NEHRU COLLEGE OF ENGINEERING AND RESEARCH CENTRE
(NAAC Accredited)
(Approved by AICTE, Affiliated to APJ Abdul Kalam Technological University, Kerala)

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

COURSE MATERIALS

ECT 305: ANALOG AND DIGITAL COMMUNICATION

VISION OF THE INSTITUTION

To mould true citizens who are millennium leaders and catalysts of change through excellence in education.

MISSION OF THE INSTITUTION

NCERC is committed to transform itself into a center of excellence in Learning and Research in Engineering and Frontier Technology and to impart quality education to mould technically competent citizens with moral integrity, social commitment and ethical values.

We intend to facilitate our students to assimilate the latest technological know-how and to imbibe discipline, culture and spiritually, and to mould them in to technological giants, dedicated research scientists and intellectual leaders of the country who can spread the beams of light and happiness among the poor and the underprivileged.

ABOUT DEPARTMENT

- Established in: 2002
- Course offered : B.Tech in Electronics and Communication Engineering

M.Tech in VLSI

- Approved by AICTE New Delhi and Accredited by NAAC
- Affiliated to the University of Dr. A P J Abdul Kalam Technological University.

DEPARTMENT VISION

To Provide Universal Communicative Electronics Engineers with corporate and social relevance towards sustainable developments through quality education.

DEPARTMENT MISSION

1) Imparting Quality education by providing excellent teaching, learning environment.

2) Transforming and adopting students in this knowledgeable era, where the electronic gadgets (things) are getting obsolete in short span.

3) To initiate multi-disciplinary activities to students at earliest and apply in their respective fields of interest later.

4) Promoting leading edge Research & Development through collaboration with academia & industry.

PROGRAMME EDUCATIONAL OBJECTIVES

PEOI. To prepare students to excel in postgraduate programmes or to succeed in industry / technical profession through global, rigorous education and prepare the students to practice and innovate recent fields in the specified program/ industry environment.

PEO2. To provide students with a solid foundation in mathematical, Scientific and engineering fundamentals required to solve engineering problems and to have strong practical knowledge required to design and test the system.

PEO3. To train students with good scientific and engineering breadth so as to comprehend, analyze, design, and create novel products and solutions for the real life problems.

PEO4. To provide student with an academic environment aware of excellence, effective communication skills, leadership, multidisciplinary approach, written ethical codes and the life-long learning needed for a successful professional career.

PROGRAM OUTCOMES (POS)

Engineering Graduates will be able to:

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSO)

PSO1: Ability to Formulate and Simulate Innovative Ideas to provide software solutions for Real-time Problems and to investigate for its future scope.

PSO2: Ability to learn and apply various methodologies for facilitating development of high quality
System Software Tools and Efficient Web Design Models with a focus on performance optimization.

PSO3: Ability to inculcate the Knowledge for developing Codes and integrating hardware/software
products in the domains of Big Data Analytics, Web Applications and Mobile Apps to create innovative career path and for the socially relevant issues.

COURSE OUTCOMES ECT 305

ECT305	ANALOG AND DIGITAL COMMUNICATION	CATEGORY	L	T	P	CREDIT
		PCC	3	1	0	4

Preamble: This course aims to develop analog and digital communication systems. **Prerequisite**: ECT 204 Signals and Systems, MAT 204 Probability, Random Process andNumerical Methods

Course Outcomes: After the completion of the course the student will be able to

CO 1	Explain the existent analog communication systems.
CO 2	Apply the concepts of random processes to LTI systems.
CO 3	Apply waveform coding techniques in digital transmission.
CO 4	Apply GS procedure to develop digital receivers.
CO 5	Apply equalizer design to counteract ISI.
CO 6	Apply digital modulation techniques in signal transmission.

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO 1	3	3										
CO 2	3	3	2	3	3							
CO 3	3	3	2	3	3	2						2
CO 4	3	3	2	3	3	2						2
CO 5	3	3	2	3	3	2						2
CO 6	3	3	2	3	3	2						2

Assessment Pattern

Bloom’s Category	**Continuous Assessment Tests**		**End Semester Examination**
	1	**2**	
Remember	10	10	20
Understand	30	30	60
Apply	10	10	20
Analyse			
Evaluate			
Create			

Mark distribution

Total Marks	CIE	ESE	ESE Duration
150	50	100	3 hours

Continuous Internal Evaluation Pattern:

Attendance	: 10 marks
Continuous Assessment Test (2 numbers)	: 25 marks
Assignment/Quiz/Course project	: 15 marks

End Semester Examination Pattern: There will be two parts; Part A and Part B. Part A contain 10 questions with 2 questions from each module, having 3 marks for each question. Students should answer all questions. Part B contains 2 questions from each module of which student should answer any one. Each question can have maximum 2 sub-divisions and carry 14 marks.

Course Level Assessment Questions

Course Outcome 1 (CO1): The existent analog communication system

1. What are the needs for analog modulation
2. Give the mathematical model of FM signal and explain its spectrum.

Course Outcome 2 (CO2): Application of random processes

1. Compute the entropy of a Gaussian random variable.
2. A six faced die is thrown by a player. He gets Rs. 100 if face 6 turns up, loses Rs. 20

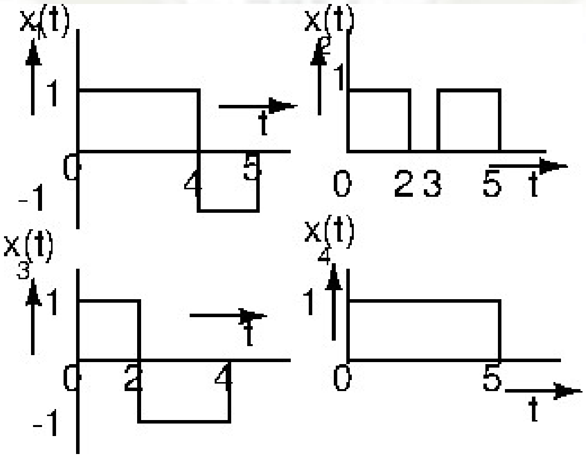
if face 3 or 4 turn up, gets Rs. 50 if face 5 turns up and loses Rs 10 if face 1 or 2 turn up. Draw thepdf and CDF for the random variable. Check if it is profitable based on statistical expectation.

Course Outcome 3 (CO3): Waveform coding

1. Compute the A and mu law quantized values of a signal that is normalized to 0.8 with A=32 and mu=255.
2. Design a 3-tap linear predictor for speech signals with the autocorrelation vector [0.95,0.85,0.7,0.6] , based on Wiener-Hopf equation. Compute the minimum mean square error.

Course Outcome 4 (CO4): G-S Procedure and effects in the channel

1. Apply G-S procedure on the following signals and plot their signal space.



2. Derive the Nyquist criterion for zero ISI.

Course Outcome 5 (CO5): Digital modulation

1. Give the mathematical model of a BPSK signal and plot its signal constellation.
2. Draw the BER-SNR plot for the BPSK system

SYLLABUS

Module 1 Analog Communication

Block diagram of a communication system. Need for analog modulation. Amplitude modulation. Equation and spectrum of AM signal. DSB-SC and SSB systems. Block diagram of SSB transmitter and receiver. Frequency and phase modulation. Narrow and wide band FM and their spectra. FM transmitter and receiver.

Module 2 Review of Random Variables and Random Processes

Review of random variables – both discrete and continuous. CDF and PDF, statistical averages. (Only definitions, computations and significance) Entropy, differential entropy. Differential entropyof a Gaussian RV. Conditional entropy, mutual information.

Stochastic processes, Stationarity. Conditions for WSS and SSS. Autocorrelation and power spectral density. LTI systems with WSS as input.

Module 3 Source Coding

Source coding theorems I and II (Statements only). Waveform coding. Sampling and Quantization. Pulse code modulation, Transmitter and receiver. Companding. Practical 15 level A and mu-law companders. DPCM transmitter and receiver. Design of linear predictor. Wiener-Hopf equation. Delta modulation. Slope overload.

Module 4 G-S Procedure and Effects in the Channel

Gram-Schmitt procedure. Signal space.

Baseband transmission through AWGN channel. Mathematical model of ISI. Nyquit criterion for zero ISI. Signal modeling for ISI, Raised cosine and Square-root raised cosine spectrum, Partial response signalling and duobinary coding. Equalization. Design of zero forcing equalizer.

Vector model of AWGN channel. Matched filter and correlation receivers. MAP receiver, Maximum likelihood receiver and probability of error.

Capacity of an AWGN channel (Expression only) -- significance in the design of communication schemes.

Module 5 Digital Modulation Schemes

Digital modulation schemes. Baseband BPSK system and the signal constellation. BPSK transmitter and receiver. Base band QPSK system and Signal constellations. Plots of BER Vs SNR with analysis. QPSK transmitter and receiver. Quadrature amplitude modulation and signal constellation.

Text Books

1. "Communication Systems", Simon Haykin, Wiley.
2. "Digital Communications: Fundamentals and Applications", Sklar, Pearson.
3. "Digital Telephony", John C. Bellamy, Wiley

References

1. "Principles of Digital Communication," R. Gallager, Oxford University Press
2. "Digital Communication", John G Proakis, Wiley.

Course Contents and Lecture Schedule

No	**Topic**	**No. of Lectures**
1	**Analog Communication**	
1.1	Block diagram of communication system, analog and digital systems , need for modulation	2
1.2	Amplitude modulation, model and spectrum and index of modulation	2
1.3	DSB-SC and SSB modulation. SSB transmitter and receiver	2
1.4	Frequency and phase modulation. Model of FM, spectrum of FM signal	2
2	**Review of Random Variables**	
2.1	Review of random variables, CDF and PDF, examples	2
2.2	Entropy of RV, Differential entropy of Gaussian RV, Expectation, conditional expectation, mutual information	4
2.3	Stochastic processes, Stationarity, WSS and SSS. Autocorrelation and power spectral density. Response of LTI systems to WSS	3
3	**Source Coding**	
3.1	Source coding theorems I and II	1
3.2	PCM,Transmitter and receiver, companding Practical A and mu law companders	4
3.3	DPCM, Linear predictor, Wiener Hopf equation	3
3.4	Delta modulator	1

4	**GS Procedure and Channel Effects**	
4.1	G-S procedure	3
4.2	ISI, Nyquist criterion, RS and SRC, PR signalling and duobinary coding	3
4.3	Equalization, design of zero forcing equalizer	3
4.4	Vector model of AWGN channel, Correlation receiver, matched filter	4
4.5	MAP receiver, ML receiver, probability of error	1
4.6	Channel capacity, capacity of Gaussian channel, Its significance in design of digital communication schemes	2
5	**Digital Modulation**	
5.1	Need of digital modulation in modern communication.	1
5.2	Baseband QPSK system, signal constellation. Effect of AWGN, probability of error (with derivation). BER-SNR curve, QPSK transmitter and receiver.	4
5.3	QAM system	1

ECT305 ADC PRACTICE QUESTIONS MODULE - I

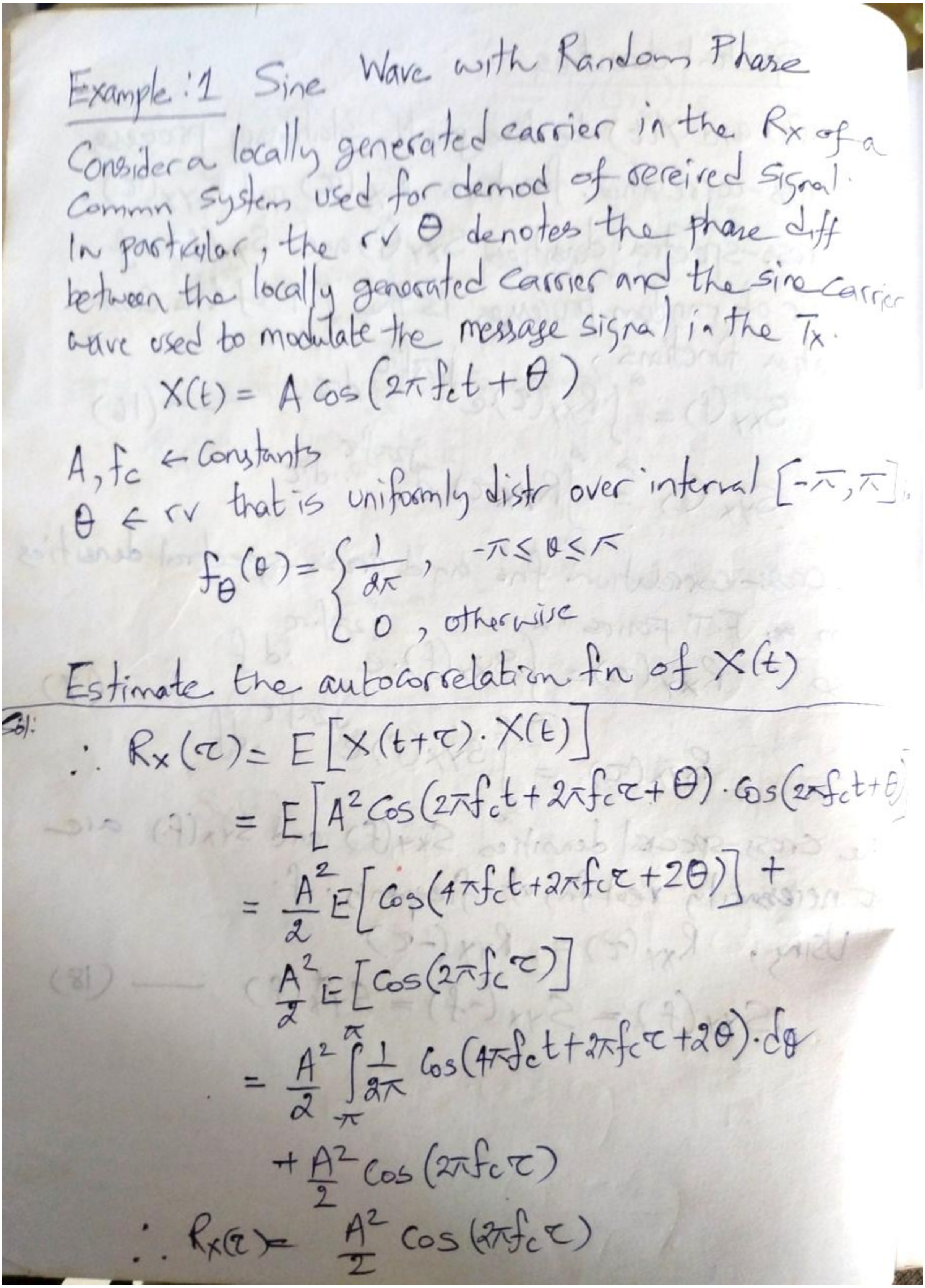
1. State and define Modulation. With the aid of example, indicate the two signals required for effective Modulation to take place.
2. For an audio signal at 15 kHz injected into free space, estimate the antenna length and justify whether it is physically feasible to implement. Provide the solution if not feasible.
3. State and define Information source, Message signal, Transducer and Carrier signal needed for a modern Communication System. Give examples.
4. A raw audio signal at 50 kHz needs to be injected into free space. Estimate the length/height of the aerial needed and justify whether it is physically practicable to implement. Provide the solution if not practicable.
5. Given three raw signals at 10 KHz, 15 KHz and 1 MHz, estimate in each case the length/height of the antenna needed for radiation into free space and verify its feasibility.
6. State and explain A3E, DSB-SC and SSB modulation.
7. State and briefly explain Phase cancellation method with the aid of a diagram.

8. With the aid of a block diagram, describe the general Communication System. Provide the functionality of each and every block in the system.
9. Analyse the need for Modulation by highlighting at least five reasons and explaining each of them.
10. Define Amplitude Modulation. From first principles, derive the equation for an AM Wave, with the aid of neat diagrams of relevant signals.
11. Draw and neatly label the time domain representation and frequency domain spectrum of an AM Signal. Hence derive the modulation index and the bandwidth of the AM wave.
12. Find the carrier and modulating frequencies, the modulation index and maximum deviation of an FM wave represented by,
    $v_{FM}(t)= 20 \sin(6 \times 10^8 t + 5 \cos 1250 t)$
    What is the power dissipated by the wave in a 100 ohms resistive load?
13. Describe the frequency spectrum of a FM wave for both the NBFM and WBFM cases. State the Carson's rule and its application.
14. Derive the Power relations in an Amplitude Modulated wave from first principles.
15. Calculate the percentage power saving when the carrier and one of the sidebands are suppressed in an AM wave to a depth of (a) 90% (b) 55% (c) 75%
16. Explain the generation of FM signal.
17. With the aid of an illustrative block diagram, explain the receiver for frequency modulated signal using the Superhetrodyne structure.
18. Validate the requirements for modulation in modern analog communication systems, delving into factors that mandate its implementation in the Transmitter.
19. Provide a comparison between Analog communication systems and Digital communication system on the basis of performance criteria.
20. Define Frequency Modulation. From first principles, derive the equation for an FM Wave, with the aid of concise diagrams of modulating signal, carrier signal and FM modulated signal.
21. With the help of diagrams, explain any two methods used for the generation of SSB signals.

ECT305 ADC PARTIAL PRACTICE QUESTIONS MODULE – II

1. State and define Random Variable. With the help of diagram relate Random Variable, Sample Space and Probability.
2. From first principles, prove that the Variance of a Random variable is equal to the difference between the Mean Square value and square of the Mean.
3. State and define expected value or mean, moments, second order moments, central moment and standard deviation of a random variable.
4. State and explain Stationary and Non-stationary processes. Give examples.
5. State and explain Stationarity in the strict sense with the aid of an example.
6. State and explain Wide Sense Stationary Random process.
7. Define Entropy of a discrete Random variable.

8. Analyse a Stochastic Process, depicting the relationship with Sample space and functions of time with the help of a diagram.
9. State and define Autocorrelation Function. Analyse its properties from first principles and provide the physical relevance of each of them.
10. Analyse a Random variable, illustrating the relationship with Sample space and probability with the help of a diagram.
11. With the aid of an illustrative diagram, define and explain the difference between stochastic process and random variable.
12. Analyse the stochastic processes from first principles and formulate the functional time independence relationships for Strictly Stationary and Weakly Stationary processes.
13. Consider a discrete memoryless source with source alphabet $\mathcal{S} = \{s_0, s_1, s_2\}$, whose three distinct symbols have the following probabilities: $p_0=1/4$, $p_1=1/4$, $p_2=1/2$. Calculate the Entropy of the Source.
14. From first principles, estimate the Differential Entropy of a continuous Gaussian random process. Comment on the specific properties for Gaussian random processes.



Example : 1 Sine Wave with Random Phase

Consider a locally generated carrier in the Rx of a Commn system used for demod of received signal. In particular, the rv θ denotes the phase diff between the locally generated carrier and the sine carrier wave used to modulate the message signal in the Tx.

$$X(t) = A\cos(2\pi f_c t + \theta)$$

$A, f_c \leftarrow$ Constants

$\theta \leftarrow$ rv that is uniformly distr over interval $[-\pi, \pi]$.

$$f_\theta(\theta) = \begin{cases} \frac{1}{2\pi}, & -\pi \leq \theta \leq \pi \\ 0, & \text{otherwise} \end{cases}$$

Estimate the autocorrelation fn of $X(t)$

Sol:

$$\therefore R_X(\tau) = E[X(t+\tau) \cdot X(t)]$$

$$= E[A^2 \cos(2\pi f_c t + 2\pi f_c \tau + \theta) \cdot \cos(2\pi f_c t + \theta)]$$

$$= \frac{A^2}{2} E[\cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)] +$$

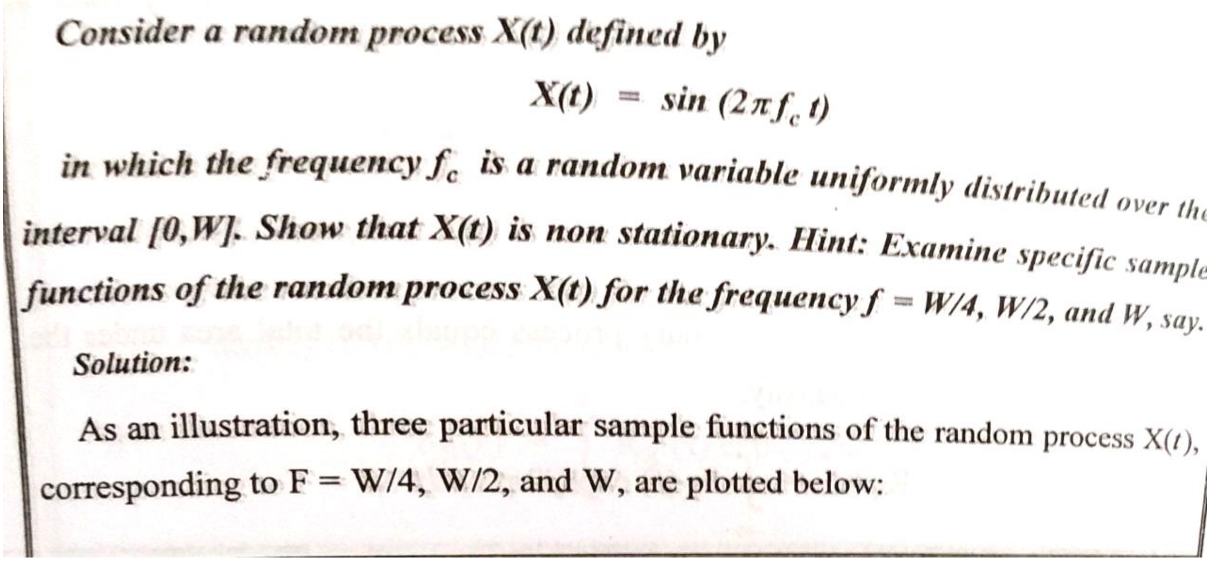
$$\frac{A^2}{2} E[\cos(2\pi f_c \tau)]$$

$$= \frac{A^2}{2} \int_{-\pi}^{\pi} \frac{1}{2\pi} \cos(4\pi f_c t + 2\pi f_c \tau + 2\theta) \cdot d\theta$$

$$+ \frac{A^2}{2} \cos(2\pi f_c \tau)$$

$$\therefore R_X(\tau) = \frac{A^2}{2} \cos(2\pi f_c \tau)$$

Example -2



Consider a random process X(t) defined by

$$X(t) = \sin(2\pi f_c t)$$

in which the frequency f_c is a random variable uniformly distributed over the interval [0,W]. Show that X(t) is non stationary. Hint: Examine specific sample functions of the random process X(t) for the frequency f = W/4, W/2, and W, say.

Solution:

As an illustration, three particular sample functions of the random process $X(t)$, corresponding to F = W/4, W/2, and W, are plotted below:

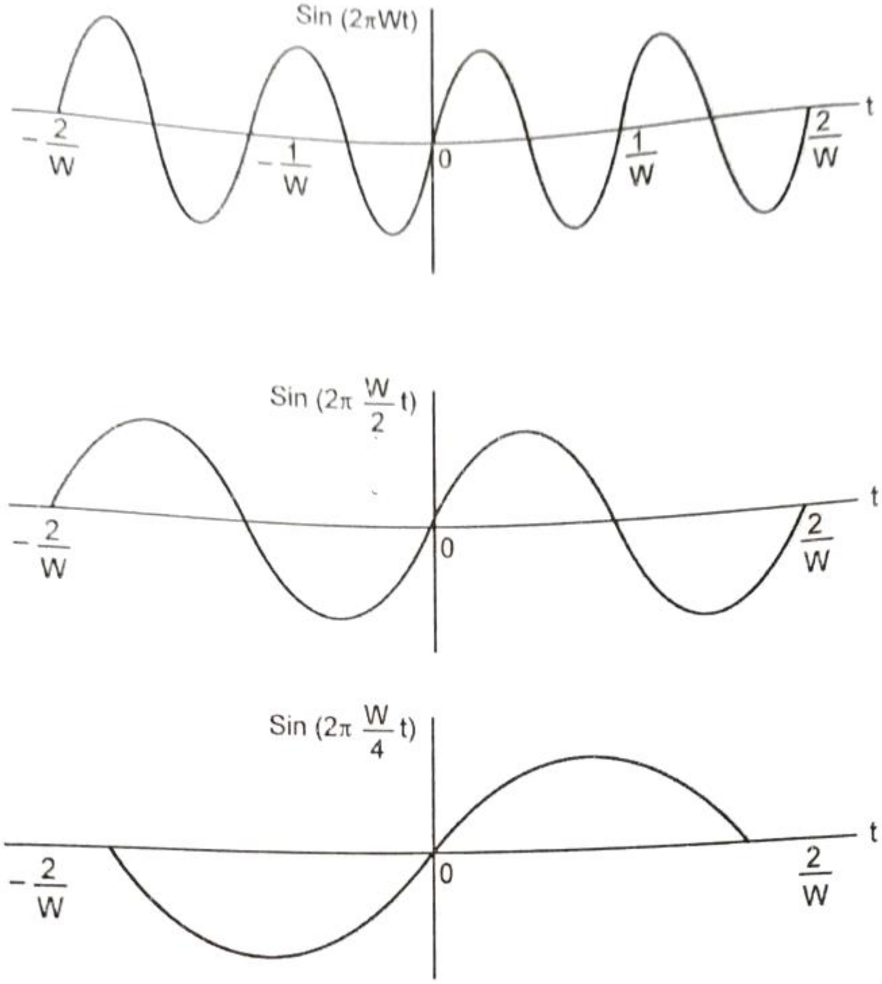


Fig. 2.6.

To show that X(t) is nonstationary, we need only observe that every waveform illustrated above is zero at $t = 0$, positive for $0 < t < 1/2W$, and negative for $-1/2$ W $< t < 0$. Thus, the probability density function of the random variable X(t_1) obtained by sampling X(t) at $t_1 = 1/4$W is identically zero for negative argument, whereas the probability density function by the random variable X(t_2) obtained by sampling X(t) at $t = -$ ¼ W is non zero only for the negative arguments. Clearly, therefore,

$f_{X(t_1)}(X_1) \neq f_{X(t_2)}(X_2)$, and the random process X(t) is non stationary.

Example: 3:- Sine wave with Random Phase (cont..)

$$X(t) = A\cos(2\pi f_c t + \theta)$$

Where $\theta \leftarrow$ uniformly distr rv over $[-\pi, \pi]$

Estimate the PSD of $X(t)$. Prove Property 1.

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau)$$

Use eqn (5)

$$\therefore S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\cdot e^{-j2\pi f\tau}\cdot d\tau$$

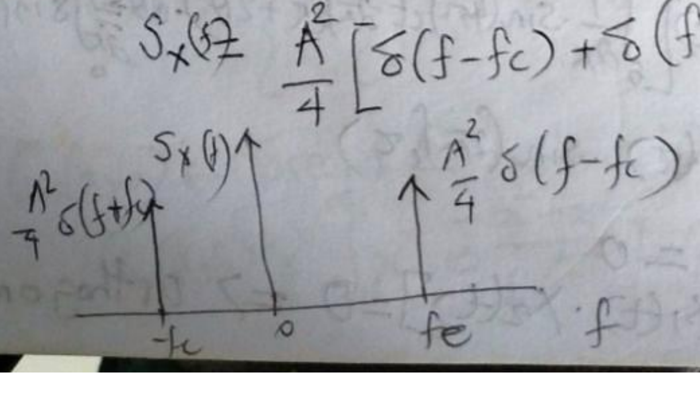
$$= \frac{A^2}{2}\int_{-\infty}^{\infty} \cos(2\pi f_c\tau)\cdot e^{-j2\pi f\tau}\cdot d\tau$$

$$= \frac{A^2}{2}\int_{\infty}^{\infty} R.P\{e^{j2\pi f_c\tau}\} e^{-j2\pi f\tau}\cdot d\tau$$

$$= \frac{A^2}{4}\int_{\infty}^{\infty} (e^{j2\pi f_c\tau} + e^{-j2\pi f_c\tau}) e^{-j2\pi f\tau}\cdot d\tau$$

$$= \frac{A^2}{4}\int_{\infty}^{\infty} e^{-j2\pi(f-f_c)\tau}\cdot d\tau + \frac{A^2}{4}\int_{-\infty}^{\infty} e^{-2\pi(f+f_c)\tau}\cdot d\tau$$

$$S_X(f) = \frac{A^2}{4}\left[\delta(f-f_c) + \delta(f+f_c)\right]$$



Exer 4 : Mixing a Random Process with a Sine Wave

$X(t), \Theta$ ← Independent rps

$$Y(t) = X(t) \cdot \cos(2\pi f_c t + \Theta)$$

Show ~~Find~~ PSD of $Y(t)$ is $\frac{1}{4}[S_X(f-f_c) + S_X(f+f_c)]$

Autocorrelation function of $Y(t)$

$$R_Y(\tau) = E[Y(t+\tau) \cdot Y(t)]$$

$$= E[X(t+\tau) \cos(2\pi f_c t + 2\pi f_c \tau + \theta) X(t) \cos(2\pi f_c t + \theta$$

$$= E[X(t+\tau) \cdot X(t)] \cdot E[$$

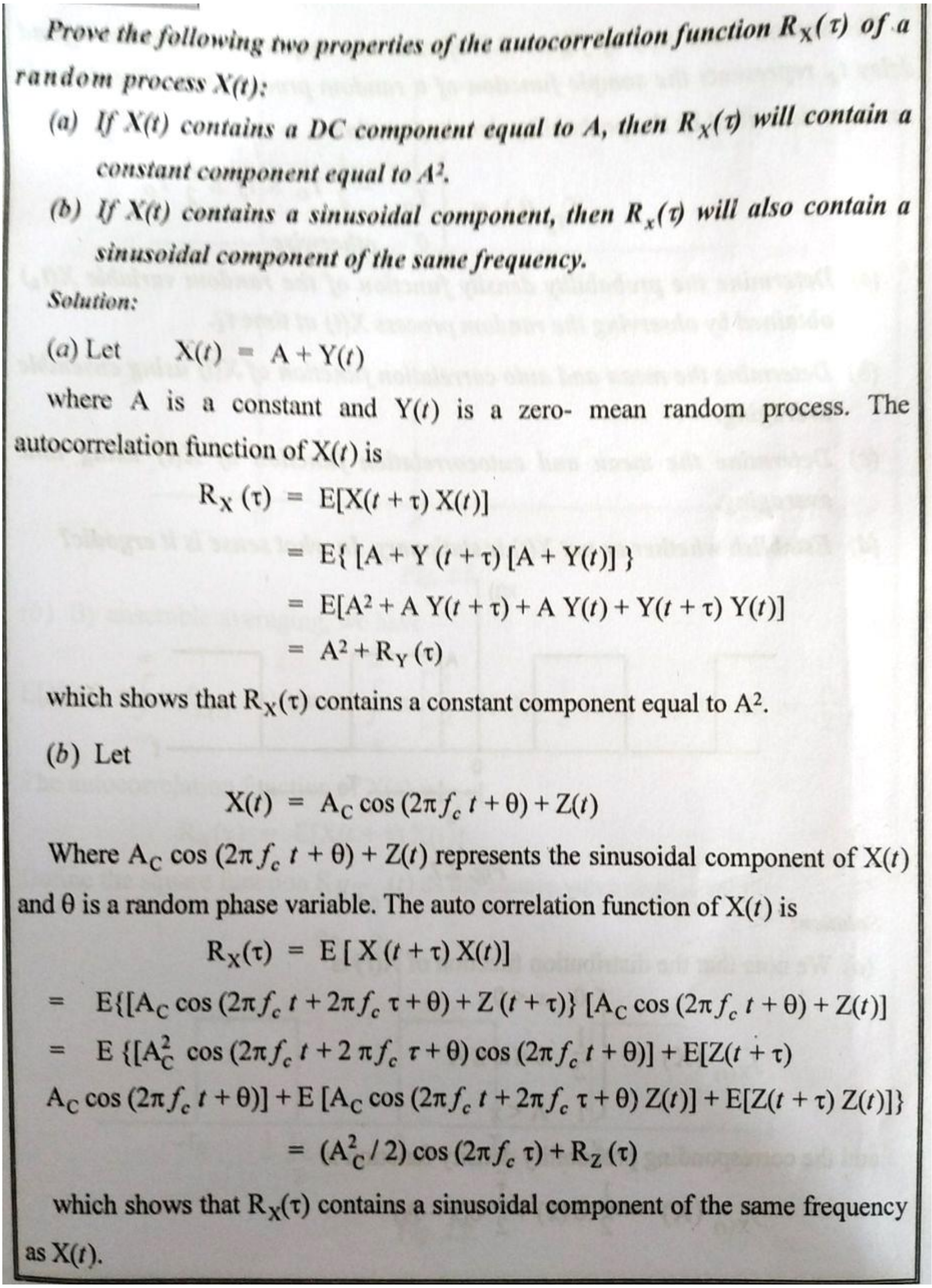
$$= \frac{1}{2} R_X(\tau) E[\cos(2\pi f_c \tau) + \cos(4\pi f_c t + 2\pi f_c \tau + 2\theta)]$$

$$= \frac{1}{2} R_X(\tau) \cos 2\pi f_c \tau$$

The PSD is the FT of autocorrelation fn

$$\therefore S_Y(f) = \frac{1}{4}[S_X(f-f_c) + S_X(f+f_c)]$$

Example -5



Prove the following two properties of the autocorrelation function $R_X(\tau)$ of a random process $X(t)$:

(a) If $X(t)$ contains a DC component equal to A, then $R_X(\tau)$ will contain a constant component equal to A^2.

(b) If $X(t)$ contains a sinusoidal component, then $R_x(\tau)$ will also contain a sinusoidal component of the same frequency.

Solution:

(*a*) Let $X(t) = A + Y(t)$

where A is a constant and $Y(t)$ is a zero- mean random process. The autocorrelation function of $X(t)$ is

$$
\begin{aligned}
R_X(\tau) &= E[X(t+\tau)\,X(t)] \\
&= E\{\,[A + Y(t+\tau)\,[A + Y(t)]\,\} \\
&= E[A^2 + A\,Y(t+\tau) + A\,Y(t) + Y(t+\tau)\,Y(t)] \\
&= A^2 + R_Y(\tau)
\end{aligned}
$$

which shows that $R_X(\tau)$ contains a constant component equal to A^2.

(*b*) Let

$$X(t) = A_C \cos(2\pi f_c t + \theta) + Z(t)$$

Where $A_C \cos(2\pi f_c t + \theta) + Z(t)$ represents the sinusoidal component of $X(t)$ and θ is a random phase variable. The auto correlation function of $X(t)$ is

$$R_X(\tau) = E[X(t+\tau)\,X(t)]$$

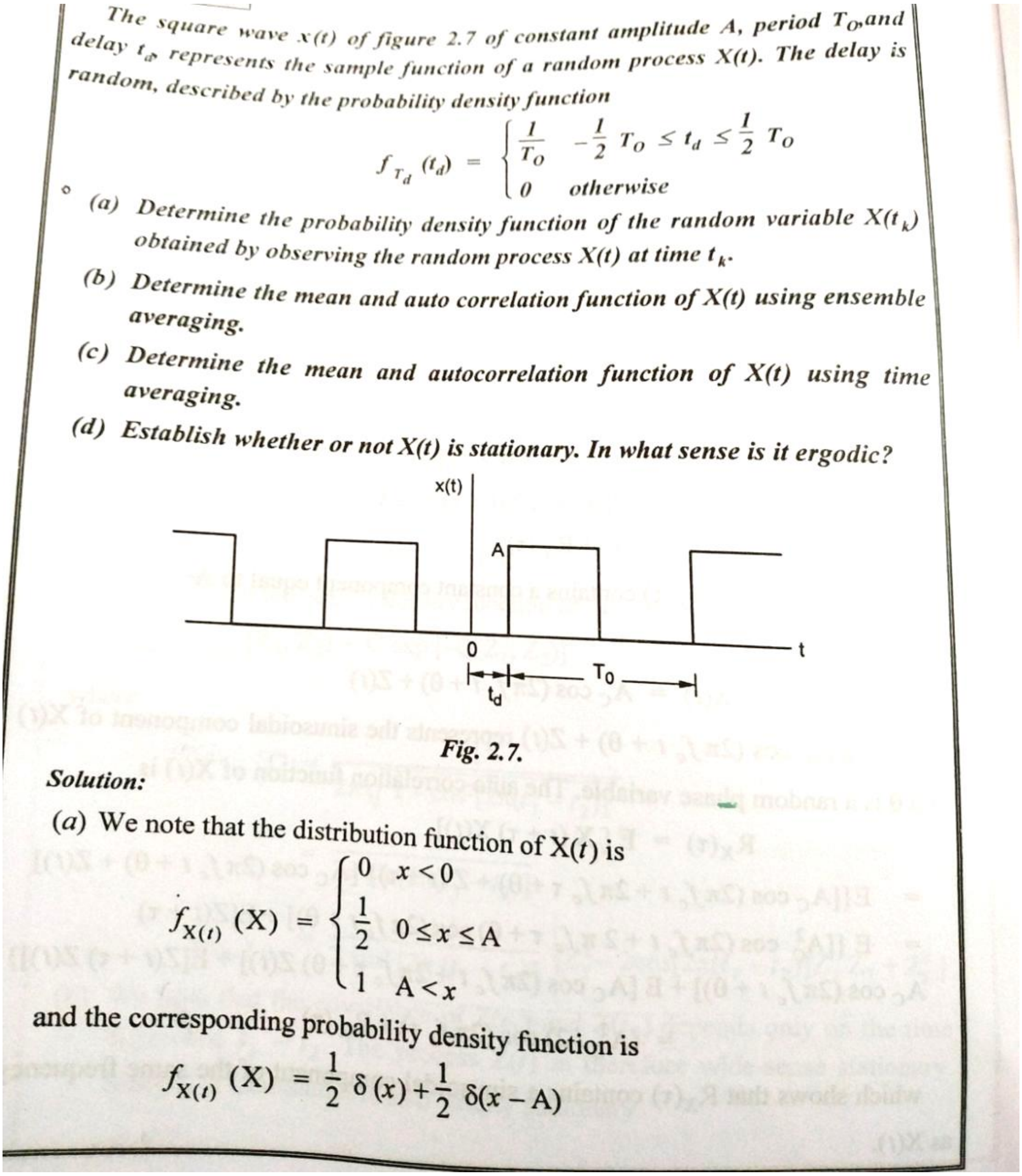
$$= E\{[A_C \cos(2\pi f_c t + 2\pi f_c \tau + \theta) + Z(t+\tau)\}\,[A_C \cos(2\pi f_c t + \theta) + Z(t)]$$

$$= E\{[A_C^2 \cos(2\pi f_c t + 2\pi f_c \tau + \theta)\cos(2\pi f_c t + \theta)] + E[Z(t+\tau)\,A_C \cos(2\pi f_c t + \theta)] + E[A_C \cos(2\pi f_c t + 2\pi f_c \tau + \theta)\,Z(t)] + E[Z(t+\tau)\,Z(t)]\}$$

$$= (A_C^2/2)\cos(2\pi f_c \tau) + R_Z(\tau)$$

which shows that $R_X(\tau)$ contains a sinusoidal component of the same frequency as $X(t)$.

Example -6



The square wave $x(t)$ of figure 2.7 of constant amplitude A, period T_O, and delay t_d, represents the sample function of a random process $X(t)$. The delay is random, described by the probability density function

$$f_{T_d}(t_d) = \begin{cases} \frac{1}{T_O} & -\frac{1}{2}T_O \le t_d \le \frac{1}{2}T_O \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine the probability density function of the random variable $X(t_k)$ obtained by observing the random process $X(t)$ at time t_k.

(b) Determine the mean and auto correlation function of $X(t)$ using ensemble averaging.

(c) Determine the mean and autocorrelation function of $X(t)$ using time averaging.

(d) Establish whether or not $X(t)$ is stationary. In what sense is it ergodic?

Fig. 2.7.

Solution:

(*a*) We note that the distribution function of $X(t)$ is

$$f_{X(t)}(X) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \le x \le A \\ 1 & A < x \end{cases}$$

and the corresponding probability density function is

$$f_{X(t)}(X) = \frac{1}{2}\,\delta(x) + \frac{1}{2}\,\delta(x - A)$$

which are illustrated below:

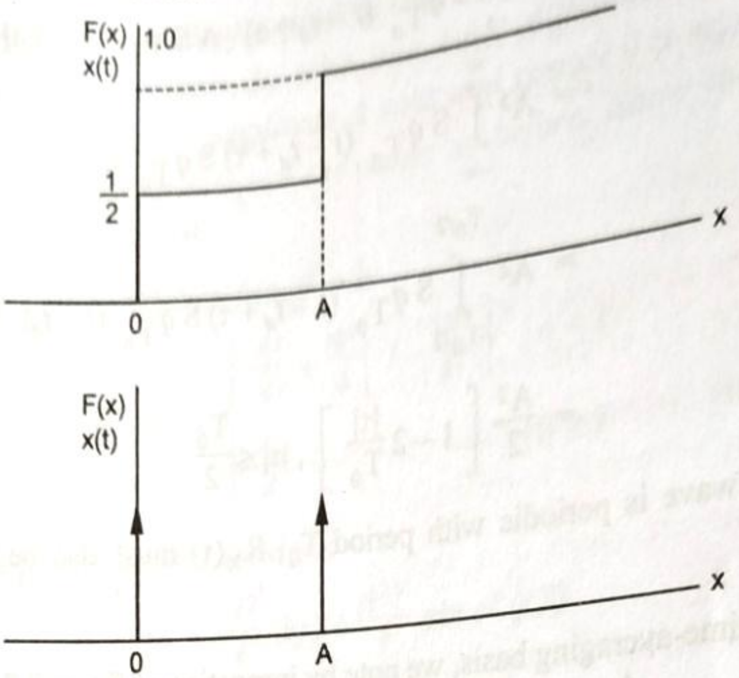


Fig. 2.8.

(*b*) By ensemble averaging, we have

$$E[X(t)] = \int_{-\infty}^{\infty} x\, f_{X(t)}\,(x)\, dx \;=\; \int_{-\infty}^{\infty} x \left[\frac{1}{2}\;\delta(x) + \frac{1}{2}\;\delta(x - A)\right] dx \;=\; \frac{A}{2}$$

The autocorrelation function of X(*t*) is

$$R_X(\tau) \;=\; E[X(t + \tau)\, X(t)]$$

Define the square function $Sq_{T_0}\,(t)$ as the square-wave shown below:

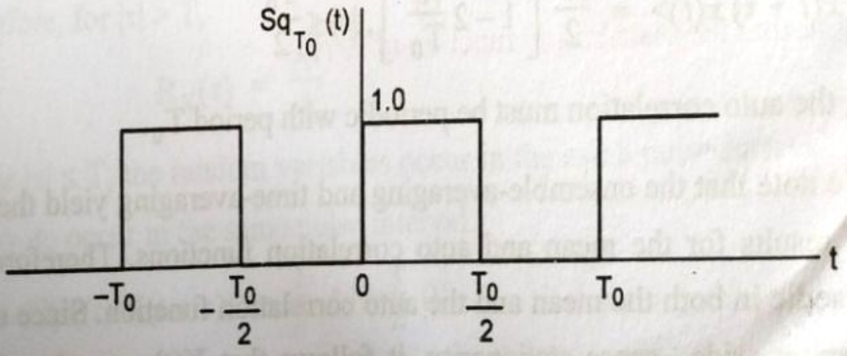


Fig. 2.9.

Then, we may write

$$R_X(\tau) = E[A\,S\,q_{T_0}\,(t - t_d + \tau) \cdot A\,S\,q_{T_0}\,(t - t_d)]$$

$$= A^2 \int_{-\infty}^{\infty} S\,q_{T_0}\,(t - t_d + \tau)\,S\,q_{T_0}\,(t - t_d)\,f_{T_d}\,(t_d)\,d\,t_d$$

$$= A^2 \int_{-T_0/2}^{T_0/2} S\,q_{T_0}\,(t - t_d + \tau)\,S\,q_{T_0}\,(t - t_d) \cdot \frac{1}{T_0}\,d\,t_d$$

$$= \frac{A^2}{2}\left[1 - 2\frac{|\tau|}{T_0}\right], |\tau| \le \frac{T_0}{2}$$

Since the wave is periodic with period T_0, $R_X(\tau)$ must also be periodic with period T_0.

(*c*) On a time-averaging basis, we note by inspection of figure 2.7 that the mean is $< x(t) > = \frac{A}{2}$

Next, the autocorrelation function

$$< x\,(t + \tau)\,x(t) > \; = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} x(t + \tau)\,x(t)\,dt$$

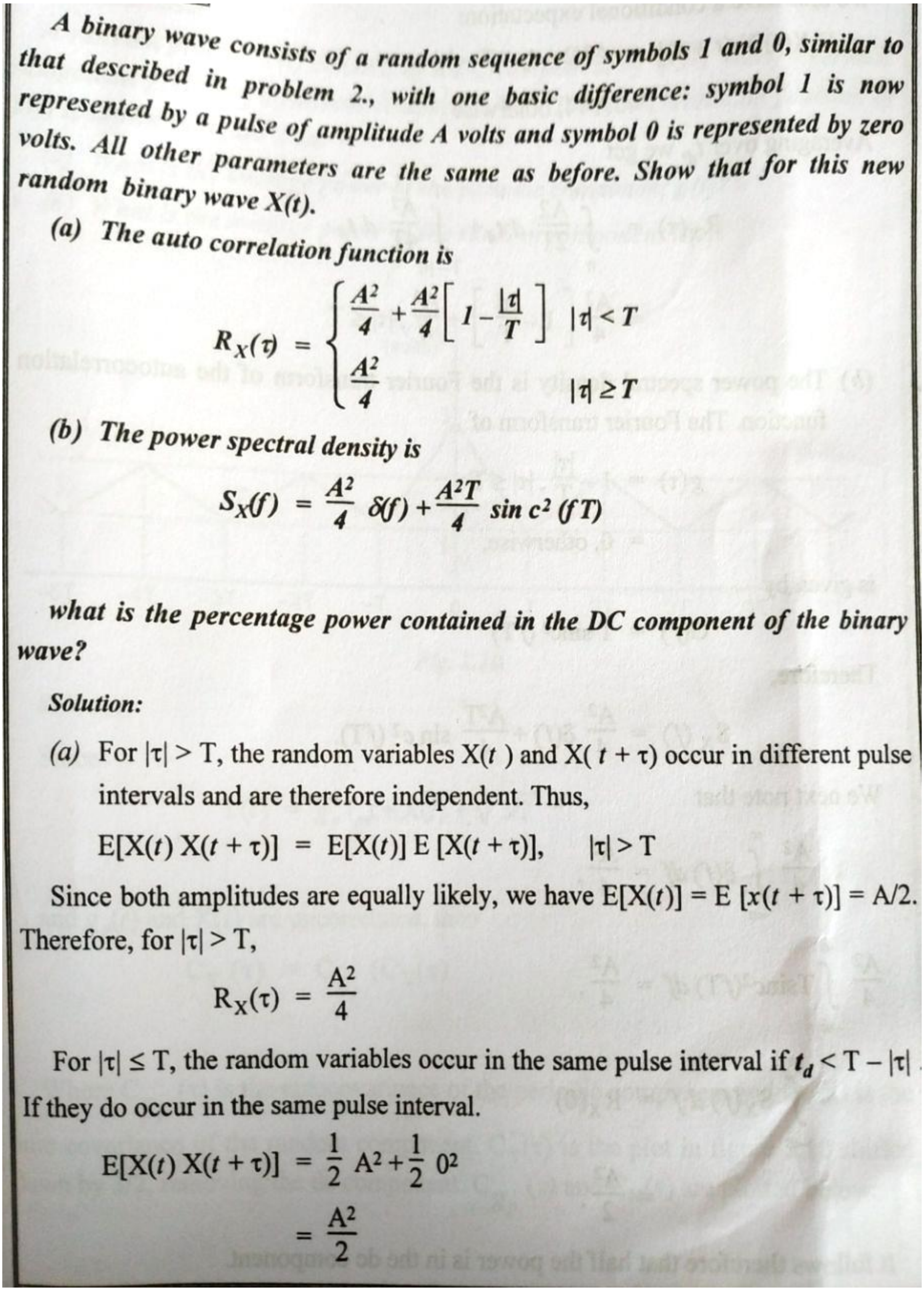
has its maximum value of $A^2/2$ at $\tau = 0$, and decreases linearly to zero at $\tau = T_0/2$. Therefore,

$$< x(t + \tau)\,x(t) > \; = \frac{A^2}{2}\left[1 - 2\frac{|\tau|}{T_0}\right], |\tau| \le \frac{T_0}{2}$$

Again, the auto correlation must be periodic with period T_0.

(*d*) We note that the ensemble-averaging and time-averaging yield the same set of results for the mean and auto correlation functions. Therefore, X(*t*) is ergodic in both the mean and the auto correlation function. Since ergodicity implies wide –sense stationarity, it follows that X(*t*) must be wide-sense stationary.

Example -7



A binary wave consists of a random sequence of symbols 1 and 0, similar to that described in problem 2., with one basic difference: symbol 1 is now represented by a pulse of amplitude A volts and symbol 0 is represented by zero volts. All other parameters are the same as before. Show that for this new random binary wave X(t).

(a) The auto correlation function is

$$R_X(\tau) = \begin{cases} \frac{A^2}{4} + \frac{A^2}{4}\left[1 - \frac{|\tau|}{T}\right] & |\tau| < T \\ \frac{A^2}{4} & |\tau| \geq T \end{cases}$$

(b) The power spectral density is

$$S_X(f) = \frac{A^2}{4}\,\delta(f) + \frac{A^2T}{4}\,\sin c^2\,(fT)$$

what is the percentage power contained in the DC component of the binary wave?

Solution:

(a) For $|\tau| > T$, the random variables $X(t)$ and $X(t + \tau)$ occur in different pulse intervals and are therefore independent. Thus,

$$E[X(t)\,X(t+\tau)] = E[X(t)]\,E\,[X(t+\tau)], \qquad |\tau| > T$$

Since both amplitudes are equally likely, we have $E[X(t)] = E\,[x(t+\tau)] = A/2$. Therefore, for $|\tau| > T$,

$$R_X(\tau) = \frac{A^2}{4}$$

For $|\tau| \leq T$, the random variables occur in the same pulse interval if $t_d < T - |\tau|$. If they do occur in the same pulse interval.

$$E[X(t)\,X(t+\tau)] = \frac{1}{2}\,A^2 + \frac{1}{2}\,0^2$$

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

We thus have a conditional expectation:

$$E[X(t)\,X(t+\tau)] = A^2/2, \quad t_d < T - |\tau|$$

$$= A^2/4, \text{ otherwise}$$

Averaging over t_d, we get

$$R_X(\tau) = \int_0^{T-|\tau|} \frac{A2}{2T}\, d\,t_d + \int_{T-|\tau|}^{T} \frac{A^2}{4T}\, d\,t_d$$

$$= \frac{A^2}{4}\left[1 - \frac{|\tau|}{T}\right] + \frac{A^2}{4},\ |\tau| \le T$$

(*b*) The power spectral density is the Fourier transform of the autocorrelation function. The Fourier transform of

$$g(\tau) = 1 - \frac{|\tau|}{T},\ |\tau| \le T$$

$$= 0, \text{ otherwise,}$$

is given by

$$G(f) = T\,\text{sinc}^2(fT)$$

Therefore,

$$S_X(f) = \frac{A^2}{4}\,\delta(f) + \frac{A^2T}{4}\,\sin c^2(fT).$$

We next note that

$$\frac{A^2}{4}\int_{-\infty}^{\infty} \delta(f)\, df = \frac{A^2}{4},$$

$$\frac{A^2}{4}\int_{-\infty}^{\infty} T\text{sinc}^2(fT)\, df = \frac{A^2}{4},$$

$$\int_{-\infty}^{\infty} S_X(f)\, d\,f = R_X(0)$$

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

It follows therefore that half the power is in the dc component.

Example -8

A random process $Y(t)$ consists of a DC component of $\sqrt{3/2}$ volts, a periodic component $g(t)$, and a random component $X(t)$. The autocorrelation function of $Y(t)$ is shown in figure 2.10.

(a) What is the average power of the periodic component $g(t)$?

(b) What is the average power of the random component $X(t)$?

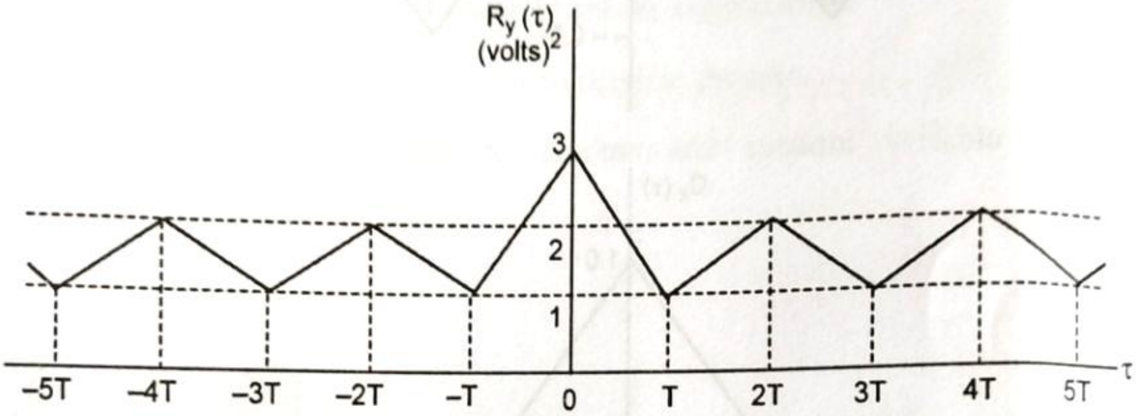


Fig. 2.10.

Solution:

Since

$$Y(t) = g_p(t) + X(t) + \sqrt{3/2}$$

and $g_p(t)$ and $X(t)$ are uncorrelated, then

$$C_Y(\tau) = C_{g_p} \; (C_X(\tau)$$

Where $C_{g_p}(\tau)$ is the autocovariance of the periodic component and $C_X(\tau)$ is the auto covariance of the random component. $C_Y(\tau)$ is the plot in figure 2.10 shifted down by 3/2, removing the dc component. $C_{g_p}(\tau)$ and $C_X(\tau)$ are plotted below:

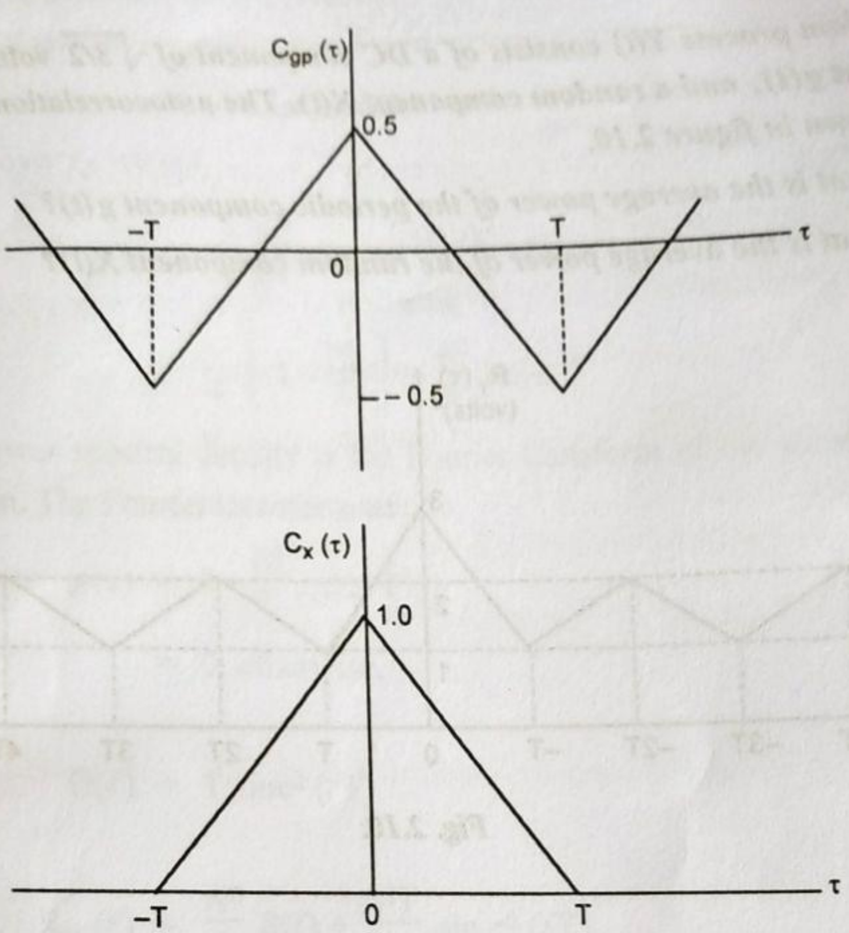


Fig. 2.11.

Both $g_p(t)$ and $X(t)$ have zero mean,

(*a*) The average power of the periodic component $g_p(t)$ is therefore,

$$\frac{1}{T_0}\int_{-T_0/2}^{T_0/2} g_p^2\,(t)\,dt = C_{g_p}\,(0) = \frac{1}{2}$$

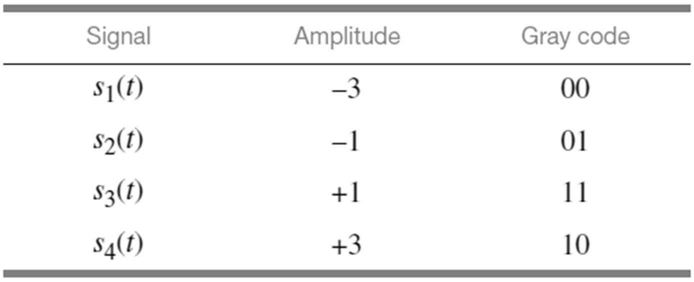
(*b*) The average power of the random component $X(t)$ is

$$E[X^2\,(t)] = C_X(0) = 1$$

Problems on Gram–Schmidt Orthogonalization Procedure

Example 1: **2B1Q Code**

The 2B1Q code is the North American line code for a special class of Modems called Digital Subscriber Lines. This code represents a Quaternary PAM signal in the Gray-encoded alphabet as shown below.

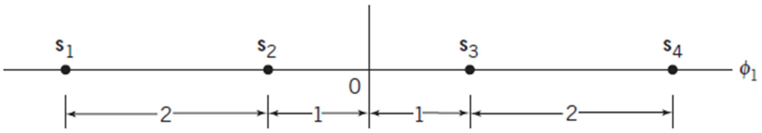


Signal	Amplitude	Gray code
$s_1(t)$	–3	00
$s_2(t)$	–1	01
$s_3(t)$	+1	11
$s_4(t)$	+3	10

Amplitude levels of the 2B1Q code

The four possible signals $s_1(t)$, $s_2(t)$, $s_3(t)$,and $s_4(t)$, are amplitude-scaled versions of a Nyquist pulse. Each signal represents a *dibit* (i.e., pair of bits). The issue of interest is to find the vector representation of the 2B1Q code.

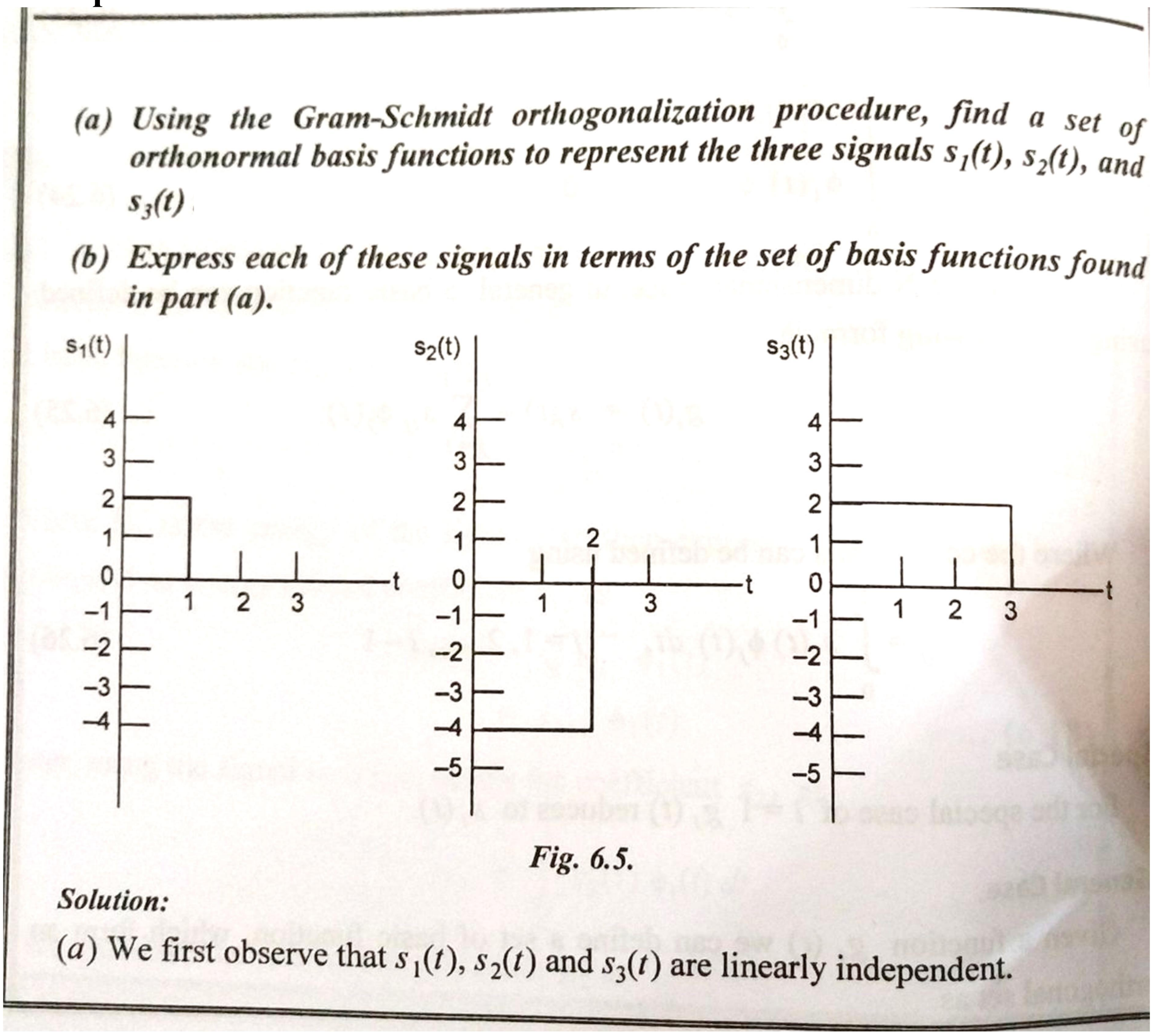
This example is simple enough to solve it by inspection. Let $\phi_1(t)$ denote a pulse normalized to have unit energy. The $\phi_1(t)$ so defined is the only basis function for the vector representation of the 2B1Q code. Accordingly, the signal-space representation of this code consists of four signal vectors $\mathbf{s}_1$, $\mathbf{s}_2$, $\mathbf{s}_3$, and $\mathbf{s}_4$, which are located on the $\phi_1(t)$ axis in a symmetric manner about the origin. In this example, $M = 4$ and $N = 1$.



Signal-space representation of the 2B1Q code

Generalize the result depicted in the diagram for the 2B1Q code as follows: the signal-space diagram of an M-ary PAM signal, in general, is one-dimensional with M signal points uniformly positioned on the only axis of the diagram.

Example 2



(a) Using the Gram-Schmidt orthogonalization procedure, find a set of orthonormal basis functions to represent the three signals $s_1(t)$, $s_2(t)$, and $s_3(t)$.

(b) Express each of these signals in terms of the set of basis functions found in part (a).

Fig. 6.5.

Solution:

(a) We first observe that $s_1(t)$, $s_2(t)$ and $s_3(t)$ are linearly independent.

The energy of $s_1(t)$ is

$$E_1 = \int_0^1 (2)^2 \, dt = 4$$

The first basis function is therefore

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$$

$$= \begin{cases} 1, & 0 \le t \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Define

$$s_{21} = \int_0^T s_2(t)\, \phi_1(t)\, dt$$

$$= \int_0^T (-4)\,(1)\, dt = -4$$

$$g_2(t) = s_2(t) - s_{21}\, \phi_1(t)$$

$$= \begin{cases} -4, & 1 \le t \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Hence, the second basis function is

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)\, dt}}$$

$$= \begin{cases} -1, & 1 \le t \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Define

$$s_{31} = \int_0^T s_3(t)\, \phi_1(t)\, dt$$

$$= \int_0^1 (3)\,(1)\, dt = 3$$

$$s_{32} = \int_{T}^{2T} s_3(t)\,\phi_2(t)\,dt$$

$$= \int_{1}^{2} (3)(-1)\,dt = -3$$

$$g_3(t) = s_3(t) - s_{31}\,\phi_1(t) - s_{32}\,\phi_2(t)$$

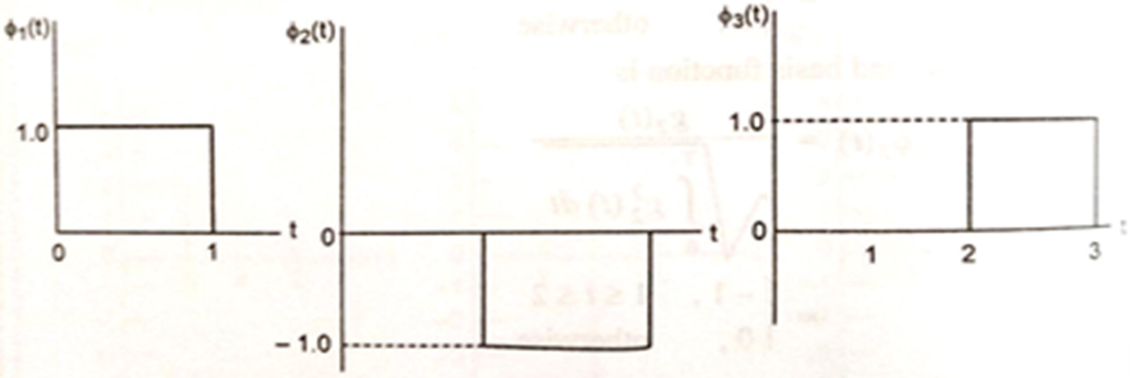
$$= \begin{cases} 3, & 2 \le t \le 3 \\ 0, & \text{otherwise} \end{cases}$$

Hence, the third basis function is

$$\phi_3(t) = \frac{g_3(t)}{\sqrt{\int_{0}^{T} g_3^2(t)\,dt}}$$

$$= \begin{cases} 1, & 2 \le t \le 3 \\ 0, & \text{otherwise} \end{cases}$$

The three basis functions are as follows (graphically)



(b)

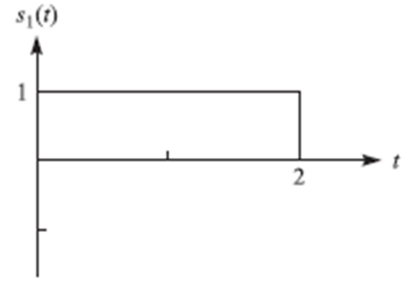
$$s_1(t) = 2\,\phi_1(t)$$

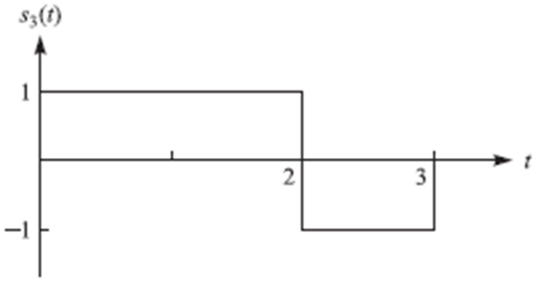
$$s_2(t) = -4\,\phi_1(t) + 4\,\phi_2(t)$$

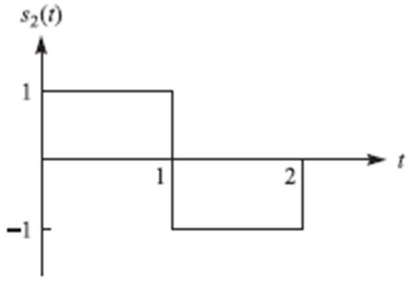
$$s_3(t) = 3\,\phi_1(t) - 3\,\phi_2(t) + 3\,\phi_3(t)$$

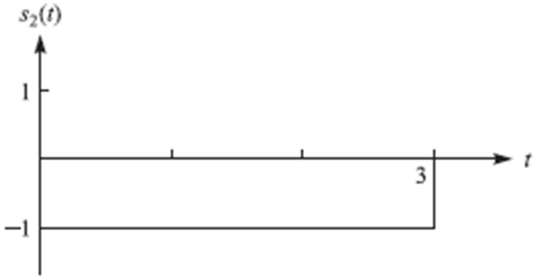
Example 3

Apply the Gram-Schmidt procedure to the set of four waveforms shown in diagram below.









<u>Sol</u>:

Inspect the waveforms and write the equations.

∴ $s_1(t) = 1 , 0 \le t \le 2$

∴ $s_2(t) = \begin{cases} 1, & 0 \le t \le 1 \\ -1, & 0 \le t \le 2 \end{cases}$

∴ $s_3(t) = \begin{cases} 1, & 0 \le t \le 2 \\ -1, & 2 \le t \le 2 \end{cases}$

∴ $s_4(t) = -1 , 0 \le t \le 3$

To find the Energy $E_1$of signal $s_1(t)$, apply eqn(8),

$$E_i = \int_0^T s_i^2(t)\, dt, \qquad i = 1, 2, \ldots, M$$

∴ $E_1 = \int_0^T s_1^2\,(t).dt = \int_0^2 1\,.dt = 2$

Apply eqn(17),

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$$

∴ $\phi_1(t) = \frac{1}{\sqrt{2}}\, s_1(t)$

$$s_1(t) = \sqrt{E_1}\,\phi_1(t) \qquad = s_{11}(t)\phi_1(t)$$

∴ $s_{11} = 2.\frac{1}{\sqrt{2}} = \sqrt{2}$

∴ $\int_0^T \phi_1^2\,(t).\,dt = \int_0^2 1/2\,.dt = 1$ → Unit Energy Signal!

Next, use signal $s_2(t)$,

$$s_{21} = \int_0^T s_2(t)\phi_1(t)\,\mathrm{d}t$$

Use eqn(18),

$$g_2(t) = s_2(t) - s_{21}\phi_1(t)$$ - Intermediate function

∴ $s_{21} = \frac{1}{\sqrt{2}}\int_0^1 1\,.\,dt + \frac{1}{\sqrt{2}}\int_0^1 -1\,.\,dt = 0$

$s_2(t)$ and $\phi_1(t)$ are orthogonal.

∴ $g_2(t) = s_2(t)$

Use eqn(20),

$$\phi_2(t) = \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{E_2 - s_{21}^2}}$$

∴ $E_2 = \int_0^1 1^2\,.\,dt + \frac{1}{\sqrt{2}}\int_0^1 -1^2\,.\,dt = 2$

∴ $\phi_2(t) = \frac{1}{\sqrt{2}}.\ s_2(t)$

∴ $\int_0^T \phi_2^2\,(t).\,dt = \int_0^1 (\frac{1}{2})\,.\,1.\,dt + \int_1^2 (\frac{1}{2})\,.\,1.\,dt = 1$ → Unit Energy functions!

∴ $\int_0^T \phi_1(t)\phi_2\,(t).\,dt = \int_0^1 \left(\frac{1}{\sqrt{2}}\right).\left(\frac{1}{\sqrt{2}}\right).\,dt - \int_0^1 \left(\frac{1}{\sqrt{2}}\right).\left(\frac{1}{\sqrt{2}}\right).\,dt = 0$ → Orthogonal functions!

Next higher order coefficients,

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,\mathrm{d}t, \qquad j = 1, 2, \ldots, i-1$$

∴ $s_{31} = \int_0^T s_3(t)\phi_1\,(t).\,dt = \int_0^2 1.\,(\frac{1}{\sqrt{2}})\,.\,dt + \int_2^3 (0)\,.\,-1.\,dt = \sqrt{2}$

∴ $s_{32} = \int_0^T s_3(t)\phi_2\,(t).\,dt = \int_0^1 1.\,(\frac{1}{\sqrt{2}})\,.\,1.\,dt + \int_1^2 1.\,(\frac{1}{\sqrt{2}})\,.\,-1.\,dt + \int_2^3 0\,.\,-1.\,dt = 0$

Use eqn(21),

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t)$$

∴ $g_3(t) = s_3(t) - s_{31}\,\phi_1(t) - s_{32}\,\phi_2(t) = s_3(t) - \sqrt{2}.\phi_1(t) = \begin{cases} -1, & 2 \le t \le 3 \\ 0, & \text{otherwise} \end{cases}$

Use eqn(22),

$$\phi_i(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)\,\mathrm{d}t}}, \qquad j = 1, 2, \ldots, N$$

∴ $\int_0^T g_3^2\,(t).\,dt = \int_0^2 [s_3(t) - \sqrt{2}\left(\frac{1}{\sqrt{2}}\right).\,1]^2\ .\,dt + \int_2^3 [s_3(t)]^2\ .\,dt$

$= \int_0^2 (1 - 1)\,.\,dt + \int_2^3 1\ .\,dt = 1$

∴ $\phi_3(t) = s_3(t) - \sqrt{2}\,\phi_2(t) = \begin{cases} -1, & 2 \le t \le 3 \\ 0, & \text{otherwise} \end{cases}$

Check that $\phi_3(t)$ is a unit energy function.

Check $\phi_3(t)$ is orthogonal to $\phi_1(t)$ and $\phi_2(t)$, respectively.

Next use signal $s_4(t)$,

Use eqn(20), eqn(21)and eqn(22),

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,\mathrm{d}t, \qquad j = 1, 2, \ldots, i-1$$

∴ $s_{41} = \int_0^T s_4(t)\phi_1(t).dt = \int_0^2 -1.(\frac{1}{\sqrt{2}}).1.dt + \int_2^3 0\ .-1.dt = -\sqrt{2}$

∴ $s_{42} = \int_0^T s_4(t)\phi_2(t).dt = \int_0^2 -1.(\frac{1}{\sqrt{2}}).1.dt + \int_2^3 0\ .-1.dt = 0$

∴ $s_{43} = \int_0^T s_4(t)\phi_3(t).dt = \int_2^3 -1\ .-1.dt = 1$

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t)$$

∴ $g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t).dt$

∴ $g_4(t) = s_4(t) - \sum_{j=1}^{3} s_{4j}\phi_j(t).dt$

$= s_4(t) - \sqrt{2}\,\phi_2(t) - \phi_3(t)$

$= (-1)_0^2 + (\sqrt{2}\ (1).(\frac{1}{\sqrt{2}})\ )_0^2 - (-1)_2^3$

$= (-1)_0^2 + (1)_0^2 + (-1)_2^3 + (1)_2^3 = 0$

∴ $\phi_4(t) = 0$

Final answers for the orthogonal basis functions,

∴ $\phi_1(t) = \frac{1}{\sqrt{2}}, 0 \le t \le 2$

∴ $\phi_2(t) = \begin{cases} \frac{1}{\sqrt{2}}, & 0 \le t \le 1 \\ -\frac{1}{\sqrt{2}}, & 0 \le t \le 2 \end{cases}$

∴ $\phi_3(t) = \begin{cases} -1, & 2 \le t \le 3 \\ 0, & \text{otherwise} \end{cases}$

∴ $\phi_4(t) = 0$

ECT305 ADC QB MODULE V

1.

Calculate the BER of BFSK system for the following parameters.
PSD of white noise $N_0/2 = 10^{-10}$ W/Hz, Amplitude of carrier is 1 mV at receiver input and Frequency of baseband NRZ signal is f_b= 1 KHz.

2.

Determine the minimum bandwidth of a BPSK modulator with a carrier frequency of 40 MHz and an input bit rate of 500 kbps. For a minimum BER of 10^{-4}, estimate the signal power.

3.

Binary data has to be transmitted over a telephone link that has a usable bandwidth of 3000Hz and a maximum achievable signal-to-noise power ratio of 6 dB at its output.

a. Determine the maximum signalling rate and probability of error if a coherent ASK scheme is used for transmitting binary data through this channel.

b. If the data is maintained at 300 bits/sec, calculate the error probability.

4.

Binary data is transmitted over an RF band pass channel with a usable bandwidth of 10 MHz at a rate of $4.8x10^{6}$ bits/sec using an ASK signalling method. The carrier amplitude at the receiver antenna is 1 mv and the noise power spectral density at the receiver input is 10^{-15} Watt/Hz. Find the error probability of a coherent and non-coherent receiver.

5.

Binary data is transmitted at a rate of 10^6 bits/sec over a microwave link having a bandwidth of 3 MHz. Assume that the noise power spectral density at the receiver input is $10^{-10}\ W/Hz$. Find the average carrier power required at the receiver input for coherent PSK and DPSK signalling schemes to maintain $P_e \leq 10^{-4}$.

6. From first principles, derive the Probability of Bit Error or the BER for Non-coherent detection of BFSK signal.

7. From first principles, derive the Probability of Bit Error or the BER for Non-coherent detection of DPSK signal.

8. From first principles, derive the Probability of Bit Error or the BER for QPSK signal.

-

9. From first principles, derive the Probability of Bit Error or the BER for QAM signal.

-

10. Draw block diagrams and signal space diagrams and explain the transmitter and receiver for BPSK.

-

11. Draw block diagrams and signal space diagrams and explain the transmitter and receiver for QPSK.

12. Draw block diagrams and signal space diagrams and explain the transmitter and receiver for QAM.

13. From first principles, derive the Probability of Bit Error or the BER for BPSK signal.

-

JOYOUS WELCOME TO SEMESTER-5 ECE!!

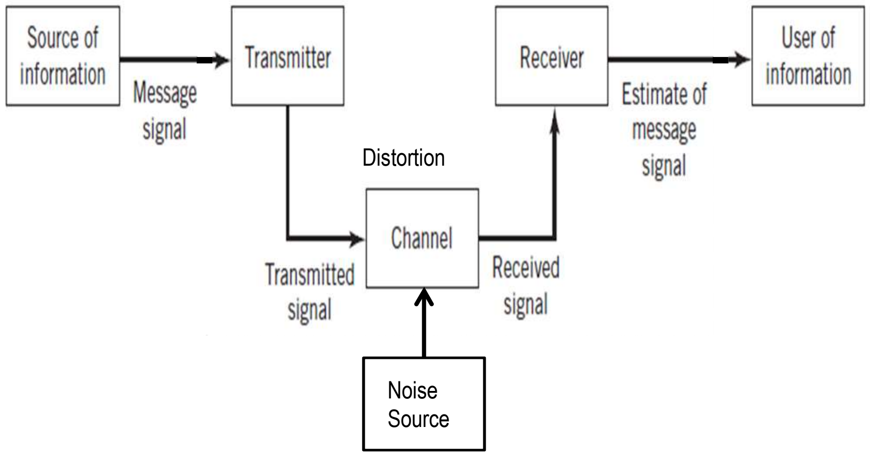
WELCOME TO THE WONDERFUL WORLD OF **ANALOG AND DIGITAL COMMUNICATION**

ECT 305 ADC
P.RAJKUMAR
PROFESSOR
ECE DEPARTMENT

Analog Communication

ECT 305 Analog and Digital Communication
Module I Part-1

Block diagram of a Communication System



Communication System

- There are three main elements to every communication system, namely, Transmitter, Channel, and Receiver.
- The Transmitter is located at one point in space, the Receiver is located at some other point separate from the transmitter, and the Channel is the physical medium that connects them together as an **integrated communication system.**
- The purpose of the Transmitter is to convert the message signal produced by the source of information into a form suitable for transmission over the Channel.
- However, as the transmitted signal propagates along the Channel, it is distorted due to channel imperfections.
- Additionally, Noise and Interfering signals are added to the channel output, so the received signal is a corrupted version of the transmitted signal.
- The Receiver works on the received signal and reconstructs a recognizable form of the original message signal for an end user of the information.

Information Source

- The information comes from the information source, which originates it.
- Information is a very generic word signifying at the abstract level anything intended for communication, which may include some thought, news, feeling, visual scene, and so on.
- The information source converts this information into a physical quantity.
- For example, speech signal, written script or picture.
- **This physical manifestation of the information is termed as message signal.**

- The message signal usually will be in the non-electrical form.
- For electrical communication purpose, need to convert the message signal to the electrical form, which is achieved using a suitable Transducer.
- **Transducer is a device which converts energy in one form to the other.**
- The speech signal is nothing but the acoustic pressure variations plotted as a function of time.
- These acoustic pressure variations are converted into electrical form using Microphone as the transducer.

Transmitter

- If the message arriving from the information source is electrical in nature, only then it will be suitable for immediate transmission.
- Even then, a lot of work must be done to make such a message suitable.
- It is necessary to convert the incoming sound signals into electrical variations, to restrict the range of the audio frequencies and then to compress their amplitude range.
- In wire telephony no processing may be required…
- …but in long-distance communications, a Transmitter is required to process, and possibly encode, the incoming information so as to make it suitable for transmission and subsequent reception.

- Eventually, in a Transmitter, the information modulates the Carrier, i.e., is superimposed on a high-frequency sine wave.
- The actual method of Modulation varies from one system to another.
- Modulation may be high level or low level, and the system itself may be amplitude modulation, frequency modulation, pulse modulation or any variation or combination of these, depending on the requirements.
- The Modulated signal from the Modulator is transmitted or radiated into the atmosphere using an Antenna as the transducer, which converts the signal energy in waveguide to free space Electromagnetic Waves.

Channel

- Channel is the physical medium which connects the Transmitter with the Receiver.
- The physical medium includes copper wire, coaxial cable, fibre optic cable, wave guide and free space or atmosphere.
- The choice of a particular channel depends on the feasibility and also the purpose of communication system.
- For speech communication among a group of people working in one physically localized place, copper wire may be the best choice.
- But, if the information needs to be sent to millions of people scattered in a geographical area like radio and television broadcasting, then free space or atmosphere is the best choice.

- The amount of message signal which finally reaches the receiver depends on the characteristics of the channel.
- The channel also refer to the frequency range allocated to a particular service or transmission, such as television channel and the allowable carrier bandwidth with modulation.

Noise Sources

- Noise is a random signal with non-deterministic amplitude and frequency range.
- Noise is defined as a random, unwanted signal that corrupts the information bearing signal.
- Noise originates in the channel.
- Examples of Noise, include Thermal noise due to Brownian motion in conductors and media, Shot Noise or Impulse Noise, the ‘dreaded’ AWGN, Shadow or ‘ghost’ noise, Fading etc…

Receiver

- The Receiver gets the incoming modified version of the message signal from the channel and processes it to recreate the original form of the message signal.
- There are many types of Receivers in Communication Systems, depending on the processing required to recreate the original message signal and also final presentation of the message to the destination.
- Most of the receivers broadly come under to the Super heterodyne type.
- Among the different processing steps employed, Demodulation is the most important one which extracts the message signal available in the received signal to the original electrical version of the message.
- So Demodulation is an inverse operation of Modulation.

- The purpose of Receiver and form of Output Display influence its construction as much as the type of modulation system used.
- Accordingly the Receiver can be a very simple Crystal Receiver, with headphones, to a far more complex RADAR Receiver, with its involved Antenna arrangements and visual display system.
- The output of a Receiver may be fed to a Loud speaker, Video display unit, Teletypewriter, various RADAR displays, Television picture tube, Pen recorder or Computer.
- For each case different arrangements must be made, each affecting the Receiver design.
- Note that the Transmitter and Receiver must be in agreement with Modulation methods used.

Destination

- The Destination is the final block in the communication system which receives the message signal and processes it to comprehend the information present in it.
- Usually, Humans will be the destination block.
- The incoming message signal via speech mode is processed by the speech perception system to comprehend the information.
- Similarly, the message signal via video or visual scene and written script is processed by the visual perception system to comprehend the information.

MODULATION

- The term modulate means regulate.
- The process of regulating is Modulation.
- Thus, for regulation we need one physical quantity which is to be regulated and another physical quantity which dictates regulation.
- In communication system, the signal to be regulated is termed as Carrier.
- The signal which dictates regulation is termed as Modulating signal.
- Message acts as Modulating signal.
- The Modulation process is the most important operation in the modern communication systems.

Example

- There is a special and rare cultural event from a reputed artist organized at a far distant place (destination city) from your actual location (source city).
- It is too far to reach the destination city by walking.
- However, you have decided to attend the event and enjoy the live performance.
- Then what will you do?
- The obvious choice is you will take the help of transportation to carry you from the source city to the destination city.
- Thus there are two important aspects to be observed in this example.
- The first one is you because you are the Message part.
- The second one is the transportation vehicle which is the Carrier.
- Once you reach the destination city, the purpose of the Carrier is served.

- Exactly similar situation is present in Electrical Communication.
- The message signal which is to be transmitted to the Receiver is like you and cannot travel for long distance by itself.
- Hence it should take the help of a Carrier which has the capacity to take the message to the Receiver.
- This is the basic aim of Modulation;
- So that Message can 'sit' on the Carrier and reach the Receiver.

Need for Modulation

1. Practical size of Antenna:

- The distance travelled by a signal in atmosphere is directly proportional to its frequency and inversely proportional to wavelength .
- Most of the message signals like Speech (300Hz – 3400Hz and Music are in the audio frequency range (20 Hz-20 KHz) and hence they can hardly travel for few meters on their own.
- For efficient Radiation and Reception, the transmitting and receiving Antennas would have to have length comparable to a Quarter-Wavelength of the Frequency used.
- For a message at 1 MHz, its wavelength is 300m ($3x10^8$/ $1x 10^6$) and hence antenna length should be about 75 m.
- Alternatively, for a signal at 15 kHz, the antenna length will be about 5000 m!
- A vertical antenna of this size is impracticable!

2. Message signals have frequency overlap

- It is not possible to transmit signal frequencies directly;
- All message is concentrated within the same range (20 Hz-20 kHz for speech and music, few MHz for video), so that all signals from the different sources would be hopelessly and inseparably mixed up.
- In any city, only one broadcasting station can operate at a given time.
- In order to separate the various signals, it is necessary to convert them all to different portions of the electromagnetic spectrum.
- Each must be given its own carrier frequency location.
- This also overcome the difficulties of poor radiation at low frequencies and reduces interference.
- Once signals have been translated, a tuned circuit is used in the front end of the receiver to make sure that the desired section of the spectrum is admitted and all unwanted ones are rejected.
- The tuning of such a circuit is normally made variable and connected to the tuning control, so that the receiver can select any desired transmission within a predetermined range.

Need for Modulation

1. Reduction in the height of antenna
2. Avoids mixing of signals
3. Increases the range of communication
4. Multiplexing is possible
5. Improves quality of reception

1. Reduction in the height of antenna

For the transmission of radio signals, the antenna height must be multiple of $\lambda/4$,where λ is the wavelength.

$\lambda = c /f$

where c : is the velocity of light

f: is the frequency of the signal to be transmitted

The minimum antenna height required to transmit a baseband signal of f = 10 kHz is calculated as follows: $3x10^8/(4x10x10^3)$= 7.5Km

The antenna of this height is practically impossible to install .

Now, let us consider a modulated signal at f = 1 MHz . The minimum antenna height is given by, $3x10^8/ 4x1x10^6$ = 75m

This antenna can be easily installed practically . Thus, modulation reduces the height of the antenna .

2. Avoids mixing of signals

- If the baseband sound signals are transmitted without using the modulation by more than one transmitter, then all the signals will be in the same frequency range i.e. 0 to 20 kHz .
- Therefore, all the signals get mixed together and a receiver can not separate them from each other .
- Hence, if each baseband sound signal is used to modulate a different carrier then they will occupy different slots in the frequency domain (different channels).
- Thus, modulation avoids mixing of signals .

3. Increase the Range of Communication

- The frequency of baseband signal is low, and the low frequency signals can not travel long distance when they are transmitted .
- They get heavily attenuated .
- The attenuation reduces with increase in frequency of the transmitted signal, and they travel longer distance .
- The modulation process increases the frequency of the signal to be transmitted.
- Therefore, it increases the range of communication.

4. Multiplexing is possible

- Multiplexing is a process in which two or more signals can be transmitted over the same communication channel simultaneously .
- This is possible only with modulation.
- The multiplexing allows the same channel to be used by many signals .
- Hence, many TV channels can use the same frequency range, without getting mixed with each other or different frequency signals can be transmitted at the same time .

5. Improves Quality of Reception

- With Single Side Band (SSB), Frequency Modulation (FM), the effect of noise is reduced to a great extent.
- This improves quality of reception.

Digital versus Analog Performance Criteria

- A principal difference between analog and digital communication systems has to do with the way in which we evaluate their performance.
- Analog systems draw their waveforms from a continuum, which therefore forms an infinite set—that is, a receiver must deal with an infinite number of possible waveshapes.
- The figure of merit for the performance of analog communication systems is a fidelity criterion, such as signal-to-noise ratio, percent distortion, or expected mean-square error between the transmitted and received waveforms.
- By contrast, a digital communication system transmits signals that represent digits.
- These digits form a finite set or alphabet, and the set is known a priori to the receiver.
- A figure of merit for digital communication systems is the probability of incorrectly detecting a digit, the probability of error (*PE*) or bit error rate (*BER*).

AMPLITUDE MODULATION

In amplitude modulation, the amplitude of a carrier signal is varied by the modulating voltage, whose frequency is invariably lower than that of the- carrier. In practice, the carrier may be high frequency (HF) while the modulation is audio. Formally, AM is defined as a system of modulation in which the amplitude of the carrier is made proportional to the instantaneous amplitude of the modulating voltage.

Let the carrier voltage and the modulating voltage, v_c and v_m respectively, be represented by,

$$v_c = V_c \sin \omega_c t \qquad \text{......... (1)}$$

$$v_m = V_m \sin \omega_m t \qquad \text{......... (2)}$$

From the definition of AM, you can see that the (maximum) amplitude V_c of the unmodulated carrier will have to be made proportional to the *instantaneous* modulating voltage $V_m \sin \omega_m t$ when the carrier is amplitude-modulated.

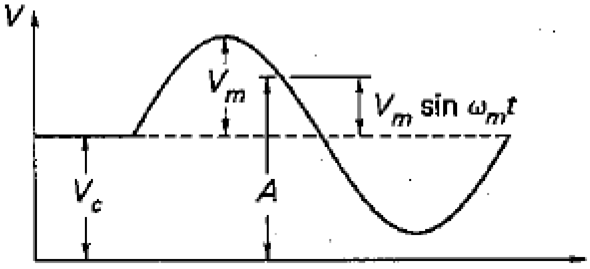


Fig-1 Amplitude of AM wave.

When a carrier is amplitude-modulated, the proportionality constant is made equal to unity, and the instantaneous modulating voltage variations are *superimposed* onto the carrier amplitude. Thus_ when there is temporarily no modulation, the amplitude of the carrier is equal to its unmodulated value. When modulation.is present, the amplitude of the carrier is varied by its instantaneous value. The situation is illustrated in Figure 3-1, which shows how the maximum amplitude of the amplitude-modulated voltage is made to vary.in accordance with modulating voltage changes. Figure 3-1 also shows that something unusual (distortion) will occur if *Vm* is greater than *Vc* (this distortion is a result of overdriving the amplifier stage). This, and the fact that the ratio *Vm/Vc* often occurs, leads to the definition of the *modulation index.*

$$m = \frac{V_m}{V_c} \qquad \text{......... (3)}$$

The modulation index is a number lying between 0 and 1, and it is very often expressed as a percentage and called the *percentage modulation.* From Fig-1 and Equation (3) it is possible to write an equation for the amplitude of the amplitude-modulated voltage.

$$A = V_c + v_m = V_c + V_m \sin \omega_m t = V_c = mV_c \sin \omega_m t$$

$$= V_c (1 + m \sin \omega_m t) \qquad \text{......... (4)}$$

The instantaneous voltage of the resulting amplitude-modulated wave is,

$$v = A \sin \theta = A \sin \omega_c t = V_c (1 + m \sin \omega_m t) \sin \omega_c t \quad \text{......... (5)}$$

Equation (3-6) may be expanded, by means of the trigonometric relation

$$\sin x \sin y = ½ [\cos (x - y) - \cos (x + y)],$$

$$v = V_c \sin \omega_c t + \frac{mV_c}{2} \cos (\omega_c - \omega_m)t - \frac{mV_c}{2} \cos (\omega_c + \omega_m)t \quad \text{......... (6)}$$

It has thus been shown that the equation of an amplitude-modulated wave contains three terms. The first term is represents the unmodulated carrier. It is apparent that the process of amplitude modulation has the effect of adding to the unmodulated wave, rather than changing it. The two additional terms produced are the two sidebands outlined. The frequency of the lower sideband (LSB) *is fc - fm* and the frequency of the upper sideband (USB) *is fc + fm.*

The very important conclusion to be made at this stage is that the bandwidth required for amplitude modulation is twice the frequency of the modulating signal. In modulation by several sine waves simultaneously, as in the AM broadcasting service, *the bandwidth required is twice the highest modulating frequency.*

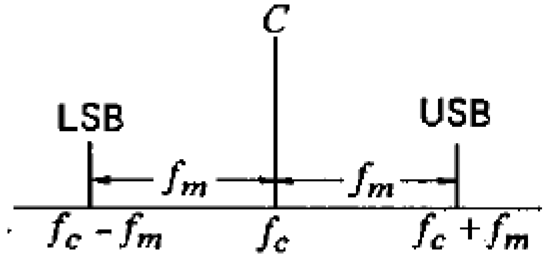


Fig-2 Frequency Spectrum of AM Wave

Representation of AM in the time domain:

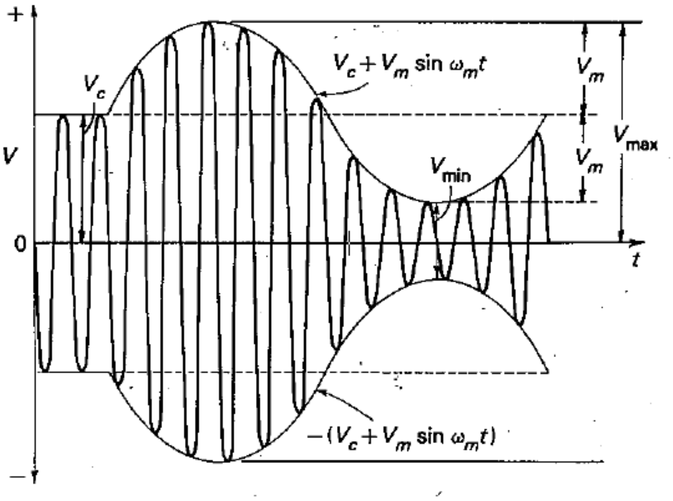


Fig-3 Amplitude Modulated Wave

Amplitude modulation may be represented in any of three ways, depending on the point of view. Fig-2 shows the frequency spectrum and so illustrates Eqn (6).

AM is shown simply as consisting of three discrete frequencies. Of these, the central frequency, i.e., the carrier, has the highest amplitude, and the other two are disposed symmetrically about it, having amplitudes which are equal to each other, but which can never exceed half the carrier amplitude [see Equation (3-7), and note that *m* ;j, l].

The appearance of the amplitude-modulated wave is of great interest, and it is shown in Fig-3 for one cycle of the modulating sine wave. It is derived from Fig-1, which showed the amplitude, or what may now be called the top *envelope* of the AM wave, given by the relation *A* = *Vc* + *Vm* sin *ωmt.* The maximum negative amplitude, or bottom envelope, is given by *-A* = *-(Vc+ Vm* sin *ωmt).* The modulated wave extends between these two limiting envelopes and has a repetition rate equal to the unmodulated carrier frequency. ·

It will be recalled that *Vm* = *mVc,* and it is now possible to use this relation to calculate the· index (or percent) of modulation from the waveform of Fig-3 as follows:

$$V_m = \frac{V_{\max} - V_{\min}}{2} \qquad \text{......... (7)}$$

and

$$V_c = V_{\max} - V_m = V_{\max} - \frac{V_{\max} - V_{\min}}{2}$$

$$= \frac{V_{\max} + V_{\min}}{2} \qquad \text{......... (8)}$$

$$m = \frac{V_m}{V_c} = \frac{(V_{\max} - V_{\min})/2}{(V_{\max} + V_{\min})/2}$$

$$= \frac{V_{\max} + V_{\min}}{V_{\max} + V_{\min}} \qquad \text{......... (9)}$$

The eqn-9 is the standard method of evaluating the modulation index when calculating from a waveform such as may be seen on an oscilloscope, i.e., when both the carrier and the modulating voltages are known. *It may not be used in any other situation.* When only the rms values ·of the carrier and the modulated voltage or current are known, or when the unmodulated and the modulated output powers are given, it is necessary to understand and use the power relations in the AM wave.

Finally, if the main interest is the instantaneous modulated voltage, the phasor diagrams depicting the three individual components of the AM wave may be drawn.

Power Relations in the AM Wave

It has been shown that the carrier component of the modulated wave has the same amplitude as the unmodulated carrier. The modulated wave contains extra energy in the two sideband components. Therefore, the modulated wave contains more power than the carrier had before modulation took place. Since the amplitude of the sidebands depends on the modulation index *VmlVc,* it is anticipated that the total power in the modulated wave will depend on the modulation index also. This relation may now be derived. ·

The total power in the modulated wave will be,

$$P_t = \frac{V_{\text{carr}}^2}{R} + \frac{V_{\text{LSB}}^2}{R} + \frac{V_{\text{USB}}^2}{R} \text{ (rms)} \qquad \text{......... (10)}$$

where all three voltages are (rms) values (V2 converted to peak), and *R* is the resistance,(e.g., antenna resistance), in which the power is dissipated. The first term of Eqn (10) is the unmodulated -carrier power and is given by,

$$P_c = \frac{V_{\text{carr}}^2}{R} = \frac{(V_c/\sqrt{2})^2}{R}$$

$$= \frac{V_c^2}{2R} \qquad \text{......... (11)}$$

Similarly,

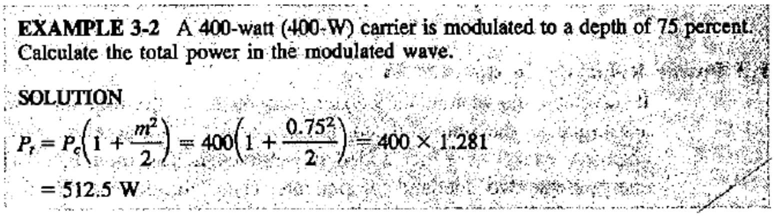
$$P_{\text{LSB}} = P_{\text{USB}} = \frac{V_{\text{SB}}^2}{R} = \left(\frac{mV_c/2}{\sqrt{2}}\right)^2 \div R = \frac{m^2V_c{}^2}{8R}$$

$$= \frac{m^2}{4}\frac{V_c^2}{2R} \qquad \text{......... (12)}$$

$$P_t = \frac{V_c^2}{2R} + \frac{m^2}{4}\frac{V_c^2}{2R} + \frac{m^2}{4}\frac{V_c^2}{2R} = P_c + \frac{m^2}{4}P_c + \frac{m^2}{4}P_c$$

$$\frac{P_t}{P_c} = 1 + \frac{m^2}{2} \qquad \text{......... (13)}$$

Eqn-13 relates the total power in the amplitude-modulated wave to the unmodulated carrier power. It is interesting to note from Eqn-13 that the maximum power in the AM wave is *Pt = 1.5P* when *m* = 1. This is important, because it is the maximum power that relevant amplifiers must be capable of handling without distortion.

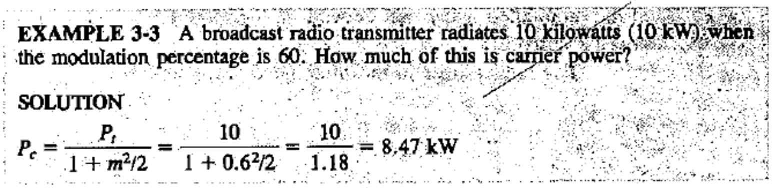


EXAMPLE 3-2 A 400-watt (400-W) carrier is modulated to a depth of 75 percent. Calculate the total power in the modulated wave.

SOLUTION

$$P_t = P_c\left(1 + \frac{m^2}{2}\right) = 400\left(1 + \frac{0.75^2}{2}\right) = 400 \times 1.281$$

$$= 512.5 \text{ W}$$



EXAMPLE 3-3 A broadcast radio transmitter radiates 10 kilowatts (10 kW) when the modulation percentage is 60. How much of this is carrier power?

SOLUTION

$$P_c = \frac{P_t}{1 + m^2/2} = \frac{10}{1 + 0.6^2/2} = \frac{10}{1.18} = 8.47 \text{ kW}$$

Double Sideband, Full Carrier (DSB-FC)

When a carrier is amplitude-modulated by a single sine wave, the resulting signal consists· of three frequencies, the original carrier frequency, the upper sideband frequency (fc+ fm) and the lower sideband frequency (fc - fm). This is an automatic consequence of the AM mixing process, and will always happen unless steps are taken to prevent it. Steps may be taken either during or after the modulation process, to remove or attenuate any combination of the components of the AM wave. It is the purpose of this chapter to deal with the factors involved in, and the advantages and disadvantages of, suppressing or removing the carrier and/or either of the sidebands. Generation of various forms of single-sideband modulation will also be considered.

The carrier of "standard" or double sideband, full carrier (DSBFC) AM (officially known as A3Emodulation) conveys no information. This is because the carrier component remains constant in amplitude and frequency, no matter what the modulating voltage does. Just as the fact that the two sidebands are images of each other, since each is affected by changes in the modulating voltage amplitude via the exponent mVc/2. All the information can be conveyed by the use of one sideband. The carrier is superfluous, and the other sideband is redundant. The main reason for the widespread· use of A3E is the relative simplicity of the modulating and demodulating equipment. Furthermore, A3E is the acceptable form used for broadcasting.

The AM power equation states that the ratio of total power to carrier power is given by (1+ m^2/2):1.

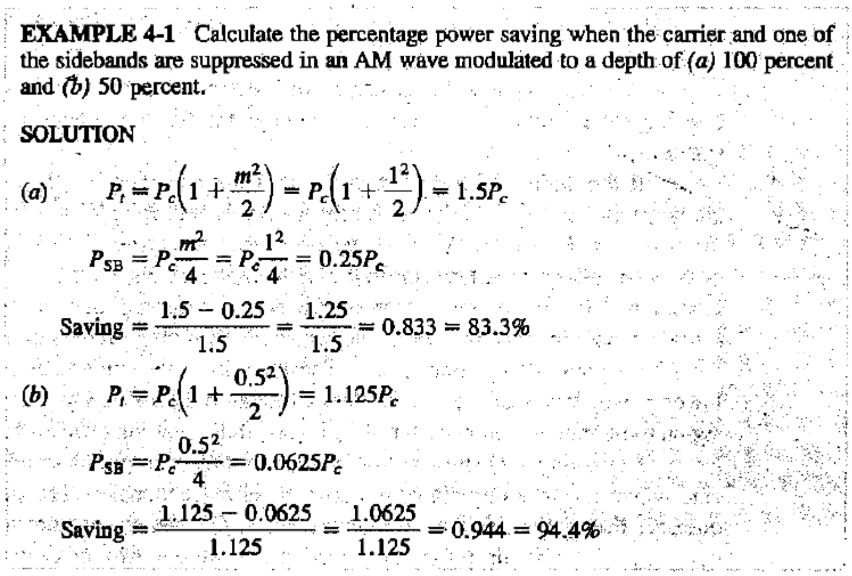
Double Sideband, Suppressed Carrier (DSB-SC)

If the carrier is suppressed, only the sideband power remains.

As this is only Pc (m^2/2), a two-thirds saving is affected at 100 percent modulation, and even more is saved as the depth of modulation is reduced.

Single Sideband (SSB)

If one of the sidebands is now also suppressed, the remaining power is Pc (m^2/4), a further saving of 50 percent over carrier-suppressed AM.



EXAMPLE 4-1 Calculate the percentage power saving when the carrier and one of the sidebands are suppressed in an AM wave modulated to a depth of *(a)* 100 percent and *(b)* 50 percent.

SOLUTION

(a)
$$P_t = P_c\left(1 + \frac{m^2}{2}\right) = P_c\left(1 + \frac{1^2}{2}\right) = 1.5P_c$$

$$P_{SB} = P_c\frac{m^2}{4} = P_c\frac{1^2}{4} = 0.25P_c$$

$$\text{Saving} = \frac{1.5 - 0.25}{1.5} = \frac{1.25}{1.5} = 0.833 = 83.3\%$$

(b)
$$P_t = P_c\left(1 + \frac{0.5^2}{2}\right) = 1.125P_c$$

$$P_{SB} = P_c\frac{0.5^2}{4} = 0.0625P_c$$

$$\text{Saving} = \frac{1.125 - 0.0625}{1.125} = \frac{1.0625}{1.125} = 0.944 = 94.4\%$$

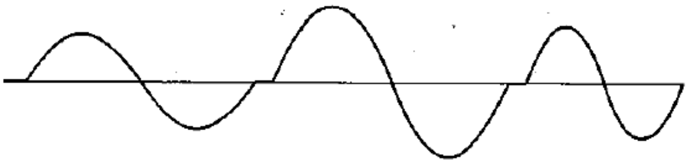
Example indicates how wasteful of power it is to send the carrier and both sidebands in situations in which only one sideband would suffice. A further check shows that the use of SSB immediately halves the bandwidth required for transmission, as compared with A3E.

In practice, SSB is used to save power in applications where such a power saving is warranted, i.e., in mobile systems, in which weight and power consumption must naturally be kept low.

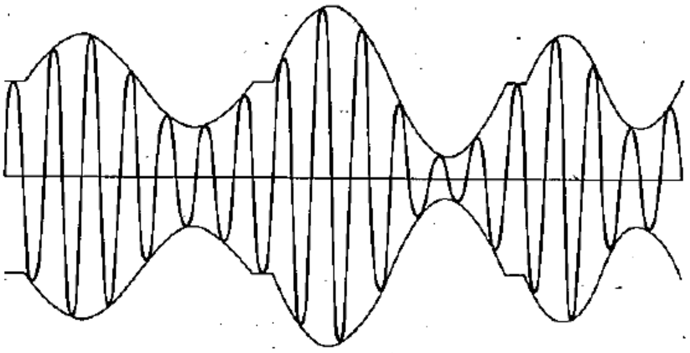
Single-sideband modulation is also used in applications in which bandwidth is at a premium. Point-to-point communications, land, air and mari**time mobile communications, television, telemetry, military communications, radio** navigation and amateur radio are the greatest users of SSB in one form or another.

Waveforms are illustrated in Fig-4, together with the modulating voltage, the corresponding AM voltage, and a wave with the carrier removed.

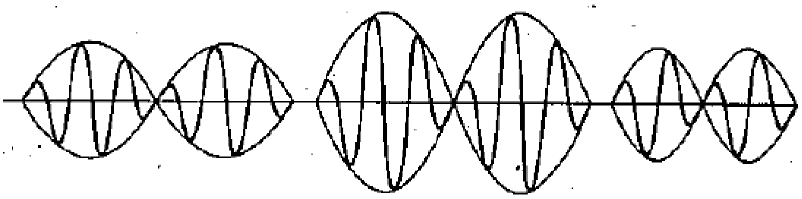
Two different modulating amplitudes and frequencies are shown for comparison. This demonstrates that here the SSB wave is a single radio frequency. Its amplitude is proportional to the amplitude of the modulating voltage, and its frequency varies with the frequency of the modulating signal. An upper sideband is shown so that its frequency increases with that of the modulation.



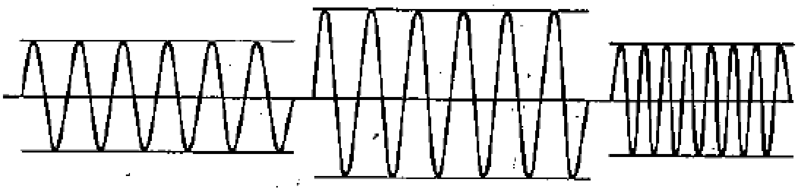
(*a*) Modulating signal



(*b*) AM wave **DSB-FC**



(*c*) Suppressed-carrier wave **DSB-SC**



(*d*) SSB suppressed-carrier wave **SSB**

Fig-4. Waveforms for various types of amplitude modulation, (a) Modulating signal; *(b)* AM wave; *(c)* Suppressed-carrier wave; (d) SSB suppressed-carrier wave.

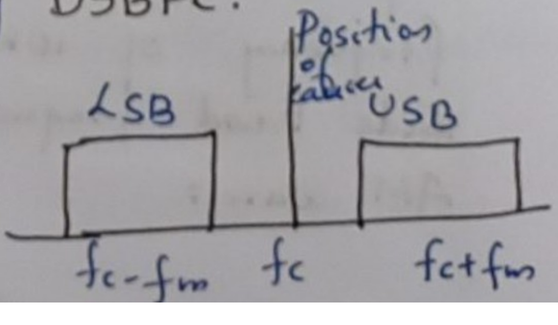
- Types of AM

1. DSB FC (A_3E): Double Side Band Full Carrier
2. DSB SC : Double Side Band Suppressed Carrier
3. SSB FC : Single Side Band Full Carrier
4. SSB SC : Single Side Band Suppressed Carrier
5. SSB RC : Single Side Band Reduced Carrier
6. ISB : Independent Side Band
7. VSB : Vestigal Side Band (used in TV broadcasting).

→ Double Sideband Suppressed Carrier (DSB-SC)s/h:

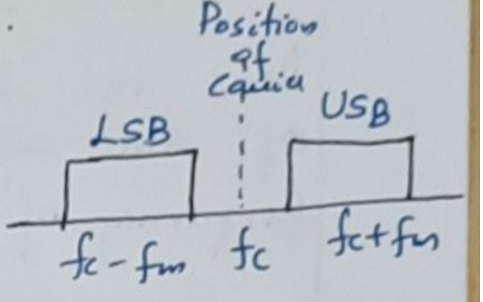
In the process of AM, the modulated wave consists of the carrier wave and two side bands. The modulated wave has the information only in the side bands. Side band is nothing but a band of frequencies, containing power, which are the lower and higher frequencies of the carrier frequency (f_{LSB} & f_{USB}).

The transmission of a signal, which contains a carrier along with two sidebands can be termed as Double Side Band Full Carrier system or simply DSBFC.



However, such a txm is inefficient. Because, 2/3 of the power is being wasted in the carrier, which carries no information.

If this carrier is suppressed and the saved power is distributed to the two sidebands, then such a process is called Double Sideband Suppressed Carrier system or simply DSBSC.



- Mathematical Expressions

Let modulating signal,

$$e_m(t) = E_m \cos(2\pi f_m t) \longrightarrow 37$$

Carrier signal, $e_c(t) = E_c \cos(2\pi f_c t) \longrightarrow (38)$

Mathematically we can represent the eq^n of DSBSC wave as the product of modulating and carrier signals.

$$s(t) = e_m(t)\, e_c(t) \longrightarrow (39)$$

$$\Rightarrow s(t) = E_m E_c \cos(2\pi f_m t) \cos(2\pi f_c t) \longrightarrow (40)$$

- Bandwidth of DSBSC Wave

Bandwidth, BW = $f_{max} - f_{min}$

Consider the eq^n of DSBSC modulated wave.

$$s(t) = E_m E_c \cos(2\pi f_m t) \cos(2\pi f_c t)$$

$$\Rightarrow s(t) = \frac{E_m E_c}{2} \cos[2\pi(f_c + f_m)t] + \frac{E_m E_c}{2} \cos[2\pi(f_c - f_m)t] \longrightarrow (41)$$

The DSB SC modulated wave has only 2 frequencies. So, the maximum and minimum frequencies are $f_c + f_m$ and $f_c - f_m$ respectively.

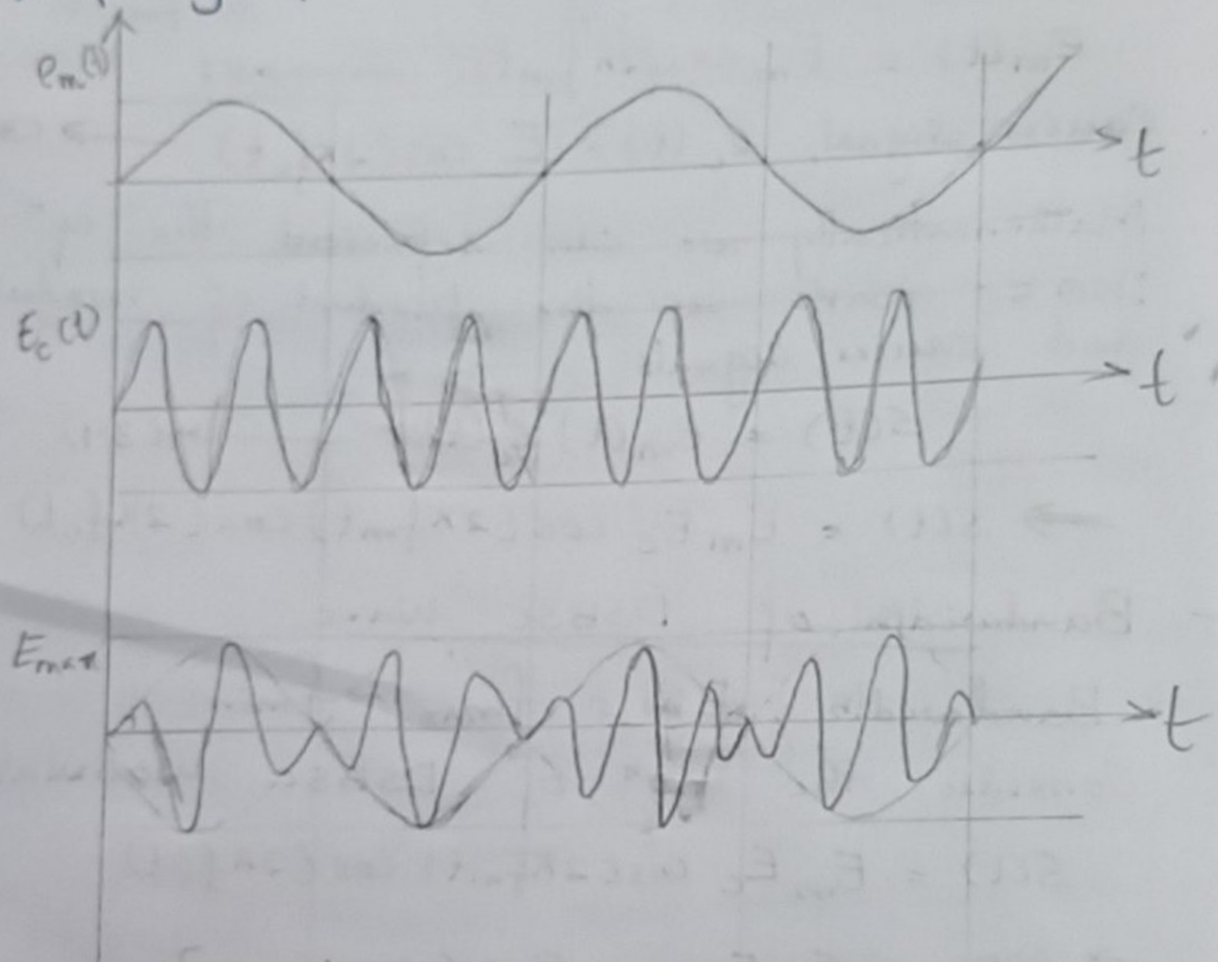
ie $f_{max} = f_c + f_m$ and $f_{min} = f_c - f_m$

Substitute f_{max} and f_{min} values in the BW formula

$$BW = (f_c + f_m) - (f_c - f_m)$$

$$BW = 2f_m \longrightarrow 42$$

Thus the BW of DSBSC wave is same as that of AM wave and it is equal to twice the frequency of modulating signal.



Few points

(1) It is obvious from the fig, that the DSBSC signal exhibits phase - reversal at zero crossings ie whenever the base-

band signal $x(t)$ crosses zero. Because of this, the envelope of a DSB-SC modulated signal is different from the message signal. This is unlike the case of an AM wave signal.

(2) From fig, it is also clear that the impulses at $\pm \omega_c$ are missing which means that the carrier term is suppressed in the spectrum and only two sideband terms, USB and LSB are left. Therefore, it is called double sideband suppressed carrier (DSB-SC) system.

(3) In fig, considering only positive side the upper side band frequency is $\omega_c + \omega_m$ whereas the lowerside band frequency is $\omega_c - \omega_m$. The difference of these two is equal to the transmission bandwidth of a DSB-SC signal i.e,

Bandwidth, $B = (\omega_c + \omega_m) - (\omega_c \mp \omega_m) = 2\omega_m$.

Note that BW of DSBSC is same as that of AM wave.

→ Single Sideband Suppressed - Carrier (SSB-SC)

If the twoside bands carry same information, DSB signal is redundant i.e in DSB, the basic information is transmitted twice, once in each side band. So one sideband may be suppressed. The carrier and one side band is completely suppressed. When only one side band is transmitted, the modulation is refered to as single side band modulation. It is also called as SSB or SSB-SC modulation.

The AM modulated signal from balanced modulator is given by

$$e(t) = K\, e_m(t) \cos \omega_c t \longrightarrow (43)$$

where K - multiple constant

$e_m(t)$ - modulating signal $E_m \cos \omega_m t$

$$\therefore\ e(t) = K E_m \cos \omega_m t \cos \omega_c t$$

$$= \frac{K E_m}{2} \left[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t \right]$$

$$e(t) = E_{max} \left[\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t \right] \rightarrow (44)$$

where $E_{max} = \frac{K E_m}{2}$

The upper side frequency (USF) signal is given by,

$$e_{USF} = E_m \cos(\omega_c + \omega_m)t \longrightarrow (45)$$

The lower side frequency (LSF) s/l is given by,

$$e_{LSF} = E_m \cos(\omega_c - \omega_m)t \longrightarrow (46)$$

One of these side band frequencies can be removed by filtering.

Demodulation of a SSB signal is achieved by multiplying it with a locally generated synchronous carrier signal. Detectors using this principle are called product detectors and balanced modulator circuits are used for this purpose.

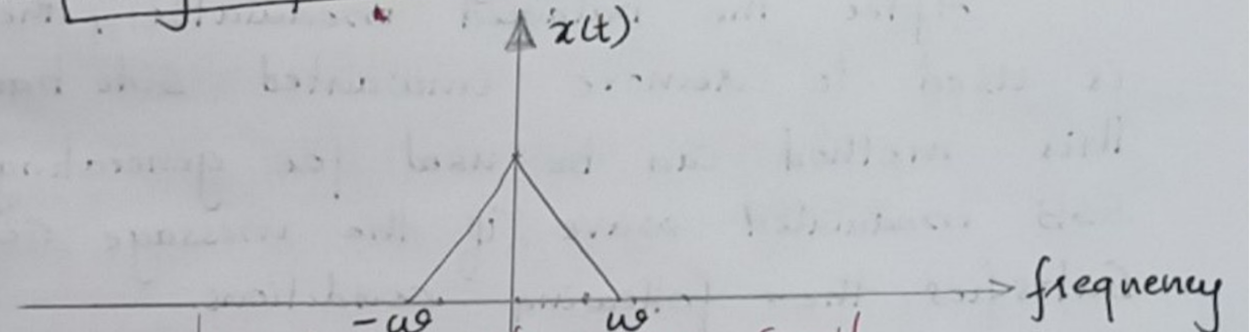
- Advantages of SSB-SC Modulation.

1. Reduction in transmission bandwidth.
2. Power saving since the high power carrier and one side band are not being transmitted.
3. Reduced Noise.

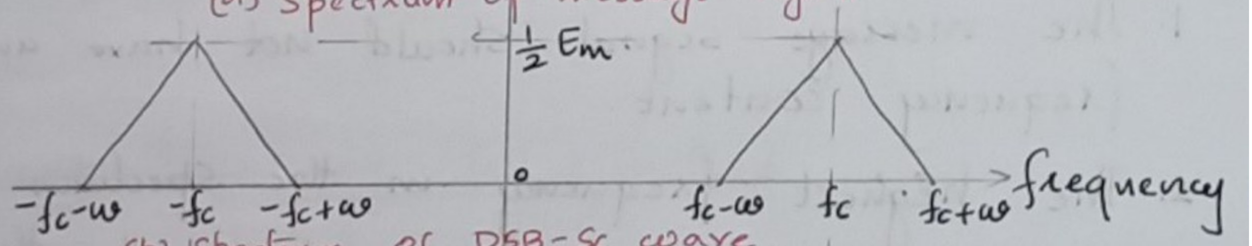
- Drawbacks of SSB-SC Modulation

1. It is expensive
2. Highly compler to implement.

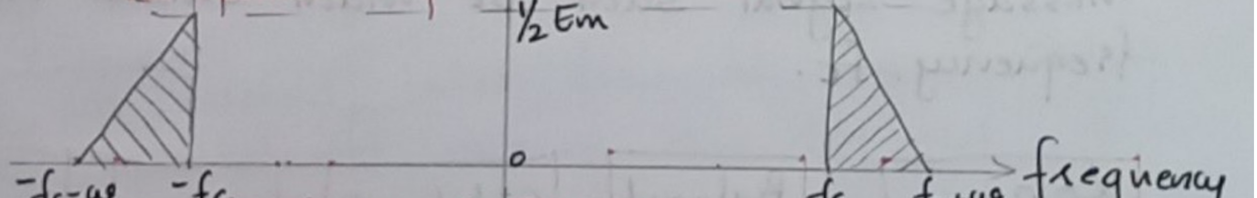
- Frequency spectrum of SSB-SC



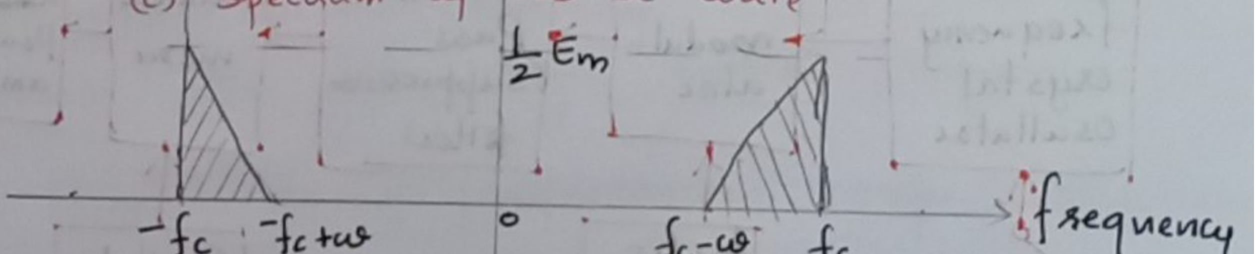
(a) spectrum of message signal



(b) Spectrum of DSB-Sc wave



(c) Spectrum of SSB-SC wave



(d) Spectrum of SSB-SC wave

→ Methods for generating SSB-SC signal.

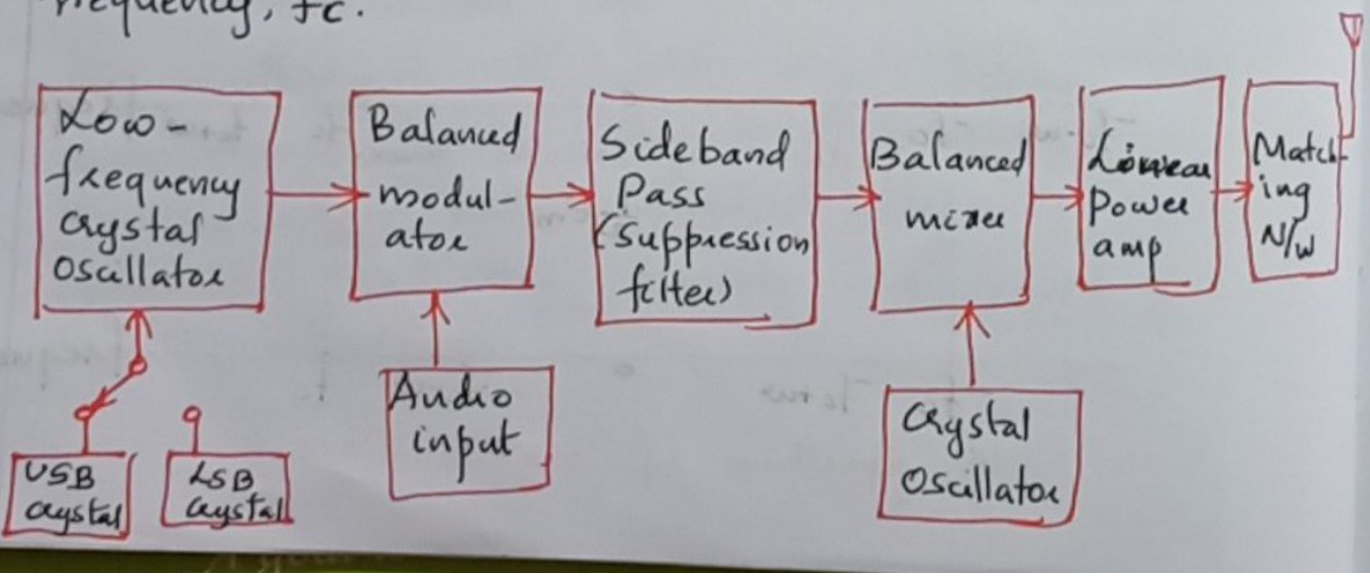
There are three practical methods to remove unwanted side band from the double side band signal to get the single side band signal. They are

1. Filter Method (Frequency discrimination method)
2. Phase shift method (Phasing method)
3. The third method (Weaver's method)

1. Filter Method: (Frequency discrimination method)

After the balanced modulator, the filter is used to remove unwanted side band. This method can be used for generating the SSB modulated wave, if the message signal satisfies the following conditions.

1. The message signal should not have any low frequency content.
2. The highest frequency in the spectrum of the message signal such be much smaller than carrier frequency, fc.



The balanced modulator is used to supress the carrier. Then filter suppresses one side band signal. The frequency of the generated single sideband signal is very low than the transmitter frequency. This frequency is boosted upto the transmitter frequency by the balanced mixer and crystal oscillator. The process of frequency boosting is also called up conversion. The sideband signal having frequency equal to the transmitter frequency is then amplified by the linear amplifier before transmission.

– Advantages of filter method

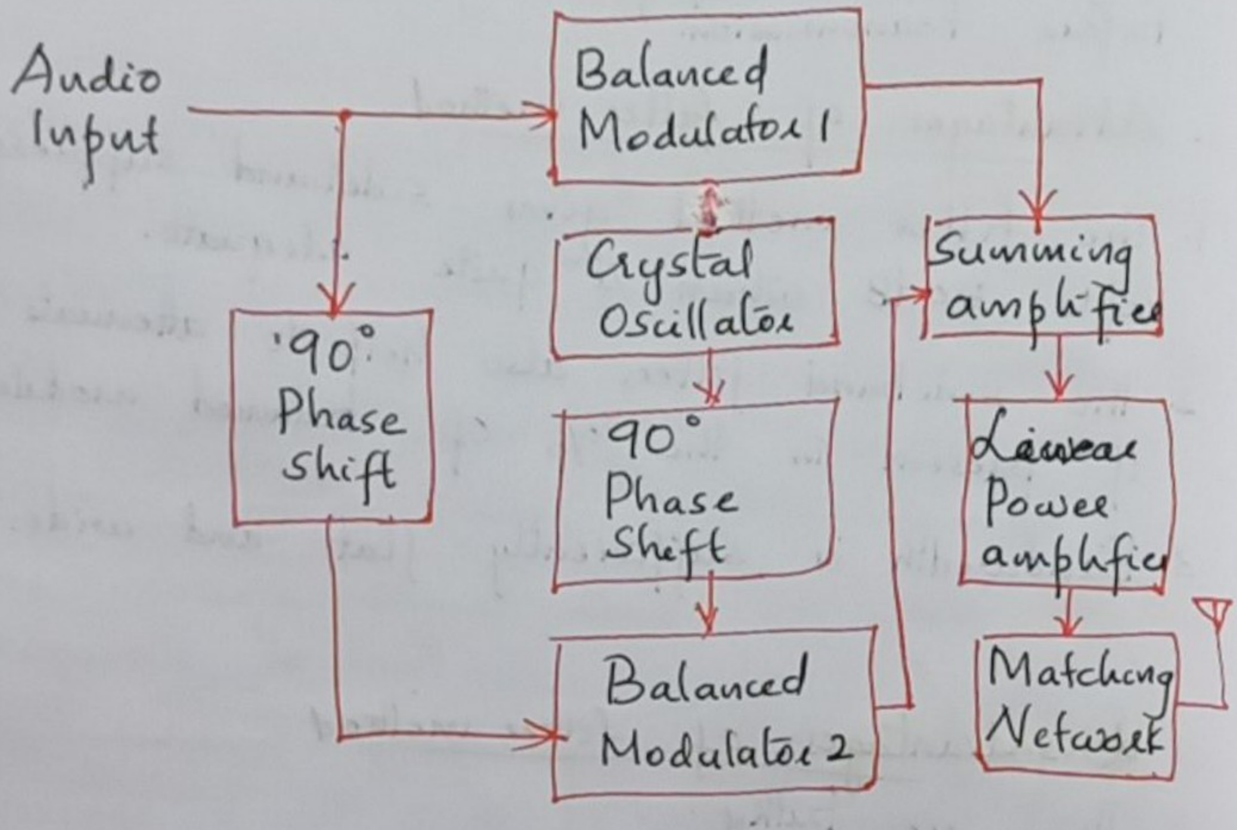
1. The filter method gives sideband suppression upto 50dB which is quite adequate.
2. The sideband filters also help to attenuate carrier if present in the o/p of balanced modulator.
3. Bandwidth is sufficiently flat and wide.

Disadvantages of filter method

1. They are bulky.
2. As modulation takes place at lower carrier frequency repeated mixing is required in conjuction with extremely stable oscillator to generate SSB at high radio frequencies.
3. At lower audio frequencies expensive filters are required.

2. Phase Shift Method (Phasing Method)

The phase shift method of SSB generation uses a phase shift technique that causes one of the sidebands to be cancelled out. The circuit doesnot have any sideband filters, and the primary modulation can be done at the transmitting frequency.



It relies on phase shifting and cancellation to eliminate the carrier and the unwanted sideband. This system uses two balanced modulators M_1 and M_2 and two 90° phase shifting ~~net~~ networks.

$$e_{LSF} = E_{L\max} \cos(\omega_c - \omega_m)t \longrightarrow (47)$$

The standard trignometric identity for the difference of two angles gives

$$e_{LSF} = E_{L\max}\left[\cos\omega_c t \cos\omega_m t + \sin\omega_c t \sin\omega_m t\right] \rightarrow (48)$$

but

$$\sin\omega_c t = \cos\left[\omega_c t - \pi/2\right] \longrightarrow (49)$$

$$\sin\omega_m t = \cos\left(\omega_m t - \pi/2\right] \longrightarrow (50)$$

Therefore,

$$e_{LSF} = E_{L\max}\left[\cos\omega_c t \cos\omega_m t + \cos(\omega_c t - \pi/2) \cos(\omega_m t - \pi/2)\right] \rightarrow (51)$$

The first term of the above eq^n (51) is the result of balanced modulator1, which multiplies the two unshifted signals. The Second term is the result of balanced modulator 2, which multiples the two signals each shifted by (-90°). The carrier signal is cancelled out in this circuit by both of the balanced modulators and the unwanted Side bands cancel at the output of the Summing amplifier.

If the two outputs are Subtracted instead of adder the upper sideband will result.

$$e_{USF} = E_{u\max} \cos(\omega_c + \omega_m)t$$

$$= E_{u\max}\left[\cos\omega_c t \cos\omega_m t - \sin\omega_c t \sin\omega_m t\right) \rightarrow (5$$

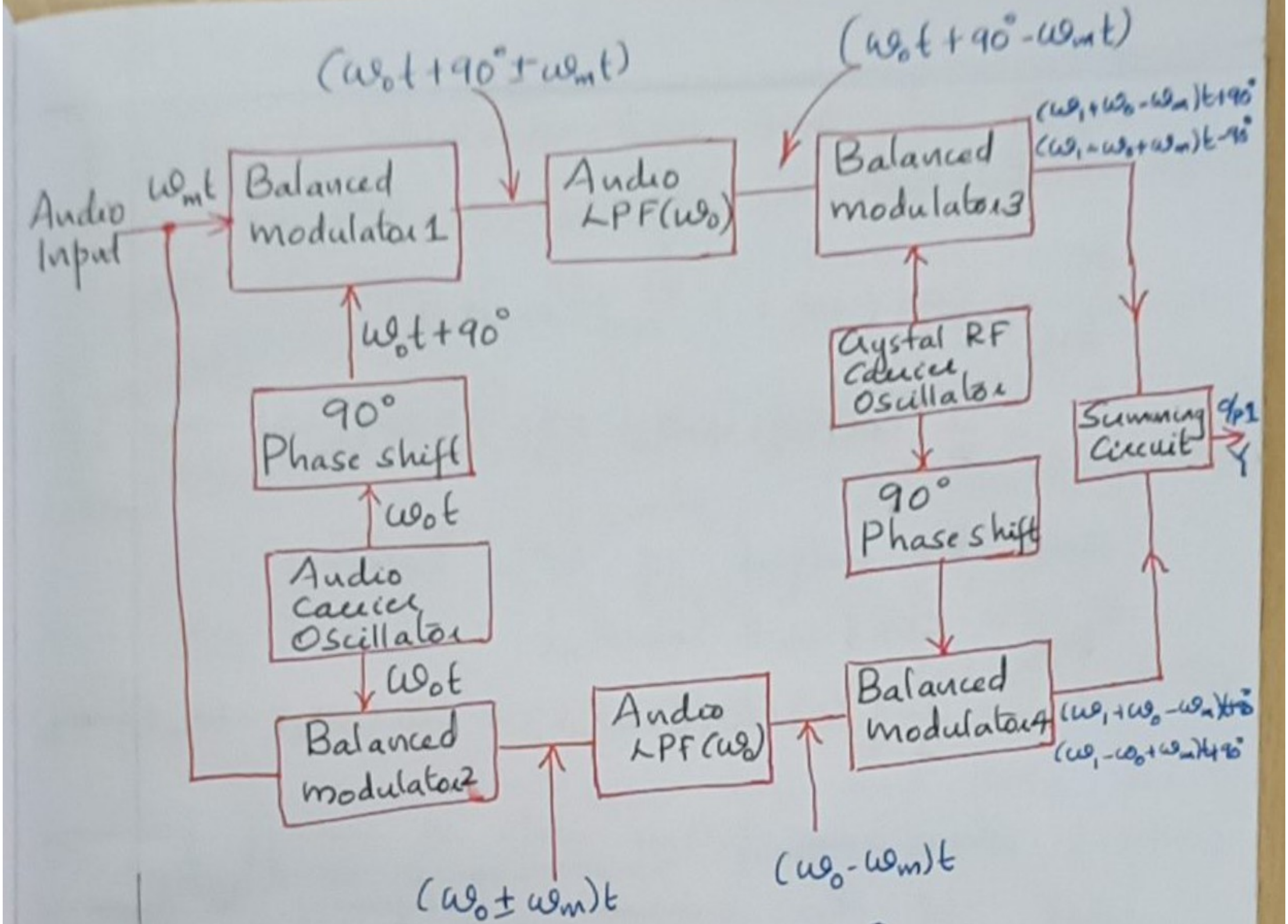
- Advantages of phase shift method

1. Bulky filters are replaced by small filters.
2. Low audio frequencies are used for modulation.
3. It can generate SSB at any frequency.
4. Easy switching from one side band to other side band is possible.

- Disadvantages

1. The design of the 90° phase shifting network for the modulating signal is extremely critical.
2. It requires complex AF phase shift network.
3. The output of two balanced modulators must be exactly same otherwise cancellation will be incomplete.

3. Third method (Weaver's method)



$Y\ (o/p\ 1) = (\omega_1 + \omega_0 - \omega_m)t + 90^\circ$

The third method of generating SSB-SC modulation is attributed to D.K Weaver and was developed during the 1950's. It is similar to the phase shifting method presenting previously, but it differs in that the modulating signal is first modulated on a low-frequency subcarrier which is then modulated onto frequency carrier.

Modulators BM_1 and BM_2 both have the unshifted modulating signal as inputs. BM_1 also takes the low-frequency subcarrier with a 90° shift introduced in it from the oscillator signal. BM_2 takes the subcarrier signal directly from the oscillator. Assuming unity magnitudes and sinusoidal single-freq. modulation, the o/p from BM_1 becomes,

$$e_{BM_1} = \cos(\omega_0 t + \pi/2)\cos\omega_m t$$

$$= \frac{1}{2}\left[\cos(\omega_0 t + \omega_m t + \frac{\pi}{2}) + \cos(\omega_0 t - \omega_m t + \frac{\pi}{2})\right] \rightarrow (53)$$

and the output of BM_2 becomes

$$e_{BM_2} = \cos\omega_0 t \cos\omega_m t$$

$$= \frac{1}{2}\left[\cos(\omega_0 t + \omega_m t) + \cos(\omega_0 t - \omega_m t)\right] \rightarrow (54)$$

Low pass filters with a cutoff frequency set at the subcarrier frequency to removes the sum (the first) term from each of the above signals, leaving only the second (difference) terms as inputs to BM_3 and BM_4. The signal applied to BM_3 is shifted by +90° from that applied to BM_4. This process eliminates the need to provide a wide band 90° phase shifting network for the base band signals.

The o/p of BM_3

$$e_{BM_3} = \cos\omega_1 t \cos\left((\omega_0 - \omega_m)t + \frac{\pi}{2}\right) \rightarrow (55)$$

$$e_{BM_3} = \frac{1}{2}\left[\cos(\omega_1 t + ((\omega_0 - \omega_m)t + \pi/2)) + \cos(\omega_1 t - (\omega_0 - \omega_m)t + \pi/2)\}\right] \rightarrow (56)$$

$$e_{BM_3} = \frac{1}{2}\Big[\cos\big((\omega_1+\omega_0)t - \omega_m t + \tfrac{\pi}{2}\big) + \cos\big((\omega_1-\omega_0)t + \omega_m t - \tfrac{\pi}{2}\big)\Big] \longrightarrow (57)$$

and the output of BM4 becomes

$$e_{BM_4} = \cos\big(\omega_1 t + \tfrac{\pi}{2}\big)\cos(\omega_0 - \omega_m)t \longrightarrow (58)$$

$$e_{BM_4} = \frac{1}{2}\Big[\cos\big((\omega_1 t + \pi/2) + (\omega_0 - \omega_m)t\big) + \cos\big((\omega_1 t + \pi/2) - (\omega_0 - \omega_m)t\big)\Big] \longrightarrow (59)$$

$$e_{BM_4} = \frac{1}{2}\Big[\cos\big((\omega_1+\omega_0)t - \omega_m t + \tfrac{\pi}{2}\big) + \cos\big((\omega_1-\omega_0)t + \omega_m t + \tfrac{\pi}{2}\big)\Big] \longrightarrow (60)$$

The outputs from BM_3 and BM_4 are added in a summing amplifier to produce the final output.

$$e_{out} = \cos\big((\omega_1 + \omega_0 - \omega_m)t + \tfrac{\pi}{2}\big) \longrightarrow (61)$$

This is the lower side band on the carrier frequency (f_1+f_0).

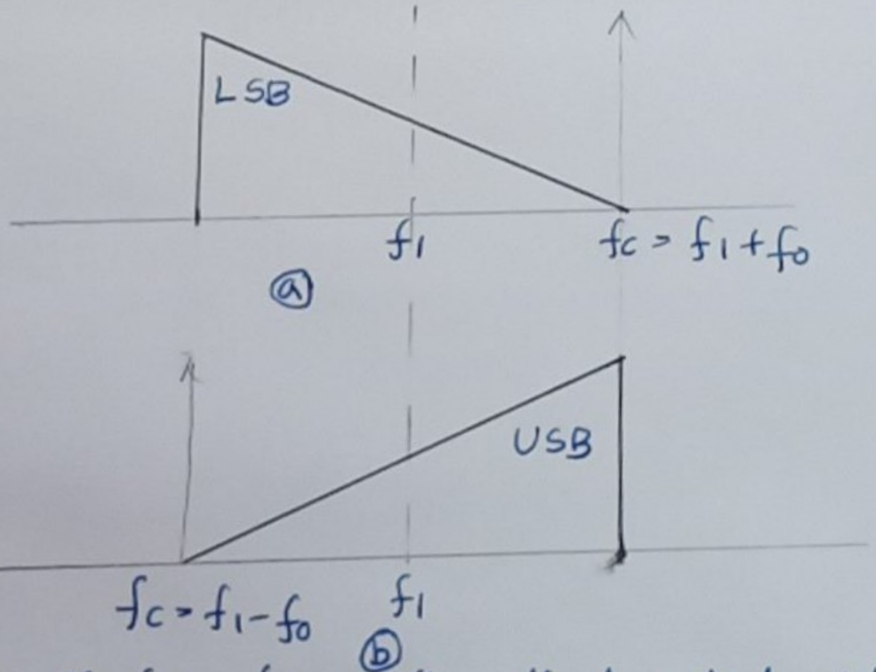
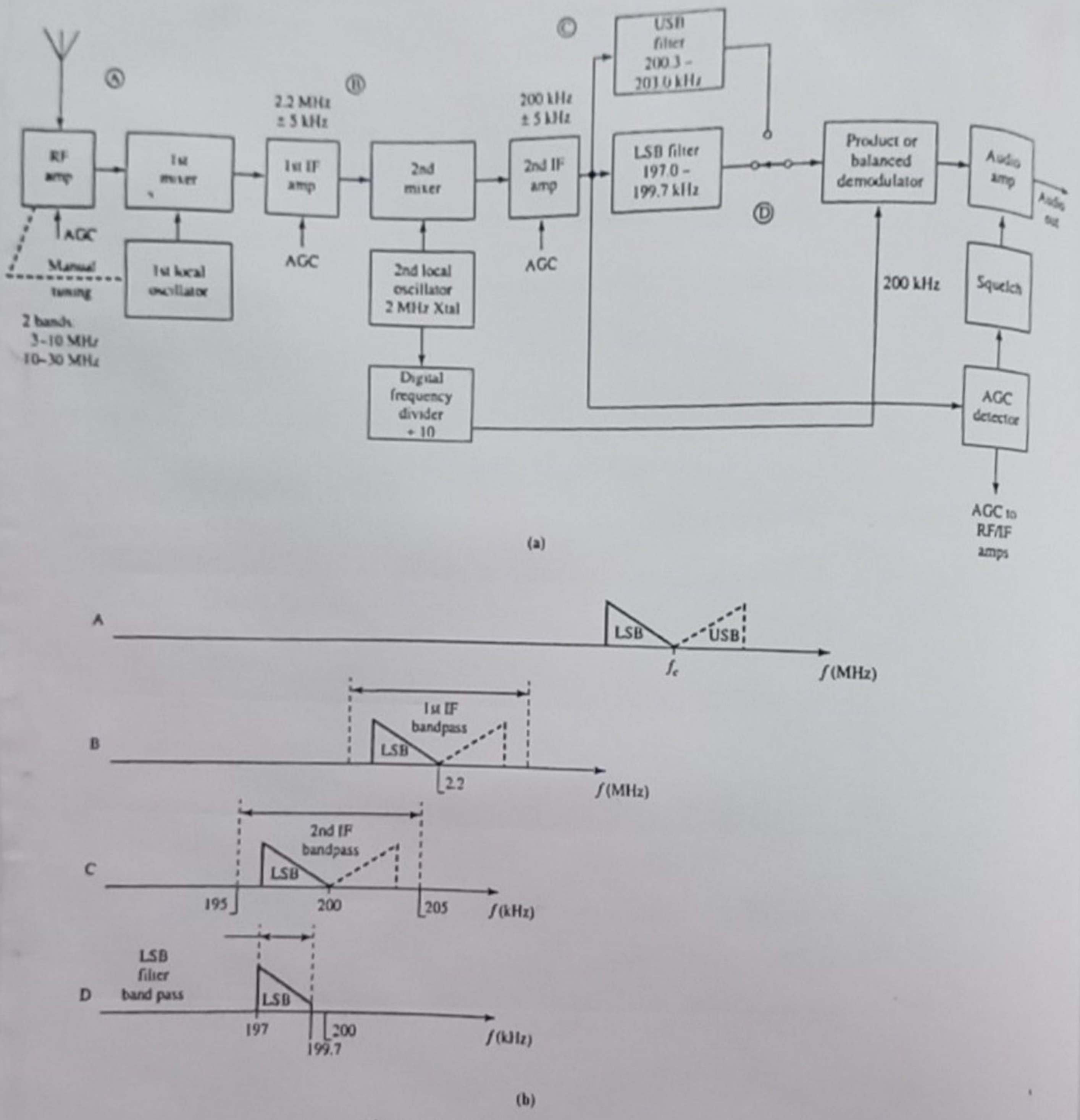


Fig o/p spectra for the third method ckt
ⓐ for LSB ⓑ for USB

→ SSB Reception

The received SSB signal is multiplied with a synchronous carrier and the result contains the original modulation signal as



(a)

(b)

Figure 9.5.1 (a) Single-sideband HF receiver. (b) Spectra in the HF receiver for an LSB signal: Ⓐ, received RF signal; Ⓑ, output of first IF amp; Ⓒ, output of second IF amp; Ⓓ, output of LSB filter.

one component. Balanced or product modulators are used for demodulation. The carrier signal for the demodulator must be locally generated if the signals are true SSB signals with the carrier completely suppressed. This requires extreme stability for the local oscillator signals for demodulating and for the superheterodyne converters.

Double conversion is often used in SSB receivers. Since very good adjacent channel selectivity must be provided since SSB signals are usually packed closely together in the frequency spectrum.

The first local oscillator and RF amplifier are manually tuned in two switched bands. The output of second crystal oscillator is divided by 10 in a digital counter to provide the carrier signal for the demodulator.

The output from the detector is passed through a gated audio amplifier that turns off the output to keep the noise down when the signal level drops below a preset threshold. This is called squelch. The amplified IF signal is rectified to provide the AGC voltage for the RF and IF amplifiers and for the squelch circuit.

The 2nd IF amplifier is followed by two filters, USB and LSB filters. USB filter passes IF upperside band and rejects the lower side band. The LSB filter passes IF lower side band. The approximate sideband is Selected by a switch that connects the o/p of the desired filter to the product detector.

One of the largest application of this type of SSB is multichannel citizen's band (CB) transceivers.

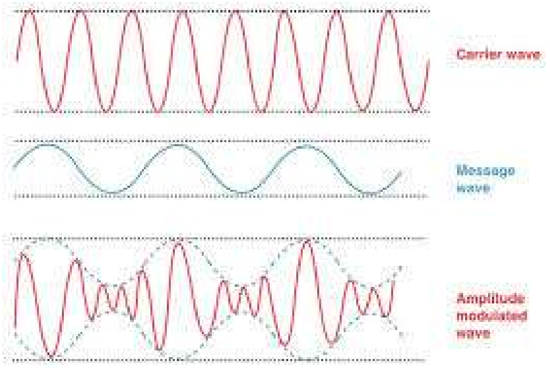
- Comparision between Phase shift method and Filter method

	Parameter	Filter Method	Phase Shift Method
1.	Method of Cancelling unwanted side band.	Using filter	Using balance modulator and 90° phase Shifter.
2.	Use of low modulating frequency	Not possible	Possible
3.	Need of up-conversion	Needed	Not needed
4.	Need of linear amplifier	Needed	Needed
5.	Possibility of SSB generation at any frequency	Not possible	Possible

AM, PM and FM modulation

What is amplitude modulation

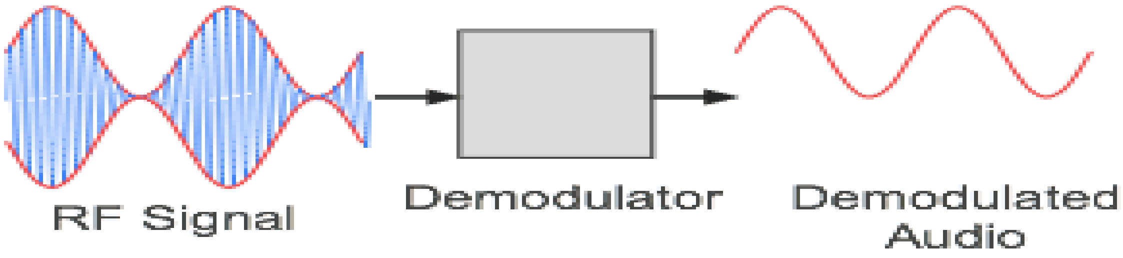
- In order that a radio signal can carry audio or other information for broadcasting or for two way radio communication, it must be modulated or changed in some way. Although there are a number of ways in which a radio signal may be modulated, one of the easiest is to change its amplitude in line with variations of the sound.



Amplitude modulation

Amplitude demodulation

- Amplitude modulation, AM, is one of the most straightforward ways of modulating a radio signal or carrier. It can be achieved in a number of ways, but the simplest uses a single diode rectifier circuit.
- Other methods of demodulating an AM signal use synchronous techniques and provide much lower levels of distortion and improved reception where selective fading is present.
- One of the main reasons for the popularity of amplitude modulation has been the simplicity of the demodulation. It enables costs to be kept low - a significant advantage in producing vast quantities of very low cost AM radios.



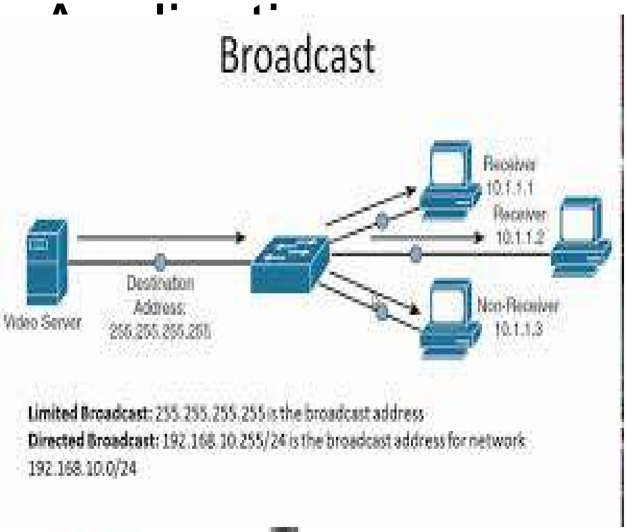
Advantages & disadvantages of AM

Advantages

- It is simple to implement
- it can be demodulated using a circuit consisting of very few components
- AM receivers are very cheap as no specialised components are needed.

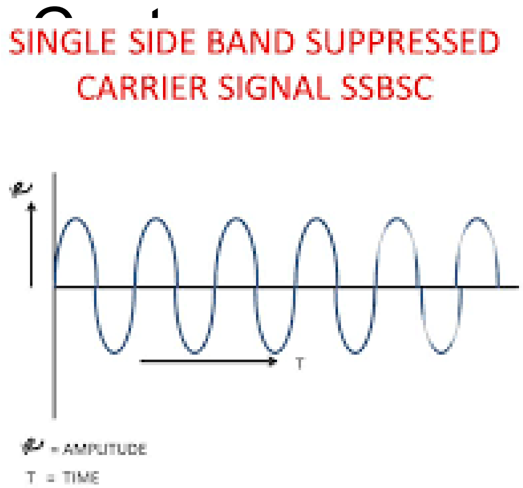
Disadvantages

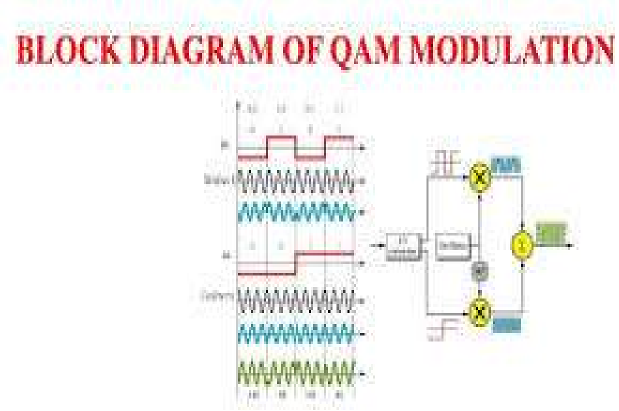
- It is not efficient in terms of its power usage
- It is not efficient in terms of its use of bandwidth, requiring a bandwidth equal to twice that of the highest audio frequency
- It is prone to high levels of noise because most noise is amplitude based and obviously AM detectors are sensitive to it.



- ***Broadcast transmissions:*** AM is still widely used for broadcasting on the long, medium and short wave bands. It is simple to demodulate and this means that radio receivers capable of demodulating amplitude modulation are cheap and simple to manufacture. Nevertheless many people are moving to high quality forms of transmission like frequency modulation, FM or digital transmissions.

- ***Air band radio:*** VHF transmissions for many airborne applications still use AM. . It is used for ground to air radio communications as well as two way radio links for ground staff as well.



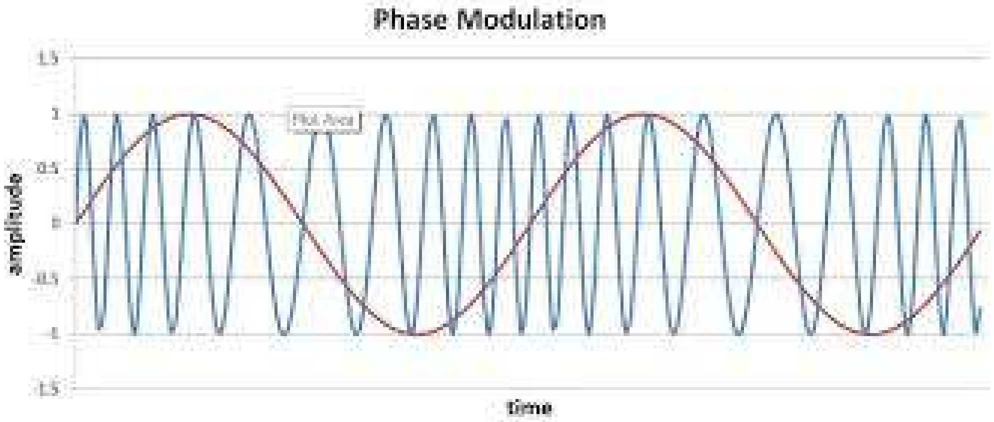


- ***Single sideband:*** Amplitude modulation in the form of single sideband is still used for HF radio links. Using a lower bandwidth and providing more effective use of the transmitted power this form of modulation is still used for many point to point HF links.

- ***Quadrature amplitude modulation:*** AM is widely used for the transmission of data in everything from short range wireless links such as Wi-Fi to cellular telecommunications and much more. Effectively it is formed by having two carriers 90° out of phase.

What is Phase Modulation

- Phase modulation, PM is sometimes used for analogue transmission, but it has become the basis for modulation schemes used for carrying data. Phase shoft keying, PSK is widely used for data communication.Phase modulation is also the basis of a form of modulation known as quadrature amplitude modulation, where both phase and amplitude are varied to provide additional capabilities.



Phase modulation basics

- A radio frequency signal consists of an oscillating carrier in the form of a sine wave is the basis of the signal. The instantaneous amplitude follows this curve moving positive and then negative, returning to the start point after one complete cycle - it follows the curve of the sine wave.

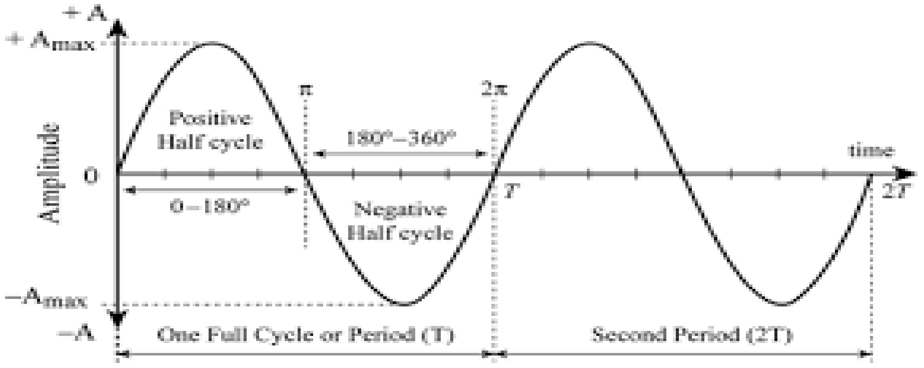
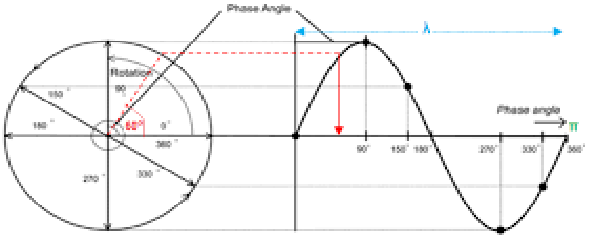


Figure 1

Cont..

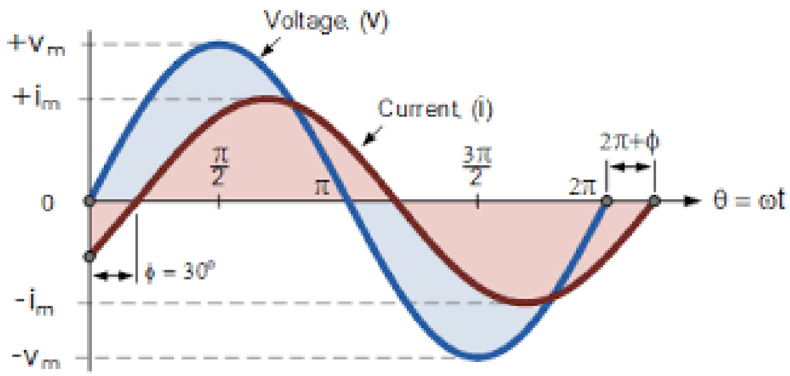
- The sine wave can also be represented by the movement of a point around a circle, the phase at any given point being the angle between the start point and the point on the waveform as shown.



- Phase angle of poin hase advances as time progresses can be said to have a phase differe

Cont..

- Phase modulation works by modulating the phase of the signal, i.e. changing the rate at which the point moves around the circle. This changes the phase of the signal from what it would have been if no modulation was applied. In other words the speed of rotation around the circle is modulated about the mean value.

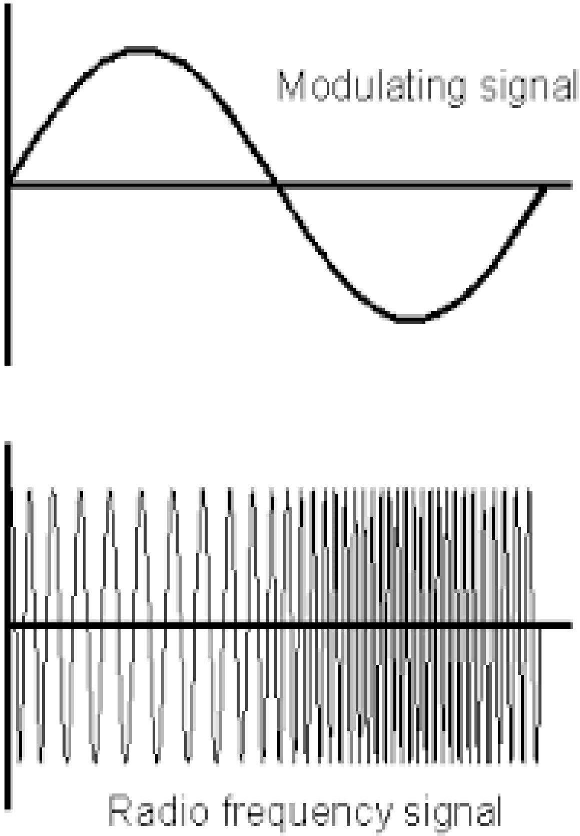


Forms of phase modulation

Although phase modulation is used for some analogue transmissions, it is far more widely used as a digital form of modulation where it switches between different phases. This is known as phase shift keying, PSK, and there are many flavours of this. It is even possible to combine phase shift keying and amplitude keying in a form of modulation known as quadrature amplitude modulation, QAM.

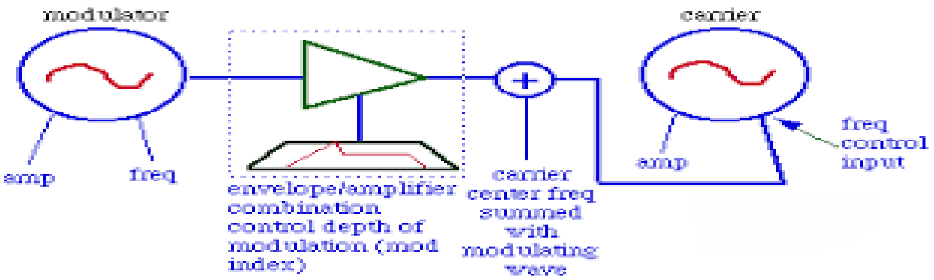
What is frequency modulation, FM

- As with any form of modulation, it is necessary to be able to successfully demodulate it and recover the original signal. The FM demodulator may be called a variety of names including FM demodulator, FM detector or an FM discriminator.
- There are a number of different types of FM demodulator, but all of them enable the frequency variations of the incoming signal to be converted into amplitude variations on the output. These are typically fed into an audio amplifier, or possibly a digital interface if data is being passed over the system.



FM modulators

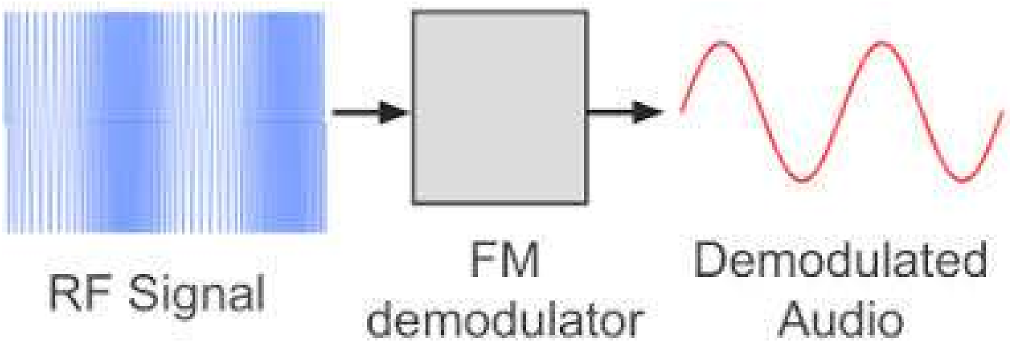
- ***Varactor diode oscillator:*** This method simply requires the use of a varactor diode placed within the tuned circuit of an oscillator circuit. It is even possible to use a varactor diode within a crystal oscillator circuit. Typically when crystal oscillators a re used the signal needs to be multiplied in frequency, and only narrow band FM is attainable.



- ***Phase locked*** cellent method of generating fre y to manage the constraints w it provides and excellent solution.

Frequency demodulation

- it is necessary to be able to successfully demodulate it and recover the original signal. The FM demodulator may be called a variety of names including FM demodulator, FM detector or an FM discriminator.
- There are a number of different types of FM demodulator, but all of them enable the frequency variations of the incoming signal to be converted into amplitude variations on the output. These are typically fed into an audio amplifier, or possibly a digital interface if data is being passed over the system.



CONT..

The list below gives some of the forms of phase shift keying that are used:

- PM - Phase Modulation
- PSK - Phase Shift Keying
- BPSK - Binary Phase Shift Keying
- QPSK - Quadrature Phase Shift Keying
- 8 PSK - 8 Point Phase Shift Keying
- 16 PSK - 16 Point Phase Shift Keying
- OPSK - Offset Phase Shift Keying

Advantages of frequency modulation, FM:

- ***Resilience to noise:*** One particular advantage of frequency modulation is its resilience to signal level variations. The modulation is carried only as variations in frequency.
- ***Easy to apply modulation at a low power stage of the transmitter:*** Another advantage of frequency modulation is associated with the transmitters.
- ***It is possible to use efficient RF amplifiers with frequency modulated signals:*** It is possible to use non-linear RF amplifiers to amplify FM signals in a transmitter and these are more efficient than the linear ones required for signals with any amplitude variations (e.g. AM and SSB).

disadvantages of frequency modulation, FM

- ***FM has poorer spectral efficiency than some other modulation formats:*** Some phase modulation and quadrature amplitude modulation formats have a higher spectral efficiency for data transmission than frequency shift keying, a form of frequency modulation.
- ***Requires more complicated demodulator:*** One of the minor disadvantages of frequency modulation is that the demodulator is a little more complicated, and hence slightly more expensive than the very simple diode detectors used for AM.

CONT..

- ***Some other modes have higher data spectral efficiency:*** Some phase modulation and quadrature amplitude modulation formats have a higher spectral efficiency for data transmission that frequency shift keying, a form of frequency modulation.
- ***Sidebands extend to infinity either side:*** The sidebands for an FM transmission theoretically extend out to infinity. They are normally significant for wideband frequency modulation transmissions, although small for narrow band FM.

Generation of FM Signals

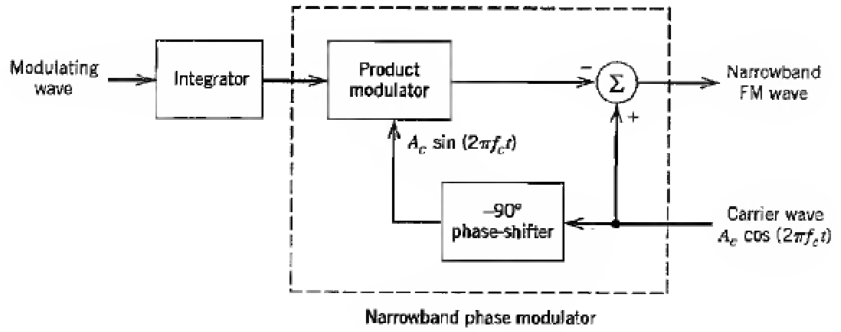
There are essentially two basic methods of generating frequency-modulated signals, namely, direct FM and indirect FM.

In the direct method the carrier frequency is directly varied in accordance with the input baseband signal, which is readily accomplished using a voltage-controlled oscillator.

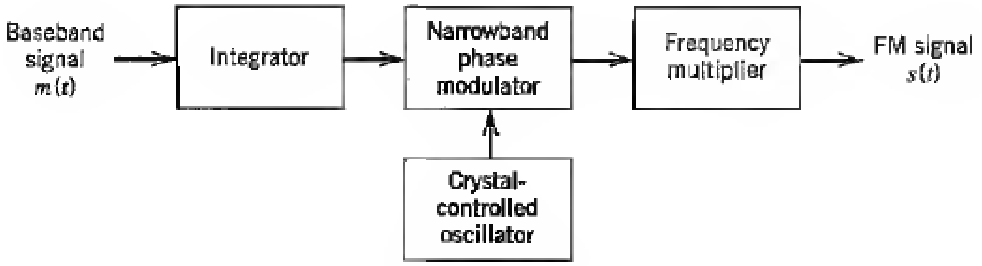
In the indirect method, the modulating signal is first used to produce a narrowband FM signal, and frequency multiplication is next used to increase the frequency deviation to the desired level.

The indirect method is the preferred choice for frequency modulation when the stability of carrier frequency is of major concern as in commercial radio broadcasting.

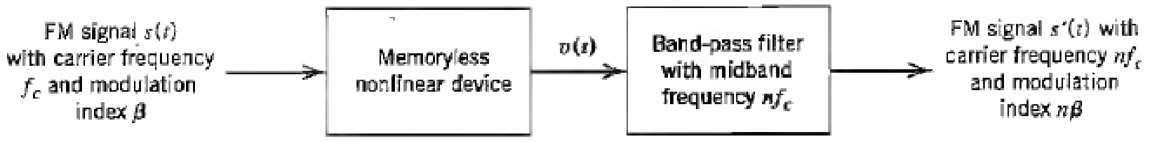
Indirect FM



Block diagram of a method for generating a narrowband FM signal.

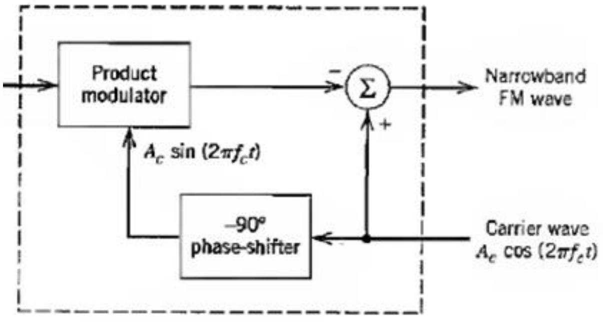


Block diagram of the indirect method of generating a wideband FM signal.



Block diagram of frequency multiplier

The message (baseband) signal m(t) is first integrated and then used to phase-modulate a crystal-controlled oscillator; the use of crystal control provides frequency stability. To minimize the distortion inherent in the phase modulator, the maximum phase deviation or modulation index is kept small, thereby resulting in a narrowband FM signal.



Narrowband phase modulator

The narrowband FM signal is next multiplied in frequency by means of a frequency multiplier so as to produce the desired wideband FM signal.

A frequency multiplier consists of a nonlinear device followed by a band-pass filter, as shown. The implication of the nonlinear device being memoryless is that it has no energy-storage elements. The mid-band frequency of the band-pass filter is set equal to nfc where fc is the carrier frequency of the incoming FM signal s(r). Moreover, the band-pass filter is designed to have a bandwidth equal to n times the transmission bandwidth of s(t).

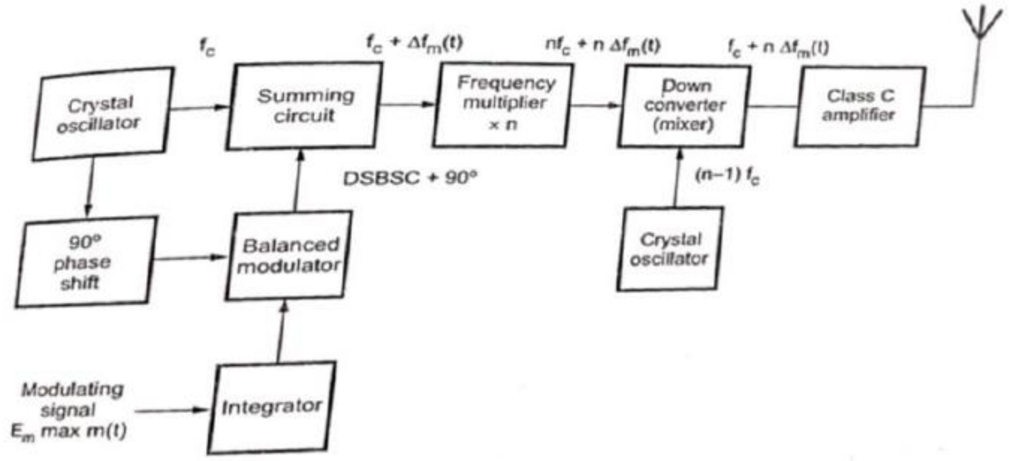


Figure: FM transmitter in which FM is achieved through Armstrong Method

In this method, the initial modulation takes place as an amplitude modulated DSBSC signal so that a crystal controlled oscillator can be used if desired. The crystal oscillator generates the subcarrier, which can be low, say on the order of 100 KHz. One output from the oscillator is phase shifted by 90 degrees to produce the sine term, which is then DSBSC modulated in the balanced modulator by Vm(t). This is combined with the direct output from the oscillator in the summing amplifier, the result then being the phase modulated signal. The modulating signal is passed through an integrator to the modulated to get the frequency modulated signal At this stage, the equivalent frequency deviation will be low

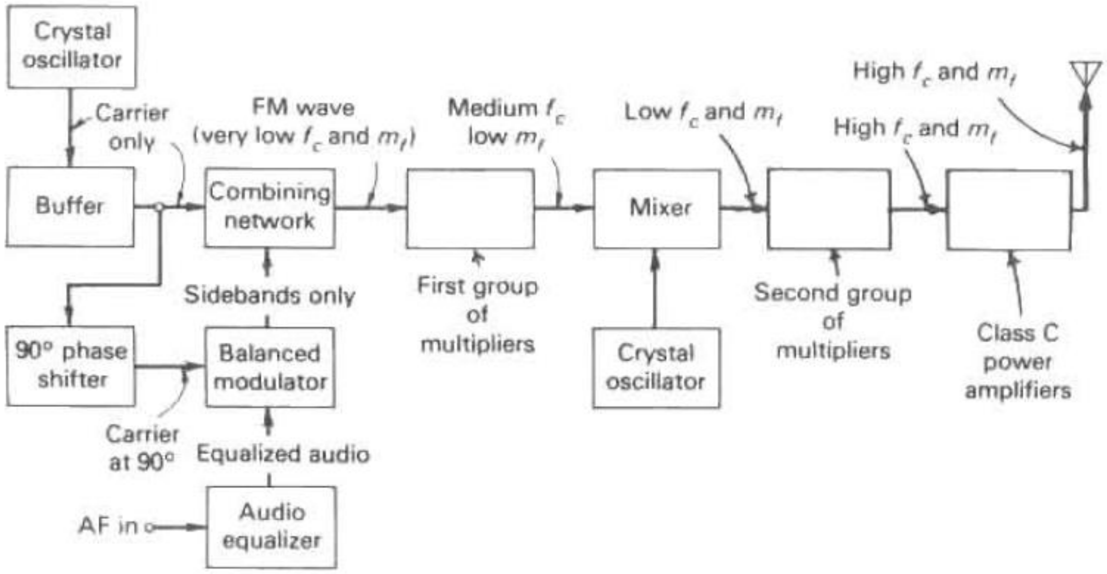
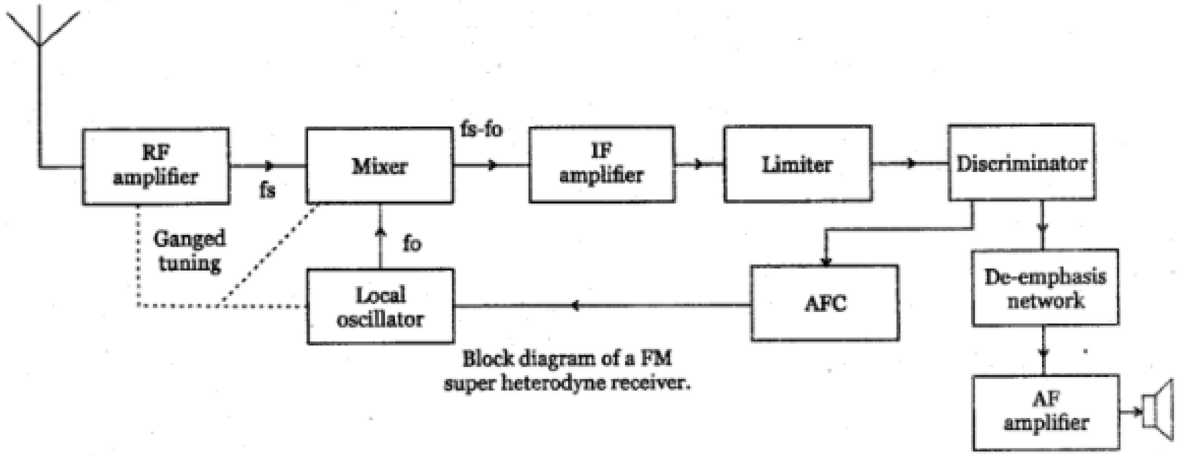


FIGURE 5-15 Block diagram of the Armstrong frequency-modulation system.

FM Receiver: Superhetrodyne structure



Block diagram of a FM super heterodyne receiver.

RF amplifier: The RF amplifier increases the signal strength before the signal is fed to mixer when turned to the desired frequency. The RF amplifier is designed to handle large bandwidth of 150 kHz.

Mixer: the incoming RF signal of frequency f_m is applied to a mixer which also receives the output from the local oscillator. A new frequency called intermediate frequency IF is produced whose value is difference of local oscillator signal f and signal frequency f.

Local oscillator: the receiver converts incoming carrier frequency to the IF by using local oscillator frequency higher than incoming tuned frequency. Colpitts oscillator is used as the local oscillator.

IF amplifier: IF signal is amplified by one or more number of amplifiers, which raises the strength of IF signal. It has multistage class A amplifier providing better selectivity and gain.

Limiter: It removes all the amplitude variation in FM signal caused by noise. Differential amplifiers are preferred for limiter.

Discriminator: It recovers the modulating signal from the IF signal. It converts frequency variation into corresponding voltage variation and produces the modulating signal.

De-emphasis network: It reduces the relative amplitude of high frequency signals that are boosted in the transmitter and brings them back to their original level.

AF amplifier: It amplifier the modulating signal recovered by the FM detector. The speaker converts the electrical signal into sound signal.

Review of Random variables and Random process:

Module-II

ECT305 Analog and Digital Communication

Review of Probability Theory

- Communication systems deal with quantities that are not deterministic and hence there is a need to apply non-deterministic (random) and probabilistic approaches.
- You should review the basics of Probability that you have already learnt, be aware that probability of an event always lies between 0 and 1.
- Also you should be aware of independent events, statistically uncorrelated quantities.
- Awareness of Conditional Probability and Bayes theorem.
- Awareness of Basic rules and formulations for the Probability of events.

Random Variables

- When a random experiment is performed, there are various outcomes possible.
- It is convenient to consider the expt and its possible outcomes as defining a space and its points.
- With the kth outcome of the expt, there is a point associated called Sample Point, denoted by s_k.
- The totality of sample points corresponding to the aggregate of all possible outcomes of the expt is called Sample Space (S).
- An Event corresponds to either a single sample point or a set of sample points.
- The outcome of an expt can be a variable that can wander over a set of sample points and whose value is determined by the expt.

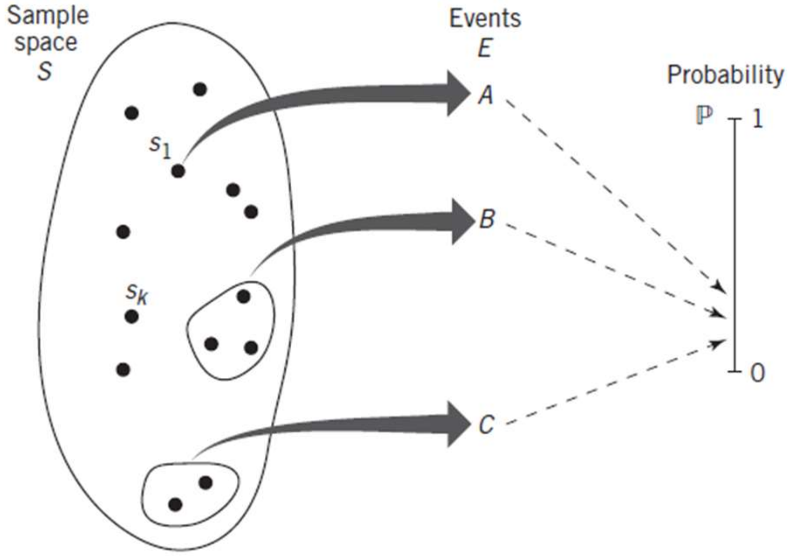


Illustration of the relationship between sample space, events, and probability

Random Variable

- A function whose domain is a sample space and whose range is some set of real numbers is called a random variable of the expt.
- When the outcome of an experiment is 's', the random variable is denoted by X(s) or X.
- The random variable (rv) can be:
  - a discrete random variable (or)
  - a continuous random variable.

Discrete Random Variable

- Consider, for example, the sample space that represents the integers 1, 2, …, 6, each one of which is the number of dots that shows uppermost when a die is thrown.
- Let the sample point s_k denote the event that *k* dots show in one throw of the die.
- The random variable used to describe the probabilistic event s_k in this experiment is said to be a Discrete Random Variable.

Continuous Random Variable

- Consider the electrical noise being observed at the front end of a communication receiver.
- The random variable, representing the amplitude of the noise voltage at a particular instant of time, occupies a continuous range of values, both positive and negative.
- The random variable representing the noise amplitude is said to be a Continuous Random Variable.

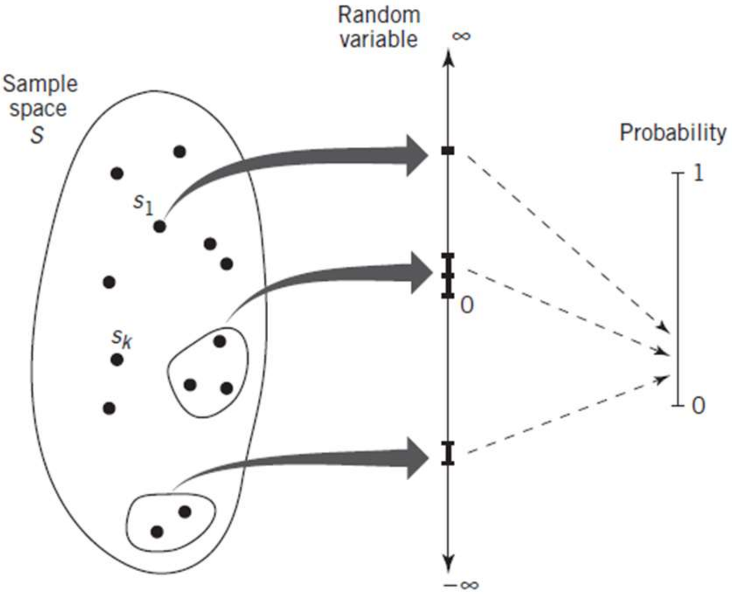


Illustration of the relationship between sample space, random variables, and probability.

- Consider the random variable X and the probability of the event $X \leq x$.
- Denote this probability by P[$X \leq x$].

  F_X(x)= P[$X \leq x$] for all x ………(1)
- The function $F_X(x)$ is called the <u>Cumulative Distribution Function</u> or the <u>Distribution Function</u> of the random variable X.
- Note that $F_X(x)$ is a function of x, not of the random variable X.
- For any point x in the sample space, the distribution function $F_X(x)$ expresses the probability of an event.

- The distribution function $F_X(x)$, applicable to both continuous and discrete random variables, has two fundamental properties:
- 1. $F_X(x)$ is a bounded function that lies between zero and one. $0 \leq F_X(x) \leq 1$
- 2. It is a monotone non-decreasing function of x

$$F_X(x_1) \leq F_X(x_2) \text{ if } x_1 \leq x_2$$

- The random variable X is said to be *continuous* if the distribution function $F_X(x)$ is differentiable with respect to x everywhere, as shown by,

$$f_X(x) = \frac{d}{dx}F_X(x) \qquad \text{for all } x \qquad \text{.........(2)}$$

- The new function $f_X(x)$ is called the Probability Density Function (pdf) of the random variable X.
- The probability of the event $x_1 < X \leq x_2$ is

$$P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$$
$$= F_X(x_2) - F_X(x_1)$$
$$= \int_{x_1}^{x_2} f_X(x).dx \qquad \text{.........(3)}$$

$\therefore$ $F_X(x) = \int_{-\infty}^{x_2} f_X(x).dx$

Also, $F_X(-\infty) = 0$, $F_X(\infty) = 1 \Longrightarrow \int_{-\infty}^{\infty} f_X(x).dx = 1$

The probability density function pdf must always be a non-negative function with a total area of unity.

Multiple Random Variables

- Consider two random variables X and Y.
- The joint distribution function $F_{X,Y}(x,y)$ is the probability that the random variable X is less than or equal to a specified value x, and that the random variable Y is less than or equal to another specified value y.
- $F_{X,Y}(x)= P[X \leq x, Y \leq y]$ for all x and y …(4)

- The Joint Probability Density Function of the random variables X and Y is given by,

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \quad \text{……….(5)}$$

- The joint distribution function $F_{X,Y}(x,y)$ is a monotone non-decreasing function of both x and y.
- The joint probability density function $f_{X,Y}(x,y)$ is always non-negative.
- Also, the total volume under the graph of a joint probability density function must be unity.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1 \quad \text{……….(6)}$$

- The *marginal* probability density functions are $f_X(x)$ and $f_Y(y)$.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy \quad \text{……….(7)}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \quad \text{……….(8)}$$

<u>Conditional Probability Density Function</u>

- Suppose that X and Y are two continuous random variables with their joint probability density function $f_{X,Y}(x, y)$.
- The *conditional probability density function* of Y, such that $X = x$, is defined by,

$$f_Y(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{……….(9)}$$

- Provided that $f_X(x) > 0$, where $f_X(x)$ is the marginal density of X;

$$f_Y(y|x) \geq 0 \qquad \int_{-\infty}^{\infty} f_Y(y|x)\,dy = 1$$

- From eqn (9) obtain multiplication rule.

$$f_{X,Y}(x, y) = f_Y(y|x)f_X(x)$$

- Suppose that knowledge of the outcome of X can, in no way, affect the distribution of Y.
- Then, the conditional probability density function $f_Y(y|x)$ reduces to the marginal density $f_Y(y)$,

$$f_Y(y|x) = f_Y(y)$$

- In such a case, express the joint probability density function of the random variables X and Y as the product of their respective marginal densities; i.e.,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Golden Rules:

1. If the joint probability density function of the random variables X and Y equals the product of their marginal densities, then X and Y are Statistically Independent.

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

2. The summation of two independent continuous random variables leads to the Convolution of their respective probability density functions.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,\mathrm{d}x$$

Statistical Average

- The *expected value* or *mean* of a continuous random variable *X* is defined by,
  μ_X=E[X]=$\int_{-\infty}^{\infty} \mathrm{x}.f_X(\mathrm{x}).\mathrm{dx}$, $\mathrm{f_X(x)}$→pdf of *x* ....(10)
- The expectation of a sum of random variables is equal to the sum of individual expectations.
- The expectation of the product of two statistically independent random variables is equal to the product of their individual expectations.

- let *X* denote a random variable and let *g*(*X*) denote a real-valued function of *X* defined on the real line.
- The quantity obtained by letting the argument of the function *g*(*X*) be a random variable is also a random variable,

Let Y=g(X)

E[Y]=$\int_{-\infty}^{\infty} \mathrm{y.fY(y).dy}$, where $\mathrm{f_Y(y)}$→pdf of *y*

Another way → E[g(X)]= $\int_{-\infty}^{\infty} \mathrm{g(x)}.f_X(\mathrm{x}).\mathrm{dx}$ ……(11)

- This is called Expected value rule;

Second-order Moments

- For the special case of $g(X) = X^n$, application of eqn(11) leads to the *n*th *moment* of prob distribution of a random variable *X*,

$E[X^n]= \int_{-\infty}^{\infty} \mathrm{x}^{\mathrm{n}}. f_X(\mathrm{x}). \mathrm{dx}$ ……(12)

- Put n=1 $\Longrightarrow$ $E[X]= \int_{-\infty}^{\infty} \mathrm{x}\,. f_X(\mathrm{x}). \mathrm{dx}$ → Mean!
- Put n=2 $\Longrightarrow$ $E[X^2]= \int_{-\infty}^{\infty} \mathrm{x}^2\,. f_X(\mathrm{x}). \mathrm{dx}$ → Mean Square Value! ……(13)

Variance

- Central Moments are the moments of the difference between a random variable *X* and its mean μ_X.
- The *n*th Central Moment of *X* is,

$E[(X- \mu_X)^n]= \int_{-\infty}^{\infty} (\mathrm{x} - \mu_X)^{\mathrm{n}}. f_X(\mathrm{x}). \mathrm{dx}$……(14)

- The Second Central Moment or Variance is,

$Var[X]=E[(X- \mu_X)^2]= \int_{-\infty}^{\infty} (\mathrm{x} - \mu_X)^2. f_X(\mathrm{x}). \mathrm{dx}$ …..(15)

- The variance of a random variable *X* is commonly denoted by σ_X^2.
- The square root of the variance, namely σ_X, is called the *standard deviation* of the random variable *X*.

$$\sigma_X^2 = \mathrm{Var}[X] = E[(X - \mu_X)^2]$$
$$= E[X^2 - 2\mu_X X + \mu_X^2]$$
$$= E[X^2] - 2\mu_X E[X] + \mu_X^2$$
$$= E[X^2] - 2\mu_X^2 + \mu_X^2$$
$$= E[X^2] - \mu_X^2$$

If the Mean is zero, the Variance and Mean Square value of the rv X are equal.

Joint Moments

- The joint moment of a pair of rvs X and Y is the expectation of X^iY^k, where *i* and *k* are positive integer values.

$\therefore$ $E[X^i Y^k] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i\, y^k . f_{X,Y}(x,y).\, dx.dy$ ......(16)

- Correlation is defined as E[XY] corresponds to i=k=1
  → $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y . f_{X,Y}(x,y).\, dx.dy$ ......(17)
- Covariance of rvs X and Y is defined as the Correlation of the centered rvs X-E[X] and Y-E[Y],

$Cov[XY] = E[(X-E[X]).(Y-E[Y])] = E[(X - \mu_X).(Y - \mu_Y)]$ ......(18)

$\therefore$ $Cov[XY] = E[XY] - \mu_X\mu_Y$......(19)

- Correlation Coefficient of rvs X and Y is the Covariance of X and Y, normalized w.r.t product of their Standard deviations.

$$\therefore\ \rho = \frac{\mathrm{Cov}[XY]}{\sigma_X \sigma_Y} \qquad \text{......(20)}$$

- Two rvs X and Y are Uncorrelated iff their Covariance is zero
- Two rvs X and Y are Orthogonal iff their Correlation is zero, i.e., E[XY] = 0
- If rvs X and Y are Statistically Independent, then they are Uncorrelated, but the converse is not necessarily true.
- If one of the rvs *X* and *Y* or both have zero mean, and if they are orthogonal, then they are uncorrelated, and vice versa.

Stochastic or Random Processes

- A Stochastic Process is a set of Random Variables indexed in time.
- In performing a random expt, the Sample space is considered and each outcome of the expt is associated with a Sample point.
- The totality of Sample points corresponds to the aggregate of all possible outcomes of the expt is called the Sample space.
- Each Sample point of the Sample space is a function of time. (Main difference compared to Random variable!)
- The Sample space or Ensemble composed of functions of time is called a Random or Stochastic Process.

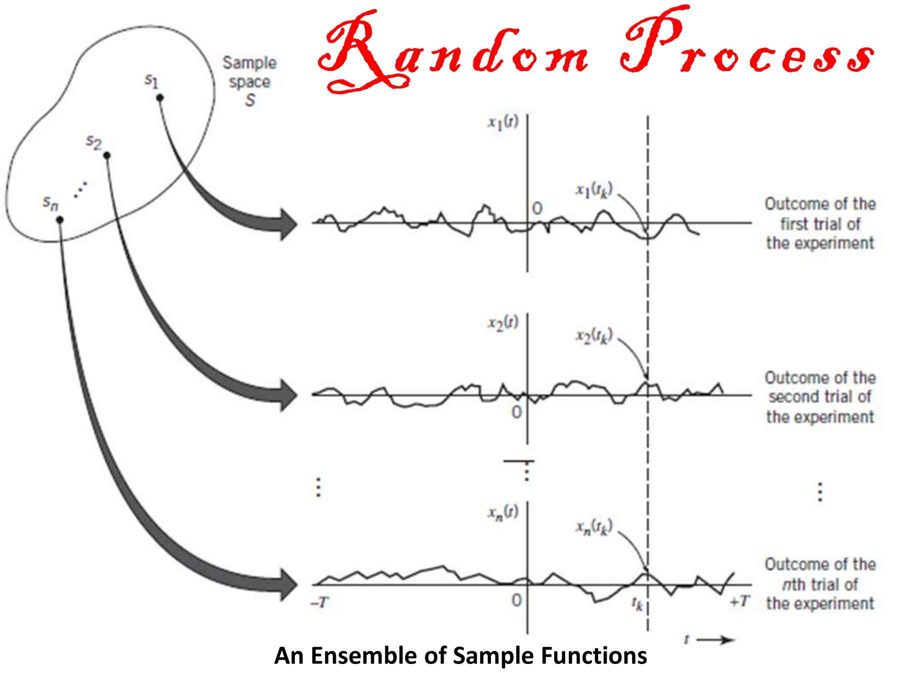
- Consider a random expt specified by the outcomes ‘s’ from sample space ‘S’.
- Assign to each Sample point ‘s’, a function of time,
  $X(t,s)$, $-T \leq t \leq T$,
  where 2T is the total observation interval.
- So, each Sample function is, $x_i(t)=X(t, s_i)$
- Consider a set of sample fns $\{x_i(t)|_{i=1,2,\ldots,n}\}$
- For a fixed time t_k, within the observation interval, the set of numbers,

$\{x_1(t_k), x_2(t_k),\ldots., x_n(t_k)\}=\{X(t_k, s_1), X(t_k, s_2),\ldots, X(t_k, s_n)\}$

→constitute a Random variable (rv).

- The Indexed Ensemble of rv, $\{X(t, s)\}$ is called Random Process.
- The Random Process is simplified as $X(t)$

Random Process



An Ensemble of Sample Functions

- The Random Process X(t) is an Ensemble of time functions together with a Probability rule that assigns a probability to any event associated with an observation of one of the sample functions of the stochastic process.
- For a Random Variable, the outcome of a random expt is mapped into a number, where as for a Random Process, the outcome of the random expt is mapped into a waveform that is a function of time.

- In dealing with random processes in the real world, it is found that the statistical characterization of a process is independent of the time at which observation of the process is initiated.
- That is, if such a process is divided into a number of time intervals, the various sections of the process exhibit essentially the same statistical properties.
- Such a stochastic process is said to be Stationary.
- Otherwise, it is said to be Non-Stationary.

- A stationary process arises from a stable phenomenon that has evolved into a steady-state mode of behavior, whereas a non-stationary process arises from an unstable phenomenon.
- Consider a random process $X(t)$ initiated at $t=-\infty$
- Let $X(t_1), X(t_2), \ldots, X(t_k)$ denote the random variables obtained by sampling the process $X(t)$ at times $t_1, t_2, \ldots, t_k$, respectively.
- The joint (cumulative) distribution function of this set of rvs is $F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k)$.

- Shift all the sampling times by a fixed amount τ denoting the *time shift*, thereby obtaining the new set of random variables: $X(t_1+\tau), X(t_2+\tau), \ldots, X(t_k+\tau)$.
- The joint distribution function of this set of rvs is $F_{X(t_1+\tau), \ldots, X(t_k+\tau)}(x_1, \ldots, x_k)$.
- The random process $X(t)$ is said to be stationary in the strict sense, or Strictly Stationary, if the invariance condition holds.

$$F_{X(t_1+\tau), \ldots, X(t_k+\tau)}(x_1, \ldots, x_k) = F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k) \quad \ldots\ldots(21)$$

- for all values of time shift τ, all positive integers k, and any possible choice of sampling times $t_1, \ldots, t_k$.

- A random process $X(t)$, initiated at time t=-∞, is Strictly Stationary if the joint distribution of any set of random variables obtained by observing the process $X(t)$ is invariant w.r.t the location of the origin $t = 0$, for all positive integer k, and all choices of the sampling times $t_1,t_2,\ldots,t_k$’.

- Two random processes $X(t)$ and $Y(t)$ are Jointly Strictly Stationary if the joint distributions of the two sets of random variables $X(t_1)$, …. , $X(t_k)$ and $Y(t_1')$,….,$Y(t_j')$ are invariant w.r.t the origin $t = 0$, for all positive integers k and j, and all choices of the sampling times $t_1,t_2,\ldots,t_k$ and $t_1',t_2',\ldots,t_j'$.

Points to Ponder

1. For $k = 1$,

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x) \qquad \text{for all } t \text{ and } \tau$$ ……(22)

The first-order distribution function of a strictly stationary random process is independent of time t.

2. For $k = 2$ and $\tau = -t_2$,

$$F_{X(t_1),X(t_2)}(x_1,x_2) = F_{X(0),X(t_1-t_2)}(x_1,x_2) \qquad \text{for all } t_1 \text{ and } t_2$$ ……(23)

The second-order distribution function of a strictly stationary random process depends only on the time difference between the sampling instants and not on the particular times at which the random process is sampled.

Weakly Stationary Processes

- Another important class of random processes is Weakly Stationary processes.
- A random process *X*(*t*) is said to be weakly stationary if its moments satisfy the following two conditions:

1. The Mean(μ_X) of process X(t) is constant for all time 't'.

2. The Autocorrelation function of the process X(t) depends solely on the difference between any two times at which the process is sampled. $R_X(t_1,t_2)=R_X(t_2-t_1)$

- Such processes are also referred to as Wide-Sense Stationary processes in the literature.
- Such a process may not be Stationary in the strict sense.
- All Stationary processes are wide sense stationary, but every wide-sense Stationary process may not be strictly Stationary.
- Truly Stationary process cannot occur in real life.

Mean of Stationary Process

- Consider a stationary random process *X*(*t*).
- The Mean of the process *is* the Expectation of the random variable obtained by sampling the process at some time *t*,

$\mu_X(t)=E[X(t)]= \int_{-\infty}^{\infty} x f_{X(t)}(x).\,\mathrm{dx}$, $f_{X_{(t)}}(x)$→pdf of X(t) ……(24)

- The pdf of a Stationary Process is independent of 't'

∴ $\mu_X(t)=\mu_X$ for all 't' ……(25)

Autocorrelation Function

- Autocorrelation Function of the random process $X(t)$ is the Expectation of the product of two random variables, $X(t_1)$ and $X(t_2)$, obtained by sampling the process $X(t)$ at times t_1 and t_2, respectively.

  ∴ $R_X(t_1,t_2) = E[X(t_1)X(t_2)]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1),X(t_2)}(x_1,x_2)\, dx_1\, dx_2 \quad \text{......(26)}$$

$f_{X(t_1),X(t_2)}(x_1,x_2)$ ←Second order pdf of random process

---

- For a stationary random process, $f_{X(t_1),X(t_2)}(x_1,x_2)$ depends only on the difference between observation times t_1 and t_2.
- The Autocorrelation function of a strictly stationary process depends only on the time difference $t_2 - t_1$.
- ∴ $R_X(t_1,t_2) = R_X(t_2-t_1)$ for all t_1 and t_2 ......(27)
- The Autocovariance function of a stationary random process X(t) is given by,

$$C_X(t_1, t_2) = E[(X(t_1)-\mu_X).(X(t_2)-\mu_X)]$$

$$= R_X(t_2-t_1) - \mu_X^2 \quad \text{......(28)}$$

- The Autocovariance function of a stationary random process depends only on the time difference (t_2-t_1).

- The autocovariance function of a weakly stationary process *X*(*t*) depends only on the time difference ($t_2 - t_1$).
- This equation also shows that knowing the mean and the autocorrelation function of the process *X*(*t*), can uniquely determine the autocovariance function.
- The mean and autocorrelation function are therefore sufficient to describe the first two moments of the process.

Properties of Autocorrelation Function

$R_X(\tau) = E[X(t+\tau)X(t)]$ for all ‘t’ ……(29)

$\tau \rightarrow$ represents time shift

PROPERTY 1 Mean-square Value

- The mean-square value of a stationary process X(t) is obtained from $R_X(\tau)$ simply by putting $\tau = 0$

$\therefore R_X(0) = E[X^2(t)]$ ……(30)

PROPERTY 2 Symmetry

- The autocorrelation function $R_X(\tau)$ of a stationary process X(t) is an even function of the time shift τ; i.e.,

i.e., $R_X(\tau) = R_X(-\tau)$ ……(31)

$\therefore R_X(\tau) = E[X(t)X(t-\tau)]$

- A graph of the autocorrelation function $R_X(\tau)$ plotted versus τ is symmetric about the origin.

PROPERTY 3 Bound on Autocorrelation Function

The autocorrelation function $R_X(\tau)$ attains its maximum magnitude at $\tau = 0$; i.e.,

$|R_X(\tau)| \leq R_X(0)$ ……(32)

PROPERTY 4 Normalization

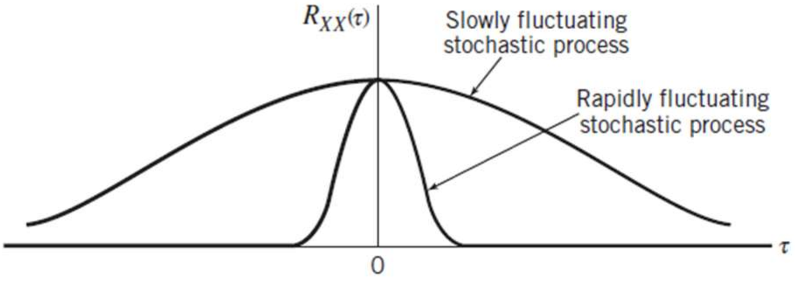
Given the normalized autocorrelation function,

$\rho_X(\tau)=\frac{R_X(\tau)}{R_X(0)}$ ……(33)

Values are confined to the range [-1, 1]

Physical Significance of the Autocorrelation Function

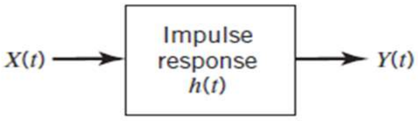
- The Autocorrelation function $R_X(\tau)$ is significant because it provides a means of describing the interdependence of two random variables obtained by sampling the stochastic process X(t) at times τ seconds apart.
- The more rapidly the random process *X*(*t*) changes with time, the more rapidly will the Autocorrelation function $R_X(\tau)$ decrease from its maximum $R_X(0)$ as τ increases.



Autocorrelation functions of slowly and rapidly fluctuating stochastic processes.

Transmission of a Random Process through a Linear Time-invariant Filter

- Let a Random process $X(t)$ be applied as input to a Linear Time-Invariant (LTI) filter of impulse response $h(t)$, to produce a new random process $Y(t)$ at the filter output.



Transmission of a random process thro' a LTI filter.

- In general, it is difficult to describe the probability distribution of the output random process $Y(t)$, even when the probability distribution of the input random process $X(t)$ is completely specified.

- Let random process X(t) be a Stationary process.
- The transmission of a process thro' a LTI filter is governed by Convolution integral, i.e., express output *Y*(*t*) as,

$$Y(t) = \int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1 \quad \tau_1 \rightarrow \text{ is a local time.}$$

- Hence, the Mean of *Y*(*t*) is,

$$\mu_Y(t) = E[Y(t)] = E[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1\ ] \quad ......(34)$$

$$= \int_{-\infty}^{\infty} h(\tau_1)\mathbb{E}[X(t-\tau_1)]\,d\tau_1$$

$$= \int_{-\infty}^{\infty} h(\tau_1)\mu_X(t-\tau_1)\,d\tau_1 \quad ......(35)$$

- When i/p *X*(*t*) is stationary, the mean is a constant,

$\therefore \mu_X(t-\tau_1) = \mu_X(t) = \mu_X =$ constant

$$\therefore \mu_Y(t) = \mu_X \int_{-\infty}^{\infty} h(\tau_1)\,d\tau_1 = \mu_X H(0) \quad ......(36)$$

- where *H*(0) is the zero-frequency (dc) response of the system.
- The mean of random process *Y*(*t*) produced at the output of a LTI filter in response to a stationary process *X*(*t*), acting as the input process, is equal to the mean of *X*(*t*) times the zero-frequency(dc) response of the filter.

- <u>Autocorrelation Function of Y(t)</u>:

$R_Y(t, u)$= E[Y(t)Y(u)]

- where t and u denote two values of the time at which the output process $Y(t)$ is sampled.

$R_Y(t, u)$=

$$\mathbb{E}\left[\int_{-\infty}^{\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1 \int_{-\infty}^{\infty} h(\tau_2)X(u-\tau_2)\,d\tau_2\right] \quad \text{......(37)}$$

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

$$= \int_{-\infty}^{\infty}\left[h(\tau_1)\int_{-\infty}^{\infty} d\tau_2\, h(\tau_2)\; R_X\; (t-\tau_1, u-\tau_2)\right] d\tau_1 \quad \text{......(38)}$$

- When the input $X(t)$ is a stationary process, the autocorrelation function of $X(t)$ is only a function of the <u>difference between the sampling times</u> $t-\tau_1$ and $u-\tau_2$.
- Put $\tau = u - t$

$$R_Y\;(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{X.}\;(\tau+\tau_1-\tau_2)\,d\tau_1\,d\tau_2 \quad \text{......(39)}$$

- which depends only on the time difference τ

From eqn (36) and eqn(39),

- If the input to a stable LTI filter is a stationary random process, then the output of the filter is also a stationary random process.
- Use Property 1 of the Autocorrelation function $R_Y(\tau)$, it follows, that the Mean-Square value of the output process *Y*(*t*) is obtained by putting $\tau = 0$ in eqn (39),

$$R_Y(0) = \mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_X(\tau_1-\tau_2)\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2 \qquad \text{......(40)}$$

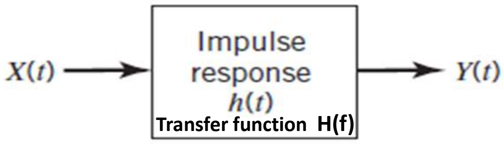
- This is a constant!

Power Spectral Density (PSD)

- Power Spectral Density (PSD) of a Wide- Sense Stationary (WSS) Random Process X(t) is the Fourier Transform of the Autocorrelation Function (ACF).
- $S_X(f)$= Fourier Transform of $R_X(\tau)$

$$= \int_{-\infty}^{\infty} R_X(\tau)\, e^{j2\pi f\tau} d\tau \qquad \text{......(41)}$$

LTI Filter



Transmission of a Random process thro' a LTI filter.

The impulse response of the Filter is the <u>inverse Fourier Transform</u> of Transfer function.

$\therefore$ $h(\tau_1) = \int_{-\infty}^{\infty} H(f).\, e^{j2\pi f\tau_1}\, df$ ......(42)

From eqn(40), $\mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1)h(\tau_2)R_{XX}(\tau_1 - \tau_2)\, d\tau_1\, d\tau_2$

$E[Y^2(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} H(f).\, e^{j2\pi f\tau_1}\, df\right] h(\tau_2)\, R_X(\tau_2 - \tau_1)\, d\tau_1\, d\tau_2$

$= \int_{-\infty}^{\infty} df.\, H(f) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2) \int_{-\infty}^{\infty} R_X(\tau_2 - \tau_1)\, e^{j2\pi f\tau_1} d\tau_1$

Let $\tau = \tau_2 - \tau_1$

$\therefore$ $E[Y^2(t)] = \int_{-\infty}^{\infty} df\, H(f) \int_{-\infty}^{\infty} d\tau_2 h(\tau_2)\, e^{+j2\pi f\tau_2} \int_{-\infty}^{\infty} R_X(\tau)\, e^{j2\pi f\tau} d\tau$

$\int_{-\infty}^{\infty} h(\tau_2)\, e^{+j2\pi f\tau_2} d\tau_2 = H^*(f)$ ← Conjugate of H(f)

$\therefore$ $E[Y^2(t)] = \int_{-\infty}^{\infty} df\, |H(f)|^2 \int_{-\infty}^{\infty} R_X(\tau)\, e^{j2\pi f\tau} d\tau$

$|H(f)|^2$ ← Square magnitude response of H(f)

$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{j2\pi f\tau} d\tau$ ← <u>Power Spectral Density</u>!

$\therefore$ $E[Y^2(t)] = \int_{-\infty}^{\infty} |H(f)|^2 S_X(f)\, df$ ......(43)

Relevance of Eqn(43)

- The Mean Square value of the output of the LTI filter, in response to a Stationary Random process at its input, is equal to the integral over all frequencies of the Power Spectral Density of the input Stationary Random process multiplied by the Squared Magnitude Response of the Filter.

Einstein-Wiener-Khintchine Relations

Plate-1

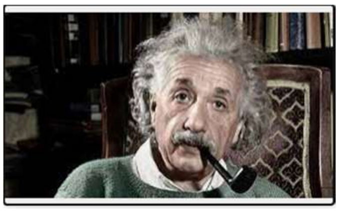


Plate-2

Plate-3

Norbert Wiener proved this theorem for the case of a deterministic function in 1930. **Guess who is this person!**

Aleksandr Khinchin later formulated an analogous result for stationary stochastic processes and published that probabilistic analogue in 1934. **Guess who is this person!**

Albert Einstein explained, without proofs, the idea in a brief two-page memo in 1914. **Guess the pioneer!!**

Einstein-Wiener-Khintchine Relations

- Einstein-Wiener-Khintchine (EWK) Relations gives the relationship between Power Spectral Density and Autocorrelation Function of a Stationary Random Process.
- The Power Spectral Density $S_X(f)$ and the Autocorrelation function $R_X(\tau)$ of a Stationary Random Process X(t) form a Fourier transform pair.

$$S_X(f)= \int_{-\infty}^{\infty} R_X(\tau)\, e^{j2\pi f\tau} d\tau \quad \text{......(44a)}$$

$$R_X(\tau)= \int_{-\infty}^{\infty} S_X(f)\, e^{-j2\pi f\tau} df \quad \text{......(44b)}$$

Properties of the PSD

Property 1:

The Power Spectral Density (PSD) of a Stationary random process for zero frequency (dc) value is equal to the total area under the graph of Autocorrelation function (ACF).

$$\therefore \ S_X(0)= \int_{-\infty}^{\infty} R_X(\tau)\, d\tau \quad \text{......(45)}$$

Property 2:

The Mean Square value of a Stationary random process is equal to the total area under the graph of Power Spectral Density (PSD).

$$\therefore \ R_X(0)=E[X^2(t)]= \int_{-\infty}^{\infty} S_X(f)\, df \quad \text{......(46)}$$

Property 3:

The Power Spectral Density (PSD) of a Stationary random process is always non-negative, i.e.,

$$S_X(f) \geq 0 \text{ for all } 'f' \quad ......(47)$$

Property 4:

The Power Spectral Density (PSD) of a real-valued random process is an even function of frequency, i.e.,

$$S_X(-f) = S_X(f) \quad ......(48)$$

Property 5:

The normalized Power Spectral Density (PSD) of a random process has properties associated with a probability density function.

$$p_X(f) = \frac{S_X(f)}{\int_{-\infty}^{\infty} S_X(f).df} \geq 0, \text{ for all 'f'} \quad ......(49)$$

- The total area under the function p_X(f) is unity.

PSD of Output Random Process

- $S_Y(f)$ is the Power Spectral Density (PSD) of the o/p random process obtained by passing random process X(t) thro’ an LTI filter of transfer function H(f).
- PSD of random process is equal to <u>Fourier transform</u> of its Autocorrelation function(ACF).

$\therefore$ $S_Y(f)= \int_{-\infty}^{\infty} R_Y(\tau)\, e^{j2\pi f\tau} d\tau$ ……(50)

Use eqn(39)

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

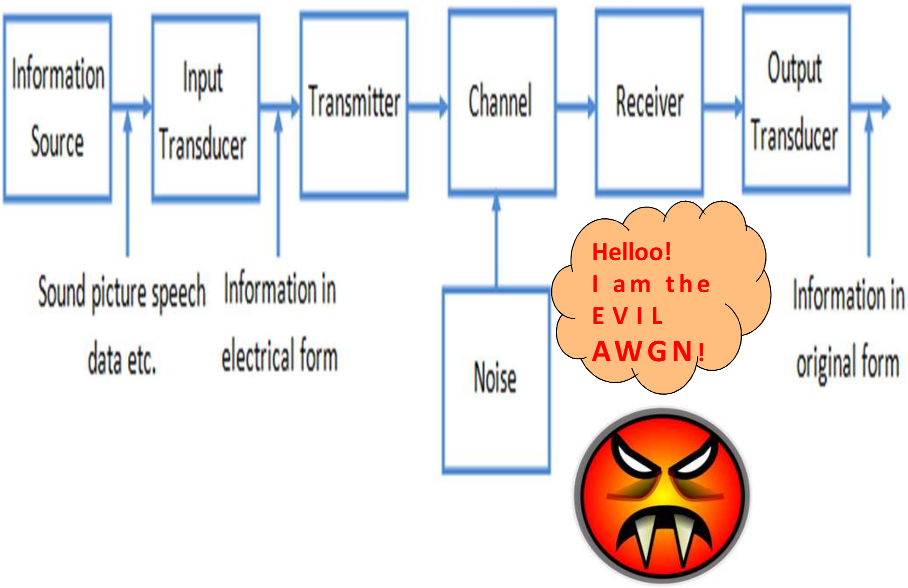
Let $\tau+\tau_1-\tau_2=\tau_0 \rightarrow \tau=\tau_0-\tau_1+\tau_2$

$S_Y(f) = H(f).\ H^*(f).\ S_X(f)$

$= |H(f)|^2.\ S_X(f)$ ……(51)

The Power Spectral Density of the o/p random process Y(t) is equal to the Power Spectral Density of the i/p random process X(t) multiplied with the squared magnitude response of the Filter.

Block Diagram of a Generic Digital Communication System



Helloo!
I am the
EVIL
AWGN!

Entropy

- The amount of information $I(s_k)$ produced by the source during an arbitrary signaling interval depends on the symbol s_k emitted by the source at the time.
- The self-information $I(s_k)$ is a discrete random variable that takes on the values $I(s_0)$, $I(s_1)$, …, $I(s_{K-1})$ with probabilities p_0, p_1, …, p_{K-1} respectively.
- The Expectation of $I(s_k)$ over all the probable values taken by the random variable S is given by:

$$H(S) = \mathbb{E}[I(s_k)] = \sum_{k=0}^{K-1} p_k I(s_k) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right) \quad \text{………(8)}$$

- The quantity *H(S)* is called the Entropy, defined as follows:
- The Entropy of a discrete random variable, representing the output of a Source of Information, is a measure of the Average Information content per source symbol.
- Note that the entropy *H*(*S*) is independent of the alphabet $\mathcal{S}$
- It depends only on the probabilities of the symbols in the alphabet $\mathcal{S}$ of the source.

Properties of Entropy

Entropy of the discrete random variable *S* is bounded as follows:

$$0 \leq H(S) \leq \log_2 K \qquad \text{………(9)}$$

where *K* is the number of symbols in the alphabet $\mathcal{S}$.

Two properties:

1. $H(S) = 0$, if, and only if, the probability $p_k = 1$ for some k, and the remaining probabilities in the set are all zero, this lower bound on entropy corresponds to no uncertainty.

2. $H(S) = \log K$, if, and only if, $p_k = 1/K$ for all k (i.e., all the symbols in the source alphabet $\mathcal{S}$ are equiprobable). this upper bound on entropy corresponds to maximum uncertainty.

Proof of eqn(9) is needed to be undertaken in Assgn-1!

Problem – 3

A source emits one of the three possible symbols during each signalling interval, with probabilities p_0 = 1/2, p_1 = 1/4 and p_2 = 1/4. Calculate the Entropy of the source.

Sol:

$$H(S) = \mathbb{E}[I(s_k)] = \sum_{k=0}^{K-1} p_k I(s_k) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$$

$$\therefore \ H(S) = p_0\log_2(1/ p_0) + p_1\log_2(1/ p_1) + p_2\log_2(1/ p_2)$$

$$\therefore \ H(S) = 1/2 \times \log_2(2) + 1/4 \times \log_2(4) + 1/4 \times \log_2(4)$$

$$= 1/2 \times 1 + 1/4 \times 2 + 1/4 \times 2 = \underline{1.5 \text{ bits/symbol}}$$

Mutual Information

- The channel output Y is a noisy version of the channel input X, and that the Entropy H(S) is a measure of the prior uncertainty about X, how can we measure the uncertainty about X after observing Y?
- Define Conditional Entropy of X, given that Y $=y_k$.

$$\therefore \quad H(X|Y = y_k) = \sum_{j=0}^{J-1} p(x_j|y_k)\log_2\left(\frac{1}{p(x_j|y_k)}\right) \qquad \text{.........(20)}$$

- This quantity is itself a random variable that takes on the values $H(X|Y = y_0), \ \dots, \ H(X|Y = y_{K-1})$ with probabilities $p(y_0), \ \dots, \ p(y_{K-1})$.

- The mean of entropy $H(X|Y = y_k)$ is,

$$\therefore \quad H(X|Y) = \sum_{k=0}^{K-1} H(X|Y = y_k)p(y_k)$$

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j|y_k)p(y_k) \log_2\left(\frac{1}{p(x_j|y_k)}\right) \quad \text{.........(21)}$$

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k) \log_2\left(\frac{1}{p(x_j|y_k)}\right)$$

$$\because \quad p(x_j, y_k) = p(x_j|y_k)p(y_k) \quad \text{.........(22)}$$

- The quantity H(X|Y) is called a Conditional Entropy.
- It represents the amount of uncertainty remaining about the channel input after the channel output has been observed.

- Since the Entropy H(X) represents our uncertainty about the channel input before observing the channel output, and the Conditional Entropy H(X|Y) represents our uncertainty about the channel input after observing the channel output, it follows that the difference H(X) - H(X|Y) must represent our uncertainty about the channel input that is resolved by observing the channel output.
- This important quantity is called the Mutual Information of the channel.

$$\therefore \quad I(X;Y) = H(X) - H(X|Y) \quad \text{.........(23)}$$

- Similarly, $I(Y;X) = H(Y) - H(Y|X)$ ....(24)
- H(Y) is entropy of channel output and H(Y|X) is conditional entropy of channel output given channel input.

Properties of Mutual Information

Property 1

- The mutual information of a channel is symmetric; i.e.,

$$I(X;Y) = I(Y;X) \quad \text{………(25)}$$

- where the mutual information I(X,Y) is a measure of the uncertainty about the channel input that is resolved by observing the channel output.
- the mutual information I(Y,X) is a measure of the uncertainty about the channel output that is resolved by sending the channel input.

PROPERTY 2 **Non-negativity**

- The mutual information is always nonnegative; i.e.,

$$I(X;Y) \geq 0 \quad \text{………(30)}$$

- To prove this property, note from eqn(18) that,

$$p(x_j|y_k) = \frac{p(x_j, y_k)}{p(y_k)} \quad \text{………(31)}$$

- Hence, substituting eqn(31) into eqn(27), express the mutual information of the channel,

$$I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{p(x_j, y_k)}{p(x_j)p(y_k)}\right) \quad \text{………(32)}$$

- If, the input and output symbols of the channel are statistically independent; that is,

$$p(x_j, y_k) = p(x_j)p(y_k) \quad \text{for all } j \text{ and } k \qquad \text{.........(33)}$$

- Substitution eqn(33) into eqn (32) yields,

$$I(X;Y) = 0$$

- We cannot lose information, on the average, by observing the output of a channel.

PROPERTY 3 **Expansion of the Mutual Information**

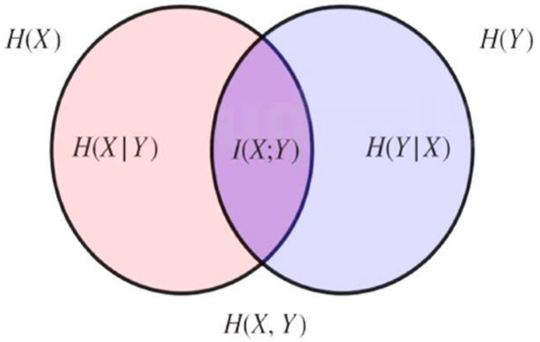
- The mutual information of a channel is related to the joint entropy of the channel input and channel output by,

$$I(X;Y) = H(X) + H(Y) - H(X, Y) \qquad \text{.........(34)}$$

- where the joint entropy H(X, Y) is defined by,

$$H(X, Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{1}{p(x_j, y_k)}\right) \qquad \text{.........(35)}$$

RELATION AMONG ENTROPIES and MI:VENN DIAGRAM



JOINT ENTROPY

$$
\begin{aligned}
H(X,Y) &= H(X)+H(Y/X) \\
&= H(Y) + H(X/Y)
\end{aligned}
$$

RELATION AMONG ENTROPIES

- **CHAIN RULE**

$$H(X,Y) = H(X) + H(Y/X)$$
$$= H(Y) + H(X/Y)$$

- If X and Y are independent each other, then
- $H(Y/X) = H(Y)$, Also $H(X/Y) = H(X)$
- Then, $H(X,Y) = H(X) + H(Y)$

Introduction

- So far we have dealt with discrete random variables and random vectors.
- In this module, the treatment will be on continuous random variables and random vectors…
- Makes sense as all real-life signal behaviour including our beloved AWGN are continuous in nature.
- All of us know the continuous versus discrete matter.
- Channel can be wireless or wired and are best modelled by continuous behaviour.
- So there’s the need to gently progress from discrete to continuous statistical characterization….

Continuous Channels

- For Radio, TV etc, the modulating messages are continuous speech or picture signal.
- This message set can be treated as equivalent to a continuous space whose sample points form a continuum, in contrast to the discrete case.
- Define a continuous channel as one whose input is a sample point from a continuous sample space and the output is a sample point belonging to the same sample space or a different one.
- A Zero Memory Continuous Channel is one in which the channel output statistically depends on the corresponding channels without memory.
- Claude Shannon has again injected a beacon of light in the dreary darkness by formulation of his third theorm– Shannon-Hartley law

Continuous Random Ensembles

- Consider a continuous random variable X with the probability density function $f_X(x)$.
- The Differential Entropy is given by,

$$h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2\left[\frac{1}{f_X(x)}\right] \mathrm{d}x$$

……….(1)

- view the continuous random variable *X* as the limiting form of a discrete random variable that…
- assumes the value $x_k = k\Delta x$, where $k = 0, \pm 1, \pm 2, \ldots$, and Δx approaches zero.
- By definition, the continuous random variable *X* assumes a value in the interval $[x_k, x_k + \Delta x]$ with probability $f_X(x_k)\,\Delta x$.
- Hence, permitting $\underline{\Delta x \text{ to approach zero}}$, the Ordinary Entropy of the Continuous Random Variable *X* takes the limiting form…

$$H(X) = \lim_{\Delta x \to 0} \sum_{k=-\infty}^{\infty} f_X(x_k)\Delta x \log_2\left(\frac{1}{f_X(x_k)\Delta x}\right)$$

$$= \lim_{\Delta x \to 0} \left(\sum_{k=-\infty}^{\infty} f_X(x_k) \log_2\left(\frac{1}{f_X(x_k)}\right)\Delta x - \log_2 \Delta x \sum_{k=-\infty}^{\infty} f_X(x_k)\Delta x \right)$$

$$= \int_{-\infty}^{\infty} f_X(x) \log_2\left(\frac{1}{f_X(x)}\right) \mathrm{d}x - \lim_{\Delta x \to 0}\left(\log_2 \Delta x \int_{-\infty}^{\infty} f_X(x_k)\, \mathrm{d}x\right)$$

$$= h(X) - \lim_{\Delta x \to 0} \log_2 \Delta x \qquad \textbf{……….(2)}$$

- In the last line of eqn(2), use has been made of **eqn(1)** and the fact that the **total area under the curve of the probability density function $f_X(x)$ is unity.**
- In the limit as **Δx approaches zero**, the term –**$\log_2\Delta x$ approaches infinity**.
- This means that the entropy of a continuous random variable is infinitely large.
- This to be true since a continuous random variable may assume a value anywhere in interval [-∞, ∞];
- There are **uncountable infinite numbers of probable outcomes**.

- To avoid the problem associated with the term **$\log_2\Delta x$**, adopt **$h(X)$** as a Differential Entropy, with the term **$-\log_2\Delta x$** serving merely as a reference.
- Since the information send over a channel is actually the difference between two Entropy terms that have a common Reference, the information will be the same as the difference between the corresponding **Differential Entropy terms**.
- So it's perfectly justified in using term **$h(X)$**,defined in eqn(1), as the **Differential Entropy** of the **continuous random variable X**.

∵

Continuous Random variable vector

- When a continuous random vector **X** consists of *n* random variables $X_1, X_2, \ldots, X_n$,define the Differential Entropy of **X** as the n-fold integral,

$$h(\mathbf{X}) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) \log_2\left[\frac{1}{f_{\mathbf{X}}(\mathbf{x})}\right] d\mathbf{x} \qquad \text{..........(3)}$$

- where $f_{\mathbf{X}}(\mathbf{x})$ is the <u>Joint Probability Density function</u> of **X**.

EXAMPLE: **Uniform Distribution**

- Consider a random variable *X* uniformly distributed over the interval (0, *a*).
- The probability density function of *X* is,

$$f_X(x) = \begin{cases} \frac{1}{a}, & 0 < x < a \\ 0, & \text{otherwise} \end{cases} \qquad \text{..........(4)}$$

∴ Differential entropy $h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2\left[\frac{1}{f_X(x)}\right] dx$

$$\therefore \quad h(X) = \int_0^a \frac{1}{a}\log(a)\, dx \;=\; \log a$$

Note that log *a* < 0 for *a* < 1. So, unlike a Discrete Random Variable, Differential entropy of a Continuous Random Variable can be <u>negative</u>.

Points to Ponder

PROPERTY -1

For any finite variance, a Gaussian random variable has the largest differential entropy attainable by any other random variable.

PROPERTY- 2

The entropy of a Gaussian random variable is uniquely determined by its variance (i.e., the entropy is independent of the mean).

Due to Property-1, the Gaussian Channel model is so widely used as a Conservative Model in the study of Digital Communication Systems.

Mutual Information

- Define the mutual information between the pair of continuous random variables X and Y as follows:

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log_2\left[\frac{f_X(x|y)}{f_X(x)}\right]dx\,dy \quad \text{..........(5)}$$

- where $f_{X,Y}(x,y)$ is the joint probability density function of X and Y and $f_X(x/y)$ is the conditional probability density function of X given $Y = y$.

Properties of Mutual Information

1. I(X;Y) = I (Y;X)
2. I(X;Y) ≥ 0
3. I(X;Y) = h(X) – h(X|Y)
4. I(X;Y) = h(Y)-h(Y|X)

h(X) → differential entropy of X

h(Y) → differential entropy of Y

h(X|Y) → conditional differential entropy of X given Y

h(Y|X) → conditional differential entropy of Y given X

$$h(X|Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\log_2\left[\frac{1}{f_X(x|y)}\right] \mathrm{d}x\,\mathrm{d}y$$

……….(6)

Differential Entropy- Gaussian Variable

Gaussian distribution with zero mean and variance σ^2.

$$\mathrm{f}_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

$$h(X) = -\int_S \mathrm{f}_X(x) \ln \mathrm{f}_X(x) dx$$ **nats/message**

$$= -\int \mathrm{f}_X(x) \ln\left[\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}\right] dx$$

$$= \int \mathrm{f}_X(x) \ln\left[\sqrt{2\pi\sigma^2}\right] dx + \int x^2/2\sigma^2\ \mathrm{f}_X(x) dx$$

$$= \frac{1}{2}\ln 2\pi\sigma^2 + \frac{E[X^2]}{2\sigma^2}$$

$$= \frac{1}{2}\ln 2\pi\sigma^2 + \frac{1}{2}$$ **nats/message**

$$= \frac{1}{2}\log_2 2\pi e\sigma^2 \text{ bits}$$ **/message**

Points to Ponder

PROPERTY -1

For any finite variance, a Gaussian random variable has the largest differential entropy attainable by any other random variable.

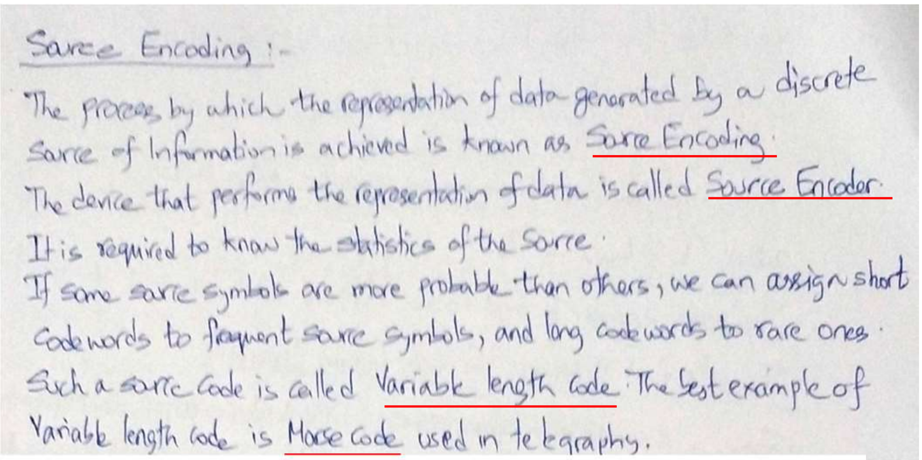
PROPERTY- 2

The entropy of a Gaussian random variable is uniquely determined by its variance (i.e., the entropy is independent of the mean).

Due to Property-1, the Gaussian Channel model is so widely used as a Conservative Model in the study of Digital Communication Systems.

Source Coding

ECT 305 Module III Part-1



Source Encoding :-

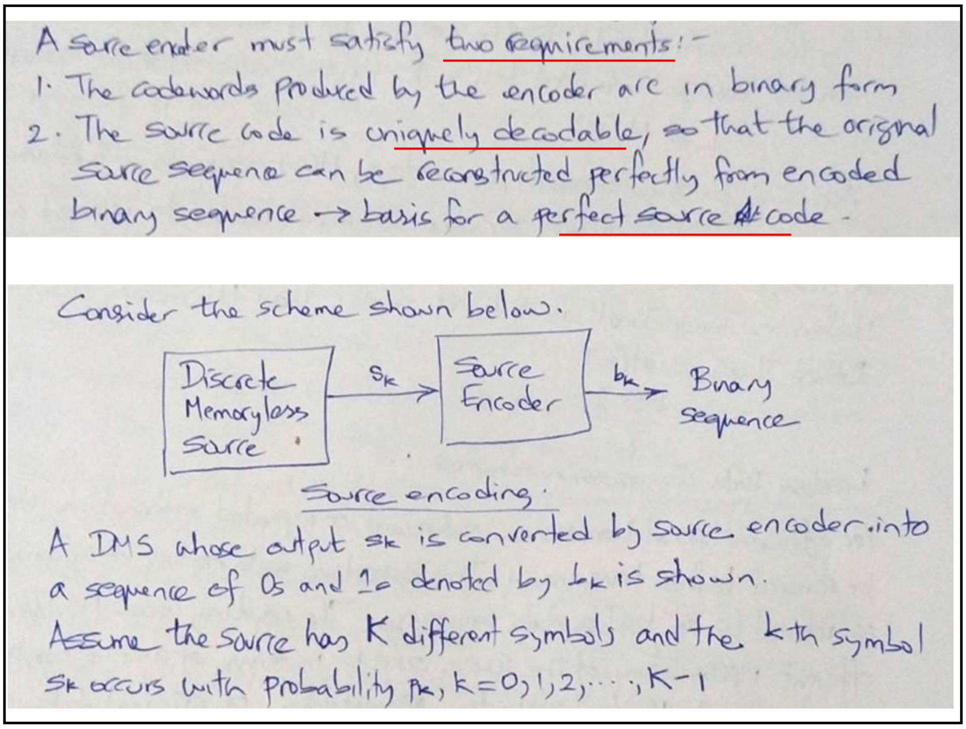
The process by which the representation of data generated by a discrete Source of Information is achieved is known as Source Encoding.

The device that performs the representation of data is called Source Encoder.

It is required to know the statistics of the Source.

If some source symbols are more probable than others, we can assign short codewords to frequent source symbols, and long codewords to rare ones.

Such a source code is called Variable length code. The best example of Variable length code is Morse code used in telegraphy.



A source encoder must satisfy two requirements:-

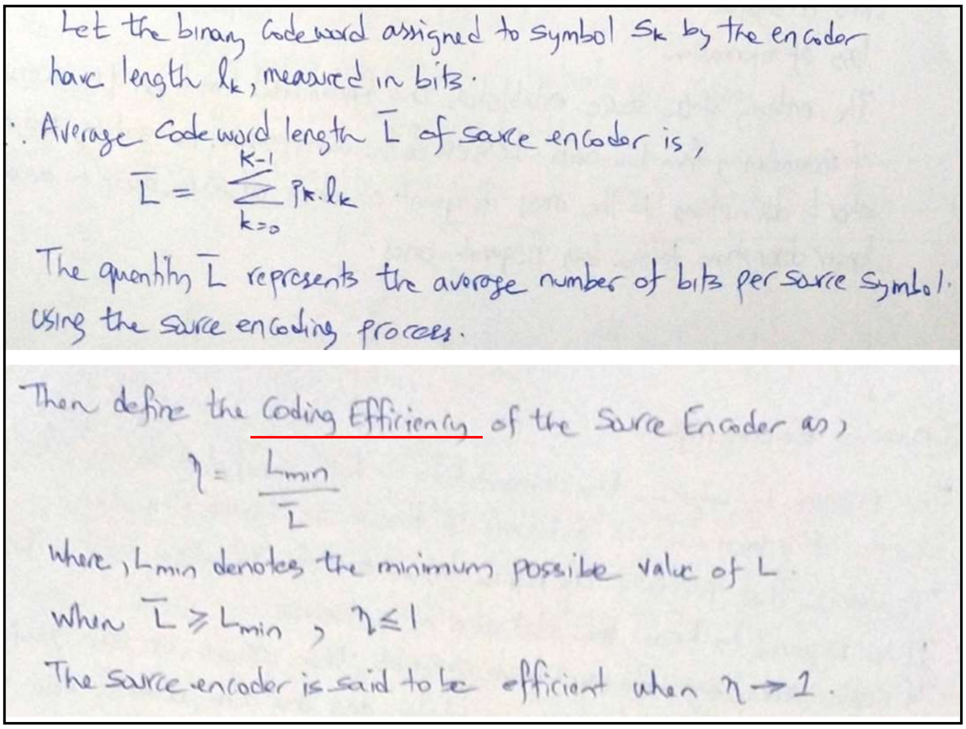
1. The codewords produced by the encoder are in binary form
2. The source code is uniquely decodable, so that the original source sequence can be reconstructed perfectly from encoded binary sequence → basis for a perfect source code.

Consider the scheme shown below.

Source encoding.

A DMS whose output s_k is converted by source encoder into a sequence of 0s and 1s denoted by b_k is shown.

Assume the source has K different symbols and the kth symbol s_k occurs with probability p_k, $k = 0, 1, 2, \cdots, K-1$



Let the binary codeword assigned to symbol s_k by the encoder have length l_k, measured in bits.

∴ Average codeword length $\bar{L}$ of source encoder is,

$$\bar{L} = \sum_{k=0}^{K-1} p_k . l_k$$

The quantity $\bar{L}$ represents the average number of bits per source symbol using the source encoding process.

Then define the Coding Efficiency of the Source Encoder as,

$$\eta = \frac{L_{min}}{\bar{L}}$$

Where, L_{min} denotes the minimum possible value of L.

When $\bar{L} \geqslant L_{min}$, $\eta \leq 1$

The source encoder is said to be efficient when $\eta \to 1$.

Shannon's first Theorem: Source Coding Theorem

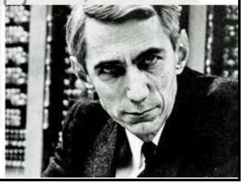
Given a discrete memoryless source (DMS) whose output is represented by the random variable S, the entropy H(S) imposes the following bound on the average codeword length $\bar{L}$ for any source encoding scheme.

$$\bar{L} \geq H(S)$$

According to this theorem, the entropy H(S) represents a fundamental limit on average number of bits per source symbol needed to represent a DMS. It can be made small as but not smaller than the entropy H(S).

Putting $L_{min} = H(S)$

$$\therefore \eta = \frac{H(S)}{\bar{L}}, \quad \eta \leq 1.$$



Channel Capacity

- Given a channel characterized by the transition probability distribution {$p(y_k|x_j)$}, the *channel capacity*, which is formally defined in terms of the Mutual Information between the channel input and output as follows:

$$C = \max_{\{p(x_j)\}} I(X;Y) \quad \text{bits per channel use} \quad \ldots\ldots\ldots\ldots\ldots(3)$$

- The maximization in (3) is done, subject to two input probabilistic constraints:

$$p(x_j) \geq 0 \quad \text{for all } j \qquad\qquad \sum_{j=0}^{J-1} p(x_j) = 1$$

Channel Capacity

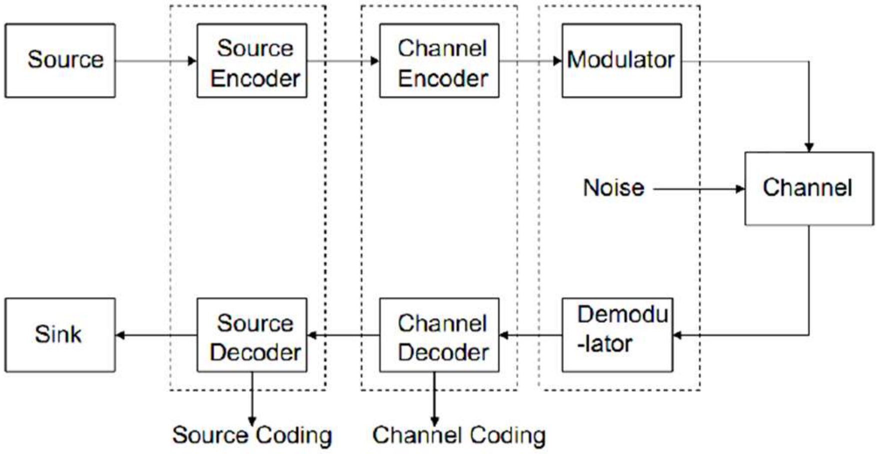
- The channel capacity of a discrete memoryless channel, C, is defined as the maximum Mutual Information $I(X;Y)$ in any single use of the channel (i.e., signaling interval), where the maximization is over all possible input probability distributions $\{p(x_j)\}$ on X.
- The channel capacity is clearly an intrinsic property of the channel.

By making the order 'n' of an extended Prefix source encoder large enough, we can make the code faithfully represent the discrete memoryless source S, as closely as desired.

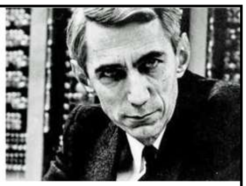
So, the average codeword length of an extended prefix code can be made as small as the Entropy of the source, provided that the extended code has a high enough order, in accordance with the Source Coding theorem.

However, the price we have to pay for decreasing the average codeword length is increased decoding complexity that is brought about by the high order of the extended prefix code.

Digital Communication System



Shannon's second theorem→ Noiseless Coding Theorem



- Let a Discrete Memoryless Source (DMS) with an alphabet *S* have entropy *H*(*S*) for random variable *S* and produce symbols once every T_s seconds.
- Let a Discrete Memoryless Channel have capacity *C* and be used once every T_c seconds, Then, if,

$$\frac{H(S)}{T_s} \leq \frac{C}{T_c} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

- There exists a coding scheme whose source output can be transmitted over the channel and be reconstructed with an arbitrarily small probability of error.
- The parameter C/T_c is called the critical rate; when eqn (1) is satisfied with the equality sign, the system is said to be signaling at the critical rate.
- Conversely, if, $\frac{H(S)}{T_s} > \frac{C}{T_c}$ ……………………….(2)
- it is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error.

The theorem specifies the channel capacity *C* as a fundamental limit on the rate at which the transmission of reliable error-free messages can take place over a Discrete Memoryless Channel.

There are two limitations of the theorem:

1. The channel-coding theorem does not show us how to construct a good code.

   Rather, view the theorem as an existence proof in that if eqn-1 is satisfied, then good codes do exist.
2. No precise result for the probability of symbol error after decoding the channel output.

   Rather, the probability of symbol error tends to zero as the length of the code increases, again provided that the condition of eqn-1 is satisfied.

Pulse Modulation

- In Continuous-Wave (CW) Modulation, some parameter of a sinusoidal carrier wave is varied continuously in accordance with the message signal.
- In Pulse Modulation, some parameter of a pulse train is varied in accordance with the message signal. Two types →
  - **1.** Analog pulse modulation, in which a periodic pulse train is used as the carrier wave and some characteristic of each pulse (e.g., amplitude, duration, or position) is varied in a continuous manner in accordance with the corresponding sample value of the message signal.
  - Thus, in analog pulse modulation, information is transmitted basically in analog form but the transmission takes place at discrete times.
  - **2.** Digital pulse modulation, in which the message signal is represented in a form that is discrete in both time and amplitude, thereby permitting transmission of the message in digital form as a sequence of coded pulses

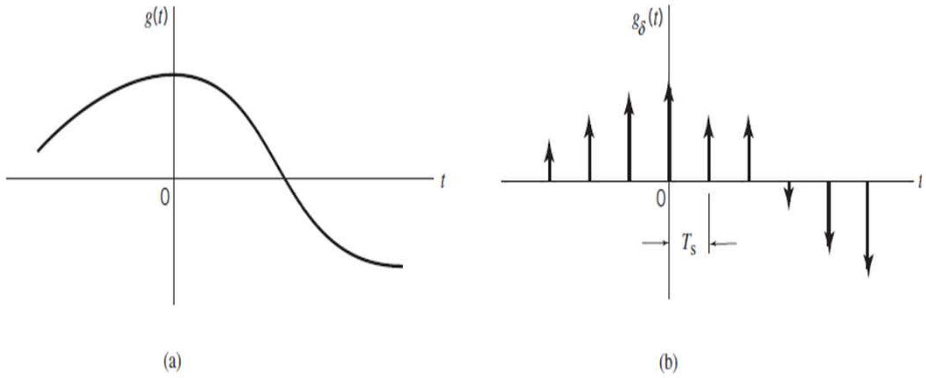
Waveform Coding

- The use of coded pulses for the transmission of analog information-bearing signals represents a basic component in digital communications.
- Digital pulse modulation, is described as the conversion of analog waveforms into coded pulses, hence Waveform Coding.
- The conversion may be viewed as the transition from Analog to Digital communications.

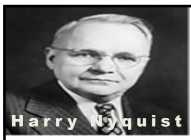
Sampling Process

- The Sampling process is analyzed in the time domain.
- It is an operation that is basic to Digital signal processing and digital communications.
- The sampling process enables an Analog signal to be converted into a corresponding sequence of Samples that are usually spaced uniformly in time.
- It is mandatory to choose the Sampling rate properly in relation to the Bandwidth of the message signal, so that the Sequence of Samples uniquely defines the original Analog signal.

- Consider an arbitrary signal $g(t)$ of finite energy, which is specified for all time t.
- A segment of the signal $g(t)$ is shown.
- Suppose that we sample the signal $g(t)$ instantaneously and at a uniform rate, once every T_s seconds.
- Obtain an infinite sequence of samples spaced T_s seconds apart and denoted by $\{g(nT_s)\}$, where n takes on all possible integer values, positive as well as negative.
- We refer to T_s as the sampling period, and to its reciprocal $f_s = 1/T_s$ as the sampling rate.
- this ideal form of sampling is called Instantaneous Sampling.



The sampling process.
(a) Analog signal.
(b) Instantaneously sampled version of the Analog signal.



The Sampling Theorem

1. A band-limited signal of finite energy that has no frequency components higher than *W* hertz is completely described by specifying the values of the signal samples at instants of time separated by 1/2*W* seconds. *Frequency-Domain Description of Sampling*

2. A band-limited signal of finite energy that has no frequency components higher than *W* hertz is completely recovered from a knowledge of its samples taken at the rate of 2*W* samples per second. *Reconstruction formula*

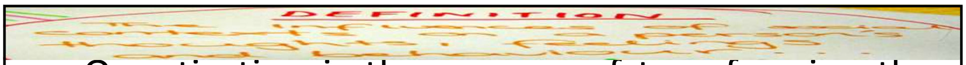
- Part 1 of the Sampling theorem, is performed in the Transmitter.
- Part 2 of the Sampling theorem, is performed in the Receiver.
- For a signal bandwidth of W hertz, the sampling rate of $2W$ samples per second is called the Nyquist rate.
- Its reciprocal $1/2W$ (measured in seconds) is called the Nyquist interval.

EXAMPLE-1-**Sampling of Voice Signals**

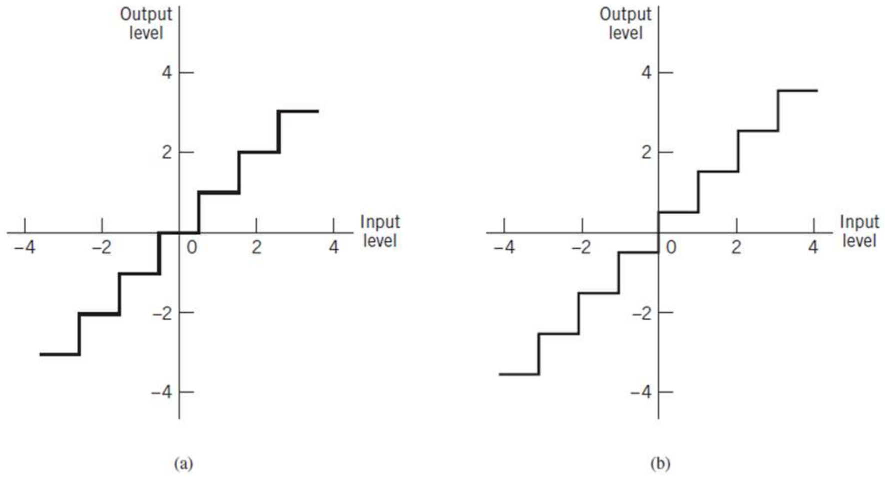
- The frequency band, from 100 Hz to 3.1 kHz, is adequate for Telephonic communication.
- This limited frequency band is accomplished by passing the voice signal through a Low-Pass Filter with its cutoff frequency set at 3.1 kHz.
- Such a Filter may be viewed as an Anti-Aliasing Filter.
- With such a cutoff frequency, the Nyquist rate is $f_s = 2 \times 3.1 = 6.2$ kHz.
- The standard Sampling rate for the waveform coding of voice signals is 8 kHz.
- Design specifications for the Reconstruction (Low-Pass) Filter in the Receiver are as follows:
  - Cut-off frequency 3.1 kHz
  - Transition band 6.2 to 8 kHz
  - Transition-band width 1.8 kHz.

Quantization

- An analog message signal (e.g., voice) has a continuous range of amplitudes and its samples have a continuous amplitude range.
- Within the finite amplitude range of the signal, there is an infinite number of amplitude levels.
- It is not necessary to transmit the exact amplitudes of the samples as any human sense (the ear or the eye) can detect only finite intensity differences.
- This means that the message signal may be approximated by a signal constructed of discrete amplitudes selected on a minimum error basis from an available set.
- The use of a finite number of discrete amplitude levels is a basic condition of waveform coding using PCM.
- Assign the discrete amplitude levels with sufficiently close spacing, makes the approximated signal practically indistinguishable from the original message signal.



- Quantization is the process of transforming the real sample amplitude $m(nT_s)$ of a message signal $m(t)$ at time $t = nT_s$ into a discrete integer value amplitude $v(nT_s)$ taken from a finite set of possible integer amplitudes.
- Quantizer, the device performing the quantization process is memoryless *and* instantaneous, which means that the transformation at time $t = nT_s$ is not affected by earlier or later samples of the message signal $m(t)$.
- This simple form of Scalar Quantization, though not optimum, is commonly used in practice.

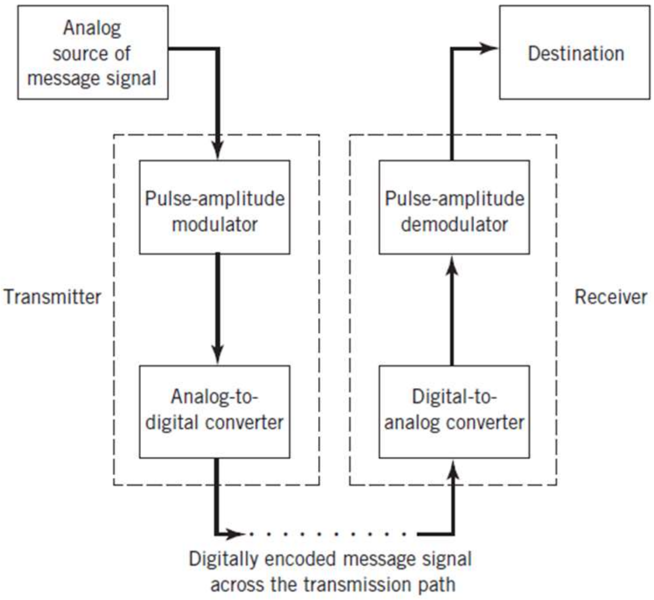


Two types of quantization: (a) Midtread and (b) Midrise.

- Quantizers can be of a uniform or nonuniform type.
- In a uniform quantizer, the representation levels are uniformly spaced; otherwise, the quantizer is nonuniform.
- The quantizer characteristic can also be of midtread or midrise type.
- Observe the input–output characteristic of a uniform quantizer of the midtread type, which is so called because the origin lies in the middle of a tread of the staircase like graph.
- Observe the corresponding input–output characteristic of a uniform quantizer of the midrise type, in which the origin lies in the middle of a rising part of the stair case like graph.
- Both the midtread and midrise types of uniform quantizers are symmetric about the origin.

Pulse-Code Modulation

- PCM is a discrete-time, discrete-amplitude waveform-coding process, by means of which an analog signal is directly represented by a sequence of coded pulses.
- The Transmitter consists of two components: a pulse-amplitude modulator followed by an analog-to-digital (A/D) converter → a quantizer followed by an encoder.
- The Receiver performs the inverse of these two operations: digital-to-analog (D/A) conversion → an decoder followed by a dequantizer. followed by pulse-amplitude demodulator.
- The Communication Channel is responsible for transporting the encoded pulses from the Transmitter to the Receiver.



Block diagram of PCM system.

Sampling in the Transmitter

- The incoming message signal is sampled with a train of rectangular pulses short enough to closely approximate the instantaneous sampling process.
- For perfect reconstruction of the message signal at the receiver, the sampling rate must be greater than twice the highest frequency component *W* of the message signal in accordance with the Sampling Theorem.
- A low-pass anti-aliasing filter is used at the front end of the pulse-amplitude modulator to exclude frequencies greater than *W* before sampling and which are of negligible practical importance.
- Sampling permits the reduction of the continuously varying message signal to a limited number of discrete values per second.

Quantization in the Transmitter

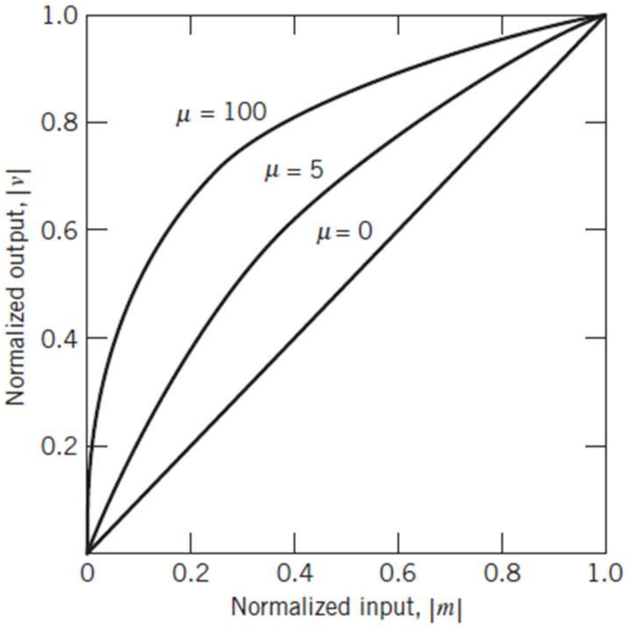
- The PAM representation of the message signal is then quantized in the analog-to-digital converter, to get a new representation of the signal that is discrete in both time and amplitude.
- The quantization process may follow a uniform law.
- In telephonic communication, however, it is preferable to use a variable separation between the representation levels for efficient utilization of the communication channel.
- Consider, for example, the quantization of voice signals.
- The range of voltages covered by voice signals, from the peaks of loud talk to the weak passages of weak talk, is on the order of 1000 to 1.

- By using a Non-uniform quantizer the step size increases as the separation from the origin of the input–output amplitude characteristic of the quantizer is increased.
- So the large end-steps of the quantizer can take care of possible excursions of the voice signal into the large amplitude ranges that occur rarely.
- So, the weak passages are favored at the expense of the loud passages.
- In this way, a nearly uniform percentage precision is achieved throughout the greater part of the amplitude range of the input signal.
- The end result is that fewer steps are needed than would be the case if a uniform quantizer were used.
- Hence the improvement in channel utilization.

- Assume Memoryless Quantization, the use of a non-uniform Quantizer is equivalent to passing the message signal through a Compressor and then applying the compressed signal to a Uniform quantizer.
- A particular form of Compression law that is used in practice is the so-called μ-*law*, which is defined by,

$$|v| = \frac{\ln(1 + \mu|m|)}{\ln(1 + \mu)} \qquad \text{................(25)}$$

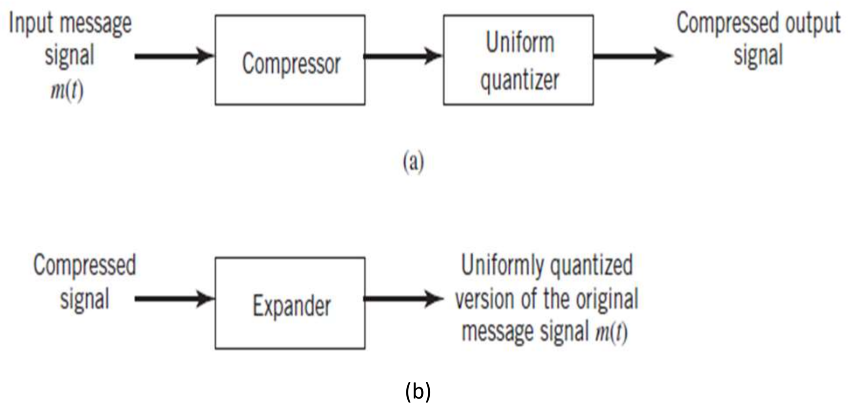
- *m* and *v* are the input and output voltages of the Compressor, and μ is a positive constant.
- Assume that *m* and *v* are scaled so that they both lie inside the interval [–1, 1].



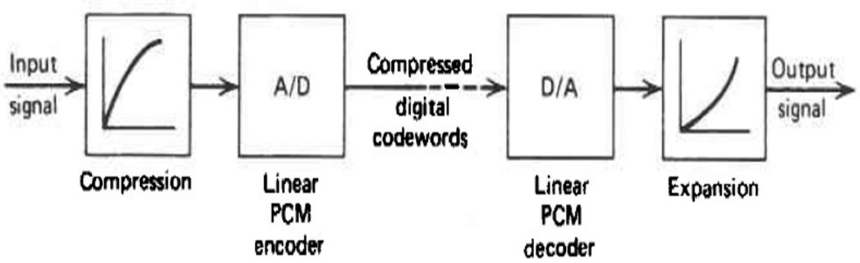
Compression law : µ-law

Companding

- To restore the signal samples to their correct relative level, use a device in the receiver with a characteristic opposite to the Compressor.
- Such a device is called an Expander.
- Ideally, the Compression and Expansion laws are exactly the inverse of each other.
- With this provision in place, except for the effect of quantization, the Expander output is equal to the Compressor input.
- The cascade combination of a Compressor and an Expander, is called a Compander.



(a) Nonuniform quantization of the message signal in the transmitter.
(b) Uniform quantization of the original message signal in the receiver.

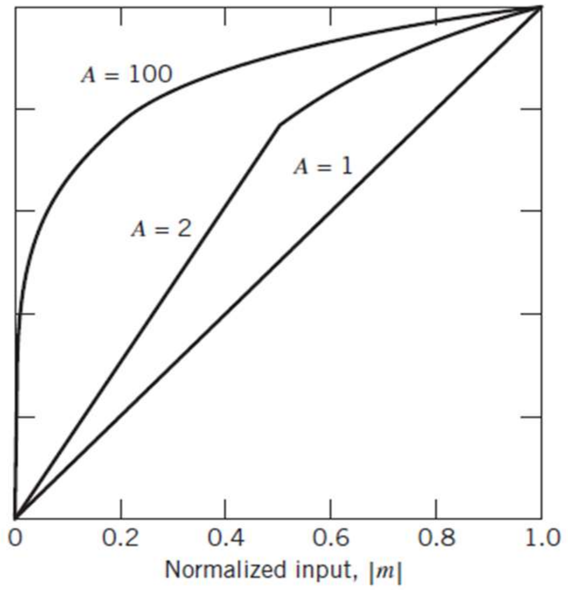


Companded PCM with Analog Compression and Expansion

- Another compression law that is used in practice is the *A-law,* defined by,

$$|v| = \begin{cases} \dfrac{A|m|}{1 + \ln A}, & 0 \le |m| \le \dfrac{1}{A} \\ \dfrac{1 + \ln(A|m|)}{1 + \ln A}, & \dfrac{1}{A} \le |m| \le 1 \end{cases} \quad \text{…………….(26)}$$

- Where *A* is another positive constant.
- Eqn(26) is plotted for varying *A*.
- The case of uniform quantization corresponds to *A* = 1.

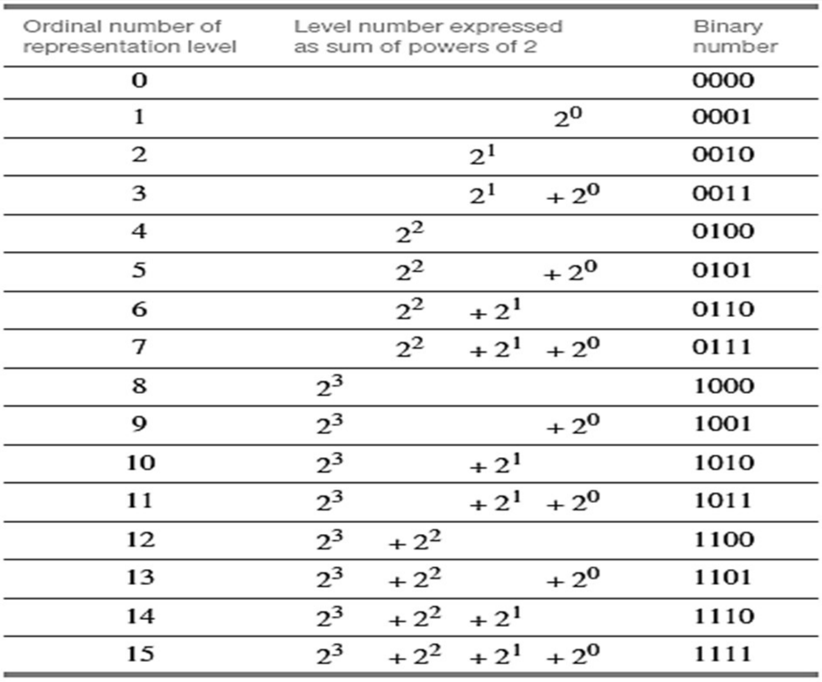


Compression law : A-law

- For both the μ-law and *A*-law, the dynamic range capability of the Compander improves with increasing μ and *A*, respectively.
- But the SNR for low-level signals increases at the expense of the SNR for high-level signals.
- A reasonable SNR for both low- and high-level signals needs a compromise in choosing the value of parameter μ for the μ-law and parameter *A* for the *A*-law.
- The typical values used in practice are μ = 255 for the μ-law and *A* = 87.6 for the *A*-law.

Encoding in the Transmitter

- To make Sampling and Quantizing of the transmitted signal more robust to Noise, Interference, and other Channel Impairments, an Encoding process is needed to translate the discrete Quantized Samples to Binary code.
- the Binary code is the preferred choice for encoding for the following reason:
- The maximum advantage over the effects of Noise encountered in a Communication System is obtained by using a Binary code because a Binary symbol withstands a relatively high level of noise and it is easy to regenerate.



Ordinal number of representation level	Level number expressed as sum of powers of 2	Binary number
0		0000
1	2^0	0001
2	2^1	0010
3	$2^1 + 2^0$	0011
4	2^2	0100
5	$2^2 + 2^0$	0101
6	$2^2 + 2^1$	0110
7	$2^2 + 2^1 + 2^0$	0111
8	2^3	1000
9	$2^3 + 2^0$	1001
10	$2^3 + 2^1$	1010
11	$2^3 + 2^1 + 2^0$	1011
12	$2^3 + 2^2$	1100
13	$2^3 + 2^2 + 2^0$	1101
14	$2^3 + 2^2 + 2^1$	1110
15	$2^3 + 2^2 + 2^1 + 2^0$	1111

Binary number system for T = 4 bits/sample

ECT 305 ADC Mod- III Part- 2

DPCM
and
DELTA MODULATION

Outline

- *Delta Modulation*
- *Delta Sigma Modulation*
- Linear Prediction
- Differential Pulse Code Modulation
- Adaptive Differential Pulse Code Modulation

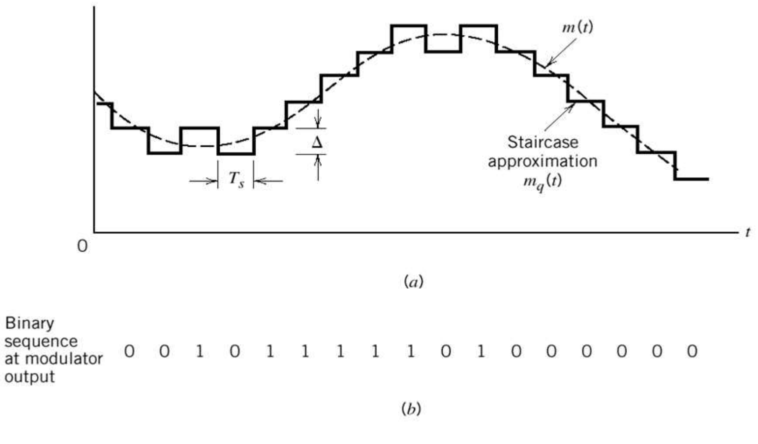
Delta Modulation

- **Definition:** Delta Modulation is a technique which provides a staircase approximation to an over-sampled version of the message signal (analog input).

- sampling is at a rate higher than the Nyquist rate – aims at increasing the correlation between adjacent samples; simplifies quantizing of the encoded signal

---

Illustration of the delta modulation process



Principle Operation

- message signal is over-sampled
- difference between the input and the approximation is quantized in two levels - +/-Δ
- these levels correspond to positive/negative differences
- provided signal does not change ***very rapidly*** the approximation remains within +/-Δ

Assumptions and model

We assume that:

- *m(t)* denotes the input message signal
- $m_q(t)$ denotes the staircase approximation
- $m[n] = m(nT_s), n = +/-1, +/-2$ ... denotes a sample of the signal *m(t)* at time $t=nT_s$, where T_S is the sampling period
- then

- we can express the basic principles of the delta modulation in a mathematical form as follow:

$$e[n] = m[n] - m_q[n-1] \qquad (3.52)$$

error signal

$$e_q = \Delta \operatorname{sgn}(e[n]) \qquad (3.53)$$

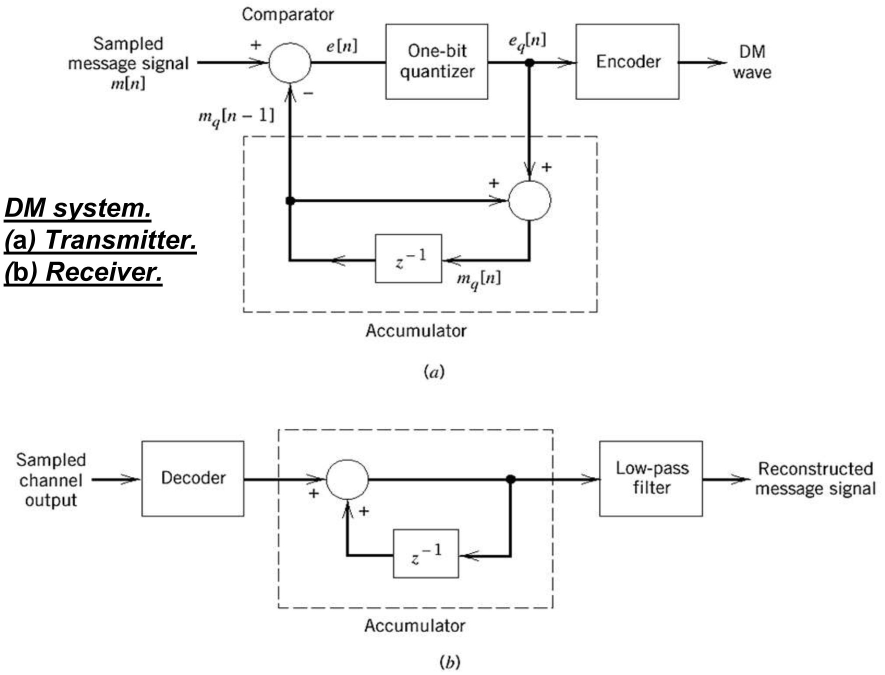
quantized error signal

$$m_q[n] = m_q[n-1] + e_q[n] \qquad (3.54)$$

quantized output

Pros and cons

- Main advantage – simplicity
- Sampled version of the message is applied to a modulator (comparator, quantizer, accumulator)
- delay in accumulator is “unit delay” = one sample period (z^{-1})



DM system.
(a) Transmitter.
(b) Receiver.

Transmitter Side

- comparator – computes difference between input signal and one interval delayed version of it
- quantizer – includes a hard-limiter with an input-output relation a scaled version of the signum function
- accumulator – produces the approximation *$m_q[n]$* (final result) at each step by adding either +Δ or –Δ
- = tracking input samples by one step at a time

$$m_q[n] = \Delta \sum_{i=1}^{n} \operatorname{sgn}(e[i])$$

$$= \sum_{i=1}^{n} e_q[i]$$

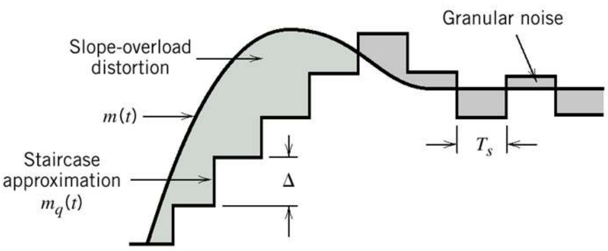
(3.55)

Receiver Side

- decoder – creates the sequence of positive or negative pulses
- accumulator – creates the staircase approximation $m_q[n]$ similar to tx side
- out-of-band noise is cut off by low-pass filter (bandwidth equal to original message bandwidth)

Noise in Delta Modulation Systems

- slope overhead distortion
- granular noise



Slope Overhead Distortion

- The quantized message signal can be represented as:

$$m_q[n] = m[n] + q[n] \qquad (3.56)$$

- where the input to the quantizer can be represented as:

$$e[n] = m[n] - m[n-1] - q[n-1]$$

(3.57)

Sample of m(T) at time nT

Quantizer input at time (n-1)T

So, (except for the quantization error) the **quantizer input** is the first backward difference (**derivative) of the input signal =** inverse of the digital integration process

Discussion

- Consider the max slope of the input signal *m(t)*
- To increase the samples *{m$_q$[n]}* as fast as the input signal in its *max slope region* the following condition should be fulfilled:

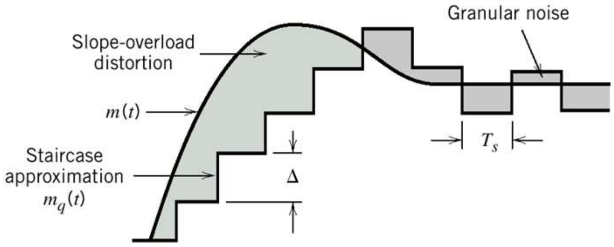
$$\frac{\Delta}{T_s} \geq \max\left|\frac{dm(t)}{dt}\right|$$

(3.58)

otherwise the step-size Δ is too small

Granular Noise

- In contrast to slope overhead
- Occurs when step size is too large
- Usually relatively flat segment of the signal
- Analogous to quantization noise in PCM systems



Conclusion:

- 1. Large step-size is necessary to accommodate a wide dynamic range
- 2. Small step-size is required for accuracy with low-level signals
- = compromise between slope overhead and granular noise
- = adaptive delta modulation, where the step size is made to vary with the input signal

Outline

- Delta Modulation
  - Delta Sigma Modulation
- Linear Prediction
- *Differential Pulse Code Modulation*
- Adaptive Differential Pulse Code Modulation

Differential PCM

- Sampling at higher then Nyquist rate creates correlation between samples (good and bad)
- Difference between samples has small variance – ***smaller than the variance of the signal itself***
- Encoded signal contains ***redundant information***
- Can be used to a positive end – remove redundancy before encoding to get a more efficient signal to be transmitted

How it works – the background

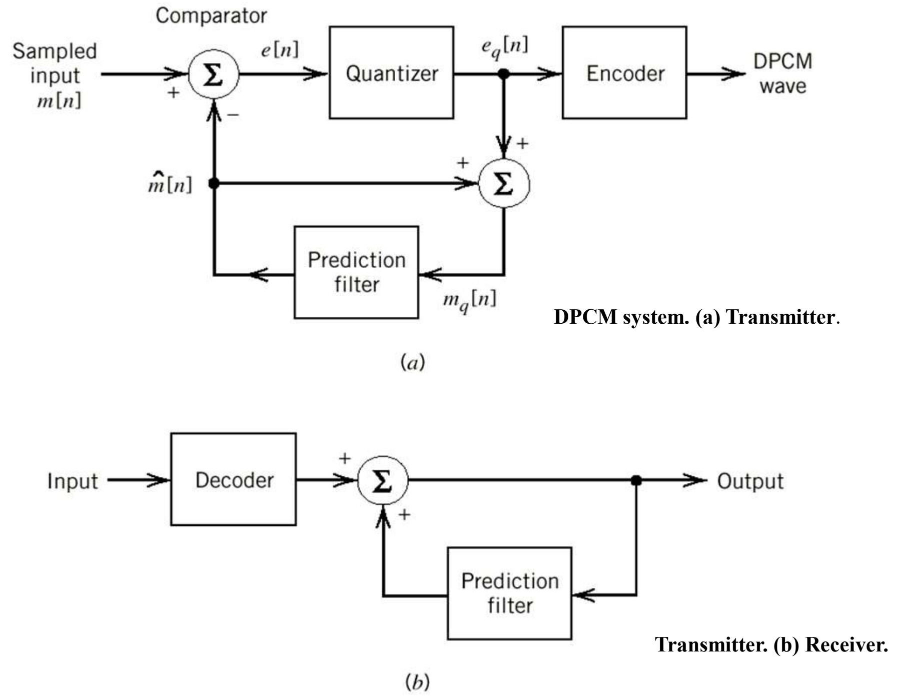
- We know the signal up to a certain time
- Use prediction to estimate future values
- Signal sampled at $f_s = 1/T_s$; sampled sequence – $\{m[n]\}$, where samples are T_s seconds apart
- Input signal to the quantizer – difference between the unquantized input signal $m(t)$ and its prediction:

$$e[n] = m[n] - \hat{m}[n]$$

prediction of the input sample

(3.74)

- Predicted value – achieved by linear prediction filter whose input is the quantized version of the input sample m[n].
- The difference e[n] is the prediction error (what we expect and what actually happens)
- By encoding the quantizer output we actually create a variation of PCM called differential PCM (DPCM).



DPCM system. (a) Transmitter.

Transmitter. (b) Receiver.

Details:

- Block scheme is very similar to DM
- quantizer input

$$e[n] = m[n] - \hat{m}[n] \qquad (3.74)$$

- quantizer output may be expressed as:

$$e_q[n] = e[n] + q[n] \qquad (3.75)$$

- prediction filter output may be expressed as:

$$m_q[n] = \hat{m}[n] + e_q[n] \qquad (3.76)$$

If we substitute 3.75 into 3.76 we get:

$$m_q[n] = \underbrace{\hat{m}[n] + e[n]}_{\text{sum is equal to input sample}} + q[n] \qquad (3.77)$$

Quantized input of the prediction filter -

$$m_q[n] = m[n] + q[n] \qquad (3.78)$$

- $m_q[n]$ is the quantized version of the input sample *m[n]*
- so, irrespective of the properties of the prediction filter the quantized sample $m_q[n]$ at the prediction filter input differs from the original sample *m[n]* with the quantization error *q[n]*.
- If the prediction filter is good, the ***variance of the prediction error e[n]*** will be smaller than the ***variance of m[n]***
- This means that if we make a very good prediction filter (adjust the number of levels) it will be possible **to produce a quantization error with a smaller variance than if the input sample *m[n]* is quantized directly** as in standard PCM

Receiver side

- decoder – constructs the quantized error signal
- quantized version of the input is recovered by using the same prediction filter as at the tx
- if there is no channel noise – encoded input to the decoder is identical to the transmitter output
- then the receiver output will be equal to $m_q[n]$ (differs from $m[n]$ by $q[n]$ caused by quantizing the prediction error $e[n]$)

- DPCM and DM
  - DPCM includes DM as a special case
  - Similarities
    - subject to slope-overhead and quantization error
  - Differences
    - DM uses a 1-bit quantizer
    - DM uses a single delay element (zero prediction order)
- DPCM and PCM
  - both DM and DPCM use feedback while PCM does not
  - all subject to quantization error

Processing Gain

- Output signal-to-noise ratio (SNR_O)

$$(SNR)_O = \frac{\sigma_M^2}{\sigma_Q^2} \qquad (3.79)$$

- σ_M^2 – variance of m[n]
- σ_Q^2 – variance of quantization error q[n]
- rewrite using variance of the prediction error σ_E^2

$$(SNR)_O = \left(\frac{\sigma_M^2}{\sigma_E^2}\right)\left(\frac{\sigma_E^2}{\sigma_Q^2}\right) = G_p (SNR)_Q$$

(3.80)

$$(SNR)_Q = \frac{\sigma_E^2}{\sigma_Q^2} \qquad (3.81)$$

signal-to-quantization noise ratio

$$G_p = \frac{\sigma_M^2}{\sigma_E^2} \qquad (3.82)$$

processing gain

- The processing gain G_p when greater than unity represents the signal-to-noise ratio that is due to the differential quantization scheme.
- For a given input message signal σ_M is fixed, so the smaller the σ_E the greater the G_p.
- This is the design objective of the prediction filter
- For voice signals – optimal main advantage of DPCM over PCM is b/n 4-11 dB
- Advantage expressed in terms of bit rate (bits)
  - 1 bit =6 dB of quantization noise
  - So for fixed SNR, sampling rate 8 kHz – DCPM provides saving of 8-16 kb/s (1 -2 bits per sample) PCM

Signal Space Analysis:
Module – IV Part-1

ECT 305 Analog and Digital Communication

Introduction

The steady transition from analog communications to digital communications has been empowered by several factors:

1. Ever-increasing advancement of digital silicon chips, digital signal processing, and computers, which, in turn, has prompted further enhancement in digital silicon chips, thereby repeating the cycle of improvement.

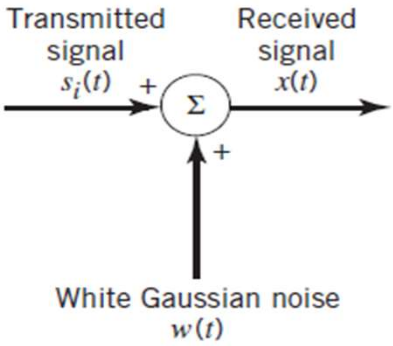
2. Improved reliability, which is afforded by digital communications to a much greater extent than is possible with analog communications.

3. Broadened range of multiplexing of users, which is enabled by the use of digital modulation techniques.

4. Communication networks, for which, in one form or another, the use of digital communications is the preferred choice.

- As an example, consider the remote connection of two digital computers, with one computer acting as the information source by calculating digital outputs based on observations and inputs fed into it.
- The other computer acts as the recipient of the information.
- The source output consists of a sequence of 1s and 0s, with each binary symbol being emitted every T_b seconds.
- The transmitting part of the digital communication system takes the 1s and 0s emitted by the source computer and encodes them into distinct signals denoted by $s_1(t)$ and $s_2(t)$, respectively, which are suitable for transmission over the analog channel.
- Both $s_1(t)$ and $s_2(t)$ are real-valued energy signals, as shown by,

$$E_i = \int_0^{T_b} s_i^2(t)\,dt, \qquad i = 1, 2 \qquad \text{.............. (1)}$$



AWGN model of a channel.

- With the analog channel represented by an AWGN model, the received signal is defined by,

$$\therefore \quad x(t) = s_i(t) + w(t), \quad \begin{cases} 0 \le t \le T_b \\ i = 1, 2 \end{cases} \quad \text{..............} \ (2)$$

- where w(t) is the channel noise.
- The receiver has the task of observing the received signal x(t) for a duration of T_b seconds and then making an estimate of the transmitted signal $s_i(t)$ or equivalently the ith symbol, i = 1, 2.
- But, due to channel noise, the receiver will make occasional errors.
- The requirement, therefore, is to design the receiver so as to minimize the average Probability of Symbol Error.

To minimize the average probability of symbol error between the receiver output and the symbol emitted by the source, make the digital communication system as reliable as possible.

To achieve this important design objective in a generic setting that involves an M-ary alphabet whose symbols are denoted by m_1, m_1, …,m_M , there are two basic issues:

1. To optimize the design of the Receiver so as to minimize the average probability of symbol error.

2. To choose the set of signals $s_1(t)$, $s_2(t)$,…, $s_M(t)$ for representing the symbols m_1, m_1, …,m_M, respectively, since this choice affects the average probability of symbol error.

Geometric Representation of Signals

- The essence of geometric representation of signals is to represent any set of M energy signals $\{s_i(t)\}$ as linear combinations of N orthonormal basis functions, where $N \leq M$.
- Given a set of real-valued energy signals, $s_1(t)$, $s_2(t),\ldots, s_M(t)$ each of duration T seconds,

$$\therefore \quad s_i(t) = \sum_{j=1}^{N} s_{ij}\phi_j(t), \qquad \begin{cases} 0 \leq t \leq T \\ i = 1, 2, \ldots, M \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots (3)$$

- where the coefficients of the expansion are defined by,

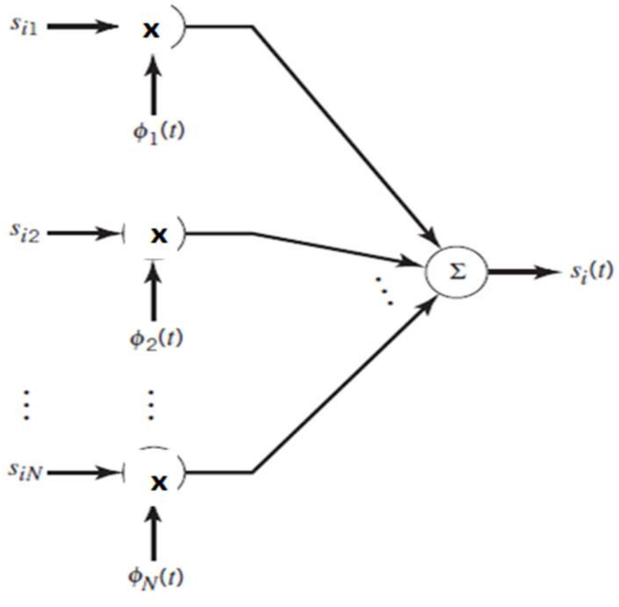
$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,\mathrm{d}t, \qquad \begin{cases} i = 1, 2, \ldots, M \\ j = 1, 2, \ldots, N \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots (4)$$

- The real-valued basis functions $\phi_1(t)$, ϕ_2 (t), …, $\phi_N(t)$ form an orthonormal set,

$$\therefore \quad \int_0^T \phi_i(t)\phi_j(t)\,\mathrm{d}t = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \qquad \ldots\ldots\ldots\ldots\ldots (5)$$

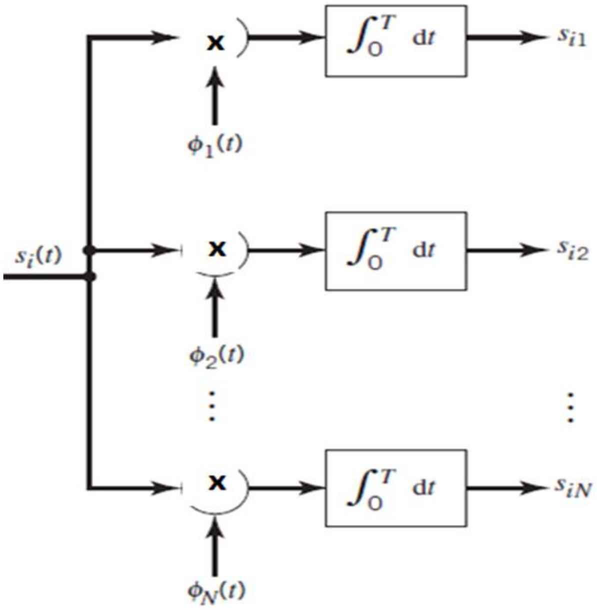
- where δ_{ij} is the Kronecker delta.
- The first condition of (5) states that each basis function is normalized to have unit energy.
- The second condition states that the basis functions $\phi_1(t)$, ϕ_2 (t), …, $\phi_N(t)$ are orthogonal w.r.t each other over the interval $0 \leq t \leq T$.

- For prescribed i, the set of coefficients $\{s_{ij}\}_{j=1}^{N}$ may be viewed as an N-dimensional signal vector, denoted by vector $\mathbf{s}_i$.
- The important point to note here is that the vector $\mathbf{s}_i$ bears a one-to-one relationship with the transmitted signal $s_i(t)$.
- Given the N elements of the vector $\mathbf{s}_i$ operating as input, use the scheme shown next to generate the signal $s_i(t)$, which follows directly from eqn(3).
- This figure consists of a bank of N multipliers with each multiplier having its own basis function followed by a summer.
- This scheme is considered as a Synthesizer.



(a) Synthesizer for generating the signal $s_i(t)$.

- Conversely, given signals $s_i(t)$, i = 1, 2, …, M, operating as input, use the scheme shown below to calculate the coefficients s_{i1}, s_{i2}, …, s_{iN} which follows directly from (4).
- This scheme consists of a bank of N product integrators or correlators with a common input, and with each one of them supplied with its own basis function.
- The scheme is considered as an Analyzer.



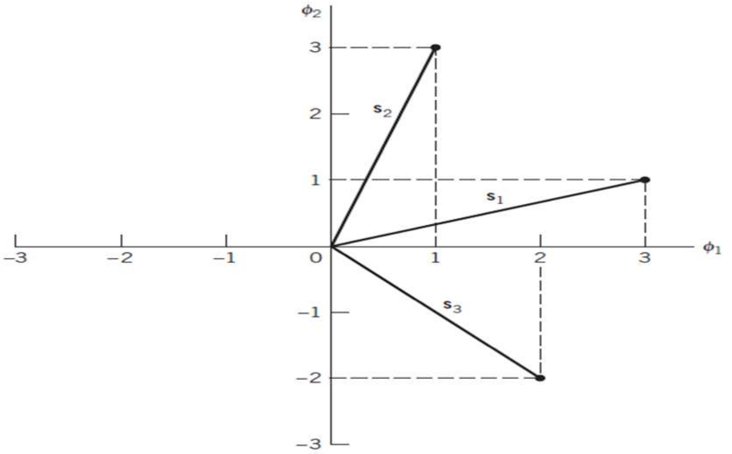
Analyzer for reconstructing the signal vector $\{s_i\}$.

- Each signal in the set $\{s_i(t)\}$ is completely determined by the signal vector,

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix}, \qquad i = 1, 2, \ldots, M \qquad \text{.............. (6)}$$

- The notion of two- and three dimensional Euclidean spaces can be extended to an N-dimensional Euclidean space, with the set of signal vectors $\{\mathbf{s}_i | i = 1, 2, \ldots, M\}$ as defining a corresponding set of M points in an N-dimensional Euclidean space, with N mutually perpendicular axes labelled $\phi_1(t)$, ϕ_2 (t), …, $\phi_N(t)$.
- This N-dimensional Euclidean space is called the Signal Space.

- This form of representation is shown below for the case of a two-dimensional signal space with three signals; that is, N = 2 and M = 3.



Geometric representation of signals for the case when N = 2 and M = 3.

- In an N-dimensional Euclidean space, define lengths of vectors and angles between vectors.
- Denote the length (also called the absolute value or norm) of a signal vector $\mathbf{s}_i$ by the symbol $\|\mathbf{s_i}\|$.
- The squared length of any signal vector $\mathbf{s}_i$ is defined to be the inner product or dot product of $\mathbf{s}_i$ with itself, as given by,

$$\|\mathbf{s}_i\|^2 = \mathbf{s}_i^{\mathbf{T}}\mathbf{s}_i = \sum_{j=1}^{N} s_{ij}^2, \qquad i = 1, 2, \ldots, M \qquad \text{..............} (7)$$

- where s_{ij} is the jth element of $\mathbf{s}_i$ and the superscript T denotes matrix transposition.

- There is an interesting relationship between the energy content of a signal and its representation as a vector.
- By definition, the energy of a signal $s_i(t)$ of duration T seconds is,

$$E_i = \int_0^T s_i^2(t)\,\mathrm{d}t, \qquad i = 1, 2, \ldots, M \qquad \text{..............} (8)$$

- Therefore, substituting eqn(3) into eqn(8),

$$E_i = \int_0^T \left[\sum_{j=1}^{N} s_{ij}\phi_j(t)\right]\left[\sum_{k=1}^{N} s_{ik}\phi_k(t)\right]\mathrm{d}t$$

- Interchanging the order of summation and integration, and then rearranging terms,

$$E_i = \sum_{j=1}^{N}\sum_{k=1}^{N} s_{ij}s_{ik}\int_0^T \phi_j(t)\,\phi_k(t)\,\mathrm{d}t \qquad \text{..............} (9)$$

- Since, by definition, the $\phi_j(t)$ form an orthonormal set in accordance with the two conditions of eqn(5), eqn(9) reduces simply to,

$$E_i = \sum_{j=1}^{N} s_{ij}^2 = \|\mathbf{s}_i\|^2 \quad \text{.............. (10)}$$

- Thus, (7) and (10) show that the energy of an energy signal $s_i(t)$ is equal to the squared length of the corresponding signal vector $\mathbf{s}_i(t)$.
- In the case of a pair of signals $s_i(t)$ and $s_k(t)$ represented by the signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$, respectively,

$$\int_0^T s_i(t)s_k(t)\,\mathrm{d}t = \mathbf{s}_i^{\mathrm{T}}\mathbf{s}_k \quad \text{.............. (11)}$$

- The inner product of the energy signals $s_i(t)$ and $s_k(t)$ over the interval $[0,T]$ is equal to the inner product of their respective vector representations $\mathbf{s}_i$ and $\mathbf{s}_k$.
- Note that the inner product is invariant to the choice of basis functions $\{\phi_j(t)\}_{j=1}^{N}$, in that it only depends on the components of the signals $s_i(t)$ and $s_k(t)$ projected onto each of the basis functions.

- Yet another useful relation involving the vector representations of the energy signals *$s_i(t)$ and $s_k(t)$* is described by,

$$\|\mathbf{s}_i - \mathbf{s}_k\|^2 = \sum_{j=1}^{N} (s_{ij} - s_{kj})^2$$

$$= \int_0^T (s_i(t) - s_k(t))^2 \, dt \qquad \text{............. (12)}$$

- where $\|\mathbf{s_i} - \mathbf{s_k}\|^2$ is the Euclidean distance d_{ik} between the points represented by the signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$.

- To complete the geometric representation of energy signals, need to have a representation for the angle θ_{ik} subtended between two signal vectors $\mathbf{s}_i$ and $\mathbf{s}_k$.
- By definition, the cosine of the angle θ_{ik} is equal to the inner product of these two vectors divided by the product of their individual norms,

$$\cos(\theta_{ik}) = \frac{\mathbf{s}_i^{\mathrm{T}} \mathbf{s}_k}{\|\mathbf{s}_i\| \|\mathbf{s}_k\|} \qquad \text{............. (13)}$$

- The two vectors $\mathbf{s}_i$ and $\mathbf{s}_k$ are thus orthogonal or perpendicular to each other if their inner product $\mathbf{s}_i^{\mathrm{T}}\mathbf{s}_k$ is zero, in which case θ_{ik} = 90°.

Gram–Schmidt Orthogonalization Procedure

- The Gram–Schmidt Orthogonalization procedure uses a complete Orthonormal set of basis functions.
- To proceed with the formulation of this procedure, suppose there is a set of *M* energy signals denoted by $s_1(t)$, $s_2(t)$, …, $s_M(t)$.
- Starting with $s_1(t)$ chosen from this set, the first basis function is defined by,

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}} \qquad \text{………….. (17)}$$

- where E_1 is the energy of the signal $s_1(t)$.

$$s_1(t) = \sqrt{E_1}\phi_1(t) = s_{11}(t)\phi_1(t)$$

$$s_{11} = \sqrt{E_1}$$

- $\phi_1(t)$ has unit energy as required.
- Next, using the signal $s_2(t)$, define the coefficient s_{21} as,

$$s_{21} = \int_0^T s_2(t)\phi_1(t)\,\mathrm{d}t$$

- Introduce a new intermediate function,

$$g_2(t) = s_2(t) - s_{21}\phi_1(t) \qquad \text{………….. (18)}$$

- which is orthogonal to $\phi_1(t)$ over interval $0 \leq t \leq T$ by the definition of s_{21} and the fact that the basis function $\phi_1(t)$ has unit energy.

- Define the second basis function as,

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)\,dt}} \qquad \text{............. (19)}$$

- Substituting eqn(18) into (19) and simplifying,

$$\phi_2(t) = \frac{s_2(t) - s_{21}\phi_1(t)}{\sqrt{E_2 - s_{21}^2}} \qquad \text{............. (20)}$$

- where E_2 is the energy of the signal $s_2(t)$.
- From eqn (19), it is seen that, $\int_0^T \phi_2^2(t)\,dt = 1$
- in which case eqn (20) yields, $\int_0^T \phi_1(t)\phi_2(t)\,dt = 0$
- Therefore, $\phi_1(t)$ and $\phi_2(t)$ form an <u>orthonormal pair</u> as required.

- Continuing the procedure in this manner, in general, define,

$$g_i(t) = s_i(t) - \sum_{j=1}^{i-1} s_{ij}\phi_j(t) \qquad \text{............. (21)}$$

- where the coefficients s_{ij} are themselves defined by,

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt, \qquad j = 1, 2, \ldots, i-1$$

- For $i = 1$, the function $g_i(t)$ reduces to $s_i(t)$.
- Given the $g_i(t)$, define the <u>set of basis functions</u>,

$$\phi_j(t) = \frac{g_i(t)}{\sqrt{\int_0^T g_i^2(t)\,dt}}, \qquad j = 1, 2, \ldots, N \qquad \text{............. (22)}$$

- which form an <u>orthonormal</u> set.

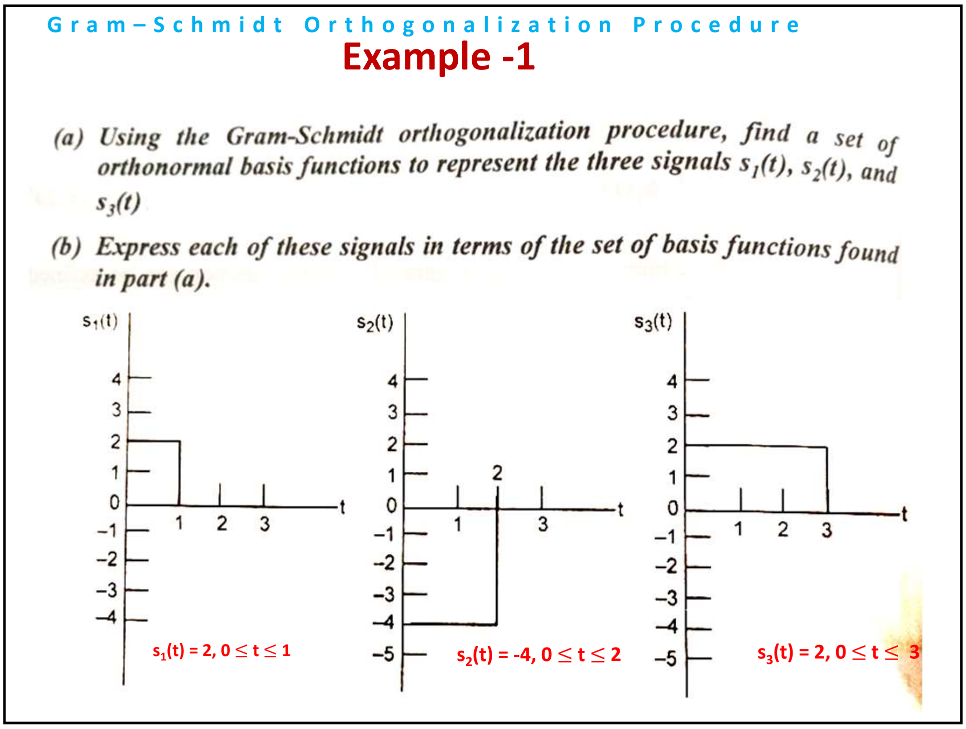
The dimension N is less than or equal to the number of given signals, M, depending on one of two possibilities:

1. The signals $s_1(t)$, $s_2(t)$, …, $s_M(t)$ form a linearly independent set, in which case $N = M$.
2. The signals $s_1(t)$, $s_2(t)$, …, $s_M(t)$ are not linearly independent, in which case $N < M$ and the intermediate function $g_i(t)$ is zero for $i > N$.

Points to Ponder

1. Unlike the Fourier series expansion of a periodic signal or the Sampled representation of a band-limited signal, the Gram–Schmidt Orthogonalization procedure is not restricted to be in terms of sinusoidal functions (as in the Fourier series) or sinc functions of time (as in the sampling process).

2. The expansion of the signal $s_i(t)$ in terms of a finite number of terms is not an approximation where only the first N terms are significant, instead it is an exact expression, where N and only N terms are significant.



Gram−Schmidt Orthogonalization Procedure

Example -1

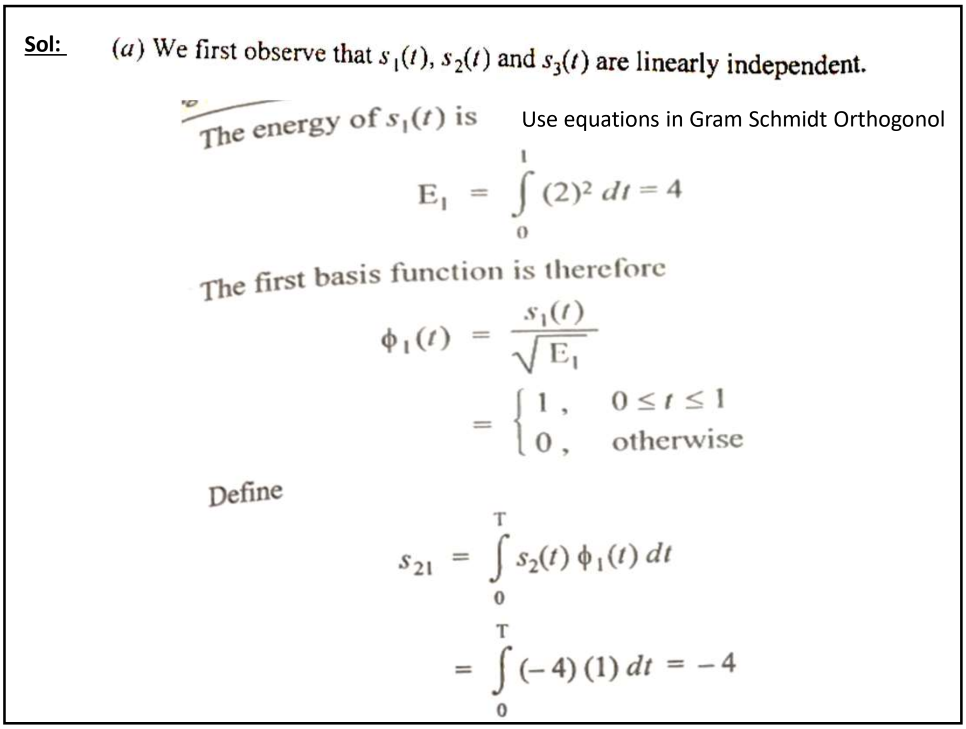
(a) Using the Gram-Schmidt orthogonalization procedure, find a set of orthonormal basis functions to represent the three signals $s_1(t)$, $s_2(t)$, and $s_3(t)$

(b) Express each of these signals in terms of the set of basis functions found in part (a).

$s_1(t) = 2, 0 \leq t \leq 1$

$s_2(t) = -4, 0 \leq t \leq 2$

$s_3(t) = 2, 0 \leq t \leq 3$



Sol: (*a*) We first observe that $s_1(t)$, $s_2(t)$ and $s_3(t)$ are linearly independent.

The energy of $s_1(t)$ is

Use equations in Gram Schmidt Orthogonol

$$E_1 = \int_0^1 (2)^2 \, dt = 4$$

The first basis function is therefore

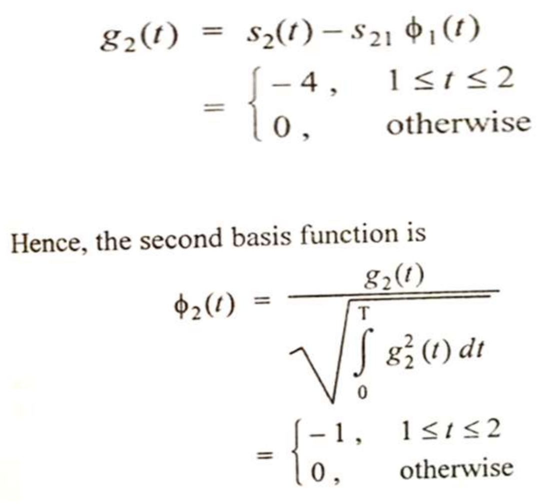
$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$$

$$= \begin{cases} 1, & 0 \leq t \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Define

$$s_{21} = \int_0^T s_2(t)\, \phi_1(t)\, dt$$

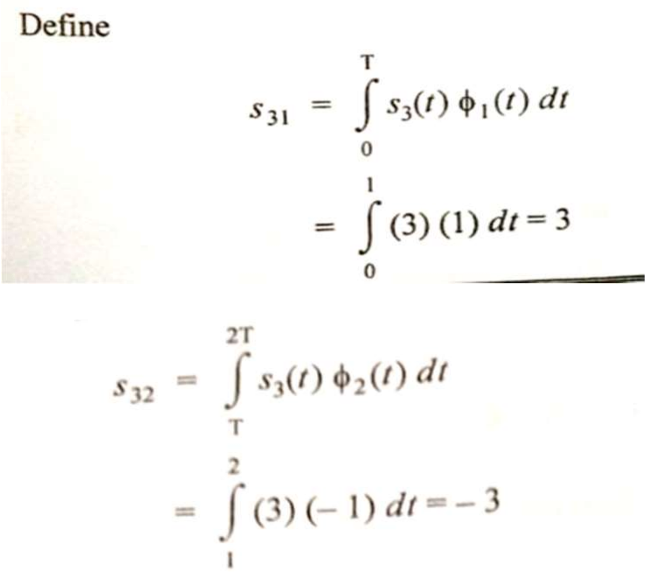
$$= \int_0^T (-4)(1)\, dt = -4$$



$$g_2(t) = s_2(t) - s_{21}\,\phi_1(t)$$
$$= \begin{cases} -4, & 1 \le t \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Hence, the second basis function is

$$\phi_2(t) = \frac{g_2(t)}{\sqrt{\int_0^T g_2^2(t)\,dt}}$$
$$= \begin{cases} -1, & 1 \le t \le 2 \\ 0, & \text{otherwise} \end{cases}$$



Define

$$s_{31} = \int_0^T s_3(t)\,\phi_1(t)\,dt$$
$$= \int_0^1 (3)\,(1)\,dt = 3$$

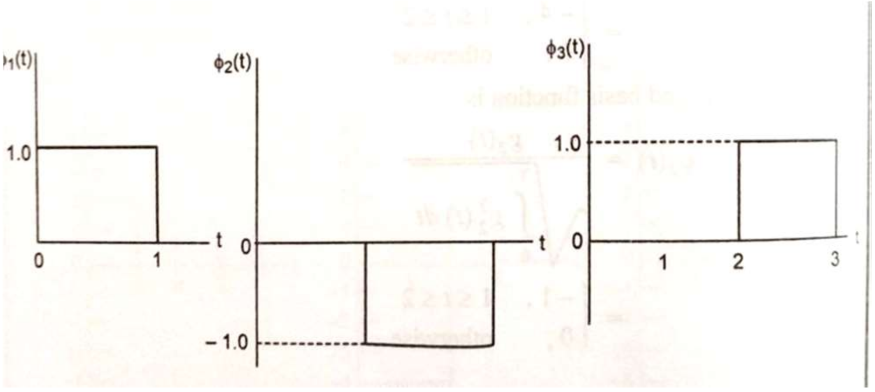
$$s_{32} = \int_T^{2T} s_3(t)\,\phi_2(t)\,dt$$
$$= \int_1^2 (3)\,(-1)\,dt = -3$$

$$
\begin{aligned}
g_3(t) &= s_3(t) - s_{31}\,\phi_1(t) - s_{32}\,\phi_2(t) \\
&= \begin{cases} 3\,, & 2 \le t \le 3 \\ 0\,, & \text{otherwise} \end{cases}
\end{aligned}
$$

Hence, the third basis function is

$$
\begin{aligned}
\phi_3(t) &= \frac{g_3(t)}{\sqrt{\int\limits_0^T g_3^2(t)\,dt}} \\
&= \begin{cases} 1\,, & 2 \le t \le 3 \\ 0\,, & \text{otherwise} \end{cases}
\end{aligned}
$$

The three basis functions are as follows (graphically)



(*b*)

$$s_1(t) = 2\,\phi_1(t)$$

$$s_2(t) = -4\,\phi_1(t) + 4\,\phi_2(t)$$

$$s_3(t) = 3\,\phi_1(t) - 3\,\phi_2(t) + 3\,\phi_3(t)$$

Baseband transmission through AWGN Channel

Module-IV Part-2

ECT305 Analog and Digital Communication

Introduction

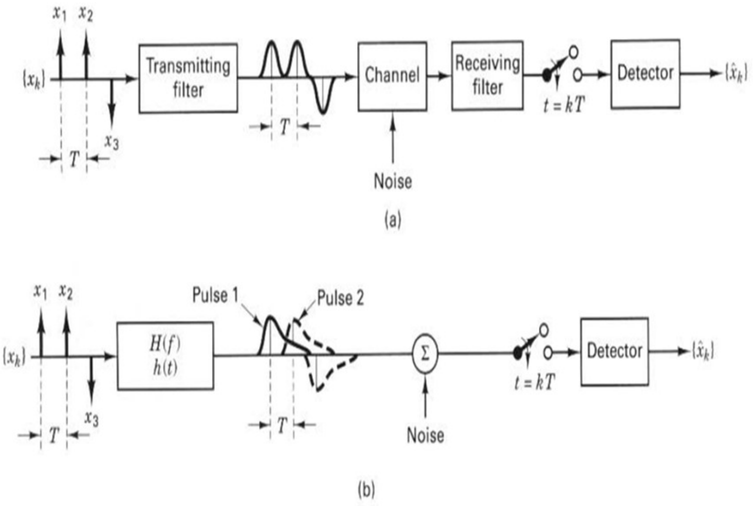
- In baseband signaling, the received waveforms are ideally in a pulse-like form.
- Alas! The arriving baseband pulses are not in the form of ideal pulse shape, each one occupying its own symbol interval.
- The filtering at the Transmitter and the Channel, cause the received pulse sequence to "suffer" from Inter Symbol Interference (ISI) and thus appear as an amorphous "smeared" signal, not quite ready for sampling and detection.
- The goal of the demodulator is to recover a baseband pulse with the best possible signal-to-noise ratio (SNR), free of any ISI.
- Equalization is a technique use to accomplish this goal, which has powerful signal-processing techniques, to make it possible to compensate for channel-induced interference.

- The task of detector is to retrieve the bit stream from the corrupted received waveform, affected by signal distortions, channel impairments and noise.
- There are two causes for error-performance degradation.
- The first is the effect of Filtering at the Transmitter, Channel and Receiver.
- A non-ideal system transfer function causes symbol “Smearing” or Inter Symbol Interference (ISI).
- Another more significant cause for error-performance degradation is Electrical Noise and interference produced by additive thermal noise in amplifiers and circuits, atmospheric noise, switching transients, intermodulation noise as well as interfering signals from other sources.

- Noise amplitude of Thermal noise is distributed according to a normal or Gaussian distribution.
- The PSD of white noise is $G_n(f)=N_0/2$ and is flat for all frequencies of interest (low frequencies up to a frequency of 10^{12}Hz).
- Due to this constant PSD, it is referred to as White Noise.
- Hence the “deadly” AWGN is used to model the noise in the detection process and in design of receivers.
- A channel that infects the information bearing signal in this manner is designated as AWGN channel.

INTERSYMBOL INTERFERENCE

- There are various filters in the transmitter, receiver and in the channel.
- At the transmitter, the information symbols, characterized as impulse or voltage levels modulate pulses that are then filtered to comply with some band width constraint.
- For baseband signals, the wired channel has distributed reactance that distorts the pulses.
- Some band pass systems, such as wireless systems are characterized by fading channels that behave like undesirable filters manifesting signal distortion.
- When the receiving filter is configured to compensate for the distortion caused by both the transmitter and channel, it is referred to as an equalizing filter or receiving/equalizing filter.
- A convenient model for the system is to lump all the filtering effects into one overall equivalent system transfer function.



Intersymbol Interference in the detection process

(a) Typical baseband digital system **(b) Equivalent model**

- $H_t(f)$ characterizes the transmitting filter, $H_c(f)$, the filtering within the channel and $H_r(f)$, the receiving/equalizing filter.

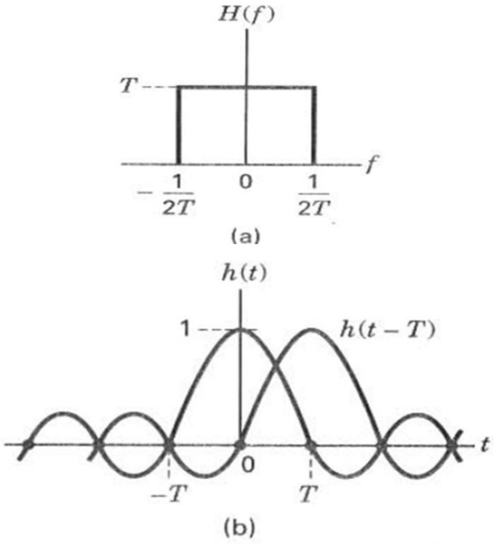
$$H(f) = H_t(f)\, H_c(f)\, H_r(f) \quad \text{…………(13)}$$

- The characteristic H(f) then represents the composite system transfer function due to all filtering at various locations throughout the transmitter/receiver chain.
- For a PCM waveform, the detector makes a symbol decision by comparing a sample of the received pulse to a threshold.
- The detector decides that a binary one was sent if the received pulse is positive and that a binary zero was sent, if the received pulse is negative.

- Due to the effects of system filtering, the received pulses can overlap one another as shown.
- The tail of a pulse can “smear” into adjacent symbol intervals, thereby interfering with the detection process and degrading the error performance.
- Such interference is termed as Inter Symbol Interference (ISI).
- Even in the absence of noise, the effects of filtering and channel induced distortion lead to ISI.
- Sometimes $H_c(f)$ is specified, and the problem remains to determine $H_t(f)$and $H_r(f)$, such that ISI is minimized at the output of $H_r(f)$.

Nyquist Criteria for zero ISI

- Nyquist investigated the problem of specifying a received pulse shape so that no ISI occurs at the detector.
- He showed that the theoretical maximum system bandwidth needed in order to detect R_s symbols/s, without ISI is $R_s/2$ Hz.
- This occurs when the System Transfer Function H(f) is made Rectangular, as shown.
- For baseband systems, when H(f) is such a filter with single-sided bandwidth 1/2T (the ideal Nyquist filter), its impulse response, the inverse Fourier transform of H(f) is of the form h(t) = sinc (t/T), as shown.
- This sinc (t/T) shaped pulse is called the ideal Nyquist pulse.



Nyquist channels for zero ISI.
(a) Rectangular system transfer function.
(b) Received pulse shape h(t) = sinc (t/T)

- The multiple lobes comprise a main lobe and side lobes called pre- and post-main lobe tails that are infinitely long.
- Nyquist established that if each pulse of a received sequence is of the form sinc (t/T), the pulses can be detected without ISI.
- As shown in figure, ISI can be avoided.
- There are two successive pulses, h(t) and h(t-T).
- Even though h (t) has long tails, the figure shows a tail passing through zero amplitude at the instant (t=T) when h (t-T) is to be sampled, and likewise all tails pass through zero amplitude when any other pulse of the sequence h(t-kT), k=$\mp$1, $\mp$2, … is to be sampled.

- Therefore, assuming that the sampling timing is perfect, there will be no ISI degradation introduced.
- For baseband systems, the bandwidth required to detect 1/T such pulses (symbols) per second is equal to 1/2T.
- So, a system with bandwidth W= 1/2T = $R_s/2$ Hz, can support a maximum transmission rate of 2W = 1/T= R_s symbols/s (Nyquist bandwidth constraint) without ISI.

Ideal solution

- Thus, for ideal Nyquist filtering and zero ISI, the maximum possible symbol transmission rate per Hz, called the Symbol Rate Packing, is 2 symbols/Hz.
- It is to be noted that from the rectangular shaped transfer function of the ideal Nyquist filter and the infinite length of its corresponding pulse, that such ideal filters are not exactly realizable , they can only be approximately realized in practice.

- Nyquist filter and Nyquist pulse are used to describe the general class of filtering and pulse shaping that satisfy zero ISI at the sampling points.
- A Nyquist filter is one whose frequency transfer function can be represented by a sinc (t/T) function multiplied by another time function.
- Hence, there are countless number of Nyquist filters and corresponding pulse shaped.
- Amongst the class of Nyquist filters, the most popular ones are Raised Cosine filter and the Root Raised Cosine filter.

- A fundamental parameter for communication systems is Bandwidth Efficiency, R/W, whose units are bits/s/Hz.
- R/W, represents a measure of Data Throughput per Hz of Bandwidth and thus measures how efficiently any signalling technique utilizes the bandwidth resource.
- Since the Nyquist bandwidth constraint dictates that the theoretical maximum symbol rate packing without ISI is 2 symbols/s/Hz, what it says about the maximum number of bits/s/Hz? →
- It says nothing about bits directly.
- The constraints deals only with pulses or symbols, and the ability to detect their amplitude values without distortion from other pulses.
- To find R/W for any signalling scheme, one needs to know how many bits each symbol represents.

- Consider an M-ary PAM signalling set.
- Each symbol (comprising k bits) is represented by one of the M-pulse amplitudes.
- For k=6 bits per symbol, the symbol set size is $M=2^k = 64$ amplitudes.
- Thus with 64-ary PAM, the theoretical maximum bandwidth efficiency that is possible without ISI is 12 bits/s/Hz.

Pulse Shaping to reduce ISI

- The more compact the signalling spectrum, the higher is the allowable data rate, or the greater is the number of users that can simultaneously be served.
- This has important implications to communication service providers, since greater utilization of the available bandwidth translates into greater revenue.
- For most communication systems, except spread-spectrum systems, the goal is to reduce the required system bandwidth as much as possible.
- Nyquist has provided the basic limitation to such bandwidth reduction.
- What happened if a system is forced to operate at smaller bandwidths than the constraint dictates? →
- The pulses would become spread in time which would degrade the system's error performance due to increased ISI.

- A judicious goal is to compress the bandwidth of the data impulses to some reasonably small bandwidth greater than the Nyquist minimum.
- This is done by Pulse Shaping with a Nyquist filter.
- If the band edge of the filter is steep, approaching the rectangle, then the signalling spectrum can be made most compact.
- But, such a filter has an impulse response duration approaching infinity.
- Each pulse extends into every pulse in the entire sequence.
- Long-time responses exhibit large amplitude tails nearest the main lobe of each pulse.
- Such tails are undesirable as they contribute zero ISI only when the sampling is performed at exactly the correct sampling time.
- When tails are large, small timing errors will result in ISI.
- Therefore, although a compact spectrum provides optimum bandwidth utilization, it is very susceptible to ISI degradation induced by timing errors.

The Raised Cosine Filter

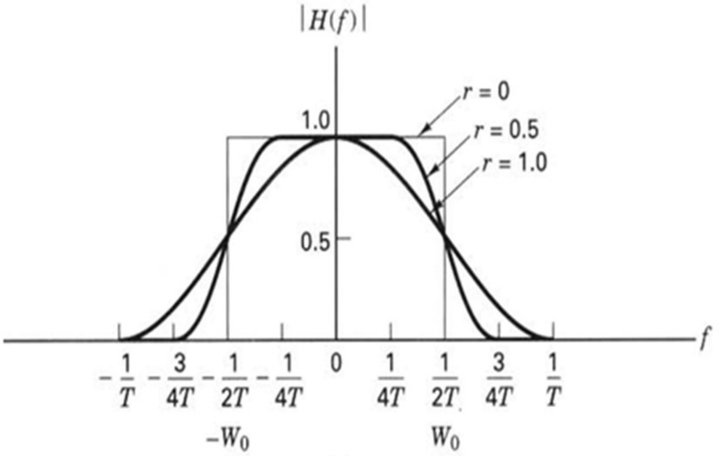
- The receiving Filter is often referred to as an Equalizing Filter as it is configured to compensate for the distortion caused by both the transmitter and the channel.
- The configuration of this Filter is chosen so as to optimize the composite system frequency transfer function H(f).
- One frequently used H(f) transfer function belonging to the Nyquist class (zero ISI at the sampling times) is called the Raised Cosine filter.

- It can be expressed as,

$$H(f) = \left\{ \begin{array}{ll} 1 & \text{for } |f| < 2W_0 - W \\ \cos^2\left(\dfrac{\pi}{4}\dfrac{|f| + W - 2W_0}{W - W_0}\right) & \text{for } 2W_0 - W < |f| < W \\ 0 & \text{for } |f| > W \end{array} \right\} \qquad \text{…………(14)}$$

- Where W is the absolute bandwidth and $W_0 = 1/2T$ represent the minimum Nyquist bandwidth for the rectangular spectrum and the -6 dB bandwidth (or half-amplitude point) for the Raised Cosine spectrum.

- The difference (W-W_0) is termed the Excess Bandwidth, which means additional bandwidth beyond the Nyquist minimum (i.e., for the rectangular spectrum, W is equal to W_0).
- The Roll-Off Factor is defined as $r=(W- W_0)/ W_0$, where $0 \leq r \leq 1$.
- It represents the Excess Bandwidth divided by the filter -6 dB Bandwidth (i.e., the fractional excess bandwidth).
- For a given W_0, the Roll-Off r specifies the required excess bandwidth as a fraction of W_0 and characterizes the steepness of the filter roll-off.

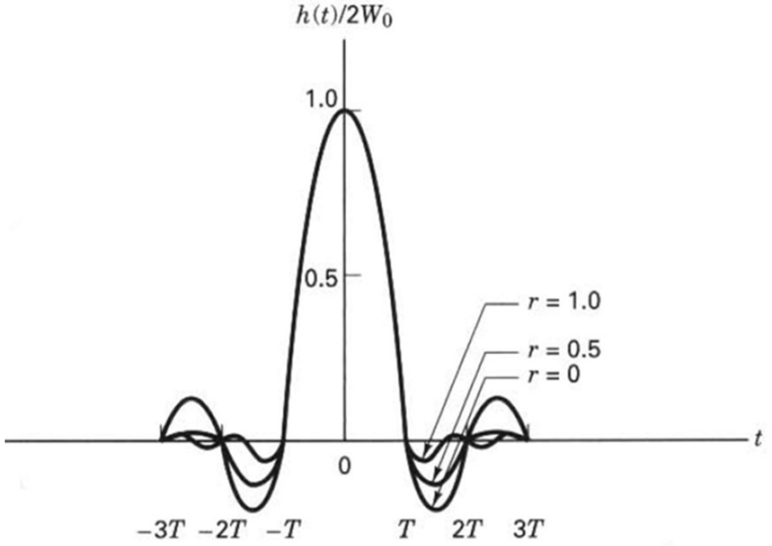


Raised-Cosine Filter characteristics - System Transfer Function

- The raised cosine characteristic is shown for a roll-off values of *r = 0*, *r = 0.5*, and *r = 1*.
- The *r = 0* roll-off is the Nyquist minimum bandwidth case.
- Note that when *r = 1*, the required excess bandwidth is 100%, and the tails are quite small.
- A system with such an overall spectral characteristic can provide a symbol rate of R_s symbols/s using a bandwidth of R_s Hz (twice the Nyquist minimum bandwidth), thus yielding a symbol rate packing of 1 symbol/s/Hz.

- The corresponding impulse response for the filter is,

$$h(t) = 2W_0(\text{sinc}\, 2W_0 t)\frac{\cos[2\pi(W - W_0)t]}{1 - [4(W - W_0)t]^2} \quad \text{............(15)}$$



Raised-Cosine Filter characteristics -System Impulse Response

- It is plotted as shown in the diagram for *r = 0*, *r = 0.5*, and *r = 1*.
- The tails have <u>zero value</u> at each pulse sampling time, regardless of the <u>roll-off</u> value.
- Only an <u>approximate implementation</u> of a filter described by eqn(14) and a pulse shape described by eqn(15) possible, since the Raised Cosine spectrum <u>is not physically realizable</u> (for the same reason that the ideal Nyquist filter is not realizable).
- A <u>realizable filter</u> must have an impulse response of finite duration and exhibit a zero output prior to the pulse turn-on time.

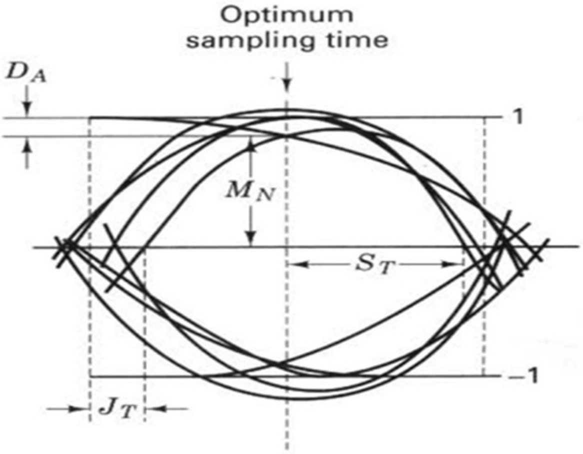
- A Pulse Shaping filter should satisfy two requirements.
- It should provide the desired roll-off, and it should be realizable, i.e., the impulse response need to be truncated to a finite length.
- Start with the Nyquist bandwidth constraint that the minimum required system bandwidth W for a symbol rate of R_s symbols/s without ISI is $R_s/2$ Hz.
- A more general relationship between required bandwidth and symbol transmission rate involves the filter roll-off factor *r* and can be stated as,

$$W = \frac{1}{2}(1 + r)R_s \qquad \text{………… (16)}$$

- So, with $r = 0$, this equation describes the minimum required bandwidth for ideal Nyquist filtering.
- For $r > 0$, there is a bandwidth expansion beyond the Nyquist minimum.
- So, for this case, R_s is now less than twice the bandwidth.
- If the demodulator outputs one sample per symbol, then the Nyquist sampling theorem has been violated, since there are too few samples left with to reconstruct the analog waveform, with presence of aliasing.
- Since the family of raised cosine filters is characterized by zero ISI at the times that the symbols are sampled, unambiguous detection can still be achieved.

Eye Pattern

- An Eye Pattern is the display that results from measuring a system’s response to baseband signals in a prescribed way.
- On the vertical plates of an oscilloscope, connect the receiver’s response to a random pulse sequence.
- On the horizontal plates, connect a saw tooth wave at the signaling frequency.
- The horizontal time base of the oscilloscope is set equal to the symbol (pulse) duration.
- This setup superimposes the waveform in each signaling interval into a family of traces in a single interval (0, T).



Eye Pattern

- The Eye Pattern that results for bipolar pulse signaling is shown.
- Since the symbols stem from a random source, they are sometimes positive and sometimes negative, and the persistence of the CRT display allowed one to see the resulting pattern shaped as an eye.
- The width of the opening indicates the time over which sampling for detection might be performed.
- It is to be noted that, the optimum sampling time corresponds to the maximum eye opening, yielding the greatest protection against noise.
- If there were no filtering in the system, that is, if the bandwidth corresponding to the transmission of these data pulses were infinite, then the system response would yield rectangular pulse shapes.

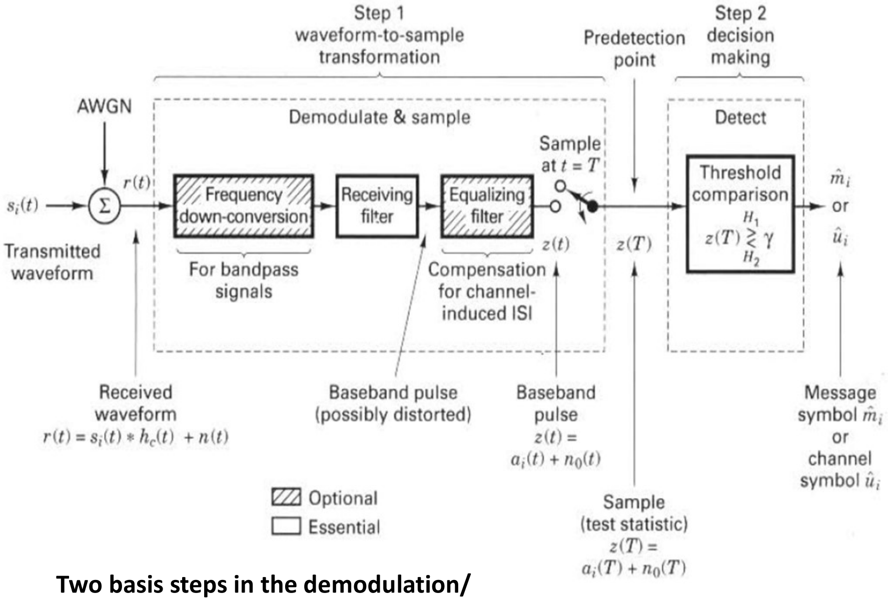
- In that case, the pattern would look like a box rather than the eye.
- The range of amplitude differences labelled D_A is a measure of distortion caused by ISI, and the range of time differences of the zero crossings labelled J_T is a measure of the timing jitter.
- Measures of noise margin M_N and sensitivity-to-timing error S_T are also shown in the diagram.
- In general, the most frequent use of the eye pattern is for qualitatively assessing the extend of the ISI.
- As the eye closes, ISI is increasing.
- As the eye opens, ISI is decreasing.

The Matched Filter

- A Matched Filter is a linear Filter designated to provide the Maximum Signal-to-Noise Power ratio at its Output for a given Transmitted Symbol Waveform.
- Consider that a known Signal s(t) plus AWGN n(t) is the Input to a LTI Receiver Filter followed by a Sampler

→as shown in the diagram to follow…

From book on DC by esteemed pioneer BERNARD SKLAR



Two basis steps in the demodulation/ detection of digital signals

Analysis of Matched Filter

- At time t=T, the sampler output z(T) consists of a signal component a_i and a noise component n_0.
- The variance of the output noise is denoted as σ_0^2, so that the ratio of instantaneous signal power to the average noise power, at time t=T, out of the sampler in *step 1* is,

$$\left(\frac{S}{N}\right)_T = \frac{a_i^2}{\sigma_0^2} \quad \text{................(1)}$$

- The aim is to find the Filter Transfer Function $H_0(f)$ that Maximizes eqn(1).

- Express the signal $a_i(t)$ at the filter output in terms of the filter transfer function H(f) and the Fourier transform of input signal, as

$$a_i(t) = \int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi ft}\,df \quad \text{................(2)}$$

- The PSD of white noise is $G_n(f)=N_0/2$ and is flat for all frequencies of interest (up to a frequency of 10^{12}Hz).
- So the output noise power can be expressed as ,

$$\sigma_0^2 = \frac{N_0}{2}\int_{-\infty}^{\infty} |H(f)|^2\,df \text{................(3)}$$

- From eqn(1),

$$\left(\frac{S}{N}\right)_T = \frac{\left|\int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT}\,df\right|^2}{N_0/2\int_{-\infty}^{\infty}|H(f)|^2\,df}$$ ……………(4)

- Next evaluate the value of H(f) = H_0(f) for which the maximum S/N is achieved, by using Schwarz's inequality as shown below,

$$\left|\int_{-\infty}^{\infty} f_1(x)f_2(x)\,dx\right|^2 \leq \int_{-\infty}^{\infty}|f_1(x)|^2\,dx\int_{-\infty}^{\infty}|f_2(x)|^2\,dx$$ ……………(5)

- The equality holds if $f_1(x) = kf_2^*(x)$, where k← arbitrary constant, *← complex conjugate

- Replace, $f_1(x)$ with H(f) and $f_2(x)$ with S(f)e $^{j2\pi ft}$,

$$\left|\int_{-\infty}^{\infty} H(f)S(f)e^{j2\pi fT}df\right|^2 \leq \int_{-\infty}^{\infty}|H(f)|^2\,df\int_{-\infty}^{\infty}|S(f)|^2\,df$$ ……………(6)

- Substituting into eqn(1),

$$\left(\frac{S}{N}\right)_T \leq \frac{2}{N_0}\int_{-\infty}^{\infty}|S(f)|^2\,df$$ ……………(7)

$$\max\left(\frac{S}{N}\right)_T = \frac{2E}{N_0}$$ ……………(8)

- Where the energy E of the input signal s(t) is,

$$E = \int_{-\infty}^{\infty}|S(f)|^2\,df$$ ……………(9)

- Thus, the maximum output $(S/N)_T$ depends on the input signal energy and the PSD of the noise, not on the particular shape of the waveform that is used.
- The equality of eqn (6) holds only if the optimum filter transfer function $H_0(f)$ is employed, such that,

$$H(f) = H_0(f) = kS^*(f)e^{-j2\pi fT} \qquad \text{................}(10)$$

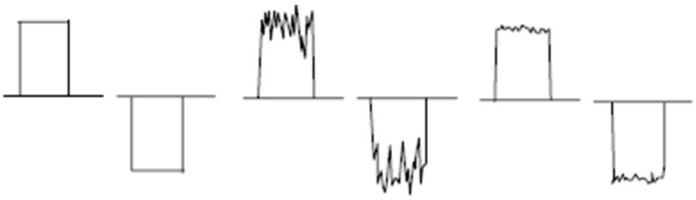
$$h(t) = \mathcal{F}^{-1}\{kS^*(f)e^{-j2\pi fT}\} \qquad \text{................}(11)$$

- Since s(t) is a real-valued signal,

$$h(t) = \begin{cases} ks(T-t) & 0 \le t \le T \\ 0 & \text{elsewhere} \end{cases} \qquad \text{................}(12)$$

- Therefore, the impulse response of a Filter that produced the Maximum output Signal-to-Noise ratio is the mirror image of the message signal s(t), delayed by the symbol time duration T.
- The delay of T seconds makes eqn(12) causal, i.e., the delay of T seconds makes h(t) a function of positive time in the interval $0 \le t \le T$.
- Without the delay of T seconds, the response s(-t) is unrealizable, since it describes a response as a function of negative time.

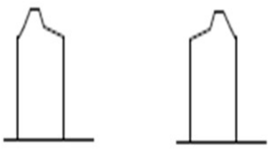
- Input signal s(t)+n(t)



(a) Transmitted signal, square pulses
(b) At the receiver, distorted by a lot of noise.
(c) After the receive filter, looking a good deal more like the transmitted signal.

Matched Filter

The signal to noise ratio is maximized when the impulse response of that filter is exactly a reversed and time delayed copy of the transmitted signal.

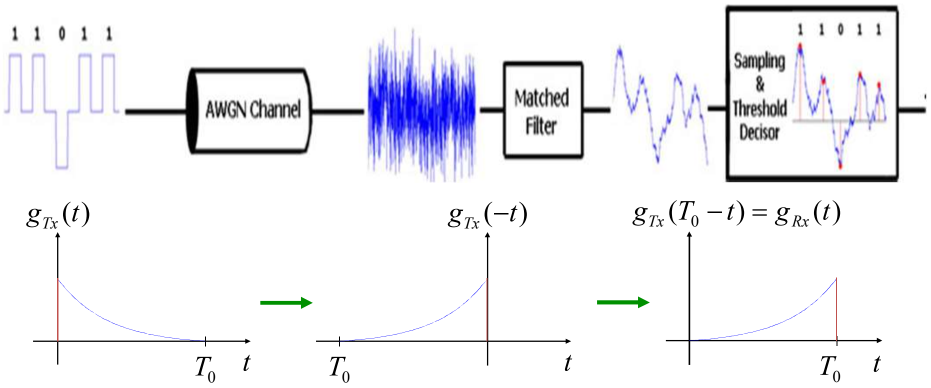


(a) Transmitted signal, (b) the required impulse response of the receive filter.

Matched filter example

- Received SNR is maximized at time T_0

Matched Filter: optimal receive filter for maximized $\frac{S}{N}$



Transmit filter

Receive filter (matched filter)

Inference

- The impulse response of the optimum filter is a **scaled, time reversed and delayed version** of the input signal g(t) i.e., it is matched to the input signal.
- A linear time invariant filter defined in this way is called a **Matched filter.**

Properties of Matched Filter

- A filter that is matched to a pulse signal g(t) of duration T, is characterized by an impulse response that is a time reversed and delayed version of the input.
- The peak pulse signal-to-noise ratio of a matched filter depends only on the ratio of the signal energy to the PSD of the white noise at the filter input.
- The **Rayleigh's energy theorem** states that the integral of the squared magnitude spectrum of a pulse signal w.r.t frequency is equal to the signal energy.

Properties of Matched Filters

Consider a known signal $g(t)$,

$$G_0(f) = H_{\text{opt}}(f)G(f)$$

$$= kG^*(f)G(f)\exp(-j2\pi fT)$$

$$= k|G(f)|^2 \exp(-j2\pi fT)$$

$$g_0(T) = \int_{-\infty}^{\infty} G_0(f)\exp(j2\pi fT)df$$

$$= k\int_{-\infty}^{\infty} |G(f)|^2 df$$

$$g_0(T) = kE$$

$$E\left[n^2(t)\right] = \int_{-\infty}^{\infty} S_N(f)\,df$$

$$= \frac{k^2 N_0}{2}\int_{-\infty}^{\infty} |G(f)|^2 df$$

$$= k^2 N_0 E/2$$

$$\eta_{\max} = \frac{(kE)^2}{(kN_0 E/2)} = \frac{2E}{N_0}$$

which is independent of waveform (E/N_0 = signal energy to noise PSD ratio)

► EXAMPLE **Matched Filter for Rectangular Pulse**

Consider a signal $g(t)$ in the form of a rectangular pulse of amplitude A and duration T, as shown in Figure . In this example, the impulse response $h(t)$ of the matched filter has exactly the same waveform as the signal itself. The output signal $g_o(t)$ of the matched filter produced in response to the input signal $g(t)$ has a triangular waveform, as shown in Figure

The maximum value of the output signal $g_o(t)$ is equal to kA^2T, which is the energy of the input signal $g(t)$ scaled by the factor k; this maximum value occurs at $t = T$, as indicated in Figure ·

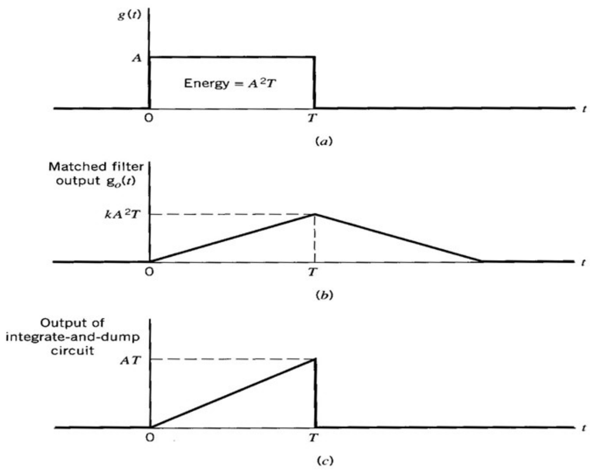
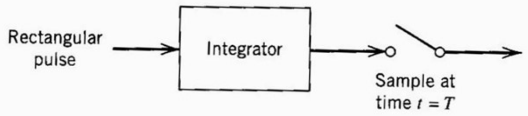


FIGURE (*a*) Rectangular pulse. (*b*) Matched filter output. (c) Integrator output.



Integrate-and-dump circuit.

For the special case of a rectangular pulse, the matched filter may be implemented using a circuit known as the *integrate-and-dump circuit*, a block diagram of which is shown in Figure The integrator computes the area under the rectangular pulse, and the resulting output is then sampled at time $t = T$, where T is the duration of the pulse. Immediately after $t = T$, the integrator is restored to its initial condition; hence the name of the circuit. Figure shows the output waveform of the integrate-and-dump circuit for the rectangular pulse of Figure We see that for $0 \leq t \leq T$, the output of this circuit has the *same waveform* as that appearing at the output of the matched filter; the difference in the notations used to describe their peak values is of no practical significance. ◀

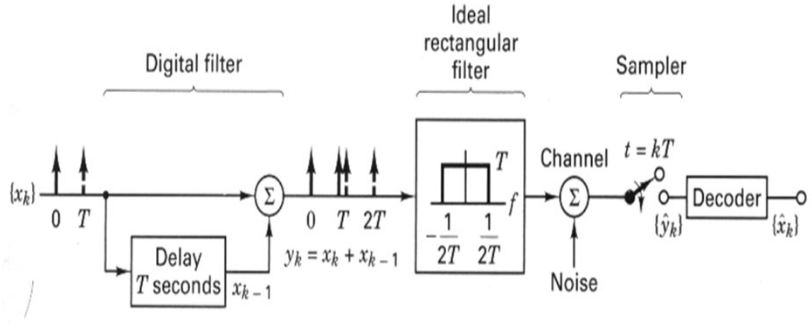
Correlative Level Coding

Introduction

- In 1963, Adam Lender showed that it is possible to transmit 2W symbols/s with zero ISI, using the theoretical minimum bandwidth of W Hz, without infinite sharp filters, using the technique of Duobinary Signaling which is also known as Correlative Coding and Partial Response Signaling.
- The basic idea behind the Duobinary technique is to introduce some controlled amount of ISI into the datastream rather than trying to eliminate it completely.
- By introducing Correlated Interference between the pulses, and by changing the detection procedure, Lender was effectively able to cancel out the interference at the detector and thereby achieve the ideal symbol-rate packing of 2 symbols/s/Hz, an amount that had been considered unrealizable.

Duobinary Signaling

- Duobinary signaling introduces controlled ISI.
- Consider the model of the Duobinary coding process.



Duobinary Signaling

- Assume that a sequence of binary symbols $\{x_k\}$ is to be transmitted at the rate of R symbols/s over a system having an ideal rectangular spectrum of bandwidth W=R/2 = 1/2T Hz.
- Initially, the pulses pass through a simple digital filter, which incorporates a one digit delay to each incoming pulse, the filter adds the current value to the previous pulse.
- So, for every pulse to the input of the digital filter, there is summation of two pulses out.
- Each pulse of the sequence $\{y_k\}$ out of the digital filter can be expresses as,

$$y_k = x_k + x_{k-1} \quad \text{……………(1)}$$

- Hence, the $\{y_k\}$ amplitudes are not independent as each y_k digit carries with it the memory of the prior digit.
- The ISI introduced to each y_k digit comes only from the preceding x_{k-1} digit.
- This correlation between the pulse amplitudes of $\{y_k\}$ can be thought of as the controlled ISI introduced by the Duobinary coding.
- Controlled interference is the essence of this novel technique, since at the detector, such controlled interference can be as easily removed as it was added.

- The sequence $\{y_k\}$ is followed by the ideal Nyquist filter that does not introduce any ISI.
- At the receiver sampler, the sequence $\{y_k\}$ can be recovered exactly in the absence of noise.
- Since all system experience noise contamination, refer to the received $\{y_k\}$ as the estimate of $\{y_k\}$ and denote it $\{\hat{y}_k\}$.
- Removing the controlled interference with the Duobinary decoder yields an estimate of $\{x_k\}$ which is denoted as $\{\hat{x}_k\}$.

Duobinary Decoding

- If the binary digit x_k is equal to ± 1, then by using eqn(1), y_k has one of the three possible values: +2, 0, or -2.
- The Duobinary code results in a three-level output.
- In general for M-ary transmission, partial response signalling results in 2M-1 output levels.
- The decoding procedure involved the inverse of the coding procedure, namely, subtracting the x_k-1 decision from the y_k digit.

Example -1.

Duobinary coding and decoding

Demonstrate Duobinary coding and decoding for the following sequence $\{x_k\}$ = 0 0 1 0 1 1 0. Let the first bit be a start-up digit, not part of the data.

Sol:

Binary digit sequence $\{x_k\}$: 0 0 1 0 1 1 0

Bipolar amplitudes $\{x_k\}$: -1 -1 +1 -1 +1 +1 -1

Coding rule $y_k = x_k + x_{k-1}$: -2 0 0 0 2 0

Decoding decision rule:

If $\hat{y}_k$ = 2, decide that $\hat{x}_k$= +1 (or binary one).

If $\hat{y}_k$ = -2, decide that $\hat{x}_k$= -1 (or binary one).

If $\hat{y}_k$ = 0, decide opposite of the previous decision.
Decoded bipolar sequence $\{\hat{x}_k\}$: -1 +1 -1 +1 +1 -1
Decoded binary sequence $\{\hat{x}_k\}$: 0 1 0 1 1 0

Points to Ponder…

- The decision rule simply implements the subtraction of each $\hat{x}_{k-1}$ decision from each $\hat{y}_{k.}$
- One drawback of this decision technique is that once an error is made, it tends to propagate, causing further errors, since present decisions depend on prior decisions.
- A means of avoiding this error propagation is known as Precoding.

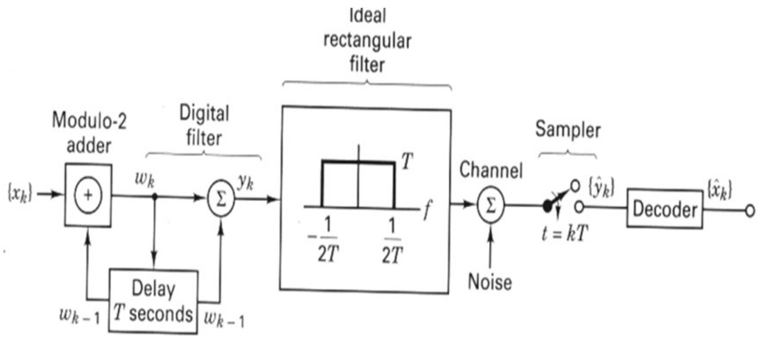
Precoding

- Precoding is accomplished by first differentially encoding the {x_k} binary sequence into a new {w_k} binary sequence by means of the equation:

$$w_k = x_k \oplus w_{k-1} \quad \text{……….. (2)}$$

- The {w_k} binary sequence is then converted to a bipolar pulse sequence, and the coding operation proceeds in the say way as before.
- However, with Precoding, the detection process is quite different from the detection of ordinary Duobinary coding.

The Precoding model is shown below in the diagram, whereby it is clear that the modulo-2 addition producing the precoded {w_k} sequence is performed on the binary digits, while the digital filtering producing the {y_k} sequence is performed on the bipolar pulses.



Precoded Duobinary Signalling

Example:2 Duobinary Precoding

Demonstrate Duobinary coding and decoding when using Precoding for the following sequence $\{x_k\}$ = 0 0 1 0 1 1 0.

Sol:

Binary digit sequence $\{x_k\}$:	0	0	1	0	1	1	0
Precoded Sequence: $w_k = x_k \oplus w_{k-1}$	0	0	1	1	0	1	1
Bipolar amplitudes $\{x_k\}$: $y_k = w_k + w_{k-1}$:	-1	-1	+1	+1	-1	+1	+1
Coding rule		-2	0	+2	0	0	+2

- Decoding decision rule:
  If $\hat{y}_k = \pm 2$, decide that $\hat{x}_k$= binary zero.
  If $\hat{y}_k = 0$, decide that $\hat{x}_k$= binary one.

- Decoded binary sequence $\{\hat{x}_k\}$:

0 1 0 1 1 0

- The differential Precoding enables to decode the $\{\hat{y}_k\}$ sequence by making a decision on each received sample singly, without resorting to prior decisions that could be in error.
- The major advantage is that in the event of a digit error to noise, such an error does not propagate to other digits.
- The first bit in the differentially precoded binary sequence $\{w_k\}$is an arbitrary choice.
- If the start-up bit in $\{w_k\}$ had been chosen to be a binary one instead of binary zero, the decoded result would have been the same.

Generalized Partial response signalling.

- The duobinary transfer function has a digital filter that incorporates a one-digit delay followed by an ideal rectangular transfer function.
- Examine an equivalent model.
- The Fourier transform of a delay can be given as $e^{-j2\pi fT}$.
- So the input digital filter can be characterized as the frequency transfer function,

$$H_1(f) = 1 + e^{-j2\pi fT} \quad \text{……….. (3)}$$

- The transfer function of the ideal rectangular filter is,

$$H_2(f) = \begin{cases} T & \text{for } |f| < \frac{1}{2T} \\ 0 & \text{elsewhere} \end{cases} \quad \text{………..} \ (4)$$

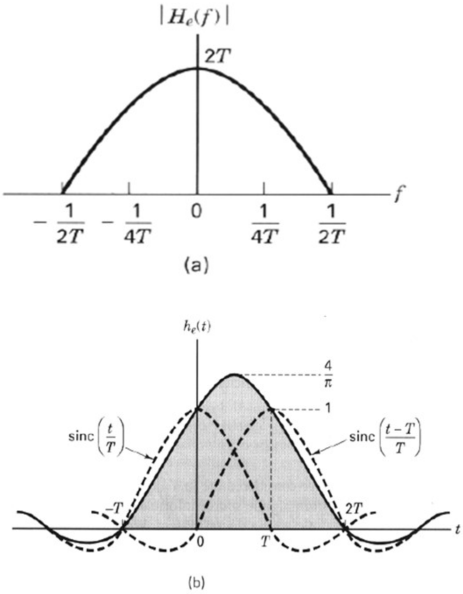
- The overall equivalent transfer function of the digital filter cascaded with the ideal rectangular filter is then given by,

$$\begin{aligned} H_e(f) &= H_1(f)H_2(f) \qquad \text{for } |f| < \frac{1}{2T} \\ &= (1 + e^{-j2\pi fT})T \qquad \text{………..} \ (5) \\ &= T(e^{j\pi fT} + e^{-j\pi fT})e^{-j\pi fT} \end{aligned}$$

$$|H_e(f)| = \begin{cases} 2T\cos \pi fT & \text{for } |f| < \frac{1}{2T} \\ 0 & \text{elsewhere} \end{cases} \quad \text{………..} \ (6)$$

- Thus the composite transfer function for the cascaded digital and rectangular filters, has a gradual roll-off to the band edge as shown below.
- The transfer function can be approximated by using realizable analog filtering, a separate digital filter is not needed.
- The duobinary equivalent $H_e(f)$ is called a Cosine Filter.
- The Cosine Filter is not the same as the Raised Cosine Filter.
- The corresponding impulse response $h_e(t)$ is got by taking the inverse Fourier transform of $H_e(f)$ and is given as,

$$h_e(t) = \text{sinc}\left(\frac{t}{T}\right) + \text{sinc}\left(\frac{t-T}{T}\right) \quad \text{………..} \ (7)$$



Duobinary Transfer Function and Pulse Shape
(a) Cosine Filter (b) Impulse response of the Cosine Filter

- For every impulse $\delta(t)$ at the input of the model, the output is $h_e(t)$ with an appropriate polarity.
- Notice that there are only two nonzero samples at T-second intervals, giving rise to controlled ISI from the adjacent bit.
- The introduced ISI is eliminated by use of the decoding procedure.
- Although the Cosine Filter is noncausal and therefore nonrealizable, it can easily be approximated.
- The implementation of the precoded duobinary technique can be accomplished by first differentially encoding the binary sequence $\{x_k\}$ into the sequence $\{w_k\}$.
- The pulse sequence $\{w_k\}$ is then filtered by the equivalent cosine characteristic.

Comparison of Binary with Duobinary Signalling

- The duobinary technique introduces correlation between pulse amplitudes, whereas the more restrictive Nyquist criterion assumes that the transmitted pulse amplitudes are independent of one another.
- Duobinary signalling exploits introduced correlation to achieve zero ISI signal transmission using a smaller system bandwidth than is otherwise possible.
- But there is trade-off involved.
- Duobinary coding required three levels, compared with the usual two levels of binary coding.
- For a fixed amount of signal power, the ease of making reliable decisions is inversely related to the number of levels that must be distinguished in each waveform.

- Therefore, although duobinary signalling achieves the zero ISI requirements with minimum bandwidth, duobinary signalling also required more power than binary signalling, for equivalent performance against noise.
- For a given probability of bit error (P_B), duobinary signalling required approximately 2.5 dB greater SNR than binary signalling, while using only 1/(1+r) the bandwidth that binary signalling requires, where r is the filter roll-off.

Transmission Over AWGN Channel:

Module – IV Part-3

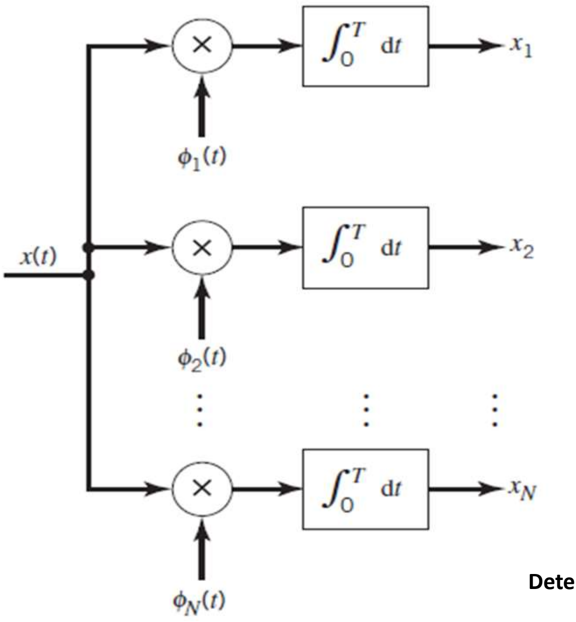
ECT305 Analog and Digital
Communication

Conversion of the Continuous AWGN Channel into a Vector Channel

- Suppose that the input to the bank of N product integrators or correlators is the received signal *x(t)* defined as:

$$x(t) = s_i(t) + w(t), \quad \begin{cases} 0 \leq t \leq T \\ i = 1, 2, \ldots, M \end{cases} \qquad \ldots(1)$$

- *w(t)* is a sample function of a white Gaussian noise process W(t) of zero mean and power spectral density $N_0/2$.
- The output of correlator j is the sample value of a random variable X_j



Detector or Demodulator

The output of correlator j is the sample value of a random variable X_j

$$x_j = \int_0^T x(t)\phi_j(t)\,dt \qquad \text{...(2)}$$

$$= s_{ij} + w_j, \qquad j = 1, 2, \ldots, N$$

$$s_{ij} = \int_0^T s_i(t)\phi_j(t)\,dt$$

Deterministic quantity!!

$$w_j = \int_0^T w(t)\phi_j(t)\,dt$$

Random quantity!!

.....(3)

$$s_{ij} = \int_0^T s_i(t)\phi_i(t)dt$$

$$w_i = \int_0^T w(t)\phi_i(t)dt$$

deterministic quantity

contributed by the transmitted signal s_i(t)

random quantity

sample value of the variable W_i due to noise

- Consider a random process *X′(t),* with *x′(t),* a sample function which is related to the received signal *x(t)* as follows:

$$x'(t) = x(t) - \sum_{j=1}^{N} x_j\phi_j(t) \qquad \textbf{.....(4)}$$

$$x'(t) = x(t) - \sum_{j=1}^{N}(s_{ij} + w_j)\phi_j(t)$$

$$= w(t) - \sum_{j=1}^{N} w_j\phi_j(t)$$

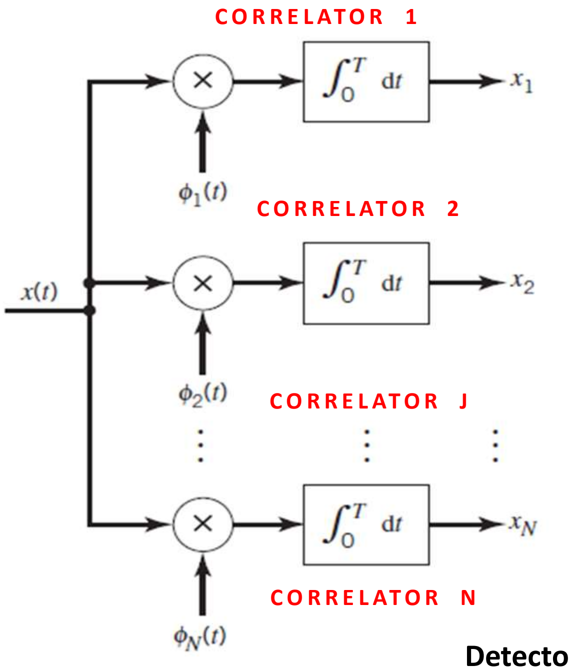
$$= w'(t) \qquad \textbf{.....(5)}$$

- *which means that* ***the sample function x′(t) depends only on the channel noise!***

- The received signal can be expressed as:

$$x(t) = \sum_{j=1}^{N} x_j \phi_j(t) + x'(t)$$

$$= \sum_{j=1}^{N} x_j \phi_j(t) + w'(t) \qquad \text{.....(6)}$$



Detector or Demodulator

Statistical Characterization

- The received signal (output of the correlator) is a random signal.
- To describe it we need to use statistical methods – mean and variance. ??

- The assumptions are:
  - X(t) denotes a random process, a sample function of which is represented by the received signal x(t).
  - $X_j(t)$ denotes a random variable whose sample value is represented by the correlator output $x_j(t)$, j = 1, 2, …N.
  - We have assumed AWGN, so the noise is Gaussian, so X(t) is a Gaussian process and being a Gaussian RV, X_j is described fully by its Mean value and Variance.

Mean Value

- Let $W_{j,}$ denote a random variable, represented by its sample value w_j, produced by the jth correlator in response to the Gaussian noise component w(t).
- So it has zero mean (by definition of the AWGN model)

$$\mu_{x_j} = E\left[X_j\right]$$

$$= E\left[s_{ij} + W_j\right]$$

$$= s_{ij} + E[W_j]$$

$$\mu_{x_j} = s_{ij} \quad \text{.....(7)}$$

- …then the ***mean of X_j*** depends only on s_{ij}

Variance

$$\begin{aligned}\sigma_{X_j}^2 &= \mathrm{var}[X_j] \\ &= \mathbb{E}[(X_j - s_{ij})^2] \\ &= \mathbb{E}[W_j^2]\end{aligned}$$ …..(8)

$$W_j = \int_0^t W(t)\phi_j(t)\,\mathrm{d}t$$

$$\begin{aligned}\sigma_{X_j}^2 &= \mathbb{E}\int_0^T W(t)\phi_j(t)\,\mathrm{d}t\int_0^T W(u)\phi_j(u)\,\mathrm{d}u \\ &= \mathbb{E}\left[\int_0^T\int_0^T \phi_j(t)\phi_j(u)W(t)W(u)\,\mathrm{d}t\,\mathrm{d}u\right]\end{aligned}$$ …..(9)

- Interchanging the order of integration and expectation,

$$\begin{aligned}\sigma_{X_j}^2 &= \int_0^T\int_0^T \phi_j(t)\phi_j(u)\mathbb{E}[W(t)W(u)]\,\mathrm{d}t\,\mathrm{d}u \\ &= \int_0^T\int_0^T \phi_j(t)\phi_j(u)R_W(t,u)\,\mathrm{d}t\,\mathrm{d}u\end{aligned}$$ (10)

- $R_w(t,u)$ →Autocorrelation function of the noise process

- Since the noise is stationary and with a constant power spectral density, it can be expressed as:

$$R_w(t,u) = \frac{N_0}{2}\delta(t-u) \qquad \text{.....(11)}$$

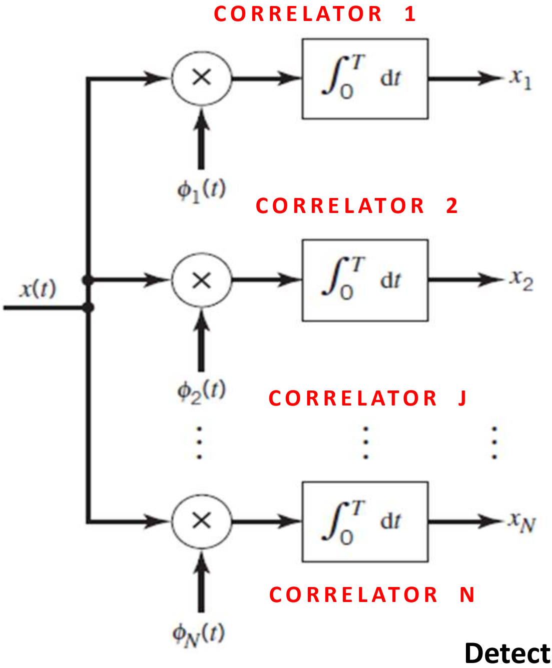
- After substitution for the variance we get:

$$\sigma_{X_j}^2 = \frac{N_0}{2}\int_0^T\int_0^T \phi_j(t)\phi_j(u)\delta(t-u)\,dt\,du$$

$$= \frac{N_0}{2}\int_0^T \phi_j^2(t)\,dt \qquad \text{.....(12)}$$

- And since $\phi_j(t)$ has unit energy for the ***variance*** we finally have:

$$\sigma_{X_j}^2 = \frac{N_0}{2}, \qquad \text{for all } j \qquad \text{.....(13)}$$



Detector or Demodulator

- ***Correlator outputs, denoted by X_j have variance equal to the power spectral density $N_0/2$ of the noise process W(t).***
- So, all the N correlator o/ps denoted by X_j with j=1,2,…N have a variance equal to PSD $N_0/2$ of W(t).
- ***Show that …***
- X_j are mutually uncorrelated (use Covariance of correlator o/ps)
- X_j are statistically independent (follows from above because X_j are Gaussian)

Define the vector

$$\mathrm{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \mathbf{14}$$

x is called the observation vector.

The elements are indep. Gaussian RVs with means s_{ij}

and $N_0/2$ variances (for sample $s_i(t)$)

The conditional pdf given that $s_i(t)$ (m_i) was transmitted is

$$f_{\mathrm{x}}(\mathrm{x}|m_i) = \prod_{j=1}^{N} f_{x_j}(x_j|m_i) \;,\; i = 1,2,\ldots,M \qquad \mathbf{15}$$

Any channel that satisfies **15** is called a memoryless channel

$$f_{x_j}(x_j|m_i) = \frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{1}{N_0}(x_j - s_{ij})^2\right], \quad \begin{matrix} j=1,2,\ldots,N \\ i=1,2,..,M \end{matrix}$$ 16

$$f_{\mathrm{x}}(\mathrm{x}|m_i) = (\pi N_0)^{-\frac{N}{2}}\exp\left[-\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2\right], i=1,2,\ldots,M$$ 17

Recall $x(t) = \sum_{j=1}^{N} x_j\phi_j(t) + w'(t)$

$w'(t)$ is a zero - mean Gaussian process and indep. of $\{X_j\}$

$$E\left[X_j W'(t_k)\right] = 0 \ , \quad \begin{cases} j=1,2,\ldots,N \\ 0 \le t_k \le T \end{cases}$$ 18

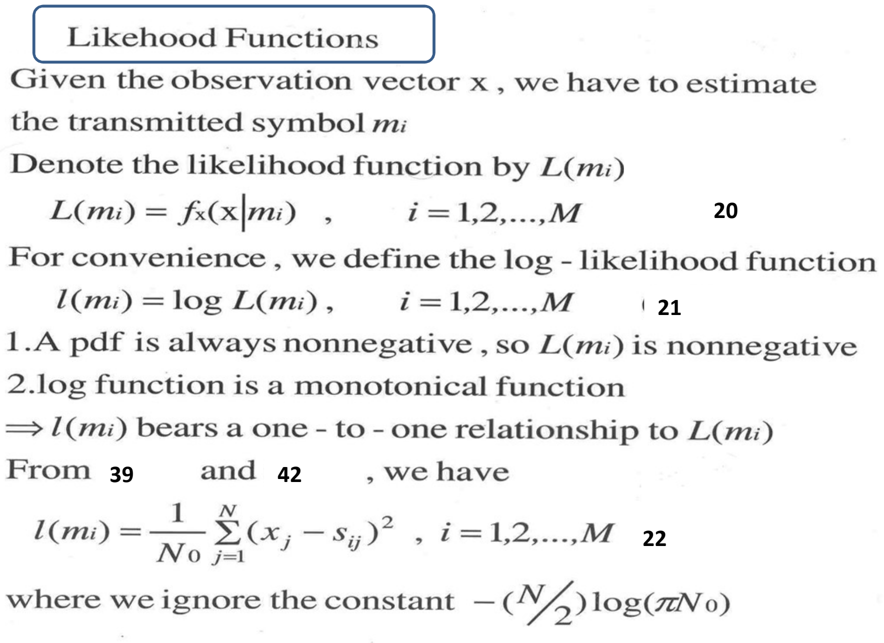
$\Rightarrow$ Theorem of irrelevance : Only the projections of the noise onto the basis functions affects the detections , the remainder is irrelevant

$\Rightarrow$ The AWGN channel is equivalent to an N - dim. vector channel

$\mathrm{x} = s_i + \mathrm{w} \qquad , \quad i=1,2,\ldots,M$ 19

Theorem of Irrelevance:

As far as signal detection in additive white Gaussian noise is concerned, only the projections of the noise onto the basis functions of the signal set affects the sufficient statistics of the detection problem, the remainder of the noise is irrelevant.



Likehood Functions

Given the observation vector x , we have to estimate the transmitted symbol m_i

Denote the likelihood function by $L(m_i)$

$$L(m_i) = f_{\mathbf{x}}(\mathbf{x}|m_i) \quad , \qquad i = 1,2,\ldots,M \qquad 20$$

For convenience , we define the log - likelihood function

$$l(m_i) = \log L(m_i) \, , \qquad i = 1,2,\ldots,M \qquad 21$$

1.A pdf is always nonnegative , so $L(m_i)$ is nonnegative

2.log function is a monotonical function

$\Rightarrow l(m_i)$ bears a one - to - one relationship to $L(m_i)$

From 39 and 42 , we have

$$l(m_i) = \frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{ij})^2 \quad , \; i = 1,2,\ldots,M \qquad 22$$

where we ignore the constant $-(N/2)\log(\pi N_0)$

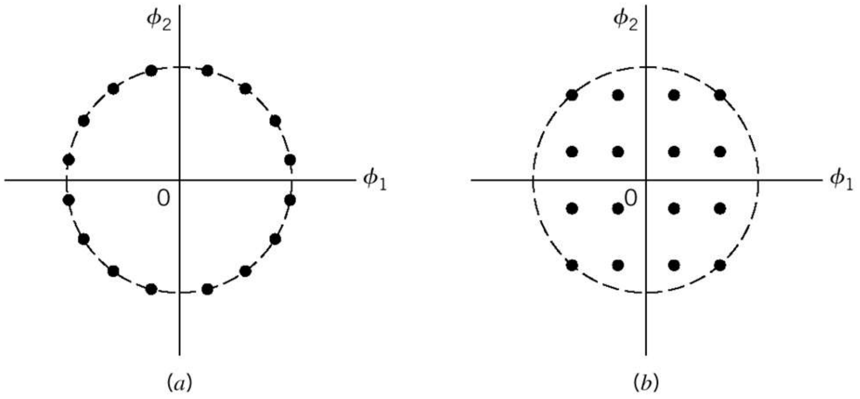
Transmitted signal representations

- Let one of the M possible signals of $s_i(t)$, i=1,2,…M be transmitted in each time slot T with equal probability 1/M.
- This is applied to a bank of correlators supplied with N orthogonal basis functions.
- The resulting correlator o/ps define the signal vector $\mathbf{s}_i$.
- $\mathbf{s}_i$ is represented as a point in Euclidean space of dimensions N<=M, called transmitted signal point.
- The set of message points corresponding to the set of transmitted signals is called Signal Constellation.

Coherent Detection of Signals in Noise:

Maximum Likelihood Decoding.

The set of transmitted signals is called a signal constellation

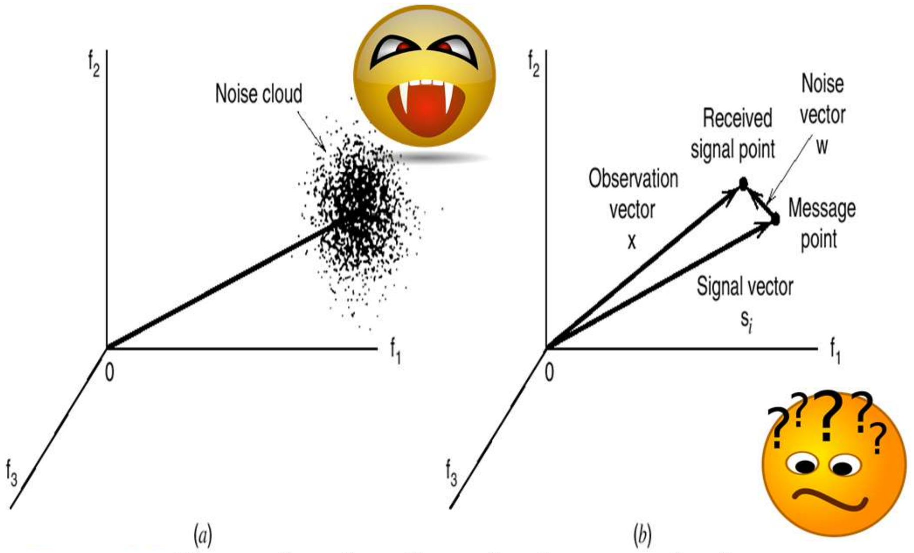


(*a*) (*b*)

ignal constellation for (a) M-ary PSK and (b) corresponding M-ary QAM, for M =16.

Representation of received signal x(t)

- Complicated due to presence of w(t).
- Correlator o/ps define observation vector **x** that may be represented in the same Euclidean space as received signal point.
- The received signal point wanders about the message point in a random way and may lie anywhere inside a Gaussian distributed “cloud” with message point as centroid.
- This is due to the **Noise Perturbation** effect.



Illustrating the effect of noise perturbation, depicted in (a), on the location of the received signal point, depicted in (b).

Detection problem:

Given x, perform a mapping from x to an estimate $\hat{m}$ of m_i, in a way that would minimize the probability of error.

The prob.of error denoted by $P_e(m_i|\mathrm{x})$ is

$$P_e(m_i|\mathrm{x}) = P(m_i \text{ not sent } |\mathrm{x})$$

$$= 1 - P(m_i \text{ sent } |\mathrm{x}) \qquad \mathbf{23}$$

The optimum decsion rule is

set $\hat{m} = m_i$ if

$$P(m_i \text{sent } |\mathrm{x}) \geq P(m_k \text{sent } |\mathrm{x}) \text{ for all } k \neq i \qquad \mathbf{24}$$

which is also called the maximum a posteriori probability (MAP) rule.

In terms to the a priori prob. of $\{m_i\}$, using Bayes' rule, we may restate the MAP rule as

set $\hat{m} = m_i$ if

$$\frac{P_k f_{\mathrm{x}}(\mathrm{x}|m_k)}{f_{\mathrm{x}}(\mathrm{x})} \text{ is maximum for } k = i \qquad 25$$

where p_k is the a priori prob. of m_k

Note that

1. $f_{\mathrm{x}}(\mathrm{x})$ is indep. of $\{m_i\}$
2. If $\{m_i\}$ are equally likely, $p_k = p_i = p$
3. $f_{\mathrm{x}}(\mathrm{x}|m_k)$ bears one - to - one relationship to $l(m_k)$

Then we can restate the decision rule as

set $\hat{m} = m_i$ if

$$l(m_k) \text{ is maximum for } k = i \qquad 26$$

-Maximum Likelihood Rule

The maximum likelihood decoder differs from the maximum a posteriori decoder (Assum. of $p_k = \text{constant}$)

Let Z denote the N - dim space (observation space).

We may partition Z into M - decision regions denoted by $Z_1, Z_2, \ldots, Z_M$

Observation vector x lies in Z_i if

$$l(m_k) \text{ is max. for } k = i \qquad 27$$

Recall $$l(m_k) = -\frac{1}{N_0}\sum_{j=1}^{N}(x_j - s_{kj})^2, \ \mathrm{i} = 1,2,\ldots,M$$

↑

Minimize this term to maximize $l(m_k)$ by the choice $i = k$

x lies in Z_i if

$$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \left\|\mathrm{x} - s_k\right\|^2 \text{ is min. for } k = i \qquad \mathbf{28}$$

$$\Rightarrow \mathrm{x} \in Z_i \quad \text{, if } \left\|\mathrm{x} - s_k\right\| \text{ is min. for } k = i \qquad \mathbf{29}$$

$\Rightarrow$ to choose the message point closest to the received signal point.

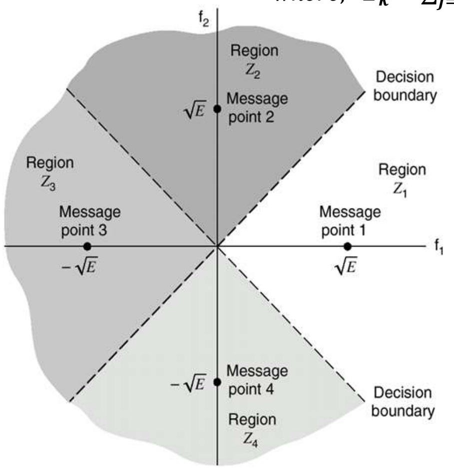
$$\sum_{j=1}^{N}(x_j - s_{kj})^2 = \underset{\uparrow \atop \text{indep. of } k}{\sum_{j=1}^{N} x_j^2} - 2\sum_{j=1}^{N} x_j s_{kj} + \underset{\uparrow \atop \text{energy of } s_k(t) = E_k}{\sum_{j=1}^{N} s_{kj}^2} \qquad \mathbf{30}$$

Equivalently we have

$\mathrm{x} \in Z_i$ if

$$\sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2}E_k \text{ is max. for } k = i \qquad \mathbf{31}$$

$$\textit{where,}\quad \boldsymbol{E_k} = \sum_{j=1}^{N} \boldsymbol{s_{kj}}^2 \qquad \mathbf{32}$$

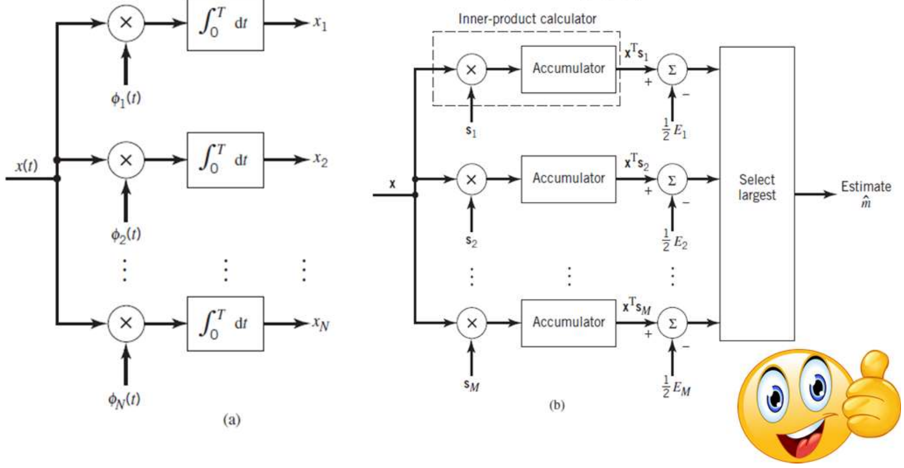


Illustrating the partitioning of the observation space into decision regions for the case when N =2 and M =4; it is assumed that the M transmitted symbols are equally likely.

Overall Correlation Receiver structure

$$\sum_{j=1}^{N} x_j s_{kj} - \frac{1}{2}E_k \text{ is max. for } k = i$$

where, $E_k = \sum_{j=1}^{N} s_{kj}^2$



Detector **Maximum Likelihood Decoder (MLD)**

Correlation Receiver

- The optimum receiver for an AWGN channel and for the case when the transmitted signals $s_1(t)$, $s_2(t)$, …, $s_M(t)$ are equally likely is called a Correlation receiver.
- It consists of two subsystems, as shown.
- **1.** Detector, which consists of *M* correlators supplied with a set of orthonormal basis functions $\phi_1(t)$, $\phi_1(t)$, …, $\phi_N(t)$, that are generated locally; this bank of correlators operates on the received signal $x(t)$, $0 \leq t \leq T$, to produce the observation vector **x**.
- **2.** Maximum-likelihood decoder, which operates on the observation vector **x** to produce an estimate of the transmitted symbol *mi, i* = 1, 2, …, *M*, in such a way that the average probability of symbol error is minimized.

- In accordance with the maximum likelihood decision rule, the decoder multiplies the N elements of the observation vector $\mathbf{x}$ by the corresponding N elements of each of the M signal vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$.
- Then, the resulting products are successively summed in accumulators to form the corresponding set of inner products $\{\mathbf{x}^T\mathbf{s}_k | k = 1, 2, \ldots, M\}$.
- Next, the inner products are corrected for the fact that the transmitted signal energies may be unequal.
- Finally, the largest one in the resulting set of numbers is selected, and an appropriate decision on the transmitted message is thereby made.

Matched Filter Receiver

- The optimum detector involves a set of correlators.
- Alternatively, use a different but equivalent structure in place of the correlators.
- To explore this alternative method of implementing the optimum receiver, consider a linear time-invariant filter with impulse response $h_j(t)$.
- With the received signal $x(t)$ operating as input, the resulting filter output is defined by the convolution integral,

$$y_j(t) = \int_{-\infty}^{\infty} x(\tau)h_j(t-\tau)\,\mathrm{d}\tau \qquad \text{....... (33)}$$

- To proceed further, evaluate this integral over the duration of a transmitted symbol, namely $0 \leq t \leq T$.
- With time t restricted in this manner, replace the variable τ with t,

$$y_j(T) = \int_0^T x(t)h_j(T-t)\,\mathrm{d}t \quad \text{....... (34)}$$

- Consider next a detector based on a bank of correlators.
- The output of the jth correlator is defined by :

$$x_j = \int_0^T x(t)\phi_j(t)\,\mathrm{d}t \quad \text{....... (35)}$$

- For $y_j(T)$ to equal x_j, this condition is satisfied provided that.

$$h_j(T-t) = \phi_j(t) \qquad \text{for } 0 \leq t \leq T \text{ and } j = 1, 2, \ldots, M$$

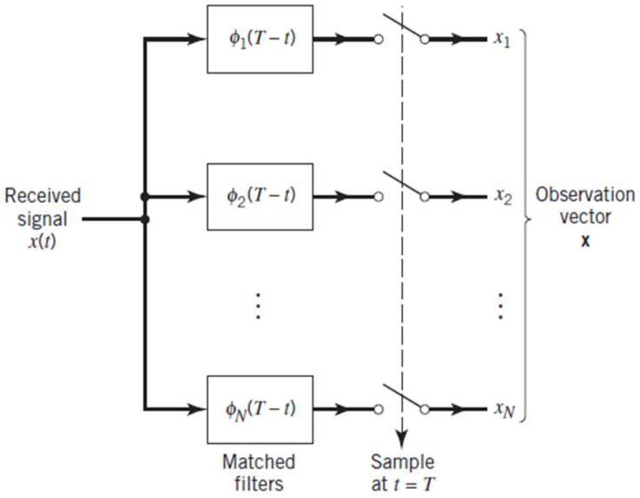
- Equivalently, express the condition imposed on the desired impulse response of the filter,

$$h_j(t) = \phi_j(T-t), \qquad \text{for } 0 \leq t \leq T \text{ and } j = 1, 2, \ldots, M \quad \text{....... (36)}$$

- Given a pulse signal $\phi(t)$ occupying the interval $0 \leq t \leq T$, a linear time-invariant filter is said to be matched to the signal $\phi(t)$ if its impulse response $h(t)$ satisfies the condition,

$$h(t) = \phi(T-t) \text{ for } 0 \leq t \leq T \quad \text{....... (37)}$$

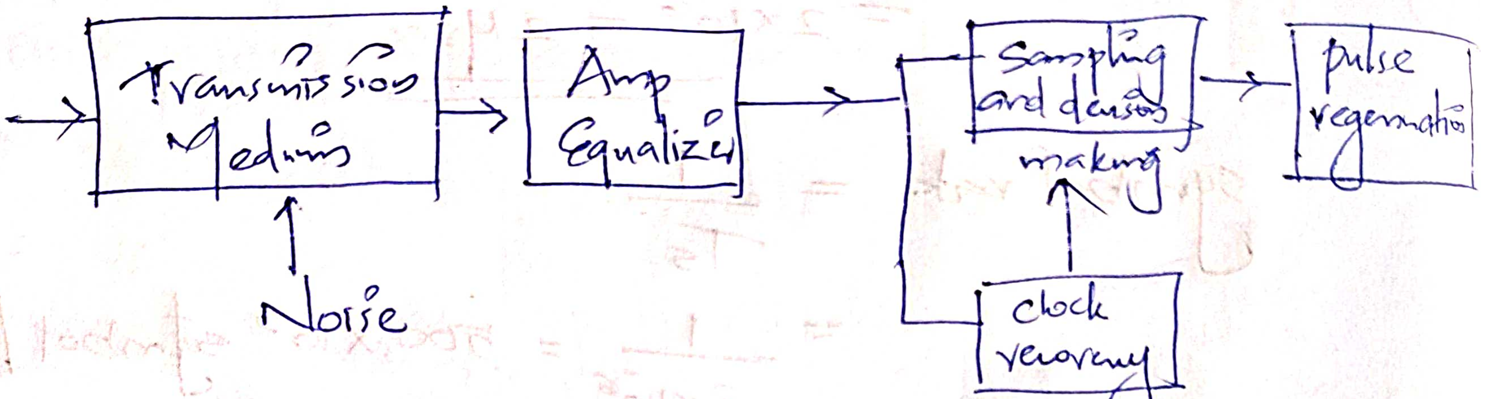
A time-invariant filter defined in this way is called a Matched filter. Correspondingly, an optimum receiver using Matched filters in place of Correlators is called a Matched-filter receiver.



Detector part of matched filter receiver

Equalizer:

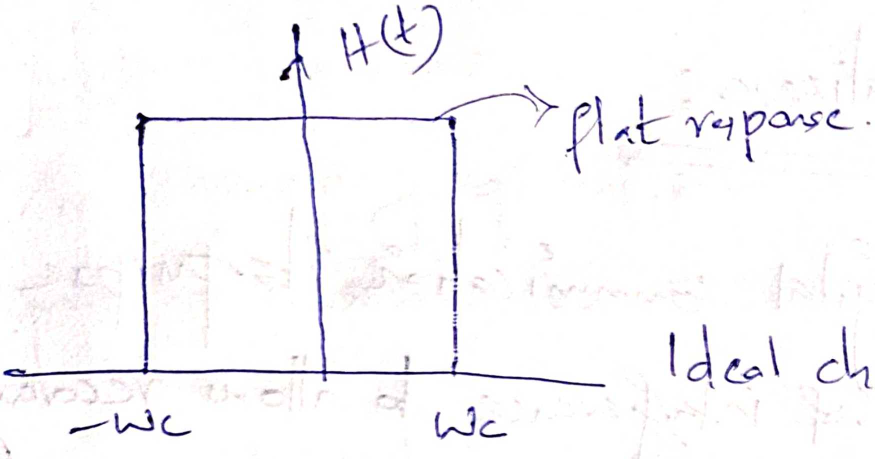
Structure of Regenerative Repeater:



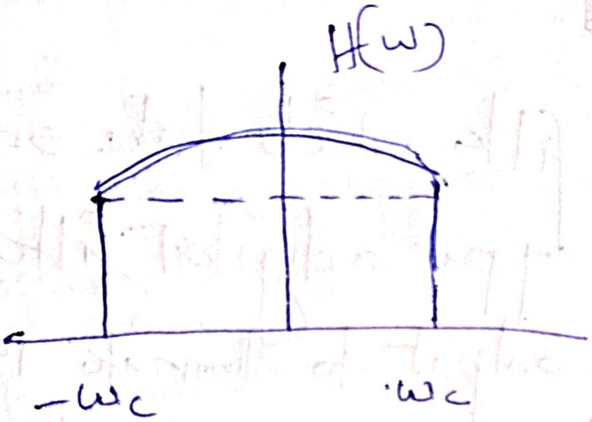
i) to know the start and end of the

ii) actual sampling instant

The signal that is coming along distance experiances attenuation noise and distortion, the attenuation can be comphensated using a amplifier while distortion can be compensated using a device called equalizer. This General form of distortion some freq are attenuated highly while rest are attenuated differently this changes the shape of pulse. To compensate this behavior of channel we use equlizer with freq response which is reciprocal to that of the channel as shown.

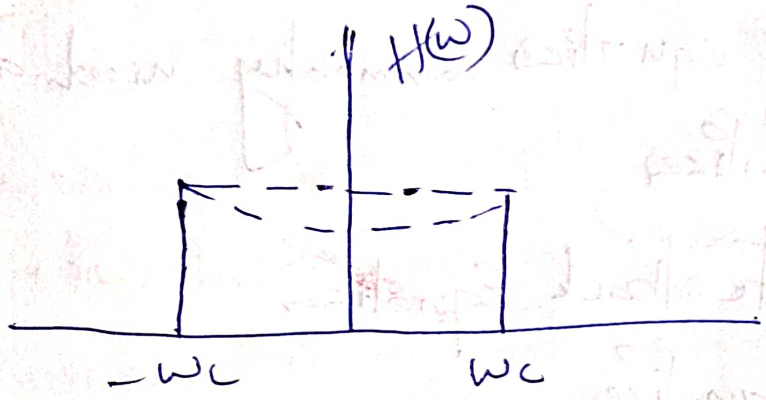


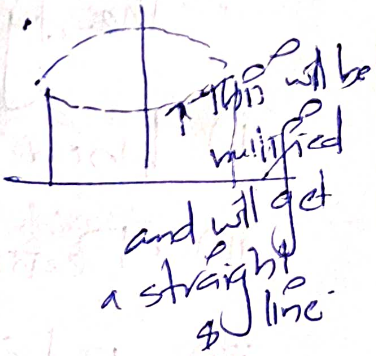
Ideal channel response



Practical channel response.

If the channel response is not a straight line the signal get distorted. ie, at $f=0$ gain is maximum as we move from 0 to wc the gain will decrease and it will change the property of the signal. This will cause distortion.





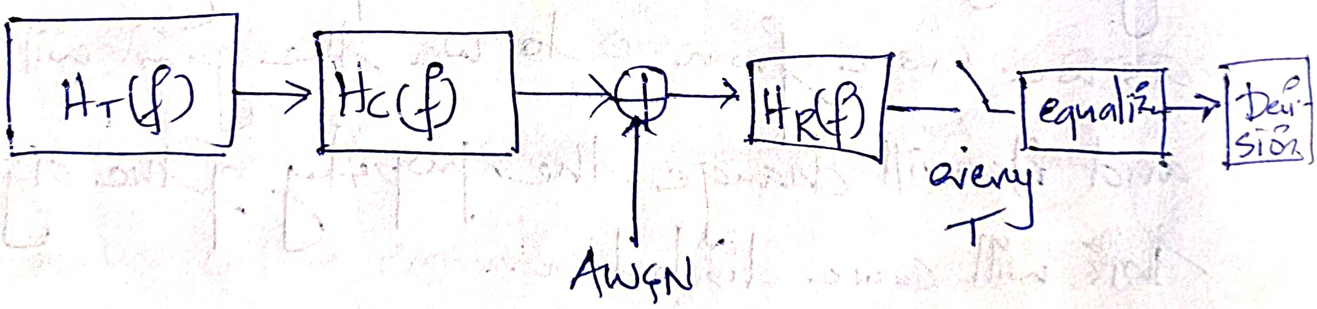
and will get a straight line.

Response of Equalizer.

Equalizers:

In digital communication, its purpose is to reduce Intersymbol interference to allow recovery of the transmit symbols.

We choose a filter which take samples at intervals T and ~~at~~ put a digital filter called equalizer at the output to eliminate ISI as shown. This approach to remove ISI is usually known as equalization.



The types of equalizer commonly used are.

i) Linear equalizer

ii) Decision Feedback Equalizer

iii) Blind Equalizer

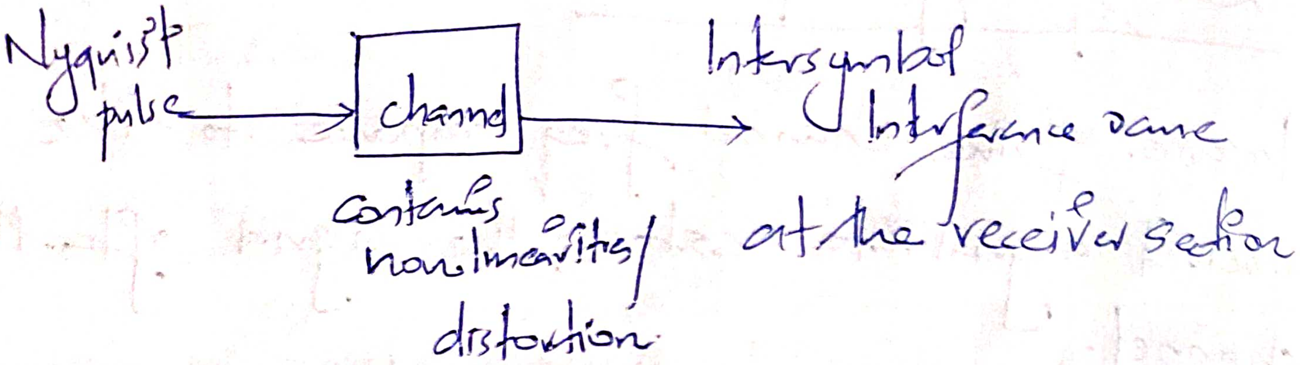
iv) Adaptive Equalizer

v) Zero forcing equalizer.

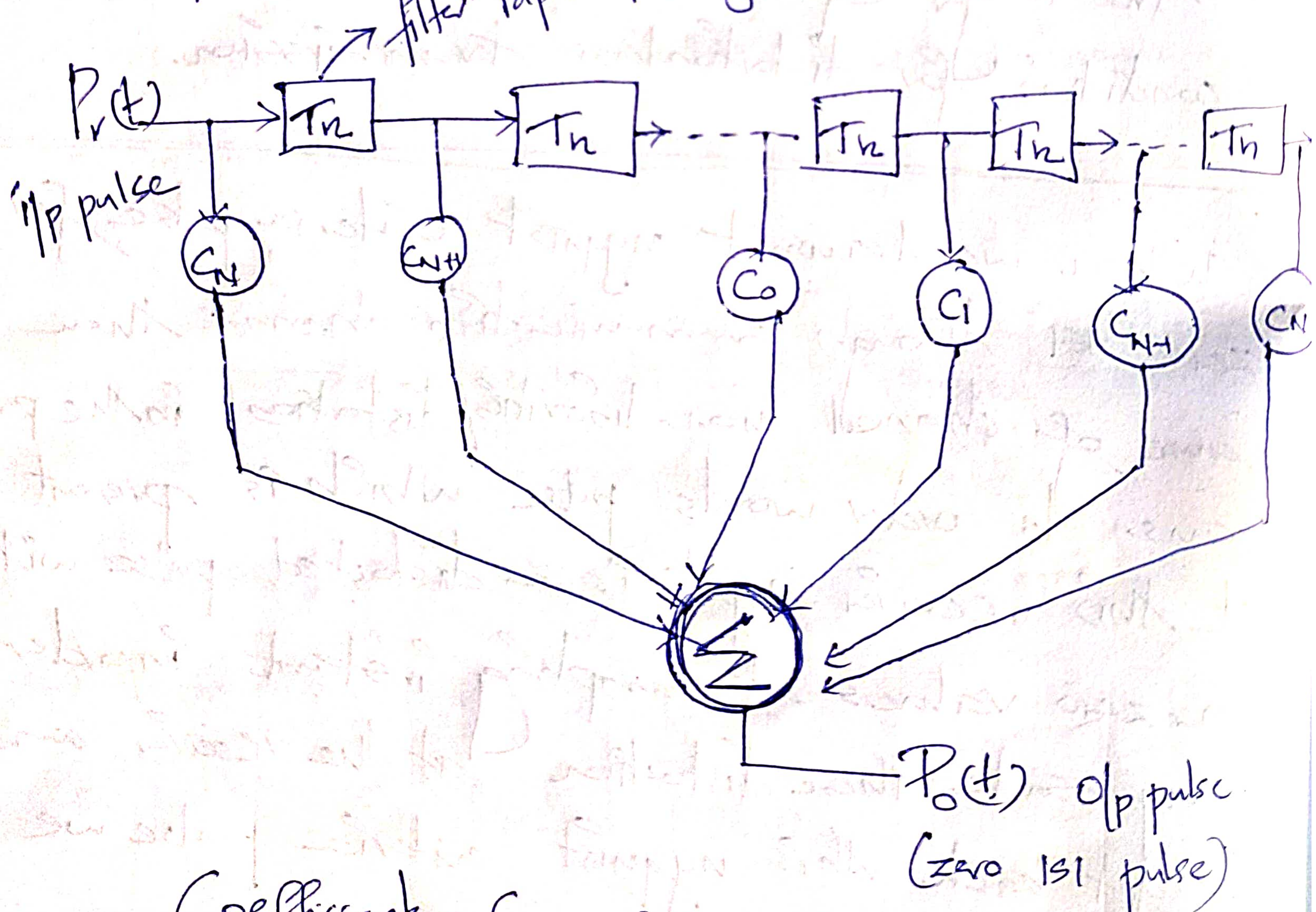
Zero forcing Equalizer:

- Inverse of the channel freq response to the received signal. to restore the signal after the channel.
- Forcing corresponds to bringing down the ISI to zero in a noise free case. This will be helpfull when ISI is significant compared to noise.
- Simple implementation
- The longer equalizer, the more the ideal conditions for distortionless transmission.

When we transmit nyquist criteria pulse for zero ISI through communication channel then because of channel non-linearities/distortions in the pulse occurs. In order words pulse which is present at this receiver input is a distorted pulse with non-zero values at sampling instants in order to compensate these distortion at the receiver and to get back this nyquist criteria pulse we use equalizer. One such equalizer is a zero forcing equalizer.



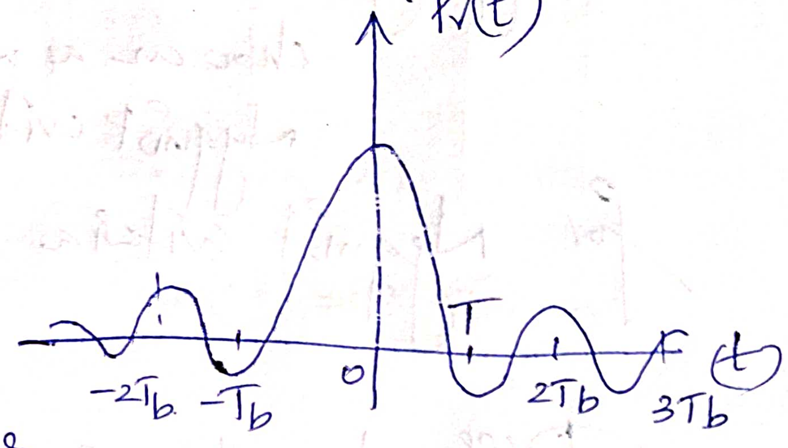
At transmitter have no ISI but at the received section will have ISI. In order to remove this ISI at the receiver section we use zero forcing equalizer.



Coefficients C_N C_{N+1} ---- $C_0, C_1,$

It is a types of filter, the no. of delay element 2N

And No: of coefficients are 2N+1.



at $T_b, 2T_b, 3T_b$ the values $\neq 0$ it is not a Nyquist criteria pulse

So we have to design a receiver section having zero ISI pulse.

$$P_o(t) = \sum_{n=-N}^{+N} C_n P_r(t - nT_b)$$

at sampling instant $t = kT_b$.

$$P_o(kT_b) = \sum_{n=-N}^{N} C_n P_r[kT_b - nT_b] \quad k = 0, \pm 1, \pm 2 \ldots$$

$$\boxed{P_o(k) = \sum_{n=-N}^{N} C_n P_r[k-n]} \text{ — } ①$$

$k = 0, \pm 1, \pm 2 \ldots$

$k \rightarrow$ is an integer

$$P_o(k) = \begin{matrix} 1, & k=0 \\ 0, & k \neq 0. \end{matrix}$$

Designing :- here the coefficients should be choosen as which satisfy the Nyquist criteria.

for Nyquist criteria $P_o(k) = 1 \quad , k=0$

$\qquad\qquad\qquad\qquad\quad 0 \quad k \neq 0$

Different values of k ie, $-N$ to N in equation ①

$$k=-N \} \quad P_o(-N) = C_{-N} P_r(0) + C_{-N+1} P_r(-1) - - - - + C_N P_v(-2N)$$

$$k=-N+1 \} \quad \underbrace{P_o(-N+1)}_{\text{design o/p}} = C_{-N} P_r(1) + C_{-N+1} P_r(0) + - - - - \underbrace{C_N}_{\text{coefficient}} P_v(-2N+1)$$

Pr (matrix)

$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \end{bmatrix}}_{(2N+1)\times 1} = \underbrace{\begin{bmatrix} P_v(0) & P_v(-1) & --- & P_v(-2N) \\ P_r(1) & P_r(0) & ---- & P_r(-2N+1) \\ \vdots & \vdots & & \\ P_r(2N) & --- & ---- & P_r(0) \end{bmatrix}}_{(2N+1)(2N+1)} \begin{bmatrix} C(-N) \\ C_{-N+1} \\ \vdots \\ C_0 \\ \vdots \\ C_N \end{bmatrix}$$

Q) Design a three tap zero equalizer if following parameter given

$P_r(0) = 1$ $P_r(2) = 0.1$

$P_r(-1) = 0.2$ $P_r(-2) = 0.05$

$\underbrace{P_r(1) = -0.3}$

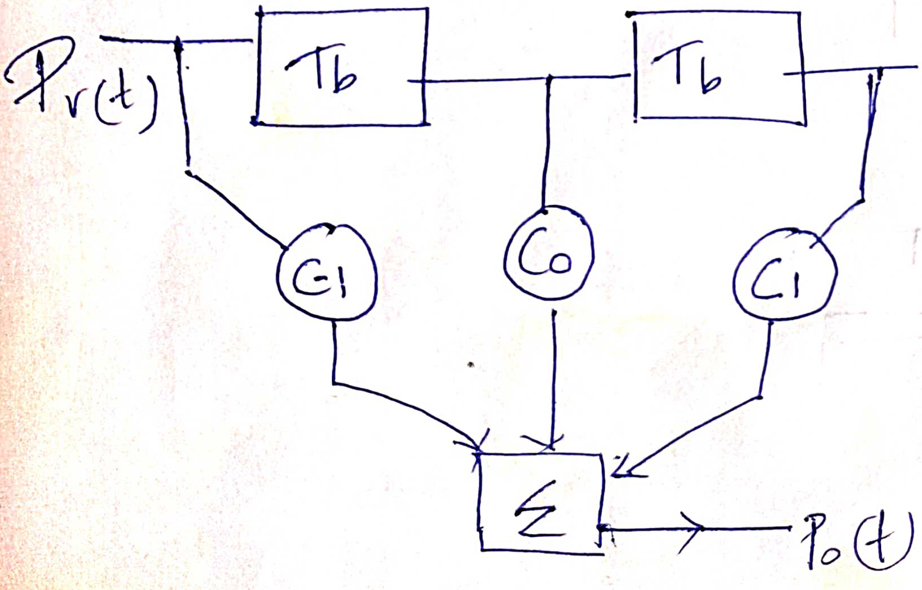
all are non-zero values ie; pulse received doesnot satisfy nyquist criteria.

Solution :

$2N + 1 = \underbrace{3}_{3\ tap}$

So; $N = 1$.

1st step : Structure of equilizer



$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} P_r(0) & P_r(-1) & P_r(-2) \\ P_r(1) & P_r(0) & P_r(-1) \\ P_r(2) & P_r(1) & P_r(0) \end{bmatrix} \begin{bmatrix} C_{-1} \\ 0 \\ C_1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & -0.2 & 0.05 \\ -0.3 & 1 & -0.2 \\ 0.1 & -0.3 & 1 \end{bmatrix} \begin{bmatrix} C_{-1} \\ C_0 \\ C_1 \end{bmatrix}$$

$$0 = C_{-1} - 0.2\,C_0 + 0.05\,C_1$$

$$1 = -0.3\,C_{-1} + C_0 - 0.2\,C_1$$

$$0 = 0.1\,C_{-1} - 0.3\,C_0 + C_1$$

$C_{-1} = 0.209$

$C_0 = 1.12$

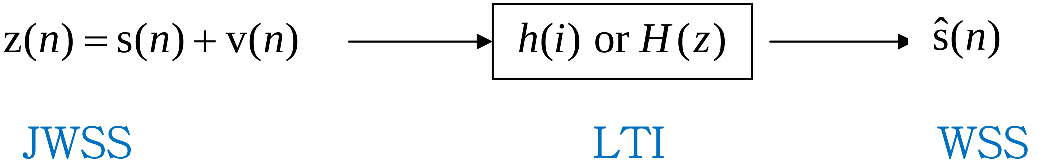
$C_1 = 0.31$

3. The Wiener Filter

3.1 The Wiener-Hopf Equation

The Wiener filter theory is characterized by:

1. The assumption that both signal and noise are random processed with known spectral characteristics or, equivalently, known auto- and cross-correlation functions.
2. The criterion for best performance is minimum mean-square error. (This is partially to make the problem mathematically tractable, but it is also a good physical criterion in many applications.)
3. A solution based on scalar methods that leads to the optimal filter weighting function (or transfer function in the stationary case).



R_s, R_v, R_z, R_{sz} or $S_s(z), S_v(z), S_z(z), S_{sz}(z)$ known

$$J = E\left\{\left[s(n) - \hat{s}(n)\right]^2\right\}$$

Fig. 3.1-1 Wiener Filter Problem

We now consider the filter optimization problem that Wiener first solved in the 1940s. Referring to Fig. 3.1-1, we assume the following:

1. The filter input is an additive combination of signal and noise, both of which are jointly wide-sense stationary (JWSS) with known auto- and cross-correlation functions (or corresponding spectral functions).
2. The filter is linear and time-invariant. No further assumption is made as to its form.
3. The output is also wide-sense stationary.
4. The performance criterion is minimum mean-square error.

The estimate $\hat{\mathrm{s}}(n)$ of a signal s(n) is given by the convolution representation

$$\hat{\mathrm{s}}(n) = h(n) * \mathrm{z}(n) = \sum_{i=-\infty}^{\infty} h(n-i)\mathrm{z}(i) = \sum_{i=-\infty}^{\infty} h(i)\mathrm{z}(n-i), \qquad (3.1\text{-}1)$$

where z(i) is the measurement and $h(n)$ is the impulse response of the estimator. Let $\mathcal{H}$ denote the region of support of $h(n)$, defined by

$$\mathcal{H} = \{n : h(n) \neq 0\}.$$

Then, Eq. (3.1-1) can be rewritten as

$$\hat{\mathrm{s}}(n) = \sum_{i \in \mathcal{H}} h(i)\mathrm{z}(n-i)\,. \tag{3.1-2}$$

Let the mean-square estimation error (MSE) J be

$$\begin{aligned} J &= E\left\{\left[\mathrm{s}(n) - \hat{\mathrm{s}}(n)\right]^2\right\} \\ &= E\left\{\left[\mathrm{s}(n) - \sum_{i \in \mathcal{H}} h(i)\mathrm{z}(n-i)\right]\left[\mathrm{s}(n) - \sum_{j \in \mathcal{H}} h(j)\mathrm{z}(n-j)\right]\right\} \\ &= E\left\{\mathrm{s}^2(n)\right\} - 2\sum_{i \in \mathcal{H}} h(i)E\left\{\mathrm{s}(n)\mathrm{z}(n-i)\right\} + \sum_{i \in \mathcal{H}}\sum_{j \in \mathcal{H}} h(i)h(j)E\left\{\mathrm{z}(n-i)\mathrm{z}(n-j)\right\}. \end{aligned} \tag{3.1-3}$$

To minimize the MSE, take the partial derivatives of J with respect to $h(i)$, for each $h(i) \neq 0$. Then, set the result equal to zero

$$\begin{aligned} \frac{\partial J}{\partial h(i)} &= -2E\left\{\mathrm{s}(n)\mathrm{z}(n-i)\right\} + 2\sum_{j \in \mathcal{H}} h(j)E\left\{\mathrm{z}(n-i)\mathrm{z}(n-j)\right\} \\ &= 0. \end{aligned} \tag{3.1-4}$$

Solving Eq. (3.1-4), we find

$$\sum_{j \in \mathcal{H}} h(j)E\left\{\mathrm{z}(n-i)\mathrm{z}(n-j)\right\} = E\left\{\mathrm{s}(n)\mathrm{z}(n-i)\right\}, \quad i \in \mathcal{H} \tag{3.1-5}$$

which we may express in the form of

$$\sum_{j \in \mathcal{H}} h(j) R_{\mathrm{z}}(j-i) = R_{\mathrm{sz}}(i), \quad i \in \mathcal{H}\ , \tag{3.1-6}$$

where $R_{\mathrm{z}}(k)$ is the autocorrelation function of z(n) and $R_{\mathrm{sz}}(k)$ is the crosscorrelation function of s(n) and z(n).

Eq. (3.1-6) is the discrete-time ***Wiener-Hopf equation***. It is the basis for the derivation of the Wiener filter.

3.2 The FIR Wiener Filter

Suppose the filter has an impulse response $h(n)$ with support $\mathcal{H}$, where $\mathcal{H}$ is a finite set, e.g., $\mathcal{H}$ = {0,1,2,…,N-1}. Then the impulse response $h(n)$ has a finite duration, and this type of filter is called a *finite impulse response* (FIR) filter.

For the FIR filter, the Wiener-Hopf equation is written

$$\sum_{j=0}^{N-1} h(j) R_{\mathrm{z}}(j-i) = R_{\mathrm{sz}}(i), \quad 0 \le i \le N-1. \tag{3.2-1}$$

This may be written in matrix form

$$\begin{bmatrix} R_{\mathrm{z}}(0) & R_{\mathrm{z}}(1) & \cdots & R_{\mathrm{z}}(N-1) \\ R_{\mathrm{z}}(1) & R_{\mathrm{z}}(0) & \cdots & R_{\mathrm{z}}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_{\mathrm{z}}(N-1) & R_{\mathrm{z}}(N-2) & \cdots & R_{\mathrm{z}}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix} = \begin{bmatrix} R_{\mathrm{sz}}(0) \\ R_{\mathrm{sz}}(1) \\ \vdots \\ R_{\mathrm{sz}}(N-1) \end{bmatrix}, \tag{3.2-2}$$

where we have used the fact that $R_{\mathrm{z}}(k) = R_{\mathrm{z}}(-k)$. Denote Eq. (3.2-2) in a convenient form

$$R_{\mathrm{z}}\underline{h} = \underline{r}_{\mathrm{sz}} \tag{3.2-3}$$

where the autocorrelation function R_{z} is symmetric and positive-definite.

Eq. (3.2-3) is solved for $\underline{h}$,

$$\underline{h} = R_{\mathrm{z}}^{-1}\underline{r}_{\mathrm{sz}}. \tag{3.2-4}$$

This is the FIR Wiener filter of order $N-1$. Note that R_{z} is a Toeplitz matrix and Eq. (3.2-4) may be solved using a computationally-efficient method such as the Levinson-Durbin algorithm.

Mean-Square Error (MSE)

We may write the MSE of the FIR Wiener filter

$$
\begin{aligned}
\text{MSE} &= E\left\{\tilde{\mathrm{s}}(n)\left[\mathrm{s}(n)-\sum_{i=0}^{N-1}h(i)\mathrm{z}(n-i)\right]\right\} \\
&= E\left\{\tilde{\mathrm{s}}(n)\mathrm{s}(n)\right\}-\sum_{i=0}^{N-1}h(i)E\left\{\tilde{\mathrm{s}}(n)\mathrm{z}(n-i)\right\}.
\end{aligned}
\tag{3.2-5}
$$

To simplify Eq. (3.2-5), rewrite Eq. (3.1-5) in the following form

$$
\begin{aligned}
E\left\{\mathrm{z}(n-i)\left[\mathrm{s}(n)-\sum_{j\in\mathcal{H}}h(j)\mathrm{z}(n-j)\right]\right\} &= 0 \\
E\left\{\mathrm{z}(n-i)\left[\mathrm{s}(n)-\hat{\mathrm{s}}(n)\right]\right\} &= 0 \\
E\left\{\mathrm{z}(n-i)\tilde{\mathrm{s}}(n)\right\} &= 0, \quad i\in\mathcal{H}.
\end{aligned}
\tag{3.2-6}
$$

This relation is known as the orthogonality principle for LTI estimators. Applying this relation to Eq. (3.2-5) gives

$$
\begin{aligned}
\mathrm{MSE} &= E\left\{\tilde{\mathrm{s}}(n)\mathrm{s}(n)\right\} \\
&= E\left\{\mathrm{s}^2(n)\right\} - \sum_{i=0}^{N-1} h(i) E\left\{\mathrm{s}(n)\mathrm{z}(n-i)\right\} \\
&= R_{\mathrm{s}}(0) - \sum_{i=0}^{N-1} h(i) R_{\mathrm{sz}}(i) \\
&= R_{\mathrm{s}}(0) - h^T \underline{r}_{\mathrm{sz}}.
\end{aligned}
\tag{3.2-7}
$$

Observe that if no filtering is performed and we simply use $\hat{\mathrm{s}}(n) = z(n)$, the MSE is

$$
\mathrm{MSE}_{\text{no filter}} = E\left\{\left[\mathrm{s}(n) - z(n)\right]^2\right\} = R_{\mathrm{s}}(0) - 2R_{\mathrm{sz}}(0) + R_{\mathrm{z}}(0).
\tag{3.2-8}
$$

We can use Eq. (3.2-8) to measure the effectiveness of the Wiener filter. The reduction in MSE due to Wiener filtering is given by

$$
\text{reduction in MSE} = 10\log_{10}\left(\frac{\mathrm{MSE}_{\text{no filter}}}{\mathrm{MSE}_{\text{filter}}}\right) dB.
\tag{3.2-9}
$$

We can inspect the MSE to determine whether the filter performance is acceptable. If the MSE is too high, we may wish to use a larger value of N.

Example 3.2-1

Suppose we have a signal $\mathrm{s}(n)$ with autocorrelation function $R_{\mathrm{s}}(k) = 0.95^{|k|}$. The signal is observed in the presence of additive white noise with variance $\sigma_{\mathrm{v}}{}^2 = 2$. Hence $R_{\mathrm{v}}(k) = 2\delta(k)$. $\mathrm{s}(n)$ and $\mathrm{v}(n)$ are uncorrelated, zero-mean, JWSS random processes. We would like to design the optimum LTI second-order filter *h*.

Since the filter is the second-order, we set $N = 3$. The matrix equation Eq. (3.2-2) to be solved is

$$\begin{bmatrix} R_{\mathrm{z}}(0) & R_{\mathrm{z}}(1) & R_{\mathrm{z}}(2) \\ R_{\mathrm{z}}(1) & R_{\mathrm{z}}(0) & R_{\mathrm{z}}(1) \\ R_{\mathrm{z}}(2) & R_{\mathrm{z}}(1) & R_{\mathrm{z}}(0) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ h(2) \end{bmatrix} = \begin{bmatrix} R_{\mathrm{sz}}(0) \\ R_{\mathrm{sz}}(1) \\ R_{\mathrm{sz}}(2) \end{bmatrix}. \tag{3.2-10}$$

For an additive-noise measurement model

$$\mathrm{z}(n) = \mathrm{s}(n) + \mathrm{v}(n),$$

the autocorrelation functions $R_{\mathrm{z}}(k)$ and $R_{\mathrm{sz}}(k)$ may be given as follows

$$\begin{aligned} R_{\mathrm{z}}(k) &= E\left\{\mathrm{z}(n)\mathrm{z}(n-k)\right\} \\ &= E\left\{\left[\mathrm{s}(n)+\mathrm{v}(n)\right]\left[\mathrm{s}(n-k)+\mathrm{v}(n-k)\right]\right\} \\ &= E\left\{\left[\mathrm{s}(n)\mathrm{s}(n-k)\right]\right\}+E\left\{\left[\mathrm{v}(n)\mathrm{v}(n-k)\right]\right\} \\ &= R_{\mathrm{s}}(k)+R_{\mathrm{v}}(k), \end{aligned} \tag{3.2-11}$$

and

$$\begin{aligned} R_{\mathrm{sz}}(k) &= E\left\{\mathrm{s}(n)\left[\mathrm{s}(n-k)+\mathrm{v}(n-k)\right]\right\} \\ &= E\left\{\left[\mathrm{s}(n)\mathrm{s}(n-k)\right]\right\}+E\left\{\left[\mathrm{s}(n)\mathrm{v}(n-k)\right]\right\} \\ &= R_{\mathrm{s}}(k). \end{aligned} \tag{3.2-12}$$

According to Eqs. (3.2-11) and (3.2-12), we have

$$\begin{aligned} R_{\mathrm{z}}(k) &= 0.95^{|k|}+2\delta(k) \\ R_{\mathrm{sz}}(k) &= 0.95^{|k|}. \end{aligned}$$

Then Eq. (3.2-10) becomes

$$\begin{bmatrix} 3 & 0.95 & 0.9025 \\ 0.95 & 3 & 0.95 \\ 0.9025 & 0.95 & 3 \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ h(2) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.95 \\ 0.9025 \end{bmatrix}. \tag{3.2-13}$$

The solution of Eq. (3.2-13) is

$$\underline{h} = \left[0.2203 \ \ 0.1919 \ \ 0.1738\right]^T .$$

The MSE of this estimator can be computed via Eq. (3.2-7),

$$\text{MSE} = R_s(0) - h^T \underline{r}_{sz} = 0.4405 . \tag{3.2-13}$$

For comparison, we remark that without any filtering,

$$\text{MSE}_{\text{no filter}} = R_s(0) - 2R_{sz}(0) + R_z(0) = R_v(0) = 2 . \tag{3.2-14}$$

The FIR Wiener filter has reduced the MSE by

$$\begin{aligned} \text{reduction in MSE} &= 10\log_{10}\left(\frac{\text{MSE}_{\text{no filter}}}{\text{MSE}_{\text{filter}}}\right) dB \\ &= 10\log_{10}\left(\frac{2}{0.4405}\right) \qquad (3.2\text{-}15) \\ &\approx 6.5 dB. \end{aligned}$$

3.3 The Noncausal Wiener Filter

Rewrite the Wiener-Hopf equation, Eq. (3.1-6),

$$\sum_{j\in\mathcal{H}} h(j)R_{\mathrm{z}}(j-i)=R_{\mathrm{sz}}(i), \quad i\in\mathcal{H}$$

and let $\mathcal{H}=\mathbb{Z}$, where $\mathbb{Z}$ is the set of integers: $\mathbb{Z}=\{\cdots,-2,-1,0,1,2,\cdots\}$. Then, the Wiener-Hopf equation becomes

$$\sum_{j=-\infty}^{\infty} h(j)R_{\mathrm{z}}(j-i)=R_{\mathrm{sz}}(i), \quad \text{for all } i\in\mathbb{Z}. \qquad (3.3\text{-}1)$$

Take the *z*-transform of Eq. (3.3-1)

$$H(z)S_{\mathrm{z}}(z)=S_{\mathrm{sz}}(z)$$

so that

$$H(z)=\frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}(z)}. \qquad (3.3\text{-}2)$$

Eq. (3.3-2) is the solution for the noncausal Wiener filter.

Mean-Square Error (MSE)

As with the FIR Wiener filter of Section 3.2, we can derive an expression for the MSE. The MSE similar to Eq. (3.2-7) can be written as

$$\mathrm{MSE} = R_{\mathrm{s}}(0) - \sum_{i=-\infty}^{\infty} h(i) R_{\mathrm{sz}}(i)\,. \tag{3.3-3}$$

Example 3.3-1

Let us apply a noncausal Wiener filter to the filtering problem of Example 3.2-1. In that example we found that

$$R_{\mathrm{z}}(k) = 0.95^{|k|} + 2\delta(k),\ \text{and}\ \ R_{\mathrm{sz}}(k) = 0.95^{|k|}\,. \tag{3.3-4}$$

Take *z*-transforms of Eq. (3.3-4)

$$\begin{aligned} S_{\mathrm{z}}(z) &= \frac{1-(0.95)^2}{(1-0.95z^{-1})(1-0.95z)} + 2 \\ &= \frac{2.3955(1-0.7931z^{-1})(1-0.7931z)}{(1-0.95z^{-1})(1-0.95z)}, \quad 0.7931 < |z| < \frac{1}{0.7931} \end{aligned}$$

and

$$S_{\mathrm{sz}}(z) = \frac{1-(0.95)^2}{(1-0.95z^{-1})(1-0.95z)}, \quad 0.95 < |z| < \frac{1}{0.95}.$$

The noncausal Wiener filter, Eq. (3.3-2), is given by

$$H(z) = \frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}(z)} = \frac{0.0975}{2.3955(1-0.7931z^{-1})(1-0.7931z)}$$
$$= \frac{0.1097(1-(0.7931)^2)}{(1-0.7931z^{-1})(1-0.7931z)}, \quad 0.95 < |z| < \frac{1}{0.95}.$$

The impulse response of the filter is

$$h(n) = 0.1097(0.7931)^{|n|}.$$

The mean-square errors associated with the noncausal Wiener filter is according to Eq. (3.3-3),

$$
\begin{aligned}
\text{MSE} &= R_s(0) - \sum_{n=-\infty}^{\infty} h(n) R_{sz}(n) \\
&= (0.95)^0 - \sum_{n=-\infty}^{\infty} 0.1097(0.7931)^{|n|}(0.95)^{|n|} \\
&= 1 - 0.1097\left[\sum_{n=0}^{\infty}(0.7931)^n(0.95)^n + \sum_{n=-1}^{-\infty}(0.7931)^{-n}(0.95)^{-n}\right] \\
&= 1 - 0.1097\left[2\sum_{n=0}^{\infty}(0.7931)^n(0.95)^n - 1\right] \\
&= 1 - 0.1097\left[\frac{2}{1-(0.7931)(0.95)} - 1\right] = 0.2195.
\end{aligned}
$$

Since the MSE with no filter is 2 according to Eq. (3.2-14), the improvement of the noncausal filter over no filtering is

$$
\begin{aligned}
\text{reduction in MSE} &= 10\log_{10}\left(\frac{\text{MSE}_{\text{no filter}}}{\text{MSE}_{\text{filter}}}\right) dB \\
&= 10\log_{10}\left(\frac{2}{0.2195}\right) \\
&\approx 9.6 dB.
\end{aligned}
$$

Compared to the second-order filter in Eq. (3.2-15) in Ex. (3.2-1), the noncausal Wiener filter reduces the estimation error by an additional 3 dB.

3.4 The Causal Wiener Filter

Rewrite the Wiener-Hopf equation, Eq. (3.1-6),

$$\sum_{j \in \mathcal{H}} h(j) R_{\mathrm{z}}(j-i) = R_{\mathrm{sz}}(i), \quad i \in \mathcal{H}$$

and let $\mathcal{H} = \mathbb{N}$, where $\mathbb{N}$ is the set of integers: $\mathbb{N} = \{0, 1, 2, \cdots\}$. Then, the Wiener-Hopf equation becomes

$$\sum_{j=0}^{\infty} h(j) R_{\mathrm{z}}(j-i) = R_{\mathrm{sz}}(i), \quad \text{for all } i \in \mathbb{N}. \tag{3.4-1}$$

Eq. (3.4-1) can not be simplified by taking the z-transform because of its causality restriction. To solve the Wiener-Hopf equation, the spectral factorization and the causal-part extraction are necessary.

Theorem 3.4-1 (Spectral Factorization)

Let $\mathrm{x}(n)$ be a real-valued, zero-mean, WSS random process with power density spectrum $S_{\mathrm{x}}(z)$, where $S_{\mathrm{x}}(z)$ is rational in *z* and has no poles on the unit circle. Then $S_{\mathrm{x}}(z)$ can be factored into the product

$$S_{\mathrm{x}}(z) = S_{\mathrm{x}}^{+}(z) S_{\mathrm{x}}^{-}(z), \qquad (3.4\text{-}2)$$

where

$S_{\mathrm{x}}^{+}(z)$ and $S_{\mathrm{x}}^{-}(z)$ are rational in *z*,
if z_i is a pole of $S_{\mathrm{x}}^{+}(z)$, then $|z_i| < 1$,
if z_i is a zero of $S_{\mathrm{x}}^{+}(z)$, then $|z_i| \leq 1$,
if z_i is a pole of $S_{\mathrm{x}}^{-}(z)$, then $|z_i| > 1$,
if z_i is a zero of $S_{\mathrm{x}}^{-}(z)$, then $|z_i| \geq 1$, and
$S_{\mathrm{x}}^{+}(z) = S_{\mathrm{x}}^{-}(z^{-1})$.

Example 3.4-1

Let $\mathrm{s}(n)$ be a random process described by the difference equation

$$\mathrm{s}(n) = 1.1\mathrm{s}(n-1) - 0.24\mathrm{s}(n-2) + 2\mathrm{w}(n) + 3\mathrm{w}(n-1), \qquad (3.4\text{-}3)$$

where w(*n*) is a zero-mean, WSS random process with autocorrelation function $R_{\mathrm{w}}(n) = 5(0.6)^{|n|}$.

We want to find the spectral factorization of $S_{\text{s}}(z)$.

Take the *z*-transform of Eq. (3.4-3)

$$S(z) = 1.1z^{-1}S(z) - 0.24z^{-2}S(z) + 2W(z) + 3z^{-1}W(z).$$

The system transfer function is

$$H(z) = \frac{S(z)}{W(z)} = \frac{2 + 3z^{-1}}{1 - 1.1z^{-1} + 0.24z^{-2}} = \frac{2(1 + 1.5z^{-1})}{(1 - 0.3z^{-1})(1 - 0.8z^{-1})}.$$

Using the input-output power spectrum relation, the power spectrum of s(*n*) is given by

$$\begin{aligned} S_{\text{s}}(z) &= H(z)H(z^{-1})S_{\text{w}}(z) \\ &= \frac{2(1 + 1.5z^{-1})}{(1 - 0.3z^{-1})(1 - 0.8z^{-1})} \frac{2(1 + 1.5z)}{(1 - 0.3z)(1 - 0.8z)} \frac{3.2}{(1 - 0.6z^{-1})(1 - 0.6z)}, \end{aligned} \tag{3.4-4}$$

where it is used that

$$S_{\text{w}}(z) = 5\frac{1 - (0.6)^2}{(1 - 0.6z^{-1})(1 - 0.6z)}.$$

Collect the terms in Eq. (3.4-4) that have poles or zeros inside the unit circle to form $S_{\mathrm{s}}^{+}(z)$,

$$S_{\mathrm{s}}^{+}(z)=\frac{2\sqrt{3.2}(1+1.5z)}{(1-0.3z^{-1})(1-0.8z^{-1})(1-0.6z^{-1})} \quad . \tag{3.4-5}$$

Collect the terms in Eq. (3.4-4) that have poles or zeros outside the unit circle to form $S_{\mathrm{s}}^{-}(z)$,

$$S_{\mathrm{s}}^{-}(z)=\frac{2\sqrt{3.2}(1+1.5z^{-1})}{(1-0.3z)(1-0.8z)(1-0.6z)} \quad . \tag{3.4-6}$$

From Eqs. (3.4-5) and (3.4-6), it is noticed that

$$S_{\mathrm{s}}^{-}(z)=S_{\mathrm{s}}^{+}(z^{-1}) .$$

The Causal-Part Extraction

Consider the impulse response $h(n)$ of a real-valued LTI system with rational *z*-transform $H(z)$. In the time domain we can split $h(n)$ into its causal and anticausal parts such that

$$h(n)=\text{causal part of } h(n)+\text{anticausal part of } h(n), \tag{3.4-7}$$

where

$$\text{causal part of } h(n) \equiv \left[h(n)\right]_+ = h(n)1(n),$$
$$\text{anticausal part of } h(n) \equiv \left[h(n)\right]_- = h(n)1(-n-1).$$

By the linearity of the *z*-transform, Eq. (3.4-7) becomes,

$$H(z) = \left[H(z)\right]_+ + \left[H(z)\right]_-, \tag{3.4-8}$$

where

$$\begin{aligned} \left[H(z)\right]_+ &= \mathcal{Z}\left\{h(n)1(n)\right\}, \\ \left[H(z)\right]_- &= \mathcal{Z}\left\{h(n)1(-n-1)\right\}. \end{aligned} \tag{3.4-9}$$

We now develop a method for determining $\left[H(z)\right]_+$ and $\left[H(z)\right]_-$. By long division in *z*, z^{-1}, or both for a rational $H(z)$, we can always convert it into the form

$$H(z) = \underbrace{\sum_{n=1}^{L} c_{-n} z^{n}}_{P_A(z)} + \underbrace{\sum_{n=0}^{M-N} c_n z^{-n}}_{P_C(z)} + \underbrace{\left(\frac{\sum_{n=0}^{N-1} b_n z^{-n}}{\sum_{n=0}^{N} a_n z^{-n}} \right)}_{Q(z)} \tag{3.4-10}$$

$H(z)$ consists of two polynomial terms, $P_A(z)$ and $P_C(z)$, and a proper fraction in z^{-1}, $Q(z)$. $P_A(z)$ corresponds to a purely anticausal sequence, and $P_C(z)$ to a purely causal sequence.

Now we examine $Q(z)$. Let $K \le N$ be the number of distinct poles of $Q(z)$, and denote these poles and their associated degrees by p_k and m_k, respectively, $k = 1, \cdots, K$. Then $Q(z)$ may be written as

$$Q(z) = \left(\frac{1}{a_0} \right) \frac{\sum_{n=0}^{N-1} b_n z^{-n}}{\prod_{k=1}^{K} (1 - p_k z^{-1})^{m_k}}.$$

Via partial fractions, we can write this expression as

$$Q(z) = \sum_{k=1}^{K} \sum_{m=1}^{m_k} \frac{q_{k,m}}{(1 - p_k z^{-1})^m} \equiv \sum_{k=1}^{K} Q_k(z). \tag{3.4-11}$$

Suppose that $H(z)$ be stable. Then its ROC includes the unit circle and the ROC of each $Q_k(z)$ must include the unit circle. Let N_A be the total order of the poles of $H(z)$ outside the unit circle, and N_C be the total order of the poles of $H(z)$ inside the unit circle with the exception of the origin. Then $N_A + N_C = N$ and we may write

$$Q_A(z) = \sum_{\substack{k=1 \\ |p_k|>1}}^{K} Q_k(z) = \frac{\sum_{n=0}^{N_A-1} \lambda_n z^{-n}}{\sum_{\substack{k=1 \\ |p_k|>1}}^{K} (1 - p_k z^{-1})^{m_k}},$$

and

$$Q_C(z) = \sum_{\substack{k=1 \\ 0<|p_k|<1}}^{K} Q_k(z) = \frac{\sum_{n=0}^{N_C-1} \mu_n z^{-n}}{\sum_{\substack{k=1 \\ 0<|p_k|<1}}^{K} (1 - p_k z^{-1})^{m_k}}.$$

$Q_A(z)$ is the anticausal, proper, rational portion of $H(z)$, and $Q_C(z)$ is the causal, proper, rational portion of $H(z)$. Then for a stable rational $H(z)$, we have

$$\left[H(z)\right]_+ = P_C(z) + Q_C(z),$$

which is the sum of a purely causal sequence and all terms in the partial fraction expansion of

$H(z)$ that have poles inside the unit circle.

Example 3.4-2

We want the causal part of

$$H(z) = \frac{6z^2 - 51z + 128 - 109z^{-1} + 197z^{-2} - 232z^{-3} + 80z^{-4}}{2 - 17z^{-1} + 40z^{-2} - 16z^{-3}},$$

with $\text{ROC} = \left\{ z : \frac{1}{2} < |z| < 4 \right\}$.

Long division in *z* gives

$$
\begin{array}{r|l}
 & 3z^2 \qquad +4 \\ \cline{2-2}
2-17z^{-1}+40z^{-2}-16z^{-3} & 6z^2-51z+128-109z^{-1}+197z^{-2}-232z^{-3}+80z^{-4} \\
 & 6z^2-51z+120-\ 48z^{-1} \\ \cline{2-2}
 & \qquad 8-\ 61z^{-1}+197z^{-2}-232z^{-3}+80z^{-4} \\
 & \qquad 8-\ 68z^{-1}+160z^{-2}-\ \ 64z^{-3} \\ \cline{2-2}
 & \qquad\quad 7z^{-1}+\ 37z^{-2}-168z^{-3}+80z^{-4}
\end{array}
$$

$$H(z)=3z^2+4+\frac{7z^{-1}+\ 37z^{-2}-168z^{-3}+80z^{-4}}{2-17z^{-1}+40z^{-2}-16z^{-3}}.$$

Then long division in z^{-1} produces

$$\begin{array}{r r}
 & -5z^{-1} \quad -2 \qquad\qquad\qquad\qquad \\
\cline{2-2}
-16z^{-3}+40z^{-2}-17z^{-1}+2 \big) & 80z^{-4}-168z^{-3}+37z^{-2}+\ 7z^{-1} \quad \\
 & 80z^{-4}-200z^{-3}+85z^{-2}-10z^{-1} \quad \\
\cline{2-2}
 & 32z^{-3}-48z^{-2}+17z^{-1} \qquad \\
 & 32z^{-3}-80z^{-2}+34z^{-1}-4 \\
\cline{2-2}
 & 32z^{-2}-17z^{-1}+4
\end{array}$$

$$H(z) = 3z^2 + 4 + (-2) - 5z^{-1} + \frac{4-17z^{-1}+\ 32z^{-2}}{2-17z^{-1}+40z^{-2}-16z^{-3}}.$$

Let $Q(z)$ be the rational term and find its partial fractions

$$Q(z) = \frac{4-17z^{-1}+\ 32z^{-2}}{2-17z^{-1}+40z^{-2}-16z^{-3}} = \frac{2-\frac{17}{2}z^{-1}+16z^{-2}}{(1-\frac{1}{2}z^{-1})(1-4z^{-1})^2}$$

$$= \frac{1}{(1-4z^{-1})^2} + \frac{1}{1-\frac{1}{2}z^{-1}}.$$

Hence,

$$H(z) = 3z^2 + 2 - 5z^{-1} + \frac{1}{(1-4z^{-1})^2} + \frac{1}{1-\frac{1}{2}z^{-1}}.$$

So the causal part of $H(z)$ is

$$\left[H(z)\right]_+ = 2 - 5z^{-1} + \frac{1}{1-\frac{1}{2}z^{-1}},$$

which corresponds to the sequence $2\delta(n) - 5\delta(n-1) + \left(\frac{1}{2}\right)^n 1(n)$. Also the anticausal part of $H(z)$ is

$$\left[H(z)\right]_- = 3z^2 + \frac{1}{(1-4z^{-1})^2},$$

which corresponds to $3\delta(n+2) + \left[(-4)^n 1(-n-1)\right] * \left[(-4)^n 1(-n-1)\right]$.

The Causal Wiener Filter

Assume that $S_{\mathrm{z}}(z)$ is rational in *z* and has no poles or zeros on the unit circle. The equation that we must solve to obtain the causal Wiener filter is as is given in Eq. (3.4-1)

$$\sum_{j=0}^{\infty} h(j) R_{\mathrm{z}}(j-i) = R_{\mathrm{sz}}(i), \quad i \geq 0 .$$

To solve the Wiener-Hopf equation, define

$$h'(i) = R_{\mathrm{sz}}(i) - \sum_{j=0}^{\infty} h(j) R_{\mathrm{z}}(j-i), \quad \text{for all } i \in \mathbb{Z} . \tag{3.4-12}$$

Since *h* is the causal Wiener filter, $h'(i)$ can be written

$$h'(i) = \begin{cases} 0, & i \geq 0 \\ R_{\mathrm{sz}}(i) - \sum_{j=-\infty}^{\infty} h(j) R_{\mathrm{z}}(j-i), & i < 0. \end{cases} \tag{3.4-13}$$

Now we can take the *z*-transform of both sides of Eq. (3.4-12) and obtain

$$\begin{aligned} H'(z) &= S_{\mathrm{sz}}(z) - H(z)S_{\mathrm{z}}(z) \\ &= S_{\mathrm{sz}}(z) - H(z)S_{\mathrm{z}}^{+}(z)S_{\mathrm{z}}^{-}(z). \end{aligned}$$

Dividing by $S_{\mathrm{z}}^{-}(z)$, we have

$$\frac{H'(z)}{S_{\mathrm{z}}^{-}(z)} = \frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}^{-}(z)} - H(z)S_{\mathrm{z}}^{+}(z).$$

Extract the causal part of both sides of this equation and find

$$\left[\frac{H'(z)}{S_{\mathrm{z}}^{-}(z)}\right]_{+} = \left[\frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}^{-}(z)}\right]_{+} - \left[H(z)S_{\mathrm{z}}^{+}(z)\right]_{+}. \qquad (3.4\text{-}14)$$

From Eq. (3.4-13), $h'(i)$ is purely anticausal. Therefore $H'(z)$ contains only poles outside the unit circle. Since the zeros of $S_{\mathrm{z}}^{-}(z)$ lie outside the unit circle, the poles of $\dfrac{1}{S_{\mathrm{z}}^{-}(z)}$ also lie outside the unit circle. We conclude that all the poles of $\dfrac{H'(z)}{S_{\mathrm{z}}^{-}(z)}$ are outside the unit circle, and thus

$$\left[\frac{H'(z)}{S_z^-(z)}\right]_+ = 0.$$

$H(z)$ is causal by definition, so its poles lie within the unit circle. The poles of $S_z^+(z)$ are also within the unit circle. Hence all the poles of $H(z)S_z^+(z)$ are inside the unit circle, and

$$\left[H(z)S_z^+(z)\right]_+ = H(z)S_z^+(z).$$

We can not conclude anything about $\dfrac{S_{sz}(z)}{S_z^-(z)}$. Therefore, Eq. (3.4-14) becomes

$$0 = \left[\frac{S_{sz}(z)}{S_z^-(z)}\right]_+ - H(z)S_z^+(z),$$

which produces the causal Wiener filter, described by its system function $H(z)$

$$H(z) = \frac{1}{S_z^+(z)}\left[\frac{S_{sz}(z)}{S_z^-(z)}\right]_+. \qquad (3.4\text{-}15)$$

<u>Example 3.4-3</u>

Consider the same situation of Examples 3.2-1 and 3.3-1 in which s(n) and v(n) are uncorrelated, zero-mean, JWSS random processes with

$$S_{\mathrm{s}}(z) = \frac{1-(0.95)^2}{(1-0.95z^{-1})(1-0.95z)},$$

and

$$S_{\mathrm{v}}(z) = 2.$$

We observe $\mathrm{z}(n) = \mathrm{s}(n) + \mathrm{v}(n)$ and estimate $\mathrm{s}(n)$.

From Example 3.3-1, we have

$$S_{\mathrm{z}}(z) = \frac{2.3955(1-0.7931z^{-1})(1-0.7931z)}{(1-0.95z^{-1})(1-0.95z)}.$$

Perform a spectral factorization on $S_{\mathrm{z}}(z)$ and obtain

$$S_{\mathrm{z}}^{+}(z) = 1.5477\frac{(1-0.7931z^{-1})}{(1-0.95z^{-1})},$$

and

$$S_{\mathrm{z}}^{-}(z) = 1.5477\frac{(1-0.7931z)}{(1-0.95z)}.$$

$S_{\mathrm{sz}}(z) = S_{\mathrm{s}}(z)$ because $\mathrm{s}(n)$ and $\mathrm{v}(n)$ are uncorrelated. Therefore, we have

$$\begin{aligned}\frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}^{-}(z)} &= \frac{0.0630}{(1-0.95z^{-1})(1-0.7931z)}\\ &= \frac{-0.0794z^{-1}}{(1-0.95z^{-1})(1-1.2608z^{-1})}\\ &= \frac{0.2555}{1-0.95z^{-1}} - \frac{0.2555}{1-1.2608z^{-1}}\end{aligned}$$

and

$$\left[\frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}^{-}(z)}\right]_{+} = \frac{0.2555}{1-0.95z^{-1}}.$$

The causal Wiener filter is then

$$H(z) = \frac{1}{S_{\mathrm{z}}^{+}(z)}\left[\frac{S_{\mathrm{sz}}(z)}{S_{\mathrm{z}}^{-}(z)}\right]_{+}$$

$$= \frac{1-0.95z^{-1}}{1.5477(1-0.7931z^{-1})} \bullet \frac{0.2555}{1-0.95z^{-1}} = \frac{0.1651}{1-0.7931z^{-1}},$$

and

$$h(n) = 0.1651(0.7931)^{n} 1(n).$$

The MSE associated with this filter is computed using Eq. (3.3-3)

$$\mathrm{MSE}_{\mathrm{C}} = 1 - 0.1651\sum_{i=0}^{\infty}(0.7931)^{i}(0.95)^{i} = 0.3302.$$

Compared to no filtering (MSE = 2), the causal Wiener filter reduces the MSE by 7.8 dB. (6.5 dB in FIR filter and 9.6 dB in noncausal filter.)

Digital Modulation Schemes

ECT305 ADC Module V

Introduction

- In baseband pulse transmission, a data stream represented in the form of a discrete pulse-amplitude modulated (PAM) signal is transmitted directly over a low-pass channel.
- In digital passband transmission, on the other hand, the incoming data stream is modulated onto a carrier (usually sinusoidal) with fixed frequency limits imposed by a band-pass channel of interest.
- The communication channel used for passband data transmission may be a microwave radio link, a satellite channel, or the like.
- The modulation process making the transmission possible involves switching (keying) the amplitude, frequency, or phase of a sinusoidal carrier in some fashion in accordance with the incoming data.
- Thus there are three basic signaling schemes, and they are amplitude-shift keying (ASK), frequency-shift keying (FSK), and phase-shift keying (PSK).
- They may be viewed as special cases of amplitude modulation, frequency modulation, and phase modulation.

- In the transmission of digital info over communication channel, Modulator is the <u>interface device</u> that maps digital information into analog waveforms that match channel characteristics.
- Mapping takes blocks of $k=\log_2 M$ bits at a time from the info sequence $\{a_n\}$ and selects one of $M=2^k$ deterministic, finite energy waveforms $\{s_m(t), m=1,2,\ldots,M\}$ for tx over the channel.

- Mapping done under constraint that a waveform transmitted in any time interval depends on one or more previously transmitted waveforms → Modulator has <u>Memory</u>.

- Otherwise, the Modulator is said to be <u>Memoryless</u>.

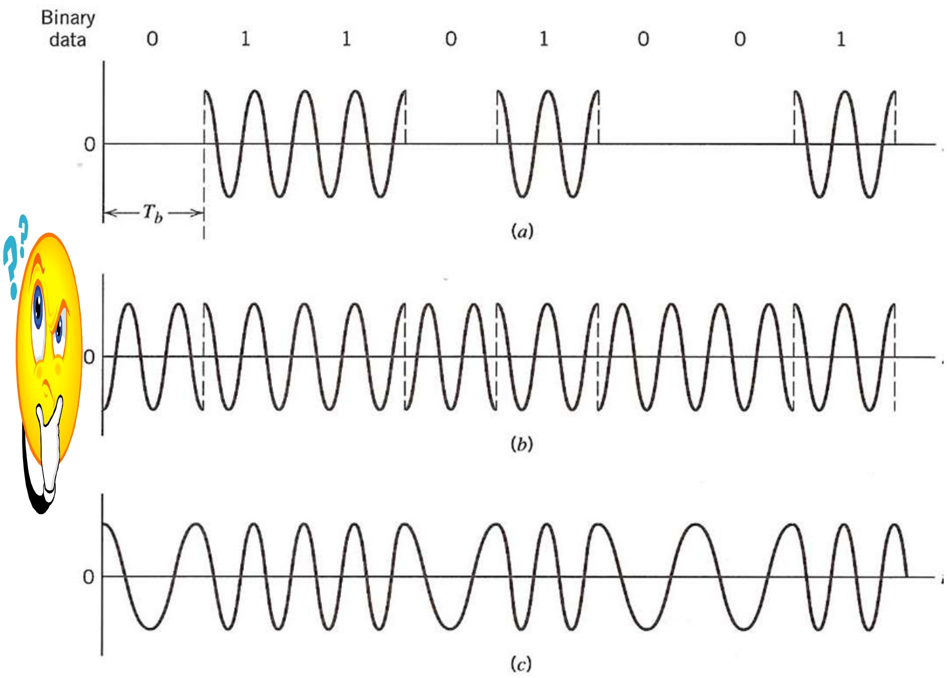
- Data transmission uses a sine carrier wave modulated by data stream.
- Modulation involves switching amplitude, frequency, or phase of a carrier in accordance with the incoming data.
- There are three basic signaling schemes:
  - Amplitude-shift keying (ASK)
  - Frequency-shift Keying (FSK)
  - Phase-shift keying (PSK)

Need for Digital Modulation in modern Communications

- Digital Modulation provides more information capacity, high data security, quicker system availability with great quality communication.
- Hence, digital modulation techniques have a greater demand, for their capacity to convey larger amounts of data than analog modulation techniques.

- The move to digital modulation provides more information capacity, compatibility with digital data services, higher data security, better quality communications, and quicker system availability.
- Developers of communications systems face these constraints:
- available bandwidth
- permissible power
- inherent noise level of the system
- The RF spectrum must be shared, yet every day there are more users for that spectrum as demand for communications services increases.
- Digital modulation schemes have greater capacity to convey large amounts of information than analog modulation schemes.

- Fundamental to all wireless communications is modulation, the process of impressing the data to be transmitted on the radio carrier.
- Most wireless transmissions today are digital, and with the limited spectrum available, the type of modulation is more critical than it has ever been.
- The main goal of modulation today is to squeeze as much data into the least amount of spectrum possible.
- That objective, known as spectral efficiency, measures how quickly data can be transmitted in an assigned bandwidth.
- The unit of measurement is bits per second per Hz (b/s/Hz).
- Multiple techniques have emerged to achieve and improve spectral efficiency.

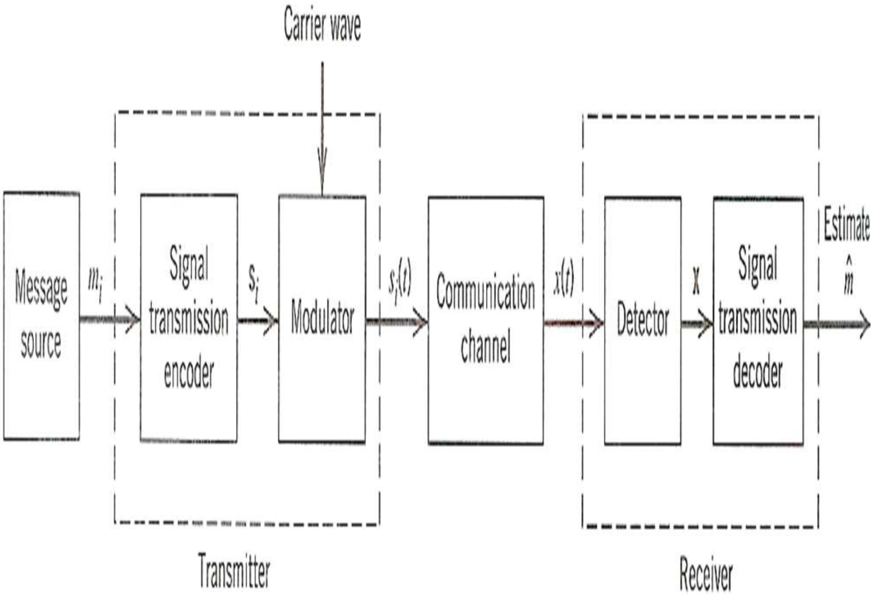


- Unlike ASK signals, both PSK and FSK signals have a const envelope.
- This property makes PSK and FSK signals *impervious to amplitude nonlinearities*.
- In practice, PSK and FSK signals are preferred to ASK signals for *passband data Tx over nonlinear channels*.
- Digital modulation techniques may be classified into *coherent* and *noncoherent* techniques, depending on whether Rx is equipped with a *phase-recovery circuit* or not.
- *Phase-recovery circuit* ensures osc supplying locally gen carrier in Rx is ***synchronized*** (in both frequency and phase) to osc supplying carrier used to originally modulate incoming data stream in Tx.

Passband Transmission Model

- In a functional sense, model a passband data transmission system as shown→
- First, a message source exists that emits one symbol every T seconds, with the symbols belonging to an alphabet of M symbols, which we denote by m_1, m_2 …m_M.
- The a priori probabilities $P(m_1)$, $P(m_2)$, …, $P(m_M)$ specify the message source output.
- When the M symbols of the alphabet are equally likely,

$$p_i = P(m_i) \quad = \frac{1}{M} \quad \text{for all } i \qquad \text{……….(1)}$$



Functional model of passband data transmission system

- The M-ary output of the message source is presented to a signal transmission encoder, producing a corresponding vector **s**$_i$ made up of N real elements, one such set for each of the M symbols of the source alphabet.
- Note that the dimension N $\leq$M.
- With the vector **s**$_i$ as input, the modulator builds a distinct signal $s_i(t)$ of duration T seconds as the representation of the symbol m_i generated by the message source.
- The signal $s_i(t)$ is necessarily an energy signal,

$$E_i = \int_0^T s_i^2(t)\,dt, \qquad i = 1, 2, \ldots, M \qquad \text{……….(2)}$$

- The bandpass communication channel, coupling the transmitter to the receiver, is assumed to have two characteristics:
  - **1**. The channel is linear, with a bandwidth that is wide enough to accommodate the transmission of the modulated signal $s_i(t)$ with negligible or no distortion.
  - **2**. The channel noise w(t) is the sample function of a white Gaussian noise process of zero mean and power spectral density $N_0/2$.
- The receiver, which consists of a detector followed by a signal transmission decoder, performs two functions:
  - **1**. It reverses the operations performed in the transmitter.
  - **2**. It minimizes the effect of channel noise on the estimate $\hat{m}$ computed for the trans mitted symbol m_i.

Binary Phase-Shift Keying

- In a coherent binary PSK system, signals $s_1(t)$ and $s_2(t)$ used to represent binary symbols 1 and 0.
- A pair of sinusoidal waves that differ only in a relative phase shift of 180^o are called antipodal signals.

$$s_1(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \qquad \ldots\ldots\ldots\ldots(3)$$

$$s_2(t) = \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \qquad \ldots\ldots\ldots\ldots(4)$$

where $0 \le t \le T_b$ and E_b is the transmitted signal energy per bit.

- To ensure that each transmitted bit contains an integral no: of cycles of carrier wave, f_c is made equal to n_c/T_b for some fixed integer n_c.
- In the case of binary PSK, there is only one basis function of unit energy:

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t), \quad 0 \le t < T_b \qquad \ldots\ldots\ldots\ldots(5)$$

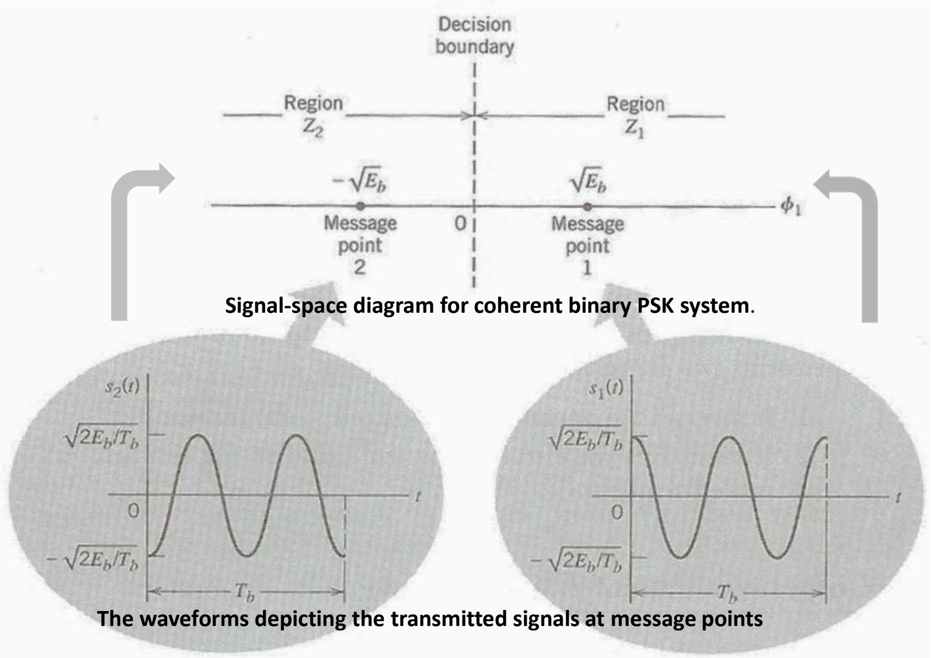
……….(6)

$$s_1(t) = \sqrt{E_b}\phi_1(t), \; s_2(t) = -\sqrt{E_b}\phi_1(t), \; 0 \le t < T_b$$

$$s_{11} = \int_0^{T_b} s_1(t)\phi_1(t)\,dt = +\sqrt{E_b} \qquad s_{21} = \int_0^{T_b} s_2(t)\phi_1(t)\,dt = -\sqrt{E_b}$$

……….(7)

Phase-modulated signals (Binary Phase-Shift Keying)



Signal-space diagram for coherent binary PSK system.

The waveforms depicting the transmitted signals at message points

Coherent detection

- Partition the signal space into two regions:
- set of points closest to message point 1 at $+\sqrt{E_b}$;
- set of points closest to message point 2 at $-\sqrt{E_b}$;
- construct the midpoint of the line joining these two message points and then marking off the appropriate decision regions.
- these two decision regions are marked Z_1 and Z_2, according to the message point around which they are constructed.

- The decision rule is now simply to decide that signal $s_1(t)$ (i.e., binary symbol 1) was transmitted if the received signal point falls in region Z_1 …
- …and to decide that signal $s_2(t)$ (i.e., binary symbol 0) was transmitted if the received signal point falls in region Z_2.
- Two kinds of erroneous decisions may, however, be made:
  - **1**. *<u>Error of the first kind</u>*. Signal $s_2(t)$ is transmitted but the noise is such that the received signal point falls inside region Z_1; so the receiver decides in favor of signal $s_1(t)$.
  - **2.** *<u>Error of the second kind</u>*. Signal $s_1(t)$ is transmitted but the noise is such that the received signal point falls inside region Z_2; so the receiver decides in favor of signal $s_2(t)$.

Phase-modulated signals (Binary Phase-Shift Keying)

- **Probability of Bit Error** is proportional to the distance between the closest points in the constellation.
  - A simple upper bound can be found using the assumption that noise is additive, white, and Gaussian.

$$\text{Prob}\{\text{bit error}\} \leq Q\left(\frac{d}{\sqrt{2N_0}}\right)$$

  - *d* is distance between nearest constellation points.

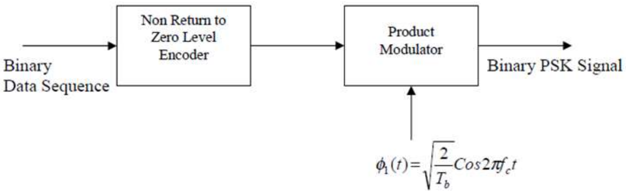
Probability of Bit Error - BPSK

- *Q(x)* is the **Q-function**, the area under a normalized Gaussian function (also called a Normal curve or a bell curve)

$$Q(z) = \int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

$$d = 2\sqrt{E_b} \quad \text{so} \quad \text{Prob}\{\text{bit error}\} \leq Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

Coherent Binary PSK:



Fig(a) Block diagram of BPSK transmitter

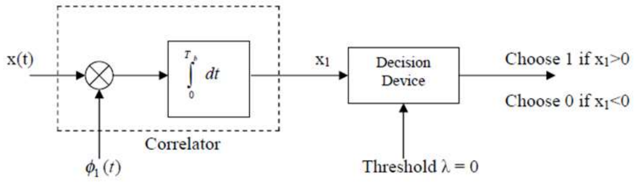


Fig (b) Coherent binary PSK receiver

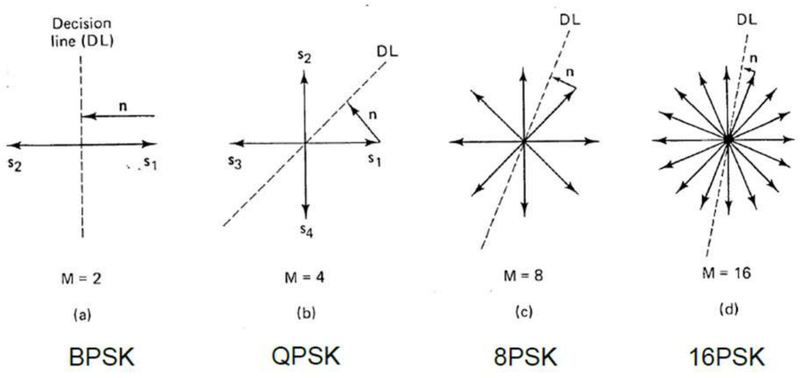
BPSK TRANSMITTER

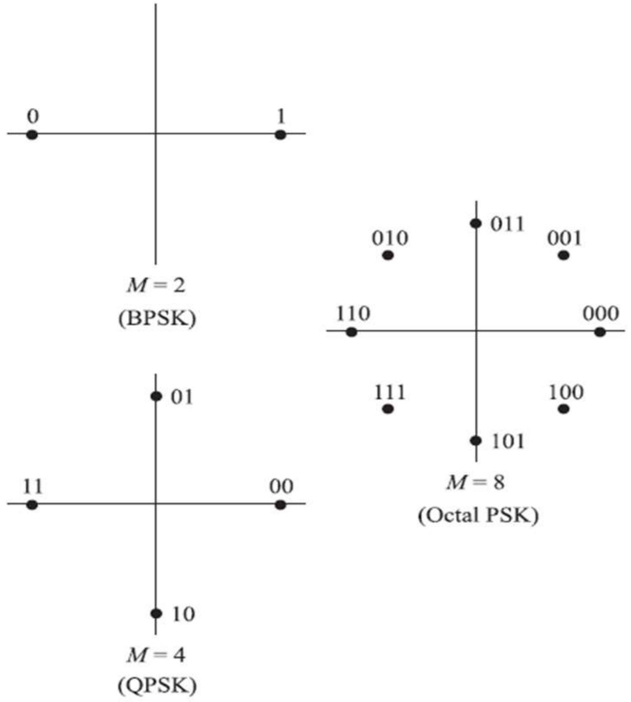
- The transmitter consists of two components:
- **1.** Polar NRZ-level encoder, which represents symbols 1 and 0 of the incoming binary sequence by amplitude levels.
- **2.** Product modulator, which multiplies the output of the polar NRZ encoder by the basis function $\phi_1(t)$; in effect, the sinusoidal $\phi_1(t)$ acts as the "carrier" of the binary PSK signal.

BPSK RECEIVER

- To make an optimum decision on the received signal $x(t)$ in favor of symbol 1 or symbol 0 (i.e., estimate the original binary sequence at the transmitter input), assume that the receiver has access to a locally generated replica of the basis function $\phi_1(t)$.
- In other words, the receiver is synchronized with the transmitter, as shown in the block diagram of Fig b.
- Identify two basic components in the binary PSK receiver:
  - **1.** Correlator, which correlates the received signal $x(t)$ with the basis function $\phi_1(t)$ on a bit-by-bit basis.
  - **2.** Decision device, which compares the correlator output against a zero-threshold, assuming that binary symbols 1 and 0 are equi-probable.
  - If the threshold is exceeded, a decision is made in favor of symbol 1; if not, the decision is made in favor of symbol 0.
- Equality of the correlator with the zero-threshold is decided by the toss of a fair coin (i.e., in a random manner).

MPSK Signal Constellations





Quadri Phase-Shift Keying

- The need to provide reliable performance, indicated by a very low probability of error, is one important goal in the design of a digital communication system.
- Another important goal is the efficient utilization of channel bandwidth.
- Quadri phase shift keying (QPSK) is a bandwidth-conserving modulation scheme, using coherent detection.

$$s_i(t) = \begin{cases} \sqrt{\frac{2E}{T}} \cos\left[2\pi f_c t + (2i-1)\frac{\pi}{4}\right], & 0 \le t \le T \\ 0, & \text{elsewhere} \end{cases}$$

where $i = 1, 2, 3, 4$; E is the transmitted signal energy per symbol, and T is the symbol duration.

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[(2i-1)\frac{\pi}{4}\right] \cos(2\pi f_c t) - \sqrt{\frac{2E}{T}} \sin\left[(2i-1)\frac{\pi}{4}\right] \sin(2\pi f_c t)$$

- Defined a pair of quadrature carriers:

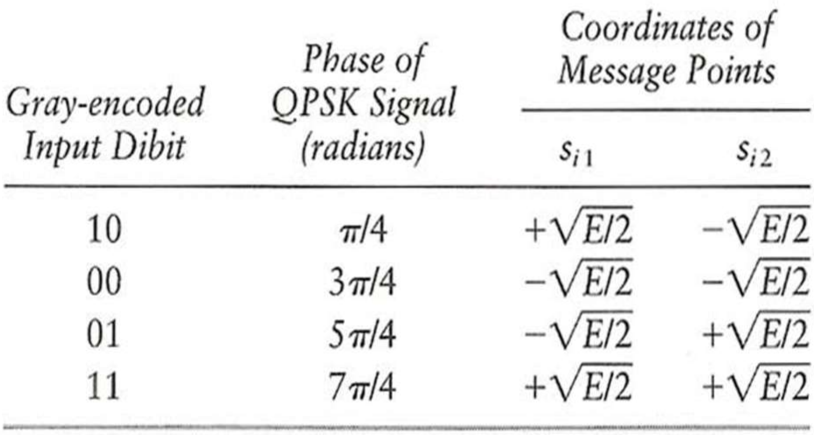
$$\phi_1(t)=\sqrt{\frac{2}{T}}\cos(2\pi f_c t), \qquad 0 \le t \le T$$

$$\phi_2(t)=\sqrt{\frac{2}{T}}\sin(2\pi f_c t), \qquad 0 \le t \le T$$

- There are **four message points**, and the associated signal vectors are defined by

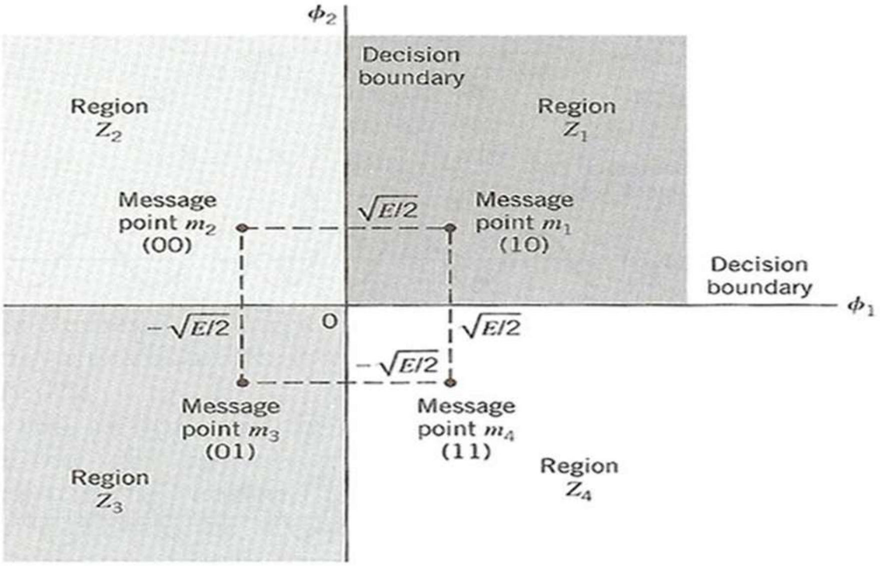
$$s_i(t)=\begin{bmatrix} \sqrt{E}\cos\left((2i-1)\frac{\pi}{4}\right) \\ -\sqrt{E}\sin\left((2i-1)\frac{\pi}{4}\right) \end{bmatrix}, \qquad i=1,2,3,4$$

- Each possible value of the phase corresponds to a unique dibit.
- The Gray coding is commonly used.



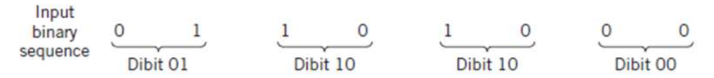
Gray-encoded Input Dibit	Phase of QPSK Signal (radians)	Coordinates of Message Points	
		s_{i1}	s_{i2}
10	$\pi/4$	$+\sqrt{E/2}$	$-\sqrt{E/2}$
00	$3\pi/4$	$-\sqrt{E/2}$	$-\sqrt{E/2}$
01	$5\pi/4$	$-\sqrt{E/2}$	$+\sqrt{E/2}$
11	$7\pi/4$	$+\sqrt{E/2}$	$+\sqrt{E/2}$

Signal space diagram of coherent QPSK system



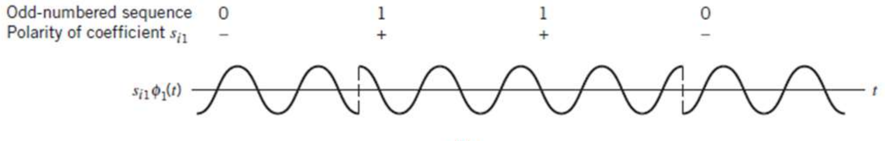
QPSK Waveforms

- Consider the input binary sequence 01101000.
- This sequence is divided into two other sequences, consisting of odd- and even-numbered bits of the input sequence.
- The waveforms representing the two components of the QPSK signal, namely $s_{i1}\phi_1(t)$ and $s_{i2}\phi_2(t)$.
- These two waveforms may individually be viewed as examples of a binary PSK signal.
- Adding them, we get the QPSK waveform.



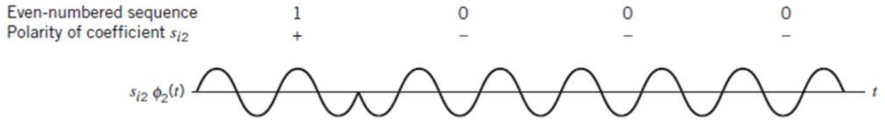
(a)

Input binary sequence



(b)

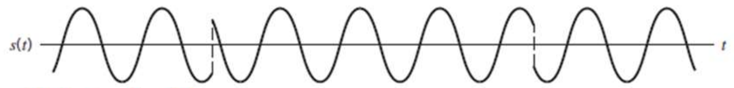
Odd-numbered dibits of input sequence and associated binary PSK signal



(c)

Even-numbered dibits of input sequence and associated binary PSK signal

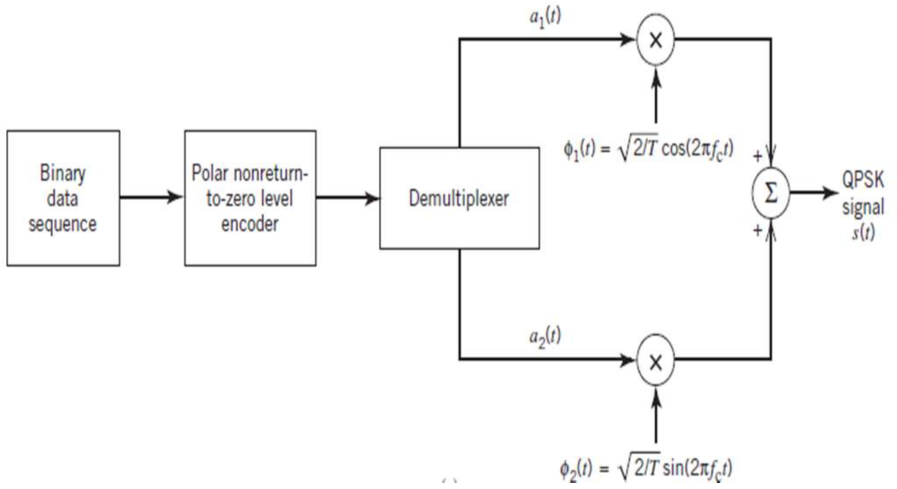
QPSK waveform defined as



$$s(t) = s_{i1}\phi_1(t) + s_{i2}\phi_2(t).$$

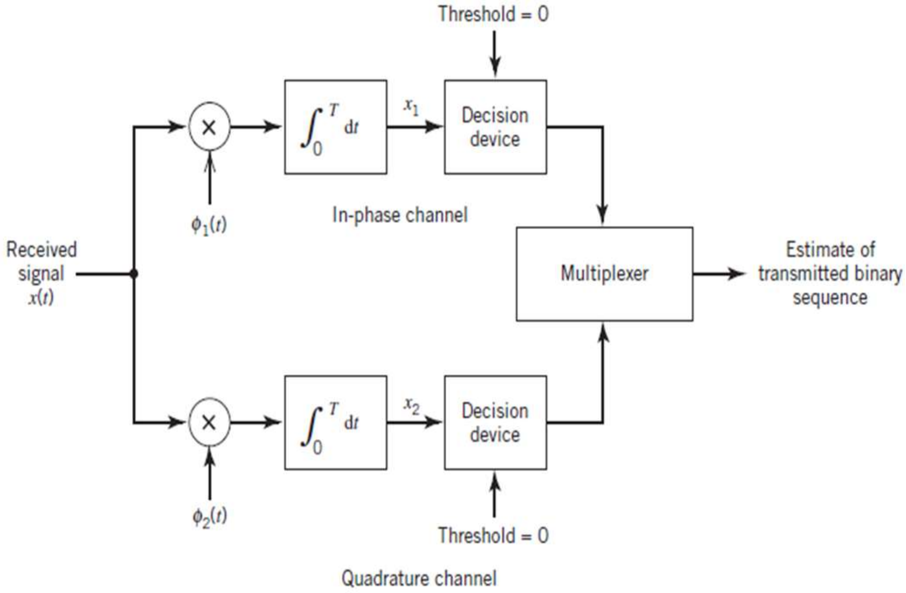
(d)

Generation of QPSK Signals



- A distinguishing feature of the QPSK transmitter is the block labeled Demultiplexer.
- The function of the Demultiplexer is to divide the binary wave produced by the polar NRZ-level encoder into two separate binary waves, one of which represents the odd-numbered dibits in the incoming binary sequence and the other represents the even-numbered dibits.
- The QPSK transmitter may be viewed as two binary PSK generators that work in parallel, each at a bit rate equal to one-half the bit rate of the original binary sequence at the QPSK transmitter input.

Coherent Detection of QPSK Signals



- The QPSK receiver is structured in the form of an *in-phase path* and a *quadrature path*, working in parallel.
- The functional composition of the QPSK receiver is as follows:
- **1.** Pair of correlators, which have a common input $x(t)$.
- The two correlators are supplied with a pair of locally generated orthonormal basis functions $\phi_1(t)$ and $\phi_2(t)$ which means that the receiver is synchronized with the transmitter.
- The correlator outputs, produced in response to the received signal $x(t)$, are denoted by x_1 and x_2, respectively.
- **2.** Pair of decision devices, which act on the correlator outputs x_1 and x_2 by comparing each one with a zero-threshold; here, it is assumed that the symbols 1 and 0 in the original binary stream at the transmitter input are equally likely.

- If $x_1 > 0$, a decision is made in favor of symbol 1 for the in-phase channel output; on the other hand, if $x_1 < 0$, then a decision is made in favor of symbol 0.
- Similar binary decisions are made for the quadrature channel.
- **3.** Multiplexer, the function of which is to combine the two binary sequences produced by the pair of decision devices.
- The resulting binary sequence so produced provides an *estimate* of the original binary stream at the transmitter input.

Probability of Symbol Error or BER

What is the Probability of **symbol error** for QPSK?

$$P_{MPSK} \approx 2Q\left(\frac{d_{min}/2}{\sqrt{N_0/2}}\right) = 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{M}\right)$$

So, $P_{sQPSK} = 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin \pi/4\right) = \mathbf{2Q}\left(\sqrt{\frac{\mathbf{E_s}}{\mathbf{N_0}}}\right)$

So, $P_{eQPSK} = \frac{1}{\log_2 4} \times 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right) = \mathbf{Q}\left(\sqrt{\frac{\boldsymbol{2E_b}}{\boldsymbol{N_0}}}\right)$

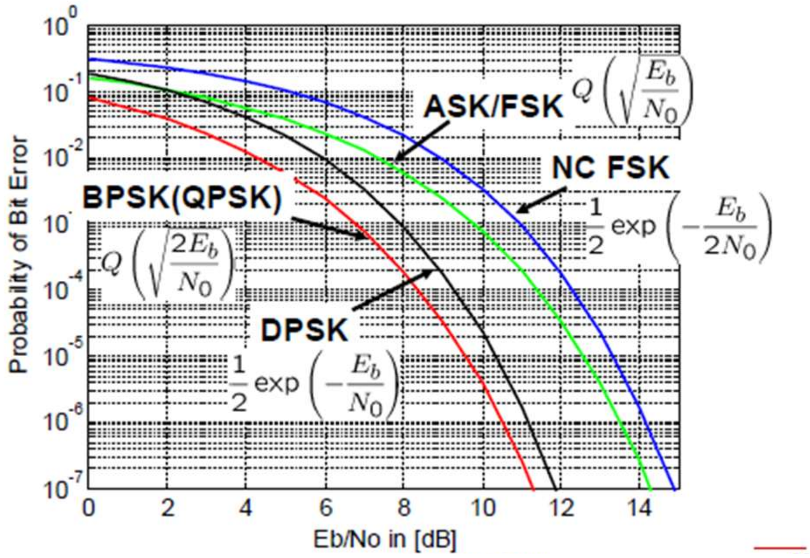
Points to Ponder

- A QPSK system achieves the <u>same average probability of bit error</u> as a binary PSK system for the same bit rate and the same E_b/N_0, but uses <u>only half the channel bandwidth</u>!
- For the same E_b/N_0 and, therefore, the same average probability of bit error, a QPSK system transmits information at <u>twice the bit rate</u> of a binary PSK system for the <u>same channel bandwidth</u>!
- For a prescribed performance, QPSK uses channel bandwidth <u>better</u> than binary PSK, which explains the <u>preferred use of QPSK over binary PSK</u> in practice.

Summary of P_e for Different Binary Modulations

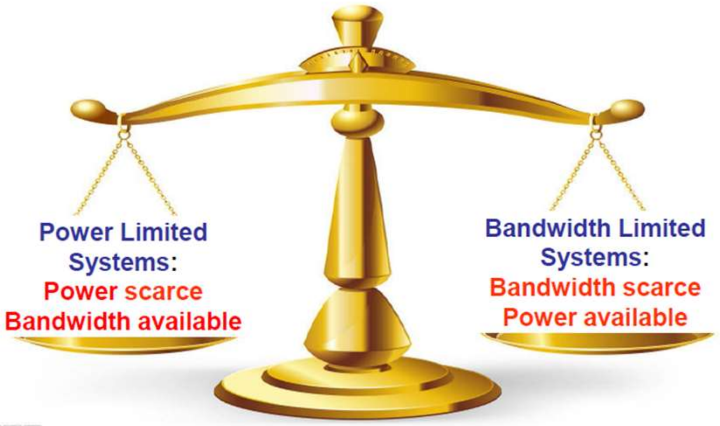
Coherent PSK	$Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$
Coherent ASK	$Q\left(\sqrt{\frac{E_b}{N_0}}\right)$
Coherent FSK	$Q\left(\sqrt{\frac{E_b}{N_0}}\right)$
Non-Coherent FSK	$\frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right)$
DPSK	$\frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right)$

P_e Plots for Different Binary Modulations



System Design Tradeoff

Which Modulation to Use ?



Power Limited Systems:
Power scarce
Bandwidth available

Bandwidth Limited Systems:
Bandwidth scarce
Power available

Practical Applications

- BPSK:
    - WLAN IEEE802.11b (1 Mbps)
- QPSK:
    - WLAN IEEE802.11b (2 Mbps, 5.5 Mbps, 11 Mbps)
    - 3G WDMA
    - DVB-T (with OFDM)
- QAM
    - Telephone modem (16QAM)
    - Downstream of Cable modem (64QAM, 256QAM)
    - WLAN IEEE802.11a/g (16QAM for 24Mbps, 36Mbps; 64QAM for 38Mbps and 54 Mbps)
    - LTE Cellular Systems
- FSK:
    - Cordless telephone
    - Paging system

8.9. QUADRATURE AMPLITUDE MODULATION (QAM)

Quadrature Amplitude Modulation or QAM is a form of modulation which is widely used for modulating data signals onto a carrier used for radio communications. It is widely used because it offers advantages over other forms of data modulation such as PSK, although many forms of data modulation operate along each other.

Quadrature Amplitude Modulation, QAM is a signal in which two carriers shifted in phase by 90 degrees are modulated and the resultant output consists of both amplitude and phase variations. In view of the fact that both amplitude and phase variations are present it may also be considered as a mixture of amplitude and phase modulation.

i.e $\boxed{\text{QAM} = \text{ASK} + \text{PSK}}$

A motivation for the use of quadrature amplitude modulation comes from the fact that a straight amplitude modulated signal, i.e. double sideband even with a suppressed carrier occupies twice the bandwidth of the modulating signal. This is very wasteful of the available frequency spectrum. QAM restores the balance by placing two independent double sideband suppressed carrier signals in the same spectrum as one ordinary double sideband suppressed carrier signal.

8.9.1. ANALOG AND DIGITAL QAM

Quadrature amplitude modulation, QAM may exist in what may be termed either analog or digital formats. The analog versions of QAM are typically used to allow multiple analog signals to be carried on a single carrier. For example it is used in PAL and NTSC television systems, where the different channels provided by QAM enable it to carry the components of chroma or colour information. In radio applications a system known as C-QUAM is used for AM stereo radio. Here the different channels enable the two channels required for stereo to be carried on the single carrier.

Digital formats of QAM are often referred to as "Quantised QAM" and they are being increasingly used for data communications often within radio communications systems. Radio communications systems ranging from cellular technology as in the case of LTE through wireless systems including WiMAX, and Wi-Fi 802.11 use a variety of forms of QAM, and the use of QAM will only increase within the field of radio communications.

8.9.2. DIGITAL / QUANTISED QAM BASICS

Quadrature amplitude modulation, QAM, when used for digital transmission for radio communications applications is able to carry higher data rates than ordinary amplitude modulated schemes and phase modulated schemes. As with phase shift keying, etc, the number of points at which the signal can rest, i.e. the number of points on the constellation is indicated in the modulation format description, e.g. 16QAM uses a 16 point constellation.

When using QAM, the constellation points are normally arranged in a square grid with equal vertical and horizontal spacing and as a result the most common forms of QAM use a constellation with the number of points equal to a power of 2 i.e. 4, 16, 64 . . . .

By using higher order modulation formats, i.e. more points on the constellation, it is possible to transmit more bits per symbol. However the points are closer together and they are therefore more susceptible to noise and data errors.

Normally a QAM constellation is square and therefore the most common forms of QAM16QAM, 64QAM and 256QAM.

The advantage of moving to the higher order formats is that there are more points within the constellation and therefore it is possible to transmit more bits per symbol. The downside is that the constellation points are closer together and therefore the link is more susceptible to noise. As a result, higher order versions of QAM are only used when there is a sufficiently high signal to noise ratio.

To provide an example of how QAM operates, the constellation diagram below table shows the values associated with the different states for a 16QAM signal. From this it can be seen that a continuous bit stream may be grouped into fours and represented as a sequence.

8.9.3. TYPES OF QAM

Name	Bits per symbol (N)	Number of symbols (M) $(M = 2^N)$
4 QAM	2	4
8 QAM	3	8
16 QAM	4	16
32 QAM	5	32
64 QAM	6	64

8.9.4. PRINCIPLE OF QAM

The general form of QAM is defined by the transmitted signal such as

$$S_i(t) = \sqrt{\frac{2E_0}{T}}\, a_i \cos(2\pi f_c t) + \sqrt{\frac{2E_0}{T}}\, b_i \sin(2\pi f_c t) \qquad 0 \le t \le T \quad ...(1)$$

Where, E_0 is the energy of the signal with the lowest amplitude a_i & b_i are a pair of independent intergers chosen in accordance with the location of the pertinent message point.

The signal S_i(t) consists of two phase quadrature carriers, each of which is modulated by a set of discrete amplitudes, hence it called quadrature amplitude modulation.

The signal S_i(t) can be expanded in terms of a pair of basis functions

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t) \qquad 0 \le t \le T \quad ...(2)$$

and

$$\phi_2(t) = \sqrt{\frac{2}{T}} \sin(2\pi f_c t) \qquad 0 \le t \le T \quad ...(3)$$

The coordinates of the i^{th} message point are $a_i\sqrt{E}$ and $b_i\sqrt{E_0}$, where (a_i, b_i) is an element of the L-by-L matrix

$$\{a_i, b_i\} = \begin{bmatrix} (-L+1, L-1) & (-L+3, L-1) & \cdots & (L-1, L-1) \\ (-L+1, L-3) & (-L+3, L-3) & \cdots & (L-1, L-3) \\ \vdots & \vdots & & \vdots \\ (-L+1, -L+1) & (-L+3, -L+1) & \cdots & (L-1, -L+1) \end{bmatrix}$$

Where, $L = \sqrt{M}$

For, example, For the 16 point QAM $L = 4$ then we have the matrix

$$\{a_i, b_i\} = \begin{bmatrix} (-3, 3) & (-1, 3) & (1, 3) & (3, 3) \\ (-3, 1) & (-1, 1) & (1, 1) & (3, 1) \\ (-3, -1) & (-1, -1) & (1, -1) & (3, -1) \\ (-3, -3) & (-1, -3) & (1, -3) & (3, -3) \end{bmatrix}$$

8.9.5. GEOMETRICAL REPRESENTATION

Case 1: M = 16

The signal constellation for QAM consists of a square lattice of message points. Below Figure shows signal constellation for $M = 16$.

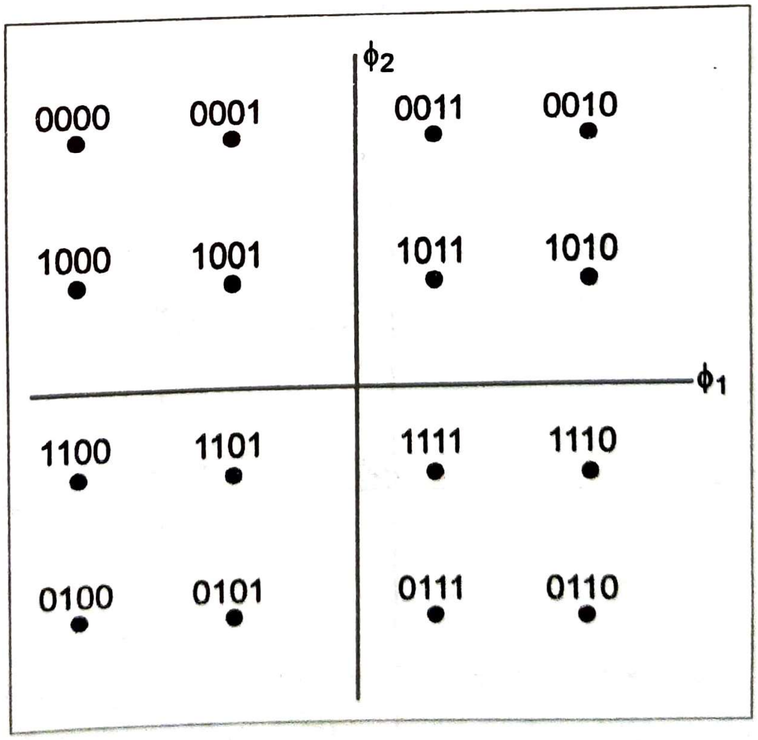


Fig. 8.20. Signal constellation of QAM for M=16

For $M = 16$, the corresponding signal constellations for the in-phase and quadrature components of the amplitude phase modulated wave are shown below.

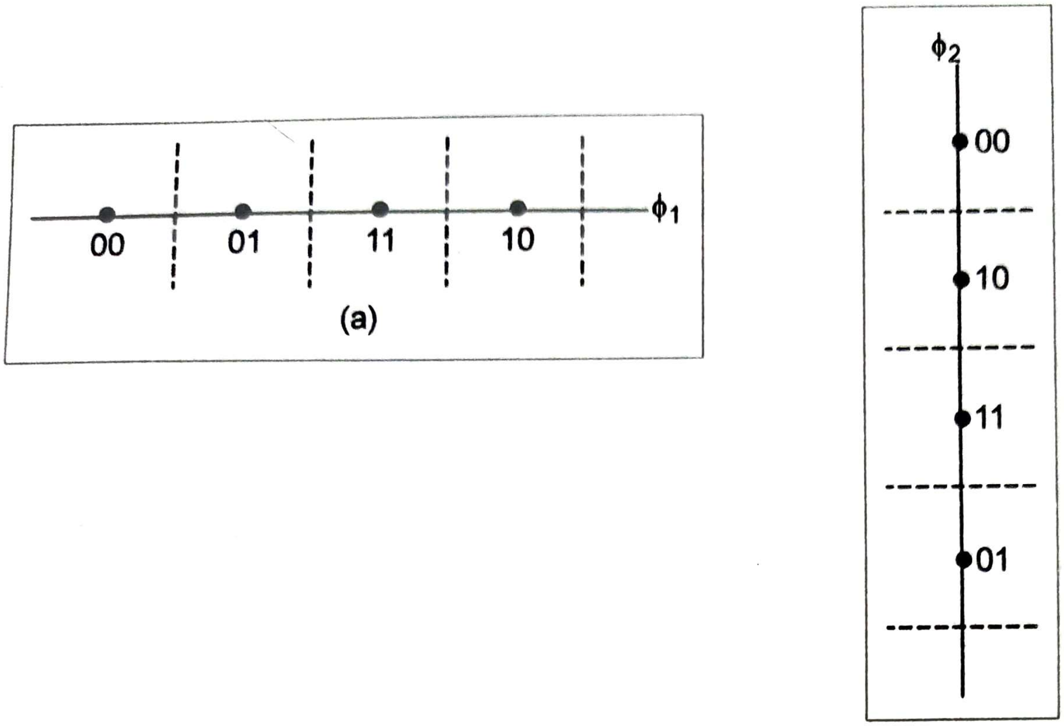


Fig. 8.21. Decomposition of signal constellation of QAM (for M=16) into two signal – space diagrams for (a) in-phase comment $\phi_1(t)$ and (b) quadrature comment $\phi_2(t)$

Case 2: M = 4

The signal constellation for M = 4 is shown in below Figure, which is recognized to be the same as QPSK.

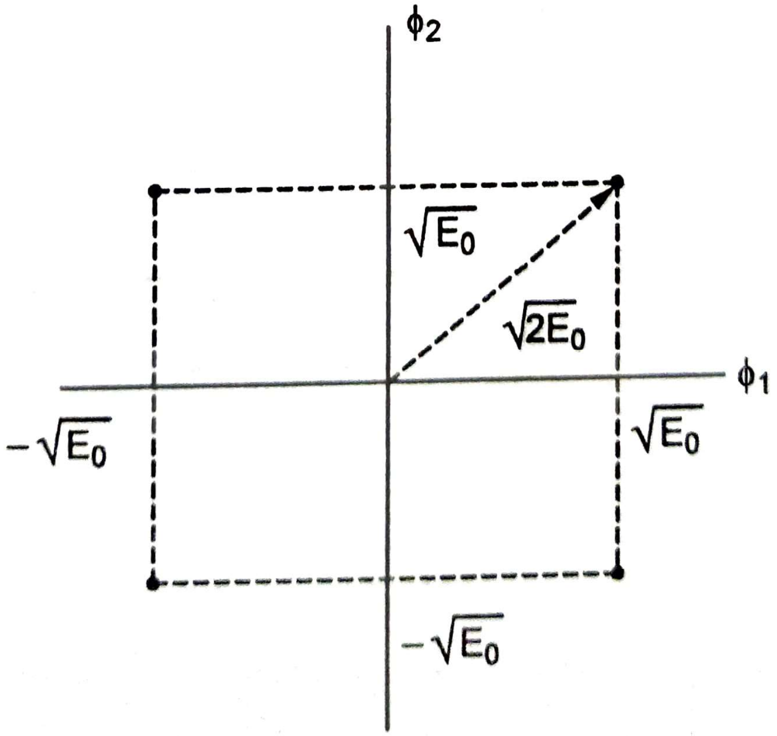


Fig. 8.22. Signal constellation for the QAM for M = 4

8.9.6. GENERATION

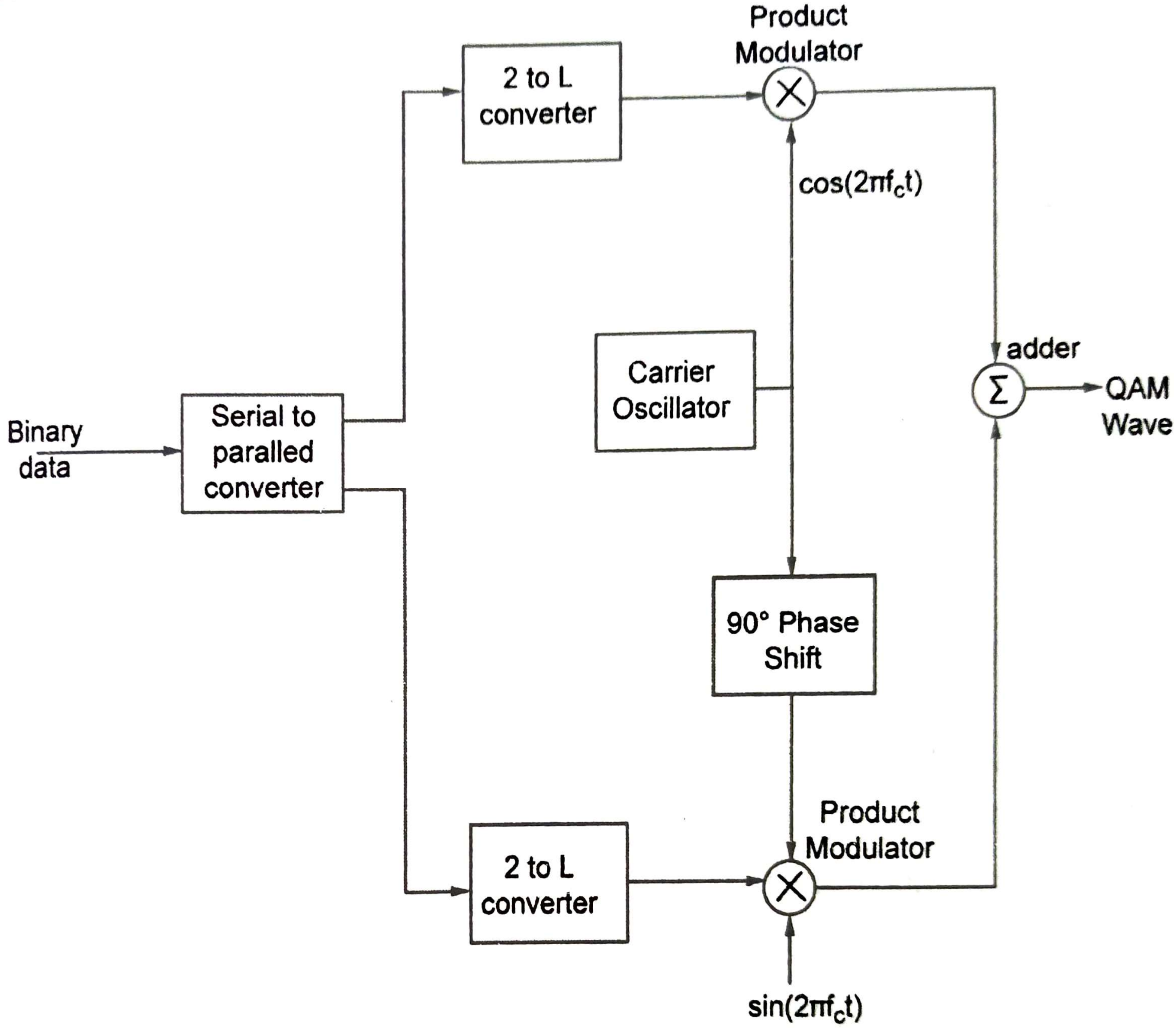


Fig. 8.23. QAM System Transmitter

Serial To paralled Converter

Serial to paralled converter accepts a binary sequence at a bit rate $R_b = 1/T_b$ and produces two paralled binary sequences whose bit rates are $R_b/2$ each,

2-to-L Converters

In 2 to L converters, where $L = \sqrt{M}$, generate polar L-Level signals in response to the respective in-phase and quadrature channel inputs.

Quadrature carrier multiplexing of the two polar L-Level signals to generate desired QAM signal.

8.9.7. DETECTION

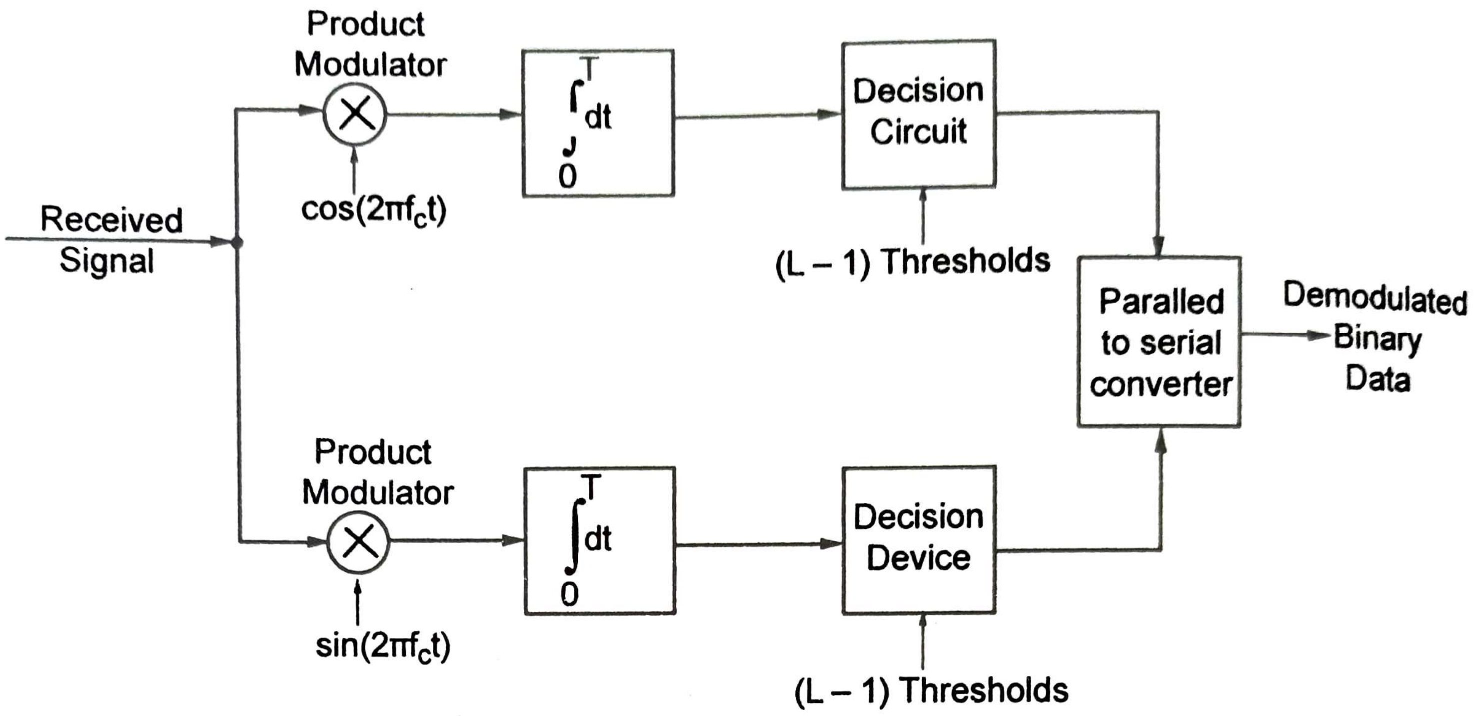


Fig. 8.24. QAM System Receiver

Decision Circuit

Decoding of each baseband channel is accomplished at the output of the decision circuit, which is designed to compare the L-Level signals against L-1 decision thresholds.

Paralled-to-Serial Converter

The two binary sequences are combined in the parallel to serial converter to Reproduce the original binary sequence.

8.9.8. BANDWIDTH

Bandwidth of QAM signal will be

$$\boxed{BW = \frac{2f_b}{W}}$$

8.9.9. POWER SPECTRAL CHARACTERISTICS

Power spectral density of baseband QAM signal will be

$$S(f) = P_s T_s \left[\frac{\sin(\pi f T_s)}{\pi f T_s}\right]^2$$

We know that $E_s = P_s T_s$

The above equation gives power spectral density of QAM when they modulate the carrier, the main lobe given by above equation is shifted at carrier frequency f_c.

$$S(f) = \frac{P_s T_s}{2}\left[\frac{\sin\pi(f-f_c)T_s}{\pi(f-f_c)T_s}\right]^2 + \frac{P_s T_s}{2}\left[\frac{\sin\pi(f+f_c)T_s}{\pi(f+f_c)T_s}\right]^2$$

This equation gives power spectral density of QAM signal.

8.9.10. PROBABILITY OF ERROR

To calculate the probability of symbol error for QAM, we proceed as follows.

1. The in-phase and quadrature components of QAM are independent so the probability of correct detection for such a scheme may be written as

$$P_c = (1 - P_e')^2 \qquad \ldots(1)$$

where P_e' is the probality of symbol error for either component.

2. The signal constellation for the in-phase or quadrature component has a geometry similar to PAM with a corresponding number of amplitude levels.

$$P_e' = \left[1 - \frac{1}{L}\right] erfc\left[\sqrt{\frac{E_0}{N_0}}\right] \qquad \ldots(2)$$

where L is the square root of M

3. The probility of symbol error for QAM is given by

$$\begin{aligned} P_e &= 1 - P_c \\ &= 1 - (1 - P_e')^2 \qquad \ldots(3) \\ &\simeq 2P_e' \end{aligned}$$

where it is assumed that P_e' is small compared to unity. The probability of symbol error for QAM is given by

$$P_e \simeq 2\left[1 - \frac{1}{\sqrt{M}}\right] erfc\left[\sqrt{\frac{E_0}{N_0}}\right] \qquad \ldots(4)$$

8.9.11. TRANSMITTED ENERGY

The trasmitted energy in QAM is variable in its instantaneous value depends on the particular symbol transmitted. So we can express P_e interms of the average value of the transmitted energy rather than E_0.

Assuming that the L amplitude levels of the in-phase or quadrature component are equally likely, we have

$$E_{av} = 2\left[\frac{2\,E_0}{L}\sum_{i=1}^{L/2}(2\,i-1)^2\right] \qquad \ldots(5)$$

Summing the series in equation (5) we get

$$E_{av} = \frac{2\,(L^2-1)\,E_0}{3}$$

$$= \frac{2\,(M-1)\,E_0}{3} \qquad \ldots(6)$$

According, we may rewrite equation (4) interms of E_{av} as

$$P_e \simeq 2\left[1-\frac{1}{\sqrt{M}}\right] erfc\left[\sqrt{\frac{3\,E_{av}}{2\,(M-1)\,N_0}}\right] \qquad \ldots(7)$$

which is the desired result

8.9.12. ADVANTAGES

- It has better noise immunity than PSK
- It is easy to design than PSK

8.9.13. DISADVANTAGES

- It has higher error probability than QPSK
- Relatively complex

QAM vs other modulation formats

As there are advantages and disadvantages of using QAM it is necessary to compare QAM with other modes before making a decision about the optimum mode. Some radio communications systems dynamically change the modulation scheme dependent upon the link conditions and requirements - signal level, noise, data rate required, etc.

The table below compares various forms of modulation.

Summary of types of modulation with data capacities

Modulation	Bits per symbol	Error margin		Complexity
OOK	1	1/2	0.5	Low
BPSK	1	1	1	Medium
QPSK	2	$1 / \sqrt{2}$	0.71	Medium
16 QAM	4	$\sqrt{2} / 6$	0.23	High
64QAM	6	$\sqrt{2} / 14$	0.1	High

Typically it is found that if data rates above those that can be achieved using 8-PSK are required, it is more usual to use quadrature amplitude modulation. This is because it has a greater distance between adjacent points in the I - Q plane and this improves its noise immunity. As a result it can achieve the same data rate at a lower signal level.

However the points no longer the same amplitude. This means that the demodulator must detect both phase and amplitude. Also the fact that the amplitude varies means that a linear amplifier is required to amplify the signal.

QAM Applications

QAM is in many radio communications and data delivery applications. However some specific variants of QAM are used in some specific applications and standards.

For domestic broadcast applications for example, 64 QAM and 256 QAM are often used in digital cable television and cable modem applications. In the UK, 16 QAM and 64 QAM are currently used for digital terrestrial television using DVB - Digital Video Broadcasting. In the US, 64 QAM and 256 QAM are the mandated modulation schemes for digital cable as standardized by the SCTE in the standard ANSI/SCTE 07 2000.

In addition to this, variants of QAM are also used for many wireless and cellular technology applications.

Model Question Paper

A P J Abdul Kalam Technological University Fifth Semester B Tech Degree

Examination Branch:Electronics and Communication

COURSE: ECT 305 ANALOG AND DIGITAL COMMUNICATION

Time: 3 Hrs **Max. Marks: 100**

PART A

Answer All Questions

1 Explain the need for modulation (3)*K2*

2 Plot the spectrum of an FM signal (3)*K2*

3 In a game a six faced die is thrown. If 1 or 2 comes the player gets Rs 30, if 3 or 4 the player gets Rs 10, if 5 comes he loses Rs. 30 and in the event of 6 he loses Rs. 100. Plot the CDF and PDF of gain or loss (3) *K3*

4 Give the conditions for WSS (3)*K2*

5 Compute the step size for a delta modulator without slope overload if the input is $A\cos 2\pi 120t$ (3)*K3*

6 State source coding theorems I and II (3)*K1*

7 Give the Nyquist criterion for zero ISI. (3)*K1*

8 Give the mathematical model of ISI (3)*K2*

9 Plot BER against SNR for a BPSK system (3)*K2*

10 Draw the signal constellation of a QPSK system with and without AWGN. *(3)* *K3*

PART B

Answer one question from each module. Each question carries 14 mark.

Module I

11(A) Give the model of AM signal and plot its spectrum (10)*K2*

11(B) If a sinusoidal is amplitude modulated by the carrier $5 \cos 2\pi 300t$ to a depth of 30 %, compute the power in the resultant AM signal. (4)*K3*

OR

12(A) Explain how SSB is transmitted and received. (10)*K2*

12(B) Compute the bandwidth of the narrow band FM signal withmodulating signal frequency of 1*kHz* and index of modulation 0.3 *(4) K3*

Module II

13(A) Compute the entropy of Gaussian random variable. (10)*K3*

13(B) Give the relation between autocorrelation and power spectral density of a WSS. (4)*K2*

OR

14(A) Test whether the random process $X(t) = A\cos 2\pi ft+\theta$ is WSS if θ is uniformly distributed in the interval $[-\pi,\pi]$ (10)*K3*

14(B) Explain mutual information. Give its relation with self information. (4)*K2*

Module III

15(A) A WSS process with autocorrelation $RX(\tau) = e-\alpha|\tau|$ is ap-plied to an LTI system with impulse response $h(t) = e-\beta t$ with $|\alpha| > 0$ and $|\beta| > 0$. Find the output power spectral density (10)K3

15(B) Give the conditions for stationarity in the strict sense. (4)*K2*

OR

16(A) Find an orthonormal basis set fot the set of signals (7)*K3*

$$s_1(t) = Asin(2\pi f_0 t); \qquad 0 \leq t \leq T$$

and

$$s_2(t) = Acos(2\pi f_0 t); \qquad 0 \leq t \leq T$$

where $f_0 = \frac{m}{T}$ where m is an integer.

16(B) Plot the above signal constellation and draw the decision region on it. Compute the probability of error. *(7)* *K3*

Module IV

17(A) Compute the probability of error for maximum likely hood detection of binary transmission. *(8) K3*

17(B) Explain the term matched filter. Plot the BER-SNR curve for a matched filter receiver (6)*K2*

OR

18(A) Design a zero forcing equalizer for the channel that is characterized by the filter taps {1,0.7,0.3} (8)*K3*

18(B) Explain partial rsponse signaling (6)*K2*

Module V

19 For a shift keying system defined by $s(t) = A_c\, k\sin(2\pi f_c t) \pm A_c k\cos(2\pi f_c t)$ plot the signal constellation. Compute the probability of error. *(14) K3*

OR

20(A) Derive the probability of error for a QPSK system with Gray coding. *(10) K3*

20(B) Draw the BER-SNR plot for a QPSK system (4)*K3*