tree for the input sequence 1 0 1 0 0 1 has been shown by a broken line on the code tree for the (2, 1, 2) convolutional encoder of Fig. 13.10. This suggests that to decode a received sequence, one can search the code tree for the path which is closest, in terms of Hamming distance, to the received sequence. However, since the number of possible paths is 2^N for an N-bit message sequence, tree-searching requires a very complex decoder.

In general, one may consider the following options for decoding of convolutional codes:

1. Feedback decoding
2. Sequential decoding
3. Maximum likelihood decoding

Feedback decoding requires simplified hardware but at the expense of good performance. For sequential decoding the complexity of the hardware goes on increasing as the performance becomes better and better. Maximum Likelihood decoding, which, for a Binary Symmetric Channel, is equivalent to minimum distance decoding, is optimal for AWGN but requires considerable search as mentioned earlier. However, Viterbi algorithm, which applies maximum likelihood principle, limits the comparison to a much smaller number compared to 2^N and has found extensive application in practice, particularly in satellite communications.

Before briefly discussing Viterbi algorithm, we shall first discuss the principle of Maximum Likelihood decoding.

Maximum likelihood decoding Having discussed the convolutional encoder and its code tree, trellis and state diagram representations, we will now proceed to discuss the decoding process. If $\boldsymbol{m}$ is the message sequence or vector, the encoder maps it into a code vector $\boldsymbol{C}$ and there is a one-to-one relation between the two. Let us assume that the code vector $\boldsymbol{C}$ produced by the encoder is transmitted over a Binary Symmetric Channel (BSC) which has a transition probability p owing to noise on the channel. Let the received vector be $\boldsymbol{r}$. The function of the receiver is to observe $\boldsymbol{r}$ and make the best possible estimate $\hat{\boldsymbol{m}}$ of $\boldsymbol{m}$. Since there is a one-to-one relationship between $\boldsymbol{C}$ and $\boldsymbol{m}$, it means that the receiver has to make the best possible estimate $\hat{\boldsymbol{C}}$ of $\boldsymbol{C}$ by observing the received vector $\boldsymbol{r}$. If its estimate $\hat{\boldsymbol{C}}$ of $\boldsymbol{C}$ is correct, i.e., if $\hat{\boldsymbol{C}} = \boldsymbol{C}$, then $\hat{\boldsymbol{m}} = \boldsymbol{m}$ and we say no decoding error has occurred. On the other hand, if $\hat{\boldsymbol{C}} \neq \boldsymbol{C}$, $\hat{\boldsymbol{m}}$ will not be equal to $\boldsymbol{m}$, the transmitted message vector, and so we say that a decoding error has occurred. The decoding rule by which the receiver chooses an estimate $\hat{C}$ of the transmitted codeword $\boldsymbol{C}$, is said to be an optimum one if that decoding rule minimizes the average probability of occurrence of a decoding error. From the maximum likelihood (ML) detection technique discussed in Chapter 4, we know that the probability of decoding error can be minimized by maximizing the log likelihood function. So, if $p(\boldsymbol{r} \mid \boldsymbol{C})$ is the conditional probability of receiving $\boldsymbol{r}$ given that $\boldsymbol{C}$ has been transmitted, then the ML decoding technique consists of choosing that estimate $\hat{C}$ of the transmitted code vector which maximizes $\log[p(\boldsymbol{r}|\boldsymbol{C})]$, the log likelihood function, and thereby minimizes the probability of decoding error.

Now, the transmitted code sequence $\boldsymbol{C}$ and the received sequence $\boldsymbol{r}$ will have the same number of digits in them, although the digits in corresponding locations might differ at a few locations owing to the effect of noise on the Binary Symmetric Channel (BSC) over which $\boldsymbol{C}$ is assumed to have been transmitted. So, if the transmitted sequence $\boldsymbol{C}$ is of length N digits, we will now try to express the log likelihood function in terms of the transition probability p of the BSC and the number of digits that are affected. Since the noise affects the various digits of $\boldsymbol{C}$ independently, if we denote the i^{th} digits of $\boldsymbol{C}$ and $\boldsymbol{r}$ by c_i and r_i respectively, we may write

$$p(\boldsymbol{r}|\boldsymbol{C}) = \prod_{i=1}^{N} p(r_i|c_i) \tag{13.69}$$

Taking logarithm on both sides,

$$\log p(\boldsymbol{r}|\boldsymbol{C}) = \sum_{i=1}^{N} \log p(r_i|c_i) \tag{13.70}$$

But,

$$p(r_i|c_i) = \begin{cases} p & \text{if } r_i \neq c_i \\ (1-p) & \text{if } r_i = c_i \end{cases} \tag{13.71}$$

So, if d be the number of digits of $\boldsymbol{C}$ that have been affected, i.e., if d be the Hamming distance between $\boldsymbol{r}$ and $\boldsymbol{C}$,

$$\begin{aligned} \log p(\boldsymbol{r}|\boldsymbol{C}) &= d \log p + (N-d)\log(1-p) \\ &= \log p^d + (N-d)\log(1-p) \end{aligned} \tag{13.72}$$

Note

(i) $N >> d$ so that $(N-d) \approx N$; and $N \log(1-p)$ is the same for all $\boldsymbol{C}$.

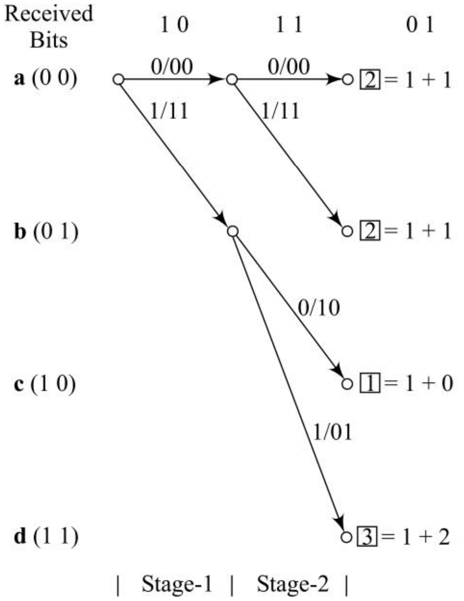
(ii) Since p, the transition probability of the BSC has to be less than 0.5 (why?) and is actually very much smaller that that, p^d will go on decreasing as d increases. Further, d, the Hamming distance between $\boldsymbol{r}$ and $\boldsymbol{C}$ will be different for the various possible code sequences. Thus, to maximize $\log[p(\boldsymbol{r}|\boldsymbol{C})]$, the log likelihood function, we have to minimize d by an appropriate choice of $\boldsymbol{C}$ as our estimate $\hat{C}$ of the transmitted code vector.

Thus, the maximum likelihood decoding is equivalent to minimum distance decoding. This means that when the receiver receives a sequence $\boldsymbol{r}$, it has to compare $\boldsymbol{r}$ with all the possible transmitted code vectors and choose that particular code vector which is closest to the received sequence $\boldsymbol{r}$ as the code vector that has been transmitted. '*closest*' here means that of all the possible code vectors, the chosen code vector should differ from the received vector $\boldsymbol{r}$ in the fewest number of locations.

Viterbi algorithm Earlier, we had stated that the code tree could be used for decoding a convolutional code by identifying a path through the code tree that differs from the received sequence in the fewest number of locations. We also stated that this is generally not feasible in practice because the number of paths in the tree code with which the received sequence will have to be compared, grows exponentially with the length of the message sequence. Since a code tree and the trellis diagram are equivalent representations, suppose we choose to use the trellis diagram rather than the code tree for decoding. Then the number of nodes at any level of the trellis diagram does not go on increasing with the length of the message sequence. Actually it remains constant at 2^{L+1} where $(L+1)$ denotes the total number of shift register stages. For example, if we consider the trellis diagram of Fig.13.17 which has been drawn for a (2, 1, 2) convolutional encoder (for which $L = 2$), there are 4 nodes at any level and at any of these four nodes, only two paths are entering. Further, these two paths are identical from that point onwards. The minimum distance decoder can, at this point, decide which path is to be retained in order to keep the path close to the received sequence. Again, at the next level too, a similar decision may be taken. This exactly, is what the Viterbi algorithm does as it proceeds through the trellis diagram. At each level, to compute the closeness of the path through the trellis with the received sequence, it uses the Hamming distance between the two as the metric. Thus, at each node, of the two paths, the one with smaller metric is retained and the other one is rejected. The retained paths are termed the survivor, or active paths. Only the survivor paths and their metrics are stored. Details of the way we have to proceed with Viterbi algorithm are best understood by carefully following the various steps in the following example.

Example 13.27 Apply Viterbi algorithm for maximum likelihood decoding of the convolutional code generated by the (2, 1, 2) convolutional encoder of Fig.13.10. Assume that the received sequence is 1 0 1 0 0 1 1 0 1 1 1 1.

Solution Initially the contents of all the shift-registers are assumed to be zero, i.e., we start from the state marked a (0 0). The procedure we adopt is: at each stage we find the optimum paths to the four states keeping in view the two received bits during that stage. From the state diagram of Fig.13.16 as well as the trellis diagram of Fig.13.17, it is clear that at any stage each state may be approached from two previous stages. For example, state **a** (0 0) may be approached from previous **a** (0 0) state, or **c** (1 0) state; state **b** (0 1) may be approached from **a** (0 0) state or **c** (1, 0) state; state **c** (1 0) may be approached from **b** (0 1) or **d** (1 1); and state **d** (1 1) may be approached from a previous **d** (1 1) state or from **b** (0 1) state. Always, the one with minimum Hamming distance is the survivor and is retained whereas the other path with larger Hamming distance is discarded. Hamming distances of the survivors are labeled.



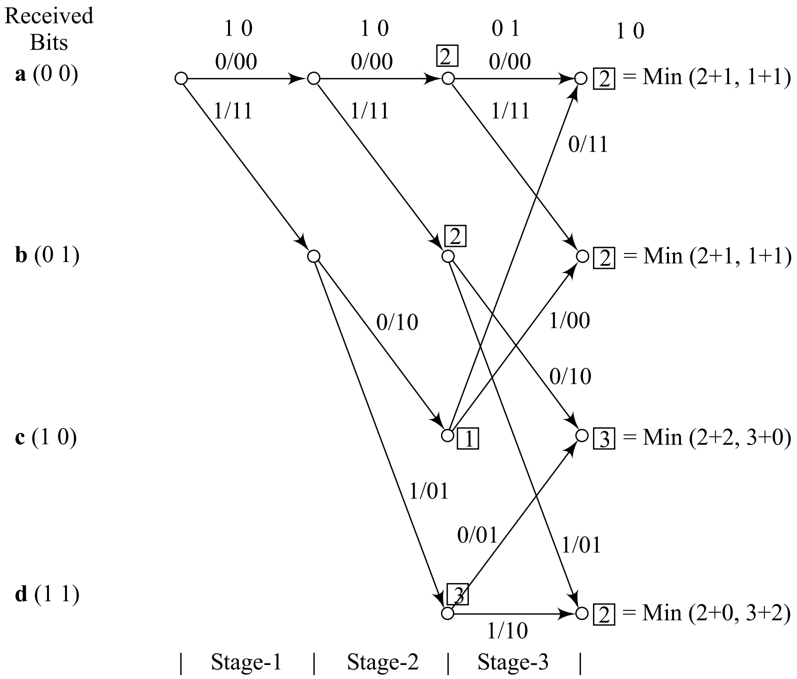
During stage 1, output 0 0 differs from the received bits 1 0 in 1 place. ∴ Hamming distance (HD) is 1. Similarly in stage 2 alos it is 1. ∴ Total 2.

Hence is **a** → **a** in stage 1. The output 0 0 differs from received bits 1 0 in one place. ∴ HD is 1. then **a** → **b** in stage 2. Output 1 1 differs from received bits 1 0 in one place. ∴ HD is 1. ∴ HD of this path to state b is 2 = 1 + 1

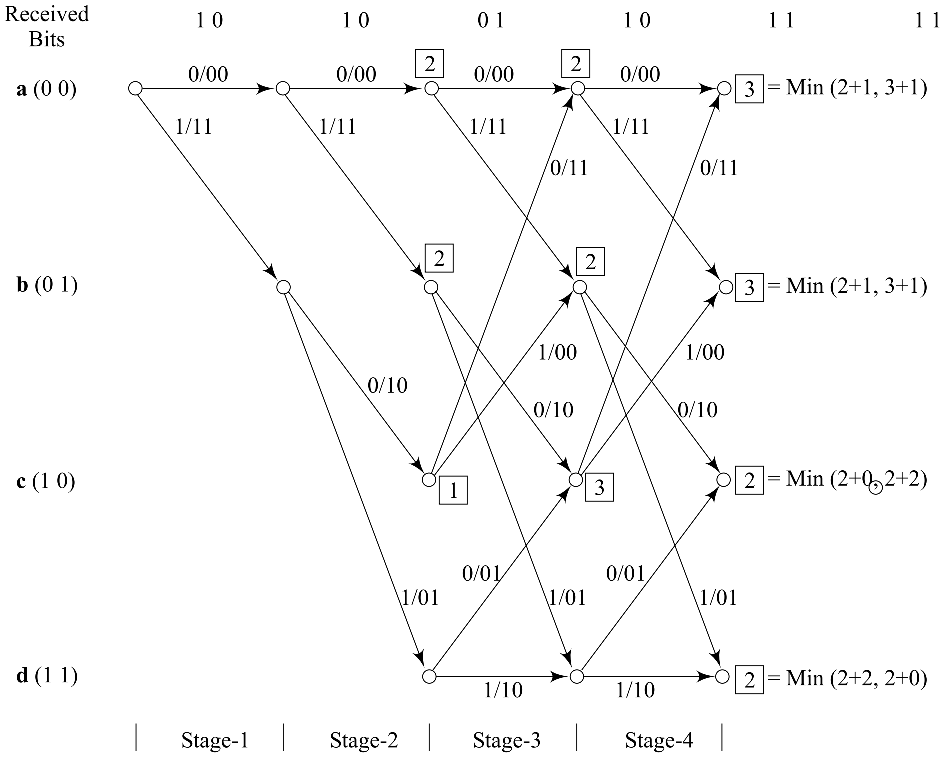
It is **a** → **b** in stage 1 and **b** → **c** in stage 2. The HD of the path in stage 1 is 1 and HD of the path from **b** to **c** is zero since output 1 0 matches with received bits. ∴ Total HD path = 1 = 1 + 0

Here it is **a** → **b** with HD = 1 and **b** to **d** with HD = 2 because output 0 1 differs from received bits in two places ∴ Total HD of the path = 1 + 2 = 3

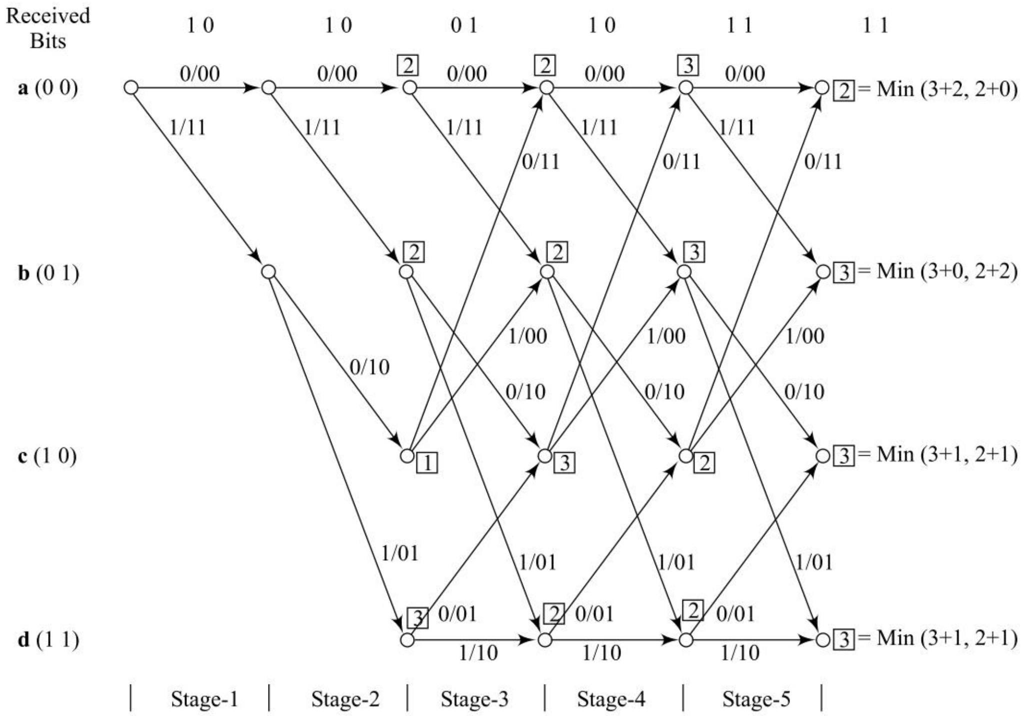
(a) Stages 1 and 2



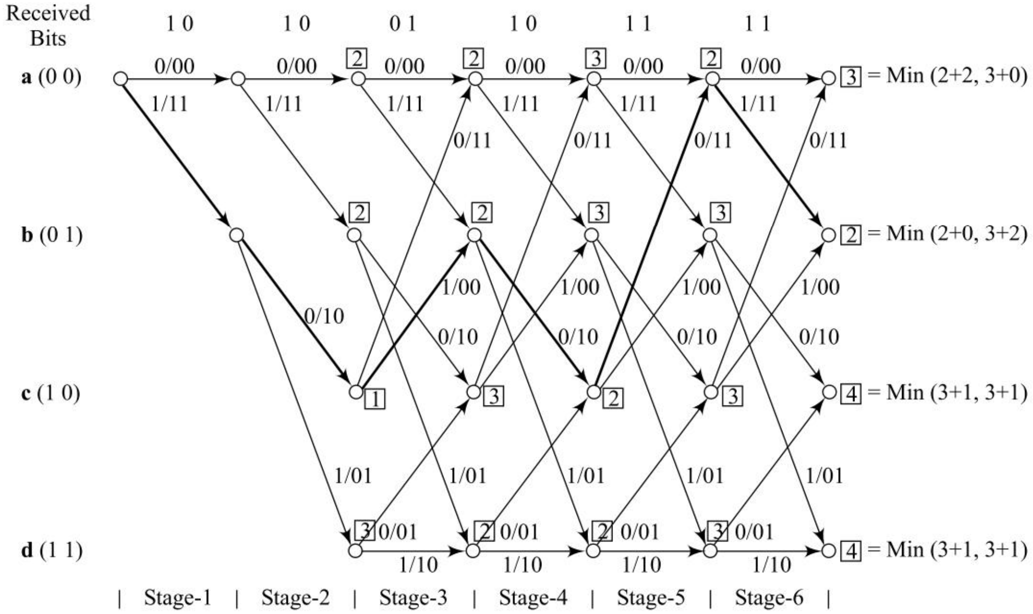
(b) Stage 3



(c) Stage 4



(d) Stage 5



(e) Stage 6

Fig. 13.21 *Illustration of the use of Viterbi algorithm*

The optimum path is identified to be the one indicated by thick, black path ending in state b. The decoded codeword and the corresponding received message sequence are:

Received bits	: 1 0 1 0 0 1 1 0 1 1 1 1
Decoded sequence	: 1 1 1 0 0 0 1 0 1 1 1 1
Received message sequence	: 1 0 1 0 0 1

Thus, the errors at the 2^{nd} and 6^{th} bits of the received sequence have been corrected and the actually transmitted sequence has been received. The message sequence used at the transmitter for encoding, has also been obtained correctly (check with the results of Example 13.21). The encoder returns to the original state **a** (0 0) and remains there after a complete codeword is outputted. Obviously, the given received sequence is not due to transmission of a complete codeword.

Free distance and coding gain of convolutional codes While discussing coding gain of block codes in Section 13.5, we found that coding gain depends upon d_{min}. For a block code, this has been defined as the minimum Hamming distance between any two codewords, or the minimum Hamming weight of any non-zero codeword of the code. But, in a convolutional code, there is nothing like a codeword. So, we regard to entire transmitted sequence corresponding to an input sequence, as a code vector and we define what is called the 'Free Distance' of a convolutional code as the minimum distance between code vectors, or the minimum Hamming weight of the non-zero code vector. This decides the error-correcting capability of the code. The free distance can be determined from the code trellis, but the procedure is quite involved and will not be described here.

Thus,

$$\text{Free distance } d_{\text{free}} \triangleq [W(X)]_{\text{min}} \tag{13.73}$$

where X is any non-zero transmitted code vector.

Just as the coding gain of a block code is equal to the product of its code rate, r and its d_{min}, for the case of convolutional codes too the coding gain depends on the product of d_{free} and r_c (where r_c is the rate of the code), and it is defined as follows.

$$\text{Coding Gain for a convolutional code } \triangleq \frac{1}{2} r_c d_{\text{free}}$$

A convolutional code improves reliability when its coding gain, $\frac{1}{2} r_c d_{\text{free}} > 1$. This also is usually expressed in dB.

Turbo codes With block codes and convolutional codes, it is not possible in practice, to approach the theoretical limit for Shannon's channel capacity because the codeword length in the case of block codes and the constraint length in the case of convolutional codes, need to be increased and that increases the complexity of the decoders very considerably, making their practical realization unfeasible. Of the various approaches proposed for overcoming these difficulties, Turbo codes and Low Density Parity Check (LDPC) codes are worth mentioning.

Turbo codes were first proposed by Berrou *et al.* in 1993. They have excellent Bit Error Rate (BER) performance, almost nearing Shannon limit and are becoming quite popular. The encoder of a turbo code consists of two systematic encoders connected together by means of an interleaver, as shown in Fig. 13.22.

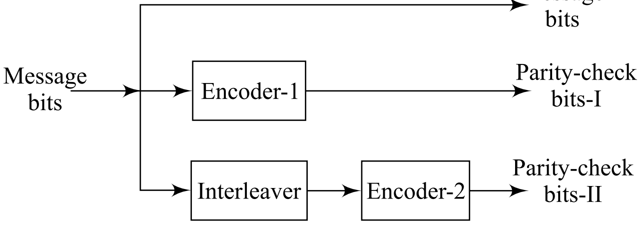


Fig. 13.22 *Encoder of a turbo code*

Encoders 1 and 2 generally use the same code and it is a short constraint length recursive convolutional code. The interleaver just permutes the input message bits and feed its output to Encoder 2. The permutation of the message bits by the interleaver is completely deterministic and may be of repetitive type or pseudo-random type.

Both turbo codes and LDPC codes make it possible to achieve very high coding gains of the order of 10 dB (coding gain may be taken as figure of merit for measuring the improvement in BER performance achieved by the use of a code).

MATLAB Example 13.1 For a (15,5) binary BCH code, find all its code words and determine its d_{min} (minimum Hamming weight of any nonzero code word). Show that it can correct up to 3 errors

MATLAB Program

```
% Show that a [15,5] BCH code can correct one error and
%  has a generator polynomial X^5 + X + 1.
clc
clear all
k = 5;% Message length
n = 15;% Code-word length
m = 4
%
nwords = 32% Number of words to encode (i.e., sequences are
randomly selected out of the 32 possible message
sequences, for encoding)
%
% Generation of messages
%
msg = gf(randint(nwords,k))
%
% Find t, the error-correction capability.
%
[genpoly,t] = bchgenpoly(n,k);
disp('Corresponding to the generator polynomial')
genpoly
disp('Error corrections')
t
disp('Therefore, It can correct up to three errors');
dmin = 2*t+1
%
%  Finding the code words...(encode the message).
%
disp('The rows represent the 32 code words')
code = bchenc(msg,n,k)
%
% adding noise to the code
%
% Corrupt up to t2 bits in each code word
disp('Code after addig noise')
noisycode =  code + randerr(nwords,n,1:t)
%
% Decode the noisy code.
%
```

```
[newmsg,err,ccode] =  bchdec(noisycode,n,k);
Hamming_weights_of_error_patterns = err
disp('The rows of this array represent the 32 message sequences obtained after
decoding the noise affected code words')
newmsg
disp('The rows represent the 32 code vectors obtained after decoding the noise
affected code vectors');
ccode
if ccode==code
  disp('All errors were corrected.')
end
if newmsg==msg
  disp('The messages have been recovered perfectly.')
end
[newmsg,err,ccode] = bchdec(noisycode,n,k);
ccode
```

Result

```
m = 4
nwords =  32
msg  array.
Array elements =
0    0    0    0    1
1    1    0    0    0
1    0    0    0    0
1    0    0    1    1
1    0    0    0    0
1    1    0    0    1
1    1    0    1    0
0    1    1    1    0
1    0    0    1    0
1    0    0    0    1
0    0    1    1    0
1    0    0    1    1
1    0    1    0    0
0    0    1    1    1
0    1    1    1    1
0    0    1    1    0
0    0    1    1    0
0    0    0    1    1
1    0    1    1    0
0    1    0    0    0
1    1    0    0    1
0    1    1    1    1
0    1    1    0    1
1    1    0    0    1
1    0    1    0    1
1    0    0    0    1
0    1    0    1    1
0    1    1    1    0
1    0    1    0    0
```

```
1    0    1    1    1
1    1    0    1    1
0    0    0    0    1
Corresponding to the generator polynomial genpoly
Array elements =
     1    0    1    0    0    1    1    0    1    1    1    0    0    0    0
Error corrections
```

  t = 3   Therefore, It can correct up to three errors d_{min} = 7

The rows represent the 32 = 2^5 code words

```
code Array elements =
0    0    0    0    1    0    1    0    0    1    1    0    1    1    1
1    1    0    0    0    0    1    0    1    0    0    1    1    0    1
1    0    0    0    0    1    0    1    0    0    1    1    0    1    1
1    0    0    1    1    0    1    1    1    0    0    0    0    1    0
1    0    0    0    0    1    0    1    0    0    1    1    0    1    1
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
1    1    0    1    0    1    1    0    0    1    0    0    0    1    1
0    1    1    1    0    0    0    0    1    0    1    0    0    1    1
1    0    0    1    0    0    0    1    1    1    1    0    1    0    1
1    0    0    0    1    1    1    1    0    1    0    1    1    0    0
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
1    0    0    1    1    0    1    1    1    0    0    0    0    1    0
1    0    1    0    0    1    1    0    1    1    1    0    0    0    0
0    0    1    1    1    1    0    1    0    1    1    0    0    1    0
0    1    1    1    1    0    1    0    1    1    0    0    1    0    0
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
0    0    0    1    1    1    1    0    1    0    1    1    0    0    1
1    0    1    1    0    0    1    0    0    0    1    1    1    1    0
0    1    0    0    0    1    1    1    1    0    1    0    1    1    0
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
0    1    1    1    1    0    1    0    1    1    0    0    1    0    0
0    1    1    0    1    1    1    0    0    0    0    1    0    1    0
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
1    0    1    0    1    1    0    0    1    0    0    0    1    1    1
1    0    0    0    1    1    1    1    0    1    0    1    1    0    0
0    1    0    1    1    0    0    1    0    0    0    1    1    1    1
0    1    1    1    0    0    0    0    1    0    1    0    0    1    1
1    0    1    0    0    1    1    0    1    1    1    0    0    0    0
1    0    1    1    1    0    0    0    0    1    0    1    0    0    1
1    1    0    1    1    1    0    0    0    0    1    0    1    0    0
0    0    0    0    1    0    1    0    0    1    1    0    1    1    1
```

Code after adding noise

```
Noisy code  array elements =
0    0    0    0    1    0    1    0    0    1    1    0    0    1    1
1    1    0    0    0    0    1    0    1    0    1    1    1    0    1
0    0    0    0    0    1    0    1    0    0    1    1    0    1    1
1    0    0    1    1    0    1    0    1    1    0    0    1    1    0
0    0    0    0    0    1    0    1    0    0    1    0    0    0    1
```

```
1    1    0    0    1    0    1    0    1    1    1    1    0    1    0
0    1    0    1    0    1    1    0    0    1    0    1    0    1    1
0    1    1    1    1    0    1    0    1    0    1    0    0    1    1
0    0    0    1    0    0    0    1    1    0    1    0    1    0    1
0    0    0    1    1    1    1    1    0    1    1    1    1    0    0
0    0    1    1    0    0    0    1    0    0    0    0    1    0    0
1    0    0    1    1    0    1    0    1    0    0    1    0    1    0
1    0    1    0    0    1    1    0    1    1    1    0    0    0    1
0    1    1    1    1    1    1    1    0    0    1    0    0    1    0
1    1    1    1    1    0    0    0    1    1    0    0    1    0    0
0    1    1    0    0    1    1    1    0    0    0    0    1    0    0
0    0    1    1    0    1    1    1    0    1    0    0    1    0    1
0    1    0    1    1    1    1    1    1    0    1    1    0    0    1
1    0    1    1    1    0    1    0    0    0    1    1    1    0    1
0    1    0    0    0    1    1    1    0    0    1    0    1    1    0
1    1    0    0    0    0    0    0    0    0    1    1    0    1    0
0    1    1    1    1    0    1    1    1    1    0    0    1    0    0
0    0    1    1    1    1    1    0    0    0    0    1    0    1    0
1    1    0    0    1    0    0    0    1    1    0    1    0    1    0
1    0    1    0    0    1    0    0    1    0    0    0    1    0    1
0    0    0    0    1    1    1    1    0    1    0    0    1    0    0
0    1    0    1    1    1    0    1    0    0    0    0    1    1    0
0    0    1    1    0    0    0    0    1    0    1    0    0    1    1
1    0    1    0    1    1    1    1    1    1    1    0    0    0    0
1    0    0    0    1    0    0    0    0    1    0    1    0    0    1
1    1    0    1    1    1    0    0    0    0    1    0    1    1    0
0    0    0    0    1    0    1    0    0    1    0    0    1    1    1
Hamming_weights_of_error_patterns =
Columns 1 through 15
1    1    1    3    3    1    2    2    2    3    3    2    1    3    2
Columns 16 through 30
3    1    2    3    1    3    1    2    1    2    2    3    1    2    2
Columns 31 through 32
1    1
Newmsg
The rows of this array represent the 32 message sequences obtained after decoding the noise-affected code words
newmsg Array elements =
0    0    0    0    1
1    1    0    0    0
1    0    0    0    0
1    0    0    1    1
1    0    0    0    0
1    1    0    0    1
1    1    0    1    0
0    1    1    1    0
1    0    0    1    0
1    0    0    0    1
0    0    1    1    0
1    0    0    1    1
```

```
1    0    1    0    0
0    0    1    1    1
0    1    1    1    1
0    0    1    1    0
0    0    1    1    0
0    0    0    1     1
1    0    1    1    0
0    1    0    0    0
1    1    0    0    1
0    1    1    1    1
0    1    1    0    1
1    1    0    0    1
1    0    1    0    1
1    0    0    0    1
0    1    0    1    1
0    1    1    1    0
1    0    1    0    0
1    0    1    1    1
1    1    0    1    1
0    0    0    0    1
The rows represent the 32 code vectors obtained after decoding the noise af-
fected code vectors
ccode  Array elements =
0    0    0    0    1    0    1    0    0    1    1    0    1    1    1
1    1    0    0    0    0    1    0    1    0    0    1    1    0    1
1    0    0    0    0    1    0    1    0    0    1    1    0    1    1
1    0    0    1    1    0    1    1    1    0    0    0    0    1    0
1    0    0    0    0    1    0    1    0    0    1    1    0    1    1
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
1    1    0    1    0    1    1    0    0    1    0    0    0    1    1
0    1    1    1    0    0    0    0    1    0    1    0    0    1    1
1    0    0    1    0    0    0    1    1    1    1    0    1    0    1
1    0    0    0    1    1    1    1    0    1    0    1    1    0    0
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
1    0    0    1    1    0    1    1    1    0    0    0    0    1    0
1    0    1    0    0    1    1    0    1    1    1    0    0    0    0
0    0    1    1    1    1    0    1    0    1    1    0    0    1    0
0    1    1    1    1    0    1    0    1    1    0    0    1    0    0
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
0    0    1    1    0    1    1    1    0    0    0    0    1    0    1
0    0    0    1    1    1    1    0    1    0    1    1    0    0    1
1    0    1    1    0    0    1    0    0    0    1    1    1    1    0
0    1    0    0    0    1    1    1    1    0    1    0    1    1    0
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
0    1    1    1    1    0    1    0    1    1    0    0    1    0    0
0    1    1    0    1    1    1    0    0    0    0    1    0    1    0
1    1    0    0    1    0    0    0    1    1    1    1    0    1    0
1    0    1    0    1    1    0    0    1    0    0    0    1    1    1
1    0    0    0    1    1    1    1    0    1    0    1    1    0    0
0    1    0    1    1    0    0    1    0    0    0    1    1    1    1
0    1    1    1    0    0    0    0    1    0    1    0    0    1    1
```

```
1    0    1    0    0    1    1    0    1    1    1    0    0    0    0
1    0    1    1    1    0    0    0    0    1    0    1    0    0    1
1    1    0    1    1    1    0    0    0    0    1    0    1    0    0
0    0    0    0    1    0    1    0    0    1    1    0    1    1    1
All errors were corrected.
The messages have been recovered perfectly.
```

Summary

- *Error control strategies:*
  - (a) Forward Error Correction (FEC)
  - (b) Automatic Repeat Request (ARQ)
- *ARQ Systems:* These use error detecting codes. On detecting an error, the receiver requests the transmitter through a feedback channel, for retransmission.
- *Error-detecting and correcting codes:* The channel encoder at the transmitter introduces controlled redundancy. The decoder of the receiver makes use of this redundancy to detect/correct the errors. Error correction needs more redundancy to be introduced than error detection.
- *Error-control codes:* They are basically of two types:
  - (a) Block codes
  - (b) Convolutional codes
- *Block codes:* The encoder of an (n, k) block code takes k message bits at a time and maps them into n encoded bits $(n > k)$. The (n, k) block code has 2^k codewords.
- *Code rate:* The code rate of an (n, k) block code is the ratio of number of message bits to the number of bits in a codeword, i.e., code rate = $(k/n) = r$ and $0 < r < 1$.
- *Channel data rate:* If the source rate is R_s and an (n, k) block code encoder supplies data to the channel, the channel data rate $= (n/k)R_s$.
- *Hamming distance:* The Hamming distance between two N-length binary sequences is defined as the number of locations in which they are different.
- *Systematic block code:* A systematic (n, k) block code is one in which each codeword consists of k unaltered message bits followed by $(n - k)$ parity check bits.
- *Generator matrix:* The generator matrix, $\boldsymbol{G}$, of an (n, k) systematic block code is given by

  $$\boldsymbol{G} = [\boldsymbol{P} \vdots \boldsymbol{I}_k]_{k \times n}$$

  where $\boldsymbol{P}$ is the $k \times (n-k)$ coefficient matrix given by

  $$\boldsymbol{P} = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,n-k-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,n-k-1} \\ \vdots & \vdots & & \vdots \\ p_{k-1,0} & p_{k-1,1} & \cdots & p_{k-1,n-k-1} \end{bmatrix}$$

- *Code vector:* If $\boldsymbol{C}$ is a $1 \times n$ code vector, $\boldsymbol{m}$ is a $1 \times k$ message vector and $\boldsymbol{G}$ is the generator matrix of an (n, k) block code, then $\boldsymbol{C} = \boldsymbol{m}[\boldsymbol{G}]$.
- *Parity-check matrix $\boldsymbol{H}$*: $\boldsymbol{H} = [\boldsymbol{I}_{n-k} \vdots \boldsymbol{P}^T]_{(n-k) \times n}$
- *Some useful relations:* $\boldsymbol{G} \cdot \boldsymbol{H}^T = \boldsymbol{0}$ and $\boldsymbol{H}\boldsymbol{G}^T = \boldsymbol{0}$
- *Repetition (n, k) code:* It is a linear (n, k) block code in which a single message bit is encoded into n identical bits to produce a $(n, 1)$ block code.
- *Syndrome:* If $\underline{r}$ is a $1 \times n$ received vector, its syndrome is a $1 \times (n–k)$ vector given by $\boldsymbol{S} = \boldsymbol{r} \cdot \boldsymbol{H}^T$ while $\boldsymbol{C} \cdot \boldsymbol{H}^T = \boldsymbol{0}$
- *Hamming code:* It is an (n, k) linear block code that has the following parameters and is a single-error correcting code:

  Block length : $n = 2^m - 1$

  No. of message bits: $k = 2^m - 1 - m$

No.of parity bits : $m = (n - k)$
and m is a positive integer and at least equal to 3.

- *Cyclic code:* It is a subclass of linear block codes. Any cyclic shift given to a non-zero codeword of a cyclic code results in another codeword.
- *Generator polynomial g(x):* The generator polynomial $g(x)$ of an (n, k) cyclic code is a polynomial of degree $(n–k)$ that is a factor of $(x^n + 1)$. It is also the polynomial of least degree in the code.
- *Parity-check polynomial h(x):* The parity-check polynomial of an (n, k) cyclic code with $g(x)$ as its generator polynomial, satisfies the relation:

$$g(x)\, h(x) \bmod (x^n + 1) = 0$$

- *Dual of a linear block code:* For a linear (n, k) block code with ***G*** as the generator matrix and ***H*** as the parity check matrix, there exists a dual code with ***H*** as its generator matrix and ***G*** as the parity-check matrix.
- *BCH codes:* Common binary BCH codes, known as primitive BCH codes are characterized by
  Block length $n = 2^m - 1$
  No. of message bits: $k \geq n - mt$ where $m \geq 3$ and $t < (2^m - 1)/2$.
  Min. distance : $d_{\min} \geq 2t + 1$
- *Convolutional encoder:* An encoder for a convolutional code generates the output codeword by modulo-2 discrete convolution of its own impulse response with the sequence formed by the present message bit and a few message bits preceding the present bit.
- *Structure of a convolutional encoder:* A convolutional encoder, in its simplest form, is a finite state machine and consists of a tapped shift register with $(L+1)$ stages whose outputs are connected to one or more modulo-2 adders through coefficient multipliers.
- *Constraint length:* For an (n, k, L) convolutional code, the quantity $n(L + 1)$ is referred to as the constraint length. It represents the number of successive encoded bits influenced by each message bit.
- *Code rate of a convolutional code:* The code rate of an (n, k, L) convolutional code is given by

$$r = \frac{K}{n(K + L)}$$

where K represents the message sequence length. If $K >> L$.

$$r \cong \frac{1}{n}$$

- *Representation of the structure of a convolutional code:* It may be represented by any of the following three equivalent methods:
  (a) Code tree (b) State diagram (c) Code trellis diagram
- *Decoding of convolutional codes:* Basically, there are three methods of decoding a convolutional code:
  (a) Sequential decoding (b) Feedback decoding
  (c) Maximum likelihood decoding (ML decoding)
  Of these, Viterbi algorithm using ML decoding has become quite popular.

References and Suggested Readings

1. Berrou, C., and A. Glavieux, Near Optimum Error Correcting Coding and Decoding: Turbo Codes, *IEEE Trans. on Communications*, Vol. 44, pp. 1261–1271, 1996.
2. Berlekamp, E.R., *Algebraic Coding Theory*, McGraw-Hill, N.Y. 1968.
3. Bhargava, V.K., Forward Error Correction Schemes for Digital Communications, *IEEE Communications Magazine*, Vol. 21, No.1, pp. 11–19, 1983.
4. Carlson, A. B., Paul B. Crilly, and Janet C. Rutledge, *Communication Systems – An Introduction to Signals and Noise in Electrical Communication*, 4th Edition (International Edition), McGraw-Hill, 2002.
5. Fehr, Dr. Kamilo, *Wireless Digital Communication – Modulation and Spread Spectrum Applications*, Eastern Economic Edition, Prentice-Hall Private Limited (India), 1995.
6. Forney, Jr., G.D., The Viterbi Algorithm, *Proceedings of the IEEE*, Vol. 61, pp. 268–278, 1973.

7. Gallager, R. G., *Information Theory and Reliable Communication*, Wiley & Sons, New York, 1968.
8. Hamming, R.W., *Coding and Information Theroy*, Prentice Hall, Englewood Cliffs, New Jersey 1980.
9. Haykin, Simon, *Communication Systems*, 4th Edition, (Wiley Student Edition) Wiley India, 2001.
10. Lathi, B. P., and Zhi Ding, *Modern Digital and Analog Communication Systems*, International 4th Edition, Oxford University Press, 2010.
11. Lee, Edward A., and David G. Messerschmidt, *Digital Communications*, 2nd Edition, Allied Publishers Limited, 1994.
12. Lin, S., and D.J. Costello Jr., *Error-control Coding – Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.
13. Proakis, John G., *Digital Communications*, 4th Edition (McGraw-Hill International Edition), McGraw-Hill, 2001.
14. Proakis, John G., and Masoud Salehi, *Fundamentals of Communication Systems*, Low Price Edition, Pearson Education, 2005.
15. Taub, Herbert, Donald L. Schilling, and Gautam Saha, *Taub's Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill, New Delhi, 2008.
16. Wells, Richard B., *Applied Coding and Information Theory for Engineers*, Low Price Edition, Pearson Education, 1999.
17. Ziemer, R.E., and W.H. Tranter, *Principles of Communications – Systems, Modulation and Noise*, 3rd Edition, Jaico Publishing House, Mumbai, 2002.

Review Questions

1. What are the two broad strategies adopted for error control in digital communications?
2. What are the different types of ARQ? Briefly discuss their features, advantages and disadvantages.
3. Compare ARQ and FEC methods of error control
4. What is meant by a 'linear code'?
5. What is a systematic block code?
6. In a linear block code, how are the parity check bits produced?
7. Why is the generator matrix of a linear block code given that name?
8. In an (n, k) linear block code, how many codewords will be there? Justify your answer.
9. Given $\boldsymbol{G}$, the generator matrix of a systematic linear block code, is it possible to construct the $\boldsymbol{H}$ matrix? If your answer is a 'yes', explain how.
10. If $\boldsymbol{G}$ and $\boldsymbol{H}$ are the generator matrix and the parity-check matrix of a linear block code, show that $\boldsymbol{G}\cdot\boldsymbol{H}^T = 0$.
11. What is a repetition code?
12. What is a 'syndrome vector'? How is it useful?
13. Briefly explain syndrome decoding of linear block codes.
14. Draw the block diagram of an encoder for a linear (n, k) block code and explain its working.
15. What is the meaning and the significance of d_{min} of a block code?
16. What is a cyclic code?
17. Show that giving i cyclic shifts to the right to the code vector of a cyclic code is equivalent to multiplying the code polynomial corresponding to that code vector by x^i, modulo – $(x^n + 1)$
18. If $b(x) = b_0 + b_1x + \ldots + b_{n-k-1}x^{n-k-1}$ is the parity bits polynomial and $m(x) = m_0 + m_1x + \ldots + m_{k-1}x^{k-1}$ is the message bits polynomial of a systematic (n, k) linear cyclic code, write down the expression for the corresponding code polynomial.
19. State the relation between the generator polynomial $g(x)$ and the parity check polynomial $h(x)$ of an (n, k) cyclic linear code.
20. Draw the block diagram of an encoder for an (n, k) linear cyclic code and explain its working.
21. Draw the block diagram of the syndrome calculator for an (n, k) cyclic code and explain its working.
22. What is a Hamming code? What are its features?
23. Describe the structure of a t-error correcting binary BCH code. State some of the important good features of BCH codes.
24. What is a convolutional code? How is it different from a block code?

25. A particular convolutional code is described as an (n, k, L) code. What do these letters n, k and L represent?
26. What is meant by the ‘constraint length’ of a convolutional encoder?
27. What are the different methods of describing the structure of a convolutional encoder?
28. Explain how you would draw the trellis diagram of a convolutional encoder given its state diagram.
29. Show that for a binary symmetric channel, the ML decoding is equivalent to minimum distance decoding.
30. Briefly describe Viterbi algorithm for maximum likelihood decoding of convolutional codes.

Problems

1. For the (7, 4) systematic Hamming code, determine
   (a) the generator matrix $\boldsymbol{G}$.
   (b) the parity-check matrix, $\boldsymbol{H}$.
   (c) all the valid codewords
   (d) the minimum distance, d_{min} of the code
2. For the dual of the systematic (7, 4) Hamming code of problem 1, find the $\boldsymbol{G}$ matrix, $\boldsymbol{H}$ matrix, all the valid codewords, and d_{min}.
3. For a (3, 1) repetition code, determine:
   (a) $\boldsymbol{G}$ matrix
   (b) $\boldsymbol{H}$ matrix
   (c) The syndrome vector $\boldsymbol{S}$ for all possible single error patterns
4. Let $\boldsymbol{H}$ be the parity-check matrix of an (n, k) code. Then the matrix $\boldsymbol{H}_e$ defined by

$$\boldsymbol{H}_e = \begin{bmatrix} & & & 0 \\ & & & 0 \\ & \boldsymbol{H} & & \vdots \\ & & & 0 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

   Is the parity-check matrix of an extended code. Show that
   (a) $\boldsymbol{H}_e$ defines an $(n + 1, k)$ code
   (b) the minimum distance of the extended code is one more than the minimum distance of the original code.
5. Determine the relationship between n and k of a Hamming code. Using this result, show that the code rate approaches 1 for large value of n.
6. For the (7, 4) cyclic code of Example 13.18 with $1 + x + x^3$ as the generator polynomial $g(x)$, show that (0 1 1 0 1 0 0) is a code vector. When this codeword is transmitted, it was received as 0 1 1 1 1 0 0. Determine the syndrome polynomial for this received word.
7. A Golay code is a cyclic (23, 12) code with

$$g(x) = 1 + x + x^5 + x^6 + x^7 + x^9 + x^{11}$$

   Show that it can correct up to 3 errors.
8. For the convolutional encoder shown in Fig. P13.8, draw the code tree.
9. For the convolutional encoder of Fig. P13.1, draw the state diagram, and the trellis diagram.
10. The received sequence corresponding to an output sequence of the encoder of Fig. P13.1 is 1 1 0 1 0 1 0 1 1 1 1 1. Using Viterbi algorithm and using the trellis diagram of Problem 9, find the decoded sequence.

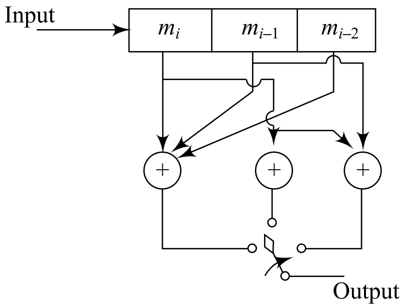


Fig. P13.8

Multiple-Choice Questions

1. The ARQ systems having highest and lowest throughput efficiency are respectively
   (a) Stop-and-wait ARQ and Goback-N ARQ
   (b) Selective repeat ARQ and Stop-and-wait ARQ
   (c) Selective repeat ARQ and Goback-N ARQ
   (d) Stop-and-wait ARQ and Selective repeat ARQ

2. For a linear (7, 4) block code, the ratio of parity bits to message bits and the code rate are respectively

(a) $\left(\frac{3}{7}, \frac{4}{7}\right)$; (b) $\left(\frac{4}{7}, \frac{3}{7}\right)$; (c) $\left(\frac{3}{4}, \frac{4}{7}\right)$; (d) $\left(\frac{4}{3}, \frac{4}{7}\right)$

3. For a linear (7, 4) block code, the code rate and the ratio of channel rate to source rate are respectively

(a) $\left(\frac{4}{7}, \frac{7}{4}\right)$; (b) $\left(\frac{4}{7}, \frac{3}{7}\right)$; (c) $\left(\frac{4}{7}, \frac{4}{7}\right)$; (d) $\left(\frac{3}{7}, \frac{7}{3}\right)$

4. Given $\boldsymbol{C}_1 = 1\ 1\ 0\ 1\ 0\ 1\ 1$ and $\boldsymbol{C}_2 = 1\ 0\ 1\ 1\ 1\ 0\ 1$, the difference in their Hamming weights and the Hamming weight of their difference, are respectively

(a) (1, 4) (b) (0, 0) (c) (4, 0) (d) (0, 4)

5. For a (7, 4) linear block code, the sizes of $\boldsymbol{G}$ and $\boldsymbol{H}^T$ matrices are respectively

(a) $(4 \times 7,\ \ 3 \times 7)$; (b) $(4 \times 7,\ \ 7 \times 3)$; (c) $(4 \times 7,\ \ 7 \times 4)$; (d) $(7 \times 4,\ \ 7 \times 3)$

6. For a (5, 1) repletion code, the sizes of the $\boldsymbol{G}$ matrix and $\boldsymbol{H}$ matrix are respectively

(a) $(4 \times 5,\ \ 5 \times 4)$; (b) $(4 \times 1,\ \ 4 \times 5)$; (c) $(1 \times 5,\ \ 4 \times 5)$; (d) $(1 \times 4,\ \ 4 \times 5)$

7. A (7, 4) systematic Hamming code has $\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$. It is known that a certain received code sequence $\boldsymbol{r}$ is in error in the second, third and fifth places. The syndrome of $\boldsymbol{r}$ is

(a) (0, 1, 0) (b) (0, 1, 1) (c) (0, 0, 1) (d) (0, 0, 0)

8. $g(x)$ and $h(x)$ are respectively the generator polynomial and the parity-check polynomial of a linear cyclic (n, k) code. The degree of $h(x)$ and the product of $g(x)$ and $h(x)$ are respectively

(a) $(k, x^n + 1)$; (b) $(k, 1)$ (c) $(n - k, 1)$ (d) $(n - k, x^n + 1)$

9. If a block code is to have a t-error correction capability, then the minimum distance $d_{\min}$ of the code should be such that

(a) $d_{\min} \geq 2t$ (b) $d_{\min} \geq t$ (c) $d_{\min} \geq 2t + 1$ (d) $d_{\min} > 2t + 1$

10. If the constraint length of an (n, k, L) convolutional code is defined as the number of encoder output bits influenced by each message bit, then the constraint length is given by

(a) $L(n + 1)$ (b) $n(L + 1)$ (c) $n(L + k)$ (d) $L(n + k)$

11. The code tree of an (n, k, L) linear convolutional code repeats after the

(a) k^{th} stage (b) L th stage (c) $(k$+1)th stage (d) $(L$+1)th stage

12. The state diagram of an (n, k, L) convolutional encoder is shown in Fig. M13.12. The number of shift register stages and the number of outputs for the encoder are respectively:

(a) (2, 2) (b) (1, 1) (c) (2, 1) (d) (1, 2)

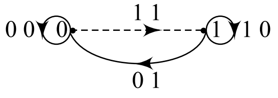


Fig. M13.12

Answers for Multiple-Choice Questions

1. (b) 2. (c) 3. (a) 4. (d) 5. (b) 6. (c) 7. (d) 8. (a)
9. (c) 10. (b) 11. (d) 12. (a)

SPREAD SPECTRUM COMMUNICATION SYSTEMS

14

> *"Most of the important things in the world have been accomplished by people who have kept on trying when there seemed to be no hope at all."*
>
> **Dale Carnegie (1888 – 1955)**
> *American writer*

Learning Objectives

After going through this chapter, students will be able to

- *explain the principle of DSSS and FHSS and describe the operation of spread spectrum-based communication systems,*
- *analyze the operation of and design PN sequence generators that produce PN sequences with good auto-correlation and cross-correlation characteristics,*
- *calculate the probability of error of DS spread spectrum systems under single-tone jamming and multi-user conditions,*
- *calculate the probability of error of FHSS systems under 'barrage jamming' and 'partial-band jamming' conditions,*
- *describe how spread spectrum systems may be used for providing Code Division Multiple Access, and*
- *explain how the PN code generators at the transmitter and the receiver can be perfectly aligned.*

Historical Background

In 1942, one, Ms. Hedy Lamarr, a scientifically talented Hollywood movie star and her co-inventor and an eccentric classical music composer, Mr. George Antheil took a USA patent for what they called, "A Secret Communication System". Conceptwise, it was the same as what we now know as a "Frequency Hopping Spread Spectrum System". The idea which they had patented was transmission of a message over a number of radio frequency carriers which followed a seemingly random pattern of sequence. The main objective of their proposed system was to have secure and secret communication between a transmitter and a receiver. Later, the US defense authorities used this concept for radio-guidance of a torpedo towards the intended target without any interference by way of jamming, etc. Because of the immense possibilities for its use in military applications, most of the research and development work on spread spectrum systems was carried out by the military establishments of various countries and most of this work was not available in open literature. The first commercial application of spread spectrum concept came up only in 1980 and it was for providing multiple access facility from the earth stations to the transponders on board a geosynchronous communication satellite.

14.1 INTRODUCTION

There are some applications, like military communications, wherein a transmitted message is to be received only by the receiver for which it is intended; others should not be able to receive it. Further, in order to make the communication reliable, it should not be possible for any one to jam the transmitted signal.

Spread spectrum systems are intended to provide such secure and reliable communication. As the name suggests, these systems spread the spectrum of the transmitted signal over a very wide bandwidth. This is achieved in these systems by modulating for a second time, an already modulated signal in such a way as to spread the power of the transmitted spread spectrum signal over a very large bandwidth. Thus, the power spectral density of this signal is so low that any ordinary AM (or FM) receiver, with its 10 kHz (or 200 kHz) front-end bandwidth, receives an amount of spread spectrum signal power that is very much lower than the thermal noise power entering the receiver. Thus, the unauthorized receiver will not be able to receive the spread spectrum (SS) signal. As will be shown later, the SS signal cannot easily be jammed. Thus, these SS systems can provide very secure and reliable communication, making them ideally suited for military communications.

One attractive feature of spread spectrum signal which makes it extremely useful, especially in civilian applications, is that it enables an increase in the number of users over a given band – a feature that is exploited for providing multiple access in satellite communications and for increasing the number of subscribers using the same band, in the case of cellular mobile communications.

Thus, spread spectrum communication provides

1. protection against eavesdropping
2. resistance to intentional jamming
3. resistance to fading caused by multipath effects
4. multi-user facility over a given channel
5. ranging facility

Basically, there are two types of spread spectrum systems. They are

1. Direct Sequence (DS) spread spectrum systems
2. Frequency Hopping (FH) spread spectrum systems

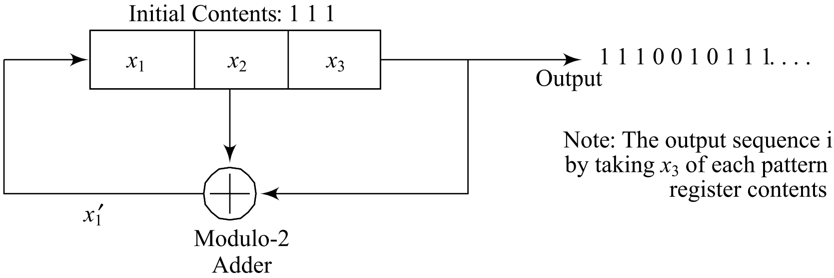
In this chapter, we will study both these types of systems, taking them in that order. Since pseudo noise (PN) sequences play an important role in the spread spectrum systems, we shall first briefly discuss PN sequences – their generation and their characteristics.

14.2 PSEUDO-NOISE SEQUENCES (PN SEQUENCES)

These are binary sequences which resemble random binary sequences. (In random binary sequences, 0s and 1s are equi-probable.) Actually these PN sequences are deterministic and they repeat after a certain length. However, because of their noise-like properties, they are widely used in applications such as synchronization, ranging, etc. Just as white noise has a delta function as its auto-correlation function, these PN sequences also have auto-correlation functions which are highly peaked for zero shift (i.e., Delta) and are almost zero even for small shifts to either side.

PN sequences can be generated easily using shift register circuits with feedback from one or more stages. A PN sequence generator using a three-stage shift register is shown in Fig. 14.1.

Since there are 3-shift-register stages and since each stage can have either a zero or a one, there can be $2^3 = 8$ distinct sets of contents, including 0 0 0. However, the 0 0 0 state is not permitted because once the shift register contents are 0 0 0, there will be no change whatever may be the number of shifts we give by clocking the circuit. For the PN sequence generator of Fig. 14.1, if we assume that the shift register contents are initially 1 1 1, with each clocking pulse, the contents will change as shown in Table 14.1.



Note: The output sequence is obtained by taking x_3 of each pattern of the shift register contents

Fig. 14.1 *A three-stage shift register, PN sequence generator*

Table 14.1 *Operation of the PN sequence generator of Fig. 14.1*

Shifts		$x_1' = x_2 * x_3$		**Shift register contents** $x_1\ x_2\ x_3$
0		-		1 1 1
1		$1 \oplus 1 = 0$		0 1 1
2		$1 \oplus 1 = 0$		0 0 1
3		$0 \oplus 1 = 1$		1 0 0
4		$0 \oplus 0 = 0$		0 1 0
5		$1 \oplus 0 = 1$		1 0 1
6		$0 \oplus 1 = 1$		1 1 0
7		$1 \oplus 0 = 1$		1 1 1

Note *After the 7th shift, the pattern of the contents will repeat.*

Thus, this PN sequence generator produces a sequence of length 7 and thereafter the same sequence will be repeated. This is to be expected since we have excluded one pattern, the all-zero pattern from the eight possible patterns. Hence, if N is the length of the sequence and m is the number of shift register stages,

$$N = 2^m - 1 \tag{14.1}$$

Every PN-sequence generator with m shift registers need not produce $(2^m - 1)$ length sequence. It depends on the feedback connections and the type of logic circuit used for combining the feedback outputs (In Fig. 14.1, the logic used was simply exclusive-OR addition). PN sequences with $2^m - 1$ length are called a *maximal length sequences*. An important property of a maximal length sequence is the number of 1s in it is always one more than the number of 0s. This property is called the ‘balance property’.

Since the PN sequence is periodic (with a period of $2^m - 1$ for maximal length sequence), its auto-correlation function defined by

$$R(\tau) = \frac{1}{T_c} \int_{-T_c/2}^{T_c/2} c(t)c(t-\tau)d\tau \tag{14.2}$$

is also periodic with the same period.

In Eq. (14.2), $c(t)$ is the time function representing the PN sequence, τ is the delay and T_c is the duration of each binary digit in the sequence.

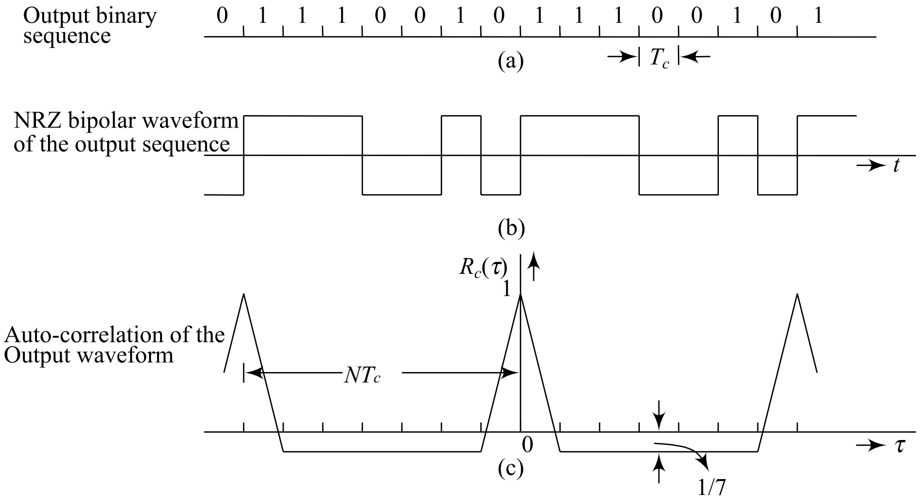


Fig. 14.2 *(a) Output binary sequence (b) NRZ Bipolar waveform representation of the output sequence (c) Auto-correlation function of the PN sequence generated by the PN sequence generator of Fig. 14.1.*

Applying Eq. (14.2) to the case of a maximal-length sequence of N bits length and denoting the bit-duration of the maximal-length sequence by T_c, we get

$$R_c(\tau) = \begin{cases} 1 - \left(\dfrac{N+1}{NT_c}\right)|\tau|; & |\tau| \le T_c \\ -\dfrac{1}{N} & \text{for the rest of the period} \end{cases} \tag{14.3}$$

Figure 14.2 shows the binary output sequence from the PN sequence generator in (a), the corresponding bipolar waveform in (b) and the auto-correlation function $R_c(\tau)$ as a function of τ (as given in Eq. (14.3)) in (c).

We know that the Fourier transform of the ACF gives the PSD of a signal. Further, we also know that since the ACF, $R_c(\tau)$, is periodic, the power spectral density of the PN sequence is discrete. It is in this respect that the maximal-length PN sequence differs from the random binary sequence. Since the maximal-length sequence is periodic, its ACF is also periodic and its spectrum is discrete. However, the random binary sequence is not periodic and its power spectrum is continuous.

Table 14.2 gives the number and location of the feedback connections that would give us maximal-length PN sequences.

Table 14.2 *Feedback details for generation of maximal-length PN sequences*

m	Sequence Length *N*	Sequence (Initial state: All ones)	Feedback digit
2	3	1 1 0	$x_1 \oplus x_2$
3	7	1 1 1 0 0 1 0	$x_2 \oplus x_2$
4	15	1 1 1 1 0 0 0 1 0 0 1 1 0 1 0	$x_3 \oplus x_4$
5	31	1 1 1 1 1 0 0 1 1 0 1 0 0 1 0 0 0 0 1 0 1 0 1 1 1 0 1 1 0 0 0	$x_2 \oplus x_5$

Example 14.1 A four-stage shift register with feedback connections taken from the outputs of stages 4 and 1 through a modulo-2 adder, is used for PN sequence generation. Assuming the initial contents of the shift register to be 0100, determine the output sequence. What is the length of the sequence?

Solution

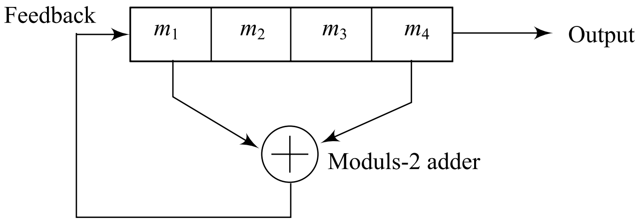


Fig. 14.3 *PN Generator circuit*

Feedback Digit	Shift Register Contents				Output Digit
	0	**1**	**0**	**0**	
0	0	0	1	0	0
0	0	0	0	1	0
1	1	0	0	0	1
1	1	1	0	0	0
1	1	1	1	0	0
1	1	1	1	1	0
0	0	1	1	1	1
1	1	0	1	1	1
0	0	1	0	1	1
1	1	0	1	0	1
1	1	1	0	1	0
0	0	1	1	0	1
0	0	0	1	1	0
1	1	0	0	1	1
0	0	1	0	0	1

The table shows the initial contents of the shift-register stages at the top. The first row shows the feedback digit $m_1 \oplus m_4$, the contents of the shift-register stages and the output digit after the first shifting. The other rows show the feedback digit, the shift-register contents and the output digit after each subsequent shifting.

After 15 shiftings, the initial contents of the shift-registers are once gain obtained. For further shiftings, the same cycle of events will repeat. Thus, the length of one period of the PN sequence is 15.

The output sequence corresponding to one period of the generator is as given in the last column and is

Output sequence: 0 0 1 0 0 0 1 1 1 1 0 1 0 1 1

Since m, the number of stages, is 4 and since the length of one period of the sequence is 15 which is equal to $2^m - 1$, the sequence is a maximal length sequence.

14.3 BASICS OF DIRECT SEQUENCE SPREAD SPECTRUM COMMUNICATION

In an actual DS spread spectrum system, the binary data to be transmitted is first carrier modulated – say using PSK and then this modulated signal is subjected to spreading by multiplying it by the PN sequence. However, in order to discuss the effect of multiplying the data sequence by the PN sequence, for the present, we shall consider only a baseband DS spread spectrum system, i.e., without initial carrier modulation by the data sequence. Let the NRZ bipolar waveform of the data sequence (before carrier modulation) be denoted by $d(t)$ and the NRZ polar waveform of the binary output sequence of the PN sequence generator be denoted by $c(t)$. Let the data digit duration be T_b sec and the PN sequence digit duration be T_c sec. In DS spread spectrum systems, it is always so arranged that

$$T_c << T_b \tag{14.4}$$

Figure 14.4 shows the waveforms of $d(t)$ and $c(t)$ and also of the product of $d(t)$ and $c(t)$.

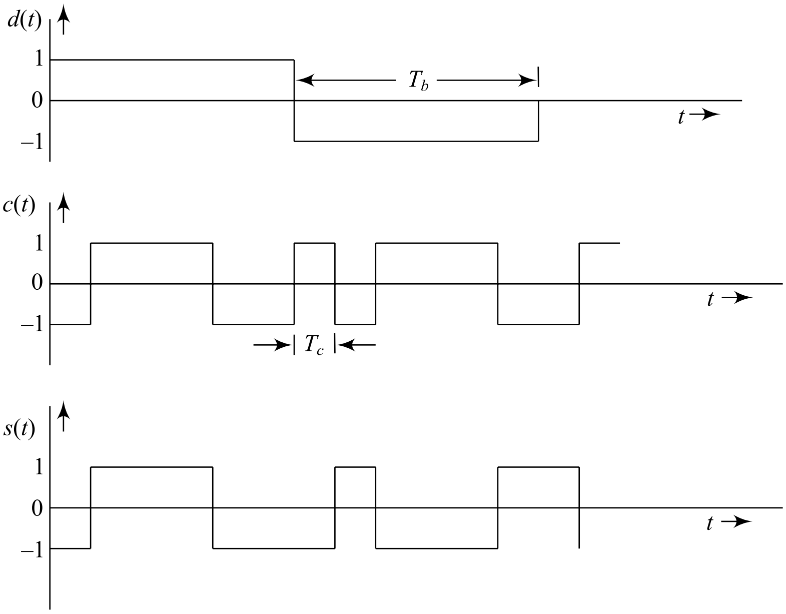


Fig. 14.4 *(a) d(t), the NRZ polar waveform of the data sequence, (b) c(t), the NRZ polar waveform of the PN sequence, (c) s(t), the product waveform, i.e., product of d(t) and c(t).*

Since $T_c << T_b$, the waveform $d(t)$ is narrow band, while the $c(t)$ waveform is a wideband signal. Product of $d(t)$ and $c(t)$ in time domain is equivalent to convolution of their spectra. But since the spectrum of $c(t)$ is very wide compared to that of $d(t)$, the product waveform $s(t)$ will have a spectrum which is almost the same as that of $c(t)$, the PN sequence waveform. That is, the spectrum of $s(t)$, the baseband DS spread spectrum signal is almost like the spectrum of $c(t)$ itself. The PN sequence waveform, with its digital period T_c very small compared to T_b, actually 'chips in' into the data waveform $d(t)$. That is why, the inverse of T_c is called the chip frequency. Figure 14.5 shows the basic operations to be performed at the transmitter for producing the *baseband* spread spectrum signal and at the receiver for recovering the data sequence from the baseband DS spread spectrum signal. In order to illustrate how the spread spectrum modulation enables us to reject

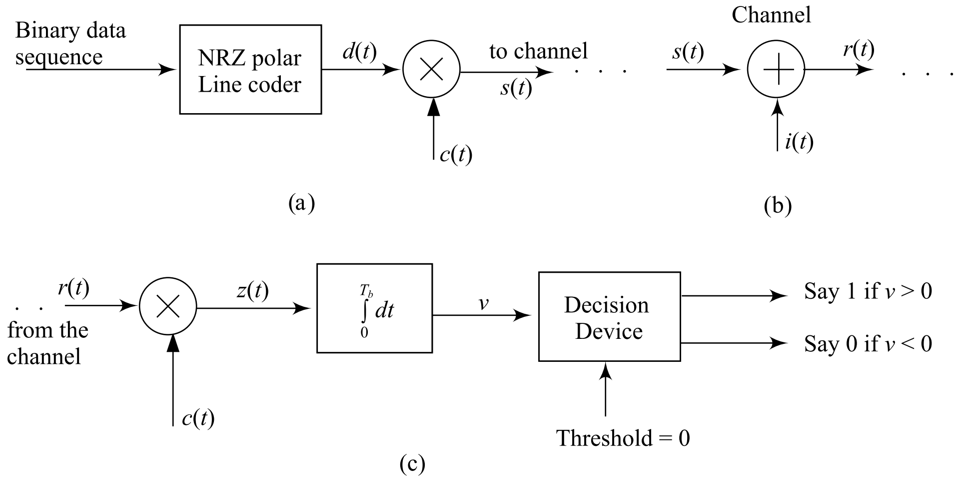


Fig. 14.5 *Baseband DS spread spectrum communication system model. (a) Transmitter, (b) Channel, (c) Receiver*

deterministic interfering signals *added* to the transmitted signal $s(t)$ during the course of its passage through the channel, we are adding the interfering signal $i(t)$ to the DS spread spectrum signal $s(t)$.

Since the interference is additive,

$$r(t) = s(t) + i(t) = d(t) \cdot c(t) + i(t) \tag{14.5}$$

The first operation to be performed at the receiver is to de-spread the received signal. For this purpose, it is multiplied by the PN sequence waveform $c(t)$, which is assumed to be in perfect synchronism with the $c(t)$ used at the transmitting end.

$$z(t) = r(t) \cdot c(t) = d(t) \cdot c^2(t) + i(t) \cdot c(t) \tag{14.6}$$

From Eq. (14.3b), we find that $c(t)$ is either 1 or –1 at any time.
So, $c^2(t) = +1$ for all t. Hence,

$$z(t) = d(t) + i(t) \cdot c(t) \tag{14.7}$$

From Eq. (14.7), we find that when we de-spread the message or data waveform $d(t)$, the interference signal $i(t)$ gets spread over a wide bandwidth by getting multiplied by the PN sequence waveform, $c(t)$. Thus, we find that $z(t)$ consists of a narrowband component $d(t)$ and a wideband component $i(t).c(t)$. As shown in Fig.14.5(c), $z(t)$ is integrated over a period of T_b, the data bit duration. The integrator acts as a low pass filter and removes the wideband component $i(t) \cdot c(t)$, thus achieving suppression of interfering signals. Further, at the end of each T_b, the output of the integrator gives a voltage v, whose value depends on whether the $d(t)$ was +1 or –1 during that interval T_b besides, of course, the period T_b itself. This voltage is given to a comparator which acts as the decision device and says that $d(t)$ was 1 during that T_b if $v > 0$ and that it was a –1 if $v < 0$. Thus, the original data sequence is recovered suppressing the additive interfering signals picked up by the channel.

The protection to the data given by the spreading sequence will improve as the PN sequence length is longer for a given data rate. The price paid for the security of communication, of course, is larger transmission bandwidth, more complexity of the system, etc.

14.4 BPSK-DS SPREAD SPECTRUM SYSTEMS AND PROBABILITY OF ERROR

We have discussed in the previous section, how a DS spread spectrum baseband system can suppress interference signals. In practice, however, the data sequence, after spreading, is carrier modulated, generally using either BPSK, QPSK, or MSK. Then it is transmitted over the channel. At the receiving end, the received signal is first subjected to coherent detection using the locally generated carrier signal that is arranged to be in phase and frequency synchronism with the carrier used at the transmitter. The output of the coherent detector is then subjected to dispreading by multiplying it with a locally generated PN sequence that is identical to and in synchronism with the one at the transmitter. After de-spreading, it is integrated over a bit duration T_b to get the observed random variable v, which is used for decision making, as shown in Fig.14.6.

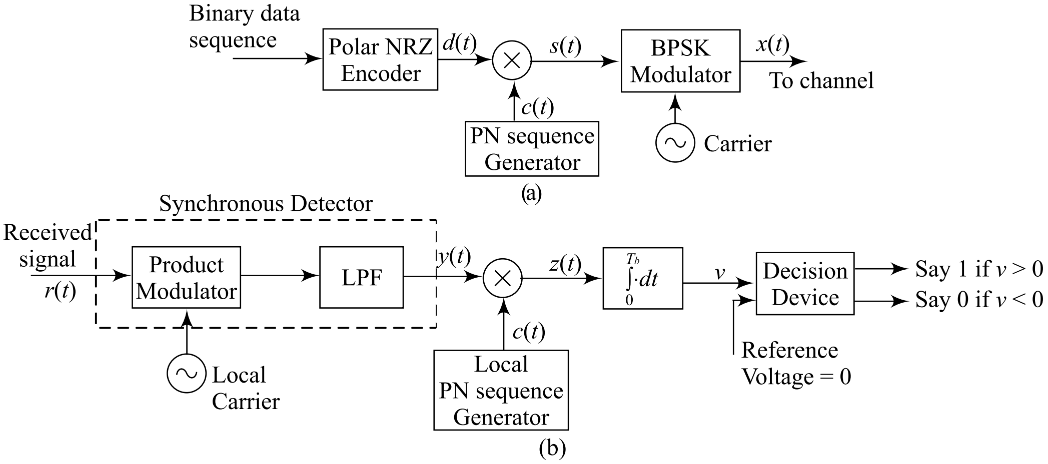


Fig. 14.6 *Direct sequence spread spectrum system using BPSK (a) Transmitter, (b) Receiver*

14.4.1 Probability of Error due to Thermal Noise on the Channel

In Section 14.3, we had shown that a deterministic interfering signal added to the baseband spread spectrum signal in the course of its passage through the channel will be suppressed by the dispreading operation and the subsequent integration in the receiver. Thus, the interference signal is not going to influence the decision made by the decision device. In other words, the interfering signal does not affect the probability of error.

From this, one may be tempted to jump to the conclusion that spread-spectrum systems will, in a similar way, suppress the random noise also added by the channel. But it is not correct to conclude like that. Random noise added to the SS signal during its passage through the channel, is unaffected by the de-spreading operation in the receiver. The de-spreading signal $c(t)$ is like a random binary wave. So when the noise is multiplied by it, all that happens is that for some randomly occurring periods, the polarity of the noise is changed. Obviously, this does not in any way affect the power spectral density, or the probability density function of the noise. Earlier, we had seen that the message data sequence is un-affected by the spreading and de-spreading. Since the signal as well as the thermal noise added by the channel are unaffected, the probability of error of a DS spread spectrum system using BPSK modulation, is the same as what a normal BPSK system gives. That is,

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_b}{\eta}} \tag{14.8}$$

14.5 RESISTANCE TO JAMMING

Jamming is resorted to in order to make a communication ineffective. It consists of radiating a large amount of RF power in a narrow band around the carrier frequency used for that communication.

We will now briefly analyze and see the effect of jamming on DS spread spectrum BPSK communication. To simplify the analysis, we shall make the following assumptions:

1. We will assume that the jamming signal is a single-tone signal at the frequency f_c which is the frequency of the carrier used for BPSK modulation at the DS spread spectrum BPSK transmitter.
2. Although in practice as shown in Fig. 14.6, the spectrum spreading operation at the transmitting-end precedes the BPSK modulation, we will, for the sake of this analysis, assume that the modulation (BPSK) is done first and the spectrum spreading is done subsequently. Similarly, at the receiver also, we will reverse the order in which phase demodulation and spectrum de-spreading are done. This is quite justified, because of the linear nature of all these operations.

So, for this analysis, we will use the following model for the DS spread-spectrum BPSK communication system.

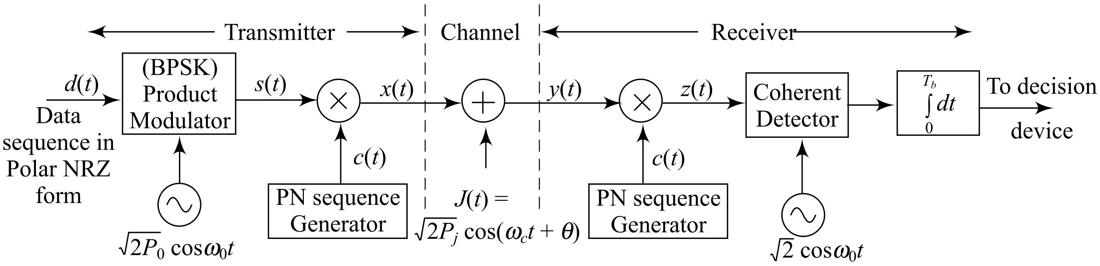


Fig. 14.7 *Model of DS spread-spectrum BPSK system used for analysis*

Let the carrier signal at the transmitter have a power of P_0 and a frequency of f_c so that the carrier signal may be represented as $\sqrt{2P_0}\cos\omega_0 t$. Let the jamming signal be of normalized power P_j and frequency f_c so that it is $J(t) = \sqrt{2P_j}\cos(\omega_0 t + \theta)$. The jammer signal phase will not have any relationship with the phase of the carrier used for BPSK modulation. θ, in general, will be a random phase and we may justifiably assume that it is uniformly distributed over $[0, 2\pi]$. We assume perfect synchronism between the two PN-sequence generators – one at the transmitter and the other at the receiver. We also assume that the locally generated carrier signal has a nominal power of unity and that it is in frequency and phase synchronism with the carrier used at the transmitter.

With the data $d(t)$ in polar NRZ format, the BPSK modulator is just a product device. Hence,

$$s(t) = \sqrt{2P_0}\,d(t)\cos\omega_0 t \tag{14.9}$$

and

$$x(t) = s(t)\cdot c(t) = \sqrt{2P_0}\,d(t)c(t)\cos\omega_0 t \tag{14.10}$$

∴

$$y(t) = \sqrt{2P_0}\,d(t)c(t)\cos\omega_0 t + \sqrt{2P_j}\cos(\omega_0 t + \theta) \tag{14.11}$$

and

$$z(t) = \sqrt{2P_0}\,d(t)c^2(t)\cos\omega_0 t + \sqrt{2P_j}\,[\cos(\omega_0 t + \theta)]c(t)$$

But $c^2(t)$, as already pointed out earlier, is equal to 1 for all t.

∴

$$z(t) = \sqrt{2P_0}\,d(t)\cos\omega_0 t + \sqrt{2P_j}\,c(t)\cos(\omega_0 t + \theta) \tag{14.12}$$

The coherent detector is again a product device followed by a low pass filter. Hence, the input, $w(t)$, to the low pass filter in the coherent detector is given by

$$\begin{aligned} w(t) &= z(t)\cdot\sqrt{2}\cos\omega_0 t \\ &= 2\sqrt{P_0}d(t)\cos^2\omega_0 t + 2\sqrt{P_j}c(t)[\cos^2\omega_0 t\cos\theta + \sin\omega_0 t\cos\omega_0 t\sin\theta)] \\ &= \sqrt{P_0}d(t)[1+\cos 2\omega_0 t] + \sqrt{P_j}c(t)[1+\cos 2\omega_0 t]\cos\theta + \sqrt{P_j}c(t)[\sin 2\omega_0 t]\sin\theta \end{aligned}$$

When $w(t)$ is fed to the low pass filter whose cut-off frequency is just equal to the bandwidth of the baseband data signal, $d(t)$, all the terms involving components with frequency around $2f_c$ will be blocked by the LPF since its cut-off frequency is very much smaller than $2f_0$. Hence, output of the LPF, viz. $v(t)$ is given by

$$v(t) = \sqrt{P_0}d(t) + \sqrt{P_j}c'(t)\cos\theta \tag{14.13}$$

Note that in the second term on the RHS of Eq. (14.13), we have used $c'(t)$ instead of $c(t)$. This is because, $c(t)$ is a wideband signal and the LPF with its cut-off frequency $f_b = 1/T_b$ (this is the cut-off frequency chosen because the LPF pass bandwidth should be just enough to pass the data signal $d(t)$, whose bandwidth is $1/T_b$) will not pass $c(t)$. It will pass only that part of the spectrum of $c(t)$ which lies within its pass band. Hence $c'(t)$ represents the time signal corresponding to that part of the spectrum of $c(t)$ which passes through the LPF. The power spectrum of $\sqrt{P_j}c(t)\cos\theta$ is given by

$$S(f) = \frac{P_jT_cE[\cos^2\theta]}{2}\left[\frac{\sin\pi T_c f}{\pi T_c f}\right]^2 \tag{14.14}$$

This is a sinc^2 function whose first zero value occurs at $f = 1/T_c = f_c$. Since f_c, the chip frequency is very much greater than f_b, the inverse of the bit duration, the value of $S(f)$ over a bandwidth of $|f| \le f_b$ can safely be taken as a constant equal to the one corresponding to the peak value of the sinc^2 function.

The PSD of the output term caused by the sinusoidal jamming signal is given by

$$S(f) = \frac{P_jT_cE[\cos^2\theta]}{2} \tag{14.15}$$

As mentioned earlier, θ is a r.v. which is uniformly distributed over $[0, 2\pi]$. Hence,

$$E[\cos^2\theta] = \frac{1}{2} \tag{14.16}$$

$$\therefore \qquad S_0(f) = (P_jT_c)/4; \quad |f| \le f_b = \frac{1}{T_b} \tag{14.17}$$

Now, from Eq. (14.12), it is clear that the signal component of the input to the coherent detector is the same as what it would have been for a simple BPSK receiver. If the interfering signal were random channel noise of two-sided PSD equal to $\eta/2$, the noise signal at the input to the LPF of the coherent detector would also have a two-sided PSD of $\eta/2$. In such a case, we know that the probability of error, P_e would be

$$P_e = \frac{1}{2}erfc\sqrt{\frac{E_b}{\eta}} \tag{14.18}$$

When we have the sinusoidal jamming signal instead of white noise as the interfering signal added by the channel, in the case of the DS spread-spectrum BPSK system, the data signal component of the input to the coherent receiver is the same but the interfering signal component, instead of being the noise signal with PSD of $\eta/2$, is a wideband signal (because of the chip signal $c(t)$) which gives at the output of the LPF a noise-like signal with a constant PSD over $|f| \le f_b$, given by $S_0(f)$ of Eq. (14.17). So, substituting $S_0(f)$ in the place of $\eta/2$ in Eq. (14.18), we get

$$P_e = \frac{1}{2} erfc \sqrt{\frac{2E_b}{P_j \cdot T_c}}$$

But E_b = bit energy = $P_0 T_b$.

∴ $$P_e = \frac{1}{2} erfc \sqrt{2\left(\frac{P_0}{P_j}\right) \cdot \left(\frac{T_b}{T_c}\right)} = \frac{1}{2} erfc \sqrt{\frac{P_0}{(P_j \cdot T_c)/2T_b}} \tag{14.19}$$

∴ $$P_e = \frac{1}{2} erfc \sqrt{\frac{P_0}{P_j/2(f_c/f_b)}} \tag{14.20}$$

The quantity $[P_j/2(f_c/f_b)]$ is called the effective jamming power

∴ $$P_{j\text{eff}} \triangleq [P_j/2(f_c/f_b)] \tag{14.21}$$

∴ $$P_e = \frac{1}{2} erfc \sqrt{\frac{P_0}{P_{j\text{eff}}}} \tag{14.22}$$

The quantity (f_c/f_b), the ratio of the chip frequency to the bit frequency, which is always very much larger than 1, is called the processing gain and denoted by G_p.

∴ $$\text{Processing gain} = G_P \triangleq (f_c/f_b) \tag{14.23}$$

The processing gain is a measure of the extent to which the jamming signal power is reduced due to the use of spread spectrum. Larger the value of G_p, smaller will be the $P_{j\,eff}$ for a given P_j. Smaller the value of $P_{j\,eff}$, smaller will be the P_e since the complementary error function is a monotonically decreasing function of its argument.

So, higher the ratio of chip frequency to bit frequency, better will be the resistance to a narrowband jamming signal.

Example 14.2 Determine the jamming margin of a DSSS/BPSK system with a processing gain $G_p = 40$ dB if $P_e = 10^{-8}$ in the presence of jamming.

Solution Assuming equi-probable 0s and 1s, the probability of error of a BPSK system is

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{\eta}} = Q \sqrt{\frac{2E_b}{\eta}}$$

where E_b is bit energy and η is the one-sided PSD of white noise.

We may treat the jammer also as another source of noise. If P_j is the jammer's power at the receiver, we may write the probability of error due to the jamming signal alone as

$$P_e = Q \sqrt{\frac{2E_b}{N_j}}$$

where $N_j \triangleq \frac{P_j}{W_c}$

W_c being the transmission bandwidth of the DSSS system. Since the value of η has not been given, we shall ignore channel noise.

$$P_e = Q(2E_b/N_j) = 10^{-8}$$

From the Q-function tables, we find that

$$\sqrt{2E_b/N_j} = 5.7 \quad \therefore (E_b/N_j) = 16.245$$

Jamming margin (J.M) is defined as

$$\underset{(\text{in dB})}{J \cdot M} \triangleq 10 \log_{10} G_p - 10 \log_{10}(E_b/N_j)$$

and it represents the ability of the system to operate in the presence of jamming/interference.

$$\therefore \qquad \underset{(\text{in dB})}{J \cdot M} = 40 - 10 \log_{10} 16.245 = 40 - 12.107$$

$$\therefore \qquad J \cdot M(\text{in dB}) = 27.9 \text{ dB}$$

14.5.1 Multiple Access Using DS Spread Spectrum

Till now we have discussed how DS spread spectrum systems enjoy some immunity against evesdropping (i.e., interception by unauthorized receivers) and narrowband jamming. However, if with all its huge transmission bandwidth, if a DSSS system can serve only one user, it will indeed be wasteful of bandwidth. DS spread spectrum systems can, in fact, provide multiple access, i.e., allow multiple users. This multiple access facility is called the Code Division Multiplexing Access (CDMA) and it has certain advantages over the other multiple access facilities like FDMA, TDMA, etc. It does not require any bandwidth allocation as in FDMA or, any time allocation and synchronization as in TDMA.

In CDMA using Direct Sequence Spread Spectrum (DSSS) each user is provided with a unique PN code and the PN codes given to different users are *almost* uncorrelated. Figure 14.8 illustrates the principle of CDMA based on DS spread spectrum.

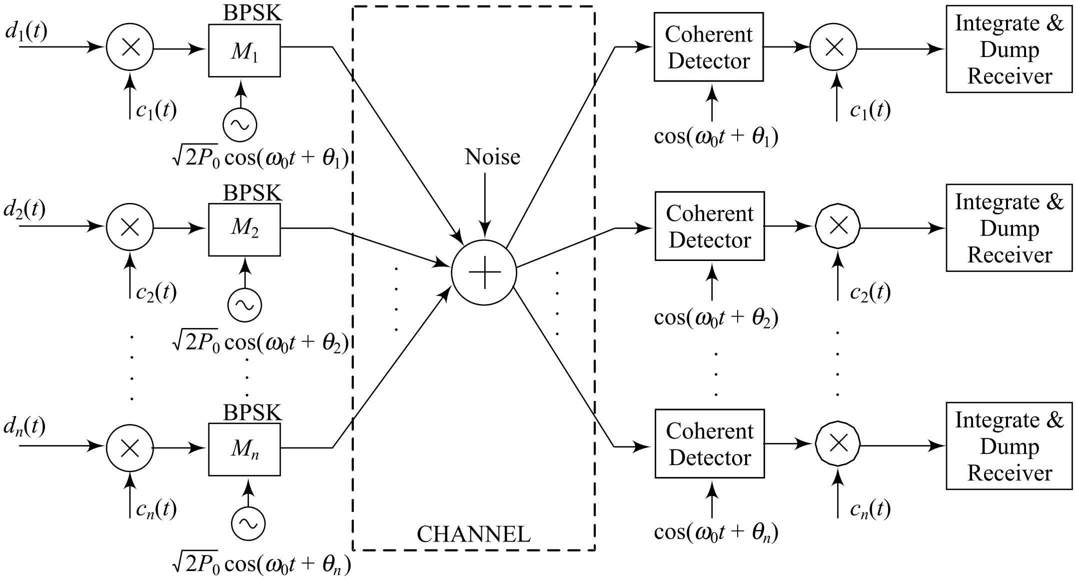


Fig. 14.8 *A DSSS-based CDMA system*

There are n users, each one transmitting data using a DS spread spectrum BPSK system; and all of them use the same carrier frequency f_0. Since there is no frequency division, or time division and since all the DSSS signals will be present at the input of each of the receivers, multiple access interference (MAI) exists at each of the receivers, i.e., at the output of any given receiver, there will be some interference caused by the remaining (n – 1) users. We will now try to analyze and see how this interference affects the probability of error. To facilitate this analysis, we will make the following simplifying assumptions:

1. The chip frequencies are the same for all the n systems.
2. The transmitted powers are the same for all the systems.
3. The data rates $f_b = 1/T_b$ are the same for all the systems.
4. Thermal noise introduced by the channel is not taken into account
5. The power presented by each DS SS signal at its receiver input is the same for all the receivers (i.e., the near-far problem is not considered).
6. The random phases of the n carrier signals are statistically independent.

Let f_0 Hz be the common carrier frequency, $d_i(t)$ be the data transmitted and $c_i(t)$ be the PN sequence signal of the i^{th} user, and P_0 be the power presented by each signal at the input of its receiver. The data rate is f_b for all users and let θ_i be the random phase of the carrier of the i^{th} user. Then the signal present at the input of each receiver is given by

$$z(t) = \sum_{i=1}^{n} \sqrt{2P_0}\, c_i(t) d_i(t) \cos(\omega_0 t + \theta_i) \tag{14.24}$$

At the k^{th} receiver, as shown in Fig. 14.8 the received signal $z(t)$ is multiplied by $c_k(t)$ for de-spreading. Then, the resultant BPSK signal is coherently detected by multiplying it by $\sqrt{2}\cos(\omega_0 t + \theta_k)$ to get the signal $v(t)$ which is then applied to the integrate and dump circuit, the output variable from which is finally applied to the decision device. Because the integrator acts like a low pass filter, not all the components present in $v(t)$ will be able to reach the output of the integrator. The signal $v_0(t)$ at the output of the integrator is

$$\left.\begin{aligned} v_0(t) &= \sum_{i=1}^{n} \sqrt{P_0}\, c_i(t) c_k(t) d_i(t) \cos(\theta_i - \theta_k) \\ &= \sqrt{P_0}\, d_k(t) + \sum_{\substack{i=1 \\ i \neq k}}^{n} \sqrt{P_0}\, c_i(t) c_k(t) d_i(t) \cos(\theta_i - \theta_k) \end{aligned}\right\} \tag{14.25}$$

In writing the last equation, an assumption has been made that all the PN sequence waveforms, $c_i(t)$s make transitions at the same time. It is then possible to put the last equation in the following form:

$$v_0(t) = \sqrt{P_0}\, d_k(t) + \sum_{\substack{i=1 \\ i \neq k}}^{n} \sqrt{P_0}\, c_{k,i}(t) \cos(\theta_{k,i}) \tag{14.26}$$

Note that this equation has the same form as Eq. (14.13), the only difference being that in Eq. (14.26) there are (n – 1) interfering signals whereas in Eq. (14.13) there is only one. The PSD of one interfering signal was found to be given by Eq. (14.17) as

$$S_0(f) = P_j T_c / 4; \quad |f| \le \frac{1}{T_b} \tag{14.27}$$

Hence in the present case, where there are (n – 1) statistically independent interfering signals, the total PSD of all these interferers is

$$S_{0T}(f) = \frac{(n-1)P_0 T_c}{4}; \quad |f| \le \frac{1}{T_b} \tag{14.28}$$

The probability of bit error is then obtained by substituting for P_j in Eq. (14.19) making use of Eqs. (14.27) and (14.28). From Eq. (14.27), we have

$$P_j = S_0(f) \cdot \frac{4}{T_c}$$

we have to replace $S_0(f)$ in this by $S_{0T}(f)$ because we have not one, but $(n - 1)$ interferers, which are having a total PSD of $S_{0T}(f)$ given by Eq. (14.28)

$$\therefore \qquad P_j = \frac{S_{0T}(f) \cdot 4}{T_c} = \frac{(n-1)P_0 T_c \cdot 4}{4 \cdot T_c} = (n-1)P_0 \qquad (14.29)$$

Substituting this in Eq. (14.19), we get

$$P_e = \frac{1}{2} erfc \sqrt{2\left(\frac{P_0}{P_j}\right) \cdot \left(\frac{T_b}{T_c}\right)} = \frac{1}{2} erfc \sqrt{2\left(\frac{1}{n-1}\right)\left(\frac{T_b}{T_c}\right)}$$

$$\therefore \qquad P_e = \frac{1}{2} erfc \sqrt{2\left(\frac{1}{n-1}\right)\left(\frac{T_b}{T_c}\right)} \qquad (14.30)$$

Since the complementary error function is a monotonically decreasing function of its argument, in order to minimize the probability of bit error, we have to maximize the quantity under the square-root sign in Eq. (14.30) by making

$$\frac{f_c}{f_b} >> (n-1)/2 \qquad (14.31)$$

Before concluding the discussion on CDMA, we would like to point out that one of the important assumptions we made in connection with CDMA systems was that the PN codes given to different users are almost uncorrelated. For this to be true it is necessary that the cross-correlation between the PN sequences assigned to any two users should be zero for all cyclic shifts. But, unfortunately, ordinary PN sequences do not satisfy this requirement.

However, there is a special class of PN sequences, called the Gold Sequences which possess excellent cross-correlation properties. Gold codes are briefly discussed in Section 14.7.

Example 14.3 In a DSSS/BPSK-based CDMA system with a processing gain of 40 dB and a probability of error $P_e = 10^{-7}$ for each user, how many users can be accommodated if all the users share equal power?

Solution From Eq. (14.30), we have

$$P_e = \frac{1}{2} erfc \sqrt{2\left(\frac{1}{(n-1)}\right)\left(\frac{T_b}{T_c}\right)} = Q\sqrt{4\left(\frac{1}{(n-1)}\right)\left(\frac{T_b}{T_c}\right)}$$

But P_e is given to be 10^{-7}. So, using the Q-function tables, we find that

$$\sqrt{4\left(\frac{1}{(n-1)}\right)\left(\frac{T_b}{T_c}\right)} = 5.2$$

$$4\left(\frac{1}{(n-1)}\right)\left(\frac{T_b}{T_c}\right) = 27.04$$

But it is given that $G_p = 40$ dB $= 10^4$ (ratio)

and we know that in a DSSS system, $G_p = \left(\frac{f_c}{f_b}\right)$

$$\therefore \qquad 4\left(\frac{1}{(n-1)}\right)\frac{1}{10^{-4}} = 27.04$$

Solving for n, we get

$$n = 1482$$

Example 14.4 In a DSSS-CDMA system, the data rate $f_b = 6$ kbps and the chip rate $f_c = 12$ Mb/s. What is the jamming margin if an output *SNR* of 10 dB is required for a $P_e = 10^{-5}$? Assume a system loss of 1.5 dB owing to imperfections in tracking and detection.

Solution The processing gain $G_p = \frac{f_c}{f_b} = \frac{12 \times 10^6}{6 \times 10^3} = 2000$

$\therefore \quad G_p$ in dB $= 10 \log_{10} 2000 = 33$ dB

$\therefore$ Jamming margin (J.M.) is given by

$$\underset{(\text{in dB})}{J \cdot M} = G_p - [\text{system loss in dB} + (S/N)_0]$$

$$= 33 - 1.5 - 10 = 21.5 \text{ dB}$$

14.5.2 Ranging Using DS Spread Spectrum Signals

To see how a DS spread spectrum signal may be used for ranging, consider a DS spread spectrum signal:

$$s(t) = \sqrt{2P_0}\,c(t)\cos(\omega_0 t + \theta) \tag{14.32}$$

For ranging, $s(t)$ is transmitted. After it impinges on the intended target, a part of the reflected signal is received. The received signal $r(t)$ may be represented as

$$r(t) = \alpha s(t - 2T) = \alpha\sqrt{2P_0}\,c(t - 2T)\cos \omega_0 (t - 2T)$$

This signal is now correlated with the chip signal $c(t)$ used in the transmitter, delayed by an adjustable and accurately known delay of τ sec, i.e., a signal $c(t - \tau)$. The output of the correlator is

$$R(\tau) = \int_0^{NT_c} c(t - 2T)c(t - \tau)dt \tag{14.33}$$

where NT_c is the total length of the PN sequence $c(t)$.

τ is adjusted till $R(\tau)$ takes a maximum value. In fact, $R(\tau)$ takes a maximum value when $\tau = 2T$ and will be negligibly small if τ is greater than $(T_c + 2T)$. If v is the velocity of electromagnetic waves ($v = 3 \times 10^8$ meters/sec).

$$\text{Range of the target} = d = \frac{v}{2}\tau = \left(\frac{1}{2} \times 3 \times 10^5 \tau\right) \text{km}$$

The accuracy of measurement is obviously dependent upon T_c. Smaller the value of T_c relative to T, the better is the accuracy.

Example 14.5 A DSSS system used for range measurement is required to give a range resolution of 0.01 km. Find the chip rate that is to be used.

Solution In a DSSS-based range measuring system, the precision of measurement = ±T_c = Chip period.

∴ if v km/sec is the velocity of light, this precision corresponds to a range resolution of vT_c km.

$$\therefore \qquad vT_c = 0.01$$

Substituting 3×10^5 km for v and solving for $(1/T_c)$, the chip rate, we get

$$\left(\frac{1}{T_c}\right) = \text{chip rate} = \frac{v}{0.01} = \frac{3 \times 10^5}{0.01} = 3 \times 10^7 \text{ Hz}$$

$$\therefore \qquad \text{chip rate} = 30 \text{ MHz.}$$

14.6 FREQUENCY HOPPING SPREAD SPECTRUM (FHSS) SYSTEMS

We have found, while discussing the resistance to jamming of a DS spread spectrum system that it depends primarily on the processing gain, G_p. The processing gain is the ratio of the chip frequency f_c to the bit frequency f_b. So, for a given data rate, the resistance to jamming of a DS spread spectrum system can be improved only by increasing the chip-rate relative to the data rate. Beyond a certain limit, practical difficulties in the design and implementation of PN sequence generators make it difficult to achieve very high chip frequencies and this puts a limit on the processing gain and the degree of resistance to jamming that can be achieved using DS spread spectrum systems.

One way of overcoming the above difficulty is to use frequency hopping spread spectrum. In this also, just like in DS spread spectrum systems, the binary digital data modulates a carrier using a traditional modulation scheme like M-ary FSK (For reasons to be discussed later, modulation schemes requiring coherent detection, like the PSK, QPSK, etc., cannot be used). This M-ary FSK modulated signal is then modulated a second time by another carrier frequency, but this carrier frequency changes its value, or rather hops, at regular intervals of T_c, the chip period, from one value to another from among a given set of values, according to a predetermined, pseudo-random pattern. This carrier frequency hopping is controlled at the transmitter by a pseudorandom code generator, as shown in the block diagram of Fig. 14.9.

As shown in Fig. 14.9(a), binary data $d(t)$ is first used to produce an M-ary FSK modulated signal. This is again modulated by a carrier produced by a frequency synthesizer that is controlled by a PN-code generator. This modulation is performed by feeding the M-ary FSK signal as well as the output of the frequency synthesizer to a mixer. The mixer produces the sum frequency and difference frequency. The BPF that follows the mixer selects only the sum frequency signal, which is the FH/MFSK signal.

At the receiving end, the received signal is fed to the mixer to which the output of a frequency synthesizer is also given. The frequency of the signal produced by this synthesizer is controlled by a PN code generator which is identical to, and is in synchronism with, the one at the transmitter. The set of frequencies produced by the frequency synthesizer and their hopping pattern, are also exactly identical to those at the transmitter. The de-spread M-ary FSK signal coming out from the BPF is then detected using a *non-coherent* M-ary FSK detector. The reason for using non-coherent detection, in spite of its poorer performance as compared to a coherent detector, is the fact that frequency synthesizers cannot maintain phase coherence over successive hops. In fact, that is the reason why FHSS systems do not make use of phase dependent modulation schemes like BPSK, M-ary PSK, QPSK, etc. which require coherent detection. The output of the M-ary FSK non-coherent detector is then fed to the decision device (not shown in figure). Although the

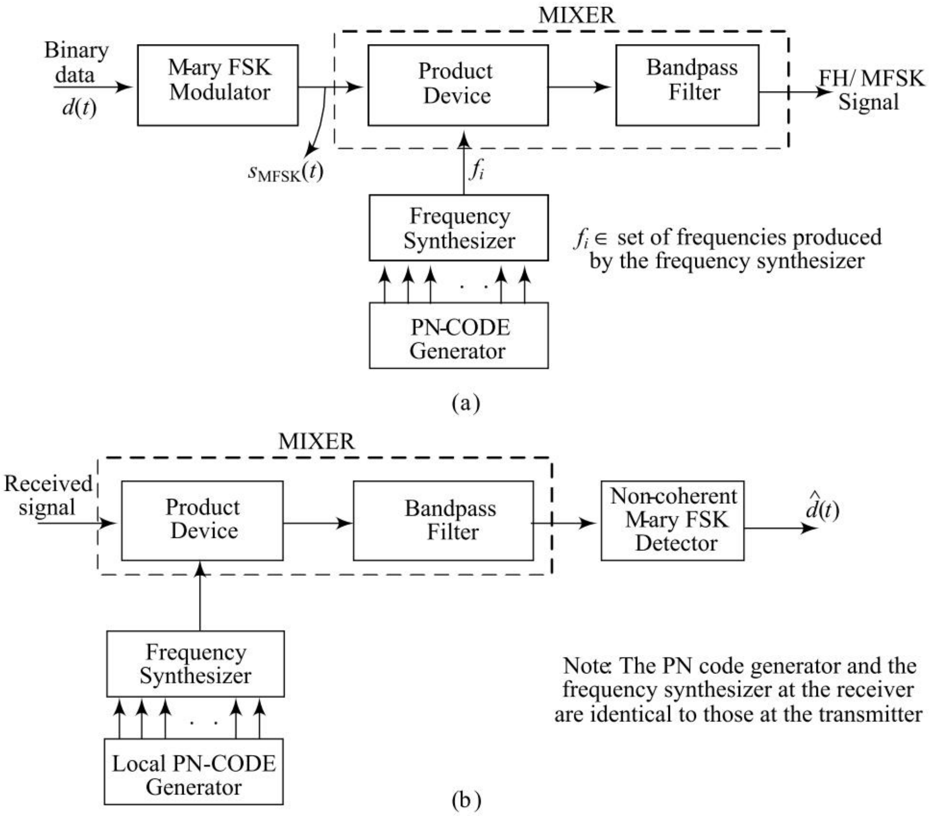


Fig. 14.9 *FHSS/M-ary FSK (a) Transmitter, (b) Receiver*

non-coherent detector of an FHSS system will have a poorer performance than the coherent detector that we can use in a DSSS system, generally, the higher processing gain attainable in FHSS systems more than compensates for the poor performance of the non-coherent detector.

14.6.1 FHSS Signal and Spreading Factor

Let T_s denote the symbol period for the M-ary FSK modulation. Let f_0 denote the unmodulated carrier frequency of the M-ary FSK modulation. Then the M-ary FSK angular frequencies are:

$$\omega_k = \omega_0 \pm \frac{1}{2}\Delta\omega, \omega_0 \pm \frac{3}{2}\Delta\omega, \omega_0 \pm \frac{5}{2}\Delta\omega, \ldots, \omega_0 \pm \frac{(M-1)}{2}\Delta\omega \tag{14.34}$$

and the M-ary FSK modulated signal itself can be written as

$$s_{\text{MFSK}}(t) = \sqrt{2P_0}\cos(\omega_k t + \phi_k) \quad \text{for} \quad kT_s \le t \le (k+1)T_s \tag{14.35}$$

After each chip period T_c, the frequency synthesizer output hops to a new value. If we denote the output signal frequency of the frequency synthesizer to be f_i during a given chip period, then, *during that chip period*, we may represent the FH/MFSK signal as

$$s_{\text{FH/MFSK}}(t) = \sqrt{2P_0}\cos[(\omega_k + \omega_i)t + \phi_k] \tag{14.36}$$

In the above equation, we have taken $(\omega_k + \omega_i)$ since the BPF following the mixer has been designed to give the sum frequency, as has already been mentioned. The mixing of the MFSK signal with the synthesizer output signal having a frequency of f_i increases the bandwidth occupancy. If the MFSK signal has a bandwidth W_s then we may write

$$W_c = \text{B.W. of the FH/MFSK signal} = LW_s \tag{14.37}$$

where $L > 1$, is called the *spreading factor* or *processing gain*.

All along, we have been talking about the symbol period, T_s and the chip period T_c, i.e., the period during which the frequency of the signal from the synthesizer remains at one of the fixed values out of the total number of frequencies to which it can 'hop'. As these two periods, T_s and T_c are not directly related, they may be chosen independently. This gives rise to two types of frequency-hopping spread spectrum systems:

1. **Slow-hopping FHSS:** In this, the symbol rate $R_s = 1/T_s$ of the MFSK signal is an integer multiple of the hop rate $R_c = 1/T_c$, where T_c is the chip period. Also, the bit rate R_b and the symbol rate R_s in a MFSK system are related by

$$R_s = \frac{R_b}{\log_2 M}$$

   Thus, since $T_c \geq T_s$ in slow hopping FHSS, there can be several symbols in one hop interval T_c.
2. **Fast-hopping FHSS:** If $T_c < T_s$, it is known as fast hopping. Thus, there will be multiple hops within each symbol of the M-ary FSK.

Figure 14.10 shows a typical hopping pattern.

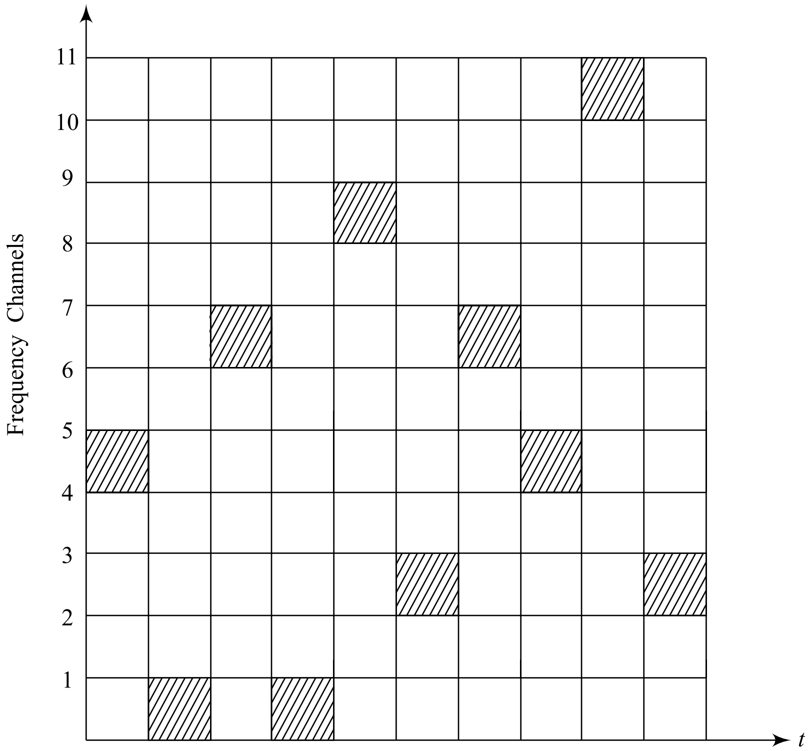


Fig. 14.10 *Typical hopping pattern*

Example 14.6 For a slow FHSS system, show that the processing gain is equal to 2^m, the number of frequency slots available for hopping.

Solution In slow FHSS, the hop rate, i.e., the chip rate f_c is less than the message bit rate f_b. Therefore, two or more baseband bits are transmitted at the same frequency.

Let Δf be the separation between adjacent frequencies of the frequency synthesizer and let B_T be the bandwidth of the modulated carrier wave (whatever be the type of digital modulation employed) before spreading. Then Δf has to be larger than or at least equal to B_T. If there are 2^m discrete frequencies among which the synthesizer frequency hops, the bandwidth of the transmitter signal (after hopping) is

$$B_{T(\text{FHSS})} = 2^m \cdot \Delta f = 2^m \cdot B_T \quad \text{(say)}$$

$$\therefore \qquad L = \frac{B_{T(\text{FHSS})}}{B_T} = G_p, \text{ the processing gain} = 2^m$$

14.6.2 Resistance to Jamming

As we did in the case of DS spread spectrum systems, for the FH spread spectrum systems too, we shall examine the effect of jamming and interferences in terms of their effect on the probability of error.

We shall focus on two cases of jamming:

1. 'White noise, or 'barrage jamming', which covers the entire bandwidth W_c of the FHSS signal.
2. Partial-band jamming in which the jamming signal power is spread over only a part of the bandwidth W_c of the FHSS signal.

White noise or barrage jamming As shown in Fig. 14.11(a), let J be the jamming power which is spread over the full bandwidth W_c of the FHSS signal.

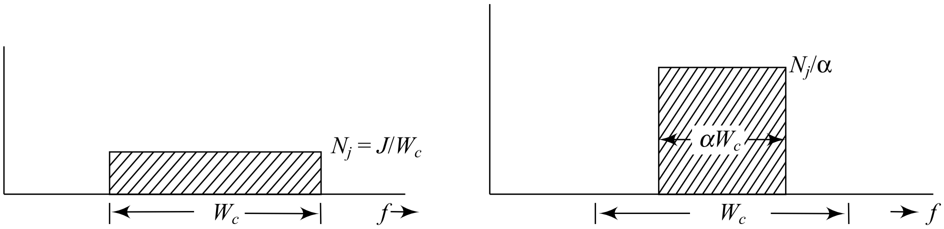


Fig. 14.11 *(a) Barrage jamming, (b) Partial band jamming*

In the barrage jamming case, the jamming signal appears as white noise. Since we are considering FHSS with M-ary FSK, the bit-error probability in the presence of white noise of two-sided PSD of $\eta/2$, is

$$P_e = \frac{1}{2} e^{-E_b/2\eta} \tag{14.38}$$

Hence, in the presence of barrage jamming signal with a PSD of N_j, the bit-error probability becomes

$$P_{ej} = \frac{1}{2} \exp\left[\frac{-E_b}{2(\eta + N_j)}\right] \tag{14.39}$$

where $N_j = J/W_c$.

Partial band jamming If α denotes the fraction of the FHSS signal bandwidth, W_c, which is covered by the jamming signal, assuming J to be the total jamming power as in the previous case, the jamming PSD is N_j/α and it has a bandwidth αW_c. So the probability of being jammed is α and the probability of not being jammed is $(1 - \alpha)$. Therefore, the probability of error may be written as

P_e = (prob. of error in case of jamming) × (Probability of jamming) + (prob. of error if there is no jamming) × (Prob. of not being jammed)

$$\therefore \qquad P_e = \frac{\alpha}{2}\exp\left[\frac{-E_b}{2(\eta + N_j/\alpha)}\right] + \frac{(1-\alpha)}{2}\exp\left[\frac{-E_b}{2\eta}\right] \tag{14.40}$$

Generally, slow-hop FHSS signal is more susceptible since one or more symbols are transmitted at the same frequency. Also, the effect of single-tone jamming will be negligible if there are a very large number of frequency slots to hop to.

Example 14.7 An FHSS/BFSK is used for transmitting binary data coming at a rate of 20 kbps. The unspread BFSK signal occupies a bandwidth of 25 kHz. The received signal power is –15 dBm. A jammer which can produce a received power of at the most –20 dBm either as a narrow band signal of 25 kHz bandwidth, or as a broadband signal occupying the full bandwidth of the FHSS system, is trying to jam the FHSS signal. If the spreading factor L of the FHSS/BFSK system is 25, find the improvement in the SNR (in dB) under broadband jamming as compared to narrowband jamming. Assume the one-sided PSD of the AWGN of the channel to be 10^{-11} W/Hz.

Solution The received power $P_R = -15 \text{ dBm} = 3.162 \times 10^{-5} W$

$$E_b = \text{bit energy} = P_R \cdot T_b = \frac{3.162 \times 10^{-5}}{20 \times 10^3} = 1.581 \times 10^{-9}$$

Jamming power = –20 dBm = 10^{-5} $W = P_j$

In the case of narrow band jamming, this jamming power is spread over a bandwidth of 25 kHz.

So the power spectral density of the narrowband jammer at the receiver =

$$\frac{P_j}{25 \times 10^3} = \frac{10^{-5}}{25 \times 10^3} = 0.4 \times 10^{-9} = \eta_{\text{JNB}}$$

∴ for narrow band jamming, the *SNR* is given by

$$\frac{E_b}{\eta_{\text{JNB}} + \eta} = \frac{1.581 \times 10^{-9}}{0.4 \times 10^{-9} + 10^{-11}} = \frac{1.581 \times 10^{-9} \times 10^{-11}}{41} = 5.86 \text{ dB}$$

Since L = 25 = spreading factor of the FHSS system, and since the bandwidth of the BFSK signal before spreading is 25 kHz, the bandwidth of the FHSS/BFSK signal is

$$\underset{\text{(FHSS/BFSK)}}{B_T} = 25 \times 25 \text{ kHz} = 625 \text{ kHz}$$

∴ the PSD of the jamming signal at the receiver in the case of wideband jamming is

$$\eta_{\text{JWB}} = \frac{P_j}{625 \times 10^3} = \frac{10^{-5}}{625 \times 10^3} = 1.6 \times 10^{-11}$$

∴ *SNR* for the wideband jamming case is

$$\frac{E_b}{\eta_{\text{JWB}}+\eta}=\frac{1.581\times10^{-9}}{1.6\times10^{-11}+10^{-11}}=0.608\times10^{2}=60\text{ (ratio)}$$

∴ $$\left(\frac{E_b}{\eta_{\text{JWB}}+\eta}\right)_{\text{dB}}=10\log_{10}60=17.78\text{ dB}$$

∴ improvement in SNR is (17.78 – 5.86) = 11.92 dB

14.6.3 CDMA with FHSS

As in DS spread spectrum, multiple access is achieved in FHSS also by assigning a unique PN code to each user, which in this case controls the frequency hopping pattern. These codes that are assigned, must be so chosen that *collisions* do not occur. Recall that the frequency produced by the frequency synthesizer during a chip period depends on the PN sequence values during that chip period. So sometimes it may so happen that two or more users have, at a given time, the same PN sequence values produced by their respective PN sequence generators. In that case, a collision is said to have occurred in the spectrum.

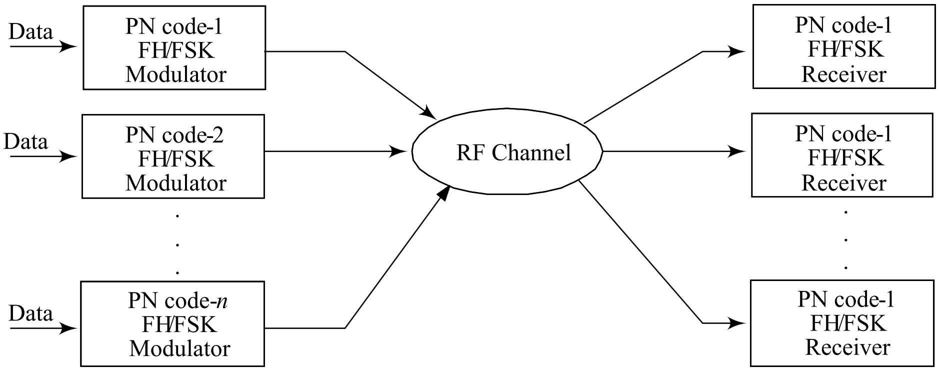


Fig. 14.12 *CDMA with FHSS*

Whether it is a slow hopping FHSS or a fast hopping FHSS, when a collision occurs, it results in considerable increase in detection errors. In slow hopping FHSS, a collision in a particular hop, will result in several consecutive symbols being erroneous since there will be several M-ary symbols in the case of M-ary FSK and several bits in the case of binary FSK during that hop period. Such burst type of errors can be corrected by employing FEC using Reed-Solomon error-correcting codes.

14.6.4 Applications of Spread Spectrum Systems

Because of its ability to reject narrow band as well as broad band jamming, FHSS systems find applications in military communications. FHSS further finds extensive use in bluetooth.

Since a DS spread spectrum signal consists of a sequence of extremely short pulses and is therefore capable of giving very good accuracies in range measurements. DSSS systems find extensive use in Global Positioning Systems (GPS). Further, the use of spread spectrum in GPS permits their use at reasonable power levels because of the processing gain G_p.

Synchronization In spread spectrum communication systems, for satisfactory operation, there should be perfect alignment between the transmitted and received PN codes. Further, if coherent detection is needed then the locally generated carrier at the receiver must be in frequency and phase synchronism with the carrier at the transmitter. It must be remembered that the carrier frequency as well as the PN clock may drift with time. Further, if there is relative motion between the transmitter and receiver, as happens in the case of mobile and satellite spread spectrum systems, the carrier and PN clock will suffer Doppler frequency shift too.

Insofar as synchronization of carrier frequency and phase are concerned, the techniques adopted for coherent reception in the case of conventional analog and digital communication are replicated. Synchronization of the PN sequence of the receiver with that of the transmitter proceeds in two steps – *acquisition* and *tracking*. Acquisition is nothing but initial coarse alignment and this process tries to bring the receiver PN sequence in alignment with that of the transmitter to within a fraction of a chip period. Once the acquisition part is completed, fine alignment has to be done and that process is referred to as tracking. Both acquisition and tracking make use of feedback loop.

Acquisition

1. For DS spread spectrum:

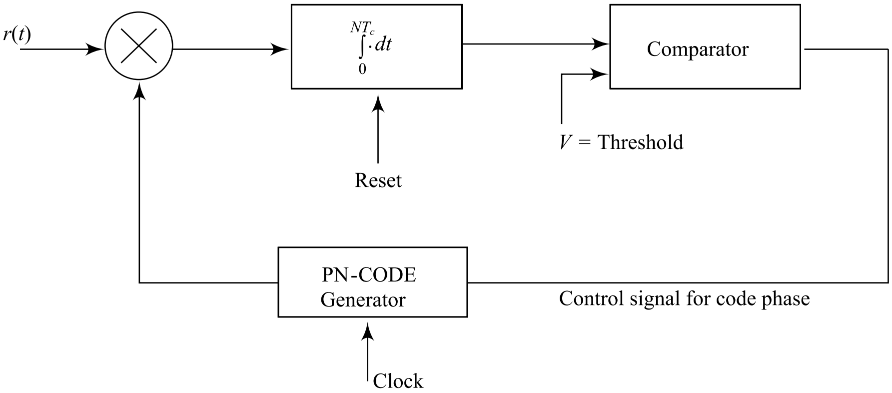


Fig. 14.13 *DS spread spectrum synchronization (Acquisition)*

Let the PN code used by the transmitter have N chips so that its total duration is NT_c where T_c is the chip period. An exact replica of the PN code used at the transmitter, which is available at the receiver, is used to multiply the received signal $r(t)$. The product is then integrated over a period of NT_c, i.e., the total duration of the code. If the codes are in exact alignment, complete de-spreading will take place and the output of the integrator will be rather large compared to the threshold V and so the phase of the PN code generator will remain unaltered. But, in case the alignment between the two codes is not correct, the control signal from the comparator advances the phase of the PN code by half a chip. This way the process continues till the alignment of the two codes is correct to within half a chip.

Thus, if initially the misalignment between the two codes was n chips, since each time the integration is performed for NT_c seconds and it results in half-a-chip correction, the total time taken for acquisition is given by

$$T_{\text{acq}} = 2nNT_c \text{ sec} \tag{14.41}$$

In the above discussion, of course, we have totally ignored the presence of additive noise present along with $r(t)$.

2. For FH spread spectrum systems: The acquisition circuit for FH spread spectrum is shown in Fig. 14.14. Note that the PN code generator and the frequency synthesizer of Fig. 14.14 are exactly identical to those at the transmitter.

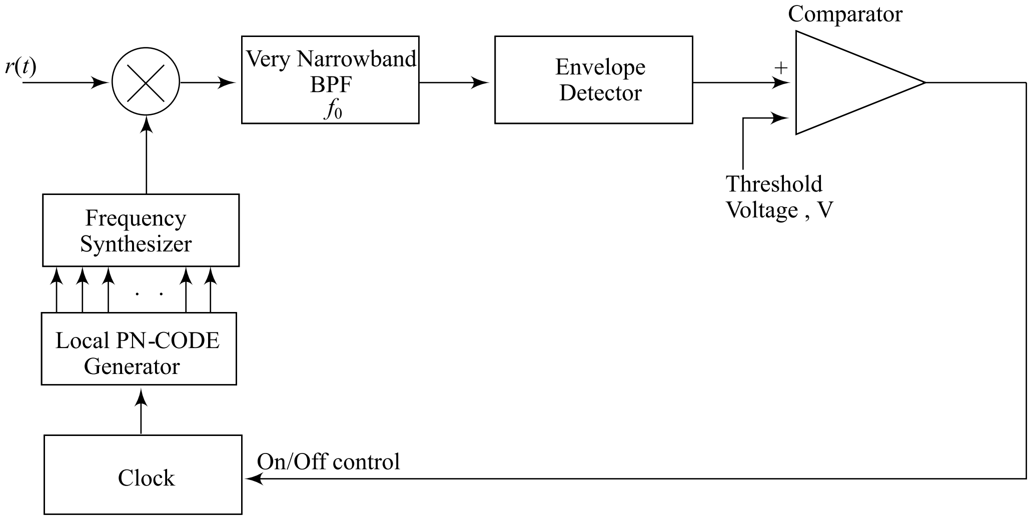


Fig. 14.14 *'Camp-and-wait' type of acquisition circuit for FHSS signals*

Let $r(t)$, the received FHSS signal have a frequency of $f_0 + f_i$ where f_0 is the carrier frequency of the first modulation and f_i is the frequency of the frequency synthesizer, at the transmitter. At the same time, suppose f_j is the frequency of the signal produced by the frequency synthesizer in the acquisition circuit of the receiver and let us assume that $f_j \neq f_i$. Since the received signal $r(t)$ and this signal from the frequency synthesizer of the acquisition circuit are multiplied, the product will have two components, the sum frequency component and the difference frequency component. The sum frequency component having a frequency of $f_0 + (f_i + f_j)$ will be rejected by the very narrow band BPF centered on f_0. The difference frequency component with a frequency of $f_0 + (f_i - f_j)$ will be able to produce only a small voltage at the output of the BPF as $f_j \neq f_i$. Since this will be less than the threshold voltage V, the output voltage of the comparator under these conditions, is arranged to turn off the controlled oscillator producing the clock pulses which are applied to the PN code generator. So the state of the code generator does not change and hence the output signal of the frequency synthesizer will remain or 'camp' at the same frequency f_j. The frequency synthesizer at the transmitter, however, continues to go on hopping from one frequency to another. Since f_j is also among the frequencies it will 'hop' to, at some later instant of time, it will hop to that value and the received signal frequency will then be $f_0 + f_j$. Then, this and the output signal from the frequency synthesizer of the acquisition circuit, which is still at the frequency f_j, will mix and the difference frequency, which is now f_0, will produce a large voltage at the output of the BPF. Since this will be larger than the threshold voltage V, the on/off control voltage will be such as to 'turn on' the clock driving the PN code generator. Then the frequency synthesizer in the acquisition circuit will start 'hopping' again. Even now, the PN code generator in the acquisition circuit may not be exactly in alignment with that in the transmitter but the misalignment will be within $\pm\frac{1}{2}T_c$, where T_c is the chip period.

Tracking Once the acquisition or coarse alignment is over, the next step is tracking, i.e., the fine alignment.

For DS spread spectrum: Tracking for DS spread spectrum signals may be done using what is called the delay-locked loop (DLL) which is shown in Fig. 14.15.

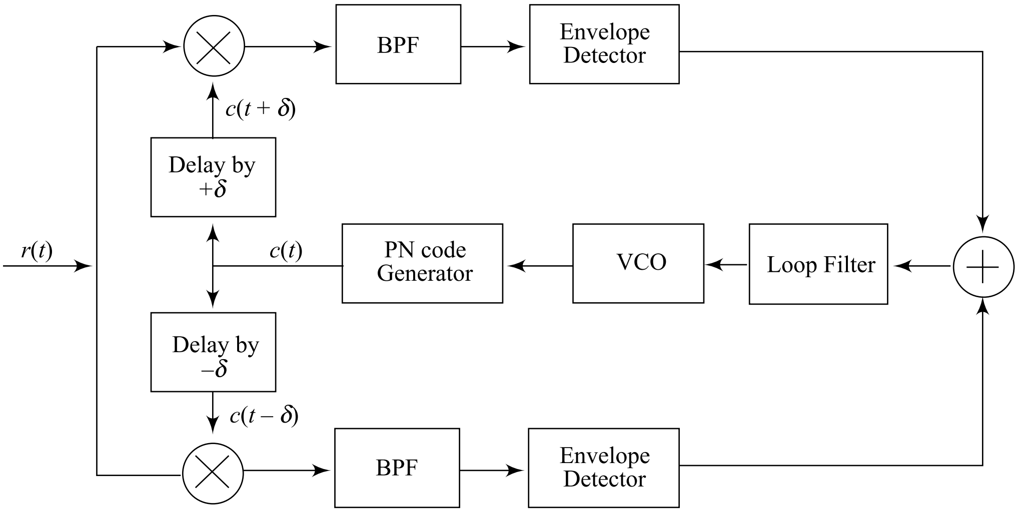


Fig. 14.15 *A delay-locked loop for DS spread spectrum signal tracking*

The received DS spread spectrum signal, $r(t)$, is applied simultaneously to two multipliers. One of the multipliers is fed with the PN code delayed by δ, a fraction of the chip interval and the other multiplier is fed with the same PN code advanced by δ. The output from each multiplier is fed to a BPF centered on f_0 and the output of each BPF is envelope detected and the envelope detector outputs are compared. The output of the summer is filtered in the loop filter and the output signal from the loop filter is used for controlling the VCO from which the clock signals for the PN code generator are derived. Here, since fine alignment is needed, a continuous control is used. In case the $c(t)$ generated in the DLL tracking circuit is in perfect alignment with the PN code used at the transmitter, the envelope detector outputs from the upper and lower arms would be equal and so the loop filter output would be zero and the phase of the PN code produced will not be changed. But if the PN code at the transmitter and the one in the DLL circuit are not perfectly aligned, the output of one envelope detector would be more than that of the other and the loop filter output, which controls the VCO, will change the clock signals in the proper direction so as to bring the two PN codes into alignment.

Since the acquisition process is already over, the PN codes at the transmitter and the receiver must have been coarsely aligned. So, the frequencies produced by their respective frequency synthesizers will be the same for most part of each hopping period as shown in (a) and (b) of Fig. 14.17. During this time, these frequencies are the same (as f_i in both, or f_j in both), the BPF output will be large, and when the frequencies are not the same, the BPF output will be very small. Actually, the bandwidth of the BPF used here in the tracking circuit is not very narrow; instead, it has a bandwidth that is sufficient to pass the data. The VCO shown in the figure is adjusted to have a nominal frequency that is equal to the frequency at which the hopping takes place, and it produces a rectangular waveform swinging between +1 V and –1 V with 50% duty cycle as shown in part (d) of Fig. 14.17. The output of the VCO is allowed to pass through the gate whenever there is output from the envelope detector and is blocked whenever the output of the envelope detector is zero as happens during those periods when the output frequencies of the synthesizers at the transmitter and the receiver are not the same. The result of this gate operation is that when the two frequencies are not the same, the input to the LPF will not be the symmetrical rectangular waveform of the VCO, instead, the durations of its positive swing and negative swing will be unequal therefore LPF output will not be zero. On the other hand, if the PN code at the transmitter and the receiver are exactly aligned, the gate allows the VCO waveform all the time and so the input to the LPF will be having equal durations for its positive and negative swings.

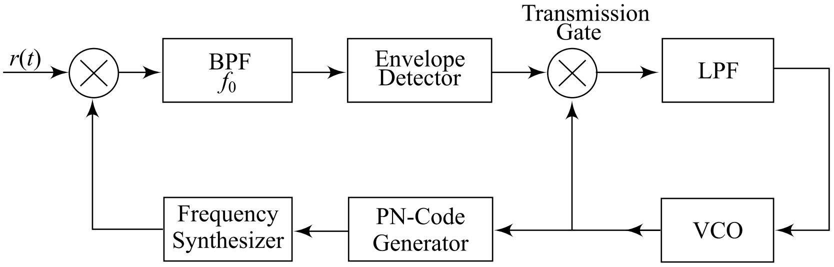


Fig. 14.16 *Early-Late Gate circuit for tracking FHSS signal*

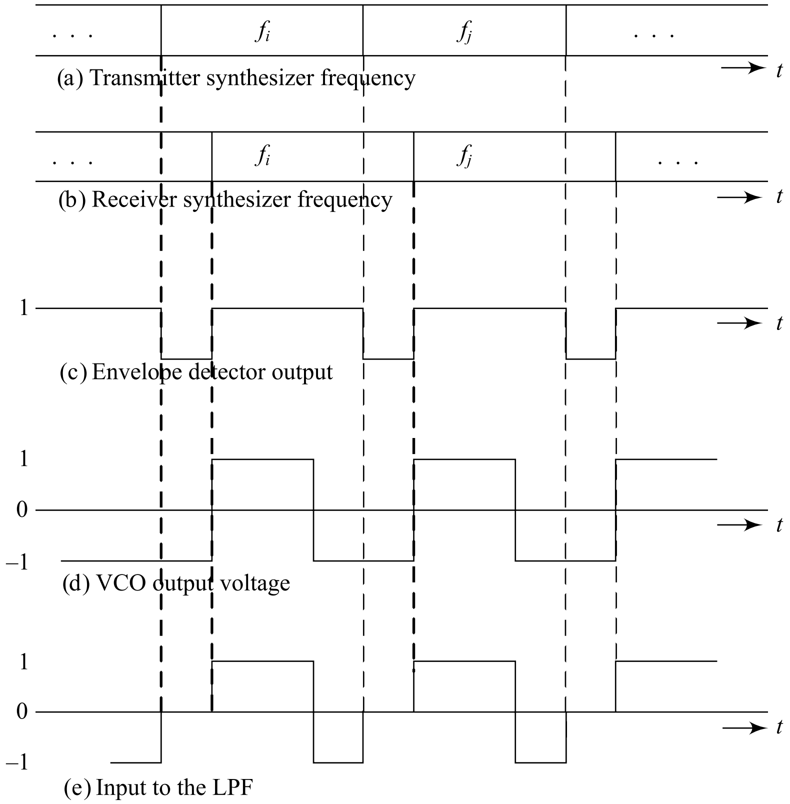


Fig. 14.17 *Early-Late gate circuit waveforms*

So in this case, the LPF output will be zero. Whenever the LPF output is not zero, the VCO, which is the clock waveform generator, will adjust in such a way as to make the LPF output equal to zero, i.e., to make the receiver PN code to get aligned with that at the transmitter.

14.7 GOLD CODES

As we have seen in Section 14.5 during our discussion on CDMA using spread spectrum, PN sequences used for CDMA applications are required to have certain desirable characteristics.

If $x(t)$ and $y(t)$ are two PN sequences, these properties are:

1. $[R_{xx}(0)/R_{xx}(\tau)]_{\tau\neq 0}$ should be as large as possible.
2. $|R_{xy}(\tau)|_{\max}$ should be as small as possible for all τ.
3. $[|R_{xy}(\tau)|_{\max}/R_{xx}(0)]$ should be small.

In the context of the above requirements, we find that

1. PN sequences generated by a single shift register do not have good cross-correlation properties so are not quite suitable for CDMA applications.
2. To generate different PN sequences with a single shift register, we have to change the feedback connection each time.
3. A shift register of given length gives very few unique output sequences.

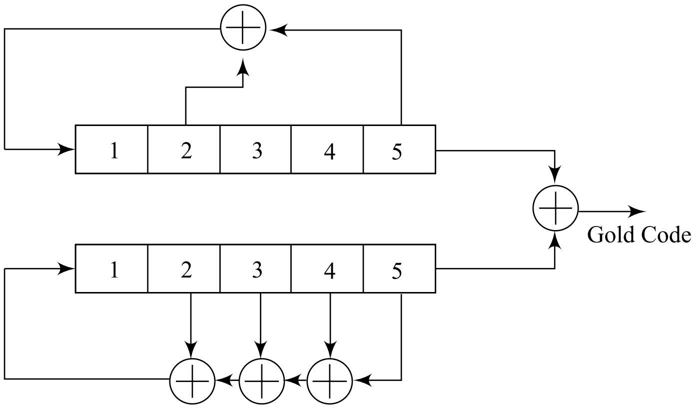


Fig. 14.18 *Gold code generator*

For Gold codes, which are generated by the modulo-2 addition of the outputs of a few m-bit shift registers (generally only two) with certain feedback connections, the cross-correlation between $N = (2^m - 1)$ length sequences has a maximum value bounded by

$$R_{xy} = \phi/N \tag{14.42}$$

where

$$\phi = \begin{cases} 2^{(m+1)/2} + 1 & ; \quad m \text{ odd} \\ 2^{(m+2)/2} + 1 & ; \quad m \text{ even} \end{cases} \tag{14.43}$$

or in some cases

$$R_{xy} = -\frac{1}{N} \tag{14.44}$$

Table 14.3 gives the preferred tap connections for a pair of shift registers and the $[|R_{xy}(\tau)|_{\max}/R_{xx}(0)]$ for the Gold codes generated.

Table 14.3 *Preferred pairs and connections for Gold codes generation*

m	Preferred tap connections	$\lvert R_{xy}(\tau)\rvert_{\max}/R_{xx}(0)$
5	(5, 2) (5, 4, 3, 2)	0.290
7	(7, 3) (7, 3, 2, 1)	0.134
8	(8, 7, 6, 5, 2, 1) (8, 7, 6, 1)	0.129
10	(10, 8, 5, 1) (10, 7, 6, 4, 2, 1)	0.064
12	(12, 9, 8, 5, 4, 3) (12, 7, 6, 4)	0.031

Gold codes generated using preferred pairs and connections, have very good cross- correlation properties although their auto-correlation values for $\tau \neq 0$ are slightly more. Further, since we get a different periodic sequence for each set of initial conditions of the registers, the two m-bit preferred register pairs will give us $(2^m - 1)$ unique output sequences. If the original sequences also are counted, we get altogether $(2^m + 1)$ unique sequences, whereas a single register configuration can give only one unique periodic output sequence.

Gold codes are very useful for CDMA applications because of their attractive cross-correlation property and the possibility of obtaining a large number of unique sequences.

For more details on Gold codes, the reader is suggested to read J.K. Holmes, *Coherent Spread Spectrum Systems*, Wiley, 1982.

MATLAB Example 14.1 Although all analysis in the book is restricted to AWGN channels, practical channels are often frequency selective (its effective impulse response is not an impulse). Wireless channels often experience deep fades. In this example, we will demonstrate how spread spectrum signal can be used to deal with frequency selective channels in a simple manner. In such channels, the signal is the weighted sum of the current and the previous samples. One implication is that the received pulse shape is modified by the channel. With chip-rate sampling, the first few chips might be contaminated by the previous symbol. To ensure simple processing, we omit these chips in the receiver, and then apply a matched filter to the modified pulse. Assuming BPSK signaling, and the same chip sequence as used in Fig. 8.2, the BER is depicted in Fig. 14.19 as a function of the average BER. We observe that simply ignoring the contaminated chips makes the receiver simple, and yields good BER performance.

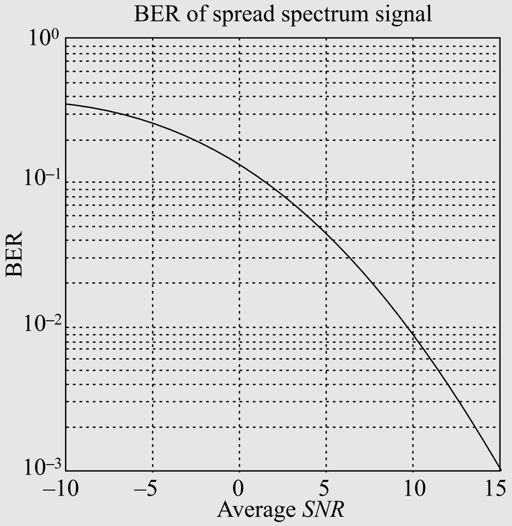


Fig. 14.19

MATLAB code:

```
c = [-1,1,1,1,-1,-1,1]; %chip sequence
N = length(c); % Processing gain
e = [sqrt(3)/2,1/2]; % average gains of the impulse response points
snr_array = [-10:1:15]; %array of SNRs that will be plotted
```

```
num_realizations=10^6; % a large number of realizations are used for better
averaging
noise_var_array = 10.^(-snr_array/20);
for snr_val=1:length(noise_var_array)
noise_level = noise_var_array(snr_val); %noise level is set to vary SNR with
signal variance fixed
error=0;
for realization=1:num_realizations
d = sign(randn(1,2)); %generate a sequence of 2 symbols
x = (kron(d,ones(1,N)).*kron(ones(1,2),c))/sqrt(N);%performs multiplication of
chip sequence and symbols
h = randn(1,2)+sqrt(-1)*randn(1,2); %generates random channel coefficients
h = h.*e;
y = h(1)*x(N+2:length(x)) + h(2)*x(N+1:length(x)-1); %generated channel output
for 1 symbol
%y = awgn(y,20);
noise = randn(1,N-1)+sqrt(-1)*randn(1,N-1); %generate noise
y = y + noise_level*noise; % add noise of the proper level
%We assume h(1) and h(2) are known
%modified chip sequence is now g = h(1)*c + h(2)*[c(7),c(1:6)];
%We use a filter matched to g as can be expected
g = h(1)*c + h(2)*[c(7),c(1:6)];
g = g/norm(g,2);
%However, we ignore the first chip as discussed earlier
%matched filter output after, sampling becomes
z = y*g(2:7)’;
%output of decision device is now
decision = sign(real(z));
if (decision~=d(2))
 error = error+1;
end
end
BER(snr_val) = error/num_realizations;
end
semilogy(snr_array,BER)
title(‘BER of Spread Spectrum Signal’)
ylabel(‘BER’)
xlabel(‘Average SNR’)
axis(‘square’)
```

MATLAB Example 14.2 Gold Codes

Take two 12-bit shift registers and the mod-2 combination of their outputs to obtain Gold sequences by giving the preferred pair of feedback connections to them, viz. (12, 7, 6, 4) and (12, 9, 8, 5, 4, 3). Find the cross-correlation to auto correlation ratio and plot the auto-correlation and cross-correlation.

MATLAB Program

```
%Gold code sequence generator
% The m-file uses two preferred pairs of m-sequences (length 2^n-1)
% chips long where n=12 with length 4095. The 4097 gold codes
%(two original m-sequences plus 2^n-1 Gold codes) produced in a 4095x4097 matrix
```

```
% Here (with n=12)
% The preferred pair used here is (12,7,6,4) and (12,9,8,5,4,3).
% The m-file will also check the cross-correlation values between any two codes
in % the matrix and should be three valued. The auto-correlation of a Gold code
can % also be checked.
clc
clear all
% Generate 1st m-sequence (12,7,6,4) of length 4095 (1+x^4+x^6+x^7+x^12)
n = 12;
N = 2^n-1;
k = 80
register1=ones(1,n); %initial fill
code1=zeros(1,N);
for i=1:N
  temp = mod(register1(4)+register1(6)+register1(7)+register1(12),2);
  code1(i) = 2*register1(n)-1;
  for j=n:-1:2
    register1(j)=register1(j-1);
  end
  register1(1) = temp;
end
% Generate 2nd m-sequence (12,9,8,5,4,3) of length 4095 (1+x^3+x^4+x^5+x^9+x^9+x^12)
register2= ones(1,n);  %initial fill code2=zeros(1,4095);
for i=1:N
  temp = mod(register2(3)+register2(4)+register2(5)+register2(8) +register2(9)+register2(12) , 2);
  code2(i) = 2*register2(n)-1;
  for j=n:-1:2
    register2(j)=register2(j-1);
  end
  register2(1) = temp;
end
m_sequence_1=code1;  %1/-1(bipolar sequence) output
m_sequence_2=code2;   %1/-1(bipolar sequence) output
%
%
m_sequence_1=m_sequence_1';  %transpose to a column
m_sequence_2=m_sequence_2';   %transpose to a column
%
%
% Generate a set of unique Gold codes in a matrix (4095x4097) which includes
% the original 1st and 2nd m-sequences plus (2^n - 1) = 4095 other Gold codes % with n=12.
%These unique codes are with initial fills of [1 1 1 1 1 1 1 1 1 1 1 1] in register1
% and register2. Other unique sets of Gold codes can be generated with
% different initial fill values.
%
%
m_sequence_1 =m_sequence_1>0;   %change 1/-1 to 1/0
m_sequence_2 =m_sequence_2>0;   %change 1/-1 to 1/0
%
Gold_code_matrix(:,1) = m_sequence_1;
Gold_code_matrix(:,2) = m_sequence_2;
```

```
%
for phase_shift=0:N-1
  shifted_code=circshift(m_sequence_2,phase_shift);
  Gold_code_matrix(:,3+phase_shift)=mod(m_sequence_1+shifted_code,2);
end
%
% Change matrix codes from 1/0 to 1/-1 and show 33 codes in command window.
  Gold_code_matrix=2*Gold_code_matrix-1; %change 1/0 to 1/-1
%
% Choose 2 codes from Gold code matrix and plot the cross-correlation
% values. If codes are from preferred pairs, the three values would be
%(-9,-1,+7).
%
codeA=Gold_code_matrix(:,9);
codeB=Gold_code_matrix(:,11);
%
% Determine cross-correlations
%
for shift=0:k
  shifted_code1 = circshift(codeA,shift);
  crosscorrelation(shift+1) = codeB'*shifted_code1;
end
%
  subplot(2,1,1)
  plot(crosscorrelation)
  grid on
  xlabel('shifts');ylabel('value of correlations');
  title('Cross-correlations of two codes')
  xlim([1 k]);
%
% Choose 1 code from the Gold code matrix and plot the autocorrelation values.
%
 codeC=Gold_code_matrix(:,17);
%
for shift=0:k
  shifted_code_A= circshift(codeC,shift);
  autocorrelation_1(shift+1) = codeC'*shifted_code_A;
end
%
% Show that all autocorrelation values of m-sequence 1 and 2
%(with nonzero shift) equals 31/-1.
%
subplot(2,1,2)
plot(auto-correlation_1)
grid on
xlabel('shifts');ylabel('value of correlations');
title('Gold code auto-correlation' )
xlim([1 k]);
ratio = abs(max(crosscorrelation(80)))/autocorrelation_1(1)
```

Result

```
ratio = 0.0154
```

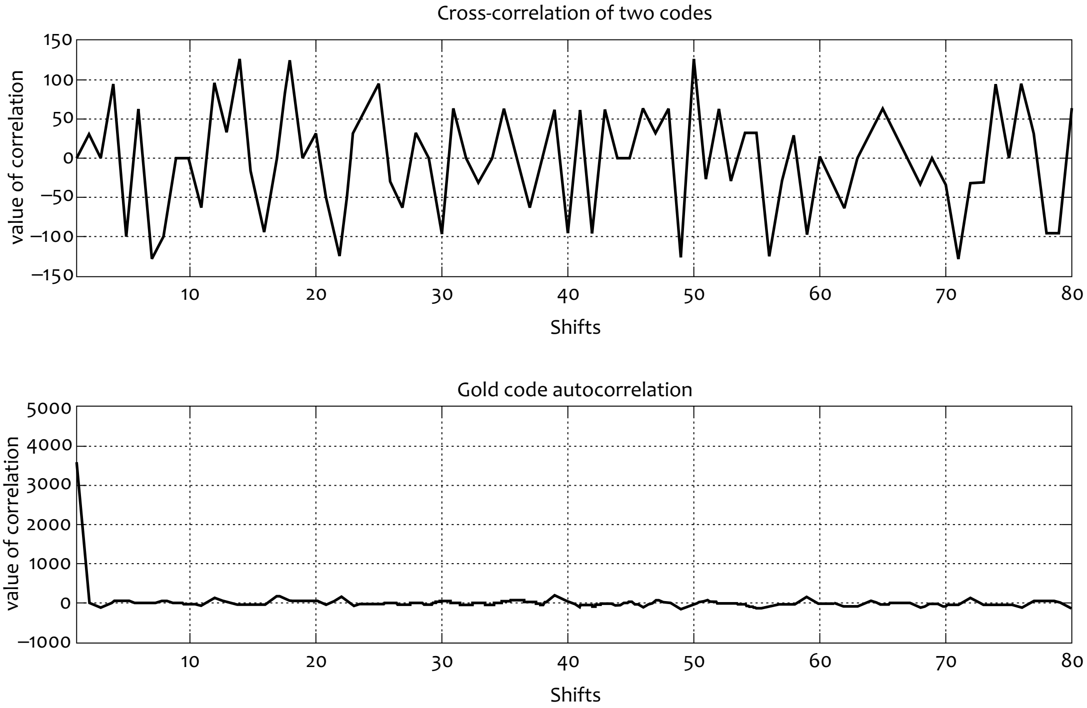


Fig. 14.20

Summary

- Spread spectrum communication provides secure and reliable communication by preventing interception and resisting jamming.
- There are basically two types of spread spectrum systems; the Direct Sequence Spread Spectrum (DSSS) and the Frequency Hopping Spread Spectrum (FHSS). Both of them relay heavily on Pseudo-random Noise sequence generators (PN code generators).
- PN code generators use shift-registers with appropriate feedback.
- An m stage shift-register with appropriate feedback can produce PN sequences of length $(2^m - 1)$. Such sequences are called maximal-length sequences.
- These PN sequences have auto-correlation functions that resemble those of white noise.
- The basic principle of DSSS is to spread the signal power over a very large bandwidth by modulating an already BPSK modulated signal, using a spreading signal, which is a maximal-length PN sequence.
- At the receiving end, the DSSS signal is first de-spread by multiplying the received signal by a PN sequence which is an exact replica of the one used for spreading at the transmitting-end and is in synchronism with it. Then the BPSK signal is detected using a coherent detector.
- If the chip frequency is f_c, the inverse of T_c, the pulse width of the PN sequence used for spreading, and f_b is the inverse of the bit rate of the data, then the ratio (f_c/f_b) is called the 'Processing Gain' of the DSSS system and it determines the degree of resistance offered by the DSSS signal to jamming.
- Both DSSS and FHSS systems provide CDMA facility.

- DSSS signals can be used for ranging and the accuracy attainable increases with the chip frequency.
- In FHSS, a BFSK or M-ary FSK signal is again modulated by a carrier signal whose frequency goes on hopping from one value to another from among a given set of values, at regular intervals of T_c, the chip period, according to a predetermined pseudo-random pattern.
- As phase coherence cannot be maintained in FHSS, only non-coherent detection is possible. Hence only BFSK or M-ary BFSK signals are used.
- In DSSS as well as FHSS, it is necessary to maintain perfect alignment between the PN sequences used at the transmitter and the receiver.
- Synchronization of the PN sequences is done in two stages. The first stage, called, 'Acquisition', achieves coarse alignment, while the second stage, called 'Tracking', achieves perfect synchronization.
- For CDMA applications, a large number of distinct PN codes must be available and they should have auto-correlation characteristics similar to those of white noise and the cross-correlation between any two must be ideally zero. Gold codes satisfy all these requirements.

References and Suggested Readings

1. Carlson, A. Bruce, Paul B. Crilly, and Janet C. Rutledge, *Communication Systems – An Introduction to Signals and Noise in Electrical Communication*, 4th Edition (International Edition), McGraw-Hill, 2002.
2. Fehr, Dr. Kamilo, *Wireless Digital Communication*, Eastern Economic Edition, Prentice-Hall of India Private Limited, 2003.
3. Haykin, Simon, *Communication Systems*, 4th Edition, Wiley India, 2001.
4. Holmes, J. K., *Coherent Spread Spectrum Systems*, Wiley, 1982.
5. Lathi, B. P., and Zhi Ding, *Modern Digital and Analog Communication Systems*, International Fourth Edition, Oxford University Press, 2010.
6. Proakis, John G., and Masoud Salehi, *Fundamentals of Communication Systems*, Low Price Edition, Pearson Education, 2005.
7. Taub, Herbert, Donald L. Schilling, and Gautam Saha, *Taub's Principles of Communication Systems*, 3rd Edition, Tata McGraw-Hill Publishing Company Limited, New Delhi, 2008.

Review Questions

1. What are advantages of spread spectrum communication?
2. What are the two basic types of spread spectrum systems? Explain the basic principle of each of them.
3. What are PN sequences? Discuss their characteristics.
4. Explain how PN sequences are generated. What are maximal length sequences? What are their properties and why are the preferred?
5. Draw the circuit diagram of a PN sequence generator for generating length 15 PN sequences. Assuming the initial contents of the shift register stages to be all ones, explain its working. What is the output sequence obtained?
6. Explain with the help of a neat block diagram, the working of a BPSK DS spread spectrum transmitter.
7. Draw the block diagram of a BPSK/DS spread spectrum receiver and explain its working.
8. Explain how a DS spread spectrum system can suppress narrow band interfering signals.
9. Define the term 'Processing Gain', of a direct sequence spread spectrum system and explain its significance.
10. With the help of a neat block diagram, explain the working of a DS spread spectrum-based CDMA system.
11. Explain the principle of ranging using a DSSS system. On what factor(s) does the accuracy of measurement depend?
12. Explain the principle of FHSS systems.
13. Draw the block diagram of a BFSK/FHSS transmitter and explain it working.
14. With the help of a neat block diagram, explain the working of a BFSK/FHSS receiver.
15. Explain how a FHSS signal resists barrage type as well as partial band jamming.
16. With the help of a block diagram, explain the working of an FHSS based CDMA system.

17. Explain the process of acquisition in the case of a DSSS system.
18. How is acquisition accomplished in the case of an FHSS system?
19. Explain how tracking is performed in the case of a DSSS system.
20. How is tracking performed in an FHSS system?
21. What are the disadvantages of a single shift-register PN sequence generators?
22. Ideally what are the characteristics required to be possessed by a PN sequence to be used in CDMA applications?
23. What are Gold sequences? How are they generated?
24. Briefly discuss some of the important characteristics of Gold sequences which make them very useful in CDMA applications.

Problems

1. If the initial state of the register is 1111, find the output sequence of the shift register. Is it an ML sequence?

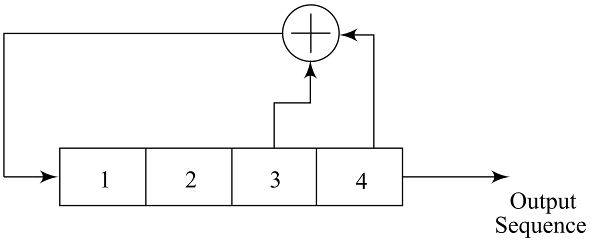


Fig. P14.1

2. A four-stage shift register is used to generate a maximal length sequence. If the chip rate is 10^6 chips per second, find the PN sequence length and PN sequence period.
3. A DSSS/BPSK system is using a shift register of 19 stages for generation of the PN sequence. Determine the processing gain if the system is to be given an average probability of error of 10^{-5} in the presence of a single-tone jamming signal which has an average power three times that of the signal of interest, at the receiver input. Ignore the effect of additive white noise of the channel.
4. A DSSS/BPSK system has an information rate of 3 kbps. At the receiver, a single-tone jamming signal is producing a jamming signal that is five times more powerful than the desired signal. If $\eta = 10^{-20}$ W/Hz and if in the absence of the jamming signal the $(SNR)_R = 60$ dB, calculate the chip rate and the transmission bandwidth if the probability of error (in the presence of the jamming signal) is to be $P_e = 10^{-7}$.
5. In a DSSS/BPSK-based CDMA system there are 10 users and each one is transmitting information at 6 kbps rate. If the probability of error is not to exceed 10^{-7}, what should be the minimum chip rate needed? Assume that additive channel noise can be ignored.
6. A slow hop BFSK/FHSS system is operating with a received SNR of 50 dB and $\eta = 10^{-19}$ W/Hz. But a jammer has started giving a 'barrage jamming' signal which at the receiver has five times the power level of the desired signal. If a probability of error, P_e, of 10^{-7} is to be maintained even in the presence of the jamming signal, what is the minimum processing gain and the corresponding transmission bandwidth required for the FHSS system?

Multiple-Choice Questions

1. The maximum length of a PN sequence that can be generated using a four-stage shift-register is
   (a) 4 (b) 8 (c) 16 (d) 15
2. The auto-correlation of the output of a PN sequence generator is
   (a) random (b) deterministic and periodic
   (c) deterministic and non-periodic (d) None of these

3. If f_b and f_c are the bit frequency of the data and the chip frequency respectively in a DSSS system, then
(a) $f_c >> f_b$ (b) $f_c << f_b$ (c) $f_c = f_b$ (d) No specific relation

4. A DSSS system transmitting binary data at the rate of 10 K bits/sec has PN sequence generators with five shift-register stages and clock frequency of 1 MHz. The processing gain of the system is
(a) 47 dB (b) 30 dB (c) 20 dB (d) 50 dB

5. An FHSS system has its frequency synthesizers controlled by five stage shift-registers with feedback connections taken from the second and fifth stages. The number of slots available for frequency hopping is
(a) 32 (b) 31 (c) 24 (d) 28

Key to Multiple-Choice Questions

1. (d) 2. (b) 3. (a) 4. (c) 5. (a)

MULTICHANNEL MODULATION—OFDM AND DMT

15

> *"The fantastic advances in the field of electronic communication constitute a greater danger to the privacy of the individual."*
>
> **Earl Warren (1891–1974)**
> *American Republican politician and judge*

Learning Objectives

After going through this chapter, students will be able to

- *explain how multipath can cause signal fading and ISI,*
- *understand the difference between conventional FDM and Orthogonal Frequency Division Multiplexing,*
- *describe how the subcarrier frequencies are chosen and explain how they are made orthogonal to each other,*
- *understand how ISI and frequency selective fading can be almost eliminated by having a large number of subcarriers,*
- *explain the meaning of cyclic prefix, why it is inserted and how it helps,*
- *explain by mathematical analysis, how OFDM transforms a wideband single-carrier ISI channel into a number of independent narrowband parallel channels without ISI and how IDFT and DFT bring about this transformation,*
- *understand that DMT is but a variant of OFDM and know in which aspects they differ, and*
- *describe some of the applications of OFDM like ASDL and DAB.*

15.1 INTRODUCTION

Besides causing attenuation, physical communication channels have two major effects on the signals transmitted through them. The signal gets distorted and also gets corrupted by noise. Signal distortion can be caused not only by the non-ideal transmission characteristic of the channel, but also by other factors such as multipath propagation in the case of some wireless channels and signal reflections caused by impedance mismatches in the case of wire line channels. In analog communication, multipath may not have serious consequences because human ears and eyes can often tolerate the resultant distortions to a good extent. But in digital communications, multipath, which creates multiple copies of a signal with different delays, can create at the input to the receiver, a situation in which the delayed versions of one symbol when added may cause the symbol to spill over beyond that time-slot and interfere with the adjacent symbols, leading to severe intersymbol interference (ISI).

As we had already seen earlier, the two main factors contributing to decoding errors in a digital communication system, are ISI and the noise (AWGN) introduced by the channel. In Chapter 10, we had discussed equalization, which is one of the very widely used techniques for combating ISI. Transmitters generally have very little or no knowledge of the transmission characteristics of the channel. Hence, it is normal for the receivers to provide equalization. However, equalization arrangements, especially the non-linear ones like the MLSE, are computationally quite complex. Simpler equalizers like the feedforward and decision feedback equalizers suffer from the disadvantage that they are too much parameter-sensitive. Similarly, it may not always be possible to force all the ISI to zero using a zero-forcing equalizer. Insofar as channel noise is concerned, in Chapter 11, we had discussed optimum filters that would minimize its effect. It should, however, be noted that ISI usually results in some amount of performance degradation even if an optimum detector is used in the receiver. Multichannel modulation which we will be discussing in this chapter, tackles the problem of high speed data transmission over a channel having severe intersymbol interference by converting this problem of serial transmission of symbols over a single channel into the relatively much simpler problem of parallel transmission of the given data stream over a large number of what are called, ‘subchannels’, each of which may be viewed as an AWGN channel. We know it is easy to handle these.

Multichannel modulation, also known as multicarrier modulation, is a bandwidth-efficient communication technique that makes use of ‘Orthogonal Frequency Division Multiplexing’ (OFDM). Simple OFDM assumes only partial *a priori* knowledge of the channel. Discrete-Time Multitone Modulation (DMT), which is a variant of OFDM, however, makes use of a complete knowledge of the channel transmission characteristics in order to selectively allocate symbols carrying more number of bits to subchannels having relatively low attenuation. This has the effect of equalizing the probability of error across the various subchannels and also maximizing the average receiver *SNR*.

OFDM, which forms the basis for multichannel modulation, offers several benefits including high spectral efficiency as well as a good degree of immunity from RF interferences and the undesirable effects arising from multipath. It thus finds a number of practical applications in wire line as well as wireless communication.

15.2 MULTIPATH AND MULTIPATH CHANNELS

In high speed wireless channels, the most attractive feature of OFDM is its high resistance to the severe problem of multipath propagation that causes considerable errors in the received data. Multipath interference at the receiver input produces two effects: ISI and frequency selective fading. Hence, in this section, we shall briefly discuss the phenomenon of multipath and how it leads to these two effects.

The phenomenon by which a transmitted signal arrives at the input to the receiver via two or more paths, is referred to as ‘***multipath***’. Figure 15.1 illustrates a typical example of multipath propagation in which the received signal is arriving via two separate paths – one, the direct path (LOS) between the transmitter and the receiver and the second, the path taken by the part of the transmitted signal that reaches the receiver after getting reflected by the earth’s surface.

For the above illustration of multipath propagation, it has been assumed that a direct path (LOS path) exists. Quite often, in cellular mobile communication, there may not be any direct path between the base station and the mobile phone. The signal from the base station will then be reaching the mobile receiver only via scattering, or diffraction over a number of buildings surrounding the mobile receiver. As another example of multipath propagation, we may consider wireless communication between two aircrafts in flight. In this case, there can be an LOS path and one or more ground reflected paths.

Thus, in a multipath scenario, the receiver receives multiple copies of the transmitted signal through various paths. These copies arrive at the receiver ***sequentially*** with different time delays (and hence with different phase shifts) because of the different path lengths. There will also be slight difference in the ampli-

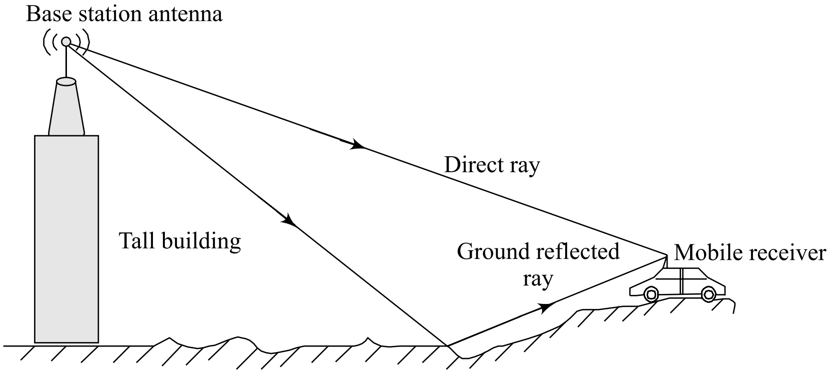


Fig. 15.1 *Illustration of multipath propagation*

tudes of the signal components arriving through different paths; but this is not of much significance. Insofar as the relative phase shifts of the different components of the received signal are concerned, the two ***extreme*** cases are:

1. They may have zero relative phase shifts and so add constructively so as to produce a strong resultant signal.
2. They may add destructively and produce a zero/very weak resultant signal.

In case the transmitter as well as the receiver are stationary, the multipath scenario at the receiver input will not vary with time and such a multipath is termed as a 'static' multipath environment. In case either the transmitter, or the receiver, or both, are moving, a 'dynamic' multipath environment that changes with time, results. Obviously, in a static multipath environment, the amplitude of the received signal does not change with time. But, in a dynamic multipath environment, the nature, lengths and even the number of paths as well as the phase relationships of the signal components arriving via those paths, will go on changing with time. Hence, the resultant received signal strength goes on changing. There may be constructive addition of the signal components arriving through various paths at certain locations and destructive addition at some others. This results in what is called, 'signal fading'. Further, this signal fading may be either '***frequency selective***', or '***non-frequency selective***', depending upon whether the bandwidth of the transmitted signal is greater than, or less than what is called the '***coherence bandwidth***' of the channel which is the signal bandwidth for which signal distortion becomes noticeable at the output of the channel.

As mentioned earlier, in a multipath scenario, the receiver receives multiple copies of the transmitted signal and these arrive sequentially with different path delays. Consider now a situation in which a transmitter is transmitting symbols, each of duration T and that these symbols are arriving at the receiver via. two paths with a delay difference of ΔT. Then, as illustrated in Fig.15.2, each symbol gets extended by ΔT and overlaps with the next symbol for a time period of ΔT and thus causes intersymbol interference (ISI).

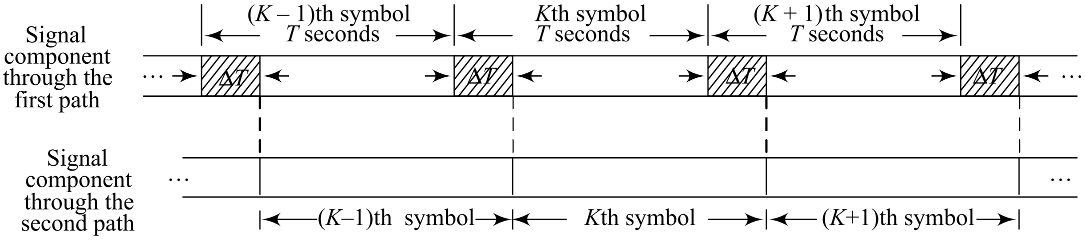


Fig. 15.2 *Two signal components with path delay difference* ΔT

In high speed data transmission, the symbol duration T must necessarily be kept small and ΔT, which is independent of T and depends only on the maximum difference in path delays, can then become comparable to T and cause considerable ISI. It must be noted that smaller the value of ΔT, the delay difference, as compared to the symbol duration T, smaller is the intersymbol interference. In fact, OFDM exploits this fact to make the ISI negligible, by making the symbol duration in each of its subchannels equal to N times the symbol duration in the original data to be transmitted, where N is the number of subchannels. As will be seen in the next section, it mitigates the other effect of multipath, namely frequency selective fading, by making the channel response of each of the N narrowband channels almost 'flat' by using a large N.

15.3 BASIC PRINCIPLE OF OFDM

Conceptually, OFDM has been known for at least the last four decades. However, its practical implementation was simply not possible with the type of semiconductor and computer techniques that were available in the 1960s and 1970s. It became a practical reality – and a highly successful one at that, from the late 1990s. Because of its superior performance in terms of spectral efficiency and considerable immunity from multipath effects, it is better suited for the present-day high-speed data transmission requirements as compared to the traditional single-carrier modulation methods.

In conventional 'Frequency Division Multiplexing' (FDM), independent signals produced by different sources are translated in frequency by subcarrier modulation and these frequency-translated signals are then arranged in adjacent frequency slots in a non-overlapping manner so that they all share the available spectrum. This frequency division multiplexed signal is then transmitted over the channel. To facilitate the separation of these various modulated signals using easily available bandpass filters, small guard bands are also provided between adjacent frequency slots. This increases the total frequency bandwidth occupancy of the multiplexed signal that is transmitted.

As against this, in the case of Orthogonal Frequency Division Multiplexing (OFDM), the subcarriers are rather tightly packed with no guard bands and adjacent subcarrier frequencies are separated by only ($1/T$) Hz, *the symbol rate in each subcarrier*. However, separation of the modulated signals having different subcarriers does not pose any problem since these subcarriers with a separation of ($1/T$) Hz are orthogonal (This was shown in connection with Sunde's FSK (See Eq. (11.104) of Section 11.4.3 in Chapter 11). This makes it possible for OFDM to have a very high spectral efficiency. Whereas independent signals from different sources modulate the various subcarriers in conventional FDM, the various independent *symbols* from the same serial data stream which is to be transmitted, are used in OFDM for modulating the various orthogonal subcarriers. This process is illustrated in the Table 15.1 assuming that the following serial bit stream is to be transmitted using *five* orthogonal subcarriers (In practice, the number of subcarriers used will be very high. It is of the order of 1536 in the case of Digital Audio Broadcasting, i.e., DAB).

Assumed data stream:

$$1, -1, -1, 1, 1, 1, -1, 1, -1, -1, -1, 1, -1, 1, 1, 1, -1, -1, 1, -1$$

Orthogonal subcarriers:

$$SC_1, \quad SC_2, \quad SC_3, \quad SC_4, \quad SC_5$$

As shown in Table 15.1, the 20 bits in the assumed serial data stream are arranged into five parallel blocks using a serial-to-parallel converter, and are used for modulating the various subcarriers.

From the above illustration it is clear that the data rate in each of the five subchannels is only (1/5) of the data rate that would have resulted, had the assumed data been transmitted using a single carrier. In general,

Table 15.1 *Allocation of bits to the subcarriers*

SC_1	SC_2	SC_3	SC_4	SC_5
1	–1	–1	1	1
1	–1	1	–1	–1
–1	1	–1	1	1
1	–1	–1	1	–1

Note

(i) Each column in the above table represents the bits that will be carried by the subcarrier marked at the top of that column.

(ii) These data-modulated subcarriers are then re-multiplexed to create the OFDM carrier.

if N parallel subchannels are used, and if the rate of the given serial data is $(1/T_s)$, the data rate in each subchannel would be

$$\frac{1}{T} = \frac{1}{NT_s} \quad \text{or} \quad T = NT_s \tag{15.1}$$

where T denotes the symbol period in each subchannel and T_s denotes the sampling period, which is equal to the symbol duration in the original data.

Now, to make the subchannels orthogonal to each other we choose Δf, the spacing between adjacent subcarriers equal to $(1/T)$, the symbol rate in each subchannel.

$$\therefore \qquad \Delta f = \frac{1}{T} = \frac{1}{NT_s} = \frac{f_s}{N} \tag{15.2}$$

This implies that if $H_c(f)$ is the frequency response of the available single-carrier channel, it is sampled in frequency domain at regular frequency intervals of $\Delta f = f_s/N$ as shown in Fig. 15.3.

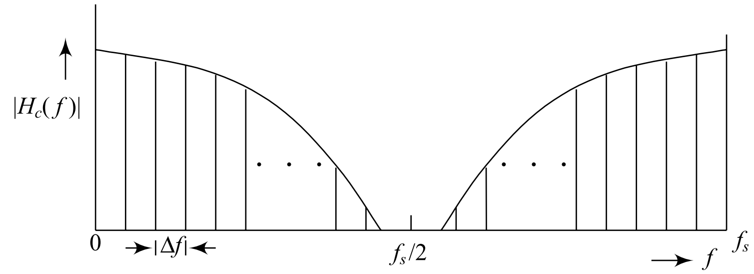


Fig. 15.3 *Sampling the frequency response of the channel at regular frequency interval of Δf*

Then, if f_k is the mid-frequency of the bandwidth of the k^{th} subchannel, we may write

$$x_k(t) = k^{\text{th}} \text{ subcarrier} = \sin(2\pi f_k t + \theta_k);\ k = 0, 1, \ldots, (N-1)$$

where θ_k is the phase angle of the k^{th} subcarrier, and the orthogonality of the subchannels means that the inner product of any two distinct subchannel carriers must be equal to zero. i.e.,

$$\int_0^T \sin(2\pi f_m t + \theta_m)\cdot \sin(2\pi f_n t + \theta_n)dt = 0 \tag{15.3}$$

$$m \neq n \text{ and } 0 \leq m, n \leq (N-1)$$

From the foregoing, we now make the following two remarks.

Remarks

(i) The symbol interval in each subchannel of the OFDM system is N times the symbol interval in the single carrier system. Hence, for a large N, the symbol interval in each subchannel can be made much larger than the path delay difference caused by multipath propagation. Thus, by choosing a large N, we can almost make the OFDM system ISI-free. Thus, OFDM converts an ISI channel into N parallel subchannels which are simple AWGN channels without ISI.

It must, however, be noted that N cannot be indefinitely increased, since the serial-to-parallel conversion of the data will then lead to very large time delays.

(ii) A large value of N will make Δf, the subcarrier spacing, very small and each subchannel will then have almost a flat frequency response over its narrow bandwidth. This mitigates, to a very large extent, the occurrence of frequency selective fading caused by multipath. However, as stated above, N cannot be increased indefinitely.

(iii) Each subchannel of the OFDM system is generally modulated using the spectrally efficient M-ary QAM.

Thus, at least conceptually, formation of an OFDM signal may be visualized as illustrated in Fig. 15.4:

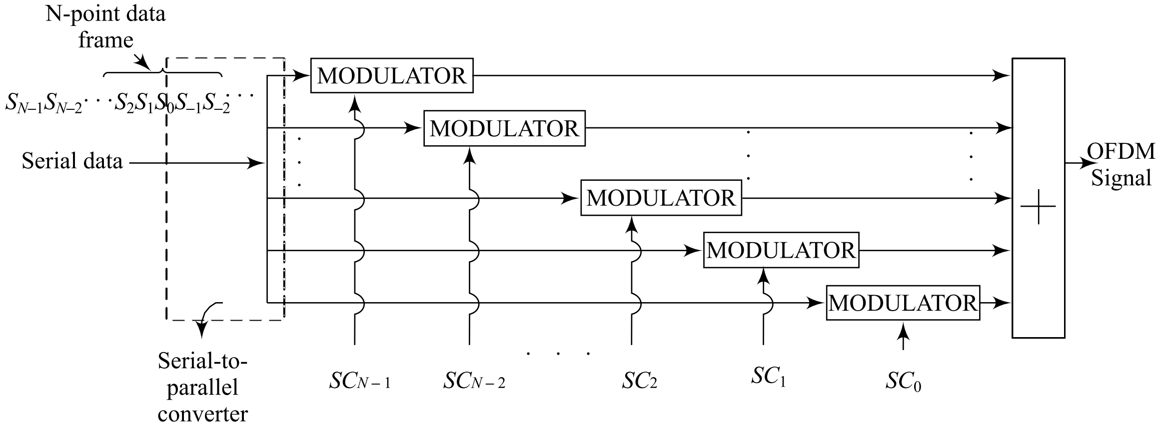


Fig. 15.4 *Illustration of the conceptual formation of an OFDM signal*

From the above, it appears that the implementation of an OFDM system requires a bank of *N* subcarrier oscillators, modulators and coherent M-ary QAM demodulators (on the receiving side). When *N* is large, as is generally the case in practical applications, it becomes quite expensive. Hence, practical implementation requires an altogether different approach. In this connection, it is interesting and useful to note the fact that the whole process of transformation of the single carrier wideband channel into *N* narrowband subchannels operating in parallel, is discrete both in time as well as frequency. This makes it possible to have a matrix representation of the transformation process. This, in turn, permits a DFT implementation of the transformation. DFT implementation can be carried out completely in the digital domain using special purpose hardware for performing FFT which is an efficient way of implementing DFT.

Hence, in what follows, we will show how DFT and its inverse operation IDFT may be used to convert a single carrier wideband ISI channel into a number of narrowband parallel subchannels without ISI. As the pertinent analysis uses a number of key basic concepts of digital signal processing, we propose to first discuss these concepts briefly in the next section before proceeding with that analysis.

15.4 SOME SIGNAL PROCESSING CONCEPTS

15.4.1 Discrete-Time Fourier Transform

A discrete-time signal is a sequence of real or complex numbers. These numbers are assumed to be occurring at regular intervals of T seconds. The n^{th} element of a discrete-time signal is generally denoted by $x(nT)$ and the signal itself is denoted by $\{x(nT)\}$. As this notation is cumbersome, $x(nT)$ is generally used to represent the sequence itself or its n^{th} element. From the context, one can easily make out which one $x(nT)$ represents.

The spectrum of a discrete-time signal $x(nT)$ is represented by its 'Discrete-Time Fourier Transform' (DTFT) defined by

$$X(e^{j\omega T}) \underline{\Delta} \sum_{n=-\infty}^{\infty} x(nT)e^{-j\omega nT} \tag{15.4}$$

As is clear from Eq. (15.4), $X(e^{j\omega T})$ is in general, a complex-valued function of the frequency variable 'f', even if the DT signal $x(nT)$ is purely real and that $X(e^{j\omega T})$ is periodic in frequency with a period $= \frac{1}{T} = f_s$ Hz.

If the time interval T between two consecutive elements of the signal $x(nT)$ is taken as unity,

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \tag{15.5}$$

$X(e^{j\omega})$ is a periodic complex-valued function of ω with a period of 2π radians.
The DTFT has an inverse and

$$x(nT) = \frac{1}{2\pi}\int_{-\omega_s/2}^{\omega_s/2} X(e^{j\omega T})e^{j\omega nT}\,d\omega \tag{15.6}$$

Note *In the above equation, $\omega_s = 2\pi f_s = 2\pi/T$*

and
$$x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega n}\,d\omega \tag{15.7}$$

Some useful properties of DTFT

1. **Linearity:** If $x_1(n) \xleftrightarrow{\text{DTFT}} X_1(e^{j\omega})$, $x_2(n) \xleftrightarrow{\text{DTFT}} X_2(e^{j\omega})$ and if a_1 and a_2 are any two arbitrary constants then:
$$[a_1x_1(n) + a_2x_2(n)] \xleftrightarrow{\text{DTFT}} a_1X_1(e^{j\omega}) + a_2X_2(e^{j\omega}) \tag{15.8}$$
2. **Hermitian symmetry:** If $x(n)$ is a purely real valued discrete-time signal, its DTFT, viz. $X(e^{j\omega})$ will have Hermitian symmetry, i.e.,
$$|X(e^{-j\omega})| = |X(e^{j\omega})| \text{ and } \angle X(e^{-j\omega}) = -\angle X(e^{j\omega}) \tag{15.9}$$
3. **Convolution theorem:** Let $x_1(n) \xleftrightarrow{\text{DTFT}} X_1(e^{j\omega})$ and $x_2(n) \xleftrightarrow{\text{DTFT}} X_2(e^{j\omega})$.

   Then
   $$\text{DTFT}[x_1(n) * x_2(n)] = X_1(e^{j\omega}) \cdot X_2(e^{j\omega}) \tag{15.10}$$
   where the symbol * denotes linear convolution.

4. **Multiplication theorem:** Let $x_1(n) \xleftrightarrow{\text{DTFT}} X_1(e^{j\omega})$ and $x_2(n) \xleftrightarrow{\text{DTFT}} X_2(e^{j\omega})$.

   Then
$$x_1(n) \cdot x_2(n) \xleftrightarrow{\text{DTFT}} X_1(e^{j\omega}) * X_2(e^{j\omega}) \tag{15.11}$$

15.4.2 Discrete Fourier Transform (DFT)

In the DTFT, given by Eq. (15.4), the parameter, *t*, is discretized but the frequency parameter *f* is not discretized; it is a continuous variable. Hence, DTFT cannot be directly used for machine computation of the spectrum of a DT signal. In the Discrete Fourier Transform (DFT), both these parameters are discretized. Time is represented by the discrete variable *n* and the frequency by the discrete variable *k*. The DFT is invertible and the inverse transform is denoted by IDFT. The DFT of an *N*-length sequence *x*(*n*), *n* = 0, 1, …,(*N*–1), is given by

DFT:
$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk};\ k = 0, 1, \ldots, (N-1) \tag{15.12}$$

X(*k*)s are called the DFT coefficients, and are in general complex even if *x*(*n*) is a real-valued sequence. As can be seen from the above equation, the DFT transforms an *N*-length sequence, *x*(*n*), into another *N*-length sequence, *X*(*k*), called the DFT coefficients sequence. We shall see the physical meaning of these DFT coefficients later. The inverse transformation is given by

IDFT:
$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j\frac{2\pi}{N}nk};\ n = 0, 1, \ldots, (N-1) \tag{15.13}$$

Remark *The (1/N) factor in the IDFT equation is only a normalization factor which is used for ensuring that if* $x(n) \xleftrightarrow{\text{DTF}} X(k)$ *then* $X(k) \xleftrightarrow{\text{IDFT}} x(n)$ *without any need for scaling. In fact, we may split it and use a* $(1/\sqrt{N})$ *factor in each of the DFT and IDFT equations.*

Matrix representation of DFT and IDFT The DFT equation (as well as the IDFT equation) actually represents a set of *N* linear equations that can be compactly represented by a matrix equation. Let us define

$$W \,\underline{\Delta}\, e^{-j\frac{2\pi}{N}} \tag{15.14}$$

The DFT Eq. (15.12) may now be expanded for *k* = 0, 1, 2, …, (*N* – 1) to get

$$\begin{aligned}
X(0) &= x(0)\cdot 1 + x(1)W^{(1)(0)} + x(2)W^{(2)(0)} + \ldots + x(N-1)W^{(N-1)(0)}\\
X(1) &= x(0)W^{(0)(1)} + x(1)W^{(1)(1)} + x(2)W^{(2)(1)} + \ldots + x(N-1)W^{(N-1)(1)}\\
X(2) &= x(0)W^{(0)(2)} + x(1)W^{(1)(2)} + x(2)W^{(2)(2)} + \ldots + x(N-1)W^{(N-1)(2)}\\
&\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots\\
X(N-1) &= x(0)W^{(0)(N-1)} + x(1)W^{(1)(N-1)} + x(2)W^{(2)(N-1)} + \ldots + x(N-1)W^{(N-1)^2}
\end{aligned}$$

These *N* equations may conveniently be written down as a single matrix equation. While doing so, we must note that W^m is *N*-periodic in the sense that as *m* takes integer values 0, 1, 2, etc., W^m will have distinct values only up to *m* = (*N* – 1) and that the values that it takes will repeat thereafter. For example, when *m* = *N*, $W^m = 1$, which is the same as W^0. Similarly, $W^{N+1} = W$ and, in general, W^{N+p}, $0<p<N$, is simply equal to W^p itself. Hence, the above *N* equations may be written down completely as

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ \vdots \\ X(N-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W & W^2 & \cdots & W^{(N-1)} \\ 1 & W^2 & W^4 & \cdots & W^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W^{(N-1)} & W^{2(N-1)} & \cdots & W^{(N-1)^2} \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{bmatrix} \tag{15.15}$$

The $N \times N$ matrix on the RHS of Eq. (15.15) is called the DFT matrix of order N.

Properties of the DFT matrix

1. It is a square matrix and is symmetrical about its principal diagonal.
2. The entries in the 0^{th} row and hence in the 0^{th} column, are all equal to 1.
3. In general, the entry in the i^{th} row and j^{th} column, where, $0 \le j \le (N\text{–}1)$, is given by $W^{i\text{–}j}$
4. The DFT matrix is a unitary matrix, i.e., if $\boldsymbol{Q} = \frac{1}{\sqrt{N}}(\text{DFT matrix})$ then $\boldsymbol{Q}^*\boldsymbol{Q} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{Q}^*$ is the Hermitian transpose (i.e., complex conjugate transpose) of $\boldsymbol{Q}$. This implies that $\boldsymbol{Q}^*$ is the inverse of $\boldsymbol{Q}$.

Similarly, the matrix equation representation of the IDFT may be written down as

$$\begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{bmatrix} = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W^{-1} & W^{-2} & \cdots & W^{-(N-1)} \\ 1 & W^{-2} & W^{-4} & \cdots & W^{-2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W^{-(N-1)} & W^{-2(N-1)} & \cdots & W^{-(N-1)^2} \end{bmatrix} \begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ \vdots \\ X(N-1) \end{bmatrix} \tag{15.16}$$

15.4.3 Linear and Circular Convolutions

The linear convolution between two sequences $x(n)$ and $y(n)$ is defined by the following equation:

$$z(n) = \sum_{m=-\infty}^{\infty} x(n-m)y(m) \tag{15.17}$$

where $z(n)$ is the n^{th} element of the resultant sequence, $\{z(n)\}$. This linear convolution of $x(n)$ and $y(n)$ is generally denoted by

$$z(n) = x(n) * y(n) \tag{15.18}$$

In case $x(n)$ and $y(n)$ are finite length sequences of lengths N_1 and N_2 respectively, it can be shown that the linear convolution of these two sequences will result in another finite length sequence, $z(n)$ of length $(N_1 + N_2 - 1)$.

The linear convolution illustrated in Fig.15.5 may be represented as a matrix multiplication given below.

$$\begin{bmatrix} z(0) \\ z(1) \\ z(2) \\ z(3) \\ z(4) \\ z(5) \\ z(6) \end{bmatrix} = \begin{bmatrix} x(0) & 0 & 0 & 0 \\ x(1) & x(0) & 0 & 0 \\ x(2) & x(1) & x(0) & 0 \\ x(3) & x(2) & x(1) & x(0) \\ 0 & x(3) & x(2) & x(1) \\ 0 & 0 & x(3) & x(2) \\ 0 & 0 & 0 & x(3) \end{bmatrix} \begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \end{bmatrix} \tag{15.19}$$

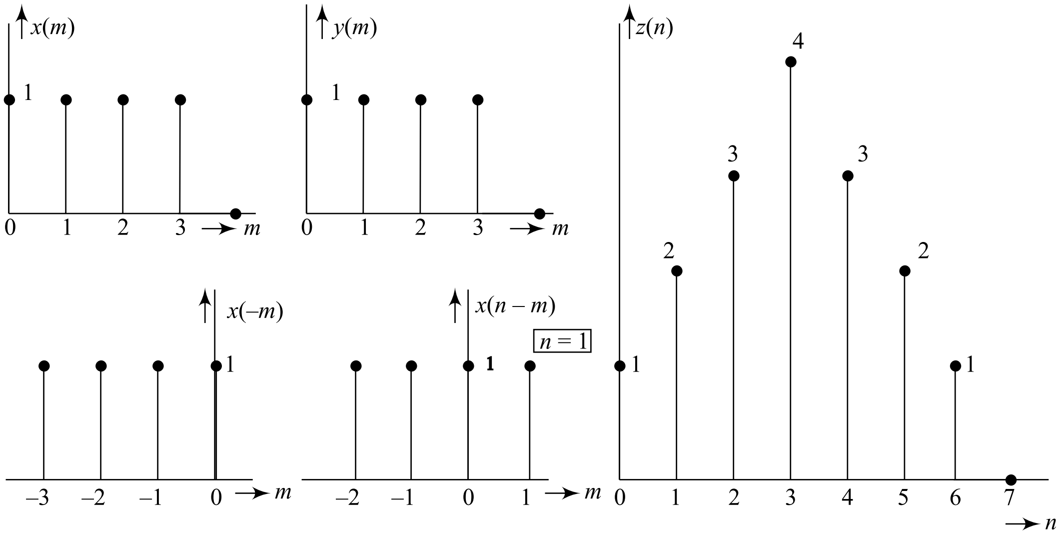


Fig. 15.5 *Graphical representation of discrete linear convolution*

Circular convolution For linear convolution, the two sequences $x(n)$ and $y(n)$ which are to be convolved, need not be of the same length. However, for circular or cyclic convolution, the two sequences which are to be convolved are to be of equal length and the resultant sequence will also be of the same length.

The circular or cyclic convolution of two N-length sequences $x(n)$ and $y(n)$ is defined by the following equation:

$$z(n) = \sum_{m=0}^{N-1} x(n-m)_N\, y(m);\ n = 0, 1, 2, \ldots, (N-1) \tag{15.20}$$

Here, $(n-m)_N$ denotes modulo-N subtraction of m from n. $\{z(n)\}$ is the discrete-time signal resulting from the circular convolution of $x(n)$ and $y(n)$. The circular convolution is often denoted by

$$z(n) = x(n) \otimes y(n) \tag{15.21}$$

Equation (15.20) may conveniently be represented as the following matrix equation, assuming $N = 5$:

$$\begin{bmatrix} z(0) \\ z(1) \\ z(2) \\ z(3) \\ z(4) \end{bmatrix} = \begin{bmatrix} x(0) & x(4) & x(3) & x(2) & x(1) \\ x(1) & x(0) & x(4) & x(3) & x(2) \\ x(2) & x(1) & x(0) & x(4) & x(3) \\ x(3) & x(2) & x(1) & x(0) & x(4) \\ x(4) & x(3) & x(2) & x(1) & x(0) \end{bmatrix} \begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \\ y(4) \end{bmatrix} \tag{15.22}$$

Some useful properties of DFT

1. **DFT is linear** in the sense that if $x_1(n) \xleftrightarrow[N]{\text{DFT}} X_1(k),\ x_2(n) \xleftrightarrow[N]{\text{DFT}} X_2(k)$ and if a_1 and a_2 are arbitrary constants,

$$[a_1x_1(n) + a_2x_2(n)] \xleftrightarrow[N]{\text{DFT}} [a_1X_1(k) + a_2X_2(k)] \tag{15.23}$$

2. From the *N*-periodicity of $e^{-j\frac{2\pi}{N}}$, it follows that the DFT and IDFT equations are *N*-periodic, i.e., they *assume* that $\{X(k)\}$ and $\{x(n)\}$ sequences are *N*-periodic. Hence, $X(k) = X(k + mN)$ for every integer *k*, $0 \le k \le (N-1)$ and for any integer *m*. Also, $x(n) = x(n + mN)$ for every *n*, $0 \le n \le (N-1)$ and any integer *m*.
3. **Convolution theorem:** If $x(n)$ and $y(n)$ are two sequences each of length *N*, with $x(n) \xleftrightarrow[N]{\text{DFT}} X(k)$ and $y(n) \xleftrightarrow[N]{\text{DFT}} Y(k)$, and if their *circular convolution* leads to $z(n)$, then
$$Z(k) = X(k) \cdot Y(k);\ k = 0, 1, 2, \ldots, (N-1) \tag{15.24}$$
Note that DFT supports circular convolution and not linear convolution.
4. **Multiplication theorem:** If $x(n)$ and $y(n)$ are two *N*-length sequences with $x(n) \xleftrightarrow[N]{\text{DFT}} X(k)$ and $y(n) \xleftrightarrow[N]{\text{DFT}} Y(k)$, and if $z(n) = x(n) \cdot y(n)$ then
$$Z(k) = \frac{1}{N}[X(k) \otimes Y(k)] \tag{15.25}$$
5. **Complex-conjugate theorem:** Let $x(n)$ be a *real-valued* sequence of length *N*. Let $x(n) \xleftrightarrow[N]{\text{DFT}} X(k)$. Then,
$$X(N-k) = \overline{X(k)};\ k = 0, 1, 2, \ldots, (N-1) \tag{15.26}$$

Note *Overbar indicates complex conjugate*

So, for any real-valued sequence of finite length *N*, the k^{th} DFT coefficient and the $(N-k)^{\text{th}}$ DFT coefficient are complex conjugate of each other.

15.4.5 Physical Meaning of the DFT Coefficients

The relationship between DFT and DTFT reveals the physical meaning of the DFT coefficients. For this purpose, consider a sequence $x(n)$ which is causal and of length *N*.

Then

$$\text{DTFT of } x(nT)\text{: } X(e^{j\omega T}) = \sum_{n=0}^{N-1} x(nT)e^{-j\omega nT} \tag{15.27}$$

and

$$\text{DFT of } x(nT)\text{: } X(k) = \sum_{n=0}^{N-1} x(nT)e^{-j\frac{2\pi}{N}nk};\ k = 0, 1, 2, \ldots, (N-1) \tag{15.28}$$

A comparison of the above two equations reveals that

$$X(k) = X(e^{j\omega T})\Big|_{\omega = \frac{2\pi}{NT}k = \left(\frac{\omega_s}{N}\right)k};\ k = 0, 1, 2, \ldots, (N-1) \tag{15.29}$$

Equation (15.29) tells us that the k^{th} DFT coefficient, $X(k)$, of an *N*-length sequence $x(n)$ is the value of the DTFT of that sequence evaluated at the frequency $\omega = \left(\frac{\omega_s}{N}\right)k$. This again suggests that the *N*-point DFT of an *N*-length sequence, $x(n)$ can be obtained by sampling the DTFT of the sequence $x(n)$ (in the frequency domain) at the frequencies

$$\omega = 0, \left(\frac{\omega_s}{N}\right), \left(\frac{\omega_s}{N}\right) \cdot 2, \left(\frac{\omega_s}{N}\right) \cdot 3, \ldots, \left(\frac{\omega_s}{N}\right) (N-1) \tag{15.30}$$

15.4.6 Diagonalization of a Matrix and Circulant Matrices

1. Diagonalization of a matrix: Let A be an $N \times N$ matrix with N linearly independent eigenvectors : $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, …, $\boldsymbol{X}_N$. Then

$$A\boldsymbol{X}_i = \lambda_i \boldsymbol{X}_i \quad ; \quad i = 1, 2, \ldots, N \tag{15.31}$$

where λ_i is the eigenvalue corresponding to the eigenvector $\boldsymbol{X}_i$.

Stacking the N equations represented by Eq. (15.31) side by side, we get

$$[A\boldsymbol{X}_1 \ A\boldsymbol{X}_2 \ . \ . \ . \ A\boldsymbol{X}_N] = [\lambda_1\boldsymbol{X}_1 \ \ \lambda_2\boldsymbol{X}_2 \ \ . \ . \ . \ \lambda_N\boldsymbol{X}_N]$$

i.e.,

$$A[\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \cdots \quad \boldsymbol{X}_N] = [\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \cdots \quad \boldsymbol{X}_N] \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & . \\ & & & \lambda_N \end{bmatrix}$$

Now, representing the diagonal matrix on the RHS of the above equation by A, and the matrix with eigenvectors $\boldsymbol{X}_1 \ \boldsymbol{X}_2 \ . \ . \ . \ \boldsymbol{X}_N$ as its columns, by U, we have

$$AU = U\Lambda \quad \therefore A = U\Lambda U^{-1} \quad \text{or} \quad \Lambda = U^{-1}AU \tag{15.33}$$

We therefore find that the matrix A is diagonalized by premultiplying it by U^{-1} and post-multiplying by U, where U is the $N \times N$ matrix whose columns are the N linearly independent eigenvectors of A. Such a matrix, U, is called a modal matrix. In fact, this diagonalization is a particular case of similarity transformation and the matrices A and Λ are similar. We know that all similar matrices have the same set of eigenvalues.

2. Circulant matrices: Before proceeding to circulant matrices, we will first discuss one very interesting and useful class of matrices, called the 'permutation matrices'.

Definition A permutation matrix, P, is a square matrix whose elements are all either 0 or 1 and which has exactly one 1 in each row and column.

As an example, a set of three permutation matrices of size 3 × 3 are given below

$$P_1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; \ P_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}; \ P_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{15.34}$$

Note that $P_1^2 = P_2$ and $P_1^3 = P_0$, which is an identity matrix of size 3 × 3.

Properties of permutation matrices

1. Let P be an $N \times N$ permutation matrix and A be any $N \times N$ matrix. Then, PA is a row-permutated version of A while AP is a column permuted version.
2. Permutation matrices are orthogonal matrices.
3. If P is a permutation matrix, $P^{-1} = P^T$, the transpose of P.

It is an easy matter to show that the above set of three 3 × 3 permutation matrices have the following set of common eigenvectors:

$$\boldsymbol{V}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}; \ \boldsymbol{V}_2 = \begin{bmatrix} 1 \\ W \\ W^2 \end{bmatrix}; \ \boldsymbol{V}_2 = \begin{bmatrix} 1 \\ W^2 \\ W \end{bmatrix} \tag{15.35}$$

where

$$W \underline{\Delta}\, e^{-j\frac{2\pi}{3}} \tag{15.36}$$

Thus, the modal matrix for a 3 × 3 permutation matrix is the following matrix which has the above three eigenvectors as its columns:

$$U_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & W & W^2 \\ 1 & W^2 & W \end{bmatrix} \tag{15.37}$$

But, we know that the above U_3 is the 3 × 3 DFT matrix.

We may generalize the above result and state that the $N \times N$ DFT matrix is the modal matrix for any $N \times N$ permutation matrix.

The general form of an $N \times N$ circulant, or cyclic matrix is

$$\underset{\substack{(N\times N)\\ \text{(Circulant)}}}{C} \;\underline{\Delta} \begin{bmatrix} c_0 & c_{N-1} & c_{N-2} & \cdots & c_1 \\ c_1 & c_0 & c_{N-1} & \cdots & c_2 \\ c_2 & c_1 & c_0 & \cdots & c_3 \\ \vdots & \vdots & \vdots & & \vdots \\ c_{N-1} & c_{N-2} & c_{N-3} & \cdots & c_0 \end{bmatrix} \tag{15.38}$$

The reader might have noticed that each column (or row) of the above matrix is obtained by giving a one step circular, or cyclic shift to the previous column (or row). A 3 × 3 size cyclic matrix is given by

$$\underset{3\times 3}{C} = \begin{bmatrix} c_0 & c_2 & c_1 \\ c_1 & c_0 & c_2 \\ c_2 & c_1 & c_0 \end{bmatrix}$$

Note that any 3 × 3 circulant matrix like the above one, is a linear combination of the three 3 × 3 permutation matrices.

i.e.,

$$\underset{3\times 3}{C} = c_0 P_0 + c_1 P_1 + c_2 P_2$$

In general, we therefore have

$$\underset{N\times N}{C} = c_0 P_0 + c_1 P_1 + \cdots + c_{N-1} P_{N-1} \tag{15.39}$$

where the P_is, $i = 0, 1, \ldots, (N-1)$ are the set of $N \times N$ permutation matrices. Since these $N \times N$ – permutation matrices have the $N \times N$ – DFT matrix as their modal matrix, their linear combination $C_{N\times N}$, viz. the cyclic matrix of size $N \times N$ also has the $N \times N$ – DFT-matrix as its modal matrix.

$$\underset{\text{N × N matrix}}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_N^1 & \cdots & W_N^{(N-1)} \\ \vdots & \vdots & & \vdots \\ 1 & W_N^{(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix}} \underset{\text{N × N circulant matrix}}{\begin{bmatrix} c_0 & c_{N-1} & \cdots & c_1 \\ c_1 & c_0 & \cdots & c_2 \\ \vdots & & & \vdots \\ c_{N-1} & c_{N-2} & \cdots & c_0 \end{bmatrix}} \underset{\text{N × N IDFT matrix}}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_N^{-1} & \cdots & W_N^{-(N-1)} \\ \vdots & \vdots & & \vdots \\ 1 & W_N^{-(N-1)} & \cdots & W_N^{-(N-1)^2} \end{bmatrix}} = \underset{\text{N × N diagonal matrix}}{\begin{bmatrix} \lambda_0 & & & & 0 \\ & \lambda_1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & \cdot & \lambda_{(N-1)} \end{bmatrix}}$$

The diagonal elements of the $N \times N$ diagonal matrix on the RHS of Eq. (15.40) are the eigenvalues (see Eq. 15.33) of the $N \times N$ cyclic matrix $C_{N\times N}$.

Remarks

It is left as an exercise to the reader to show that these N eigenvalues are just the DFT coefficients of the generating vector of the cyclic matrix, i.e., DFT coefficients of the first column of the cyclic matrix $C_{N\times N}$. This means that

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & W_N^1 & \cdots & W_N^{(N-1)} \\ \vdots & \vdots & & \vdots \\ 1 & W_N^{(N-1)} & \cdots & W_N^{(N-1)^2} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix} = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_{N-1} \end{bmatrix} \tag{15.41}$$

15.5 SINGLE CARRIER CHANNEL TO MULTICARRIER CHANNEL TRANSFORMATION – ROLE OF DFT AND IDFT

As has been stated in Section 15.1, OFDM assumes partial information about the transmission channel. If the channel could be modeled as an FIR system of order K, the OFDM system should have an *a priori* knowledge of K. Of course, its knowledge about the channel would be complete if the channel's impulse response coefficients, $h(0), h(1), \ldots, h(K)$ are also known. But OFDM system implementation does not require knowledge of these impulse response coefficients of the FIR channel.

Suppose a data sequence $\{s(n)\}$ is given as input to an FIR channel of order K. Then, we know that the linear convolution of the channel impulse response sequence and the data sequence will give us the output samples. That is,

$$z(n) = \sum_{i=0}^{K} h(i)s(n-i) + w(n) \tag{15.42}$$

where $w(n)$ is the n^{th} sample of the additive white noise process on the channel, and $z(n)$ is the n^{th} output sample. Then, a vector of N output samples is given by writing down the N equations corresponding to $n = 0, 1, \ldots, (N-1)$ in Eq. (15.42)

$$\begin{aligned} z(0) &= h(0)s(0) + h(1)s(-1) + \cdots + h(K)s(-K) + w(0) \\ z(1) &= h(0)s(1) + h(1)s(0) + \cdots + h(K)s(-K+1) + w(1) \\ &\vdots \\ z(K) &= h(0)s(K) + h(1)s(K-1) + \cdots + h(K)s(0) + w(K) \\ &\vdots \\ z(N-2) &= h(0)s(N-2) + h(1)s(N-3) + \cdots + h(K)s(N-2-K) + w(N-2) \\ z(N-1) &= h(0)s(N-1) + h(1)s(N-2) + \cdots + h(K)s(N-1-K) + w(N-1) \end{aligned} \tag{15.43}$$

From the above set, we find that in order to determine N output samples $z(0), z(1), \ldots, z(N-1)$, as many as $N + K$ input sample values $s(-K), s(-K+1), \ldots, (s(-1), s(0), s(1), \ldots, s(N-1)$ are being used. Insertion of what is called the 'cyclic prefix', a step that is equivalent to zero-padding that we normally resort to for converting a linear convolution into a cyclic convolution, makes the input data vector also to be of the same length as the output vector. Further, the $N \times N$ matrix that we get by writing Eq. (15.43) as a matrix equation, now becomes a cyclic $N \times N$ matrix. Imagine the serial data to be as shown below:

$$\ldots, s(-K), s(K+1), \ldots, s(-2), s(-1), s(0), s(1), \ldots, s(N-\overline{K+1}), s(N-K),$$
$$\ldots, s(N-2), s(N-1), s(0), s(1), \ldots$$

Insertion of cyclic prefix then consists of replacing $\{s(-K), s(-K+1), \ldots, s(-2), s(-1)\}$ by $\{s(N-K), \ldots, s(N-2), s(N-1)\}$. Once this is done the N-equations of Eq. (15.43) may be written down as a matrix equation involving a cyclic or circular matrix, as shown below.

$$\begin{bmatrix} z(0) \\ z(1) \\ z(2) \\ \cdot \\ \cdot \\ \cdot \\ z(K) \\ \cdot \\ \cdot \\ z(N-2) \\ z(N-1) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & 0 & \cdots & h(K) & h(K-1) & \cdots & h(1) \\ h(1) & h(0) & 0 & \cdots & 0 & h(K) & \cdots & h(2) \\ h(2) & h(1) & h(0) & 0 \cdots & 0 & h(K) & \cdots & h(3) \\ \cdot & \ddots & \ddots & \ddots & \ddots & \ddots & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ h(K) & h(K-1) & \cdot & \cdots & h(1) & h(0) & 0 \cdots & h(K) \\ 0 & h(K) & h(K-1) & \cdots & h(2) & h(1) & h(0) \cdots & 0 \\ \cdot & \ddots & \ddots & & & \ddots & \ddots & 0 \\ \cdot & & & & & & & 0 \\ 0 & 0 & \cdot & \cdots h(K) & h(K-1) & \cdot & \cdots & h(1) \; h(0) \end{bmatrix} \begin{bmatrix} s(0) \\ s(1) \\ s(2) \\ \cdot \\ \cdot \\ \cdot \\ s(K) \\ \cdot \\ \cdot \\ s(N-2) \\ s(N-1) \end{bmatrix} + \begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \cdot \\ \cdot \\ \cdot \\ w(K) \\ \cdot \\ \cdot \\ w(N-2) \\ w(N-1) \end{bmatrix} \tag{15.44}$$

This $N \times N$ circulant, or cyclic matrix, H_c, can be diagonalized using the DFT and IDFT matrices of size $N \times N$ (see Eq. (15.40).

$$\therefore \qquad W_N H_c W_N^{-1} = \Lambda_H \tag{15.45}$$

where Λ_H is a diagonal matrix with the eigenvalues of the circulant matrix, H_c, as its diagonal elements. But, we know from Eq. (15.41) and the remark preceding it that these eigenvalues are the DFT coefficients of the N-length generator vector of H_c, viz., the DFT of $[h(0)\ \ h(1)\ \ h(2)\ \ .\ .\ .\ \ h(K)\ \ 0\ \ 0\ \ .\ .\ .\ \ 0]^T$. This vector is the impulse response sequence of the channel and so its N-point DFT is given by the N-samples of its DTFT (see Section 15.4.5 and in particular Eq. (15.29). That is, they are $H(0), H(1), H(2), \ldots, H(N-1)$.

$$\Lambda_H = \begin{bmatrix} H(0) & & & \text{\Large 0} \\ & H(1) & & \\ \text{\Large 0} & & \cdot & \\ & & & \cdot \\ & & & H(N-1) \end{bmatrix} \tag{15.46}$$

From Eq. (15.45), we have

$$H_c = W_N^{-1} \Lambda_H W_N = \left(\frac{1}{\sqrt{N}} W_N \right)^{-1} \Lambda_H \left(\frac{1}{\sqrt{N}} W_N \right) \tag{15.47}$$

Note that as stated earlier in the remark following Eq. (15.13), the normalization factor $(1/N)$ is now split into two factors and each of these, viz. $(1/\sqrt{N})$, is attached to the DFT and IDFT.

Substituting for H_c in Eq. (15.44) using Eq. (15.47), we get

$$\begin{bmatrix} z(0) \\ z(1) \\ \cdot \\ \cdot \\ \cdot \\ z(K) \\ \cdot \\ \cdot \\ z(N-1) \end{bmatrix} = \left(\frac{1}{\sqrt{N}}W_N\right)^{-1} \Lambda_H \left(\frac{1}{\sqrt{N}}W_N\right) \begin{bmatrix} s(0) \\ s(1) \\ \cdot \\ \cdot \\ \cdot \\ s(K) \\ \cdot \\ \cdot \\ s(N-1) \end{bmatrix} + \begin{bmatrix} w(0) \\ w(1) \\ \cdot \\ \cdot \\ \cdot \\ w(K) \\ \cdot \\ \cdot \\ w(N-1) \end{bmatrix} \tag{15.48}$$

Let us now define the following vectors:

$$\tilde{\boldsymbol{s}} \underline{\Delta} \begin{bmatrix} \tilde{s}(0) \\ \tilde{s}(1) \\ \cdot \\ \cdot \\ \cdot \\ \tilde{s}(K) \\ \cdot \\ \cdot \\ \tilde{s}(N-1) \end{bmatrix} = \left(\frac{1}{\sqrt{N}}W_N\right) \boldsymbol{s} = \left(\frac{1}{\sqrt{N}}W_N\right) \begin{bmatrix} s(0) \\ s(1) \\ \cdot \\ \cdot \\ \cdot \\ s(K) \\ \cdot \\ \cdot \\ s(N-1) \end{bmatrix} \tag{15.49a}$$

$$\tilde{\boldsymbol{z}} \underline{\Delta} \begin{bmatrix} \tilde{z}(0) \\ \tilde{z}(1) \\ \cdot \\ \cdot \\ \cdot \\ \tilde{z}(K) \\ \cdot \\ \cdot \\ \tilde{z}(N-1) \end{bmatrix} = \left(\frac{1}{\sqrt{N}}W_N\right) \begin{bmatrix} z(0) \\ z(1) \\ \cdot \\ \cdot \\ \cdot \\ z(K) \\ \cdot \\ \cdot \\ z(N-1) \end{bmatrix} \tag{15.49b}$$

and

$$\tilde{\boldsymbol{w}} \triangleq \begin{bmatrix} \tilde{w}(0) \\ \tilde{w}(1) \\ \cdot \\ \cdot \\ \cdot \\ \tilde{w}(K) \\ \cdot \\ \cdot \\ \tilde{w}(N-1) \end{bmatrix} = \left(\frac{1}{\sqrt{N}} W_N\right) \begin{bmatrix} w(0) \\ w(1) \\ \cdot \\ \cdot \\ \cdot \\ w(K) \\ \cdot \\ \cdot \\ w(N-1) \end{bmatrix} \tag{15.49c}$$

Now, premultiplying both sides of Eq. (15.48) by $\left(\frac{1}{\sqrt{N}} W_N\right)$, and then making use of Eq. (15.49), we get

$$\tilde{\boldsymbol{z}} = \Lambda_H \tilde{\boldsymbol{s}} + \tilde{\boldsymbol{w}} \tag{15.50}$$

Since Λ_H is the diagonal matrix given by Eq. (15.46), the above equation may be written down as the following set of N equations *in terms of the* n^{th} *elements* of the various vectors involved:

$$\tilde{\boldsymbol{z}}(n) = H(n)\,\tilde{\boldsymbol{s}}(n) + \tilde{\boldsymbol{w}}(n);\, n = 0, 1, \ldots, (N-1) \tag{15.51}$$

For a proper interpretation of this extremely useful equation, one should clearly understand the role of $\tilde{s}(n)$ and the difference between $\tilde{s}(n)$ and $s(n)$. As per Eq. (15.42), the data sequence fed to the FIR channel is, of course, $s(n)$. We now distinguish this from the source data which is of interest at the destination (i.e., at the output of the receiver). Equation (15.49a) tells us that this source data vector $\tilde{\boldsymbol{s}}(n)$ is the DFT of the data vector $\boldsymbol{s}(n)$. That is, the data vector $\boldsymbol{s}(n)$ that is actually transmitted over the FIR channel is obtained by taking the IDFT of the source data vector $\tilde{\boldsymbol{s}}(n)$. Equation (15.43) then tells us that the data part of the channel output $z(n)$ corresponding to the data $s(n)$ as input, can be obtained without recourse to the complex linear convolution of Eq. (15.42) by the insertion of the cyclic prefix. From Eq. (15.49b), we find that $\tilde{\boldsymbol{z}}$ vector can be obtained (in the receiver) from the vector $\boldsymbol{z}$ by taking its DFT. Then Eq. (15.51) tells us that the n^{th} element, $n = 0, 1, \ldots, (N-1)$ of $\tilde{\boldsymbol{z}}$ vector is equal to a scalar complex number $H(n)$ (that represents the complex gain of the n^{th} subchannel) times $\tilde{\boldsymbol{s}}(n)$ plus the n^{th} sample of the modified noise sample $\tilde{\boldsymbol{w}}(n)$.

In fact, Eq. (15.51) shows that the single FIR channel is in effect, replaced by N parallel channels with gains $H(0)$, $H(1)$, …, $H(N-1)$ and transmitting $\tilde{\boldsymbol{s}}(0), \tilde{\boldsymbol{s}}(1) \ldots, \tilde{\boldsymbol{s}}(N-1)$, respectively. It further shows that by taking the IDFT of the source data vector and the DFT of the channel output vector, OFDM converts a single ISI channel of order K into N parallel subchannels that are devoid of ISI. And, what is interesting is, that this is achieved without having to generate and modulate the N subcarriers. Also, these subchannels are independent since the noise samples on them are independent. This is because $w(n)$s being samples of white Gaussian noise, are statistically independent. So $\tilde{\boldsymbol{w}}(n)$s are also independent since $\tilde{\boldsymbol{w}}$ vector is obtained by a linear transformation (DFT) of the $\tilde{\boldsymbol{w}}$ vector. Figure 15.6 shows the block diagram of an OFDM system implemented using an N-point IDFT at the transmitter and an N-point DFT at the receiver.

Equation (15.51) suggests that the action of all the OFDM stages for the IDFT onwards can as well be equivalently represented as shown in Fig. 15.7.

Since $H(0)$, $H(1)$, …, $H(N-1)$ are the DFT coefficients of the N-length vector $[h(0)\ h(1) \ldots h(N-1)\ 0 \ldots 0]^T$, they are just complex numbers independent of frequency. Thus, the subchannel gains are frequency independent.

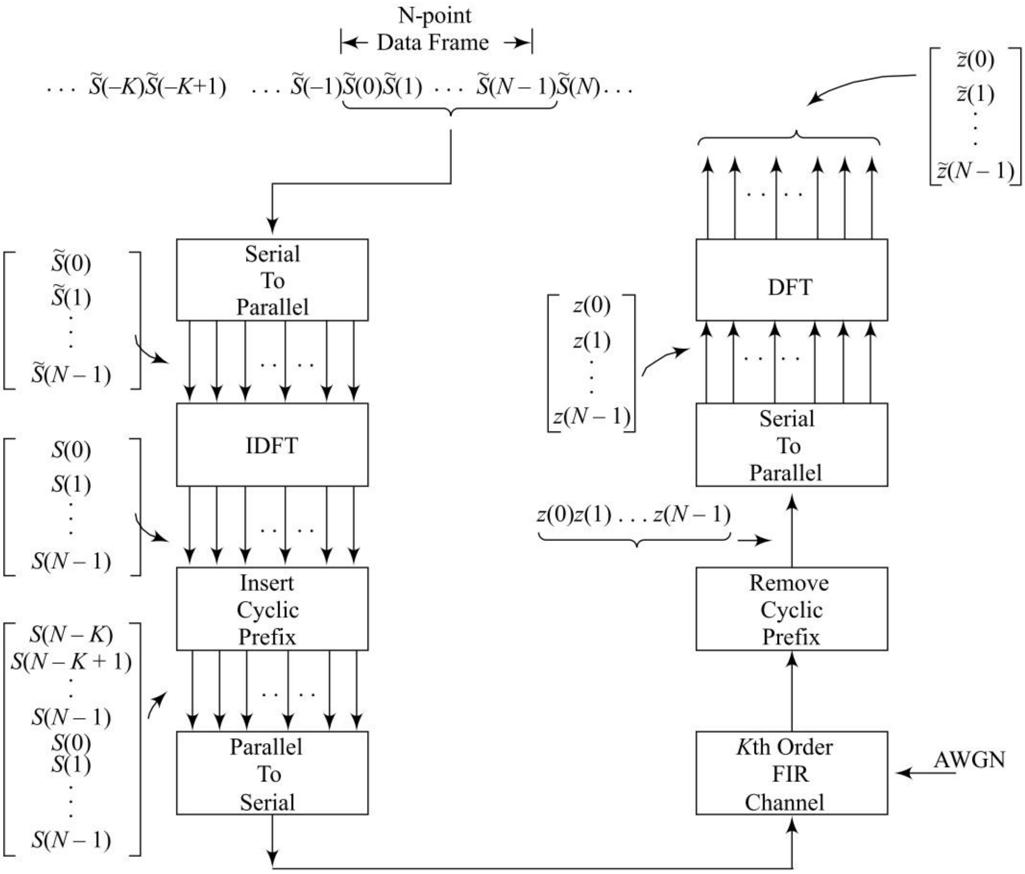


Fig. 15.6 *N-point DFT implementation of an OFDM transmission system*

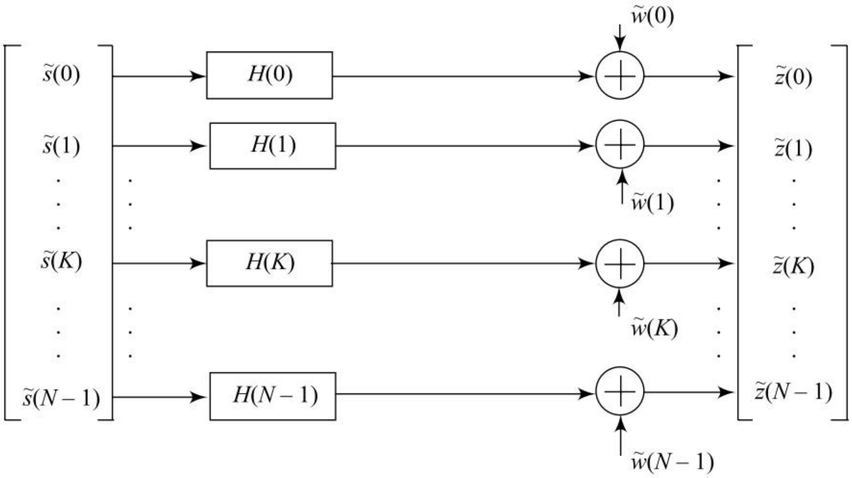


Fig. 15.7 *Diagrammatic representation of Eq. (15.51) illustrating the generation of N-independent AWGN channels by an OFDM system*

At this stage, we would like to make the following remarks regarding K, the channel filter order and the cyclic prefix:

Remarks

(i) The OFDM transmitter must have a priori knowledge of the order K of the channel filter.

(ii) The length of the cyclic prefix being equal to K, use of an over-estimated value of K will lead to a reduction of the effective data rate and wastage of the channel bandwidth. This is because, insertion of the K-length cyclic prefix symbols at the transmitter and their removal at the output of the channel amounts to transmitting effectively N symbols in a time duration of (N + K)T.

To understand more clearly how the subcarrier modulation is implemented using IDFT at the transmitter, let us closely examine what exactly the 'source data symbols', denoted by $\tilde{s}(n)$, $n = 0, 1, \ldots, (N - 1)$ represent. For this purpose, let us revisit the example pertaining to Table 15.1. For the sake of simplicity, it was assumed there that in each frame, each subcarrier was getting modulated by a single bit. But as stated in point iii. of Remark 1, M-ary QAM with its high spectral efficiency, is generally used for subcarrier modulation. Suppose we want to use 16-ary QAM for this purpose. Then, in that example, we can modify Table 15.1 and arrange all the 20 serial bits in one block and allocate a group of 4 bits to each subcarrier as shown below.

SC_1	SC_2	SC_3	SC_4	SC_5
$\{1,-1,-1,1\}$	$\{1,1,-1,1\}$	$\{-1,-1,-1,1\}$	$\{-1,1,1,1\}$	$\{-1,-1,1,-1\}$

Now, each of the above 4-bit groups can have 2^4 possible states each of which can be mapped uniquely into one of the 16 possible message states in the 16-ary QAM constellation. We know that each of these message states can be specified by its two orthogonal coordinates in the signal space of 16-ary QAM, i.e., by a complex number. *Thus, each 4-bit sequence that is to be transmitted over a subchannel by 16-ary QAM, is uniquely mapped into a complex number. Hence, the actual serial binary input data which is to be transmitted, is first segmented into groups of 4-bits each and these groups are mapped into the complex numbers corresponding to the 16-ary QAM message points in the constellation*. It is the sequence of these complex numbers which constitutes what was earlier termed as 'source data' and we represented an N-length sequence of these complex numbers by the $[\tilde{s}(0)\ \tilde{s}(1)\ \tilde{s}(2)\ldots\tilde{s}(N-1)]^T$ vector in Eq. (15.49a), where N subcarriers were assumed. Hence, each of the complex numbers $\tilde{s}(i), 0 \le i \le (N-1)$, represents the message point (in the 16-ary QAM constellation) corresponding to the 4-bit sequence that is used for quadrature amplitude modulating the i^{th} subcarrier. Equation (15.51) now tells us that the output from the n^{th} subcarrier is nothing but a zero-mean AWGN sample plus a scaled version of the complex number $\tilde{s}(n)$ representing the message point in the QAM constellation corresponding to the 4-bit sequence given to the n^{th} subcarrier in that OFDM frame, the scaling factor being the gain of the n^{th} subchannel.

The main task before the OFDM receiver is to recover $\tilde{s}(n), 0 \le n \le (N-1)$ from $\tilde{z}(n), 0 \le n \le (N-1)$ as accurately as possible and then use these in the QAM decision device.

Multiplying both sides of Eq. (15.51) by $[H(n)]^{-1}$, we get

$$[H(n)]^{-1}\,\tilde{z}(n) = \tilde{s}(n) + [H(n)]^{-1}\,\tilde{w}(n) \tag{15.52}$$

Since $w(n)$ is a Gaussian zero-mean random variable, the recovered message point $[H(n)]^{-1}\,\tilde{z}(n)$ is also a random variable. We use this in the QAM decision device which identifies that QAM message point $\tilde{s}(n)$ which is nearest to this and decides that message point as the one that has been transmitted. The unique message sequence corresponding to that message point is then produced. The message sequence transmitted over the n^{th} subchannel is thus decoded. Since $H(n)$ represents the gain of the n^{th} subchannel, $[H(n)]^{-1}$ in Eq. (15.52) is a gain compensation factor used so as to make the total gain of each message point equal to unity before feeding it to the QAM decision device. It has to be noted however that since $[H(n)]^{-1}$ multiplies both the signal term as well as the noise term of Eq. (15.51), the gain compensation process does not alter the *SNR*. *The foregoing however points to the need for the receiver to have knowledge of the actual gains of each of the N subchannels for optimum detection of the transmitted message points*. These subchannel gains are usually estimated by the receiver using the received signals corresponding to unmodulated subcarriers initially transmitted by the OFDM transmitter.

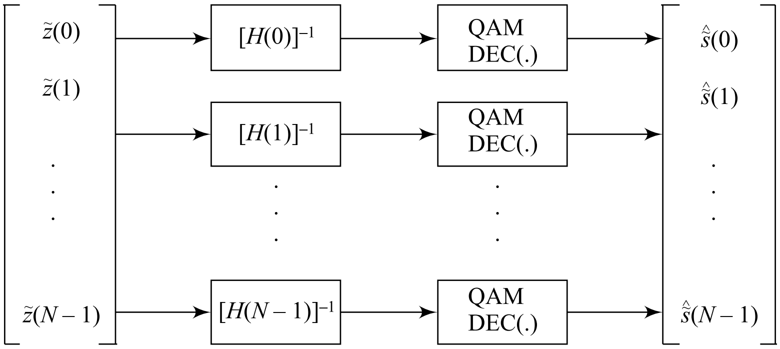


Fig. 15.8 *Gain compensation arrangement in the N subchannels*

15.6 DISCRETE TIME MULTITONE (DMT) MODULATION

It is evident from Eq. (15.51) that the *SNR* will be different in the various subchannels. This is because while $\tilde{w}(n)$, the noise term has the same variance for all n and is equal to the $\eta/2$, the two-sided PSD of the zero-mean AWGN on the channel, $H(n)$, the subchannel gain is different for various subchannels.

As has been stated earlier in this chapter, DMT is a variant of OFDM in which subchannels with high gain that yield a high *SNR* are modulated to carry more bits per symbol than the subchannels with low gain. This is called ***bit loading***. This enables the transmitter to optimally choose different constellations for the various subchannels so as to maximize the average receiver *SNR*. Obviously, this will be possible only if the transmitter has *a priori* knowledge of the gains $H(n)$s for $n = 0, 1, \ldots, (N-1)$. Generally, the receiver estimates these $H(n)$s using the unmodulated subcarrier signals initially transmitted by the transmitter. In DMT, therefore, the receiver must make available this information to the transmitter. That is why, we say that *while an OFDM transmitter does not need to have complete knowledge of the channel response, a DMT transmitter needs that knowledge*.

A DMT system block diagram differs from the OFDM block diagram shown in Fig. 15.5 on the transmitter side as it will have what is called a ‘constellation encoder’ included between the serial-to-parallel block and the IDFT block. This constellation encoder allocates bits among the N-parallel channels as per a bit-loading algorithm. It also represents each subchannel by a QAM constellation.

15.7 PEAK-TO-AVERAGE POWER RATIO (PAR)

A serious problem encountered in OFDM systems (this includes DMT systems too) is that the transmitted signal has occasional large peaks whenever the subchannel signals add constructively in phase. In fact, it has a noise-like amplitude with a large dynamic range. These large peaks can cause the power amplifiers of the transmitter to saturate leading to the generation of inter-modulation distortion. Therefore to avoid this, it becomes necessary to use power amplifiers with a large peak-to-average power ratio; but then, it leads to inefficiency. Several methods are in use to overcome, or at least reduce, this PAR problem. In some OFDM systems, the peaks in the transmitted signal are limited. This introduces distortion and can lead to higher levels of data errors. However, error-correcting codes can be used to eliminate those errors, although, of course, their use reduces the data rate to some extent. Another approach to overcome the PAR problem is to introduce pseudorandom phase shifts into the subchannel signals and thus reduce the possibility of occurrence of large peaks. Information regarding the set of pseudorandom phase shifts given during a given signal interval can be passed on to the receiver using one of the *N* subchannels.

15.8 ADVANTAGES OF OFDM AND DMT

1. It allows overlapping of the subcarrier spectra and thus has high spectral efficiency.
2. Close spacing of the subcarrier frequencies makes the subchannels to be flat narrowband fading channels. *Because of this, OFDM is more robust against frequency selective fading as compared to the single-carrier systems*.
3. OFDM is free of ISI because of the use of cyclic prefix.
4. Through the use of error-correcting codes, the effect of frequency selective fading can be completely eliminated.
5. It is quite robust against impulsive noise.

15.9 DISADVANTAGES OF OFDM AND DMT

1. OFDM signals have a high peak-to-average power ratio. This causes problems in the operation of the RF power amplifiers.
2. As compared to a single-carrier system, *an OFDM system is more sensitive to carrier frequency drift and offsets.*

15.10 APPLICATIONS OF OFDM AND DMT

Because of the advantages listed above, OFDM has found applications in a variety of digital communication systems. These include Digital Audio Broadcasting (DAB), Digital Video Broadcasting (DVB), Wireless LAN's etc. In fact, it is useful in all direct broadcasting and terrestrial applications which encounter multipath distortion. DMT, on the other hand, is particularly useful in transmission of digital data over two-way channels and hence is widely used in Asymmetric Digital Subcarrier Line (ADSL). In what follows, we briefly describe a few of these applications.

15.11 ASYMMETRIC DIGITAL SUBSCRIBER LINE (ADSL)

The twisted-pair wire line that connects a subscriber's premises to the telephone central office is called the ***subscriber's local loop***. With the advent of internet, the need arose for providing high speed internet access to the telephone subscriber over the subscriber's local loop. The dial-up type of voice band modems that were in use for this purpose in the earlier days of internet, could not however go beyond a data rate of 56 kbps as they were using only the relatively very narrow voice band for digital transmission. That is why ADSL, which can have data rates up to 6.8 Mbps, has mostly replaced the voice band modems.

The bandwidth that the subscriber's local loop can support decreases with length of the loop and is 1.2 MHz for a typical 3 km length loop. Transmission from the central office to the subscriber is called the down link and transmission from the subscriber to the central office is called the uplink. Generally, transmission on the subscriber's local loop is asymmetrical since high speed digital transmission at the rate of a few mega-bits per second takes place along the down link whereas data is transmitted along the uplink at a much lower rate of a few hundred kilo-bits per second. That is why, it is called Asymmetrical Digital Subscriber Link (ADSL).

Of the available bandwidth of 1.2 MHz, the lower 0–25 kHz bandwidth obtained at the output of a low pass filter with a cut-off frequency of 25 kHz, is earmarked for voice transmission and the rest, i.e., 25 kHz to 1.2 MHz, obtained at the output of a high pass filter with a 25 kHz cut-off frequency, is for digital transmission. The 25 kHz to 1.104 MHz bandwidth supports 256 parallel DMT subchannels. These 256 available subchannels are divided between the downlink and the uplink. Subcarriers 6 to 32 corresponding to 25 kHz to 138 kHz are usually allocated to the transmission of data along the uplink, while the rest, i.e., 33 to 256, corresponding to 138 kHz to 1.104 MHz, are allocated for downlink transmission.

The receiver informs the transmitter via the uplink about the SNRs on the various subchannels and depending on the *SNR* the transmitter selects the number of bits/symbol (i.e., the constellation size) for each subchannel, which may range anywhere from two bits/symbol corresponding to QPSK to six bits/symbol corresponding to 64-QAM.

15.12 DIGITAL AUDIO BROADCASTING (DAB)

Digital Audio Broadcasting (DAB) evolved during the period 1981 to 1990. A consortium of European nations started a research project named Eureka 147 in the year 1987 for developing DAB. The project was successful and the first trial broadcasts were made in 1990. The DAB standards were finalized and adopted by the International Telecommunication Union (ITU) in 1994 and by the European Telecommunication Standards Institute (ETSI) in 1997. However, subsequently different countries adopted different standards. As of 2006, approximately 1000 DAB stations were operating worldwide. It is expected that in course of time, DAB will replace the conventional analog AM and FM broadcasting which are presently in use. Below are some basic OFDM-related features of Eureka 147 DAB which are useful for terrestrial applications, or satellite applications, or for both.
It operates in four different modes, as listed below:

Mode-I in BAND-III (174 – 240 MHz) for terrestrial.
Mode-II in the L-BAND (1452 – 1492 MHz) for both terrestrial as well as satellite applications.
Mode-III for frequencies below 3 GHz for both terrestrial and satellite applications.
Mode-IV in the L-BAND for terrestrial as well as satellite applications.

In all the modes, OFDM is used and differential QPSK modulation is made use of on all the subchannels. Mode-I is meant for terrestrial broadcasting and it uses 1536 subcarriers with a spacing of 1 kHz between adjacent subcarriers. This corresponds to a symbol duration of 1 ms on each subchannel. It uses a frame duration of 96 ms and a cyclic prefix of 246 μs. For conserving the bandwidth, Eureka 147 DAB employs MPEG audio compression.

An upgrade to DAB, called the DAB+, was introduced in 2006. It makes use of a better audio codec HE-AACV2, 'MPEG Surround' audio format and Reed-Solomon error correcting code.

Summary

- The phenomenon by which a transmitted signal arrives at the receiver via two or more paths is called 'multipath'.
- In static multipath the received signal amplitude does not change with time.
- When the transmitter or the receiver, or both are in motion, the multipath environment at the receiver is dynamic and goes on changing with time causing signal fading.
- ISI is small if ΔT, the time-delay difference is small compared to the symbol duration T.
- OFDM stands for Orthogonal Frequency Division Multiplexing.
- It uses a large number of subcarriers which are orthogonal to each other and are tightly packed.
- Instead of independent signal, independent symbols from the same serial data which is to be transmitted are used in OFDM to modulate the various subcarriers.

- If N is the number of subcarriers, T_s is the sampling period for the given serial data to be transmitted and T is the symbol duration on each subcarrier, then $T = NT_s$.
- If Δf is the spacing between adjacent subcarriers then $\Delta f = 1/T$. This makes the subcarriers orthogonal to each other.
- Since $T = NT_s$, for a given sampling rate of the data, making N very large will help in making $T >> \Delta T$, where ΔT is the time-delay difference. Thus a large N can make the channel ISI free.
- A large value of N will make the frequency response of each subchannel to be almost flat. This helps in reducing frequency selective fading caused by multipath.
- OFDM transforms a wideband single-carrier channel with ISI into N narrowband AWGN parallel subchannels which are ISI free.
- The subcarrier modulation and demodulation as well as the above-mentioned transformation of a wideband ISI channel into N narrowband parallel AWGN channels, is implemented in practice, using IDFT and DFT by using dedicated FFT hardware.
- Insertion of cyclic prefix consists of replacing the K symbols $\{s(-K)\, s(-K+1) \ldots s(-1)\}$ by $\{s(N-K)\, s(N-K+1) \ldots s(N-1)\}$.
- OFDM transmitter must have an *a priori* knowledge of the order K of the FIR channel.
- For optimum detection of the transmitted message points, an OFDM receiver must have knowledge of the actual gains of each of the subchannels.
- DMT stands for Discrete-time Multi-Tone Modulation. It is a form of OFDM in which different subchannels are modulated using different constellations so as to improve the overall *SNR*.
- DMT transmitter requires complete knowledge of the channel.
- The main problem with OFDM and DMT systems is that their signal have a very large peak-to-average value resulting in saturation of the final power amplifiers of the transmitter.
- OFDM has emerged as a very popular digital communication scheme and has found several applications in wire line as well as wireless digital communications. These include ADSL, wireless LAN, DAB and DVB, etc.

References and Suggested Reading

1. Bahai, Ahmad R.S., and B.R. Saltzberg, *Multicarrier Digital Communications – Theory and Applications of OFDM*, Klewer Academic Publications; 2002.
2. Haykin, Simon, *Communication Systems*, 4th Edition, Wiley & Sons (India), 2001.
3. Hoeg, W. and Lauterback, T, *Digital Audio Broadcasting: Principles and Applications of Digital Radio*, John Wiley & Sons, New York, 2001.
4. Lathi, B. P., and Zhi Ding, *Modern Digital and Analog Communication Systems*, International Fourth Edition, Oxford University Press, 2009.
5. Proakis, John G., and Masoud Salehi, *Fundamentals of Communication Systems*, Low Price Edition, Pearson Education, 2005.
6. Rao, P. Ramakrishna, *Signals and Systems*, Tata McGraw-Hill, 2008 (For basics of DTFT, DFT, etc.).

MATHEMATICAL FORMULAE

Appendix

A

1. Cramer's Method of Solving a System of Linear Equations

Let $$a_1x + b_1y = c_1$$
$$a_2x + b_2y = c_2$$

Then $$x = \frac{\Delta x}{\Delta} \text{ and } y = \frac{\Delta y}{\Delta}$$

where, $$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$$

$$\Delta_x = \begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = c_1b_2 - c_2b_1$$

and $$\Delta_y = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = a_1c_2 - a_2c_1$$

2. Geometric Progressions

(i) Let n^{th} term of a geometric progression with 'a' as the first term and 'r' as the common ratio, be t_n. Then

$$t_n = a \cdot r^{n-1}$$

(ii) Sum of n terms = $$\begin{cases} \dfrac{a(1-r^n)}{(1-r)} & \text{if } |r| < 1 \\ \dfrac{a(r^n - 1)}{(r-1)} & \text{if } |r| > 1 \end{cases}$$

(iii) Sum of an infinite geometric progression $= \dfrac{a}{(1-r)} \quad \text{if } |r| < 1$

3. Series Expansion

(i) $e^x = 1 + \dfrac{x}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots$

(ii) $\cos\theta = 1 - \dfrac{\theta^2}{2!} + \dfrac{\theta^4}{4!} + \cdots$

(iii) $\sin\theta = \theta - \dfrac{\theta^3}{3!} + \dfrac{\theta^5}{5!} + \cdots$

(iv) $\tan\theta = \theta + \dfrac{1}{3}\theta^3 + \dfrac{2}{15}\theta^5 + \cdots$

(v) $\sin^{-1}\theta = \theta + \dfrac{1}{6}\theta^3 + \dfrac{3}{40}\theta^5 + \cdots$

(vi) $\tan^{-1}\theta = \theta - \dfrac{1}{3}\theta^3 + \dfrac{1}{5}\theta^5 + \cdots |\theta| < 1$

(vii) $\text{sinc } x = 1 - \dfrac{1}{3!}(\pi x)^2 + \dfrac{1}{5}(\pi x)^5 - \cdots$

(viii) $\log(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} + \cdots \text{ if } |x| < 1$

4. Some Useful Limits

(i) $\operatorname*{Lim}_{\theta\to 0} \cos\theta = 1$

(ii) $\operatorname*{Lim}_{\theta\to 0} \sin\theta = 0$

(iii) $\operatorname*{Lim}_{\theta\to 0} \dfrac{\sin\theta}{\theta} = 1$ (θ in radians)

(iv) $\operatorname*{Lim}_{x\to a} \dfrac{x^n - a^n}{x - a} = n \cdot a^{n-1}$

(v) $\operatorname*{Lim}_{x\to a} \dfrac{f(x)}{g(x)} = \operatorname*{Lim}_{x\to a} \dfrac{f'(x)}{g'(x)} = \operatorname*{Lim}_{x\to a} \dfrac{f''(x)}{g''(x)} = \cdots$

5. Differentiation

(i) $\dfrac{d}{dx}(x^n) = n \cdot x^{n-1}$

(ii) $\dfrac{d}{dx}(e^x) = e^x$

(iii) $\dfrac{d}{dx}(a^x) = a^x \log a$

(iv) $\dfrac{d}{dx}(\sin x) = \cos x$

(v) $\dfrac{d}{dx}(\cos x) = -\sin x$

(vi) $\dfrac{d}{dx}(\tan x) = \sec^2 x$

(vii) $\dfrac{d}{dx}(\cot x) = -\text{cosec}^2 x$

(viii) $\dfrac{d}{dx}(\sec x) = \sec x \tan x$

(ix) $\frac{d}{dx}(\text{cosec } x) = -\text{cosec } x \cot x$

(x) $\frac{d}{dx}(\log x) = \frac{1}{x}$

(xi) $\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}$

(xii) $\frac{d}{dx}(\cos^{-1} x) = \frac{-1}{\sqrt{1-x^2}}$

(xiii) $\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2}$

(xiv) $\frac{d}{dx}(\cot^{-1} x) = \frac{-1}{1+x^2}$

(xv) chain rule: $\frac{d}{dx} f(y) = \left[\frac{d}{dy} f(y)\right] \cdot \frac{dy}{dx}$

6. Integration

(i) $\int e^x dx = e^x + c$

(ii) $\int a^x dx = \frac{a^x}{\log a} + c$

(iii) $\int\limits_{n \neq -1} x^n \, dx = \frac{x^{n+1}}{n+1} + c$

(iv) $\int x^{-1} dx = \log x + c$

(v) $\int K \, dx = Kx + c$

(vi) $\int \sin xdx = -\cos x + c$

(vii) $\int \cos xdx = \sin x + c$

(viii) $\int \tan xdx = -\log \cos x + c$

(ix) $\int \sec xdx = \log(\sec x + \tan x) + c$

(x) $\int \text{cosec } xdx = \log(\text{cosec } x - \cot x) + c$

(xi) $\int \cot xdx = \log \sin x + c$

(xii) $\int \sec^2 xdx = \tan x + c$

(xiii) $\int \text{cosec}^2 xdx = -\cot x + c$

(xiv) $\int \sec x \tan xdx = \sec x + c$

(xv) $\int \text{cosec } x \cdot \cot xdx = -\text{cosec } x + c$

(xvi) $\int \frac{dx}{\sqrt{1-x^2}} = \sin^{-1} x + c = -\cos^{-1} x + c$

(xvii) $\int \frac{dx}{1+x^2} = \tan^{-1} x + c = -\cot^{-1} x + c$

(xviii) $\int \frac{dx}{x^2+a^2} = \frac{1}{a}\tan^{-1}\frac{x}{a} + c$

(xix) $\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\log\left|\frac{x-a}{x+a}\right| + c$

(xx) $\int \frac{dx}{\sqrt{x^2-a^2}} = \log\left[x+\sqrt{x^2-a^2}\right] + c$

(xxi) $\int \frac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\left[\frac{x}{a}\right] + c$

(xxii) $\int \frac{dx}{\sqrt{a^2+x^2}} = \log\left[x+\sqrt{x^2+a^2}\right] + c$

(xxiii) $\int x\sin(ax)dx = \frac{1}{a^2}[\sin(ax) - ax\cos(ax)] + c$

(xxiv) $\int x\cos(ax)dx = \frac{1}{a^2}[\cos(ax) + ax\sin(ax)] + c$

(xxv) $\int xe^{ax}\,dx = \frac{1}{a^2}e^{ax}(ax-1) + c$

(xxvi) $\int xe^{ax^2}\,dx = \frac{1}{2a}e^{ax^2} + c$

7. Some Useful Definite Integrals

(i) $\int_0^\infty e^{-ax^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}; \quad a > 0$

(ii) $\int_0^\infty x^2e^{-ax^2}dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}}; \quad a > 0$

(iii) $\int_0^\infty \text{sinc}\; xdx = \int_0^\infty \text{sinc}^2 xdx = \frac{1}{2}$

(iv) $\int_0^\infty \frac{x\sin(ax)}{(b^2+x^2)}dx = \frac{\pi}{2}e^{-ab}; \quad a \text{ and } b > 0$

(v) $\int_0^\infty \frac{\cos(ax)}{(b^2+x^2)}dx = \frac{\pi}{2b}e^{-ab} \quad a \text{ and } b > 0$

(vi) $\int_0^\infty e^{-ax}\cos(bx)dx = \frac{a}{a^2+b^2}; \quad a > 0$

(vii) $\int_0^\infty e^{-ax}\sin(bx)dx = \frac{b}{a^2+b^2}; \quad a > 0$

(viii) $\int_0^\infty e^{\pm j2\pi yx}dx = \delta(y)$

(ix) $\int_0^\infty \frac{x^{m-1}}{1+x^n}dx = \frac{\pi/n}{\sin(m\pi/n)}; \quad n > m > 0$

8. Trigonometric Identities

(i) $e^{\pm jx} = \cos x \pm j \sin x$

(ii) $\cos x = \dfrac{e^{jx} + e^{-jx}}{2}$

(iii) $\sin x = \dfrac{1}{2j}(e^{jx} - e^{-jx})$

(iv) $\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$

(v) $\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$

(vi) $\cos 2x = \cos^2 x - \sin^2 x$

(vii) $\sin 2x = 2 \sin x \cos x$

(viii) $\cos^2 x = \dfrac{1}{2}(1 + \cos 2x)$

(ix) $\sin^2 x = \dfrac{1}{2}(1 - \cos 2x)$

(x) $\cos 3x = 4 \cos^3 x - 3 \cos x$

(xi) $\sin 3x = 3 \sin x - 4 \sin^3 x$

USEFUL MATHEMATICAL AND PHYSICAL CONSTANTS

Appendix

B

1. Mathematical Constants

Pi (π)	$\pi = 3.1415927$
Base of natural logarithm	$e = 2.7182818$
Logarithm of *e* to base 2	$\log_2 e = 1.442695$
Logarithm of 2 to base 10	$\log_{10} 2 = 0.30103$

2. Physical Constants

Boltzmann's constant	$k = 1.38 \times 10^{-23}$ Joule/degree Kelvin
Plank's constant	$h = 6.625 \times 10^{-34}$ Joule-second.
Charge of an electron	$e = 1.602 \times 10^{-19}$ coulomb
Speed of light in vacuum	$c = 2.998 \times 10^{8}$ meters/second
Thermal energy kT_0 at standard room temperature of $T_0 = 273°\text{K}$	$kT_0 = 3.77 \times 10^{-21}$ Joule

FOURIER TRANSFORM THEOREMS AND TRANSFORM PAIRS

Appendix C

1. Useful Theorems

Theorem	Function	Transform
Linearity	$a_1x_1(t)+a_2x_2(t)$	$a_1X_1(f)+a_2X_2(f)$
Time-delay	$x(t-\tau)$	$X(f)e^{-j\omega\tau}$
Scale change	$x(at)$	$\frac{1}{\lvert a\rvert}X(f/a)$
Conjugation	$\overline{x}(t)$	$\overline{X}(-f)$
Duality	$X(t)$	$x(-f)$
Modulation	$x(t)e^{j2\pi f_c t}$	$X(f-f_c)$
Differentiation	$\frac{d}{dt}x(t)$	$j2\pi fX(f)$
Integration	$\int_{-\infty}^{t} x(\tau)d\tau$	$\frac{1}{j2\pi f}X(f)+\frac{1}{2}X(0)\delta(f)$
Convolution	$x(t) * y(t)$	$X(f)\cdot Y(f)$
Multiplication	$x(t)\cdot y(t)$	$X(f) * Y(f)$
Parseval's or Rayleigh's theorem	$\int_{-\infty}^{\infty} \lvert x(t)\rvert^2 dt = E_x = \int_{-\infty}^{\infty} \lvert X(f)\rvert^2 df$	
Generalized Parseval's theorem	$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}df$	

2. Basic Fourier Transform Pairs

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt \quad \text{and} \quad x(t) = \int_{-\infty}^{\infty} X(f)e^{j\omega t}\,df$$

S. No.	Signal in time domain	Signal in frequency domain
1.	$x(t) = \delta(t)$	$X(f) = 1$
2.	$x(t) = 1$	$X(f) = \delta(f)$
3.	$x(t) = u(t)$	$X(f) = \frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$
4.	$x(t) = e^{j(\omega_0 t + \phi)}$	$X(f) = e^{j\phi}\delta(f - f_0)$
5.	$x(t) = \text{sgn}(t)$	$X(f) = \frac{1}{j\pi f}$
6.	$x(t) = \cos(\omega_0 t + \phi)$	$X(f) = \frac{1}{2}\left[e^{j\phi}\delta(f - f_0) + e^{-j\phi}\delta(f + f_0)\right]$
7.	$x(t) = e^{-at}u(t)$	$X(f) = \frac{1}{a + j2\pi f}$
8.	$x(t) = e^{-a\|t\|}$	$X(f) = \frac{2a}{a^2 + (2\pi f)^2}$
9.	$x(t) = A\Pi(t/\tau)$	$X(f) = A\tau \text{ sinc } f\tau$
10.	$x(t) = \text{sinc } 2Wt$	$X(f) = \frac{1}{2W}\Pi(f/2W)$
11.	$x(t) = A\Lambda(t/\tau)$	$X(f) = A\tau \text{ sinc}^2 f\tau$
12.	$x(t) = \text{sinc}^2 2Wt$	$X(f) = \frac{1}{2W}\Lambda(f/2W)$

HILBERT TRANSFORM PAIRS

Appendix

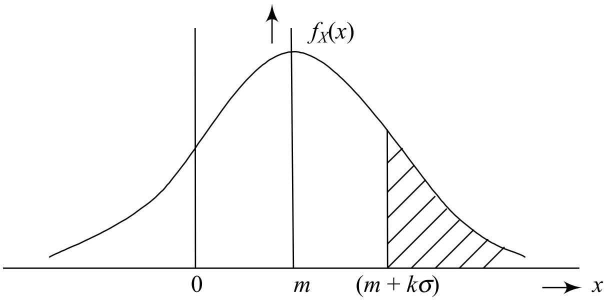
D

	Time Function	Hilbert Transform
1.	$\cos 2\pi f_c t$	$\sin 2\pi f_c t$
2.	$\sin 2\pi f_c t$	$-\cos 2\pi f_c t$
3.	$x(t)\cos 2\pi f_c t$ (When $f_c >> W$, the band limiting frequency of $x(t)$)	$x(t)\sin 2\pi f_c t$
4.	$x(t)\sin 2\pi f_c t$ (When $f_c >> W$)	$-x(t)\cos 2\pi f_c t$
5.	$1/t$	$-\pi\delta(t)$
6.	$(\sin t)/t$	$(1 - \cos t)/t$
7.	$\delta(t)$	$(1/\pi t)$

ERROR FUNCTIONS AND Q–FUNCTIONS

Appendix E

If ***X*** is Gaussian with mean *m* and variance σ^2,



$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2}$$

Probability of ***X*** taking a value greater than $(m + k\sigma)$ is the area under the shaded region and is given by

$$Q(k) \,\underline{\Delta}\, \frac{1}{\sqrt{2\pi}} \int_k^{\infty} e^{-\lambda^2/2} d\lambda ,$$

where $Q(\cdot)$ is called the Q – function.
The error function and complementary error function are defined as follows:

$$erf(k) \,\underline{\Delta}\, \frac{2}{\sqrt{\pi}} \int_0^{k} e^{-\lambda^2} d\lambda = 1 - 2Q(\sqrt{2}k)$$

and

$$erfc(k) \,\underline{\Delta}\, \frac{2}{\sqrt{\pi}} \int_k^{\infty} e^{-\lambda^2} d\lambda = 1 - erf(k) = 2Q(\sqrt{2}k)$$

For $k = 3$, $Q(k)$ may be approximated by

$$Q(k) \cong \frac{1}{\sqrt{2\pi}k} e^{-k^2/2}$$

Error Function Values

k	$erf(k)$	k	$erf(k)$	k	$erf(k)$
0.00	0.0000	0.80	0.74210	1.60	0.97635
0.05	0.05637	0.85	0.77067	1.65	0.98083
0.10	0.11246	0.90	0.79691	1.70	0.98379
0.15	0.16800	0.95	0.82089	1.75	0.98667
0.20	0.22270	1.00	0.84270	1.80	0.98909
0.25	0.27633	1.05	0.86244	1.85	0.99111
0.30	0.32863	1.10	0.88021	1.90	0.99279
0.35	0.37938	1.15	0.89612	1.95	0.99418
0.40	0.42839	1.20	0.91031	2.00	0.99532
0.45	0.47548	1.25	0.92290	2.50	0.99959
0.50	0.52050	1.30	0.93401	3.00	0.99998
0.55	0.56332	1.35	0.94376		
0.60	0.60386	1.40	0.95229		
0.65	0.64203	1.45	0.95970		
0.70	0.67780	1.50	0.96611		
0.75	0.71116	1.55	0.97162		

CONSTRAINED OPTIMIZATION USING LAGRANGE MULTIPLIERS

Appendix F

Constrained optimization is employed quite frequently in all branches of engineering. Here, we briefly explain the way Lagrange multipliers method may be used for constrained optimization.

Suppose the function $x(t)$ is to be minimized or maximized subject to the constraint $c(t) = 0$. We proceed by first forming the function

$$J = x(t) + \lambda c(t),$$

where, λ is called a Lagrange multiplier. Then, sufficient conditions for optimal value of $x(t)$ are:

$$\frac{\partial J}{\partial t} = 0 \text{ and } \frac{\partial J}{\partial \lambda} = 0$$

Note

(i) In case there are several variables, we take a partial derivatives of J with respect to each of those variables.

(ii) In case there are several constraints, we use one Lagrange multiplier to introduce each of the constraints. For example, if $c_1(t) = 0, c_2(t) = 0, \ldots, c_n(t) = 0$ *are n constraints under which the function x(t) has to be optimized, we form the following function*

$$J = x(t) + \lambda_1 c_1(t) + \lambda_2 c_2(t) + \ldots + \lambda_n c_n(t)$$

and the sufficient conditions for optimality of x(t) are given by

$$\frac{\partial J}{\partial t} = 0,\ \frac{\partial J}{\partial \lambda_1} = 0,\ \frac{\partial J}{\partial \lambda_2} = 0, \ldots,\ \frac{\partial J}{\partial \lambda_n} = 0$$

Example F.1 Suppose, we have a wire of length L and using it we want to form a rectangle that encloses maximum possible area.

Let x and y be the length and breadth of the rectangle.

So, we have to maximize its area xy under the constraint that $2(x + y) = L$. So, let us form the function

$$J = xy + \lambda(2x + 2y - L)$$

Then,

$$\frac{\partial J}{\partial x} = y + 2\lambda = 0 \quad \Rightarrow \quad \lambda = -y/2 \tag{i}$$

and

$$\frac{\partial J}{\partial y} = x + 2\lambda = 0 \quad \Rightarrow \quad \lambda = -x/2 \tag{ii}$$

Since $\lambda = -y/2,$ substituting for λ in (ii)

$$x - y = 0 \quad \Rightarrow \quad x = y$$

$$\frac{\partial J}{\partial \lambda} = 2x + 2y - L = 0 \quad \Rightarrow \quad 4x = L \text{ or } x = L/4$$

Hence, we should form a square with each side = $\frac{L}{4}$

SIGNAL FADING AND DIVERSITY RECEPTION

Appendix G

Random variations in the amplitude and phase of a received signal are referred to as 'fading'. If a fading is frequency dependent, such a fading is called 'frequency-selective fading'. Fading is caused most often by the variation with respect to time, of the difference in path lengths of various rays reaching the receiver. In this appendix, we will discuss the phenomenon of fading and the techniques for reducing/eliminating its effect.

G.1 Propagation Characteristics and Fading

We shall first briefly examine how the propagation characteristics sometimes lead to fading in the various frequency ranges.

1. Medium Wave Band (550–1600 kHz): At these frequencies, it is entirely ground wave propagation during the daytime since the sky wave at these frequencies is totally absorbed by the D–layer of the ionosphere during the daytime. Hence, there will not be any fading and a steady signal strength is obtained in the service area during the day. However, at night, the D–layer disappears and a fairly strong sky wave component is present in addition to the ground wave, especially at frequencies near the higher-end of the band. With increase in distance from the transmitter, the ground wave component becomes weaker due to attenuation. But, the sky wave component becomes stronger. Thus, there exists a region in which the two components are approximately equal in strength. Since these two components have traveled by different paths, interference between them can produce a resultant signal whose strength may vary between their sum and difference depending upon their relative phase relationship.

If the ionosphere were time-invariant, the phase relation between the two would be constant and a resultant signal of steady amplitude would be obtained. However, electron density of the ionosphere goes on fluctuating and because of this, the height from which the sky wave component gets reflected, also goes on fluctuating. Thus, the path length of the sky wave component goes on changing continually, affecting the phase relationship between the two components in a random way. Hence, the amplitude and phase of the resultant signal fluctuate causing considerable fading of the received signal. Since phase change of the sky wave component is directly proportional to the change in path length and inversely proportional to the wavelength, the frequencies at the higher end of the medium wave band suffer deeper fading. The duration of the fade will be of the order of a few seconds to tens of seconds.

2. Shortwave Band (1600 kHz–30 MHz): At these frequencies, the ground wave attenuates within a short distance from the transmitter and so is of no consequence. All long distance shortwave communication is therefore only by ionospheric reflections.

As in the case of frequencies near the upper edge of the medium wave band, in the shortwave band too, fading is caused by the interference of wave components. However, while it is the interference between the ground wave component and the sky wave component in the case of the medium wave band, it is the interference between two or more sky wave components (of the same transmitted signal) in the case of the shortwave band.

Figure G.1 shows one possibility of two ionospheric reflected rays taking different paths for arriving at the receiver. Since the height of the reflection point in the ionosphere goes on changing continually, the path lengths of the two rays and also the phase relation between the two rays arriving at the receiver will go on changing randomly. Thus, there will be fading of the received signal and the fading will be more severe at the higher frequencies. This frequency-selective fading can terribly distort an AM signal. However, for SSB signals the distortion will be much less because of their smaller bandwidth and so they will be intelligible.

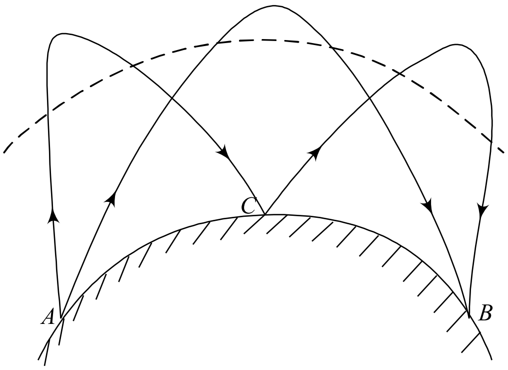


Fig. G.1 *Fading caused by interference between rays with different number of hops*

Although it is a rare occurrence, fading of a sky wave signal may be caused by '***Faraday fading***'. Under certain favorable conditions, a plane polarized wave entering the ionosphere is changed into an elliptically polarized wave due to the earth's magnetic field. Owing to random fluctuations in the electron density in the ionosphere, the direction of the major axis of the ellipse goes on changing causing fading of the signal induced in a vertical receiving antenna.

3. Fading at frequencies beyond 30 MHz: At these frequencies, the ionosphere does not reflect the waves except occasionally due to the presence of the sporadic E-layer which reflects waves at frequencies in the range of 30–60 MHz. Propagation at frequencies above 30 MHz is therefore only by tropospheric or space wave propagation, which is essentially a line-of-sight (LOS) propagation. Elevated antennas are used for extending the range. These frequencies are extensively used for FM broadcasting, TV, terrestrial microwave relays and point-to-point communications including mobile communications.

In LOS propagation, the received signal is the vector sum of the direct-ray and the ground-reflected ray. Sometimes, there may be a third ray too – the one caused by reflection/scattering by irregularities in the troposphere. Local irregularities can change the phase relation between the direct-ray and the ground-reflected ray and this may cause fading.

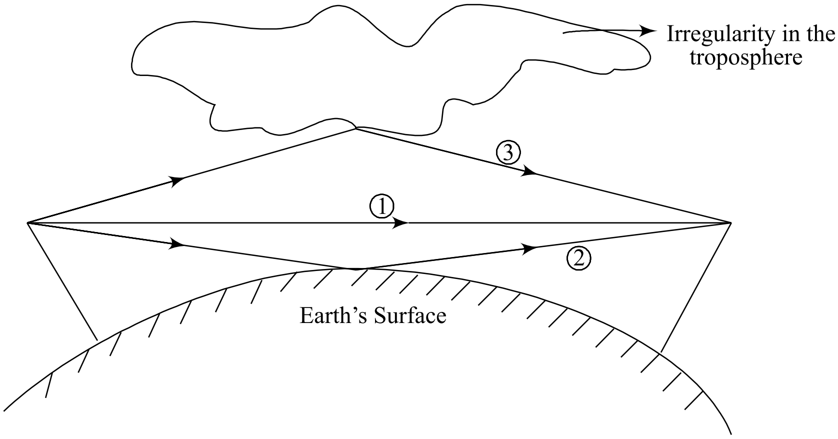


Fig. G.2 *Multipath interference and fading in LOS propagation*

Further, the position of the irregularity in the troposphere that gave rise to the third ray, will go on changing rapidly, causing the path length and so the phase-on-arrival of the third ray to change. This produces rapid fluctuations in the strength of the resultant received signal. Fading is generally most pronounced near the radio horizon and in the shadow region (i.e., beyond the radio horizon).

G.2 Diversity Systems for Combating Fading

Automatic Gain Control (AGC) in AM receivers is meant to minimize fading and it can reduce signal variations from about 30–40 dB to about 3–4 dB. However, the ability of AGC is limited. If the received signal suffers a deep fade that brings it to the same level as noise, the AGC simply raises this noise level by increasing the gain of the RF and IF stages. On the other hand, in such a situation (i.e., received signal suffering a deep fade) a diversity reception system tries to maintain a good signal level at the input to the receiver and so will be quite effective.

FM systems have a certain amount of immunity against fading since the intelligence is carried not by the amplitude of the modulated signal, but its frequency. AGC, if provided in the receiver, is primarily for improving the limiting action and it can provide further immunity from fading. However, here too, if a deep fade takes the received signal strength to such a low level that the receiver operates below the threshold, only noise will be obtained at the output. Further, because of their large bandwidth, wideband FM signals are more susceptible to distortion caused by frequency-selective fading than say, AM or SSB signals.

Diversity reception relies on the principle that if n number of replicas of a given transmitted signal are obtained from n independently fading channels, the probability of all of them fading simultaneously will be p^n where p is the probability of fading of any one of the channels. Different diversity systems differ in the way they try to provide the n independently fading channels. Different types of diversity are:

1. Space diversity
2. Frequency diversity
3. Time diversity
4. Polarization diversity

Of these, the first two, i.e., space diversity and frequency diversity are most widely used. We will now discuss briefly the salient features of these two types.

1. Space Diversity: In this type of diversity, independently fading channels are obtained by using a number of receiving antennas spaced about 3 to 30 wavelengths apart. For the shortwave band of 3 MHz to 30 MHz, the antennas are spaced about 3 to 10 wavelengths apart. Each antenna is connected to a separate receiver; but all these receivers share a common local oscillator, a detector, an AGC system and the audio amplifiers. The receiver with the maximum signal at its input at a given time, produces maximum AGC bias and hence, at that instant, it contributes maximum signal to the combined output; and the contribution of all the other receivers is negligible. Even if there is severe fading with a single antenna system (i.e., no diversity), a space diversity system with two antennas will ensure that there is no noticeable fading. By using a common local oscillator, we are ensuring that all the receivers are tuned to the same transmitted signal. For microwave links using space diversity, the two antennas are generally mounted on the same tower, but at different heights, keeping adequate separation between the two.

2. Frequency Diversity: In frequency diversity, a given message signal at the transmitter is made to modulate two separate carrier frequencies whose frequency separation is more than the ***coherence bandwidth****[*] of the

---

Coherence bandwidth B_c of a multipath channel is the inverse of T_m, where T_m is the multipath time-spread of the channel which represents the time delay between the arrival of the first and the last multipath signal components. For HF ionospheric channel, B_c, is typically 10^2 to 10^3 Hz. Coherence time of a channel, T_c, on the other hand, is the inverse of the Doppler spread B_d of the channel. A typical value of the coherence time of HF ionospheric channel is 1 to 10 seconds.

channel. This frequency separation ensures that the two channels fade independently. The single receiving antenna that is used at the receiving end feeds two separate receivers having separate RF amplifiers, mixers and local oscillators and a common set of IF amplifier, detector, AGC system and audio amplifiers.

Both these diversity systems are widely used in commercial as well as military communications. However, since frequency diversity is wasteful of the frequency spectrum, space diversity is generally preferred wherever it is feasible to use it. Because of space constraint, frequency diversity however, has to be used in all ship-to-ship and ship-to-shore HF communications. For certain types of communications like the troposcatter communication links which suffer severe fading, both space and frequency diversity are simultaneously used. Such an arrangement is called '***quadruple diversity***' system.

G.3 Digital Communication through Multipath Fading Channels

As we have already seen, fading is caused by multipath propagation because of which the signal arrives at the receiver through paths of different lengths. The different path lengths cause the various multipath components to arrive at the receiver with different phases. The number of paths, the path lengths and so the phases of the multipath components will change with time owing to changes in the medium. The resultant signal at the receiver, which is the vector sum of the various multipath components can therefore, at times, take one of the two possible extreme limits – one corresponding to the constructive interference of the multipath components, and the other corresponding to their destructive interference. This results in the phenomenon called fading. Such channels are therefore characterized by time-varying impulse response models. Fading always leads to higher values of average probability of error and thus deteriorates the performance of any digital communication system. As in the case of analog communication systems operating over such channels, here too, diversity reception can be used to improve the system performance. Examples of multipath fading channels are the ones involved in line-of-sight microwave links, cellular mobile communications and underwater acoustic communication.

Characterization of multipath fading channels In a multipath fading channel, changes in the medium and therefore the changes in the impulse response of the channel are rather unpredictable making it necessary to characterize the channels statistically.

Assume that an unmodulated carrier, $A\cos\omega_c t$, is transmitted through a multipath fading channel. It will then travel by different paths and each multipath component suffers attenuation as well as time-delay, both of which are time-dependent. Hence, ignoring for a moment, the additive noise, the received signal may be expressed as

$$\begin{aligned} x(t) &= A\sum_n a_n(t)\cos[2\pi f_c t - \tau_n(t)] \\ &= A\,\mathrm{Re}\left[\left\{\sum_n a_n(t)e^{-j2\pi f_c\tau_n(t)}\right\}e^{j2\pi f_c t}\right] \end{aligned} \tag{G.1}$$

where $a_n(t)$ and $\tau_n(t)$ are respectively the time-varying attenuation factor and the time-varying propagation delay for the n^{th} path. From Eq. (G.1) it is clear that the equivalent lowpass complex-valued received signal is

$$\begin{aligned} y(t) &= \sum_n a_n(t)e^{-j2\pi f_c\tau_n(t)} \\ &= \sum_n a_n(t)e^{-j\theta_n(t)} \end{aligned} \tag{G.2}$$

Equation (G.2) reveals two things:

1. Since $a_n(t)$ and $\theta_n(t)$ are both time-varying, $y(t)$, the response of the channel to $\exp(j2\pi f_c t)$ has many different frequency components, although $\exp(j2\pi f_c t)$ itself is a single-frequency signal. The bandwidth of $y(t)$ is called the Doppler frequency spread, B_d, of the channel and it indicates how rapidly $y(t)$ is changing with time.
2. The equivalent lowpass channel has a time-varying impulse response given by

$$h(\tau; t) = \sum_n a_n(t) e^{-j2\pi f_c \tau_n(t)} \delta(\tau - \tau_n(t)) \tag{G.3}$$

where, $h(\tau; t)$ represents the response of the channel at time t due to an impulse applied at $(t - \tau)$. Since $a_n(t)$ and $\tau_n(t)$ vary randomly with time, $y(t)$ as well as $h(\tau; t)$ are complex-valued random processes. If n, the number of propagation paths is very large, central limit theorem can be applied. In that case, $h(\tau; t)$ can be modeled as a complex-valued Gaussian random process. Then, at any instant of time, $|h(\tau; t)|$, the envelope of $h(\tau; t)$, will be Rayleigh distributed and the channel itself is referred to as a ***Rayleigh fading channel***. However, if in addition to the randomly moving ones, there are some fixed scatterers too in the medium, as sometimes happens in cellular mobile communication, then $h(\tau; t)$ will have a non-zero mean and it will have an envelope with Ricean distribution. In that case, the channel is referred to as a ***Ricean channel***.

Channel parameters Because of the difference in path lengths, the various multipath signal components arrive at the receiver at different points in time. The difference in arrival times of the first-to-arrive and the last-to-arrive multipath signal components is called the multipath delay spread and is denoted by T_m. Besides the ***Doppler frequency spread***, B_d, which we had discussed earlier and the multipath delay spread T_m, there are two more parameters that are useful in characterizing multipath fading channels. These parameters are the ***coherence time*** T_c and the ***coherence bandwidth*** B_c of the channel.

The coherence time T_c is defined as the inverse of the Doppler frequency spread B_d.

i.e.,

$$T_c \triangleq \frac{1}{B_d} \tag{G.4}$$

Coherence bandwidth, B_c, of the channel is defined as the inverse of the multipath delay spread, T_m.

i.e.,

$$B_c \triangleq \frac{1}{T_m} \tag{G.5}$$

If a signal transmitted through a multipath fading channel has a bandwidth less than the coherence bandwidth, B_c of the channel, then all the different frequency components of the signal fade simultaneously and similarly and hence the channel is said to be '***frequency non-selective***'. On the other hand, if B_c is less than the signal bandwidth, signal components separated in frequency by more than B_c will be attenuated and phase-shifted differently, and the channel is said to be ***frequency selective***. If the coherence time T_c is larger than the symbol period, the channel is said to be a ***slow-fading channel*** and if T_c is less than the symbol period, the channel is said to be a ***fast-fading channel***.

Modeling of multipath fading channels A time-variant multipath channel is generally modeled as shown in Fig. G.3 as a tapped delay-line with uniformly spaced taps. If W is the bandwidth of the signal transmitted through the channel, then the tap spacing is $(1/W)$. The tap coefficients are modeled as complex-valued Gaussian processes which are mutually uncorrelated, and

$$c_n(t) = a_n(t) e^{-j\theta_n(t)} \tag{G.6}$$

The total length of the tapped delay-line is the multipath delay spread, T_m, and is equal to (L/W) where L is the number of signal paths.

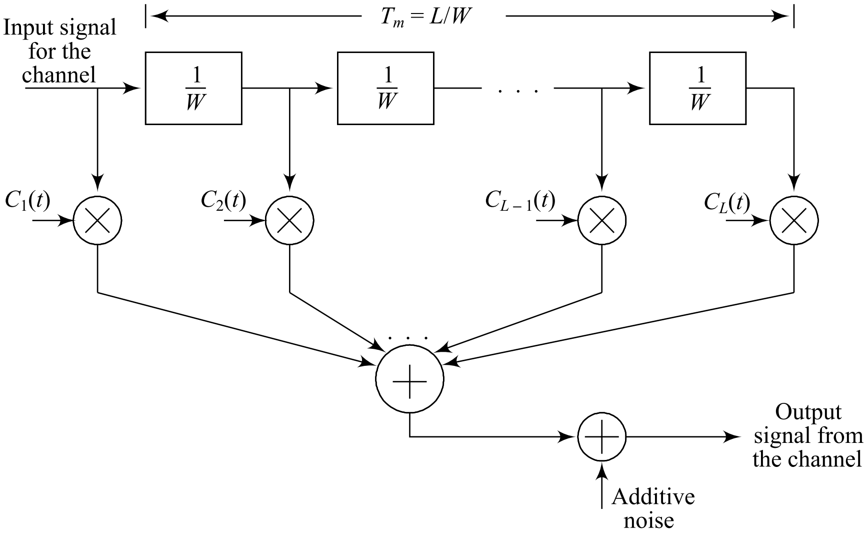


Fig. G.3 *Tapped delay-line model for a time variant multipath fading channel*

Digital modulations for transmission over multipath fading channels Although a number of different types of digital modulation schemes are available and at least theoretically can be used on any channel, use of ASK and QAM is generally avoided for transmission over multipath fading channels. This is because, it is extremely difficult to distinguish between two adjacent amplitude levels when the received signal amplitude itself is having large fluctuations due to fading. For this reason, only FSK and PSK are used for signals to be transmitted over fading multipath channels.

Performance of BPSK and orthogonal BFSK Analysis of the performance of BPSK and orthogonal BFSK modulation when used for transmission over a frequency non-selective (i.e., $B_c > W$) Rayleigh fading channel shows that the average probabilities of error work out to

$$P_e = \frac{1}{2}\left[1 - \sqrt{\frac{\overline{\rho_b}}{1+\overline{\rho_b}}}\right] \qquad \text{for BPSK} \tag{G.7}$$

and

$$P_e = \frac{1}{2}\left[1 + \sqrt{\frac{\overline{\rho_b}}{2+\overline{\rho_b}}}\right] \qquad \text{for orthogonal BFSK} \tag{G.8}$$

where $\overline{\rho_b}$ is the average *SNR*/bit. For large values of $\overline{\rho_b}$, these probabilities of error can be approximated by

$$P_e \approx \frac{1}{4\overline{\rho_b}} \qquad \text{for BPSK} \tag{G.9}$$

and

$$P_e \approx \frac{1}{2\overline{\rho_b}} \qquad \text{for orthogonal BFSK} \tag{G.10}$$

Note that orthogonal BFSK is inferior to BPSK by 3 dB. Further while the probabilities of error for these modulations decreased exponentially with the *SNR*/bit in the case of a purely AWGN channel, in the case of transmission over Rayleigh fading channels, the probabilities of error are decreasing only inversely with the *SNR*/bit. This shows a clear deterioration of performance due to fading.

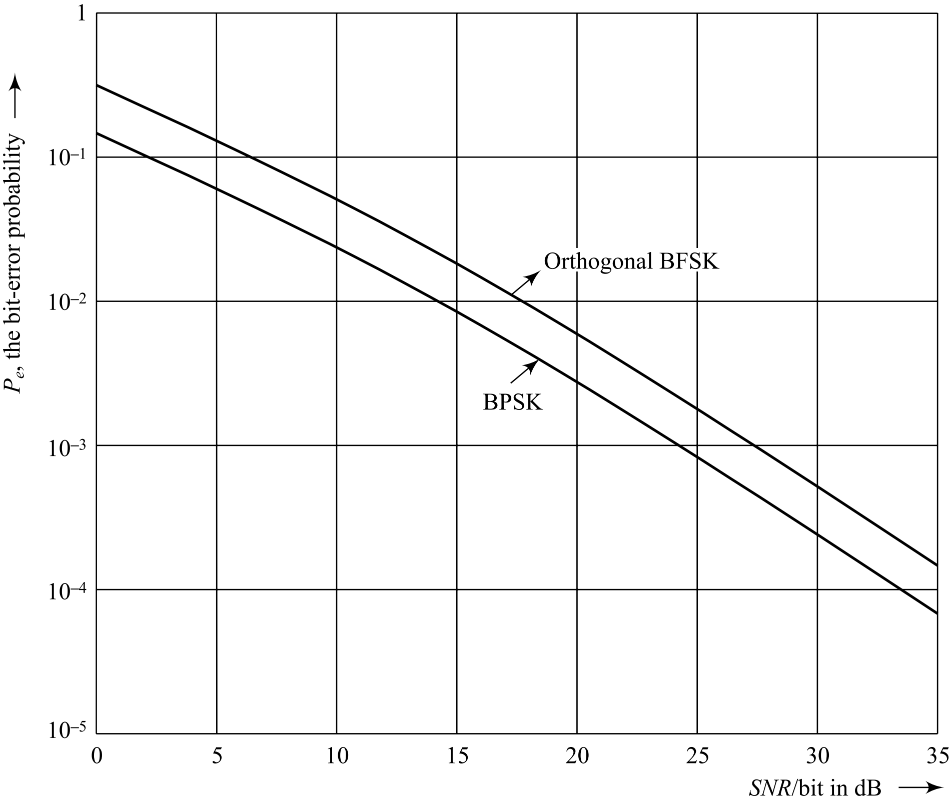


Fig. G.4 *Performance of BPSK and Orthogonal BFSK on a frequency non-selective Rayleigh fading channel*

Diversity reception for performance improvement Fading always causes performance degradation. However, performance improvement is possible through the use of diversity reception, in which we supply to the receiver *n* number of replicas of the signal transmitted through *n* independently fading channels, as described earlier. While one may use different methods like space diversity or frequency diversity, etc. to obtain *n* independently fading channels, the way these independently fading signals are used/combined in the receiver, is also important.

There are different methods of combining these signals. What we described earlier in connection with space diversity, is a simple one which is easy to implement. In that, we arranged matters so that at any instant of time, only the received signal that was strongest among all at that moment was allowed to contribute to the output of the receiver. However, there are better methods of combining, although they are quite complex. These are:

1. Equal gain combiner: In this type of combiner, which is quite suitable for coherent demodulation and detection, the receiver estimates the phase offsets of the *n* received signals after they are demodulated. These phase corrected signals are then summed up and their sum is applied as input to the detector.

2. Maximal ratio combiner: As in the equal-gain combiner, in this combiner also the phase offsets are estimated and corrected after demodulation. In addition, the received signal powers are also estimated and the phase-corrected signals from the demodulators are then weighted in proportion to their respective signal strengths (square roots of the powers) and then the sum of these weighted signals is applied as input to the detector. This method also is quite suitable for coherent demodulation and detection.

3. Square-law combiner: If the receiver is using non-coherent demodulation, as for instance when orthogonal signals are used for transmitting the information over a number of independently fading channels, the square-law combiner may be used. In this, the outputs of the non-coherent demodulators are squared, summed up and then given as input to the detector.

Improvement in performance with diversity ($n > 1$), is seen in terms of the reduction in *SNR*/bit achieved as compared to the case with no diversity (i.e., $n = 1$), for any specified error probability. It is true that as n, the number of independently fading channels used in the diversity system is increased, the improvement in performance also increases. However, while $n = 2$ gives considerable improvement and $n = 4$ gives some more, the additional improvement achieved by using still larger values of n is quite small and is perhaps not worth it, keeping in view the additional complexity involved. Figure G.5 illustrates the performance improvement with n for BPSK with maximal-ratio combiner and for orthogonal BFSK with square-law combiner. In fact, all these combiners provide an exponential decrease in bit-error probability with n.

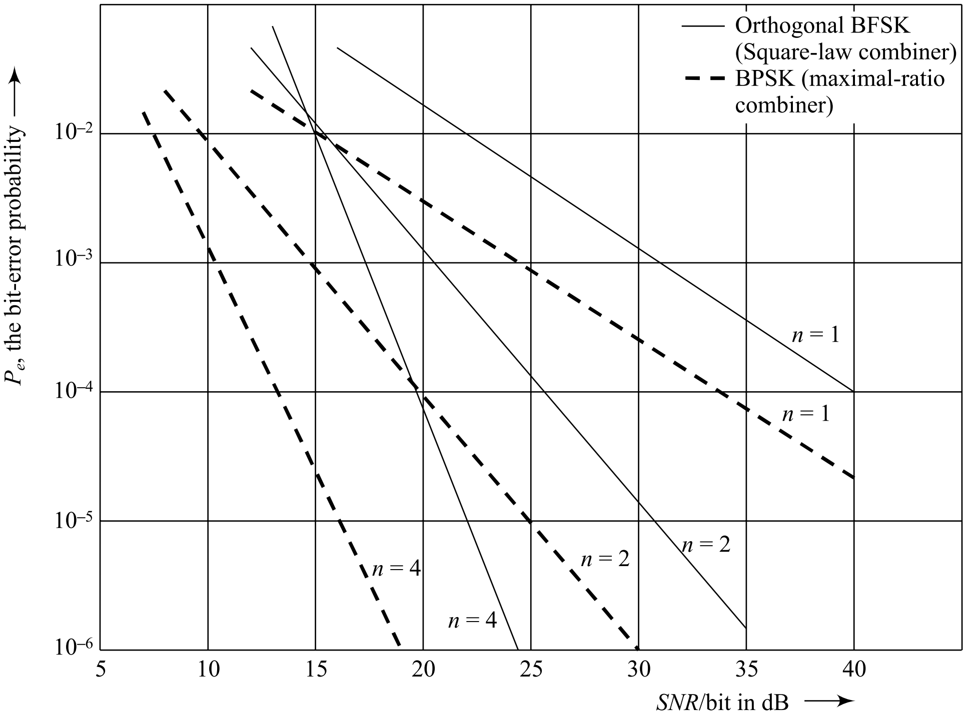


Fig. G.5 *Performance of BPSK and orthogonal BFSK with diversity*

Suggested Reading

1. Price, R., and Green, P.E., Jr., A Communication Technique for Multipath Channels, *Proc. IRE, 1958*, Vol. 46, pp. 555–570.
2. Proakis, John G., and Masoud Salehi, *Communication Systems Engineering*, 2nd Edition, Prentice-Hall of India Pvt. Ltd. New Delhi, 2006.
3. Schwartz, M., Bennet, W.R., and Stein, S., *Communication Systems and Techniques*, McGraw-Hill, New York, 1966.

INDEX