Analog Communication

About the Author

P Ramakrishna Rao is presently Advisor, Raghu Engineering College, Visakhapatnam. He holds a BE (Telecommunications) degree from Madras University, an ME degree from IISc, Bangalore, and a PhD from IIT Kanpur. He retired as a professor from REC (now NIT) Tiruchirapalli, and has over 45 years of teaching and 25 years of research experience. At REC Tiruchirapalli, he had carried out several DOE and DRDO-funded research projects.

Dr Rao has written over 25 research publications in reputed international journals and proceedings of international conferences. He was a member of the National Radar Council's Working Group on Sonar and Underwater Electronics from 1987 to 1993. He was also the recipient of the 'Academic Excellence Award' at REC Tiruchirapalli in 1997. Apart from this, he was a reviewer for IEEE Transactions on ASSP and IEE Part-F.

Dr Rao's areas of research interest include Array Signal Processing and Communication. His book *Signals and Systems* was published under the TMH banner in 2008 and has done extremely well.

Analog Communication

P Ramakrishna Rao

Formerly Professor
Department of Electronics and Communication Engineering
National Institute of Technology
Tiruchirapalli, Tamil Nadu

Tata McGraw Hill Education Private Limited
NEW DELHI

McGraw-Hill 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

Tata McGraw-Hill

Published by Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008.

Analog Communication

ISBN (13): 978-0-07-070480-0
ISBN (10): 0-07-070480-5

Vice President and Managing Director—McGraw-Hill Education, Asia-Pacific Region: *Ajay Shukla*

Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Publishing Manager (SEM & Tech. Ed.): *Shalini Jha*
Senior Editorial Researcher: *Koyel Ghosh*
Executive—Editorial Services: *Sohini Mukherjee*
Senior Production Managers: *Satinder S Baveja, P L Pandita*
Marketing Manager (SEM & Tech. Ed.): *Biju Ganesan*
Senior Product Specialist (SEM & Tech. Ed.): *John Mathews*

General Manager—Production: *Rajender P Ghansela*
Assistant General Manager—Production: *B L Dogra*

Typeset at Tulyasys Technologies, No. 1, Arulananthammal Nagar, Thanjavur 613 007, and printed at
Magic International Pvt Ltd., Plot No 26E, Sector 31, Industrial Site IV, Greater Noida.

Cover Printer: Magic International Pvt Ltd.

RYLQCRCCRQBLY

The McGraw·Hill Companies

Dedicated to...

My wife

Kameshwari,

And our children

Ramani, Shankar and Ramana Murthy

CONTENTS

Preface xi
Acknowledgements xiii
Tata McGraw-Hill's 360° Development Process xv

1. Introduction **1**
1.1 Historical Perspective *1*
1.2 An Overview of Communication Systems *2*
1.3 Electromagnetic Spectrum, Ranges and Application Areas *4*
1.4 Channel Types, Characteristics and Modelling *4*
References and Suggested Reading 12

2. Signals, Transforms and Spectral Analysis **13**
2.1 Introduction *13*
2.2 Signals *14*
2.3 Unit Step and Unit Impulse Functions *17*
2.4 Continuous-Time Fourier Series *20*
2.5 The Complex-Exponential Fourier Series *24*
2.6 Trignometric Fourier Series *30*
2.7 Convolution and Correlation of Signals *35*
2.8 The Continuous-Time Fourier Transform *41*
2.9 Energy Spectral Density and Power Spectral Density *59*
2.10 Hilbert Transform *66*
Summary 76
References and Suggested Reading 81
Review Questions 81
Fill in the Blanks 82
Problems 82
Multiple Choice Questions 88
Key to Multiple Choice Questions 91

3. Signal Transmission through Systems **92**
3.1 Introduction *92*
3.2 Review of LTI System Theory *93*
3.3 Signal Transmission through LTI Systems *105*
3.4 Paley–Wiener Criterion for Physical Realizability *111*
3.5 System Bandwidth and Rise Time *112*
Summary 115
References and Suggested Reading 117
Review Questions 117
Problems 118
Objective Type Questions 119
Multiple Choice Questions 120
Key to Multiple Choice Questions 122

4. Amplitude Modulation **123**
4.1 Introduction *123*
4.2 Amplitude Modulation *124*
4.3 Demodulation of AM Signals *146*
4.4 Double Sideband Suppressed Carrier (DSB-SC) Modulation *156*

4.5 Single Sideband Modulation *168*
4.6 Vestigial Sideband Modulation *182*
Summary 190
References and Suggested Reading 193
Review Questions 194
Fill in the Blanks 195
Multiple Choice Questions 196
Problems 197
Key to Multiple Choice Questions 198

5. Angle Modulation **199**
5.1 Introduction *199*
5.2 Angle-Modulated Signals *200*
5.3 Narrowband Angle Modulation *210*
5.4 Spectrum of an Angle-Modulated Signal *213*
5.5 Power of an Angle-Modulated Signal and Effective Bandwidth *218*
5.6 Generation of Wideband Angle-Modulated Signals *227*
5.7 Effects of Channel Non-Linearities on FM Signals *234*
5.8 Detection of FM Signals *236*
5.9 FM Broadcasting *247*
Summary 249
References and Suggested Reading 251
Review Questions 251
Fill in the Blanks 252
Multiple Choice Questions 253
Problems 254
Key to Multiple Choice Questions 256

6. AM and FM Transmitters and Receivers **257**
6.1 Introduction *257*
6.2 AM Broadcast Transmitters *259*
6.3 AM Broadcast Receivers *262*
6.4 SSB Transmitters *276*
6.5 SSB Receivers *278*
6.6 FM Transmitters and Receivers *281*
Summary 284
References and Suggested Reading 288
Review Questions 288
Fill in the Blanks 289
Multiple Choice Questions 290
Problems 291
Key to Multiple Choice Questions 292

7. Probability and Random Processes **293**
7.1 Introduction *293*
7.2 Basics of Probability *294*
7.3 Conditional Probability *294*
7.4 Random Variables *296*
7.5 Functions of a Single Random Variable *302*
7.6 Mean, Variance and Characteristic Function *303*
7.7 Functions of Two Random Variables *306*
7.8 Jointly Gaussian Random Variables *309*

7.9 Central Limit Theorem *309*
7.10 Random Processes *311*
7.11 Stationarity, Auto-correlation and Power Spectrum *313*
7.12 LTI Systems with Random Processes as Inputs *317*
7.13 Representation of Band-Limited and Bandpass Processes *321*
Summary 324
References and Suggested Reading 328
Review Questions 328
Fill in the Blanks 329
Multiple Choice Questions 330
Problems 332
Key to Multiple Choice Questions 334

8. Noise **335**
8.1 Introduction *335*
8.2 Noise Sources and Classification *336*
8.3 Atmospheric Noise *336*
8.4 Thermal Noise *337*
8.5 Shot Noise *345*
8.6 Noise Equivalent Bandwidth of a Filter *349*
8.7 Noise Figure and Equivalent Noise Temperature of 2-Port Networks *353*
8.8 Narrow-Band Noise Representation *364*
Summary 365
References and Suggested Reading 367
Review Questions 367
Fill in the Blanks 367
Multiple Choice Questions 368
Problems 369
Key to Multiple Choice Questions 370

9. Noise Performance of AM and FM Systems **371**
9.1 Introduction *371*
9.2 Destination SNR of a Base-Band System *372*
9.3 Model for Linear Modulation Systems *372*
9.4 $(S/N)_D$ for SSB-SC Systems *374*
9.5 DSB-SC Systems *375*
9.6 AM Systems *377*
9.7 Noise Performance of Frequency Modulated Systems *389*
9.8 Pre-emphasis and De-emphasis *396*
9.9 Threshold Effect in FM *401*
Summary 409
References and Suggested Reading 411
Review Questions 411
Fill in the Blanks 411
Multiple Choice Questions 412
Problems 413
Key to Multiple Choice Questions 415

10. Sampling and Analog Pulse Modulation **416**
10.1 Introduction *416*
10.2 Sampling of Band-Limited Lowpass Signals *417*
10.3 Lowpass Sampling Theorem *419*

10.4 Ideal or Impulse Sampling *421*
10.5 Reconstruction *427*
10.6 Sampling Using a Sequence of Pulses *428*
10.7 Practical Sampling *431*
10.8 Anti-Aliasing and Reconstruction Filters *434*
10.9 PAM and Time-Division Multiplexing *438*
10.10 Pulse-Time Modulation Systems *444*
10.11 Noise Performance of Analog Pulse Modulation Systems *452*
Summary 459
References and Suggested Reading 460
Review Questions 461
Fill in the Blanks 461
Multiple Choice Questions 462
Problems 463
Key to Multiple Choice Questions 464

Appendix–A *Matlab Programs 465*
Appendix–B *Mathematical Formulae 485*
Appendix–C *Values of Useful Mathematical and Physical Constants 487*
Appendix–D *Hilbert Transform Pairs 488*
Appendix–E *Fourier Transform 488*
Appendix–F *Error Functions and Q–Functions 490*
Index **491**

PREFACE

'Communication Systems' is a core subject for all Electronics and Communication Engineering (ECE) students at the undergraduate level. In most Indian universities, it is generally in the third year of a four-year undergraduate programme that the subject is taught. It is covered in two semesters—analog communication is taught first, and digital communication thereafter. In keeping with the course trend, the content that makes up communication systems has been spread over two stand-alone texts, one on *Analog communication* and the other on *Digital Communication*. Both are designed to serve as comprehensive textbooks at the undergraduate level and cover the respective courses as taught in most Indian universities.

I had the privilege of teaching one-semester courses on analog communication as well as digital communication at NIT Tiruchirapalli (formerly REC) for several years and also for a few years at colleges affiliated to Osmania University, Hyderabad and Karnataka University, Dharwar. The present book, *Analog Communication*, has evolved out of the lecture notes prepared while teaching those courses.

A subject like analog communication systems can be treated at different levels. At one end, it may be dealt with in a purely theoretical way, while at the other end, it may be discussed at the circuit level. I have, in this book, adopted an approach, which I feel, is best suited at the undergraduate level—a suitable mix of theory and circuit diagrams. Concepts are clearly explained using simple and lucid language; mathematical analysis has been used wherever it was found necessary, and the results and their implications have been clearly explained. The block-diagram approach has been used to explain the operation of systems, and circuit diagrams have been used wherever necessary.

Linear system theory, spectral analysis, probability and random processes are some of the essential pre-requisites for a course on analog communication. In most of the universities, these are covered through separate one-semester courses on 'Signals and Systems' and 'Probability and Random Processes'. However, for the sake of completeness and for ready reference, 'Signals, Transforms and Spectral Analysis', 'Signal Transmission through Systems' and 'Probability and Random Processes' have been covered in this book in a review-like approach, by devoting one chapter separately to each one of them.

Scope of the Book

The contents of the book have been so designed that the book covers, almost fully, the prescribed syllabus for a one-semester course on analog communication of almost all Indian universities. Further, three full chapters have been devoted to cover the topics that provide the necessary background to the reader who might not have been exposed to those topics earlier.

The book will therefore be useful for

(i) All engineering undergraduate students specializing in Electronics and Communication Engineering, or Electrical and Electronics Engineering, or Electronics and Instrumentation Engineering, or Computer Science and Engineering
(ii) All candidates preparing for IETE examinations
(iii) All candidates preparing for Institution of Engineers (India) examinations
(iv) All those preparing for GATE and similar competitive examinations
(v) Practicing engineers, as a reference

Chapter Organization

Chapter 1 gives a brief history of the evolution of communication systems and an overview of analog communication systems together with characteristics of various types of communication channels and their modeling. Signal analysis, transforms and spectral analysis are reviewed in *Chapter 2*. *Chapter 3* discusses the topics of linear system theory and the transmission of signals through systems, in a review like fashion. AM, DSB-SC, SSB-SC and VSB methods of modulation, time-domain and frequency-domain representation of these modulated signals, their methods of generation and detection as well as fields of application are discussed in detail in *Chapter 4*. *Chapter 5* discusses the time and frequency-domain representation, methods of generation and detection and fields of application of frequency and phase-modulated signals.

Transmitters and receivers, their types, specifications, principles of operation, and operational characteristics have been discussed in *Chapter 6*. A review of probability theory and random processes is presented in *Chapter 7*. *Chapter 8* discusses various sources of noise, their spectral characteristics and techniques for noise calculations in active and passive circuits. Performance of AM, DSB-SC, SSB-SC and FM communication systems in the presence of noise, and comparison of these systems on the basis of their noise performance, have been presented along with certain techniques for reducing the effect of noise on the performance of FM systems, in *Chapter 9*. In the last chapter, i.e., *Chapter 10*, various analog pulse modulation systems like PAM, PDM and PPM, have been presented along with their noise performance.

Appendix A is dedicated to MATLAB problems and examples. *Appendix B* gives useful mathematical identities and formulae and *Appendix C* gives values of commonly used physical and mathematical constants. *Appendix D* provides a table of Hilbert transform pairs; and *Appendix E* gives tables of Fourier transform properties and pairs. *Appendix F* presents tables of error function.

How to use the Book

The book can serve as a good textbook for a one-semester course on 'Analog Communication'. Those who have had good exposure to signals and systems, probability theory and random processes through a full-fledged one-semester course for each, may revisit chapters 2, 3 and 7 and refresh the concepts. For the others, these chapters, together with the books recommended for reference at the end of these chapters, lay a solid foundation for a proper understanding of the material presented in the remaining chapters.

Salient Features

From the pedagogical point of view, several useful features have been used in this book. These include

1. Clear explanation of concepts using easily understandable *simple language and style*
2. Focused coverage of *Transmitters and Receivers*
3. In-depth discussion on *noise* and *noise performance of AM and FM systems*
4. Inclusion of relevant MATLAB examples
5. About 165 *worked-out examples*, carefully selected to reinforce the understanding of concepts and to illustrate the way the tools developed can be used for solving problems
6. Clearly stated *learning objectives* at the beginning of each chapter to guide and motivate the student
7. *Summary* at the end of each chapter to reinforce the learning objectives and summarize the concepts
8. A large number of appropriately selected *problems* at the end of each chapter, close to 150, to enable the student apply the techniques learned
9. *Review Questions* at the end of each chapter, totaling over 190, to test the student's understanding of the key concepts
10. *Objective Questions* and *Multiple Choice Questions* (with key) at the end of each chapter, numbering over 360, to drill in the concepts and tools

Web Supplements

The text is accompanied by a wealth of web supplements and may be accessed at https://www.mhhe.com/rao/ac

ACKNOWLEDGEMENTS

A lot of effort had to be put in during the last 18 months in writing this book. During this, time, I received help, guidance and support from many and am deeply indebted to all of them. I am grateful to Prof. Prasad Gandikota of ANITS, Visakhapatnam, and Prof. P Parvatheesam for helping with the MATLAB examples. I thank Mr Y Rajani Kanth who has typed the manuscript copy and generated all the line diagrams neatly and accurately.

I am thankful to Tata McGraw-Hill for encouraging me to write this book. The constant cooperation and helpful suggestions from the staff of Tata McGraw-Hill have been largely responsible in shaping the book to its present form and in the timely completion of the project. In this context, a particular mention must be made of the very helpful role played by Ms Koyel Ghosh, the editor for this book.

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

All efforts have been made to make the book error-free; nevertheless, some errors might have crept in. Suggestions / criticisms from readers for improvement of the book are most welcome at prakriya37@hotmail.com.

P Ramakrishna Rao
Visakhapatnam

Publisher's Note

The best ideas come from you! So, please send your comments, views and suggestions to *tmh.elefeedback@gmail.com,* mentioning the title and author name in the subject line. Please feel free to report any piracy of the book spotted by you!

Tata McGraw-Hill's 360° Development Process

At Tata McGraw-Hill, product development is intensive and ensures that only the best and the most accurate content reaches the users in a manner that enhances the learning process. The 360° Development Process is a continuous, market-oriented approach to building accurate and innovative products. This process is designed to provide a broad, comprehensive spectrum of feedback for refinement and innovation of our learning tools for both students and instructors.

Here is a brief overview of the initiatives included in the 360° Development Process.

Editorial Activities

The span and depth of the content is assessed and ratified by the editors through their careful research and understanding of the market.

Manuscript Review Panels

To conceptually asses the content, the manuscript is continuously subjected to the scrutiny of a panel of subject experts from around the country.

Expert Panel Discussions

To gain course insights, discussions and feedback are invited from experts across the country.

Student Activities

Our ultimate focus is to serve the students and we involve them closely within our product development processes.

Reviewers and Expert Panel

Subhendhu Kr Sahoo	*Birla Institute of Technology and Science, Pilani, Rajasthan*
B D Patel	*College of Engineering, Roorkee, Uttar Pradesh*
Savina Bansal	*Giani Zail Singh College of Engineering and Technology, Bathinda, Punjab*
Vandana Tak	*University College of Engineering Kota, Rajasthan*
S S Pathak	*Indian Institute of Technology Kharagpur, West Bengal*
P R Sahu	*Indian Institute of Technology Guwahati, Assam*
Swadhin Mishra	*Roland Institute of Technology, Behrampur, Odisha*
Tanushree Bose	*Sikkim Manipal Institute of Technology, Sikkim, West Bengal*
Sarat Kumar Patra	*National Institute of Technology, Rourkela, Odisha*
Prajakta Ashtekar	*Maharashtra Academy of Engineering, Alandi, Maharashtra*
Anushka Kadage	*Textile and Engineering Institute, Kolhapur, Maharashtra*
V N Patil	*P.E. Society's Modern College of Engineering, Pune, Maharashtra*
Manisha Kuveskar	*Savitribai Phule Polytechnic, Pune, Maharashtra*
Jennifer S Raj	*Karunya University, Coimbatore, Tamil Nadu*
N Gunavathi	*National Institute of Technology, Tiruchirapalli, Tamil Nadu*

R Harikumar	*Bannari Amman Institute of Technology, Erode, Tamil Nadu*
Vineeta Gejji	*KLS Gogte Institute of Technology, Belgaum, Karnataka*
B N Bhandari	*JNTU College of Engineering, Kukatpally, Hyderabad*
S Varadarajan	*Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh*
Ranjan Mishra	*ICFAI University, Dehradun, Uttarakhand*
Dinesh Chandra	*JSS Academy of Technical Education, Noida, Uttar Pradesh*
K K Tripathi	*Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh*
N P Gupta	*Shree Ganpati Institute of Technology, Ghaziabad, Uttar Pradesh*
A K Singh	*Institute of Technology and Sciences, Ghaziabad, Uttar Pradesh*
M R Tripathy	*Amity Institute of Information Technology, Noida, Uttar Pradesh*
Ashish Tayal	*Bharat Institute of Technology, Meerut, Uttar Pradesh*
Anuranjan Kansal	*JSS Academy of Technical Education, Noida, Uttar Pradesh*
Yatheesh K C	*JSS Academy of Technical Education, Noida, Uttar Pradesh*
A K Panda	*National Institute Of Science & Technology, Berhmapur, Odisha*
P Kabisapathy	*College Of Engineering & Technology, Bhubaneswar, Odisha*
P K Nanda	*Institute Of Technological Education & Research, Bhubaneswar, Odisha*
Ajit Das	*International Institute of Information and Technology, Bhubaneswar, Odisha*
Badri Narayan Sahoo	*Institute Of Technological Education & Research, Bhubaneswar Odisha*
R N Mishra	*Centurion Institute Of Technology, Bhubaneswar, Odisha*
Sanjit Dash	*Krupajal Engineering College, Bhubaneswar, Odisha*
Priyadarshini Kanungo	*C V Raman College Of Engineering, Bhubaneswar, Odisha*
Aruna Tripathy	*College Of Engineering & Technology, Bhubaneswar, Odisha*
Tapas Kumar	*College Of Engineering & Technology, Bhubaneswar, Odisha*

Student Panel

Amrik Singh
Arvinder Kaur
A Sunil Kumar
Balaji A
B Sreenivas
Cherry
D V Soundari
E Kishore
G Anil Kumar
Garima
C Ramya Swathi
D Anu Kiruthiga
D Deenadhayalan
Deepak
Dinesh Chandra
D Nagaraju
Priyadarshini Ravi
P S Karthik
P Sreenithi
P Tejaswani

G Gowtham
G Janardhan Reddy
G Ramya
G Raviteja
G S Sathish
G Vijaya Lakshmi
Jyoti Sumalatha
K Dinesh Kumar
K Divyakrishna
K Krishnakanth
K M V Prasad
K N Manasa
K S Prabakaran
K Tirumala
Lipsi
Manish Kumar Anand
M Bharathi
Meenakshi
Mehak
M Kathirvel
M Nagarjuna
M Surendra Praveen Kumar
M Suresh
N Dheeraj
Nipun
Nishant Jain
Nitasha
N Kiranmye
N Nagaraja
N P Saranya
N Ragasudha
Nupur
Preeti
Rahul
R Madhusudanan
R Tamilvannan
Sagar Sharma
S Bhargav
S Haripriya
Shaweta
S Hema Latha
Sivaprakash S
S Mahesh
S Manohar
S Nagaraju
Srinivas
S Salma Azmee
Sunpreet Singh
S Vijay
Tanya Khanna
T Rahul Kumar Reddy
T Shangeetha
T S Harivikaram
T Sudhakar
T Sundar Pavan Kumar
T Tejpawan
T Yamini
Vaishnavi P
Vanita Singla
V Manaswitha
V Poojitha
V S Anand
V Vasanth
V Yaswanth
Y Phaneendra Kumar
Y Sathyadheeraj

1 Introduction

In this chapter, the student,

- *becomes familiar with the history of evolution of modern electrical communications*
- *identifies the basic elements of an analog communication system and understands the functions performed by each of them*
- *understands the characteristics and limitations of various communication channels*
- *comes to know the way different communication channels are modeled*

1.1 HISTORICAL PERSPECTIVE

Various forms of communication like the telephone, television, radio, FAX, e-mail, etc., have all become an integral part of our daily lives and we dread to imagine what our lives would be like without them. The history of the development of electrical communications over the last hundred and fifty years makes for a really interesting reading.

The era of electrical communications began with the successful commissioning, in 1844, of the first telegraphic link between Washington and Baltimore, by Samuel Morse. The telegraph code, invented by him and subsequently named after him as the 'Morse Code', consisting of different combinations of dots and dashes for representing the various letters of the English alphabet, is indeed a remarkable invention and reveals his inventive genius. It is a variable-length code in which the most frequently occurring letters of the English alphabet are assigned codes of shortest length—for example, the letter 'e' which occurs most frequently, is represented by simply a dot.

The invention of the 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 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 Strowger switches by 'Solid-state Switches' which are faster and also more reliable as they do not have any moving parts.

There has been a remarkable progress in telephone communication, especially with the advent of wireless communication. Further, because of the availability of communication satellites (from 1965 onwards) and optical fibres, it has become possible to simultaneously transmit an extremely large number of telephone conversations over very long distances. Wireless communication also give rise to 'wireless telegraphy' for long-distance transmission of telegraph signals.

That wireless communication over long distances using electromagnetic waves is possible, 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 super heterodyne radio receiver by Armstrong during the first world war, AM radio broadcasting became firmly established. Subsequently, around 1930, Armstrong invented 'frequency modulation' and FM radio broadcasting also started.

Zworykin of the United States of America demonstrated the first 'television system' in 1929 and commercial TV broadcasting was started a few years later.

During the last three decades, there has been a tremendous growth in communications. Launching of the first communication satellite, 'Early Bird', in 1965, marked the beginning of commercial satellite communications. Optical fibres which can support extremely large bandwidths are increasingly being used for high-speed transmission of voice, data and video. Cellular mobile communication is enabling us to communicate instantly with anyone anywhere in the world. Besides, personal communications, providing various types of services—voice, video and data—are being made available.

1.2 AN OVERVIEW OF COMMUNICATION SYSTEMS

Communication systems may be broadly divided into two types—analog communication systems and digital communication systems. In analog communication systems, the information to be transmitted is in the form of an analog signal—a continuous-time waveform. In digital communication systems, the message, or information to be transmitted is digital in form. It may be noted here that in digital communication systems, even though the message is in digital form, the transmitted signal may still be an analog waveform, as we may use a sinusoid of one frequency to represent the binary signal 1 of the message and a sinusoid of a different frequency to represent the binary 0 of the message signal. The 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 information source might have been sampled and encoded to put it in digital form before transmission. We shall discuss the analog communication systems in this book.

1.2.1 Analog Communication Systems—An Overview

Any analog communication system basically consists of a source, a transmitter, a channel and a receiver, as shown in Fig. 1.1.

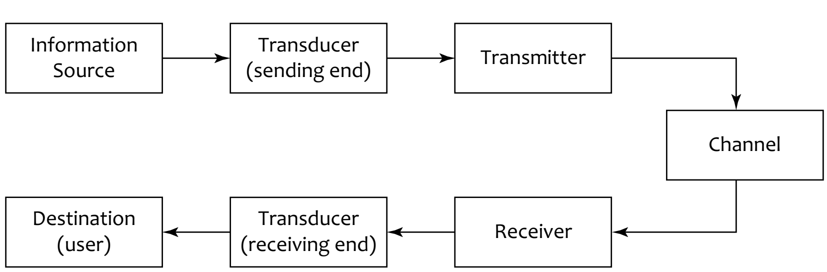


Fig. 1.1 *Basic elements of an analog communication system*

The information source produces information, which may be in the form of speech, or music, or possibly 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.

Transmitter The job of the transmitter is to put this 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. 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. Secondly, 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, freespace, 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.

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 fibre, 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.

During the course of its travel through the channel, the signal may be distorted and attenuated. Distortion can arise due to inadequate bandwidth of the channel. Attenuation in the channel may be attributed to signal spreading and absorption in the case of electromagnetic waves propagating in the atmosphere or free space, and to losses in the case of twisted wires, coaxial cables and optical fibres. Attenuation depends upon frequency, path length and the medium of propagation, and can be quite high. For instance, for satellite communication, the total attenuation (uplink and downlink together) is typically of the order of about 200 db. During its transmission through the channel, the signal may be corrupted by additive noise. This noise may be thermal noise arising from the resistors and the electron devices in the front-end of the receiver. It can also be due to natural phenomenon like lightning discharges during thunderstorms, or may be due to man-made disturbances like automobile ignition, etc. Further, the received signal may also exhibit 'fading' (wide fluctuations in its amplitude) due to multipath effects. These multipath effects arise in a situation wherein a signal reaches the receiver by travelling via more than one path, taking different propagation times for the different paths. In some channels, the channel characteristics may vary significantly with time. In such cases adaptive equalization techniques may have to be resorted to.

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 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.

1.3 ELECTROMAGNETIC SPECTRUM, RANGES AND APPLICATION AREAS

The available electromagnetic spectrum may be conveniently divided into ten 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*	*Applications*
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, Amateur radio, mobile
7	30 MHz – 300 MHz	Very High Frequencies (VHF)	TV, FM, mobile communications
8	300 MHz – 3 GHz	Ultra High Frequencies (UHF)	TV, Radar, satellite communications
9	3 GHz – 30 GHz	Super High Frequencies (SHF)	Terrestrial microwave and satellite communications
10	10^5 GHz – 10^6 GHz	Optical Frequencies (OF)	Optical communications

1.4 CHANNEL TYPES, CHARACTERISTICS AND MODELLING

Communication channels may take a variety of forms. In certain communication systems, we may provide a physical connection between the transmitter and receiver say, using a transmission line; in some others, no such physical connectivity may exist. Based on this, we may broadly classify communication channels into the following two types.

(i) Wire-line channels
(ii) Wireless channels

1.4.1 Wire-line Channels

Wire-line channels may make use of twisted pairs of wires, co-axial cables, optical fibres, or a combination of some of these.

Twisted-pair wires are extensively used for connecting telephone subscribers to the local exchange; and provide a modest bandwidth of a few hundred kilo Hertz. However, they suffer from cross-talk interference and induced additive noise. Co-axial cables, on the other hand, are used for carrying signals of frequencies ranging from a few hundred kilo Hertz to about one giga-Hertz. Solid dielectric-filled, flexible co-axial cables typically provide a bandwidth of several megahertz and give an attenuation of approximately 200 db per 100 metres at 1 GHz. Co-axial cables are relatively immune to cross-talk interference and additive induced noise.

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

An optical fibre consists of a central *core* surrounded by another layer, called the '*cladding*'. Both the core and cladding are made of silica while the 'jacket', which in turn surrounds the cladding and protects it, is made of a 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. Optical fibres support extremely large bandwidths—almost 10% of the centre frequency, amounting to nearly 10^{13} Hz. Modern optical fibres provide very little attenuation, of the order of 0.2 dB/km. They have lot of advantages—immunity to interferences and induced noises, small size and light weight, 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.

1.4.2 Wireless Channels

They are primarily of two types.

(i) Underwater acoustic channels
(ii) Wireless electromagnetic channels

Underwater Acoustic Channels Because of the good conductivity of sea water, electromagnetic waves get very much attenuated within a short distance. The skin depth for sea water at 10 kHz being about 2 to 3 metres, electromagnetic waves even at 10 kHz get attenuated to $(1/e)$ of their value within 2 to 3 metres. However, acoustic waves at that frequency can propagate up to even several hundreds of kilometres. That is why, sonars (underwater counterpart of radars) and all underwater communication systems make use of acoustic waves. These channels, however, are characterized by high levels of noise. This noise is partly contributed by the movement of underwater living organisms like fish, etc., and partly by distant shipping, cavitation, etc. Further, ocean bottom reflections and reflections from the air–ocean interface at the top, create multipath effects and fading. In addition, these channels are highly time-varying. Nevertheless, reliable narrow-band digital data transmission is possible even over long distances.

Wireless Electromagnetic Channels We know that for an antenna to act as an efficient radiator of electromagnetic waves, its physical size should be at least of the order of $\lambda/10$. Going by this rule of thumb, we find that even at the highest frequency in the band 30 Hz to 3 kHz, the minimum antenna size required turns out to be 10 km since λ equals 100 km at that frequency. As it is impossible to have an antenna of that size, the use of 30 Hz–3 kHz band is ruled out.

Once radiated by the transmitting antenna, by which mode the electromagnetic waves will travel, will depend on their frequency.

(i) 3 kHz–30 kHz At these very low frequencies and low frequencies, the wavelengths are such that the ionosphere and the earth's surface effectively act as a waveguide and guide these waves with very little attenuation around the earth. Thus, these frequency bands are useful for navigational purposes. Low-speed digital data requiring only a narrow bandwidth, can be transmitted.

(ii) 300 kHz–3 MHz Waves in this frequency range are called *medium frequency waves* and the most dominant mode of propagation for these waves is by what is called the *ground wave propagation*. This nomenclature comes about from the fact that these waves propagate along the surface of the earth, almost gliding over the earth's surface. Hence, these waves must necessarily be vertically polarized, since the horizontal component, if any, of the wave's electric field gets short-circuited by the earth's surface. The vertical electric field component induces charges on the earth's surface and these charges move along with the wave, thus creating induced currents in the earth's surface. The flow of these induced currents results in power loss owing to the finite conductivity of the earth. This power loss has to be supplied by the electromagnetic wave only and therefore, it gets weakened as it propagates. Sommerfield has shown that when the earth's surface is assumed to be flat, the ground-wave field strength is given by

$$\text{Ground-wave field strength } E_d = A\left(\frac{E_0}{d}\right) \tag{1.1}$$

where,

d is the distance from the transmitting antenna

A is a factor that takes into account the ground losses and depends in a complex way upon the conductivity, the dielectric constant of the earth, the frequency of the wave and the distance from the transmitting antenna

E_0 is the field strength at a unit distance from the transmitting antenna

As the curvature of the earth can be totally ignored up to distances of

$$l_{(\text{km})} = \frac{80}{\sqrt[3]{f_{\text{MHz}}}} \tag{1.2}$$

where, f_{MHz} is the frequency in megahertz, Eq. (1.1) is reasonably accurate up to distances of $d = 2l$. It should also be noted that the waves follow the curvature of the earth to some extent because of diffraction.

Medium waves are used for AM radio broadcasting. Man-made noise and atmospheric noise limit the primary service area of these broadcasting stations to 150 to 200 km. During night time, however, due to disappearance of the D-layer (of the ionosphere) which absorbs frequencies below about 3 MHz, these waves propagate to longer distances partly by the sky-wave mode, using the F-layer.

(iii) 3 MHz–30 MHz This range of frequencies is referred to as the *high-frequency band* (HF band). At these frequencies the ground-wave attenuates rapidly. The dominant mode of propagation for waves in this band is by what is called as the *sky-wave*, i.e., by bending of the rays back to the earth by the layers of the ionosphere through a process of refraction.

The ionosphere is a region surrounding the earth at a height ranging from 60 km to 400 km from the earth's surface, where the earth's atmosphere is partly ionized due to solar radiations and consists of free electrons, ions and gas molecules. Local maxima of the electron density in this region result in the formation of layers as shown in Fig. 1.2.

The typical day-time structure of the ionosphere is shown in this figure. During night time, however, due to the absence of the ionizing radiations from the sun, the electron densities will, in general, be less and the F_1 and F_2 layers merge to form a single F-layer. Further, the D-layer disappears, so that only the E and F layers exist. In addition to this diurnal variation, there is a long-term variation of the ionospheric electronic densities associated with the 11-year sun-spot cycle.

The ionized regions will have dielectric constant and refractive index values different from those of free space. The frequency, electron density and refractive index, are related as

$$n = \sqrt{k} = \sqrt{1 - \frac{81N}{f^2}} \tag{1.3}$$

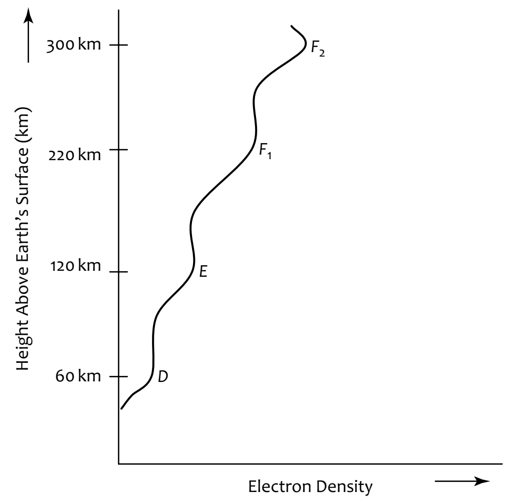


Fig. 1.2 *Variation of electron density with height*

where,

n = refractive index
k = dielectric constant relative to that of free space
N = number of electron per cubic centimeter
F = frequency in kHz

Note that when $f^2 < 81N$, the refractive index becomes imaginary and the medium will not be able to transmit a wave having such a frequency; it will be highly attenuated—similar to what happens in a wave-guide when the frequency is below the cutoff frequency.

The difference in the refractive index between the free space and the ionized region makes an incident ray to change its direction as per Snell's law:

$$n \sin\phi = \sin\phi_0 \tag{1.4}$$

where, ϕ_0 is the angle of incidence at the free-space-ionosphere interface and ϕ is the angle of the ray path at some point Q in the ionized medium. At the point R,

$$n = \sin\phi_0 \tag{1.5}$$

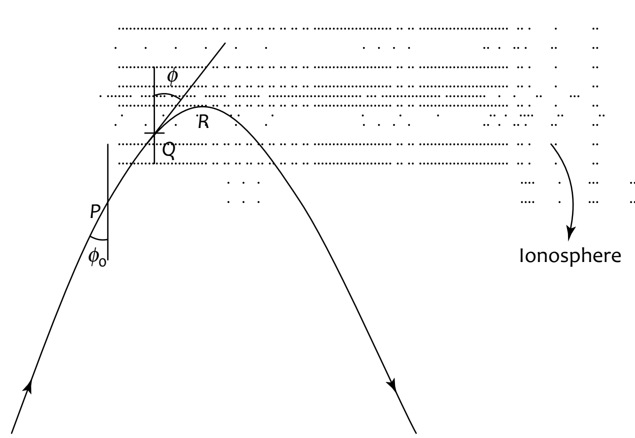


Fig. 1.3 *An incident ray returning back to the earth*

since ϕ_0 is 90° at that point. If $\phi_0 = 0$ as it would be for vertical incidence, reflection can take place only if $n = 0$ at some point. That is, the wave penetrates the ionosphere till that point is reached, where the electron density reaches the value

$$N = \frac{81}{f^2} \tag{1.6}$$

If N corresponds to the maximum electron-density of any layer, waves with the corresponding frequency f_c, if incident vertically on that layer, will not return back to earth—they will penetrate through the ionosphere. Waves of frequency lower than f_c for the layer, will return back to earth whatever may be their angle of incidence. Further, waves of frequency greater than f_c can return back to earth only if the angle of incidence is large, i.e., when the ray is incident at a grazing angle.

The above discussion leads to the concept of '*skip distance*'. It can easily be seen that for a transmitter operating at a certain frequency, there is a certain minimum distance below which no sky wave can be received. This minimum distance is called the *skip distance* at that frequency. Putting it in a different way, we may say that for a given transmitter–receiver distance, there is a particular maximum frequency, called the '*Maximum Usable Frequency*' (MUF), above which the receiver cannot receive any sky-wave emanating from that transmitter.

When a ray gets reflected from the ionosphere once and reaches the earth, it is said to have made one '*hop*'. The maximum single-hop distances are of the order of 2000 and 4000 km when the E and F-layers respectively are considered. Still longer distances may be covered if '*multi-hop*' transmission is considered.

Attenuation of a wave propagating in the sky-wave mode occurs mainly because of two effects—spreading of the wavefront with distance and energy absorption from the wave during its passage through the ionosphere, especially the D-layer region. Electron density in the D-region is high during day time and very low during night time. Further, it has been found that the D-layer absorption per unit distance is inversely proportional to the square of the wave frequency. Thus, ionospheric absorption will be negligible during night time at all frequencies, while in the day time, it increases as the frequency is reduced down to the order of about 1 MHz. Thus, during the night time a sky-wave of good intensity is obtained due to reflection from the E-layer.

Further, at these frequencies, atmospheric noise and man-made noise do not cause much disturbance. Thus, this band is used for A.M. broadcasting, amateur radio and mobile communication.

(iv) Above 30 MHz At frequencies above 30 MHz, the ground-wave attenuates to negligible levels in a very short distance and the ionosphere cannot reflect the waves back to the earth. However, waves at these frequencies can propagate by a mode variously known as '*Line-of-Sight* (LOS) *mode*', '*space-wave mode*', and '*tropospheric propagation mode*'; the last name arising out of the fact that these waves propagate through the lowest layer of the earth's atmosphere, known as the '*troposphere*'.

This type of propagation takes place between elevated transmitting and receiving antennas, A_T and A_R, as shown in Fig. 1.4.

The signal picked up by the receiving antenna has generally two components—one, the direct component R_1 and the other, a ground-reflected component, R_2. Here, for simplicity, the earth's surface is considered to be flat.

For a given d, the distance between the transmitting and receiving antennas, if h_T and h_R are not too large, the difference in the path lengths of the direct ray R_1 and the ground-reflected ray R_2 will not be much and so the signal strengths of these two components will be almost equal. However, they will differ in phase. This phase difference is partly due to the difference in path lengths and partly due to the phase-shift suffered by the ground reflected ray due to the reflection at the earth's surface. Hence, when we add these two components vectorially, they may interfere constructively to give a large intensity value,

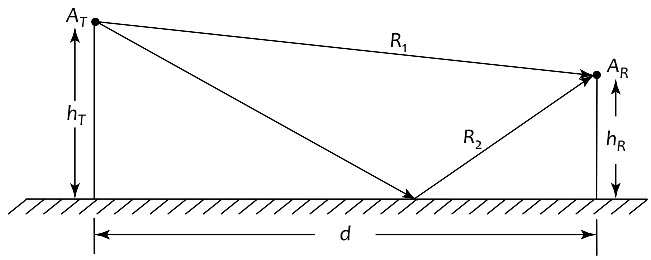


Fig. 1.4 *Basic Space-wave propagation model with a flat earth*

or they may interfere destructively so as to give a very low value. A simple analysis, with the earth's surface assumed flat, yields a received signal field strength given by

$$E_r = \left(\frac{2E_0}{d}\right)\sin\left(\frac{2\pi h_T h_R}{\lambda d}\right) \tag{1.7}$$

where,

E_0 = field intensity at a unit distance from the transmitting antenna, (the value of this depends upon the radiated power and the directive gain of the transmitting antenna)

d = transmitter-to-receiver distance

λ = wavelength of the transmitted wave

h_T, h_R = heights of the transmitting and receiving antennas above the ground

Note: Distances, heights and λ are all to be in the same units.

A plot of E_r (in dB) vs d, using Eq. (1.7), is shown in Fig. 1.5.

From this figure, it is clear that the received signal strength varies appreciably with distance from the transmitting antenna, taking a number of maximum and minimum values. This causes signal fading in the case of mobile communication.

Till now, for simplicity, we have been considering the earth's surface to be flat. When we take into account the curvature of the earth's surface, several new issues crop up.

(1) For given height h_T and h_R of the transmitting and receiving antennas, there is a limit imposed on the range (i.e., distance d_h between the antennas) for terrestrial communication because of radio horizon.

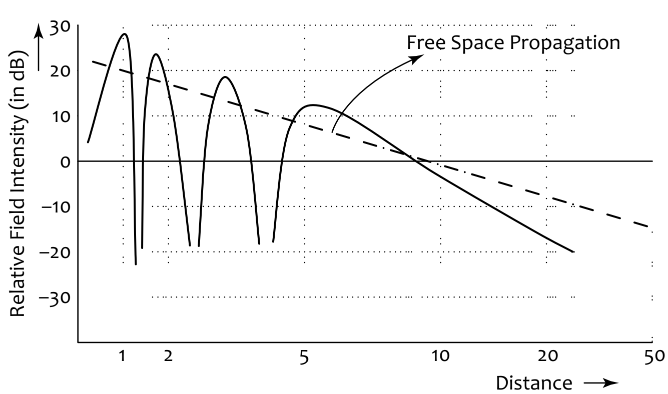


Fig. 1.5 *Relative field intensity variation with distance*

$$d_h = \begin{cases} \text{Max. distance between} \\ \text{transmitting and receiving antennas} \end{cases} = \left[\sqrt{17h_T} + \sqrt{17h_R}\right] \text{km} \tag{1.8}$$

where h_T and h_R are in metres, and d_h is in km.

Since the refractive index of the atmosphere decreases with height due to decrease of atmospheric pressure, radio waves in the troposphere are bent towards the earth. Because of this the radio horizon is farther by about 30% than the horizon as seen by the eye.

(2) Equation (1.7), valid for the flat-earth case, may still be used, but with h_T and h_R in it replaced by h'_T and h'_R respectively, where h'_T and h'_R are the effective heights as shown in Fig. 1.6(b).

(3) The number of maxima and minima, as well as their locations (in Fig. 1.5) will change.

(4) Because of the curvature of the earth, the ground-reflected wave will not be a plane wave. This results in the received signal being slightly weaker than in the flat-earth case.

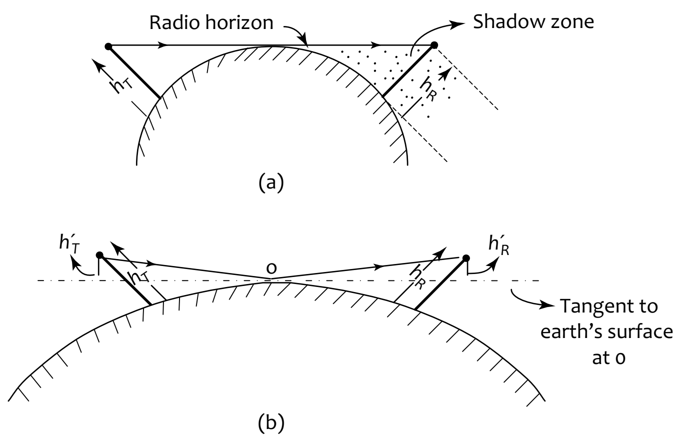


Fig. 1.6 **(a)** *Radio horizon and shadow zone* **(b)** *Effective heights of the antennas when earth's curvature is considered*

(5) If the distance between the transmitting and receiving antennas is quite large, i.e., much greater than the radio horizon, the receiving antenna may be in the shadow region (see Fig. 1.6(a)) and may not be able to receive either the direct ray, or the ground-reflected ray, unless it is of considerable height. *Generally, however, because of diffraction some signal strength will be available even in the shadow region*.

VHF and UHF bands, forming the lower-end of this range of frequencies being considered, are useful for FM broadcasting, TV and mobile communications. As the propagation is by LOS, earth's curvature makes it necessary to keep the antennas for FM and TV broadcasting at considerable heights so as to get a reasonable coverage area of at least 40 to 50 km Atmospheric noise and man-made noise do not cause any problems at these frequencies; but the thermal noise, originating from electronics of the receiver front-end causes some disturbance. Multipath and absorption due to rain are the other issues to be taken care of. Frequencies in the microwave frequency range (3 GHz to 30 GHz) are used for terrestrial microwave relay links and for satellite links. The main problem at these frequencies is the heavy transmission loss due to absorption by rain and fog.

Channel Modeling The foregoing classification of communication channels is based entirely on the physical nature of the transmission medium. However, 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 behaviour of the channel as possible, so as to make it represent the actual situation as accurately as possible, subject, of course, 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 various commonly used channels in the light of the above, 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 fibre channel is time-invariant. Thus, most of the commonly used communication channels can be generally represented by one of the following three models.

(i) Additive Gaussian Noise Channel

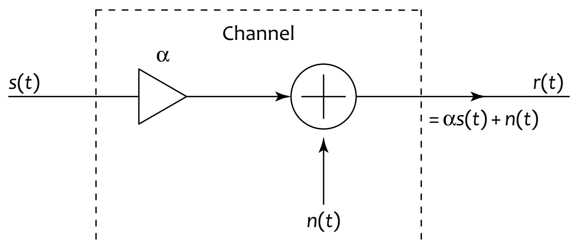


Fig. 1.7 *Additive Gaussian noise channel*

A channel model that is most extensively used is the *additive Gaussian channel* which portrays the channel as one which, as shown in Fig. 1.7, simply attenuates the signal by a factor $\alpha(0 < \alpha < 1)$, and introduces 'additive noise', which itself is modeled as Gaussian

$$r(t) = \alpha s(t) + n(t) \tag{1.9}$$

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

(ii) 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. 1.8.

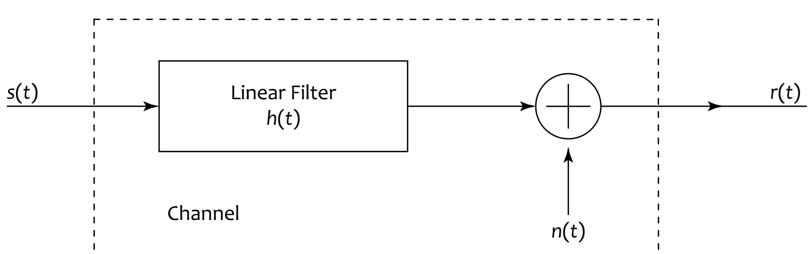


Fig. 1.8 *Bandwidth-limited linear channel*

These channels are time-invariant and so the filter shown in Fig. 1.8 is an LTI system with an impulse response function, $h(t)$. Thus,

$$\begin{aligned} r(t) &= s(t) * h(t) + n(t) \\ &= \int_{-\infty}^{\infty} s(t-\tau)h(\tau)d\tau + n(t) \end{aligned} \tag{1.10}$$

(iii) Linear Time-Variant Channels Channels like the underwater 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. 1.9.

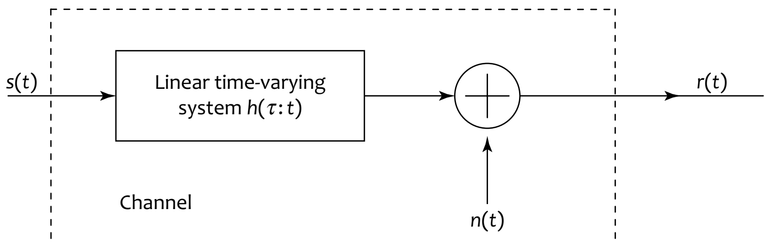


Fig. 1.9 *Linear time-variant channel*

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,

$$r(t) = \int_{-\infty}^{\infty} h(\tau : t)s(t-\tau)d\tau + n(t) \tag{1.11}$$

REFERENCES AND SUGGESTED READING

1. M A McMahon, *The Making of a Profession—A Century of Electrical Engineering in America*, IEEE Press, 1984.
2. John G Proakis and M Salehi: *Communication Systems Engineering*, Second Edition, Prentice-Hall of India, Indian Reprint, 2006.
3. Simon Haykin: *Communication Systems*, John Wiley & Sons, Third edition. 1995.
4. F E Terman: *Electronic and Radio Engineering*, Fourth edition, McGraw-Hill Book Company, Inc. 1955.
5. M Dolukhanov: *Propagation of Radio Waves*, MIR Publications, Moscow, 1971.

2 Signals, Transforms and Spectral Analysis

On going through this chapter, the student 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*
- *expand a given periodic signal in terms of its complex exponential and trigonometric Fourier series and sketch its magnitude and phase spectra*
- *find the auto-correlation of a given signal, or the cross-correlation and convolution of two given signals*
- *determine 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*
- *apply Hilbert transform techniques to determine the lowpass equivalent signal (or, system) for any given bandpass signal (or, system)*

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.

While dealing with signals, the entity which is of paramount importance to a communication engineer is its frequency content. 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 and the Fourier transform provides this information in the case of a non-periodic signal. We shall be discussing these in this chapter.

As we are going to see in the next chapter, an important class of systems, called the Linear Time-Invariant (LTI) systems, possess the property that a system belonging to this class is completely characterized by (or described by) a single function of time, called the impulse response function of the system. We find that the output $y(t)$ of an LTI system with an impulse response function $h(t)$, to an input signal $x(t)$, is given by the 'Convolution of $x(t)$ with $h(t)$'. Thus, 'convolution', which is an operation between two signals, plays a key role in the analysis of systems. We shall discuss this 'convolution' operation in some detail, in this chapter.

'Correlation' between two signals, $x(t)$ and $y(t)$, gives a measure of the degree of similarity between the two signals. Correlation operation is widely used in communication engineering to detect the presence of a 'known signal' in a given received signal. This correlation operation is again closely related to the convolution operation and correlation too will be discussed in this chapter.

Two important classes of signals are the energy signals and the power signals. Energy signals possess finite energy, while power signals possess finite average power. One important aspect of these signals, which is of interest to a communication engineer, is the way the energy/power of the signal is distributed with respect to frequency. Thus, we discuss about the '*Energy Spectral Density*' (ESD) of the energy signals and the '*Power Spectral Density*' (PSD) of the power signal; and see how these are related to the Auto-correlation Function (ACF) of the energy signal or the power signal, as the case may be. We also examine how the ESD of an energy signal, or the PSD of a power signal gets modified during its passage through an LTI system.

Hilbert transform is very useful in the study of communication engineering, as it plays a key role in the representation of a single sideband suppressed carrier modulated signal, in the representation of bandpass signals and in the analysis of bandpass systems. Hence, we briefly discuss this transform, its properties and applications.

It should be noted that the reader is assumed to have had a prior exposure to all the topics covered in this chapter. Hence, they are not covered in a very exhaustive fashion; instead, they are presented in a review-like fashion, presenting all the essential details. For a detailed discussion, the interested reader may refer to the references given.

SIGNALS 2.2

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 and say that a signal is a single-valued function of one or more independent variables and carries some information.

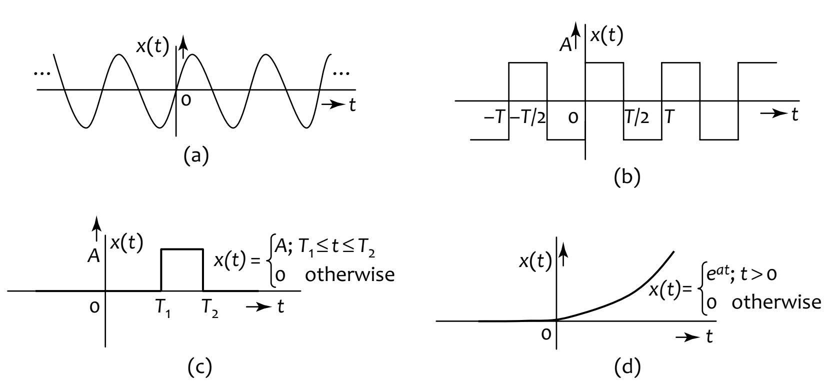


Fig. 2.1 *Examples of continuous-time signals*

2.2.1 Types of Signals

(a) 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.

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.

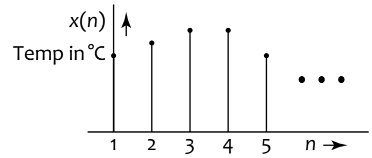


Fig. 2.2 *A discrete-time signal*

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.

(b) 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.

(c) 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 $x^2(t).R$. 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 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

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

Similarly the average power of $x(t)$ is given by

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

Definitions

1. *A signal whose total energy is finite and non-zero, is called an energy signal. If E is the energy,* $0 < E < \infty$.
2. *A signal whose average power is finite and non-zero is called a power signal. If* P_{av} *is the average power,* $0 < P_{av} < \infty$.

Note

1. 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.
2. 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.
3. Signals with a 'finite' or 'asymptotically finite' duration in time, such as the ones given below, are energy signals.

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

   (b) $x(t) = Ae^{-a|t|} \quad ; \quad a > 0$
4. In general, all periodic signals are power signals. (But every power signal need not be a periodic signal.)
5. 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: $x(t) = 5e^{-t} \quad ; \quad -\infty < t < \infty$

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.*

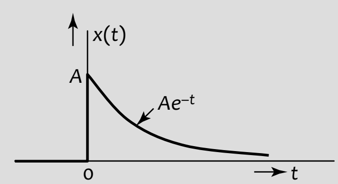


Fig. 2.3 $x(t) = Ae^{-t}$

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

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

Example 2.2

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

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

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

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

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

But the second integral is zero and so

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

(d) Deterministic Signals and Random Signals A signal whose value at any instant of time, $-\infty < t < \infty$, is known a priori, 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. We will be discussing in detail about such signals in Chapter 8.

2.3 UNIT STEP AND UNIT IMPULSE FUNCTIONS

(a) The unit step function, denoted by $u(t)$, is defined by the following:

$$u(t) \triangleq \begin{cases} 1 \text{ for } t \geq 0 \\ 0 \text{ for } t < 0 \end{cases} \tag{2.4}$$

$u(t)$ 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.

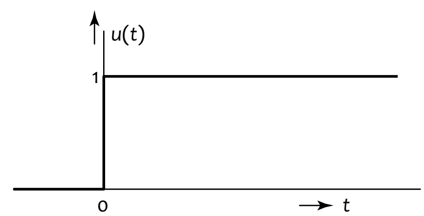


Fig. 2.4 *A unit step function*

2.3.1 The Unit Impulse Function: δ(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 at least 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

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

This implies that 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 $\int_{0-\in}^{0+\in} 1.\delta(t)dt = 1.$ 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 a strength (area) of one.

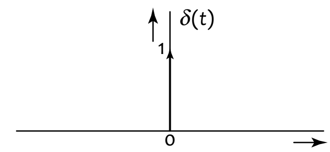


Fig. 2.5 *A unit impulse*

Since $\delta(t)$ represents a unit impulse occurring at $t = 0$, using 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 the above figure 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 $t = \tau$.

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 seconds and we get $x(t - \tau)$.

This is discussed in more detail and proved in Section 2.5 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 $\Delta \to 0$.

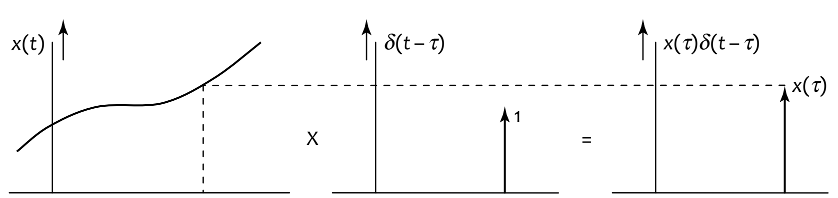


Fig. 2.6 *A diagrammatic representation of Eq. (2.6)*

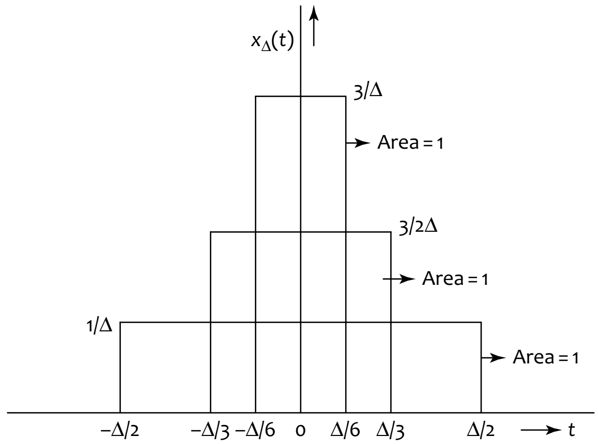


Fig. 2.7 *Unit impulse as limiting case of a rectangular pulse*

2.3.2 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 $t \geq 0$, the area is equal 1. Hence

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

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

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

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

2.3.3 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 Sec. 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.

(a) 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* $x(t)$.

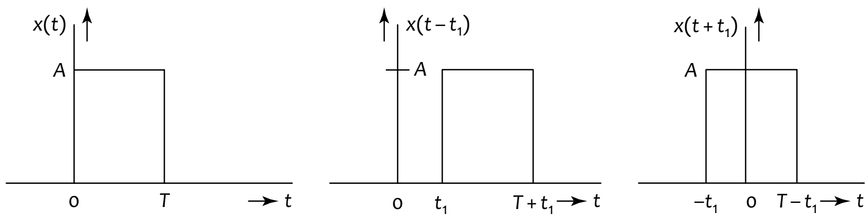


Fig. 2.8 **(a)** *x(t)* **(b)** *For $t_1 > 0$, $x(t - t_1)$ is obtained by shifting x(t) to the right by t_1 seconds* **(c)** *$x(t + t_1)$ is obtained by shifting x(t) to the left by t_1 seconds if $t_1 > 0$*

(b) 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.

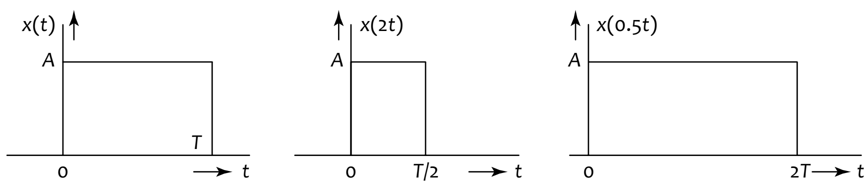


Fig. 2.9 **(a)** *x(t)* **(b)** *x(2t), a compressed version of x(t) and* **(c)** *x(0.5t), an expanded version of x(t)*

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 $x(2t - 3)$ form $x(t)$, we should note that we have to do time-shifting first *and then only* do the time-scaling. This is because

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

whereas,

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

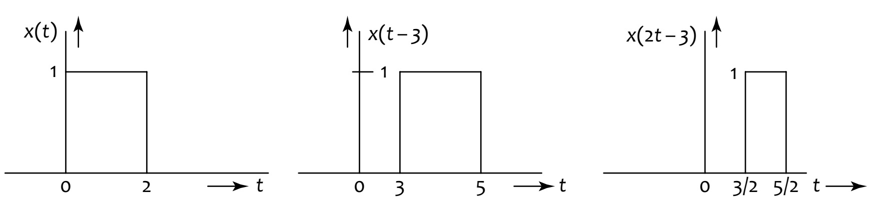


Fig. 2.10 **(a)** *x(t)* **(b)** *x(t – 3)* **(c)** *x(2t – 3)*

2.4 CONTINUOUS-TIME FOURIER SERIES

We had stated that a signal is a single-valued function of time and that the spectral content or frequency content of periodic signals can be obtained using Fourier series expansion of the signals.

As will become clear later, Fourier series expansion of a signal $x(t)$, is nothing but an orthogonal expansion of the signal using what is called a '*complete set*' of orthogonal functions. That is, we express the signal as a linear combination of the members of a complete set of orthogonal functions, as follows:

$$x(t) = c_1 f_1(t) + c_2 f_2(t) + c_3 f_3(t) + \dots \tag{2.9}$$

where $f_i(t) \in S$, a complete set of orthogonal functions, and c_i's are constants.

As the reader must have noticed, there is a similarity between this and the way we express a vector $\boldsymbol{A}$ in Euclidian space in terms of its components along the three orthogonal axes, X, Y and Z.

$$\boldsymbol{A} = x\ \boldsymbol{i} + y\ \boldsymbol{j} + z\ \boldsymbol{k} \tag{2.10}$$

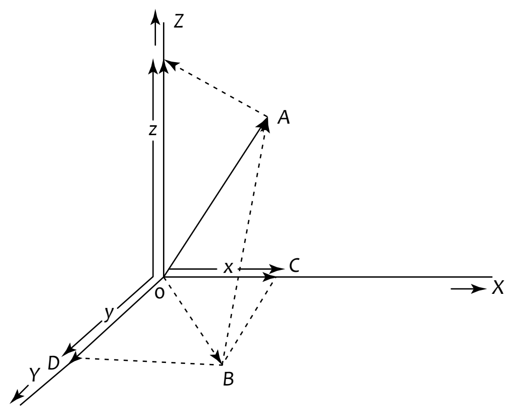


Fig. 2.11 *Vector OA and its components along X, Y and Z directions*

where $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are respectively the unit vectors along X, Y and Z and x, y and z are the coordinates of $\boldsymbol{A}$ along X, Y and Z respectively.

But then, what exactly do we mean by saying that two functions, or signals are orthogonal. For this purpose let us examine the similarity or analogy between signals and vectors a little more deeply.

2.4.1 Analogy between Signals and Vectors

We know that the dot product of two vectors **A** and **B** is a scalar and is given by

$$\mathbf{A}.\mathbf{B} = |A|\,|B| \cos\ \theta \tag{2.11}$$

where, θ is the angle between **A** and **B** and $|A|$ and $|B|$ are the magnitudes respectively of **A** and **B.**

$$\therefore \text{ Component of } \mathbf{A} \text{ along } \mathbf{B} = |A| \cos\ \theta = \left[\frac{\mathbf{A}.\mathbf{B}}{|B|^2}\right]\mathbf{B} \tag{2.12}$$

$$\text{And component of } \mathbf{B} \text{ along } \mathbf{A} = |B| \cos\ \theta = \left[\frac{\mathbf{A}.\mathbf{B}}{|A|^2}\right]\mathbf{A} \tag{2.13}$$

(i) Further, if **A** and **B** are orthogonal, $\theta = \pi/2$ and hence,

$$\mathbf{A}.\mathbf{B} = |A|\ |B| \cos\ \pi/2 = 0 \tag{2.14}$$

(ii) Coversly if $\qquad \mathbf{A}.\mathbf{B} = 0 \qquad (2.15)$

it implies that **A** and **B** are orthogonal vectors

We now use Eq. (2.15) and say that two distinct functions, $f_1(t)$ and $f_2(t)$ are orthogonal to each other over an interval t_1 to t_2 if

$$\int_{t_1}^{t_2} f_1(t) f_2(t) dt = 0 \tag{2.16}$$

If both $f_1(t)$ and $f_2(t)$ are purely real valued. In general, if they are complex-valued, they are defined to be orthogonal over t_1 to t_2 if

$$\int_{t_1}^{t_2} f_1(t) f_2^*(t) dt = 0 \tag{2.17}$$

where $f_2^*(t)$ denotes the complex conjugate of $f_2(t)$.

Thus, a set S of functions $f_i(t)$, $i = 1, 2, \ldots$ is said to be forming a set of orthogonal functions over the interval t_1 to t_2 if

$$\int_{t_1}^{t_2} f_i(t) f_j^*(t) dt = \begin{cases} 0 \quad ; & i \neq j \\ k_i \quad ; & \text{if} \quad i = j \end{cases} \tag{2.18}$$

and

This set S is said to be a *complete set of orthogonal functions* over the interval t_1 to t_2 if *any function* $x(t)$ defined over t_1 to t_2 can be expressed, *without any error*, as a linear combination of the memebrs of the set S.

Let
$$x(t) = c_1 f_1(t) + c_2 f_2(t) + \ldots + c_i f_i(t) + \ldots + c_n f_n(t) \tag{2.19}$$

where $f_i(t)$, $i = 1$ to n are mutually orthogonal '*non-zero functions*'. What we mean by saying that $f_i(t)$ is a non-zero function over the interval t_1 to t_2, is that

$$\int_{t_1}^{t_2} | f_i(t) |^2 \ dt \neq 0$$

Multiplying by $f_i^*(t)$ on both sides, and integrating over t_1 to t_2,

$$\int_{t_1}^{t_2} x(t) f_i^*(t) dt = c_1 \int_{t_1}^{t_2} f_1(t) f_i^*(t) dt + c_2 \int_{t_1}^{t_2} f_2(t) f_i^*(t) dt + \ldots + c_i \int_{t_1}^{t_2} | f_i(t) |^2 dt + c_n \int_{t_1}^{t_2} f_n(t) f_i^*(t) dt$$

Since $f_i(t)$, $i = 1$ to n are orthogonal, only one term, viz.,

$c_i \int_{t_1}^{t_2} | f_i(t) |^2 \ dt$ will be non-zero on the RHS of the the above equation

$$\therefore \qquad \int_{t_1}^{t_2} x(t) f_i^*(t) \, dt = c_i \int_{t_1}^{t_2} | f_i(t) |^2 \ dt$$

Since $f_i(t)$ is a non-zero function over t_1 to t_2, we may write

$$c_i = \frac{\int_{t_1}^{t_2} x(t) f_i^*(t) dt}{\int_{t_1}^{t_2} | f_i(t) |^2 dt} \tag{2.20}$$

In Eq. (2.19), since $c_i f_i(t)$ is the component of $x(t)$ along the signal $f_i(t)$, we may now write

$$\text{component of } x(t) \text{ along the signal } f_i(t) = \left[\frac{\int_{t_1}^{t_2} x(t) f_i^*(t) dt}{\int_{t_1}^{t_2} | f_i(t) |^2 dt} \right] f_i(t) \tag{2.21}$$

In Eq. (2.18), if

$$\int_{t_1}^{t_2} f_i(t) f_j^*(t)\, dt = \begin{cases} 0 & for \;\; i \neq j \\ 1 & for \;\; i = j \end{cases} \tag{2.22}$$

then the set of functions is referred to as a set of orthonormal functions; and incase it is complete in the sense that any arbitary x(t) defined over t_1 to t_2 can be expressed as a linear combination of the elements of this set, it is a complete set of orthonormal functions.

Note: To get a set of orthonormal functions $g_i(t)$'s from a set of orthogonal functions, $f_i(t)$'s, we normalize each function as follows.

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

2.4.2 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

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

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

forms 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.

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

2.5 THE COMPLEX-EXPONENTIAL FOURIER SERIES

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

$$x_n(t) = e^{\frac{j2\pi nt}{T}}, \; n = 0, \pm 1, \pm 2, \ldots$$

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

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

(i) Assume $m \neq n$.

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

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

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

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

(ii) Assume $m = n$. Then $m - n = 0$ and hence

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

Thus, $x_n(t) = e^{j\frac{2\pi}{T}nt}$, $n = 0, \pm 1, \pm 2, \ldots$, 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)

$$= \sum_{n=-\infty}^{\infty} c_n e^{j2\pi nf_0 t}; \quad -\frac{T}{2} \le t \le +\frac{T}{2} \tag{2.24}$$

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

$$\boxed{\underset{\substack{\text{Periodic}\\ \text{with period } T}}{x(t)} = \sum_{n=-\infty}^{n=\infty} c_n e^{j2\pi nf_0 t}; \quad f_0 = \frac{1}{T}; \quad -\infty < t < \infty} \tag{2.25}$$

Now, to determine the constants c_n, $n = 0, \pm1, \pm2, \ldots$, which are called the complex-exponential Fourier series coefficients of $x(t)$, we use Eq. (2.20) and write

$$c_n = \frac{\int_{-T/2}^{T/2} x(t) e^{-j2\pi nf_0 t}\, dt}{\int_{-T/2}^{T/2} \left|e^{j2\pi nf_0 t}\right|^2 dt} = \frac{1}{T}\int_{-T/2}^{T/2} x(t) e^{-j2\pi nf_0 t}\, dt \tag{2.26}$$

We may summarise 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

where

$$\boxed{\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} c_n e^{j2\pi nf_0 t}\ ; \quad f_0 \triangleq \frac{1}{T}; \quad -\infty < t < \infty, \\ c_n &= \frac{1}{T}\int_{-T/2}^{T/2} x(t) e^{-j2\pi nf_0 t} dt \end{aligned}} \tag{2.27}$$

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

$$c_n = |c_n| e^{j\theta_n} \tag{2.28}$$

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.

A close look at Eq. (2.27) reveals that the Fourier series expasion 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 dc 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. (2.27), if we substitute for c_n using Eq. (2.28), we get

$$x(t) = \sum_{n=-\infty}^{\infty} |c_n| e^{j(2\pi n f_0 t + \theta_n)} \quad ; \quad f_0 \triangleq \frac{1}{T} \quad ; \quad -\infty < t < \infty \tag{2.29}$$

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 2.3

Determine the complex-exponential Fourier series expansion of the periodic signal shown.

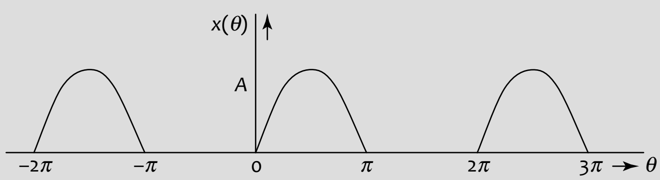


Fig. 2.12 *Signal of Example 2.3*

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

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

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

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

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

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

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

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

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

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

$\begin{matrix}\text{for } n \text{ odd} \\ n \neq 1 \text{ or } -1\end{matrix}$: $(1 - n)$ 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 $e^{-j\pi(1+n)}$ and 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

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

Example 2.4

x(t) is a periodic signal and is shown in Fig. 2.13. Find its complex exponential Fourier series expansion and plot its magnitude and phase spectra.

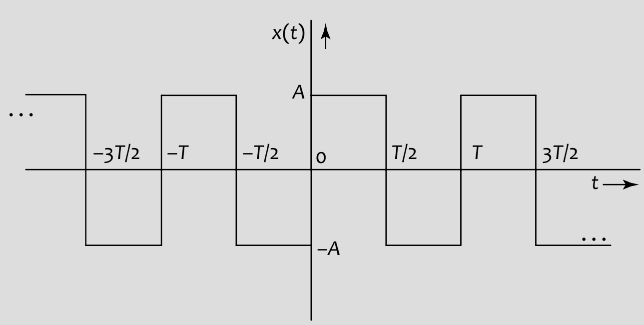


Fig. 2.13 *Signal of Example 2.4*

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

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

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

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

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

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. 2.14 and 2.15 respectively.

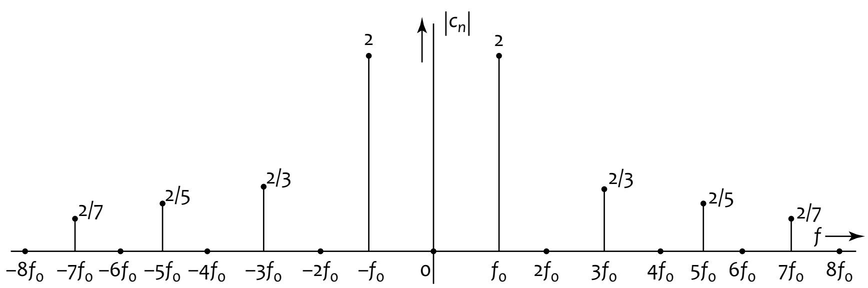


Fig. 2.14 *Magnitude spectrum*

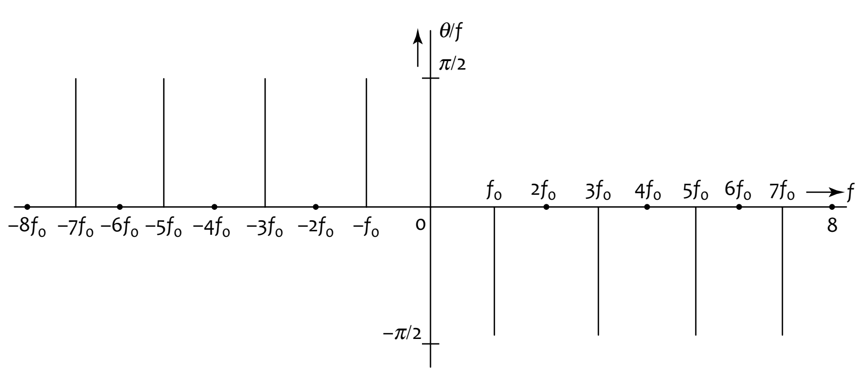


Fig. 2.15 *Phase spectrum*

Properties of Complex-Exponetial Fourier Series Coefficients We now give a list of important theorems and properties of the complex-exponential Fourier series coefficients (CEFSC's), c_n's. The reader is expected to work out the proofs.

1. If $x(t)$ and $y(t)$ are two periodic signals with the same fundamental period, and if their CEFSC's 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

$$\boxed{c_n^z = ac_n^x + bc_n^y} \tag{2.30}$$

This is called the *linearity theorem*.

2. If $x(t)$ is a periodic signal with fundamental frequency f_0 and if $y(t) \triangleq x(t-t_0)$ then

$$\boxed{c_n^y = e^{-j2\pi n f_0 t_0} c_n^x} \tag{2.31}$$

This is called the *time-shift theorem*.

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

$$\boxed{c_n^y = c_{n-k}^x} \tag{2.32}$$

This is called the *frequency-shift theorem*.

4. $x(t)$ and $y(t)$ are periodic signals with the same fundamental period T, and if $z(t) = x(t)*y(t) = \int_{-\infty}^{\infty} x(t-\tau)y(\tau)d\tau$ then

$$\boxed{c_n^z = T \cdot c_n^x \cdot c_n^y} \tag{2.33}$$

This is called the *convolution theorem*.

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

$$\boxed{c_n^z = c_n^x * c_n^y = \sum_{k=-\infty}^{\infty} c_k^x \cdot c_{n-k}^y} \tag{2.34}$$

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

$$\boxed{c_n^y = j2\pi n f_0 c_n^x} \tag{2.35}$$

This is known as the *differentiation theorem.*

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

$$c_n^y = c_n^x \tag{2.36}$$

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_n's as its CEFSC's. Then

$$\text{Average power of } x(t) = \frac{1}{T}\int_{-T/2}^{T/2} |x(t)|^2\,dt = \sum_{n=-\infty}^{\infty} |c_n|^2 \tag{2.37}$$

This is called *Parseval's theorem* pertaining to the complex-exponential Fourier series. Since the average power of $c_n e^{j2\pi n f_0 t}$ is equal to $|c_n|^2$, Eq. (2.37) 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 CEFSC's are represented by c_n's then

$$\boxed{c_{-n} = c_n^*, \text{ if } x(t) \text{ is real valued}} \tag{2.38}$$

where the * indicates complex-conjugation.

10. If a periodic signal $x(t)$ with c_n's, as its CEFSC's is real valued and even with respect to t, then c_n's are also real and are even with respect to n.

$$\boxed{c_n\text{'s are real } c_{-n} = c_n, \text{ if } x(t) \text{ is real an even}} \tag{2.39}$$

11. If a periodic signal $x(t)$ with c_n's, as its CEFSC's, is real valued and has odd symmetry with respect to t, then c_n's are purely imaginary and have odd symmetry with respect to n.

$$\boxed{c_n\text{'s purely imaginary } c_{-n} = c_n^*, = -c_n, \text{ if } x(t) \text{ is real and odd}} \tag{2.40}$$

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. (2.22).

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

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

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

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

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

(ii) We will now show that any two distinct functions in the above set are orthogonal to each other, For this, using Eq. (2.22), we find that

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

Also,

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

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

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 $$\omega_0(t) \;\Delta\; \frac{2\pi}{T} \tag{2.41}$$

Making use of Eq. (2.20) and noting that all the functions of the set are normalized, i.e., that in Eq. (2.20)

$$\int_{-T/2}^{T/2} |f_i(t)|^2 \; dt = 1$$

we have

$$\alpha_0 = \int_{-T/2}^{T/2} x(t)\frac{1}{\sqrt{T}}\; dt = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} x(t)\; dt \tag{2.42}$$

$$\alpha_n = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt \tag{2.43}$$

and $$\beta_n = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt \tag{2.44}$$

If we now define

$$\alpha_0 \frac{1}{\sqrt{T}} \;\underline{\Delta}\; a_0 \quad a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt \tag{2.45}$$

$$\alpha_n \sqrt{\frac{2}{T}} \;\underline{\Delta}\; a_n, \text{ then} \quad a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\cos n\omega_0 t dt \tag{2.46}$$

and

$$\beta_n \sqrt{\frac{2}{T}} \;\underline{\Delta}\; b_n, \text{ then} \quad b_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\sin n\omega_0 t dt, \tag{2.47}$$

Using the above Eq. (2.41) may now be written as

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t; \quad -\frac{T}{2} \le t \le +\frac{T}{2}; \quad \omega_0 \;\underline{\Delta}\; \frac{2\pi}{T} \tag{2.48}$$

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

$$\boxed{\underset{\text{periodic}}{x(t)} = a_0 + \sum_{n=1}^{\infty} a_n \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n \sin n\omega_0 t\,; -\infty \le t \le \infty\,; \omega_0 \;\underline{\Delta}\; \frac{2\pi}{T}} \tag{2.49}$$

where,

$$a_0 = \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,dt$$

$$a_n = \frac{2}{T}\int_{-T/2}^{T/2} x(t)\,\cos n\omega_0 t dt \tag{2.50}$$

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

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, if we put

$$c_0 = a_0; \qquad c_n = \frac{1}{2}\left(a_n - jb_n\right) \text{ and } c_{-n} = \frac{1}{2}\left(a_n + jb_n\right) \tag{2.51}$$

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,

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

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

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

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

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.

(i) 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.

(ii) 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.

(iii) A periodic function $x(t)$ with period T is said to be having rotational, or, half-wave symmetry, if

$$x(t \pm T/2) = -x(t) \text{ for all } t$$

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

$$\int_{-T/2}^{T/2} |x(t)|\ dt < \infty \tag{2.52}$$

The above condition for the existance of Fourier series is called the *weak Dirichlet's condition*.

It should however, be noted that existance of Fourier series doesn't guarantee their convergence at all points and that for convergence, the following conditions, known as *strong Dirichlet's conditions* must be satisfied.

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

(ii) $x(t)$ must have only a finite number of maxima and minima in one period.

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

Example 2.5

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

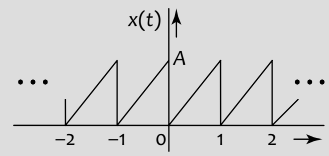


Fig. 2.16 Signal of Example 2.5

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

(a) Trigonometric Fourier Series

$\therefore$
$$a_0 = \frac{A}{1}\int_0^1 t dt = \frac{A}{2}$$

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

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

(b) Complex-exponential Fourier Series

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

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

Example 2.6

For the periodic signal shown in Fig. 2.17, determine the complex-exponential and trigonometric Fourier series.

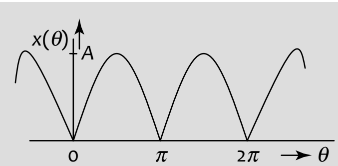


Fig. 2.17 Signal of Example 2.6

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

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

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

On simplification, this gives

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

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

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

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 $\sin\theta.\cos n\theta = \frac{1}{2}[\sin(\theta+n\theta)+\sin(\theta-n\theta)]$

And simplifying, we get

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

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

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

2.7 CONVOLUTION AND CORRELATION OF SIGNALS

Before we proceed further with the various Fourier transform theorems, it is necessary for us to discuss in some detail, about two important operations—convolution and correlation of two signals. 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.

2.7.1 Convolution

The convolution of two continuous-time signals $x(t)$ and $y(t)$, represented by the notation $x(t)* y(t)$ is given by

$$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau) y(t-\tau) d\tau \tag{2.53}$$

By a change of variable, the above integral, generally referred to as the convolution integral, may also be written as

$$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(t-\tau) y(\tau) d\tau \tag{2.54}$$

In Eqs (2.53) and (2.54), 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

(i) Commutative Property

$$x(t) * y(t) = y(t) * x(t)$$

(ii) Associative Property

$$[x(t) * y(t)] * z(t) = x(t) * [y(t) * z(t)]$$

(iii) Distributive Property

$$x(t) * [y(t) + z(t)] = x(t) * y(t) + x(t) * z(t)$$

(iv) 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

$$x(t) * [ay(t) + bz(t)] = aw(t) + br(t)$$

(v) $x(t) * \delta(t-\tau) = x(t-\tau)\ldots$ [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 equation]

This is a very useful result and is used quite often.

(vi) 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 2.7

Given $x(t) = 5\cos t$ and $y(t) = 2e^{-|t|}$, find $x(t)*y(t)$.

Let $z(t) = x(t) * y(t)$

$$\therefore \qquad z(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau$$

But, $$y(t) = 2e^{-|t|} \text{ so that } y(t-\tau) = 2e^{-|t-\tau|} = \begin{cases} 2e^{t-\tau} & \text{for } t < \tau \\ 2e^{-(t-\tau)} & \text{for } t > \tau \end{cases}$$

$$\therefore \qquad z(t) = \int_{-\infty}^{t} (5\cos\tau)2e^{-(t-\tau)}d\tau + \int_{t}^{\infty} (5\cos\tau)2e^{(t-\tau)}d\tau$$

$$= 5[\cos t + \sin t] + 5[\cos t - \sin t] = 10\cos t \;; \qquad -\infty < t < \infty$$

2.7.2 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 Section 2.4, Eq. (2.21) we had shown that the component of $x(t)$ along the signal $y(t)$, may be written as

$$\frac{\int_{t_1}^{t_2} y^*(t)x(t)dt}{\int_{t_1}^{t_2} |y(t)|^2 \, dt} \cdot y(t) \tag{2.55}$$

Note: In case $x(t)$ and $y(t)$ are real valued signals, the complex-conjugation, represented by * in the above equations, may be ignored.

If the two energy signals, $x(t)$ and $y(t)$ are such that

$$\int_{t_1}^{t_2} y^*(t)x(t)dt = 0 \tag{2.56}$$

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:

$$\int_{-\infty}^{\infty} x(t)y^*(t)dt \tag{2.57}$$

But there is a problem in straightway using the above equation. To understand this problem, consider $x(t)$ and $y(t)$ shown in the following figures, i.e., Figs 2.18 and 2.19. The two signals are exactly identical, except that $y(t)$ is a time-delayed version of $x(t)$.

Fig. 2.18 *Signal x(t)*

If we straightaway apply Eq. (2.57) 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 the Eq. (2.57) as follows.

$$\int_{-\infty}^{+\infty} x(t)y^*(t-\tau)dt \tag{2.58}$$

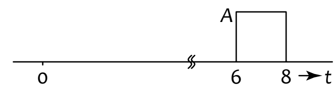


Fig. 2.19 *Signal y(t)*

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

$$R_{xy}(\tau) = \int_{-\infty}^{+\infty} x(t)y^*(t-\tau)dt \tag{2.59}$$

τ has been called as 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. (2.59) 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 2.18 and 2.19, this happens when $\tau = -6$.

By putting $(t - \tau) = \lambda$, Eq. (2.59) may be re-written as,

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)dt \tag{2.60}$$

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 2.18 and 2.19, 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 say that the two signals have no correlation, or, that they are un-correlated.

2.7.3 Symmetry Properties of Cross-Correlation

From Eq. (2.60), if $x(t)$ and $y(t)$, are complex-valued signals,

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)dt \tag{2.61}$$

Replacing τ by $-\tau$ in Eq. (2.60), we have

$$R_{xy}(-\tau) = \int_{-\infty}^{\infty} x(t-\tau)y^*(t)dt$$

$$= \int_{-\infty}^{\infty} x(t)y^*(t+\tau)dt \tag{2.62}$$

But

$$R_{yx}(\tau) = \int_{-\infty}^{\infty} y(t+\tau)x^*(t)dt \tag{2.63}$$

Comparing Eq. (2.62) with Eq. (2.63), we find that

$$\boxed{R_{yx}(\tau) = R^*_{xy}(-\tau) \quad \text{for complex-valued signals}} \tag{2.64}$$

In case the two signals are real valued, it is clear that

$$\boxed{R_{yx}(\tau) = R_{xy}(-\tau) \quad \text{for real valued signals}} \tag{2.65}$$

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

$$\left.\begin{aligned} R_{xx}(\tau) &= \int_{-\infty}^{\infty} x(t)x(t-\tau)dt \\ &= \int_{-\infty}^{\infty} x(t+\tau)x(t)dt \end{aligned}\right\} \text{ if } x(t) \text{ is real valued} \tag{2.66}$$

and,

$$\left.\begin{aligned} R_{xx}(\tau) &= \int_{-\infty}^{\infty} x(t)x^*(t-\tau)dt \\ &= \int_{-\infty}^{\infty} x(t+\tau)x^*(t)dt \end{aligned}\right\} \text{ if } x(t) \text{ is complex valued} \tag{2.67}$$

From the above, we find that

$$\boxed{R_{xx}(-\tau) = R_{xx}(\tau) \quad \text{if } x(t) \text{ is real}} \tag{2.68}$$

and

$$\boxed{R_{xx}(-\tau) = R^*_{xx}(\tau) \quad \text{if } x(t) \text{ is complex}} \tag{2.69}$$

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

$$\boxed{R_{xx}(0) > |R_{xx}(\tau)|; \tau \neq 0} \tag{2.70}$$

(for a rigorous proof, refer to Ref.6)

Equation (2.70) 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 $x(t)$.

2.7.4 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

$$R_{xy}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) y^*(t-\tau) dt \tag{2.71}$$

and auto-correlation of $x(t)$ as

$$R_{xx}(\tau) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) x^*(t-\tau) dt \tag{2.72}$$

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:

$$R_{xy}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t) y^*(t-\tau) dt \tag{2.73}$$

$$R_{xx}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t) x^*(t-\tau) dt \tag{2.74}$$

Remark: Since $x(t)$ and $y(t)$ both have a period T_0, the integrands of Eqs (2.73) and (2.74) 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 2.8

Find $R_{xx}(\tau)$ if $x(t) = e^{-t}u(t)$.

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt = \int_{-\infty}^{\infty} e^{-t}u(t)e^{-(t-\tau)}u(t-\tau)dt$$

(i) when $\tau > 0$.

$$u(t)u(t-\tau) = \begin{cases} 1 & \text{for } t > \tau \\ 0 & \text{for } t < \tau \end{cases}$$

$\therefore$

$$R_{xx}(\tau) = \int_{\tau}^{\infty} e^{-t}e^{-(t-\tau)}dt = e^{\tau}\int_{\tau}^{\infty} e^{-2t}dt = \frac{1}{2}e^{-\tau}u(\tau)$$

(ii) 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 2.9

If $x(t) = A\cos(\omega_0 t + \theta)$, find $R_{xx}(\tau)$.

$$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} A^2 \cos(\omega_0 t + \theta)\cos[\omega_0(t-\tau)+\theta]\, dt$$

Put $\phi = \omega_0 t + \theta$

$$R_{xx}(\tau) = \frac{A^2}{T_0}\cdot\frac{1}{\omega_0}\int_{-\pi+\theta}^{\pi+\theta} \cos\phi[\cos\phi\cos\omega_0\tau + \sin\phi\sin\omega_0\tau]\, d\phi$$

$$= \frac{A^2}{4\pi}\cos\omega_0\tau\int_{-\pi+\theta}^{\pi+\theta}(1+\cos 2\phi)d\phi + \frac{A^2}{4\pi}\sin\omega_0\tau\int_{-\pi+\theta}^{\pi+\theta}\sin 2\phi\, d\phi$$

$$= \frac{A^2}{4\pi}\cos\omega_0\tau[\pi+\theta+\pi-\theta] = \frac{A^2}{2}\cos\omega_0\tau$$

$\therefore$ 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.

2.8 THE CONTINUOUS-TIME FOURIER TRANSFORM

In the last few sections we have 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 just repeats, even though the signal doesn't.

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:

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \quad ; \quad -\infty < t < \infty \quad ; \quad f_0 \triangleq \frac{1}{T}, \tag{2.75}$$

where,
$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-jn\omega_0 t}\, dt, \tag{2.76}$$

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. (2.76), it is clear that c_n becomes a function of ω. Hence, representing c_n as $c_n(\omega)$, we may re-write Eq. (2.76) as

$$Tc_n(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \tag{2.77}$$

Since the RHS (and therefore the LHS too) of Eq. (2.77) 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

$$X(\omega) \triangleq \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \tag{2.78}$$

Further, from Eq. (2.53), it follows that

$$x(t) = \frac{1}{T}\sum_{\omega=-\infty}^{\infty} Tc_n(\omega)e^{j\omega t} \tag{2.79}$$

However,

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

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\, d\omega; -\infty < t < \infty \tag{2.80}$$

Equation (2.78) 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. (2.80), 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 be forming a '*Fourier transform pair*'. Their relationship is symbolically represented using the following notation.

$$x(t) \xleftrightarrow{F.T} X(f)$$

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 (2.78) and (2.80) and write the Fourier and Inverse Fourier transforms respectively as follows:

$$\boxed{X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi f t}\, dt} \tag{2.81}$$

and

$$\boxed{x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi f t}\, df} \tag{2.82}$$

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 $x(t)$.

2.8.1 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,

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

Since $\left|e^{-j2\pi f t}\right| = 1$ and $\left|x(t)e^{-j2\pi f t}\right| = |x(t)|\left|e^{-j2\pi f t}\right|$,

it follows that

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \tag{2.83}$$

is the condition required to be satisfied for $|X(f)|$ to be finite. Thus, Eq. (2.83) 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 be seeing 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. (2.83).

The Fourier transform integral given by Eq. (2.81) and the inverse Fourier transform integral given by Eq. (2.82) 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 of this book, we simply state here that if a non-periodic signal $x(t)$ satisfies the Dirichlet's conditions, then the point-wise convergence of the integral

$$\int_{-\infty}^{\infty} X(f)e^{j2\pi f t}\, df$$

is guarenteed for all values of t except those corresponding to discontinuities. The Dirichlet's conditions are the following:

(a) $x(t)$ should be absolutely integrable.
(b) $x(t)$ should have only a finite number of maxima and minima in any finite interval of time.
(c) In any finite interval of time, the number of discontinuities of $x(t)$ should be finite.
(d) 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.

2.8.2 Some Simple Properties of Fourier Transform

We now give, without proof, a list of some simple, but very useful, properties of the Fourier transform. The reader is urged to work out the proof using Eqs (2.81) and (2.82).

1. $X(0)$ is equal to the area under $x(t)$. This is because $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi f t} dt$

   $\therefore \quad X(0) = \int_{-\infty}^{\infty} x(t)dt =$ 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

$$|X(-f)| = |X(f)| \text{ while } \angle X(-f) = -\angle X(f) \tag{2.84}$$

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

$$X(f) = 2\int_0^{\infty} x(t)\cos\omega t dt \tag{2.85}$$

(b) If the signal $x(t)$ has odd symmetry then its Fourier transform $X(f)$ is given by

$$X(f) = -2j\int_0^{\infty} x(t)\sin\omega t dt \tag{2.86}$$

2.8.3 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 2.10

$$x(t) = \begin{cases} A \quad ; \quad |t| \le \tau/2 \\ 0 \quad ; \quad otherwise \end{cases}$$

Determine and plot the magnitude and phase spectra of x(t).

The given signal is a rectangular pulse and its plot is as shown in Fig. 2.20.

This being a commonly used signal, it is given a special symbol.

$$x(t) = A\prod(t/\tau) \quad \text{or} \quad A\ \text{rect}(t/\tau)$$

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.

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*** $-\dfrac{\tau}{2}$ ***to*** $+\dfrac{\tau}{2}$

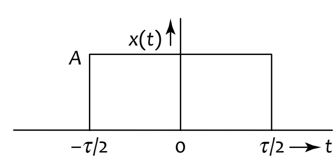


Fig. 2.20 *A rectangular pulse*

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
$$\boxed{X(f) = \mathcal{F}\left[A\Pi(t/\tau)\right] = A\tau \text{ sinc} f\tau} \tag{2.87}$$

Plots of the magnitude and phase spectra of the signal $x(t)$ are shown in the following Figs 2.21(a) and 2.21(b) respectively.

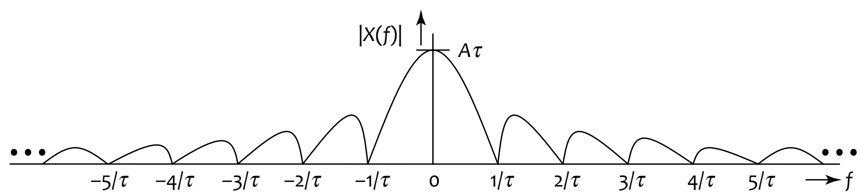


Fig. 2.21(a) *Magnitude spectrum of* $A\Pi(t/\tau)$

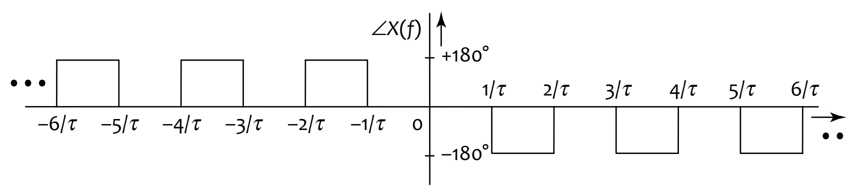


Fig. 2.21(b) *Phase spectrum of* $A\Pi(t/\tau)$

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 $\pm 1/\tau$, $\pm 2/\tau$, $\pm 3/\tau$. 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. 2.19(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.

2.8.4 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 over bar 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

$$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} \overline{Y(f)}\left[\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\, dt\right]df$$

$$= \int_{-\infty}^{\infty} \overline{Y(f)}.X(f)df$$

Thus,
$$\boxed{\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}df} \tag{2.88}$$

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. (2.88) becomes:

$$\boxed{\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2\, df} \tag{2.89}$$

In the above equation, we know that the LHS represents the energy E of the signal $x(t)$. Hence, Eq. (2.89) 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 (2.89) 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 2.11

If the signal $x(t) = Ae^{-t/T}u(t)$ is given as input to an ideal lowpass 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?

We have to first find the spectrum of $X(f)$ of the signal $x(t)$. For this we note:

$$X(f) = \int_{-\infty}^{\infty} Ae^{-t/T}u(t)e^{-j2\pi ft}dt = A\int_{0}^{\infty} e^{-\frac{(1+j2\pi fT)t}{T}}\, dt$$

$$X(f) = \frac{AT}{1+j2\pi fT} \quad \text{and} \quad |X(f)|^2 = \frac{A^2T^2}{1+4\pi^2 f^2T^2}$$

Putting $2\pi fT = \tan\theta$ in the above and noting that $df = (1/2\pi T)\sec^2\theta\, d\theta$, we have,

$$E_x = \text{Total energy in the signal } x(t) = \int_{-\infty}^{\infty} \frac{A^2T^2}{1+4\pi^2 f^2T^2}\,df$$

$$= \int_{-\pi/2}^{\pi/2} \left(\frac{A^2T}{2\pi}\right) d\theta = \frac{A^2T}{2}$$

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. (2.67) 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,

$$p = \frac{E_0}{E_x} \times 100\% = \frac{(A^2T/4)}{(A^2T/2)} \times 100 = 50\%$$

2.8.5 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.

(a) 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

$$\boxed{\mathcal{F}\left[\alpha x(t) + \beta y(t)\right] = \alpha X(f) + \beta Y(f)} \qquad (2.90)$$

The proof of this theorem is trivial and left as an exercise to the reader.

(b) Time-delay Theorem This theorem gives us the Fourier transform of $x(t - \tau)$, the time-delayed version of a $x(t)$ in terms of $X(f)$, the Fourier transform of $x(t)$. It says that:

If $\quad x(t) \xleftrightarrow{FT} X(f)$ Then, $\quad x(t-\tau) \xleftrightarrow{FT} X(f)e^{-j2\pi f\tau}$

Proof

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

$$\therefore \quad \mathcal{F}[x(t-\tau)] = \int_{-\infty}^{\infty} x(t-\tau)e^{-j2\pi ft}\,dt$$

Putting $t - \tau = t'$, $t = t' + \tau$ and $dt = dt'$

$$\therefore \quad \mathcal{F}[x(t-\tau)] = \int_{-\infty}^{\infty} x(t')e^{-j2\pi f(t'+\tau)}\,dt'$$

$$= \left[\int_{-\infty}^{\infty} x(t')e^{-j2\pi ft'}\,dt'\right] e^{-j2\pi f\tau} = X(f)e^{-j2\pi f\tau}$$

$$\therefore \qquad \boxed{x(t-\tau) \xleftrightarrow{F.T} X(f)e^{-j2\pi f\tau}} \qquad (2.91)$$

Note: Since $\left|X(f)e^{-j2\pi f\tau}\right| = \left|X(f)\right|$, it follows that shifting of a signal along the time axis changes only the phase spectrum but not the magnitude spectrum.

(c) 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 $x(t) \xleftrightarrow{F.T} X(f)$ then,

$$x(t)e^{j2\pi f_c t} \xleftrightarrow{F.T} X(f-f_c)$$

Proof

$$\mathcal{F}[x(t)e^{j2\pi f_c t}] = \int_{-\infty}^{\infty} \{x(t)e^{j2\pi f_c t}\}e^{-j2\pi ft}\,dt$$

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

$$\therefore \qquad \boxed{x(t)e^{j2\pi f_c t} = X(f-f_c)} \qquad (2.92)$$

Equation (2.92) 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. 2.22(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. (2.84), $X(f)$ must have a magnitude which has even symmetry. The spectrum of $x(t)e^{j\omega_c t}$, as given by Eq. (2.92), is plotted in Fig. 2.22(b).

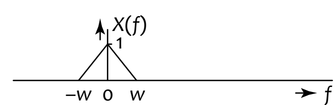


Fig. 2.22(a) *Spectrum of x(t)*

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.

From Eq. (2.92), we may write

$$x(t)e^{j\omega_c t} \xleftrightarrow{F.T} X(f-f_c)$$

$$\therefore \qquad x(t)e^{-j\omega_c t} \xleftrightarrow{F.T} X(f+f_c)$$

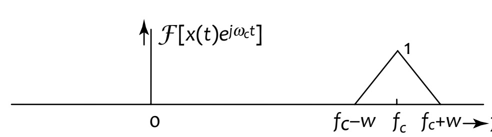


Fig. 2.22(b) *Spectrum of* $x(t)e^{j\omega_c t}$

Adding these two and invoking the linearity property of the Fourier transform, we get

$$\therefore \qquad \boxed{x(t)\cos\omega_c t \xleftrightarrow{F.T} \frac{1}{2}\left[X(f-f_c)+X(f+f_c)\right]} \qquad (2.93)$$

Hence, the spectrum of $x(t)$ $\cos\omega_c t$, the modulated signal, would appear as shown in Fig. (2.23).

It may be noted that whereas the spectrum of $x(t)$ cos $\omega_c t$ is having 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.

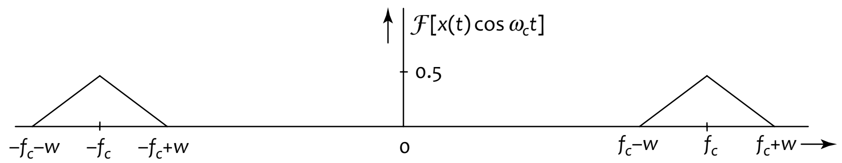


Fig. 2.23 *Spectrum of x(t) cosω$_c$t*

(d) 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 Section 2.2 of this chapter, 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 $x(t) \xleftrightarrow{F.T} X(f)$ then

$$x(at) \xleftrightarrow{F.T} \frac{1}{|a|} X(f/a)$$

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

$$\boxed{x(at) \xleftrightarrow{F.T} \frac{1}{|a|}X(f/a)} \tag{2.94}$$

Remarks

1. 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.
2. 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.
3. Since compression in time leads to expansion in frequency, and vice versa, a signal can not be compressed/expanded simultaneously in time as well as in frequency.

4. 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; and a female voice recorded at some speed, would sound like a male voice, when played back at a much lower speed.

(e) Duality Theorem This theorem enables us to write down the spectra of certain signals just by inspection, as illustrated in Example 2.12.

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 $f' = -f$

$$X(t) = -\int_{+\infty}^{-\infty} x(-f')e^{j2\pi f't}df' = \int_{-\infty}^{\infty} x(-f')e^{j2\pi f't}df' = \mathcal{F}^{-1}[x(-f)]$$

$$\therefore \qquad \boxed{X(t) \xleftrightarrow{F.T.} x(-f)} \tag{2.95}$$

(f) 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 $x(t) \xleftrightarrow{F.T.} X(f)$, $y(t) \xleftrightarrow{F.T.} Y(f)$ 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

$$Z(f) = \int_{-\infty}^{\infty} z(t)e^{-j2\pi f t}dt = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x(\lambda)y(t-\lambda)d\lambda\right\}e^{-j2\pi f t}dt$$

$$= \int_{-\infty}^{\infty} x(\lambda)\left\{\int_{-\infty}^{\infty} y(t-\lambda)e^{-j2\pi f t}dt\right\}d\lambda$$

$$= \int_{-\infty}^{\infty} x(\lambda)Y(f)e^{-j2\pi\lambda f} d\lambda \quad \text{(by applying Time-delay theorem)}$$

$$= Y(f)\int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi\lambda f} d\lambda \quad = Y(f).X(f)$$

$$\therefore \quad \boxed{Z(f) = Y(f).X(f) \quad \text{if } z(t) = x(t) * y(t)} \tag{2.96}$$

(g) 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 $z(t) = x(t) \cdot y(t)$, 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 $\quad z(t) = x(t).y(t)$

$$\therefore \quad Z(f) = \int_{-\infty}^{\infty} \{x(t).y(t)\}e^{-j2\pi f t} dt$$

Now, writing $y(t)$ as the inverse Fourier transform of $Y(\lambda)$, where λ is a dummy frequency parameter,

$$Z(f) = \int_{-\infty}^{\infty} x(t)\{Y(\lambda)e^{j2\pi\lambda t} d\lambda\}e^{-j2\pi f t} dt$$

$$= \int_{-\infty}^{\infty} Y(\lambda)\left\{\int_{-\infty}^{\infty} [x(t)e^{j2\pi\lambda t}]e^{-j2\pi f t} dt\right\} d\lambda$$

Now, using the modulation theorem, we may write

$$Z(f) = \int_{-\infty}^{\infty} Y(\lambda)X(f-\lambda)d\lambda = X(f) * Y(f)$$

$$\therefore \quad \boxed{Z(f) = X(f) * Y(f) \quad \text{if } z(t) = x(t).y(t)} \tag{2.97}$$

Example 2.12

Determine the energy contained in the signal x(t) = 20 sinc 10t.

We shall solve the problem by making use of Parseval's theorem. Earlier, we had seen (see Eq. 2.87) that

$$A\Pi(t/\tau) \leftrightarrow A\tau \text{ sinc } f\tau$$

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 AWsinc 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 sinc $10t = AW$ sinc Wt

Thus $AW = 20$ and $W = 10$ $\quad \therefore A = 2$

We know from the duality theorem that

$$AW \text{ sinc } Wt \leftrightarrow A\Pi(-f/w) = A\Pi(f/w)$$

Hence
$$20 \text{ sinc } 10t \xleftrightarrow{FT} 2\Pi(f/10) = X(f)$$

This $X(f)$ is a rectangular pulse in frequency domain with an amplitude of 2 and base width of 10.

$\therefore$ applying Parseval's theorem,

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

$$= \int_{-5}^{5} 2^2 \, df = 4 \times 10 = 40 \text{ units}$$

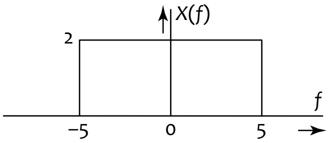


Fig. 2.24 *FT of x(t)*

Example 2.13

Find the Fourier transform of $x(t) = \begin{cases} \cos \pi t & ; \; -\frac{1}{2} \le t \le \frac{1}{2} \\ 0 & ; \; \textit{otherwise} \end{cases}$

We can solve it either by using the defining equation of the Fourier transform, or by using the convolution theorem.

(i) 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.

$$X(f) = \int_{-1/2}^{1/2} \cos \pi t \cdot \cos 2\pi ftdt$$

$$= \frac{1}{2} \int_{-1/2}^{1/2} \{\cos \pi(2f+1)t + \cos \pi(2f-1)t\} tdt$$

$$= \frac{1}{2} \left[\frac{\sin \pi(2f+1)}{\pi(2f+1)} \Bigg|_{t=-1/2}^{t=1/2} + \frac{\sin \pi(2f-1)}{\pi(2f-1)} \Bigg|_{t=-1/2}^{t=1/2} \right]$$

$$= \frac{1}{2} \left[\text{sinc}\left(f + \frac{1}{2} \right) + \text{sinc}\left(f - \frac{1}{2} \right) \right]$$

(ii) Using the convolution theorem of Fourier transform

The given $x(t)$ is shown in Fig. 2.25. 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 $|t| \leq ½$ and zero outside.

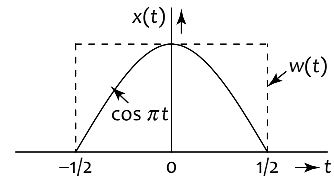


Fig. 2.25 *The given x(t)*

$\therefore \qquad x(t) = \cos \pi t \cdot \omega(t)$

Hence, $\qquad X(f) = \mathcal{F}[\cos \pi t] * W(f)$

But $\qquad \mathcal{F}[\cos \pi t] = \frac{1}{2}\left[\delta\left(f - \frac{1}{2}\right) + \delta\left(f + \frac{1}{2}\right)\right]$

$$\therefore \qquad X(f) = \frac{1}{2}\left[\delta\left(f - \frac{1}{2}\right) + \delta\left(f + \frac{1}{2}\right)\right] * W(f)$$

$$= \frac{1}{2}\left[W\left(f - \frac{1}{2}\right) + W\left(f + \frac{1}{2}\right)\right] \quad \text{(Replication property of an impulse)}$$

But $\qquad w(t) = \Pi(t/1) \qquad \therefore \quad W(f) = \operatorname{sinc} f$

$$\therefore \qquad X(f) = \frac{1}{2}\left[\operatorname{sinc}\left(f - \frac{1}{2}\right) + \operatorname{sinc}\left(f + \frac{1}{2}\right)\right]$$

(h) Differentiation-in-time Theorem This theorem enables us to straightaway 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 $\dot{x}(t) = j2\pi f X(f)$

Proof

$$x(t) = \mathcal{F}^{-1}[X(f)] = \int_{-\infty}^{\infty} X(f) e^{j2\pi f t} df \tag{2.98}$$

$$\therefore \qquad \frac{dx(t)}{dt} = \frac{d}{dt}\left[\int_{-\infty}^{\infty} X(f) e^{j2\pi f t} df\right] = \int_{-\infty}^{\infty} \{j2\pi f X(f)\} e^{j2\pi f t} df \tag{2.99}$$

Comparing Eqs (2.98) with (2.99), we have

$$\boxed{\dot{x}(t) \; = j2\pi f \; X(f)} \tag{2.100}$$

n iterations of the above process yields

$$\boxed{\frac{d^n}{dt^n} x(t) \xleftrightarrow{F \cdot T \cdot} (j2\pi f)^n \cdot X(f)} \tag{2.101}$$

Remarks

(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 $2\pi f$ 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.

(i) Differentiation-in-frequency Theorem This theorem can be considered as the dual of the differentiaion-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

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

$$\therefore \qquad \frac{d}{df}X(f) = \int_{-\infty}^{\infty} x(t)\frac{d}{df}[e^{-j2\pi ft}]dt = \int_{-\infty}^{\infty} \{x(t)(-j2\pi t\}e^{-j2\pi ft}dt$$

Comparing the LHS's and RHS's of the above two equations, we may state that

$$\boxed{(-j2\pi t)x(t) \xleftrightarrow{F.T} \frac{d}{df}X(f)} \qquad (2.102)$$

2.8.6 Fourier Transforms using Impulses

We shall now derive the Fourier transforms of certain functions like the sine and cosine, the unit-step and signum function, etc., which are not absolutely integrable. This we do by using impulses.

(i) 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 2.1, 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, $$\boxed{\mathcal{F}[\delta(t)] = 1} \qquad (2.103)$$

Equation (2.103) 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. 2.26.

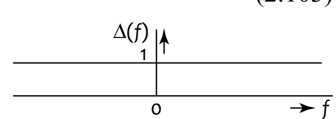


Fig. 2.26 *Spectrum of a unit impulse function in time, δ (t)*

(ii) Fourier Transform of x(t) = 1 Applying duality theorem to the transform given by Eq. (2.103), we get

$$\mathcal{F}[1] = \delta(-f) = \delta(f) = \text{Unit impulse in the frequency domain}$$

$$\therefore \qquad \boxed{1 \xleftrightarrow{F.T.} \delta(f)} \qquad (2.104)$$

Since $\delta(t) \xleftrightarrow{F.T.} 1$, if we apply the time-delay theorem, we get

$$\boxed{\delta(t-\tau) \xleftrightarrow{F.T.} e^{-j2\pi f\tau}} \qquad (2.105)$$

(iii) Transform of $e^{j2\pi f_0 t}$ From modulation theorem, we know that

$$x(t)e^{j2\pi f_0 t} \xleftrightarrow{F.T.} X(f-f_0)$$

In the above, if we take $x(t)$ to be equal to 1,

$$\boxed{e^{j2\pi f_0 t} \xleftrightarrow{F.T.} \delta(f-f_0)} \tag{2.106}$$

(iv) Transformer of cos $2\pi f_0 t$ We have noted that,

$$e^{j2\pi f_0 t} \xleftrightarrow{F.T.} \delta(f-f_0)$$

$$e^{-j2\pi f_0 t} \xleftrightarrow{F.T.} \delta(f+f_0)$$

Combining the two and invoking the linearity theorem,

$$\boxed{\cos 2\pi f_0 t \xleftrightarrow{F.T} \frac{1}{2}[\delta(f-f_0)+\delta(f+f_0)]} \tag{2.107}$$

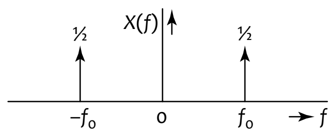


Fig. 2.27 *Spectrum of cos $2\pi f_0 t$*

The spectrum of cos $2\pi f_0 t$, as given Eq. (2.107) is shown in Fig. 2.27.

(v) Transform of the Signum Function The signum function in time, denoted by sgn(t), is defined as

$$\text{sgn}(t) = \begin{cases} 1 & \text{if } t > 0 \\ -1 & \text{if } t < 0 \end{cases} \tag{2.108}$$

We shall derive the Fourier transform of the signum function by making use of the differentiation theorem.

From Fig. 2.28, we find that

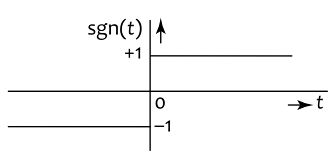


Fig. 2.28 *A signum function*

$$\frac{d}{dt}[\text{sgn}(t)] = 2\delta(t)$$

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)

$$\therefore \qquad S(f) = \frac{1}{j\pi f} \tag{2.109}$$

Note that at $f = 0$ the Fourier transform of sgn(t) appears to become infinitely large and therefore indeterminate, as per Eq. (2.109). 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,

$$\boxed{\text{Sgn}(t) \xleftrightarrow{F.T} \begin{cases} \dfrac{1}{j\pi f}; & f \neq 0 \\ o\quad ; & f = 0 \end{cases}} \tag{2.110}$$

(vi) Transform of u(t) From Fig. 2.28, it is clear that

$$1+\operatorname{sgn}(t) = 2u(t)$$

i.e.,
$$u(t) = \frac{1}{2}[1+\operatorname{sgn}(t)] \tag{2.111}$$

Now taking Fourier transform on both sides, noting that the FT of 1 is $\delta(f)$ and invoking the linearity theorem of the Fourier transform,

$$u(t) \xleftrightarrow{F.T} U(f) = \frac{1}{2}\left[\delta(f) + \frac{1}{j\pi f}\right]$$

$\therefore$
$$\boxed{U(f) = \frac{1}{2}\left[\delta(f) + \frac{1}{j\pi f}\right]} \tag{2.112}$$

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
$$u(t-\tau) = \begin{cases} 1 \text{ for } \tau < t \\ 0 \text{ for } \tau > t \end{cases}$$

$\therefore$
$$x(t) * u(t) = \int_{-\infty}^{t} x(\tau)d\tau = y(t)$$

$\therefore$
$$Y(f) = X(f).U(f) = \frac{1}{2}\left[X(f)\delta(f) + \frac{X(f)}{j\pi f}\right] \quad \text{(from Eq. 2.112)}$$

Making use of the sampling property of the impulse function, we have

$$X(f)\delta(f) = X(0)\delta(f)$$

Hence,

$$\boxed{\int_{-\infty}^{t} x(\tau)d\tau \xleftrightarrow{F.T} \frac{1}{2}\left[X(0)\delta(f) + \frac{X(f)}{j\pi f}\right]}$$

Example 2.14

Find the Fourier transform of the signal x(t) shown in Fig. 2.29 (a).

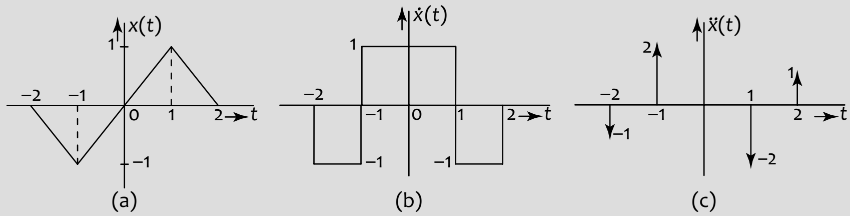


Fig. 2.29 **(a)** $x(t)$, **(b)** $\dot{x}(t)$, *and* **(c)** $\ddot{x}(t)$

We shall use the differentiation theorem of Fourier transform to find the *F.T.* of $x(t)$.

From Fig. 2.29(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

$$x(t) \xleftrightarrow{F.T} X(f). \text{ Then}$$

$$\dot{x}(t) \xleftrightarrow{F.T} j2\pi fX(f) \text{ and } \ddot{x}(t) \xleftrightarrow{F.T} -4\pi^2 f^2 X(f)$$

Hence, $$\mathcal{F}\left[-\delta(t+2) + 2\delta(t+1) - 2\delta(t-1) + \delta(t-2)\right] = -4\pi^2 f^2 X(f)$$

But, the LHS of the above is

$$= -e^{-j4\pi f} + 2e^{-j4\pi f} - 2e^{j4\pi f} + e^{j4\pi f}$$

$\therefore$ $$4\pi^2 f^2 X(f) = \left(e^{j4\pi f} - e^{-j4\pi f}\right) - 2\left(e^{j2\pi f} - e^{-j2\pi f}\right)$$

$$= 2j\sin 4\pi f - 4j\sin 2\pi f$$

$\therefore$ $$X(f) = \frac{-1}{j2\pi^2 f^2}[\sin 4\pi f - 2\sin 2\pi f]$$

Example 2.15

Find the signal f(t) if its Fourier transform F(ω) is as shown in Figs 2.30(a) and 2.30(b).

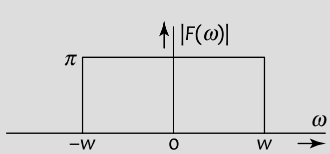


Fig. 2.30 **(a)** $F(\omega)$

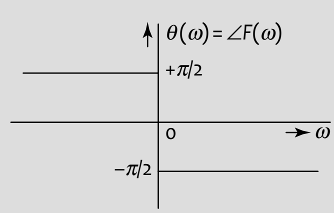


Fig. 2.30 **(b)** $\theta(\omega)$

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

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega) e^{j2\pi ft}\, d\omega$$

$$\frac{1}{2\pi}\left[\int_{-w}^{0} \pi e^{j\pi/2}.\, e^{j\omega t}\, d\omega + \int_{o}^{w} \pi e^{-j\pi/2}.\, e^{\omega t}\, d\omega\right]$$

On simplification, this gives

$$f(t) = \left[\frac{1 - \cos Wt}{t}\right]$$

Example 2.16

If $x(t) \xleftrightarrow{F.T} X(f)$, find the Fourier transforms of the following signals in terms of $X(f)$.

(a) $x(t-2)e^{jt}$ (b) $x(1 - t)$ (c) $x\left(\frac{t}{2} - 2\right)$

(a) Let $x_1(t) = x(t-2)e^{jt}$. Then $x(t-2) \xleftrightarrow{F.T} X(f)e^{-j4\pi f}$ and $x_1(t) = x(t-2)e^{j2\pi(1/2\pi)t}$

$\therefore$ $$x_1(t) \xleftrightarrow{F.T} X\left(f - \frac{1}{2\pi}\right)e^{-j4\pi(f - \frac{1}{2\pi})} \text{ (from modulation theorem)}$$

(b) Let $x_2(t) = x(1-t)$. Now, $x(t) \xleftrightarrow{F.T} X(f)$. Hence,

$x(-t) \xleftrightarrow{F.T} X(-f)$ and $x(-t+1) \xleftrightarrow{F.T} X(-f)e^{-j2\pi f}$ (Time-delay theorem) (from scaling theorem)

(c) Let $x_3(t) = x\left(\frac{t}{2} - 2\right)$. Since $x(t) \xleftrightarrow{F.T} X(f)$,

From time-delay theorem we have $x(t-2) \xleftrightarrow{F.T} X(f)e^{-j4\pi f}$

And from scaling theorem, $x\left(\frac{t}{2} - 2\right) \xleftrightarrow{F.T} 2X(2f)e^{-j8\pi f}$

Example 2.17

Find the Fourier transform of $x(t) = \left(\frac{1}{1+t^2}\right)$.

We know that $e^{-|t|} \xleftrightarrow{F.T} \dfrac{2}{1+4\pi^2 f^2}$

Now, applying duality theorem,

$$\frac{2}{1+4\pi^2 t^2} \xleftrightarrow{F.T} e^{-|f|}$$

If we let $y(t) = \dfrac{2}{1+4\pi^2 t^2}$ and $y(at) = \dfrac{2}{1+t^2} = 2x(t)$

$$y(at) = \frac{2}{1+4\pi^2 a^2 t^2} = \frac{2}{1+t^2}$$

$$4\pi^2 a^2 = 1 \text{ and } a = \frac{1}{2\pi}$$

$\therefore$ If $Y(f) = e^{-|f|}, \quad 2X(f) = \mathcal{F}[y(at)] = \dfrac{1}{a} Y(f/a)$

$\therefore$ $2X(f) = 2\pi Y(2\pi f) \quad \therefore \; X(f) = \pi e^{-|\omega|}$

2.9 ENERGY SPECTRAL DENSITY AND POWER SPECTRAL DENSITY

2.9.1 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: $$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) y(t-\tau) dt \tag{2.113}$$

Convolution: $$z(t) = x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau) y(t-\tau) d\tau \tag{2.114}$$

If we now define $$\therefore w(t) \triangleq x(t) * y(-t) = \int_{-\infty}^{\infty} x(\tau) y(\tau - t) d\tau \tag{2.115}$$

Then, replacing the dummy variable τ in Eq. (2.115) by u, we have,

$$w(t) = \int_{-\infty}^{\infty} x(u) y(u-t) du \tag{2.116}$$

In the cross-correlation Eq. (2.113), if t is replaced by u,

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(u) y(u-\tau) du \tag{2.117}$$

A comparison of Eqs (2.116) and (2.117) reveals that

$$R_{xy}(\tau) = [x(t) * y(-t)]\Big|_{t \to \tau} \tag{2.118}$$

and $$R_{xx}(\tau)=[x(t)*x(-t)]\Big|_{t\to\tau} \tag{2.119}$$

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.

2.9.2 Energy Spectral Density

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,

Since $$z(t)=x(t)*y(t) \tag{2.120}$$

$$Z(f)=X(f).Y(f) \tag{2.121}$$

Then, from Eq. (2.118), we know that

$$\mathcal{F}\left[R_{xy}(\tau)\right]=\mathcal{F}[x(t)].\mathcal{F}[y(-t)] \tag{2.122}$$

i.e., $$\mathcal{F}[R_{xy}(\tau)] \ = \ X(f).Y(-f) \ \triangleq \ S_{xy}(f) \tag{2.123}$$

In a similar manner, we have

$$\mathcal{F}[R_{yx}(\tau)] \ \triangleq \ S_{yx}(f)=Y(f).X(-f) \tag{2.124}$$

$\therefore$ if $R_{xx}(\tau)$ is the auto-correlation function of $x(t)$, then

$$\mathcal{F}[R_{xx}(\tau)] \ = X(f).X(-f)=|X(f)|^2 \tag{2.125}$$

But, from Parseval's theorem for Fourier transform, we know that (refer to Eq. 2.89) $|X(f)|^2$ 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 $S_{xx}(f)$.

$\therefore$ $$\mathcal{F}[R_{xx}(\tau)] \ = \ S_{xx}(f)=|X(f)|^2 \tag{2.126}$$

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., $$\boxed{R_{xx}(\tau)\xleftrightarrow{F.T}S_{xx}(f)} \tag{2.127}$$

This relationship is generally referred to as the auto-correlation theorem and may be derived directly as follows.

$$\mathcal{F}\left[R_{xx}(\tau)\right]=\int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j2\pi f\tau}d\tau=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[x(t)x(t-\tau)dt\right]e^{-j2\pi f\tau}d\tau$$

$$=\int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt\cdot\int_{-\infty}^{\infty} x(t-\tau)e^{-j2\pi f(\tau-t)}d\tau$$

putting $\lambda = (t - \tau)$

$$\mathcal{F}\left[R_{xx}(\tau)\right] = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt . \int_{-\infty}^{\infty} x(\lambda)e^{-j2\pi(-f)\lambda}d\tau$$

$$= X(f).X(-f) = |X(f)|^2 = S_{xx}(f)$$

Since $$R_{xx}(\tau) \xleftrightarrow{F.T} S_{xx}(f), \text{ we have,}$$

$$\int_{-\infty}^{\infty} S_{xx}(f)e^{j2\pi f\tau}df = R_{xx}(\tau)$$

Putting $\tau = 0$ on both sides of the above equation

$$R_{xx}(0) = \text{ACF for zero lag} = \int_{-\infty}^{\infty} S_{xx}(f)df$$

$$= \text{Area under the energy spectral density function}$$

$$= \text{Energy of } x(t).$$

$$\therefore \quad \boxed{R_{xx}(0) = \text{Energy of } x(t), \text{ an energy signal}} \tag{2.128}$$

2.9.3 Power Spectral Density

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

$$x_T(t) \triangleq \begin{cases} x(t) & \text{for } |t| \leq T/2 \\ 0 \ ; & \text{otherwise} \end{cases} \tag{2.129}$$

Thus, $x_T(t)$ is a finite energy signal and hence, is Fourier transformable.

Let $$x_T(t) \xleftrightarrow{F.T} X_T(f) \tag{2.130}$$

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

$$P_{x_T x_T}(f) = \frac{|X_T(f)|^2}{T} \tag{2.131}$$

and the average power spectral density of $x(t)$ as

$$P_{xx}(f) = \underset{T\to\infty}{\text{Lt}} \left[\frac{|X_T(f)|^2}{T}\right] \tag{2.132}$$

Recalling (from Eq. 2.72) 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,

$$\mathcal{F}[R_{xx}(\tau)] = \mathcal{F}\left[\underset{T\to\infty}{\text{Lt}} \frac{1}{T}\int_{-T/2}^{T/2} x(t)x(t-\tau)dt\right]$$

$$= \underset{T\to\infty}{\text{Lt}}\left[\mathcal{F}\left\{\frac{1}{T}\int_{-\infty}^{\infty} x_T(t)x_T(t-\tau)dt\right\}\right]$$

$$= \underset{T\to\infty}{\text{Lt}}\left[\frac{1}{T}\mathcal{F}\left\{\int_{-\infty}^{+\infty} x_T(t)x_T(t-\tau)dt\right\}\right] = \underset{T\to\infty}{\text{Lt}} \frac{1}{T}\mathcal{F}\left[R_{x_T x_T}(\tau)\right]$$

$$= \underset{T\to\infty}{\text{Lt}}\left[\frac{|X_T(f)|^2}{T}\right] = P_{xx}(f) \quad \text{(from Eq. 2.132)}$$

$\therefore$ power spectral density of a power signal is the Fourier transform of the auto-correlation of the signal.

$$\boxed{R_{xx}(\tau) \xleftrightarrow{F\cdot T\cdot} P_{xx}(f)} \tag{2.133}$$

Example 2.18

Find the ACF and ESD of the signal $x(t) = e^{-t}u(t)$.

We know, from Eq. (2.119) that

$$R_{xx}(\tau) = [x(t) * x(-t)]\Big|_{t\to\tau}$$

$\therefore$

$$S_{xx}(f) = X(f).X(-f) = |X(f)|^2$$

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 2.19

Find the ACF and ESD of the signal $x(t) = A\prod(t/2T)$.

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt$$

and,

$$R_{xx}(\tau) = [x(t) * x(-t)]\Big|_{t\to\tau}$$

Since $x(t)$ has even symmetry,

$$R_{xx}(\tau) = [x(t) * x(t)]\Big|_{t \to \tau}$$

$$\therefore \quad S_{xx}(f) = \mathcal{F}\left[x(t) * x(t)\right] = |X(f)|^2 = 4A^2T^2 \text{ sinc}^2 (2fT)$$

$$\therefore \quad R_{xx}(f) = \mathcal{F}^{-1}\left[4A^2T^2 \text{ sinc}^2 (2fT)\right] = A^2\left[2T - |\tau|\right]$$

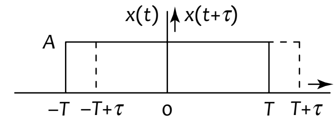


Fig. 2.31 *x(t) of Example 2.19*

$R_{xx}(\tau)$ 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.

$$\therefore \quad R_{xx}(\tau) = A^2(2T - |\tau|)$$

Hence, $R_{xx}(\tau)$ is a triangular waveform as shown in Fig. 2.32.

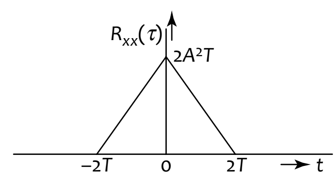


Fig. 2.32 $R_{xx}(\tau)$ *of AΠ(t/2T)*

Properties of Power Spectral Density

(i) $P_{xx}(f)$ of a signal is always non-negative, since

$$P_{xx}(f) = \underset{T \to \infty}{\text{Lt}} \left[\frac{|X_T(f)|^2}{T}\right]$$

(ii) $P_{xx}(f)$ is the Fourier transform of $R_{xx}(\tau)$

(iii) The total area under the PSD curve of a signal equals the average power of the signal.

Since $$\therefore R_{xx}(\tau) = \mathcal{F}^{-1}\left[P_{xx}(f)\right] = \int_{-\infty}^{\infty} P_{xx}(f) e^{j2\pi f\tau} df$$

$$\therefore \quad R_{xx}(0) = \int_{-\infty}^{\infty} P_{xx}(f)df = \text{area under the PSD curve.}$$

But, $$R_{xx}(\tau) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t)x(t - \tau)\, dt$$

$$\therefore \quad R_{xx}(0) = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x^2(t)\, dt = \text{Av. power of } x(t).$$

$$\therefore \quad \int_{-\infty}^{\infty} P_{xx}(f)\, df = \text{Average power } P_{\text{av}} \text{ of } x(t).$$

(iv) 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 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 F·T· on both sides, we have:

$$Y(f) = X(f) \cdot H(f)$$

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.

Relation Between Input and Output ESDs Let $x(t)$ be an energy signal. From Eq. (2.135), 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 ESD's of output and input signals.

$$\therefore \qquad |Y(f)|^2 = S_{yy}(f) = |X(f)|^2 \cdot |H(f)|^2 = S_{xx}(f) \cdot |H(f)|^2$$

Hence,
$$S_{yy}(f) = |H(f)|^2 \cdot S_{xx}(f) \tag{2.136}$$

Relation Between Input and Output PSDs Let us now assume that the input signal $x(t)$, is a power signal. Also, let

$$x_T(t) = w(t) \cdot x(t) \tag{2.137}$$

where, $w(t)$ is a rectangular window function defined by

$$w(t) = \begin{cases} 1 & \text{for} \quad |t| \le T \\ 0 & \text{otherwise} \end{cases} \tag{2.138}$$

Then, we know that

$$P_{xx}(f) = \underset{T\to\infty}{\mathrm{Lt}} \left[\frac{|X_T(f)|^2}{T} \right]$$

where, $X_T(f) = \mathcal{F}[x_T(t)]$

and that
$$P_{yy}(f) = \underset{T\to\infty}{\mathrm{Lt}} \left[\frac{|Y_T(f)|^2}{T} \right]$$

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

$$P_{yy}(f) = |H(f)|^2 \cdot P_{xx}(f) \tag{2.139}$$

Example 2.20

$x(t) = e^{-t/\tau} u(t)$ is applied as input to an L-section highpass RC filter with a time constant of τ seconds. Find the ESD of the output of the filter. Express the output signal energy as a percentage of the input signal energy.

Transfer function $H(f)$ of the RC filter is

$$H(f) = \frac{j2\pi fRC}{1+j2\pi fRC} = \frac{j\omega\tau}{1+j\omega\tau} \quad \text{since } \tau = RC$$

$$\therefore \qquad |H(f)|^2 = \frac{\omega^2\tau^2}{1+\omega^2\tau^2}$$

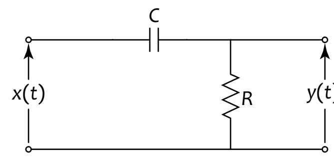


Fig. 2.33 *An L-section highpass RC filter*

$\therefore$ 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 $\quad S_{xx}(f) = |X(f)|^2, \quad \text{where} \quad X(f) = \int_{-\infty}^{\infty} e^{-t/\tau} u(t) e^{-j\omega t}\, dt = \frac{\tau}{1+j\omega\tau}$

$$\therefore \qquad |X(f)|^2 = \frac{\tau^2}{1+\omega^2\tau^2} \quad \text{and} \quad E_x = \int_{-\infty}^{\infty} |X(f)|^2 df = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\tau^2}{1+\omega^2\tau^2} d\omega$$

Putting $\omega\tau = \tan\theta, \quad 1+\omega^2\tau^2 = \sec^2\theta \quad \text{and} \quad d\omega = \frac{1}{\tau}\sec^2\theta\, d\theta$

$\therefore$ $E_x =$ energy in the input signal $= \dfrac{\tau^2}{2\pi}\displaystyle\int_{-\pi/2}^{+\pi/2} \frac{1}{\tau}\, d\theta = \frac{\tau}{2}$

$$S_{xx}(f) = |X(f)|^2 |H(f)|^2 = \frac{\omega^2\tau^4}{(1+\omega^2\tau^2)^2}$$

$\therefore$ $E_y =$ total energy at the output of the filter $= \displaystyle\int_{-\infty}^{\infty} S_{xx}(f) df$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\omega^2\tau^4}{(1+\omega^2\tau^2)^2} d\omega$$

Substituting $\tan\theta$ 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 2.21

The signal $x(t) = 10\cos(4\pi \times 10^3)\, t$ is given as input to an L-section lowpass RC filter having 3 dB cutoff frequency of 10^3 Hz. Determine and sketch the output PSD.

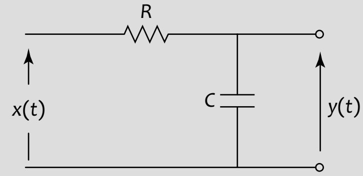


Fig. 2.34 *An L-section RC lowpass filter*

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

$$R_{xx}(\tau) = \frac{A^2}{2} \cos \omega_0 t = \frac{10^2}{2} \cos(4\pi \times 10^3 \tau)$$

$\therefore$
$$P_{xx}(f) = \mathcal{F}[50 \cos(4\pi \times 10^3)\tau] = 25[\delta(f - f_0) + \delta(f + f_0)]$$

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
$$H(f) = \frac{1}{1 + j\omega RC}$$

$$|H(f)|^2 = \frac{1}{1 + \omega^2 R^2 C^2}$$

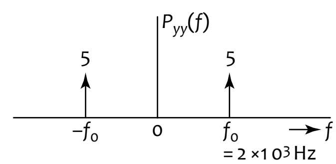


Fig. 2.35 *A sketch of $P_{yy}(f)$*

Since 3 dB frequency for an RC lowpass filter $= \dfrac{1}{2\pi RC}$

$$\frac{1}{2\pi RC} = 10^3 \quad \therefore R^2C^2 = \frac{1}{4\pi^2 \times 10^6}$$

$\therefore$
$$P_{yy}(f) = \frac{25}{1 + \omega^2 \left(\dfrac{1}{4\pi^2 \times 10^6} \right)} [\delta(f - f_0) + \delta(f + f_0)]$$

$$= 5[\delta(f - f_0) + \delta(f + f_0)] \quad \text{(from sampling property of delta functions)}$$

where, $f_0 = 2 \times 10^3$ Hz

2.10 HILBERT TRANSFORM

Hilbert transform is different from the other transforms like the Fourier transform in the sense that Hilbert transforming a signal $x(t)$ does not bring about a change of domain. $\hat{x}(t)$, the Hilbert transform of $x(t)$, is also a time signal, just like $x(t)$. Hilbert transforming a signal $x(t)$ produces only a phase shift of –90° for all the frequency components of $x(t)$. This property of the Hilbert transform makes it extremely useful—in the representation of bandpass signals, especially the single sideband modulated signals; the bandpass-to-lowpass transformation of signals and systems and in the implementation of certain modulator circuits like the phase-shift modulators.

2.10.1 Definition and Frequency-domain Interpretation

The Hilbert transform, $\hat{x}(t)$, of a signal $x(t)$, is defined as the signal obtained by convolving $x(t)$ with $1/(\pi t)$.

$\therefore$
$$\hat{x}(t) \triangleq x(t) * \frac{1}{\pi t} \tag{2.140}$$

i.e.,
$$\hat{x}(t) = \int_{-\infty}^{\infty} \frac{x(\tau)}{\pi(t-\tau)} d\tau = \int_{-\infty}^{\infty} \frac{x(t-\tau)}{\pi\tau} d\tau \tag{2.141}$$

Note:

1. This definition of Hilbert transform is applicable to all signals that are Fourier transformable.
2. 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. (2.140), 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. (2.110), we have

$$\operatorname{sgn}(t) \xleftrightarrow{F.T} \frac{1}{j\pi f} \tag{2.142}$$

Hence, from duality theorem of Fourier transforms,

$$\frac{1}{j\pi t} = \operatorname{sgn}(-f) = -\operatorname{sgn}(f), \text{ since } \operatorname{sgn}(f) \text{ is an odd function of } f.$$

$$\therefore \qquad \frac{1}{\pi\, t} \xleftrightarrow{F.T} -j\operatorname{sgn}(f) \tag{2.143}$$

Going back to Eq. (2.142), we therefore have

$$\hat{X}(f) = -j\operatorname{sgn}(f)X(f) \tag{2.144}$$

But
$$\operatorname{sgn}(f) = \begin{cases} 1 & \text{for} \quad f > 0 \\ -1 & \text{for} \quad f < 0 \end{cases}$$

$$\therefore \qquad \frac{\hat{X}(f)}{X(f)} = \begin{cases} -j & \text{for} \quad f > 0 \\ j & \text{for} \quad f < 0 \end{cases} \tag{2.145}$$

Since $\hat{X}(f)$ is the spectrum of $\hat{x}(t)$ while $X(f)$ is the spectrum of $x(t)$, it follows from Eq. (2.145) 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,

$$\left|\hat{X}(f)\right| = |X(f)| \tag{2.146}$$

i.e., a Hilbert transform does not alter the magnitude spectrum.

From Eq. (2.140), 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

$$h(t) = \frac{1}{\pi t}, \tag{2.147}$$

and whose input signal is $x(t)$. Such an LTI system, called as the Hilbert Transformer, will have a transfer function $H(f)$ given by

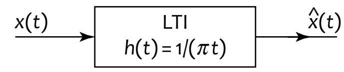


Fig. 2.36 *A Hilbert transformer*

$$H(f) = \frac{X(f)}{X(f)} = -j\,\mathrm{sgn}(f) = \begin{cases} -j & \text{for } f > 0 \\ j & \text{for } f < 0 \end{cases} \tag{2.148}$$

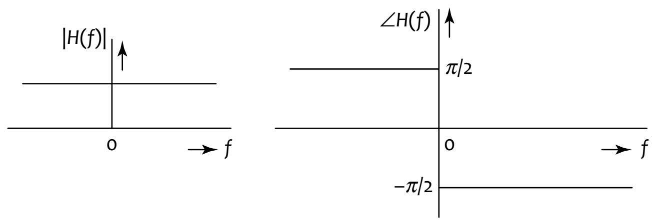


Fig. 2.37 **(a)** *Magnitude response of a Hilbert transformer* **(b)** *Phase response of a Hilbert transformer*

2.10.2 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 $x(t) \xleftrightarrow{H.T} \hat{x}(t)$, then $\hat{x}(t) \xleftrightarrow{H.T} -x(t) = \hat{\hat{x}}(t)$.

Proof $\hat{x}(t) \xleftrightarrow{F.T} -j\ \mathrm{Sgn}\ (f)X(f)$.

Hence $\hat{\hat{x}}(t) \xleftrightarrow{F.T} \{-j\ \mathrm{Sgn}\ (f)\}\{-j\ \mathrm{Sgn}\ (f)\}X(f)$

$\therefore$
$$F\left[\hat{\hat{x}}(t)\right] = -\mathrm{sgn}^2(f) \cdot X(f) = -X(f)$$

$\therefore$
$$\boxed{\hat{\hat{x}}(t) = -x(t)}$$

(4) A signal and its Hilbert transform are orthogonal to each other.

i.e.,
$$\int_{-\infty}^{\infty} x(t)x(t)dt = 0$$

Proof of the above property is left as an exercise to the reader (see Example 2.21).

Remark: From the –90° phase-shift property of Hilbert transform, it follows that

(i) $\sin \omega_0 t \xleftrightarrow{H.T} -\cos \omega_0 t$ (2.149)

(ii) $\cos \omega_0 t \xleftrightarrow{H.T} \sin \omega_0 t$ (2.150)

(5) If $x(t)$ is a lowpass signal and $y(t)$ is a highpass signal, and if their spectra are non-overlapping then

$$\widehat{x(t)y(t)} = x(t) \cdot \hat{y}(t) \tag{2.151}$$

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,

$$Z(f) = X(f) * Y(f) = \int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)d\lambda$$

$$\therefore \qquad \hat{Z}(f) = \mathcal{F}\left[\hat{z}(t)\right] = -j\,\text{sgn}(f)Z(f)$$

$$= -j\int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)\,\text{sgn}(f)d\lambda$$

Hence $$\hat{z}(t) = \mathcal{F}^{-1}\left[\hat{Z}(f)\right] = -j\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(\lambda)Y(f-\lambda)\,\text{sgn}(f)e^{j2\pi ft}d\lambda df$$

$$= -j\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X(\lambda)e^{j2\pi\lambda t}\cdot Y(f-\lambda)e^{j2\pi(f-\lambda)t}\,\text{sgn}(f)dfd\lambda \qquad (2.152)$$

Here, $x(t)$ is a lowpass signal, bandlimited 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 highpass 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. (2.152), 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. (2.152) 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 lowpass signal and $y(t)$ is highpass signal, and if their spectra are non-overlapping, then

$$\widehat{x(t)y(t)} = x(t)\cdot\hat{y}(t)$$

2.10.3 Analytic Signal or Pre-envelope

If $x(t)$ is a real valued signal, its analytic signal or pre-envelope is defined as

$$x_+(t) \;\triangleq\; x(t) + j\hat{x}(t) \qquad (2.153)$$

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 bandpass signals and in the analysis of bandpass 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. (2.153) and denote the Fourier transform of $x_+(t)$ by $X_+(f)$, we have

$$X_+(f) = X(f) + j\left[-j\,\text{sgn}(f)X(f)\right]$$

$$= \begin{cases} X(f) + X(f) & ; \quad f > 0 \\ X(f) - X(f) & ; \quad f < 0 \end{cases}$$

$$X_{+}(f) = \begin{cases} 2X(f) & \text{for } f > 0 \\ 0 & \text{for } f < 0 \end{cases} \tag{2.154}$$

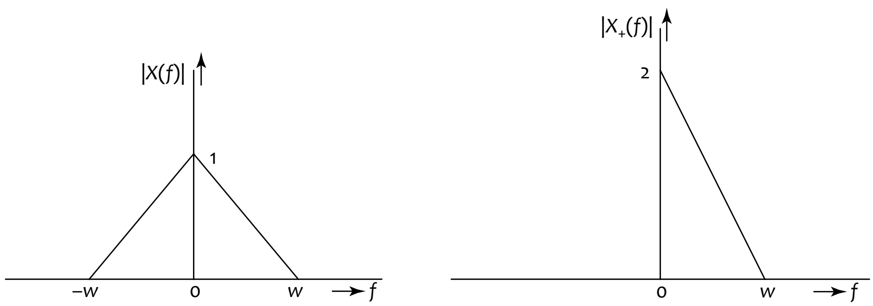


Fig. 2.38 **(a)** *Magnitude spectrum of x(t);* **(b)** *Magnitude spectrum of* $x_+(t)$

2.10.4 Complex-envelope Representation of Bandpass Signals

A bandpass 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 bandpass 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 around a carrier frequency of say a few megahertz.

Consider a real-valued bandpass signal with amplitude spectrum as shown in Fig. 2.39(a). The amplitude spectrum of the pre-envelope of $x(t)$ is shown in Fig. 2.39(b). If the pre-envelope signal $x(t)$

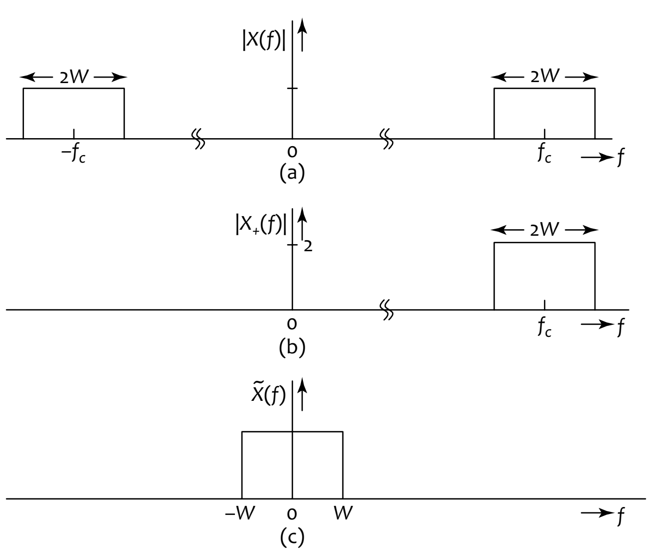


Fig. 2.39 **(a)** *Amplitude spectrum of the bandpass signal x(t)* **(b)** *Amplitude spectrum of pre-envelope of x(t)* **(c)** *Amplitude spectrum of complex envelope of x(t)*

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 $$\tilde{X}(f) = X_+(f + f_c) \tag{2.155}$$

then $$\tilde{x}(t) = x_+(t)e^{-j2\pi f_c t} \tag{2.156}$$

Hence, $$x_+(t) = \tilde{x}(t)e^{j2\pi f_c t} \tag{2.157}$$

Now, since $x_+(t) = x(t) + j\hat{x}(t)$, we have,

$$x(t) = \text{Re}[x_+(t)] = \text{Re}[\tilde{x}(t)e^{j2\pi f_c t}] \tag{2.158}$$

Because of Eq. (2.158), $\tilde{x}(t)$ *is called the complex-envelope of the bandpass signal* $x(t)$. Note that while $x(t)$ is bandpass signal, its complex-envelope $\tilde{x}(t)$ is a complex-valued lowpass signal. The reason for calling this lowpass signal, $\tilde{x}(t)$ as the complex-envelope of real valued bandpass signal, $x(t)$ is as follows. Suppose

$$x(t) = a(t)\cos[\omega_c t + \theta(t)] \tag{2.159}$$

where $a(t)$ and $\theta(t)$ are real valued lowpass signals. Then, we may write

$$x(t) = a(t)\cos[\omega_c t + \theta(t)] = \text{Re}\left[\left\{a(t)e^{j\theta(t)}\right\}e^{j\omega_c t}\right] \tag{2.160}$$

In Eq. (2.160), $\left\{a(t)e^{j\theta(t)}\right\}$ is obviously the complex envelope with $e^{j\omega_c t}$ being the complex carrier. A comparison of Eqs (2.158) and (2.160) reveals that

$$\tilde{x}(t) = a(t)e^{j\theta(t)} \tag{2.161}$$

The complex-envelope representation of a bandpass 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 bandpass systems.

2.10.5 In-phase and Quadrature Component Representation

Using complex-envelope, we shall now derive the 'in-phase and quadrature component' representation of a real-valued bandpass signal $x(t)$ with centre frequency f_c. Let $\tilde{x}(t)$ be the complex-envelope of $x(t)$. Since $\tilde{x}(t)$ is complex-valued function, let

$$\tilde{x}(t) = x_I(t) + jx_Q(t) \tag{2.162}$$

Since $\tilde{x}(t)$ is a lowpass signal of bandwidth, say, W, $x_I(t)$ and $x_Q(t)$ are also lowpass signals of the same bandwidth W, but are real valued. From Eq. (2.158), we have

$$x(t) = \text{Re}\left[\tilde{x}(t)e^{j\omega_c t}\right] = \text{Re}\left[\left\{x_I(t) + jx_Q(t)\right\}\left\{\cos\omega_c t + j\sin\omega_c t\right\}\right]$$

$$\therefore \quad \boxed{x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t} \tag{2.163}$$

This representation of the bandpass signal $x(t)$, is called the *canonical* representation of $x(t)$. The lowpass real valued signal, $x_I(t)$ is called the 'in-phase' component of the bandpass signal $x(t)$, while the real valued lowpass signal, $x_Q(t)$, is called the 'quadrature' component of the bandpass 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 $\omega_c t$.

As the reader might have noticed, in the foregoing discussion, we have used three different representations of the real valued bandpass signal, $x(t)$, with centre frequency f_c. These different representations are

$$x(t) = a(t)\cos\left[\omega_c t + \theta(t)\right] \tag{2.164}$$

$$x(t) = \text{Re}\left[\tilde{x}(t)e^{j\omega_c t}\right] = \text{Re}\left[x_+(t)\right] \tag{2.165}$$

and
$$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \tag{2.166}$$

The entities used in these three representations are obviously related. By expanding RHS of Eq. (2.164) and comparing with RHS of Eq. (2.166), we get

$$\left.\begin{aligned} x_I(t) &= a(t)\cos\theta(t) \\ x_Q(t) &= a(t)\sin\theta(t) \end{aligned}\right\} \tag{2.167}$$

By writing $\cos\left[\omega_c t + \theta(t)\right]$ of Eq. (2.164) as $\text{Re}\left[e^{j\{\omega_c t + \theta(t)\}}\right]$ and comparing with RHS of Eq. (2.165), we get

$$\tilde{x}(t) = \text{Complex Envelope} = a(t)e^{j\theta(t)} \tag{2.168}$$

so that
$$a(t) = |\tilde{x}(t)| \tag{2.169}$$

Further, from Eq. (2.157), we have

$$x_+(t) = \text{Pre} - \text{Envelope of } x(t) = \tilde{x}(t)e^{j2\pi f_c t}$$

$$|x_+(t)| = |\tilde{x}(t)| = a(t) \tag{2.170}$$

Also, from Eq. (2.167), we have

$$a(t) = \left[x_I^2(t) + x_Q^2(t)\right]^{1/2} \tag{2.171}$$

and
$$\theta(t) = \tan^{-1}\left[\frac{x_Q(t)}{x_I(t)}\right] \tag{2.172}$$

2.10.6 Lowpass Representation of Bandpass Systems

We have, in the previous section, defined a bandpass signal as one whose spectrum is non-negligible in a band of frequencies of width, say, $2W$, centered on a frequency, say, f_c.

A linear, time-invariant bandpass system is one which accepts an input signal, $x(t)$, processes it in some manner, depending upon its impulse response function, $h(t)$, and gives a bandpass signal $y(t)$, as the output signal. However, in practice, the input signal given to a bandpass system is generally a bandpass signal which is such that the passband of the bandpass system and the spectrum of the bandpass input signal are both centered on the same centre frequency, f_c. The bandwidth of the passband of the system is generally designed to be the same or slightly less than the bandwidth of the input bandpass signal, in order to prevent the out-of-band noise from entering the system along with the desired signal. For convenience in mathematical analysis, we shall, however, assume that they are equal.

In the analysis of bandpass linear time-invariant systems, our interest is invariably to determine the output signal for a given input bandpass signal. As we know, for LTI systems, the output signal can be determined by convolving the given input signal with the impulse response function of the system. But, convolution of bandpass signals is a very tedious process.

We have seen that the concept of the complex-envelope of a bandpass signal is useful as it permits us to conveniently use a lowpass representation for a bandpass signal. Hence, if we can extend that concept for lowpass description of bandpass systems too, we can considerably simplify the analysis of bandpass systems, since we can then treat them as lowpass systems and insert the carrier signal in the end result. In this connection, we first note that the impulse response function, $h(t)$, of a bandpass system is also a bandpass signal, since it is, after all, the output signal of the bandpass system under certain conditions, for a certain input signal (a unit impulse function). Thus, the input signal, $x(t)$, the impulse response function, $h(t)$, and the output signal, $y(t)$ are all bandpass signals centered on the same centre frequency f_c, and having the same bandwidth. Hence, we can represent each one of them in terms of its complex-envelope.

$$\begin{aligned} x(t) &= \mathrm{Re}\left[\tilde{x}(t)e^{j2\pi f_c t}\right] \\ &= \frac{1}{2}\left[\tilde{x}(t)e^{j2\pi f_c t} + \tilde{x}^*(t)e^{-j2\pi f_c t}\right] \end{aligned} \tag{2.173}$$

$$\begin{aligned} h(t) &= \mathrm{Re}[\tilde{h}(t)e^{j2\pi f_c t}] \\ &= \frac{1}{2}\left[\tilde{h}(t)e^{j2\pi f_c t} + \tilde{h}^*(t)e^{-j2\pi f_c t}\right] \end{aligned} \tag{2.174}$$

and

$$\begin{aligned} y(t) &= \mathrm{Re}[\tilde{y}(t)e^{j2\pi f_c t}] \\ &= \frac{1}{2}\left[\tilde{y}(t)e^{j2\pi f_c t} + \tilde{y}^*(t)e^{-j2\pi f_c t}\right] \end{aligned} \tag{2.175}$$

In the above equations, * is used to denote complex conjugation. We know that for an LTI system,

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

Substituting in this equation for $x(t)$ and $h(t)$ by using Eqs (2.173) and (2.174), we get

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} \frac{1}{4}\left\{\tilde{x}(t-\tau)e^{j2\pi f_c(t-\tau)} + \tilde{x}^*(t-\tau)e^{-j2\pi f_c(t-\tau)}\right\}\left\{\tilde{h}(\tau)e^{j2\pi f_c \tau} + \tilde{h}^*(\tau)e^{-j2\pi f_c \tau}\right\}d\tau \\ &= \frac{1}{4}\int_{-\infty}^{\infty}\left\{\tilde{x}(t-\tau)e^{j2\pi f_c(t-\tau)}\tilde{h}(\tau)e^{j2\pi f_c\tau} + \tilde{x}^*(t-\tau)e^{-j2\pi f_c(t-\tau)}\tilde{h}^*(\tau)e^{-j2\pi f_c\tau}\right\}d\tau \\ &\quad + \frac{1}{4}\int_{-\infty}^{\infty}\left\{\tilde{x}(t-\tau)e^{j2\pi f_c(t-\tau)}\tilde{h}^*(\tau)e^{-j2\pi f_c\tau} + \tilde{x}^*(t-\tau)e^{-j2\pi f_c(t-\tau)}\tilde{h}(\tau)e^{j2\pi f_c\tau}\right\}d\tau \end{aligned}$$

$$\begin{aligned} \therefore \quad y(t) &= \frac{1}{4}\int_{-\infty}^{\infty}\left\{\tilde{x}(t-\tau)e^{j2\pi f_c t}\tilde{h}(\tau) + \tilde{x}^*(t-\tau)e^{-j2\pi f_c t}\tilde{h}^*(\tau)\right\}d\tau \\ &\quad + \frac{1}{4}\int_{-\infty}^{\infty}\left\{\tilde{x}(t-\tau)e^{j2\pi f_c t}\tilde{h}^*(\tau)e^{-j4\pi f_c\tau} + \tilde{x}^*(t-\tau)e^{-j2\pi f_c t}\tilde{h}(\tau)e^{j4\pi f_c\tau}\right\}d\tau \end{aligned} \tag{2.176}$$

The integrand of the last integral contains terms with factors like $e^{-j4\pi f_c\tau}$ and $e^{j4\pi f_c\tau}$, which are of very high-frequency. However, $\tilde{x}(\tau)$ as well as $\tilde{h}(\tau)$ which are multiplying them, are low-frequency signals. Hence when this integrand is integrated for all values of τ from $-\infty$ to $+\infty$, the integral will be almost zero.

Thus, we find that Eq. (2.176) may be written as

$$y(t) = \left[\frac{1}{4}\int_{-\infty}^{\infty} \tilde{x}(t-\tau)\tilde{h}(\tau)d\tau + \frac{1}{4}\int_{-\infty}^{\infty} \left\{\tilde{x}(t-\tau)\tilde{h}(\tau)\right\}^* d\tau\right] e^{j2\pi f_c t} \tag{2.177}$$

If we now define

$$\tilde{y}(t) = \frac{1}{2}\int_{-\infty}^{\infty} \tilde{x}(t-\tau)\tilde{h}(\tau)d\tau \tag{2.178}$$

$$y(t) = \frac{1}{2}\left[\tilde{y}(t)e^{j2\pi f_c t} + \tilde{y}^*(t)e^{-j2\pi f_c t}\right] = \text{Re}\left[\tilde{y}(t)e^{j2\pi f_c t}\right] \tag{2.179}$$

Equation (2.179) therefore tells us that $\tilde{y}(t)$, defined as in Eq. (2.178) is the complex-envelope of the output signal $y(t)$, and is given by

$$\boxed{\tilde{y}(t) = \frac{1}{2}\int_{-\infty}^{\infty} \tilde{x}(t-\tau)\tilde{h}(\tau)d\tau = \frac{1}{2}\left[\tilde{x}(t) * \tilde{h}(t)\right]} \tag{2.180}$$

Equation (2.180) gives us a lowpass interpretation of the bandpass system's response to an input bandpass signal. It says that

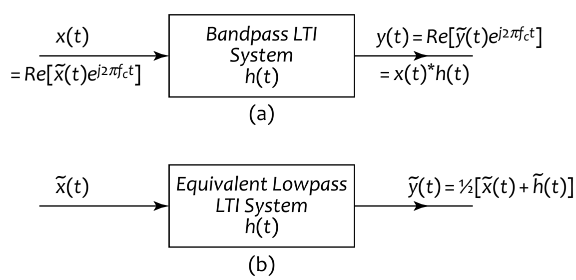


Fig. 2.40 **(a)** *Bandpass LTI system with impulse response h(t)* **(b)** *Equivalent lowpass LTI system with complex-valued impulse response, h(t)*

Thus, for a bandpass LTI system with passband centre frequency f_c and impulse response $h(t)$, the procedure for determining the output $y(t)$ for a specified bandpass input signal $x(t)$ having a spectrum with centre frequency of its bandwidth equal to f_c, using the equivalent lowpass LTI system, is as follows.

(a) Replace the bandpass LTI system with impulse response $h(t)$ by an equivalent lowpass LTI system having a complex-valued impulse response, $\tilde{h}(t)$ related to $h(t)$ by

$$h(t) = \text{Re}[\tilde{h}(t)e^{j2\pi f_c t}] \tag{2.181}$$

(b) Replace the bandpass input signal $x(t)$ by a lowpass input signal $\tilde{x}(t)$, the complex-envelope of $x(t)$, and related to $x(t)$ by

$$x(t) = \text{Re}[\tilde{x}(t)e^{j2\pi f_c t}] \tag{2.182}$$

(c) Determine $\tilde{y}(t)$, the complex envelope of the output bandpass signal $y(t)$, by convolving $\tilde{x}(t)$ and $\tilde{h}(t)$ and using the relation

$$\tilde{y}(t) = \frac{1}{2}\left[\tilde{x}(t) * \tilde{h}(t)\right] \tag{2.183}$$

(d) Finally, determine $y(t)$ from its complex-envelope using the relation

$$y(t) = \text{Re}[\tilde{y}(t)e^{j2\pi f_c t}] \tag{2.184}$$

Example 2.22

If x(t) is an energy signal, show that x(t) and $\hat{x}(t)$ are orthogonal to each other over the interval $-\infty < t < \infty$.

To show that $x(t)$ and $\hat{x}(t)$ are orthogonal over $-\infty < t < \infty$ we have to prove that

$$\int_{-\infty}^{\infty} x(t)x(t)dt = 0$$

From the generalized Parseval's theorem of FT (see Eq. 2.88), 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 $1/(\pi t)$.

$$\therefore \qquad \int_{-\infty}^{\infty} x(t)\hat{x}(t)dt = \int_{-\infty}^{\infty} x(t)\overline{\hat{x}(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{\widehat{X}(f)}df$$

$$= \int_{-\infty}^{\infty} X(f)\left[j\,\text{sgn}(f)\overline{X(f)}\right] df = j\int_{-\infty}^{\infty} \text{sgn}(f)\left|X(f)\right|^2 df$$

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.

$$\therefore \qquad \int_{-\infty}^{\infty} x(t)\hat{x}(t)dt = 0$$

Example 2.23

Find the Hilbert transform of the rectangular pulse $x(t) = A\Pi(t/\tau)$.

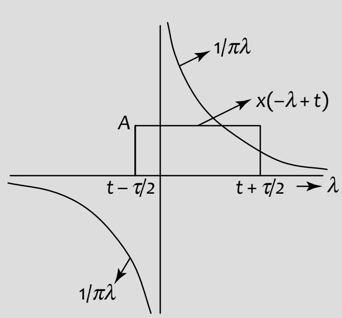


Fig. 2.41 *Convolution of x(t) with 1/(πt)*

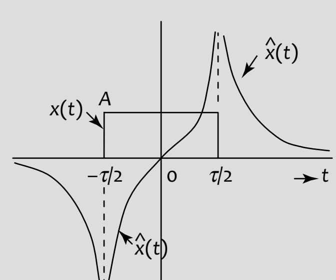


Fig. 2.42 *Signal x(t) and its Hilbert transform*

$$\hat{x}(t) = x(t) * \frac{1}{\pi t} = A\int_{-\infty}^{\infty} \frac{\Pi(t-\tau)}{\pi\lambda} d\lambda$$

$$= A\int_{t-\tau/2}^{t+\tau/2} \frac{1}{\pi\lambda} d\lambda$$

$$\therefore \qquad \hat{x}(t) = \frac{A}{\pi}\left[\log_e \left|\frac{t+\tau/2}{t-\tau/2}\right|\right]$$

Figure 2.39 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 2.24

Given a signal $x(t) = A\Pi(t/T)\cos(\omega_c t + \theta)$, find (i) its analytic signal (ii) spectrum of its analytic signal, (iii) complex envelope, and (iv) the natural envelope, a(t). Assume that fcT >> 1.

$$\hat{x}(t) = A\Pi(t/T)\sin(\omega_c t + \theta) \quad \text{(from Eq. (2.151)}$$

(i) Analytic signal or Pre-envelope of $x(t)$

$$= x_+(t) = x(t) + jx(t) = A\Pi(t/T)\left[\cos(\omega_c t + \theta) + j\sin(\omega_c t + \theta)\right]$$

$$= A\Pi(t/T)e^{i(\omega_c t + \theta)}$$

(ii) $$X_+(f) = \begin{cases} ATe^{j\theta}\text{sinc}\,(f - f_c)T; & f > 0 \\ 0 & ; f < 0 \end{cases}$$

(iii) $\tilde{x}(t) =$ complex envelope of $x(t) = x_+(t)e^{-j2\pi f_c t}$

$$= A\Pi(t/T)e^{j\theta}$$

(iv) Natural envelope of $x(t)$, i.e., $a(t) = |\tilde{x}(t)| = A\Pi(t/T)$.

SUMMARY

1. A signal is a single-valued function of one or more variables and carries some information.
2. A continuous-time signal is one whose value is defined at all instants of time, e.g., a sine wave.
3. A discrete-time signal is one whose values are defined only at a discrete set of points in time, e.g., a sequence of numbers representing the temperature at a fixed time, taken on a daily basis.
4. 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.
5. A signal whose total energy is finite and non-zero, is called an energy signal, e.g., a rectangular pulse of finite duration: $x(t) = Ae^{-|t|/T}$
6. A signal whose average power is finite and non-zero, is called a power signal, e.g., a sine wave.
7. (a) A unit impulse function is defined 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

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

8. $u(t) = \int_{-\infty}^{t} \delta(\lambda)d\lambda$ and $\frac{d}{dt}u(t) = \delta(t)$

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

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}; \quad -\infty < t < \infty$$

where, c_n's are called the complex-exponential Fourier series coefficients and are given by

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

Note: c_n's are in general complex numbers, even if $x(t)$ is a real valued function.

10. (i) 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 $x(t)$.

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

11. (a) If $x(t)$ is periodic in t with a period $t = 1/f_0$, 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,

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

and

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

a_0, a_n's and b_n's are called trigonometric Fourier series coefficients of $x(t)$.

(b) For an $x(t)$ which has even symmetry, all b_n's are zero. For a $x(t)$ which has odd symmetry, all a_n's are zero. For a $x(t)$ which is symmetry about the time axis, $a_0 = 0$.

12. (a) *Weak Dirichlet's condition* For a 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.

13. The convolution of two continuous-time signals, $x(t)$ and $y(t)$, is given by

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

14. The correlation $R_{xy}(\tau)$ between two continuous-time energy signals is given by

$$R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y^*(t-\tau)dt = \int_{-\infty}^{\infty} x(t+\tau)y^*(t)dt$$

15. The correlation between two continuous-time power signals is given by

$$R_{xy}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)y^*(t-\tau)dt = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t+\tau)y^*(t)dt$$

16. The auto-correlation of a periodic signal $x(t)$ with period T_0 is given by:

$$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)x^*(t-\tau)dt$$

17. *Fourier transform*

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

Inverse Fourier transform of $X(f) = x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi f t}\, df$

18. (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.

19. *Properties of Fourier transform*

(i) If $x(t) \xleftrightarrow{F.T} X(f)$, then $X(0) = \int_{-\infty}^{\infty} x(t)dt$ = Area under $x(t)$

(ii) $X(f)$ is in general, a complex function of frequency, even if $x(t)$ is a real valued function.
(iii) If $x(t)$ is real valued, $X(f)$ will have Hermitian symmetry.

20. 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 $x(t)$.

21. *Fourier transform theorems*

(i) $\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df = \text{Energy of } x(t)$: Parseval's Theorem

(ii) $[\alpha x(t) + \beta y(t)] \xleftrightarrow{F.T} \alpha X(f) + \beta Y(f)$: Linearity Theorem

(iii) $x(t-\tau) \xleftrightarrow{F.T} X(f) e^{-j2\pi f\tau}$: Time-delay Theorem

(iv) $x(t) e^{j2\pi f_c t} \xleftrightarrow{F.T} X(f - f_c)$. : Modulation Theorem

(v) $x(at) \xleftrightarrow{F.T} \frac{1}{|a|} X(f/a)$: Scaling Theorem

(vi) $X(t) \xleftrightarrow{F.T} x(-f)$: Duality Theorem

(vii) If $z(t) = x(t) * y(t)$, then $Z(f) = Y(f).X(f)$: Convolution Theorem

(viii) If $z(t) = x(t).y(t)$, then $Z(f) = X(f) * Y(f)$: Multiplication Theorem

(ix) $\dot{x}(t) \xleftrightarrow{F.T} j2\pi fX(f)$: Differentiation in time Theorem

(x) $-j2\pi t\, x(t) \xleftrightarrow{F.T} \frac{d}{df} X(f)$: Differentiation in Frequency Theorem

(xi) $\int_{-\infty}^{t} x(\tau) d\tau \xleftrightarrow{F.T} \frac{1}{2}\left[X(0)\delta(f) + \frac{X(f)}{j\pi f} \right]$: Integration Theorem.

22. Relationship between convolution and correlation

$$R_{xx}(\tau) = [x(t) * x(-t)]\big|_{t\to\tau}$$

23. If $x(t)$ is an energy signal, its energy spectral density (ESD) is given by the FT of its auto-correlation function

$$R_{xx}(\tau) \xleftrightarrow{F.T} S_{xx}(f)$$

and $R_{xx}(0) = \text{energy of } x(t) = \int_{-\infty}^{\infty} S_{xx}(f) df$

24. 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 $P_{xx}(f) = \text{PSD of } x(t) = F[R_{xx}(\tau)]$

25. *Properties of PSD*

(i) $P_{xx}(f)$, the PSD of a signal $x(t)$, is always non-negative.

(ii) $P_{xx}(f)$ is the Fourier transform of $R_{xx}(\tau)$

(iii) The total area under the PSD curve of a signal equals the average power of the signal.

(iv) PSD of a real valued power signal, $x(t)$, is an even function of frequency.

26. *Relationship between input and output spectral densities of an LTI system*:

(i) ESD: $S_{yy}(f) = |H(f)|^2 \cdot S_{xx}(f)$

(ii) PSD: $P_{yy}(f) = |H(f)|^2 \cdot P_{xx}(f)$

27. (a) *Hilbert Transform*: $x(t) \xleftrightarrow{F.T} \hat{x}(t) \;=\; x(t) * \dfrac{1}{\pi t}$

(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 $x(t) \xleftrightarrow{H.T} \hat{x}(t)$, then $\hat{x}(t) \xleftrightarrow{H.T} -x(t)$. $\therefore\ \hat{\hat{x}} = -x(t)$.

(iv) A signal and its Hilbert transform are Orthogonal to each other.

28. (a) *Analytic Signal* If $x(t)$ is a real valued signal, its analytic signal, or pre-envelope is defined as $x_+(t) \;\Delta\; x(t) + j\hat{x}(t)$

(b) Spectrum of analytic signal is $X_+(f) = \begin{cases} 2X(f) & \text{for } f > 0 \\ 0 & \text{for } f < 0 \end{cases}$

29. *Different Representations of Bandpass Signals*

(i) $x(t) = a(t)\cos[\omega_c t + \theta(t)]$, where $a(t)$ and $\theta(t)$ are lowpass signals—Envelope and phase representation

(ii) $x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t$: $x_I(t)$ and $x_Q(t)$ are In-phase and Quadrature lowpass signals.

Here, $a(t) = \sqrt{x_I^2(t) + x_Q^2(t)}$

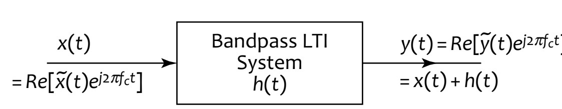
and $\theta(t) = \tan^{-1}\left[\dfrac{x_Q(t)}{x_I(t)}\right]$

(iii) $x(t) = \mathrm{Re}\left[\tilde{x}(t)e^{j2\pi f_c t}\right] = \mathrm{Re}\left[x_+(t)\right]$

where, $\tilde{x}(t) =$ complex envelope of $x(t) = a(t)e^{j\theta(t)}$

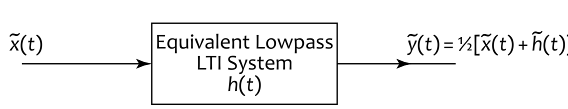
30. *Equivalent lowpass system for a bandpass system*

If:



(a)

Then:



(b)

REFERENCES AND SUGGESTED READING

1. Simon Haykin, Barry Van Veen: *Signals and Systems*, Second Edition, John Wiley & Sons (Asia) Pvt. Ltd. 2004.
2. M J Roberts, *Signals and Systems*; Tata McGraw-Hill Edition, New Delhi, 2003.
3. B P Lathi, Signals, systems and communication, B S Publications, 2001.
4. R Ziemer, W Tranter and D Fannen: *Signals and Systems: Continuous and Discrete*; Prentice-Hall, 1998.
5. Bruce Carlson et. al: *Communication Systems—An Introduction to Signals and Noise in Electrical Communication*, 4th Edition, McGraw-Hill International Edition, 2002.
6. P Ramakrishna Rao: *Signals and Systems*, Tata McGraw-Hill, 2008.

REVIEW QUESTIONS

1. Define and give an example for each of the following:
   (a) Continuous-time signals (b) Discrete-time signals
2. Define and give an example for each of the following:
   (a) Energy signals (b) Power signals
3. Determine the values of the following integrals:

   (a) $\displaystyle\int_{t=-3}^{t=3} \delta(t-4)dt$ (b) $\displaystyle\int_{t=-2}^{t=2} e^{-5t}\delta(t-1)dt$
4. When do you say that the two signals $x(t)$ and $y(t)$ are orthogonal and orthonormal?
5. What is meant by a complete set of orthonormal functions? Give an example of such a set of functions.
6. State and explain Dirichlet's conditions for convergence of Fourier series.
7. Write down the complex-exponential Fourier series expansion of the signal $x(t) = 5 \cos 10\pi t$.
8. If $X(f)$, the Fourier transform of $x(t)$, has Hermitian symmetry, comment on the nature of $x(t)$.
9. Write down the Fourier transform of $x(t) = 5 \sin (\omega_0 t + \theta)$.
10. If $X(f)$ is the Fourier transform of $x(t)$, what does $|x(f)|^2$ represent in relation to the signal $x(t)$?
11. Sketch the magnitude and phase spectra of the signal $x(t) = 20 \sin (50\pi t + 45°)$.
12. Explain the usefulness of the convolution theorem of Fourier transform in determining the convolution of two continuous-time signals.
13. Explain, graphically, the difference between convolution and correlation of two continuous-time signals.
14. Show that the ACF of two real valued continuous-time signals is an even function of τ, the lag parameter.
15. Show that the power spectral density of a power signal, $x(t)$, is the Fourier transform of its auto-correlation function.
16. Derive the relation between the output signal and input signal power spectral densities of an LTI system.
17. Sketch the magnitude and phase responses of a Hilbert transformer.
18. Show that $\hat{\hat{x}}(t) = -x(t)$.
19. Define the 'analytic signal' of a real valued signal $x(t)$.
20. Define and explain the significance of the 'Complex Envelope' of a real valued bandpass signal $x(t)$.

FILL IN THE BLANKS

1. Energy signals are those having __________ energy and __________ average power.
2. Power signals are those having __________ energy and __________ average power.
3. $x(t) = e^{-|t|}, -\infty < t < \infty$ is an example of a ______ signal (energy/power).
4. $u(t)$ is ________ signal (energy/power).
5. Area under the signal $5\delta\,(t-3)$ is _____________.
6. The value of $\int_{-5}^{5} e^{j\omega t}\delta(t)dt$ is ___________.
7. The relationship between $\delta\,(t)$ and $u(t)$ is ___________.
8. If a periodic signal $x(t)$ is purely real valued, then its complex-exponential Fourier series coefficients will have ________ symmetry.
9. The spectrum of a continuous-time periodic signal is __________ (continuous/discrete).
10. If $x(t) \xleftrightarrow{F.S} c_n^x$ then $x(t-\tau) \xleftrightarrow{F.S}$ ___________.
11. If $x(t) \xleftrightarrow{F.S} c_n^x$ then $e^{j2\pi kf_0 t}x(t) \xleftrightarrow{F.S}$ _________.
12. A periodic signal $x(t)$ is said to have half-wave symmetry (or rotational symmetry) if $x(t) =$ ________.
13. If a periodic signal $x(t)$ possesses half-wave symmetry, then its spectrum will have only _________. (even/odd) harmonics.
14. The Fourier transform of a continuous-time signal $x(t)$ exists if _____________.
15. When a signal is shifted in time, its _____________ spectrum alone changes.
16. If $x(t) \xleftrightarrow{F.T} X(f)$ then $|X(f)|^2$ represents _______ of $x(t)$ with respect to frequency, at a frequency f.
17. If the phase spectrum of $x(t)$ is $\theta\,(f)$, the phase spectrum of the signal $x(t-t_0)$ is __________.
18. Convolution is a ___________ (linear/non-linear) operation.
19. $x(t) * \delta(t-\tau) =$ ___________.
20. For real valued signals, $R_{yx}(\tau) =$ __________.
21. For complex-valued signals, $R_{yx}(\tau) =$ __________.
22. $R_{xx}(\tau) \leq$ ________ for any τ.
23. For an energy signal, $R_{xx}(0)$ represents the ___________ of the signal $x(t)$.
24. The area under the PSD curve is equal to ______ of the signal.
25. The area under the ACF curve of a power signal is equal to the _________ value of its PSD.
26. If $x(t) \xleftrightarrow{H.T} \hat{x}(t)$ then $\hat{x}(t) =$ ______* ______.
27. If $x(t) \xleftrightarrow{H.T} \hat{x}(t)$ then $\left|\hat{X}(f)\right| =$ ___________.
28. For a Hilbert transformer, the impulse response is ______________.
29. If $x_+(t)$ is the pre-envelope of a real valued signal $x(t)$, $X_+(f) = 0$ for ____________.
30. A bandpass signal $x(t) = a(t)\cos\left[\omega_c t + \theta(t)\right]$. Then, in terms of $a(t)$ and $\theta\,(t)$, the in-phase and quadrature component representation of $x(t)$ is given by ____________.

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) $x(t) = \cos tu(t)$ (f) $x(t) = \sin 3t + \cos \pi t$

2. If $x(t)$ is as shown in Fig. P-2.1, sketch and label each of the following signals.
   (a) $x(t-3)$ (b) $x(2t)$
   (c) $x(t/2)$ (d) $x(-2t)$
   (e) $x(3t-2)$

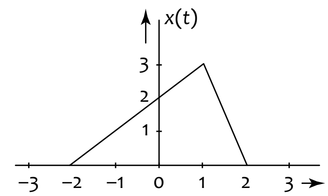


Fig. P-2.1

3. Which of the following signals are power signals, and which of them are energy signals? Are there some signals which are neither power signals nor energy signals? In each case, justify your answer. 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*) $e^{-|t|}\Pi\left(\frac{(t+1)}{6}\right)$
   (g) $te^{-2t}u(t-1)$
4. The signal $x(t)$ given by

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

is called the raised cosine pulse; and is sketched in Fig. P-2.2. Determine the total energy of this signal.

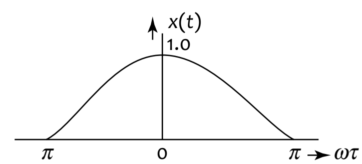


Fig. P-2.2

5. For the signal $x(t)$ shown in Fig. P-2.3(a), determine the following using

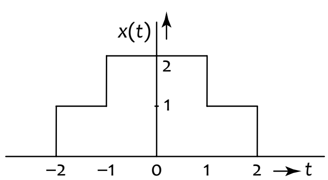


Fig. P-2.3(a)

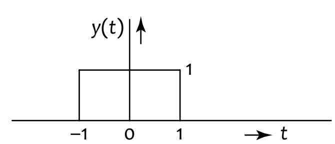


Fig. P-2.3(b)

   (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. Functions $f_1(t)$, $f_2(t)$ and $f_3(t)$ are as shown in Fig. P-2.4 (a), (b) and (c).

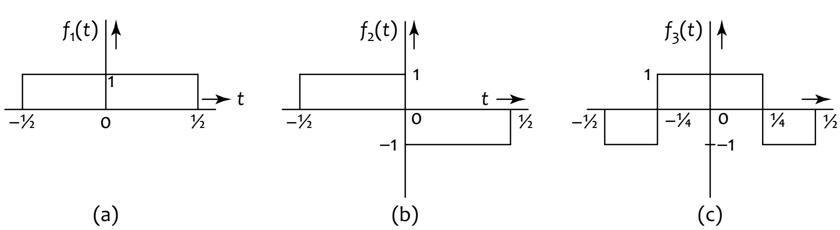


Fig. P-2.4

   (i) Show that the functions $f_1(t)$, $f_2(t)$ and $f_3(t)$ are orthogonal over the interval $(-1/2,\ 1/2)$
   (ii) If the signal $x(t)=2\sin 2\pi t$ is expanded in terms of these functions, find the integral squared error of such a representation of $x(t)$.

7. A signal $x(t)$ is as shown in Fig. P-2.5.
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 $(0, 2\pi)$.

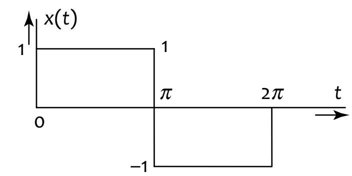


Fig. P-2.5

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

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

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, $i = 1$ to n.

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

Expand $x(t)$ over the interval (0,1) by
   a. trigonometric Fourier series
   b. Complex exponential Fourier series

10. Expand the periodic function $x(\theta)$ shown in Fig. P-2.6 using Trigonometric Fourier series.

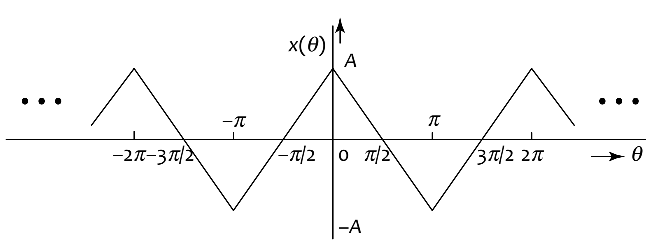


Fig. P-2.6

11. Expand the periodic waveform $x(t)$ shown in Fig. P-2.7 by complex exponential as well as trigonometric Fourier series.

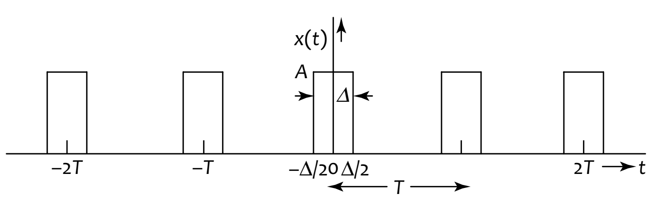


Fig. P-2.7

12. For the periodic waveform shown in Fig. P-2.8, determine the complex exponential and trigonometric Fourier series expansions.

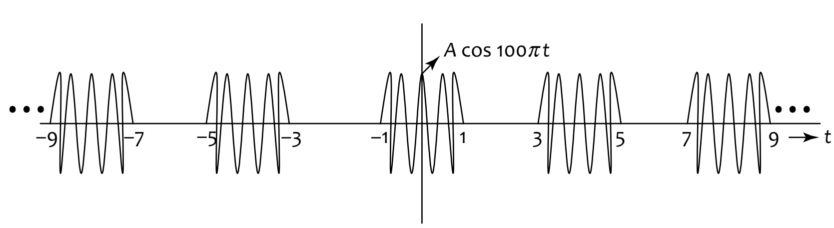


Fig. P-2.8

13. 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.
14. (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 seconds. 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.
15. 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.
16. Find the Fourier transforms of the following signals:
    (i) $x(t) = e^{-3t}u(t-2)$
    (ii) $x(t) = e^{-2|t|}$
    (iii) $x(t) = 2te^{-2t}u(t)$
    (iv) $x(t)$ shown in Fig. P-2.9
    (v) $x(t) = \left[\exp\{j2\pi(t-1) - (t-1)\}\right]u(t-1)$

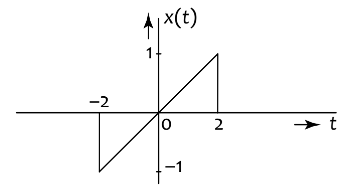


Fig. P-2.9

17. Find the signal $x(t)$ whose Fourier transform $X(f)$ is given in (i) Fig. P-2.10(a) (ii) Fig. P-2.10 (b)

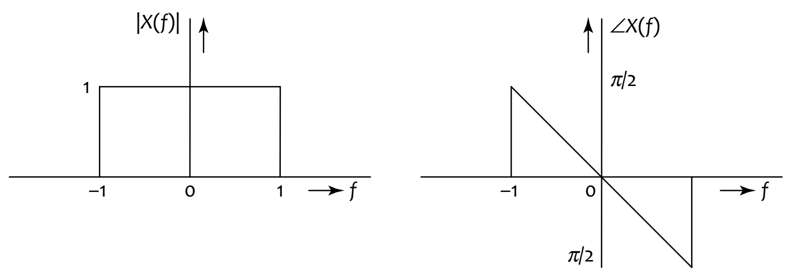


Fig. P-2.10(a)

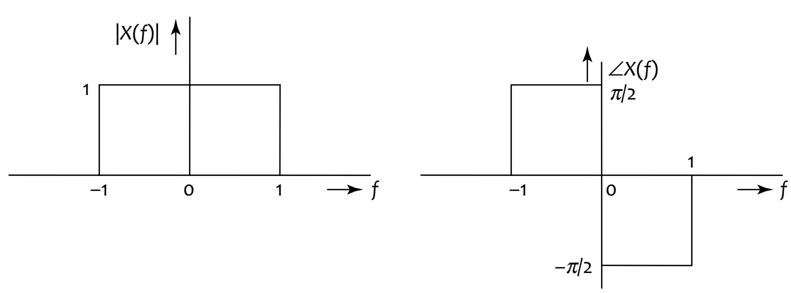


Fig. P-2.10(b)

18. Use Parseval's theorem to calculate the energy in the signal $x(t) = 4\,\text{sinc}\ 40t$.
19. Calculate the energy contained in the signal in problem 18 for $|f| \le \dfrac{3}{2\pi}$. Express it as a percentage of the total energy of the signal.
20. Find the convolution of $x(t) = 5\Pi(t/4)$ with $y(t) = 5\Pi(t/4)$.
21. Find the Fourier transform of $z(t) = 100\wedge(t/8)$ where $100\wedge(t/8)$ is a triangular pulse symmetrical about the $t = 0$ axis and having a peak amplitude of 100 and a total base width of 8 secs. (*Hint*: Use the result of Problem 20 and the convolution theorem of Fourier transform).

22. Given that $X(f)$ is the Fourier transform of $x(t)$, find the Fourier transforms of the following
    (i) $y(t) = 2x(3t-2)$

    (ii) $y(t) = x(\frac{t}{2}-1)e^{j200\pi t}$

    (iii) $y(t) = x(1-2t)$

23. Find the Fourier transforms of the signals shown in Fig. P-2.11(a) to (e).

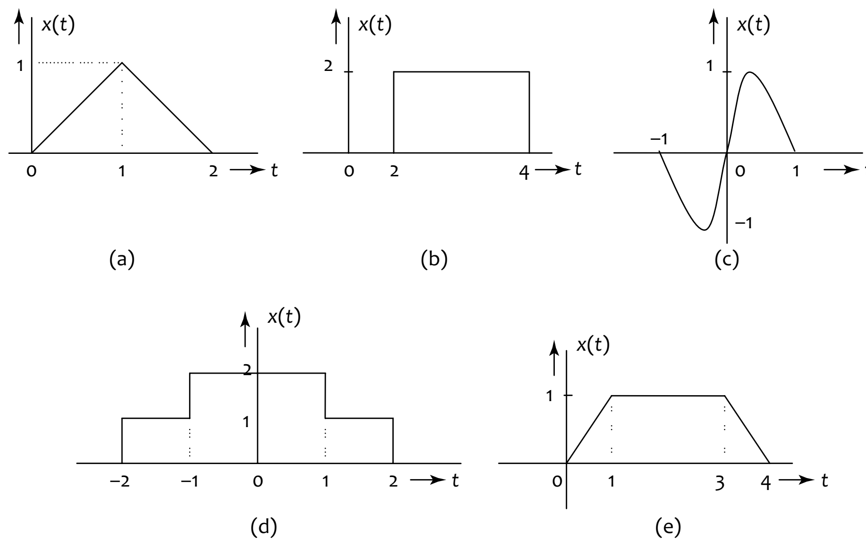


Fig. P-2.11

24. 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)

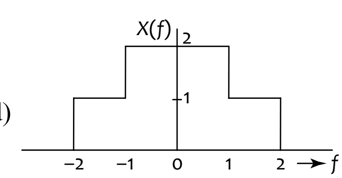


Fig. P-2.12 *of problem (d)*

25. If the signal shown in Figs P-2.11(a), (b) and (d) of Problem 23 are multiplied by cos $50\pi t$, determine and sketch the magnitudes of the Fourier transforms of the resulting signals.
26. Determine the Fourier transform of the $x(t)$ shown in Fig. P-2.13.
    (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.

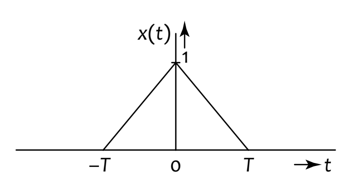


Fig. P-2.13

27. Find $R_{xy}(\tau)$ given that $x(t) = 3\cos\omega_0 t$ and $y(t) = 2\cos\omega_0 t$.
28. Determine the ACF $R_{xx}(\tau)$ for the signal $x(t)$ of Fig. P-2.14. Take its FT and determine its power spectral density.
29. Find the cross-correlation $R_{xy}(\tau)$ of the periodic signals $x(t)$ and $y(t)$ shown in Figs P-2.14 and 2.15.

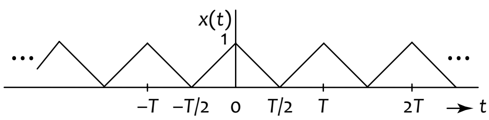


Fig. P-2.14

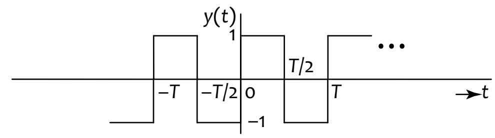


Fig. P-2.15

30. Find the ACF $R_{xx}(\tau)$ and energy spectral density, $S_{xx}(f)$ of the rectangular pulse shown in Fig. P-2.16.
31. 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

$$z(t) = \begin{cases} 10e^{2t} & \text{for } t < 0 \\ 0 & \text{for } t > 0 \end{cases}$$

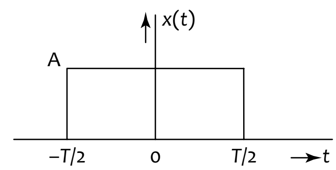


Fig. P-2.16

32. For the signals shown in Fig. P-2.17 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) $R_{zw}(\tau)$

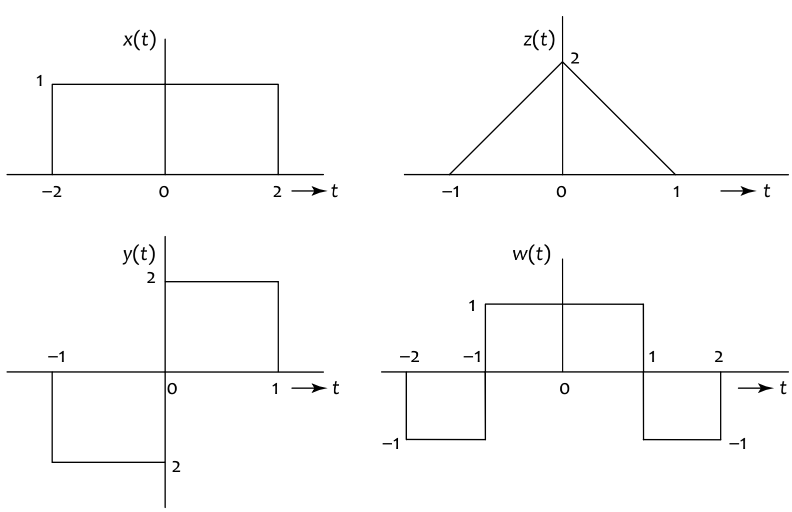


Fig. P-2.17

33. Find the power spectral density of $x(t) = 10 \cos 20\pi t$. What will be the power spectral density of $10 \sin 20\pi t$?
34. Referring to Fig. P-2.18, determine (a) $R_{yy}(\tau)$, (b) $R_{yx}(\tau)$ in terms of $R_{xx}(\tau)$

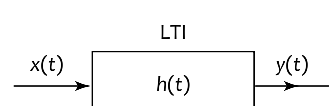


Fig. P-2.18

35. 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 r.m.s value of the signal?

36. White noise is defined as the noise for which the power spectral density is constant, as shown in Fig. P-2.19, What is the auto-correlation function of this noise? If such a noise is passed through an ideal lowpass filter with a passband gain of 1, and cutoff frequency of 500 Hz, at what intervals should we take samples of the output of the LPF if the output samples are to be uncorrelated?

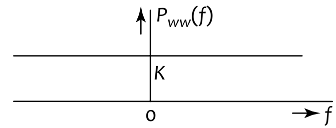


Fig. P-2.19

37. If $x(t) = (1/t)$ sin t, show that $\hat{x}(t) = (1/t)(1 - \cos t)$

38. $x_1(t)$ 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 $x_1(t)$.

39. 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) $x(t) = 5\cos 60\,\pi t . \cos 6 \times 10^4 \pi\, t$

40. $X(f)$ shown in Fig. P-2.20 is the Fourier transform of a signal $x(t)$ and is real. Determine and sketch the spectrum of each of the following signals.
(a) $y(t) = \frac{1}{2}\left[x(t) + j\hat{x}(t)\right]$
(b) $z(t) = \left[x(t) + j\hat{x}(t)\right]e^{j2\pi f_c t}$ where, $f_c >> W$
(c) $w(t) = \left[x(t) - j\hat{x}(t)\right]e^{-j2\pi f_c t}$ where, $f_c >> W$

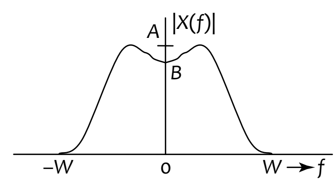


Fig. P-2.20

41. Sketch the signals:
(a) $x(t) = 200\ \mathrm{sinc}(200t)\cos 2\pi 10^4 t$
(b) $y(t) = \left[x(t) + j\hat{x}(t)\right]$ where $x(t)$ is as given in part (a)
(c) Determine $\tilde{x}(t)$, the complex envelope of $x(t)$ and sketch its spectrum.

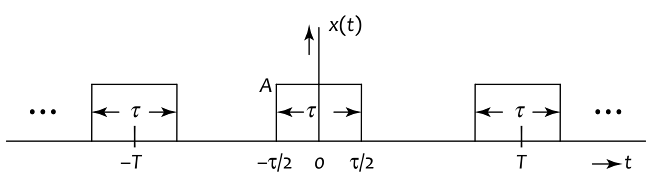
MULTIPLE CHOICE QUESTIONS

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 ant 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. The average power of the periodic signal $c_n e^{j2\pi nf_0 t}$ is
   (a) c_n^2 (b) $|c_n^2|$ (c) $|c_n|^2$ (d) $c_n^2 e^{j4\pi nf_0 t}$
6. 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, $e^{j2\pi nf_0 t}$, $n=0, \pm1, \pm2,\ldots$
   (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, $e^{j2\pi nf_0 t}$, $n=0, \pm1, \pm2,\ldots$
   (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, $e^{j2\pi nf_0 t}$, $n=0, \pm1, \pm2,\ldots$
   (d) energy of the signal may be obtained in the time-domain or from the frequency domain
7. If $x(t)=\sum_{n=-\infty}^{\infty} c_n e^{j2\pi nf_0 t}$; $-\infty<t<\infty$ and $f_0 \Delta \frac{1}{T}$ 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) $c_n=-c_n^*$
8. 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
9. 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
10. 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
11. In the discrete spectrum of the periodic signal $x(t)$ shown in the figure, the harmonic component having zero amplitude is



T = 10 seconds, τ = 0.2 seconds

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

12. Strictly speaking, one of the following signals is not Fourier transformable—which one?
(a) $e^{-|t|}$ (b) rect(t/τ) (c) tr(t/τ) (d) $\sin \omega_0 t$
13. 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
14. 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
15. 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
16. If $x(t) = 10$ sinc $5t$, the energy contained in the signal is
(a) 100 (b) 50 (c) 10 (d) 20
17. 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
18. 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) $\frac{j}{20\pi}[\delta(f+5)-\delta(f-5)]$
19. 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) $X(f)e^{j4\pi f}$
20. 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) $\frac{-1}{a-j\omega}$
21. $x(t) = 10\,\text{sinc}\,2t$ 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) $5\left[\Pi\left(\frac{(f+200)}{2}\right)+\Pi\left(\frac{(f-200)}{2}\right)\right]$
22. 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
23. If $x(t) = 10\Pi(t/10)$, $S_{xx}(f)$ is
(a) a sinc function (b) a triangular function
(c) a sinc -square function (d) a rectangular function
24. $x(t) = 5\Pi(t/10)$, the maximum value of $R_{xx}(\tau)$ is
(a) 250 (b) 50 (c) 500 (d) 25
25. $x(t) = 10\Pi(t/10)$, the maximum value of $S_{xx}(0)$ is
(a) 100 (b) 1000 (c) 500 (d) 5000

26. $x(t) = 10\Pi(t/10)$. The total area under the $S_{xx}(f)$ curve is
(a) 1000 (b) 500 (c) 100 (d) 10000
27. The signal $e^{-t}u(t)$ is applied as input to an L-section RC lowpass filter with time-constant equal to 1. The energy spectral density at the output of the filter at the 3-db cutoff frequency of the filter is
(a) 1 (b) 0.5 (c) 0.25 (d) 1.5
28. If $x(t) \xleftrightarrow{H.T} x(t)$, 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) $\hat{X}(f) = j\,\text{sgn}(f)X(-f)$
29. If $x(t) \xleftrightarrow{H.T} \hat{x}(t)$; $\hat{X}(f) = \mathcal{F}[\hat{x}(t)]$, and $\hat{X}(f) = \left|\hat{X}(f)\right| e^{j\hat{\theta}(f)}$ then,
(a) $\left|\hat{X}(f)\right| = -|X(f)|$ (b) $\hat{X}(f) = \overline{X(f)}$
(c) $\left|\hat{X}(f)\right| = |X(f)|$ and $\hat{\theta}(f) = 90°$ (d) $\left|\hat{X}(f)\right| = |X(f)|$ and $\hat{\theta}(f) = -90°$
30. $x(t) \xleftrightarrow{H.T} \hat{x}(t)$, then $\hat{\hat{x}}(t)$ equals
(a) $-x(t)$ (b) $x(t)$ (c) $\overline{x(t)}$ (d) $x(-t)$
31. $\cos 20\pi\, t.\cos 2000\pi\, t \;\underline{\Delta}\; x(t)$. Then $\hat{x}(t)$ is
(a) $\sin 20\pi\, t.\sin 2000\pi\, t$ (b) $\sin 20\pi\, t.\cos 2000\pi\, t$
(c) $\cos 20\pi\, t.\sin 2000\pi\, t$ (d) None of the above
32. If $x_+(t)$ is the analytic signal corresponding to the real valued signal $x(t)$, and if $X_+(f) = \mathcal{F}\left[x_+(t)\right]$, then $X_+(f)u(-f)$ is given by
(a) 0 (b) $2X(f)$ (c) $2X(-f)$ (d) none of the above
33. If $x(t) \xleftrightarrow{H.T} \hat{x}(t)$ and $y(t) \;\underline{\Delta}\; \hat{x}(t) - jx(t)$, then $Y(f)u(-f)$ is
(a) 0 (b) $2X(f)$ (c) $-2X(f)$ (d) none of the above
34. If $\tilde{x}(t)$ is the complex-envelope of a real valued bandpass 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) $x_1(t)\cos\omega_c t - x_2(t)\sin\omega_c t$
35. If $x(t)$, a real valued bandpass 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

Key to Multiple Choice Questions

1. (a)	2. (d)	3. (a)	4. (d)	5. (c)	6. (b)
7. (c)	8. (b)	9. (a)	10. (c)	11. (c)	12. (d)
13. (b)	14. (d)	15. (c)	16. (d)	17. (b)	18. (d)
19. (a)	20. (a)	21. (b)	22. (d)	23. (c)	24. (a)
25. (d)	26. (a)	27. (c)	28. (b)	29. (d)	30. (a)
31. (c)	32. (a)	33. (c)	34. (d)	35. (a)	

3 Signal Transmission through Systems

In this chapter, the student

- ➢ *understands the meaning and significance of the terms 'linearity', 'time-invariance', as applied to systems*
- ➢ *realizes that any LTI system is completely characterized by its impulse response or its transfer function*
- ➢ *learns to determine the impulse response, transfer function and step response of an LTI system given its electrical equivalent circuit and also comment on its stability*
- ➢ *will understand the conditions for distortionless transmission of a signal through an LTI system*
- ➢ *can determine, by applying Paley–Wiener criterion, whether a given transfer function is physically realizable or not*
- ➢ *understands the relationship between the bandwidth of an LTI system and the minimum rise time of any pulse output from that system*

3.1 INTRODUCTION

In the previous chapter we had discussed about the classification of signals and the representation of periodic signals in terms of their discrete spectra using Fourier series, and aperiodic ones in terms of their continuous spectra by using the Fourier transform. In this chapter, we will be presenting a brief review of the theory of linear time-invariant systems, as these play an important role in communication engineering.

In the discussion on signal transmission through systems, one important topic that merits serious consideration is 'distortionless transmission of a signal through a linear time-invariant system'. As we are going to see, two conditions are to be satisfied by an LTI system for distortionless transmission of a signal through it. These conditions are that the system should have the same magnitude of gain for all frequencies and the phase-shift introduced by the system must be proportional to frequency; i.e., it should have a flat magnitude response and a linear phase response over the entire frequency range from $-\infty$ to $+\infty$. However, no practical system can satisfy these two conditions over the entire frequency range from $-\infty$ to $+\infty$. Further, as we had seen in the last chapter, most of the signals, periodic or aperiodic, that we encounter in practice, will have their spectra extending from $-\infty$ to $+\infty$ of frequency, although most of their energy/average power, may be concentrated only over a certain finite band of frequencies. The foregoing facts throw up the very pertinent question: How should we define signal bandwidth and system bandwidth? It also underscores the need for an interpretation of the conditions for distortionless transmission in terms of signal bandwidth and system bandwidth. In this chapter, we will be discussing these aspects as well as the distortion suffered by pulse-like signals with very sharp leading and trailing edges when such signals are transmitted through channels with inadequate bandwidth (at the high-frequency side).

3.2 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. 3.1(a) or (b)

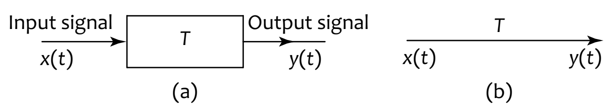


Fig. 3.1 *Diagrammatic representation of a system*

Systems may be broadly classified into

(a) Continuous-time systems
(b) 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:

(i) Static (i.e., memory-less) or dynamic (with memory)
(ii) Linear or non-linear
(iii) Time-varying or Time-invariant

3.2.1 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.

3.2.2 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.*

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.

3.2.3 Time Varying and Time-Invariant Systems

Time invariance is the property of a system which makes the behaviour 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 3.1

A certain continuous-time system, is described by the following input–output relation $y(t) = x(2t)$

Is this system

(i) Static or dynamic?

(ii) Linear or nonlinear?

(iii) Fixed or time-varying?

Justify your answers.

(i) 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.

(ii) $x(t) \xrightarrow{T} x(2t)$

$$\therefore \quad x_1(t) \xrightarrow{T} x_1(2t) = y_1(t) \text{ and } x_2(t) \xrightarrow{T} x_2(2t) = y_2(t)$$

$$\text{Then} \quad \left[a_1x_1(t) + a_2x_2(t)\right] \xrightarrow{T} \left[a_1x_1(2t) + a_2x_2(2t)\right]$$

$$= a_1y_1(t) + a_2y_2(t)$$

$\therefore$ it is a linear system.

(iii) $x(t) \xrightarrow{T} x(2t) = y(t)$

$\therefore$ $$x(t-\tau) \xrightarrow{T} x(2t-\tau) \neq y(t-\tau) \text{ since}$$

$$y(t-\tau) = x(2t-2\tau)$$

$\therefore$ the system is not time-invariant.

Example 3.2

Show that an ideal differentiator with input x(t) and output y(t) related by $y(t) = \dfrac{dx(t)}{dt}$, is a linear time-invariant system.

We are given that $$x(t) \xrightarrow{T} y(t) = \frac{dx(t)}{dt}$$

Hence, if $$x_1(t) \xrightarrow{T} y_1(t) \text{ then } y_1(t) = \frac{dx_1(t)}{dt}$$

and if $$x_2(t) \xrightarrow{T} y_2(t), \text{ then } y_2(t) = \frac{dx_2(t)}{dt}$$

also, if $\left[a_1 x_1(t) + a_2 x_2(t)\right]$ is given as the input,

$$\left[a_1 x_1(t) + a_2 x_2(t)\right] \xrightarrow{T} y(t) = \frac{d}{dt}\left[a_1 x_1(t) + a_2 x_2(t)\right] = a_1 \frac{dx_1(t)}{dt} + a_2 \frac{dx_2(t)}{dt}$$

$\therefore$ $$y(t) = a_1 y_1(t) + a_2 y_2(t)$$

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 $\dfrac{dx(t-\tau)}{dt} = \dfrac{dx(\lambda)}{d\lambda} = y(\lambda) = y(t-\tau)$

$\therefore$ the ideal differentiator is time-invariant.

3.2.4 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.

3.2.5 Impulse Response, *h*(*t*), of an LTI System

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.

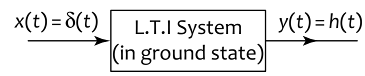


Fig. 3.2 *Impulse response of a system*

Impulse Response Characterization of an LTI System

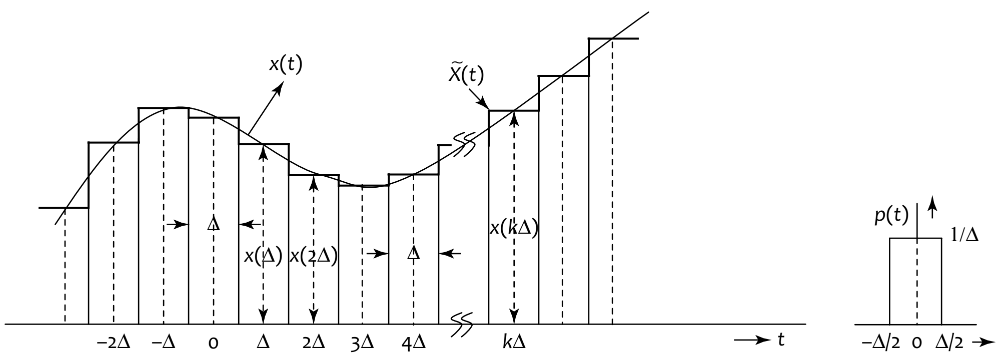


Fig. 3.3 **(a)** *Signal x(t) and its approximation* **(b)** *Rectangular pulse p(t)*

In Fig. 3.3(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

$$\tilde{x}(t) = \sum_{k=-\infty}^{\infty} x(k\Delta)p(t-k\Delta)\Delta \tag{3.2}$$ TQ1

$$x(t) = \lim_{\Delta\to 0} \tilde{x}(t) = \lim_{\Delta\to 0}\left[\sum_{k=-\infty}^{\infty} x(k\Delta)p(t-k\Delta)\Delta\right] \tag{3.3}$$

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 $d\tau$.

$\therefore$
$$x(t) = \lim_{\Delta\to 0} \tilde{x}(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau \tag{3.4}$$

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

$$x(t) \xrightarrow{T} y(t) = T[x(t)]$$

$$\therefore \quad y(t) = T\left[\int_{-\infty}^{\infty} x(\tau)\delta(t-\tau)d\tau\right] = \int_{-\infty}^{\infty} x(\tau)T[\delta(t-\tau)]d\tau$$

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

$$\boxed{y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau} \tag{3.5}$$

The integrals in Eq. (3.5) are called '*convolution integrals*', or the '*superposition integrals*'.

From Eq. (3.5), we find that a 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 say that an LTI system is completely characterized by its impulse response function $h(t)$.*

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 $t = 0$.

$\therefore$ 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.

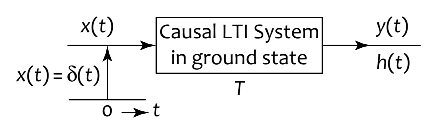


Fig. 3.4 *Impulse response of a causal system*

$$\boxed{\therefore \textit{ For a causal LTI system, } h(t) = 0 \text{ for } t < 0.} \tag{3.6}$$

In the light of Eq. (3.6), the convolution integrals, for a *causal* LTI system, can be written as

$$\boxed{y(t) = \int_{0}^{\infty} x(t-\tau)h(\tau)d\tau = \int_{-\infty}^{t} x(\tau)h(t-\tau)d\tau} \tag{3.7}$$

Step Response of an LTI System *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 an LTI system, the step response and impulse response are related through the following equation

$$g(t) = \int_{-\infty}^{t} h(\lambda)d\lambda \quad \therefore \quad h(t) = \frac{dg(t)}{dt} \tag{3.8}$$

Example 3.3

An R-C lowpass filter is shown in Fig. 3.5. Find its impulse response and step response.

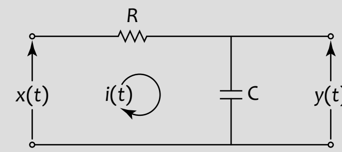


Fig. 3.5 *R-C lowpass filter*

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

$$RC\left[sY(s) - y(0^-)\right] + Y(s) = 1 \quad \text{for} \quad t > 0$$

$$\therefore \qquad Y(s) = \frac{1/RC}{s + 1/RC}$$

Now, taking the inverse Laplace transform of the above,

$$y(t) = \begin{Bmatrix} \text{Response of the system to a unit impulse} \\ \textit{function} \text{ when the system in ground state} \end{Bmatrix} = h(t) = \frac{1}{RC}e^{-t/RC}.u(t).$$

To find the step response, we note that

$$g(t) = \int_{-\infty}^{t} h(\lambda)\, d\lambda = \int_{-\infty}^{t} \frac{1}{RC}e^{-\lambda/RC}\ u(\lambda)\, d\lambda$$

$$\frac{1}{RC}\int_{0}^{t} e^{-\lambda/Rc} d\lambda = \left[1 - e^{-t/RC}\right]\ u(t)$$

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 $h(t)$.

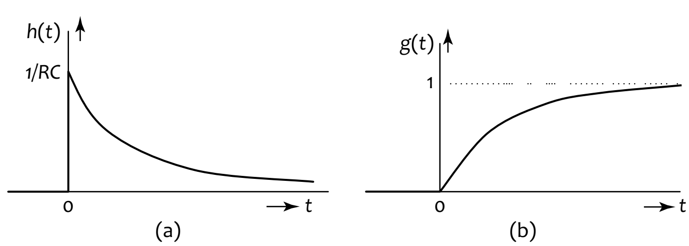


Fig. 3.6 **(a)** *Impulse response and* **(b)** *Step-response of an R-C lowpass filter*

The impulse response and the step response of the given lowpass *R*-*C* filter are plotted in Figs 3.6(a) and (b) respectively.

Example 3.4

A particular LTI system has $h(t) = e^{-2t}\,u(t)$. Determine its output signal $y(t)$ corresponding to an input signal $x(t) = u(t)$.

$$y(t) = x(t) * h(t) = \int_0^{\infty} e^{-2t} u(t) u(t-\tau)\, d\tau$$

since $\quad u(\tau) = 0$ for $\tau < 0$ and $u(t-\tau) = 0$ for $\tau > t$

$$y(t) = \int_0^{t} e^{-2t}\, d\tau = -\frac{1}{2} e^{-2t}\Big|_{\tau=0}^{t} = \frac{1}{2}\left(1 - e^{-2t}\right)$$

Example 3.5

If $x(t)$ and $y(t)$ are as shown in Fig. 3.7(a) and Fig. 3.7(b) determine graphically, the signal $z(t) = x(t) * y(t)$.

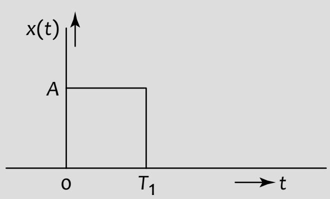


Fig. 3.7 **(a)** $x(t)$

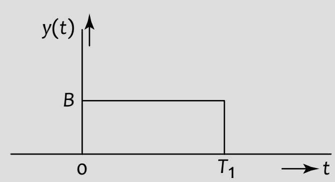


Fig. 3.7 **(b)** $y(t)$

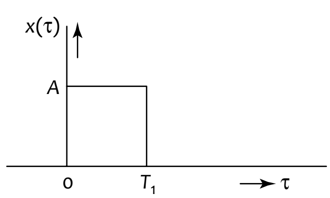


Fig. 3.8 **(a)** $x(\tau)$

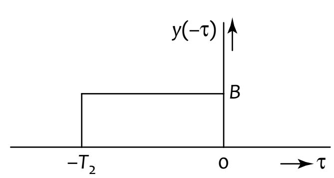


Fig. 3.8 **(b)** $y(-\tau)$

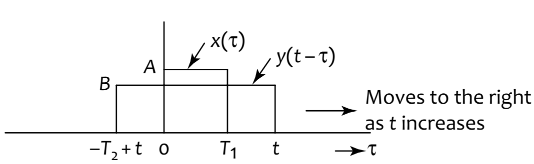


Fig. 3.8 **(c)** $x(\tau)$, $y(t-\tau)$

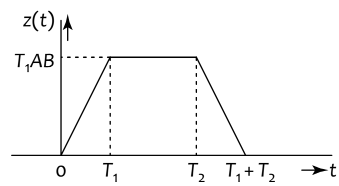


Fig. 3.8 **(d)** $z(t)$

$z(t) = \displaystyle\int_{-\infty}^{\infty} x(\tau) y(t-\tau) d\tau$ = area under the product of $x(\tau)$ and $y(t-\tau)$ for t.

From Fig. 3.8(c) the following points are evident:

(i) When $t \le 0$, the product of $x(\tau)$ and $y(t-\tau)$ is zero, as there is no overlap of the two.

(ii) 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.

(iii) 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.

(iv) 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 $t > (T_1 + T_2)$.

(v) 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 $= T_1 + T_2$.

(vi) 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 3.6

The input x(t) and the corresponding output y(t) of a causal LTI system T are as shown in Fig. 3.9(a) and (b) respectively. Find the impulse response function h(t) of the system.

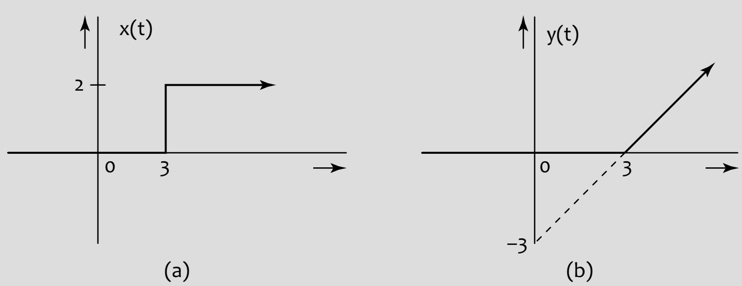


Fig. 3.9 **(a)** *Input x(t)* **(b)** *Output y(t)*

We know that for an LTI system, if $x(n) \xrightarrow{T} y(n)$, then $\dot{x}(n) \xrightarrow{T} \dot{y}(n)$

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,

$$\dot{y}(t) = u(t-3)$$

$\therefore$
$$2\delta(t-3) \xrightarrow{T} u(t-3)$$

or
$$\delta(t) \xrightarrow{T} \frac{1}{2}u(t) \quad (\because \text{ system is LTI})$$

$\therefore$ impulse response function $h(t) = \dfrac{1}{2}u(t)$.

3.2.6 Stability

One important way of defining stability of a system, is in terms of the '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 3.1 *An LTI system T with impulse response h(t), is stable in the BIBO sense iff h(t), is absolutely integrable.*

i.e., $$iff \int_{-\infty}^{\infty} |h(t)|dt < \infty$$

Proof (i) **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.

(ii) 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

$$x(\tau) = \begin{cases} 1 & if \quad h(-\tau) > 0 \\ -1 & if \quad h(-\tau) < 0 \\ 0 & if \quad h(-\tau) = 0 \end{cases} \tag{3.9}$$

Since $|x(t)|$ is either 1 or zero, $x(t)$ is obviously a bounded signal, bound to a value 1. When it is given as input to T, the output is given by

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

$$\therefore \qquad y(0) = \int_{-\infty}^{\infty} x(\tau)h(-\tau)d\tau = \int_{-\infty}^{\infty} |h(-\tau)|\,d\tau \quad \text{from Eq. (3.9)}$$

$$\therefore \qquad y(0) = \int_{-\infty}^{\infty} |h(-\tau)|\,d\tau = \int_{-\infty}^{\infty} |h(\tau)|\,d\tau$$

But

$$\int_{-\infty}^{\infty} |h(\tau)|\,d\tau \quad \text{is not finite (given)}$$

$\therefore$ $y(0)$ is not finite. $\therefore$ $y(t)$ is not a bounded signal even though $x(t)$ is. Hence, T is not stable.

$\therefore$ $\boxed{\text{An LTI system with impulse response } h(t) \text{ is stable in the BIBO sense iff } \int_{-\infty}^{\infty} |h(t)|\,dt < \infty}$

Example 3.7

Examine the stability of the system shown in Fig. 3.10.

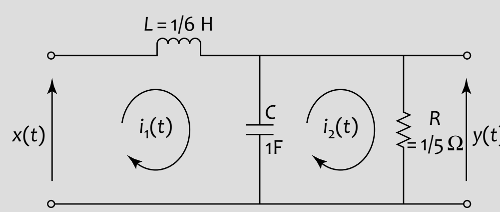


Fig. 3.10 *System for Example 3.7*

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

$$LC\frac{d^2y(t)}{dt^2} + \frac{L}{R}\frac{dy(t)}{dt} + y(t) = x(t)$$

To examine the BIBO stability, we have to first find $h(t)$, of the system. So, 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.

$$LC\left[s^2Y(s) - \cancel{sy(0^-)} - \cancel{\dot{y}(0^-)}\right] + \frac{Ls}{R}Y(s) - \cancel{\frac{L}{R}y(0^-)} + Y(s) = 1$$

i.e., $$Y(s)\left[LCs^2 + \frac{L}{R}s + 1\right] = 1$$

$$\therefore \quad Y(s) = \frac{1}{LCs^2 + \frac{L}{R}s + 1} = \frac{1}{\frac{1}{6}s^2 + \frac{5}{6}s + 1} = \frac{6}{s^2 + 5s + 6}$$

$$= \frac{-6}{s+3} + \frac{6}{s+2}$$

Now, taking the inverse Laplace transform on both sides,

$$y(t) = h(t) = 6\left[e^{-2t} - e^{-3t}\right] u(t)$$

$$\therefore \quad \int_{-\infty}^{\infty} |h(t)|\,dt = 6\int_{0}^{\infty} \left|e^{-2t} - e^{-3t}\right| dt = 6\int_{0}^{\infty} \left(e^{-2t} - e^{-3t}\right) dt = 1$$

$\therefore$ $\int_{-\infty}^{\infty} |h(t)|\,dt$ is finite and the given system is stable in the BIBO sense.

3.2.7 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, $$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau = \int_{-\infty}^{\infty} e^{j\omega_0(t-\tau)}h(\tau)d\tau$$

$$\therefore \quad y(t) = e^{j\omega_0 t}\int_{-\infty}^{\infty} h(\tau)e^{-j\omega_0\tau}d\tau = e^{j\omega_0 t}H(f)\Big|_{f=f_0}, \qquad (3.10)$$

where $H(f) = \mathrm{F}\left[h(t)\right]$, 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. (3.10), 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. (3.10) we may write:

$$\left.\frac{y(t)}{x(t)}\right|_{x(t)=e^{j\omega_0 t}} = H(f_0) \tag{3.11}$$

$H(f_0)$ 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:

$$H(f) = |H(f)|e^{j\theta(f)} \tag{3.12}$$

In Eq. (3.12), $|H(f)|$ 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 3.8

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.

Taking the Laplace transform on both sides with $x(t)$ equal to $\delta(t)$ and all initial conditions as zero, we get

$$sY(s) + 6Y(s) = X(s) = 1$$

$$Y(s)[s+6] = 1 \quad \therefore \quad Y(s) = \frac{1}{s+6}$$

$$h(t) = y(t) = e^{-6t}u(t) \quad \therefore \quad H(f) = \frac{1}{6 + j2\pi f}$$

$$|H(f)| = \frac{1}{\sqrt{36 + 4\pi^2 f^2}} \text{ and } \theta(f) = -\tan^{-1}\left(\frac{2\pi f}{6}\right) = -\tan^{-1}\left(\frac{\pi}{3} f\right)$$

The magnitude and phase responses are as shown in Fig. 3.11(a) and (b) respectively.

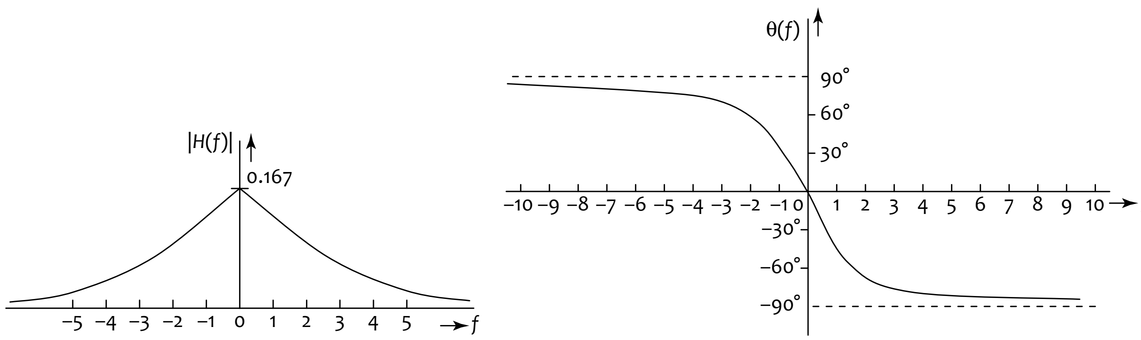


Fig. 3.11 **(a)** *Magnitude response* **(b)** *Phase response*

SIGNAL TRANSMISSION THROUGH LTI SYSTEMS **3.3**

In this section, we shall discuss two specific aspects of transmission of a signal through an LTI system: (i) un-distorted transmission of a signal through an LTI system, and (ii) 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.

3.3.1 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

$$y(t) = Ax(t-\tau) \tag{3.13}$$

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. (3.13), we get

$$Y(f) = AX(f)e^{-j2\pi f\tau} \tag{3.14}$$

But we know that $y(t) = x(t) * h(t)$

and hence $$Y(f) = X(f)\cdot H(f)$$

or, $$\frac{Y(f)}{X(f)} = H(f) = \text{Transfer function of the system}$$

$\therefore$ from Eq. (3.14), we have,

$$\frac{Y(f)}{X(f)} = H(f) = Ae^{-j2\pi f\tau} \tag{3.15}$$

From Eq. (3.15), 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

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 3.12(a) and (b) respectively.

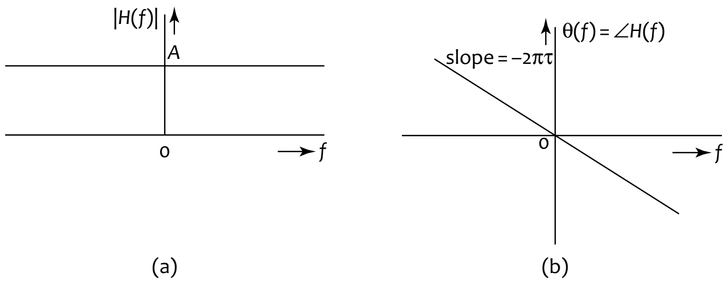


Fig. 3.12 **(a)** *Magnitude response* **(b)** *Phase response of a distortionless transmission system*

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 $-\infty$ to $+\infty$, 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 upto 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 eventhough 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. 3.13.

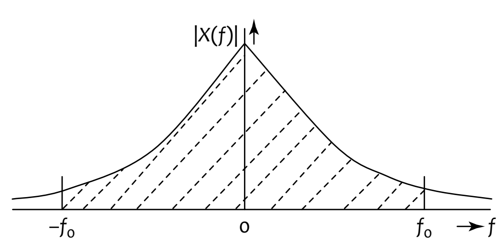


Fig. 3.13 *Two-sided spectrum of an arbitrary signal, the bandwidth of which is specified as f_0 Hz*

Thus, for the signal whose magnitude spectrum is shown in Fig. 3.10, f_0 is called the signal bandwidth, since

$$\frac{\int_{-f_0}^{f_0} |X(f)|^2 df}{\int_{-\infty}^{\infty} |X(f)|^2 df} = 0.95$$

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 3.14 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. 3.14 it is the frequency range from f_l to f_h. This bandwidth $(f_h - f_l)$, is called as the half-power bandwidth, or the 3-db bandwidth.

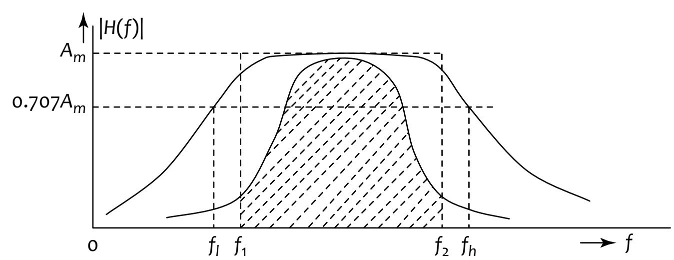


Fig. 3.14 *Frequency response characteristic of a system and the spectrum of an input signal*

In Fig. 3.14 the signal bandwidth f_1 to f_2 lies within the system bandwidth, f_l 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, in so far 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. In so far 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 3.9

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 = τ and that it has the maximum value at that point.

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,

$$h(t) = \int_{-\infty}^{\infty} Ae^{-j2\pi f\tau} . e^{j2\pi f\, t} df$$

$$= A\int_{-\infty}^{\infty} e^{-j2\pi f(\tau - t)} df = A\int_{-\infty}^{0} e^{-j2\pi f(\tau - t)} df + A\int_{0}^{\infty} e^{-j2\pi f(\tau - t)} df$$

$$= A\int_{0}^{\infty} e^{j2\pi f(\tau - t)} df + A\int_{0}^{\infty} e^{-j2\pi f(\tau - t)} df$$

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
$$h(t)\Big|_{t=\tau-t_1} = 2A\int_0^{\infty} \cos 2\pi f t_1 df$$

∴
$$h(t)\Big|_{t=(\tau+t_1)} = h(t)\Big|_{t=(\tau-t_1)}$$

Thus, $h(t)$ has even symmetry about $t = \tau$. Also, since

$$h(t) = 2A\int_{-\infty}^{\infty} \cos 2\pi f(\tau - t) df$$

and since $\cos 2\pi f(\tau - t) = \cos 0 = 1$ for $t = \tau$, $h(t)$ takes the maximum value at $t = \tau$.

3.3.2 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,

$$Y(f) = X(f) \cdot H(f) \tag{3.16}$$

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. (3.16) we may write

$$|Y(f)| e^{j\theta_y(f)} = |X(f)| e^{j\theta_x(f)} \cdot |H(f)| e^{j\theta_H(f)} \tag{3.17}$$

∴
$$|Y(f)| = |X(f)| \cdot |H(f)| \tag{3.18}$$

and
$$\theta_y(f) = \theta_x(f) + \theta_H(f) \tag{3.19}$$

Equations (3.18) and (3.19) show respectively, how the transfer function of the system modifies magnitude spectrum and the phase spectrum of the input signal. 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 passbands and the bands of frequencies which are highly attenuated, are called the stopbands of the filter. Depending on the type of filter, there may be one or more passbands and stopbands.

A filter which transmits, without any attenuation, all frequencies of the input signal that are less than a certain frequency, called the cutoff frequency and rejects all frequencies above it, is called a *lowpass filter*. A filter whose stopband is below a certain cutoff frequency and its passband above that frequency, is called a *highpass filter*. Other types of filters which are of interest are the *bandpass 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.

Ideal Lowpass Filter Consider an ideal lowpass filter with a passband gain A, passband width B Hz and a linear phase response with a slope of $-2\pi\tau$. Then its transfer function is

$$H(f) = A\prod(f/2B)e^{-j2\pi f\tau} \tag{3.20}$$

Taking the inverse Fourier transform, we get its impulse response function as

$$h(t) = 2AB \text{ sinc } 2B(t - \tau) \tag{3.21}$$

The magnitude response, phase response and impulse response functions of this ideal LPF are shown in Fig. 3.15.

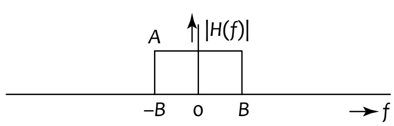


Fig. 3.15(a) *Magnitude response of an ideal LPF*

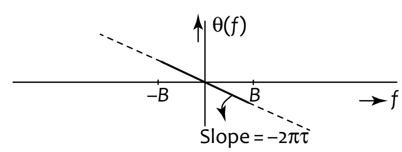


Fig. 3.15(b) *Phase response of an ideal LPF*

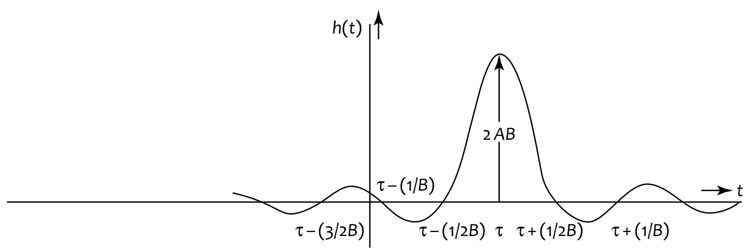


Fig. 3.15(c) *Impulse response h(t) of an ideal LPF*

Highpass Filter Consider an ideal high pass filter with passband gain A, cutoff frequency f_c and time-delay t_0 second. Its transfer function may be written as

$$H(f) = |H(f)|e^{j\theta(f)} \tag{3.22}$$

where $$|H(f)| = A[1 - \prod(f/2f_c)] \tag{3.23}$$

and $$\theta(f) = -2\pi f t_0 \tag{3.24}$$

The magnitude response and phase response of this Ideal HPF are shown in Fig. 3.16.

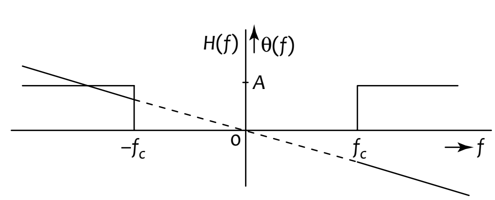


Fig. 3.16 *Magnitude and phase response of an ideal HPF*

Example 3.10

Find the impulse response $h_{HP}(t)$ of an ideal highpass filter with a passband gain of A, cutoff frequency of f_c Hz and a linear phase response with a slope of $-2\pi\tau_o$.

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
$$\theta(f) = -2\pi t_0 f$$

$\therefore$
$$h(t) = F^{-1}[H(f)] = F^{-1}\{[A - A\Pi(f/2f_c]e^{-j2\pi t_0 f}\}$$
$$= F^{-1}\left[Ae^{-j2\pi t_0 f} - A\Pi(f/2f_c e^{-j2\pi t_0 f}\right]$$
$$= AF^{-1}\left[e^{-j2\pi t_0 f}\right] - AF^{-1}\left[\Pi(f/2f_c)\right] * F^{-1}\left[e^{-j2\pi t_0 f}\right]$$
$$= A\delta(t - t_0) - A[2f_c \text{ sinc } 2f_c t)] * \delta(t - t_0)$$
$$h(t) = A\delta(t - t_0) - 2Af_c \text{ sinc } 2f(t - t_0)$$

Since the sinc function extends in time from $-\infty$ to $+\infty$, the ideal HPF is also not a causal system and hence, is not physically realizable.

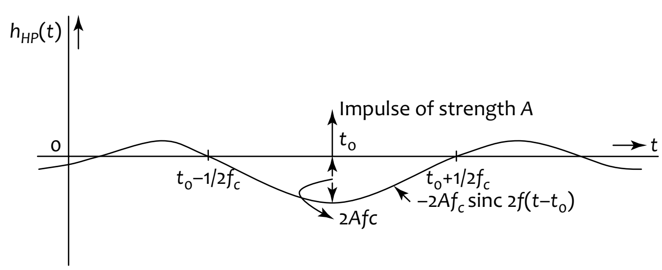


Fig. 3.17 *Impulse response of an ideal highpass filter of passband gain A and cutoff frequency f_c*

Ideal Bandpass Filter Consider an ideal bandpass filter with passband from f_1 to f_2, passband gain A and a time delay t_0 second.

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,
$$|H(f)| = A\Pi[(f + f_0)/B] + A\Pi[(f - f_0)/B] \tag{3.25}$$

and
$$\theta(f) = -2\pi f t_0 \tag{3.26}$$

Taking the inverse Fourier transform of $H(f)$, we get

$$h(t) = 2AB[\text{sinc } B(t - t_0)]\cos 2\pi f_0(t - t_0) \tag{3.27}$$

The magnitude and phase responses of this ideal BPF are shown in Fig. 3.18(a) while its impulse response function is shown in Fig. 3.18(b).

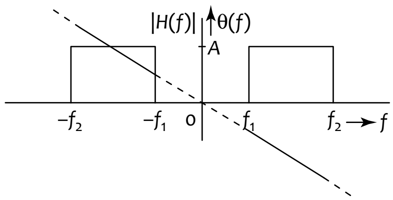


Fig. 3.18(a) *Magnitude and phase response of the Ideal BPF*

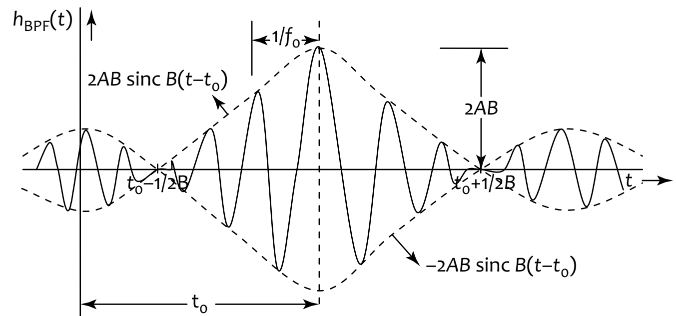


Fig. 3.18(b) *Impulse response of an ideal BPF*

Example 3.11

Determine the impulse response function, h(t) of an ideal BPF with passband gain of A and passband bandwidth of B Hz centered on f_0 Hz and having a linear phase response.

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 $$|H(f)| = A\Pi[(f+f_0)/\mathrm{B}] + A\Pi[(f-f_0)/\mathrm{B}]$$

and $$\theta(f) = -2\pi f t_0$$

Here, $(-2\pi t_0$ is the gradient of the linear phase response)

$\therefore$ $$h(t) = \mathrm{F}^{-1}\left\{\left[A\Pi\{(f+f_0)/B\} + A\Pi\{(f-f_0)/B\}\right]e^{-j2\pi f t_0}\right\}$$

$$= \mathrm{F}^{-1}\left[A\Pi\{(f+f_0)/B\}e^{-j2\pi f t_0} + A\Pi\{(f-f_0)/B\}e^{-j2\pi f t_0}\right]$$

$$= A\{(B\,\mathrm{sinc}\,Bt\}e^{-j2\pi f t_0} * \delta(t-t_0) + A\{(B\,\mathrm{sinc}\,Bt\}e^{j2\pi f t_0} * \delta(t-t_0)$$

$$= \left\{A[B\,\mathrm{sinc}\,B(t-t_0)]e^{-j2\pi f_0(t-t_0)} + [B\,\mathrm{sinc}\,B(t-t_0)]e^{j2\pi f_0(t-t_0)}\right\}$$

$$= 2AB\,\mathrm{sinc}\,B(t-t_0)]\cos 2\pi\ f_0(t-t_0)$$

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 $t = -\infty$ to $t = +\infty$. Thus, $h(0) \neq 0$, for $t < 0$ for all these filters and hence they are non-causal and cannot be physically realized.

3.4 PALEY–WIENER CRITERION FOR PHYSICAL REALIZABILITY

Till now, we have been discussing the question of physically 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

$$\int_{-\infty}^{\infty} \frac{|\log_e |H(f)||}{1+f^2} df < \infty \tag{3.28}$$

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. (3.28) 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. (3.28), 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. (3.28), 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.

SYSTEM BANDWIDTH AND RISE TIME 3.5

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 lowpass filter, say, a first-order RC lowpass filter, or an ideal LPF.

Example 3.12

Find the relation between bandwidth and the rise-time of a pulse in the case of the first-order R-C lowpass filter shown in Fig. 3.19.

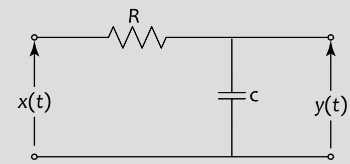


Fig. 3.19 *First-order R-C lowpass filter*

The differential equation is

$$RC\frac{dy(t)}{dt} + y(t) = x(t)$$

Taking Fourier transform on both sides, we get

$$\frac{Y(f)}{X(f)} = H(f) = \frac{1}{1 + j2\pi fRC} \tag{3.29}$$

In Example 3.3, we have shown that the unit-step response, $g(t)$ of the system, is given by

$$g(t) = (1 - e^{-t/RC})u(t) \tag{3.30}$$

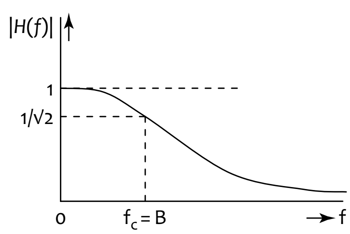


Fig. 3.20 *Magnitude response of a first-order RC lowpass filter*

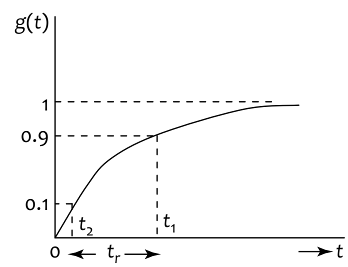


Fig. 3.21 *Unit-step response of a first-order RC lowpass filter*

The 3db frequency f_c of this filter is such that

$$|H(f)|\Big|_{f=f_c} = \frac{1}{\sqrt{2}}$$

This gives $$f_c = B = \frac{1}{2\pi RC} \quad \therefore RC = \frac{1}{2\pi B}$$

Substituting this for RC in Eq. (3.30) we get

$$g(t) = (1 - e^{-2\pi Bt})u(t) \tag{3.31}$$

Referring to Fig. 3.21,

$$g(t)\Big|_{t=t_1} = (1 - e^{-2\pi Bt_1}) = 0.9 \tag{3.32}$$

$$g(t)\Big|_{t=t_2} = (1 - e^{-2\pi Bt_2}) = 0.1 \tag{3.33}$$

From Eq. (3.32) and (3.33), we have

$$e^{-2\pi Bt_1} = 0.1 \quad \text{and} \quad e^{-2\pi Bt_2} = 0.9$$

$\therefore$ $$e^{2\pi B(t_1 - t_2)} = 0.9$$

Taking logarithm to the base e on both sides,

$$(t_1 - t_2) = t_r = \text{Rise time} = \frac{\log_e 9}{6.28B} = 0.35\big/B$$

Thus,

$$\boxed{t_r = \frac{0.35}{B} \text{ for a first order } RC \text{ lowpass filter}} \tag{3.34}$$

Rise Time with an Ideal LPF We will now model the cable or transmission line as an ideal LPF with cutoff frequency B Hz. Let the magnitude response be as shown in Fig. 3.22.

Without loss of generality, we shall further assume that the time-delay τ, of the ideal LPF is zero. Then, from Eq. (3.21), we have

$$h(t) = 2B\,\text{sinc}\,2Bt \tag{3.35}$$

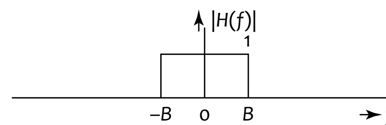


Fig. 3.22 *Magnitude response of an ideal LPF*

The step-response $g(t)$ is given by

$$g(t) = \int_{-\infty}^{t} h(\lambda)d\lambda = \int_{-\infty}^{t} 2B\ \text{sinc}\ 2B\lambda\ d\lambda \tag{3.36}$$

Put $\tau = 2B\lambda$ we get,

$$g(t) = \int_{-\infty}^{0} \text{sinc}\ \tau\ d\tau + \int_{0}^{2Bt} \text{sinc}\ \tau\ d\tau \tag{3.37}$$

But,

$$\int_{-\infty}^{0} \text{sinc}\ \tau\ d\tau = \int_{0}^{\infty} \text{sinc}\ \tau\ d\tau = \frac{1}{2} \tag{3.38}$$

The other integral in Eq. (3.37) has to be evaluated numerically or by referring to the table of 'sine integral function', $Si(\theta)$, where

$$Si(\theta)\ \underline{\Delta}\ \int_{0}^{\theta} \frac{\sin x}{x}\ dx \tag{3.39}$$

Since

$$\text{sinc}\ \tau\ \underline{\Delta}\ \frac{\sin \pi\tau}{\pi\tau}$$

Putting $x = \pi\tau$, we get

$$Si(\theta)\ \underline{\Delta}\ \int_{0}^{\theta} \frac{\sin x}{x}\ dx = \pi \int_{0}^{\theta/\pi} \text{sinc}\ \tau\ d\tau \tag{3.40}$$

A plot of the above sine integral function, $Si(\theta)$, is given in Fig. 3.23.

From Eq. (3.39), we have

$$\frac{1}{\pi} Si(\theta) = \int_{0}^{\theta/\pi} \text{sinc}\tau\ d\tau \tag{3.41}$$

Now, reverting to Eq. (3.37), and recalling that our interest is in evaluating $g(t)$,

$$g(t) = \frac{1}{2} + \int_{0}^{2Bt} \text{sinc}\ \tau\ d\tau = \frac{1}{2} + \frac{1}{\pi} Si(\theta) \tag{3.42}$$

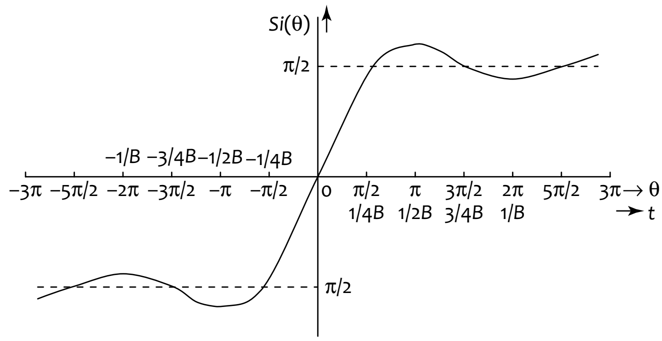


Fig. 3.23 *The sine integral function*

$\therefore$ putting $\frac{\theta}{\pi} = 2Bt$, we have

$$t = \frac{\theta}{2\pi B} \tag{3.43}$$

Using Eq. (3.42) and Eq. (3.43), we plot $g(t)$ vs t; and this is shown in Fig. 3.24.

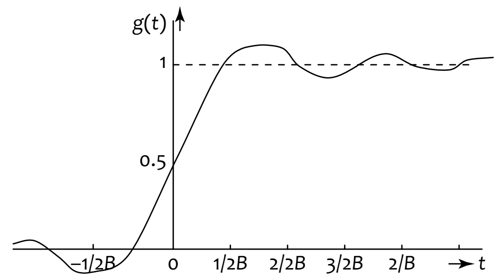


Fig. 3.24 *Response of an ideal LPF to a unit step*

Now, let us find the slope of $g(t)$ at $t = 0$. From Eq. (3.42),

$$\frac{d}{dt} g(t)\bigg|_{t=0} = \frac{d}{dt}\left[\frac{1}{2} + \int_0^{2Bt} \text{sinc } \tau \, d\tau\right]\bigg|_{t=0} = 2B \tag{3.44}$$

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 $\left(\frac{1}{2B}\right)$.

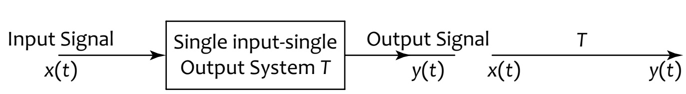
Thus, the time required by $g(t)$ to increase from a value of 0.1 to 0.9, which is the rise time t_r, is given by

$$t_r \cong \frac{0.8}{2B} \cong \frac{0.4}{B}$$

$$\therefore \qquad \boxed{t_r = \frac{0.4}{B} \text{ for an ideal LPF of bandwidth } B \text{ Hz}} \tag{3.45}$$

SUMMARY

1. A system may be defined as an entity which acts on one or more inputs (or excitations) and produces one or more responses.



2. **Continuous-time systems** These are defined as those systems which take continuous-time signals as input and produce continuous-time signals as output.

3. **Discrete-time systems** It takes a discrete-time signals as input and produces another discrete-time signal as output.
4. **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.
5. **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.
6. **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'.
7. **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.
8. **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.
   All physically realizable real-time systems must be causal.
9. **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.
10. **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.
11. ***h*(*t*) of a causal system** For a causal LTI system, $h(t) = 0$ for $t < 0$.
12. **Relation between step response and impulse response**

    If $g(t)$ is unit step response, then $h(t) = \dfrac{dg(t)}{dt}$
13. **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.
14. **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.
15. **BIBO stability theorem** An LTI system T is stable in the BIBO sense iff its impulse response, $h(t)$ is absolutely integrable, i.e., if and only if $\int_{-\infty}^{\infty} |h(t)|\,dt < \infty$.
16. **Eigensignals of an LTI system** An eigensignal 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 eigensignal.

17. **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 eigensignal $\exp(j2\pi ft)$.
18. **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.
19. For any real system, magnitude response will have even symmetry and the phase response will have odd symmetry.
20. **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.
21. **Ideal lowpass filter** For an ideal LPF with passband gain A, passband 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 $H(f) = A\Pi(f/2B)e^{-j2\pi f\tau}$.
22. **Ideal bandpass filter** For an ideal BPF with passband from f_1 to f_2, passband 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

$$h(t) = 2AB\left[\operatorname{sinc} B(t-t_0)\right]\cos 2\pi f_0(t-t_0)$$

23. **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

$$\int_{-\infty}^{\infty} \frac{\left|\log_e |H(f)|\right|}{(1+f^2)} df < \infty$$

24. **Rise-time and bandwidth** (i) For a first-order R-C lowpass filter, Rise time $t_r = \dfrac{0.35}{B}$ where, B is its half-power bandwidth.
    (ii) For an ideal LPF of bandwidth B Hz, the rise time $t_r = \dfrac{0.4}{B}$.

REFERENCES AND SUGGESTED READING

1. Simon Haykin and B Van Veen: *Signals and Systems*, John Wiley & Sons (Asia) Pvt Ltd, Singapore; 2004.
2. R Ziemer; W H Tranter; D Ronald Fannin: *Signals and Systems: Continuous and Discrete*; second edition, Mcmillan Publishing Company, New York, 1990.
3. A Oppenheim; A S Wilsky; S Hamid Nawab: *Signals and Systems*, Pearson Education (Singapore), Delhi, 2004.
4. B P Lathi : *Signals, Systems and Communication*, B.S. Publications, Hyderabad, 2001.
5. P Ramakrishma Rao: *Signals and Systems*: Tata McGraw-Hill Publishing Company Ltd; New Delhi; 2008.

REVIEW QUESTIONS

1. Define the terms: 'static system' and 'dynamic system'. Give one example for each of these.
2. How do you define 'linearity property' of a system?

3. The input-output relationship of a system is as shown in Fig. RQ-1.
   (a) Is this system linear? Justify your answer.
   (b) Is this system static or dynamic? Why?

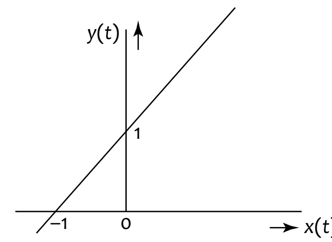
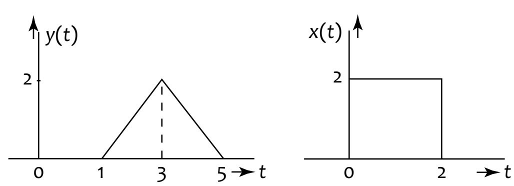


Fig. RQ-1

4. The impulse response of a certain system is $h(t) = A\Pi\left(\dfrac{t - T/2}{T}\right)$.
   Is this system static or dynamic? Why? What is the input–output relationship for the system?
5. Define the 'time-invariance' property of a continuous-time system? Give examples of a time-varying system.
6. Define the terms: (a) Impulse response (b) Causality, as applied to systems.
7. What is the relationship between 'impulse response' and 'step response' of an LTI system?
8. What is the condition on the impulse response of an LTI system for the system to be stable in the BIBO sense?
9. Define the terms: 'eigensignal' and 'transfer function' of a stable LTI system.
10. State the two conditions required to be satisfied by an LTI system for an input signal to pass through it without any distortion?
11. Why are the ideal LPF, HPF and BPF not physically realizable?

PROBLEMS

1. 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) $y(t) = \dfrac{d}{dt}x(t)$
   (h) $y(t) = 5x(2t+3)$ (i) $y(t) = \log_e |x(t)|$
2. Determine whether the following systems are static or dynamic, linear or non-linear, time-invariant or time-varying, causal or non-causal.
   (a) $2\dfrac{dy(t)}{dt} + 4y(t) + 6 = 3x(t)$ (b) $3\dfrac{dy(t)}{dt} + 10y(t) = 2x(t)$
   (c) $2\dfrac{dy(t)}{dt} + ty(t) = 4x(t)$ (d) $5\dfrac{dy(t)}{dt} + 2y^2(t) = 3x(t)$

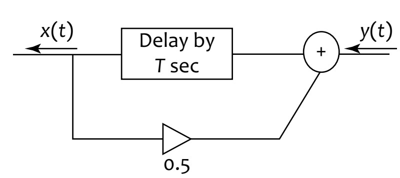
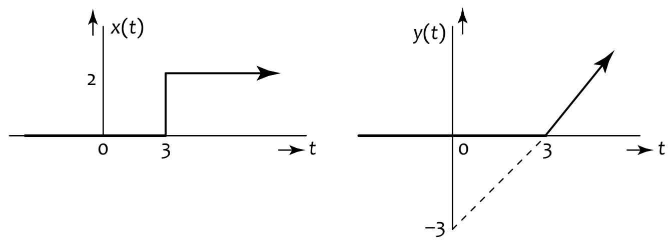


Fig. P-3.1

3. Show that an ideal integrator is an LTI system.
4. Obtain the impulse response and the transfer function $H(f)$ of the following system. Plot its magnitude response.
5. Find the impulse responses of the systems given in Problems 2(a) and 2(b).
6. The input $x(t)$ and the corresponding output $y(t)$ of a causal LTI system T are as shown in Fig. P-3.2. Find the impulse response function $h(t)$ of the system.



(a) Input x(t)

(b) Output y(t)

Fig. P-3.2

7. 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 $x(t) = \cos(100\pi t)u(t)$.
8. Show that a sinusoid/cosinusoid of any frequency is an eigensignal of any LTI system.
9. 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.
10. Find the frequency response, $H(f)$, of the system described by

$$\frac{d^3y(t)}{dt^3} + 0.5\frac{d^2y(t)}{dt^2} + 0.75\frac{dy(t)}{dt} + 2y(t) = x(t)$$

11. For the system shown in Fig. P-3.3, find the frequency response.
12. The impulse response of a system is $h(t) = 10e^{-3t}u(t)$. Find and plot the response of the system to an input $x(t) = \Pi\left(\frac{t-1}{2}\right)$.
13. Complete the derivation leading to the result given by Eq. (3.21).
14. Complete the derivation leading to the result given by Eq. (3.25).
15. Complete the derivation leading to the result given by Eq. (3.27).

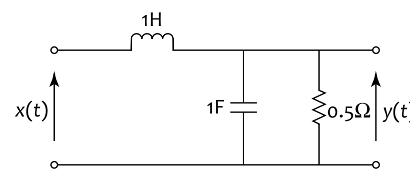


Fig. P-3.3

OBJECTIVE TYPE QUESTIONS

Fill in the blanks

1. Dynamic systems are systems having ____________________.
2. Electrical systems made up of purely resistive networks are ____________.
3. The input–output relationship of static systems of the continuous-time type is described by ____________________ equations.
4. For continuous-time type dynamic systems, the input-output relationship is described by __________________ equations.
5. A continuous-time type dynamic systems is said to be in ‘ground state’ or ‘at rest’, if ______________________.
6. A system is said to be linear if it satisfies the __________ and _________ properties.
7. For a causal system, the value of the present output depends only on the _________ and _________ inputs.
8. All physically realizable systems must be ____________.
9. A continuous-time LTI system is completely characterized by its ________________.
10. Two LTI systems with impulse responses $h_1(t)$ and $h_2(t)$ are connected in cascade. Impulse response of the overall system is ________________.

11. Two LTI systems with impulse responses $h_1(t)$ and $h_2(t)$ are connected in parallel. Impulse response of the overall system is ________________.
12. The relation between the step-response $g(t)$ and the impulse response $h(t)$ of an LTI system is given by $h(t) =$ ____________________.
13. For a causal CT system, $h(t)$ should satisfy the condition ___________________.
14. A CT signal $x(t)$ is said to be a bounded signal if __________________.
15. The BI-BO criterion for stability is ________________.
16. For an LTI system, the condition on $h(t)$ for the system to be stable in the BI-BO sense, is ________________.
17. The meaning of 'eigensignal of a system', is _________________.
18. _______________, ______________ are the eigensignals of LTI systems.
19. Frequency response function of an LTI system is the _____________ of its impulse response.
20. When the input signal to an LTI system is an eigensignal at a frequency ω_0, the ratio of the output signal to input signal gives _________________.
21. The conditions for distortionless transmission of signals through an LTI system, are (i) ______________ (ii) _______________.
22. For the phase response of an LTI system with an impulse response $h(t)$, to be linear, $h(t)$ should be symmetrical about ______________, where ______________ is the _____________ introduced by the system. Also $h(t)$ has to have a maximum value at _____________ .
23. The impulse response of an ideal LPF is a __________________ function.
24. Paley–Wiener criterion for physical realizability of a square-integrable magnitude response $|H(f)|$ is _______________.
25. The rise time depends on __________ of the system and is __________ proportional to it.

MULTIPLE CHOICE QUESTIONS

1. (Choose the incorrect answer). A system composed of purely resistive networks is
   (a) dynamic (b) linear (c) time-invariant (d) static
2. The system with $y(t) = x(3t)$ is
   (a) static (b) linear (c) fixed (d) causal
3. The LTI system with $h(t) = e^{-t}, -\infty < t < \infty$, is
   (a) causal and stable (b) causal and unstable
   (c) non-causal and stable (d) non-causal and unstable
4. 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.
5. 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) $h_1(t)\cdot h_2(t)$
6. When two LTI systems with impulses 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) $h_1(t)\cdot h_2(t)$

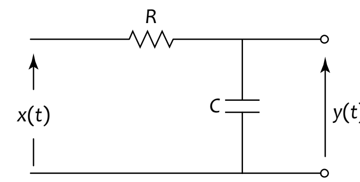
7. $x_1(t) = A[u(t) - u(t-10)]$ and $x_2(t) = B[u(t) - u(t-5)]$. $x_1(t) * x_2(t)$ is a
(a) triangular pulse (b) rectangular pulse
(c) trapezoidal pulse (d) sinc pulse

8. $x_1(t) = A[u(t) - u(t-5)]$ 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

9. The transfer function $H(f)$ of the *R-C* lowpass filter shown in the figure, is given by

(a) $\dfrac{1}{1 + j2\pi fRC}$ (b) $\dfrac{1/RC}{1 + j2\pi fRC}$

(c) $\dfrac{1}{1 - j2\pi fRC}$ (d) $j2\pi fRC$



10. The 3-db cutoff frequency for the filter of MCQ 9 is

(a) $\dfrac{1}{RC}$ (b) RC (c) $\dfrac{1}{2\pi RC}$ (d) none of the above

11. 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 the above

12. If * denotes convolution operation and overbar 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 the above

13. 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

14. 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.

15. Signal transmission through an LTI system cannot be distortion less 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

16. 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.
17. 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 $t = \tau$.
18. 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 the above
19. 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) $2AB\,\text{sinc}B(t-\tau)$
(c) $AB\,\text{sinc}\,2B(t-\tau)$
(d) $B\,\text{sinc}\,2B(t-\tau)$
20. For a first-order *R-C* lowpass 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) $\frac{3.5}{B}$
21. 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. (a)	2. (b)	3. (d)	4. (b)	5. (c)	6. (a)
7. (c)	8. (a)	9. (a)	10. (c)	11. (a)	12. (b)
13. (a)	14. (a)	15. (b)	16. (b)	17. (c)	18. (b)
19. (c)	20. (a)	21. (d)			

4 Amplitude Modulation

After going through this chapter, the student will be

- *able to form a clear idea of the meaning of 'modulation' and the need for it*
- *able to give the time-domain representation, spectrum, and the methods of generation and detection of AM signals*
- *familiar with the operation of an envelope detector, the types of distortions it can give rise to and the reason behind each type of distortion*
- *able to give the time and frequency-domain representation as well as the methods of generation and coherent detection of DSB-SC signals, and SSB-SC and VSB signals*

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 speech, music or images, etc. These signals cannot be transmitted directly and need some preprocessing. This preprocessing needed for making them suitable for transmission, is called *modulation*. The need for modulation arises because of several reasons.

4.1.1 Need for Modulation

(i) Antenna Size Long-distance communication takes place 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 be 3 km, which is definitely not practical. 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 referred to as 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.

(ii) 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 the desired speech signal only 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?

(iii) Multiplexing Multiplexing is the technique used for transmitting several information-bearing signals simultaneously over the same physical channel. The modulation process makes it possible to multiplex several message signals and transmit them simultaneously by using different carrier frequencies for the various message signals.

(iv) Hardware Problem The modulation process enables us to avoid many hardware problems which would be encountered if there were to be no modulation.

Having seen the need for modulation, we shall now try to clearly understand what this modulation process consists of. The modulation process consists of varying any one of the parameters of a high-frequency sinusoidal signal, called the carrier signal, in accordance with the variations in the amplitude of the 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: 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 amplitude of the message signal, the modulation is called *amplitude modulation*, *frequency modulation*, or *phase modulation*. Since the rate of change of the phase represents frequency, phase modulation and frequency modulation are closely related and are together called 'angle modulation'.

AMPLITUDE MODULATION 4.2

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 types 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

$$x_c(t) = A\cos(\omega_c t + \theta)$$

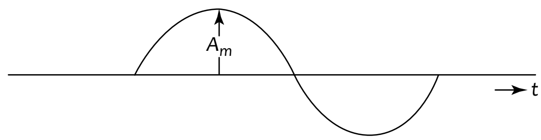
where
$$A = A_c + A_m x(t)$$

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., $|x(t)| \le 1$.

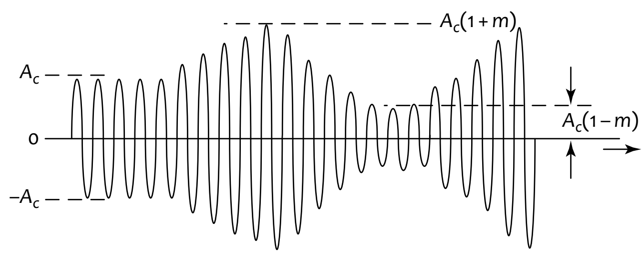
Let
$$m \triangleq \frac{A_m}{A_c}$$

Then,
$$A = A_c + A_m x(t) = A_c\left[1 + mx(t)\right]$$

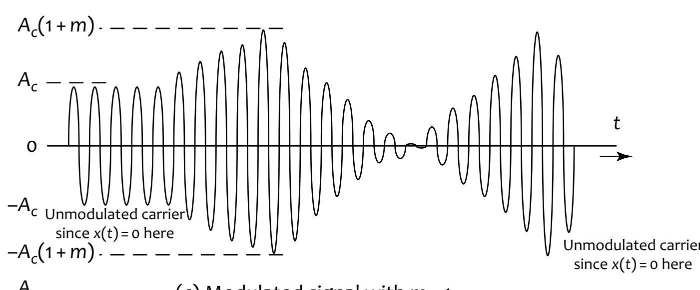
Hence,
$$x_c(t) = A_c\left[1 + mx(t)\right]\cos(\omega_c t + \theta)$$



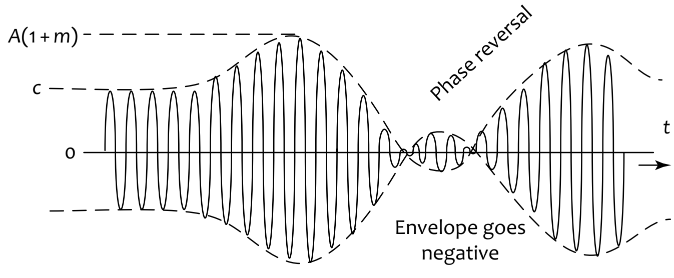
(a) Modulating signal $x(t) = \sin \omega_m t$



(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*

Without loss of generality, we may take $\theta = 0$ so that

$$\boxed{x_c(t) = A_c\left[1 + mx(t)\right]\cos\omega_c t} \tag{4.2}$$

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 signal 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

$$x_c(t) = A_c\left[1 + m\cos\omega_m t\right]\cos\omega_c t$$

$$= A_c\cos\omega_c t + mA_c\cos\omega_c t\cdot\cos\omega_m t$$

$$\boxed{\therefore \quad x_c(t) = A_c\cos\omega_c t + \frac{1}{2}mA_c\cos(\omega_c + \omega_m)t + \frac{1}{2}mA_c\cos(\omega_c - \omega_m)t} \tag{4.4}$$

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.

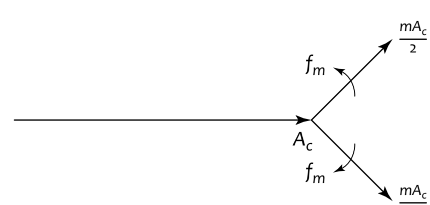


Fig. 4.2 *Phasor diagram of a single-tone modulated AM signal*

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

$$\text{Power in the carrier component} = \frac{1}{2}A_c^2 = \text{say } P_c$$

$$\text{Power in the upper side-frequency component} = \frac{1}{8}m^2 A_c^2$$

$$\text{Power in the lower side-frequency component} = \frac{1}{8}m^2 A_c^2$$

$$\boxed{\therefore \quad \text{total power in the AM signal} = P_T = P_c\left[1+\frac{m^2}{2}\right]} \tag{4.5}$$

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, this part of the 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 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 5 V peak amplitude and 100 kHz frequency is amplitude modulated by a 5 kHz signal of peak amplitude of 3 V. What is the modulation index? Draw the two-sided spectrum of the modulated signal.

$$\begin{aligned} x_c(t) &= A_c\left[1+m\cos\omega_m t\right]\cos\omega_c t \\ &= 5\left[1+mx(t)\right]\cos\omega_c t \\ &= 5\cos\omega_c t + m5x(t)\cos\omega_c t \end{aligned}$$

$\therefore$ $A_c = 5$ and $5m = 3$ since $|x(t)| \le 1$ $\quad\therefore$ $m = 3/5 = 0.6$

The side-frequencies are

$$(-f_c - f_m), (-f_c + f_m); (f_c - f_m) \text{ and } (f_c + f_m)$$

i.e., –105 kHz, –95 kHz; 95 kHz and 105 kHz.

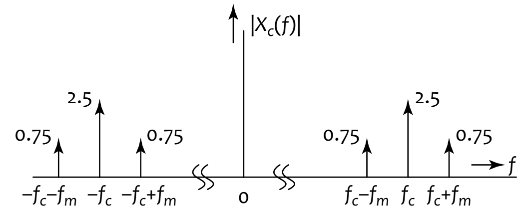


Fig. 4.3 *Spectrum of AM signal of Example 4.1*

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.

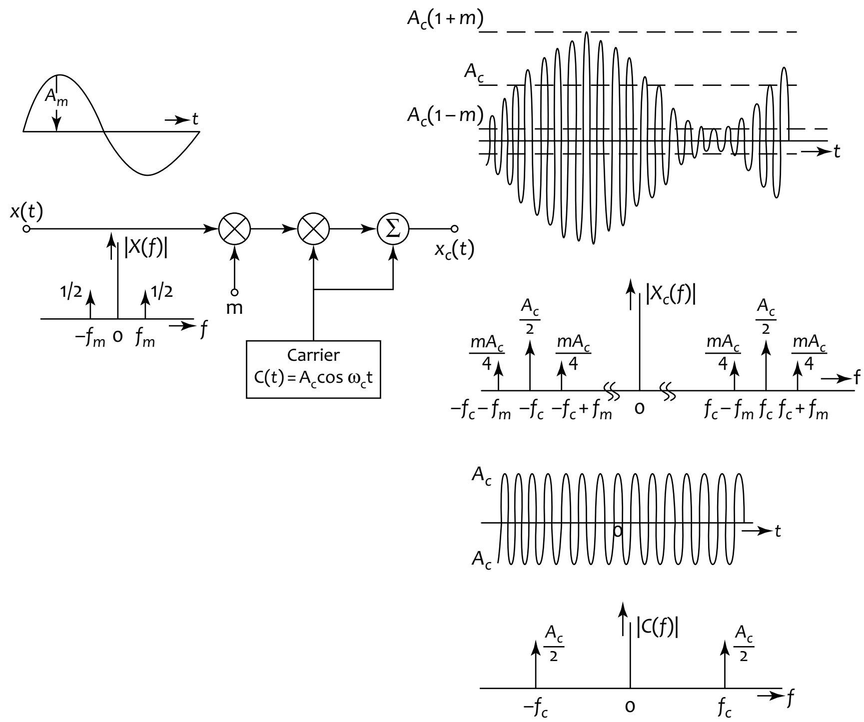


Fig. 4.4 *Amplitude modulation: Waveforms and Spectra of message, carrier and modulated original*

Example 4.2

A carrier wave of 10 MHz frequency and peak value of 10 V is amplitude modulated by a 5 kHz sine-wave of 6 V amplitude. Determine the modulation index and draw the one-sided spectrum of the modulated wave. *(University Question)*

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

$$x_c(t) = A_c\left[1 + m\cos\omega_m t\right]\cos\omega_c t$$

where, $A_c = 10\,\text{V}$, $m = 0.6$, $\omega_m = 2\pi f_m = 2\pi \times 5 \times 10^3$ rad/s and

$\omega_m = 2\pi f_c = 2\pi \times 10 \times 10^6$ rad/s

$x_c(t)$ may therefore be written in an expanded form as

$$x_c(t) = A_c \cos\omega_c t + mA_c(\cos\omega_c t . \cos\omega_m t)$$

$$= A_c \cos\omega_c t + \frac{1}{2} mA_c \cos(\omega_c + \omega_m)t + \frac{1}{2} mA_c \cos(\omega_c - \omega_m)t$$

Thus, $x_c(t)$ is made up of 3 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 $0.3 \times 10 = 3$ 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.

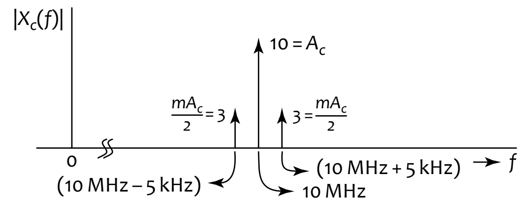


Fig. 4.5 *One-sided spectrum of AM signal of Example 4.2*

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 side-bands—an upper side-band (USB) extending from f_c to $(f_c + f_m)$ Hz and a lower side-band (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.

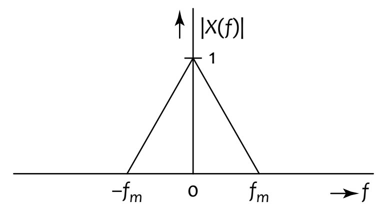


Fig. 4.6 *Amplitude spectrum of the message signal x(t)*

From Eq. (4.2), we have

$$x_c(t) = A_c\left[1 + mx(t)\right]\cos\omega_c t$$

$$= A_c\cos\omega_c t + mA_c x(t)\cos\omega_c t \tag{4.7}$$

Since $$\cos\omega_c t \xleftrightarrow{F.T} \frac{1}{2}\left[\delta(f+f_c) + \delta(f-f_c)\right],$$

taking the FT of Eq. (4.7) on both sides,

$$X_c(f) = \frac{A_c}{2}\left[\delta(f+f_c) + \delta(f-f_c)\right] + \frac{mA_c}{2}\left[X(f-f_c) + X(f+f_c)\right] \tag{4.8}$$

Here, we have made use of the FT pair

$$x(t)\cos\omega_c t \xleftrightarrow{F.T} \frac{1}{2}\left[X(f-f_c) + X(f+f_c)\right]$$

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].

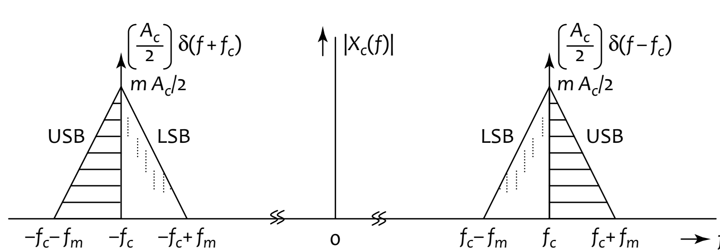


Fig. 4.7 *Amplitude spectrum of an AM signal*

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 employing AM 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.

Upper side-frequency $= f_c + f_m = 5005\,\text{kHz}$

Lower side-frequency $= f_c - f_m = 4995\,\text{kHz}$

Adding these two, $2f_c = 10{,}000$ kHz $\quad \therefore\ f_c = 5000 \text{ kHz} = 5 \text{ MHz}$

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-wave band?

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 $= 2 \times 5$ 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

$$\langle x_c^2(t)\rangle = \frac{1}{2}\left[\left\langle A_c^2 \{1 + mx(t)\}^2 \right\rangle\right]$$

where, the symbol $<z>$ is used to represent the average value of z.

$$\therefore \qquad \langle x_c^2(t)\rangle = \frac{1}{2}\left[A_c^2 + 2m\langle x(t)\rangle A_c^2 + m^2 A_c^2 \langle x(t)\rangle\right]$$

Assuming $\langle x(t)\rangle = 0$, 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,

$$\langle x_c^2(t)\rangle = \text{ Average power of } x_c(t) = \frac{1}{2}\left[A_c^2 + m^2 A_c^2 \langle x^2(t)\rangle\right] \tag{4.10}$$

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 $x(t)$ has been 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 RHS 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.

Disadvantages of AM

(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 an unmodulated carrier alone is transmitted, the antenna current is 9 amperes. When sinusoidal modulation is present, the antenna current is found to be 11 amperes. What is the percentage of modulation used?

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?

$$P_T = 100 \times 10^3 = P_c \left[1 + \frac{0.64}{2}\right]$$

$$\therefore \quad P_c = \frac{100 \times 10^3}{1.32} = 75.8 \text{ kW}$$

$\therefore$ 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?

$$P_{T_1} = P_c \left[1 + \frac{0.64}{2}\right] = 1.32 P_c = 132 \text{ kW}$$

$$\therefore \quad P_c = 100 \text{ kW}$$

Let the modulation index of the second sinusoid be m.

$$\therefore \quad P_{T_2} = 150 \times 10^3 = 100 \times 10^3 \left[1 + 0.32 + \frac{m^2}{2}\right]$$

$$\therefore \quad 50 \times 10^3 = 100 \times 10^3 \left[0.32 + \frac{m^2}{2}\right]$$

$$\therefore \quad (50 - 32) \times 10^3 = m^2 \times 50 \times 10^3$$

or
$$m^2 = \left(\frac{18}{50}\right) = 0.36 \quad \therefore m = 0.6$$

Example 4.8

Determine the overall percentage of modulation in the above example when both the sinusoidal audio signals are simultaneously modulating the carrier.

$$m = \sqrt{m_1^2 + m_2^2} = \sqrt{0.64 + 0.36} = 1$$

$\therefore$ 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)

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.

$\therefore$ carrier component of power in the AM signal $= P_c = 2/3$ kW.

Total sideband power in the AM signal (with 100% modulation) $= \left(1-\frac{2}{3}\right)\text{kW} = \frac{1}{3}\text{kW}$

$\therefore$ power in each sideband $= \frac{1}{6}\text{kW}$

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?

It is given that the unmodulated carrier power $= P_c = 10$ kW

Max. depth of modulation without overloading $= 40\%$ $\therefore$ $m_1 = 0.4$

Total power in the AM signal $= P_T = P_C\left[1+\frac{m_1^2}{2}\right] = 10^4 \times 1.08 = 10.8$ kW

$\therefore$ 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 maximum unmodulated carrier power that would make the total power in the AM signal reach the value 10.8 kW.

$$\therefore \qquad P_T = 10.8\times 10^3 = P_C^1\left[1+\frac{(0.3)^2}{2}\right] = 1.045P_C^1$$

$$\therefore \qquad P_C^1 = \frac{10.8\times 10^3}{1.045} = 10.33\text{ kW}$$

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 (i) 100%, and (ii) 50%.

(JNTU, May 2007)

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

$$P_T' = P_c \frac{m^2}{4}$$

$$\therefore \ \% \text{ saving in power} = \frac{100(P_T - P_T')}{P_T}$$

$$= \frac{100\left\{P_c\left[1+\frac{m^2}{2}\right] - P_c\frac{m^2}{4}\right\}}{P_c\left[1+\frac{m^2}{2}\right]} = \frac{100\left[1+\frac{m^2}{4}\right]}{\left[1+\frac{m^2}{2}\right]}$$

(i) When $m = 1$ corresponding to 100% modulation,

$$\% \text{ saving in power } = 100\left[\frac{1+1/4}{1+1/2}\right] = 83.3\%$$

(ii) 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.

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

$= P_c\left\lfloor 1+m^2\overline{x^2}\right\rfloor$ 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 $0 \le m \le 1$.

$$\eta = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} = \frac{1}{1+\frac{1}{m^2\overline{x^2}}} \quad \text{(since } m^2\overline{x^2} = 0 \text{ is ruled out.)}$$

$$\therefore \qquad \eta_{\max} = \frac{1}{1+\left(\frac{1}{m^2\overline{x^2}}\right)_{\min}}$$

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.

In the following discussion, we have 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.

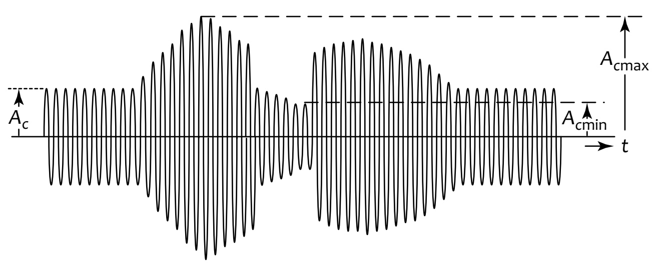


Fig. 4.8 *Modulated signal when the modulating signal is some complex waveform*

In a case like this, we define two indices of modulation:

(i) The positive peak modulation index $\triangleq \dfrac{A_{c_{\max}} - A_c}{A_c}$

(ii) The negative peak modulation index $\triangleq \dfrac{A_c - A_{c_{\min}}}{A_c}$

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.

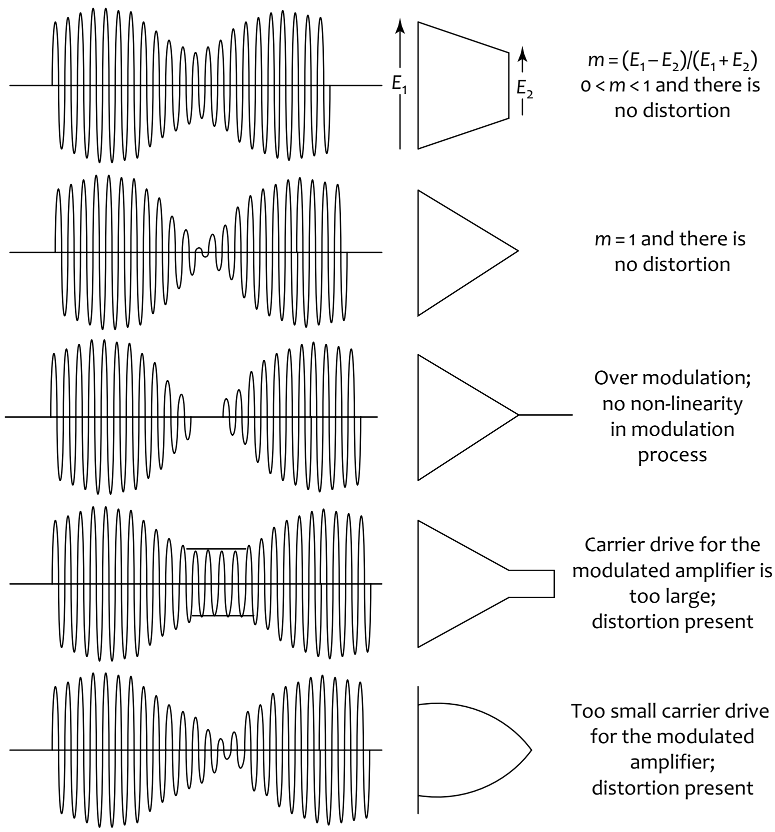


Fig. 4.9 *Trapezoidal patterns under different conditions*

Example 4.13

An AM (double sideband plus full carrier) signal waveform is as shown in Fig.4.10. Determine modulation index m. Write down the expression for the modulated signal, the total power, carrier power and sideband power.

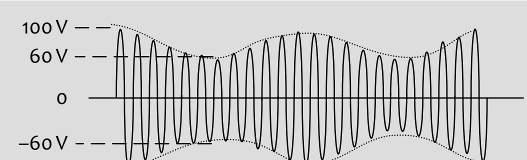


Fig. 4.10 *Signal for Example 4.13*

(i) $A_c[1+m] = 100 \qquad \therefore 2A_c = 160$

$A_c[1-m] = 60$

$$\therefore A_m = 100 - A_c = 20 \quad \therefore m = \frac{A_m}{A_c} = 0.25$$

(ii) $x_c(t) = A_c\left[1+mx(t)\right]\cos\omega_c t$

$$= 80\left[1+0.25\cos\omega_m t\right]\cos\omega_c t$$

(iii) $P_c = \dfrac{1}{2}A_c^2 = \dfrac{1}{2}\times 6400 = 3200$ watts

$$\therefore\ P_T = 3200\left[1+\left(\frac{1}{4}\right)^2/2\right] = 3200+\frac{3200}{32} = 3300\text{ W}$$

$\therefore$ 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 10 V peak voltage. Calculate the modulation index and the ratio of the side-lengths (L_1/L_2) of the corresponding trapezoidal pattern. *(University Question)*

A sketch of the modulated signal is shown in Fig. 4.11.

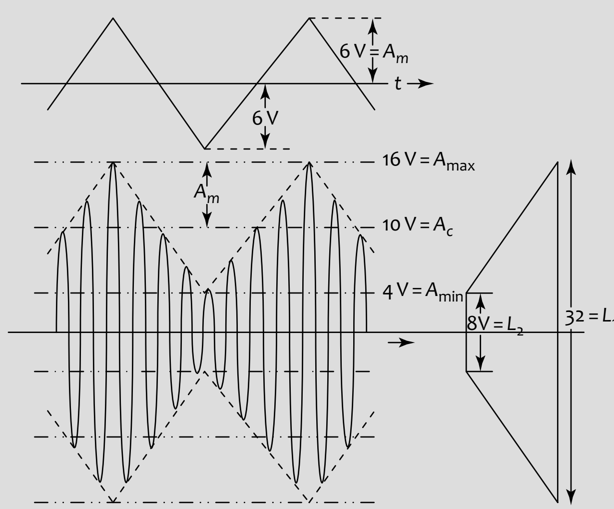


Fig. 4.11 *Modulated Signal*

$$\text{Modulation index } m = \frac{A_{max}-A_{min}}{A_{max}+A_{min}} = \frac{16-4}{16+4} = \frac{12}{20} = 0.6$$

$$L_1 = 2(A_c + A_m) = 16\times 2 = 32$$

$$L_2 = 2(A_c - A_m) = 2(10-6) = 8$$

$$\therefore \qquad \frac{L_1}{L_2} = \frac{32}{8} = 4$$

4.2.7 Generation of AM

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:

(i) Modulators using non-linear devices
(ii) Modulators using product devices
(iii) 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

$$e_{out} = a_0 + a_1 e_{in} + a_2 e_{in}^2, \tag{4.13}$$

where, the constants a_0, a_1, a_2 depend on the shape of the input–output characteristic of the device. Suppose we make

$$e_{in}(t) = x(t) + E_c \cos \omega_c t \tag{4.14}$$

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 $e_{in}(t)$ in Eq. (4.13) using Eq. (4.14), we get

$$e_{out}(t) = \left(a_0 + \frac{E_c^2}{2}\right) + \left[a_1 x(t) + a_2 x^2(t)\right] + \frac{E_c^2}{2}\cos 2\omega_c t + a_1 E_c \left[1 + \left(\frac{2a_2}{a_1}\right) x(t)\right] \cos \omega_c t$$

In the above, the first term is a dc term which can always be suppressed by using a coupling capacitor. The second term $a_1 x(t) + a_2 x^2(t)$ 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 $a_1 E_c \left[1 + \left(\frac{2a_2}{a_1}\right) x(t)\right] \cos \omega_c t$ 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 bandpass 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.

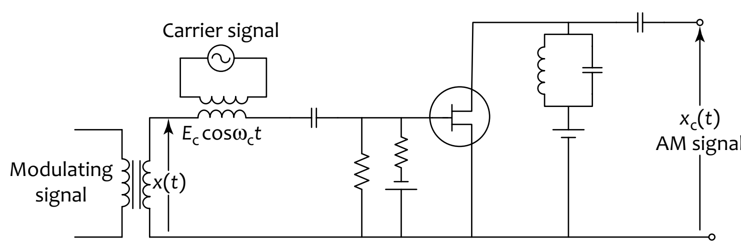


Fig. 4.12 *A square-law amplitude modulator*

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 $e_{out}(t)$.

$$E_{\text{out}}(f) = \left(a_0 + \frac{E_c^2}{2}\right)\delta(f) + a_1 X(f) + a_2\left[X(f) * X(f)\right] + \frac{E_c^2}{4}\left[\delta(f+2f_c) + \delta(f-2f_c)\right]$$

$$+\frac{a_1 E_c}{4}\left[\delta(f+f_c) + \delta(f-f_c)\right] + a_2 E_c\left[X(f+f_c) + X(f-f_c)\right] \tag{4.15}$$

A sketch of $E_{\text{out}}(f)$ is shown in Fig. 4.13.

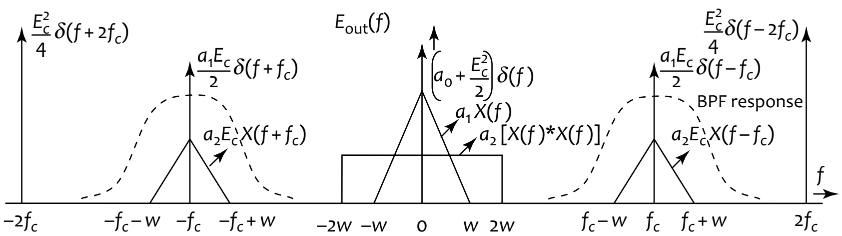


Fig. 4.13 *Spectrum of $e_{\text{out}}(t)$ of the square-law modulator*

4.2.9 Modulators Using Product Devices

These are based on Eq. (4.2) which states that an amplitude-modulated signal is given by

$$x_c(t) = A_c\left[1 + mx(t)\right]\cos\omega_c t$$

$$= A_c\cos\omega_c t + mA_c x(t)\cos\omega_c t$$

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 adder part of Fig. 4.14 may of course be realized using an op-amp.

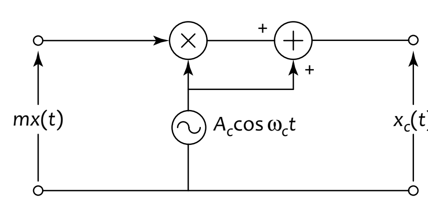


Fig. 4.14 *Block diagram of a product modulator*

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 If we assume that

(i) the forward resistance of the diode is extremely small compared to R_L

and

(ii) $|x(t)| \le 1$ and $A_c >> 1$

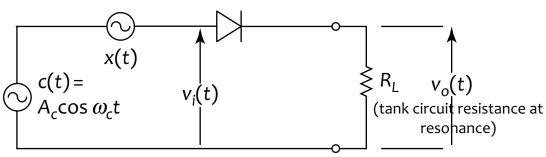


Fig. 4.15 *A switching modulator using a diode*

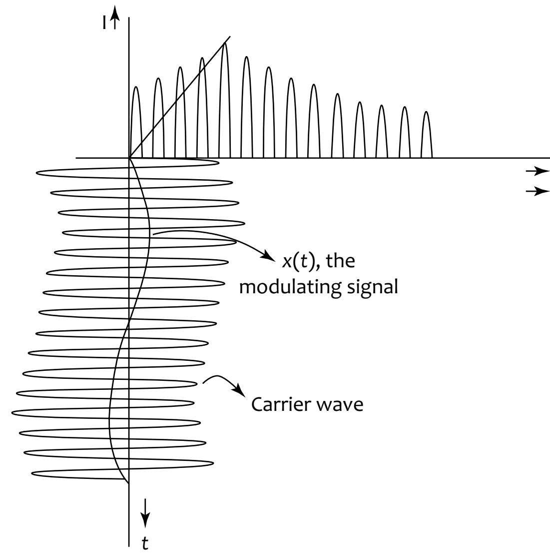


Fig. 4.16 *Diode switching modulator working principle*

then we may state that

$$v_0(t) = \begin{cases} v_i(t) \ \text{ whenever } \ A_c \cos \omega_c t > 0 \\ 0 \ \text{ otherwise } \ (A_c \cos \omega_c t < 0) \end{cases} \tag{4.16}$$

Since

$$v_i(t) = x(t) + A_c \cos \omega_c t \tag{4.17}$$

it means that

$$v_0(t) = v_i(t)g(t) = \left[x(t) + A_c \cos \omega_c t\right] g(t) \tag{4.18}$$

where, $g(t)$ is a gate waveform with a period $T_0 = (1/f_c)$ as shown in Fig. 4.17.

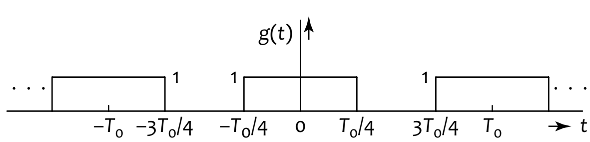


Fig. 4.17 *The periodic gate waveform g(t)*

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,

$$a_n = \frac{2}{T_0}\int_{-T_0/2}^{+T_0/2} x(t)\cos n\omega_c t dt = \frac{4}{T_0}\int_0^{T_0/2} x(t)\cos n\omega_c t dt$$

$$= \frac{4}{T_0}\int_0^{T_0/4} \cos n\omega_c t\ dt = \frac{4}{T_0}\left(\frac{1}{n\omega_c}\right)\sin n\omega_c t\Big|_0^{T_0/4}$$

$$= \frac{2}{n\pi}\left[\sin\left(\frac{n\pi}{2}\right)\right] = \begin{cases} 0 \text{ if } n \text{ is even} \\ \dfrac{2}{n\pi}(-1)^{\frac{n-1}{2}} \text{ if } n \text{ is odd} \end{cases}$$

$\therefore$
$$g(t) = \frac{1}{2} + \sum_{n=1}^{\infty}\left[\frac{2}{n\pi}\sin\left(\frac{n\pi}{2}\right)\right]\cos n\omega_c t$$

$\therefore$
$$g(t) = \frac{1}{2} + \frac{2}{\pi}\cos\omega_c t - \frac{2}{3\pi}\cos 3\omega_c t + \frac{2}{5\pi}\cos 5\omega_c t - \frac{2}{7\pi}\cos 7\omega_c t + ... \tag{4.19}$$

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 bandlimiting 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.

The base bias supply V_{BB} reverse biases the base-emitter junction beyond cutoff 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.

There will be one current pulse for each r.f. 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 r.f. 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 r.f. cycle (i.e., carrier cycle) will, however, be varying from one r.f. cycle to the next. This component of current is marked in Fig. 4.19 as $i_c(t)$.

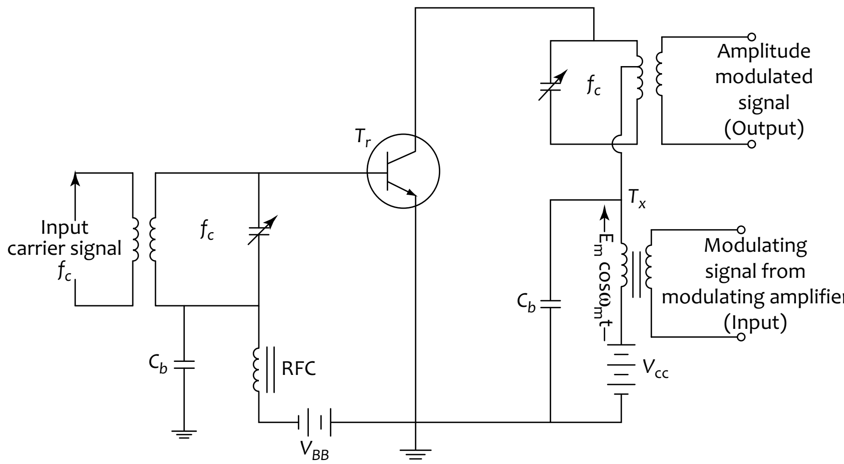


Fig. 4.18 *A collector-modulated class-C amplifier*

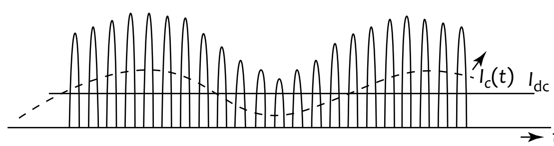


Fig. 4.19 *Collector current pulses in a class-C collector modulated amplifier*

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. 4.15.

If, $$m \triangleq \frac{E_m}{V_{CC}} = \text{modulation index}, \tag{4.23}$$

$$i_c(t) = I_{DC}(1 + m \cos_m t) \tag{4.24}$$

$P_T(t) \triangleq$ Total power input into the collector circuit (averaged over an r.f. cycle)

$$= V_{CC}\left[1 + m \cos \omega_m t\right] \; i_c(t) \tag{4.25}$$

But $$i_c(t) = I_{DC}(1 + m \cos_m t) \quad \text{from Eq. (4.24)}$$

Thus,

$$P_T(t) = V_{CC}\left[1 + m \cos \omega_m t\right] \; I_{dc}\left[1 + m \cos \omega_m t\right] \tag{4.26}$$

We define

$$P_B \triangleq V_{CC} \cdot I_{DC} \tag{4.27}$$

Then P_B represents the dc power supplied by the V_{CC} supply to the collector circuit.

$$\therefore \quad P_T(t) = P_B\left[1 + 2m \cos \omega_m t + m^2 \; \cos^2 \omega_m t\right]$$

$$\therefore \quad P_T(t) = P_B\left[1 + \frac{1}{2}m^2 \cos 2\omega_m t + 2m \cos \omega_m t + \frac{1}{2}m^2\right] \tag{4.28}$$

When we average $P_T(t)$ over a modulating signal cycle, the second and third terms on the RHS of Eq. (4.28) vanish.

$$\therefore \qquad P_{T\text{av}} = P_B\left[1+\frac{m^2}{2}\right] \tag{4.29}$$

$P_{T\text{av}}$ represents the total average power supplied to the collector circuit. Of this, $P_B = V_{CC}.\ I_{DC}$ 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

$\eta P_{T\text{av}}$ = Total average power in the amplitude modulated output signal = P_0

$$P_0 = \eta P_{T\text{av}} = \eta P_B + \eta P_B\frac{m^2}{2} \tag{4.30}$$

ηP_B = Carrier component of $P_0 = P_c$

$\eta P_B\dfrac{m^2}{2} = P_C\dfrac{m^2}{2}$ = 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 $A_{c\ max} = 75$, and $A_{c\ min} = 15$, 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%, and (ii) 90%.

(a) $m = \dfrac{A_{c\max} - A_{c\min}}{A_{c\max} + A_{c\min}} = \dfrac{75-15}{75+15} = \dfrac{60}{90} = 66.7\%$

(b) $P_T = P_c\left[1+\dfrac{m^2}{2}\right]$. Here $P_c = \dfrac{A_c^2}{2}$ but $A_c = 15 + \dfrac{75-15}{2} = 45$

$$\therefore \qquad P_c = \frac{45^2}{2} = \frac{2025}{2} = 1012.5 \text{ W}$$

$\therefore$ sideband power (total) $= P_c \cdot \dfrac{m^2}{2} = \dfrac{45^2}{2} \times \dfrac{4}{9} \times \dfrac{1}{2} = 225$ W

(c) $x_c(t) = 45\left[1+\dfrac{2}{3}\cos\omega_m t\right]\cos\omega_c t + A\cos\omega_c t$

$$= (45+A)\left[1+\left(\frac{30}{(45+A)}\right)\cos\omega_m t\right]\cos\omega_c t$$

(i) $\dfrac{30}{45+A} = 0.5 \quad \therefore 60 = 45 + A \quad \therefore A = 15$

$\therefore$ carrier to be added = $15\angle 0°$

(ii) $\dfrac{30}{45+A} = 0.9 \quad \therefore A = -11.67$

$\therefore$ 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%. (i) What is the carrier power in the output modulated wave? (ii) What will be the class-C power amplifier collector dissipation for 100% modulation? (iii) What should be the modulating amplifier transistor rating in watts for this depth of modulation? (iv) What is the overall efficiency of the circuit (including class-C and class-A power amplifiers)?

Let dissipation in the transistor (class-C power amplifier) be P_d.

At $$m = 0.6: \quad P_d = P_{in} - P_0 = P_{\text{in}}[1-\eta] = 0.4P_{\text{in}} = 24$$

$\therefore$ $$P_{in} = \frac{24}{0.4} = 60 \text{ W}$$

But $$P_{in} = P_c\left[1 + \frac{0.6^2}{2}\right] = 1.18P_c = 60$$

$\therefore$ $$P_c = \frac{60}{1.18} = 50.85 \text{ W}$$

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 $= 50.85 \times 0.6 = 30.5$ W

$\therefore$ answer for part (i) is 30.5 W.

(ii) $m = 1$

$$P_0 = 30.5(1 + 0.5) = 45.75\text{W} = \text{Total output power with } m = 1.$$

$\therefore$ the corresponding input power $= P_{\text{in}} = 45.75/0.6 = 76.25$ W

$\therefore$ $\underset{(m=1)}{P_d} = 76.25 - 45.75 = 30.5$ W

(iii) $m = 1$ 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

$$P_{SB} = P_c \times 0.5 = 50.85 \times 0.5 = 25.425 \text{ W}$$

(iv) Rating of the transistor used for the class-A modulating amplifier:
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.

$\therefore$ $$P_{d\max} = P_{\text{in}} = \frac{25.425}{0.2} = 127.125 \text{ W}$$

(v) Overall efficiency at $m = 0.6$:

$$\underset{m=0.6}{\eta} = \frac{\text{Total output power with } m = 0.6}{\text{Total input power (for the class-C and class-A amplifiers) with } m = 0.6}$$

$$= \frac{36}{50.85 + 127.125} = \frac{36}{180} = 0.2$$

$\therefore$ 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.

From Eq. (4.21), we have

$$m = \frac{4}{A_c \pi}$$

$$A_c = \frac{4}{m\pi} = \frac{4}{0.85\pi} = 1.498 \text{ volts}$$

Example 4.18

A collector-modulated class-C power amplifier is giving an amplitude-modulated signal of 100 W average power 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.

$$P_0 = P_c\left(1 + \frac{m^2}{2}\right) = 100 \text{ W} \quad \therefore P_c = \frac{100}{1.32} = 75.75 \text{ W}$$

$\therefore$ 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, Then

$$P_D = P_{\text{in}} - P_0 = P_0\left(\frac{1}{\eta}\right) - P_0 = P_0\left(\frac{1-\eta}{\eta}\right) = 100 \times \frac{0.2}{0.8} = 25 \text{ W}$$

4.2.11 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* and 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 watts; 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.

To summarize, the advantages and disadvantages of these two types of modulation are the following:

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

1. 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

1. 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

1. Large amounts of modulating signal power will be needed.

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

(i) Coherent/synchronous detection
(ii) Square-law detection
(iii) 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

$$x_c(t) = A_c\left[1 + mx(t)\right]\cos\omega_c t \quad \text{(refer Eq. 4.2)}$$

As shown in Fig. 4.20, coherent/synchronous detection consists of

(a) Generating the carrier signal, correct in frequency and phase, at the receiver
(b) Multiplexing $x_c(t)$, the received signal, by this locally generated carrier signal
(c) Lowpass filtering the above product of the two signals

$$\therefore \quad x_c(t)\cos\omega_c t = A_c\cos^2\omega_c t + mA_c x(t)\cos^2\omega_c t$$

$$= \frac{A_c}{2} + \frac{A_c}{2}\cos 2\omega_c t + \frac{mA_c}{2}x(t)\left[1+\cos 2\omega_c t\right]$$

If the highest frequency component present in $x(t)$ is W Hz, let the cutoff frequency of the lowpass filter be W Hz. Then at the output of the 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.

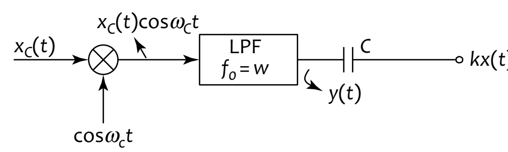


Fig. 4.20 *Block diagram of a coherent detector for AM signals*

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 Side Band Suppressed Carrier (DSB-SC) signal. It is sufficient 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

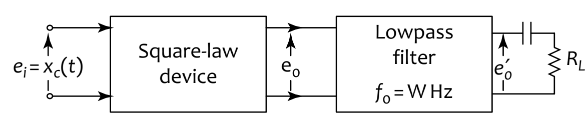


Fig. 4.21 *A square-law detector*

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 $$e_i = x_c(t) = A_c\left[1+mx(t)\right]\cos\omega_c t$$

$\therefore$ substituting this in Eq. (4.31), we have

$$e_0 = a_0 + a_1 A_c\left[1+mx(t)\right]\cos\omega_c t + a_2 A_c^2\left[1+mx(t)\right]^2\cos^2\omega_c t$$

$$= \left(a_0 + \frac{a_2 A_c^2}{2}\right) + \left[a_1 A_c + ma_1 A_c x(t)\right]\cos\omega_c t + \frac{a_2}{2}A_c^2\cos 2\omega_c t$$

$$+a_2 A_c^2 mx(t) + a_2 A_c^2 mx(t)\cos 2\omega_c t + \frac{a_2}{2}m^2 A_c^2 x^2(t) + \frac{a_2}{2}m^2 A_c^2 x^2(t)\cos 2\omega_c t$$

Since the lowpass filter has a cutoff frequency $f_0 = W$ Hz which is very small compared to the carrier frequency f_c, the output of the lowpass 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 cutoff frequency of the lowpass 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 $v_o = v_{in}^2$. If the Fourier transform of the signal m(t) is constant at the value M_o 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 Question)*

The square-law device of the square-law detector has an input–output relationship $v_0 = v_{\text{in}}^2$

$\therefore$ 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 cutoff 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 cutoff frequency of the LPF in the detector is f_m Hz.

$\therefore$ when $v_0(t)$ is lowpass 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.

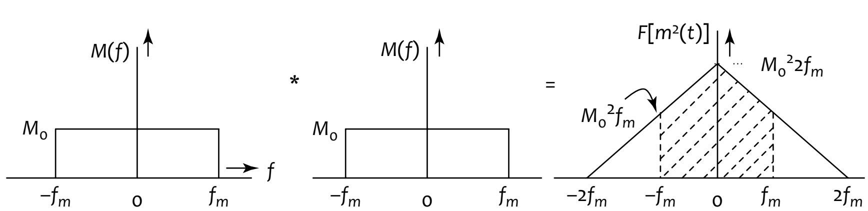


Fig. 4.22(a) *Signal for Example 4.19*

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

$$\frac{1}{2}\delta(t)+M(f)+\frac{1}{2}F[m(t).m(t)]$$
$$=\frac{1}{2}\delta(t)+M(f)+\frac{1}{2}[M(f).M(f)]$$

Before sketching $v_D(f)$, let us see the shape of $[M(f)*M(f)]$.

Because the LPF has a cutoff 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 $m^2(t)$].

$\therefore\ V_D(f)=\frac{1}{2}\delta(t)+M(f)+$ that part of $\mathcal{F}[m^2(t)]$ which is from $-f_m$ to f_m

When we sketch this we get Fig. 4.22(b).

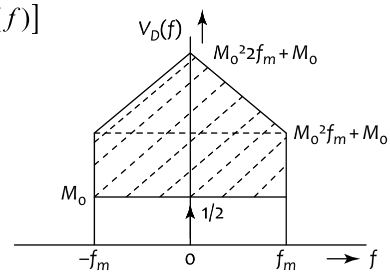


Fig. 4.22(b) *Fourier transform of the output*

4.3.3 The 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. 4.23.

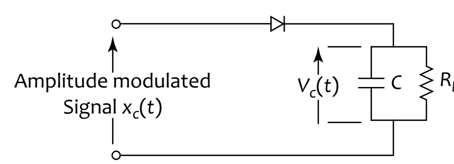


Fig. 4.23 *Basic circuit of an envelope detector*

Principle of Operation During the positive half-cycle of the r.f., the diode is forward biased and it conducts, charging the capacitor C. At the peak of an r.f. cycle, say point A, the capacitor gets charged to that peak value. Then onwards, the r.f. 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

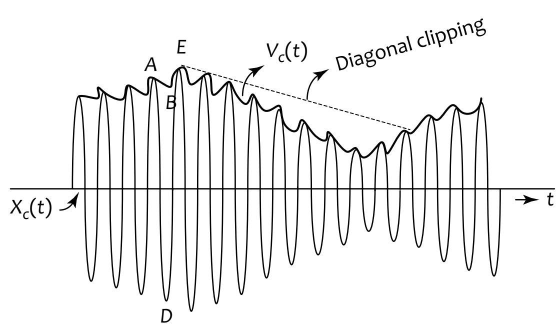


Fig. 4.24 *Working of an envelope detector*

discharging through the resistor R_L. While this process is going on, the r.f. voltage of the AM wave goes through the portion *ADB*. At the point *B*, the instantaneous voltage across the capacitor and the r.f. voltage of the AM wave are equal. After this instant corresponding to *B*, while the r.f. 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 r.f. cycle at which the diode stops conducting. This cycle of events will go on repeating in all the subsequent r.f. cycles. The voltage across the capacitor therefore follows the variations as shown by the thick line in Fig. 4.24. It is readily seen that $v_c(t)$, the voltage across the capacitor approximately follows the envelope of $x_c(t)$, the AM signal. Lowpass filtering the $v_c(t)$ removes the r.f. 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 the following:

(i) The charging of the capacitor takes place almost instantaneously so that the voltage across the capacitor can almost follow the portion of the r.f. 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

$$R_sC << \frac{1}{f_c} \tag{4.32}$$

(ii) The time-constant for the discharge of the capacitor, viz., R_LC, should be quite large compared the period of the r.f.,

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.

(iii) The discharge time-constant, although quite large compared to the r.f. period, $(1/f_c)$, 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. 4.24.

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

$$e(t) = A_c\left[1 + m\cos\omega_m t\right] \tag{4.36}$$

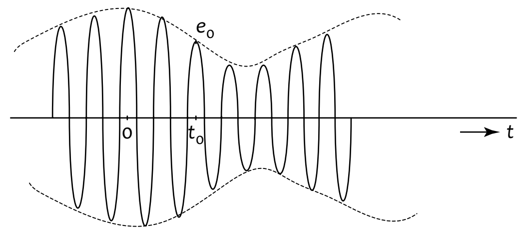


Fig. 4.25 *Amplitude-modulated signal*

$\therefore$ 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

$$e_0 = A_c\left[1 + m\cos\omega_m t_0\right] \tag{4.38}$$

At any time t ($t > t_0$), the voltage across the capacitor of Fig. 4.23 is given by

$$e_c(t) = e_0 e^{-(t-t_0)/R_L C}$$

$\therefore$ 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 $t = t_0$.

$$\therefore \text{ maximum rate of change of the capacitor voltage} = \left.\frac{de_c(t)}{dt}\right|_{t=t_0} = \frac{-e_0}{R_L C} \tag{4.38(a)}$$

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.

$$\therefore \qquad \frac{e_0}{R_L C} \ge A_c m\omega_m \sin\omega_m t_0 \tag{4.39}$$

If we now substitute for e_0 in the above equation by using Eq. (4.38), we have

$$\frac{A_c\left[1 + m\cos\omega_m t_0\right]}{R_L C} \ge A_c m\omega_m \sin\omega_m t_0$$

or,

$$R_L C \le \frac{1}{\omega_m\left[\dfrac{m\sin\omega_m t_0}{1 + m\cos\omega_m t_0}\right]} \tag{4.40}$$

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

$$\cos \omega_m t_0 = -m$$

$\therefore$ corresponding to this worst-case condition,

$$\boxed{R_L C \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]} \tag{4.41}$$

Equation (4.41) 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 p.f 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?

$$R_L C = 400 \times 10^3 \times 100 \times 10^{-12} = 4 \times 10^{-5} \text{ s}$$

$\therefore$ from Eq. (4.41), we have

$$R_L C \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]$$

$$\therefore \qquad 4 \times 10^{-5} \le \frac{1}{6.28 f_m}\left[\frac{\sqrt{1-(0.75)^2}}{0.75}\right]$$

$$\therefore \qquad f_m \le \frac{10^5}{4 \times 6.28}\left[\frac{0.6614}{0.75}\right] = 3510 \text{ Hz}$$

$$\therefore \qquad f_m \le 3510 \text{ Hz} \quad \therefore \text{ max. frequency} = 3510 \text{ Hz}$$

4.3.4 A Practical Diode Detector

The circuit of a practical diode detector is shown in Fig. 4.26.

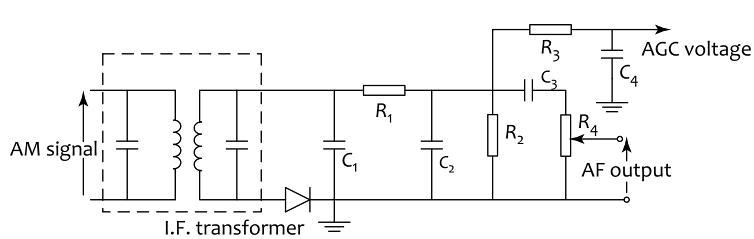


Fig. 4.26 *A practical diode detector*

In this circuit, C_1 and C_2 are provided for r.f. 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 r.f., 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 filters 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 now make the following simplifying assumptions.

(i) C_1 and C_2 act as short circuits for the carrier (intermediate) frequency and as open circuits for dc and audio frequencies.

(ii) 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. 4.26, 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 be seeing presently, a distortion, referred to as *negative peak clipping*, results 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. 4.26, we find that

- The detector load for dc $\underline{\Delta}$ $R_{DC} = R_1 + R_2$ (4.42)
- The detector load for audio frequencies $\underline{\Delta}$ $R_{AC} = R_1 + (R_2 \parallel R_3 \parallel R_4)$ (4.43)

where $(R_2 \parallel R_3 \parallel R_4)$ denotes parallel combination of R_2, R_3 and R_4.

Since $R_2 > (R_2 \parallel R_3 \parallel R_4)$, it follows that the ac load (i.e., at audio frequencies) of the detector is always less than the dc load.

$$\therefore \quad R_{ac} < R_{dc} \tag{4.44}$$

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,
$$I_m = \frac{A_m}{R_{ac}} \quad \text{and} \quad I_c = \frac{A_c}{R_{dc}}$$

where,

A_m = Peak of the audio component of the envelope

A_c = Peak of the unmodulated carrier wave

$$\therefore \quad m_d = \frac{I_m}{I_c} = \frac{A_m / R_{ac}}{A_c / R_{dc}} = \left(\frac{A_m}{A_c}\right)\left(\frac{R_{dc}}{R_{ac}}\right) = m\left(\frac{R_{dc}}{R_{ac}}\right) \tag{4.45}$$

$$\therefore \quad m_{max} = m_{d\,max}\left(\frac{R_{ac}}{R_{dc}}\right)$$

But, $m_{d\,max} \le 1$

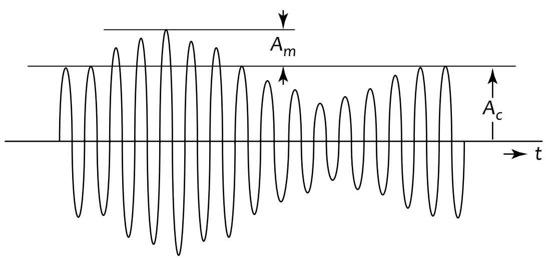


Fig. 4.27 *Received AM signal*

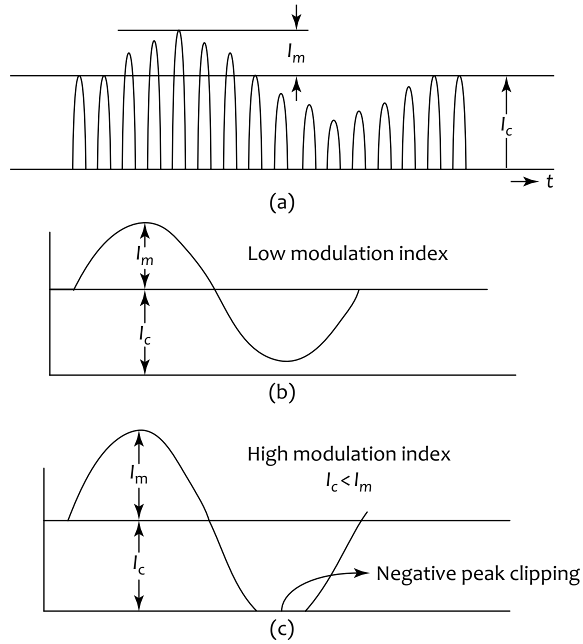


Fig. 4.28 **(a)** *Actual diode current* **(b)** *dc and audio components of current for low value of m* **(c)** *dc and audio components of current for high value of m*

$$\boxed{\therefore m_{\max} = \left(\frac{R_{ac}}{R_{dc}}\right)} \tag{4.46}$$

and,

$$\boxed{\text{negative peak clipping occurs if } m > \left(\frac{R_{ac}}{R_{dc}}\right)} \tag{4.47}$$

The above result is only an approximation since the assumptions listed in (i) and (ii) 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, 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 *Electronic and Radio Engineering* by F E Terman, McGraw-Hill.

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.

(University Question)

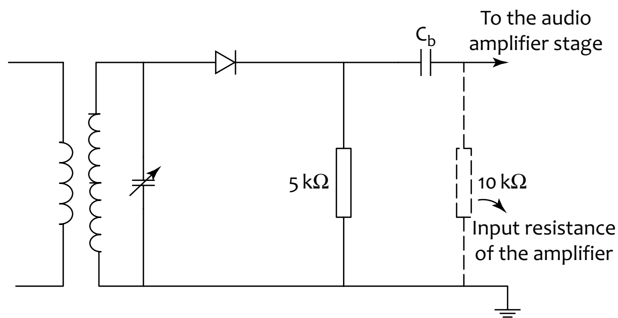


Fig. 4.29 *Circuit for Example 4.21*

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.

i.e.,
$$R_{ac} = \frac{5\times 10}{(5+10)}\text{ k}\Omega = 3.3\text{ k}\Omega$$

$$R_{dc} = 5\text{ k}\Omega$$

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

$$m_{max} = \frac{R_{ac}}{R_{dc}} = \frac{3.3}{5} = 0.66$$

Example 4.22

A signal $x_c(t) = 5[1 + 2\cos\omega_c(t)]$ is to be demodulated. Check whether some of the following detectors can be used: (i) an envelope detector, (ii) a square-law detector, and (iii) a synchronous detector or coherent detector.

Given,
$$x_c(t) = 5[1+2\cos\omega_m t]\cos\omega_c t$$

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 $= ax_c^2(t)$

$$= a\left\{25\left[1+4\cos\omega_m t+4\cos^2\omega_m t\right]\left[\frac{1}{2}\left(1+\cos 2\omega_c t\right)\right]\right\}$$

$$= \left\{\frac{25}{2}+50\cos\omega_m t+\frac{50}{2}\left(1+\cos 2\omega_m t\right)+\frac{25}{2}\cos 2\omega_c t+50\cos 2\omega_c t.\cos\omega_m t+50\cos^2\omega_m t.\cos 2\omega_c t\right\}a$$

If the dc component is blocked by a coupling condenser and the high frequency components are removed by using an LPF of cutoff frequency f_m after the square-law device, the final output will be $z(t) = a.0.5\cos\omega_m t$, 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 cutoff frequency of W Hz, the bandwidth of the modulating signal.

$$x_c(t)\cos\omega_c t = 5\left[1+2\cos\omega_m t\right]\cos^2\omega_c t = \frac{5}{2}\left[1+2\cos\omega_m t\right]\left[1+2\cos\omega_c t\right]$$

$$= \frac{5}{2}+\frac{5}{2}\cos 2\omega_c t + 5\cos\omega_m t + \frac{5}{4}\cos(2\omega_c+\omega_m)t + \frac{5}{4}\cos(2\omega_c-\omega_m)t$$

$\therefore$ the output of the LPF $= z(t) = \left(5\cos\omega_m t + \dfrac{5}{2}\right)$

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 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 DSB-SC 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

$$x_c(t) = A_c x(t)\cos\omega_c t \tag{4.48}$$

(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 in power as well as bandwidth. It is in forming the chrominance signal in the NTSC and PAL colour 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 an 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.

4.4.1 Time-Domain Representation of DSB-SC Signals

From Eq. (4.48), we have
$$\underset{\text{(DSB-SC)}}{x_c(t)} = x(t)A_c \cos\omega_c t \tag{4.48a}$$

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. 4.30, simple envelope detector using a diode cannot be used for recovering the message signal from a DSB-SC signal.

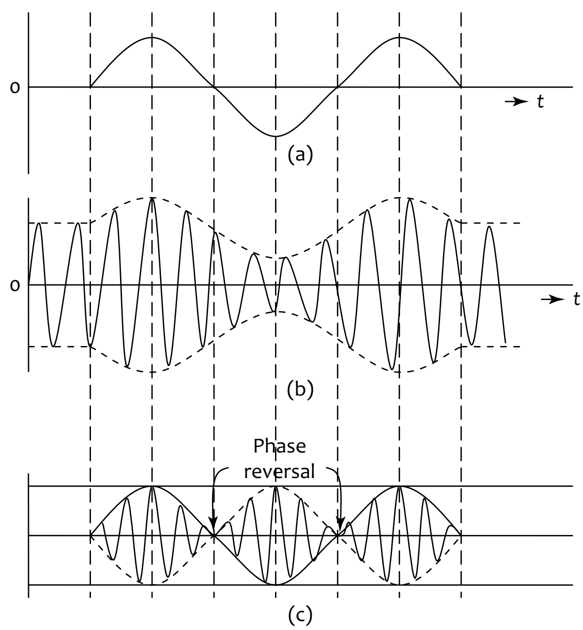


Fig. 4.30 *Waveform of* **(a)** *modulating signal,* **(b)** *AM signal (m < 1), and* **(c)** *the DSB-SC signal (product of x(t) and c(t))*

Power in a DSB-SC Signal

From Eq. (4.48) 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

$$P_{x_c} = \underset{T\to\infty}{\mathrm{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} x_c^2(t)\,dt$$

$$= \underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} \frac{1}{2} A_c^2 x^2(t)\left[1+\cos 2\omega_c t\right]dt$$

$$= \frac{A_c^2}{2} \underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt + \frac{A_c^2}{2} \underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} \cos 2\omega_c t dt$$

The second integral is zero since it is the area under a cosine curve. Further,

$$\underset{T\to\infty}{\text{Lt}} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t)dt = P_x = \text{Average power in the modulating signal.}$$

$$\boxed{\therefore \quad P_{x_c} = \frac{A_c^2}{2} \cdot P_x} \tag{4.49}$$

4.4.2 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. 4.31. 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. (4.48), we have

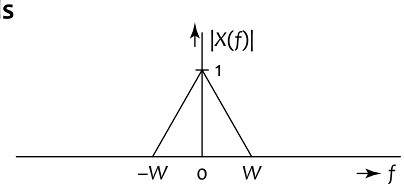


Fig. 4.31 *Spectrum of the modulating signal*

$$X_c(f) = \mathcal{F}\left[x(t)\cdot A_c \cos\omega_c t\right] = \frac{1}{2} A_c \left[X(f-f_c) + X(f+f_c)\right] \tag{4.49a}$$

Figure 4.32 gives a plot of $|X_c(f)|$ making use of the $X(f)$ that we have assumed earlier.

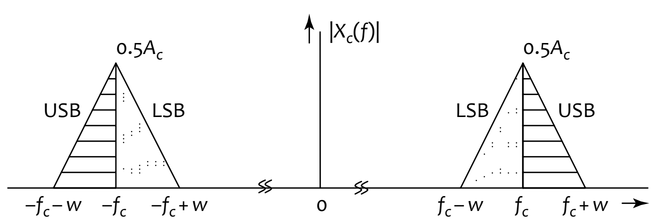


Fig. 4.32 *Amplitude spectrum of a DSB-SC signal*

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 4.23

The modulating signal in an AM-SC system is a multiple-tone signal by $m(t) = A_1\cos\omega_1 t + A_2\cos\omega_1 t + A_3\cos\omega_1 t$. 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 $A_3 > A_2 > A_3$. *(University Question)*

An AM-SC system is nothing but a 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

$$x_c(t) = m(t)\left[A_c \cos\omega_c t\right] = \left[A_1 \cos\omega_1 t + A_2 \cos\omega_2 t + A_3 \cos\omega_3 t\right] A_c \cos\omega_c t$$

$$= A_1 A_c \cos\omega_c t \cos\omega_1 t + +A_2 \; A_c \cos\omega_c t \cos\omega_2 t + A_3 \; A_c \cos\omega_c t \cos\omega_3 t$$

$$= \frac{1}{2} A_1 A_c \left[\cos(\omega_c + \omega_1)t + \cos(\omega_c - \omega_1)\right] + \frac{1}{2} A_2 \; A_c \left[\cos(\omega_c + \omega_2)t + \cos(\omega_c - \omega_2)\right]$$

$$+ \frac{1}{2} A_3 \; A_c \left[\cos(\omega_c + \omega_3)t + \cos(\omega_c - \omega_3)\right]$$

Noting that $\omega_3 > \omega_2 > \omega_1$; $A_1 > A_2 > A_3$ and $\omega_c >> \omega_1$, ω_2 and ω_3, we may plot the one-sided spectrum of $x_c(t)$ as shown in Fig. 4.33.

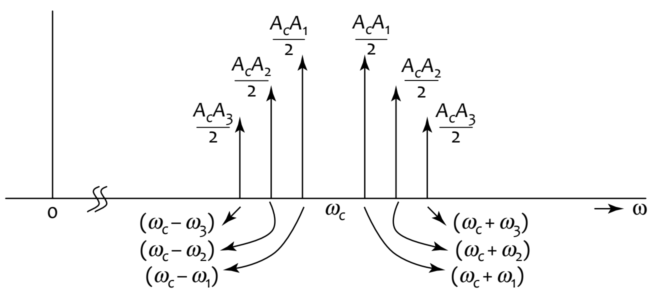


Fig. 4.33 *One-sided output spectrum for Example 4.23*

Example 4.24

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.

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 in Fig. 4.34.

Hence, the spectrum of the DSB-SC modulated signal is as shown in Fig. 4.35.

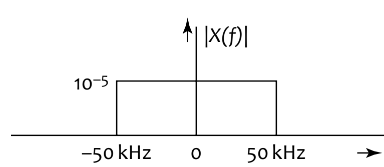


Fig. 4.34 *Fourier transform of x(t)*

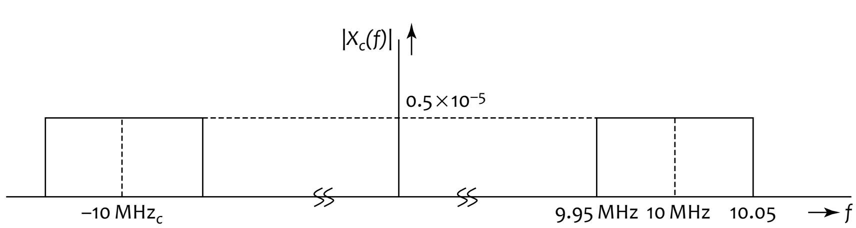


Fig. 4.35 *Spectrum of $x_c(t)$*

4.4.3 Generation of DSB-SC Signals

(i) 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

$$x_{c_1}(t) = A_c[1 + mx(t)]\cos\omega_c t \tag{4.50}$$

and

$$x_{c_2}(t) = A_c[1 - mx(t)]\cos\omega_c t \tag{4.51}$$

Subtracting x_{c_2} from x_{c_1}, we have

$$x_{c_3}(t) = x_{c_1}(t) - x_{c_2}(t) = 2mA_c x(t)\cos\omega_c t \tag{4.52}$$

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. 4.36 in which we have used two identical amplitude modulators using the non-linearity of FETs (see Fig. 4.12).

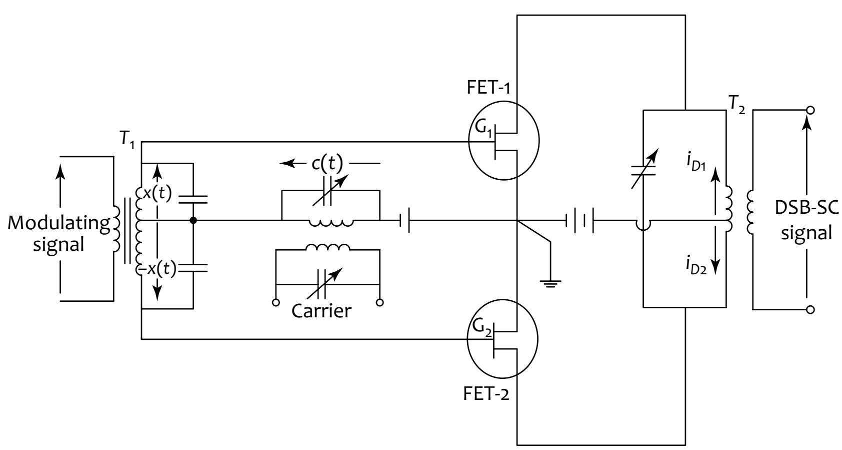


Fig. 4.36 *A balanced modulator using FETs*

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 signals 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. 4.12.

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 $$e_{g_1} = x(t) + A_c \cos\omega_c t$$

$$\therefore \quad i_{D_1} = a_0 + a_1\left[x(t) + A_c \cos\omega_c t\right] + a_2\left[x(t) + A_c \cos\omega_c t\right]^2$$

$$= \left[a_0 + \frac{a_2}{2}A_c^2\right] + a_1 x(t) + a_1 A_c \cos\omega_c t + a_2 x^2(t) + \frac{a_2}{2}A_c^2 \cos 2\omega_c t + 2a_2 x(t) A_c \cos\omega_c t \quad (4.53)$$

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 $e_{g_2} = -x(t) + A_c \cos\omega_c t$

$$\therefore \quad i_{D_2} = \left[a_0 + \frac{a_2}{2}A_c^2\right] - a_1 x(t) + a_1 A_c \cos\omega_c t + a_2 x^2(t) + \frac{a_2}{2}A_c^2 \cos 2\omega_c t - 2a_2 x(t) A_c \cos\omega_c t$$

$$\therefore \quad (i_{D_1} - i_{D_2}) = 2a_1 x(t) + 4a_2 A_c x(t) \cos\omega_c t$$

The output tank circuit, which is tuned to the carrier frequency, eliminates the first term and a voltage $e_0(t)$ where,

$$e_0(t) = kx(t)\cos\omega_c t \quad (4.54)$$

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 practice, 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.

(ii) The Ring Modulator (Balanced Modulator using Diodes) One popular type of balanced modulator (especially in telephone circuits), is the ring modulator shown in Fig. 4.37. 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 centre-taps of the secondary of transformer T_1 and the centre-tap of the primary of transformer T_2, as shown in the figure. First let us assume that the modulating signal is absent.

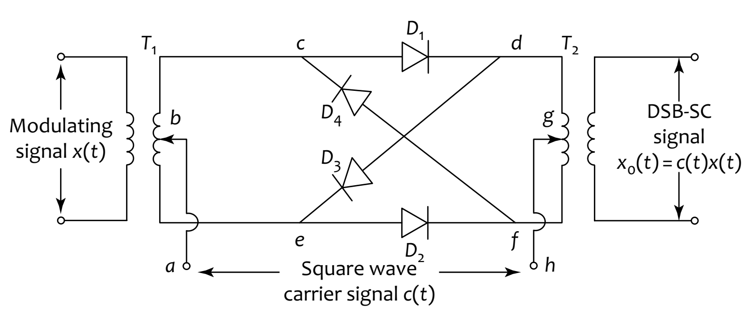


Fig. 4.37 *A ring modulator*

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 centre taps b and g are true centre-taps and the diodes are identical in their behaviour, 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.

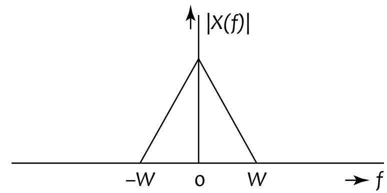


Fig. 4.38 *Spectrum of the modulating signal*

It may be noted that just like the case of the DSB-SC modulator of Fig. 4.30, 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. 4.38 with a bandwidth of W Hz, then the spectrum of the ring modulator output would be as shown in Fig. 4.39.

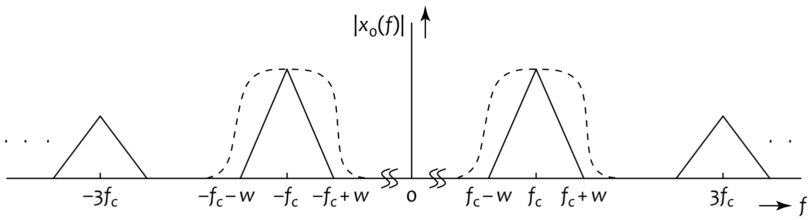


Fig. 4.39 *Spectrum of the output of the ring modulator*

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 centre frequency f_c and bandwidth $2W$, at the output of the ring modulator, we can obtain the desired DSB-SC modulated signal.

Example 4.25

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.

(University Question)

Referring to Figs. 4.37 and 4.38, and noting that, as mentioned underneath Fig. 4.39, 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, because of the bandpass filter with centre 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 3.4 kHz frequency, the output of the ring modulator (with the above mentioned bandpass 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.

4.4.4 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 4.30(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

$$x_c(t) \cos \omega_c t = A_c x(t) \cos^2 \omega_c t = A_c x(t) \left[\frac{1 + \cos 2\omega_c t}{2} \right]$$

$$= \frac{1}{2} A_c x(t) + \frac{1}{2} A_c x(t) \cos 2\omega_c t \tag{4.55}$$

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 lowpass filtering. Synchronous detection, or coherent detection may therefore be represented by the following block diagram in Fig. 4.40.

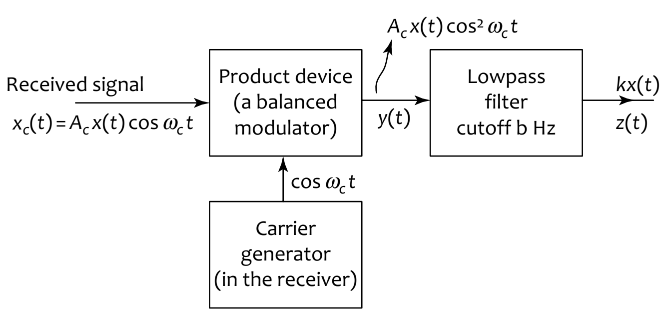


Fig. 4.40 *A coherent detector*

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. 4.40, the output $y(t)$ of the product device will now be

$$y(t) = A_c x(t) \cos \omega_c t \cdot \cos(\omega_c t + \theta)$$

$$\therefore \quad y(t) = \frac{1}{2} A_c x(t) \cos(2\omega_c t + \theta) + \frac{1}{2} A_c x(t) \cos \theta \tag{4.56}$$

The lowpass filter following the product device has a cutoff 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. (4.56) 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 lowpass filter and we have:

$$z(t) = \frac{1}{2} A_c x(t) \cos\theta \tag{4.57}$$

Thus, the signal output of the coherent detector is $x(t)\cos\theta$ instead of $x(t)$. This has the following consequences.

(i) Even if θ remains constant, which of course, is not true in practice, the cos θ factor tends to reduce the output message signal.

(ii) 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 4.26

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

(i) $\phi = 0$ and (ii) $\phi \neq 0$ and $|m(t)| << A_c/2$ (University Question)

The input to the envelope detector is given by

$$y(t) = \left[A_c \cos(\omega_c t + \phi) + m(t)\cos\omega_c t\right]$$

$$= \left[A_c \cos\phi + m(t)\right]\cos\omega_c t - \left[A_c \sin\omega_c t\right]\sin\omega_c t \tag{A}$$

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,

$$y(t) = R(t)(\cos\omega_c t)\cos\theta(t) - R(t)(\sin\omega_c t)\sin\theta(t)$$

$$= \left[R(t)\cos\theta(t)\right]\cos\omega_c t - \left[R(t)\sin\theta(t)\right]\sin\omega_c t \tag{B}$$

Comparing Eqs (a) and (b),

$$R(t)\cos\theta(t) = \left[A_c \cos\phi + m(t)\right]$$

and $R(t)\sin\theta(t) = A_c \sin\phi$

Hence,
$$R(t) = \sqrt{R^2(t)\cos^2\theta(t) + R^2(t)\sin^2\theta(t)}$$

$$= \sqrt{\left[A_c\cos\phi + m(t)\right]^2 + \left[A_c \sin\phi\right]^2}$$

∴
$$R(t) = \sqrt{A_c^2 + m^2(t) + 2m(t)A_c\cos\phi}$$

and $$\theta(t) = \tan^{-1}\left[\frac{A_c \sin\phi}{A_c \cos\phi + m(t)}\right]$$

(i) 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.

(ii) 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)

$$\therefore \quad R(t) \cong \sqrt{A_c^2 + 2A_c m(t)\cos\phi} = A_c\sqrt{1 + \frac{2m(t)}{A_c}\cos\phi}$$

Now $A_c >> 2|m(t)|$ $\quad \therefore \quad \frac{2m(t)}{A_c}\cos\phi << 1$

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 f 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.

4.4.5 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 *In-phase 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. 4.41, both these outputs are fed to the phase discriminator, which consists of a product device followed by a lowpass filter. For θ values that are quite small, we know that cos $\theta \approx 1$ and sin $\theta \approx 0$. Thus, the output of the product device in the phase discriminator is of the form $A_c^2 x^2(t)\theta$. The lowpass filter, which has a very low cutoff 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 $x^2(t)$.

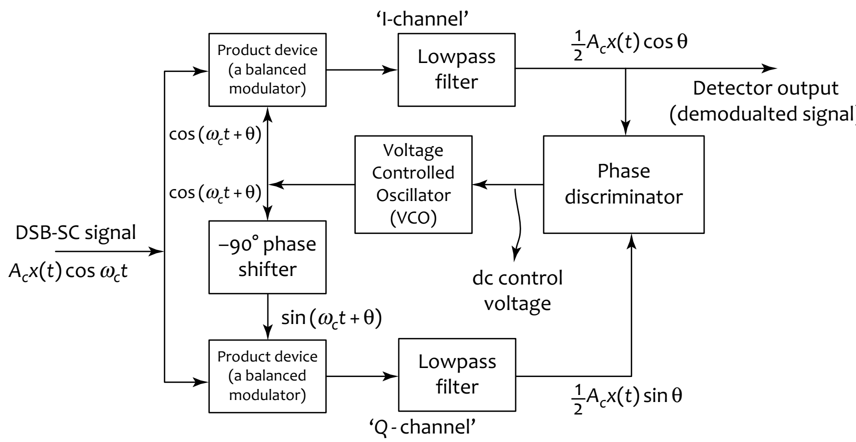


Fig. 4.41 *Costas receiver or Costas loop*

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.

4.4.6 The 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.

$$y(t) = x_c^2(t) = A_c^2 x^2(t)\cos^2\omega_c t = \frac{1}{2}A_c^2 x^2(t)\left[1+\cos 2\omega_c t\right]$$

$$z(t) = \frac{1}{2}A_c^2 x^2(t)\cos 2\omega_c t$$

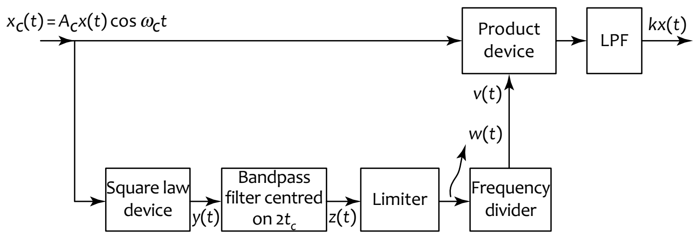


Fig. 4.42 *The squaring loop for carrier recovery*

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 lowpass filtering this product using a lowpass filter with a cutoff frequency B Hz such that $W < B < (2f_c - W)$, where W Hz is the bandlimiting frequency of the modulating signal.

Just like the Costas loop, the squaring loop also suffers from the disadvantage of 180° phase ambiguity in so far as the demodulated signal $x(t)$ is concerned.

4.4.7 Quadrature Carrier Multiplexing of DSB-SC Signals (QAM)

Quadrature carrier multiplexing (also called Quadrature Amplitude Modulation, 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,

$$x_c(t) = x_{c_1}(t) + x_{c_2}(t) = A_c \left[x_1(t) \cos \omega_c t + x_2(t) \sin \omega_c t\right],$$

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, signals $x_1(t)$ and $x_2(t)$ can be recovered from the multiplexed signal $x_c(t)$ by coherent detection as shown in Fig. 4.43, wherein the balanced modulators act as product devices.

At the receiving end, the message signals $x_1(t)$ and $x_2(t)$ are recovered in the following manner.

$$x_c(t) = A_c x_1(t) \cos \omega_c t + A_c x_2(t) \sin \omega_c t$$

$$\therefore \quad x_c(t) \cos \omega_c t = \frac{1}{2} A_c x_1(t)\left[1 + \cos 2\omega_c t\right] + \frac{1}{2} A_c x_2(t) \sin 2\omega_c t$$

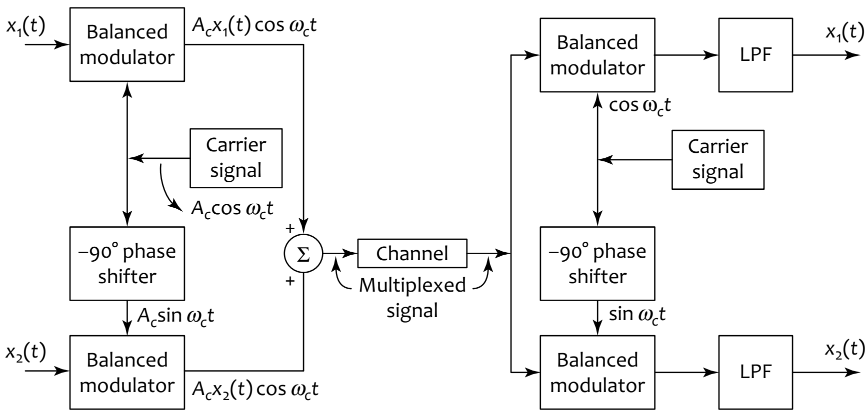


Fig. 4.43 *Quadrature carrier multiplexed system*

Subsequent lowpass 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 lowpass 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 4.27

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?

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. 4.40)

$$y(t) = A_c x(t)\cos\omega_c t \cdot \cos(\omega_c + \Delta\omega)t$$

$$= \frac{1}{2}A_c x(t)\cos(\Delta\omega)t \cdot \cos(2\omega_c + \Delta\omega)t$$

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.

4.5 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 Side Band Suppressed Carrier or SSB-SC modulation.

4.5.1 Frequency-Domain and Time-Domain Representation of SSB-SC Signals

In Fig. 4.32 we had sketched the amplitude spectrum of a typical DSB-SC signal. Figure 4.44 shows the same with a scaled version of the amplitude spectrum of the message signal itself super imposed on it.

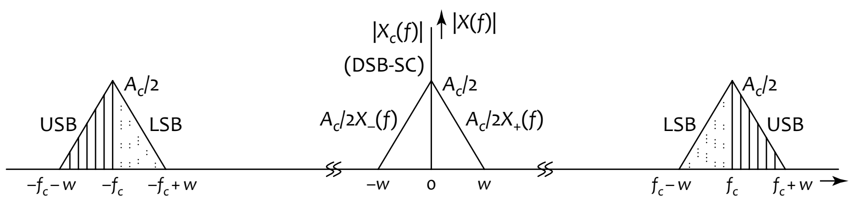


Fig. 4.44 *Amplitude spectrum of DSB-SC signal and amplitude spectrum of the message signal (scaled)*

From Fig. 4.44, 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 Figs 4.45(a) and (b) respectively.

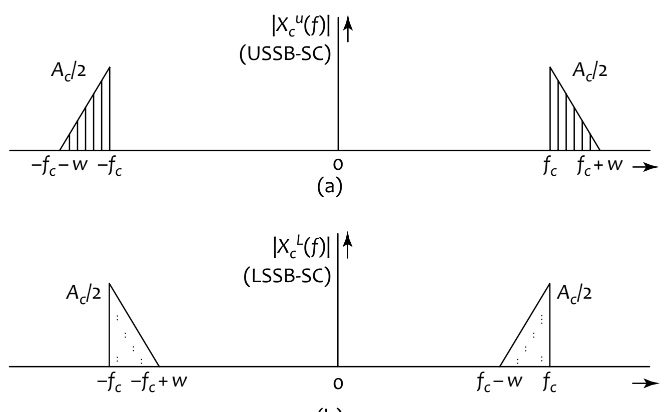


Fig. 4.45 **(a)** *Spectrum of a USSB-SC signal* **(b)** *Spectrum of a LSSB-SC signal*

The message spectrum shown in Fig. 4.45 may be visualized as the sum of $\frac{A_c}{2}X_+(f)$ and $\frac{A_c}{2}X_-(f)$ where $\frac{A_c}{2}X_+(f)$ is the positive frequency part and $\frac{A_c}{2}X_-(f)$ is the negative frequency part. From Eq. (2.154) of Section 2.10, we know that

$$\mathcal{F}^{-1}\left[\frac{A_c}{2}X_+(f)\right] = \frac{A_c}{4}x_+(t) \tag{4.58}$$

and $$\mathcal{F}^{-1}\left[\frac{A_c}{2}X_-(f)\right]=\frac{A_c}{4}x_-(t) \tag{4.59}$$

where,

$$x_+(t)=x(t)+j\hat{x}(t) \tag{4.60}$$

is the pre-envelope of $x(t)$ for positive frequencies,
and

$$x_-(t)=x(t)-j\hat{x}(t) \tag{4.61}$$

is the pre-envelope of $x(t)$ for negative frequencies.

Then, from Fig. 4.44, we may write the following:

If spectrum of the USSB-SC signal $= X_c^u(f)$,

then, $$X_c^u(f)=\frac{A_c}{2}\left[X_+(f-f_c)+X_-(f+f_c)\right] \tag{4.62}$$

Taking the inverse Fourier transform on both sides of the above, we get

$$x_c^u(t)= \text{ USSB-SC signal } =\frac{A_c}{4}x_+(t)e^{-j\omega_c t}+\frac{A_c}{4}x_-(t)e^{j\omega_c t}$$

$$=\frac{A_c}{4}\left[x(t)+j\hat{x}(t)\right]e^{-j\omega_c t}+\frac{A_c}{4}\left[x(t)-j\hat{x}(t)\right]e^{+j\omega_c t}$$

$$=\frac{A_c}{4}\left[x(t)\left\{e^{j\omega_c t}+e^{-j\omega_c t}\right\}\right]-j\frac{A_c}{4}\left[\hat{x}(t)\left\{e^{j\omega_c t}-e^{-j\omega_c t}\right\}\right]$$

$\therefore$ $$x_c^u(t)=\frac{A_c}{2}\left[x(t)\cos\omega_c t-\hat{x}(t)\sin\omega_c t\right] \tag{4.63}$$

Equation (4.63) 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

$$x_c^L(t)=\frac{A_c}{2}\left[x(t)\cos\omega_c t+\hat{x}(t)\sin\omega_c t\right] \tag{4.64}$$

Example 4.28

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.

From Eq. (4.63), we have

$$x_c^u(t)=\frac{A_c}{2}\left[x(t)\cos\omega_c t-\hat{x}(t)\sin\omega_c t\right]$$

$$=\frac{A_c}{2}\left[\cos\omega_m t\cdot\cos\omega_c t-\sin\omega_m t\cdot\sin\omega_c t\right]$$

$$=\frac{A_c}{2}\cos\left(\omega_c+\omega_m\right)t$$

$\therefore$ $$X_c^u(f)=\frac{A_c}{4}\left[\delta\left\{f-(f_c+f_m)\right\}+\delta\left\{f+(f_c+f_m)\right\}\right]$$

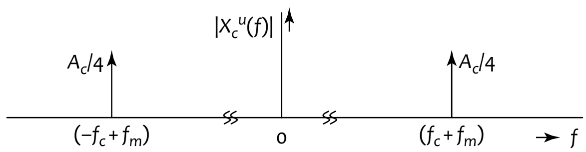


Fig. 4.46 *Spectrum of the USSB-SC modulated signal of Example 4.28*

4.5.2 Methods of Generation of SSB-SC Signals

There are *mainly* two methods of generation of SSB-SC modulated signals.

(i) **Filter method** or the **Balanced Modulator-filter Method:** In this method, we first generate a DSB-SC signal and then filter out from it the unwanted sideband.

(ii) **Phasing Method**: This method of generation of SSB-SC signals is based on direct implementation of the Eqs (4.63) and (4.64) 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 Generation of 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. We will now elaborate this.

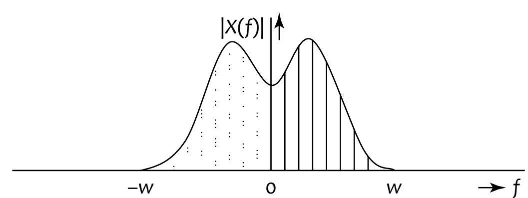


Fig. 4.47 *Spectrum of x(t)*

Suppose the modulating signal $x(t)$ has a spectrum as shown in Fig. 4.47. Then the DSB-SC signal will have a spectrum as shown in Fig. 4.48.

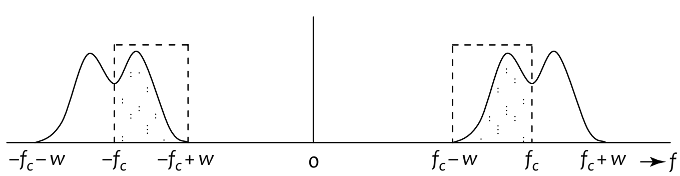


Fig. 4.48 *Spectrum of the DSB-SC signal. The dotted lines show the passband of the filter for obtaining LSSB-SC signal.*

To obtain an LSSB-SC signal from the DSB-SC signal, the passband 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. 4.49 with

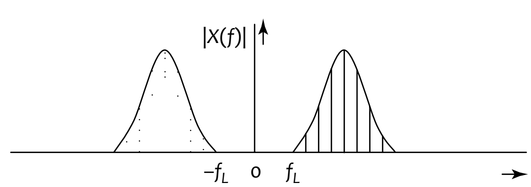


Fig. 4.49 *Spectrum of a modulating signal with an energy gap*

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. 4.50. As may be seen from the figure, it is now possible to make use of a practical narrow bandpass filter with a finite transition bandwidth for suppressing the unwanted upper sideband of the DSB-SC signal.

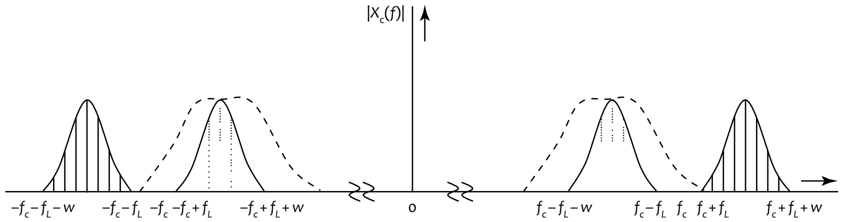


Fig. 4.50 *Spectrum of a DSB-SC signal when the modulating signal has a gap in its spectrum, as shown in Fig. 4.45. Filter passband is shown in dotted lines.*

The filter, a narrow bandpass 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 upto about 300 Hz; i.e., these signals possess a spectrum of the type shown in Fig. 4.49, 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 atleast 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. 4.51.

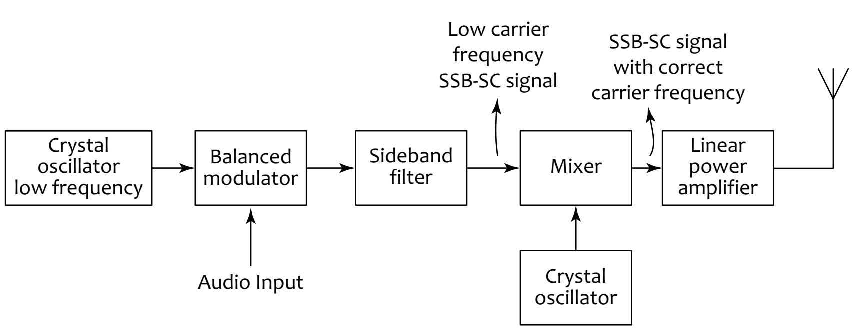


Fig. 4.51 *Block diagram of a SSB-SC transmitter*

With regard to the above block diagram, the following points may be noted.

(i) 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.

(ii) 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.

(iii) The sideband filter must *attenuate* the unwanted sideband at least up to 60 dB relative to the desired sideband.

Alternatively, we may use a 2-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. 4.52.

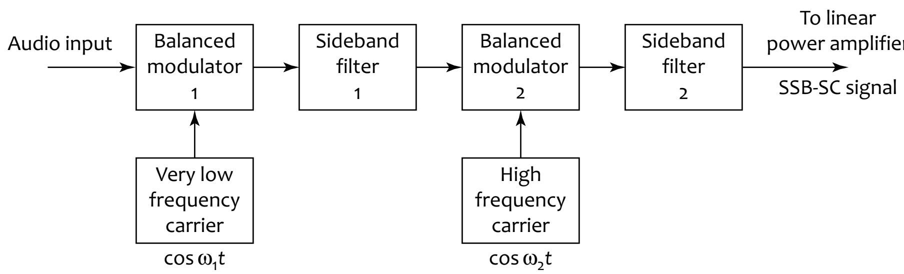


Fig. 4.52 *A two-stage SSB-SC modulator*

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 Generation of SSB-SC Signals This method is based on direct implementation of Eq. (4.63) or (4.64) 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. 4.53. 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.

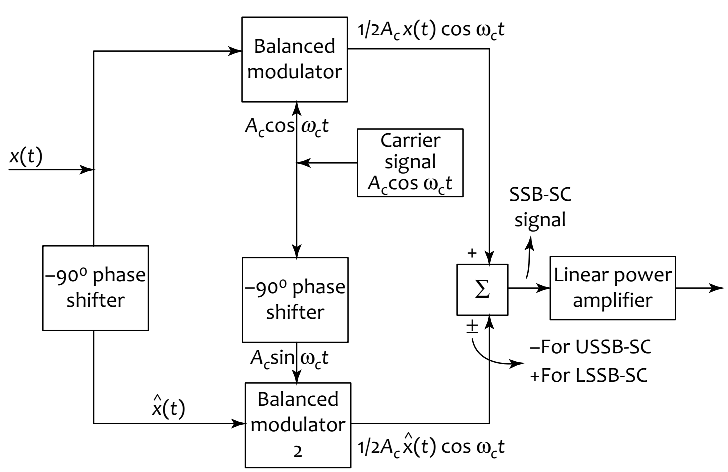


Fig. 4.53 *The phasing method of generation of an SSB-SC signal*

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. The 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 vive-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, or The Third Method This method, a variant of the phasing method, was invented in 1950 by DK Weaver. It avoids the need for wideband phase-shifters which are difficult to construct and expensive and instead, uses an AF subcarrier at an audio frequency, say f_0.

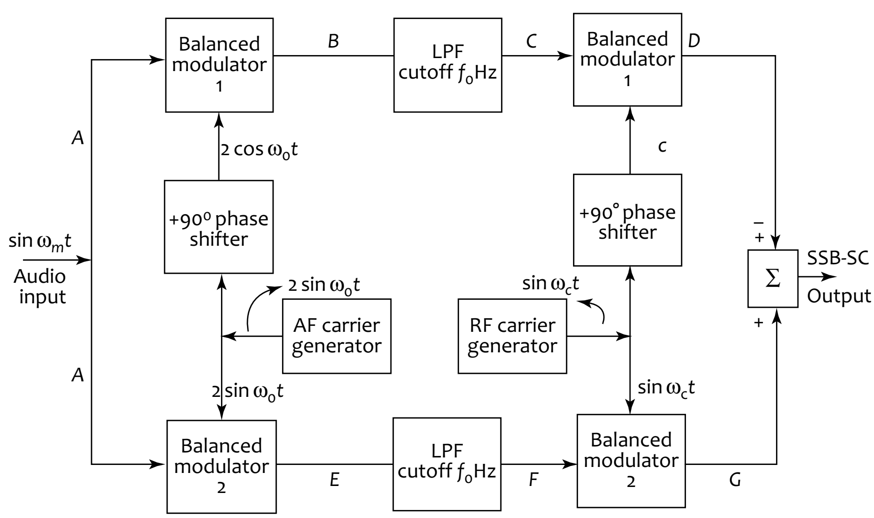


Fig. 4.54 *Weaver's method of generation of SSB-SC signals*

A: $\sin \omega_m t$

B: $2 \sin \omega_m t \sin \omega_0 t$

C: $\sin(\omega_m - \omega_0)t$

D: $\sin(\omega_m - \omega_0)t \cos \omega_c t$

$$= \frac{1}{2}\left[\sin(\omega_c - \omega_0 + \omega_m)t - \sin(\omega_c + \omega_0 - \omega_m)t\right]$$

E: $2 \sin \omega_m t \sin \omega_0 t$

F: $\cos(\omega_m - \omega_0)t$

G: $\sin\omega_c t \cos(\omega_m - \omega_0)t$

$$= \frac{1}{2}\left[\sin(\omega_c + \omega_0 - \omega_m)t + \sin(\omega_c - \omega_0 + \omega_m)t\right]$$

$\therefore\ D + G$: $\sin(\overline{\omega_c - \omega_0} + \omega_m)$

$-D + G$: $\sin(\overline{\omega_c + \omega_0} + \omega_m)$

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.

4.5.3 Detection of SSB-SC Signals

SSB-SC signals can be demodulated using coherent detection, as shown in Fig. 4.55:

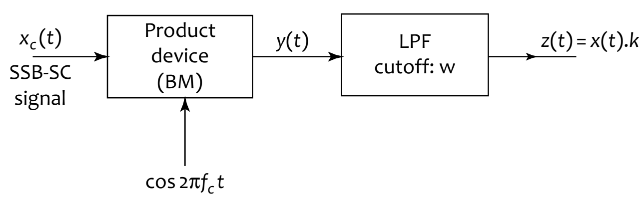


Fig. 4.55 *Coherent detection of SSB-SC signals*

Let the SSB-SC signal be $x_c(t)$. Then from Eqs (4.63) and (4.64),

$$x_c(t) = \frac{A_c}{2}\left[x(t)\cos\omega_c t \pm \hat{x}(t)\sin\omega_c t \cos\omega_c t\right]$$

$$\therefore \qquad y(t) = x_c(t)\cos\omega_c t$$

$$= \frac{A_c}{2}x(t)\cos^2\omega_c t \pm \frac{A_c}{2}\hat{x}(t)\sin\omega_c t\cos\omega_c t$$

$$= \frac{A_c}{4}x(t) + \frac{A_c}{4}x(t)\cos 2\omega_c t \pm \frac{A_c}{4}\hat{x}(t)\sin 2\omega_c t$$

The second and third terms are high-frequency terms and will be rejected by the LPF whose cutoff frequency is W Hz, the band-limiting frequency of $x(t)$.

Hence, $$z(t) = \frac{1}{4}A_c x(t) = k \cdot x(t) \tag{4.65}$$

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. 4.56(a) and (b).

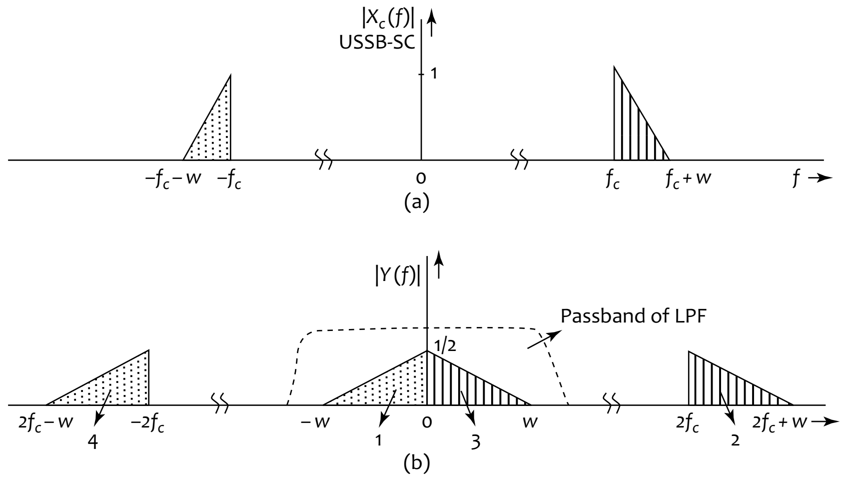


Fig. 4.56 *Frequency-domain interpretation of coherent detection of a USSB-SC signal*

Figure 4.56(a) shows the spectrum of the USSB-SC signal, $x_c(t)$.

Now, $$y(t) = x_c(t) \cdot \cos 2\pi f_c t$$

$\therefore$ $$Y(f) = X_c(f) * \frac{1}{2}\left[\delta(f-f_c)+\delta(f+f_c)\right]$$

$$= \frac{1}{2}X_c(f-f_c)+\frac{1}{2}X_c(f+f_c) \tag{4.66}$$

1 and 2 of Fig. 4.56(b) represent $\frac{1}{2}X(f-f_c)$ while 3 and 4 represent $\frac{1}{2}X(f+f_c)$. Thus, Fig. 4.56(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 passband 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.

4.5.4 Effect of Phase and Frequency Errors of the Local Carrier

(i) When the Local Carrier has a Phase Error θ

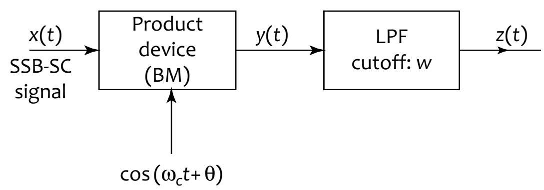


Fig. 4.57 *Effect of phase error of the local carrier*

Let the received SSB-SC signal be represented by

$$x_c(t) = \frac{A_c}{2}\left[x(t)\cos\omega_c t \mp \hat{x}(t)\sin\omega_c t\right]$$

where, as we know, the minus sign applies for USSB-SC signals and plus sign for the LSSB-SC signals

$\therefore$
$$y(t) = x_c(t)\cdot\cos(\omega_c t+\theta)$$

$$= \frac{A_c}{4}\left[x(t)\left\{\cos(2\omega_c t+\theta)+\cos\theta\right\}\right] \mp \frac{A_c}{4}\left[\left\{\hat{x}(t)\sin 2\omega_c t\cos\theta\right\} - \left\{\hat{x}(t)\overline{1-\cos 2\omega_c t}\sin\theta\right\}\right]$$

So, after lowpass filtering

$$z(t) = \frac{A_c}{4}\left[\cos\theta x(t) \pm (\sin\theta)\hat{x}(t)\right]$$

Note that in this equation, the plus sign applies for USSB-SC.

Taking Fourier transform on both sides

$$Z(f) = \frac{A_c}{4}\left[X(f)\cos\theta \pm (-j\,\mathrm{sgn}\,f)X(f)\sin\theta\right]$$

But
$$\mathrm{sgn}\,f = \begin{cases} 1 & \text{for } f > 0 \\ -1 & \text{for } f < 0 \end{cases}$$

$\therefore$ on simplification,

$$Z(f) = \frac{A_c}{4}X(f)e^{\pm j\theta} \tag{4.67}$$

where, the –ve sign applies for USSB.

Equation (4.67) 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 as far 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.

(ii) 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. 4.57, 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,

$$y(t) = \frac{1}{2}A_c\cos 2\pi(f_c+\Delta f)t\left[x(t)\cos 2\pi f_c t \pm \hat{x}(t)\sin\omega_c t\right]$$

$$= \frac{1}{4}A_c x(t)\left[\cos 2\pi(\Delta f)t + \cos 2\pi(2f_c+\Delta f)t\right] - \frac{1}{4}A_c\hat{x}(t)\left[\sin 2\pi(2f_c+\Delta f)t - \sin 2\pi(\Delta f)t\right]$$

Since the LPF has a cutoff frequency $W << f_c$, we have

$$z(t) = \frac{1}{4}A_c\left[x(t)\cos 2\pi(\Delta f)t \pm \hat{x}(t)\sin 2\pi(\Delta f)t\right] \tag{4.68}$$

(with +ve 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 an SSB-SC signal for which $x(t)$ is the modulating signal and (Δf) is the carrier signal. From Eq. (4.68), 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. 4.58.

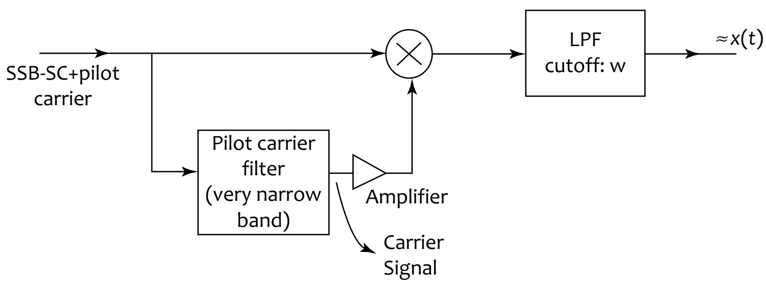


Fig. 4.58 *Homodyne detection*

Equation (4.68) 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. (4.68), 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 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 4.29

A synchronous detection of SSB signal shows phase and frequency discrepancy. Consider

$$s(t) = \sum_{i=1}^{N} \left[\cos\omega_c t \cos(\omega_i t + \phi_i) - \sin\omega_c t \sin(\omega_i t + \phi_i) \right]$$

is an SSB signal. This signal is multiplied by the locally generated carrier $\cos\omega_c t$ and then passed through a lowpass filter.

(a) Prove that the modulating signal can be completely recovered if the cutoff frequency of the filter is $f_N < f_0 < 2f_c$.

(b) Determine the recovered signal when the multiplying 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 $\Delta\omega = 2\pi\Delta f$. *(University Question)*

Synchronous detection is another name for coherent detection. Although it has not been explicitly mentioned, this question assumes that $f_N > f_{N-1} > f_{N-2} \ldots > f_1$, 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 lowpass filtered.

$$s(t)\cos\omega_c t = \sum_{i=1}^{N}\left[\cos^2 \omega_c t\cos(\omega_i t + \phi_i) - \sin\omega_c t\cos\omega_c t\sin(\omega_i t + \phi_i)\right]$$

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 $\frac{1}{2}[\sin 2\omega_c t]$

$$s(t)\cos\omega_c t = \frac{1}{2}\sum_{i=1}^{N}\left\{\cos(\omega_i t + \phi_i)\left[1 + \cos 2\omega_c t\right] - \sin(\omega_i t + \phi_i)\left[\sin 2\omega_c t\right]\right\}$$

When we lowpass filter this using a LPF whose cutoff 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 frequencies close to $2f_c$, they 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,

$$s(t)\cos\omega_c t = \frac{1}{2}\sum_{i=1}^{N}\left[\cos(\omega_i t + \phi_i)\left\{\cos(2\omega_c t + \phi) + \cos\phi\right\} - \left[\sin(\omega_i t + \phi_i)\right]\left\{\sin(2\omega_c t + \phi) - \sin\phi\right\}\right]$$

When this is lowpass filtered by an LPF whose cutoff frequency f_0 is s.t. $f_N < f_0 < 2f_c$

$$z(t) = \text{output of the filter } = \frac{1}{2}\sum_{i=1}^{N}\left\{\cos\phi\cos(\omega_i t + \phi_i) + \sin\phi\sin(\omega_i t + \phi_i)\right\}$$

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

$$z(t) = \text{output of the filter } = \frac{1}{2}\sum_{i=1}^{N}\left[\cos(\Delta\omega)t\cos(\omega_i t + \phi_i) + \sin(\Delta\omega)t\sin(\omega_i t + \phi_i)\right]$$

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 4.30

What is known as a phase-shift SSB-SC demodulator, is shown in the figure. Show that it demodulates an SSB-SC signal.

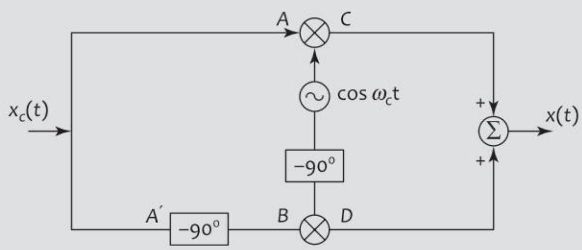


Fig. 4.59 *Phase-shift SSB-SC detector*

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, $x(t)$.

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 HT)

$\therefore$ 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.

4.5.5 Applications of SSB-SC Modulation

From the previous discussion, it is clear that SSB-SC modulation cannot be used for transmission of music and video signals 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 gives saving in power as well as bandwidth and thereby reduces the cost of the transmitter, SSB-SC receivers are quite complex and expensive. It just doesn't make sense to make millions of receivers expensive in order to save a little in 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 9.

4.5.6 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 guardbands in the multiplexed signal. These guardbands facilitate the recovery of the individual messages by making the specifications for the bandpass filters used for recovering 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. 4.60, N sub-carriers, each differing from its adjacent one by 4 kHz, are used. These sub-carriers are LSSB-SC modulated by the telephone message signals. Before modulation, each telephone message is first passed through a lowpass filter to ensure that it is strictly bandlimited to 3.2 kHz. After LSSB-SC modulation, the modulated signal is passed through a bandpass filter. The i th message channel, having a sub-carrier frequency of f_{ci}, will have a BPF whose passband extends from (f_{ci} -4 kHz) to f_{ci}.

In Fig. 4.60, 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 BPF's which separate the various SSB-SC sub-carrier modulated message signals. These are then coherently demodulated using the various sub-carriers and the detected message signals are then passed through LPFs and recovered.

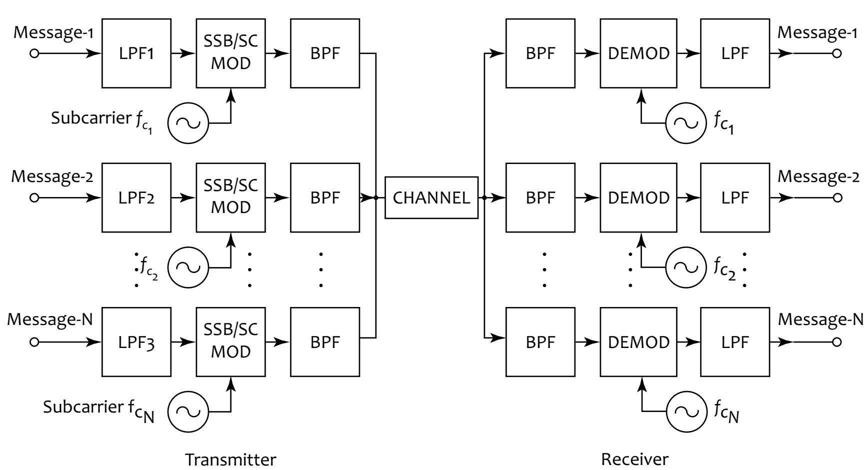


Fig. 4.60 *An FDM system*

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 sub-carriers 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. They are then 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.

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

4.5.7 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-I 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.

4.6 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.

(i) If we employ SSB-SC, the receiver becomes quite complex and expensive, as we have to use coherent detection.
(ii) 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.
(iii) 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.

(iv) Further, the phase response of the bandpass filters used in the filter method will not be linear near the passband edges. This will make the received video signal 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 as 'Vestigial Sideband Modulation; or VSB modulation.

Consider a video signal $x(t)$. Let its spectrum be as shown in Fig. 4.61(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 $y(t) = x(t).\cos 2\pi f_c t$. The spectrum $Y(f)$ of $y(t)$ is given by

$$Y(f) = \frac{1}{2}\left[X(f + f_c) + X(f - f_c)\right] \tag{4.69}$$

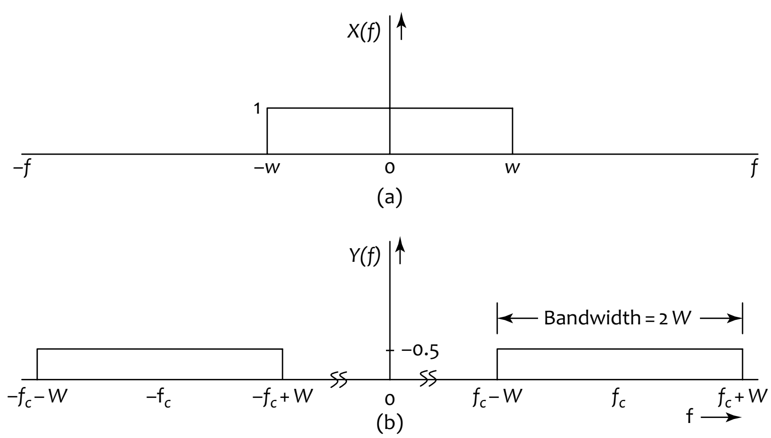


Fig. 4.61 **(a)** *Spectrum of x(t)* **(b)** *Spectrum of y(t)*

$y(t)$ is a DSB-SC signal and its spectrum $Y(f)$ is shown in Fig. 4.61(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).

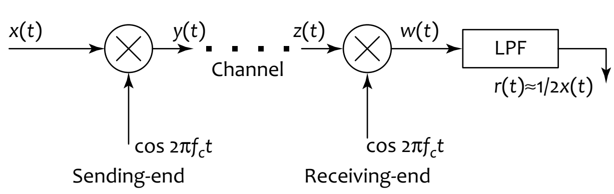


Fig. 4.62 *Coherent reception of y(t)*

$$\therefore \qquad w(t) = y(t).\cos 2\pi f_c t = x(t)\cos^2 2\pi f_c t = \frac{1}{2}x(t)\left[1 + \cos 4\pi f_c t\right]$$

$$\therefore \qquad w(t) = 0.5x(t) + 0.5x(t)\cos 4\pi f_c t$$

$$\therefore \qquad W(f) = 0.5X(f) + \frac{0.5}{2}\left[X(f + 2f_c) + X(f - 2f_c)\right]$$

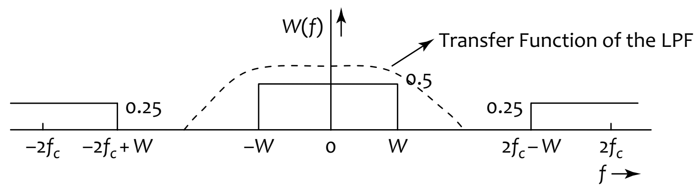


Fig. 4.63 *Spectrum of w(t) and transfer function of the LPF*

As shown in Fig. 4.63, the LPF which has a cutoff 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. 4.61(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. 4.64(a)

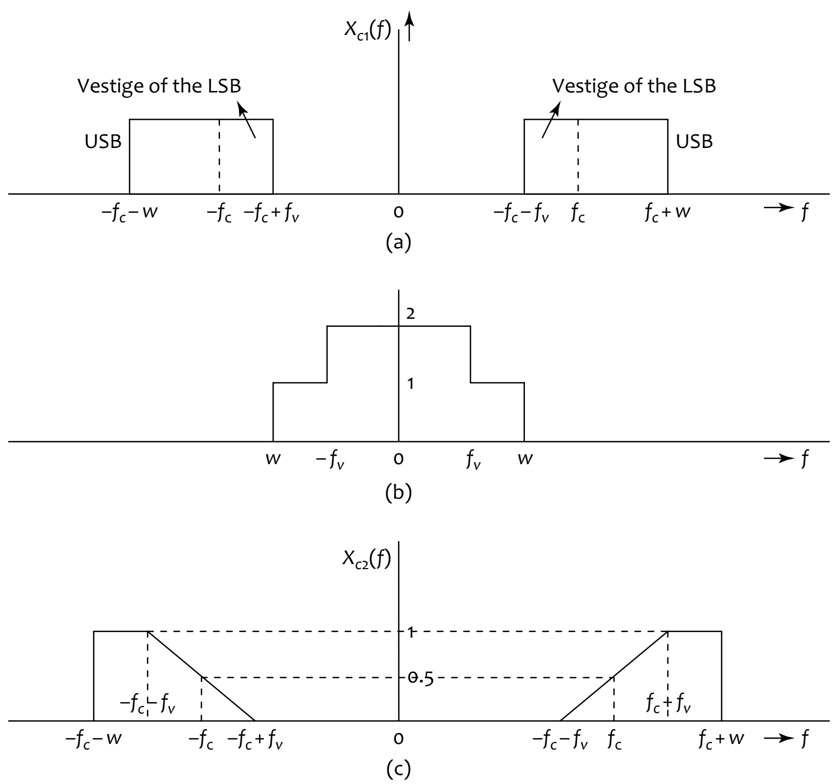


Fig. 4.64 **(a)** *Full upper sideband plus a portion of LSB transmitted* **(b)** *The demodulated signal resulting from* **(a)** **(c)** *USB and LSB suitably shaped to avoid the distortion shown in* **(b)**

Figure 4.64(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 4.63(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. 4.65.

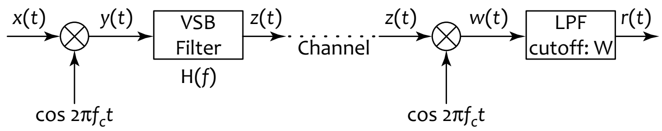


Fig. 4.65 *Generation of VSB signal and recovery of x(t)*

4.6.1 Frequency-domain Representation of a VSB Signal

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

$$Y(f) = \frac{1}{2}\left[X(f+f_c) + X(f-f_c)\right] \tag{4.70}$$

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

$$Z(f) = H(f).Y(f) = \frac{1}{2}H(f)\left[X(f+f_c) + X(f-f_c)\right] \tag{4.71}$$

4.6.2 Transfer Function of the VSB Filter

The spectrum of $w(t)$ is

$$\begin{aligned} W(f) &= \frac{1}{2}H(f)\left[X(f+f_c) + X(f-f_c)\right] * \frac{1}{2}\left[\delta(f+f_c) + \delta(f-f_c)\right] \\ &= \frac{1}{4}H(f+f_c)\left[X(f+2f_c) + X(f)\right] + \frac{1}{4}H(f-f_c)\left[X(f-2f_c) + X(f)\right] \end{aligned}$$

Since the LPF has a cutoff 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 lowpass filtering. We may therefore write the spectrum $R(f)$ of the demodulated signal $r(t)$ as

$$R(f) = \frac{1}{4}\left[H(f+f_c) + H(f-f_c)\right]X(f)$$

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*.

$$\therefore\ H(f+f_c) + H(f-f_c) = 1\ ; \quad -W \le f \le W \tag{4.72}$$

Since the shape of the spectrum of $y(t)$, the DSB-SC signal is as shown in Fig. 4.66(a), it follows that the shape of the transfer function $H(f)$ of the VSB filter should be as shown in Fig. 4.66(b) so that $[H(f).Y(f)]$ which is $Z(f)$, will have the desired shape as shown in Fig. 4.66(c).

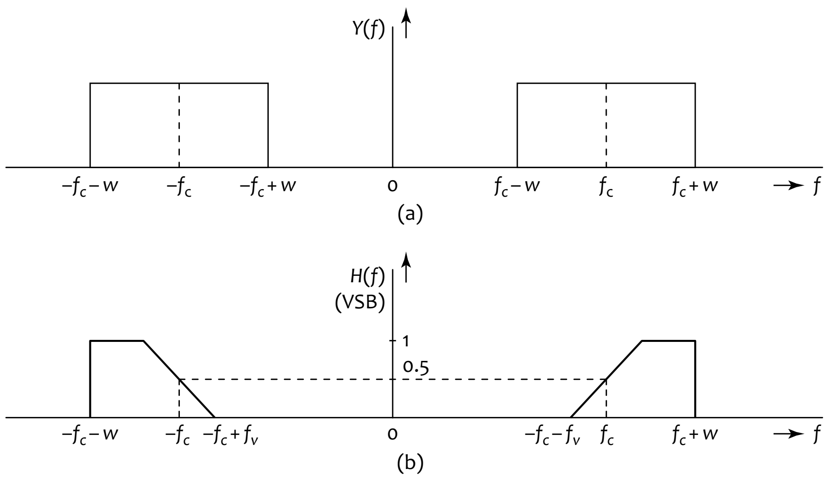


Fig. 4.66 **(a)** *DSB signal y(t)* **(b)** *H(f), the transfer function of the VSB filter (In* **(a)** *the part outlined with dark lines is the spectrum of the VSB signal)*

$H(f+f_c)$ 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 Fig. 4.67(a) and (b).

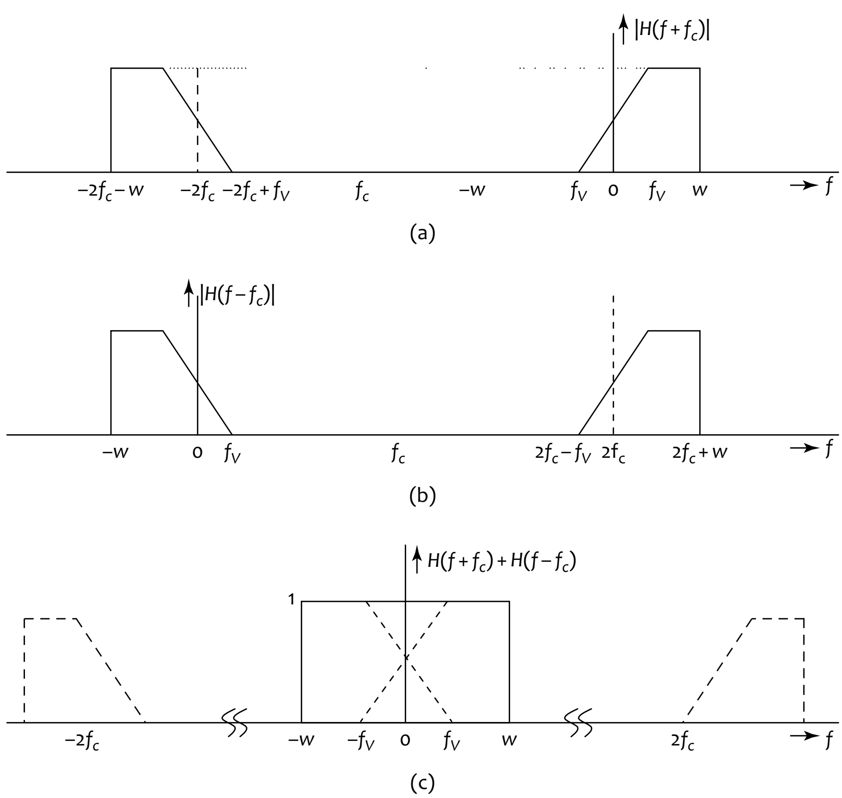


Fig. 4.67 *Sketches of* **(a)** $H(f+f_c)$; **(b)** $H(f-f_c)$; **(c)** $H(f+f_c)+H(f-f_c)$

4.6.3 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.

$$\therefore \quad z(t) = \underset{\text{VSB}}{x_c(t)} = F\left[H(f)\{X(f+f_c)+X(f-f_c)\}\right]$$

$$= \int_{-\infty}^{\infty} H(f)\left[X(f+f_c)+X(f-f_c)\right] e^{j2\pi ft}df$$

$$\therefore \quad \underset{\text{VSB}}{x_c(t)} = \int_{-\infty}^{\infty} H(f)X(f+f_c)\, e^{j2\pi ft}df + \int_{-\infty}^{\infty} H(f)X(f-f_c)\, e^{j2\pi ft}df \tag{4.73}$$

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:

$$H_v(f) = H(f_c) - H(f-f_c) = H(f+f_c) - H(f_c) \quad for \quad -W \le f \le W \tag{4.74}$$

where, $$H(f_c) = H(f)\Big|_{f=f_c}$$

Since $H(f_c) = \frac{1}{2}$, we find from Eq. (4.74) and Fig. 4.66(a) that $H_v(f)$ has a shape as shown in the Fig. 4.68.

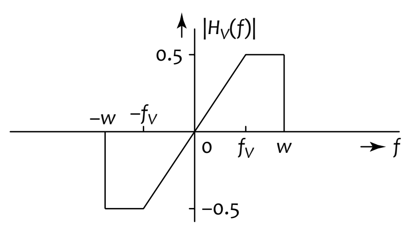


Fig. 4.68 *The function $H_V(f)$ of Eq. (4.74)*

Figure 4.68 clearly brings out the fact that the function $H_v(f)$ is an odd function of frequency,

i.e., $$H_v(-f) = -H_v(f) \tag{4.75}$$

Now, reverting to Eq. (4.73), and making the following substitutions, i.e.,

$$\alpha = (f - f_c)$$

and $$\beta = (f + f_c),$$

we get

$$\underset{\text{VSB}}{x_c(t)} = \int_{-\infty}^{\infty} H(\alpha+f_c)X(\alpha)\, e^{j2\pi(\alpha+f_c)t}d\alpha + \int_{-\infty}^{\infty} H(\beta-f_c)X(\beta)\, e^{j2\pi(\beta-f_c)t}d\beta \tag{4.76}$$

But, from Eq. (4.74), we find that

$$H(\alpha+f_c) = H(f_c) + H_V(\alpha)$$

and $$H(\beta-f_c) = H(f_c) - H_V(\beta)$$

Substituting these in Eq. (4.76), using the fact that $H(f_c) = 0.5$ and simplifying, we get

$$\underset{\text{VSB}}{x_c(t)} = x(t)\cos 2\pi f_c t - g(t)\sin 2\pi f_c t \tag{4.77}$$

where,

$$g(t) \;\underline{\Delta}\; -2j\int_{-\infty}^{\infty} X(f)H_V(f)\, e^{j2\pi ft}df \tag{4.78}$$

As is to be expected, if $f_V \to 0$, $x_{c_{\text{VSB}}}(t) \to x_{c_{\text{SSB-SC}}}(t)$

i.e., $g(t) \to \hat{x}(t)$ since $H_V(f) \to \text{sgn}(f)$ as $f_V \to 0$ (see Fig. 4.68)

4.6.4 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 Fig. 4.66(a) and (b).

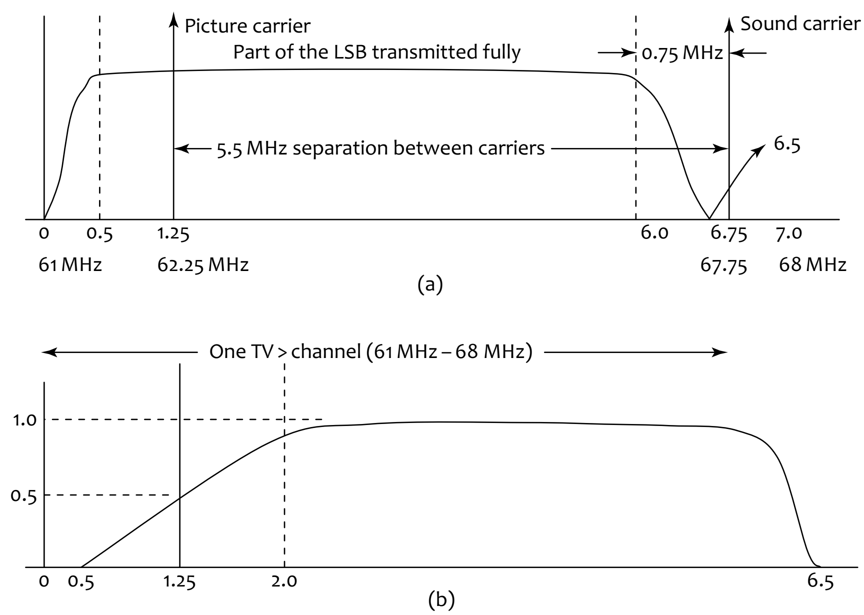


Fig. 4.69 **(a)** *Spectrum of the transmitted TV signal (CCIR-B, Monochrome)* **(b)** *Typical response characteristic of the video amplifier in the receiver*

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. 4.69(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. 4.69(a).

The fully transmitted part of the lower sideband leads to the effect shown in Fig. 4.64(b). This distortion is avoided by shaping the *response characteristic of the receiver* as shown in Fig. 4.69(b), on the lines suggested in Fig. 4.64(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. 4.64, only one-sided spectrum and response are shown in Fig. 4.69(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 is zero for these frequencies.

4.6.5 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. (4.77) by a factor m ($0 < m < 1$), the modulation index and then add the carrier term $\cos 2\pi f_c t$, to get

$$s(t) = [1 + mx(t)]\cos 2\pi f_c t - mg(t)\sin 2\pi f_c t \tag{4.79}$$

As the envelope detector extracts the envelope of this signal given to it as input, the detector output is (see Eq. 2.171)

$$a(t) = \left\{[1 + mx(t)]^2 + [mg(t)]^2\right\}^{1/2}$$

$$\therefore \quad a(t) = [1 + mx(t)]\left\{1 + \left[\frac{mg(t)}{1 + mx(t)}\right]^2\right\}^{1/2} \tag{4.80}$$

$[1 + mx(t)]$ 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. 4.69(a)].

Table 4.2 *Shows a comparison of various varieties of amplitude modulation*

System	*Useful part of transmitted power*	*B.W*	*Carrier suppression*	*Sideband suppression*	*Figure of merit*	*Receiver complexity*	*Application*
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 fully suppressed	1	Complex	Used for long-haul point-to-point communication
VSB	Moderate	B.W < 2W	No	One sideband is fully transmitted while the other is partially transmitted	–	Simple	TV broadcasting

SUMMARY

1. *Modulation* is the process of translating a low-frequency information-bearing signal to a high-frequency slot.
2. Modulation is necessary for (i) keeping the antenna size small, (ii) making it possible for the receiver to select the desired message signal, and (iii) multiplexing and transmitting several information-bearing signals simultaneously.
3. 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.
4. **Amplitude Modulation** (carrier plus both sidebands), **i.e., 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 message/modulating signal.
5. **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

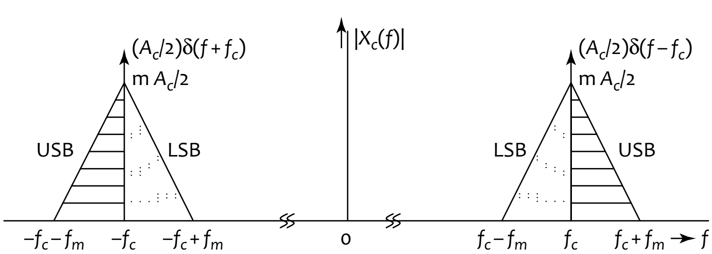
$$x_c(t) = A_c\left[1 + m\cos\omega_m t\right]\cos\omega_c t = A_c\cos\omega_c t + \frac{mA_c}{2}\cos(\omega_c + \omega_m)t + \frac{mA_c}{2}\cos(\omega_c - \omega_m)t$$

6. **Frequency-domain description of AM**

$$X_c(f) = \frac{A_c}{2}\left[\delta(f + f_c) + \delta(f - f_c)\right] + \frac{mA_c}{2}\left[X(f - f_c) + X(f + f_c)\right]$$

where, for single-tone modulation, $X(f) = \frac{1}{2}\left[\delta(f + f_m) + \delta(f - f_m)\right]$

7. **Amplitude spectrum of an AM signal**



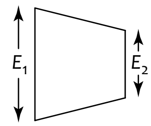
8. **Carrier and sideband power components in AM**
   (i) When a general message signal $x(t)$, with $|x(t)| \le 1$ is used,

$$\overline{x_c^2(t)} = \text{Average power in an AM signal } = \frac{1}{2}A_c^2\left[1 + m^2\overline{x^2(t)}\right]$$

   where, $\overline{x_c^2(t)}$ = Average power of the message signal.
   (ii) For single-tone modulation,

$$\overline{x_c^2(t)} = \text{Average power in an AM signal } = \frac{1}{2}A_c^2\left[1 + \frac{m^2}{2}\right]$$

9. **Trapezoidal pattern** When $0 \leq m \leq 1$ and there is no distortion.



$$m = \left(\frac{E_1 - E_2}{E_1 + E_2} \right)$$

10. **Generation of AM**
    (i) By the use of non-linear devices
    (ii) By the use of product devices
    (iii) By the use of switching devices
11. **Plate/collector—modulated class-C amplifier**
    Total average power in the AM output signal $= P_0 = \eta P_{Tav}$

    $= \eta P_B + \eta P_B \left(\frac{m^2}{2} \right)$; where, P_{Tav} = 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.

    $\therefore$ P_0 = carrier power + total sideband power

    Carrier power is supplied by E_{bb}/V_{cc} supply and sideband power is supplied by the final stage of the modulating amplifier.
12. **High-level modulation and low-level modulation** In an AM transmitter, if the modulation of the carrier is carried out at a high power level, using plate/collector modulated class-C amplifiers, it is called high-level modulation. If the modulation is carrier out at any point before the plate/collector of the final carrier power amplifier stage, it is called low-level modulation.
13. **Advantages and disadvantages of high-level modulation** As modulation is performed only at the plate/collector of the final power amplifier stage of the carrier chain, there is no need to use class-A or class-AB tuned power amplifiers whose efficiency is low.

    However, high-level modulation requires large amounts of modulating signal power since the entire sideband power of the AM signal to be radiated, has to be supplied by the final stage of the modulating signal amplifier chain.
14. **Detection of AM signals** An AM signal may be detected by (i) coherent detection, (ii) square-law detection, or (iii) envelope detection.
15. **Coherent detector** The received signal is multiplied by a locally generated carrier signal and the product is lowpass filtered using an LPF with a cutoff frequency of W Hz, the baseband signal bandwidth.
16. **Square-law detector** The received AM signal is fed to a square law device and then its output is lowpass filtered using an LPF with a cutoff frequency of W Hz.
17. **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

$$R_s C << \frac{1}{f_c} << R_L C << \frac{1}{f_m}$$

where, R_s is the source resistance, and f_m is the highest modulating signal frequency.

18. **Distortions in envelope detection**

(i) *Diagonal clipping* To avoid this, it must be ensured that

$$R_L C \le \frac{1}{\omega_m}\left[\frac{\sqrt{1-m^2}}{m}\right]$$

(ii) *Negative peak clipping* To avoid this, it must be ensured that

$$m_{\max} \le \left(\frac{R_{AC}}{R_{DC}}\right);\ R_{AC} = \text{ac load res. of the envelope detector}$$

$$R_{DC} = \text{dc load res. of the envelope detector}$$

19. **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.*
20. **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: $\underset{\text{DSB-SC}}{x_c(t)} = \frac{A_c}{2}\left[\cos(\omega_c+\omega_m)t+\cos(\omega_c-\omega_m)t\right]$

21. **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.
22. DSB-SC signals can be detected only by synchronous or coherent detection only.
23. 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.
24. **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.
25. **Single sideband suppressed carrier (SSB-SC) modulation** is an amplitude modulation process in which the carrier as well as one of the sidebands is suppressed and only one sideband is transmitted.
26. **Frequency domain representation of SSB-SC signals**

USSB-SC signal: $X_C^U(f) = \frac{A_c}{2}\left[X_+(f-f_c)+X_-(f+f_c)\right];\quad \begin{matrix} X_+(f)=F[x_+(t)] \\ x_+(t)=x(t)+j\hat{x}(t)\end{matrix}$

LSSB-SC signal: $X_C^L(f) = \frac{A_c}{2}\left[X_-(f-f_c)+X_+(f+f_c)\right];\quad \begin{matrix} X_-(f)=F[x_-(t)] \\ x_-(t)=x(t)-j\hat{x}(t)\end{matrix}$

27. **Time-domain representation of SSB-SC signals**

$x_C^U(t) =$ USSB-SC signal $= \frac{A_c}{2}\left[x(t)\cos\omega_c t - \hat{x}(t)\sin\omega_c t\right]$

and $x_C^L(t) =$ LSSB-SC signal $= \frac{A_c}{2}\left[x(t)\cos\omega_c t + \hat{x}(t)\sin\omega_c t\right]$

28. **Generation of SSB-SC signals** There are three methods.
    (i) *Filter method* In this method, first a DSB-SC signal is generated. From this, the unwanted sideband is suppressed using a filter. Advantages: very stable; used in commercial circuits.
    (ii) *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 are fed with $A_c \sin \omega_c t$ and $\hat{x}(t)$. Its output is the product of these two. Outputs of BM_2 are either added or subtracted from the output of BM_1. Addition gives LSSB-SC while subtraction gives USSB-SC signal. Advantages: used by radio amateurs, needs frequent adjustment.
    (iii) *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.
29. **Detection of SSB-SC signals**
    (i) 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.
    (ii) If the locally generated carrier has a frequency error (Δf), then the detector output will not be $x(t)$; instead, it will be a SSB-SC signal with $x(t)$ as modulating signal and (Δf) as the carrier.
30. **Applications and advantages of SSB-SC transmission**
    (i) Useful for point-to-point communication for speech but not for audio broadcasting
    (ii) Bulk transmission of telephone conversations using FDM
    (iii) Gives better (S/N) at the destination as compared to AM
31. **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.
32. **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

33. **Independent sideband transmission** It is a variant of SSB-SC transmission in which two sidebands are transmitted without the carrier, with the two sidebands carrying different speech signals.
34. **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 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 READING

1. Simon Haykin: *Communication Systems*, Third Edition, John Wiley & Sons, 1995.
2. John G Proakis and M Salehi: *Communication Systems Engineering*, Second Edition, Prentice-Hall of India, Indian Reprint, 2006.

3. Leon W Couch II: *Digital and Analog Communication Systems*, Sixth Edition, Pearson Education Publication.
4. A Bruce Carlson, Paul B Crilly and Janet C Rutledge: *Communication Systems*, Fourth Edition (International edition), McGraw-Hill, Singapore, 2002.

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, $A_m < A_c$. (i) Write down the expression for the modulated signal. (ii) Write down the expression for the carrier component and the side-frequency components. (iii) 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 fro the r.m.s. 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. Discuss the advantages and disadvantages of SSB-SC transmission.
15. In the filter method of generation of an SSB-SC signal, why do we have to use initially a low-frequency carrier?
16. 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?
17. How is an SSB-SC signal demodulated?
18. With reference to SSB-SC signal modulation, discuss the effect of an error in the locally generated carrier signal's (i) frequency, and (ii) phase.
19. State the applications of SSB transmission.
20. Draw the spectrum of an LSSB-SC signal. Write down an expression for this spectrum in terms of that of the message signal.
21. How does the 2-stage SSB-SC modulation overcome the problems associated with the design of the sideband suppression filter?
22. Critically compare the filter method and the phasing method of generation of SSB-SC signals.
23. With the help of block schematic diagram, clearly explain homodyne detection of an SSB signal transmitted with a pilot carrier.
24. Explain briefly the basic principle of FDM.
25. What is independent sideband transmission?
26. Explain why SSB transmission even with a pilot carrier is not feasible in the case of TV.
27. Sketch the typical spectrum of the VSB signal that is given as input to the video detector of a TV receiver.

28. Write down an expression for the time-domain representation of a VSB signal.
29. Sketch the spectrum of typical TV signal.
30. Sketch the typical response characteristic of the video amplifier section of a TV receiver.
31. 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?

FILL IN THE BLANKS

1. A carrier of frequency f_c is amplitude modulated (AM) using a message signal of frequency f_m. The frequencies of the two side-frequencies are _______ and _______, and the bandwidth of the modulated signal is ________ Hz.
2. A carrier $A_c \cos \omega_c t$ is amplitude modulated by a message signal $x(t)$, depth of modulation being m. If $|x(t)| \le 1$, the expression for the modulated signal in the time domain is ___________; and in the frequency domain it is ___________. The envelope of the modulated signal is given by __________.
3. A single-tone message signal amplitude modulates a carrier and the average power in the modulated signal is 118 watts, when the depth of modulation is 0.6. The carrier power is ________ watts.
4. The source of power for the carrier component of the power in the output modulated signal of a collector-modulated class-C amplifier is _________ and that of the sidebands component is ___________.
5. The two important distortions that can appear in the demodulated output of an envelope detector are __________ and ____________.
6. Out of these two distortions, ____________ is due to the rate of decrease of voltage across the capacitor-resistor parallel combination being slower compared to the rate of decrease of the envelope, and the _________ arises due to the ac load of the detector diode being not equal to its dc load.
7. A DSB-SC signal may be generated by using a _________ modulator.
8. Quadrature-carrier multiplexing permits the transmission of _______ (two, four) different messages simultaneously using a single carrier frequency.
9. SSB-SC transmission is not used for audio broadcasting because __________.
10. In commercial point-to-point communications, for the generation of SSB-SC signal, the __________ (filter, phasing) method is used.
11. The sideband filter design is made very simple if we use the _________ SSB-SC modulation.
12. In the ________ method of generation of SSB-SC signal, it is quite easy to change from LSSB-SC to USSB-SC.
13. Phasing method uses _________ (wideband, single-frequency) phase shifters whereas Weaver's method uses ________ (wideband, single-frequency) phase-shifters.
14. The time-domain representation of a USSB_SC signal is $x_c^u(t) =$ ______________.
15. The frequency-domain representation of a USSB-SC signal is $x_c^u(t) =$ __________.
16. The filter used in the filter method of generation of SSB-SC signals can have a transition bandwidth of not more than ___________ Hz.
17. The filter method of generation of SSB-SC signals can be used for speech signals but not for music because _________.
18. In an FDM system (used for multiplexing telephone channels) producing the basic group, the separation in frequency between adjacent sub-carriers is ___________ Hz.
19. A TV transmitter transmits the full _________ sideband and a vestige of the ________ sideband.
20. The VSB shaping filter is used in the TV __________ (Transmitter, Receiver).
21. The TV receivers use __________ detectors for the video.

MULTIPLE CHOICE QUESTIONS

1. In an amplitude modulated wave obtained by sinusoidal modulation of the carrier, the positive peak amplitude of the r.f. 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 r.m.s. 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 (r.m.s. value) is also halved. The type of modulation used is
(a) AM (carrier plus both sideband) (b) Single sideband plus carrier
(c) SSB-SC (d) Vestigial sideband
8. 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
9. 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 signal 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
10. 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.
11. 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
12. Vestigial sideband modulation is generally used for
(a) TV broadcasting (b) point-to-point communications
(c) telemetering (d) stereo broadcasting

13. 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
14. An SSB-SC signal may be demodulated using a
    (a) diode envelope detector
    (b) synchronous detector
    (c) ratio detector
    (d) none of these

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 dc power for the two amplifiers, (b) the dissipation in power 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 $e_0 = a_1 e_{in} + a_2 e_{in}^2$. 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 bandpass filtering. What should be the centre 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 $e_{in}(t) = x_c(t) = A_c[1 + mx(t)]\cos 2\pi f_c t$ (i) Determine $e_0(t)$. (ii) 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. In a quadrature-carrier multiplex system using a carrier frequency f_c and two message signals $x_1(t)$ and $x_2(t)$, each of bandwidth 0 to W Hz, the multiplexed signal is transmitted to the receiver through a communication channel whose transfer function is $H(f)$. Show that the $H(f)$ should satisfy the condition
   $$H(f_c + f) = H^*(f_c - f) \quad \text{for} \quad 0 \le f \le W$$
   for the receiver to recover the two message signals.
9. 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 $x_c(t)$.

10. 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?
11. 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.
12. 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 watt 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 $m = 0.8$
13. A two-stage SSB-SC modulator is shown in Fig. 4.52. 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.
14. The carrier component is added to a USSB-SC signal and the resulting signal is

$$x_c(t) = A_c \cos 2\pi f_c t + x(t)\cos\omega_c t - \hat{x}(t)\sin\omega_c t$$

    where $x(t)$ is the message signal and the $\hat{x}(t)$, its Hilbert transform. Is it possible to use an envelope detector to recover a reasonably good approximation of $x(t)$ from this modulated signal, $x_c(t)$? If it is possible what are the conditions to be satisfied?
15. Repeat Problem 13 if the final output signal of the 2-stage SSB-SC modulator is to be an LSSB-SC signal.
16. Equation (4.63) 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.
17. A scrambler is a system used for privacy of communication. In the 2-stage SSB-SC generator of Fig. 4.48, assume that the message signal has an amplitude spectrum as shown in Fig. P-4.1; 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 lowpass 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.

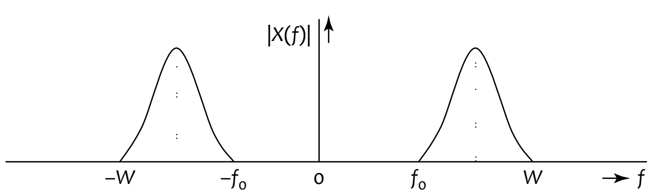


Fig. P-4.1

Key to Multiple Choice Questions

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

5

Angle Modulation

By going through this chapter, the student

- *understands the concept of angle modulation and learns the difference between frequency modulation and phase modulation*
- *learns to derive expressions for frequency-modulated waves and phase-modulated waves*
- *learns the way the modulation index and deviation ratio are defined for FM*
- *can 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*
- *can 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*
- *can draw the block diagram of a superheterodyne FM broadcast receiver and explain the function of each block*
- *can 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.

(i) As we are going to see in Chapter 9, 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.

(ii) 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.

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

$$x_c(t) = A_c \cos \theta(t)$$

$$\theta(t) = \omega_c t + \phi(t) \tag{5.2}$$

$$\therefore \quad x_c(t) = A_c \cos\left[\omega_c(t) + \phi(t)\right] \tag{5.3}$$

$\phi(t)$, 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, $k_p \;\Delta\;$ phase-deviation constant

It represents the change in phase angle per unit amplitude of the modulating signal $x(t)$ and has the unit of radians per volt.

The phase-modulated signal may therefore be written as

$$\underset{P.M}{x_c(t)} = A_c \cos\left[\omega_c t + k_p x(t)\right] \tag{5.5}$$

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

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = \frac{1}{2\pi}\frac{d}{dt}\left[\omega_c t + \phi(t)\right] = f_c + \frac{1}{2\pi}\frac{d}{dt}\phi(t) \tag{5.6}$$

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

$$\frac{1}{2\pi}\frac{d}{dt}\phi(t) = k_f x(t) \tag{5.7}$$

$\therefore$ $$f_i(t) = f_c + k_f x(t) \tag{5.8}$$

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 $$\theta(t) = \int_0^t 2\pi f_i(\alpha)d\alpha + \phi_0 \tag{5.9}$$

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

$$\underset{\text{F.M}}{x_c(t)} = A_c \cos\left[\int_0^t 2\pi f_c d\alpha + \int_0^t 2\pi k_f x(\alpha)d\alpha\right]$$

Thus,
$$\underset{\text{F.M}}{x_c(t)} = A_c \cos\left[\omega_c t + 2\pi k_f \int_0^t x(\alpha)d\alpha\right] \tag{5.10}$$

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.

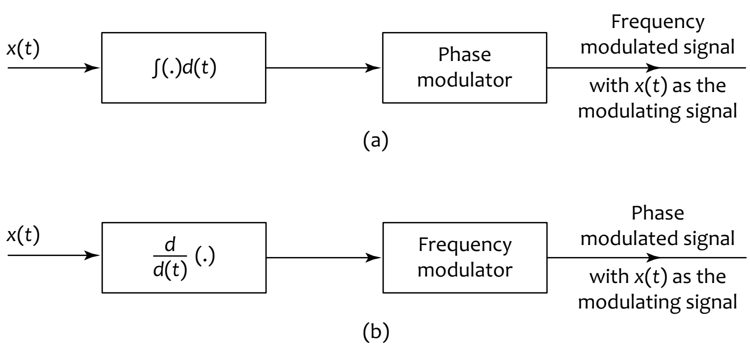


Fig. 5.1 **(a)** *Frequency modulation using a phase modulator*
(b) *Phase modulation using a frequency modulator*

Example 5.1

An angle-modulated signal is given by

$$x_c(t) = 6\cos[2\pi \times 10^7 t + 0.2 \ \sin(10^4)\pi t]$$

(i) If $x_c(t)$ is a peak modulated signal with $k_p = 5$ rad/volt; and
(ii) 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).

$$\underset{\text{F.M}}{x_c(t)} = A_c \cos[\omega_c t + 2\pi k_f \int_0^t x(\alpha)d\alpha] \quad \text{from Eq. (5.10)}$$

$$\underset{\text{P.M}}{x_c(t)} = A_c \cos[\omega_c t + k_p x(t)] \quad \text{from Eq. (5.5)}$$

(i) *For PM*: Compare the above equation for $\underset{\text{P.M}}{x_c(t)}$ with the given equation of the angle-modulated signal. If we take $x(t) = A_m \sin 10^4 \pi t$, it means that $k_p.A_m = 0.2$

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

$$x(t) = 0.04 \sin 10^4 \pi t$$

(ii) *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 $A_m \cos 2\pi \times 5 \times 10^3 t$.

$$\therefore \quad 2\pi k_f \int_0^t A_m \cos 10^4 \pi\alpha d\alpha = \left(\frac{k_f A_m}{f_m}\right) \sin 10^4 \pi t, \text{ where } f_m = 5 \times 10^3$$

From the expression for $x_c(t)$, we therefore have

$$\frac{k_f A_m}{f_m} . \sin 10^4 \pi t = 0.2 \sin 10^4 \pi t$$

$$\therefore \quad \frac{k_f A_m}{f_m} = 0.2.$$

Substituting values of k_f and f_m, 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 Fig. 5.2 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?

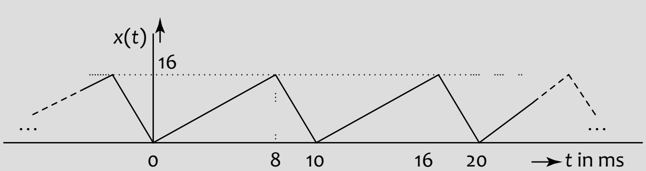


Fig. 5.2 *Signal for Example 5.2*

The modulated signal $x_c(t)$ is given by

$$x_c(t) = A_c \cos(\omega_c t + k_p x(t))$$

$\therefore$ instantaneous frequency,
$$f_i = \frac{1}{2\pi} \frac{d}{dt}\left[\omega_c t + k_p x(t)\right]$$

$$\therefore \quad f_i = f_c + \frac{1}{2\pi} k_p \frac{d}{dt} x(t)$$

$$\therefore \quad (f_i - f_c) = \text{maximum frequency deviation}$$

$$= \frac{1}{2\pi} k_p \left|\frac{d}{dt} x(t)\right|_{\max}$$

$$\therefore \qquad \left|\frac{d}{dt}x(t)\right|_{\max} = \frac{16}{2\times10^{-3}} \quad \text{(from the waveform of } x(t))$$

$$= 8000 \text{ v/sec}$$

$\therefore$ maximum frequency deviation $= \dfrac{1}{2\pi}.k_p.8000 = 80\times10^3$

$$\therefore \qquad k_p = \frac{80\times10^3\times2\pi}{8000} = 20\pi \text{ rad/volt.}$$

From $t = 0$ to $t = 8$ m.s $\quad \dfrac{d}{dt}x(t) = 2v/\text{m.s} = 2000\, v/\text{s}$

$\therefore$ during this period, frequency deviation

$$= \frac{1}{2\pi}k_p\frac{d}{dt}x(t) = \frac{1}{2\pi}\times20\pi\times2000 = 20 \text{ kHz}$$

Hence, from 0 ms to 8 ms, the frequency of the modulated signal is

$$1000 \text{ kHz} + 20 \text{ kHz} = 1020 \text{ kHz}$$

From 8 ms to 10 ms, the frequency deviation is negative and has a value of 80 kHz. Hence, during this period the frequency of the modulated wave is

$$1000 \text{ kHz} - 80 \text{ kHz} = 920 \text{ kHz}$$

The frequency of the modulated signal varies between 920 kHz and 1020 kHz.

5.2.3 Angle-Modulated Signals for Some Simple Modulating Waveforms

(i) Sinusoidal Modulating Signal

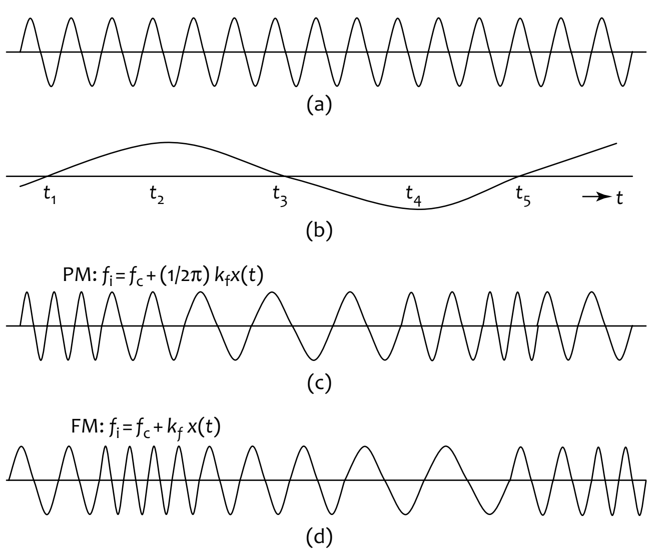


Fig. 5.3 **(a)** *The carrier signal* **(b)** *The modulating sinusoidal signal* **(c)** *The phase-modulated signal* **(d)** *The frequency-modulated signal*

(ii) A Unit-step Function

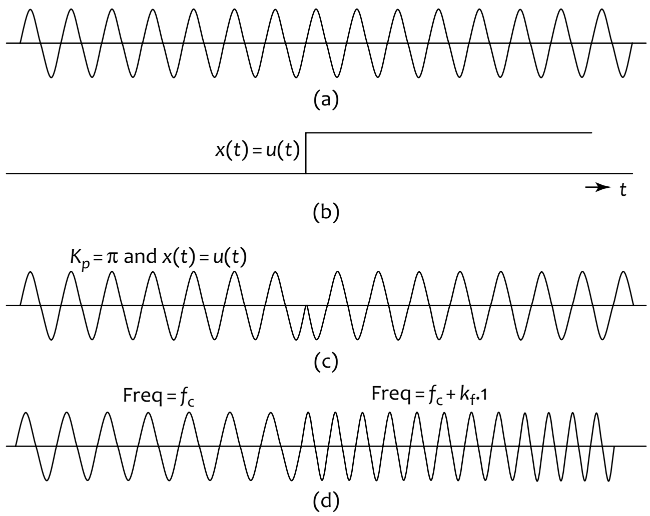


Fig. 5.4 **(a)** *The carrier signal* **(b)** *The unit-step modulating signal* **(c)** *The phase-modulated signal* **(d)** *The frequency- modulated signal*

5.2.4 Modulation Indices for FM and PM

(i) For a Single-tone Message Signal

Let
$$x(t) = A_m \cos(2\pi f_m t) \tag{5.12}$$

Then from Eq. (5.11) we have, for PM
$$\phi(t) = k_p x(t) = k_p A_m \cos(2\pi f_m t) \tag{5.13}$$

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

PM:
$$x_c(t) = A_c \cos\left[2\pi f_c t + k_p A_m \cos(2\pi f_m t)\right] \tag{5.15}$$

FM:
$$x_c(t) = A_c \cos\left[2\pi f_c t + \frac{k_f A_m}{f_m}\sin(2\pi f_m t)\right] \tag{5.16}$$

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

$$\underset{\text{PM}}{x_c(t)} = A_c \cos\left[2\pi f_c t + \beta_p \cos(2\pi f_m t)\right] \tag{5.19}$$

$$\underset{\text{FM}}{x_c(t)} = A_c \cos\left[2\pi f_c t + \beta_f \sin(2\pi f_m t)\right] \tag{5.20}$$

Note:

1. 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.
2. 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*.

(ii) For a General Modulating Signal Having seen the physical meaning of the modulating 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:

$$\beta_p \triangleq k_p \max\left[|x(t)|\right] = (\Delta\phi)_{\max} \tag{5.21}$$

and

$$\beta_f \triangleq \frac{k_f \max\left[|x(t)|\right]}{W} = \frac{(\Delta f)_{\max}}{W} \tag{5.22}$$

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 rms value 2 volts and a frequency of 1 kHz, is applied. Determine the peak frequency deviation and the deviation ratio.

Peak amplitude of the modulating signal = $2\sqrt{2}$ volts

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

$$x_c(t) = 10\cos\left[2\pi \times 10^8 t + 5\sin 2\pi \times 200t\right]$$

Determine (i) the carrier frequency, (ii) the modulating signal frequency, (iii) the peak frequency deviation, and (iv) the modulation index.

(i) $f_c = 100$ MHz $= 10^8$ Hz

(ii) $f_m = 200$ Hz

(iii) $\beta_f = \dfrac{\text{Peak freq. deviation}}{\text{Modulating signal frequency}} = 5$

$\therefore$ peak frequency deviation $= 5 f_m = 5 \times 200 = 1$ kHz.

(iv) Modulation index $\beta_f = 5$ as stated in (iii)

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? (b) 20 kHz?

(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.

From Carson's rule, BW $= 2(\beta_f + 1) f_m = 50 \times 10^3$

$$\therefore \quad 2\beta_f + 2 = \frac{50 \times 10^3}{f_m}, \text{ but } \beta_f = \frac{\Delta f}{f_m} = \frac{15 \times 10^3}{f_m}$$

$$\therefore \quad \frac{2 \times 15 \times 10^3}{f_m} + 2 = \frac{50 \times 10^3}{f_m}$$

Multiplying throughout by f_m, we get

$$2 f_m = 50 \times 10^3 - 30 \times 10^3 = 20 \times 10^3$$

$$\therefore \quad f_m = 10 \times 10^3 = 10\,\text{kHz}$$

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 (i) an FM system with $k_f = 12$ kHz/volt, and (ii) a PM system with $k_p = 1.0$ rad/volt.

(i) $\beta_f =$ Modulation index $= \left(\dfrac{k_f \cdot A_m}{f_m}\right) = \dfrac{12 \times 10^3 \times 5}{10^4} = 6$

$\therefore$ bandwidth $B_T = 2(k_f \cdot A_m + f_m) = 2(\beta_f + 1) f_m$ (Carson's rule)

$$= 2 \times 7 \times 10 \times 10^3 = 140\,\text{kHz}$$

(ii) β_p = Modulation index = $k_p.A_m$ = 1 × 5 = 5

$\therefore$ bandwidth $B_T = 2(k_f\,.\,A_m + 1)f_m = 2(\beta_p + 1)f_m$

$$= 2 \times 6 \times 10^4 = 120\,\text{kHz}$$

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?

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 = $x_c(t) = A_c \cos[\omega_c t + \phi(t)]$, the instantaneous frequency f_i is given by

$$f_i = \frac{1}{2\pi}\frac{d}{dt}[\omega_c + \phi(t)] = f_c + \frac{1}{2\pi}\frac{d}{dt}[\phi(t)]$$

$$\therefore \qquad f_i = f_c + \frac{1}{2\pi}\frac{d}{dt}[12 \sin(4\pi \times 10^3)t] = f_c + 24 \times 10^3 \cos(4\pi \times 10^3)t$$

$\therefore$ 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

(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.

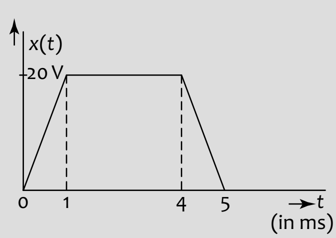


Fig. 5.5 *Signal for Example 5.9*

(a) Instantaneous frequency $f_i = f_c + k_f.x(t)$

$\therefore$ maximum instantaneous frequency = $(f_i)_{\max}$

$$= f_c + k_f \cdot [x(t)]_{\max} = 2 \times 10^6 + 4 \times 10^3 \times 20 = 2.08 \text{ MHz}$$

(b) Instantaneous frequency $f_i = f_c + \frac{1}{2\pi}\frac{d}{dt}[\phi(t)]$ where, $\phi(t)$ = 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, $\frac{d}{dt}x(t) = \frac{20}{1\times 10^{-3}} = 20{,}000$ in the interval 0 to 1 ms.

It is zero from 1 ms to 4 ms and –20,000 from 4 ms to 5 ms. Beyond 5 ms, it is always zero. Hence,

$$\left[\frac{d}{dt}x(t)\right]_{\max} = 20{,}000$$

$$\therefore \qquad (f_i)_{\max} = f_c + \frac{k_p}{2\pi}\left[\frac{d}{dt}x(t)\right]_{\max} = 2\times 10^6 + \frac{4}{2\pi}\times 20\times 10^3 = 2012.74 \text{ kHz}$$

$$= 2.012746 \text{ MHz}$$

This value is obtained in the interval 0 ms to 1 ms.

Example 5.10

A particular modulated signal is given by

$$x_c(t) = 2\cos\omega_c t + 0.4\cos 2\pi f_m t.\sin\omega_c t$$

Comment on the nature/type of modulation.

$$x_c(t) = 2\cos\omega_c t + 0.4\cos 2\pi f_m t.\sin\omega_c t$$

$$= \sqrt{2^2 + (0.4\cos 2\pi f_m t)^2}\cos[\omega_c t + \theta(t)]$$

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 tan $\theta \approx \theta$ 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

$$\sqrt{1+x} \approx \left(1 + \frac{1}{2}x\right) \textit{ if } x << 1$$

$$\therefore \qquad x_c(t) = 2[1 + \frac{0.08}{2}\cos^2 2\pi f_m t]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

$$= 2\left[1 + \frac{0.08}{2}\{1 + \cos 4\pi f_m t\}\right]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

$$= 2[1.02 + 0.02\cos 4\pi f_m t]\cos[\omega_c t - 0.2\cos 2\pi f_m t]$$

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. In what follows, we will be examining this aspect.

Referring to Eq. (5.3), we know that an angle-modulated signal could be represented by

$$x_c(t) = A_c \cos\left[\omega_c(t) + \phi(t)\right]$$

where, as stated in Eq. (5.11),

$$\phi(t) = k_p x(t) \text{ for PM}$$

and

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha)d\alpha \text{ for FM}$$

Expanding the above equation for $x_c(t)$, we get

$$x_c(t) = A_c\left[\cos\omega_c(t)\cos\phi(t) - \sin\omega_c(t)\sin\phi(t)\right] \tag{5.23}$$

Now, if $\phi(t)$ is quite small, say $\phi(t) \le 0.2$ radian, 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

$$x_c(t) = A_c\left[\cos\omega_c(t) - \phi(t)\sin\omega_c(t)\right] \tag{5.24}$$

Let us now consider single-tone modulation and let

$$x(t) = A_m \cos\omega_m t$$

5.3.1 In the Case of Phase Modulation

For this case $\phi(t) = k_p x(t)$

Hence, referring to Eq. (5.24), we have

$$\therefore \quad x_c(t) = A_c\left[\cos\omega_c(t) - \phi(t)\sin\omega_c(t)\right] = A_c\left[\cos\omega_c(t) - k_p A_m \cos\omega_m(t)\cdot\sin\omega_c(t)\right]$$

But

$$\cos\omega_m(t)\cdot\sin\omega_c(t) = \frac{1}{2}\left[\sin(\omega_c + \omega_m)t + \sin(\omega_c - \omega_m)t\right]$$

$$\therefore \quad x_c(t) = A_c\cos\omega_c(t) - \frac{A_c k_p A_m}{2}\left[\sin(\omega_c + \omega_m)t + \sin(\omega_c - \omega_m)t\right]$$

$$\therefore \quad \underset{\text{NBPM}}{x_c(t)} = A_c\cos\omega_c(t) + \frac{k_p A_c A_m}{2}\left[\cos\{(\omega_c + \omega_m)t + \pi/2\} + \cos\{(\omega_c - \omega_m)t + \pi/2\}\right] \tag{5.25}$$

The above equation 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.

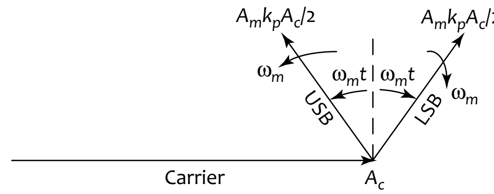


Fig. 5.6 *Phasor diagram of a narrowband phase modulated signal*

5.3.2 In the 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

$$x_c(t) = A_c\left[\cos\omega_c(t) - \left(\frac{k_f A_m}{f_m}\right)\sin\omega_m(t)\sin\omega_c(t)\right]$$
$$= A_c\cos\omega_c(t) + \frac{A_c}{2}\left(\frac{k_f A_m}{f_m}\right)\left[\cos(\omega_c+\omega_m)t - \cos(\omega_c-\omega_m)t\right] \quad (5.26)$$

Hence, the components of a NBFM signal may be represented by the following phasor diagram.

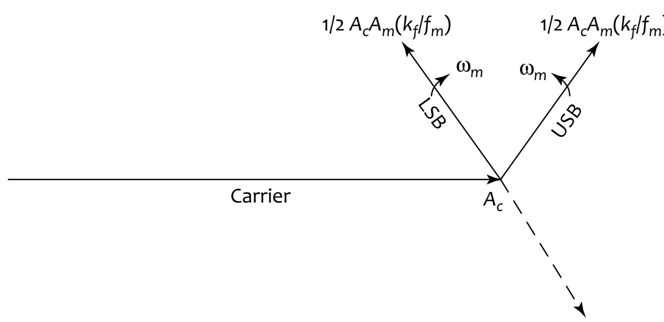


Fig. 5.7 *Phasor diagram of a narrowband frequency modulated signal*

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,

$$x_c(t) = A_c\left[1 + m\cos\omega_m(t)\right]\cos\omega_c(t)$$
$$= A_c\cos\omega_c(t) + \frac{mA_c}{2}\left[\cos(\omega_c+\omega_m)t + \cos(\omega_c-\omega_m)t\right]$$

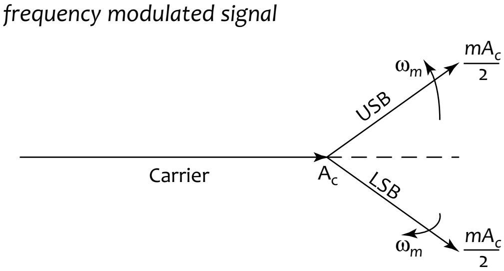


Fig. 5.8 *Phasor diagram of a single-tone modulated AM signal*

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 on Fig. 5.9.

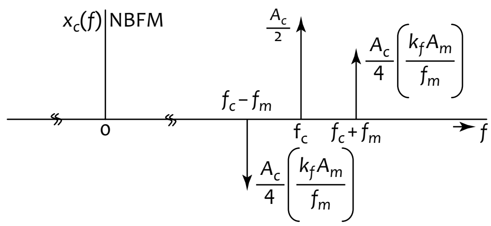


Fig. 5.9 *Two-sided spectrum of an NBFM signal*

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.

From Eq. (5.20), we have

$$\underset{\text{FM}}{x_c(t)} = A_c \cos\left[\omega_c t + \beta_f \sin 2\pi f_m t\right]$$

$\therefore$ 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 $\beta_f \triangleq \left(\frac{\Delta f}{f_m}\right)$

$\therefore$ in the first case, $\beta_{f_1} = \dfrac{50\times10^3}{5\times10^3} = 10$ 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.

$\therefore$ in the second case, $\beta_{f_2} = \dfrac{50\times10^3}{1\times10^3} = 50$ radians.

5.3.4 Generation of Narrowband PM/FM

Equation (5.24) tells us that a narrowband angle modulated signal can be represented as

$$x_c(t) = A_c\left[\cos\omega_c(t) - \phi\sin\omega_c t\right]$$

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.

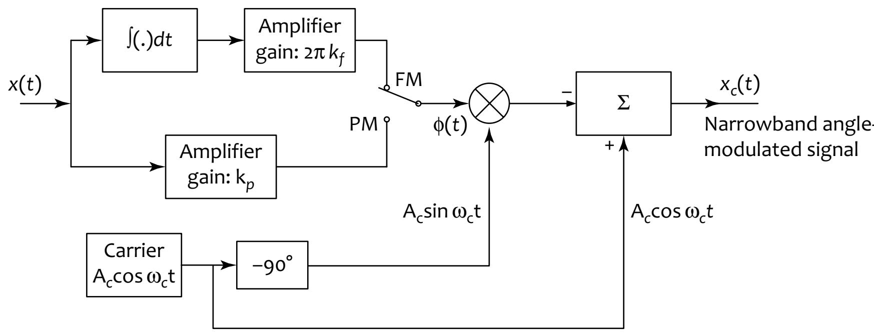


Fig. 5.10 *Generation of narrowband angle-modulated signal*

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 the 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)

$$x_c(t) = A_c \cos\left[\omega_c(t) + \phi(t)\right]$$

where,

$$\phi(t) = \begin{cases} k_p x(t) & \text{for PM} \\ 2\pi k_f \int_0^t x(\alpha) d\alpha & \text{for FM} \end{cases}$$

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

$$\beta_f \;\Delta\; \frac{A_m k_f}{f_m} = \frac{\text{Peak frequency deviation}}{\text{Modulating signal frequency}} \tag{5.29}$$

$$\beta_f \;\Delta\; m_f = \text{Modulation index for FM}$$

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

$$\beta_p \ \Delta \text{modulation index for PM } \Delta \ k_p A_m \ \Delta \ m_p \tag{5.32}$$

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

$$\phi(t) = \beta \sin \omega_m t \tag{5.33}$$

where, 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

$$x_c(t) = A_c \cos\left[\omega_c(t) + \phi(t)\right] = A_c \cos\left[\omega_c(t) + \beta \sin \omega_m(t)\right]$$

$$\therefore \quad x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot e^{j\beta \sin \omega_m t}\right] \tag{5.36}$$

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,
$$dt = \left(\frac{1}{\omega_m}\right) dx; \quad \text{when } t = -\pi / \omega_m, \ x = -\pi$$

and when $$t = \pi / \omega_m, \quad x = +\pi$$

$$\therefore \qquad c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-j(nx - \beta \sin x)} dx \tag{5.40}$$

The above integral is a function of n and β and is known as *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

$$x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \sum_{n=-\infty}^{+\infty} J_n(\beta) e^{jn\omega_m t} \right]$$

$$\therefore \qquad \boxed{x_c(t) = A_c \sum_{n=-\infty}^{+\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t} \tag{5.43}$$

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 $(\omega_c + \omega_m)$, $(\omega_c + 2\omega_m)$, $(\omega_c + 3\omega_m)$, . . . are obtained by putting $n = 1, 2, 3, \ldots$ and the lower side-frequency components having frequencies of $(\omega_c - \omega_m)$, $(\omega_c - 2\omega_m)$, $(\omega_c - 3\omega_m)$, . . . 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).

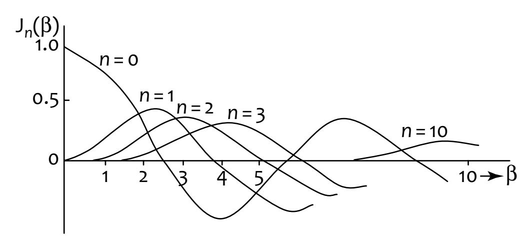


Fig. 5.11 *$J_n(\beta)$ for various values of n*

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

$$J_n(\beta) = \left(\frac{\beta}{2}\right)^n \Big/ n! \tag{5.45}$$

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

$$\begin{aligned} x_c(t) = A_c J_0(\beta)\cos\omega_c t + A_c[&J_1(\beta)\{\cos(\omega_c+\omega_m)t - \cos(\omega_c-\omega_m)t\} \\ &+J_2(\beta)\{\cos(\omega_c+2\omega_m)t + \cos(\omega_c-2\omega_m)t\} \\ &+J_3(\beta)\{\cos(\omega_c+3\omega_m)t - \cos(\omega_c-3\omega_m)t\} \\ &+ \ldots \ldots \ldots \end{aligned} \tag{5.46}$$

(an infinite number of such terms)

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.

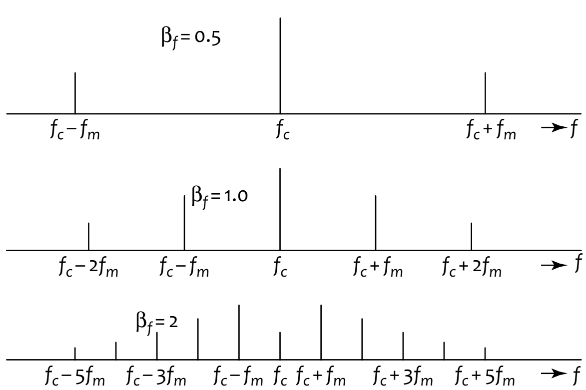


Fig. 5.12 *Amplitude spectra of an FM signal with single-tone modulation for different modulation indices*

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 $T_0 = 1/f_0$.

$$\therefore \qquad x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] = A_c \cos[\omega_c(t) + \beta_p x(t)]$$

$$= A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot e^{j\beta_p x(t)}\right] \tag{5.48}$$

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 $$e^{j\beta_p x(t)} = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t} \quad ; \quad -\infty < t < \infty$$

where,

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j\beta_p x(t)} \cdot e^{-j2\pi n f_0 t} dt \tag{5.49}$$

$$\therefore \qquad x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t}\right]$$

$$= A_c \sum_{n=-\infty}^{\infty} |c_n| \cos\{2\pi(f_c + nf_0)t + \angle c_n\} \tag{5.50}$$

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.

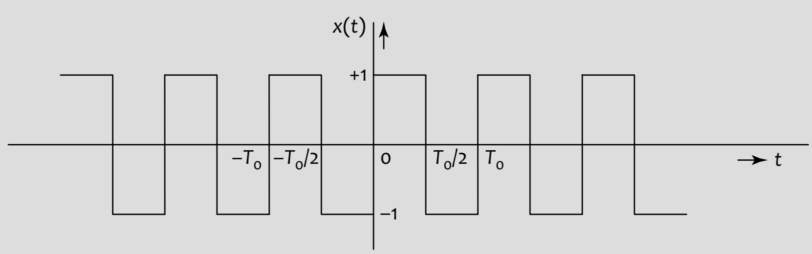


Fig. 5.13 *A square-wave modulating signal*

Let β_p be the modulation index.

Then $$x_c(t) = A_c \cos[\omega_c(t) + \phi(t)] = A_c \cos[\omega_c(t) + \beta_p x(t)]$$

$$\therefore \qquad x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot e^{j\beta_p x(t)}\right]$$

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
$$e^{j\beta_p x(t)} = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t} \quad ; \quad -\infty < t < \infty$$

$$\therefore \qquad c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} e^{j\beta_p x(t)} \cdot e^{-j2\pi n f_0 t} dt$$

Evaluating the integral and simplifying the result, we get

$$c_n = \begin{cases} 0 \ \text{ for } \ n \text{ even} \\ 2n\pi \sin\beta_p \ \text{ for } n \text{ odd} \end{cases}$$

$$\therefore \qquad e^{j\beta_p x(t)} = \sum_{\substack{n=-\infty \\ n \ \text{odd}}}^{+\infty} 2n\pi \sin\beta_p e^{-j2\pi n f_0 t} \quad ; \quad -\infty < t < \infty$$

$$\therefore \qquad x_c(t) = A_c \operatorname{Re}\left[e^{j\omega_c t} \cdot \sum_{\substack{n=-\infty \\ n \ \text{odd}}}^{+\infty} 2n\pi \sin\beta_p e^{-j2\pi n f_0 t}\right]$$

$$= A_c \operatorname{Re}\left[2\pi \sin\beta_p \sum_{\substack{n=-\infty \\ n \ \text{odd}}}^{+\infty} n e^{j(\omega_c t - 2\pi n f_0 t)}\right]$$

$$= 2\pi(\sin\beta_p) A_c \left[\sum_{\substack{n=-\infty \\ n \ \text{odd}}}^{+\infty} n \cos(\omega_c - 2\pi n f_0)t\right]$$

5.5 POWER OF AN ANGLE-MODULATED SIGNAL AND EFFECTIVE BANDWIDTH

As we had already seen, an angle-modulated signal may be represented as

$$x_c(t) = A_c \cos\left[\omega_c(t) + \phi(t)\right]$$

Hence, the average power in the angle-modulated signal is

$$\left\langle x_c^2(t)\right\rangle = \left\langle A_c^2 \cos^2\left[\omega_c t + \phi(t)\right]\right\rangle$$

$$= \frac{1}{2}A_c^2 + \frac{1}{2}A_c^2\left\langle \cos\left[2\omega_c t + 2\phi(t)\right]\right\rangle$$

$$\therefore \qquad \left\langle x_c^2(t)\right\rangle = \frac{1}{2}A_c^2 \tag{5.51}$$

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, from a theoretical point of view, this is correct. But, as has been pointed out earlier, since the amplitude of the n^{th} side-frequency component is $A_c J_n(\beta)$ and as

$$J_n(\beta) \to 0 \text{ as } n \to \infty \quad \text{(refer to Table 5.1)}$$

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.

$$\left.\begin{array}{l}\text{Now, power in the first } k\\ \text{side frequency components}\\ \text{and the carrier}\end{array}\right\} = \frac{1}{2}A_c^2 \sum_{n=-k}^{k} J_n^2(\beta)$$

$$\therefore \qquad \frac{\frac{1}{2}A_c^2 \sum_{n=-k}^{k} J_n^2(\beta)}{\frac{1}{2}A_c^2} = 0.98 = J_0^2(\beta) + 2\sum_{n=1}^{k} J_n^2(\beta)$$

$\therefore$ 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	$\beta = 0.1$	$\beta = 0.2$	$\beta = 0.5$	$\beta = 1.0$	$\beta = 2.0$	$\beta = 5.0$	$\beta = 8.0$	$\beta = 10.0$
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 $(1 + \beta)$. 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 re-write 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, is 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

$$P\left[x \le x(t) < (x+dx)\right] = f_X(x)dx$$

$$= \text{Probability of } x(t) \text{ lying between } x \text{ and } (x+dx).$$

But, we know (from the expression for f_i) that when $x(t)$ lies between x and $(x+dx)$, f_i lies between $f_c + \dfrac{1}{2\pi}k_f x$ and $f_c + \dfrac{1}{2\pi}k_f(x+dx)$.

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

$$P_{x_c}(f) = Kf_X(x), \text{ where } K \text{ is a constant.}$$

Thus, we have the important result that $P_{xc}(f)$, 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 practice, 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 $P_{xc}(f)$ 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

$$P_{x_c}(f) = \frac{A_c^2}{4\sqrt{2\pi}(\Delta f)_{\text{rms}}}\left[e^{-(f-f_c)^2/2(\Delta f)^2_{\text{rms}}} + e^{-(f+f_c)^2/2(\Delta f)^2_{\text{rms}}}\right]$$

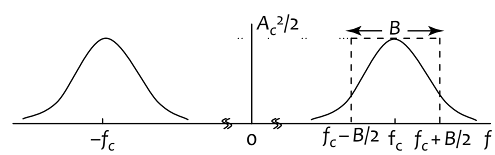


Fig. 5.14 *PSD of a Gaussian modulated FM signal*

A sketch of this is shown in Fig. 5.14.

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 $f = f_c + \frac{B}{2}$, $\mu = B/2$

$\therefore$ 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
$$y = \frac{\mu}{\sqrt{2}(\Delta f)_{\text{rms}}}, \; dy = \frac{d\mu}{\sqrt{2}(\Delta f)_{\text{rms}}}$$

$\therefore$
$$0.98\frac{A_c^2}{2} = \frac{A_c^2}{2}\frac{2\sqrt{2\pi}(\Delta f)_{\text{rms}}}{\sqrt{2\pi}(\Delta f)_{\text{rms}}} \cdot \int_0^{\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}} e^{-y^2} dy$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}} e^{-y^2} dy = \text{erf}\left[\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}}\right]$$

From error function tables,

$$\frac{B}{2\sqrt{2}(\Delta f)_{\text{rms}}} = 1.645 \quad \therefore \quad B = 4.6(\Delta f)_{\text{rms}}$$

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 $k_f A_m$ which is the peak frequency deviation, in the case of FM. But in the case of PM, f_m multiplies $(1 + k_p A_m)$. 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 Remark-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 practice, 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 $\beta_f = 1$.*

(i) 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

(ii) 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 volts, 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 volts, what is the value of the modulation index?

$$\beta = \frac{k_f \cdot A_m}{f_m} = 60 = \frac{k_f \times 3.6}{600} \quad \therefore \quad k_f = \frac{60 \times 600}{3.6} = 10^4 \text{ Hz/volt}$$

$\therefore$ 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)*

Frequency deviation Δf = 75 kHz

Audio bandwidth = 10 kHz $\quad \therefore \quad \beta_f$ = modulation index = 75/10 = 7.5

$\therefore$ max. audio frequency = f_m = 10 kHz = W

$\therefore$ using Carson's Rule, the required bandwidth is given by

Bandwidth = $2(\beta_f + 1)W = 2(7.5+1)10 \times 10^3$ = 170 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)*

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

$$B.W_1 = 2(\beta_{f_1} + 1)f_m = 2(10+1) \times 3 \times 10^3 = 66 \text{ kHz}$$

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

$$B.W_2 = 2(\beta_{f_2} + 1)f_m = 2(5+1) \times 3 \times 10^3 = 36 \text{ kHz}$$

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?

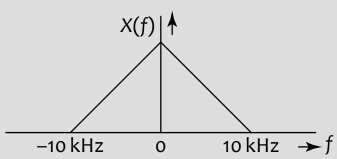


Fig. 5.15 *Spectrum of the signal of Example 5.17*

Here, the bandwidth W of the modulating signal is

$$W = 10^4 \text{ Hz}$$

k_f is given to be 60 kHz/volt and $A_m = 1$ volt since $|x(t)| \le 1$

$$\therefore \qquad \beta_f = \frac{k_f \cdot A_m}{W} = \frac{60 \times 10^3 \times 1}{10 \times 10^3} = 6.$$

$$\therefore \qquad B_T = 2\left(\beta_f + 1\right)W = 2 \times 7 \times 10^4 = 140 \text{ kHz}$$

Example 5.18

An angle-modulated signal is of the form

$$x_c(t) = 50\cos\left\lfloor 2\pi \times 10^7 \times t + 5\sin 2\pi \times 1.5 \times 10^3 t \right\rfloor$$

(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 (a), if the frequency of the modulating signal is doubled, what will be the modulation index and the transmission bandwidth?

(d) In (b), if the frequency of the modulating signal is doubled, what will be the modulation index and the transmission bandwidth?

(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) $\beta_p = k_p A_m = 5$ 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

$$v(t) = 100\cos\left[2\pi f_c t + 4 \sin 2000\pi t\right] \quad \text{where, } f_c = 10 \text{ MHz}$$

(i) Determine the average transmitted power.

(ii) Determine the peak phase deviation.

(iii) Determine the peak frequency deviation.

(iv) Is this FM or a PM signal? Explain. *(University Question)*

(i) Average transmitted power $= \dfrac{(100)^2}{2} = 500$ W

(ii) *Peak phase deviation*

Since $4 \sin 2000\,\pi t$ represents the phase deviation at any instant t and since $\sin 2000\,\pi t$ has a peak value 1, the peak phase deviation is equal to 4 radians.

(iii) *Peak frequency deviation* The instantaneous frequency is given by

$$f_i(t) = \frac{1}{2\pi}\left\{\frac{d}{dt}\left[2\pi f_c t + 4\sin 2000\pi t\right]\right\}$$

$$= f_c + \frac{4}{2\pi} \cdot 2000\pi \cdot \cos 2000\pi t = f_c + 4000\pi \cdot \cos 2000\pi$$

$\therefore$ the frequency deviation at the instant t is $4000 \cos 2000\,\pi t$ and the peak frequency deviation is 4000 Hz.

(iv) It can be considered to be a PM signal with $\beta_p = 4$ and a modulating signal of $\sin 2000\,\pi$t, or it can be considered to be an FM signal with $\beta_f = 4$ and a modulating signal of $\cos 2000\,\pi t$.

Example 5.20

An FM wave with modulation index $\beta = 1$ is transmitted through an ideal bandpass filter with midband 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.

From Eq. (5.46), the spectrum of an FM wave $x_c(t)$, with β as the modulation index, is given by

$$\begin{aligned} x_c(t) = A_c J_0(\beta)\cos\omega_c t &+ A_c\left[J_1(\beta)\{\cos(\omega_c+\omega_m)t - \cos(\omega_c-\omega_m)t\}\right] \\ &+ A_c\left[J_2(\beta)\{\cos(\omega_c+2\omega_m)t + \cos(\omega_c-2\omega_m)t\}\right] \\ &+ A_c\left[J_3(\beta)\{\cos(\omega_c+3\omega_m)t - \cos(\omega_c-3\omega_m)t\}\right] \\ &+ \ldots\ldots \end{aligned}$$

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 passband of the BPF.

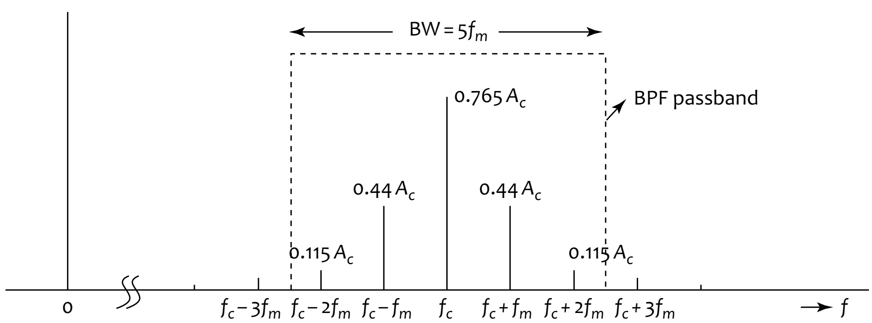


Fig. 5.16 *Amplitude spectrum for Example 5.20*

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.

Example 5.21

Express the carrier power as a fraction or percentage of the total power in an FM signal being transmitted with β_f = 2.

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.

The given angle-modulated signal may be written as

$$x_c(t) = A_c \operatorname{Re}[e^{j\omega_c t} . e^{j\beta_1 \sin\omega_1 t} . e^{j\beta_2 \sin\omega_2 t}]$$

From Eq. (5.42), we have

$$e^{j\beta_1 \sin\omega_1 t} = \sum_{n=-\infty}^{\infty} J_n(\beta_1) e^{jn\omega_1 t} \quad \text{and} \quad e^{j\beta_2 \sin\omega_2 t} = \sum_{m=-\infty}^{\infty} J_m(\beta_2) e^{jm\omega_2 t}$$

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.*

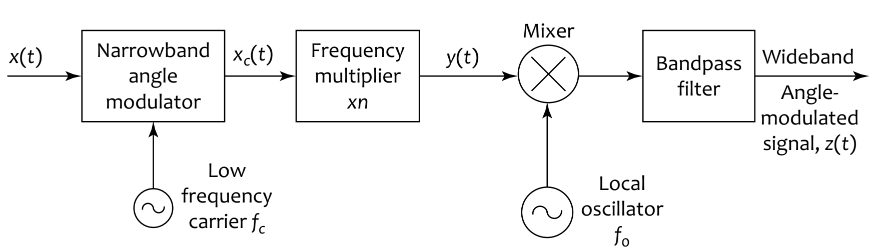


Fig. 5.17 *Armstrong or indirect method of generation of wideband angle-modulated signals*

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 a '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 centre frequency equal to either the sum frequency, or the difference frequency (whichever is needed) and whose passband 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 X_n 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.

Final carrier frequency = f_{c_4} = 91.2 × 10⁶ Hz

Initial carrier frequency = f_{c_1} = 200 × 10³ Hz

$$\therefore \qquad \left.\begin{array}{r}\text{frequency multiplication needed}\\ \text{for the carrier frequency}\end{array}\right\} = \frac{91.2\times10^6}{200\times10^3} = 456$$

Initial frequency deviation = 49 Hz = $(\Delta f)_1$

Final frequency deviation = 75 × 10³ Hz = $(\Delta f)_3$

$$\therefore \qquad \left.\begin{array}{r}\text{frequency multiplication needed}\\ \text{for the carrier frequency}\end{array}\right\} = \frac{(\Delta f)_3}{(\Delta f)_1} = \frac{75\times10^3}{49} = 1530.6$$

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.

$$\therefore \qquad f_{c_2} = f_{c_1} \times 64 = 200 \times 10^3 \times 64 = 12.8 \text{ MHz}$$

$$(\Delta f)_2 = (\Delta f)_1 \times 64 = 49 \times 64 = 3.136 \text{ kHz}$$

(Δf) at the output of the mixer = (Δf) at the input to the mixer = $(\Delta f)_2$

$$= 3.136 \text{ kHz}$$

Final carrier frequency required = 91.2 MHz = f_{c_4}

$$\therefore \text{ carrier frequency } f_{c_3} \text{ at the output of the mixer} = \frac{91.2\times10^6}{24} = 3.8 \text{ MHz}$$

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.

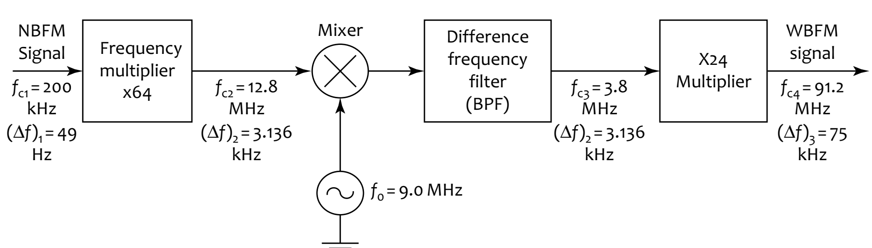


Fig. 5.18 *Indirect method of generation of WBFM of Example 5.23*

Example 5.24

In a WBFM generator of the Armstrong type shown in Fig. 5.19, the initial low-frequency carrier is of 200 kHz frequency. The maximum frequency deviation range from 100 Hz to 15 kHz, and 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.

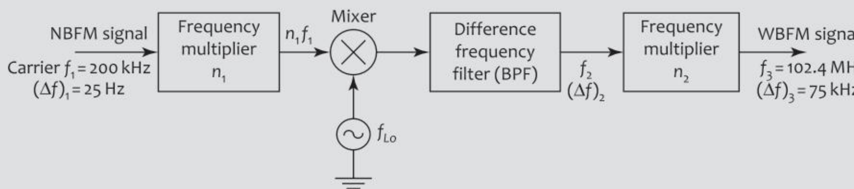


Fig. 5.19 *WBFM generator of Example 5.24*

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.

$\therefore$ total frequency multiplication needed = $n_1 n_2 = \dfrac{(\Delta f)_3}{(\Delta f)_1} = \dfrac{75\times 10^3}{25} = 3000$

Now,
$$\frac{f_3}{n_2} = f_2 = f_{LO} - n_1 f_1 \quad \text{or,} \quad f_3 = n_2 f_{LO} - n_1 n_2 f_1$$

$\therefore$
$$102.4\times 10^6 = n_2 f_{LO} - 3000\times 200\times 10^3$$

$\therefore$
$$n_2 f_{LO} = (102.4+600)\times 10^6 = 702.4 \text{ MHz}$$

Now, let us choose $n_2 = 100$ and $n_1 = 30$, so that $n_1 n_2 = 3000$

$\therefore$
$$f_{LO} = \frac{702.4\times 10^6}{n_2} = \frac{702.4\times 10^6}{100} = 7.024 \text{ MHz}$$

Hence,
$$f_2 = f_{LO} - n_1 f_1 = 7.024\times 10^6 - n_1 f_1 = (7.024-6)\times 10^6 \text{ Hz}$$

$\therefore \quad f_2 = 1.024$ MHz, and $f_3 = f_2 \times n_2 = 102.4$ MHz, 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 *L-C* 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 vary with the amplitude of the message. There are two methods for creating this variable reactance. One is to use a 'varactor', or a reactance diode and the other is to use a reactance modulator.

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 $= c(t) = C_0 + \Delta C x(t)$; $|x(t)| \le 1$

The instantaneous frequency of the oscillator at the instant 't' is given by

$$f_i(t) = \frac{1}{2\pi\sqrt{L.C(t)}} = \frac{1}{2\pi\sqrt{LC_0\left[1+\left(\frac{\Delta C}{C_0}\right)x(t)\right]}}$$

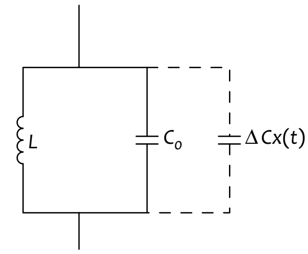


Fig. 5.20 *Variable capacitor across the tank circuit*

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 practice (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 *L-C* 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 *L-C* 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 *L-C* oscillator. Figure 5.21 shows a tuned-collector *L-C* oscillator across whose tank circuit, a varactor diode is connected. The RFC (RF choke) together with the bypass capacitor C_b ensures that the r.f 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 r.f signal. It also prevents the dc bias supply of the varactor from reaching the oscillator.

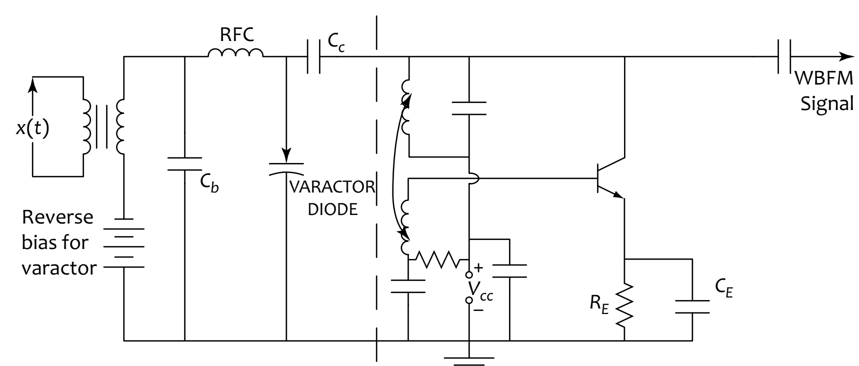


Fig. 5.21 *Typical arrangement for putting the varactor diode across the tank circuit of an L-C oscillator*

5.6.4 Using a Reactance Modulator

Principle of Operation Let the r.f voltage generated by the *L-C* oscillator be applied across the terminals *A–A*.

In the analysis that follows, we shall make the following assumptions.

(i) $i_1 << i_d$
(ii) $X_c >> R$

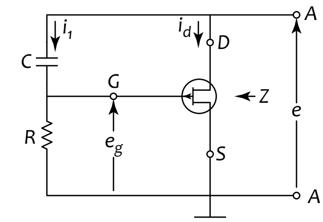


Fig. 5.22 *A FET-based reactance modulator*

The gate voltage $$e_g = i_1 R = \frac{e}{(R - jX_c)} \cdot R \qquad (5.65)$$

The drain current $$i_d = g_m \cdot e_g = g_m \cdot \frac{e.R}{(R - jX_c)} \qquad (5.66)$$

Because of our first assumption that $i_1 << i_d$, we may write the impedance seen across the terminals A–A as

$$z = \frac{e}{i_d} = \frac{e(R - jX_c)}{e.g_m.R} = \frac{R - jX_c}{g_m.R} = \frac{1}{g_m}\left[1 - j\frac{X_c}{R}\right] \tag{5.67}$$

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

$$X_{\text{eq}} = \frac{X_c}{g_m \cdot R} = \frac{1}{2\pi f_c C g_m R} = \frac{1}{2\pi f_c C_{\text{eq}}} \tag{5.69}$$

where, f_c is the frequency of the oscillator voltage and

$$C_{\text{eq}} = g_m \cdot R \cdot C \tag{5.70}$$

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 $C_{\text{eq}} = \Delta C \cdot x(t)$ 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

$$g_m \Delta \frac{\partial i_d}{\partial e_g} \Big|_{v_{ds}\ \text{constant}} \tag{5.71}$$

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.

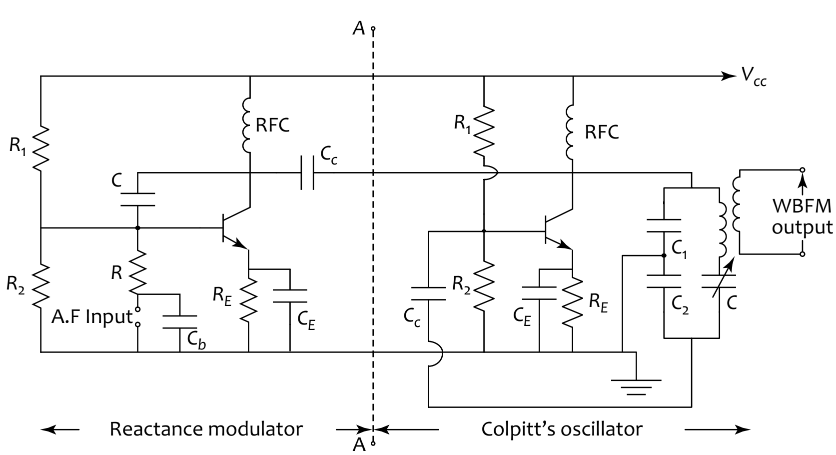


Fig. 5.23 *Direct method of generation of wideband FM using a reactance modulator*

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 of Fig. 5.22, it can be shown that the circuit to the left of terminals A–A will appear as an inductive reactance.

A 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.

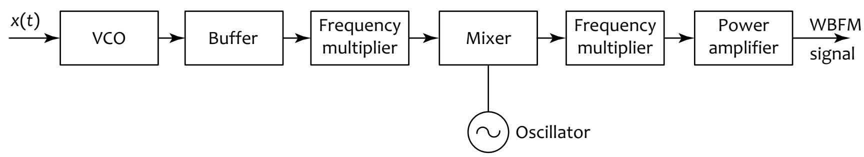


Fig. 5.24 *Varactor diode direct method of generation of a WBFM signal*

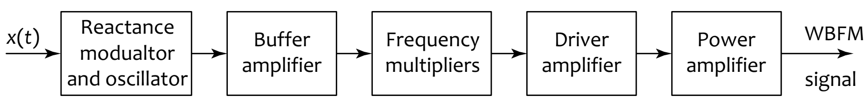


Fig. 5.25 *Reactance-modulator method of generation of WBFM signal*

As mentioned earlier at the beginning of this section, both the methods suffer from the disadvantage that the carrier frequency is obtained from an L-C 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 6 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 in so far 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

$$x_c(t) = A_c \cos\left[2\pi f_c t + \phi(t)\right] \tag{5.72}$$

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 re-arranging the terms, we get

$$\begin{aligned} &e_0(t) - (a_0 + \frac{1}{2}a_2 A_c^2) + (a_1 A_c + \frac{3}{4}a_3 A_c^3)\cos\left[2\pi f_c t + \phi(t)\right] \\ &+ \frac{1}{2}a_2 A_c^2 \cos\left[4\pi f_c t + 2\phi(t)\right] + \frac{1}{4}a_3 A_c^3 \cos\left[6\pi f_c t + 3\phi(t)\right] \end{aligned} \tag{5.75}$$

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 centre frequency f_c and bandwidth equal to

$$2\left[k_f \left|x(t)\right|_{\max} + W\right] \tag{5.76}$$

where, W is the bandwidth of the message signal, $x(t)$. The BPF output is

$$(a_1 A_c + \frac{3}{4}a_3 A_c^3)\cos\left[2\pi f_c t + \phi(t)\right] \tag{5.77}$$

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 problems, 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 Ω resistor and a 10 p.f 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.

$$\begin{aligned} c_{eq} &= g_m RC = 0.6\times10^{-3}\times1500\times10\times10^{-12} \\ &= 9 \text{ p.f} = \Delta C \end{aligned}$$

$\therefore$ 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
$$f_0 = \frac{1}{2\pi\sqrt{LC_0}}, \quad C_0 = \frac{1}{4\pi^2 \times f_0^2 \times L}$$

$$C_0 = \frac{1}{4\pi^2 \times 25 \times 10^{10} \times 10^{-3}} = 10^{-10}\,\text{F}.$$

$$\therefore \text{ peak frequency deviation } = \frac{0.5 \times 10^6 \times 9 \times 10^{-12}}{2 \times 10^{-10}} = 22.5 \text{ kHz}.$$

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 The 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, $f_{i.f}$, 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.

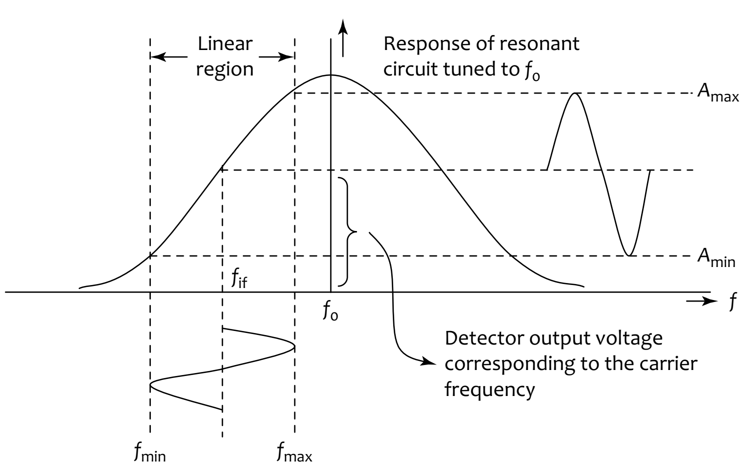


Fig. 5.26 *Principle of slope detector*

As can be seen from this figure, 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 $f_{i.f} = 10.7$ 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.

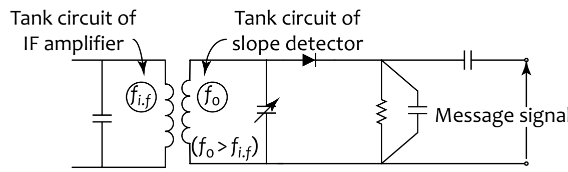


Fig. 5.27 *Frequency-to-amplitude converter followed by an envelope detector*

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).

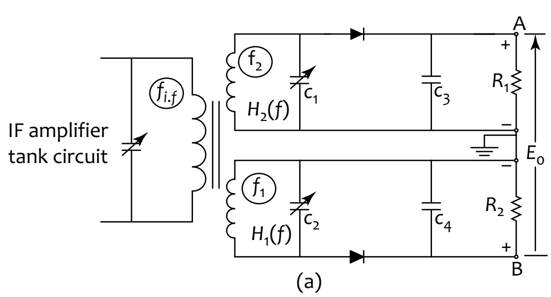


Fig. 5.28 **(a)** *Dual-slope detector circuit*

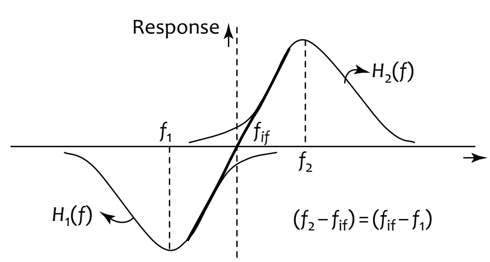


Fig. 5.28 **(b)** *Technique for larger linear range*

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, the terminal A will be at a higher potential than the 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 sub-system of an automatic frequency control unit, this FM detector is known as Foster–Seeley discriminator, phase-shift discriminator and centre-tuned discriminator.

A tank circuit consisting of a centre-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.

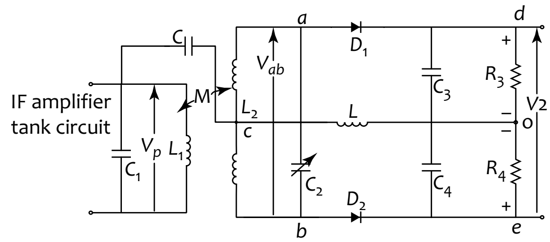


Fig. 5.29 *Foster–Seeley discriminator*

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 r.f., the circuit comprising C, L and C_4 is effectively coming across L_1. Since the reactance of the r.f. 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

$$V_s = \pm j\omega M I_p \tag{5.79}$$

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, are 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).

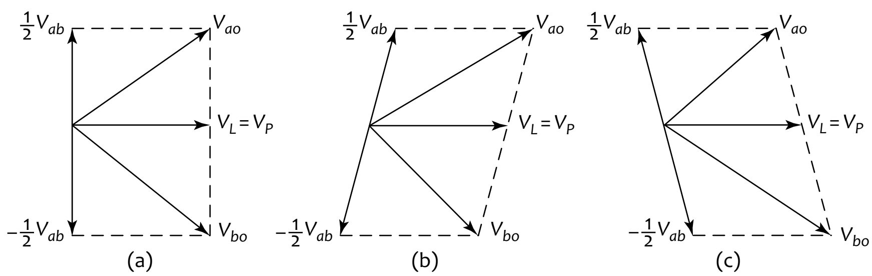


Fig. 5.30 *Phasor diagrams showing voltages across D_1 and D_2 for* **(a)** $f = f_c$ **(b)** $f > f_c$ **(c)** $f < f_c$

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 5.30(b) and (c) respectively and that

(i) For $f > f_c$: $|V_{ao}| > |V_{bo}|$ $\therefore$ V_2 is positive and equal to $|V_{ao}| - |V_{bo}|$

(ii) 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.

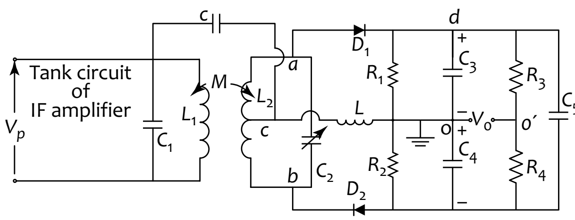


Fig. 5.31 *The ratio detector*

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.

(i) Diode D_2 is reversed in direction.
(ii) A large capacitor C_5 is connected across the output voltage of the two diodes.
(iii) 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 $R_3 = R_4, V_{0'e} = \frac{1}{2}(V_{de})$

$$\therefore \quad V_0 = \frac{V_{de}}{2} - V_{0e} = \frac{V_{d0} + V_{0e}}{2} - V_{0e} = \frac{V_{d0} - V_{0e}}{2} \tag{5.86}$$

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 un-modulated 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 cutoff frequency of W Hz, which is the message bandwidth.

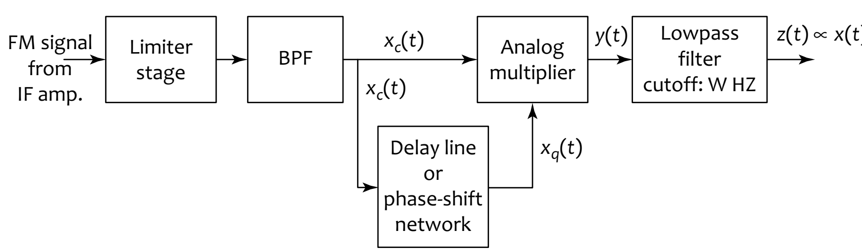


Fig. 5.32 *A quadrature FM detector*

Let the FM signal from the IF amplifier fed to the delay-line and the analog multiplier be represented as

$$\left.\begin{aligned} x_c(t) &= A_c \cos\left[\omega_c t + \phi(t)\right] \\ \text{where,}\quad \phi(t) &= 2\pi k_f \int_0^t x(\alpha)d\alpha \end{aligned}\right\} \tag{5.87}$$

k_f being the deviation constant and $x(t)$, the normalized message signal. Then the quadrature signal is

$$\begin{aligned} x_q(t) &= A_c \cos\left[\omega_c t - 90° + \phi(t-t_1)\right] \\ &= A_c \sin\left[\omega_c t + \phi(t-t_1)\right] \end{aligned} \tag{5.88}$$

$\therefore$ output of the analog multiplier is given by

$$y(t) = \sin\left[\omega_c t + \phi(t-t_1)\right]\cos\left[\omega_c t + \phi(t)\right]$$

$$\therefore \qquad \begin{aligned} y(t) &= \left[\sin\omega_c t\cos\phi(t-t_1) + \cos\omega_c t\sin\phi(t-t_1)\right]\left[\cos\omega_c t\cos\phi(t) - \sin\omega_c t\sin\phi(t)\right] \\ &= \frac{1}{2}\sin 2\omega_c t\cos\phi(t)\cos\phi(t-t_1) + \cos^2\omega_c t\cdot\cos\phi(t)\sin\phi(t-t_1) \\ &= \sin^2\omega_c t\cdot\sin\phi(t)\cos\phi(t-t_1) - \frac{1}{2}\sin 2\omega_c t\cdot\sin\phi(t)\sin\phi(t-t_1) \end{aligned}$$

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 cancelling all high-frequency terms involving $\sin 2\omega_c t$ and $\cos 2\omega_c t$, we get the output of the lowpass filter as

$$z(t) = K_1 \sin\left[\phi(t) - \phi(t - t_1)\right] \tag{5.89}$$

where K_1 is a constant.

If $\left[\phi(t) - \phi(t - t_1)\right]$ 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

$$\sin\left[\phi(t) - \phi(t - t_1)\right] \cong \left[\phi(t) - \phi(t - t_1)\right] \tag{5.90}$$

Again,

$$\frac{d\phi(t)}{dt} \approx \left[\frac{\left[\phi(t) - \phi(t - t_1)\right]}{t_1}\right] \quad \text{(meaning of a derivative)} \tag{5.91}$$

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.

$$\therefore \quad z(t) = K_1\left[\phi(t) - \phi(t - t_1)\right] = K_1 t_1 \frac{d\phi(t)}{dt} = K_2 \frac{d\phi(t)}{dt} \tag{5.92}$$

where, K_2 includes t_1.

But

$$\frac{d\phi(t)}{dt} = 2\pi k_f x(t) \quad \text{(from Eq. 5.87)} \tag{5.93}$$

$$\therefore \quad z(t) = K_3 x(t) \tag{5.94}$$

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. 5.32 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 monostable multivibrator which is designed to get triggered by the rising edges of this rectangular waveform, produces rectangular pulses of fixed duration t as shown in Fig. 5.33(d).

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

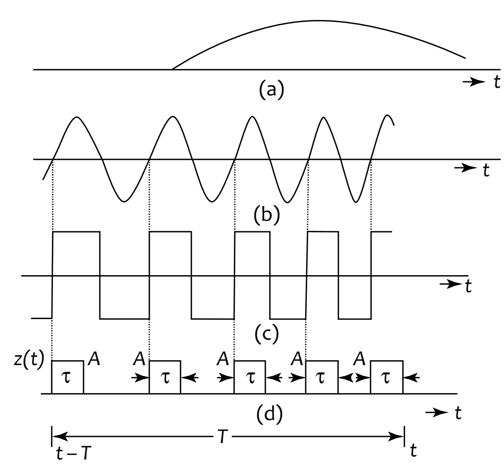


Fig. 5.33 *Waveforms to illustrate the principle of working of a zero-crossing detector* **(a)** *Modulating signal (assumed to be singe-tone)* **(b)** *FM signal* **(c)** *Hard-limited FM signal* **(d)** *Output of the monostable multivibrator*

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.

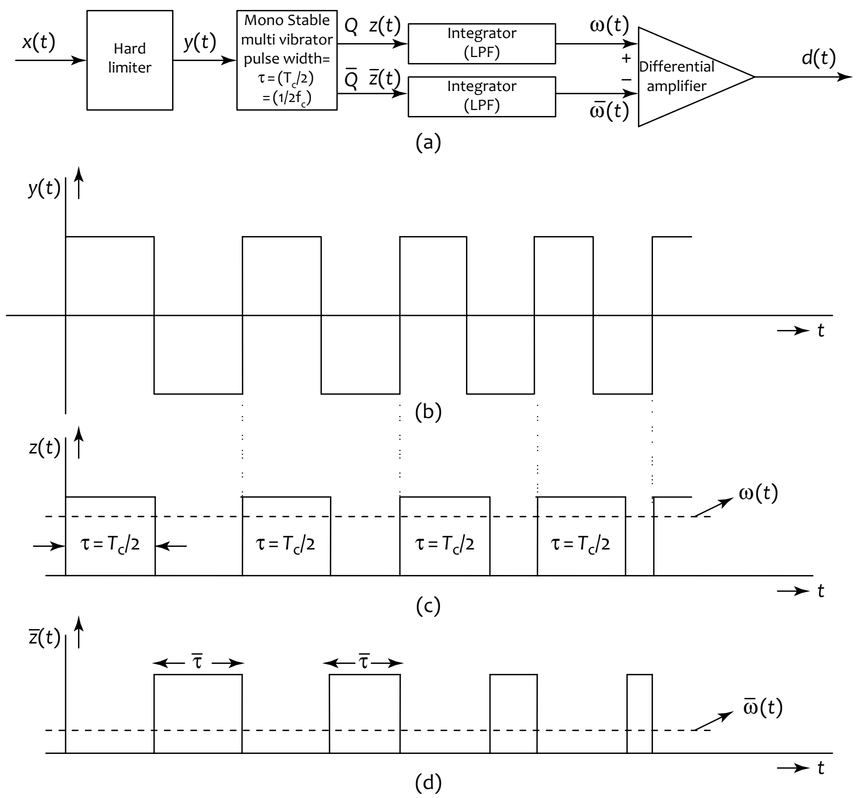


Fig. 5.34 *Balanced zero-crossing detector* **(a),** *circuit* **(b), (c), (d)** *waveforms*

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 upto 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.

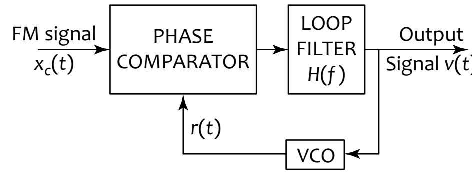


Fig. 5.35 *Block diagram of a phase-locked loop*

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

(i) 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

(ii) 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

$$x_c(t) = A_c \sin\left[2\pi f_c t + \phi(t)\right] \tag{5.96}$$

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,

$$r(t) = A_r \cos\left[2\pi f_c t + \phi_r(t)\right] \tag{5.98}$$

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,

$$x_c(t)\cdot r(t) = A_c A_r \sin\left[2\pi f_c t + \phi(t)\right]\cos\left[2\pi f_c t + \phi_r(t)\right]$$

$$= \frac{1}{2} A_c A_r \left\{\sin\left[4\pi f_c t + \phi(t) + \phi_r(t)\right]\sin\left[\phi(t) - \phi_r(t)\right]\right\}$$

The filter eliminates the high frequency component at the frequency of $2f_c$.

$\therefore$ output of the loop filter = $v(t) = \frac{1}{2} A_c A_r \sin\left[\phi(t) - \phi_r(t)\right]$ (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 very small.

Hence, we may make the approximation $\sin\phi_e \cong \phi_e$

$$\therefore \qquad v(t) = \frac{1}{2}A_c A_r \phi_e \tag{5.102}$$

and,

$$\phi_e(t) = \phi(t) - 2\pi k_v \int_0^t v(\tau)d\tau \quad \text{(from Eqs 5.99 and 5.101)}$$

$$\therefore \qquad \frac{d}{dt}\phi_e(t) + 2\pi k_v v(t) = \frac{d}{dt}\phi(t); \qquad v(t) = \int_0^\infty \phi_e(t)h(t-\tau)d\tau \tag{5.103}$$

(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.

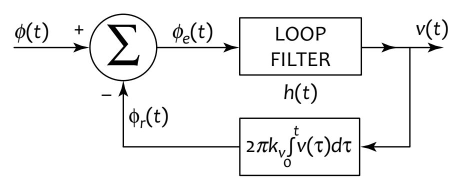


Fig. 5.36 *Linearized equivalent circuit of the PLL*

Taking the Fourier transform of the differential equation, we have

$$j2\pi f\Phi_e(f) + 2\pi k_v \Phi_e(f)H(f) = j2\pi f\Phi(f)$$

$$\Phi_e(f) = \frac{\Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)}$$

But $\qquad V(f) = \Phi_e(f)H(f), \quad (\text{since } v(t) = \phi_e(t) * h(t))$

$$V(f) = \frac{H(f)\cdot\Phi(f)}{1 + \left(\frac{k_v}{jf}\right)H(f)} \tag{5.104}$$

If the gain of the loop filter is high enough so that

$$\left(\frac{k_v}{jf}\right)H(f) >> 1 \quad \text{for} \quad |f| < W \tag{5.105}$$

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 $|f| \geq W$. $\therefore$ $H(f)$ 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. 5.34(a) is approximately proportional to the amplitude of the normalized modulating signal x(t) of the input FM signal, $x_c(t)$.

Referring to Fig. 5.34, 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.

$$\langle \omega(t) \rangle = \frac{nA\tau}{T} = \frac{\left[f_c + k_f x(t)\right]TA\tau}{T} = A\tau\left[f_c + k_f x(t)\right]$$

where $\left[f_c + k_f x(t)\right] = f_i$ = instantaneous frequency = n, the number of *positive going zero-crossings* of $y(t)$.

But,
$$\tau = \frac{T_c}{2} = \frac{1}{2f_c} \quad \therefore \quad \langle \omega(t) \rangle = A\tau\left[f_c + k_f x(t)\right] = \frac{A}{2} + \frac{A}{2}\left(\frac{\Delta f}{f_c}\right) \tag{5.108}$$

$\bar{\tau}$ changes with the amplitude of $x(t)$ and is given by

$$\bar{\tau} = \left[\frac{1}{f_c + k_f x(t)_c} - \frac{1}{2f_c}\right] = \frac{1}{f_c}\left[\frac{1}{1+(k_f / f_c)x(t)} - \frac{1}{2}\right] \approx \frac{1}{f_c}\left(1 - \frac{1}{2}\frac{k_f}{f_c}x(t) - \frac{1}{2}\right)$$

$$= \frac{1}{2f_c}\left(1 - \left(\frac{k_f}{f_c}\right)x(t)\right)$$

$$\langle \bar{\omega}(t) \rangle = \frac{nA\bar{\tau}}{T} = \frac{\left[f_c + k_f x(t)\right]TA\left(\frac{1}{2f_c}\right)\left[1 - \left(\frac{k_f}{f_c}\right)x(t)\right]}{T} = \frac{A}{2}\left[1 - \left(\frac{k_f}{f_c}\right)^2 x^2(t)\right]$$

$$\therefore \qquad \langle d(t) \rangle = \frac{nA}{T}[\tau - \bar{\tau}] = \left[1 - \left(\frac{\Delta f}{f_c}\right)x^2(t)\right]\left(\frac{\Delta f}{f_c}\right)$$

Since $\left(\frac{\Delta f}{f_c}\right) << 1$ and $x(t)$ has been normalized such that $|x(t)| \le 1$,

$$\langle d(t) \rangle \cong \left(\frac{\Delta f}{f_c}\right) = \left(\frac{k_f}{f_c}\right) x(t) \quad \therefore \quad \langle d(t) \rangle \propto x(t)$$

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 r.f. 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 a 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, lowpass filtered for removing the out-of-band noise, if any, and then finally fed to the loudspeaker,

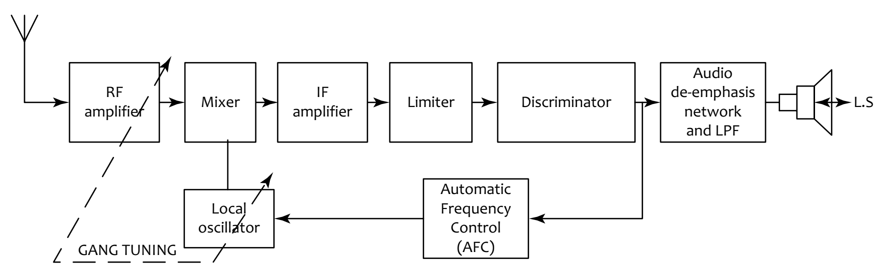


Fig. 5.37 *Block diagram of superheterodyne FM broadcast receiver*

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

$$r(t) = A\cos\omega_c t + B\cos(\omega_c + \Delta\omega)t$$
$$= (A + B\cos\Delta\omega t)\cos\omega_c t - (B\sin\Delta\omega t)\sin\omega_c t$$

Hence $r(t)$ may be written as

$$r(t) = R(t)\cos[\omega_c t + \theta(t)]$$

where, $$R(t) = \left[(A + B\cos\Delta\omega t)^2 + (B\sin\Delta\omega t)^2\right]^{1/2}$$

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 consequence. $\theta(t)$, 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.

5.9.2 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. 5.38. 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

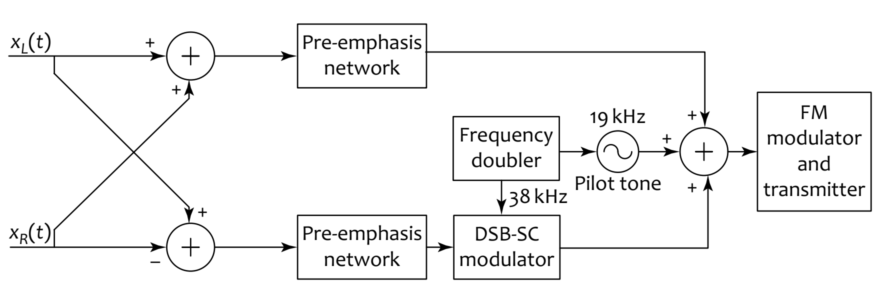


Fig. 5.38 *An FM stereo transmitter*

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.

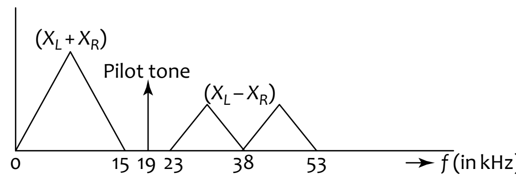


Fig. 5.39 *One-sided spectrum of the baseband signal used for frequency modulating the final carrier*

From the foregoing, it is clear that functionally, the receiver should first recover the baseband signal (whose spectrum is shown in Fig. 5.39). 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. 5.40.

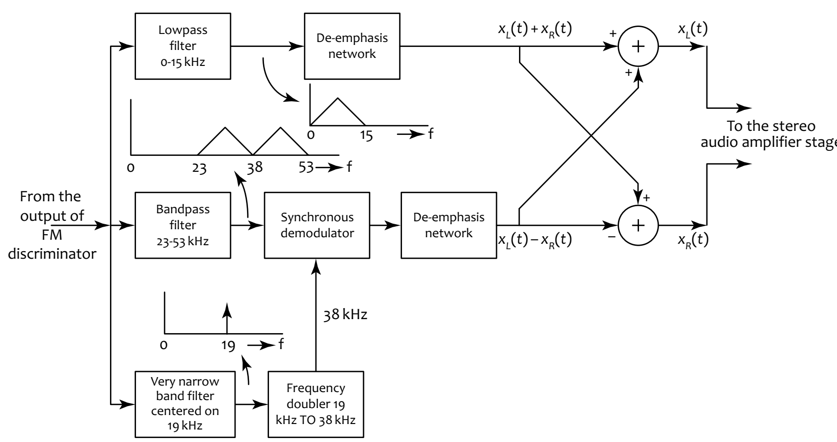


Fig. 5.40 *FM stereo receiver after the discriminator stage*

The output of the discriminator is the baseband signal whose spectrum is as in Fig. 5.39. This is fed simultaneously to a lowpass 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. 5.40. These are then fed to the (stereo) audio amplifier and they finally drive the dual loudspeakers.

SUMMARY

1. Frequency modulation and phase modulation are together known as angle modulation.
2. FM and PM both change the phase angle, but in different ways.
3. **PM:** $x_c(t) = A_c \cos\left[\omega_c t + k_p x(t)\right]$ where, k_p is called the phase deviation constant.

4. **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.
5. When $x(t)$, the modulating signal $= A_m \cos 2\pi f_m t$,
   (a) $\underset{\text{PM}}{x_c}(t) = A_c \cos\left[2\pi f_c t + k_p A_m \cos 2\pi f_m t\right]$
   (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 $\triangleq k_p A_m$
   (d) β_f = Modulation index for FM $\triangleq \frac{k_f A_m}{f_m}$
6. Deviation ratio $= \triangleq \left[\frac{\text{Peak frequency deviation}}{\text{Modulating signal frequency}}\right]$
7. (a) If $f(t)$, 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 $= 2f_m$.
8. (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.
9. (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.
10. (a) The average power of an angle-modulated signal is $P_{av} = \frac{1}{2}A_c^2$ 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:

$$B_T = 2(\lfloor \beta \rfloor + 1)W, \text{ where, } \beta \triangleq \begin{cases} k_p \max|x(t)| & \text{for PM} \\ \frac{k_f \max|x(t)|}{W} & \text{for FM} \end{cases}$$

11. (a) B_T of a 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
12. WBFM may be generated either by the indirect (or, Armstrong method), or by the direct method.
13. (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.

14. 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).
15. 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.
16. 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.

REFERENCES AND SUGGESTED READING

1. Simon Haykin: *Communication Systems*, John Wiley & Sons, New York, Fourth Edition, 2001.
2. Taub and Schilling: *Principles of Communication Systems*, Tata McGraw-Hill, New Delhi, 1995.
3. Bruce Carlson et.al: *Communication Systems*, McGraw-Hill International Edition (Fourth Edition), 2002.
4. J G Proakis and M Salehi: *Communication Systems Engineering*, Second Edition, Prentice-Hall of India, New Delhi, 2006.
5. K K Clarke and D T Hess: *Communication Circuits: Analysis and Design*, Addison-Wesley, Mass, 1971.
6. L W Couch, II: *Digital and Analog Communication Systems*, Fourth Edition, Macmillan, New York, 1993.
7. R E Ziemer and W H Tranter: *Principles of communications: Systems, Modulation and Noise*, Houghton, Miffin, Boston, 1990.

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 $\beta =$ smallest value of β for which the carrier component vanishes. Sketch the 2-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. With the help of block schematic diagrams and sketches of the spectra of appropriate signals, explain the principle of stereo FM transmission and reception.

FILL IN THE BLANKS

1. In FM, the carrier ___________ (amplitude/frequency) is varied in accordance with the variations in the ___________ (amplitude/frequency) of the modulating signal.
2. The frequency deviation constant k_f has units of ________.
3. The modulating signal $x(t)$ has to be ________ (differentiated/integrated) before being fed to a phase modulator in order to obtain a FM signal.
4. The modulating signal $x(t)$ has to be ________ (differentiated/integrated) before being fed to a frequency modulator if a PM signal is to be obtained.
5. Deviation ratio in the case of FM, is the ratio of ________ to _______.
6. If the modulating signal is a single-tone with peak amplitude A_m and frequency f_m, β_p = ______ and β_f = _____.
7. For single-tone modulation with modulation index β, the peak amplitude of the n th side-frequency component of an angle-modulated signal is given by _______.
8. $J_n(\beta)$ represents _____ ______ of the ______ _______ of order _____ and with argument ______.
9. For small values of β, $J_n(\beta) \approx$ ____________.
10. If n is a non-zero integer, $J_n(0)$ = _________.
11. If n is odd, $J_{-n}(\beta)$ = __________.
12. A carrier $A_c \cos \omega_c t$ is frequency modulated by a single-tone of frequency f_m and peak amplitude A_m to a modulation index of β. The average power of the modulated signal is ___________.
13. The effective bandwidth of an angle-modulated signal is defined as ___________.
14. Consider angle-modulation by a single-tone of frequency f_m. The effective bandwidth of the modulated signal is found to be increasing linearly with f_m. The modulation is ________ (FM / PM).
15. For commercial FM broadcasting $\beta \approx$ _______, $(\Delta f)_{max}$ = _______ kHz and adjacent carriers are to be separated by _________ kHz.
16. A FM signal with carrier frequency f_c and peak frequency deviation $(\Delta f)_{max}$ is given as input to a mixer with a local oscillator frequency f_0 and followed by a BPF that selects the difference frequency. The new carrier frequency is ________ and the new peak frequency deviation is __________.
17. A frequency discriminator may be thought of as a ________ to _________ converter.
18. In a superheterodyne FM broadcast receiver, the IF is ________.
19. In a superheterodyne FM broadcast receiver, the function of the limiter stage is to _______.
20. In stereo FM transmission, the difference signal is made to _________ modulate a 38 kHz tone.

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) $\dfrac{2\pi k_p}{W}$ (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 narrow band 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 narrow band 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

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 P-5.1, if the modulation is (a) frequency modulation, or (b) phase modulation.

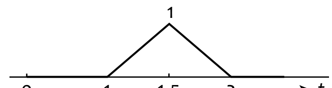


Fig. P-5.1

2. An FM signal is of the form $x_c(t) = 75\cos\lfloor 2\pi \times 5 \times 10^6 t + 6\sin 200\pi t \rfloor$
(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 $x_{1c}(t) \cdot x_{2c}(t)$ 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. A 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. P-5.2. 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.

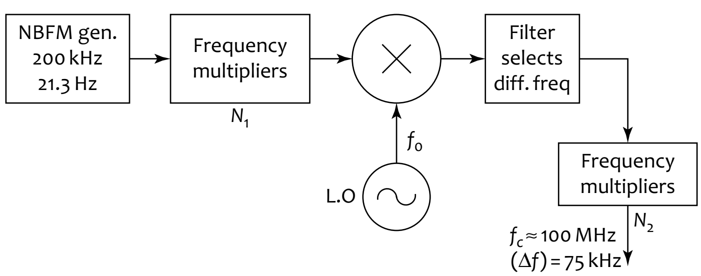


Fig. P-5.2

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. P-5.3.
   (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?
8. An angle-modulated signal is given to be

$$x_c(t) = 75\cos\lfloor 2\pi \times 10^7 t + 6\sin 2\pi \times 2 \times 10^3 t \rfloor$$

   (a) If it is an FM signal, what are its frequency deviation constant, modulation index β_f and transmission bandwidth?

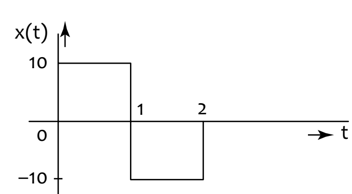


Fig. P-5.3

(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 1V 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. A 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 5 doublers and then a frequency multiplier chain consisting of 3 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 P-5.4 shows an arrangement used frequently as an FM demodulator at microwave frequencies. The delay line produces a delay of T seconds 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

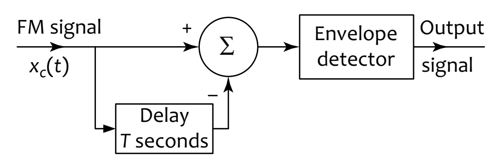


Fig. P-5.4

$$x_c(t) = A_c \cos\left[2\pi f_c t + \beta_f \sin(2\pi f_m t)\right]; \ \beta_f < 1$$

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.

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)

6

AM and FM Transmitters and Receivers

This chapter helps the student to

- *clearly understand the various ways transmitters are classified*
- *recognize the key specifications for AM and FM audio broadcast transmitters, draw their block diagrams and explain their working*
- *understand the difference between and the merits and demerits of high-level and low-level modulation*
- *understand clearly the problems like image-frequency interference, and adjacent-channel interference and the effect that the choice of IF has on these*
- *understand the importance of broadcast receiver parameters like selectivity, sensitivity, fidelity and output SNR and know the typical ranges of their values*
- *draw the block diagrams of AM and FM broadcast, SSB-SC, SSB-pilot carrier and ISB transmitters and their receivers and explain their working*

6.1 INTRODUCTION

In the fourth chapter, we had discussed various types of amplitude modulations like AM, DSB-SC and SSB-SC. We had also discussed the methods of demodulation to be adopted for these different types of modulations. Similarly, in the fifth chapter, we discussed frequency and phase modulation and the demodulation techniques for them. Thus, till now, our attention was focused mainly 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, necessary to have a good understanding of the various processes taking place both at the transmitter and at the receiver, all of which *together* make 'communication' possible. This chapter therefore is devoted to a study of AM and FM transmitters and receivers, in detail. It is neither necessary, nor is it possible, for us to go into the details at the circuit level; we will mostly confine the discussion only to the block schematic diagrams level.

6.1.1 Functions of a Transmitter

Any transmitter, whatever may be the type of modulation employed, has three basic processes taking place in it. These are

1. Generating the carrier signal at the specified frequency conforming to the frequency stability criteria as laid down by an appropriate regulatory authority and to raise the power of this carrier signal to the required level

2. Processing the message signal and raising its power level
3. Modulating the carrier signal at an appropriate power level using the message signal as the modulating signal

6.1.2 Sections of a Transmitter

Three distinct sections can clearly be identified in any transmitter (these do not exactly correspond to the three processes listed above). These are the following:

1. *The exciter section* This section comprises the carrier oscillator, frequency stabilization arrangements and the buffer amplifier, etc.
2. *The modulation section* Audio pre-amplifiers, voltage amplifiers and power amplifiers come under this section.
3. *RF power section* This section, in which RF power amplification takes place, comprises the driver amplifier and the subsequent RF power amplifiers, which are class-C power amplifiers, the final stage being the modulated class-C amplifier if the AM transmitter employs high-level modulation. In case of low-level modulation, this section consists of class-A or class-AB power amplifiers.

6.1.3 Some Definitions

(i) Effective Radiated Power (ERP) The effective radiated power from a transmitter is the average rf power from the transmitter multiplied by the gain (i.e., loss) of the transmission lines from the transmitter to the antenna and the gain of the antenna itself.

(ii) Primary Service Area It is the area around a transmitting antenna comprising all points at which the field strength due to the signal radiated by the antenna is not below a certain prescribed value, which is generally 5 to 10 mV per metre.

6.1.4 Classification of Transmitters

Transmitters may be classified in several different ways.

(i) On the basis of the frequency band in which they operate—such as medium-wave transmitters, short-wave transmitters, VHF transmitters, etc.
(ii) On the basis of the modulation employed—such as AM transmitters, FM transmitters, SSB transmitters, and so on.
(iii) On the basis of the service provided—such as broadcast transmitters, radio telephone transmitters, TV and radar transmitters, etc.
(iv) On the basis of their power—as low-power transmitters, typically less than 1 kW, medium-power transmitters, typically less than 5 kW and high-power transmitters, typically 5 kW to several hundreds of kilowatts.

A transmitter is generally designed for some specific service or application and all of its parameters such as its carrier frequency, type of modulation, power rating, etc., will be dependent upon this.

(i) Transmitters for Audio Broadcasting These transmitters handle speech, music, etc., and each of them is meant to serve a very large number of receivers. Hence, they use either AM with full carrier, or FM.

(a) AM broadcast transmitters generally handle audio frequencies only up to 5 kHz and use carrier frequencies in the medium-wave band of 550 kHz to 1650 kHz, or in the short-wave band of 3 MHz to 30 MHz. Transmitters operating in the medium-wave band primarily depend on ground wave propagation and because of the attenuation inherent in ground-wave propagation, they have limited service area. Those operating in the short-wave band primarily depend upon sky-wave propagation and cover very large areas. AM broadcast transmitters use carrier powers of the order of 1 kW to 100 kW, or more.

(b) 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 to 108 MHz in the VHF band, and make use of powers of the order of 100 kW. As propagation of radio waves in the VHF takes place by line-of-sight, the service area of these transmitters will depend not only on the carrier power radiated, but also on the height of the antennas.

(ii) Transmitters for Radio Telephony These are essentially for transmission of telephone signals from one point to another over long distances, using radio waves. AM/FM and short waves are used. Highly directional antennas are employed as the objective is only point-to-point communication and not broadcasting. Because of the use of highly directional antennas, transmitter powers need be only of the order of a few kilowatts.

(iii) Transmitters for TV Transmitters for analog television broadcasting make use of FM for the sound and vestigial sideband modulation for the video signal. They operate in the VHF and UHF bands and are assigned 7 MHz wide channels.

6.1.5 Functions of a Receiver

Any receiver, in general, has the following functions to perform.

1. To enable one to pick up the signal emanating from any *desired* transmitter when the receiver is in the service area of that transmitter, with minimum possible, or no interference caused by the signals emanating from the other transmitters whose service area also covers the receiver location
2. To amplify the picked-up signal sufficiently so that it can be demodulated
3. To extract the message signal by demodulating the picked-up RF signal
4. To raise the message signal obtained by demodulation to a power level sufficient to operate the output device (usually a loud speaker)

6.1.6 Classification of Receivers

Receivers may be classified in several different ways. One way is to classify them according to the type of modulation of the incoming signal that the receiver is capable of detecting, e.g., AM receivers, FM receivers and SSB receivers, and so on. Another way is to classify them based on the frequency range in which they operate: RF receivers, VHF receivers, microwave receivers, etc. Yet another way, and one which is based on the type of configuration of the receiver, is to classify them as Tuned Radio Frequency (TRF) receivers and superheterodyne receivers. TRF receivers were in vogue during the early stages of development of radio communication, i.e., approximately up to about 1930. Although it was invented much earlier, i.e., in 1920's, the superheterodyne configuration became popular only in the thirties. Now, it is the standard configuration for *a* almost all receivers—AM, FM, TV and even for radars.

6.2 AM BROADCAST TRANSMITTERS

AM broadcasting makes use of the 550 kHz to 1605 kHz medium-wave band which depends on ground-wave propagation and the 3 MHz to 30 MHz short-wave band which depends on the sky-wave propagation. It is meant for voice communication and audio frequencies up to 5 kHz only are used, making the transmitter bandwidth equal to 10 kHz. Thus, carrier frequencies are allocated with 10 kHz separation between adjacent channels. This means that the transmitter carrier frequency must be extremely stable, as otherwise the higher side-frequencies are likely to drift into the adjacent channel and cause interference. That is why stringent stability requirements like a drift of not more than ±20 parts per million i.e., ±0.02% of the assigned carrier frequency, are to be complied with.

Transistors are used for low power transmitters, say up to 1 kW; but in the case of high-power transmitters, the last, or, even the last few RF power amplifiers stages use only vacuum tubes.

6.2.1 High-level and Low-level Modulation

AM transmitters are generally categorized into two types—those with high-level modulation and those with low-level modulation.

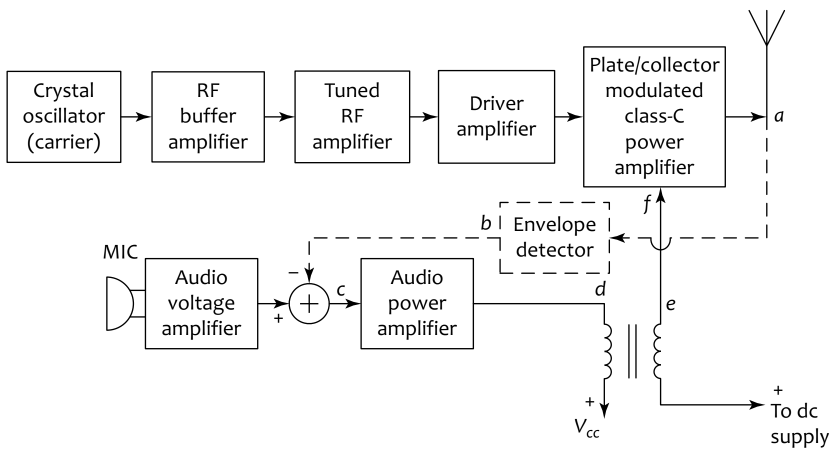


Fig. 6.1 *AM transmitter with high-level modulation*

Definitions 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 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.

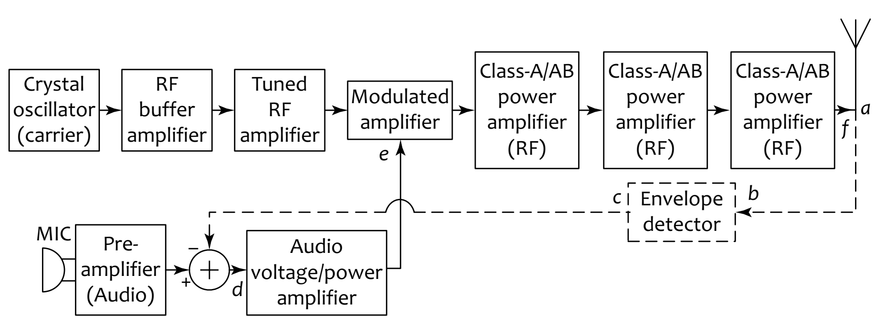


Fig. 6.2 *AM transmitter with low-level modulation*

Comparison The advantage of high-level modulation is that all the RF power amplifiers can be class-C power amplifiers, which can be designed to have very high power efficiencies of the order of 80 to 90%.

However, since the final class-C power amplifier is collector/plate modulated, as shown in Section 4.2, the power to be supplied by the final audio power amplifier will have to be very large, since the total sideband power in the AM signal at the output of the modulated amplifier is derived from the audio power supplied by the final audio power amplifier. Noting that at 100% modulation, this is 50% of the carrier power, the final audio power amplifier of a high-level transmitter with a carrier power of 10 kW will have to supply 5 kW of audio power to the modulated class-C amplifier. In the case of low-level modulation, the audio power needed to be supplied by the final audio power amplifier is very little and therefore is not a problem. However, since the modulation takes place at a stage much earlier than the final RF power amplifier, all the RF power amplifiers subsequent to the modulated stage will have to be either class-A or class-AB tuned power amplifiers, and these have very low power efficiency.

6.2.2 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

(i) 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 6.1 and 6.2., it must have a very high input impedance and a low output impedance
(ii) keep the crystal used in the carrier oscillator circuit at a constant temperature, as temperature variations can cause frequency drift
(iii) 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 picofarads) 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, the Hazeltine method and Rice method are worth mentioning.

(i) Hazeltine Method We first note that points A and B of the collector tank circuit 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}.

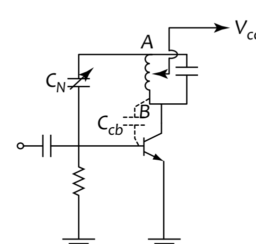


Fig. 6.3 *Hazeltine method of neutralization*

(ii) 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. 6.4. Because the centre-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}.

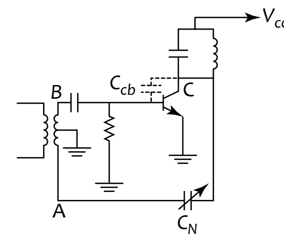


Fig. 6.4 *Rice neutralization*

6.2.3 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 incase low-level modulation is employed, if the class-A/AB tuned power amplifiers do not cause any distortion of the envelope.

This negative feedback is provided as shown in Figs 6.1 and 6.2. 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 the Figs 6.1 and 6.2. 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.

6.3 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 Second World War. However, the superheterodyne type of receiver, actually invented by Major Armstrong some time during World War I, became popular by about mid 1930's 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, although briefly, and then discuss the superheterodyne receiver in some detail.

6.3.1 Tuned Radio Frequency (TRF) Receiver

As shown in Fig. 6.5, a TRF receiver simply consists of a chain of two or three single-tuned r.f. 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.

These TRF receivers are quite simple and inexpensive. But they suffer from several severe disadvantages, chief among them being 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 Q's 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.

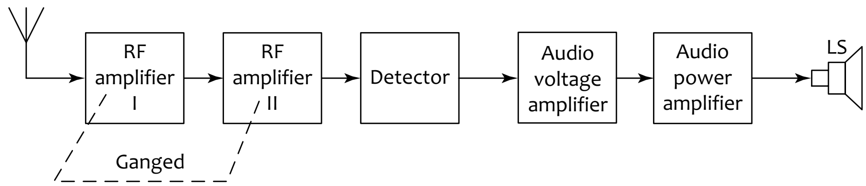


Fig. 6.5 *A tuned radio frequency receiver*

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.

6.3.2 The 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. 6.6.

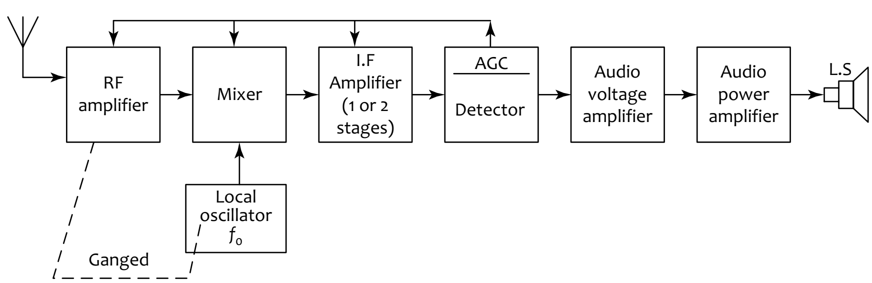


Fig. 6.6 *Block diagram of an AM superheterodyne broadcast receiver*

We shall now discuss briefly the salient features and the functions of each block in the above block diagram.

(a) 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 the following:

(i) 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 (refer Chapter 8), this will not be a desirable arrangement.

(ii) To give good image frequency rejection and IF rejection capability to the receiver.
(iii) To give some amount of adjacent channel selectivity.

(b) Local Oscillator This is an *LC* oscillator which produces a sinusoidal signal of frequency f_0 which is such that $f_0 - f_c = \text{IF}$, the pre-determined 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.

(c) 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 $f_0 = f_c + f_{if}$. 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.

(d) IF Amplifier(s) One or two stages of IF amplifiers are generally used. 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.

(e) 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. 4.24, 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. 6.6, this voltage is used for biasing the preceding stages so as to control their gains and thus provide AGC.

(f) Automatic Gain Control (AGC) An arrangement for automatic gain control, or AGC, is necessary in radio receivers for the following reasons.

(1) 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 re-tune the receiver, is however, cumbersome.
(2) 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 r.f. 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. 4.24. This AGC voltage will be high for stronger r.f. input signals and low for weaker signals. We therefore apply this as a bias voltage to the r.f. 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 r.f. 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.

(g) 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 pre-determined level, as shown in Fig. 6.7, which depicts the AGC characteristics.

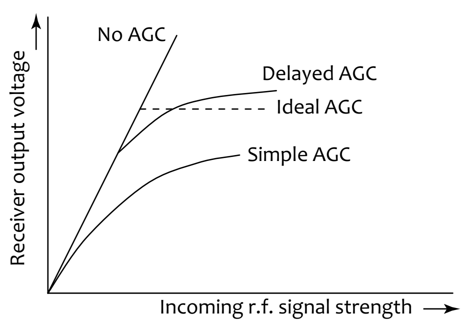


Fig. 6.7 *AGC characteristics*

Delayed AGC is generally obtained by having a separate diode rectifier circuit for producing the AGC voltage, and applying a positive bias of pre-determined 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.

Audio Voltage and 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 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.

Choice of Local Oscillator Frequency We had 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 superheterodyne receiver meant for the medium-wave band, 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) $f_o > f_c$ $\therefore \quad f_0 = f_c + f_{if}$

Since f_c ranges from 555 kHz to 1605 kHz,

$$f_0 \text{ ranges from } (555 + 455) \text{ kHz to } (1605 + 455) \text{ kHz}$$

i.e., from 1010 kHz to 2060 kHz

$$\therefore \quad f_{0\,\max} = 2060 \text{ kHz} \quad \text{and} \quad f_{0\,\min} = 1010 \text{ kHz}$$

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 practical.

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 picofarads. Hence, $C_{\min} \neq 0$.

(b) $f_o < f_c$ $\therefore \quad f_0 = f_c - f_{if}$

$$\therefore \quad \text{in this case,} \quad f_{0\,\max} = (1605 - 455) \text{ kHz} = 1150 \text{ kHz}$$

$$f_{0\,\min} = (555 - 455) \text{ kHz} = 100 \text{ kHz}$$

$$\therefore \quad \left(\frac{C_{\max}}{C_{\min}}\right) = \left(\frac{f_{0\max}}{f_{0\min}}\right)^2 = \left(\frac{1150}{100}\right)^2 = 132.25$$

This is an impractical value since it implies that $C_{\max}$ should be of the order of a few thousand picofarads!

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.

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. 6.8. However, no practical filter can give such a frequency response.

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 FM 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 MF band, say 1600 kHz, an adjacent channel signal of 1610 kHz will not

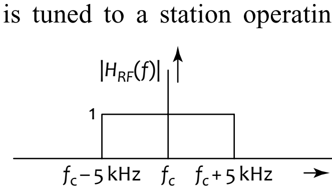


Fig. 6.8 *Ideal selectivity curve*

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 centre 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. 6.8, 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.

(i) We can use 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. 6.9.

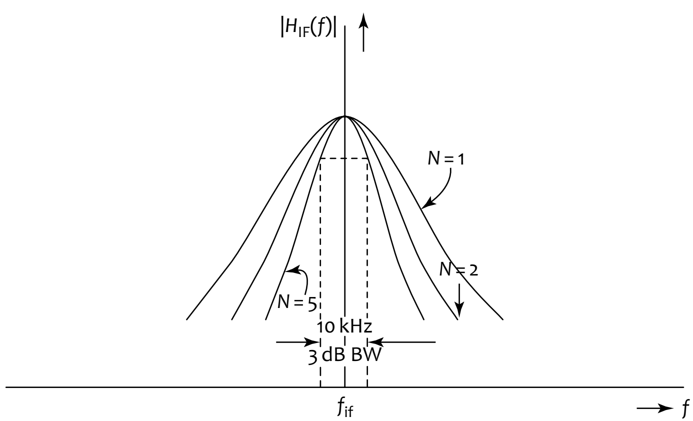


Fig. 6.9 *Selectivity curves for 1, 2 and 5 stages*

(ii) We can use 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. 6.10. As N increases, the ripple amplitude becomes smaller and the skirts become sharper.

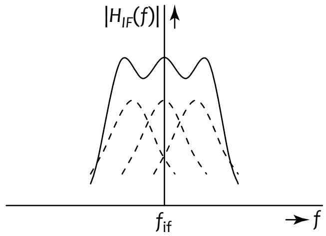


Fig. 6.10 *Selectivity curve for an IF amplifier comprising 3 stagger-tuned stages*

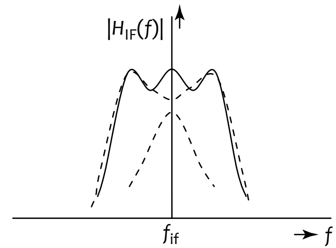


Fig. 6.11 *Using over-coupled transformers in the IF stage*

(iii) We can use 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 3 humps—one at the centre and one each on either side of it.

Nowadays, high-frequency op-amp based single-chip IF amplifiers are available.

Image Frequency Rejection and IFRR 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 $(f_c + f_{if})$. Now, if there is another station operating with a carrier frequency of $(f_0 + f_{if}) = (f_c + 2f_{if})$ 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 $(f_c + 2f_{if})$ 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

(i) The Q of the tuned circuits in the RF stages. (higher the values of Q, better is the image frequency rejection)

(ii) The value of the IF for the receiver—higher the value of the IF, better is the image frequency rejection

(iii) 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:

$$\boxed{\text{IFRR} \triangleq 10\log_{10}\left|\frac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f_c')}\right|^2} \tag{6.1}$$

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 behaviour of a parallel resonant circuit. We shall now examine this briefly.

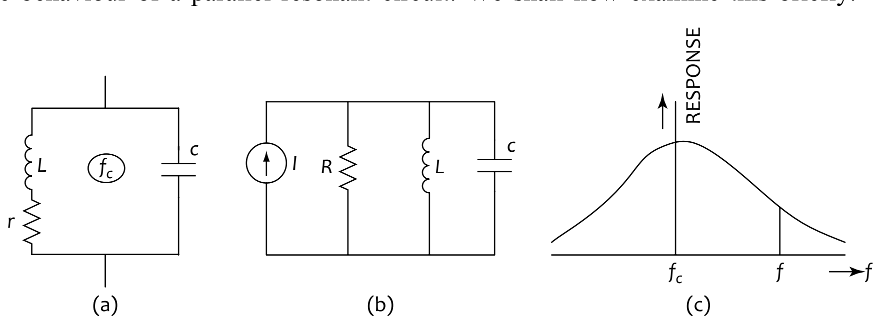


Fig. 6.12 **(a)** *A parallel resonant circuit* **(b)** *Its equivalent circuit* **(c)** *Its frequency response*

Let R be the equivalent parallel resistance which takes care of the small series resistance r associated with the coil of inductance L. Then

$$\text{admittance at resonance} \quad Y_{f_c} = \frac{1}{R}$$

At some frequency $f \neq f_c$, the admittance of the circuit is given by

$$Y_f = \frac{1}{R} + j\left(\omega C - \frac{1}{\omega L}\right) \tag{6.2}$$

If Q is the figure of merit of the tuned circuit at resonance, and if

$$\frac{1}{Y_{f_c}} \triangleq A_c \quad \text{and} \quad \frac{1}{Y_f} \triangleq A \tag{6.3}$$

We may write

$$\frac{A_c}{A} = 1 - jQ\left(\frac{\omega_c}{\omega}\right) + j\left(\frac{\omega}{\omega_c}\right)Q$$

$$= 1 + jQ\left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right) \tag{6.4}$$

$$\therefore \quad \left|\frac{A}{A_c}\right|^2 = \frac{1}{1 + Q^2\left(\dfrac{\omega}{\omega_c} - \dfrac{\omega_c}{\omega}\right)^2} = \frac{1}{1 + x^2Q^2} \tag{6.5}$$

where, $$x \triangleq \left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right) \tag{6.6}$$

$$\therefore \quad \left|\frac{A}{A_c}\right| = \frac{\text{Off-resonance response}}{\text{Response at resonance}} = \frac{1}{\sqrt{1 + x^2Q^2}} \tag{6.7}$$

Equation (6.7) 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. When we consider the image frequency, $\omega = \omega_c + 2\omega_{if}$. 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 $|A/A_c|_{\omega=\omega_c+2\omega_{if}}$ decreases. Note that Eq. (6.7) is for the case of a single stage of RF amplifier. *For multi-stage case, the relative responses get multiplied.*

From Eqs (6.1) and (6.7), it follows that

$$\text{IFRR} \triangleq 20\log_{10}\sqrt{1 + x^2Q^2}, \text{ where } \quad x = \left(\frac{\omega'}{\omega_c} - \frac{\omega_c}{\omega'}\right),$$

ω' being 2π times the image frequency. (6.7a)

Example 6.1

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?

$$\left|\frac{A}{A_c}\right| = \frac{1}{\sqrt{1+x^2Q^2}}; \quad x = \left(\frac{\omega}{\omega_c} - \frac{\omega_c}{\omega}\right)$$

For the image signal, $\omega = \omega_c + 2\omega_{if}$ $\quad\therefore\quad f = 600\times10^3 + 2\times455\times10^3$

$\therefore$ image frequency $= 1510\times10^3$ Hz when $f_c = 600\times10^3$ Hz

$$\therefore \qquad x = \left(\frac{1510}{600} - \frac{600}{1510}\right) = 2.5166 - 0.3973 = 2.1193$$

$$\therefore \qquad 1 + x^2Q^2 = 1 + 4.49\times(60)^2 = 16.165 \quad \therefore \quad \sqrt{1+x^2Q^2} = \sqrt{16165} = 127.14$$

$$\therefore \qquad \left|\frac{A}{A_c}\right| = \frac{1}{\sqrt{1+x^2Q^2}} = \frac{1}{127.14}$$

$$\therefore \qquad \left|\frac{A_c}{A_c}\right|_{\text{dB}} = \text{Image rejection in dB} = 20\log_{10}127.14 = 42 \text{ dB}$$

If a 2-stage RF amplifier is used, image rejection = 84 dB

Example 6.2

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?

We know that image frequency $f' = f_c + 2f_{if}$

and that local oscillator frequency $f_0 = f_c + f_{if}$

$$\therefore \qquad f_0 = f_c + f_{if} = 1010 \text{ kHz} = 555 \text{ kHz} + f_{if} \text{ kHz}$$

$\therefore$ f_{if} = intermediate frequency = 455 kHz

$\therefore$ image frequency $= f' = 555 \text{ kHz} + 2\times455 \text{ kHz} = 1465$ kHz

$\therefore$ from Eq. (7.7a), we have:

$$\text{IFRR} = \text{Image Frequency Rejection Ratio} = 10\log_{10}\left|\frac{H_{\text{RF}}(f_c)}{H_{\text{RF}}(f')}\right|^2 = 20\log_{10}\sqrt{1+x^2Q^2}$$

But $$x = \left(\frac{f'}{f_c} - \frac{f_c}{f'}\right) = \left(\frac{1465}{555} - \frac{555}{1465}\right) = 2.2608$$

and $$Q^2 = 40 \times 40 = 1600$$

$\therefore$ IFRR $= 20\log_{10}\sqrt{1+(2.26.8)^2 40^2} = 20\log_{10} 90.5 = 39.132$ dB

Example 6.3

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

(i) the image frequency and its rejection ratio for tuning at 1100 kHz

(ii) the image frequency and its rejection ratio for tuning at 25 MHz

$$\text{Image frequency} \quad f' = f_c + 2f_{if}$$

(i) 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

$$f_{if} = 455 \text{ kHz}$$

$\therefore$ f', the image frequency $= f_c + 2f_{if} = (1100 + 910)\text{kHz} = 2010$ kHz

From Eq. (6.7a), we have

Image Frequency Rejection Ratio = IFRR $= 10\log_{10}\left|\dfrac{H_{RF}(f_c)}{H_{RF}(f')}\right|^2$ 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 $Q^2 = 100 \times 100 = 10^4$

$\therefore$ $$\text{IFRR} = 20\log_{10}\sqrt{1+(1.28)^2 10^4} = 20\log_{10} 128 = 42.14 \text{ dB}$$

(ii) Assuming the same Q for the antenna coupling circuit when the receiver is tuned to 25 MHz, we have, in this case,

$$f_c = 25 \text{ MHz}; \quad f_{if} = 455 \text{ kHz}$$

$\therefore$ f' = Image frequency $= f_c + 2f_{if} = 25 \text{ MHz} + 910 \text{ kHz} = 25{,}910$ kHz

$$x = \left(\frac{f'}{f_c} - \frac{f_c}{f'}\right) = \left(\frac{25.910}{25} - \frac{25}{25.910}\right) = (1.0364 - 0.9648) = 0.0716$$

$\therefore$ $$\text{IFRR} = 20\log_{10}\sqrt{1+10^4(0.0716)^2} = 20\log_{10}\sqrt{52.2656} = 17.18 \text{ dB}$$

6.3.3 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 $f_{01} = (f_{s1} + f_{if})$. 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 $f_{02} = (f_{s1} - f_{if})$. 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 $f_{s2} = (f_{02} - f_{if})$ and that the signal frequency f_{s1} which is equal to $(f_{02} + f_{if})$ 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 now discuss the various factors affecting the choice of the value of the intermediate frequency, f_{if}, of the receiver.

6.3.4 Choice of the Value of IF

The following are the factors governing the choice of the IF of a superheterodyne receiver.

(i) 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.

(ii) A lower value of IF improves the selectivity of the receiver and reduces the adjacent channel interference.

(iii) 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 ($f_c + 2f_{if}$), 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.

6.3.5 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-wave bands, the two sections of the ganged condenser are made exactly identical; the minimum and maximum values of capacitance in each section being about 50 p.f. and 500 p.f. 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 2-*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 3-point tracking. These various conditions are shown in Fig. 6.13(a), (b) and (c).

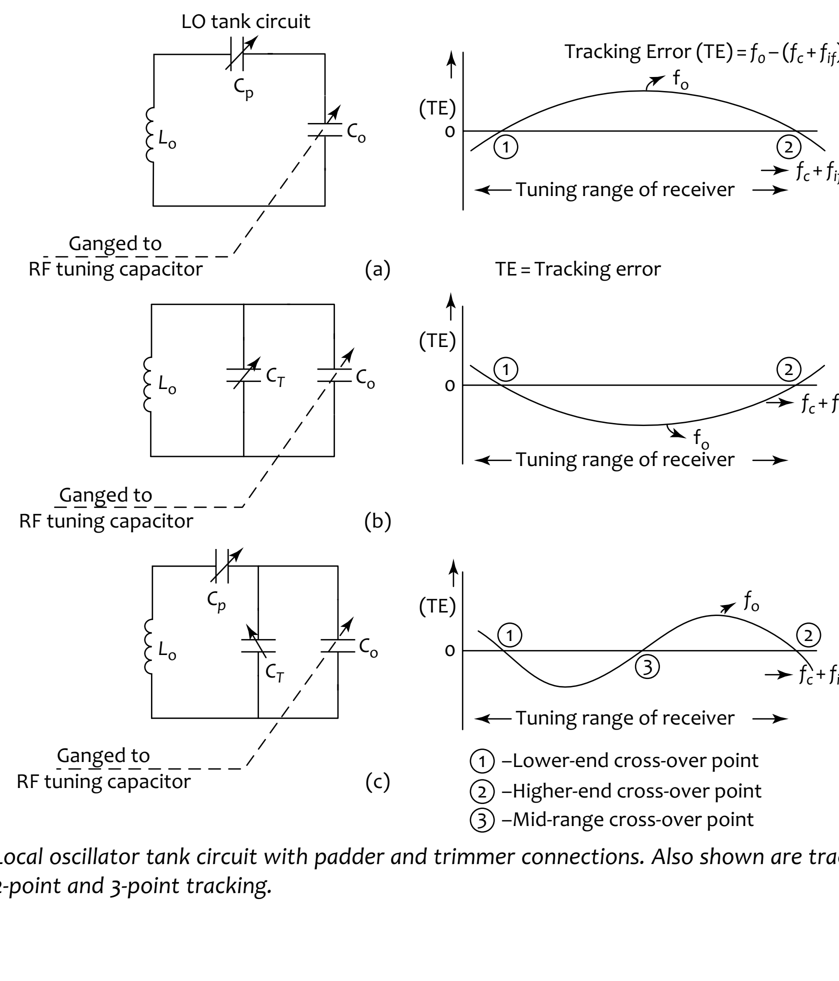


Fig. 6.13 *Local oscillator tank circuit with padder and trimmer connections. Also shown are tracking curves for 2-point and 3-point tracking.*

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 some three or four times to get correct tracking. It may be noted in this connection that to a large extent the trimmer capacitor determines the higher-end cross-over point while the padder capacitor determines the lower-end cross-over point. The mid-range cross-over point is determined by the inductance L_0.

6.3.6 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 Q's using L and C, to get an IF bandwidth of 200 kHz at a centre 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).

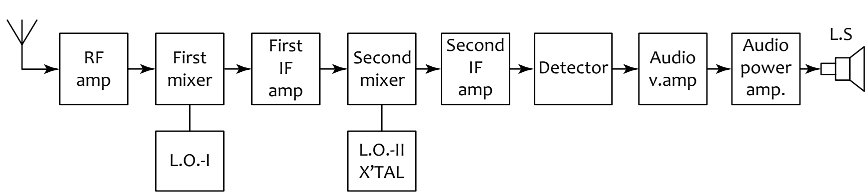


Fig. 6.14 *Double heterodyne receiver*

6.3.7 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.

(a) 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 microvolts or less). However, the loudspeaker needs about 1W 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 volt 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 watt in a resistance of appropriate value connected in the place of the loudspeaker'.

Note: The artificial antenna, comprising an inductance of 20 microhenries in series with a 200 p.f. 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. 6.15. 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.

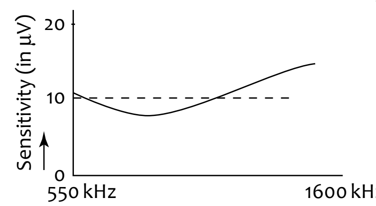


Fig. 6.15 *Typical sensitivity curve of a standard AM broadcast receiver*

(b) 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 6.16 shows the typical selectivity curve of a receiver.

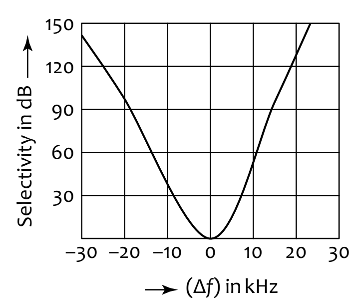


Fig. 6.16 *Typical selectivity curve of an AM broadcast receiver. '0' on X-axis corresponds to desired frequency*

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.*

(c) 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.

(i) Inter-modulation frequency components may be generated when the desired signal mixes with an interfering signal in a non-linear way.

(ii) Inter-modulation frequency components may be produced even by the non-linearities, if any, present in the detector stage.

(iii) 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.

(iv) Distortion of the message signal can arise also due to the IF amplifier frequency response being not constant over its bandwidth of 10 kHz.

(v) Poor low frequency and/or high frequency response as well as non-flat midband 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. 6.17. 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).

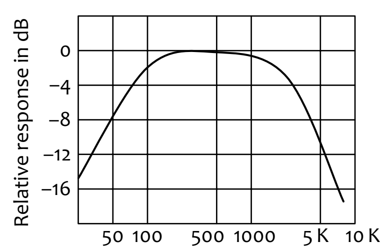


Fig. 6.17 *Typical fidelity curve of a standard broadcast receiver*

(d) 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, itself 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.

6.4 SSB TRANSMITTERS

Single sideband transmission has, as mentioned earlier in Chapter 4, 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.

6.4.1 SSB-SC Transmitter

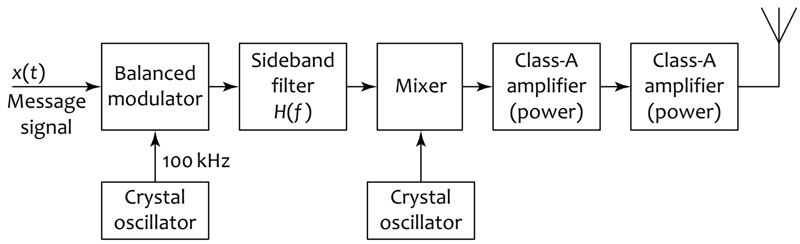


Fig. 6.18 *SSB-SC transmitter using filter method of generating the SSB-SC signal*

As mentioned earlier in Chapter 4, 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 atleast 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. 6.19.

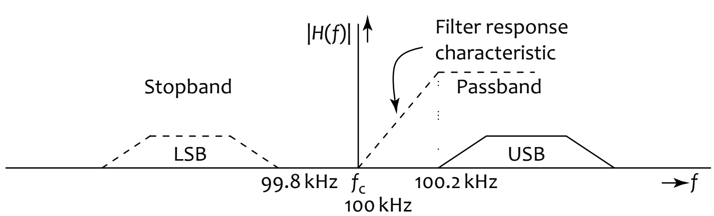


Fig. 6.19 *DSB-SC signal spectrum and the sideband filter response characteristic*

If the lower sideband is to be suppressed, the filter response can change over from passband to stopband over the 100.2 kHz to 100 kHz frequency interval; 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 *R-L-C* filters cannot provide.

6.4.2 Sideband Filters

The Q-value required for the sideband suppression filter depends on (i) the centre frequency (ii) the amount of attenuation needed for the unwanted sideband, and (iii) the transition bandwidth available. It is given by

$$Q = \frac{f_c}{(\Delta f)}\left[0.25\sqrt{A}\right] \tag{6.8}$$

where,

f_c = Centre 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, $\Delta f = 200$ 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.

6.4.3 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 ay 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.

6.4.4 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.

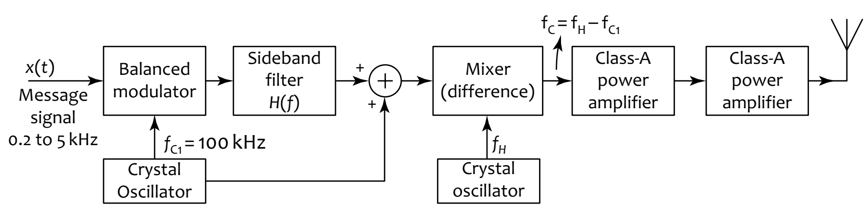


Fig. 6.20 *Block diagram of a pilot-carrier SSB transmitter*

6.5 SSB RECEIVERS

6.5.1 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 it have 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 6.21 shows the block diagram of a communication receiver meant to receive SSB-SC signals in the HF range by employing double conversion.

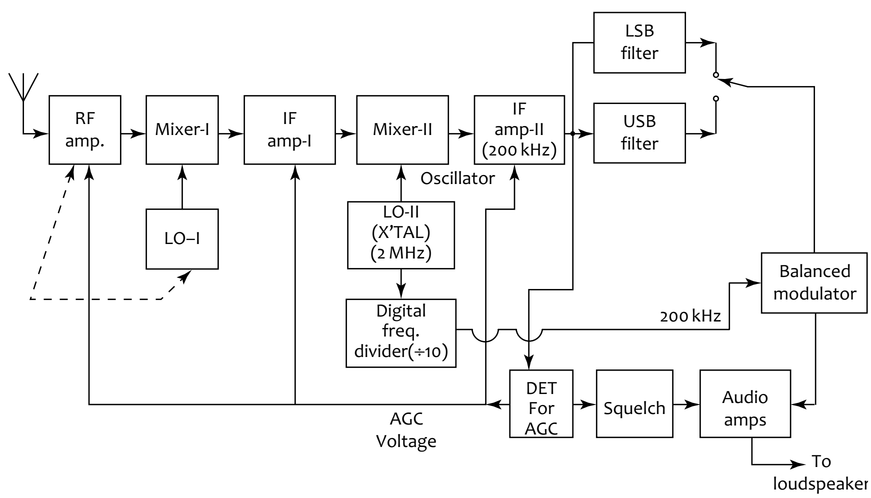


Fig. 6.21 *Block diagram of a SSB-SC receiver*

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 *n* order to avoid annoying sounds being produced by the loudspeaker in the absence of a strong desired signal.

6.5.2 SSB-Pilot Carrier Receiver

These receivers are of the double-conversion type and they make use of the pilot carrier to ensure frequency synchronization with the transmitted carrier.

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 a 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.

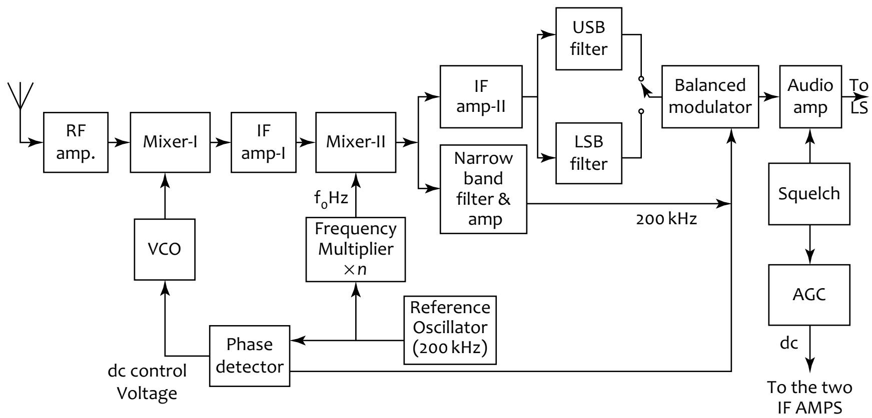


Fig. 6.22 *Block diagram of a pilot-carrier SSB-receiver*

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 narrowband filter has its passband 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 lowpass filter) acts as a coherent detector whose output is the modulating audio signal. After voltage and power amplification, this goes to the loudspeaker.

6.5.3 ISB Transmitter

As explained in Section 4.5, 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. 6.23, 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.

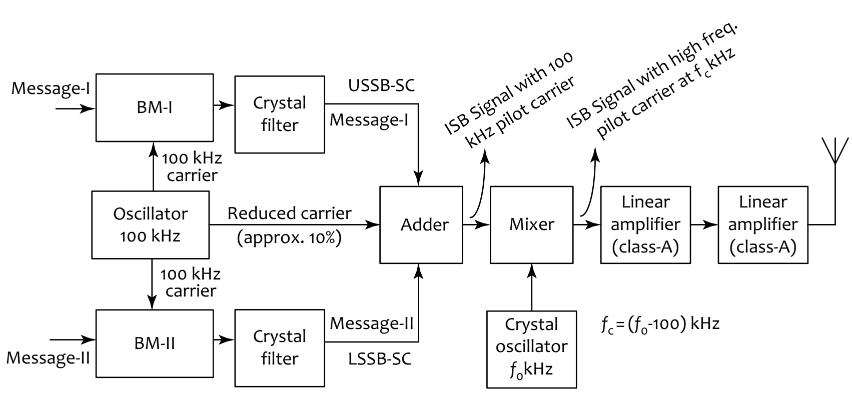


Fig. 6.23 *Block schematic diagram of a pilot carrier ISB transmitter*

6.5.4 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 100 kHz frequency. As shown in Fig. 6.24, the IF amp-II output is simultaneously applied to (i) very narrowband 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.

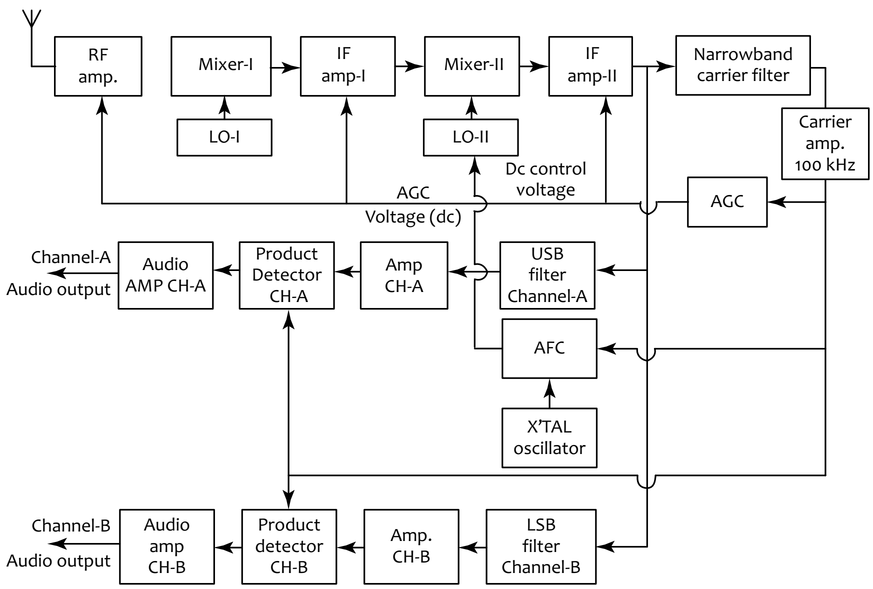


Fig. 6.24 *Block diagram of a pilot carrier ISB receiver*

The 100 kHz carrier from the narrowband 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.

6.6 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.

As discussed in Chapter 5, a WBFM signal may be generated either by the indirect method, or the direct method.

6.6.1 FM Transmitter Based on Indirect Method of Generation of WBFM (Armstrong Method)

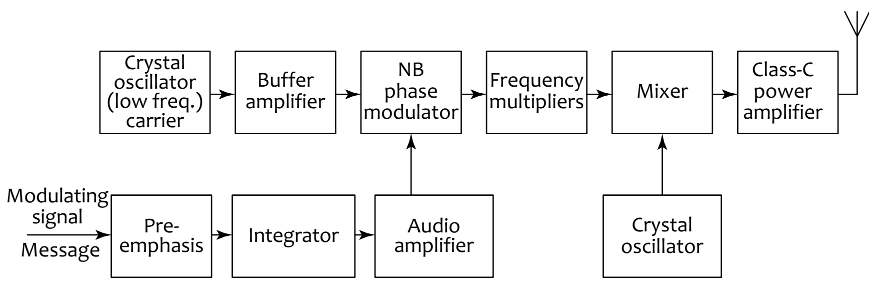


Fig. 6.25 *FM transmitter based on the indirect method*

As explained in Chapter 5, initially a low-frequency carrier is used. A narrow band phase modulator of the type shown in Fig. 5.8 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.

6.6.2 FM Transmitter Based on Direct Method of Generation

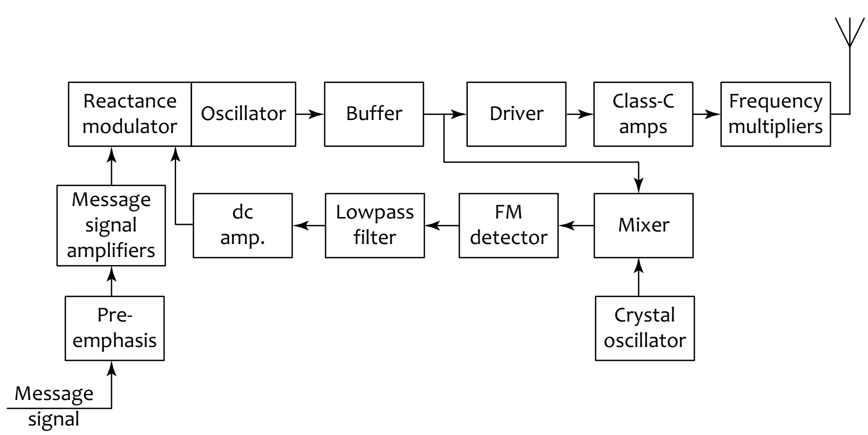


Fig. 6.26 *Block diagram of an FM transmitter using direct method of generation of WBFM*

As mentioned in Section 6.5, 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 lowpass 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.

6.6.3 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 to 108 MHz. The standard value of the intermediate frequency for these receivers is 10.7 MHz. Figure 6.27 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 r.f. 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 cascode 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 $2 \times$ 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).

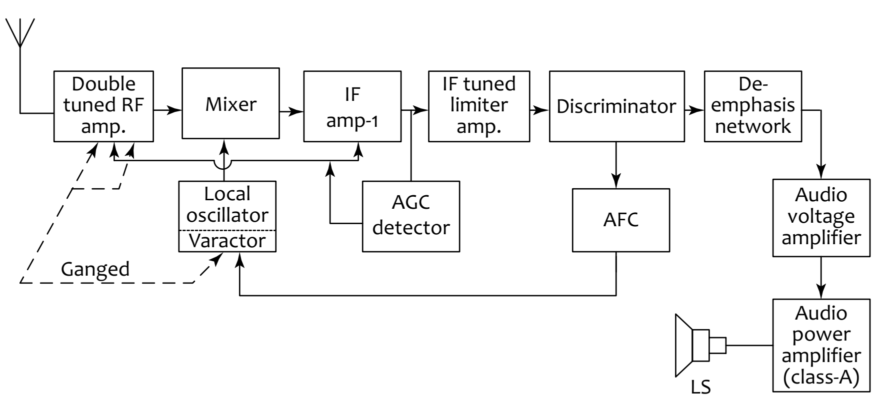


Fig. 6.27 *Block diagram of a superheterodyne FM broadcast receiver*

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 cutoff, 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 9 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-emphasising, 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.

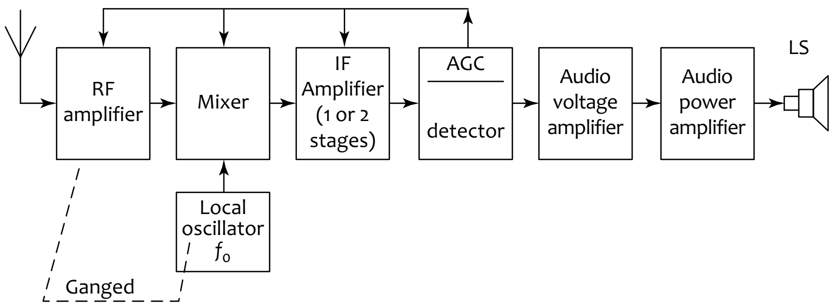
Generally, AFC is provided to keep the frequency of the local oscillator at the value that produces the correct intermediate frequency. If the average value of the intermediate frequency differs from the centre frequency of the dual-slope discriminator then, a dc voltage will be developed at the output of the discriminator. The polarity of this dc voltage will depend on the direction of deviation of the IF with respect to the centre 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.

SUMMARY

1. 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.
2. **Sections of a transmitter** (a) the exciter section—the carrier oscillator, frequency *stabilization systems*, buffer amplifier; (b) Modulation section—audio pre-amplifiers, voltage amplifiers and power amplifiers (c) RF section—RF power amplifiers (class-C, if high-level modulation is used; class-A or class-AB if low-level modulation is used).
3. **Effective Radiated Power (ERP)** The effective radiated power from a transmitter is the average RF power from the transmitter multiplied by the loss of the transmission lines and the gain of the antenna.
4. **Primary Service area of a transmitter** It is the area around the transmitting antenna comprising all points at which the field strength due to the signal radiated by the antenna is not below a certain prescribed value—generally 5 to 10 mV per metre.

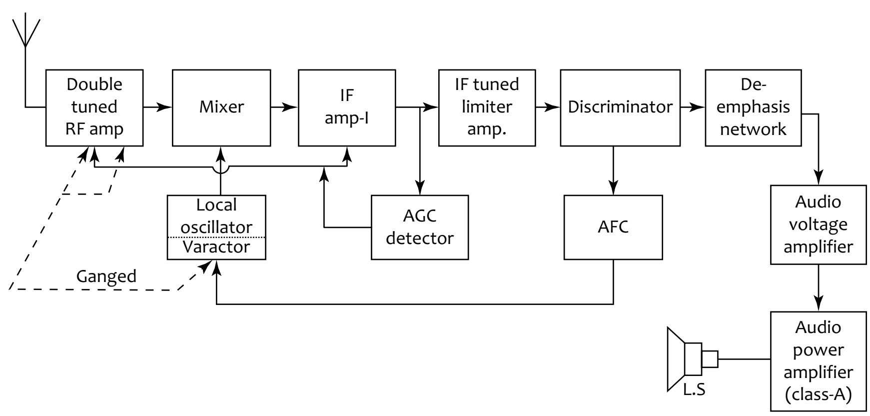
5. **Classification of transmitters** Transmitters are classified in several different ways—on the basis of operating frequency band, on the basis of modulation employed, on the basis of service provided, and on the basis of transmitted power.
6. **Functions of a receiver** To pick up any desired signal, 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.
7. **Classification of receivers** They are classified in different ways: (i) according to the type of modulation of the received signal, (ii) according to the frequency range of operation, and (iii) According to the configuration of the receiver—TRF, superheterodyne, etc.
8. **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 sky-wave 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.
9. **FM broadcast transmitters** They handle audio frequencies 06 up to 15 kHz; used mostly for high-quality speech and music. They operate in the VHF frequency band from 88 MHz to 108 MHz and depend on line-of-sight propagation and so service area is limited to about 40–80 km. Carrier frequency stability of ± 2kHz needed. Maximum frequency deviation is ±75 kHz. Adjacent carrier separation is 200 kHz. Power of the order of 100 kW are used.
10. **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.
11. **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.
12. **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.
13. **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.
14. **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.
15. **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.
16. **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.

17. **Superheterodyne broadcast receiver**



18. **Why the local oscillator frequency f_0 is kept greater than the carrier frequency f_c** In a superheterodyne receiver, the difference between f_0 and f_c should be equal to $f_{i.f}$ of the receiver. Thus, f_0 may be greater than f_c or less than f_c. But it is always arranged to be grater than f_c as otherwise, the tuning capacitor range required will be far greater than what can be obtained in practice.
19. **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 3 or more stagger-tuned stages, or 3 or more identically tuned IF stages with loose coupling of the inter-stage transformers.
20. **Image frequency** If a receiver with intermediate frequency $f_{i.f}$ is tuned to a carrier frequency f_c, the corresponding image frequency is $f' = \left(f_c + 2f_{if}\right)$
21. **Image Frequency Rejection Ratio (IFRR)** IFRR $\underline{\Delta}\, 10\log_{10}\left|\dfrac{H_{\mathrm{RF}}(f_c)}{H_{\mathrm{RF}}(f')}\right|^2$. 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. Should be atleast 40 dB.
22. **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.
23. **Choice of IF** (i) IF should be outside the tuning range of the receiver.
    (ii) Lower value of IF reduces adjacent channel interference.
    (iii) Higher value of IF improves image rejection.
    Usual Values 455 to 465 kHz for AM receivers and 10.7 MHz for FM receivers.
24. **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 the IF. This is referred to as tracking. In practice, perfect tracking cannot be achieved exactly over the entire tuning range of the receiver.
25. **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.
26. **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.

27. **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.
28. **Receiver parameters** (i) Sensitivity (ii) Selectivity (iii) Fidelity (iv) Noise figure
29. **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.
30. **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.
31. **FM transmitter based on Indirect Method** Refer Fig. 6.24.
32. **FM transmitter based on Direct Method** Refer Fig. 6.25. Note that it is imperative to make use of a carrier frequency stabilization circuit for FM transmitters based on the direct method.
33. **FM broadcast receiver block diagram**



34. **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.
35. **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. Herbert Taub and Donald Schilling: *Principles of Communication Systems*, Tata McGraw-Hill Kogakusha International Student Edition, 1971.
2. Robert J Schoenbeck: *Electronic Communications—Modulation and Transmission*, Second Edition, Prentice Hall Career & Technology, Englewood Cliffs, New Jersey, USA, 1992.
3. Simon Haykin: *Communication Systems*, Third Edition, John Wiley & Sons, 1995.
4. Leon W Couch, II: *Digital and Analog Communication Systems*, Sixth Edition, Pearson Education Asia, 2001.
5. John G Proakis and Masoud Salehi: *Communication Systems Engineering*, Second Edition, (Eastern Economy Edition), Prentice-Hall of India Pvt. Ltd, New Delhi, 2002.
6. Dennis Roddy and John Coolen: *Electronic Communications*, Third Edition (Eastern Economy Edition), Prentice-Hall of India Pvt. Ltd, New Delhi, 1994.
7. Roy Blake: *Electronic Communication Systems*, Second Edition, Thomson-Delmar, Singapore, 2005.
8. Kennedy and Davis: *Electronic Communication Systems*, Fourth Edition, McGraw-Hill International Editions, 1992.
9. A Bruce Carlson and Paul B Crilly: *Communication Systems*, Fourth Edition, McGraw-Hill, Singapore, 2002.
10. F E Terman: *Electronic and Radio Engineering*, Fourth Edition, McGraw-Hill Book Company, 1955.

REVIEW QUESTIONS

1. What are the functions of a transmitter?
2. Name the important sections of a transmitter.
3. Define the terms: ERP and Primary Service Area.
4. Define 'high-level modulation' and 'low-level modulation', and discuss the advantages and disadvantages of each.
5. Draw the block schematic diagram of an AM broadcast transmitter and explain the function of each block.
6. Explain the neutralization techniques adopted in the RF amplifiers of a transmitter.
7. Draw the block schematic diagram of a TRF type of AM broadcast receiver. Explain its functioning and its deficiencies.
8. What is the basic principle of a superheterodyne broadcast receiver? How does it overcome the limitations noted in the case of a TRF receiver?
9. Draw the block schematic diagram of a superheterodyne AM broadcast receiver and with ist help, explain the working of the receiver.
10. 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.
11. What is meant by an image signal? What are the steps generally taken to minimize image signal interference?
12. With reference to a superheterodyne broadcast receiver, explain what is meant by tracking. How is it ensured?
13. Distinguish between 2-point and 3-point tracking.
14. Discuss the factors governing the choice of IF for a superheterodyne receiver.

15. Clearly justify the following statements.
    (i) Good image signal suppression requires that the IF be high.
    (ii) Good adjacent channel selectivity can be obtained by choosing a low value of IF.
16. Define and explain the terms: 'Sensitivity', 'Selectivity', and 'Fidelity'. What are the various factors that influence these parameters?
17. With the help of a neat block schematic diagram, explain the working of an SSB-SC transmitter.
18. Clearly explain the need for the following: 'SSB transmitters use a low-frequency carrier initially'.
19. 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?
20. Draw the block schematic of a pilot-carrier SSB transmitter.
21. With the help of a neat block schematic diagram, explain the working of an SSB-SC receiver.
22. Draw the block diagram of a pilot-carrier SSB receiver and explain its working.
23. Clearly explain how an ISB signal is generated by drawing the block diagram of the relevant portion of a pilot-carrier ISB transmitter.
24. With the help of a neat block schematic diagram, explain the working of a pilot-carrier ISB receiver.
25. Explain the working of a 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.
26. Explain the working of a FM broadcast transmitter employing the indirect method of generation of WBFM by drawing the block diagram.
27. Draw the block schematic diagram of a FM broadcast receiver, and explain its working.

FILL IN THE BLANKS

1. The important sections of any transmitter are _________, _________ and _________.
2. An AM broadcast transmitter is said to be employing high-level modulation if _______.
3. An AM broadcast transmitter is said to be employing low-level modulation if ________.
4. In an AM transmitter, the RF amplifier to be used subsequent to the modulator stage are _________.
5. _________ level modulation requires large amounts of modulating signal power.
6. The main disadvantage of the TRF receivers is __________.
7. In a superheterodyne AM broadcast receiver, the RF amplifier helps in improving the ______ of the receiver and also in suppressing ______ ______.
8. In a superheterodyne receiver, most of the gain of the receiver is obtained in the _______ ,________stage.
9. The audio power amplifier of an ordinary AM broadcast superheterodyne receiver is a class __ ________ amplifier.
10. Double spotting in a superheterodyne receiver is the phenomenon of ____________.
11. In an AM broadcast superheterodyne receiver, the local oscillator frequency is arranged to be _______ (higher/lower) than the signal frequency to which the receiver is tuned.
12. In a double conversion superheterodyne receiver, the first IF is ________ (higher/lower) than the second IF.
13. Double conversion is used in certain receivers in order to achieve good _____ as well as good ____.
14. In filter method of generation of SSB signals, the sideband filters used are of _____, or ______, or ______ type.
15. A limiter is used in an FM broadcast receiver in order to _________.
16. ISB receivers use ______ type of detectors for recovering the audio.
17. FM transmitters handle audio frequencies upto _______ kHz.

MULTIPLE CHOICE QUESTIONS

1. 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
2. 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
3. A pre-emphasis circuit provides extra noise immunity by
   (a) boosting the bass frequencies
   (b) amplifying the higher audio frequencies
   (c) pre-amplifying the whole audio band
   (d) converting PM to FM
4. 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
5. 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
6. 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
7. 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
8. 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
9. Harmonic generators use
   (a) class-A amplifiers (b) class-AB amplifiers
   (c) class-B amplifiers (d) class-C amplifiers

10. 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
11. 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
12. 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.
13. 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
14. 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
15. 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
16. 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
17. An ISB transmitter is generating the ISB signal initially with a low carrier frequency of 100 kHz. To increase the carrier frequency to the final value required, we may use
    (a) a mixer
    (b) a chain of frequency multipliers
    (c) a combination of mixer and frequency multiplier chain
    (d) none of the above
18. A squelch circuit acts as
    (a) demodulator (b) a switch which acts at a set level.
    (c) an oscillator (d) a filter

PROBLEMS

1. A superheterodyne receiver has an IF of 460 kHz. Its RF amplifier is tuned to an incoming signal of 700 kHz carrier frequency. If at this frequency the tuned circuit of the RF amplifier has a Q of 60, determine the image frequency rejection in dB.
2. 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.
3. A 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?

4. A SSB transmitter uses a set-up of the form shown in Fig. P-6.1 to generate the SSB signal using filter method.

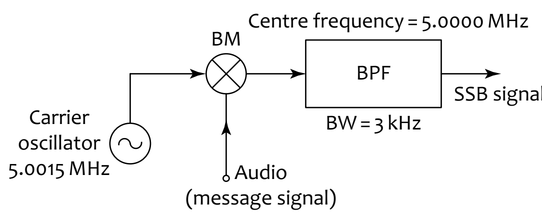


Fig. P-6.1

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

Key to Multiple Choice Questions

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

7 Probability and Random Processes

By going through this review chapter, the student

- *can thoroughly revise all the key concepts in probability and random processes*
- *will be in a position to apply the results of this chapter to the study of noise in the next chapter and the study of noise performance of AM and FM systems in Chapter 9*

7.1 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 apriori 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 intersymbol 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—possibly in a full one-semester course—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.

7.2 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.

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 σ-field, 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, 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 7.2, it is possible to derive the following basic properties of P.

(i) $P(A) = 1 - P(\overline{A})$. (ii) $P(\phi) = 0$ (iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

7.3 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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \begin{cases} \dfrac{P(AB)}{P(B)} & ; \quad P(B) \neq 0 \\ 0 & ; \quad \text{otherwise} \end{cases} \tag{7.1}$$

Remarks

1. If A and B are mutually exclusive, i.e., if $A \cap B = 0; P(A \mid B) = 0$.
2. If $P(A \mid B) = P(A)$, i.e., the occurrence of B does not affect the probability of A, events A and B are said to be statistically independent. In this case,

$$P(A \mid B) = \frac{P(AB)}{P(B)} = P(A) \Rightarrow P(AB) = P(A) \cdot P(B) \tag{7.2}$$

Example 7.1

A die is thrown and you are told that the outcome is even. Then what is the probability that the result is 2?

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 $\quad A \cap B = A, P(A \mid B) = \dfrac{P(AB)}{P(B)} = \dfrac{P(A)}{P(B)} = \dfrac{1/6}{1/2} = \dfrac{1}{3}$

$\therefore \quad P(2 \mid \text{even}) = 1/3$

Total Probability Theorem Let the events $A_1, A_2, \ldots, A_n$ belong to the same probability space. Let them be such that

$$A_1 \cup A_2 \cup A_b \ldots\ldots \cup A_n = S$$

If B is any arbitrary event, also belonging to the same probability space then

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots\ldots + P(B \mid A_n)P(A_n)$$

$$= \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

This is called the total probability theorem.

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.

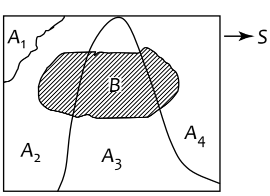


Fig. 7.1 *Partitioning of S*

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

$$\therefore \quad P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)} \tag{7.3}$$

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.

(i) $P(AB) = P(A)\cdot P(B)$; $P(BC) = P(B)\cdot P(C)$ and $P(AC) = P(A)\cdot P(C)$

(ii) $P(ABC) = P(A)\cdot P(B)\cdot P(C)$

In general, n events, $A_1, A_2, ..., A_n$ are said to be independent if for every $k < n$ the events $A_1, A_2, ..., A_k$ are independent and further, if

$$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2)\ldots P(A_n)$$

Example 7.2

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 the box B_1?

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 =

$$P(D) = P(D \mid B_1)P(B_1) + P(D \mid B_2)P(B_2) + P(D \mid B_3)P(B_3) + P(D \mid B_4)P(B_4) + P(D \mid B_5)P(B_5)$$

$$= 0.2[0.05 + 0.2 + 0.03 + 0.1 + 0.14] = 0.104$$

$\therefore$ 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, $P(B_1 \mid D)$.

$$P(B_1 \mid D) = \frac{P(B_1, D)}{P(D)} = \frac{P(D \mid B_1)P(B_1)}{P(D)}$$

But $P(D \mid B_1) = 0.05$, $P(D) = 0.104$ and $P(B_1) = 0.2$

$$P(B_1 \mid D) = \frac{0.05 \times 0.2}{0.104} = \frac{0.01}{0.104} = 0.0961$$

7.4 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:

(i) $\{\boldsymbol{X} \le x\}$, i.e., $\{\boldsymbol{X}(\xi) \le x\}$ is an event for every real number x

(ii) $P\{\boldsymbol{X}(\xi) = +\infty\} = 0 = P\{\boldsymbol{X}(\xi) = -\infty\}$

The mapping referred to in the above definition, therefore, has S, the set of all outcomes as its domain and R, the set of real numbers, as its range. 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

$$\boldsymbol{X}(\text{heads}) = 1 \text{ and } \boldsymbol{X}(\text{tails}) = 0$$

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

$$\boldsymbol{Z} = \boldsymbol{X} + j\boldsymbol{Y} \tag{7.4}$$

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

$$F_X(x) \;\underline{\Delta}\; P\{\boldsymbol{X} \le x\} \tag{7.5}$$

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 7.3

For the random experiment of tossing of a coin, let us define a random variable $\boldsymbol{X}$ *by saying that*

$$\boldsymbol{X}(\xi = \text{Heads}) = 1 \quad \text{and} \quad \boldsymbol{X}(\xi = \text{Tails}) = 0.$$

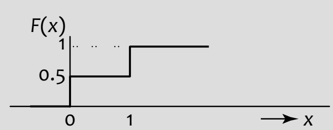


Fig. 7.2 *CDF of* **X**

The CDF of the random variable (r.v) **X** *will be as shown in Fig. 7.2, since P{Heads} = P{Tails} = 0.5. Since* $F(x) = P\{\boldsymbol{X} \le x\}$*, when* $x < 0$*,* $F(x) = 0$*. when* $x \le 0$*, but less than one,* $F(x) = 0.5$ *and for* $x \ge 1$*,* $F(x) = 1$*. Thus, we get a staircase type of CDF.*

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., $\lim_{\epsilon \to 0} F(x + \epsilon) = F(x); \epsilon > 0$
5. $F(b)$-$F(a) = P[a < \boldsymbol{X} \le b]$
6. $P[\boldsymbol{X} = x_1] = F(x_1) - F(x_1^-)$, where, $F(x_1^-) \,\underline{\Delta}\, \lim_{\epsilon \to 0} F(x_1 - \epsilon);\ \epsilon > 0$

7.4.1 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 has 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 r.v. nor a continuous r.v. is called a *mixed random variable.*

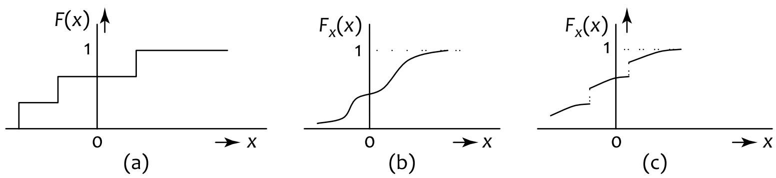


Fig. 7.3 **(a)** *CDF of a discrete r.v* **(b)** *CDF of a continuous r.v.* **(c)** *CDF of a mixed r.v.*

Definition The Probability Density Function (PDF) of a r.v. $\boldsymbol{X}$ is defined as the derivative with respect to x of its CDF, viz., $F_x(x)$

$$\therefore \qquad f_x(x) = \frac{dF_x(x)}{dx} \tag{7.6}$$

If $\boldsymbol{X}$ is a discrete r.v., we know that its $F_X(x)$ will be of the staircase type. Hence, as shown in Fig. 7.3, its probability density function (PDF) will be zero everywhere except at the points of discontinuity, where it will have impulses.

The PDF, $f_X(x)$ of a continuous r.v. $\boldsymbol{X}$ will be a continuous function of x. The PDF of a mixed r.v. 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 r.v., 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 $\sum_i p_i = 1$.

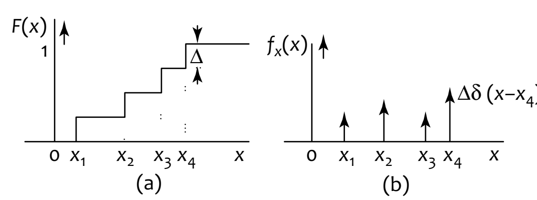


Fig. 7.4 **(a)** *CDF, and* **(b)** *PDF of a discrete random variable*

7.4.2 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):

(a) Continuous Random Variables

(i) Uniform Random Variable A random variable $\boldsymbol{X}$ is called a uniform r.v. if its probability density function $f_X(x)$ is given by

$$f_X(x) = \begin{cases} \dfrac{1}{(x_2 - x_1)} & ; \; x_1 \leq x \leq x_2 \\ 0 \;\; \text{otherwise} \end{cases} \quad ; \; -\infty < x_1 < x_2 < \infty \tag{7.7}$$

Such an r.v. is generally denoted by $\boldsymbol{U}(x_1, x_2)$. The cumulative distribution function (CDF) of this r.v. is as shown in Fig. 7.5(b) and is given by

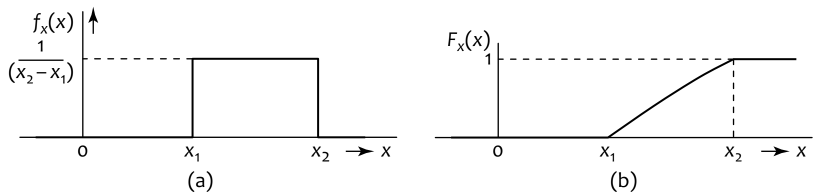


Fig. 7.5 **(a)** *PDF and* **(b)** *CDF of a uniformly distributed r.v.*

$$F_X(x) = \begin{cases} 1 \text{ for } x \geq x_2 \\ \dfrac{(x-x_1)}{(x_2-x_1)} \quad \text{for } x_1 \leq x \leq x_2 \\ 0 \text{ for } x < x_1 \end{cases} \tag{7.8}$$

The uniform r.v. 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 r.v., its range of values being from 0 to 2π.

(ii) Gaussian or Normal Random Variable The r.v. $\boldsymbol{X}$ 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

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} \tag{7.9}$$

This density function has a shape as shown in Fig. 7.6 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 $N(0, 1)$.

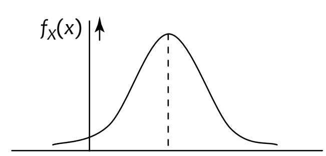


Fig. 7.6 *Density function of a Gaussian r.v. with mean m and variance σ^2*

Gaussian distribution function, $F_X(x)$ is given by

$$F_X(x) = P[\boldsymbol{X} \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-m)^2/2\sigma^2} dy$$

$$= \int_{-\infty}^{\left(\frac{x-m}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{7.10}$$

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 $N(0, 1)$ r.v., Eq. (7.10) reduces to

$$g_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = P[\boldsymbol{X} \leq x] \tag{7.11}$$

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

$$Q(x) = 1 - g_X(x) = P[X > x] \tag{7.12}$$

This Q-function, which is extensively tabulated, has the following important properties.

$$Q(-x) = 1 - Q(x) \tag{7.13}$$

$$Q(0) = \frac{1}{2} \tag{7.14}$$

$$Q(\infty) = 0 \tag{7.15}$$

(iii) Rayleigh Random Variable A random variable $\boldsymbol{X}$ is said to have a Rayleigh distribution with parameter σ^2 if its density function is

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2}\ e^{-x^2/2\sigma^2}; & x \ge 0 \\ 0 & ; x < 0 \end{cases} \tag{7.16}$$

The Rayleigh distributed r.v. has a mean value of $\sigma\sqrt{\dfrac{\pi}{2}}$ and a variance of $\left(2 - \dfrac{\pi}{2}\right)\sigma^2$. The shape of Rayleigh density function is shown in Fig. 7.7(a)

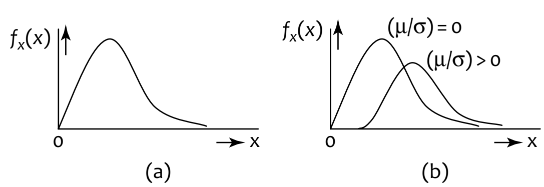


Fig. 7.7 **(a)** *Rayleigh density function* **(b)** *Ricean distribution*

If a bandpass signal has an identically distributed Gaussian zero–mean random processes as its in-phase 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.

(iv) 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

$$f_X(x) = \left[\frac{1}{\sigma^2}\ x e^{-(x^2+\mu^2)/2\sigma^2}\right] \cdot I_0\left(\frac{\mu\, x}{\sigma^2}\right) \tag{7.17}$$

where,

$$I_0(\alpha) \triangleq \frac{1}{\pi}\int_0^{\pi} e^{\alpha\cos\theta}\, d\theta \tag{7.18}$$

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. 7.7(b)) is somewhat similar to that of a Rayleigh density function. In fact, as can be seen from Eqs (7.16) and (7.17), the Rice density function simplifies into the Rayleigh density function when the parameter $\mu = 0$.

If a bandpass signal has Gaussian random processes with the same variance but *different* non-zero mean values as its in-phase and quadrature components, it can be shown that its envelope will have a 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 a Ricean distribution for its envelope.

(b) Discrete Random Variables

(i) 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.

(ii) Binomial Random Variable A discrete random variable, $\boldsymbol{X}$, is said to be a binomial random variable with parameters n and p if

$$P[\boldsymbol{X}=k]=\binom{n}{k}p^k q^{n-k} \quad ; \quad 0 \le k \le n \tag{7.19}$$

In fact, this gives the number of 1's in a sequence of 1's and 0's 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.

Example 7.4

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.*

Let $\quad (x-m) \triangleq z \quad \therefore \quad dx = dz$

$$f_X(z)=\left[\int_{-\infty}^{\infty} e^{-z^2/2\sigma^2}dz\right]\frac{1}{\sqrt{2\pi\sigma^2}}$$

Now, let $\quad A \triangleq \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2}dz.$

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 $\dfrac{r^2}{2\sigma^2} \triangleq v$ Then $dv = \dfrac{2rdr}{2\sigma^2} = \dfrac{rdr}{\sigma^2}$

$$\therefore \quad A^2 = 2\pi\sigma^2 \int_0^\infty e^{-v} dv = 2\pi\sigma^2 \qquad \therefore \quad A = \sqrt{2\pi\sigma^2}$$

Hence the factor $\dfrac{1}{\sqrt{2\pi\sigma^2}}$ in $f_X(x)$ is a normalization factor.

7.5 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 r.v. $\boldsymbol{X}$. If we can call this as another random variable $\boldsymbol{Y}$ then

$$\boldsymbol{Y} = g(\boldsymbol{X}) \tag{7.20}$$

We can then talk of the CDF, $F_Y(y)$ of the r.v. $\boldsymbol{Y}$.

$$F_Y(y) = P\left[\xi \in S : g(X(\xi)) \le y\right] \tag{7.21}$$

Now, for $\boldsymbol{Y}$ to be a r.v. 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

$$g(\boldsymbol{X}(x)) \le y$$

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

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \tag{7.22}$$

Example 7.5

If $\boldsymbol{Y} = \boldsymbol{X}^2$, find $f_Y(y)$ in terms of $f_X(x)$.

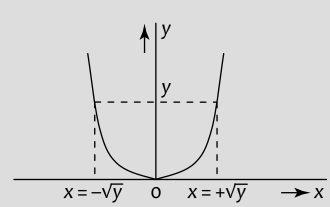


Fig. 7.8 $y = x^2$

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

$$-\sqrt{y} \le x \le \sqrt{y}.$$

$$\therefore \quad F_Y(y) = P\left[-\sqrt{y} \le \boldsymbol{X} \le \sqrt{y}\right] = P\left[\boldsymbol{X} \le \sqrt{y}\right] - P\left[\boldsymbol{X} \le -\sqrt{y}\right]$$

$$\therefore \quad F_Y(y) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right); y > 0$$

To get the corresponding density function, $f_Y(y)$, we have

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right] & \text{for } \ y > 0 \\ 0 & \text{for } \ y < 0 \end{cases}$$

MEAN, VARIANCE AND CHARACTERISTIC FUNCTION 7.6

7.6.1 Mean

The mean, or the expected value of a random variable $\boldsymbol{X}$ with the density function $f_X(x)$, is defined as

$$E\{\boldsymbol{X}\} \triangleq \int_{-\infty}^{\infty} x f_X(x)dx \tag{7.23}$$

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 r.v., we had already seen that

$$f_X(x) = \sum_i p_i \delta(x - x_i) \tag{7.24}$$

Substituting this for $f_X(x)$ in Eq. (7.23), we get

$$E\{\boldsymbol{X}\} = \sum_i p_i x_i, \quad \text{where } \ p_i = P\left[\boldsymbol{X} = x_i\right] \tag{7.25}$$

Example 7.6

*A r.v. **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.

$$\eta_X = \int_{-\infty}^{\infty} x.2e^{-2x}u(x)dx = \int_0^{\infty} 2xe^{-2x}dx = \frac{1}{2}$$

Example 7.7

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.

$$\eta = \sum_{i=1}^{6} i.p_i = (1\times 0.10) + 2(0.12) + 3(0.12) + 4(0.14) + 5(0.20) + 6(0.32)$$

$$= 4.18$$

7.6.2 Variance

The variance of a random variable $\boldsymbol{X}$ with expected value η_X, is defined as

$$\operatorname{Var}[\boldsymbol{X}] = \sigma^2 = E\lfloor(\boldsymbol{X}-\eta_X)^2\rfloor = E\lfloor\boldsymbol{X}^2\rfloor - 2\eta_X E[\boldsymbol{X}] + \eta_X^2 = E[\boldsymbol{X}^2] - \eta_X^2$$

$$\therefore \qquad \sigma^2 = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2 \tag{7.26}$$

Since $\qquad E\lfloor(\boldsymbol{X}-\eta_X)^2\rfloor \geq 0,$ it follows that

$$E[\boldsymbol{X}^2] \geq \{E[\boldsymbol{X}]\}$$

Discrete r.v In this case,

$$\sigma^2 \triangleq \sum_i p_i(x_i - \eta_X)^2 \text{ where, } \quad p_i = P[\boldsymbol{X} = x_i] \tag{7.27}$$

Note: The positive square-root of the variance is referred to as the 'standard deviation'.

Example 7.8

*Find the mean value and the variance of a random variable **X** which is uniformly distributed between x = a to x = b.*

(a) Mean $\qquad \eta_X = \int_{-\infty}^{\infty} x.f_X(x)dx = \int_a^b x.\frac{1}{(b-a)}dx = \frac{b+a}{2}$

(b) Variance $\quad \sigma^2 = \int_{-\infty}^{\infty} E[\boldsymbol{X}-\eta_X]^2 dx = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2$

$$= \left[\left\{\int_a^b x^2.\frac{1}{(b-a)}dx\right\} - \frac{(b+a)^2}{4}\right] = \frac{(b-a)^2}{12}$$

$$\therefore \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

Example 7.9

Find the variance of a Bernoulli random variable.

A Bernoulli r.v. takes the values 1 and 0 with probabilities p and $(1 - p)$

$$\therefore \qquad \eta_X = 1.p + 0.(1-p) = p$$

$$E\lfloor\boldsymbol{X}^2\rfloor = p.1^2 + (1-p).0^2 = p$$

$$\therefore \qquad \sigma^2 = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2 = p - p^2 = p(1-p)$$

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 r.v. $\boldsymbol{X}$ has a variance σ_X^2 and if c is a constant, the following are the properties of variance.

1. $\text{Var}[c.\boldsymbol{X}] = c^2\text{Var}[\boldsymbol{X}] = c^2\sigma_X^2$
2. $\text{Var}\ [c] = 0$
3. $\text{Var}[\boldsymbol{X} + c] = \text{Var}[\boldsymbol{X}] = \sigma_X^2$

7.6.3 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

$$\phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}dx \tag{7.28}$$

If we now define $$\Phi(s) \triangleq \int_{-\infty}^{\infty} f_X(x)e^{sx}dx$$

So that $$\Phi(j\omega) = \phi_X(\omega) \text{ of Eq. (7.28)}$$

Taking the first derivative of $\Phi(s)$ with respect to s, we get

$$\Phi^{(1)}(s) = \int_{-\infty}^{\infty} xf_X(x)e^{sx}dx = E\left[\boldsymbol{X}e^{s\boldsymbol{X}}\right] \tag{7.29}$$

If we take the n^{th} derivative with respect to s, we get

$$\Phi^{(n)}(s) = \int_{-\infty}^{\infty} (x)^n f_X(x)e^{sx}dx = E\left[(\boldsymbol{X})^n e^{s\boldsymbol{X}}\right] \tag{7.30}$$

If we put $s = 0$ in Eqs (7.29) and (7.30), we find that

$$\Phi^{(1)}(0) = \text{first derivative of } \Phi(s) \text{ with respect to } s \text{ at the origin } = E[\boldsymbol{X}] \tag{7.31}$$

$$\Phi^{(n)}(0) = n^{\text{th}} \text{ derivative of } \Phi(s) \text{ with respect to } s \text{ at the origin } = E[\boldsymbol{X}^n] \tag{7.32}$$

i.e., derivatives of various orders of the moment generating function $\Phi(s)$ at the origin give the moments of various orders for the r.v. $\boldsymbol{X}$.

Note: Thus, the characteristic function of an r.v. $\boldsymbol{X}$ helps us in determining moments of various order for $\boldsymbol{X}$ in an easy manner.

Discrete Random Variable If $\boldsymbol{X}$ is a discrete r.v. which takes values x_i with probabilities $p_i = 1, 2, 3, ...,$ then Eq. (7.28) reduces to

$$\phi_X(\omega) = \sum_i p_i e^{j\omega x_i} \tag{7.33}$$

Example 7.10

Show that the characteristic function of a Gaussian r.v. ***X*** *with mean value m and variance* σ^2*, is given by*

$$\phi_X(\omega) = e^{(jm\omega - 0.5\sigma^2\omega^2)}$$

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 r.v. $\boldsymbol{X}$ into another r.v. $\boldsymbol{Y}$ by putting

$$\boldsymbol{Y} = (\boldsymbol{X}\text{-}m)/\sigma$$

Then, $$E[\boldsymbol{Y}] = \frac{1}{\sigma}E[\boldsymbol{X}] - \frac{m}{\sigma} = 0. \quad \therefore \quad \text{Mean of } \boldsymbol{Y} \text{ is zero}$$

$$\text{Var}[\boldsymbol{Y}] = E[\boldsymbol{Y}^2] - \{E[\boldsymbol{Y}]\}^2 = E[\boldsymbol{Y}^2] = E\left[\frac{(\boldsymbol{X}-m)}{\sigma^2}\right] = 1$$

$\therefore$ $\boldsymbol{Y} = N(0, 1)$; and its density function is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}dy$$

$$\Phi_Y(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}.e^{sy}dy = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{(sy-y^2/2)}dy$$

But, $$sy - y^2/2 = \frac{s^2}{2} - \frac{1}{2}(y-s)^2.$$

$\therefore$ substituting this in the RHS of the above equation,

$$\Phi_Y(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\left[\frac{s^2}{2}-\frac{1}{2}(y-s)^2\right]}dy = e^{s^2/2}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(y-s)^2/2)}dy\right] = e^{s^2/2} \tag{7.34}$$

But $$\boldsymbol{Y} = \frac{\boldsymbol{X}-m}{\sigma} \text{ or } \boldsymbol{X} = \sigma\boldsymbol{Y} + m$$

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x)e^{j\omega(\sigma y+m)}dx = e^{j\omega m}\int_{-\infty}^{\infty} f_X(x)e^{j\omega\sigma y}dx = E\left[e^{j\omega\sigma\boldsymbol{Y}}\right]e^{j\omega m} \tag{7.35}$$

Since $\Phi_X(\omega) = E\left[e^{j\omega\boldsymbol{X}}\right]$, we may write: $E\left[e^{j\omega\sigma\boldsymbol{Y}}\right] = \Phi_{\sigma y}(\omega)$

But $\Phi_Y(\omega) = e^{-\omega^2/2}$ from Eq. (7.34)

$$\Phi_{\sigma Y}(\omega) = e^{-\sigma^2\omega^2/2}$$

$\therefore$ from Eq. (7.35), we have

$$\Phi_X(\omega) = e^{j\omega m}.e^{-\sigma^2\omega^2/2} = e^{(j\omega m - 0.5\sigma^2\omega^2)}$$

Example 7.11

Find the characteristic function of a Bernoulli random variable.

The Bernoulli r.v. takes the values 1 and 0 with probabilities p and $(1-p)$ respectively.

$\therefore$ $$\Phi_X(\omega) = p.e^{j\omega.1} + (1-p).e^{j\omega.0} = 1 + p\left[e^{j\omega} - 1\right]$$

7.7 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 r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ as

$$F(x,y) \;\Delta\; P[\boldsymbol{X} \le x, \boldsymbol{Y} \le y] \tag{7.36}$$

The *joint density function* may be defined as

$$f(x,y) \;\Delta\; \frac{\partial^2 F(x,y)}{\partial x \partial y} \tag{7.37}$$

Then, since $F(x,-\infty) = F(-\infty, y) = 0$ and $F(+\infty,+\infty) = 1$, we have

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha,\beta) d\alpha d\beta \tag{7.38}$$

The marginal distribution functions and density functions can be obtained from the joint distribution and density functions respectively as follows

$$F_x(x) = F(x) = F_{XY}(x,\infty) \quad \text{and} \quad F_y(y) = F(y) = F_{XY}(\infty, y) \tag{7.39}$$

and

$$f(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy; \quad f(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \tag{7.40}$$

Discrete Random Variables If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two discrete random variables defined on a certain probability space, and if they take values x_i and y_k with probabilities p_i and q_k respectively and given by

$$p_i = P[\boldsymbol{X} = x_i] \text{ and } q_k = P[\boldsymbol{Y} = y_k] \tag{7.41}$$

then their joint probability p_{ik} is given by

$$p_{ik} = P[\boldsymbol{X} = x_i \quad \text{and} \quad \boldsymbol{Y} = y_k] \tag{7.42}$$

Of course, just like the marginal probabilities, the joint probabilities also add up to a value 1,

i.e.,
$$\sum_i \sum_k p_{ik} = 1 \tag{7.43}$$

Also,

$$p_i = \sum_k p_{ik} \quad \text{and} \quad p_k = \sum_i p_{ik} \tag{7.44}$$

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 $F_Y(y \mid X \le x)$ and is defined by

$$F_Y(y \mid \boldsymbol{X} \le x) \Delta \frac{P\{\boldsymbol{X} \le x, \boldsymbol{Y} \le y\}}{\boldsymbol{P}\{\boldsymbol{X} \le x\}} = \frac{F(x,y)}{F_X(x)} \tag{7.45}$$

$\therefore$
$$f_Y(y \mid \boldsymbol{X} \le x) = \frac{\frac{\partial}{\partial y}[F(x,y)]}{F_X(x)} \tag{7.46}$$

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

$$f(y\,|\,x) = \frac{f(x,y)}{f(x)} \tag{7.47}$$

Similarly,

$$f(x\,|\,y) = \frac{f(x,y)}{f(y)} \tag{7.48}$$

If the r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent,

$$f(y\,|\,x) = \frac{f(x,y)}{f(x)} = f(y) \qquad \therefore \quad f(x,y) = f(x)\cdot f(y) \tag{7.49}$$

Discrete r.vs If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are discrete type of random variables with $P\left[\boldsymbol{X} = x_i\right] = p_i$ and $P\left[\boldsymbol{Y} = y_k\right] = q_k$,

and $P\left\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_k\right\} = p_{ik}$ with say $i = 1$ to M and $k = 1$ to N, then

$$P\left\{\boldsymbol{Y} = y_k\,|\,\boldsymbol{X} = x_i\right\} = \frac{P\left\{\boldsymbol{X} = x_i, \boldsymbol{Y} = y_k\right\}}{P\left\{\boldsymbol{X} = x_i\right\}} = \frac{p_{ik}}{p_i} \tag{7.50}$$

Conditional Mean and Variance The conditional mean of the r.v. $\boldsymbol{Y}$ given that $\boldsymbol{X} = x$, is represented by $\eta_{y|x}$ and is given by

$$\eta_{y|x} = E\left[\boldsymbol{Y}\,|\,x\right] = \int_{-\infty}^{\infty} y f(y\,|\,x)dy \tag{7.51}$$

The conditional variance is represented by $\sigma^2_{y|x}$ and is given by

$$\sigma^2_{y|x} = E\left[(\boldsymbol{Y} - \eta_{y|x})^2\,|\,x\right] = \int_{-\infty}^{\infty} (y - \eta_{y|x})^2 f_{y|x}(y\,|\,x)dy \tag{7.52}$$

Independence, Uncorrelatedness and Orthogonality

(i) If two r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent,

$$f(y\,|\,x) = f(y);\;\; f(x\,|\,y) = f(x) \text{ and } f(x,y) = f(x).f(y) \tag{7.53}$$

(ii) The co-variance of the two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as

$$\begin{aligned} C_{XY} &= E\left[(\boldsymbol{X} - \eta_X)(\boldsymbol{Y} - \eta_Y)\right] \\ &= E\left[\boldsymbol{XY}\right] - E\left[\boldsymbol{X}\right]E\left[\boldsymbol{Y}\right] \end{aligned} \tag{7.54}$$

The correlation coefficient ρ_{XY} of two r.v.s $\boldsymbol{X}$ and $\boldsymbol{Y}$ is

$$\rho_{XY} \,\Delta\, \frac{C_{XY}}{\sigma_X \sigma_Y} \tag{7.55}$$

The random variables x and y are said to be uncorrelated if their covarience is zero, i.e.,

$$C_{XY} = 0 \text{ i.e., when } \rho_{XY} = 0 \text{ or } E\left[\boldsymbol{XY}\right] = E\left[\boldsymbol{X}\right]E\left[\boldsymbol{Y}\right] \tag{7.56}$$

(iii) Random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be orthogonal, if

$$E\left[\boldsymbol{XY}\right] = 0$$

7.8 JOINTLY GAUSSIAN RANDOM VARIABLES

Definition Two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be jointly Gaussian if their joint density function is of the form

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2}+\frac{(y-m_2)^2}{\sigma_2^2}-\frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right\}\right] \quad (7.58)$$

Properties

1. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are jointly Gaussian then (a) they are individually Gaussian, and (b) the conditional densities $f(x \mid y)$ and $f(y \mid x)$ are also Gaussian.
2. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are individually Gaussian, they need not necessarily be jointly Gaussian.
3. Jointly Gaussian r.v.s are completely characterized by their mean vector and covariance matrix.
4. For jointly Gaussian r.v.s, uncorrelatedness implies statistical independence.

7.9 CENTRAL LIMIT THEOREM

The central limit theorem states that if $(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n)$ are independent random variables with means $(m_1, m_2, ..., m_n)$ and variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2, ..., \sigma_n^2)$, then the cumulative distribution function of the random variable

$$\sum_{i=1}^{n}\left(\frac{\boldsymbol{X}_i - m_i}{\sqrt{n}\ \sigma_i}\right)$$

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 $= \sigma^2$ 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 a central limit theorem that the sum of the noises produced by a very large number of independent sources tends to have Gaussian distribution.

Example 7.12

*Random variable **Y** = sin **X**, where **X** is uniformly distributed between -π/2 to +π/2. Find the density function of **Y**.*

$$\Phi_Y(\omega)=E\left[e^{j\omega y}\right]=E\left[e^{j\omega \sin x}\right]=\int_{-\infty}^{\infty} e^{j\omega \sin x} f_X(x)dx$$

Since $\boldsymbol{X}$ is uniformly distributed over -π/2 to + π/, we have,

$$f_X(x)=\begin{cases}\dfrac{1}{\pi} & ; \quad -\dfrac{\pi}{2}\le x \le \dfrac{\pi}{2} \\ 0 & ; \text{ otherwise}\end{cases}$$

$$\therefore \qquad \Phi_Y(\omega)=\frac{1}{\pi}\int_{-\pi/2}^{\pi/2} e^{j\omega \sin x}dx$$

Since $y = \sin x$, $dy = \cos x dx$ and when $x = -\pi/2$, $y = -1$ and when $x = \pi/2$, $y = 1$

$$\therefore \qquad \Phi_Y(\omega) = \frac{1}{\pi}\int_{-1}^{1} e^{j\omega y}\frac{1}{\sqrt{1-y^2}}.dy \quad \text{But} \quad \Phi_Y(\omega) = E\left[e^{j\omega y}\right] = \int_{-\infty}^{\infty} e^{j\omega y} f_Y(y)dy$$

$\therefore$ we find that $f_Y(y) = \dfrac{1}{\pi\sqrt{1-y^2}}$ for $|y| \le 1$ and zero otherwise.

Example 7.13

***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*

$$\mathbf{r} = \sqrt{X^2 + Y^2}\ ;\ \theta = \tan^{-1}(\mathbf{Y}/\mathbf{X}) \text{ where, } |\theta| < \pi$$

*Obtain the joint density function of **r** and θ. Also, obtain their marginal densities.*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \text{ and } f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2}$$

Since ***X*** and ***Y*** are given to be independent r.v.s, 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
$$J(r,\theta) = \begin{vmatrix} \dfrac{\partial x_1}{\partial r} & \dfrac{\partial x_1}{\partial \theta} \\ \dfrac{\partial y_1}{\partial r} & \dfrac{\partial y_1}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

$$\therefore \qquad f_{r,\theta}(r,\theta) = r f_{xy}(x_1, y_1) = \frac{r}{2\pi\sigma^2} e^{-r^2/2\sigma^2};0 < r < \infty \text{ and } |\theta| < \pi$$

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.to r from $r = 0$ to $r = \infty$.

$$\therefore \qquad f_r(r) = \int_{-\pi}^{\pi} f_{r,\theta}(r,\theta)d\theta = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2};0 < r < \infty$$

$$f_\theta(\theta) = \int_0^\infty f_{r,\theta}(r,\theta)dr = \frac{1}{2\pi};|\theta| < \pi$$

$\therefore$ ***r*** has Rayleigh density function while θ is uniformly distributed between $-\pi$ and $+\pi$. Also, since $f_{r,\theta}(r,\theta) = f_r(r).f_\theta(\theta)$, we find that ***r*** and θ are statistically independent.

RANDOM PROCESSES 7.10

Earlier, in Section 7.4, we had defined a random variable $\boldsymbol{X}$ 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 $\boldsymbol{X}(\xi)$, or simply, by $\boldsymbol{X}$. 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. 7.9.

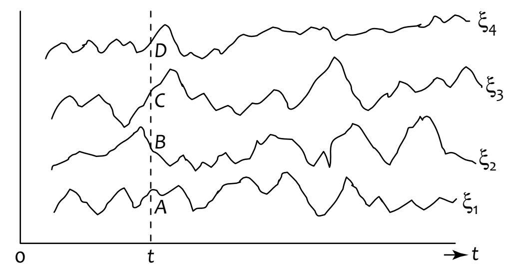


Fig. 7.9 *Random processes as family of time signals, one for each ξ*

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, the following is clear:

(a) 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.
(b) When t is fixed and ξ is a variable, $\boldsymbol{X}(t, \xi)$ represents a set of real numbers (as shown in Fig. 7.9), one for each ξ and hence $\boldsymbol{X}(t, \xi)$ in this case, is just a random variable.
(c) When both t and ξ are fixed, $\boldsymbol{X}(t, \xi)$ represents a mere number.
(d) When both t and ξ are variables, $X(t, \xi)$ 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.

7.10.1 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 r.v. and its distribution is

$$F\,(x, t) = P[\boldsymbol{X}(t) \le x] \tag{7.59}$$

The derivative with respect to x of this first-order distribution function, $F(x, t)$ of the process $\boldsymbol{X}(t)$,

$$f(x,t) = \frac{\partial}{\partial x}\left[F(x,t)\right] \tag{7.60}$$

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

$$F\,(x_1,\, x_2;\, t_1,\, t_2) = P[\boldsymbol{X}(t_1) \le x_1,\, \boldsymbol{X}(t_2) \le x_2] \tag{7.61}$$

The second-order density function is

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2}{\partial x_1 \partial x_2}\left[F(x_1, x_2; t_1, t_2)\right] \tag{7.62}$$

Of course, as usual, we must have the first-order statistics from the second-order statistics; i.e.,

$$f(x_1,t_1)=[F(x_1,\infty;t_1,t_2)] \quad \text{and} \quad f(x_1,t_1)=\int_{-\infty}^{\infty} f(x_1,x_2;t_1,t_2)dx_2 \tag{7.63}$$

The Mean Proceeding on the same lines, we define the mean of the random process $\boldsymbol{X}(t)$ as the mean of the r.v. $\boldsymbol{X}(t)$

$$\therefore \qquad \eta_X(t)=E[\boldsymbol{X}(t)]=\int_{-\infty}^{\infty} x.f(x,t)dx \tag{7.64}$$

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 $\boldsymbol{X}(t_0)$.

The 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 $\boldsymbol{X}(t_2)$.

$$R_X(t_1,t_2)=E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)] \text{ if } \boldsymbol{X}(t) \text{ is a real process}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 f(x_1,x_2;t_1,t_2)dx_1dx_2 \tag{7.65}$$

And, if $\boldsymbol{X}(t)$ is a complex valued process,

$$R_X(t_1,t_2)=E[\boldsymbol{X}(t_1)\boldsymbol{X}^*(t_2)] \tag{7.66}$$

where, the * indicates complex-conjugation.

The 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 $C_X(t_1,t_2)$

$$\therefore \qquad C_X(t_1,t_2)=E\{[\boldsymbol{X}(t_1)-\eta_X(t_1)][\boldsymbol{X}(t_2)-\eta_X(t_2)]\} \text{ for a real process,}$$

$$=R_X(t_1,t_2)-\eta_X\eta_Y \text{ for a real process}$$

and

$$C_X(t_1,t_2)=E\{[\boldsymbol{X}(t_1)-\eta_X(t_1)][\boldsymbol{X}(t_2)-\eta_X(t_2)]^*\} \text{ for a complex process,}$$

$$\therefore \qquad C_X(t_1,t_2)=R_X(t_1,t_2)-\eta_X(t_1)\eta_X^*(t_2) \quad \text{for a complex process} \tag{7.67}$$

Note:

1. $R_X(t,t)=E\left\lfloor |\boldsymbol{X}(t)|^2 \right\rfloor$= Average power in $\boldsymbol{X}(t)$
2. $C_X(t,t)=E[\boldsymbol{X}^2(t)]-\{E[\boldsymbol{X}(t)]\}^2$= 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

$$R_{XY}(t_1,t_2)=E\left[\boldsymbol{X}(t_1)\boldsymbol{Y}^*(t_2)\right]=R_{YX}^*(t_2,t_1) \tag{7.68}$$

The cross-covariance of the two processes is defined as

$$C_{XY}(t_1,t_2)=E\{[\boldsymbol{X}(t_1)-\eta_X(t_1)][\boldsymbol{Y}(t_2)-\eta_Y(t_2)]^*\}$$

$$\therefore \qquad =R_{XY}(t_1,t_2)-\eta_X(t_1)\eta_Y^*(t_2) \tag{7.69}$$

(In the above two equations, complex conjugation can be ignored if the processes are real valued).

7.10.2 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 $\{\mathbf{Y}(t_1'), \mathbf{Y}(t_2'), \ldots \mathbf{Y}(t_n')\}$ 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

$$C_{XY}(t_1, t_2) = 0 \text{ for all values of } t_1 \text{ and } t_2 \tag{7.70}$$

Orthogonal Processes Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be orthogonal processes if

$$R_X(t_1, t_2) = 0 \text{ for all values of } t_1 \text{ and } t_2 \tag{7.71}$$

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 7.14

$\mathbf{X}(t) = A\cos(\omega t + \phi)$ where ϕ is a r.v. uniformly distributed between $-\pi$ and $+\pi$. Determine the mean and auto-correlation of $\mathbf{X}(t)$.

(i) $\eta_X(t) = E[\mathbf{X}(t)] = \int_{-\infty}^{\infty} \mathbf{X}(t).f_\phi(\phi)d\phi$

$$= \frac{A}{2\pi}\int_{-\infty}^{\infty}\cos(\omega t + \phi)d\phi = 0$$

(ii) $R_X(t_1, t_2) = E\left[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)\right] = E\left[A^2\cos(\omega t_1 + \phi) + \cos(\omega t_2 + \phi)\right]$

$$= A^2 E\left[\cos(\omega t_1 + \phi) + \cos(\omega t_2 + \phi)\right]$$

$$= \frac{1}{2}A^2 E\left[\cos\omega(t_1 - t_2) + \cos(\omega t_1 + \omega t_2 + 2\phi)\right]$$

$$= \frac{1}{2}A^2\cos\omega(t_1 - t_2)$$

Example 7.15

If $\mathbf{X}(t) = ae^{j\omega t}$, determine its auto-correlation.

$$R_X(t_1, t_2) = E\left[\boldsymbol{X}(t_1)\boldsymbol{X}^*(t_2)\right] = E\left[\boldsymbol{a}e^{j\omega t_1}.\boldsymbol{a}^*\ e^{-j\omega t_2}\right]$$

$$= E\left\lfloor |\boldsymbol{a}|^2 \right\rfloor e^{j\omega(t_1 - t_2)}$$

7.11 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.,

$$f_X(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = f_X(x_1, x_2, ..., x_n; t_1 + \in, t_2 + \in, ..., t_n + \in) \tag{7.72}$$

For any integer n and any real number $\in$.

If Eq. (7.72) is true only up to $n \leq K$ then the process $\boldsymbol{X}(t)$ is said to be Kth 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:

(i) Its mean, $\eta_X(t) = E[\boldsymbol{X}(t)]$ is independent of time.
(ii) 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.

$$\therefore \qquad R_X(t_1, t_2) = R_X(\tau) = E[X(t+\tau)X^*(t)]$$

When $\tau = 0$, $R_X(0) = E\left[|\boldsymbol{X}(t)|^2\right]$ = 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

(i) It is deterministic.
(ii) It takes maximum value when $t = 0$
(iii) $R_X(0)$ = average power of the process.
(iv) For real, process $\boldsymbol{X}(t)$, $R_X(-\tau) = R_X(+\tau)$, i.e., $R_X(\tau)$ has even symmetry
(v) For a complex process $R_X(-\tau) = R_X^*(\tau)$ where * denotes complex conjugation.

Example 7.16

Show that the random process $\boldsymbol{X}(t) = A\cos(\omega t + \phi)$, where ϕ is a r.v. uniformly distributed over $-\pi$ to $+\pi$, is WSS.

It has already been shown in Example 7.14 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.

7.11.1 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. 7.8, 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 *atleast 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 7.17

Show that the process $\mathbf{X}(t) = A_c\cos(\omega_0 t + \theta)$ where θ is uniformly distributed over $-\pi$ to $+\pi$, is ergodic in mean and auto-correlation.

We have already seen in Example 7.16 that it is WSS and that

$$\eta_X(t) = E[\mathbf{X}(t)]$$

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 $[\mathbf{X}(t+\tau)\mathbf{X}(t)]$ are same.

$\therefore$ the given $\mathbf{X}(t)$ is ergodic in auto-correlation

7.11.2 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.)

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau \tag{7.73}$$

$\therefore$
$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df \tag{7.74}$$

Since in general, $R_X(-\tau) = R_X^*(\tau)$, $S_X(f)$, the PSD is always a real valued function of frequency. Further, if $\boldsymbol{X}(t)$ 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:

$$\text{Cross-correlation} = R_{XY}(\tau) \ \underline{\Delta} \ E\lfloor \boldsymbol{X}(t+\tau)\boldsymbol{Y}^*(t) \rfloor \tag{7.75}$$

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,
$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f\tau} d\tau \tag{7.76}$$

and
$$R_{XY}(\tau) = \int_{-\infty}^{\infty} S_{XY}(f) e^{j2\pi f\tau} df \tag{7.77}$$

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 7.18

Determine the power spectrum of the processes $\boldsymbol{X}(t) = A_c\cos(\omega_0 t + \theta)$; *$\theta$ is uniformly distributed over $(-\pi, +\pi)$.*

In Example 7.17 we obtained the ACF of this $\boldsymbol{X}(t)$ as

$$R_X(\tau) = \frac{A_c^2}{2}\cos\omega\tau$$

$\therefore$ the power spectrum, which is the Fourier transform of $R_X(\tau)$ is

$$S_X(f) = \frac{A_c^2}{2}\int_{-\infty}^{\infty} \cos\omega_0 t e^{-j2\pi f\tau} d\tau$$

But we know that $\mathcal{F}[\cos\omega_0\tau] = \frac{1}{2}[\delta(f+f_0) + \delta(f-f_0)]$

$\therefore$
$$S_X(f) = \frac{A_c^2}{4}[\delta(f+f_0) + \delta(f-f_0)]$$

7.11.3 Gaussian Processes

A random process $X(t)$ is said to be a Gaussian random process, if the random variables $X(t_1)$, $X(t_2)$,..., $X(t_n)$ are jointly Gaussian for all n and all t_1, t_2,..., t_n.

Properties of Gaussian Processes

(i) A Gaussian process is completely described by its mean and auto-correlation.
(ii) If a Gaussian process is wide-sense stationary then it is stationary in the strict sense too.
(iii) If a Gaussian process is given as input to a LTI system, the output process also is Gaussian.
(iv) 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 $X(t)$ 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. 7.10. As shown in the figure, it has a constant value $N_0/2$ for all frequencies.

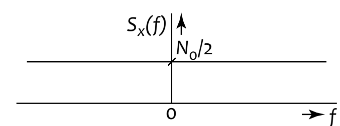


Fig. 7.10 *PSD of a white process*

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 very high frequency which is far beyond the frequencies at which any practical communication system operates, then in so far as our communication systems are concerned, we can safely assume that the PSD of the process is absolutely constant, i.e., that 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

$$R_n(\tau) = \frac{N_0}{2}\delta(\tau) \tag{7.78}$$

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 time) the two samples may be. That is why we call the white process as '*white noise*'.

7.12 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.

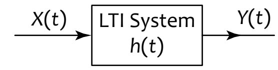


Fig. 7.11 *An LTI system with a random process as input*

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

(i) Mean of the output $= \eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)dt$ a constant independent of time. (7.79)

(ii) Cross-correlation of input and output processes $= R_{XY}(\tau) = R_X(\tau) * h(-\tau)$ (7.80)

(iii) Correlation of output process $= R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$ (7.81)

(i) 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,

$$= E[\boldsymbol{X}(t-u)] = E[\boldsymbol{X}(t)] = \eta_X = \text{ a constant}$$

$$\boxed{\therefore \quad E[\boldsymbol{Y}(t)] = \eta_X \int_{-\infty}^{\infty} h(t)dt = \text{a constant independent of } t.}$$

(ii) $R_{XY}(t_1,t_2) \;\triangleq\; E[\boldsymbol{X}(t_1)\boldsymbol{Y}(t_2)]$

$$R_{XY}(t_1,t_2) = E\left[\boldsymbol{X}(t_1)\int_{-\infty}^{\infty} \boldsymbol{X}(t_2-u)h(u)\,du\right]$$

$$= \int_{-\infty}^{\infty} E[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2-u)]h(u)du$$

If we now put $u = -\lambda$,

$$R_{XY}(t_1,t_2) = \int_{-\infty}^{\infty} R_X(t_2 - t_2 + u)h(u)du = \int_{-\infty}^{\infty} R_X(\tau+u)h(u)du$$

$$= \int_{-\infty}^{\infty} R_X(\tau-\lambda)h(\lambda)\,d\lambda = R_X(\tau) * h(-\tau)$$

$$\boxed{\therefore \quad R_{XY}(t_1, t_2) = R_X(\tau) * h(-\tau)} \tag{7.82}$$

$\therefore$ the cross-correlation is a function only of $\tau = (t_1 - t_2)$

(iii) 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 $$\boldsymbol{Y}(t_1) = \int_{-\infty}^{\infty} \boldsymbol{X}(u)h(t_1 - u)du$$

$\therefore$ $$R_Y(t_1, t_2) = E\left[\left\{\int_{-\infty}^{\infty} \boldsymbol{X}(u)h(t_1 - u)du\right\}\boldsymbol{Y}(t_2)\right]$$

$$= \int_{-\infty}^{\infty} E\{\boldsymbol{X}(u)\boldsymbol{Y}(t_2)\}h(t_1 - u)du$$

$$= \int_{-\infty}^{\infty} R_{XY}(u - t_2)h(t_1 - u)du$$

If we put $(t_1 - u) = \lambda$, we get $u = t_1 - \lambda; du = -d\lambda$

$\therefore$ $$R_Y(t_1, t_2) = \int_{-\infty}^{\infty} R_{XY}(t_1 - t_2 - \lambda)h(\lambda)d\lambda$$

$\therefore$ $$R_Y(\tau) = \int_{-\infty}^{\infty} R_{XY}(\tau - \lambda)h(\lambda)d\lambda = R_{XY}(\tau) * h(\tau) \tag{7.83}$$

$\therefore R_Y(t_1, t_2)$, 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 already 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. (7.83) using Eq. (7.82), we get

$$\boxed{R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)}$$

7.12.1 Input and Output Spectra and Cross-power Spectrum

Equations (7.79), (7.81) and (7.80) 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.

(i) From Eq. (7.79), we have

$$\eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)dt = \eta_X \text{. area under the impulse response.}$$

But we know that

$$\int_{-\infty}^{\infty} h(t)dt = \left[\int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt\right]\Bigg|_{f=0} = H(f)\Big|_{f=0} = H(0)$$

$$\boxed{\therefore \quad \eta_Y = \eta_X H(0)} \tag{7.84}$$

(ii) From Eq. (7.80), we have

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

Taking Fourier transform on both sides and noting that

$$\mathcal{F}[h(-\tau)] = H^*(f) \tag{7.85}$$

$$\boxed{S_{XY}(f) = S_X(f).H^*(f)} \tag{7.86}$$

(iii) From Eq. (7.81), we have

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$$

$\therefore$ taking Fourier transform of this on both sides,

$$S_Y(f) = S_X(f).H(f).H^*(f)$$

$$\boxed{\therefore \quad S_Y(f) = S_X(f).|H(f)|^2} \tag{7.87}$$

This is a very important result and is used quite frequently in communication engineering.

Example 7.19

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.*

From Eq. (7.84), 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. (7.87), we have

$$S_Y(f) = S_X(f).|H(f)|^2$$

$$|H(f)|^2 = H(f)H^*(f) = j2\pi f.(-j2\pi f) = 4\pi^2 f^2$$

$$S_Y(f) = 4\pi^2 f^2 . S_X(f)$$

Here, $$S_X(f) = \mathcal{F}[R_X(\tau)]$$

$\therefore$ $$S_y(f) = 4\pi^2 f^2 \left[\mathcal{F}\{R_X(\tau)\}\right] = 4\pi^2 f^2 S_X(f)$$

7.13 REPRESENTATION OF BAND-LIMITED AND BANDPASS PROCESSES

7.13.1 Band-limited Processes

In the case of a deterministic signal $x(t)$ which is lowpass and band-limited to say W Hz, i.e., $X(f) = 0$ for $|f| \geq W$, we know from the lowpass sampling theorem for deterministic signals, that if $x(t)$ is sampled at regular intervals of T_s where $T_s \leq \dfrac{1}{2W}$, 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.

$$x(t) = \frac{2W}{f_s} \sum_{k=-\infty}^{+\infty} x(kT_s) \text{ sinc } 2W(t - kT_s); \quad f_s = 1/T_s, \ -\infty < t < \infty \tag{7.88}$$

In the particular case when $T_s = \dfrac{1}{2W}$, this equation reduces to

$$x(t) = \sum_{k=-\infty}^{+\infty} x(k/2W) \text{ sinc } (2Wt - k); \quad -\infty < t < \infty \tag{7.89}$$

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 lowpass process also could be represented by its samples, or, in short, whether a similar lowpass sampling theorem exists in the case of random processes too. Fortunately, there is a *similar* theorem applicable to stationary lowpass band-limited processes, and it states as follows.

Theorem If $\boldsymbol{X}(t)$ is a stationary lowpass 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 $T_s = \dfrac{1}{2W}$, then

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \text{ sinc } 2W(t - kT_s)\right|^2\right] = 0 \tag{7.90}$$

What Eq. (7.90) says is that under the conditions stated in the theorem, $\boldsymbol{X}(t)$ is equal, *in the mean-square sense*, to

$$\sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \text{ sinc } 2W(t - kT_s)$$

Proof To prove Eq. (7.90), 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

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \text{ sinc } 2W(t - kT_s)\right|^2\right] = E\left[\boldsymbol{X}^2(t)\right] - 2\sum_{k=-\infty}^{+\infty} E\left[\boldsymbol{X}(T)\boldsymbol{X}(kT_s)\right] \text{ sinc } 2W(t - kT_s)$$

$$+\sum_k \sum_l E\left[\boldsymbol{X}(kT_s)\boldsymbol{X}(lT_s)\right] \text{ sinc } 2W(t - kT_s) \text{sinc } 2W(t - lT_s)$$

But $E\left[\boldsymbol{X}^2(t)\right]=R_X(0)$ and $E\left[\boldsymbol{X}(kT_s)\boldsymbol{X}(lT_s)\right]=R_X\left(kT_s-lT_s\right)$.

Now, if we put $m=l-k$, $l=m+k$ and LHS of Eq. (7.90) equal,

$$R_X(0)-2\sum_{k=-\infty}^{+\infty} R_X(t-kT_s)\text{ sinc }2W(t-kT_s)$$

$$+\sum_k\sum_m R_X(-mT_s)\text{ sinc }2W(t-kT_s)\text{ sinc }2W(t-kT_s-mT_s) \tag{7.91}$$

But
$$\sum_k\sum_m R_X(-mT_s)\text{ sinc }2W(t-kT_s)\text{ sinc }2W(t-kT_s-mT_s)$$

$$=\sum_k \text{sinc }2W(t-kT_s)\sum_m R_X(-mT_s)\text{ sinc }2W(t-kT_s-mT_s) \tag{7.92}$$

However, since $\boldsymbol{X}(t)$ is a real process, its ACF has even symmetry

$$\therefore \qquad R_X(-mT_s)=R_X(mT_s) \tag{7.93}$$

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 lowpass sampling theorem for bandlimited deterministic signals, using Eq. (7.89), we may expand $R_X(t)$ in terms of its samples, as follows:

$$R_X(t)=\sum_{m=-\infty}^{+\infty} R_X(mT_s)\text{ sinc }2W(t-mT_s) \tag{7.94}$$

$$\therefore \qquad R_X(t-kT_s)=\sum_{m=-\infty}^{+\infty} R_X(mT_s)\text{ sinc }2W(t-kT_s-mT_s) \tag{7.95}$$

$\therefore$ using Eqs (7.93) and (7.95), the RHS of Eq. (7.92) may be written as

$$=\sum_{k=-\infty}^{\infty} R_X(t-kT_s)\text{ sinc }2W(t-kT_s)$$

Hence, Eq. (7.91) may be modified as

$$R_X(0)-2\sum_{k=-\infty}^{+\infty} R_X(t-kT_s)\text{ sinc }2W(t-kT_s)+\sum_{k=-\infty}^{\infty} R_X(t-kT_s)\text{ sinc }2W(t-kT_s)$$

$$\therefore \text{ LHS of Eq. (7.90)}=R_X(0)-\sum_{k=-\infty}^{+\infty} R_X(t-kT_s)\text{ sinc }2W(t-kT_s) \tag{7.96}$$

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 lowpass sampling theorem for deterministic signals that

$$R_X(t)=\sum_{k=-\infty}^{+\infty} R_X(kT_s)\text{ sinc }2W(t-kT_s) \tag{7.97}$$

In Eq. (7.97), 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 $t = 0$ second. Instead, if we have a sample $t = t_0$ and every T_s sec on either side of it, Eq. (7.97) gets modified and from that modified equation we may write (by putting $t = 0$) as

$$R_X(0) = \sum_{k=-\infty}^{+\infty} R_X(t_0 - kT_s) \text{ sinc } \left[2W(t_0 - kT_s)\right]$$

Since t_0 can take any value, we may write

$$R_X(0) = \sum_{k=-\infty}^{+\infty} R_X(t - kT_s) \text{ sinc } \left[2W(t - kT_s)\right]$$

Substituting this on the RHS of Eq. (7.91), we get

$$E\left|x(t) - \sum_{k=-\infty}^{+\infty} X(t - kT_s) \text{ sinc } \left[2W(t - kT_s)\right]\right|^2 = 0$$

Thus, the theorem is proved.

7.13.2 Bandpass Processes

In Section 2.8, we had discussed in detail, the in-phase and quadrature component representation of a deterministic bandpass signal $x(t)$

i.e., $$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \tag{7.98}$$

where, $x_I(t)$ and $x_Q(t)$ are lowpass signals

Let $X(t)$ be a stationary, zero-mean bandpass process, whose power spectral density is of the form shown in Fig. 7.12.

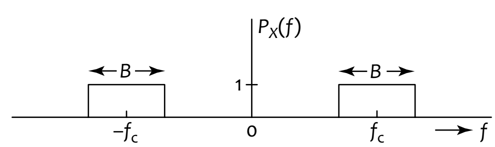


Fig. 7.12 *PSD of a bandpass process X(t)*

We encounter such a bandpass noise process when we say white noise is filtered by a bandpass filter. Similar to the bandpass signal case, we shall now define two lowpass processes $X_I(t)$ and $X_Q(t)$, where

$$X_I(t) = X(t)\cos 2\pi f_c t + \hat{X}(t)\sin 2\pi f_c t \tag{7.99}$$

and $$X_Q(t) = \hat{X}(t)\cos 2\pi f_c t - X(t)\sin 2\pi f_c t \tag{7.100}$$

Here $\hat{X}(t)$ is the Hilbert transform of $X(t)$.

From Eq. (7.99) and (7.100), it is easy to verify that

$$X_I(t)\cos 2\pi f_c t - X_Q(t)\sin 2\pi f_c t = X(t) \tag{7.101}$$

When $X(t)$ is a stationary, zero-mean bandpass 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 had 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, lowpass, jointly stationary processes. If in addition $X(t)$ is Gaussian, then $X_I(t)$ and $X_Q(t)$ will be jointly Gaussian.
2. The in-phase and quadrature components $X_I(t)$ and $X_Q(t)$ have the same average power as the process $X(t)$ itself,

   i.e., $$P_{X_I} = P_{X_Q} = P_X = \int_{-\infty}^{\infty} P_X(f)df \quad (7.102)$$

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.

SUMMARY

1. Modern probability theory is based on the following axioms:
   (i) If A is an event, $P(A) \geq 0$.
   (ii) If S is the certain event, $P(S) = 1$.
   (iii) If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.
2. Sample space is the set of all possible outcomes of a random experiment.
3. Events are defined in terms of subsets of the sample space forming a Borel field σ.
4. 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.
5. Conditional probability of A given B is $P(A \mid B) = \dfrac{P(AB)}{P(B)}$; where $P(AB)$ = Probability of joint occurrence of A and B and $P(B) \neq 0$.
6. Bayes' theorem: $P(A_i \mid B) = \dfrac{P(B / A_i)P(A_i)}{\sum_{i=1}^{n} P(B / A_i)P(A_i)}$
7. (i) Events A and B are said to be independent events if $P(AB) = P(A) \cdot P(B)$.
   (ii) In general, n events, $A_1, A_2, ..., A_n$ are said to be independent, if for every $k < n$, the events $A_1, A_2, ..., A_k$ are independent and further, if
   $$P(A_1, A_2, A_3, ..., A_n) = P(A_1)P(A_2).....P(A_n)$$
8. A real random variable is a mapping of the outcomes of a random experiment to the real line and satisfying the following two conditions:
   (i) $\{\boldsymbol{X} \leq x\}$, i.e., $\{\boldsymbol{X}(\xi) \leq x\}$ is an event for $\forall$ real number x.
   (ii) $P\{\boldsymbol{X}(\xi) = +\infty\} = P\{\boldsymbol{X}(\xi) = -\infty\} = 0$
9. The cumulative distribution function, CDF of a r.v. $\boldsymbol{X}$ is denoted by $F_X(x)$ and is defined as: $F_X(x) \triangleq P\{\boldsymbol{X} \leq x\}$
10. **Properties of CDF**
    (i) $F_X(x)$ lies between 0 and 1.
    (ii) $F_X(\infty) = 1$ and $F_X(-\infty) = 0$
    (iii) $F_X(x)$ is a non-decreasing function of x.

(iv) $F_X(x)$ is continuous from the right

(v) $F_X(b) - F_X(a) = P[a < \boldsymbol{X} \le b]$

11. Random variables are of three types—continuous, discrete and mixed types. R.Vs whose CDF is a continuous function is called a continuous r.v.s. A r.v. whose CDF has a staircase shape is called a discrete r.v. A r.v. which is neither discrete, nor continuous, is called a mixed r.v.
12. The probability density function PDF is defined as $f_X(x) = \frac{d}{dx}\left[F_X(x)\right]$
13. **Properties of PDF**

(i) $f_X(x) \ge 0$

(ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

(iii) $\int_{x_1}^{x_2} f_X(x)dx = P\left[x_1 < X \le x_2\right]$

(iv) $f_X(x) = \int_{-\infty}^{x} f(\alpha)d\alpha$

14. (i) Uniform random variable is one whose PDF is constant over a certain interval or range of x.

$$\therefore \quad f_X(x) = \begin{cases} \dfrac{1}{(x_2 - x_1)} & ; \; x_1 \le x \le x_2 \\ 0 & ; \; \text{elsewhere} \end{cases}$$

(ii) 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 r.v. $\boldsymbol{X}$.

(iv) A Rayleigh r.v is one which has PDF $f_X(x)$ given by

$$f_X(x) = \begin{cases} \left(\dfrac{x}{\sigma^2}\right) e^{-x^2/2\sigma^2}; & x \ge 0 \\ 0 \;\; ; x < 0 \end{cases}$$

(v) A Rician r.v. 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.

15. A Bernoulli random variable is a discrete r.v. which takes the values 1 and 0 with probabilities of P and $(1 - P)$.
16. A discrete r.v. $\boldsymbol{X}$ is said to be a binomial random variable with parameters n and p if

$$P\left[\boldsymbol{X} = k\right] = \binom{n}{k} p^k q^{n-k} \quad ; \quad 0 \le k \le n$$

17. If $\boldsymbol{X}$ is a r.v and if $\boldsymbol{Y} = g(\boldsymbol{X})$, then $f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|}$

18. The 'mean' or 'expected' value of a r.v. $\boldsymbol{X}$ is $E\{\boldsymbol{X}\} = \int_{-\infty}^{\infty} x f_X(x) dx$, if $\boldsymbol{X}$ is continuous.

19. If $\boldsymbol{X}$ is discrete r.v., $E\{\boldsymbol{X}\} = \sum_i p_i x_i$, where $p_i = P[\boldsymbol{X} = x_i]$

20. The variance of a continuous r.v. $\boldsymbol{X}$ with expected value η_x is define as:
Variance of $\boldsymbol{X} = \mathrm{Var}[\boldsymbol{X}] = \sigma_X^2 = E\left[(\boldsymbol{X} - \eta_X)^2\right] = E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2$

21. For a discrete r.v $\boldsymbol{X}$, $\sigma_X^2 = \sum_i p_i (x_i - \eta_X)^2$ where, $p_i = P[\boldsymbol{X} = x_i]$

22. The positive square root of variance is called 'standard deviation'.
The characteristic function of a continuous r.v. $\boldsymbol{X}$ is defined as

$$\phi_X(\omega) \triangleq \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx$$

23. The characteristic function of a discrete r.v. $\boldsymbol{X}$ which takes values x_i, $i = 1, 2, \ldots.$ with probabilities p_i, is given by

$$\phi_X(\omega) = \sum_i p_i e^{j\omega x_i}$$

24. The joint, or bi-variate distribution function $F_{X,Y}(x, y)$ is

$$F_{X,Y}(x, y) \;\triangleq\; P[\boldsymbol{X} \le x, \boldsymbol{Y} \le y]$$

25. The joint density function of twp r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ is

$$f(x, y) \;\triangleq\; \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

26. $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\alpha, \beta) d\alpha d\beta$; $F_X(x) = F_{X,Y}(x, \infty)$

$F_Y(y) = F_{X,Y}(\infty, y)$; $f(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$ and $f(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$

27. (i) if two r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are statistically independent,

$$f(y \mid x) = f(y);\; f(x \mid y) = f(x) \text{ and } f(x, y) = f(x) \cdot f(y)$$

(ii) r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be uncorrelated if their covariance is zero; i.e., if

$$C_{XY} \triangleq E[(\boldsymbol{X} - \eta_X)(\boldsymbol{Y} - \eta_Y)] = 0. \text{ Then}$$

$$E[\boldsymbol{XY}] = E[\boldsymbol{X}]E[\boldsymbol{Y}]$$

(iii) Two r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be orthogonal if $E[\boldsymbol{XY}] = 0$

28. Two r.vs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be jointly Gaussian if

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right\}\right]$$

29. 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 r.v.
30. A random process is a function of two variables—time t and outcome ξ and is denoted by $\boldsymbol{X}(t, \xi)$.
31. **The mean** of a random process at the instant $t = t_1$ is defined as the expected value (or mean) of the random variable $\boldsymbol{X}(t_1)$.
32. **The ACF** of a random process is defined as the expected value of the product of $\boldsymbol{X}(t_1)$ and $\boldsymbol{X}(t_2)$

$$R_X(t_1,t_2) \triangleq E\left[\boldsymbol{X}(t_1)\cdot \boldsymbol{X}(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1,x_2;t_1,t_2)dx_1 dx_2$$

33. **The auto co-variance** of a random process $\boldsymbol{X}(t)$ is

$$C_X(t_1,t_2) = E\left[\left\{\boldsymbol{X}(t_1) - \eta_X(t_1)\right\}\left\{\boldsymbol{X}(t_2) - \eta_X(t_2)\right\}\right]$$

34. **The auto-correlation** of a r.p. $\boldsymbol{X}(t)$ is $R_X(t_1,t_2) = E\left[\boldsymbol{X}(t_1)\boldsymbol{X}(t_2)\right]$

    The average power in $\boldsymbol{X}(\text{t}) = R_X(t,t) = E\left[|\boldsymbol{X}(t)|^2\right]$

    Variance of $\boldsymbol{X}(\text{t}) = C_X(t,t) = E\left[\boldsymbol{X}^2(t)\right] - \left\{E\left[\boldsymbol{X}(t)\right]\right\}^2$

35. **Cross covariance** $C_{XY}(t_1,t_2) = R_{XY}(t_1,t_2) - \eta_X(t_1)\eta_Y(t_2)$

    Cross correlation $R_{XY}(t_1,t_2) = E\left[\boldsymbol{X}(t_1)\boldsymbol{Y}(t_2)\right]$
36. **Independent processes** Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be statistically independent, if the set of r.vs $\{\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), .., \boldsymbol{X}(t_n)\}$ and $\{\boldsymbol{Y}(t_1'), \boldsymbol{Y}(t_2'), ... \boldsymbol{Y}(t_n')\}$ are mutually independent for all values of $t_1, t_2, ..., t_n$ and $t_1', t_2', ... t_n'$ and all integer values of n.
37. **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.
38. **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.
39. 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.
40. **Stationarity** Random processes, whose statistical properties like mean, ACF, etc., are independent of time, are called stationary processes.
41. **WSS** A process $\boldsymbol{X}(t)$ is said to be stationary in the side 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.
42. **Ergodocity** Random processes for which the time averages equal the ensemble averages, are known as ergodic processes.
43. **Wiener–Khinchine theorem** The PSD of a random process is the Fourier transform of its auto-correlation.
44. **Gaussian random process** $\boldsymbol{X}(t)$ is a Gaussian random process if the r.vs $\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), ..., \boldsymbol{X}(t_n)$ are jointly Gaussian for all values of $t_1, t_2, ..., t_n$ and all integer values of n.
45. **White noise process** A process $\boldsymbol{X}(t)$ whose PSD is a constant for all frequencies, is called a white noise process.
46. **ACF of a white noise process** For a white noise process with a PSD of $N_0/2$, the ACF is

$$R_n(\tau) = \frac{N_0}{2}\delta(t)$$

47. **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

(i) Mean of the output $= \eta_Y = \eta_X \int_{-\infty}^{\infty} h(t)dt$

(ii) Input and output cross-correlation $= R_{XY}(\tau) = R_X(\tau) * h(-\tau)$

(iii) Correlation of output process $= R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau)$

(iv) From (iii), it follows that $S_Y(f) = S_X(f).|H(f)|^2$

48. **Lowpass 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

$$E\left[\left|\boldsymbol{X}(t) - \sum_{k=-\infty}^{+\infty} \boldsymbol{X}(kT_s) \text{ sinc } 2W(t - kT_s)\right|^2\right] = 0$$

49. **Canonical representation of bandpass processes** A stationary bandpass process $\boldsymbol{X}(t) = \boldsymbol{R}(t)$ $\cos[\omega_c t + \theta(t)]$ can be represented in the canonical form, or the in-phase 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 lowpass 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, $\overline{\boldsymbol{X}^2(t)} = \overline{\boldsymbol{X}_I^2(t)} = \overline{\boldsymbol{X}_Q^2(t)}$.

REFERENCES AND SUGGESTED READING

1. A Papoulis and S Unnikrishna Pillai: *Probability, Random Variables and Stochastic Processes*, Fourth Edition, Tata-McGraw-Hill Publishing Company Limited, New Delhi, 2002.
2. W B Davenport Jr and W L Root: *Random Signals and Noise*, McGraw-Hill, New York, 1958.
3. R M Gray and L D Davisson: *Random Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
4. H Stark and J W Woods: *Probability, Random Processes and Estimation Theory for Engineers*, Second Edition, Prentice-Hall, N.J., 1994.
5. John G, Proakis and Masoud, Salehi: *Communication Systems Engineering*, second edition, Eastern Economy Edition, Prentice-Hall of India, New Delhi, 2006.
6. Simmon Haykins: *Communication Systems*, John Wiley & Sons, New York, 4th Edition, 2001.

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 variable.
14. Explain the meaning of 'conditional PDF of $\boldsymbol{Y}$ given $\boldsymbol{X}$'.
15. When do you say the r.v.s $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent? Uncorrelated? Orthogonal?
16. Define joint Gaussianity of two random variables.
17. State the properties of jointly Gaussian r.v.s.
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 (i) time variable is fixed, and (ii) outcome ξ is fixed.
21. Define 'first and second-order distribution functions' of a random process.
22. Explain what do 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? Uncorrelated? 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 lowpass band-limited processes.
32. How are the average powers of the 'in-phase' and 'quadrature' components related to the average power of a bandpass process?

FILL IN THE BLANKS

1. In probability theory, the 'sample space' consists of ____________.
2. $\{S, \mathcal{B}, P\}$ is called the ___________.
3. $P(A \cup B) =$ ___________.
4. $\dfrac{P(A,B)}{P(A)} =$ _____________.
5. Two events A and B are said to be independent events if $P(B) =$ _____________.
6. The probability density function of a discrete r.v. will be having ___________.
7. The cumulative distribution function of a random variable has a maximum value of __________.
8. ___________ and __________ completely specify a Gaussian density function.
9. ___________ random variable is quite useful in modeling a binary data generator.
10. If $f_X(x)$ is the density function of a continuous r.v. $\boldsymbol{X}$, its expectation is given by __________.
11. If $\boldsymbol{X}$ is a random variable, its characteristic function is the expectation of __________.
12. If $F_{XY}(x, y)$ is the joint distribution function of two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, the marginal distribution function of $\boldsymbol{Y}$ is given by __________.

13. The co-variance, C_{XY}, of two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as __________.
14. Random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be orthogonal if ___________.
15. Two random variables, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be jointly Gaussian if their joint density function is of the form $f_{XY}(x,y) =$ __________.
16. Jointly Gaussian random variables are completely characterized by their ______, ________ and _________, ____________.
17. For jointly Gaussian random variables _________ implies statistical independence.
18. The CDF of the average of n i.i.d. random variables, each with a mean of m and variance σ^2 will converge to $N($__, __$)$
19. A ________ _________ will be obtained when a random process is observed at a fixed instant of time.
20. A random process $X(t, \xi)$ is a continuous random process if $t \in$ __________ and is a discrete random process if $t \in$ __________.
21. The second-order distribution function of a random process $\boldsymbol{X}(t)$ is denoted by $F(x_1, x_2; t_1, t_2)$ and is given by $P[$ ________, ________$]$.
22. The auto-covariance $C_X(t_1, t_2)$ of a random process $\boldsymbol{X}(t)$ is given by $C_X(t_1, t_2) =$ _____________.
23. The random process $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be uncorrelated if their ________ for all t_1 and t_2.
24. A random process $\boldsymbol{X}(t)$ is said to be wide-sense stationary if ________ and _______.
25. Random processes whose time averages equal the ensemble averages, are known as _________ processes.
26. For any real or complex stationary process, $R_X(-\tau) =$ ________.
27. If a Gaussian process is WSS, it is _________.
28. The auto-correlation function of a white-noise process is an ______ _________.
29. If a WSS process with power spectrum $S_X(f)$ is given as input to an LTI system with transfer function $H(f)$, the cross-spectral density $S_{XY}(f)$ of the input and output processes is given by $S_{XY}(f) =$ _______.
30. If $X(t)$ is a zero-mean, stationary bandpass process, then $X_I(t)$ and $X_Q(t)$, its in-phase and quadrature components, are zero-mean, __________ stationary _________ (bandpass/lowpass) processes.

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 6 red balls. Box B contains 8 white balls and 2 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. Figure MCQ-4 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

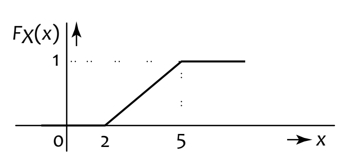


Fig. MCQ-4

5. A zero-mean bandpass signal has identically distributed Gaussian processes as its in-phase and quadrature components. The envelope of the bandpass process has a
   (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) $E[\boldsymbol{X}^2]$ (b) $\{E[\boldsymbol{X}]\}^2$ (c) $E[\boldsymbol{X}^2] - \{E[\boldsymbol{X}]\}^2$ (d) $E[\boldsymbol{X}^2] + \{E[\boldsymbol{X}]\}^2$
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) $0.6 + \left(e^{j\omega} - 1\right)$ (b) $0.6 - \left(e^{j\omega} - 1\right)$
   (c) $1 - 0.6\left(e^{j\omega} - 1\right)$ (d) $1 + 0.6\left(e^{j\omega} - 1\right)$
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 $E[\boldsymbol{X}.\boldsymbol{Y}] = E[\boldsymbol{X}]E[\boldsymbol{Y}]$. 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 $R_X(t)$ of a stationary process $\boldsymbol{X}(t)$ is a
   (a) deterministic function with maximum value at $\tau = 0$
   (b) deterministic function which is periodic
   (c) stationary random process
   (d) 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 they are
    (a) orthogonal
    (b) uncorrelated
    (c) orthogonal and one of them has zero-mean
    (d) 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) $\dfrac{2N_0}{4+\omega^2}$

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 second-to18 second interval? What is the probability of his making the call in the 10-to-18 second 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 section *A*, 10% of the students from Section *B* and 15% of the students from Section *C*, 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 $P[\ 2 \leq x \leq 3]$. Also, find $P[\ 2 \leq x \leq 3;$ Given $\boldsymbol{X} \geq 1]$.
11. Find the mean, variance and the density function of r.v. $\boldsymbol{Y}$ 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. A zero-mean Gaussian noise with a variance of 10^{-6} is rectified using a full-wave rectifier. Assuming the rectifier to be an ideal one, determine the density function of the rectified noise. What is its expected value?
14. $\boldsymbol{X}$ 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 r.v. $\boldsymbol{Z} = \boldsymbol{X} + \boldsymbol{Y}$.
15. $\boldsymbol{X}$ 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 r.v. $\boldsymbol{Z} = \sqrt{\boldsymbol{X}^2 + \boldsymbol{Y}^2}$.
16. $\boldsymbol{X}(t)$ is a stationary random process and $\boldsymbol{X}'(t)$ is its derivative. Show that for any fixed t_1, the random variable $\boldsymbol{X}(t_1)$ and $\boldsymbol{X}'(t_1)$ are orthogonal. Are they also uncorrelated?
17. Two processes $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are said to be jointly stationary if they are individually stationary and their cross-correlation is a function only of τ. Is it possible for $R_{XY}(t_1, t_2)$ to be a function only of $\tau \triangleq t_1 - t_2$ without $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ being non-stationary individually?
18. 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.
19. 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?
20. If $S_X(f)$ is the power spectrum of a stationary random process, $\boldsymbol{X}(t)$, find the PSD's of the following processes.

    (a) $\boldsymbol{X}(t - T)$ where T is a constant

    (b) $\boldsymbol{X}(t) - \boldsymbol{X}(t - T)$
21. In Section 7.12, 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?
22. A white-noise process of PSD $= N_0/2$ is the input to an ideal LPF having a cutoff frequency of 2 kHz. If uncorrelated samples are required, at what rate should the output of the filter be sampled?
23. A zero-mean white Gaussian noise with power spectral density $N_0/2$ is passed through an ideal bandpass filter of centre frequency f_c and bandwidth $2W$. If the output process is $\boldsymbol{n}(t)$, determine

    (i) the density function of the envelope of $\boldsymbol{n}(t)$

    (ii) the density function of the envelope of the process $\boldsymbol{X}(t) = A\cos 2\pi f_0 t + \boldsymbol{n}(t)$, where A is a constant

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)	

8 Noise

On going through this chapter, the student

- *understands how noise degrades the quality of communication*
- *knows the various sources of noise and the characteristics of the noise generated by each of those sources*
- *can calculate the thermal noise voltage across a 2-terminal network of only resistors, or of resistors and reactive elements connected in some manner*
- *can 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*
- *can experimentally determine the noise figure of a given 2-port network*
- *can calculate the in-phase and quadrature components of a bandpass noise process given its envelope and phase-angle representation*

8.1 INTRODUCTION

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.

If there were to be no noise, perfect communication would be possible even with very little transmitter power since 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. Further, as these amplifiers do not produce any noise, the amplified signal will be an exact replica of the one that was transmitted, in so far as the shape of the waveform is concerned. But, in practice, noise is always present and it corrupts the received signal. Amplifying the received signal does not help, as the amplifiers amplify both the signal and noise components of the received signal equally and further they add some more noise power. Thus, the signal-to-noise ratio at the output of an amplifier can only be worse than what it is at the input.

We cannot even remove the noise by filtering because the noise, as we will be seeing in this chapter, is often of very large bandwidth—much more than that of the signal. We may by using filters, easily remove the out-of-band noise (i.e., noise outside the signal bandwidth) but not the in-band noise.

Thus, noise degrades the performance of a communication system and hence it is important that we study its characteristics and take all possible steps in the design of a communication system, to reduce its effect.

8.2 NOISE SOURCES AND CLASSIFICATION

As shown in Fig. 8.1, 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*.

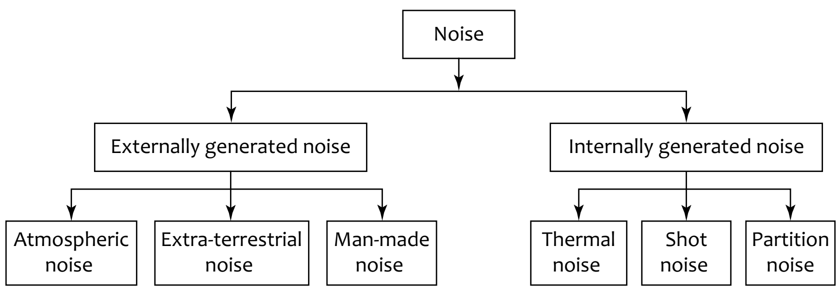


Fig. 8.1 *Types of noise*

8.3 ATMOSPHERIC NOISE

Atmospheric noise (also referred to as 'static') arises from lightning discharges (cloud-to-cloud, or cloud-to-earth), caused by thunderstorms. 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 day time, 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 day time) at all frequencies. Its intensity at night time becomes very low or insignificant, beyond about 10 MHz.

From this, 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.

8.3.1 Extra-terrestrial Noise

This has two components—'solar noise', and 'galactic noise'.

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 be having a cyclic variation with a 11-year period, called the *11-year sun-spot cycle*.

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.*

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 1 MHz to 500 MHz range. 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.

8.4 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 within 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.

Analysing 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

$$P(f) = \frac{2Rh|f|}{\left(e^{\frac{h|f|}{kT}} - 1\right)} \text{ volt}^2/\text{Hertz} \tag{8.1}$$

where,

h = Planck's constant = 6.6×10^{-34} Joule-second,

k = Boltzmann's constant = 1.38×10^{-23} J/K

In Eq. (8.1),

$$e^{\frac{h|f|}{kT}} = 1 \quad \text{at } |f| = 0$$

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'Hospital's rule and is given by

$$P(f)\Big|_{f=0} = 2kTR \text{ volts}^2/\text{Hz} \tag{8.2}$$

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

$$P(f) = 2\ kTR \ \text{volt}^2/\text{Hz} \tag{8.3}$$

Note: It must be noted that $P(f)$, as given in Eq. (8.3) represents the two-sided power spectral density as shown in Fig. 8.2.

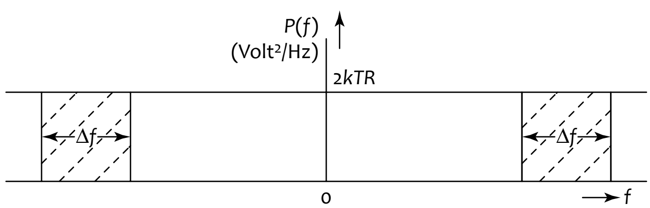


Fig. 8.2 *PSD of thermal noise*

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. 8.2, we find that

$$\text{Mean-squared value of noise in } R = (2\Delta f)\ 2kTR = 4\ kTR\Delta f \ \text{volt}^2 \tag{8.4}$$

$$\boxed{\left.\begin{array}{l}\text{Hence r.m.s. value of voltage}\\ \text{across the resistor } R\end{array}\right\} = 2\sqrt{kTR\Delta f}\ \text{volts}} \tag{8.5}$$

From the foregoing, it is clear that in so far as noise calculations are concerned, we may model a resistor of R ohms at temperature T K as follows.

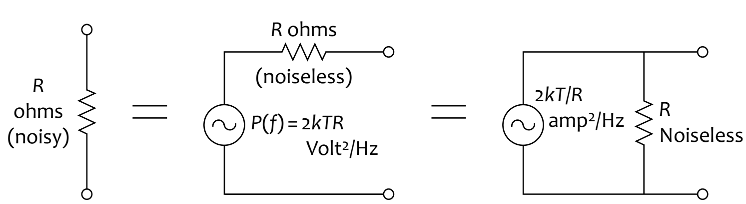


Fig. 8.3 *Modeling a noisy resistor*

We shall now make use of the noise model of a resistor shown in Fig. 8.3 to obtain the noise-equivalent circuits of resistances in series and in parallel.

8.4.1 Resistors in Series and in Parallel

(i) Series Connection

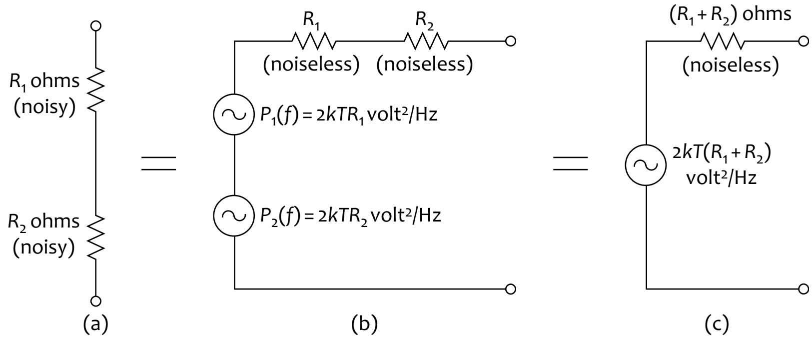


Fig. 8.4 *Resistors in series*

Superposition of PSD's In Fig. 8.4, 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 $\left[n_1(t)+n_2(t)\right]$ is given by

$$P_{12} = E\left[n_1(t)+n_2(t)\right]^2 = E\left[n_1^2(t)\right] + E\left[n_2^2(t)\right] + 2E\left[n_1(t)n_2(t)\right]$$

But since the noise processes produced in R_1 and R_2 are independent and zero-mean processes,

$$E\left[n_1(t)n_2(t)\right] = 0$$

Further

$$E\left[n_1^2\ (t)\right] = P_1 = \text{Average power of the noise process } n_1(t)$$

and
$$E\left[n_2^2\ (t)\right] = P_2 = \text{Average power of the noise process } n_2(t)$$

∴
$$P_{12} = P_1 + P_2$$

Thus, it is their powers (or the mean squared values) which get added, and not the voltages. *This means that, as shown in Fig. 8.4, 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.*

(ii) Parallel Connection

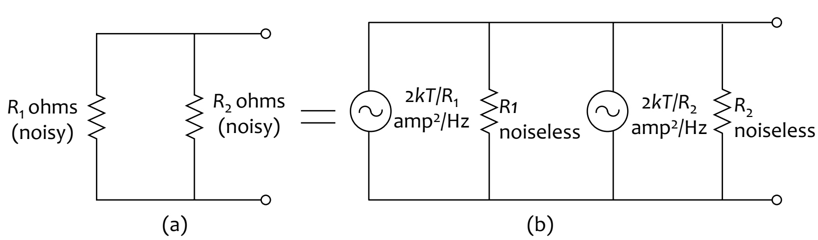


Fig. 8.5 *Resistances in parallel*

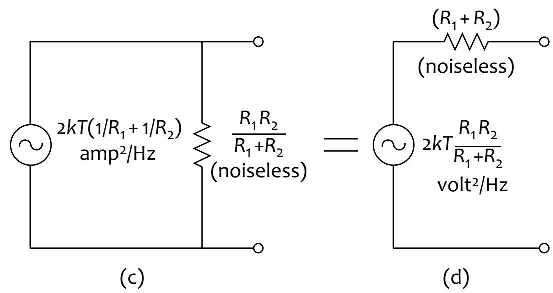


Fig. 8.5 *Resistances in parallel*

Note:

1. ***When two resistors are in series, it is their noise powers, or their noise power spectral densities (PSD's) in volt²/Hz, which can be added, and not their noise voltages.***
2. ***When two resistors are in parallel, it is their noise powers, or their noise PSD's 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 8.1

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.

From Eq. (8.5) we have

$$e_{\text{r.m.s.}} = \sqrt{4kTR(\Delta f)} \text{ volts}$$

$$= \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 10^6 \times 10^4}$$

$$= \sqrt{12 \times 1.38 \times 10^{-11}} = 12.868 \ \mu\text{V}$$

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 typically a few microvolts. 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 8.2

A 10 kΩ and a 20 kΩ resistor 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 (i) each one of them, (ii) their series combination, and (iii) their parallel combination.

(i) (a) Across the 10 kΩ resistor
From Eq. (8.5), 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) (b) Across the 20 kΩ resistor

$$e_{\text{r.m.s.}} = \sqrt{4kTR(\Delta f)} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 2 \times 10^4 \times 10^5}$$

$$\therefore \quad e_{\text{r.m.s.}} = 5.75\ \mu\text{V}$$

(iii) 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

$$e_{\text{r.m.s.}} = \sqrt{(4.07)^2 + (5.75)^2} = 7.04\ \mu\text{V}$$

(iv) With the two resistors in parallel

$$\text{Resistance of parallel combination } \frac{10 \times 20}{(10+20)} = 6.67\ \text{k}\Omega$$

$$\therefore \quad e_{\text{r.m.s.}} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 6.67 \times 10^3 \times 10^5} = \sqrt{110.4 \times 10^{-13}}$$

$$\therefore \quad e_{\text{r.m.s.}} = 3.32\ \mu\text{V}$$

8.4.2 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 8.3

A resistor of 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.

Representing the resistance R by its noise equivalent circuit, we have the following:

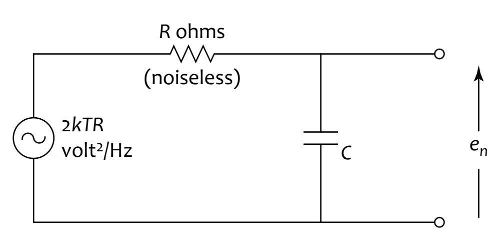


Fig. 8.6 *Noise equivalent circuit of Example 8.3*

For the R-C lowpass filter of Fig. 8.6, the transfer function is given by

$$H(f) = \frac{1}{1 + j\omega CR}$$

$$\therefore \quad |H(f)|^2 = \frac{1}{1 + 4\pi^2 f^2 C^2 R^2}$$

$$\therefore \qquad P_0(f) = \left.\begin{matrix}\text{PSD of the noise voltage}\\ \text{across the capacitor } C\end{matrix}\right\} = 2kTR.|H(f)|^2 = \frac{2kTR}{1+4\pi^2 f^2C^2R^2} \text{ volt}^2/\text{Hz}$$

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 $f = +\infty$.

$$\therefore \qquad P_0 = \int_{-\infty}^{\infty} \left(\frac{2kTR}{1+4\pi^2 f^2C^2R^2}\right) df$$

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 R-C lowpass filter goes on decreasing with R as the cutoff 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 8.4

The input circuit of an RF amplifier is a tuned circuit comprising of a coil having a resistance 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.

Let the tuned circuit be at resonance. Consider a small bandwidth Δf around the resonance frequency. The 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

$$e_{\text{r.m.s.}} = \sqrt{4kTr(\Delta f)} \qquad (8.6)$$

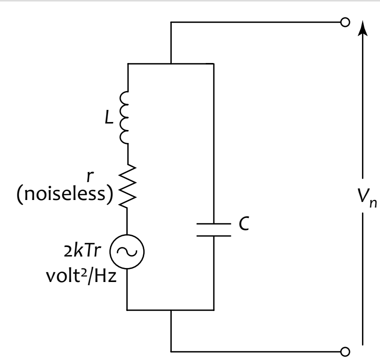


Fig. 8.7 *Tuned circuit of Example 8.4*

$$\therefore \qquad \left.\begin{matrix}\text{At resonance, the r.m.s. value}\\ \text{of voltage across the capacitor } C\\ \text{over a bandwidth of } \Delta f\end{matrix}\right\} = \sqrt{4kT(Q^2r)(\Delta f)}$$

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.

$$\therefore \qquad \left.\begin{array}{l}\text{at resonance, input noise voltage}\\ \text{(r.m.s. value) for the r.f amplifier}\end{array}\right\} = \sqrt{4kTR_d(\Delta f)} \qquad (8.7)$$

Equation (8.7) 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 8.5

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?

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 resistance 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

$$= \sqrt{4kTR_d(\Delta f)} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 20 \times 10^3 \times 2925}$$

$$= \sqrt{96876 \times 10^{-17}} = 98.42 \times 10^{-8}\ \text{V} = 0.9842\ \mu\text{V}$$

8.4.3 Available Noise Power

The maximum power transfer theorem tells us that maximum power will be delivered by the source to the load resistance R_L of Fig. 8.8, 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

$$\left(\frac{V}{2R}\right)^2 \cdot R = \frac{V^2}{4R} \triangleq \text{available power from the source}$$

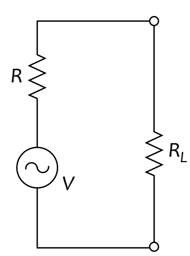


Fig. 8.8 *Maximum power transfer*

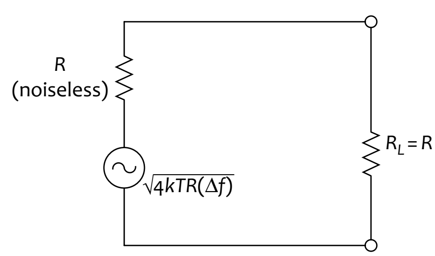


Fig. 8.9 *Maximum noise power transfer from a resistor*

Considering a resistor of R ohms as a thermal noise source as shown in Fig. 8.9, we have,

$$\text{Available noise power} = \left(\frac{\sqrt{4kTR(\Delta f)}}{2R}\right)^2 \cdot R = kT(\Delta f)$$

$$\boxed{\therefore \text{ available noise power} = kT(\Delta f)\text{watts}} \tag{8.8}$$

Noise Temperature of a Source The noise temperature of a source is defined as T:

$$T \triangleq \frac{p}{k(\Delta f)} \tag{8.8a}$$

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.

8.4.4 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. 8.10, is called *white noise*.

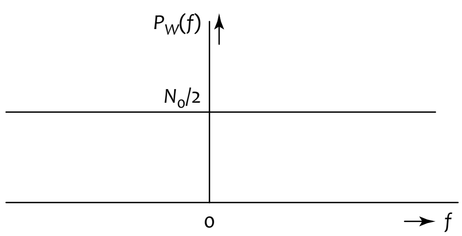


Fig. 8.10 *PSD of white noise*

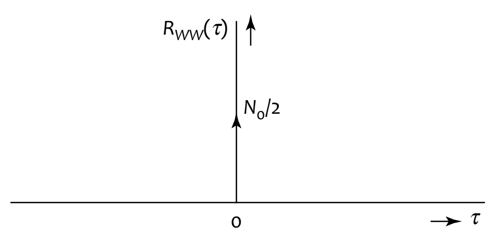


Fig. 8.11 *ACF of white noise*

Thus, the auto-correlation function of white noise is given by

$$R_{WW}(\tau) = \mathcal{F}^{-1}\left[\frac{N_0}{2}\right] = \frac{N_0}{2}\delta(\tau) \tag{8.9}$$

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 8.6

White noise with PSD of $N_0/2$ is filtered using an ideal LPF whose cutoff 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?

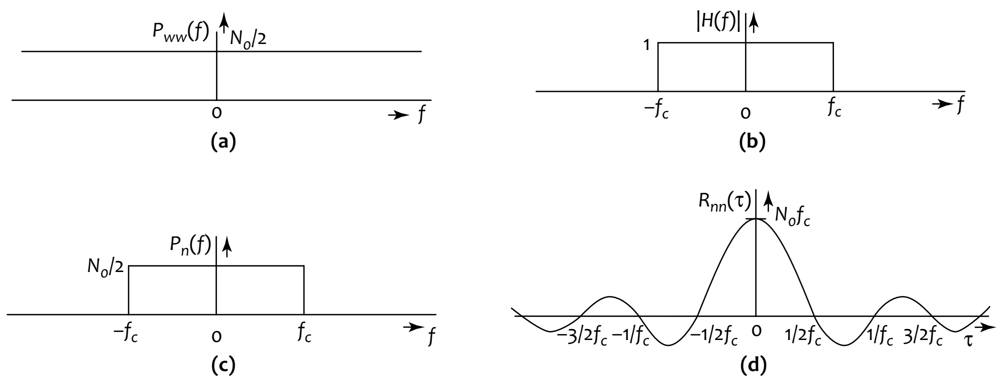


Fig. 8.12 **(a)** *PSD of white noise* **(b)** *Magnitude response of the ideal LPF* **(c)** *PSD of noise at output of LPF* **(d)** *ACF of output noise, $R_{nn}(\tau)$*

PSD of noise at output of LPF $= |H(f)|^2 \times$ 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. 8.12(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. 8.12(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.

SHOT NOISE 8.5

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 as '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. 8.13. 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.

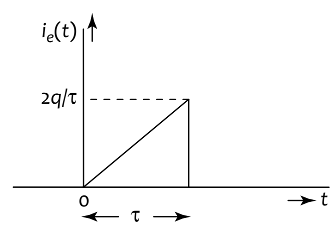


Fig. 8.13 *Current waveform created by a single electron*

Suppose a steady current of 1 milli-ampere is flowing through the diode under temperature-limited condition. A current of 1 m.a. 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 microsecond, or that exactly 6×10^3 electrons reach the anode every 10^{-12} second. 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. 8.14, 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.

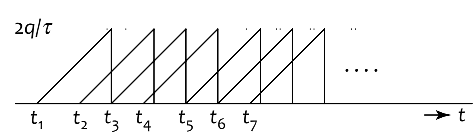


Fig. 8.14 *Waveforms of current contributed by randomly emmited electrons*

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. 8.14 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

$$i_e(t) \xleftrightarrow{\text{FT}} I_e(f) \tag{8.10}$$

But $$i_e(t) \approx q\delta(t)$$

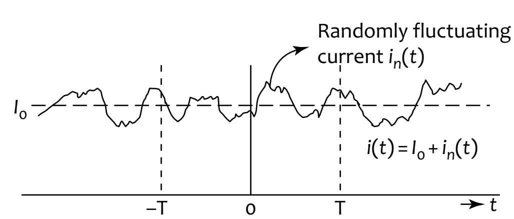


Fig. 8.15 *Anode current waveform obtained by summing up the various triangular current pulses of Fig. 8.14*

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

$$i_{nT}(t) \triangleq \begin{cases} i_n(t); & |t| \le T \\ 0 & ;\text{otherwise} \end{cases} \tag{8.11}$$

$i_{nT}(t)$ is thus a segment of $i_n(t)$ and is of duration $2T$. Hence $i_{nT}(t)$ is a finite energy signal. Let

$$i_{nT}(t) \xleftrightarrow{\text{FT}} I_{nT}(f)$$

Following the arguments similar to those of Section 2.9.3 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

$$P_{nI}(f) = \underset{T\to\infty}{\text{Lt}}\ E\left[\frac{|I_{nT}(f)|^2}{2T}\right] \tag{8.12}$$

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.

$$\boxed{\therefore P_{nI}(f) = I_0 q \text{ amp}^2/\text{Hz}} \tag{8.13}$$

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. (8.13) 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

$$I_n^2 = 2I_0 q(\Delta f)\ \text{amp}^2 \tag{8.14}$$

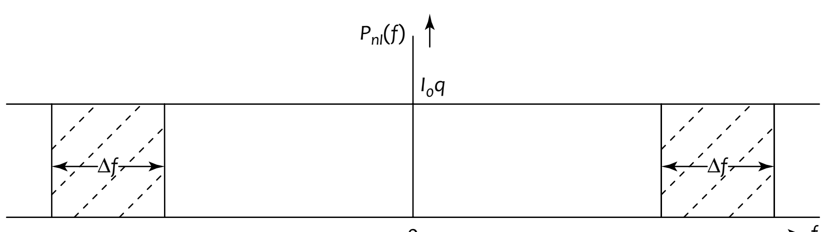


Fig. 8.16 *Showing that $P_{nl}(f)$ given by Eq. (4.13) is a 2-sided power spectrum*

8.5.1 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. (8.13) is modified as

$$P_{nI}(f) = \alpha I_0 q \ \text{amp}^2/\text{Hz} \tag{8.15}$$

In this equation, 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)

$$\alpha = \frac{1.28kT_c g_d}{qI_0} \tag{8.16}$$

where, k = Boltzmann's constant = 1.38×10^{-23} Joules per degree Kelvin
T_c = Cathode temperature in degree Kelvin
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

8.5.2 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:

$$I_n^2 = 2(I + 2I_0)q(\Delta f) \ \text{amp}^2 \tag{8.17}$$

In Eq. (8.17), 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 the '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 8.7

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.

(i) Determine the PSD of the output noise neglecting the effect of the coupling condenser.
(ii) Find the ratio of mean-squared value of the thermal noise to the mean-squared value of the total noise at the output.

With a direct current of I_0 amperes flowing through it, the mean-squared value of the shot noise generated by the diode in temperature-limited condition is given by

$$I_{\text{sh}}^2 = 2I_0 q(\Delta f) \ \text{amp}^2 \quad \text{(from Eq. 8.14)}$$

The thermal noise generated by the resistance of R ohms has a mean-squared value given by

$$I_{\text{th}}^2 = \frac{4kT(\Delta f)}{R} \text{ amp}^2 \quad \text{(from Fig. 8.3)}$$

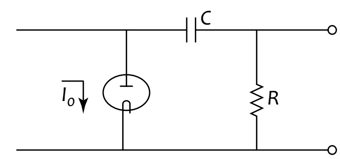


Fig. 8.17 *Circuit of Example 8.7*

$\therefore$ PSD of the total output noise $= \left(2I_0 q + \dfrac{4kT}{R}\right)$ amp²/Hz

and,
$$\left(\frac{\text{Mean-squared value of the thermal noise}}{\text{Mean-squared value of total noise}}\right) = \frac{4kT(\Delta f)/R}{\left(2I_0 q + \dfrac{4kT}{R}\right)\Delta f}$$

$$= \left(\frac{4kT}{4kT + 2I_0 qR}\right)$$

8.6 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 lowpass filter whose passband gain is $|H(0)| = |H(f)|\Big|_{f=0}$, 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.*

With White Noise of PSD Equal to $N_0/2$ Applied as Input

(i) Output noise power of the filter $= \dfrac{N_0}{2}\displaystyle\int_{-\infty}^{\infty} |H(f)|^2\, df = N_0 \int_0^{\infty} |H(f)|^2 df$

Note: With $h(t)$ real-valued, $|H(f)|$ must have even symmetry.

(ii) Output noise power of the ideal LPF $= \dfrac{N_0}{2}\displaystyle\int_{-B}^{B} |H(0)|^2\, df$

$$= N_0 B|H(0)|^2$$

$$N_0 B|H(0)|^2 = N_0 \int_0^{\infty} |H(f)|^2\, df$$

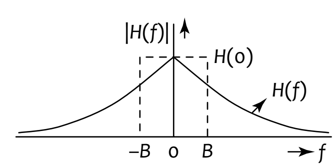


Fig. 8.18 *Transfer functions of T and the ideal LPF of passband gain H(0) and bandwidth B Hz*

$\therefore$ B = Noise-equivalent bandwidth of the filter T

$$= \frac{\displaystyle\int_0^{\infty} |H(f)|^2\, df}{|H(0)|^2}$$

$\therefore$ noise-equivalent bandwidth $B = \dfrac{\displaystyle\int_0^{\infty} |H(f)|^2\, df}{|H(0)|^2}$, for a filter with transfer function $H(f)$ (8.18)

Example 8.8

Determine the noise equivalent bandwidth of the R-C lowpass filter shown in Fig 8.19.

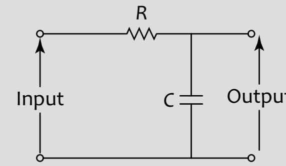


Fig. 8.19 *Circuit of Example 8.8*

For the *R-C* lowpass filter,

$$H(f) = \frac{1}{1 + j\omega CR} \qquad \therefore \quad |H(f)|^2 = \frac{1}{1 + 4\pi^2 f^2 C^2 R^2}$$

$\therefore$ when white noise of PSD equal to $N_0/2$ is applied as input to the *R-C* lowpass filter, the output noise power is

$$P_1 = \frac{N_0}{2}\int_{-\infty}^{\infty} \frac{1}{1 + 4\pi^2 f^2 C^2 R^2} df = N_0 \int_0^{\infty} \frac{1}{1 + 4\pi^2 f^2 C^2 R^2} df$$

Put $2\pi fCR = \tan\theta \quad \therefore \quad 2\pi CRdf = \sec^2 d\theta$

$\therefore$
$$P_{no} = N_0 \int_0^{\pi/2} \frac{(1/2\pi CR)\sec^2\theta}{\sec^2\theta} d\theta = \frac{N_0}{4RC}$$

When white noise of PSD equal to $N_0/2$ is applied as input to an ideal LPF of bandwidth B and passband gain $= H(0) = 1$, the corresponding output noise power is

$$P_2 = 2B\left(\frac{N_0}{2}\right) \cdot |H(0)|^2 = N_0 B$$

$\therefore$
$$N_0 B = \frac{N_0}{4RC} \quad \text{i.e.,} \quad B = \frac{1}{4RC}$$

Example 8.9

If zero-mean white noise of 2-sided PSD η/2 W/Hz is applied as input to the lowpass R-C filter of Fig. 8.19, determine and sketch the PSD and auto-correlation function of the filtered noise.

The transfer function $H(f)$ of this filter is

$$H(f) = \frac{1}{1 + j\omega RC}$$

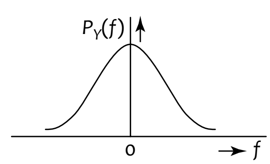


Fig. 8.20(a) *$P_y(f)$ of output noise process*

The PSD of the input white noise process is

$$P_X(f) = \eta/2$$

$\therefore$ from Eq. 7.87, 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

$$R_Y(\tau) = \frac{\eta}{4RC} e^{-\left(\frac{|\tau|}{RC}\right)}$$

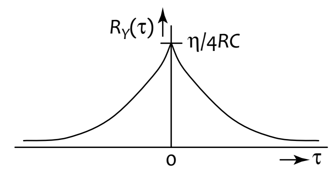


Fig. 8.20(b) *IFT of $P_y(f)$*

Example 8.10

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?

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 $= Q^2 \cdot r = 1600 \times 1.9894 = 3183\Omega = R_d$

Since this is an R-C lowpass filter, the noise equivalent bandwidth is given by

$$B_N = \frac{1}{4R_dC} = \frac{10^{12}}{4 \times 3483 \times 20}$$

$$= \frac{10^6}{25464}$$

$$= 39.27\ \text{MHz}$$

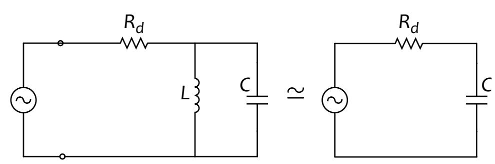


Fig. 8.21 *Noise model of the parallel resonant circuit and its approximate equivalent circuit beyond resonance frequency*

Example 8.11

Determine the noise equivalent bandwidth of a normalized lowpass Butterworth filter of order 2.

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 cutoff frequency. Hence, for a normalized second-order Butterworth filter, putting $n = 2$ and $B = 1$, we get

$$|H_2(f)|^2 = \frac{1}{1 + f^4}$$

$$\therefore \qquad |H_2(f)|^2\Big|_{f=0} = \frac{1}{1} = 1 = |H_2(0)|^2$$

$\therefore$ the noise equivalent BW of a second-order Butterworth filter is given by (refer to Eq. (8.17)

$$B_N = \text{Noise-equivalent bandwidth}$$

$$= \frac{\int_0^\infty |H_2(f)|^2 df}{|H_2(0)|^2} = \int_0^\infty \frac{1}{1+f^4} df \quad \text{since } |H_2(0)|=1$$

But $$\int_0^\infty \frac{x^{m-1}}{1+x^n} dx = \frac{\pi/n}{\sin(m\pi/n)} \quad \text{for } n > m > 0. \quad \text{(refer Appendix A)}$$

$$\therefore \quad \int_0^\infty \frac{1}{1+f^4} df = \frac{\pi/4}{\sin(\pi/4)} = \frac{1}{\sqrt{2}} \cdot \frac{\pi}{4} \quad \therefore \quad B = \text{Noise-equivalent bandwidth} = \frac{1}{\sqrt{2}} \cdot \frac{\pi}{4}$$

8.6.1 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 8.12

For a 2-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

We shall start from the output side of stage 2 and work backwards.

Step (i) A resistance of 500k at the output of the 2nd stage is equivalent, in so far as noise contribution is concerned, to a resistor of $\frac{500 \times 10^3}{(20)^2} = 1.25k$ at the input of the 2nd stage.

Step (ii) The resistor of the first stage (30k) and the input resistor of the second stage (90k) are in parallel and this parallel combination is in series with the noise equivalent resistance (10k) of the second stage and the 1.25 kΩ obtained in the step (i)

i.e., $$\left(\frac{30 \times 90}{30+90} + 10 + 1.25)\right) \times 10^3\,\Omega = 33.75\text{ k}\Omega$$

Step (iii) The resistance of 33.75 kΩ obtained at the output of the first stage will be equivalent, in so far 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

$$R_{eq} = 500 + 1000 + 337.5 = 1837.5\ \Omega$$

8.7 NOISE FIGURE AND EQUIVALENT NOISE TEMPERATURE OF 2-PORT NETWORKS

8.7.1 Signal-to-Noise Ratio

As mentioned earlier in the discussion on the result of Example 8.1, 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

$$\boxed{\text{SNR} \quad \triangleq \quad \frac{\text{Signal power}}{\text{Noise power}}} \tag{8.19}$$

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.

$$\boxed{(\text{SNR})_{\text{dB}} \quad = \quad 10\log_{10}\left[\frac{\text{Signal power}}{\text{Noise power}}\right]} \tag{8.20}$$

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

$$\text{Signal power} = G.S_i$$

$$\text{Noise power} = (G.N_i + N_a)$$

$$\text{Thus, the SNR at the output} = \left(\frac{S}{N}\right)_0 = \frac{G \cdot S_i}{(G.N_i + N_a)} < \frac{S_i}{N_i} \tag{8.21}$$

Therefore, for any amplifier, or, for that matter, for any 2-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. (8.21), 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. A 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/2-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/2-port linear network is. This ratio is called the *noise figure* of the amplifier.

$$\boxed{F = \text{ Noise figure } \underline{\Delta} \frac{(S/N)_i}{(S/N)_0} = \frac{N_0}{GN_i} = \frac{GN_i + N_a}{GN_i}} \tag{8.22}$$

With regard to the 'noise figure', there are a few points that need to be noted.

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 2-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

(i) From Eq. (8.21), we have available output noise power $N_0 = F.GN_i$. Now making use of Eq. (8.8) for N_i, we have

$$N_0 = \text{Available output noise power}$$
$$= FGkT_0(\Delta f) \tag{8.23}$$

where, T_0 is the room temperature. In RHS of Eq. (8.22), $GkT_0(\Delta f)$ 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 $F(>1)$

(ii) From Eq. (8.22),

$$\frac{N_0}{G} = FkT_0(\Delta f)$$

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

$$N'_a = \text{Internally generated noise referred to input} = (F-1)\ kT_0(\Delta f) \tag{8.24}$$

Example 8.13

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.

From Eq. (8.23), internally generated noise, referred to the input

$$= (F-1)kT_0(\Delta f)$$

From Eq. (8.22), available output noise power $= FGKT_0(\Delta f)$

Now, internally generated noise referred to the output $= G(F-1)kT_0(\Delta f)$

$\therefore$ 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 SNR's and not in decibles.

$\therefore$ 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 8.14

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.

From Eq. 8.22, 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 $27°\text{C} = 300$ K.

$$\therefore \qquad F = \frac{N_0}{GkT_0(\Delta f)} = \frac{80\times10^{-9}}{10^4\times1.38\times10^{-23}\times25\times10^6\times300}$$

$$= \frac{2318}{300} = 7.7267$$

$$\therefore \qquad 10\log_{10} F = 10\log_{10} 7.7267. = 8.879 \text{ dB}$$

8.7.2 Equivalent Noise Temperature

Although the noise figure F gives a good measure of the degree of noisiness of a device, amplifier, or any 2-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 2-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 2-port network referred to its input.

From Eq. (8.22), 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

$$N_i = kT_0(\Delta f); \quad T_0 = \text{room temperature}$$

Hence,

$$F = \frac{GN_i + N_a}{GN_i} = \frac{GkT_0(\Delta f) + GkT_e(\Delta f)}{GkT_0(\Delta f)} = \frac{T_0 + T_e}{T_0}$$

$$\boxed{\therefore \quad F = 1 + \left(\frac{T_e}{T_0}\right)} \tag{8.25}$$

or

$$\boxed{T_e = (F-1)T_0} \tag{8.26}$$

From Eq. (8.26) it is clear as to why the use of T_e is preferable for lownoise 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).

8.7.3 Noise Figure of Amplifiers in Cascade

In communication engineering, quite often we come across a number of amplifiers or 2-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 2-ports.

In this connection, let us recapitulate the following

(i) From Eq. (8.22), we have

$$F = \frac{GN_i + N_a}{GN_i} = \frac{\text{Actual output noise power}}{\text{Noise output power if the amplifier is noise-free}} \tag{8.27}$$

(ii) 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 $FG\ kT_0(s\Delta f)$ 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. (8.24).

Now consider the cascade connection of two amplifiers as shown.

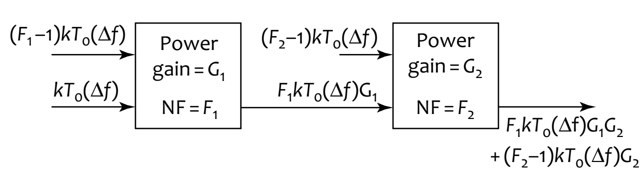


Fig. 8.22 *Cascade connection of two amplifiers*

Then the overall noise figure F is given by

$$F = \frac{\text{Actual output noise power}}{\text{Output noise power assuming the amplifiers to be noise-free}}$$

$$= \frac{F_1 k T_0 (\Delta f) G_1 G_2 + (F_2 - 1) k T_0 (\Delta f) G_2}{k T_0 (\Delta f) G_1 G_2}$$

$$\therefore \qquad F = F_1 + \frac{(F_2 - 1)}{G_1} \tag{8.28}$$

This is known as *Friis's formula*. It may be extended to any number of amplifiers connected in cascade.

$$\boxed{F = F_1 + \frac{(F_2 - 1)}{G_1} + \frac{(F_3 - 1)}{G_1 G_2} + \frac{(F_4 - 1)}{G_1 G_2 G_3} + \ldots..} \tag{8.29}$$

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.

8.7.4 Improvement of Overall Noise Figure

(1) From Eq. (8.29), 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.

8.7.5 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 (8.28) and (8.25), we have

$$1 + \frac{T_e}{T_0} = 1 + \frac{T_{e1}}{T_0} + \frac{T_{e2}}{G_1 T_0} + \frac{T_{e3}}{G_1 G_2 T_0} + \ldots$$

$$\boxed{\therefore T_e = T_{e_1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} + \ldots.} \tag{8.30}$$

Example 8.15

A source with an internal resistance of 50 Ω and an internal e.m.f. 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.

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.

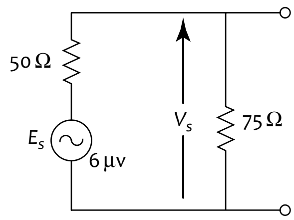


Fig. 8.23 *Signal source voltage circuit*

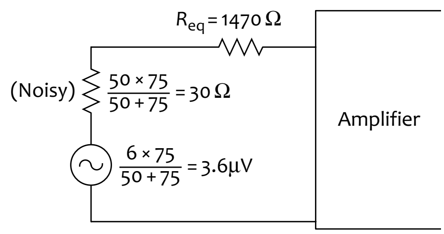


Fig. 8.24 *Input circuit of the amplifier, noise equivalent resistance of the amplifier is included*

Mean-squared value of the noise voltage

$$= 4kT_0(R_{th} + R_{eq})(\Delta f)$$

$$= 4 \times 1.38 \times 10^{-23} \times 290 \times 5 \times 10^3 (30 + 1470)$$

$$= 12 \times 10^{-14} \text{ volt}^2$$

Mean-squared value of the signal voltage $= (3.6)^2 \times 10^{-12}$ volt^2

$$= 12.96 \times 10^{-12} \text{ volt}^2$$

Hence (S/N) ratio $= \dfrac{12.96 \times 10^{-12}}{12 \times 10^{-14}} = 108$

$$\therefore \qquad \left(\frac{S}{N}\right)_{dB} = 10 \log_{10} 108 = 20.3 \text{ dB}$$

Example 8.16

(i) Determine the noise figure F of the amplifier of Example 8.15 in dB. (ii) Also determine its equivalent noise temperature.

(i) We know from Eq. (8.27) 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

$$= \left[4kT_0\left(R_{th} + R_{eq}\right)(\Delta f)\right]G = G \times 12 \times 10^{-14} \text{ volt}^2$$

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

$$= \left[4kT_0 R_{th}(\Delta f)\right]G = G \times 2.4 \times 10^{-15}$$

$$\therefore \qquad F = \frac{G \times 12 \times 10^{-14}}{G \times 2.4 \times 10^{-15}} = 50 = 10 \log_{10} 50 = 16.9 \text{ dB}$$

Note: As can be seen from the above steps, F is in fact, given by

$$F = \frac{R_{\text{th}} + R_{\text{eq}}}{R_{\text{th}}} = 1 + \frac{R_{\text{eq}}}{R_{\text{th}}} = 1 + \frac{1470}{30} = 1 + 49 = 50$$

(ii) Since $F = 1 + \dfrac{T_e}{T_0}$, we have $T_e = (50-1)290 = 14,210$ K

Example 8.17

A low-noise amplifier of 30 K equivalent noise temperature 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?

$$T_e = T_{e_1} + \frac{T_{e_2}}{G_1} \qquad \text{(see Eq. 8.29)}$$

$T_{e_1} = 30$K, $G_1 = 20$ dB $= 100$;

$F_2 =$ Noise figure of the microwave receiver $= 25$ dB $= 10^{2.5} = 316.228$

$\therefore$ $T_{e_2} =$ Equivalent Noise temperature of the microwave receiver $= (F_2 - 1)T_0$

$$= (315.228) \times (273 + 27) = 315.228 \times 300 = 945.684 \text{ K}$$

$\therefore$ overall noise equivalent temperature $= T_e = T_{e_1} + \dfrac{T_{e_2}}{G_1}$

$$= 30° + \frac{316.228}{100} = 33.16228 \text{ K}$$

8.7.6 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. 8.25.

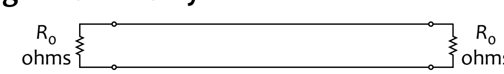


Fig. 8.25 *Lossy transmission line terminated in its R_0*

For simplicity, let us assume that the line and the resistances of R_0 ohms each, are all at the ambient temperature T_0K. 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

$$\left.\begin{array}{l}\text{Available internally generated} \\ \text{noise power at the output}\end{array}\right\} = kT_e g_L(\Delta f)$$

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

$$kT_e g_L(\Delta f) = kT_0(\Delta f) - kT_0 g_L(\Delta f)$$

$$\therefore \quad T_e = \left(\frac{T_0}{g_L} - T_0\right) = T_0(L-1)$$

$$\therefore \quad T_e = T_0(L-1)$$

But, we know, from Eq. (8.26) 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 8.18

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. 8.26(a).

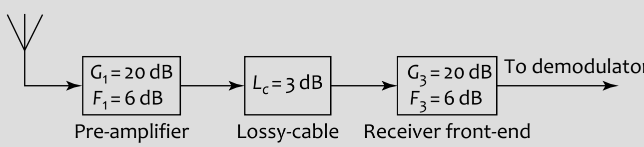


Fig. 8.26(a) *Antenna with pre-amplifier*

(i) Find the overall noise figure of the system.
(ii) 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. *(University Question)*

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 = Noise figure of the lossy cable = L (ratio) = 2

(i) Applying Friis's formula for the overall noise figure,

$$F = F_1 + \frac{(F_C - 1)}{G_1} + \frac{(F_3 - 1)}{G_1 . G_C}$$

where $G_C = (1/LC)$ = Gain of the cable

$$\therefore \quad F = 3.981 + \frac{(2-1)}{100} + \frac{(39.8-1)}{(100 \times 1/2)}$$

$$= 4.767 = 6.782 \text{ dB}$$

$6\text{ dB} = 10^{0.6} = 3.981$
$3\text{ dB} = 10^{0.3} = 2$
$16\text{ dB} = 10^{1.6} = 39.8$
$60\text{ dB} = 10^{6} = 10^{6}$
$20\text{ dB} = 10^{2} = 100$

(ii) 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. 8.26(b).

$\therefore$ overall noise figure F is now given by

$$F = F_C + \frac{(F_3 - 1)}{(1/LC)} = 2 + \frac{(39.8-1)}{(1/2)} = 79.6$$

$$\therefore \qquad F_{\text{dB}} = 10\log_{10} 79.6 = 19.01 \text{ dB}$$

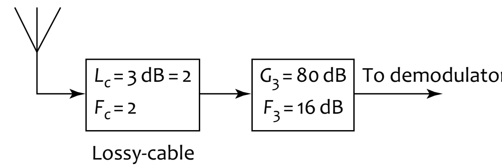


Fig. 8.26(b) *System configuration for part (ii)*

Example 8.19

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.

(i) LNA at the input followed by the cable connecting to the main receiver.
(ii) The input direct to the cable, which is then connected to the LNA, which, in turn, is connected to the main receiver. *(University Question)*

As the value of the ambient temperature, T_0 is not given, let us conveniently assume it as 17°C = 290 K.

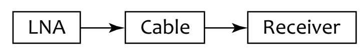


Fig. 8.27(a) *First configuration for Example 8.19*

(i) For the first case, configuration is as follows:

$$G_1 = 47\text{ dB} = 10^{4.7} = 50118.72 \qquad L_C = \text{loss} = 6.5 \text{ dB} = 4.4668 = 1/G_c \qquad F_3 = 7 \text{ dB} = 10^{0.7} = 5.0118$$

$$\therefore \quad T_{e_1} = 120^\circ\text{K} \qquad F_C = 4.4668 \qquad T_{e_c} = (F_3 - 1)T_0$$

$$T_{e_c} = (F_c - 1) = 1005.38 \qquad = 4.0118 \times 290 = 1163.442$$

Applying Friis's formula for the overall equivalent noise temperature T_e (see Eq. 8.29)

$$T_e = T_{e_1} + \frac{T_{e_c}}{G_1} + \frac{T_{e_3}}{G_1.G_C} = 120 + \frac{1005.38}{50118.72} + \frac{1163.442}{(50118.72/4.4668)}$$

$$= 120 + 0.0200 + 0.10369 = 120.12 \text{ K}$$

(ii) For the second case, the configuration is as follows:

Fig. 8.27(b) *Second configuration for Example 8.19*

$$L_C = \text{loss} = 6.5 \text{ dB} = 10^{0.65} = 4.4668 \qquad G_1 = 47\text{dB} = 10^{4.7} = 50118.72 \qquad F_3 = 7 \text{ dB} = 10^{0.7} = 5.0118$$

$$F_C = 10^{0.65} = 4.4668 \qquad T_{e_c} = (F_3 - 1)T_0$$

$$T_{e_c} = 1005.38; G_C = 1/L_C \qquad T_{e_1} = 120 \text{ K} \qquad = 4.0118 \times 290 = 1163.442$$

$\therefore$ applying Friis'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 2-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 2-port, *referred to its input*.

Example 8.20

A coil having an inductance of 2 henry and an internal resistance of 1 ohm is shunted by a capacitor of 2 farads. Determine the power density spectrum of the thermal noise at the network terminals. *(University Question)*

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. 8.28.

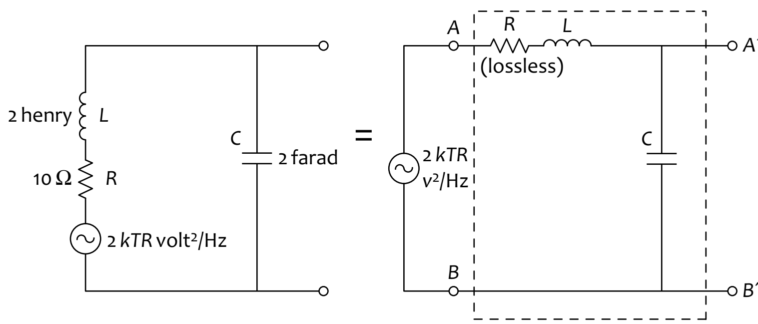


Fig. 8.28 *Circuit for Example 8.20*

The 2-port network shown inside the dotted line box has input power spectral density of 2 kTR volt^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 2-port.

Now,

$$H(f) = \frac{1/j\omega C}{R + j\omega L + \left(1/j\omega C\right)} = \frac{1}{(1-\omega^2 LC) + j\omega RC}$$

$$\therefore \quad |H(f)|^2 = \frac{1}{(1-\omega^2 LC)^2 + \omega^2 R^2 C^2}$$

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 $= 2kTR|H(f)|^2$ volt^2/Hz

$$= \frac{2kT}{(1+16\omega^4 - 4\omega^2)} \quad \text{since } R = 1\Omega$$

8.7.7 Measurement of Noise Figure

In Section 8.5 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. (8.14) 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 upto 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 set-up 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. 8.29.

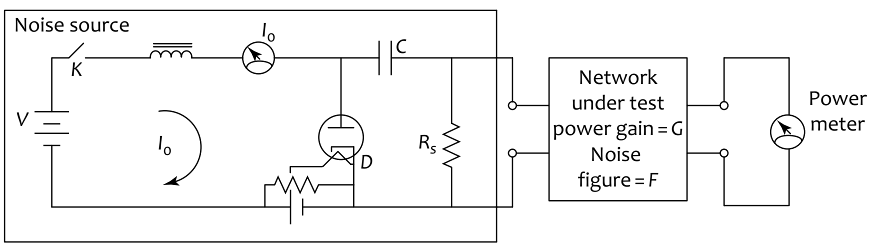


Fig. 8.29 *Measurement of noise figure*

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 Eq. (8.4) and (8.27) we have

$$P_1 = 4kTR_S(\Delta f)GF \tag{8.31}$$

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.

$$\therefore \quad P_2 = 4kTR_S(\Delta f)GF + 2qI_0R_S^2G(\Delta f) \tag{8.32}$$

$$\therefore \quad \frac{P_2}{P_1} = 1 + \frac{2qI_0R_S^2G(\Delta f)}{4kTR_S(\Delta f)GF} = 1 + \frac{qI_0R_S}{2kTF}$$

Note: In Eq. (8.32), 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

$$2 = 1 + \frac{qI_0'R_S}{2kTF} \quad \text{i.e.,} \quad F = \frac{qI_0'R_S}{2kT}$$

Now, if $T = 290$K,

$$\therefore \quad F = \frac{1.6\times 10^{-19} I_0' R_S}{2\times 1.38\times 10^{-23}\times 290} = 20 I_0' R_S$$

$$\therefore \quad F = 20 I_0' R_S$$

Thus, we will be able to determine the value of the noise figure of the 2-port network from the values of I_0' and R_S.

8.8 NARROW-BAND 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 narrow-band filter with centre frequency f_c, the received noise will then be a narrow-band noise process with centre frequency f_c and its PSD will be some what as shown in Fig. 8.30.

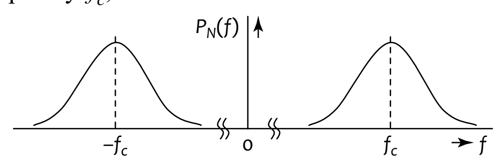


Fig. 8.30 *PSD of noise entering the receiver*

Earlier, in Section 2.8 of Chapter 2, we had shown that it is possible to represent a narrowband signal, $x(t)$, with centre frequency f_c in terms of its inphase and quadrature components as

$$x(t) = x_I(t)\cos\omega_c t - x_Q(t)\sin\omega_c t \quad \text{(ref to Eq. 2.163)}$$

where, the lowpass signal $x_I(t)$ and $x_Q(t)$ were respectively called the in phase and quadrature components of the signal $x(t)$.

In the present case, we are not dealing with a narrowband 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

$$\text{Pre-envelope of } \; n(t) = n_+(t) = n(t) + j\hat{n}(t) \tag{8.33}$$

and complex-envelope of $n(t) = \tilde{n}(t) = n_+(t)\exp[-j2\pi f_c t]$ (8.34)

Let the lowpass complex process $\tilde{n}(t)$ be represented as

$$\tilde{n}(t) = n_I(t) + jn_Q(t) \tag{8.35}$$

Since $\tilde{n}(t)$ is a lowpass process of bandwidth, say W Hz, $n_I(t)$ and $n_Q(t)$ are also lowpass of the same bandwidth.

From Eq. (8.33), we have

$$n(t) = \text{Re}\left[n_+(t)\right] = \text{Re}\left[\tilde{n}(t)e^{j\omega_c t}\right] \quad \text{(from Eq. 8.34)}$$

Now, substituting for $\tilde{n}(t)$ using Eq. (8.35), we get

$$\therefore \qquad n(t) = \text{Re}\left[\left\{n_I(t) + jn_Q(t)\right\}\ \left\{\cos\omega_c t + j\sin\omega_c t\right\}\right]$$

$$\boxed{\therefore n(t) = n_I(t)\cos\omega_c t - n_Q(t)\sin\omega_c t} \tag{8.36}$$

RHS of Eq. (8.36) is called the in-phase 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 in-phase and quadrature components, viz., $n_I(t)$ and $n_Q(t)$. These are stated in the chapter on probability and random processes.

SUMMARY

1. Noise degrades the performance of communication system.
2. 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.
3. 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.
4. Extra-terrestrial noise can cause disturbance to communications in the frequency range 20 MHz to 1.5 GHz.
5. Man-made noise causes disturbance to communications in the 1 MHz–500 MHz frequency range.
6. 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 upto 200 GHz. Hence, for all practical purposes it can be considered as a zero-mean white noise.
7. For thermal noise, $P(f) = 2KTR$ volts^2/Hz; where K is Boltzmann's constant, T is absolute temperature in K and R is the resistance in ohms.
8. $\left.\begin{array}{l}\text{r.m.s. value of noise voltage across a resistor}\\ \text{of } R \text{ ohms in a bandwdith of } \Delta\text{f Hz, at } T°\text{K}\end{array}\right\} = 2\sqrt{KTR(\Delta f)}$ volts.
9. 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.
10. 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.
11. 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.
12. *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.
13. *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.

14. In multi-electrode electron devices like triodes, pentodes, BJT's 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.
15. The **noise equivalent bandwidth** of a filter with transfer function $H(f)$ is defined as the bandwidth B of an ideal LPF whose passband 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.
16. 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.
17. SNR = Signal-to-noise ratio $\triangleq \dfrac{\text{Signal power}}{\text{Noise power}}$

$$(\text{SNR})_{\text{dB}} = 10\log_{10}\left[\frac{\text{Signal power}}{\text{Noise power}}\right]$$

18. 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.
19. Available output noise power = $FGkT_0(\Delta f)$

$$\left.\begin{array}{l}\text{Internally generated noise of a system referred to}\\ \text{the input of that system}\end{array}\right\} = (F-1)kT_0(\Delta f)$$

20. For low-noise amplifiers and devices, it is more convenient to use noise temperature instead of noise figure.

    The **equivalent noise temperature** of a device, or a 2-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 2-port network, referred to its input.
21. $T_e = (F-1)T_0$, where T_e is noise equivalent temperature of a devices/network whose noise figure is F and T_0 is the room temperature.
22. **Friis's formula** for noise figure of amplifiers in cascade

$$F = F_1 + \frac{(F_2-1)}{G_1} + \frac{(F_3-1)}{G_1G_2} + \frac{(F_4-1)}{G_1G_2G_3} + ...$$

    For the overall noise temperature of amplifiers in cascade:

$$T_e = T_e + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1G_2} + ...$$

23. A bandpass noise $n(t) = R(t)\cos\left[\omega_c t + \theta_n(t)\right]$ can be represented in the canonical form or the 'in-phase 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 in-phase and quadrature components respectively, are such that they are lowpass processes and

$$\overline{n_i^2(t)} = \overline{n_q^2(t)} = \overline{n^2(t)}$$

REFERENCES AND SUGGESTED READING

1. John G Proakis; Masoud Salehi: *Communication Systems Engineering*, Second Edition, Prentice-Hall of India Private Limited, 2006.
2. J J Freeman, : *Noise*, John Wiley & Sons, Inc, New York, 1958.
3. John C Hancock: *An Introduction to the Principles of Communication Theory*, McGraw-Hill Book Company, Inc. New York, 1961.
4. Simon Haykin: *Communication Systems*, 3/e, John Wiley & Sons, Singapore, 1995.

REVIEW QUESTIONS

1. In a communication scenario, what is meant by noise?
2. Name the important components of external noise and internal noise.
3. In which bands of the electromagnetic spectrum is communication affected by atmospheric noise? Why?
4. What are the sources of 'galactic noise'? What is the range of frequencies over which this noise has its spectral components?
5. What is the origin of thermal noise? Comment on its power spectral density.
6. Explain the meaning of the term: 'available noise power'.
7. 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?
8. How does 'shot noise' originate? Comment the power spectrum of shot-noise current.
9. What is partition noise?
10. Define and explain the term: 'noise equivalent bandwidth of a filter'.
11. What is meant by the 'equivalent noise resistance' of an amplifier?
12. The 'signal-to-noise ratio' at the output of an amplifier is given to be 200. What is its value in decibels?
13. Define and explain the terms: noise figure and noise temperature of a 2-port network? How are they related?
14. 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.

FILL IN THE BLANKS

1. Noise may be broadly classified as __________ and _________.
2. Atmospheric noise, ______________ and ________ come under the category of _______ while Thermal noise, __________ and __________ come under the category of ____________.
3. Disturbance to communication caused by atmospheric noise is more severe in the __________ band.
4. Extra-terrestrial radiation has spectral components from _______ to _______.
5. The two-sided PSD of the thermal noise from a resistor of R Ω at a temperature of T K is _______ volt^2/Hz.
6. The r.m.s. value of the thermal noise voltage across a resistor of R Ω at a temperature of T K, measured over a bandwidth of (Δf) is __________volts.
7. The available noise power from a resistor of R Ω at a temperature of T K over a bandwidth (Δf) is _________ watts.
8. White noise has a PSD which is __________.

9. The ACF of white noise with $(N_0/2)$ as its PSD is _________.
10. We cannot have perfect white noise source in practice because ___________.
11. Thermal noise may be considered as ____________ mean _________ Gaussian noise.
12. In a temperature-limited vacuum diode carrying a direct current of I_0 amperes, the PSD of shot noise is _______ amp^2/Hz.
13. In a temperature-limited vacuum diode carrying a direct current of I_0 amperes, the r.m.s. value of the shot-noise current, measured over a bandwidth (Δf) Hz is ______ amperes.
14. For any amplifier, (S_0/N_0) __________ (S_i/N_i). $(\geq , \leq)$
15. For any amplifier, the noise figure is always ___________ 1 $(\geq , \leq)$.
16. In the case of lownoise amplifiers, it is better to use ________ rather than the noise figure.
17. At a room temperature of 290 K, the noise figure of an amplifier was found to be 1.01. What is its equivalent noise temperature?
18. A number of amplifiers with available power gains G_1, G_2, . . . and noise figures F_1, F_2, . . . are connected in cascade in that order, If $G_1 >> 1$, the overall noise figure is approximately equal to _________.

MULTIPLE CHOICE QUESTIONS

1. The effect of atmospheric noise is most severe in
(a) medium wave band (b) shortwave band
(c) VHF band (d) microwave region
2. 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
3. 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
4. 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
5. 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
6. The r.m.s. value of the thermal noise voltage across a resistor of R Ω 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
7. The r.m.s. value of the thermal noise voltage across resistors R_1 and R_2 are 3 microvolts and 4 microvolts respectively. The r.m.s. value of the thermal noise across their series combination is
(a) 10 microvolts (b) 7 microvolts
(c) 5 microvolts (d) none of the above
8. 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
9. 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) $2kTR(\Delta f)$

10. 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
11. 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 amperes is
(a) 2 I_0q amp²/Hz (b) I_0q amp²/Hz
(c) ½ I_0q amp²/Hz (d) none of the above
12. 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
13. The noise equivalent bandwidth of an L-section *RC*-lowpass 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
14. 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
15. 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
16. When a number of amplifiers are connected in cascade, the overall noise figure is approximately equal to the
(a) noise figure of the most noisy amplifier (b) noise figure of the least noisy amplifier
(c) sum of noise figures of all the amplifiers (d) noise figure of the first amplifier

PROBLEMS

1. Thermal noise voltage (r.m.s.) across a resistor has been found to be 10 microvolts at a temperature of 27°C and over some bandwidth *B* Hz. What will be the r.m.s. thermal noise voltage at (i) 77°C with bandwidth *B* Hz, and (ii) 77°C with a bandwidth 2*B* Hz?
2. 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.
3. 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 (i) they are all connected in series, and (ii) when they are all connected in parallel.
4. 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?
5. 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.

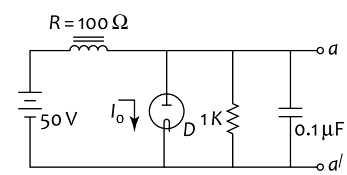


Fig. P-8.1

6. 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?

7. 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 (i) the output *(S/N)* ratio in dB, and (ii) the noise figure of the amplifier.
8. 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 needed?
9. 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.
10. 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 the amplifier has a noise figure $F_4 = 6$ dB and a gain of 60 dB,
    (i) Find the overall noise figure and noise temperature of the system
    (ii) Find the noise figure and noise temperature of the system with pre-amplifier and cable interchanged *(VTU August 2002)*.

Key to Multiple Choice Questions

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

9 Noise Performance of AM and FM Systems

Through this chapter, the student

- *learns the way the channel and the receiver are modeled for a study of the noise performance of the system*
- *will be able to analyze the noise performance of FM and the various types of AM systems and compare them*
- *learns that an FM system offers the possibility of power bandwidth trade-off and understands that there is a limit for this trade-off*
- *understands clearly how and why it is possible to improve the destination SNR for FM systems by employing pre-emphasis and de-emphasis*
- *learns that there is a threshold effect in FM receivers, understands how it arises and studies the various methods of threshold extension*

9.1 INTRODUCTION

In Chapter 8, we had discussed various types of noise, their sources and characteristics and noted that thermal noise and shot noise are both white so far as the frequencies and bandwidths used in practical communication systems are concerned. In the fourth chapter, 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 the fifth chapter, we studied angle modulation and discussed the modulation and demodulation methods for FM and PM. In the sixth chapter, we discussed the details of AM and FM transmitters and receivers. In the seventh, we had reviewed probability and random processes.

In the present chapter, we will make use of the material covered in the previous five chapters and examine, by deriving the necessary expressions, the noise performance of 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 continuous-wave 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 demodulation 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 PSD of $\eta/2$. For each type of modulation and demodulation 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 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.

9.2 DESTINATION SNR OF A BASE-BAND 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 lowpass filter with a cutoff 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 $\eta/2$. Hence, the shaded area of Fig. 9.1 represents the average noise power lying within the bandwidth of the baseband and corrupting the signal.

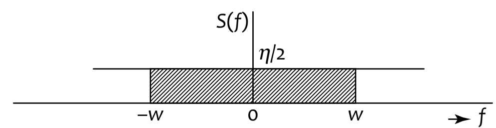


Fig. 9.1 *Noise power corrupting the baseband signal*

$$\therefore \qquad N = \begin{Bmatrix} \text{Average noise power} \\ \text{within the bandwidth } W \end{Bmatrix} = 2W.\eta/2 = \eta W$$

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 cutoff frequency equal to W, the destination signal-to-noise ratio is

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} \triangleq \gamma \tag{9.1}$$

The $\left(\frac{S}{N}\right)_D$ of other systems are generally compared with this.

9.3 MODEL FOR LINEAR MODULATION SYSTEMS

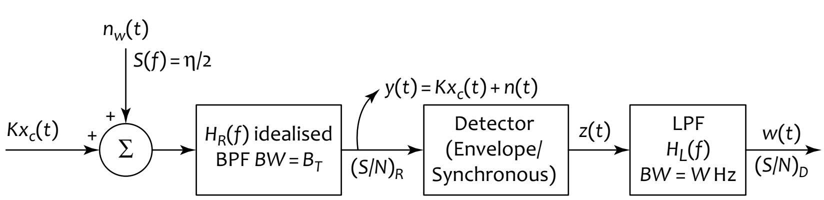


Fig. 9.2 *Model for linear modulation systems*

$x_c(t)$ is the transmitted modulated signal. k is the attenuation factor so that $K.x_c(t)$ is the received signal assumed to be having a carrier frequency $f_c = f_{i.f}$, the intermediate frequency. $n_W(t)$ is the zero-mean white Gaussian noise of the channel with a two-sided power spectral density of $\eta/2$. It gets added to the received signal. $H_R(f)$ is an ideal bandpass 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 $f_c = f_{i.f}$, the intermediate frequency, and the gain of the filter is unity in its passband. 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 bandpass noise having an average power $= (\eta/2).(2B_T) = \eta B_T$. 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 cutoff 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 $(S/N)_D$.

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

$$\left(\frac{S}{N}\right)_C = \text{Channel SNR} \triangleq \frac{\text{Average power of the modulated signal}}{\text{Average power of noise in the message bandwidth } (W)}$$

$$= \frac{S_R}{2W(\eta/2)} = \frac{S_R}{\eta W} = \gamma$$

$$\therefore \text{ figure of merit} = \left[\left(\frac{S}{N}\right)_D \cdot \frac{1}{\gamma}\right] \tag{9.1a}$$

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. The higher the figure of merit as compared to 1, the better it is.

Pre-detection Signal-to-Noise Ratio

$$y(t) = \text{Detector input} = Kx_c(t) + n(t) \tag{9.2}$$

where, $n(t)$ is a bandpass noise with an average power of ηB_T where, $B_T = 2W$ or W, depending on the type of modulation.

$$S_R = \text{Average signal power at the input to the detector}$$
$$= K^2 \overline{x_c^2(t)} \tag{9.3}$$

where, the overbar on $x_c^2(t)$ denotes its average value.

$$N_R = \text{Average noise power at the input to the detector}$$
$$= \overline{n^2(t)} = (\eta/2)(2B_T) = \eta B_T \tag{9.4}$$

$$\therefore \text{ pre-detection signal-to-noise ratio } = \left(\frac{S}{N}\right)_R = \frac{S_R}{N_R} = \frac{S_R}{\eta B_T} \tag{9.5}$$

But, we have already seen that in the case of baseband transmission, the destination SNR is given by

$$\gamma = \frac{S_R}{\eta W} \tag{9.6}$$

Hence, we may express the pre-detection SNR, viz $(S/N)_R$ as

$$\left(\frac{S}{N}\right)_R = \frac{S_R}{\eta B_T} = \frac{S_R}{\eta W}\left(\frac{W}{B_T}\right) = \gamma\left(\frac{W}{B_T}\right) \tag{9.7}$$

With this, we are now ready to proceed with the determination of the destination signal-to-noise ratios for various linear modulation schemes.

9.4 $(S/N)_D$ FOR SSB-SC SYSTEMS

For SSB-SC systems, the detector is a synchronous detector and the transmitted signal is given by

$$x_c(t) = \frac{1}{2}A_c\left[x(t)\cos\omega_c t \mp \hat{x}(t)\sin\omega_c t\right] \text{ (refer Eqs 5.63 and 5.64)} \tag{9.8}$$

with (−) sign for the USSB and (+) sign for the LSSB

But $$S_R = K^2\overline{x_c^2(t)} = \frac{1}{4}K^2A_c^2\left[\overline{x^2(t)\cos^2\omega_c t \mp 2x(t)\hat{x}(t)\sin\omega_c t\cos\omega_c t + \hat{x}^2(t)\sin^2\omega_c t}\right]$$

$$= \frac{1}{2}\frac{1}{4}K^2A_c^2\overline{x^2(t)} + \frac{1}{2}\left(\frac{1}{4}\right)K^2A_c^2\overline{\hat{x}^2(t)}$$

But $\overline{x^2(t)} = \overline{\hat{x}^2(t)}$ since the Hilbert transform doesn't alter the power.

$$\therefore \quad S_R = \frac{1}{4}A_R^2\overline{x^2(t)} \tag{9.9}$$

where $$A_R = KA_c$$

Now, input to the synchronous detector is given by

$$y(t) = Kx_c(t) + n(t) \tag{9.10}$$

If we assume that it is a USSB system and substitute for $x_c(t)$ using Eq. (9.8), and further, if we substitute for the bandpass noise $n(t)$ in Eq. (9.10) by its in-phase and quadrature component representation, we get

$$y(t) = \frac{1}{2}A_Rx(t)\cos\omega_c t - \frac{1}{2}A_R\hat{x}(t)\sin\omega_c t + n_i(t)\cos\omega_c t - n_q(t)\sin\omega_c t \tag{9.11}$$

In the synchronous detector, $y(t)$ is multiplied by $\cos\omega_c t$ and lowpass filtered to give $w(t)$ (see Fig. 9.2)

$$\therefore \quad z(t) = y(t)\cos\omega_c t = \frac{1}{2}A_Rx(t)\cos^2\omega_c t - \frac{1}{4}A_R\hat{x}(t)\sin 2\omega_c t + n_i(t)\cos^2\omega_c t - \frac{1}{2}n_q(t)\sin 2\omega_c t$$

When this $z(t)$ is lowpass filtered using an ideal LPF with a cutoff frequency of W Hz, all high-frequency components are rejected.

$$\therefore \quad w(t) = \frac{1}{4}A_Rx(t) + \frac{1}{2}n_i(t) \tag{9.12}$$

In the above, the first term is the signal term and the second term is the noise term.

$$\left(\frac{S}{N}\right)_D - \frac{(1/4)^2 A_R^2 \overline{x^2(t)}}{(1/4)\overline{n_i^2(t)}} = \frac{(1/4) A_R^2 \overline{x^2(t)}}{\eta B_T} = \frac{S_R}{\eta B_T} = \frac{S_R}{\eta W} = \gamma \tag{9.13}$$

(since $B_T = W$ for SSB)

In the above, we have made use of the properties of the in-phase and quadrature components of a zero-mean bandpass 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. (9.4). Since it is an SSB-SC system, $B_T = W$ and so $\overline{n_i^2(t)} = \eta W$.

$$\boxed{\therefore \left(\frac{S}{N}\right)_{\substack{D \\ \text{SSB–SC}}} = \gamma} \tag{9.14}$$

Thus, the 'figure of merit' for an SSB-SC system is (from equations 9.14 and 9.7)

$$\underset{\text{SSB-SC}}{\text{Figure of merit}} = \frac{(S/N)_D}{(S/N)_C}$$

$$\therefore \text{ figure of Merit for SSB-SC} = \left(\frac{\gamma}{\gamma}\right) = 1 \tag{9.14a}$$

9.5 DSB-SC SYSTEMS

For a DSB-SC system, a coherent or synchronous demodulator will be used and the modulated signal is given by

$$x_c(t) = A_c x(t) \cos \omega_c t \tag{9.15}$$

$\therefore$ received signal $= Kx_c(t)$

and
$$Kx_c(t) = KA_c x(t) \cos \omega_c t = A_R x(t) \cos \omega_c t \tag{9.16}$$

$$\therefore \quad S_R = \text{Received signal power} = K^2 \overline{x_c^2(t)} = \frac{1}{2}\overline{x^2(t)} \cdot A_R^2 \tag{9.17}$$

Also,
$$B_T = 2W \tag{9.18}$$

Input to the detector $= y(t) = A_R x(t) \cos \omega_c t + n(t)$ (9.19)

But
$$n(t) = n_i(t) \cos \omega_c t - n_q(t) \sin \omega_c t$$

$$\therefore \quad y(t) = \left[A_R x(t) + n_i(t)\right] \cos \omega_c t - n_q(t) \sin \omega_c t \tag{9.20}$$

The synchronous detector multiplies $y(t)$ by $\cos \omega_c t$

$\therefore$ $z(t)$ = Output of the multiplier in the detector

$$= \left[A_R x(t) + n_i(t)\right] \cos^2 \omega_c t - n_q(t) \sin \omega_c t \cos \omega_c t$$

$$= \frac{1}{2}\left[A_R x(t) + n_i(t)\right] + \frac{1}{2}\left[A_R x(t) + n_i(t)\right] \cos 2\omega_c t - \frac{1}{2} n_q(t) \sin 2\omega_c t$$

$\therefore$ $w(t)$, the output of the lowpass filter is given by

$$w(t) = \frac{1}{2}\left[A_R x(t) + n_i(t)\right] \tag{9.21}$$

In this, the message signal component is $\frac{1}{2}A_R x(t)$ and the noise component is $\frac{1}{2}n_i(t)$.

$$\therefore \quad \left(\frac{S}{N}\right)_D = \frac{A_R^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} \tag{9.22}$$

But $\frac{1}{2}A_R^2 \overline{x^2(t)}$ = Received signal power = S_R (from Eq. 9.17) 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)
$\therefore$ substituting these values in Eq. (9.22), we get

$$\left(\frac{S}{N}\right)_D = \frac{2S_R}{2\eta W} = \frac{S_R}{\eta W} = \gamma$$

$$\boxed{\therefore \left(\frac{S}{N}\right)_{D \atop \text{DSB–SC}} = \gamma} \tag{9.23}$$

Thus, the 'figure of merit' for a DSB-SC system is (from equations 9.7 and 9.23)

$$\underset{\text{(DSB-SC)}}{\text{Figure of Merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{\gamma}{\gamma} = 1 \tag{9.23a}$$

Example 9.1

A DSB-SC signal is transmitted over a noisy channel, with the power spectral density of the noise being as shown in Fig. 9.3(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 watts, find the output signal-to-noise ratio of the receiver.

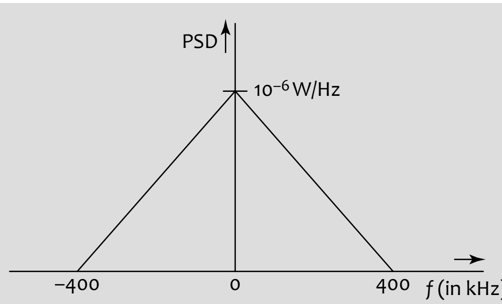


Fig. 9.3(a) *Noise PSD for Example 9.1*

Here, the additive noise on the channel is *not* white. It has a triangular shaped 2-sided power spectral density as shown in the figure. 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 bandpass noise entering the receiver is as shown by the shaded area in Fig. 9.3(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 8 kHz width and 0.5×10^{-6} W/Hz height.

$\therefore$ average power of the bandpass noise entering the receiver $= \overline{n^2(t)}$

$= $ total area of the shaded portion

$$= 2 \times \frac{1}{2} \times 10^{-6} \times 8 \times 10^3 = 8 \times 10^{-3}\text{W}.$$

But, from the properties of bandpass noise, we know that

$$\overline{n^2(t)} = \overline{n_i^2(t)}$$

$$\therefore \qquad \overline{n_i^2(t)} = 8 \times 10^{-3}\,\text{W}$$

Average power of the received DSB-SC signal

$$= \frac{1}{2} A_R^2 \overline{x^2(t)} \qquad \text{(from Eq. 9.17)}$$

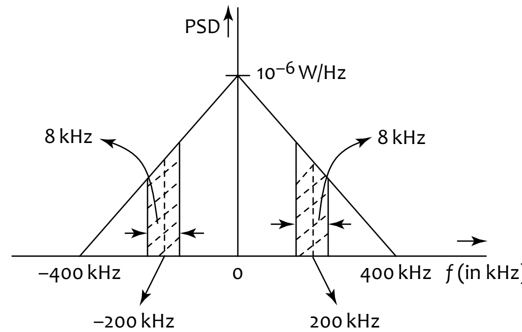


Fig. 9.3(b) *Two-sided noise power spectrum for Example 9.1*

$\therefore$ $S_R = $ received signal power $= \frac{1}{2} A_R^2 \overline{x^2(t)} = 10\text{W}$

But, from Eq. (9.22), we know that

$$\left(\frac{S}{N}\right)_D = \text{destination SNR} = \frac{A_R^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} = \frac{20}{8 \times 10^{-3}} = 2.5 \times 10^3$$

$$\therefore \qquad \left(\frac{S}{N}\right)_D = 2.5 \times 10^3 \text{ or } \left(\frac{S}{N}\right)_D \text{ in dB} = 10 \log_{10} 2.5 \times 10^3 = 33.97\text{dB}$$

9.6 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

$$B_T = 2W \tag{9.24}$$

where, of course, W is the bandwidth of $x(t)$, the message signal. The transmitted signal, $x_c(t)$ is given by

$$x_c(t) = A_c\left[1 + mx(t)\right]\cos\omega_c t \tag{9.25}$$

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.

9.6.1 AM System with a Synchronous Detector

$$\text{The received signal} = Kx_c(t) = KA_c\left[1 + mx(t)\right]\cos\omega_c t$$

$$= A_R\left[1 + mx(t)\right]\cos\omega_c t \tag{9.26}$$

$\therefore$ $S_R = $ Average received signal power $= K^2\overline{x_c^2(t)}$

$$\therefore \qquad S_R = A_R^2\overline{\left[1 + mx(t)\right]^2 \cos^2\omega_c t}$$

Since $x(t)$ is zero mean, the above expression reduces to

$$S_R = \frac{1}{2} A_R^2 \left[1 + m^2 \overline{x^2(t)}\right] \tag{9.27}$$

$y(t)$, the input to the synchronous detector is given by

$$y(t) = A_R\left[1 + mx(t)\right]\cos\omega_c t + n(t) \tag{9.28}$$

Replacing $n(t)$ in the above by its in-phase and quadrature component representation, we get

$$y(t) = \left\{A_R\left[1 + mx(t)\right] + n_i(t)\right\}\cos\omega_c t - n_q(t)\sin\omega_c t \tag{9.29}$$

The synchronous detector multiplies this by the carrier, i.e., $\cos\omega_c t$.

$$\therefore \qquad z(t) = \left\{A_R\left[1 + mx(t)\right] + n_i(t)\right\}\cos^2\omega_c t - \frac{1}{2} n_q(t)\sin 2\omega_c t \tag{9.30}$$

The lowpass filter removes all the high-frequency components as its cutoff 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

$$w(t) = \frac{1}{2}\left\{A_R\left[1 + mx(t)\right] + n_i(t)\right\} \tag{9.31}$$

In the above equation, $\frac{1}{2}A_R$ represents a dc component, $\frac{1}{2}A_R mx(t)$ 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

$$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} = \frac{S_D}{N_D} \tag{9.32}$$

But, $$\overline{n_i^2} = \overline{n^2} = \eta B_T = 2\eta W = N_D \tag{9.33}$$

(since $n(t)$ is of zero mean, variances of $n_i(t)$ and $n(t)$ will be equal)

Also, $$S_R = \frac{1}{2} A_R^2 \left[1 + m^2 \overline{x^2(t)}\right] \qquad \text{(from Eq. 9.27)}$$

$\therefore$ we may write

$$\left(\frac{S}{N}\right)_D = \frac{S_D}{S_R}\frac{S_R}{N_D} = \frac{A_R^2 m^2 \overline{x^2(t)}}{\frac{A_R^2}{2}[1 + m^2 \overline{x^2(t)}]} \cdot \frac{S_R}{2\eta W}$$

$$= \frac{m^2 \overline{x^2(t)}}{(1 + m^2 \overline{x^2(t)})} \cdot \frac{S_R}{\eta W} = \left[\frac{m^2 \overline{x^2(t)}}{(1 + m^2 \overline{x^2(t)})}\right] \cdot \gamma$$

$$\therefore \qquad \left(\frac{S}{N}\right)_{D\,\substack{\text{AM}\\ \text{Sync-det}}} = \left[\frac{m^2 \overline{x^2(t)}}{(1 + m^2 \overline{x^2(t)})}\right]\gamma \tag{9.34}$$

9.6.2 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. 9.2 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., $$z(t) = A_R[1 + mx(t)], \text{ where } A_R \triangleq KA_c.$$

However, with the channel noise, the detector input is

$$y(t) = KA_c[1 + mx(t)]\cos\omega_c t + n(t) \quad (9.35)$$

where, $n(t)$ is bandpass 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 in-phase and quadrature components.

$$\therefore \quad y(t) = A_R[1 + mx(t)]\cos\omega_c t + n_i(t)\cos\omega_c t - n_q(t)\sin\omega_c t$$

$$= \{A_R[1 + mx(t)] + n_i(t)\}\cos\omega_c t - n_q(t)\sin\omega_c t \quad (9.36)$$

$$\therefore \quad R_y(t) = \text{envelope of } y(t)$$

$$= \left[\{A_R[1 + mx(t)] + n_i(t)\}^2 + n_q^2(t)\right]^{1/2} \quad (9.37)$$

and the phase angle $\theta_y(t)$ is

$$\theta_y(t) = \tan^{-1}\left[\frac{n_q(t)}{A_R[1 + mx(t)] + n_i(t)}\right] \quad (9.38)$$

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

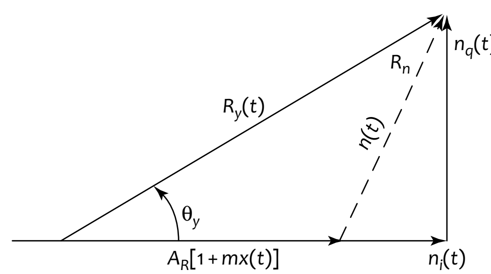


Fig. 9.4 *Phasor diagram of the components of $y(t)$, the input to the detector. $A_R^2 >> \overline{n^2}$ is assumed*

$$A_R^2 >> \overline{n^2(t)} \quad (9.39)$$

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. 9.3, in view of the assumption of Eq. (9.39), we may say that

$$P\left[A_R\ \{1 + mx(t)\} >> n_q(t)\right] \text{ is almost equal to unity.}$$

$$\therefore \quad R_y(t) \approx A_R[1 + mx(t)] + n_i(t) \quad (9.40)$$

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

$$w(t) = A_R mx(t) + n_i(t) \quad (9.41)$$

Since $A_R m x(t)$ represents the signal component and $n_i(t)$, the noise component of this output signal, we have

$$\left(\frac{S}{N}\right)_D = \frac{A_R m^2 \overline{x^2(t)}}{\overline{n_i^2(t)}} \tag{9.42}$$

But,

$$\overline{n_i^2(t)} = \overline{n^2(t)} = \eta B_T = 2\eta W \tag{9.43}$$

(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. (9.42) 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, $$S_R = \frac{1}{2} A_R^2 \left[1 + m^2 \overline{x^2(t)}\right] \quad \text{(see Eq. (9.27)}$$

$\therefore$ $$\left(\frac{S}{N}\right)_D = \frac{A_R^2 m^2 \overline{x^2(t)}}{\frac{1}{2} A_R^2 \left[1 + m^2 \overline{x^2(t)}\right]} \cdot \left(\frac{S_R}{\eta B_T}\right)$$

$$= \frac{m^2 \overline{x^2(t)}}{\left[1 + m^2 \overline{x^2(t)}\right]} \left(\frac{S_R}{\eta W}\right) \quad \text{since } B_T = 2W$$

$\therefore$ $$\left(\frac{S}{N}\right)_{\substack{D \\ \text{AM-Env.det}}} = \left[\frac{m^2 \overline{x^2(t)}}{\left[1 + m^2 \overline{x^2(t)}\right]}\right] . \gamma \tag{9.44}$$

Comparing Eqs (9.34) and (9.44), we find they are exactly the same. *However, it must be noted that Eq. (9.44) 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. (9.34) to be valid.*

The figure of merit for AM is, therefore, obtained from Eqs (9.7) and (9.44).

$$\underset{\text{(AM)}}{\text{Figure of merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{m^2 \overline{x^2} \gamma}{1 + m^2 \overline{x^2}} \cdot \frac{1}{\gamma}$$

$$= \frac{m^2 \overline{x^2}}{1 + m^2 \overline{x^2}} \tag{9.44a}$$

Example 9.2

Find the figure of merit of an AM system when the depth of modulation is (i) 100 % (ii) 50 %, and (iii) 30 %. *(University Question)*

$$\text{Figure of merit (FOM) of an AM system} = \frac{m^2 \overline{x^2}}{1 + m^2 \overline{x^2}} \quad \text{(from Eq. 9.44a)}$$

(i) $m = 1$, 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, $\overline{x^2} = 1/2$

$$\underset{m=1}{\text{FOM}} = \frac{1/2}{1+1/2} = \frac{1}{2}\times\frac{2}{3} = \frac{1}{3}$$

(ii) $m = 0.5$ i.e., 50%: FOM $= \dfrac{(0.5)^2\overline{x^2}}{1+(0.5)^2\overline{x^2}} = \dfrac{0.25\overline{x^2}}{1+0.25\overline{x^2}}$

For an $x(t)$ which is a single-tone, $\overline{x^2} = 1/2$

$$\therefore \qquad \underset{m=0.5}{\text{FOM}} = \frac{0.25\times 0.5}{1+0.25\times 0.5} = \frac{0.125}{1+0.125} = 0.111$$

(iii) $m = 0.3$ i.e., 30% modulation: FOM $= \dfrac{0.09\overline{x^2}}{1+0.09\overline{x^2}}$

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 9.3

Prove that the figure of merit of an AM system for single-tone modulation with 100% modulation is 1/3.

$$\underset{\text{of an AM system}}{\text{Figure of merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}\gamma\cdot\frac{1}{\gamma} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}$$

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 r.m.s. value is $\dfrac{1}{\sqrt{2}}$ and the mean-squared value $\overline{x^2}$ is 1/2. Further, for 100% modulation, $m = 1$

$$\therefore \qquad \left.\begin{array}{l}\text{Figure of merit for AM}\\ \text{with } m = 1 \text{ and single-tone}\\ \text{modulating signal}\end{array}\right\} = \frac{m^2\overline{x^2}}{1+m^2\overline{x^2}} = \frac{1/2}{3/2} = 1/3$$

9.6.3 Threshold Effect for AM with Envelope Detector

In case the SNR at the input to the envelope detector becomes very much less than unity, noise completely dominates over the signal and the behaviour 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. (9.37), we may write the expression for the envelope of the detector input as

$$R_y(t) = \sqrt{\{A_R[1+mx(t)]+n_i(t)\}^2 + n_q^2(t)}$$

$$= \sqrt{A_R^2(1+mx(t))^2 + n_i^2(t) + n_q^2(t) + 2A_R n_i(t)[1+mx(t)]}$$

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

$$R_y(t)=\sqrt{\left[n_i^2(t)+n_q^2(t)\right]\left[1+\frac{2A_R n_i(t)}{\left[n_i^2(t)+n_q^2(t)\right]}[1+mx(t)]\right]} \tag{9.45}$$

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. (9.45), we may make use of the approximation that when $\in$ << 1,

$$\sqrt{1+\in}\approx\left(1+\frac{\in}{2}\right)$$

$$\therefore \qquad R_y(t)=\sqrt{n_i^2(t)+n_q^2(t)}\left[1+\frac{A_R n_i(t)[1+mx(t)]}{n_i^2(t)+n_q^2(t)}\right] \tag{9.46}$$

(compare this with the *Ry*(*t*) given by Eq. (9.40) 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 9.4

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?

(University Question)

For an AM system with modulation index, m, the output SNR is given by

$$\left(\frac{S}{N}\right)_D=\frac{A_R^2 m^2\overline{x^2(t)}}{\overline{n_i^2(t)}} \qquad \text{(see Eq. 9.32)}$$

This is given to be 30 dB $= 10^3$; $m = 0.8$ and $\overline{x^2(t)}=\frac{1}{2}$ (single-tone)

$$\therefore \qquad \frac{A_R^2\times 0.64\times 1/2}{\overline{n_i^2(t)}}=1000. \quad \therefore \frac{A_R^2}{\overline{n_i^2(t)}}=\frac{1000}{0.32}.$$

But, we know that $\overline{n_i^2(t)}=\overline{n^2(t)}$ and that the carrier-to-noise ratio (CNR) is defined as

$$\text{CNR}=\frac{A_R^2/2}{\overline{n^2(t)}} \quad \therefore \text{CNR}=\frac{1000}{0.64}=1562.5$$

or, $$(\text{CNR})_{\text{dB}}=10\log_{10}1562.5=31.9\text{ dB}$$

Example 9.5

A message signal $x(t)$ of 5 kHz bandwidth and having an amplitude probability density as shown, 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/Hz(one-sided). The received signal is demodulated using an envelope detector.

(i) 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?

(ii) Assuming $(S/N)_{th}$ for envelope detection to be 10 dB, determine the threshold value of A_c.

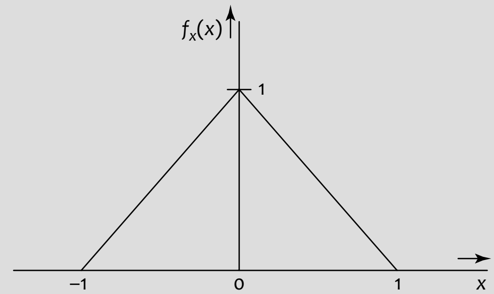


Fig. 9.5 *PDF of Signal for Example 9.5*

For AM systems, the destination SNR is given by

$$\left(\frac{S}{N}\right)_D = \left(\frac{m^2 \overline{x^2}}{1+m^2 \overline{x^2}}\right).\gamma. \quad \text{Here, } m = 0.8$$

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 $$\overline{x^2} = \int_{-\infty}^{\infty} x^2 f_X(x)dx \quad \therefore \quad \overline{x^2} = 2\int_0^1 x^2(1-x)dx = \frac{1}{6}$$

$$\therefore \qquad \left(\frac{m^2 \overline{x^2}}{1+m^2 \overline{x^2}}\right).\gamma \geq 40 \text{ dB}$$

Since 40 dB = A ratio of 10^4, substituting for m and $\overline{x^2}$, we get

$$\left(\frac{(0.8)^2\left(\frac{1}{6}\right)}{1+(0.8)^2\left(\frac{1}{6}\right)}\right).\gamma \geq 10^4 \quad \therefore \gamma \geq \frac{10^4 \times (1+0.1066)}{0.1066}$$

$$\therefore \gamma \geq 103808.63 \quad \therefore \text{ min.value of } \gamma = 103808.63$$

$\therefore$ 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

$$\overline{x(t)} = \int_{-1}^{1} xf_X(x)dx = \int_{-1}^{0} x(x+1)dx + \int_0^1 x(1-x)dx = 0$$

$$\therefore \qquad S_R = \frac{A_c^2}{2}\left[1+0.64\times\left(\frac{1}{6}\right)\right] = \frac{A_c^2}{2}(1.1066)$$

Since $\gamma_{\min} = 103808.63 = \dfrac{S_R}{\eta W} = \dfrac{A_c^2 \times 1.1066}{2 \times 2 \times 10^{-12} \times 5 \times 10^3}$

$$A_{c_{\min}}^2 = \frac{103808.63 \times 20 \times 10^{-9}}{1.1066} = 187617.26 \times 10^{-8}$$

$$\therefore \quad A_{c_{\min}} = \sqrt{187617.26 \times 10^{-8}} = 433 \times 10^{-4} = 43.3 \text{ mV}$$

$$\therefore \quad A_{c_{\min}} = 43.3 \text{ mV}$$

(ii) We are given that $\left(\dfrac{S}{N}\right)_{i\ \text{th}} = 10 \text{ dB} = 10(\text{ratio})$

$$\text{For AM, } \left(\frac{S}{N}\right)_i = \frac{S_R}{N_R} = \frac{\frac{A_c^2}{2}(1 + m^2\overline{x^2})}{(\eta / 2B_T)} = \frac{A_c^2(1 + m^2\overline{x^2})}{2\eta W}$$

$$\frac{A_c^2\left[1 + 0.64 \times \left(\frac{1}{6}\right)\right]}{2 \times 2 \times 10^{-12} \times 5 \times 10^3} = \frac{1.1066 A_c^2}{2 \times 10^{-8}}$$

$\therefore$ at threshold, $\dfrac{S_R}{N_R} = 10$

$$\therefore \quad \frac{1.1066 A_c^2}{2 \times 10^{-8}} = 10 \text{ or, } A_c^2 = \frac{2 \times 10^{-7}}{1.1066} = 0.18073 \times 10^{-8}$$

$$\therefore \quad A_{c_{\text{th}}} = 0.425 \times 10^{-4} \text{ volt or } 0.0425 \text{ mV}$$

Example 9.6

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.*

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

$$\frac{S_{R_1}}{\eta B_T} = \frac{S_{R_1}}{2\eta W} = 10 \quad \therefore \frac{S_{R_1}}{\eta W} = 20; \quad \text{or, } S_{R_1} = 20\eta W$$

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.

$$\text{Then, } \quad \left(\frac{S}{N}\right)_{D\,\text{AM}} = \left(\frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}}\right) \cdot \left(\frac{S_{R_2}}{\eta W}\right) = 10^4 = 10{,}000 \quad (\text{since } 40 \text{ dB} = 10^4)$$

Since $m = 1$ and it is tone modulation,

$$\frac{m^2\overline{x^2}}{1 + m^2\overline{x^2}} = \left(\frac{1.(1/2)}{1 + 1.(1/2)}\right) = \frac{1}{3}$$

$$\therefore \qquad \left(\frac{S}{N}\right)_{\substack{D\\ \text{AM}}} = \frac{1}{3}\cdot\left(\frac{S_{R_2}}{\eta W}\right) = 10{,}000 \quad S_{R_2} = 30{,}000\ \eta\text{W}$$

$$\therefore \qquad \left(\frac{S_{R_2}}{S_{R_1}}\right) = \left(\frac{30000\eta\text{W}}{20\eta\text{W}}\right) = 1500$$

$$\therefore \qquad \left.\begin{array}{l}\text{increase in transmitter power (in decibels)}\\ \text{required to have a destination SNR of 40 dB}\end{array}\right\} = 10\log_{10}1500 = 31.76\,\text{dB}$$

9.6.4 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 (9.14) and (9.23), 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 $\left(1/4\right)A_R^2\overline{x^2(t)}$ for SSB-SC and $\left(1/2\right)A_R^2\ \overline{x^2(t)}$ 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. (9.27), it is clear that the total sideband power in the AM is $\left(1/2\right)A_R^2 m^2\ \overline{x^2(t)}$, 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

$$(S/N)_{\substack{D\\ \text{AM}\\ m=1}} = \left(\frac{\overline{x^2(t)}}{1+\overline{x^2(t)}}\right)\cdot\gamma \tag{9.47}$$

Note: $\overline{x^2(t)}$ is always less than or equal to 1 since $x(t)$ is the normalized signal, normalized so that $|x(t)| \le 1$.

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. (9.47) 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.

(i) For Tone Modulation

$$\overline{x^2(t)} = 1/2$$

$$(S/N)_{\substack{D\\ \text{AM}\\ m=1}} = \left(\frac{1/2}{1+1/2}\right)\cdot\gamma = \gamma/3$$

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.

(ii) 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 $m = 1$,

$$(S/N)_{D \atop {\text{AM} \atop m^2\overline{x^2}(t)=1}} = \left(\frac{1}{1+1}\right)\cdot\gamma = \gamma/2$$

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 9.7

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?

(a)
$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{2\times10^{-12}\times15\times10^3} = \frac{10^8 S_R}{3}$$

Channel attenuation = 80 dB $\quad\therefore\ 10\log_{10}(S_T/S_R) = 80$

where, S_T is the average transmitted power and S_R is the average received power.

$$\therefore \qquad S_T = 10^8 S_R \quad \text{or} \quad S_R = 10^{-8} S_T.$$

$$\therefore \qquad \gamma = \frac{10^8\times10^{-8}\times S_T}{3} = \frac{S_T}{3} = 40\text{ dB} = 10^4$$

$\therefore \qquad S_T = 3\times10^4 = 30\,\text{kW}$. Since it is SSB-SC, $B_T = W = 15$ kHz.

(b) For DSB-SC, $\qquad (S/N)_D = \gamma = \dfrac{10^8 S_R}{3} \quad$ and $\quad S_R = 10^{-8} S_T$

$$\therefore \qquad \frac{S_T}{3} = 10^4 \quad \text{or} \quad S_T = 30\text{ kW}$$

Since both the sidebands are transmitted in DSB-SC, the bandwidth B_T required is $2W = 30$ kHz.

Example 9.8

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.

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{2\times10^{-11}\times15\times10^3} = \frac{S_R\times10^7}{3}$$

Average power in the message signal (before normalization) $= \int_{-5}^{5} x^2 \frac{1}{10}dx = \frac{25}{3}$

$\overline{x^2(t)}$ = Average power in the message signal after normalization so that $|x(t)| \le 1$

$$= \frac{25}{3}\times\frac{1}{5^2} = \frac{1}{3}$$

$$\therefore \qquad m^2\,\overline{x^2(t)} = 0.36\times\frac{1}{3} = 0.12 \quad \therefore \quad \frac{m^2\,\overline{x^2(t)}}{1+m^2\,\overline{x^2(t)}} = \frac{0.12}{1.12}$$

Since channel attenuation = 60 dB $= 10\log_{10}(S_T/S_R)$

$$\therefore \qquad \log_{10}\left(\frac{S_T}{S_R}\right) = 6 \quad \therefore \quad S_R = 10^{-6}S_T$$

$$\therefore \qquad \left[\frac{m^2\,\overline{x^2(t)}}{1+m^2\,\overline{x^2(t)}}\right]\cdot\gamma = \left(\frac{S}{N}\right)_D = \frac{10^{-6}S_T.10^7}{3}\cdot\frac{0.12}{1.12} = \frac{0.04S_T}{1.12}$$

But $\left(\frac{S}{N}\right)_{D_{AM}}$ has to be 40 dB ($= 10^4$)

$$\therefore \qquad \frac{0.04S_T}{1.12} = 10^4 \quad \text{or} \quad S_T = \frac{10^4\times1.12}{0.04} = 280\text{ kW};\ B_T = 2\times15\text{ kHz} = 30\text{ kHz}.$$

Example 9.9

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)| ≤ 1. The modulated signal is transmitted over a channel whose attenuation is 60 dB and has an additive white noise with two-sided power spectral density of 10^{-12} W/Hz. Determine the pre-detection and post-detection SNRs at the receiver.

P_c = Unmodulated carrier power $= 20\times10^3$ W

$m = 0.8$

$\overline{x^2(t)} = 0.78$ W when $|x(t)| \le 1$

$\therefore$ average total power of the modulated signal $= P_T$

where, $$P_T = P_c\left[1+m^2\overline{x^2}\right] = 20\times10^3\left[1+0.64\times0.78\right] = 30\times10^3\text{ W}$$

$\therefore$ S_T = Transmitted power $= 30\times10^3$ W

Attenuation of the channel = 60 dB (= a ratio of 10^{-6})

$$\therefore \qquad \frac{S_T}{S_R} = 10^6 = \frac{30\times10^3}{S_R} \quad \therefore \quad S_R = 30\times10^{-3}\text{ W}$$

$$\eta B_T = \left(2\times10^{-12}\right)\times\left(2\times15\times10^3\right) = 6\times10^{-8}\text{ W}$$

$\therefore$ noise power at the input to the detector $= N_R = 6 \times 10^{-8}$ W

$$\therefore \text{ pre-detection SNR} = \left(\frac{S}{N}\right)_R = \frac{S_R}{\eta B_T} = \frac{30\times10^{-3}}{6\times10^{-8}} = 5\times10^5 = \frac{1}{2}\times10^6$$

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{\eta(B_T/2)} = 10\times10^5$$

$$\frac{m^2\,\overline{x^2(t)}}{1+m^2\,\overline{x^2(t)}} = \frac{0.64\times0.78}{1+0.64\times0.78} = \frac{0.5}{1+0.5} = \frac{1}{2}\times\frac{2}{3} = \frac{1}{3}$$

$$\therefore \text{ post-detection SNR} = \left[\frac{m^2\,\overline{x^2(t)}}{1+m^2\,\overline{x^2(t)}}\right]\gamma = \frac{1}{3}\times10^6$$

Example 9.10

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 2-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.

For AM,
$$(S/N)_D = \left(\frac{m^2\overline{x^2}}{1+m^2\overline{x^2}}\right)\gamma = \frac{1.\gamma}{1+1} = 0.5\gamma$$

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{10^{-15}\times20\times10^3} = \frac{S_R}{2\times10^{-11}}$$

$$\therefore \quad \left(\frac{S}{N}\right)_D = \frac{0.5S_R}{2\times10^{-11}} = 0.25S_R\times10^{11} = 10^4$$

$$\therefore \quad S_R = 4\times10^{-7}\,\text{W}. \quad L = 100\text{ dB} = 10^{10}$$

$$\therefore \quad S_T = 4\times10^{-7}\times10^{10} = 4\text{ kW}$$

Example 9.11

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 2-sided PSD $\eta/2 = 0.5\times10^{-15}$ W/Hz and a transmission loss of 100 dB. If a destination SNR of 40 dB is required, determine the average transmitter power required.

$$\left(\frac{S}{N}\right)_{D_{\text{SSB}}} = \gamma = \frac{S_R}{\eta W} = 10^4$$

$$\therefore \quad S_R = 10^{-15}\times5\times10^3\times10^4 = 10^{-8}\,\text{W. Also, 100 dB = a ratio of } 10^{10}$$

$$\therefore \quad S_R = 10^{-8}\,\text{W. But } \frac{S_T}{S_R} = 10^{10} \quad \therefore \quad S_T = 10^{-8}\times10^{10} = 10^2 = 100\text{ W}$$

9.7 NOISE PERFORMANCE OF FREQUENCY MODULATED SYSTEMS

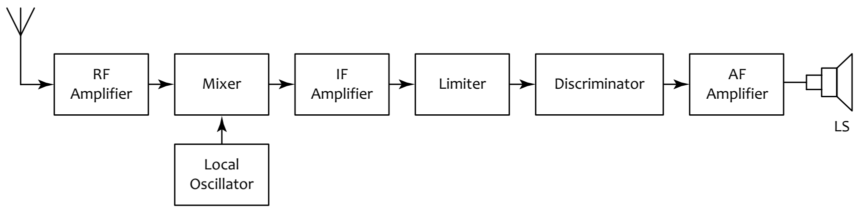


Fig. 9.6 *Block diagram of an FM broadcast superheterodyne receiver*

The block diagram of an FM broadcast superheterodyne receiver is shown in Fig. 9.6. For the purpose noise performance evaluation, we model the receiver as shown in Fig. 9.7.

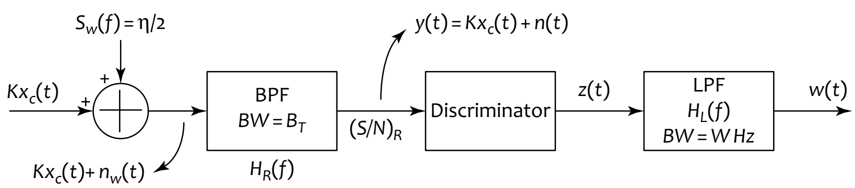


Fig. 9.7 *Receiver model for noise-performance evaluation*

Additive noise of the channel is modeled as zero-mean white Gaussian noise of a two-sided power spectral density $\eta/2$. K represents the channel attenuation. The modulated signal in this case, is given by

$$x_c(t) = A_c \cos[\omega_c t + \phi(t)], \quad f_c = f_{i.f} \text{ of the receiver} \tag{9.48}$$

where,

$$\phi(t) = 2\pi k_f \int_0^t x(\alpha) d\alpha \tag{9.49}$$

Note: In Eq. (9.49), 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. 9.8, is used to represent the combined effect of the RF and IF amplifiers.

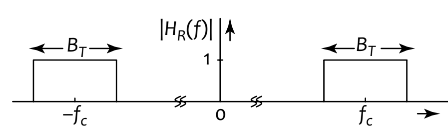


Fig. 9.8 *Response characteristic of the BPF*

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 bandpass noise centered on f_c and obtained by filtering the white noise using the BPF of bandwidth B_T, with centre frequency $f_c = f_{i.f}$, 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 bandpass 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. And the additive noise does not *directly* affect the frequency deviation of the incoming FM signal. It affects it only indirectly, as we will be seeing presently. 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

$$\overline{n^2(t)} = \frac{\eta}{2} \times 2B_T = \eta B_T \triangleq N_R \tag{9.50}$$

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

$$S_R = \frac{(KA_c)^2}{2} = \frac{A_R^2}{2} \tag{9.51}$$

$\therefore$ the pre-detection SNR is given by

$$\left(\frac{S}{N}\right)_R = \frac{S_R}{N_R} = \frac{A_R^2}{2} \cdot \frac{1}{\eta B_T} = \frac{A_R^2}{2\eta B_T} \tag{9.52}$$

As mentioned earlier, $n(t)$ is bandpass noise centered on f_c and we may represent it by its inphase and quadrature components as

$$n(t) = n_i(t)\cos\omega_c t - n_q(t)\sin\omega_c t \tag{9.53}$$

Alternatively, we may use the envelope and phase-angle representation (See Section 2.8, Eq. 2.164) and write as

$$n(t) = R_n(t)\cos\left[\omega_c t + \phi_n(t)\right] \tag{9.54}$$

where, $R_n(t)$, the envelope is related to $n_i(t)$ and $n_q(t)$ by

$$R_n(t) = \sqrt{n_i^2(t) + n_q^2(t)} \tag{9.55}$$

and is Rayleigh distributed. The phase angle, $\phi_n(t)$ is given by

$$\phi_n(t) = \tan^{-1}\left[\frac{n_q(t)}{n_i(t)}\right] \tag{9.56}$$

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

$$\begin{aligned} y(t) &= A_R\cos\left[\omega_c t + \phi(t)\right] + n(t) \\ &= A_R\cos\left[\omega_c t + \phi(t)\right] + R_n(t)\cos\left[\omega_c t + \phi_n(t)\right] \end{aligned} \tag{9.57}$$

We shall make use of Eq. (9.57) 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., $$\left(\frac{S}{N}\right)_R >> 1 \tag{9.58}$$

If under this assumption, we draw the phasor diagram for Eq. (9.57), it will appear as shown in Fig. 9.9.

For the bandpass signal $y(t)$, if $R_y(t)$ is the envelope and $\phi_y(t)$, the phase angle, we may write

$$y(t) = R_y(t)\cos\left[\omega_c t + \phi_y(t)\right] \tag{9.59}$$

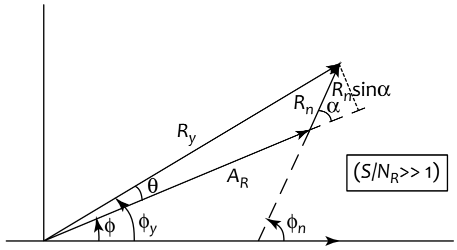


Fig. 9.9 *Phasor diagram of Eq. (9.54) when $(S/N)_R >> 1$*

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

$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\phi_y(t)\right] \tag{9.60}$$

So, let

$$z(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\phi_y(t)\right] \tag{9.61}$$

The phasor diagram of Fig. 9.7 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 bandpass 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,

$$P\left[R_n(t) << A_R\right] \text{ is almost equal to 1} \tag{9.62}$$

But, from Fig. 9.7,

$$\sin\theta(t) = \left[R_n(t)\sin\alpha(t)\right] / R_y(t) \tag{9.63}$$

However, from Eq. (9.62), it follows that the following small-angle approximation can be made so that

$$\sin\theta(t) \cong \theta(t)$$

and hence, Eq. (9.63) may be re-written as

$$\theta(t) = \frac{R_n(t)\sin\alpha(t)}{R_y(t)} \tag{9.64}$$

Thus, since

$$\phi_y(t) = \phi(t) + \theta(t) \tag{9.65}$$

We have

$$\phi_y(t) = \phi(t) + \frac{R_n(t)\sin\alpha(t)}{R_y(t)} \tag{9.66}$$

But because of Eq. (9.62), we may make the following approximation

$$R_y(t) \cong A_R \tag{9.67}$$

Hence, from Eqs (9.61) and (9.65) we have

$z(t)$ = Discriminator output signal

$$= \frac{1}{2\pi}\frac{d}{dt}\left[\phi_y(t)\right] = \frac{1}{2\pi}\frac{d}{dt}\phi(t) + \frac{1}{2\pi}\frac{d}{dt}\theta(t) = k_f x(t) + n_d(t) \tag{9.68}$$

Since $\phi(t)$ is the phase angle caused due to frequency modulating the carrier by the message signal $x(t)$, from Eqs (9.48) and (9.49) the first term in Eq. (9.68) 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. (9.68) 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 lowpass filter and reaches the destination, we have to examine the spectrum of the noise term in Eq. (9.68). For this purpose, let us re-write it as follows:

$$\frac{1}{2\pi}\frac{d}{dt}\theta(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\frac{R_n(t)\sin\alpha(t)}{A_R}\right] \text{ from Eq. (9.64) and (9.67)} \tag{9.69}$$

From the phasor diagram of Fig. 9.7, we find that

$$\alpha(t) = \phi_n(t) - \phi(t) \tag{9.70}$$

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 bandpass 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 $\left[\phi_n(t) - \phi(t)\right]$, 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. 9.10 (since $\phi(t) = 0$ when there is no modulation).

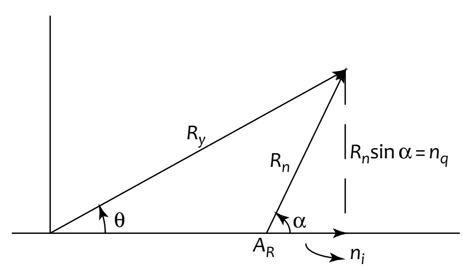


Fig. 9.10 *Phasor diagram with no modulation $(S/N)_R >> 1$*

Since $\phi(t) = 0$, $\alpha(t) = \phi_n(t)$ and so

$$R_n(t)\sin\alpha t = R_n(t)\sin\phi_n(t) = n_q(t) \tag{9.71}$$

Hence, Eq. (9.69) may be re-written as

$$\frac{1}{2\pi}\frac{d}{dt}\theta(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\frac{R_n(t)\cdot\sin\alpha t}{A_R}\right]$$

$$\therefore \quad n_d(t) = \frac{1}{A_R}\frac{1}{2\pi}\frac{d}{dt}\left[n_q(t)\right] \tag{9.72}$$

Therefore, to determine how much power of this post-detection noise goes past the lowpass filter with a cutoff 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 lowpass equivalent of the bandpass 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

$$S_n(f) = S_{n_w}(f)\left|H_R(f)\right|^2 = \frac{\eta}{2}\left|H_R(f)\right|^2 \tag{9.73}$$

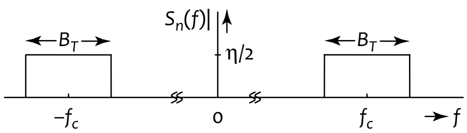


Fig. 9.11 *Power spectral density of n(t)*

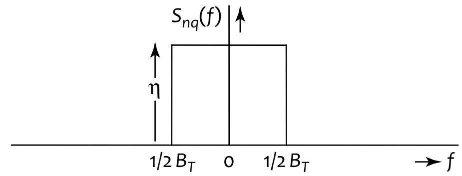


Fig. 9.12 *Power spectral density of $n_q(t)$, lowpass equivalent of n(t)*

The PSD of the bandpass noise, $n(t)$, is shown in Fig. 9.11, and that of its lowpass equivalent, $n_q(t)$, is shown in Fig. 9.12.

$$\therefore \qquad S_{nq}(f) = \eta\Pi(f/B_T) \tag{9.74}$$

The power spectrum of $\dfrac{1}{A_R}\cdot\dfrac{1}{2\pi}n_q(t)$ is then given by

$$\frac{1}{(2\pi)^2}\cdot\frac{1}{A_R^2}\cdot S_{nq}(f) \tag{9.75}$$

To find the power spectrum of the post-detection noise $n_d(t)$, in view of Eq. (9.72), we proceed as in Fig. 9.13.

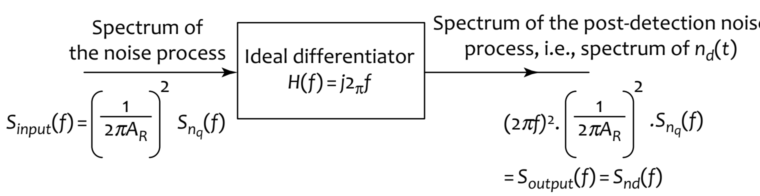


Fig. 9.13 *Deriving the spectrum of post-detection noise*

Substituting for $S_{nq}(f)$ in the expression for the $S_{nd}(f)$ and simplifying, we get

$$S_{n_d}(f) = \left(\frac{\eta f^2}{A_R^2}\right)\Pi\left(f/B_T\right)$$

$$= \left(\frac{\eta f^2}{2S_R}\right)\Pi\left(f/B_T\right) \tag{9.76}$$

A sketch of the post-detection noise spectrum is given in Fig. 9.14. 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 lowpass filter having a cutoff frequency of W Hz.

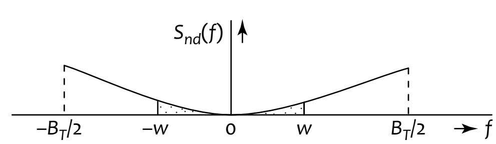


Fig. 9.14 *Power spectral density of post-detection noise*

The average power of the noise at the output of the lowpass filter = N_D = Destination noise power = Area of the shaded region

$$= \int_{-W}^{W} \left(\frac{\eta f^2}{2S_R}\right) \Pi\left(f/B_T\right) df$$

$$= \int_{-W}^{W} \frac{\eta f^2}{2S_R} df = \frac{\eta W^3}{3S_R}$$

$$\therefore \qquad N_D = \frac{\eta W^3}{3S_R} \tag{9.77}$$

The message signal component at the output of the discriminator has been found [refer Eq. (9.68)] to be $k_f\, x(t)$. Since this has a bandwidth of W, all of it passes through the lowpass filter. Hence the destination signal power is given by

$$S_D = k_f^2\, \overline{x^2(t)} \tag{9.78}$$

$\therefore$ 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. (9.49)]. 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.

$$\boxed{\therefore \left(\frac{S}{N}\right)_{D \atop \text{FM}} = 3D^2\, \overline{x^2(t)}\, \gamma} \tag{9.79}$$

Hence, from Eqs (9.79) and (9.7), the figure of merit for FM systems may be written down as:

$$\boxed{\underset{\text{(FM)}}{\text{Figure of merit}} = \frac{(S/N)_D}{(S/N)_C} = \frac{3D^2\, \overline{x^2}\gamma}{\gamma} = 3D^2\, \overline{x^2}} \tag{9.79a}$$

Example 9.12

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.

$D =$ 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. (9.79a) gives

$$\text{Figure of merit} = \frac{3}{2}\left(\frac{k_f A_m}{f_m}\right) = \frac{3}{2}\beta_f^2 \quad \text{(see Eq. 5.18)}$$

Remarks

1. Although derived under certain assumptions, the result represented by Eq. (9.79) 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. (9.79) 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.
2. 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 doesn't, 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. (9.79), will no longer be valid.
3. 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 9.13

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 Figure of merit for PM $= \frac{1}{2}m_f^2$ where, m_f is the modulation index for FM.

We know that in the case of phase modulation,

$$\phi(t) = k_p x(t) = k_p E_m \cos\omega_m t$$

$\therefore$ 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

$$(\text{FOM})_{\text{PM}} = k_p^2 \overline{x^2(t)} = k_p^2 . \frac{E_m^2}{2}; \quad \text{But } k_p E_m = m_f$$

$$\therefore \quad (\text{FOM})_{\text{PM}} = \frac{1}{2} m_f^2$$

Example 9.14

Show that narrow-band FM does not offer any better destination signal-to-noise ratio than AM.

For AM: $\left(\frac{S}{N}\right)_D = \left(\frac{m^2 \overline{x^2(t)}}{1+m^2\overline{x^2(t)}}\right).\gamma$

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 $$\left(\frac{S}{N}\right)_D = \frac{1}{2}.\gamma$$

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,

$$\left(\frac{S}{N}\right)_D = 3\beta_f^2\gamma$$

$\therefore$ if this is to be better than the $\left(\frac{S}{N}\right)_D$ for AM,

$$3\beta_f^2\gamma = \frac{1}{2}\gamma \quad \text{or} \quad \beta_f^2 = \frac{1}{6} \quad \therefore \quad \beta_f > 0.408$$

i.e., for FM to be better than AM, β_f should be greater than 0.408. But, for NBFM, $\beta_f < 0.2$ $\therefore$NBFM is no better than AM.

9.8 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. 9.15.

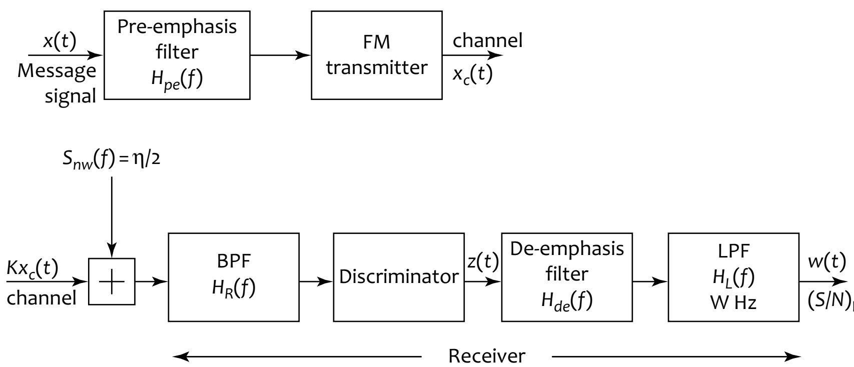


Fig. 9.15 *Pre-emphasis and de-emphasis in an FM system*

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, atleast over the message bandwidth, W.

$$H_{de}(f) = \frac{1}{H_{pe}(f)}; \quad |f| \le W \tag{9.80}$$

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.

$$H_{pe}(f) = \left(1 + j\frac{f}{f_0}\right) \tag{9.81}$$

In this, f_0 is a fixed frequency. The corresponding $H_{de}(f)$ is

$$H_{de}(f) = \frac{1}{\left(1 + j\dfrac{f}{f_0}\right)} \tag{9.82}$$

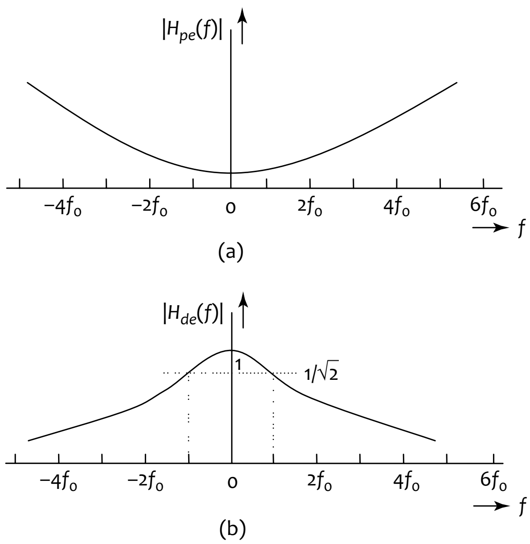


Fig. 9.16 *Magnitude responses:* **(a)** *Pre-emphasis, and* **(b)** *De-emphasis filter (linear frequency scale)*

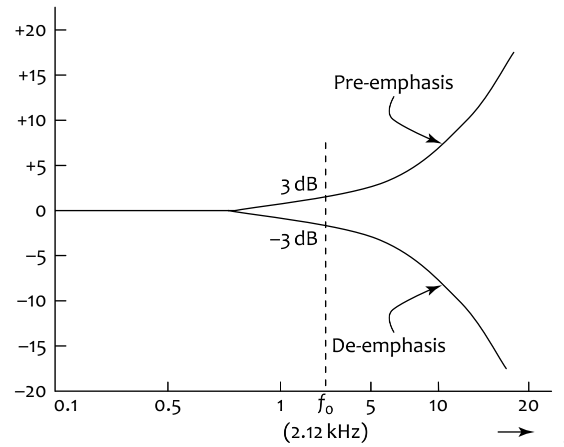


Fig. 9.17 *Magnitude responses (in decibels) of pre-emphasis and de-emphasis filters plotted using logarithmic scale for frequency*

Figures 9.16 and 9.17 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 *R-C* filter shown in Fig. 9.18(a) and $H_{de}(f)$, which is *essentially* the response of an integrator, can be closely realized by the simple *R-C* filter shown in Fig. 9.18(b).

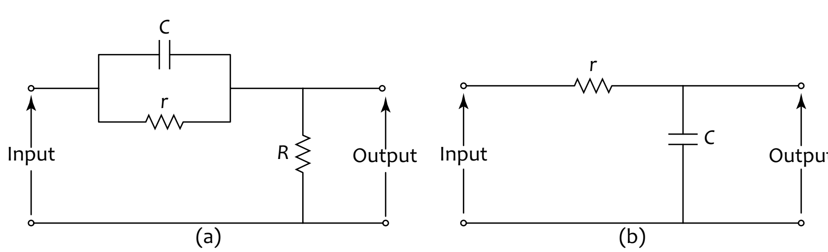


Fig. 9.18 **(a)** *Pre-emphasis filter* **(b)** *De-emphasis filter*

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 components 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.

9.8.1 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.

$$I \triangleq \left(\frac{\text{Noise power output at the destination without de-emphasis}}{\text{Noise power output at the destination with de-emphasis}} \right) \tag{9.83}$$

Earlier, we found (see Eq. 9.76) 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 lowpass 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

$$H_c(f) = H_{de}(f) \cdot H_L(f) \tag{9.84}$$

and,
$$H_L(f) = \Pi(f/2W) \tag{9.85}$$

since the baseband filter has been modeled as an ideal LPF with a cutoff frequency of W Hz.

Hence, noise PSD at the destination is given by

$$S_D(f) = S_{nd}(f) \cdot |H(f)|^2 \tag{9.86}$$

$\therefore$ average noise power at the destination, i.e., at the output of the receiver *with* pre-emphasis and de-emphasis is given by

$$N_D = \int_{-W}^{W} S_D(f)\,df = \int_{-W}^{W} S_{nd}(f)|H_c(f)|^2 df$$

$$= \int_{-W}^{W} \frac{\eta f^2}{2S_R} \cdot \Pi(f/B_T)\Pi(f/2W)|H_{de}(f)|^2\,df$$

$$= \int_{-W}^{W} \frac{\eta f^2}{2S_R} |H_{de}(f)|^2\,df = \frac{\eta}{A_R^2} \int_{-W}^{W} f^2\,|H_{de}(f)|^2\,df \tag{9.87}$$

Average noise power at the output of the receiver *without* the pre-emphasis and de-emphasis, is given by

$$N_0 = \int_{-W}^{W} S_{nd}(f)\,df = \int_{-W}^{W} \frac{\eta f^2}{2S_R} df$$

$$= \frac{\eta}{A_R^2} \int_{-W}^{W} f^2 df = \frac{2\eta W^3}{3A_R^2} \tag{9.88}$$

Note: The limits for the integral are $-W$ and $+W$ because of the baseband filter $H_L(f)$.

Substituting these in Eq. (9.83), we have

$$\boxed{I = \frac{N_0}{N_D} = \frac{2W^3}{3\int_{-W}^{W} f^2 |H_{de}(f)|^2 df}} \tag{9.89}$$

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 9.15

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.

From Eq. (9.82), we have

$$|H_{de}(f)|^2 = \left(\frac{f_0^2}{f_0^2 + f^2}\right)$$

∴ from Eq. (9.89), 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,

$I \approx 22$ ∴ improvement of $(S/N)_D$ in dB $= 10\log_{10} 22 \approx 13$ dB.

Example 9.16

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.

For an FM system, $\left(\frac{S}{N}\right)_D = 3\beta^2 \overline{x^2(t)}\,\gamma$

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 an AM system, $\left(\frac{S}{N}\right)_D = \left[\frac{m^2 \overline{x^2}}{1 + m^2 \overline{x^2}}\right]\gamma$

$$\therefore \qquad F = \frac{(m^2/2)}{1 + \frac{1}{2}m^2} = \left[\frac{m^2}{2 + m^2}\right]$$

9.9 THRESHOLD EFFECT IN FM

For wideband FM, we have seen that (refer to Eq. 9.79) the destination SNR is given by

$$\left(\frac{S}{N}\right)_{\substack{D\\ \text{WBFM}}} = 3D^2\overline{x^2}\ \gamma \tag{9.90}$$

As has already been observed in Remark 2 under that equation, it then follows that just by increasing D, the deviation ratio, 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 $\left(A_c^2/2\right)$ constant, the destination signal-to-noise ratio at the output of the receiver will, ofcourse, 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 $\left(A_c^2/2\right)$, the average transmitted power is held constant, the signal power entering the receiver, given by $\left(K^2A_c^2/2\right) = \left(A_R^2/2\right)$, is also constant, while the average noise power entering the receiver is increasing. 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 is so low that the basis on which we had derived Eq. (9.79), viz; that the $(S/N)_R$ is quite high, is violated, making the application of Eq. (9.79) no longer appropriate.

This fact, that Eq. (9.79) 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

$$x(t) = \cos\omega_m t \tag{9.91}$$

Since $x(t)$ has been normalized in such a way that $|x(t)| \le 1$,
$\therefore$ from Eq. (9.91), we have

$$\overline{x^2(t)} = \frac{1}{2} \tag{9.92}$$

Further, since $x(t)$ is single-tone, and has been normalized,

$$D = \beta \tag{9.93}$$

From the above, Eq. (9.90) may, therefore, be written as

$$\left(\frac{S}{N}\right)_{\substack{D\\ \text{WBFM}}} = \frac{3}{2}\beta^2\ \gamma \tag{9.94}$$

or,

$$\left(\frac{S}{N}\right)_{D;\ \text{dB}} = 10\log_{10}\left(\frac{3}{2}\beta^2\right) + 10\log_{10}\ \gamma$$

$$\therefore \quad \boxed{\left(\frac{S}{N}\right)_{D;\ \text{dB}} = 10\log_{10}\left(\frac{3}{2}\beta^2\right) + \gamma_{\text{dB}}} \tag{9.95}$$

Note: Relationship between ρ and γ

$$\rho \triangleq \frac{S_R}{\eta B_T} \quad \text{whereas} \quad \gamma \triangleq \frac{S_R}{\eta W}$$

$$\therefore \qquad \gamma = \rho\left(\frac{B_T}{W}\right) \quad \text{but} \quad B_T = 2W(1+\beta)$$

$$\therefore \qquad \gamma = 2\rho(1+\beta)$$

$$\therefore \qquad 10\log_{10}\gamma = 10\log_{10}2 + 10\log_{10}\rho + 10\log_{10}(1+\beta)$$

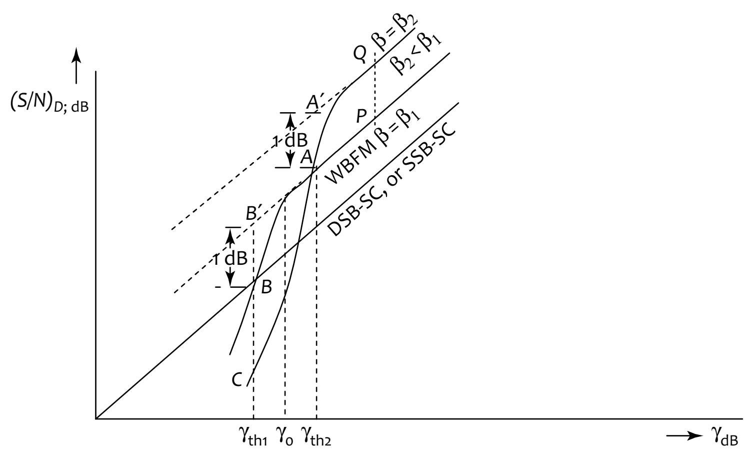


Fig. 9.19 *Plots of output SNR to input SNR for WBFM and DSB-SC or SSB-SC*

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. 9.19. For WBFM too, as per Eq. (9.95), 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. (9.79) 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. (9.79) and therefore Eq. (9.95) 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 the figure, this happens at the 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 is happening at the 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 the 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; but 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 the points *B* and *C*. Another interesting thing that we observe 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 above Fig. 9.19).

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.

9.9.1 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. (9.54)

$\therefore$
$$n(t) = R_n(t)\cos\left[\omega_c t + \phi_n(t)\right] \tag{9.96}$$

where,
$$\phi_n = \tan^{-1}\left(\frac{n_q(t)}{n_i(t)}\right) \tag{9.97}$$

$\therefore$ the input to the discriminator is (see Fig. 9.5)

$$y(t) = K.A_c\cos\omega_c t + n(t) = A_R\cos\omega_c t + n(t) \tag{9.98}$$

Combining Eqs (9.96) and (9.98), we may write $y(t)$ as

$$y(t) = R_y(t)\cos\left[\omega_c t + \xi(t)\right] \tag{9.99}$$

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. 9.20.

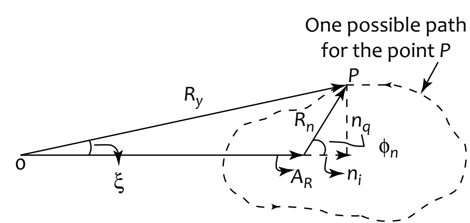


Fig. 9.20 *Phasor diagram when $(S/N)_R >> 1$*

Since $\left(\frac{S}{N}\right)_R >> 1$, $R_n << A_R$ with a high probability and so

$$\xi \approx \tan^{-1}\left(\frac{n_q}{A_R}\right) \tag{9.100}$$

Further, since $\left(\frac{n_q}{A_R}\right) << 1$ for most of the time, we may write

$$\xi \approx \left(\frac{n_q}{A_R}\right) \tag{9.101}$$

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, the 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, the point P, while moving along such random paths, will be close 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 the point P may enclose the point O as shown in Fig. 9.21. 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. 9.22(a) the discriminator output $z(t)$ which is given by

$$z(t) = \frac{d}{dt}\xi(t) \tag{9.102}$$

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. 9.22 (b)]. However, since $A_R >> R_n$ for most of the time, this phenomenon occurs only 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.

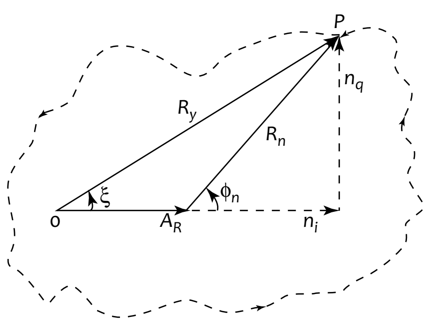


Fig. 9.21 *Phasor diagram when $(S/N)_R$ << 1 showing one possible path traversed by P*

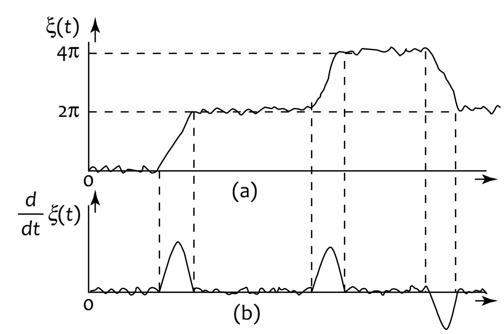


Fig. 9.22 **(a)** *$\xi(t)$ vs t* **(b)** *$\xi(t)$ vs t showing spikes in the discriminator output, z(t)*

Example 9.17

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 be 40 dB. If the signal attenuation during its passage through the channel is 40 dB, find the minimum transmitter power required.

As stated in Section 9.9, the value of β to be used may be restricted either by power considerations, or bandwidth considerations. We shall first examine this.

$$\gamma = \frac{S_R}{\eta W} \quad \text{and} \quad \rho = \frac{S_R}{\eta B_T} \quad \therefore \quad \gamma = \rho.\left(\frac{B_T}{W}\right) = \rho\frac{2W(\beta+1)}{W} = 2\rho(\beta+1)$$

$$\therefore \qquad \gamma_{th} = 2\rho_{th}(\beta+1)$$

But

$$\rho_{th} = 10\,\text{dB} = 10 \quad \therefore \quad \gamma_{th} = 20(\beta+1)$$

But

$$\left(\frac{S}{N}\right)_D = 3\beta^2 \overline{x^2}\gamma \quad \therefore \quad \left(\frac{S}{N}\right)_{D,\text{th}} = 3\beta^2 \overline{x^2}\gamma_{\text{th}}$$

$$= 3\beta^2 \overline{x^2}\, 20(\beta+1)$$

$$\therefore \quad \left(\frac{S}{N}\right)_{D,\text{th}} = 60\beta^2(1+\beta)\overline{x^2}$$

$$\therefore \quad 10^4 = 60\beta^2(1+\beta)\cdot\frac{1}{2} \quad \therefore \quad \beta^2(1+\beta) = \frac{10^4}{30} = 333$$

$\therefore \ \beta^3 + \beta^2 - 333 = 0 \qquad \therefore \ \beta \approx 6.6$ from power considerations.

Now, if we look at it from bandwidth point of view:
$W = 15$ kHz, $B_T = 2(\beta+1)W = 30\times10^3\times(\beta+1) = 200$ kHz

$\therefore \qquad \beta + 1 = 6.6$ 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

$$\gamma = \frac{S_R}{\eta W} = \frac{S_R}{10^{-8}\times10^3\times15} = \frac{S_R}{15\times10^{-5}} = 212.76$$

$$\therefore \quad S_R = 3191.5\times10^{-5} = 0.031915; \quad \text{but} \quad \left(\frac{S_T}{S_R}\right) = 10^4$$

$$\therefore \quad S_T = 3191.5\times10^{-5}\times10^4 = 3191.5\times10^{-1}$$

$$= 319.15 \text{ W}$$

Example 9.18

A message signal normalized so that $|x(t)| \le 1$ and having an average power of 1 watt and a bandwidth of 15 kHz, is to be transmitted using WBFM with $\beta = 5$, over a channel with additive noise of 2-sided PSD $= \eta/2 = 0.5\times10^{-13}$ W/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.

$$\left(\frac{S}{N}\right)_D = 3\beta^2\overline{x^2}\gamma = 3\times25\times1\times\frac{S_R}{15\times10^3\times10^{-13}} = 10^4$$

$$\therefore \quad S_R = 2\times10^{-7}\,\text{W}$$

But $$\left(\frac{S_T}{S_R}\right) = 10^{10} \quad \therefore S_T = 2\times10^{-7}\times10^{10} = 2 \text{ kW}$$

$$\therefore \quad \frac{S_R}{\eta W} = \frac{10^{-7}\times 2}{10^{-13}\times15\times10^3} = \frac{400}{3} = 21.3 \text{ dB} = \gamma$$

$$\left(\frac{S}{N}\right)_{D,\text{th}} = 60\beta^2(\beta+1)\overline{x^2} = 60\times25(5+1)\times1 = 60\times150 = 9000$$

$$= 3\times25\times\gamma_{\text{th}}\times1$$

$$\therefore \quad \gamma_{\text{th}} = \frac{9000}{3\times25} = \frac{3000}{25} = 120 \quad \text{and} \quad 10\log_{10}120 = 20.8 \text{ dB}$$

$\therefore$ the system is operating above the threshold since $\gamma > \gamma_{\text{th}}$

9.9.2 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.

9.9.3 Threshold Extension Techniques

Basically there are two 'threshold extension techniques' available. These are

(i) The 'Frequency Modulation FeedBack' (FMFB) technique
(ii) The 'Phase Lock Loop' or PLL technique

Actually, these two techniques work on similar lines and are equally effective in lowering the threshold. However, the PLL method is simple 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.

(i) The FMFB Technique As we have already discussed, the onset of threshold condition 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,

$$B = \text{Bandwidth of the noise at the input of the discriminator}$$

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 not employing any threshold extension techniques, the bandwidth of the noise at the input to the receiver's discriminator is B_T.

Hence, 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.

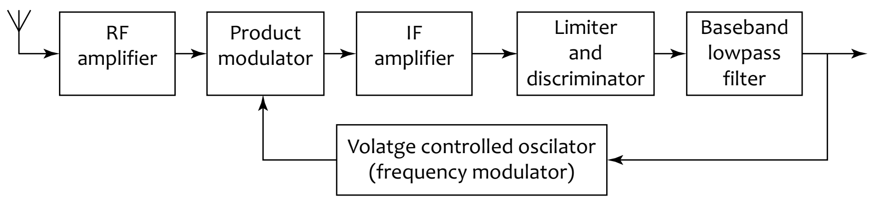


Fig. 9.23 *FMFB method for threshold extension*

Referring to Fig. 9.23, in this method for threshold extension, the normal local oscillator of the receiver is replaced by a voltage controlled oscillator, 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 a 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.

(ii) PLL Method of Threshold Extension Earlier, in Section 5.8, we had seen how a PLL could be used as a FM demodulator.

For convenience, Fig. 5.36 showing the linearized equivalent circuit of the PLL has been reproduced here as Fig. 9.24. This circuit was analyzed in Section 5.8. It was shown there that

$$\Phi_e(f) = \frac{\Phi(f)}{1 + \left(\frac{k_v}{jf}\right) H(f)}, \qquad (9.103)$$

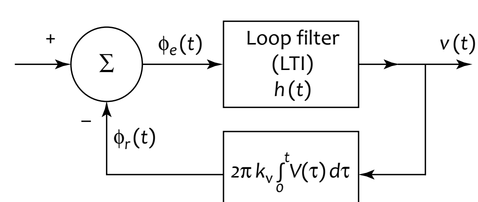


Fig. 9.24 *Linearized Equivalent circuit of the PLL*

and that

$$V(f) = \frac{H(f)\cdot\Phi(f)}{1+\left(\frac{k_v}{jf}\right)H(f)} \tag{9.104}$$

As pointed out there, if the gain of the loop filter is high enough, so that

$$\left(\frac{k_v}{jf}\right)H(f) >> 1 \text{ for } |f| < W$$

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 $\phi_e(f)$ 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 bandlimited 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 bandlimited 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. 9.25, is used as the loop filter. Table 9.1 gives a comparison of AM and FM.

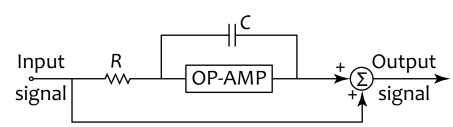


Fig. 9.25 *Loop filter for a second-order PLL*

9.9.4 Comparison of AM and FM

Table 9.1 *Comparison of AM and FM*

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 a 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.

(continued)

Table 9.1 (continued)

S.No.	*Amplitude Modulation*	*Frequency Modulation*
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 have 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

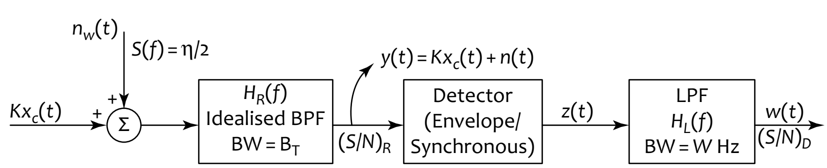
1. For a baseband transmission system, i.e., when the baseband or message signal is transmitted without any modulation,

$$\left(\frac{S}{N}\right)_D = \frac{S_R}{\eta W} \triangleq \gamma, \text{ where } S_R = \text{Received signal power}$$

$$\eta/2 = \text{PSD of white noise on the channel}$$

$$W = \text{Bandwidth of baseband signal}$$

2. Model used for linear modulation systems

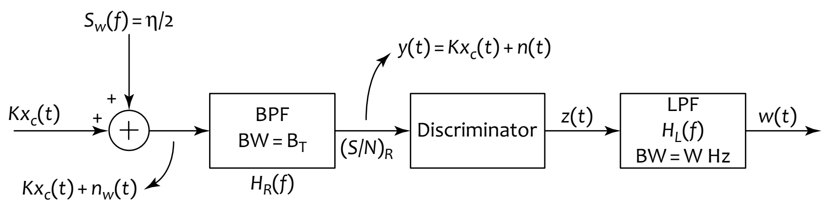


3. Pre-detection $SNR = \left(\frac{S}{N}\right)_R = \gamma\left(\frac{W}{B_T}\right)$, where B_T is the bandwidth of the transmitted signal.

4. (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.

5. 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.
6. Model used for FM systems:



7. The PSD of the noise at the output of the discriminator in the FM receiver varies as the square of the frequency.
8. $\underset{\text{FM}}{\left(\dfrac{S}{N}\right)_D} = 3D^2\overline{x^2(t)}\gamma$ provided $\left(\dfrac{S}{N}\right)_R >> 1.$
9. There is a power-bandwidth trade-off possible in WBFM as shown by $\left(\dfrac{S}{N}\right)_D$ for FM.
10. (a) **Pre-emphasis** consists of boosting up the higher message frequencies *before modulation* at the FM *transmitter.*
    (b) **De-emphasis** consists of de-emphasising 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.
11. 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.
12. For WBFM receivers, the threshold value of the input SNR is approximately 10 dB.
13. 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.
14. Figure of Merit (FOM) of a communication system is defined as

$$\text{FOM} \triangleq \frac{\text{(SNR at the destination)}}{\text{(SNR at the input to the detector but with the noise considered only over message bandwidth)}}$$

$$= \frac{(S/N)_D}{(S_R/\eta W)} = \frac{1}{\gamma}\left[\left(\frac{S}{N}\right)_D\right] \text{ since } \gamma = \left(\frac{S_R}{\eta W}\right)$$

15. 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. John G Proakis and Masoud Salehi: *Communication Systems Engineering*, Second Edition (Eastern Economy Edition), Prentice-Hall of India Private Ltd, New Delhi, 2006.
2. Simon Haykin: *Communication Systems*, Third Edition, John Wiley & Sons, 1995.
3. A Bruce Carlson, P B Crilly and Janet C Rutledge: *Communication Systems*, International (Fourth Edition) Edition, McGraw-Hill, 2002.
4. Roy Blake: *Electronic Communication Systems*, Second Edition, Thomson-Delmar, 2002.
5. Leon W Couch, II: *Digital and Analog Communication Systems*, Sixth Edition, Pearson Education Asia, 2001.
6. Herbert Taub and Donald L Schilling: *Principles of Communication Systems*, International Student Edition, McGraw-Hill Kogakusha, Ltd. 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 a 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 a FM signal.
10. Derive an expression for the PSD of noise at the output of the discriminator of a 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. Explain clearly 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?

FILL IN THE BLANKS

1. For determining the destination SNR of a baseband system, we model the receiver as an ideal ____, ____ with a cutoff frequency of _____ Hz.
2. If the additive white noise on the channel has a two-sided PSD of $\eta/2$, and if the transmission bandwidth of the modulated signal being received is B_T Hz, the noise power entering the receiver is _______.

3. A synchronous detector is a product modulator followed by a _____. The two inputs given to the product modulator are _____ and ________.
4. The destination SNR of a DSB-SC system is _________.
5. The destination SNR of a SSB-SC system is _________.
6. The peak amplitude of the FM signal at the input to the discriminator of a FM receiver is A_R. The PSD (2-sided) of the white noise on the channel is $\eta/2$ and the transmission bandwidth of the FM signal being received, is B_T. The pre-detection SNR is _______.
7. The PSD of the noise at the output of the discriminator is proportional to the ____ of the frequency.
8. The destination SNR of a FM system is directly proportional to the ________ (modulation index/ square of the modulation index).
9. In an FM system, the transmission bandwidth __________ (increases/decreases) with the modulation index, β.
10. Pre-emphasis _________ (boosts up/attenuates) the high frequency components of the modulating signal.
11. Pre-emphasis and de-emphasis, used in a FM system, help in improving the _____, ___.
12. The threshold effect in a FM receiver manifests in the form of __________.
13. The threshold value of $(S/N)_R$ for most of the FM receivers is _________.
14. Threshold extension may be obtained by using _____ or a _______.

MULTIPLE CHOICE QUESTIONS

1. The channel noise has a two-sided PSD of $\eta/2$ 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) bandpass filter and lowpass filter (b) highpass filter and lowpass filter
   (c) lowpass filter and lowpass filter (d) bandpass filter and highpass 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) γ/2 (b) γ (c) 2γ (d) γ/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 a 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

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 $(\eta/2) = 0.5\times10^{-12}$ W/Hz. If the received signal power is $S_R = 20\times10^{-9}$ W and the message bandwidth $W = 5\times10^6$ 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, bandlimited 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 (2-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. P-9.1.

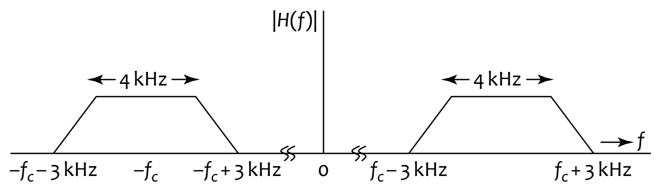


Fig. P-9.1

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 cutoff frequency of 2 kHz. Determine the pre-detection and destination SNR's.

6. A transmitter is producing an average transmitted power of 20 kW. The channel with an additive white noise of PSD (2-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 cutoff 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 cutoff frequency of W Hz.

8. A message signal, $x(t)$, normalized so that $|x(t)| \le 1$, 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 100 kHz bandwidth. Find the ratio of the destination SNRs obtained for the following two cases.
(i) The message frequency modulates the carrier and the modulated signal fully utilizes the full bandwidth of the channel.
(ii) 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 cutoff 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 microseconds, 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 2-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 ms time constant are used
12. An FM receiver employing FMFB for threshold extension, is shown in Fig. P-9.2.

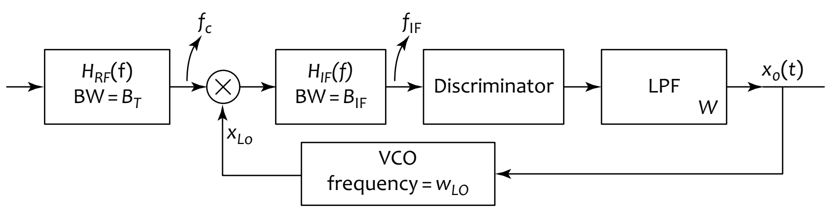


Fig. P-9.2

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?

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)

10 Sampling and Analog Pulse Modulation

In this chapter, the student

- *learns the lowpass sampling theorem and its implications, understands the meaning of terms like 'band-limited' signals, Nyquist rate, aliasing, etc.*
- *understands the usefulness and limitations of various methods of sampling by a study of the spectrum of the sampled signal in each case*
- *realizes, from the lowpass sampling theorem as well as the spectra of the sampled signals, that an LPF may be used for reconstruction of the lowpass band-limited signal from its samples*
- *learns the basic concept of time-division-multiplexing*
- *understands the way the amplitude of each sample of a continuous-time band-limited signal, is represented in PAM, PDM and PPM*
- *recognizes that bandwidth deficiency of the channel results in cross-talk in the transmission of PAM, PDM and PPM signals*
- *will be in a position to mathematically analyze the noise performance of PAM, PDM and PPM systems and compare their noise performance*

10.1 INTRODUCTION

In this chapter, we will be first discussing the lowpass sampling theorem. In essence, this theorem tells us that a lowpass signal $x(t)$, band-limited to W Hz, i.e., one which does not have any frequency components at and above W Hz, can be completely recovered *for all time* from its samples taken at regular intervals T_s, provided $T_s \le \frac{1}{2W}$ seconds. 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 lowpass filter having an appropriate cutoff 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 lowpass 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 book.

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.

10.2 SAMPLING OF BAND-LIMITED LOWPASS 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.

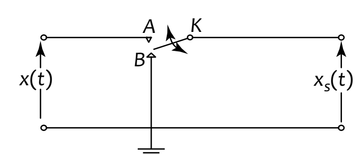


Fig. 10.1 *A switch used for sampling*

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. 10.1. Although a mechanical switch is shown in Fig. 10.1, 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.

Let the switch make contact with A once every T seconds. Then $x_s(t)$ consists of samples of $x(t)$ taken every T seconds, provided the switch makes contact with A instantaneously. However, in practice, the contact will be made for a finite amount of time, say, τ seconds.

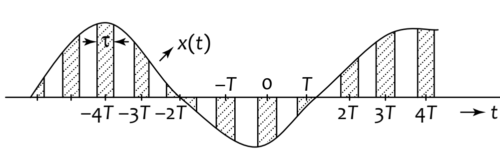


Fig. 10.2 *Sampling waveform $x_s(t)$*

Then the sampled version is as shown in Fig. 10.2. This consists of strips of $x(t)$ of width τ occurring at regular intervals of T seconds; and may be visualized as the waveform that results when $x(t)$ is multiplied by a 'sampling function' shown in Fig. 10.3.

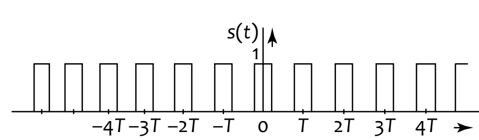


Fig. 10.3 *The sampling function*

This sampling function may be expanded using Fourier series, as it is a periodic function with period T

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi f_0 nt}\,;\; T = \text{sampling period} = \frac{1}{f_s} \tag{10.1}$$

where,

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} s(t)\ e^{-j2\pi nf_0 t}\,dt \tag{10.2}$$

However, over the interval $-T/2$ to $T/2$,

$$s(t) = \Pi(t/\tau),\ \text{since}\ \tau << T$$

Hence, we may write Eq. (10.2) as

$$c_n = f_s \int_{-\tau/2}^{\tau/2} \Pi(t/\tau)\, e^{-j2\pi nf_0 t}\,dt \tag{10.3}$$

As $\Pi(t/\tau) = 0$ outside the limits of integration, we may write

$$c_n = f_s \int_{-\infty}^{\infty} \Pi(t/\tau)\ e^{-j2\pi(nf_s)t}\,dt \quad = \quad f_s\left[\mathcal{F}\{\Pi(t/\tau)\}\right]\Big|_{f=nf_s}$$

since
$$\mathcal{F}[\Pi(t/\tau)] = \tau\,\text{sinc}\, f\tau,$$

$$c_n = f_s\tau\ \text{sinc}\ nf_s\tau \tag{10.4}$$

$$\therefore \qquad s(t) = \sum_{n=-\infty}^{\infty} f_s\tau\ \text{sinc}\ nf_s\tau\ e^{+j2\pi nf_s t} = \left[2f_s\tau\sum_{n=1}^{\infty}\text{sinc}\ (nf_s\tau)\cos 2\pi nf_s t\right] + f_s\tau \tag{10.5}$$

since $f_s = \dfrac{1}{T}$ and $\dfrac{\tau}{\mathrm{T}}$ = duty ratio of the sample function $= \alpha$,

$$s(t) = \alpha + 2\alpha\ \text{sinc}\ (\alpha)\cos 2\pi f_s t + 2\alpha\ \text{sinc}\ (2\alpha)\cos 2\pi 2f_s t + 2\alpha\ \text{sinc}\ (3\alpha)\cos 2\pi 3f_s t + \ldots \tag{10.6}$$

$$\therefore \qquad x_s(t) = x(t).s(t) = \alpha x(t) + 2\alpha\left[x(t)\ \text{sinc}\ \alpha.\cos\omega_s t + x(t)\ \text{sinc}\ (2\alpha).\cos\omega_s t + \ldots\right] \tag{10.7}$$

Taking Fourier transform on both sides, we get

$$X_s(f) = \alpha X(f) + c_1\left[X(f-f_s) + X(f+f_s)\right] + c_2\left[X(f-2f_s) + X(f+2f_s)\right] + \ldots \tag{10.8}$$

where,
$$c_k = \alpha\ \text{sinc}\ k\alpha \tag{10.9}$$

If the signal $x(t)$ has a spectrum as shown in Fig. 10.4, the spectrum of $X_s(f)$, the sampled version of $x(t)$, will be as shown in Fig. 10.5.

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 lowpass filter with a cutoff frequency of B Hz, where B is such that $W < B < (f_s - W)$ and whose gain is constant atleast up to W Hz, then, in time-domain, it

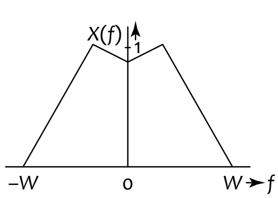


Fig. 10.4 *Spectrum of x(t)*

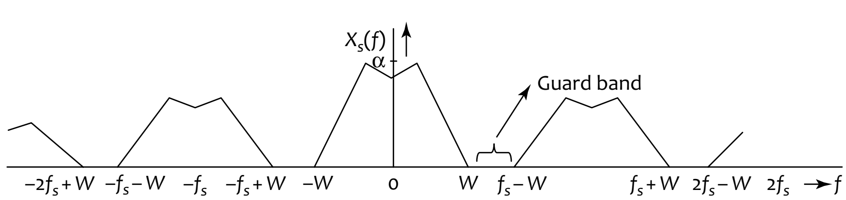


Fig. 10.5 *Spectrum of $x_s(t)$*

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 cutoff frequency equal to W.

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. 10.6. In this case, we find that there is no guard band.

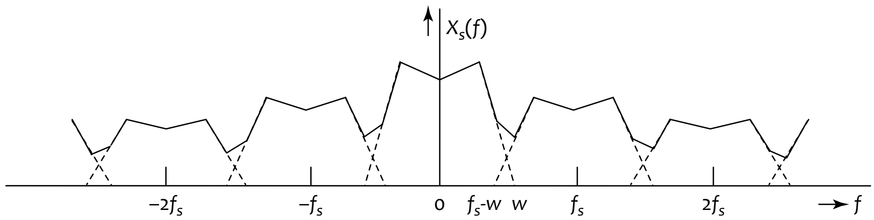


Fig. 10.6 *Spectrum of $x_s(t)$ when $f_s < 2W$*

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

(i) $x(t)$ should be band-limited to some frequency, W, and
(ii) the sampling frequency should be atleast 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 lowpass 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.

10.3 LOWPASS SAMPLING THEOREM

Statement Let $x(t)$ be a band-limited lowpass 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:

$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s)\,\text{sinc}\,2B(t - kT_s) \qquad (10.10)$$

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. 10.7.

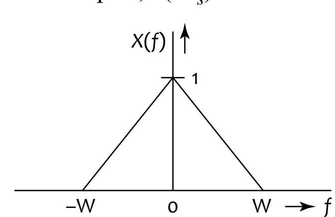


Fig. 10.7 *Spectrum of x(t)*

Consider $\tilde{X}(f)$, shown in Fig. 10.8, which is a periodic repetition of $X(f)$ at regular intervals of frequency equal to f_s, where $f_s > 2W$.

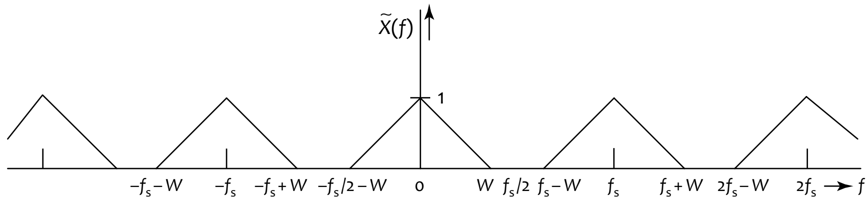


Fig. 10.8 *Spectrum of $x_s(t)$*

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

$$W < \frac{f_s}{2} < (f_s - W) \tag{10.11}$$

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

$$\tilde{X}(f) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi nT_s f}; \qquad T_s = \frac{1}{f_s} \tag{10.12}$$

where,

$$c_n = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} \tilde{X}(f) e^{-j2\pi nf T_s} df \tag{10.13}$$

Since

$$\tilde{X}(f) = X(f) \quad \text{for} \quad |f| < \frac{f_s}{2} \quad \text{and} \quad X(f) = 0 \quad \text{for} \quad |f| \geq W$$

Equation (10.13) may be written as

$$c_n = T_s \int_{-\infty}^{\infty} X(f) e^{-j2\pi nfT_s} df \tag{10.14}$$

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. (10.14) may be written as

$$c_n = T_s x(t) \Big|_{t=-nT_s} = T_s x(-nT_s)$$

Substituting this in Eq. (10.12), we have

$$\tilde{X}(f) = \sum_{n=-\infty}^{\infty} T_s x(-nT_s) e^{j2\pi nfT_s}$$

If we put $k = -n$

$$\tilde{X}(f) = T_s \sum_{k=-\infty}^{\infty} x(kT_s) e^{-j2\pi f\, kT_s} \tag{10.15}$$

If we now define a gate pulse $W_{2B}(f)$ of $2B$ width Hz in the frequency domain, i.e., if

$$W_{2B}(f) \triangleq \Pi\left(\frac{f}{2B}\right) \tag{10.16}$$

then
$$W_{2B}\tilde{X}(f) = X(f) \tag{10.17}$$

$$\therefore \quad X(f) = T_s \sum_{k=-\infty}^{\infty} x(kT_s)\, e^{-j2\pi fkT_s}.W_{2B}(f) \tag{10.18}$$

But
$$x(t) = \mathcal{F}^{-1}[X(f)] = T_s\mathcal{F}^{-1}\left[\sum_{k=-\infty}^{\infty} x(kT_s)e^{-j2\pi fkT_s}.W_{2B}(f)\right]$$

$$= T_s \sum_{k=-\infty}^{\infty} x(kT_s)\left[\mathcal{F}^{-1}\left\{e^{-j2\pi fkT_s}.W_{2B}(f)\right\}\right]$$

$$= T_s \sum_{k=-\infty}^{\infty} x(kT_s)\left\{\mathcal{F}^{-1}\left[e^{-j2\pi fkT_s}\right]\right\}*\left\{\mathcal{F}^{-1}\left[W_{2B}(f)\right]\right\} \tag{10.19}$$

Making use of the convolution theorem of Fourier transform and noting that

$$\mathrm{F}^{-1}\left\lfloor e^{-j2\pi fkT_s}\right\rfloor = \delta(t-kT_s) \tag{10.20}$$

and
$$\mathcal{F}^{-1}\left[W_{2B}(f)\right] = \mathcal{F}^{-1}[\Pi(f/2B)] = 2B \text{ sinc } 2Bt, \tag{10.21}$$

we have,
$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s)\left[\delta(t-kT_s) * \text{sinc } 2Bt\right]$$

$$\boxed{\therefore \quad x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s) \text{ sinc } 2B(t-kT_s)} \tag{10.22}$$

Equation (10.22) 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 lowpass filter whose cutoff frequency is B Hz and whose passband gain is 1, Eq. (10.22) 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 as 'ideal sampling', 'impulse sampling' or 'instantaneous sampling'.

10.4 IDEAL OR IMPULSE SAMPLING

Earlier, in Section 10.2, 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. (10.7).

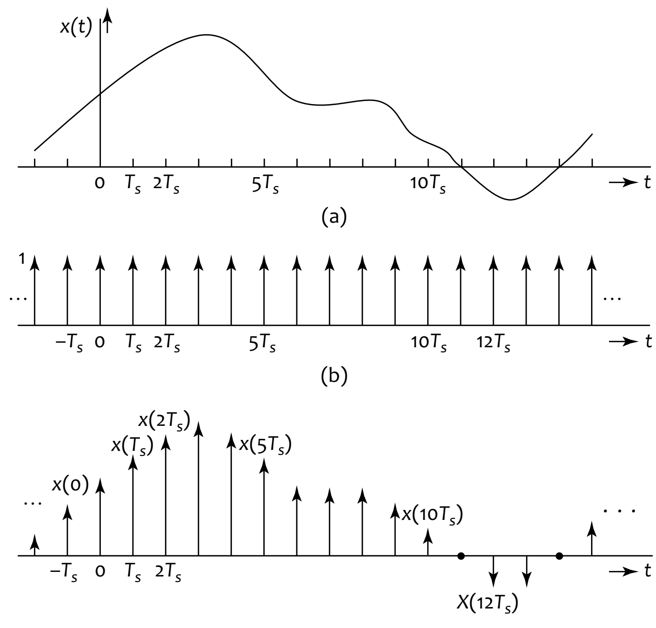


Fig. 10.9 **(a)** $x(t)$ **(b)** $s(t)$ **(c)** $x_\delta(t)$

To overcome this difficulty, we will consider a sampling function that is a sequence of unit impulses as shown in Fig. 10.9(b) instead of a sequence of unit amplitude pulses of zero width. This $s(t)$ may be expressed as

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{10.23}$$

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

$$x_\delta(t) = x(t) \,.\, s(t) \tag{10.24}$$

or,

$$X_\delta(f) = X(f) * S(f) \tag{10.25}$$

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

$$s(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t}; \quad f_s = \frac{1}{T_s}; \quad -\infty < t < \infty \tag{10.26}$$

where,

$$c_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} s(t) e^{-j2\pi n f_s t} dt \tag{10.27}$$

However, for $-\frac{T_s}{2} \le t \le \frac{T_s}{2}, s(t) = \delta(t)$, as may be seen from Fig. 10.9(b).

Hence, $$c_n = \frac{1}{T_s}\int_{-T_s/2}^{T_s/2} \delta(t)e^{-j2\pi n f_s t}dt = \frac{1}{T_s}e^{-j2\pi n f_s t}\bigg|_{t=0} = f_s.1 = f_s \tag{10.28}$$

Since $c_n = f_s$ for all values of n, substituting this in Eq. (10.26), we get

$$s(t) = f_s \sum_{n=-\infty}^{\infty} e^{j2\pi n f_s t}$$

Taking Fourier transform on both sides,

$$S(f) = f_s\mathcal{F}\left[\sum_{n=-\infty}^{\infty} e^{j2\pi n f_s t}\right] = f_s \sum_{n=-\infty}^{\infty} \mathcal{F}\left[e^{j2\pi n f_s t}\right]$$

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

$$\therefore \qquad S(f) = f_s \sum_{n=-\infty}^{\infty} \delta(f - nf_s) \tag{10.29}$$

Substituting this in Eq. (10.25) and realizing that

$$X(f) * \delta(f - nf_s) = X(f - nf_s)$$

and invoking the linearity theorem of FT, we have

$$\boxed{X_\delta(f) = f_s \sum_{n=-\infty}^{\infty} X(f - nf_s)} \tag{10.30}$$

Equation (10.30) 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 lowpass signal band-limited to W, with a spectrum as shown in Fig. 10.10(a) then $X_\delta(f)$ would be as shown in Fig. 10.10(b).

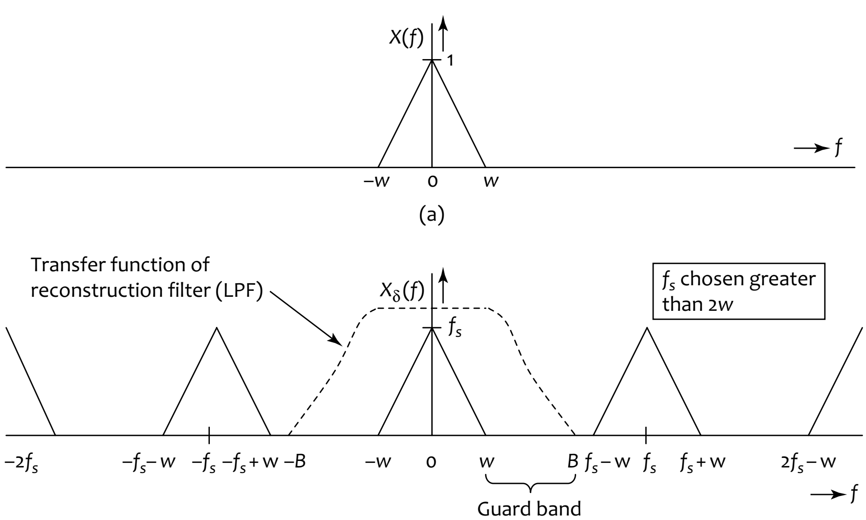


Fig. 10.10 **(a)** $X(f)$ **(b)** $X_\delta(f)$

For the sake of drawing Fig. 10.10(b), it has *been assumed that $f_s > 2W$, i.e., that 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 guardband, 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 lowpass filter whose passband 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. 10.11.

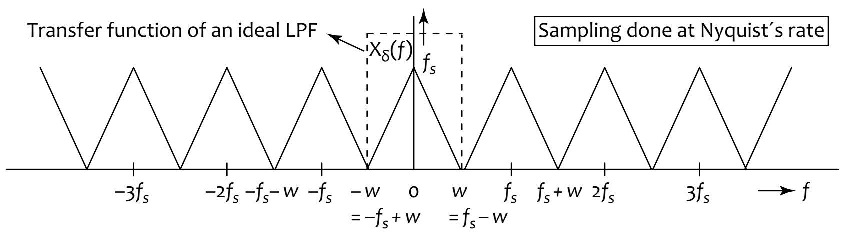


Fig. 10.11 *Spectrum of $x_\delta(t)$ when $f_s = 2W$*

As before, a lowpass filter may be employed to separate out $X(f)$ from the rest of the spectrum of $x_\delta(t)$. However, as the passband 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 guardband in the present case, only an ideal lowpass filter with a cutoff frequency of W will have to be used, as shown in dotted lines in Fig. 10.11.

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. 10.12.

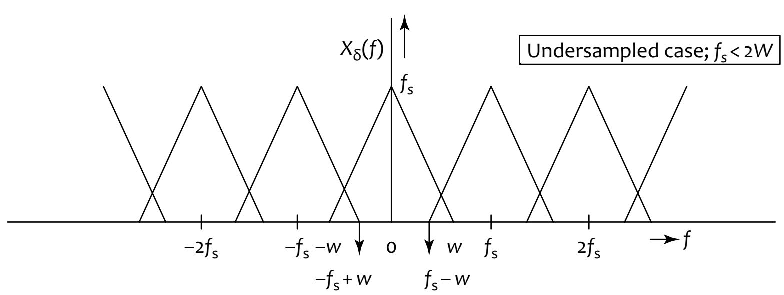


Fig. 10.12 *Spectrum of $x_\delta(t)$ when $f_s < 2W$*

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. 10.12, because of this overlapping, the high frequency components of $x(t)$ re-appear 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.

(a) $X_\delta(f)$ is a repetitive version of $X(f)$, with $X(f)$ repeating itself at regular intervals of f_s, the sampling frequency.

(b) 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.

(c) If $f_s = 2W$, i.e., Nyquist rate, no guard-band exists and an ideal LPF is needed to recover $x(t)$ from $x_\delta(t)$.
(d) If $f_s < 2W$, aliasing takes place and it is not possible to recover $x(t)$ from $x_\delta(t)$ without distortion.
(e) To avoid aliasing, it should be ensured that
   (i) $x(t)$ is strictly band-limited
   (ii) 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 $-\infty$ of time to $+\infty$ 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. 10.13, it is necessary to first band-limit $x(t)$ to some appropriate frequency W such that most part of the energy is retained.

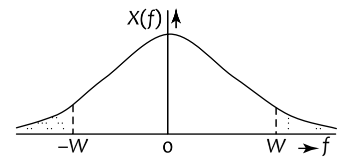


Fig. 10.13 *Spectrum of an x(t)*

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 10.1

The signal $x(t) = 10\cos 150\pi t$ is ideally sampled at a frequency $f_s = 200$ samples per second (sps). Sketch the spectrum of $x_\delta(t)$.

$$X(f) = \mathcal{F}[x(t)] = \mathcal{F}[10\cos 150\pi t]$$
$$= 5[\delta(f-75) + \delta(f+75)]$$

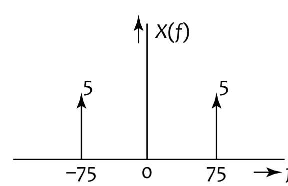


Fig. 10.14 *Sketch of X(f)*

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:

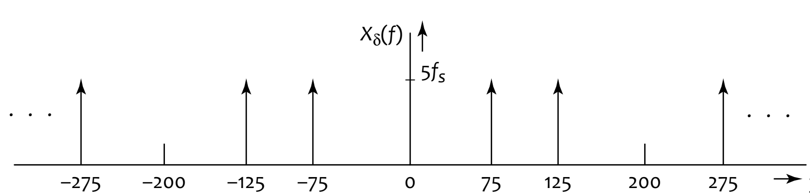


Fig. 10.15 *Sketch of $X_\delta(f)$ with $f_s = 200$ sps*

Example 10.2

For the x(t) of Example 10.1, 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.

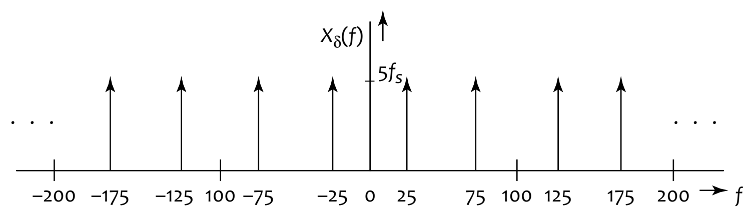


Fig. 10.16 *Sketch of $X_\delta(f)$ of Example 10.2 with $f_s = 100$ sps*

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 sps while the frequency of the signal is 75 Hz. Thus, the sampling rate of 100 sps 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 re-appearing as a low-frequency component at 25 Hz because of aliasing.

Example 10.3

How many minimum number of samples are required to exactly describe the following signal:

$$x(t) = 10\cos(6\pi t) + 4\sin(8\pi t)$$

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.

$$T_1 = \text{period of } 10\cos 6\pi t = \frac{2\pi}{6\pi} = \frac{1}{3}$$

$$T_2 = \text{period of } 4\sin 8\pi t = \frac{2\pi}{8\pi} = \frac{1}{4}$$

$$\therefore \quad \frac{T_1}{T_2} = \frac{1}{3} \times \frac{4}{1} = \frac{4}{3}, \text{ which is a rational number.}$$

Hence, $x(t)$ is periodic. Now, to determine its period T,

$$T = \text{LCM}\left(\frac{1}{3}, \frac{1}{4}\right) = 1 \quad \therefore \quad T = 3T_1 = 4T_2$$

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 sps.
$\therefore$ the number of samples in one period of $x(t)$ is equal to 8 since $T = 1$ second and the sampling frequency is 8 samples per second.

Example 10.4

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.

$$x(t) = 100 \text{ sinc}^2 100t = (10 \text{ sinc } 100t).(10 \text{ sinc } 100t)$$

We know that $10 \text{ sinc } 100t \xleftrightarrow{\text{FT}} 0.1\Pi(f/100)$

$$\therefore \qquad \mathcal{F}\lfloor 100 \text{ sinc}^2 100t \rfloor = [0.1\Pi(f/100)] * 0.1\Pi(f/100)$$

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.

$$\therefore \qquad \mathcal{F}\lfloor 100 \text{ sinc}^2 \ 100t \rfloor = [0.1\Pi(f/100)] * [0.1\Pi(f/100)]$$

$$= 0.01 \times 100\Lambda(f/200)$$

where $\Lambda(f/200)$ denotes a triangular pulse as shown.

Thus, the signal $x(t) = 100 \text{ sinc}^2 100t$ is a lowpass signal band-limited to 100 Hz.

Hence, the Nyquist rate for it is 200 sps.

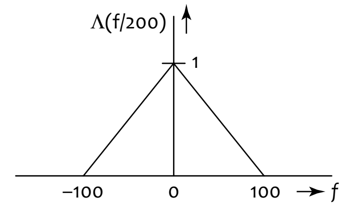


Fig. 10.17 *Triangular spectrum of the x(t) = 100 sinc² 100t signal*

RECONSTRUCTION 10.5

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. 10.11, 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 passband and let it introduce τ second time-delay. Then, we can write down its transfer function $H(f)$ as

$$H(f) = T_s\Pi(f / f_s)e^{-j\omega\tau} \tag{10.31}$$

Hence the spectrum of the output of the filter is

$$Y(f) = X_\delta(f).H(f) = T_s f_s X(f)e^{-j\omega\tau}$$

$$\therefore \qquad Y(f) = X(f)e^{-j\omega\tau} \tag{10.32}$$

or, taking the inverse Fourier transform on both sides,

$$y(t) = x(t - \tau) \tag{10.33}$$

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

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) \tag{10.34}$$

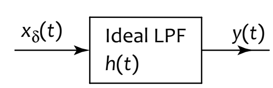


Fig. 10.18 *Recovering x(t) from x_δ(t), the sampled version*

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

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT_s)h(t - nT_s) \tag{10.35}$$

where $h(t)$, the impulse response of the ideal LPF is given by

$$h(t) = \mathcal{F}^{-1}[H(f)] = \mathcal{F}^{-1}\left[T_s \Pi(f/2W)e^{-j\omega\tau}\right]$$
$$= 2BT_s \operatorname{sinc} 2B(t-\tau) \quad (10.36)$$

In our case, the cutoff frequency B of the LPF = $W = f_s/2$.

$$\therefore \quad h(t) = 2\frac{f_s}{2}T_s \operatorname{sinc} 2B(t-\tau)$$
$$= \operatorname{sinc} 2B(t-\tau) \quad (10.37)$$

Taking the time delay τ introduced by the LPF equal to zero, and substituting for $h(t)$ in Eq. (10.35) using Eq. (10.37), we have

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \operatorname{sinc} 2B(t-nT_s) \quad (10.38)$$

But, from Eq. (10.22), we realize that RHS of Eq. (10.38) is nothing but $x(t)$, since $B = \frac{f_s}{2} = \frac{1}{2T_s}$ in this case.

$$\therefore \quad x(t) = \sum_{n=-\infty}^{\infty} x(nT_s) \operatorname{sinc} 2B(t-nT_s) \quad (10.39)$$

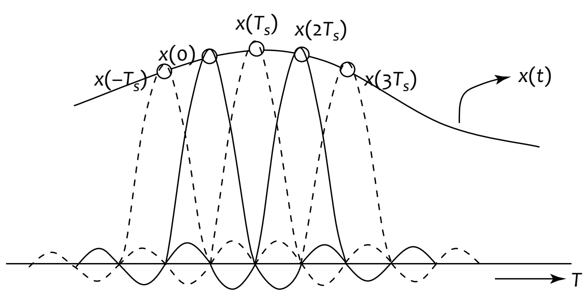


Fig. 10.19 *Reconstruction of x(t) from its samples (A sketch of RHS of Eq. 10.39)*

As explained earlier in Section 10.3 in connection with the lowpass sampling theorem and Eq. (10.22), 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. 10.19. When these are all added, together with their precursors and post-cursors, Eq. (10.39) 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'.

10.6 SAMPLING USING A SEQUENCE OF PULSES

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. 10.20(b). Again, modeling the sampling process as multiplication of $x(t)$ by $s(t)$, we have

$$x_s(t) = x(t).s(t) \quad (10.40)$$

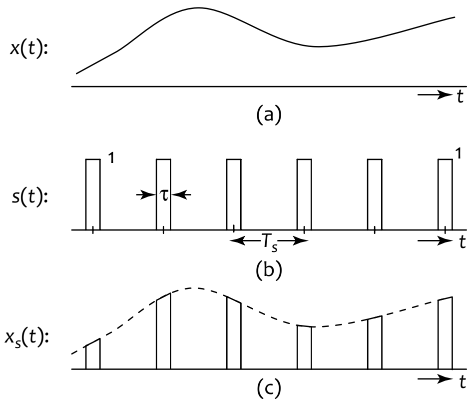


Fig. 10.20 **(a)** *Signal x(t),* **(b)** *Sampling function s(t)* **(c)** *Sampled version of x(t), i.e., $x_s(t)$*

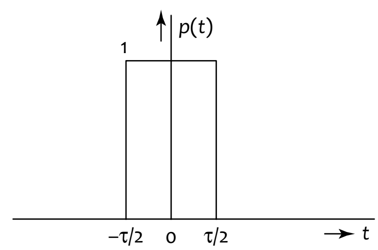


Fig. 10.21 *Pulse p(t)*

where,

$$s(t) = \sum_{k=-\infty}^{+\infty} p(t - kT_s) \tag{10.41}$$

As $s(t)$ is a periodic pulse train, let us write its Fourier series expansion

$$s(t) = \sum_{k=-\infty}^{+\infty} p(t - kT_s) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_s t} \tag{10.42}$$

where,

$$c_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} s(t) e^{-j2\pi n f_s t}\, dt$$

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

$$c_n = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} s(t) e^{-j2\pi n f_s t}\, dt = \frac{1}{T_s} \int_{-\infty}^{\infty} p(t) e^{-j2\pi n f_s t}\, dt$$

$$c_n = f_s P(nf_s) \tag{10.43}$$

where,

$$P(nf_s) = \mathcal{F}[p(t)]\Big|_{f = nf_s} \tag{10.44}$$

$\therefore$
$$s(t) = f_s \sum_{n=-\infty}^{\infty} P(nf_s)\, e^{j2\pi n f_s t} \tag{10.45}$$

and

$$X_s(f) = \mathcal{F}[x_s(t)] = \mathcal{F}\left[f_s \sum_{n=-\infty}^{\infty} P(nf_s) x(t) e^{j2\pi n f_s t} \right]$$

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

Since $$\mathcal{F}\lfloor e^{j2\pi nf_s t}\rfloor = \delta(f - nf_s)$$

hence,

$$X_s(f) = f_s \sum_{n=-\infty}^{\infty} P(nf_s)X(f - nf_s) \tag{10.46}$$

If $x(t)$ has a spectrum as shown in Fig. 10.10(a), $X_s(f)$, the spectrum of the sampled version of $x(t)$ will appear as shown in Fig. 10.22.

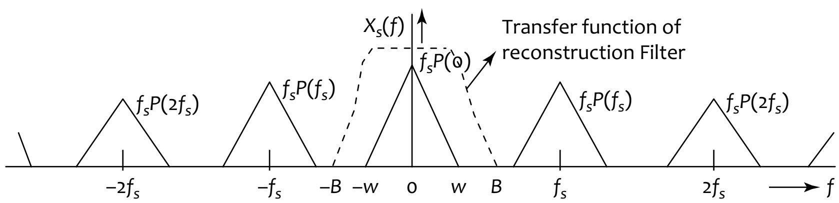


Fig. 10.22 *Spectrum of the sampled version of x(t)*

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 lowpass filter whose passband gain is constant at least up to W Hz and whose cutoff 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 10.5

The signal $x(t) = 2\cos 200\pi t + 6\cos 180\pi t$ is ideally sampled at a frequency of 150 samples per second. The sampled version $x_\delta(t)$ is passed through a unit gain ideal LPF with a cutoff frequency of 110 Hz. What frequency components will be present in the output of the LPF. Write down an expression for its output signal.

$$x(t) = 2\cos 200\pi t + 6\cos 180\pi t$$

$$= 2\cos 2\pi(100)t + 6\cos 2\pi(90)t$$

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. 10.23.

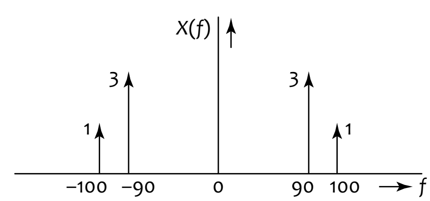


Fig. 10.23 *Spectrum of x(t)*

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, i.e., 150 Hz; and will be as shown in Fig. 10.24.

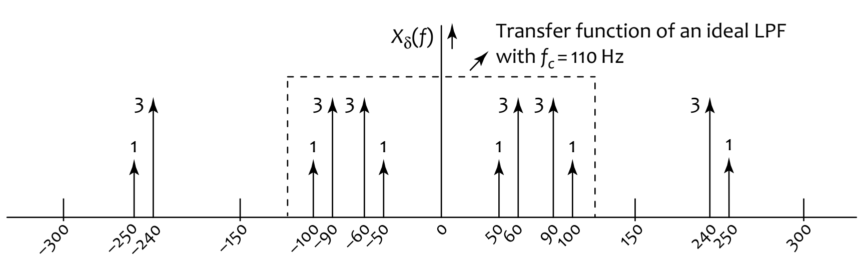


Fig. 10.24 *Spectrum of sampled version of x(t)*

From Fig. 10.24, 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 10.6

The signal $x(t) = 12\cos(800\pi t)\cos^2(1800\pi t)$ is ideally sampled at 4600 sps. What is the minimum allowable sampling frequency? What is the range of permissible cutoff frequencies for the ideal lowpass filter to be used for reconstructing the signal?

$$x(t) = 12\cos(800\pi t)\left[\frac{1}{2}\{1+\cos 3600\pi t\}\right]$$

$$= 6\cos 800\pi t + 6\cos 800\pi t.\cos 3600\pi t$$

$$= 6\cos 800\pi t + 3\cos 4400\pi t + 3\cos 2800\pi t$$

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 sps.

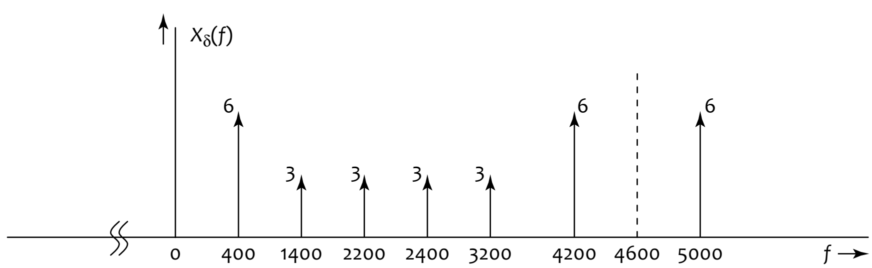


Fig. 10.25 *Spectrum of ideally sampled version of x(t) (only one-sided spectrum drawn)*

From Fig. 10.25, 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 cutoff frequency of the ideal LPF should be above 2200 Hz but less than 2400 Hz.

10.7 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. 10.26 and a typical output waveform from a S/H circuit is shown in Fig. 10.27.

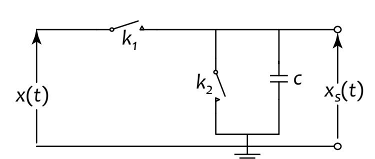


Fig. 10.26 *Schematic of an S/H circuit*

The S/H circuit essentially consists of two switches k_1 and k_2 and a capacitor c, connected as shown. 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. 10.27.

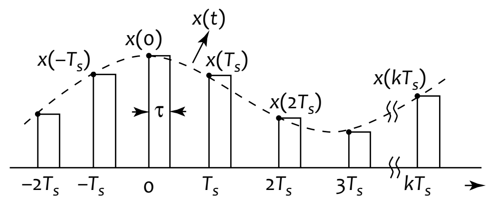


Fig. 10.27 *Signal x(t) and output of S/H circuit*

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

$$x_s(t) = \sum_{k=-\infty}^{\infty} x(kT_s)p(t - kT_s), \tag{10.47}$$

where,
$$p(t) \triangleq \Pi\left(\frac{t - \tau/2}{\tau}\right) \tag{10.48}$$

and is as shown in Fig. 10.28.

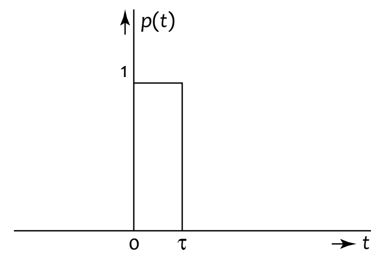


Fig. 10.28 *Pulse p(t)*

Since $\qquad p(t - kT_s) = p(t) * \delta(t - kT_s),$

we may write Eq. (10.47) as

$$x_s(t) = p(t) * \sum_{k=-\infty}^{\infty} x(kT_s)\,\delta(t - kT_s) \tag{10.49}$$

Now, taking Fourier transform on both sides,

$$X_s(f) = \mathcal{F}[x_s(t)] = \mathcal{F}[p(t)].\mathcal{F}\left[\sum_{k=-\infty}^{\infty} x(kT_s)\delta(t - kT_s)\right]$$

$\therefore$
$$X_s(f) = p(f).X_\delta(f) \tag{10.50}$$

where, $X_\delta(f)$ is the Fourier transform of $x_\delta(t)$ the ideally sampled version of $x(t)$.

Note that
$$P(f) = \mathcal{F}[p(t)]$$
$$= (\tau \operatorname{sinc} f\tau)e^{-j2\pi f\tau/2} = \tau\left(\frac{\sin \pi f\tau}{\pi f\tau}\right)e^{-j\pi f\tau} \tag{10.51}$$

We shall now assume that $x(t)$ has a spectrum as shown in Fig. 10.29 (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. 10.30(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 $X_s(f) = P(f).X_\delta(f)$, its plot will be as shown in Fig. 10.30(b).

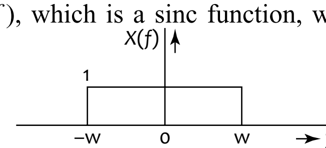


Fig. 10.29 *Assumed shape of X(f)*

As before, if we pass the sampled version $x_s(t)$ through the reconstruction filter (a 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

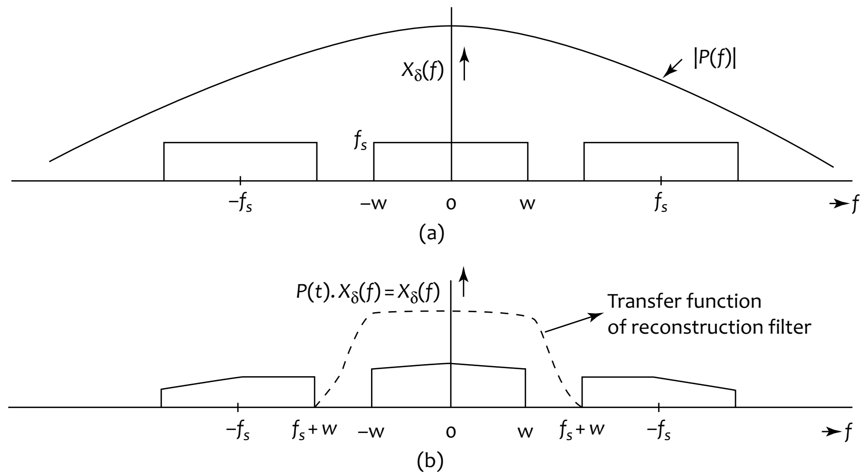


Fig. 10.30 **(a)** *Plot of P(f) and Xδ(f);* **(b)** *Plot of Xδ(f) = P(f).Xδ(f)*

of the low-frequency components, as can be seen in Fig. 10.30(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. 10.30(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

$$H_e(f) = \frac{1}{P(f)}; \qquad |f| \le W \tag{10.52}$$

Example 10.7

Figure 10.31 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.s duration and unit amplitude, determine and sketch the spectrum of the PAM signal that results.

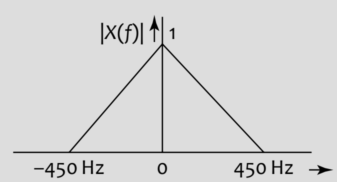


Fig. 10.31 *Spectrum of x(t)*

From Eq. (10.50) 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

$$P(f) = \tau \sin c f \tau e^{-j\pi f \tau} = \tau \left(\frac{\sin \pi f \tau}{\pi f \tau} \right) e^{-j\pi f \tau} \text{ (see Eq. 10.51)}$$

Here, τ is the width of the pulse and is given to be 10^{-4} s.

Since it is the relative attenuation of the high-frequency components of $x(t)$ relative to the low-requency components that causes the distortion, the constant factor τ and the phase factor $e^{-j\pi f \tau}$ can be ignored attention can be focused only on $\left[(\sin \pi f \tau)/(\pi f \tau)\right]$ 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 $\left(\dfrac{\sin \pi f \tau}{\pi f \tau} \right) = 1$ for $f = 0$ Hz.

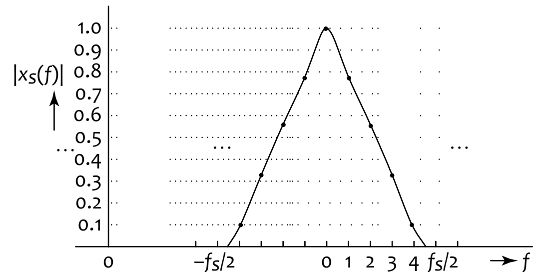


Fig. 10.32 *Spectrum of the PAM signal*

f	0	100	200	300	400	450
$\|X(f)\|$	1	0.7777	0.5555	0.3333	0.1111	0
$\left(\dfrac{\sin \pi f \tau}{\pi f \tau} \right) = 1$	1	0.96639	0.9836	0.9629	0.93547	0.91878
$\|X(f)\|P(f)\| = \|X_s(f)\|$	1	0.7749	0.54638	0.3209	0.10393	0

10.8 ANTI-ALIASING AND RECONSTRUCTION FILTERS

From the lowpass 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 cutoff 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 passband edge f_p to some desired value at the stop band edge f_s, where

$$f_p < f_{st} \leq \frac{f_s}{2} \tag{10.53}$$

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 passband, 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 passband 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 passband 0 to f_p Hz, which is the useful part of the baseband, as shown in Fig. 10.33.

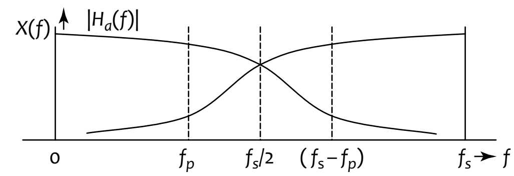


Fig. 10.33 *Anti-aliasing filter response and frequency components aliasing into the baseband*

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 passband, 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 passband, the other frequency components between $(f_s - f_p)$ and f_s which re-appear in the passband between f_p and 0 Hz, will suffer more than 60 dB attenuation.

Butterworth, Chebyshev, elliptic or Bessel type of analog lowpass 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 passband edge, typically about 4 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 cutoff lowpass 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 the compact disk encoding of audio signals, sampling frequencies as high as 3175.2 kHz are used.

10.8.1 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. 10.34, 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.

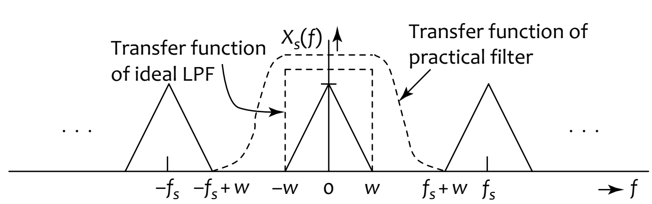


Fig. 10.34 *Action of the reconstruction filter in the frequency domain*

In principle, an ideal LPF with a cutoff frequency of W as shown in Fig. 10.34 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 lowpass filter with a flat amplitude response up to W Hz and whose gain reduces to zero before $(f_s - W)$ may be used.

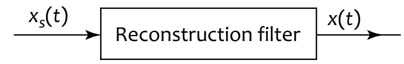


Fig. 10.35 *Action of reconstruction filter in the time domain*

The action of the reconstruction filter when viewed in the time domain, is shown in Fig. 10.35. 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 10.36(a) shows the block schematic of a ZOH while Figure 10.36(b) shows its interpolating action.

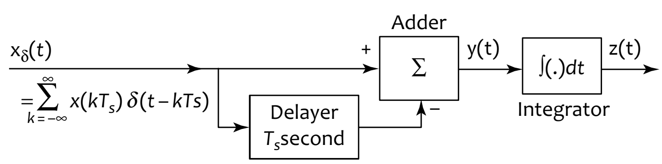


Fig. 10.36 (a) *A Zero-Order-Hold (ZOH) circuit*

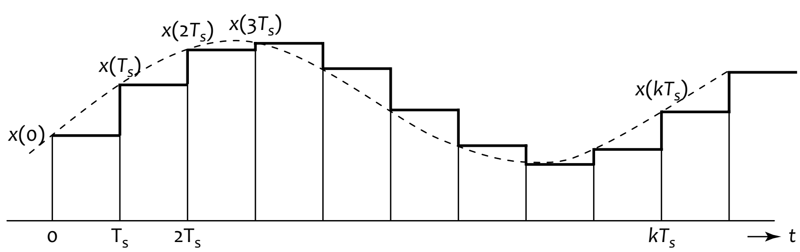


Fig. 10.36 (b) *Interpolation using a ZOH*

It is easy to find the impulse response of a zero-order-hold as can be seen from the following example.

Example 10.8

Determine the impulse response h(t) and the transfer function H(f), for a ZOH.

If a unit impulse, $\delta(t)$, is given as input to the system,

$$y(t) = \delta(t) - \delta(t - T_s)$$

$\therefore$
$$z(t) = \int y(t)dt = u(t) - u(t - T_s) = p(t) \tag{10.54}$$

where, $p(t)$ = impulse response $h(t)$ and is as shown in Fig. 10.35.

The transfer function $H(f)$ is, therefore, given by

$$H(f) = \mathcal{F}[h(t)] = T_s \text{ sinc}(fT_s)e^{-j\pi f T_s} \tag{10.55}$$

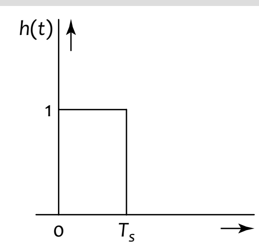


Fig. 10.37 *Impulse response of ZOH*

Hence, the output of the ZOH for an input of

$$x_\delta(t) = \sum_{k=-\infty}^{\infty} x(kT_s)\delta(t - kT_s)$$

is a staircase waveform as shown in Fig. 10.34(b). This contains several high-frequency components outside the baseband. Hence, the ZOH is generally followed by a LPF with a cutoff 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. (10.55) which gives the transfer function of the ZOH, two things are quite clear.

(i) ZOH gives a linear phase shift corresponding to a time delay of $T_s/2$.

(ii) Since the spectrum of the reconstructed signal is equal to $X_\delta(f).H(f)$ for $|f| \le W$, and since $H(f)$ is a sinc function while $X_\delta(f) = X(f)$ for $|f| \le W$ when $f_s \le 2W$, 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 10.9

An L-section RC lowpass filter with a 30 dB cutoff 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 passband, i.e., at f_c is to be at least 30 dB below the response at f_c?

For L-section RC lowpass filter, the transfer function is

$$H(f) = \frac{1}{1 + j(f/f_c)}$$

where, $f_c = 3$ dB cutoff frequency $= \dfrac{1}{2\pi RC}$

$$(\text{Response at } f_c) = \frac{1}{1+j} \quad \therefore |H(f)|\Big|_{f=f_c} = \frac{1}{\sqrt{2}}$$

Now, referring to Fig. 10.32, 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, $$x \triangleq \left(\frac{f_s}{f_c}\right)$$

$\therefore$ $$\frac{|H(f)|\Big|_{f=f_c}}{|H(f)|\Big|_{f=(f_s-f_c)}} = \left(\frac{\dfrac{1}{\sqrt{2}}}{\left(\dfrac{1}{\sqrt{1+(x-1)^2}}\right)}\right)$$

$\therefore$ $$10\log_{10}\left(\frac{\dfrac{1}{2}}{\left(\dfrac{1}{1+(x-1)^2}\right)}\right) \ge 30 \quad \text{or} \quad \frac{1+(x-1)^2}{2} \ge 10^3$$

i.e., $$x \ge \left[\sqrt{2000-1}+1\right] = 45.2$$

$\therefore$ $$f_s \ge 45.2 f_c$$

10.9 PAM AND TIME-DIVISION MULTIPLEXING

PAM signals may be generated staight away by flat-top sampling discussed in section 10.7 (see Figs 10.26 and 10.27).

The PAM signal of Fig. 10.27 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

$$\underset{\text{PAM}}{x_s(t)} = \sum_k A_0\left[1 + mx(kT_s)\right] p(t - kT_s) \tag{10.56}$$

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. 10.28; 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

$$1 + mx(kT_s) \tag{10.57}$$

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. 10.29, the spectrum of the unipolar PAM signal of Eq. (10.56) will be similar to what has been shown in Fig. 10.30 except that there will be impulses in the spectrum at $f = 0$, $\pm f_s$, $\pm 2f_s$, ... As shown in Fig. 10.30 we may use a lowpass 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.

10.9.1 Time-Division Multiplexing (TDM)

The lowpass 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 bandlimited 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.

Messages $x_1(t)$, $x_2(t)$, . . , $x_N(t)$ which are all to be time-division multiplexed, are first band-limited using lowpass filters. These band-limited signals are then sequentially sampled by the arm of the commutator at the sendingend. 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 $x_i(t)$.

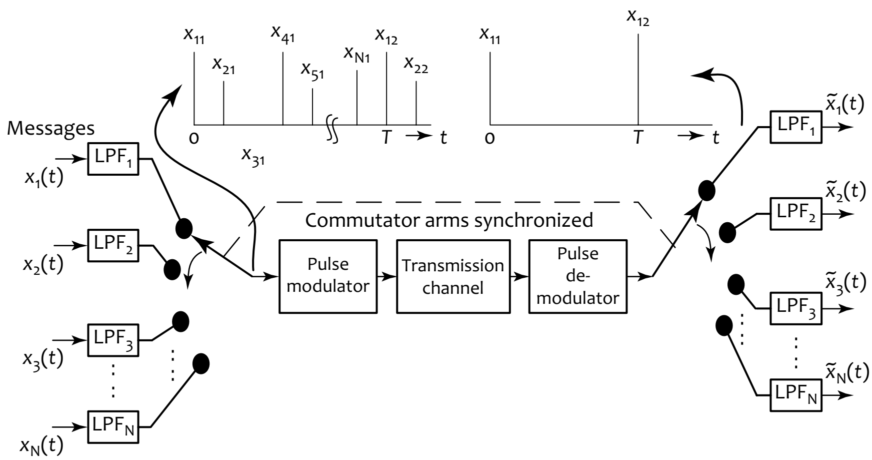


Fig. 10.38 *A PAM/TDM system*

10.9.2 Bandwidth of TDM-ed Baseband PAM signals

Assume that messages $x_1(t)$, $x_2(t)$, . . . , $x_N(t)$, each band-limited to W Hz, have been flat-top sampled by narrow pulses (which can be approximated by impulses). Assume that this baseband TDM-ed signal is transmitted over a channel with finite bandwidth. For convenience, let us model the channel by an ideal LPF with a cutoff 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 $(T_s/N) = (1/2f_c)$ 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. 10.38, 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.

$$\therefore \qquad \frac{T_s}{N} = \frac{1}{2f_c}$$

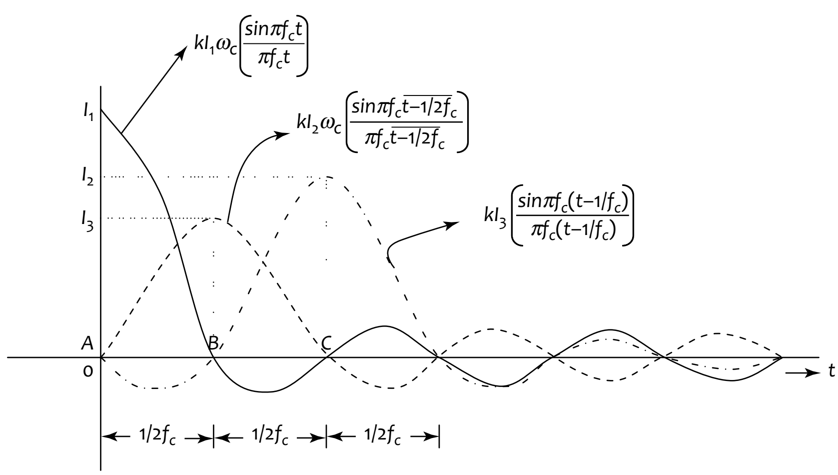


Fig. 10.39 *Response of the channel (ideal LPF) to successive samples (impulses) fed to it at t = 0, 1/2f_c, 1/f_c, etc.*

Hence,

$$f_c = \frac{N}{2T_s}$$

Since $$T_s \le \frac{1}{2W}, \text{ we have}$$

$$f_c \ge \frac{N}{\left(2.\frac{1}{2W}\right)} = NW$$

$\therefore$ $$f_c \ge NW \tag{10.58}$$

Remarks

1. 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.
2. 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 sub-carrier modulation (with no guard bands).

Example 10.10

Twenty-four 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?

(i) FDM with SSB modulation (ii) TDM with pulse amplitude modulation.

(i) *FDM-SSB*: With SSB, each message channel occupies 4 kHz and the 24 messages can be accommodated in 24 non-overlapping frequency slots, each of 4 kHz width. Hence, total bandwidth required for the frequency division multiplexed signal, is $24 \times 4 = 96$ kHz. It is assumed here that no guardbands have been provided. Since we are required to find the minimum bandwidth.

(ii) *TDM-PAM*: Equation 10.58 tells us that for TDM-PAM of N different messages, each of W Hz bandwidth, the minimum bandwidth required is NW Hz.

Example 10.11

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 all the three signals are samples at their respective Nyquist rates.

Since $x_1(t)$ and $x_2(t)$ have bandwidths of 10 kHz each, the Nyquist rate of sampling for them is 20 kilosamples/s. The Nyquist rate of sampling for $x_3(t)$ with 15 kHz bandwidth, is 30 kilosamples/s. So the commutator arrangement shown satisfies the requirement.

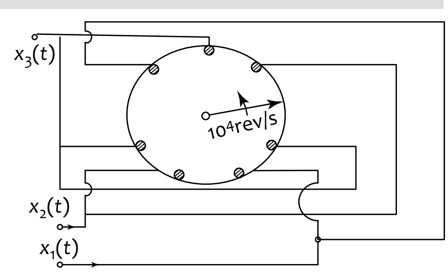


Fig. 10.40 *Commutator for Example 10.11*

10.9.3 Cross-talk in PAM

As shown in Fig. 10.37, 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 re-constructed from received samples.

Cross-talk can occur due to the following reasons:

(i) High-frequency limitation of the channel
(ii) Low-frequency limitation of the channel

10.9.4 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. In so far 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 sometimes 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.

(i) 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 $RC = \tau_c << \tau$.

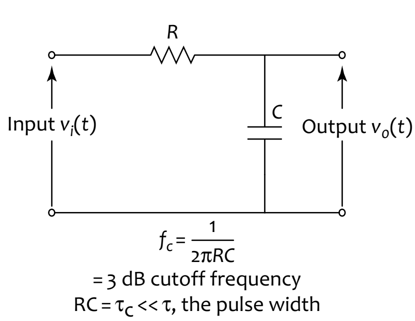


Fig. 10.41 *A lowpass R-C filter*

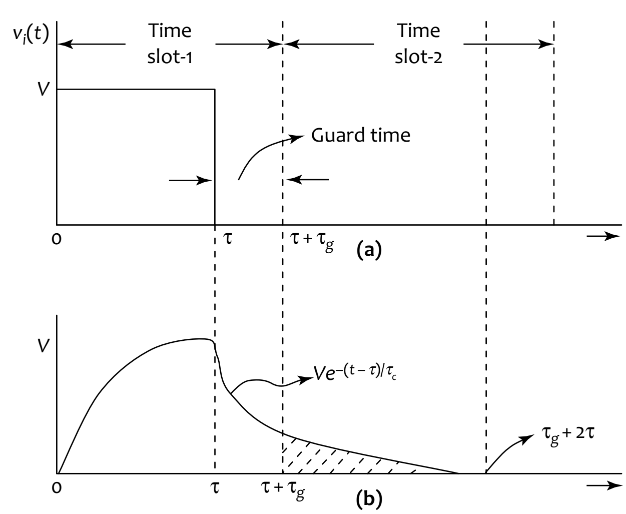


Fig. 10.42 **(a)** *Waveform of the transmitted pulse* **(b)** *Waveform of the received pulse*

Figure 10.42(a) shows the waveform of the input pulse and Fig. 10.42(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. 10.42(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

$$K \triangleq \frac{A_{12}}{A_2} \tag{10.59}$$

where, A_{12} = Shaded area in Fig. 10.42(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,

$$K = \frac{A_{12}}{A_1} \tag{10.60}$$

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 $t = 2\tau + \tau_g$, $\lambda = \tau + \tau_g$

$$\therefore \quad A_{12} = \int_{\lambda=\tau_g}^{\tau+\tau_g} Ve^{-\lambda/\tau_c}\,d\lambda = V\tau_c e^{-\tau_g/\tau_c}\left[1 - e^{-\tau/\tau_c}\right]$$

However, since $\tau >> \tau_c$, $e^{-\tau/\tau_c} \approx 0$ and so we have

$$A_{12} \approx V\tau_c e^{-\tau_g/\tau_c} \tag{10.61}$$

Substituting this for A_{12} in Eq. (10.60), we get

$$K = \frac{A_{12}}{A_1} = \frac{V\tau_c e^{-\tau_g/\tau_c}}{V\tau} = \left(\frac{\tau_c}{\tau}\right)e^{-\tau_g/\tau_c}$$

$$\boxed{\therefore \quad K = \left(\frac{\tau_c}{\tau}\right) e^{-\tau_g/\tau_c}} \tag{10.62}$$

From this equation, we find that

1. when $\tau_g >> \tau_c$, $K \cong 0$ and there is no cross-talk, and
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.

(ii) 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 *R-C* filter, but of the high-pass type, as shown in Fig. 10.43.

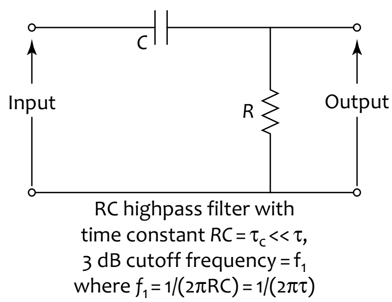


Fig. 10.43 *RC highpass filter used for modeling the channel*

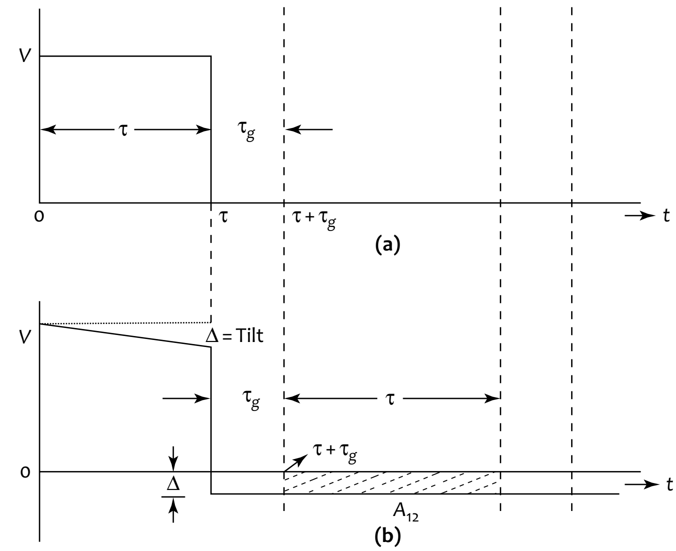


Fig. 10.44 **(a)** *Waveform of the transmitted pulse* **(b)** *Waveform of the recieved pulse*

As in the previous case, here too

$$A_1 \approx V\tau \quad (\because \quad \Delta << V) \tag{10.63}$$

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. 10.44(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,

$$\Delta = V - \left[Ve^{-t/\tau_c}\right]_{t=\tau} = V - Ve^{-\tau/\tau_c}$$

$$= V - V\left[1 - \frac{\tau}{\tau_c} + \frac{\tau^2}{\tau_c^2} - ....\right] \cong \frac{V\tau}{\tau_c}$$

(since $\tau << \tau_c$)

$$\therefore \qquad A_{12} \cong \Delta\tau = \frac{V\tau^2}{\tau_c}$$

Hence,

$$K = \frac{A_{12}}{A_1} = \frac{V\tau^2}{\tau_c} \cdot \frac{1}{V\tau} = \frac{\tau}{\tau_c}$$

$$\boxed{\therefore \quad K = \frac{\tau}{\tau_c}} \tag{10.64}$$

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.

Example 10.12

Twelve 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}.

f_s = Sampling rate = 8000 samples/second

$\therefore$ T_s, the sampling period $= \dfrac{1}{f_s} = \dfrac{1}{8000} = 125\ \mu s$

Since there are 12 message signals to be TDM-ed, the duration of the time-slot for each, i.e., τ is given by

$$\tau = \frac{125}{12} \text{ microseconds} = 10.41 \text{ microseconds.}$$

Since time-slot includes pulse width and guard time, and since guard time is given to be half of the pulse width,

$$\text{Pulse width} = \tau = 10.41 \times \frac{2}{3} = 6.94 \text{ microseconds}$$

$$\therefore \quad \text{Guard time} = \tau_g = 10.41 - 6.94 = 3.47 \text{ microseconds}$$

We know that
$$K = \left(\frac{\tau_c}{\tau}\right) e^{-\tau_g/\tau_c}$$

In the above equation, we know K, τ and τ_g. $\therefore$ solving for τ_c, we get

$$\tau_c = 0.75 \text{ microsecond} = RC$$

$\therefore$ upper 3-dB cutoff frequency of the channel $= \dfrac{1}{2\pi RC} = \dfrac{10^6}{2\pi \times 0.75} = 212.314$ kHz

$\therefore$ the minimum bandwidth of the channel = 212.314 kHz.

10.10 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) that '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) that '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, generally they are clubbed together under the common name, 'pulse-time modulation'.

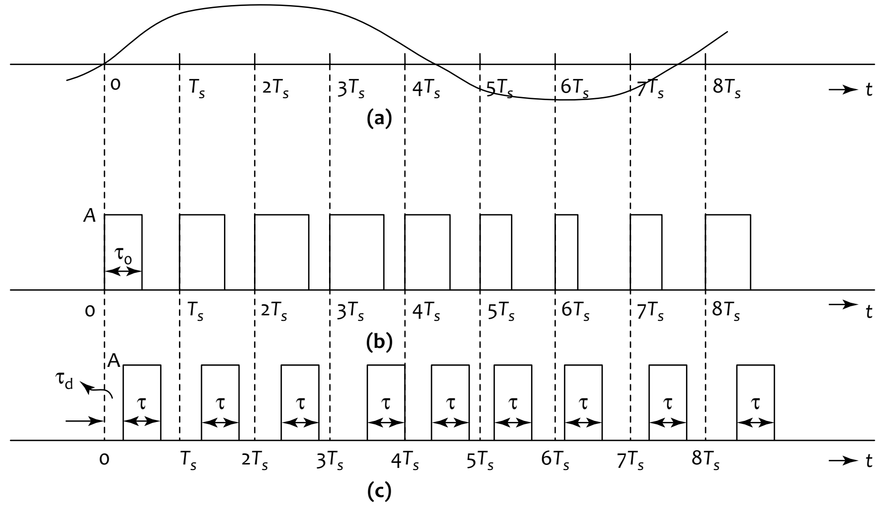


Fig. 10.45 **(a)** *Message signal x(t)* **(b)** *PDM signal* **(c)** *PPM signal*

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. 10.45(b), is given by

$$\tau_k = \tau_0 \left[1 + mx(kT_s)\right] \tag{10.65}$$

where, τ_0 is the un-modulated 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

$$\tau_0 < \frac{T_s}{2} \tag{10.66}$$

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. (10.66), 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 lowpass 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. (10.65), 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. 10.45 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

$$\tau_k = kT_s + \tau_d + t_0 x(kT_s) \tag{10.67}$$

where, t_0 is a proportionality constant having units of seconds per volt.

10.10.1 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. 10.46. In this method, the PAM signal and an inverse ramp signal are generated synchronously,

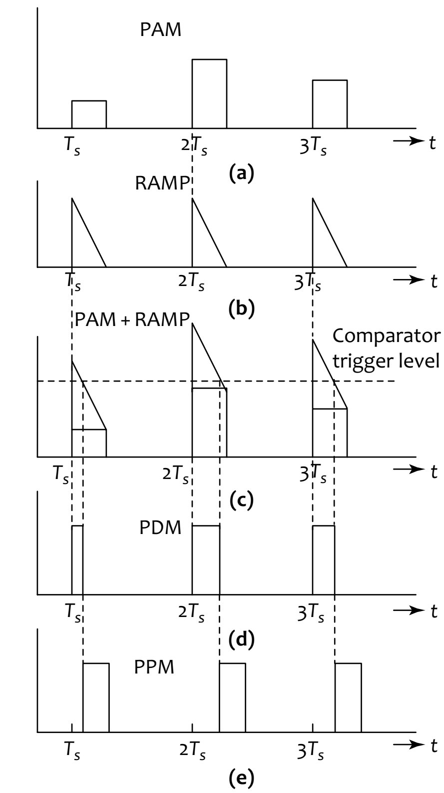


Fig. 10.46 **(a)** *PPM* **(b)** *RAMP* **(c)** *PPM + AMP* **(d)** *PDM* **(e)** *PPM*

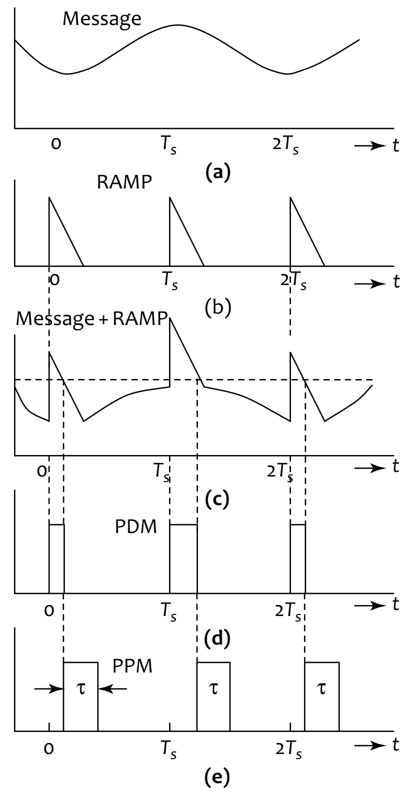


Fig. 10.47 **(a)** *Message Signal* **(b)** *Inverse ramp sequence* **(c)** *x(t) + ramp* **(d)** *PDM* **(e)** *PPM*

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. 10.47, 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. 10.48 gives yet another direct method of generation of PDM and PPM signals.

This circuit is an emitter-coupled one-shot or monostable multivibrator. In its stable state, transistor T_1 is in cutoff condition and T_2 is in conducting state. However, when a trigger pulse of sufficient amplitude

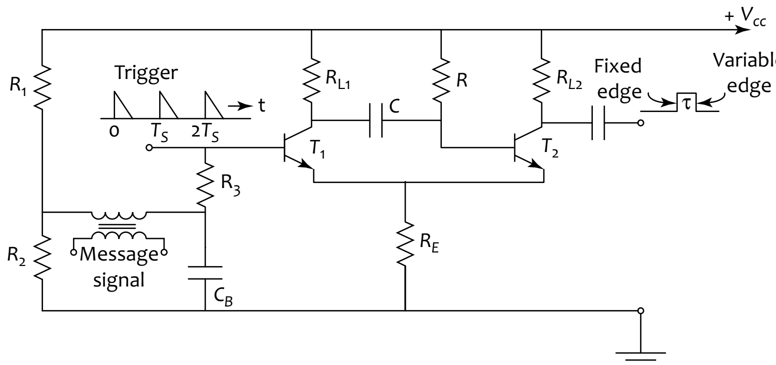


Fig. 10.48 *A circuit for generating PDM (PWM) signals*

is applied to its base, T_1 suddenly goes into conduction and T_2 is temporarily cutoff. 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 monostable multivibrator with a fixed bias for its normally cutoff transistor, one can obtain a PPM signal.

10.10.2 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 lowpass filtering is possible but it results in some distortion.

Apart from direct lowpass 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 lowpass filtering and equalization. Actually, if the pulse width in this PAM signal is quite small, equalization may not be necessary at all; simple lowpass filtering will suffice. This approach is applicable to the detection of PPM signals also. This method is illustrated in Fig. 10.49.

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. 10.49(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. 10.49(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 the leading edges of the PPM pulses. Once the PAM signal is obtained, it can be lowpass 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 lowpass filtering of the PAM signal.

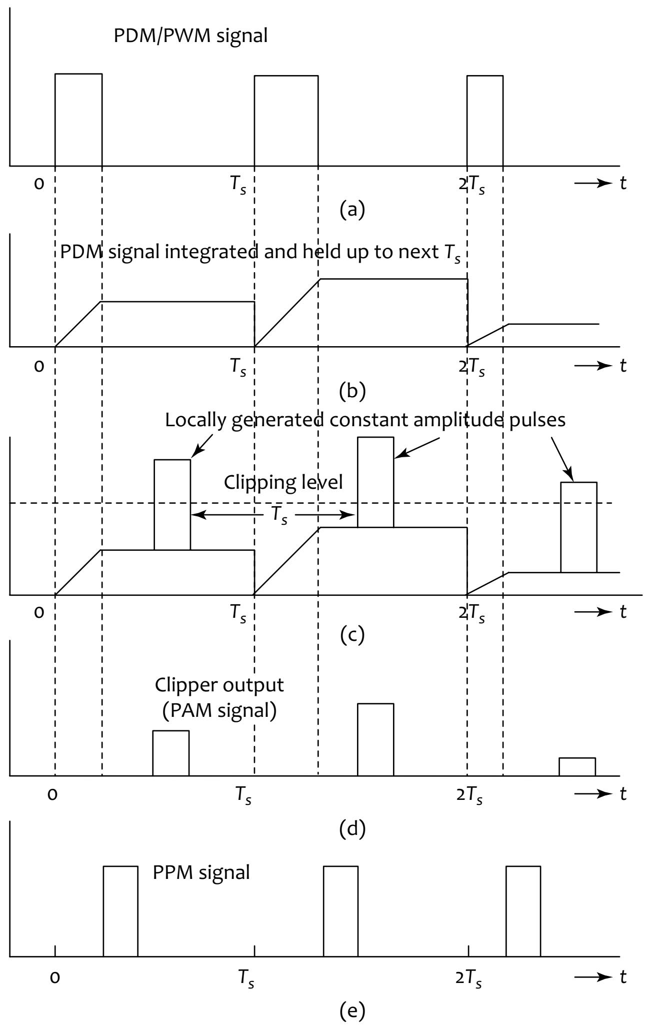


Fig. 10.49 **(a)** *PDM/PWM signal* **(b)** *PDM signal integrated and held up to next sampling instant* **(c)** *Locally generated constant amplitude pulses added on the pedestal* **(d)** *Output of the clipper* **(e)** *PPM signal*

10.10.3 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.

(i) Cross-talk due to High-Frequency Limitation of the Channel

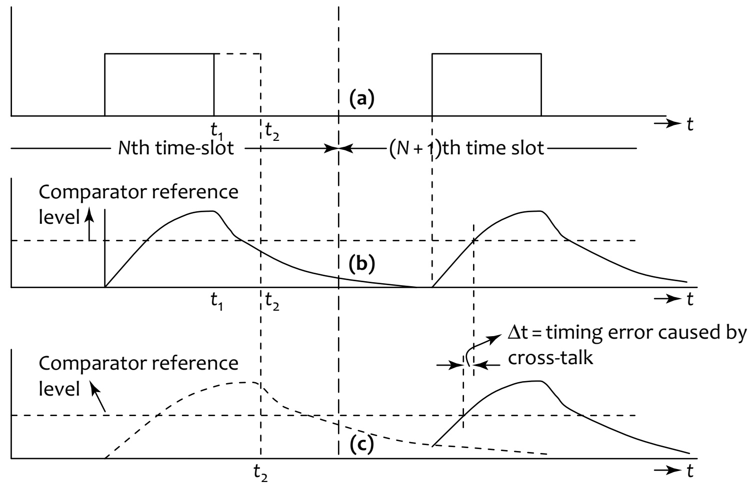


Fig. 10.50 **(a)** *Transmitted PTM signal without modulation* **(b)** *Received signal without modulation* **(c)** *Received signal when modulation is present*

In Fig. 10.49 the transmitted pulses in the N^{th} and $(N + 1)^{\text{th}}$ 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. Fig. 10.50(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 10.50(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.

(ii) Cross-talk due to Low-Frequency Deficiency

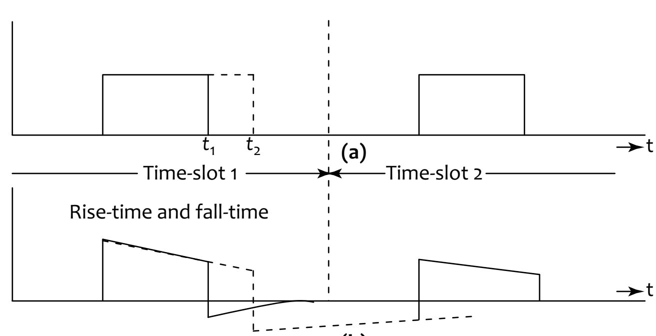


Fig. 10.51 **(a)** *Transmitted pulses (dashed line shows change due to modulation)* **(b)** *Received pulses (assuming excellent high frequency response for channel)*

Figure 10.51(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.

10.10.4 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.

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 a 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.

For PDM, the synch pulse is made to have a much larger width than any of the signal pulses.

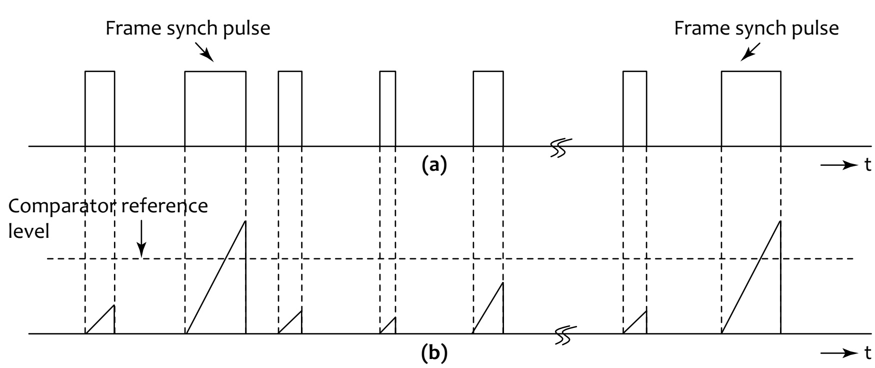


Fig. 10.52 **(a)** *TDM-ed PDM pulses along with one synch pulse per frame* **(b)** *Output of an integrator circuit*

The pulses shown in Fig. 10.52(a) are inverted and as shown in Fig. 10.53, 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 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 cutoff state and so the capacitor 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. 10.52(b). The synch pulses have a very large width and so the corresponding sawtooth pulses will have very large amplitudes compared to the amplitudes of the sawtooth pulses produced by the signal pulses. A comparator, whose reference level is adjusted to be far above the smaller sawtooth pulses, produces an output trigger pulse whenever the large sawtooth 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.

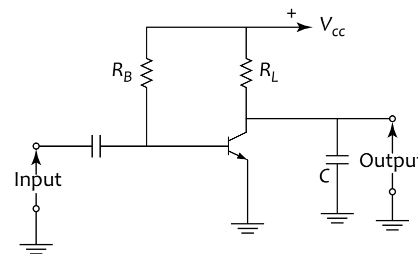


Fig. 10.53 *Sawtooth waveform generating circuit*

10.10.5 Comparison of TDM and FDM

1. TDM hardware is much simpler than that required for FDM, as there is no need for sub-carrier modulators, bandpass filters, etc.
2. In FDM cross-talk occurs mainly due to non-linear cross-modulation and imperfect bandpass 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 10.13

Five lowpass 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?

Since f_s = Sampling frequency = 5×10^3 sps,

$$T_s = \frac{1}{f_s} = \frac{1}{5 \times 10^3} \text{ sec} = 0.2 \text{ ms}$$

Hence, the interval between successive samples of a particular message signal, is 0.2 ms. If 5 such message signals are to be TDM-ed, it means that 5 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. 10.54.

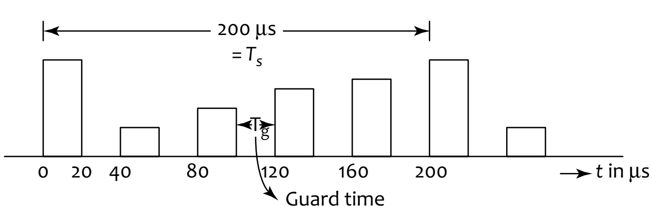


Fig. 10.54 *PAM/TDM-ed signal for Example 10.13*

Therefore, as can be seen from the figure, the guard time between adjacent pulses in the PAM/TDM-ed signal is 20 μs.

Example 10.14

A TDM signal is shown in Fig. 10.55. Show that it is possible to detect it using a time-averaging lowpass filter.

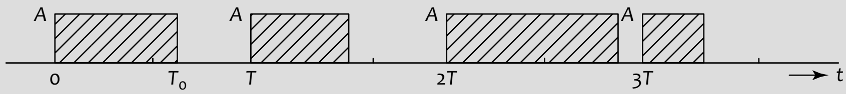


Fig. 10.55 *Signal for Example 10.14*

A time-averaging filter takes the average value over each period T.

$\therefore$ output of the filter at $t = T$ is given by

$$V_0(t) = \frac{1}{T}\int_0^{T_D} A dt + \frac{1}{T}\int_{T_D}^{T} 0 \cdot dt = \frac{AT_D}{T}$$

$\therefore$ $V_0(T)$ is proportional to T_D, the width of the pulse. Hence, a time-averaging filter can be employed to detect a TDM signal.

10.11 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

$$B_T >> \frac{1}{\tau} \tag{10.68}$$

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

$$B_T \geq \frac{1}{2\tau_{\min}}, \tag{10.69}$$

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 seconds. 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

$$B_T = \frac{1}{2\tau} \tag{10.70}$$

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. (10.70), 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

$$B_T \geq \frac{1}{2\tau_{r_{\min}}} \tag{10.71}$$

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.

(a) The pulse-modulated signals (PAM and PTM) that we consider are *baseband signals and have no high frequency carrier*.
(b) Because there is no carrier modulation, the *noise entering the receiving system is lowpass noise and not bandpass noise* as was the case when we considered the noise performance of continuous-wave modulations like AM and FM.
(c) 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.
(d) 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 demodulation, 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. 10.56 for studying the noise performance of analog pulse modulated systems.

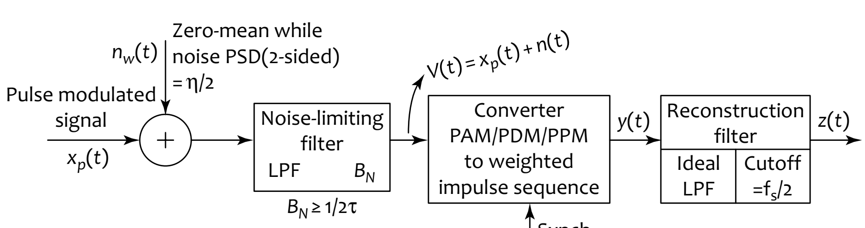


Fig. 10.56 *Model for an analog pulse modulation receiver*

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.

10.11.1 Pulse Amplitude Modulation

Pulse amplitude modulated signal plus white noise tries to enter the receiver. The noise limiting lowpass filter has a cutoff frequency $B_N \geq \dfrac{1}{2\tau}$, where τ is the pulse width of the PAM signal. Its output therefore is

$$v(t) = x_p(t) + n(t) \tag{10.72}$$

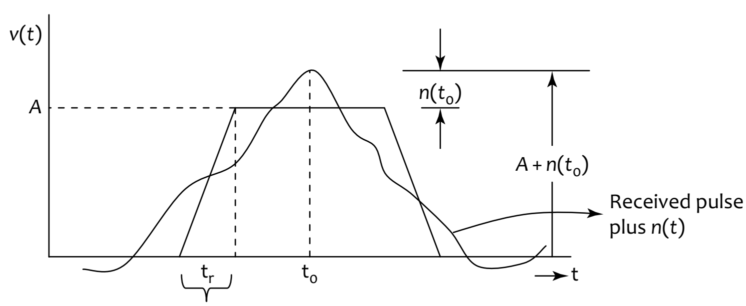


Fig. 10.57 *Received pulse plus lowpass filtered noise pulse*

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

$$v(t)\big|_{t=t_0} = v(t_0) = A + n(t_0) = A + \in_0 \tag{10.73}$$

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

$$\sigma_0^2 = \overline{n^2} = \eta B_N \tag{10.74}$$

The output of the converter, $y(t)$, which is a train of weighted impulses spaced T_s second apart, may therefore be written as

$$y(t) = \sum_k \left[A_c mx(kT_s) + \in_k\right] \delta(t - kT_s) \tag{10.75}$$

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 cutoff frequency of $f_s/2$, a passband 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

$$\begin{aligned} z(t) &= \sum_k \left[A_c mx(kT_s) + \in_k\right] \text{sinc}\left[f_s(t - kT_s)\right] \\ &= \sum_k A_c mx(kT_s) \text{sinc}\left[f_s(t - kT_s)\right] + \sum_k \in_k \text{sinc}\left[f_s(t - kT_s)\right] \end{aligned} \tag{10.76}$$

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

$$z(t) = A_0 mx(t) + n_0(t) \tag{10.77}$$

As shown in Fig. 10.56, $B_N \geq \frac{1}{2\tau}$. Since $\tau << T_s$, it follows that

$$B_N > \frac{1}{T_s} \tag{10.78}$$

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 a 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

$$N_D = \overline{n_D^2(t)} = \overline{\in_k^2} \tag{10.79}$$

But, we have already shown that the variance of the measurement error (see Eq. 10.74) is equal to the average noise power at the output of the noise limiting filter and that this is given by ηB_N.

$\therefore$
$$N_D = \eta B_N \tag{10.80}$$

Now, to determine the average signal power at destination, we proceed as follows.

$$\left.\begin{array}{l}\text{Average energy per pulse in}\\ \text{the PAM signal}\end{array}\right\} = \tau A_0^2 \overline{\left[1 + mx(kT_s)\right]^2} = E_p \tag{10.81}$$

$$\text{Number of pulses per seconds in the PAM signal} = f_s \tag{10.82}$$

$\therefore$
$$\text{received average signal power} = S_R = \tau A_0^2 \overline{\left[1 + mx(kT_s)\right]^2} \cdot f_s \tag{10.83}$$

But,
$$\overline{\left[1 + mx(kT_s)\right]^2} = 1 + m^2 \overline{x^2(t)} \tag{10.84}$$

since $x(t)$ is assumed to be of zero-mean so that

$$m\overline{x(kT_s)} = 0 \tag{10.85}$$

$\therefore$ we may rewrite Eq. (10.83) as follows:

$\therefore$
$$S_R = A_0^2 f_s \tau [1 + m^2 \ \overline{x^2}] \tag{10.86}$$

From Eq. (10.77), average signal power at the destination is given by

$$S_D = A_0^2 m^2 \overline{x^2(t)} \tag{10.87}$$

$\therefore$ using Eq. (10.80) and Eq. (10.87), 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.

$$\left(\frac{S}{N}\right)_{D;\,\max} = \frac{2A_0^2 m^2 \overline{x^2(t)} \cdot \tau}{\eta} \tag{10.88}$$

But, as per Eq. (10.84), 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. (10.88),

$$\left(\frac{S}{N}\right)_{\substack{D;\max\\ \text{PAM}}} = \frac{WA_0^2 m^2 \overline{x^2(t)} \cdot 2(S_R/\eta W)}{A_0^2 f_s \left[1 + m^2 \overline{x^2(t)}\right]} = \left[\frac{m^2 \overline{x^2(t)}}{1 + m^2 \overline{x^2(t)}}\right] \left(\frac{2W}{f_s}\right)\gamma \tag{10.89}$$

This can be further maximized by choosing $m = 1$. Then

$$\left(\frac{S}{N}\right)_{\substack{D;\max\\ \text{PAM}}} = \left[\frac{\overline{x^2}}{1 + \overline{x^2}}\right] \left(\frac{2W}{f_s}\right)\gamma \tag{10.90}$$

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 that 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, it is not for its good noise performance but only for its simplicity and for time division multiplexing.

10.11.2 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.

$$\tau_k = \tau_0 \left[1 + mx(kT_s)\right]$$

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. 10.58.

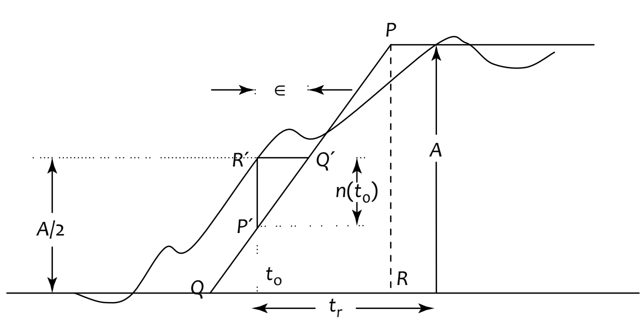


Fig. 10.58 *Received pulse and position error* $\in$

The triangles PQR and $P'Q'R'$ are similar

$$\therefore \qquad (\in / n(t_0)) = (t_r / A) \quad \therefore \quad \in = \left(\frac{t_r}{A}\right) \cdot n(t_0) \tag{10.91}$$

And, $$\sigma^2 = \overline{\epsilon^2} = \left(\frac{t_r}{A}\right)^2 \overline{n^2} = \left(\frac{t_r}{A}\right)^2 \cdot \eta B_N \tag{10.92}$$

But, $t_r \cong \dfrac{1}{2B_N}$ and $A^2 = \dfrac{E_p}{\tau_0}$ for PDM and $\dfrac{E_p}{\tau^2}$ for PPM.

where, E_p = 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

$$\therefore \qquad \sigma^2 = \frac{\eta}{4B_N A^2} = \begin{cases} \dfrac{\eta\tau_0}{4B_N E_p} & \text{for PDM} \\[2ex] \dfrac{\eta\tau}{4B_N E_p} & \text{for PPM} \end{cases} \tag{10.93}$$

$$S_R = E_p \cdot f_s = \begin{cases} A^2\tau_0 f_s & \text{for PDM} \\ A^2\tau\ f_s & \text{for PPM} \end{cases} \tag{10.94}$$

$$S_D = \begin{cases} m^2\tau_0^2\overline{x^2} & \text{for PDM; where } \tau_0 \text{ is unmodulated pulse width} \\ m^2 t_0^2\overline{x^2} & \text{for PPM; where } t_0 \text{ is the proportionality constant for conversion} \\ & \text{from amplitude to time in second/volt} \end{cases}$$

$N_D = \sigma^2$ and is as given by Eq. (10.93)

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, $A^2\tau_0 f_s = S_R$ and $\dfrac{S_R}{\eta W} = \gamma$ and $B_N \approx B_T$

$$\therefore \qquad \left(\frac{S}{N}\right)_D = 4m^2\tau_0 B_T\overline{x^2}\left(\frac{W}{f_s}\right)\gamma \tag{10.95}$$

To maximize $\left(\dfrac{S}{N}\right)_D$, we note that

$$f_{s\,\min} = 2W; \quad \tau_{0\ \max} = \frac{T_{s\,\max}}{2} = \frac{1}{2f_{s\ \min}} = \frac{1}{4W} \quad \text{and} \quad m_{\max} = 1$$

$$\therefore \qquad \left(\frac{S}{N}\right)_D \le \frac{1}{2}\left(\frac{B_T}{W}\right)\overline{x^2}\ \gamma \tag{10.96}$$

For PPM

$$\left(\frac{S}{N}\right)_D = \frac{S_D}{N_D} = \frac{m^2 t_0^2 \overline{x^2}}{\sigma^2} = \frac{m^2 t_0^2 \overline{x^2}\, 4 B_N A^2}{\eta}$$

But, $A^2 \tau\, f_s = S_R$ and $\frac{S_R}{\eta W} = \gamma$. Also, $B_N \approx B_T$

$$\therefore \quad \left(\frac{S}{N}\right)_D = \frac{4 m^2 \overline{x^2} B_T t_0^2 S_R}{\eta \tau f_s} \tag{10.97}$$

Note that $f_{s\,\min} = 2W$; $t_0 \le \frac{T_s}{2}$ so that $t_{0\ \max} = \frac{T_s}{2} = \frac{1}{4W}$

Also, the pulse width, $\tau \ge 2t_r = 1/B_T \quad \therefore \quad \tau_{\min} = \frac{1}{B_T}$

Substituting the above in Eq. (10.96) in order to maximize $(S/N)_D$, we get

$$\left(\frac{S}{N}\right)_{D;\max} = \frac{1}{8} m^2 \overline{x^2} \left(\frac{B_T}{W}\right)^2 \gamma \tag{10.98}$$

i.e.,

$$\left(\frac{S}{N}\right)_D \le \frac{1}{8} m^2 \overline{x^2} \left(\frac{B_T}{W}\right)^2 \gamma \tag{10.99}$$

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. (10.99) 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. (10.96) and (10.99), 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 10.15

A message signal has $\overline{x^2} = 0.1$ and is band-limited to 100 Hz. It is sampled at a rate of 250 sps and converted into a PDM signal with m = 0.2 and an unmodulated pulse width of 80μs, 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.

From Eq. (10.95), we have

$$\left(\frac{S}{N}\right)_D = 4 m^2 \tau_0 B_T \overline{x^2} \left(\frac{W}{f_s}\right)\gamma = 4 \times (0.2)^2 \times 80 \times 10^{-6} \times 0.1 \left(\frac{100}{250}\right)\gamma$$

$$\therefore \quad \gamma = \left(\frac{S}{N}\right)_D \cdot \frac{250 \times 10^6}{4 \times 0.04 \times 80 \times 0.1 \times 100} = \left(\frac{S}{N}\right)_D \cdot \frac{250 \times 10^6}{128} = 1953125 \left(\frac{S}{N}\right)_D$$

But $\left(\frac{S}{N}\right)_D$ should be at least 40 dB, i.e., 10^4. $\therefore$ $\left(\frac{S}{N}\right)_D \geq 10^4$

$$\therefore \quad \gamma \geq 1953125 \times 10^4 \text{. But } \gamma = \frac{S_R}{\eta W} \quad \therefore \quad S_R = \gamma \eta W$$

$$\therefore \quad S_R \geq 1953125 \times 10^4 \times 10^{-12} \times 100 = 1953125 \times 10^{-6} \text{ watts}$$

$$\therefore \quad S_R \geq 1.953125 \text{ watts}$$

10.11.3 Comparison of 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.

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 sub-carrier 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, narrowband fading of the channel may affect at the most one or two FDM channels only.	Slow, narrowband 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.

SUMMARY

1. Statement of lowpass sampling theorem: if $x(t)$ is a lowpass signal, bandlimited 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
$$x(t) = 2BT_s \sum_{k=-\infty}^{\infty} x(kT_s) \operatorname{sinc} 2B(t - kT_s)$$
2. The lowpass sampling theorem provides the basis for all analog pulse modulation systems as well as all digital communication systems.
3. $f_s = 2W$ represents the minimum sampling rate that can be used, for sampling a lowpass signal bandlimited to W Hz, if the signal is to be recovered from its samples. This minimum sampling rate is called 'Nyquist rate'.
4. 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.

5. It is possible to recover $x(t)$, without any distortion, from its samples in the case of ideal sampling as well as natural sampling.
6. 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.
7. ‘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 $x(t)$ re-appearing as low frequency components in the spectrum of the sampled signal.
8. ‘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.
9. A zero-order-hold (ZOH) may be used to reconstruct the message from its samples—it gives a stair-case approximation of the message.
10. 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.
11. A PPM signal may be obtained by directly flat-top sampling an analog signal. It can be detected by making use of a lowpass filter, followed by, if necessary, an equalizer.
12. 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 monostable multivibrator.
13. 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.
14. A PDM signal may be detected either by first converting into a PAM signal and lowpass filtering this PAM signal, or by directly lowpass filtering the PDM signal itself.
15. 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.
16. 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.
17. In TDM-ed transmission of PAM, PPM, or PDM, synchronization of the commutators at the two ends is necessary.
18. 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.
19. As far as noise performance is concerned, PAM is at least 3 dB inferior direct baseband analog message signal transmission.
20. 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 tradeoff. However PPM is inferior to WBFM by about 10 dB.

REFERENCES AND SUGGESTED READING

1. P Ramakrishna Rao: *Signal and Systems*, Tata McGraw-Hill Publishing company Ltd., New Delhi, 2008.
2. Herbert Taub and Donald L Schilling: *Principles of Communication Systems*, International Student Edition, McGraw-Hill Kogakusha, Ltd., Tokyo, 1971.

3. A Bruce Carlson, Paul B Crilly and Janet C Rutledge: *Communication Systems: An introduction to Signals and Noise in Electrical Communication*, Fourth Edition, International Edition, McGraw-Hill, 2002.
4. Leon W Couch, II: *Digital and Analog Communication Systems*, Sixth Edition, Pearson Education Asia, 2001.
5. Simon Haykin: *Communication Systems*, Third Edition, John Wiley & Sons, 1995.

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 lowpass 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 demodulate 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 highpass *R-C* filter.
14. By considering a PAM signal and using a lowpass *R-C* 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.

FILL IN THE BLANKS

1. An analog lowpass signal, $x(t)$, is band-limited to 1500 Hz. The maximum sampling period that can be used in order to recover $x(t)$ from its samples without distortion, is ________.
2. A continuous-time lowpass signal, $x(t)$, is being sampled. If the signal $x(t)$ is to be recovered from the samples, the conditions to be satisfied are (i) ____, and (ii) ______.
3. $x(t)$ is a lowpass signal band-limited to W Hz and having a spectrum $X(f)$. The ideally sampled version, $x_g(t)$, of $x(t)$, will have a spectrum $X_\delta(f)$ which is a periodic repetition of $X(f)$ with a period of ______.
4. The minimum sampling rate f_s to be used for sampling a continuous-time lowpass band-limited signal $x(t)$ for distortion-less recovery of $x(t)$ from its samples, is called _______ and it is equal to _____.
5. A lowpass band-limited signal, sampled at a frequency higher than the Nyquist rate, may be recovered from its samples by passing them through a _______.

6. For a guard band to exist in the spectrum of $x_\delta(t)$, it is necessary that the _____ be _______ than the _______.
7. Aliasing is the name given to the phenomenon where by the _______ frequency components of a signal $x(t)$ reappear as ________ frequency components in the spectrum of the sampled version of $x(t)$.
8. Anti-aliasing filter is a _______ (lowpass/highpass) filter used _______ (before/after) the sampler.
9. The distortion in the recovered signal caused by flat-top sampling, is called _____. It attenuates the ______ frequency components relative to the ______ frequency components.
10. The impulse response of a ZOH circuit has a ____ shape.
11. A PAM signal may be generated using a ______ and _____ circuit.
12. The width of an unmodulated pulse in the case of PDM can at the most be _______.
13. PAM is at least ______dB inferior to direct baseband transmission in so far as $(S/N)_D$ is concerned.
14. _________ (PDM/PPM) offers a better trade-off than ______ (PPM/PDM) between transmitter power and transmission bandwidth.

MULTIPLE CHOICE QUESTIONS

1. A band-limited lowpass signal is sampled at twice its Nyqyist rate with $f_s = 2000$ sps. The signal is bandlimited to
   (a) 250 Hz (b) 1000 Hz (c) 500 Hz (d) 2000 Hz
2. A certain lowpass signal $x(t)$ is sampled and the spectrum of the sampled version has guardband from 1500 Hz to 1900 Hz. The sampling frequency used is
   (a) 1500 sps (b) 1900 sps (c) 1700 sps (d) 3400 sps
3. A lowpass signal bandlimited to 1200 Hz was sampled and it was found that the 1000 Hz frequency component was re-appearing in the recovered signal, because of aliasing, as 400 Hz component. The sampling frequency used is
   (a) 1400 sps (b) 1600 sps (c) 2200 sps (d) 800 sps
4. $x(t) = 3\cos^2 250\pi t$. This signal is sampled at regular intervals of T seconds. The maximum value of T for which $x(t)$ may be recovered from the sampled version without any distortion, is equal to
   (a) 1 ms (b) 2 ms (c) 4 ms (d) 0.5 ms
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 sps (b) 350 sps (c) 130 sps (d) 65 sps
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 an 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) periodic version of $x(t)$ with period of T seconds
   (c) 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 the above
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 lowpass filter (b) a differentiator followed by a lowpass filter
(c) an integrator (d) a lowpass 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 C.W. 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 a FDM system
(b) affects all the message channels of a TDM system
(c) affects all the message channels of a FDM system
(d) does not have much effect on both TDM and FDM systems

PROBLEMS

1. Determine the Nyquist rate of sampling for the following signals.
(a) $x(t) = 10 \text{ sinc } 100t$ (b) $x(t) = 10\cos^2(100\pi t)$ (c) $x(t) = 10 \text{ sinc}^2(100t)$
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 seconds 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. P-10.1

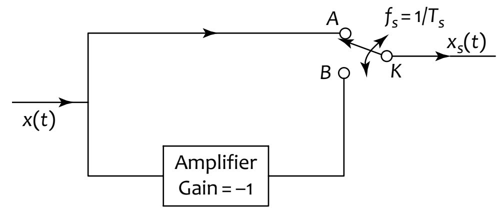


Fig. P-10.1

   (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 $T_s/2$ seconds.
   (b) If the spectrum of $x(t)$ is as shown in Fig. P-10.2.
      (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 $x_s(t)$.

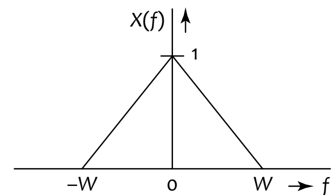


Fig. P-10.2

7. For the lowpass sampling theorem of Section 10.2, 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

$$x_p(t) = \sum_k A_0 \left[1 + mx(kT_s)\right] p(t - kT_s)$$

   (a) Show that its spectrum is given by

$$X_p(f) = A_0 f_s P(f) \left[\sum_k \left\{ \delta(f - nf_s) + mx(f - nf_s) \right\} \right]$$

   (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. Is it possible to detect a PAM signal using a product demodulator? If your answer is in the affirmative, give details of the local oscillator frequency and the cutoff frequency of the LPF.
10. 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.
11. Fifteen 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.
12. Ten 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 (i) PAM/AM with 25% duty cycle (ii) PAM/FM with baseband filtering and a peak frequency deviation of ± 75 kHz.
13. Twenty-five voice channels, each sampled at 8 kHz, are transmitted via PPM/TDM. If $\tau_0 = \tau = 1$ micro second, determine the channel bandwidth required to keep the cross-talk at –40 dB or less.

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)

Appendix–A

MATLAB PROGRAMS

The following functions are used in the programs.

```
%FFTSEQ Function 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.
%
function [M,m,df] = fftseq(m,ts,df)
%   [M,m,df] = fftseq(m,ts,df)
%   [M,m,df] = fftseq(m,ts,df)
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


%
% FSERIES Returns the Fourier series coefficients.
% funfcn s the defined function.
% It would depend on three parameters p1, p2, p3.
% The function is given over one period extending from 'a' to 'b'.
% 'fsc' is the vector of length n+1 of Fourier series coefficients
% xx0,xx1,..xxn.
% 'tol' is the error level.
%
function xx = fseries(funfcn,a,b,n,tol,p1,p2,p3)
j = sqrt(-1);
args0 = [];
for nn = 1:nargin-5
    args0 =[args0,',p',int2str(nn)]
end
args =[args0,')']
a
b
t = b-a;
xx(1) = eval(['1/(',num2str(t),').*quad(funfcn,a,b,tol',args]);
```

```
for i = 1:n
    new_fun = 'exp_fnct';
    args=[',',num2str(i),',',num2str(t),args0,')'];
    xx(i+1) = eval(['1/)',num2str(t),').*quad(new_fun,a,b,tol',args]);
end
```

```
% '
% 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.
%
function [gv1 gv2] = gengauss(m,sgma)
% [gv1,grv2] = gengauss(m,sgma)
% [gv1,grv2] = gengauss(sgma)
% [v1,gsrv2] = gengauss
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);
```

```
%
% [Nx] = Nx_ext(X,M)
% NX_EST estimates the autocorrelation of the sequence of random
% variables given in X, only Nx(0), Nx(1)...Nx(M) are computed.
% Note that Nx(m) acutally means Nx(m-1).
%
function [Nx] = Nx_ext(X,M)
N = length (X)
Nx = zeros(1,M+1);
for m = 1:M+1
    for n = 1:N-m+1
        Nx(m) = Nx(m)+X(n)*X(n+m-1);
    end
    Nx(m) = Nx(m)/(N-m+1)
end
```

Problem 1

A periodic signal x(t) with period $T_0 = 6$ is defined by x(t) = II(t/2) for abs t < = 3. This signal is passed through an LTI system with an impulse response given by

$$h(t) = \{e^{-t/2} \quad \text{for} \quad 0 \le t \le 4$$

$$= 0 \qquad \text{otherwise}$$

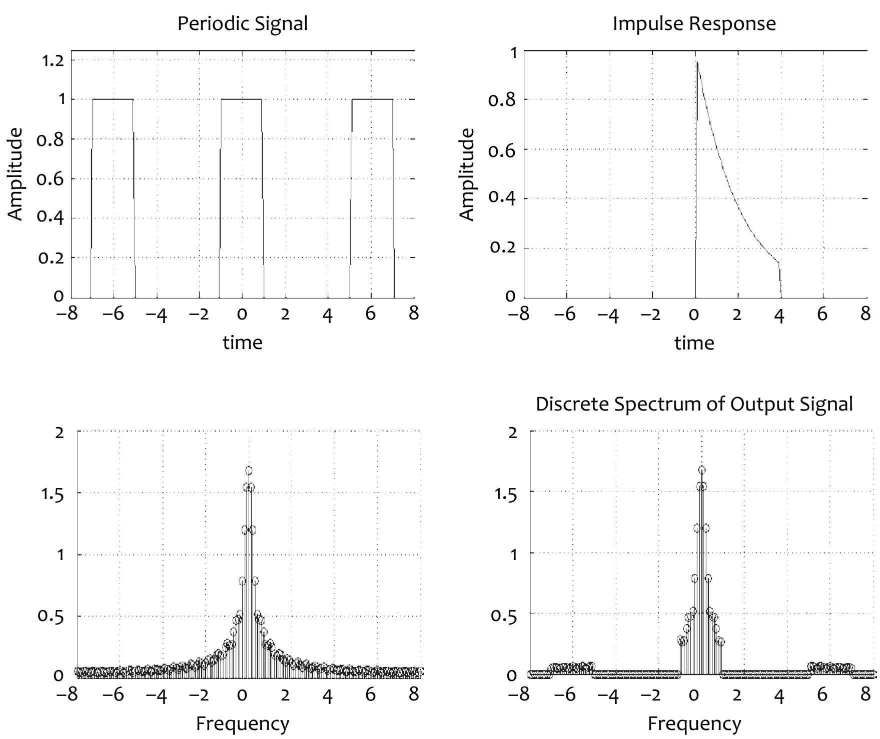
Determine the discrete spectrum of the output signal numerically using MATLAB.

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');
```

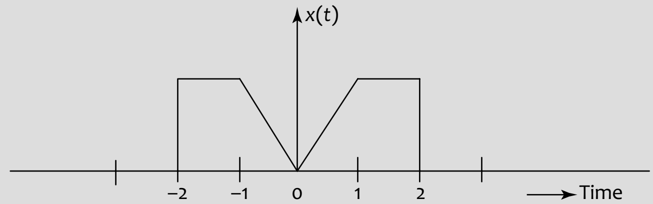
```
%
% Transfer function
%
H = fft(h)/fs;               % frequency resolution
f = [0:df:fs];
H1 = fftshift(H)             % rearrange H
subplot (2,2,3)
stem (t,abs(H1))
xlabel ('Frequency')
grid on
y = x.*H1
subplot (2,2,4)
stem (t,abs(y))
grid on
xlabel ('Frequency');
title ('discrete spectrum of output signal');
```

Results



Problem 2

Plot the magnitude spectrum and phase spectrum of the nonperiodic signal shown in the figure.

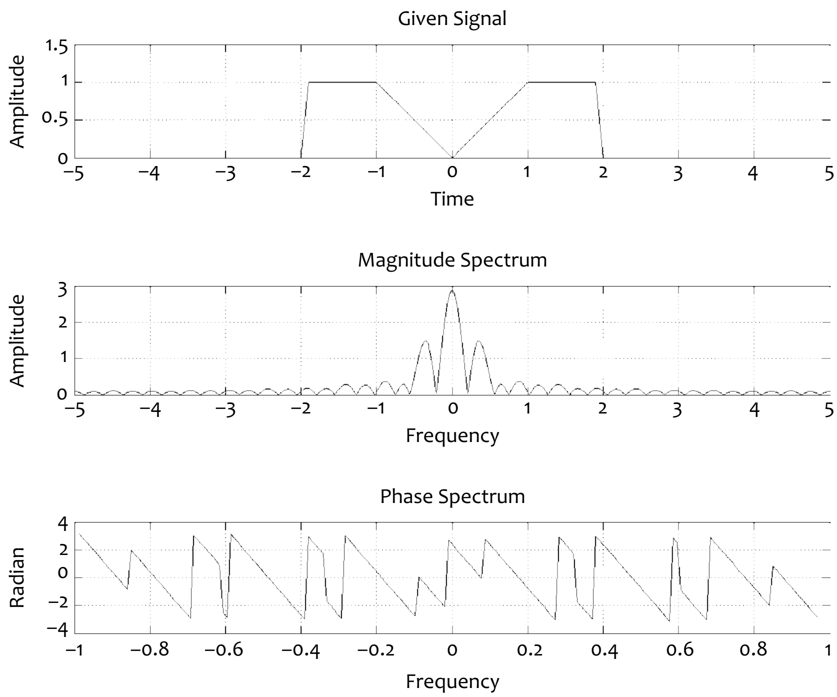


MATLAB Program

```
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



Problem 3

*Determine the spectra of the message signal m(t) and the amplitude-modulated signal xc(t) (AM with carrier + both side bands) and plot them. Plot also the waveform of the message signal. Carrier signal is cos (2*π*250*t) and modulation index m = 0.85. The message signal is a sinusoidal signal of 6.67 Hz.*

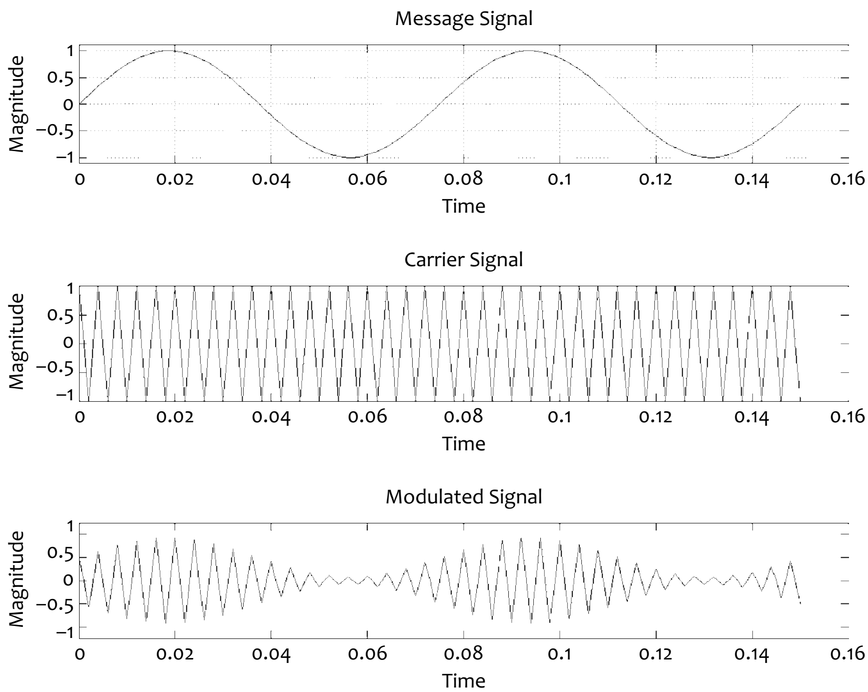
MATLAB program

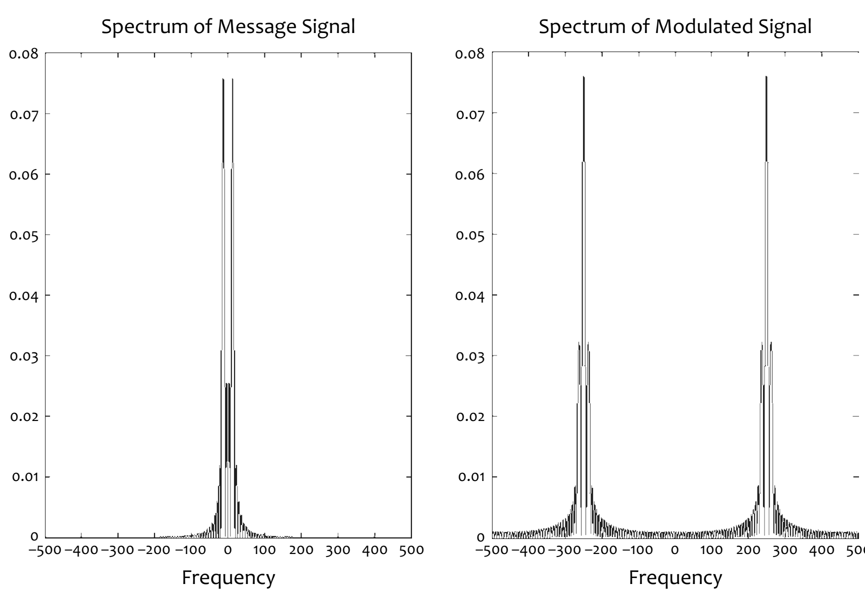
```
%
% Amplitude modulation
%
t0 = 0.15;           % signal duration
f = 1/0.15;
ts = 0.001;          % sampling interval
fc = 250;            % carrier frequency
fs = 1/ts;           % sampling frequency
t = [0:ts:t0];       % time vector
```

```
a = 0.85;        % modulation index
df = 0.5;        % required freqency 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');
```

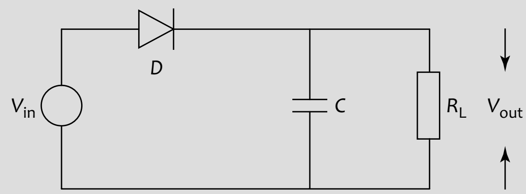
Results





Problem 4

Using MATLAB 9 generate an amplitude-modulated wave and detect it using the simple envelop detector shown in the figure:



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)*

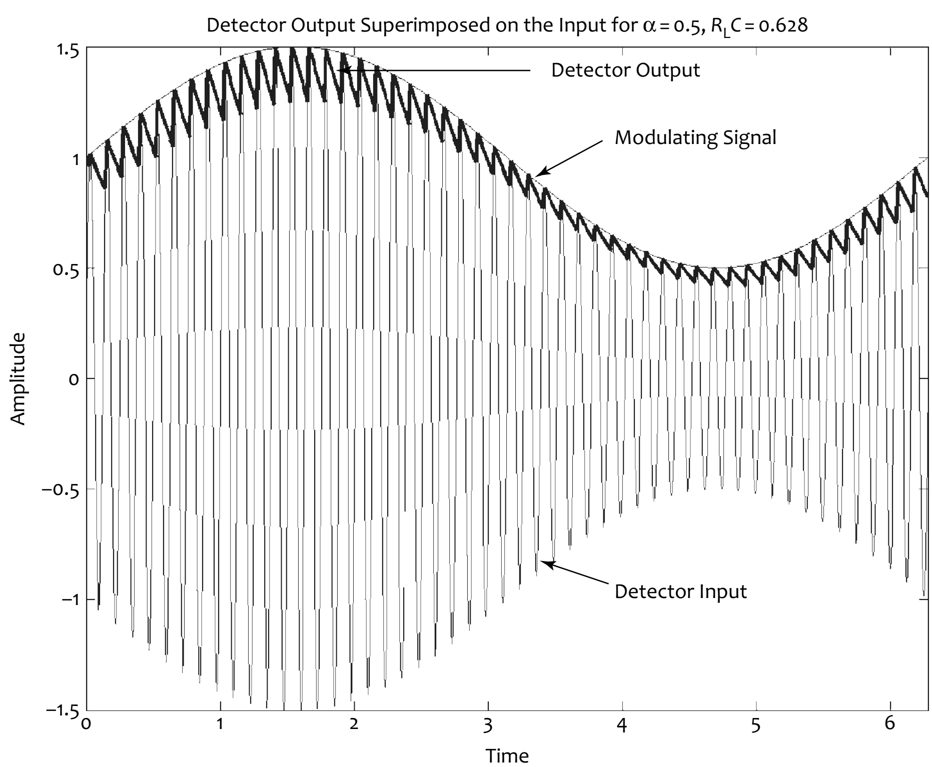
MATLAB Program

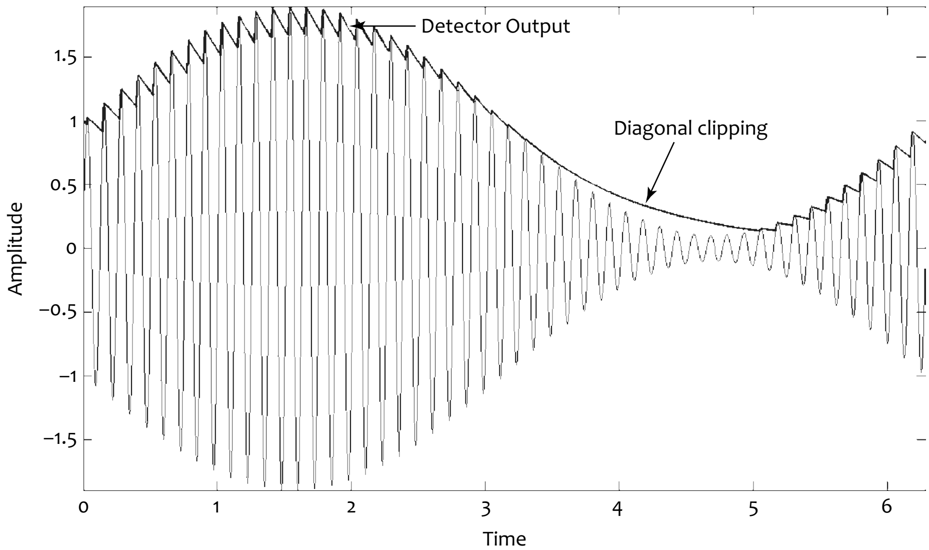
```
%
% Envelope.m detects AM waveform
%
% Part 1 of the problem
RL C = 2*pi/10;
alpha = 0.5;
Dt = 2*pi/1000;
W = 50;
global RLC, alpha, W, Dt;
t = 0:2*pi/1000:2*pi;
%
% Allocation of memory for input and output arrays
%
Vin = zeros(1,1001);
Vout = zeros(1,1001);
%
% Define input array
%
V = 1+alpha*sin(t); % 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/R_L 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





Problem 5

(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:

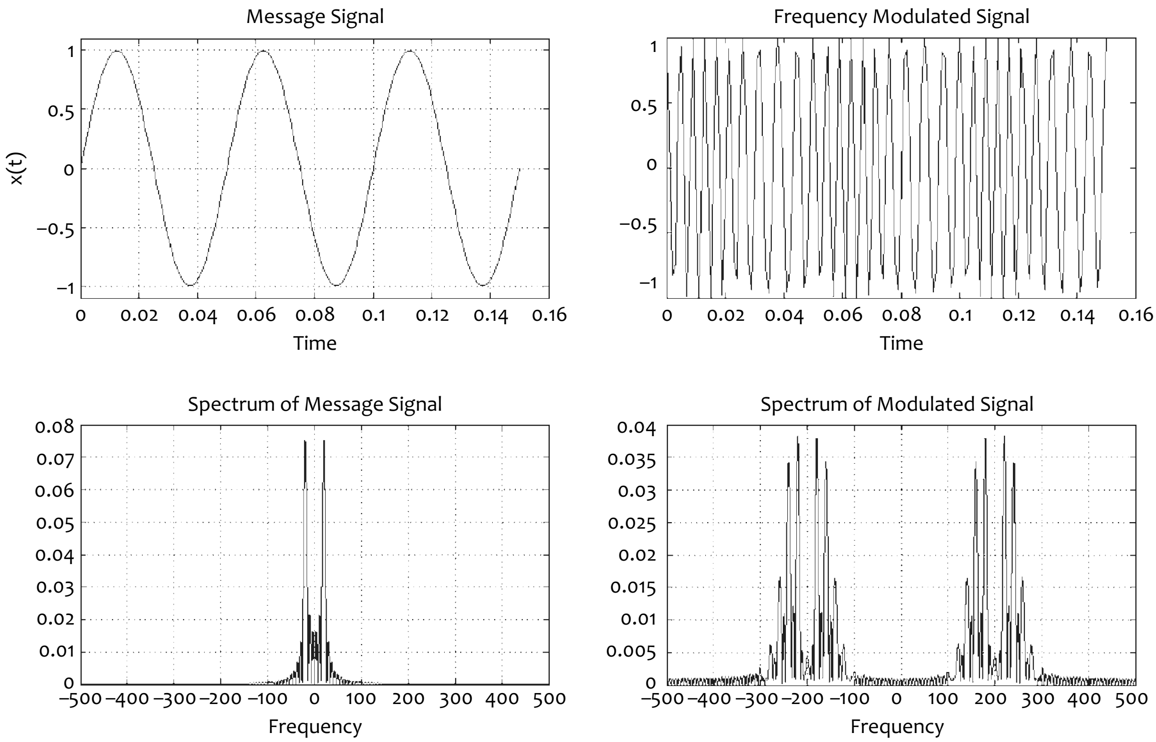
(a) Plot the message signal.
(b) Plot the modulated signal.
(c) Determine and plot the spectrum of the message signal.
(d) Plot the spectrum of the frequency-modulated signal.

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



Problem 6

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.

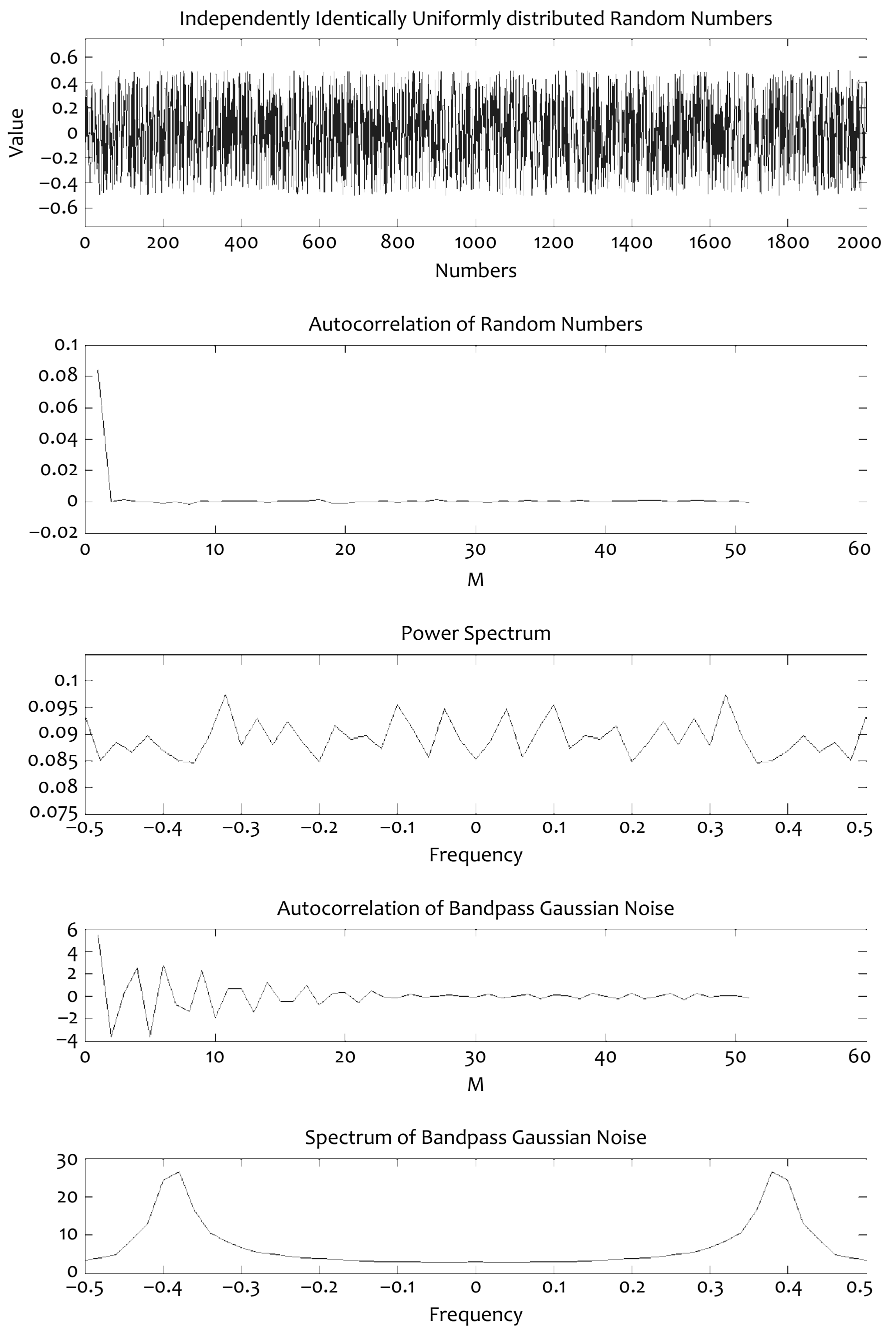
MATLAB Program

```
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');
```

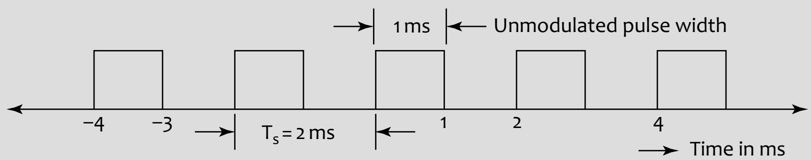
Results



Problem 7

(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 below:



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.

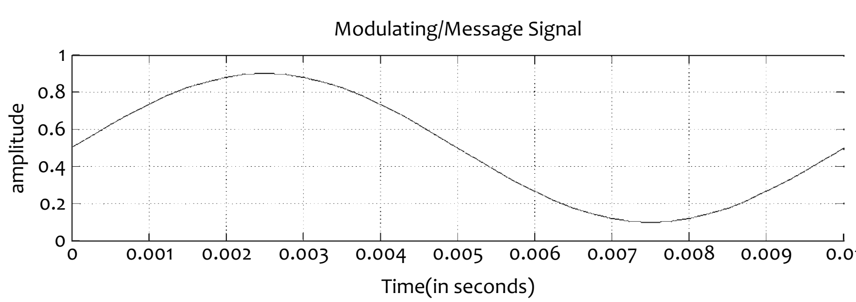
MATLAB Program

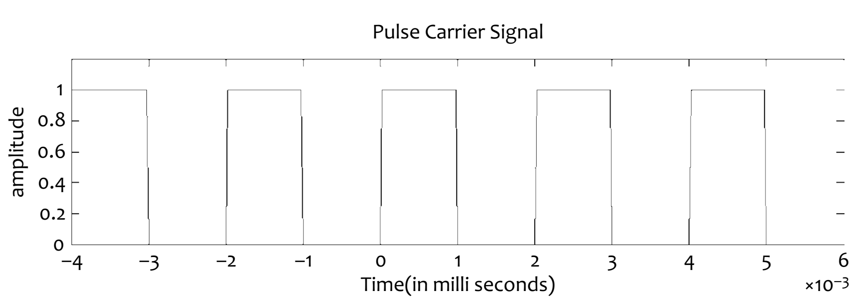
```
clc
Fc = 100;            % modulating signal frequency
Fs = 4000;           % sampling frequency
ts = 1/Fs            % sampling time 0.0025 (0.25 milli seconds)
t = [0:ts:1/Fc];     % a total 4500 samples for 9 milli seconds (500 samples per millisecond)
size(t)
%
% Generation of message wave
%
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)
```

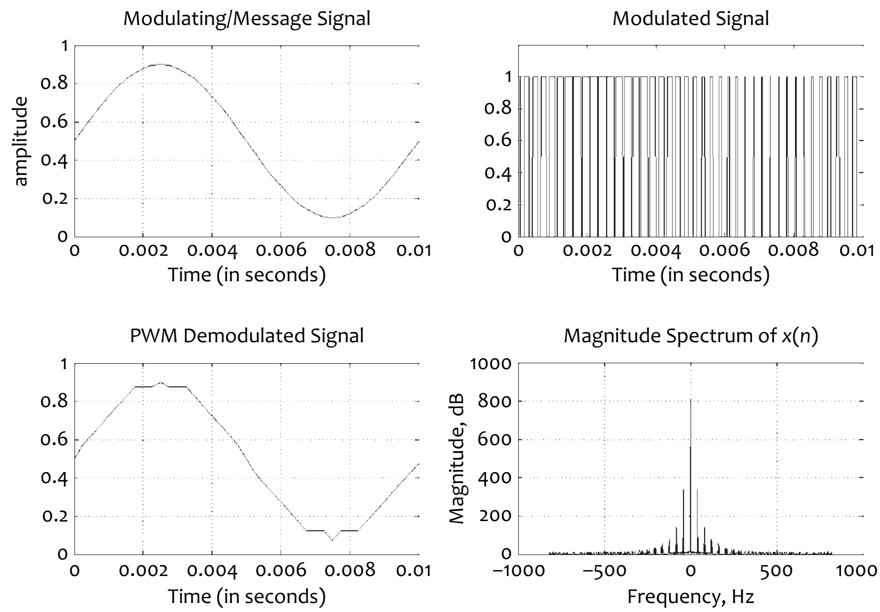
```
title('Pulse position Modulated Signal');
ylim ([0 1.2])
subplot (2,2,1)
plot(t,x)
grid on
ylabel ('amplitude');
xlabel ('time(in secs)')
title ('modulating/message signal');
%
% Demodulated signal of PPM
%
m2 = demod(y1,Fc,Fs,'ppm');
subplot (2,2,3)
plot (t,m2);
grid on
ylabel ('amplitude');
xlabel ('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

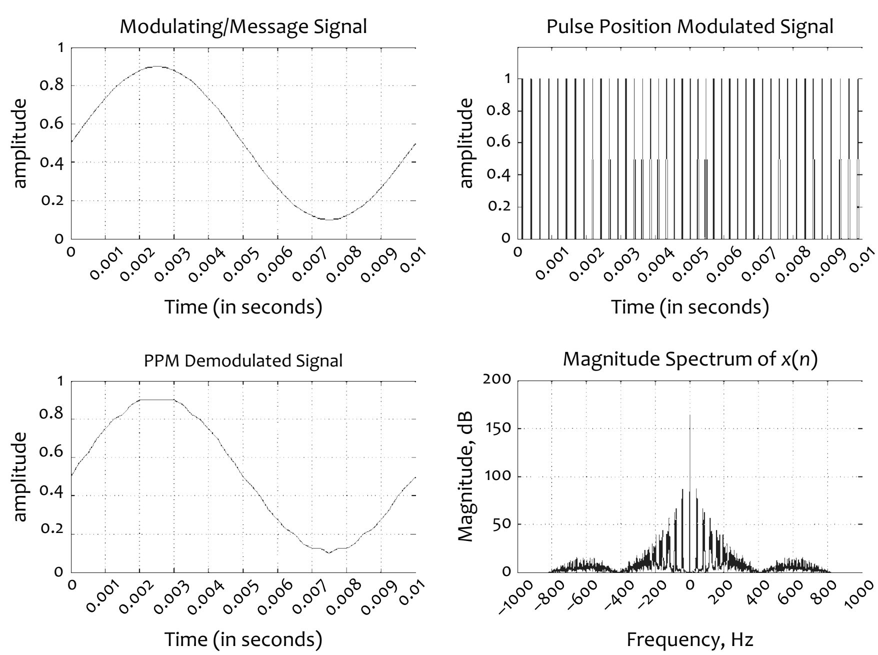




Pulse Width Modulation



Pulse Position Modulation



Appendix–B

MATHEMATICAL FORMULAE

Cramer's Method of Solving a System of Linear Equations

Let
$$a_1x + b_1y = c_1$$
$$a_2x + b_2y = c_2$$

Then
$$x = \frac{\Delta x}{\Delta} \quad \text{and} \quad y = \frac{\Delta y}{\Delta}$$

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
$$\Delta_y = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = a_1c_2 - a_2c_1$$

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
$$t_n = a.r^{n-1}$$

(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 $= \dfrac{a}{(1-r)}$ if $|r| < 1$

Series Expansion

(i) $e^x = 1 + \dfrac{x}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + ...$

(ii) $\cos\theta = 1 - \dfrac{\theta^2}{2!} + \dfrac{\theta^4}{4!} + ...$

(iii) $\sin\theta = \theta - \dfrac{\theta^3}{3!} + \dfrac{\theta^5}{5!} + ...$

(iv) $\tan\theta = \theta + \dfrac{1}{3}\theta^3 + \dfrac{2}{15}\theta^5 + ...$

(v) $\sin^{-1}\theta = \theta + \dfrac{1}{6}\theta^3 + \dfrac{3}{40}\theta^5 + ...$

(vi) $\tan^{-1}\theta = \theta - \dfrac{1}{3}\theta^3 + \dfrac{1}{5}\theta^5 + ...|\theta| < 1$

(vii) $\operatorname{sinc} x = 1 - \dfrac{1}{3!}(\pi x)^2 + \dfrac{1}{5}(\pi x)^5 - .....$

(viii) $\log(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} + ...$ if $|x| < 1$

Some Useful Limits

(i) $\lim\limits_{\theta \to 0} \cos\theta = 1$

(ii) $\lim\limits_{\theta \to 0} \sin\theta = 0$

(iii) $\lim\limits_{\theta \to 0} \dfrac{\sin\theta}{\theta} = 1$ (θ in radians)

(iv) $\lim\limits_{x \to a} \dfrac{x^n - a^n}{x - a} = n.a^{n-1}$

(v) $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = \lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)} = \lim\limits_{x \to a} \dfrac{f''(x)}{g''(x)} = ...$

Differentiation

(i) $\frac{d}{dx}(x^n) = n.x^{n-1}$

(ii) $\frac{d}{dx}(e^x) = e^x$

(iii) $\frac{d}{dx}(a^x) = a^x \log a$

(iv) $\frac{d}{dx}(\sin x) = \cos x$

(v) $\frac{d}{dx}(\cos x) = -\sin x$

(vi) $\frac{d}{dx}(\tan x) = \sec^2 x$

(vii) $\frac{d}{dx}(\cot x) = -\text{cosec}^2 x$

(viii) $\frac{d}{dx}(\sec x) = \sec x \tan x$

(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: $\frac{d}{dx} f(y) = \left[\frac{d}{dy} f(y)\right] . \frac{dy}{dx}$

Integration

(i) $\int e^x dx = e^x + c$

(ii) $\int a^x dx = \frac{a^x}{\log a} + c$

(iii) $\int_{n \neq -1} x^n \, dx = \frac{x^{n+1}}{n+1} + c$

(iv) $\int x^{-1} dx = \log x + c$

(v) $\int K dx = Kx + c$

(vi) $\int \sin x dx = -\cos x + c$

(vii) $\int \cos x dx = \sin x + c$

(viii) $\int \tan x dx = -\log \cos x + c$

(ix) $\int \sec x dx = \log(\sec x + \tan x) + c$

(x) $\int \text{cosec}\, x dx = \log(\text{cosec}\, x - \cot x) + c$

(xi) $\int \cot x dx = \log \sin x + c$

(xii) $\int \sec^2 x dx = \tan x + c$

(xiii) $\int \text{cosec}^2 x dx = -\cot x + c$

(xiv) $\int \sec x \tan x dx = \sec x + c$

(xv) $\int \text{cosec}\, x . \cot x dx = -\text{cosec}\, x + c$

(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) $\int x\sin(ax)dx = \frac{1}{a^2}\left[\sin(ax) - ax\cos(ax)\right] + c$

(xxiv) $\int x\cos(ax)dx = \frac{1}{a^2}\left[\cos(ax) + ax\sin(ax)\right] + c$

(xxv) $\int xe^{ax}dx = \frac{1}{a^2}e^{ax}(ax-1) + c$

(xxvi) $\int xe^{ax^2}dx = \frac{1}{2a}e^{ax^2} + c$

Some Useful Definite Integrals

(i) $\int_0^\infty e^{-ax^2}dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}$; $a > 0$

(ii) $\int_0^\infty x^2e^{-ax^2}dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}}$; $a > 0$

(iii) $\int_0^\infty \mathrm{sinc}x\,dx = \int_0^\infty \mathrm{sinc}^2x\,dx = \frac{1}{2}$

(iv) $\int_0^\infty \frac{x\sin(ax)}{(b^2+x^2)}dx = \frac{\pi}{2}e^{-ab}$; a and $b > 0$

(v) $\int_0^\infty \frac{\cos(ax)}{(b^2+x^2)}dx = \frac{\pi}{2b}e^{-ab}$; a and $b > 0$

(vi) $\int_0^\infty e^{-ax}\cos(bx)dx = \frac{a}{a^2+b^2}$; $a > 0$

(vii) $\int_0^\infty e^{-ax}\sin(bx)dx = \frac{b}{a^2+b^2}$; $a > 0$

(viii) $\int_0^\infty e^{\pm j2\pi yx}dx = \delta(y)$

(ix) $\int_0^\infty \frac{x^{m-1}}{1+x^n}dx = \frac{\pi/n}{\sin(m\pi/n)}$; $n > m > 0$

Trigonometric Identities

(i) $e^{\pm jx} = \cos x \pm j\sin x$

(ii) $\cos x = \frac{e^{jx}+e^{-jx}}{2}$

(iii) $\sin x = \frac{1}{2j}(e^{jx} - e^{-jx})$

(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) $\cos^2 x = \frac{1}{2}(1+\cos 2x)$

(ix) $\sin^2 x = \frac{1}{2}(1-\cos 2x)$

(x) $\cos 3x = 4\cos^3 x - 3\cos x$

(xi) $\sin 3x = 3\sin x - 4\sin^3 x$

Appendix–C

VALUES OF USEFUL MATHEMATICAL AND PHYSICAL CONSTANTS

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	$\log_{10} 2 = 0.30103$

Physical Constants

Boltzmann's constant	$k = 1.38 \times 10^{-23}$ Joule/degree Kelvin
Planck'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

$$\left.\begin{array}{l}\text{Thermal energy } kT_0 \text{ at} \\ \text{standard room temperature of} \\ T_0 = 273^\circ\text{K}\end{array}\right\} \quad kT_0 = 3.77 \times 10^{-21} \text{ Joule}$$

Appendix–D

HILBERT TRANSFORM PAIRS

	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.	$\delta(t)$	$(1/\pi t)$

Appendix–E

FOURIER TRANSFORM

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	$x(t) = \text{sgn}(t)$	$X(f) = \frac{1}{j\pi f}$

S.No.	Signal in time domain	Signal in frequency domain
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	$x(t) = \text{sinc}^2 2Wt$	$X(f) = \frac{1}{2W}\Lambda(f / 2W)$

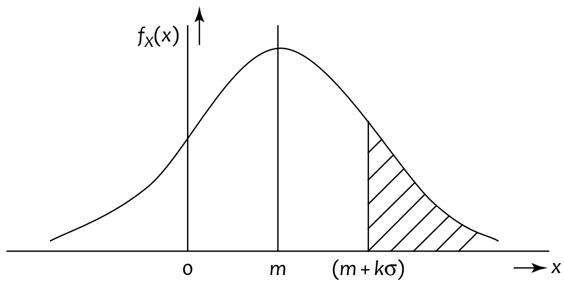
Useful Theorems

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	$x(at)$	$\frac{1}{\|a\|}X(f / a)$
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	$x(t) * y(t)$	$X(f).Y(f)$
Multiplication	$x(t).y(t)$	$X(f) * Y(f)$
Parseval's or Rayleigh's theorem	$\int_{-\infty}^{\infty} \|x(t)\|^2 dt = E_x = \int_{-\infty}^{\infty} \|X(f)\|^2 df$	
Generalized Parseval's theorem	$\int_{-\infty}^{\infty} x(t)\overline{y(t)}dt = \int_{-\infty}^{\infty} X(f)\overline{Y(f)}\, df$	

Appendix–F

ERROR FUNCTIONS AND Q-FUNCTIONS

If X is Gaussian with mean m and variance σ^2,



$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/\sigma^2}$$

Probabiltiy of X taking a value greater than $(m + k\sigma)$ is the area under the shaded region and is given by

$$Q(k) \Delta \frac{1}{\sqrt{2\pi}} \int_k^\infty e^{-\lambda^2/2} d\lambda,$$

where $Q(.)$ is called the Q-function.

The error function and complementary error function are defined as follows:

$$\text{erf}(k) \Delta \frac{2}{\sqrt{\pi}} \int_0^k e^{-\lambda^2} d\lambda = 1 - 2Q(\sqrt{2}k)$$

and $$\text{erfc}(k) \Delta \frac{2}{\sqrt{\pi}} \int_k^\infty e^{-\lambda^2} d\lambda = 1 - \text{erf}(k) = 2Q(\sqrt{2}k)$$

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)*	*k*	*erf(k)*	*k*	*erf(k)*	*k*	*erf(k)*	*k*	*erf(k)*
0.00	0.0000	0.35	0.37938	0.70	0.67780	1.05	0.86244	1.40	0.95229	1.75	0.98667	3.00	0.99998
0.05	0.05637	0.40	0.42839	0.75	0.71116	1.10	0.88021	1.45	0.95970	1.80	0.98909		
0.10	0.11246	0.45	0.47548	0.80	0.74210	1.15	0.89612	1.50	0.96611	1.85	0.99111		
0.15	0.16800	0.50	0.52050	0.85	0.77067	1.20	0.91031	1.55	0.97162	1.90	0.99279		
0.20	0.22270	0.55	0.56332	0.90	0.79691	1.25	0.92290	1.60	0.97635	1.95	0.99418		
0.25	0.27633	0.60	0.60386	0.95	0.82089	1.30	0.93401	1.65	0.98083	2.00	0.99532		
0.30	0.32863	0.65	0.64203	1.00	0.84270	1.35	0.94376	1.70	0.98379	2.50	0.99959		

Index

A

Acoustic channels 5, 12
Acoustic signal 14
Additive Gaussian channel 11
Adjacent channel interference 262, 266
Adjacent channel selectivity 262, 266, 275, 279, 286
Adjacent channel 260
Aliasing 424, 425, 431, 460, 461
Alignment 272
AM superheterodyne receiver 265
AM systems 377
AM transmitter 258
Amplitude modulated signal 129, 138
Amplitude modulation 3, 124
Amplitude probability density function 220
Amplitude spectrum 25, 68
Amplitude-limiting 284
Analog multiplier 241
Analog pulse modulation 453, 459
Analytic signal 69, 76, 80, 81
Angle modulated signals 228
Angle modulation 124, 199
Antenna 5, 123
Anti-aliasing filter 425, 434
Aperture effect 433, 447, 460, 461
Armstrong method 213, 227, 250, 282
Artificial antenna 275
Atmospheric noise 336, 365
Attenuation 108
Audio frequencies 129, 222
Audio voltage amplifier 262
Auto co-variance 327
Auto-correlation function 14, 60, 314, 330, 344
Auto-correlation 39, 40, 78, 79, 312, 319, 327
Auto-covariance 312, 330
Automatic frequency control 232
Automatic gain control (AGC) 264
Available noise power 343, 344, 356, 367, 372
Available output noise power 355
Average power 130
Average signal power 372

B

Balanced discriminator 237, 247
Balanced modulator 160, 167, 277, 280
Balanced zero-crossing detector 243
Band-limited processes 321
Band-limiting frequency 419
Bandpass filter 109, 138
Bandpass noise 389, 391
Bandpass processes 323
Bandpass signal 70, 71, 80
Band-rejection filter 109
Band-stop 109
Bandwidth limited 11
Bandwidth 106, 130
Baseband signal 440
Baseband system 372
Baseband transmission system 409
Basic group 182
Bayes' theorem 295, 324
Bernoulli random variable 301, 325
Bernoulli trials 301
Bessel function of the first kind 215
Bessel function 251
Bessel 435
BIBO criterion 100
BIBO stability criterion 101, 116
BIBO stability theorem 116
BIBO stability 102
Binomial random variable 301, 325
Boltzmann's constant 337, 365
Borel field 324
Bounded signal 116
Bounded-input 100

Bounded-output 100
Butterworth filter 351
Butterworth 435

C

Canonical representation 328
Capture effect 247, 409
Carrier frequency stability 234, 261
Carrier 3, 124
Carrier-to-noise power ratio 379
Carrier-to-noise ratio (CNR) 382
Carson's rule 207, 220, 222, 224, 250
Causal system 95, 116
Causality 95
Cellular mobile communication 2
Central limit theorem 309, 327, 329, 345
Centre-tuned discriminator 238
Channel attenuation 389
Channel signal-to-noise ratio 373
Channel 3, 146
Characteristic function of a random variable 305
Characteristic function 326
Chebyshev 435
Cladding 5
Class-A amplifier 144
Class-AB tuned power amplifiers 261
Class-C amplifier 141, 228
Class-C power amplifiers 260
Co-axial cables 5
Coherent detection 167, 175, 192, 278
Coherent detector 163, 178, 191
Coherent/synchronous detection 147
Collector-modulated class-C amplifier 141
Colpitt's oscillator 233
Communication receivers 279
Commutator 440, 450
Compact disk 435
Complete set of orthogonal functions 22, 23
Complete set 20
Complex gain 104
Complex-envelope 70, 71, 72, 73, 80, 81
Complex-exponential Fourier series 24, 84, 218
Complex-exponential 103
Conditional mean 308
Conditional probability 294
Conditional variance 308
Continuous r.v.s 325
Continuous random process 311
Continuous random variable 297, 298
Continuous-time signals 14, 15, 93, 115
Continuous-time systems 115
Convolution integral 36, 97
Convolution theorem 29, 50, 53, 421
Convolution 14, 35, 36, 59, 78, 81, 108
Core 5
Correlation coefficient 308
Correlation 35, 37, 59, 78, 81
Costas loop 164, 165
Coupling capacitor 148
Covariance matrix 309
Cross-correlation 38, 59, 60, 87, 312, 316, 318, 319, 327
Cross-covariance 312, 327
Cross-power spectrum 319
Cross-talk factor 442
Cross-talk in PAM 441
Cross-talk in PTM 448
Cross-talk 439, 442, 460
Crystal filters 274, 278
Crystal oscillators 228
Cumulative distribution function (CDF) 297, 306, 324

D

De-commutator 439, 450, 451
De-emphasis filter 397, 398, 400
De-emphasis 234, 284, 287, 397, 410, 411
Delayed AGC 265
Delay-line 241
Delta-function 36
Demodulation 146
Depth of modulation 126
Destination signal-to-noise ratio $(S/N)_D$ 371
Destination signal-to-noise ratio 394, 396
Destination SNR 380, 385, 395, 396, 411
Detection of DSB-SC signals 163
Detection of VSB signals 189
Detection 146
Detector 130, 135
Deterministic signals 17
Deviation ratio 206, 250, 326, 395, 401
Deviation 326
Diagonal clipping 150, 192, 264
Differential amplifier 139
Differential equation 98, 113
Differentiation theorem 29, 57
Differentiation-in-frequency theorem 54
Differentiation-in-time theorem 53
Differentiator 398

Diffraction 10
Diffusion 345
Diode detector 149
Direct method 230, 283, 250
Directional antennas 259
Dirichlet's condition 43
Discrete r.v 325
Discrete random process 311
Discrete random variables 297, 298
Discrete spectrum 26
Discrete-time signals 15, 116
Discrete-time systems 93, 116
Discriminator 236, 389, 390, 404
Distortion 3, 148
Distortionless transmission 92, 106, 107, 117
Distributions 17
D-layer 6, 8
Double conversion receivers 274
Double heterodyne receivers 279, 287
Double sideband suppressed carrier modulation 156
Double spotting 272, 286
Doublers 228
Drift 345
DSB-SC modulation 156
DSB-SC system 375
Duality theorem 50
Dual-slope detector 237
Dynamic systems 93, 116

E

Effective bandwidth 199, 219, 221, 228, 250, 251
Effective radiated power (ERP) 258, 282, 284
Eigensignal 103, 105, 116
E-layer 8
Electromagnetic spectrum 4
Elementary outcomes 294
Elliptic 435
Energy signal 14, 15, 79, 81, 82, 347
Energy spectral density 14, 60, 79
Ensemble average 314, 347
Envelope and phase-angle representation 390
Envelope detector 146, 149, 189, 191, 237, 264, 373, 377, 379, 389, 411
Envelope 126, 135
Equalizer 447
Equivalent lowpass LTI system 74
Equivalent lowpass system 80
Equivalent noise resistance 352, 358, 366
Equivalent noise temperature 355, 356, 357, 361, 366
Ergodic processes 327
Euclidian space 21
Even symmetry 33, 44, 108, 117
Event 294
Exciter section 258
Expected value of a random variable 303
Exponential order 112
External noise 336
Extra-terrestrial noise 336, 337

F

Fading 3, 5, 9
Fall-time 453
FDM hierarchy 193
Fidelity 274, 276, 287
Figure of merit for a DSB-SC system 376
Figure of merit for a SSB-SC system 375
Figure of merit for AM 380
Figure of merit for FM 394
Figure of merit for PM 395
Figure of merit 269, 373, 410
Filter method 171, 193, 277
Filter 108
Filtering 105
Flat-top sampling 431, 438, 459
F-layer 6, 7
FM demodulator with feedback (FMFB) 244
FM feedback detector 236
FM mobile communication 234
FM radio broadcasting 247
FM receivers 283
FM stereo broadcasting 248
FM transmitters 258
Folding-over effect 460
Foster-Seeley detector 251, 252
Foster-Seeley discriminator 238
Fourier series coefficients 29, 77
Fourier series 13, 20, 417, 420, 422
Fourier transform theorems 47, 79
Fourier transform 13, 41, 78, 82, 105, 113, 170, 320, 418, 423, 432
Free space 123
Frequency demodulator 398
Frequency deviation constant 201, 220, 223, 250, 389
Frequency deviation 208, 390, 201, 206
Frequency divider 167
Frequency division multiplexing (FDM) 180, 181, 193, 440
Frequency error 177

Frequency folding effect 424
Frequency modulation feedback (FMFB) 406
Frequency modulation 3, 124, 182, 199
Frequency multipliers 250
Frequency response function 103, 104
Frequency response 117, 268
Frequency stability 232, 257
Frequency translation 3
Frequency-shift theorem 29
Frequency-to-voltage conversion 237
Friis's formula 357, 361, 366
Function of a random variable 305

G

Galactic noise 336
Ganged condenser 273
Gate waveform 140
Gaussian density function 299, 329
Gaussian density 221
Gaussian modulated signal 221
Gaussian processes 317
Gaussian random variable 299, 325
Generalized Fourier series 24
Generalized functions 17
Ground state 94
Ground wave 6, 258, 336
Group delay 242
Guard band 419, 424
Guard time 444

H

Half-power bandwidth 107, 117
Half-wave symmetry 33
Hazeltine method 261
Hermitian symmetry 43, 81
High-frequency band 6
High-level modulation 145, 191, 260, 285
Highpass filter 109
Hilbert transform 14, 66, 75, 80, 88, 108, 180, 323, 374
Homodyne detection 178
Homogeneity 47

I

Ideal bandpass filter 110
Ideal high pass filter 109
Ideal lowpass filter 109, 421, 424
Ideal sampling 421, 459
IF amplifier bandwidth 407
IF amplifier 241, 247, 263, 264
IF bandwidth 276
Image frequency rejection ratio (IFRR) 268
Image frequency rejection 264, 268
Image frequency 270, 286
Image rejection 274
Image signal rejection 279
Image signal 268
Impulse function 317
Impulse response function 13, 64
Impulse response 96, 116, 317, 436
Impulse sampling 421
Independent processes 327
Independent sideband transmission (ISB) 182, 193
Indirect method 227, 252
In-phase and quadrature components 379
In-phase channel 165
In-phase component 71, 323
Instantaneous frequency 200, 208, 221, 231, 389
Instantaneous sampling 421
Integrated noise figure 354
Integration theorem of Fourier transform 56
Integrator 398
Interfering signal 247
Intermediate frequency (IF) 236, 247, 263, 283, 373
Inter-modulation frequency 276
Internal noise 336
Interpolating functions 421
Intersymbol interference 293
Inverse Fourier transform 42, 51, 78, 109 317, 427, 436
Ionosphere 6, 337
ISB receiver 281
ISB transmitter 280

J

Jacobian 310
Johnson noise 337, 365
Joint density function 307, 326
Jointly Gaussian 309, 324, 326
Jointly stationary processes 324
Jointly stationary 318
Jumbo group 182

K

Kirchhoff's mesh equations 102

L

Laplace transform 98, 104
Laser 417
Leading edge 446
L'Hospital's rule 337
Limit theorem 327, 329
Limiter stage 247, 252
Limiter 239
Linear phase response 92, 106
Linear system 94
Linear time-invariant (LTI) systems 13, 92
Linearity theorem 29, 47
Line-of-sight propagation 282
Line-of-sight 8, 259
Local carrier 176
Local oscillator frequency 265, 272
Local oscillator 178, 228
Loop filter 244, 408
Lossy transmission line 359
Loud speaker 259
Lower side-band 129, 169
Low-level modulation 145, 191, 260, 285
Low-noise amplifier 359, 361
Lowpass equivalent 392
Lowpass filter 109, 418
Lowpass process 321
Lowpass representation 72
Lowpass sampling theorem 322, 328, 416, 419, 428, 438, 459
LSSB-SC signal 170, 175, 192
LTI system 317, 328

M

Magnetron 417
Magnification factor 342
Magnitude response 104, 107, 117
Magnitude spectrum 44, 77, 106, 108
Man-made noise 336
Marginal density functions 307
Master groups 182
Matching transformer 265
Maximum power transfer 343
Maximum usable frequency 8
Mean of the random process 312
Mean of the random variable 312
Mean square sense 321
Mean squared value 347
Mean 299, 327
Measurement of noise figure 363
Mechanical filters 278
Medium frequency 6
Medium waves 6
Medium-wave band 258, 336
Microwave 10
Milky way 337
Mixed r.v 325
Mixed random variable 298
Mixed-type random variables 297
Mixer stage 348
Mixer 228, 264
Modified Bessel function 300
Modulated signal 124
Modulating signal 138
Modulation index for FM 205, 250
Modulation index for PM 205, 250
Modulation index 126, 133, 135, 153, 217, 220, 224, 251, 377, 445
Modulation section 258
Modulation theorem 29, 48, 58
Modulation 3, 123
Modulator 138
Monostable multivibrator 242, 447
Morse code 1
Multipath propagation 12
Multipath 3, 5
Multiplexing 3, 124, 180, 181
Multiplication theorem 29, 51
Mutual conductance 235, 238
Mutual inductance 238

N

Narrowband angle modulation 210
Narrowband angle modulator 228
Narrowband noise 364
Natural sampling 459
Negative feedback 262, 285
Negative frequencies 170
Negative peak clipping 153, 192, 264
Neutralization 261, 285
Noise equivalent bandwidth 351, 366, 367
Noise equivalent circuit 341
Noise figure of the lossy line 360
Noise figure 263, 287, 353, 356, 364, 366, 274, 276, 354
Noise power spectral densities 339
Noise power spectrum 341

Noise power 340
Noise temperature of a source 344
Noise temperature 361, 367
Noise 243, 293, 335
Noise-equivalent bandwidth of the filter 349
Noise-equivalent circuits 338
Non-linear system 94
Normal random variable 299
Notch filters 335
Nyquist rate 419, 435, 424, 459

O

Odd symmetry 33, 44, 117
Op-amp 139
Optical fibre 3, 5
Optical frequencies 3
Orthogonal expansion 20, 41
Orthogonal functions 21, 23
Orthogonal process 313, 327
Orthogonal vectors 21
Orthogonal 308
Orthonormal functions 23, 24, 31, 81
Oscilloscope 135
Out-of-band noise 336, 425
Over modulation 126, 135

P

Padder 273
Pahse spectrum 25
Paley-Wiener criterion 112, 117
PAM 438
Parseval's theorem 30, 45, 51, 60, 85
Partition noise 348, 365
Passbands 108
Peak frequency deviation 222, 223, 225, 236, 394
Peak phase deviation 225
Periodic signal 40
Personal communications 2
Phase ambiguity 166, 167
Phase demodulator 398
Phase deviation constant 200, 208, 220, 249
Phase deviation 200, 206, 208
Phase discriminator 236
Phase distortion 177
Phase error 163, 176
Phase lock loop or PLL 406
Phase modulation 3, 124, 199
Phase modulator 228
Phase response 104, 117
Phase shifters 175
Phase spectrum 44, 77, 108
Phase synchronism 166
Phase-locked loop (PLL) 244, 251
Phase-locked loop detector 236
Phase-shift discriminator 238
Phase-shift SSB-SC detector 180
Phasing method 171, 193
Physical realizability 111
Pilot carrier SSB transmitter 278
Pilot carrier 178, 249
Planck's constant 337
Point-to-point communication 182, 259
Positive frequencies 170
Post-detection noise spectrum 393
Post-detection noise 393, 397
Post-detection signal-to-noise ratio 386
Post-detection SNR 388, 413
Power gain 353
Power in a DSB-SC signal 157
Power signal 14, 15, 39, 62, 81, 82
Power spectral density of the thermal noise 337
Power spectral density 14, 61, 62, 79, 220, 316, 324 , 346, 349
Power spectrum 313, 320, 329, 393
Power-bandwidth trade-off 395, 410
Power-limited 11
Practical diode detector 152
Pre-detection signal-to-noise ratio 373
Pre-detection SNR 388, 390, 409, 413
Pre-emphasis filter 397, 398
Pre-emphasis 234, 247, 248, 284, 287, 396, 411
Pre-envelope 69, 70, 76, 80, 170
Primary service area 258, 282, 284
Probability density function (PDF) 298, 325
Probability masses 298
Probability space 294, 306
Product modulator 139
PSD of an FM signal 221
Pulse amplitude modulation (PAM) 416, 454
Pulse code modulation (PCM) 416
Pulse duration modulation (PDM) 416
Pulse modulation systems 417
Pulse position modulation (PPM) 416, 444
Pulse width modulation (PWM) 444
Pulse-time modulation 444

Q

Q-factors 238
Q-function 300
Quadrature amplitude modulation 167
Quadrature carrier multiplexing 167, 192
Quadrature channel 165
Quadrature component 71, 323, 365
Quadrature detector 251
Quadrature FM detector 236, 241

R

Radio horizon 9
Raised cosine pulse 83
Random processes 293, 311
Random signals 17
Random variable 296
Ratio detector 236, 239, 251
Rayleigh density function 310
Rayleigh density 403
Rayleigh distributed 392
Rayleigh distribution 300
Rayleigh r.v 325
Rayleigh's theorem 45
Reactance modulator 230, 232, 234, 252
Reactive elements 341
Realization of the random signal 347
Received signal power 376
Receiver 3
Reconstruction filter 428, 432, 435
Reconstruction 427
Replication property 18, 53
RF power section 258
Rice method 261
Ricean random variable 300
Rician r.v 325
Ring modulator 161
Rise time 112, 117, 453

S

Sample space 294, 324, 329
Sample and hold 431
Sampling function 417, 422
Sampling property 18, 66
Satellite communication 3
SAW filters 278
Sawtooth waveform 451
Scaling theorem 29, 49, 58
Scrambler 198
Selectivity 274, 275, 287
Sensitivity 274, 275, 287
Service areas 129
Shielding 335
Shifting in time 19
Short-wave band 258, 336
Shot noise in semiconductor diodes 348
Shot noise 336, 345, 347, 365
Sideband filters 277
Sideband power 145, 385
Sideband suppression filters 173
Sideband 131, 133
Signal bandwidth 92, 106
Signal 293
Signal-to-noise ratio 181, 222, 335, 353, 366, 373
Signum function 55
Simple AGC 265
Sine integral function 114
Single frequency noise figure 354
Single sideband suppressed carrier 156, 169
Skin depth 5
Skip distance 8
Sky-wave 6, 8, 258, 336
Slope detector 236
Snell's law 7
SNR improvement 399
Solar noise 336
Sonars 5
Source 2
Space 329
Space-charge-limited region 348
Space-wave 8
Spectrum of the DSB-SC 159
Spectrum 48, 106, 124, 217, 329
Speech signals 124
Spot noise figure 354
Square-law detection 147
Square-law detector 191
Square-law modulators 138
Squaring loop 164, 166
Squelch circuit 279
SSB transmitters 258
SSB-pilot carrier receiver 279
SSB-SC modulation 169
SSB-SC signal 157
SSB-SC systems 374
Stability 100
Stagger-tuned IF stages 267
Standard deviation 326

Static systems 93, 116
Static 336
Stationarity 327
Stationary in the wide sense 327
Stationary processes 313
Statistically independent 295, 326
Step response 97, 98
Stereophonic FM receiver 249
Stop-bands 108
Strong Dirichlet's conditions 33, 77
Subcarriers 168, 181
Sub-carriers 181
Sun-spot cycle 336
Super group 182
Superheterodyne receiver 251, 259, 263, 283, 285, 389
Superheterodyne 247
Superposition integrals 97
Superposition principle 339
Superposition 47
Surface acoustic wave 278
Switching modulator 139
Synchronism 166
Synchronization 438, 450
Synchronous demodulator 156, 375
Synchronous detection 147
Synchronous detector 373, 374, 377, 411
System bandwidth 92, 106, 107
System 93

T

Telemetry 417
Television 182
Temperature-limited condition 346
Thermal noise 299, 336, 337, 345, 365
Thermionic emission 345
Thevinin's equivalent circuit 358
Third method 171, 174
Three-point tracking 273, 286
Threshold effect for AM 410
Threshold effect in envelope detection of AM 382
Threshold effect in FM 410
Threshold effect 246
Threshold extension 406, 407, 410
Threshold in WBFM 402
Threshold phenomenon 404
Time division multiplexing (TDM) 181, 417, 438, 456
Time invariance 94
Time-constant 152, 442
Time-delay theorem 47, 58
Time-invariant system 116
Time-shift theorem 29
Time-varying system 94
Tracking error 273
Tracking 272, 286
Trailing edge 446, 449
Transfer function 87, 103, 104, 105, 117, 341, 427, 436
Transition bandwidth 172
Transit-time 346
Transmission bandwidth 200, 222, 395, 458
Transmitter 3, 257
Trapezoid method 135
Trapezoidal pattern 191
TRF receivers 259, 285
Trigonometric Fourier series 30, 77, 84, 140
Trimmer 273
Triplers 228
Troposphere 10
Tropospheric propagation mode 8
Tuned circuit 342
Tuned power amplifiers 145
Tuned radio frequency (TRF) receiver 262
Two-point tracking 273, 286
TWT amplifier 235

U

Uncorrelated processes 313, 327
Uncorrelated 338
Uniform distribution 403
Uniform random variable 298, 325
Uniformly distributed 392
Unit impulse function 17, 19, 97
Unit-step function 17, 19, 97, 112, 116
Unit-step response 113, 116
Unscrambling 198
Upper side-band 129, 169
USSB-SC signal 170, 175, 192

V

Varactor diode 232, 234
Variable capacitor 231
Variable reactance 230
Variance of a random variable 304
Variance 299
Very large groups 182
Vestigial sideband filter 185
Vestigial sideband modulation 183, 193
Video signal 177

Voltage-controlled oscillator 165, 244
VSB filter 185
VSB shaping filter 195

W

Waveguide 3, 5
Wave-guide 7
Wavelength 123
Weak Dirichlet's condition 33, 77
Weaver's method 171, 174, 193
White noise process 317, 327, 330
White noise 88, 338, 344, 347
Wideband –90° phase shifters 173
Wideband FM 213
Wide-sense stationarity 313
Wiener–Khinchine theorem 327
WSS process 330

Z

Zero-crossing detector 236, 251
Zero-crossing FM detector 242
Zero-crossings 243
Zero-mean white noise 350, 365
Zero-order-hold 436, 461