Sukkur Institute of Business Administration
Department of Telecommunication Engineering.

Analog & Digital Communication

course lecture notes

Lecturer
Dr. Muhammad Asim Ali

Fall 2015

Contents

Chapter 1

Introduction and Preliminaries

Outline

The objective of this chapter is to revisit the elementary concepts essential for this module. It is imperative that students be able to express linear systems and signals mathematically and have basic understanding of signal processing operations such as convolution and Fourier transforms. The fundamental concepts covered in this chapter are

- Ⓐ Introduction
- Ⓑ Historical background
- Ⓒ Fourier Transforms
- Ⓓ Signal and Systems
- Ⓔ Sampling & Quantization

Communication systems are designed to transmit and receive information. This field has its roots in many areas of EE. Such as computers, electrical and electronic circuits, electronic devices, digital signal processing, electromagnetics and photonics.

Noise limits our ability to communicate. If there were no noise, we could communicate messages to the outer limits of universe using finitely small amount of power. This has been known since the early days of radio. However, the theory that describes noise and the effects of noise on the transmission of information was not developed until 1940's by people like Rice[1944], Shannon [1948] and Wiener [1949].

1.1 Historical Perspective

A time chart showing the historical development of communication is given in the table below

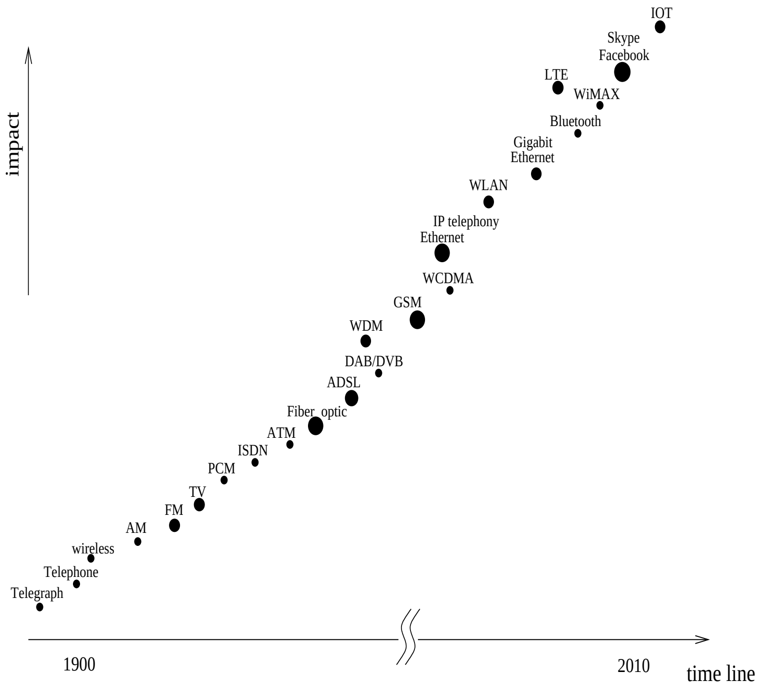


Figure 1.1: Evolution of communication technology.

Just for the interest a chronological order of significant milestones provided here

- 1837 Morse perfected the Morse code.
- 1864 Maxwell discovered radio waves.
- 1875 Alexander Bell invented telephone.

- 1901 Marconi did trans-atlantic radio transmission.
- 1918 Edwin Armstrong invented super-hytrodyne radio receiver.
- 1926 First all electric TV was developed by Philo T Farnsworth.
- 1928 Harry Nyquist provided mathematical foundation of dispersive comm. system.
- 1933 Armstrong developed frequency modulation scheme.
- 1937 Alec Reeved Invented of PCM.
- 1943 D. O. North Matched Filter.
- 1947 Sampling theorem
- 1948 Shannon's ground breaking paper on information theory.
- 1962 T-1 link devised at Bell labs.
- 1985 DARPA project $\rightarrow$ foundation of internet.

Many other significantly achievements in the field of computer, optical, satellite, wireless communications, electronics and signal processing have contributed to the exponential rise of communication systems.

The field of communication has come a long way over the past century.

1.1.1 Digital & Analog Sources and Systems

Definition:*A digital information source produces finite set of possible messages.*
A typewriter is a good example of digital source. There is a finite number of characters that can be emitted from the source.
Definition:*An analog information source produces messages that are defined on a continuum.*

A microphone is good example of an analog source. The output voltage describes the information in the sound and it is distributed over a continuous range of values.

Definition:*A digital communication system transfers information from a digital source to a digital sink.*

Definition:*A analog communication system transfers information from a analog source to an analog sink.*

Strictly speaking, a digital signal is defined usually as a function of time that can have only discrete set of values. If the digital signal is binary, only two values are allowed. An analog signal is a function of time that has continuous range of values. An electronic digital communication system usually has voltages and currents that have a digital waveform; However it may also have analog waveforms. For example, the information from a binary source many be transmitted to the sink by using a sine wave of 100Hz to represent a binary 1 and a sine wave of 500Hz to represent a binary 0. Here digital source information is transmitted to the sink by use of analog waveform, but this is still called a digital communication system.

Digital communication system has a number of advantages.

a. Relatively inexpensive digital circuits may be used.

b. Privacy is preserved by using data encryption.

c. Greater dynamic range (the difference between largest possible and smallest value) is possible.

d. Data from voice, video and data sources may be merged and transmitted over a common digital transmission system.

e. In long distance systems, noise does not accumulate from repeater to repeater.

f. Error in detected data may be small, even when there is large amount of noise is received signal.

g. Errors may often be corrected by the use of coding.

Digital communication system also has disadvantages

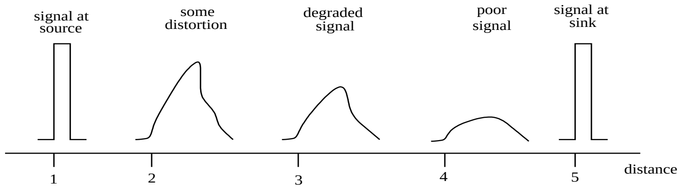


Figure 1.2: Regeneration of digital signal.

a. Generally more bandwidth is required than that for analog system.

b. Synchronization is required.

The advantages of digital communication systems by far outweigh its disadvantages.

1.1.2 Classification of Signals

Deterministic/Stochastic:

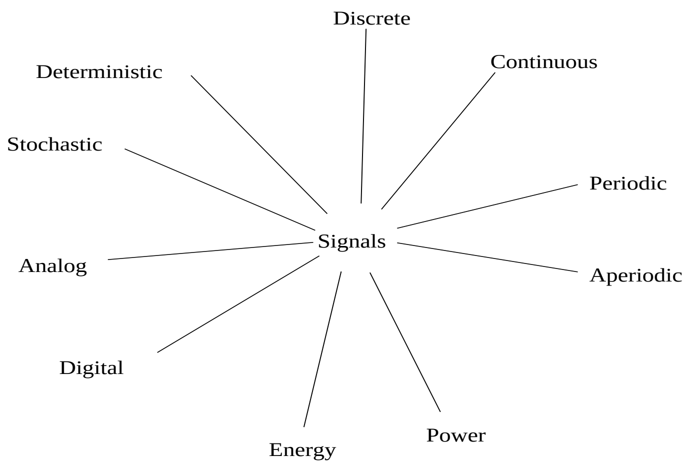
A deterministic waveform can be modelled as completely specified function in time. For example

$$x(t) = A \cdot \cos(\omega_o t + \phi_o)$$

describes a waveform where, A, ω_o and ϕ_o are known constants, this waveform is said to be deterministic.
A random waveform or (stochastic waveform) cannot be completely specified as a function of time and must be modelled probabilistically.

$$n \sim \mathcal{N}(0, \sigma_n^2) \quad \text{Gaussian distribution and} \quad u \sim \mathcal{U}(a, b) \quad \text{Uniform distribution}$$

It is evident that the signals which represent a source can not be deterministic. For example, in a digital communication system we might send information corresponding to any of the letters of english alphabets. Each letter might be represented by a deterministic waveform, but when we examine the waveform that is emitted from the source, we find that it is random wavefrom because we do not know exactly which character will be transmitted. Consequently we need to design the communication system using random waveforms, and any noise that is introduced would be described by a random waveform. This technique requires the use of probabilistic



modeling of source.

Periodic/Aperiodic:

$$x(t) = x(t + T_0) \qquad \text{for } -\infty < t < \infty \text{ and } \ T_0 > 0$$

Analog/Digital:

An analog signal $x(t)$ can take any possible value, while digital signals are limited to only certain pre-defined values.

Power/Energy:

A signal $x(t)$ is defined as an *Energy* signal if and only if it has non-zero but finite energy ($0 < E_x < \infty$).

$$E_x = \lim_{T \to \infty} \int_{t=-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)dt$$
$$E_x = \int_{t=-\infty}^{+\infty} x^2(t)dt$$

A signal is defined as an *Power* signal if and only if it has non-zero but finite power ($0 < P_x < \infty$).

$$P_x = \lim_{T \to \infty} \frac{1}{T} \int_{t=-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)dt$$

1.1.3 Block Diagram of Communication System

Communication system can be described by the block diagram Fig. 1.3. Regardless of the particular application, all communication systems involve three main subsystems, the transmitter, the channel and the receiver. The message from the source is represented by waveform $m(t)$. The message is delivered to the sink is denoted by $\tilde{m}(t)$. The $\sim$ indicates that the message received may be corrupted by noise in the channel or there may be other impairment such as undesired filtering and undesired nonlinearities. The message information may be in analog or digital format. It may be audio or video or other type of information. In multiplexed systems, there may be multiples of sources and sinks operating simultaneously. If the spectrum of signal $m(t)$ or $\tilde{m}(t)$ is located around dc $\rightarrow f = 0$. Then such signal is called as *baseband* signal. The signal processing block at the transmitter conditions the source for more

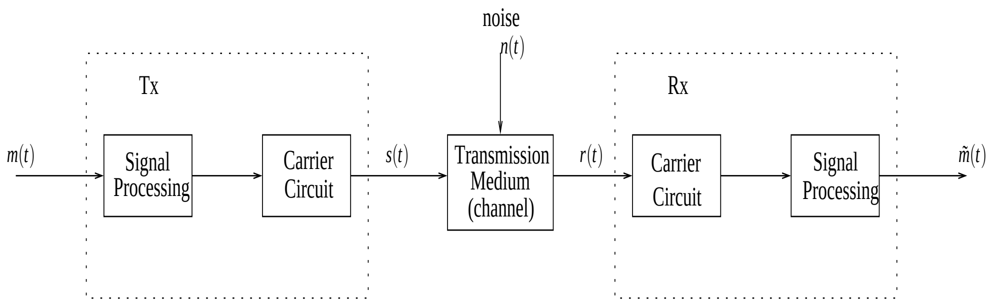


Figure 1.3: General transmission system.

efficient transmission. For example in analog system, the signal processor may be an analog lowpass filter that is used to restrict the bandwidth of $m(t)$. In hybrid systems, the signal processor may be analog-to-digital convertor (ADC). This produces a digital word ; more on this will follow later. The signal processor could add parity bits to digital word to provide channel coding so that error detection and correction can be used in the signal processor in the receiver side to reduce or eliminate bit errors that are caused by noise in the channel.

The transmitter carrier circuit converts the processed baseband signal into a frequency band that is appropriate for the transmission medium of the channel.For example, if the channel consists of fiber optic cable, the carrier circuit converts the baseband input to light frequencies$\Rightarrow$ $s(t)$ is light. If channel propagates baseband signal, no carrier circuit is needed and $s(t)$ can be output of the processing unit. A carrier circuit is needed when the operating frequency of the medium is $f_c \gg 0$. In this case $s(t)$ is called bandpass signal. For example an amplitude modulated broadcast. The mapping of a baseband input information waveform $m(t)$ into bandpass signal $s(t)$ is called modulation. A typical bandpass signal can be represented as

$$s(t) = A(t)\text{cos}[\omega_c t + \theta(t)]$$

where ω_c=$2\pi f_c$, if $A(t)$=1 and $\theta(t)$=0, $s(t)$ would be a pure sinusoid of frequency f_c with bandwidth 0. In the modulation process provided by the carrier circuit, the baseband input waveform $m(t)$ causes $A(t)$ and/or $\theta(t)$ to change as a function of $m(t)$. These fluctuations in $A(t)$ and $\theta(t)$ cause $s(t)$ to have non-zero bandwidth, more on this stuff later.

Channel may be classified as wired or wireless channels. Examples of wired channel are twisted pair cable, coaxial cable, wave guides and fiber optic cable. Typical examples of wireless channel are air, vacuum and seawater. The principles of digital communication apply to all type of channels, although certain media might favour a certain type of signalling. In general, the channel medium attenuates the signal so that the noise level of the channel and/or the noise introduced by imperfect receiver. The channel may contain active amplifying devices such as repeaters in telephone systems , satellite transponders in space communication systems. These devices are necessary to keep the signal level from noise floor.In addition the channel may provide multipath propagation between input and output that have different delays and attenuation characteristics, even worse the channel properties might change from time to time. The incurring costs and practical limitations are a major motivating factor for error detection and correction to be discussed in later sections.

The receiver takes the corrupted signal at the output of the channel and converts it to the baseband channel that can be handled by the receiver baseband processor. The baseband processor cleans up the signal and delivers the estimate of the source information $\tilde{m}(t)$ to the receiver.

The goal of the engineer is to design a communication system such that the signal is transmitted to the receiver with as little deterioration as possible, while satisfying the limited power constraints and using limited bandwidth. The deterioration in performance of quality is measured in *Bit Error Rate* in short **BER** of the delivered data $\tilde{m}(t)$.

1.2 Representation of signals

1.2.1 Some Elementary Signals

Dirac Delta Function and Unit Step Function

The dirac delta functions $\delta(t)$ is defined as

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{1.1}$$

it is also important to note that

$$\int_{t:-\infty}^{+\infty} \delta(t) dt = 1$$

The Dirac delta function is not a true function, so it is said to be a singular function.

Some important properties of the δ function are presented below

Sifting: The sifting property of the δ function is

$$\int_{t:-\infty}^{+\infty} x(t)\delta(t - t_0) dt = x(t_0)$$

Some additional properties

$$\delta(at) = \frac{1}{|a|}\delta(t)$$
$$\delta(-t) = \delta(t)$$
$$x(t) \cdot \delta(t) = x(0)\delta(t)$$
$$\delta(t) = \frac{du(t)}{dt}$$

Definition The unit step function $u(t)$ is

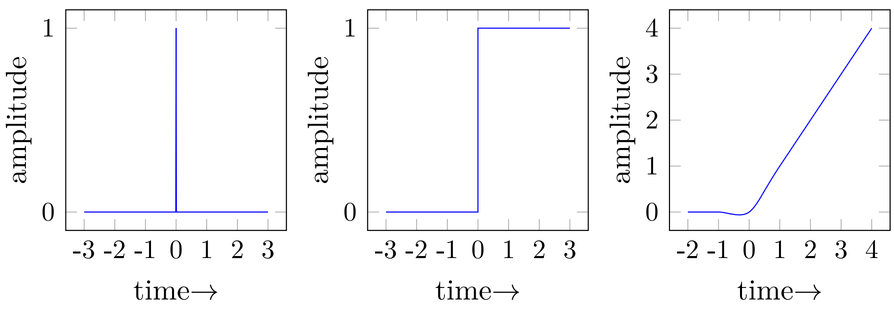


Figure 1.4: Impulse, unit step and ramp functions

$$u(t) = \begin{cases} 1 & , \quad t > 0 \\ & , \\ 0 & , \quad t < 0 \end{cases}$$

By definition $\delta(t)$ is zero everywhere except for $t = 0$, the dirac delta function is related to the unit step function as

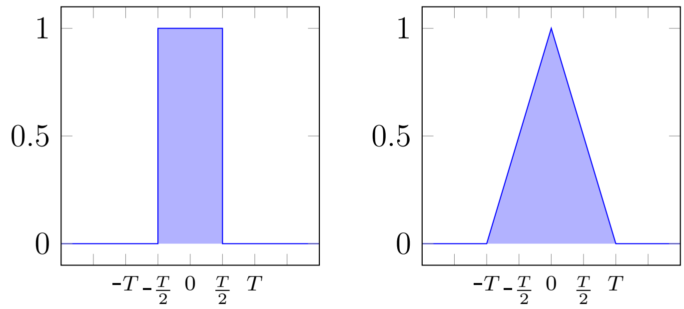
$$u(t) = \int_{\tau:-\infty}^{t} \delta(\tau) d\tau$$
$$\frac{du(t)}{dt} = \delta(t)$$

Rectangular and Triangular Pulses

The following waveforms frequently occur in communication, special symbols are defined

Definition Let $\Pi(\cdot)$ denotes a single rectangular pulse. Then

$$\Pi(\frac{t}{T}) = \begin{cases} 1, & |t| \leq \frac{T}{2} \\ 0, & |t| \geq \frac{T}{2} \end{cases}$$
$$\Lambda(\frac{t}{T}) = \begin{cases} 1 - \frac{|t|}{T}, & |t| \leq T \\ 0, & |t| \geq T \end{cases}$$



(a) Rect function (b) Tri function

Figure 1.5: Rectangular & Triangular functions.

1.2.2 Properties of Systems & Convolution

The convolution operation is used to evaluate the output of a linear system.

Definition: The convolution of a waveform $w_1(t)$ with a waveform $w_2(t)$ to produce a third waveform $w_3(t)$

$$x_3(t) = x_1(t) * x_2(t) = \int_{\tau:-\infty}^{+\infty} x_1(\tau)x_2(t-\tau)d\tau$$

The convolution is also referred to as convolution sum or superposition sum. Upon careful observation we see that t is the parameter of output function and τ is the variable of integration. The convolution sum for discrete case is

$$x_3[k] = x_1[k] * x_2[k] = \sum_{n:-\infty}^{+\infty} x_1[n]x_2[k-n]$$

The convolved output is obtained by

1. Time reversal of x_2 to obtain $x_2(-\tau)$.
2. Time shifting of x_2 by t to obtain $x_2(t-\tau)$.
3. Multiplying this result by x_1 to form the integrand $x_1(\tau)x_2(t-\tau)$

More details in **examples**

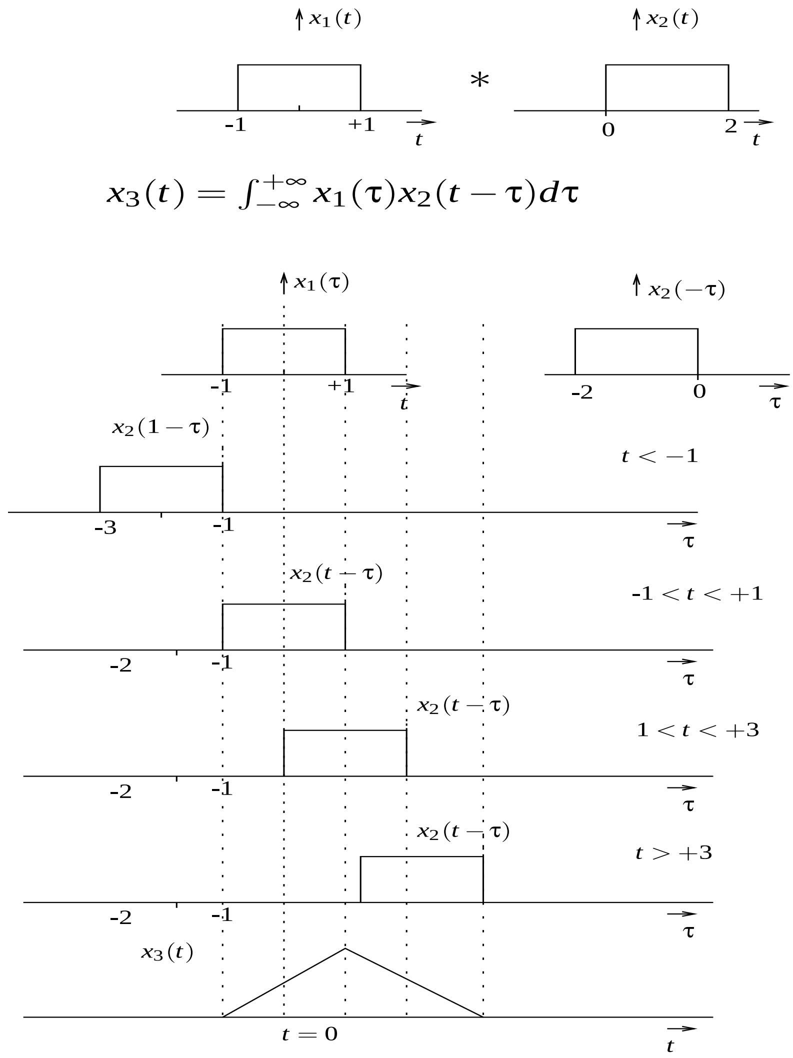
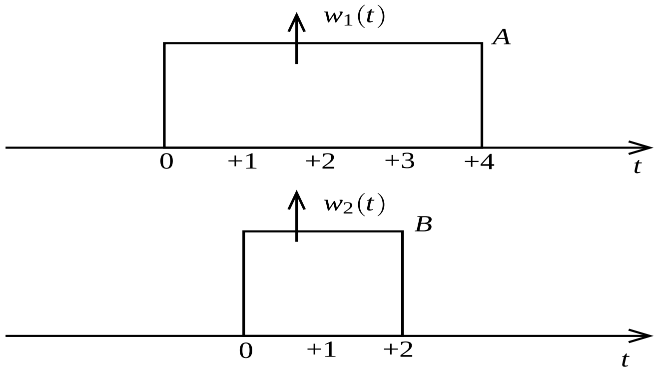
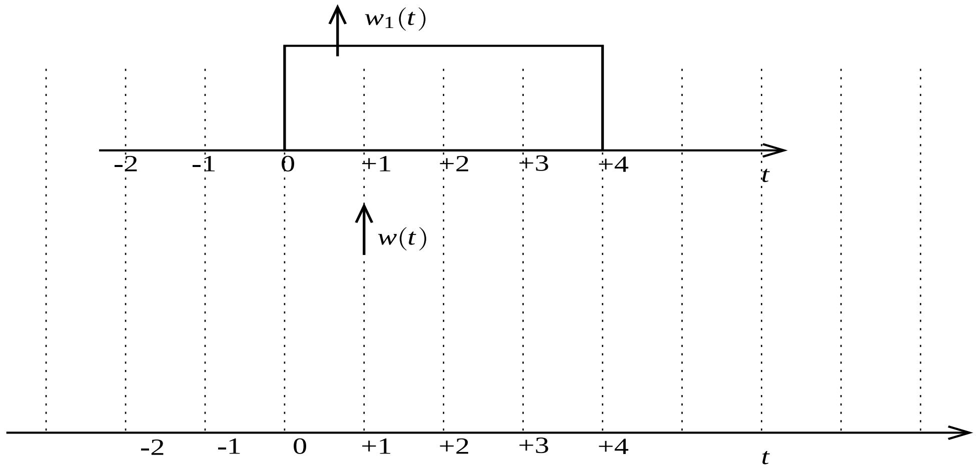
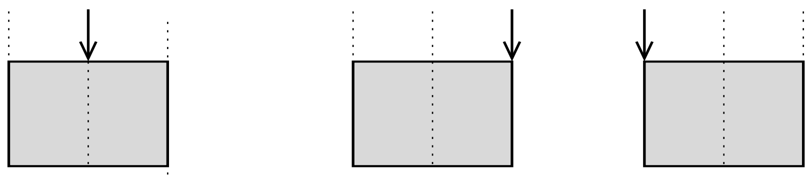


Figure 1.6: Convolution illustrated

Convolve the signals $w_1(t)$ and $w_2(t)$ as illustrated in the figure below



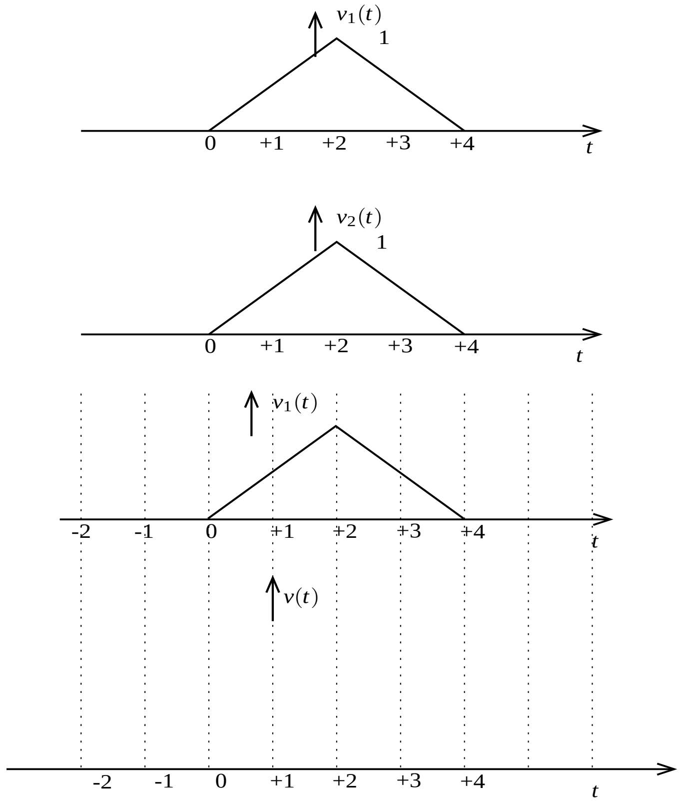


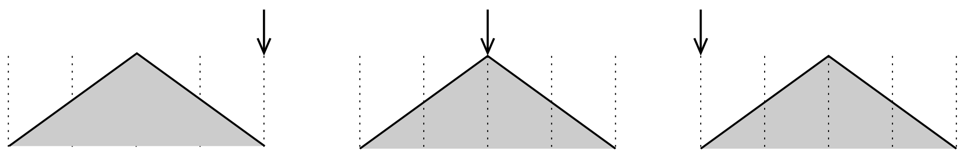


Solution:

$$w(t) = \begin{cases} & , -1 \leq t < +1 \\ AB & , +1 \leq t < +3 \\ & , +3 \leq t < 5 \end{cases}$$

Convolve the signals $w_1(t)$ and $w_2(t)$ as illustrated in the figure below

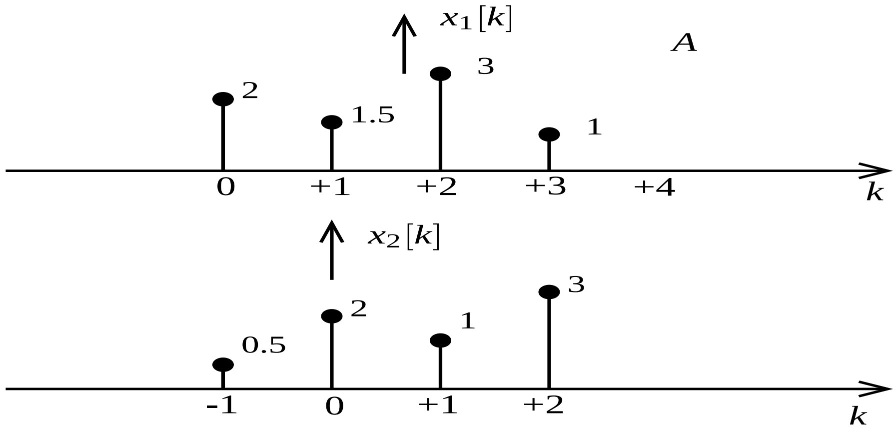


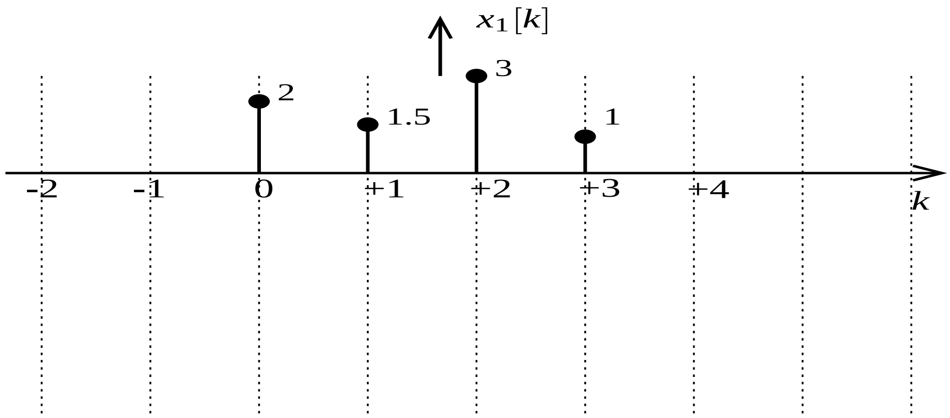


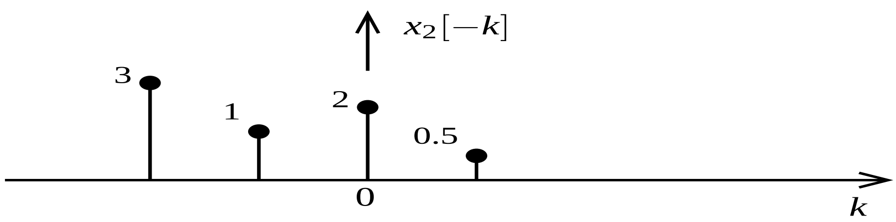
Solution:

$$w(t) = \begin{cases} & , -2 \leq t < +2 \\ 2 & , t = +2 \\ & , +2 \leq t < 6 \end{cases}$$

Convolve the signals $x_1[k]$ and $x_2[k]$ as illustrated in the figure below







Solution: This problem can be verified from the examples studied earlier.

1.2.3 Matrix representation of convolution operation

The discrete time convolution can also be conveniently expressed as a matrix operation.

$$\begin{bmatrix} y[-3] \\ y[-2] \\ y[-1] \\ y[0] \\ y[1] \\ y[2] \\ y[3] \end{bmatrix} = \begin{bmatrix} h[0] & 0 & & & & & \\ h[1] & h[0] & \ddots & & & & \\ h[2] & h[1] & h[0] & \ddots & & & \\ h[3] & h[2] & h[1] & h[0] & & & \\ 0 & h[3] & h[2] & h[1] & h[0] & & \\ & \ddots & h[3] & h[2] & h[1] & h[0] & \\ & & & h[3] & h[2] & h[1] & h[0] \end{bmatrix} \begin{bmatrix} x[-3] \\ x[-2] \\ x[-1] \\ x[0] \\ x[1] \\ x[2] \\ x[3] \end{bmatrix} \tag{1.2}$$

The above operation can be compactly expressed as vector notation as

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{1.3}$$

1.3 Sampling of Analog Signals

Sampling theorem is one of the most useful theorems it applies to digital communication and processing systems.

Theorem: Any bandlimited signal can be accurately reconstructed it is sampled at frequency which is atleast twice the maximum frequency of the signal.

Note:
Not true for bandpass signals.

$$F_s \geq 2F_{max}$$

The sampled signal $x(t)$ can be represented as

$$x(t) = \sum_{n:-\infty}^{+\infty} a_n \frac{\sin\{\pi f_s(t - (n/f_s))\}}{\pi f_s[t - (n/f_s)]}$$

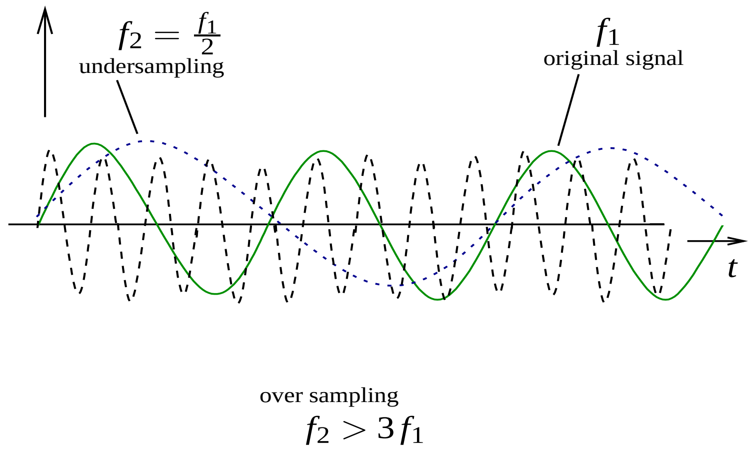


Figure 1.7: Illustration of under and over sampling or a sinusoidal waveform.

where

$$a_n = f_s \int_{-\infty}^{+\infty} x(t) \frac{\sin\{\pi f_s(t - (n/f_s))\}}{\pi f_s[t - (n/f_s)]} dt$$

Extra Reading

For further detail refer Couch section 2.7

Sampling is in fact multiplication of our continuous time domain signal with periodic impulses. The switching devices yields instantaneous values if $\Delta t \to 0$ and we have

$$x_s(t) = s(t) \cdot x(t)$$

here $s(t)$ is defined as

$$s(t) = \sum_{k:-\infty}^{+\infty} \delta(t - kT_s)$$

which is good for theoretical considerations and practically we have

$$s(t) = \sum_{k:-\infty}^{+\infty} \Pi\left(\frac{t - kT_s}{T_s}\right)$$

The ideal sampled wave is then

$$
\begin{aligned}
x_s(t) &\triangleq x(t)s_\delta(t) \\
&= x(t)\sum_{k:-\infty}^{+\infty}\delta(t-kT_s) \\
&= \sum_{k:-\infty}^{+\infty}x(kT_s)\delta(t-kT_s)
\end{aligned}
$$

To obtain the corresponding spectrum $X_\delta(f) = \mathcal{F}[x_s(t)]$

$$X_s(j\Omega) = \sum_{n:-\infty}^{+\infty} x_a(nT_s)e^{-jn\Omega T_s}$$

Another expression for $X(j\Omega)$ can be determined by noting that the Fourier transform of $s(t)$ is

$$S(j\Omega) = \frac{2\pi}{T}\sum_{k:-\infty}^{+\infty}\delta(\Omega - k\Omega_s)$$

where $\Omega_s = 2\pi/T_s$ is sampling frequency in radians per second. Therefore

$$X(e^{j\omega}) = \frac{1}{2\pi}X_a(j\Omega) * S_a(j\Omega) = \frac{1}{T}_s\sum_{k:-\infty}^{+\infty}X_a(j\Omega - jk\Omega_s)$$

1.3.1 Fourier transforms and properties

Definition: Spectrum is the frequency domain representation of time domain signals. Fourier transform is a general method of finding frequencies of a waveform.

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

where $\mathcal{F}[\cdot]$ denotes the Fourier transform of $[\cdot]$ and f is the frequency parameter with units Hz. This defines the term frequency which is a parameter in the Fourier

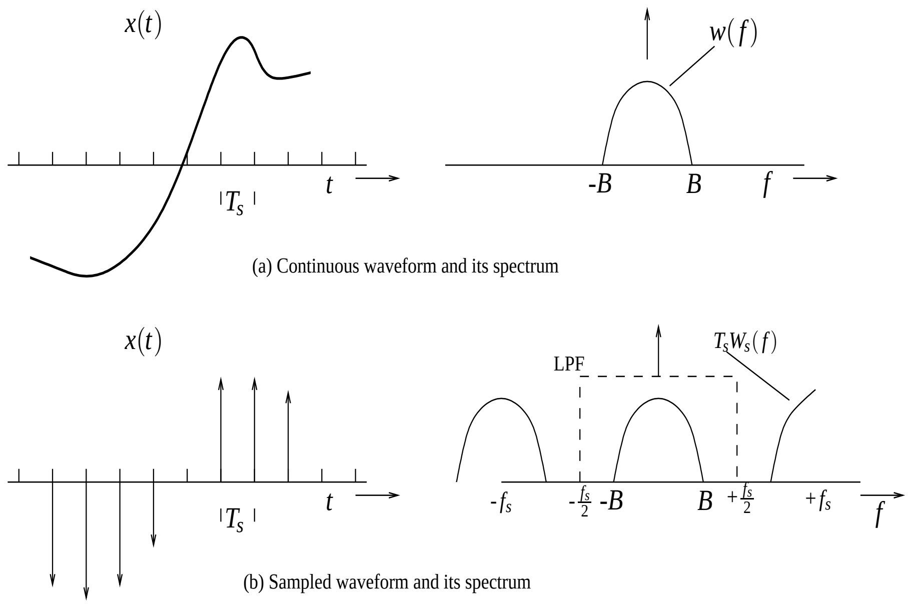


Figure 1.8: Ideal sampling illustrated

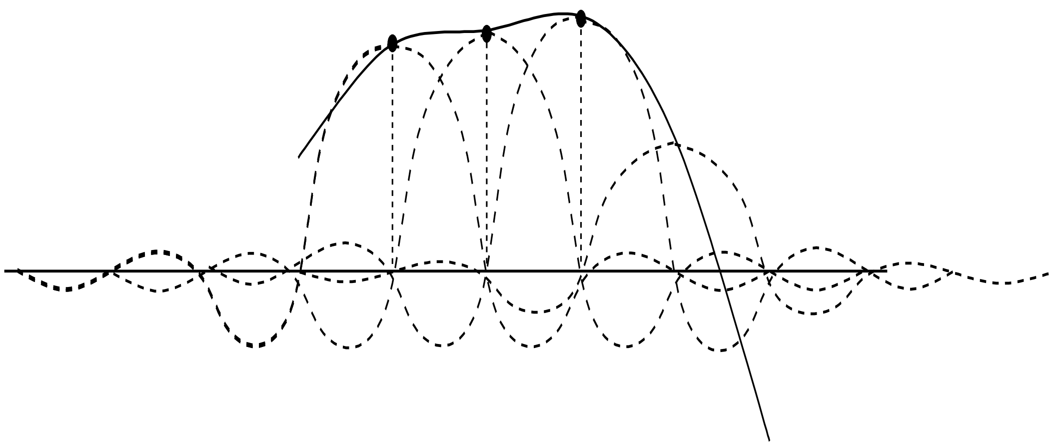


Figure 1.9: Ideal reconstruction / interpolation illustrated

transform.

$X(f)$ is called the two sided spectrum of $x(t)$, because both positive and negative frequency components are obtained by a mathematical calculation and do not have physical interpretation.

Since the base of FT is complex exponential $e^{-j2\pi ft}$, $X(f)$ is a complex function of frequencies. $X(f)$ may be decomposed into two real functions $X_r(f)$ and $X_i(f)$ such that

$$X(f) = X_r(f) + jX_i(f)$$

Which is identical to writing a complex number is cartesian system, the equivalent polar representation would be

$$X(f) = |X(f)|e^{j\theta(f)}$$

that is

$$|X(f)| = \sqrt{X_r^2(f) + X_i^2(f)} \quad \text{and} \qquad \theta(f) = \tan^{-1}\left(\frac{X_i(f)}{X_r(f)}\right)$$

This is called *magnitude-phase* form or polar form. To determine whether certain frequency components are present, one would examine the magnitude spectrum $|X(f)|$.

The time domain waveform can be calculated from the spectrum by using the inverse Fourier transform.

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

The functions $x(t)$ and $X(f)$ are said to be Fourier transform pair. $x(t) \circ\!\!-\!\!- X(f)$.

A waveform is Fourier transformable, if it satisfies both *Dirichlet conditions*

- Over any time interval of finite width, the function $w(t)$ is single valued with a finite number of maxima and minima and number of discontinuities if any is finite.

- $x(t)$ is absolutely integrable. That is

$$\int_{t=-\infty}^{+\infty} |x(t)|dt < \infty$$

Type	Analysis equation	Synthesis equation
Exponential Form	$c_k = \frac{1}{T_0}\int_{T_0} x(t)e^{j\omega_0 t}dt$	$x(t) = \sum_{k=-\infty}^{+\infty} c_k e^{j\omega_0 t}$
Trigonometric Series	$a_k = \frac{2}{T_0}\int_{T_0} x(t)\cos k\omega_0 t dt$ $b_k = \frac{2}{T_0}\int_{T_0} x(t)\cos k\omega_0 t dt$	$x(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos k\omega_0 t$ $+ b_k \sin \omega_0 t)$
Fourier integral	$X(\omega) = \int_{t=-\infty}^{+\infty} x(t)e^{-j\omega t}dt$	$x(t) = \int_{\omega=-\infty}^{+\infty} X(\omega)e^{j\omega t}d\omega$
Discrete time Fourier transform	$X(e^{j\Omega}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-jk\omega_0 n}$	$x[n] = \int_{-\pi}^{\pi} X(e^{j\Omega})e^{-j\Omega n}$
Discrete Fourier transform	$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi nk}{N}}$	$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j\frac{2\pi nk}{N}}$

Table 1.1: Standard pairs of Fourier transforms.

A weaker sufficient conditions for the existence of the Fourier transform is

$$E_x = \int_{t=-\infty}^{+\infty} |x(t)|^2 dt < \infty$$

1.3.2 Properties of Fourier transform

Linearity: If

$$x(t) \circ\!\!- X(f)$$

and

$$y(t) \circ\!\!- Y(f)$$

then

$$ax(t) + by(t) \circ\!\!-\!\!-\; aX(f) + bY(f)$$

This can be verified directly from the definition of Fourier transform.

Time Shift: If

$$x(t) \circ\!\!-\!\!-\; X(f)$$

then

$$x(t - t_0) \circ\!\!-\!\!-\; e^{j\omega t_0} X(f)$$

We verify this property

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

By replacing t with $t - t_0$, we obtain

$$\begin{aligned} x(t - t_0) &= \int_{f=-\infty}^{+\infty} X(f) e^{j2\pi f(t-t_0)} df \\ &= \int_{f=-\infty}^{+\infty} e^{-j2\pi f t_0} X(f) e^{j2\pi ft} df \end{aligned}$$

which is

$$\mathcal{F}\{x(t - t_0)\} = e^{-j\omega t_0} X(f)$$

Delay / advance in time on a signal introduces a phase shift to its transform namely $e^{-j\omega t_0}$, which is a linear function of ω.

Symmetry: Spectral symmetry of a real signal. If $w(t)$ is real then

$$W(-f) = W^*(f)$$

Proof: from the definition of Fourier transform

$$\begin{aligned} X(-f) &= \int_{t=-\infty}^{+\infty} x(t) e^{j2\pi ft} dt \\ X^*(f) &= \int_{t=-\infty}^{+\infty} x^*(t) e^{j2\pi ft} dt \end{aligned}$$

But since $x^*(t) = x(t)$, the right hand side matches the left hand side. An other important consequence of this proof is that for the spectrum of a real signal $x(t)$, the magnitude spectrum is even about origin i.e. $f = 0$

$$\begin{aligned} |X(-f)| &= |X(f)| \\ \theta(-f) &= -\theta(f) \end{aligned}$$

Time Reversal:

$$x(-t) \circ\!\!-\; X(-f)$$

Proof in the next property.

Scaling:

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

letting $t_1 = at$ as assuming $a > 0$ we obtain

$$\mathcal{F}[x(at)] = \int_{t_1=-\infty}^{+\infty} \frac{1}{a} x(t_1) e^{-j2\pi(f/a)t_1} dt_1 = \frac{1}{a} X\left(\frac{f}{a}\right)$$

For $a < 0$ we have

$$\mathcal{F}[x(at)] = \int_{t_1=-\infty}^{+\infty} -\frac{1}{a} x(t_1) e^{j2\pi(f/a)t_1} dt_1 = \frac{1}{|a|} X\left(\frac{f}{a}\right)$$

Thus for $a < 0$ or $a > 0$ we get

$$x(at) \circ\!\!-\; \frac{1}{|a|} X\left(\frac{f}{a}\right)$$

Convolution

$$x_1(t) * x_2(t) \circ\!\!-\!\!- X_1(f) \cdot X_2(f)$$

This equation is referred to as convolution theorem, which states that convolution in time domain becomes multiplication in frequency domain. Proof is not discussed here.

Parseval's Theorem

$$\int_{t=-\infty}^{+\infty} x_1(t)x_2^*(t)dt = \int_{f=-\infty}^{+\infty} X_1(f)X_2^*(f)df$$

Proof:

$$\begin{aligned} E = \int_{t=-\infty}^{+\infty} |x(t)|^2 dt &= \int_{f=-\infty}^{+\infty} |X(f)|^2 df \\ \int_{t=-\infty}^{+\infty} x_1(t)x_2^*(t)dt &= \int_{t=-\infty}^{+\infty} \left[\int_{f=-\infty}^{+\infty} X_1(f)e^{j2\pi ft}df\right] x_2^*(t)dt \\ &= \int_{t=-\infty}^{+\infty}\int_{f=-\infty}^{+\infty} X_1(f)X_2^*(t)e^{j2\pi ft}df\,dt \end{aligned}$$

interchanging the order of integration between f and t

$$\int_{t=-\infty}^{+\infty} x_1(t)x_2^*(t) = \int_{f=-\infty}^{+\infty} X_1(f)\left[\int_{t=-\infty}^{+\infty} x_2(t)e^{-j2\pi ft}dt\right]^* df$$

Additional Properties

If $x(t)$ is real and

$$x(t) = x_e(t) + x_o(t)$$

where $x_e(t)$ and $x_o(t)$ are even and odd components of $x(t)$, Thus

$$x(t) \circ\!\!-\!\!-\, X(f) = A(f) + jB(f)$$
$$X(-f) = X^*(f)$$
$$x_e(t) \circ\!\!-\!\!-\, \Re e\{X(f)\} = A\{f\}$$
$$x_o(t) \circ\!\!-\!\!-\, \Im mg\{X(f)\} = jB\{f\}$$

1.3.3 Discrete Fourier Transform

With convenience of PC and availability of DSPs, the spectrum of a waveform can be easily approximated by using the discrete fourier transform (DFT).

Definition: The discrete Fourier transform is defined as

$$X[n] = \sum_{k=0}^{N-1} x[k]e^{-j\frac{2\pi nk}{N}} \quad n = 0, \cdots, N-1$$

The inverse Fourier transform

$$x[k] = \sum_{n=0}^{N-1} X[n]e^{j\frac{2\pi nk}{N}} \quad k = 0, \cdots, N-1$$

where k and n are respectively the frequency and time indices. The DFT and IDFT calculate discrete frequency components.

Important Fourier transform pairs

Fourier Transform of Cosine wave

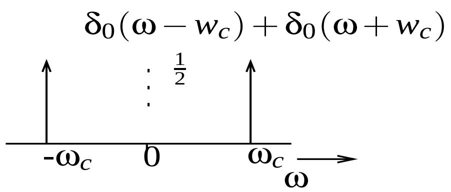


Figure 1.10: magnitude spectrum of cosine wave

$$X(j\omega) = \int_{-\infty}^{+\infty} \frac{e^{j(\omega_c t+\phi)} + e^{-j(\omega_c t+\phi)}}{2} e^{-j\omega t} dt = \frac{1}{2} e^{j\phi} \left[\int_{-\infty}^{+\infty} e^{-j(\omega-\omega_c)t} + e^{-j(\omega+\omega_c)t} dt \right]$$
$$= \frac{1}{2} e^{j\phi} [\delta_0(\omega - \omega_c) + \delta_0(\omega + \omega_c)]$$

Fourier Transform of Rectangular Pulse

$$X(f) = \int_{\frac{-T}{2}}^{\frac{+T}{2}} 1 e^{-j\omega t} dt = \frac{e^{j\omega T/2} - e^{j\omega T/2}}{j\omega}$$
$$= T \frac{\sin(\omega T/2)}{\omega T/2} = T \cdot \mathrm{sinc}(\frac{\omega T}{2}) \tag{1.4}$$

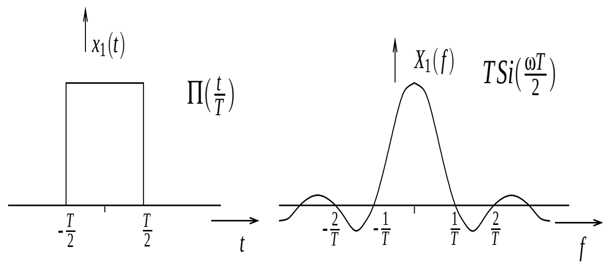


Figure 1.11: magnitude spectrum of Rect function

1.3.4 LTI Systems

Systems in the broadest sense are interconnection of components. In context of circuit theory and communication systems system can be viewed as a process in which input signals are transformed by the system or cause the system to respond in a way to produce other signals at the output.

A continuous-time system is a system in which continuous-time input signals are applied to produce continuous time output signals. A discrete-time system is a system which transforms discrete time input into discrete-time output.

$$x(t) \to y(t)$$
$$x[k] \to y[k]$$

Important properties of any continuous or discrete time system are linearity, causal-

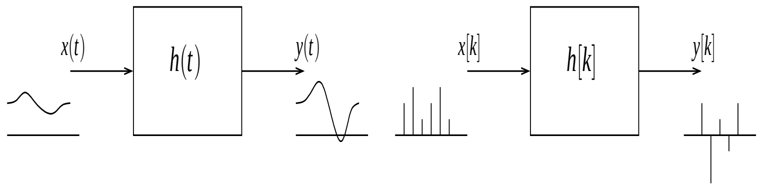


Figure 1.12: Discrete and continuous time systems.

ity, stability, time invariance. A detailed discussion of these issues in considered beyond the scope of this material.

Ideally we would like our communication channel to be an LTI but practical channels like underwater acoustic or wireless channels do not satisfy this criteria.

1.4 Correlation & Spectral Densities

From Parseval's theorem we know that the energy of a signal can be related to spectrum of signal as

$$
\begin{aligned}
E_x = \int_{-\infty}^{+\infty} x(t)x^*(t)dt &= \int_{-\infty}^{+\infty} x(t)\left[\frac{1}{2\pi}\int_{-\infty}^{+\infty} X^*(\omega)e^{-j\omega t}d\omega\right]dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} X^*(\omega)\left[\int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt\right]d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)X^*(\omega)d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega
\end{aligned}
$$

The contribution of a spectral component of frequency ω is proportional to $|G(\omega)|^2$. The total energy of output signal $Y(\omega)$ which is obtained by filtering $X(\omega)$ through a filter with frequency response $H(\omega)$ is

$$
E_y = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |X(\omega)H(\omega)|^2 d\omega
$$

Correlation function

Correlation is a way to compare two signals $x(t)$ and $y(t) \rightarrow r_{xy}(\tau)$. The correlation of a signal with itself is defined as *auto-correlation function*$\rightarrow r_{xx}(t)$ and correlation between different signals is defined as *cross-correlation function*$\rightarrow r_{xy}(t)$.

Auto-correlation function is defined as

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

setting $\nu = t + \tau$, we get

$$r_{xx}(\tau) = \int_{-\infty}^{+\infty} x^*(\nu - \tau)x(\nu)d\nu$$

In the above equation ν is a dummy variable and could be replaced by t, therefore

$$= \int_{t:-\infty}^{+\infty} x^*(t)x(t \pm \tau)dt$$

This shows that for a real $x(t)$ the autocorrelation is an even function of τ,i.e.

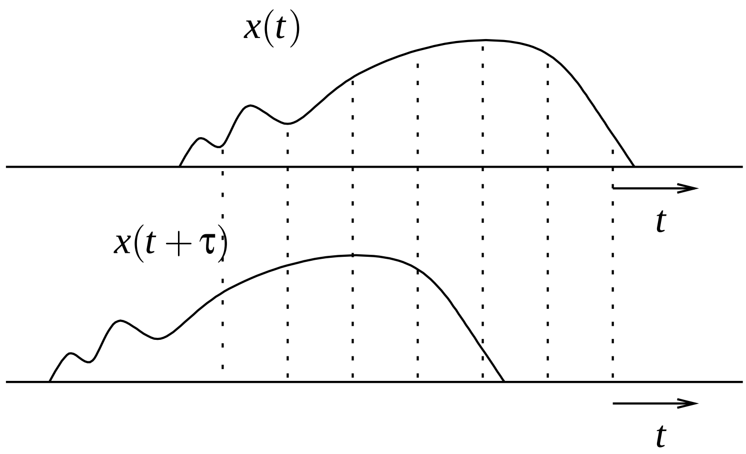


Figure 1.13: Autocorrelation is multiplication and integration with a time shifted version

$$r_{xx}(\tau) = r_{xx}(-\tau)$$

The ESD $\psi(\omega) = |X(\omega)|^2$ is the fourier transform of auto-correlation function $r_{xx}(\tau)$.

$$\begin{aligned}\mathcal{F}[r_{xx}(\tau)] &= \int_{\tau=-\infty}^{+\infty} e^{-j\omega\tau}\left[\int_{t=-\infty}^{+\infty} x^*(t)x(t+\tau)dt\right]d\tau \\ &= \int_{t=-\infty}^{+\infty} x^*(t)\left[\int_{\tau=-\infty}^{+\infty} x(t+\tau)e^{-j\omega\tau}d\tau\right]dt\end{aligned}$$

The inner integral is the Fourier transform of $x(t+\tau)$ which is $x(\tau)$ shifted by t

$$\mathcal{F}[r_{xx}(\tau)] = X(\omega)\int_{t=-\infty}^{+\infty} x^*(t)e^{j\omega t}dt = X(\omega)X^*(\omega) = |X(\omega)|^2$$

$$r_{xx}(\tau) \circ\!\!- \psi_{xx}(\omega) = |X(\omega)|^2$$

A careful observation of the correlation operation shows close connection to convolution. Indeed auto-correlation function $r_{xx}(\tau)$ is convolution of $g(\tau)$ with $g^*(-\tau)$ because,

$$g(\tau) * g^*(-\tau) = \int_{x=-\infty}^{+\infty} g(x)g^*[-(\tau - x)]dx = \int_{x=-\infty}^{+\infty} g(x)g^*(x-\tau)dx = r_{gg}(\tau)$$

For a power signal, a meaningful measure of its size is its power →'*time average of the signal energy averaged over the finite time interval real signal $x(t)$ is given by*'

$$P_x = \lim_{T\to\infty}\frac{1}{T}\int_{t=-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)dt$$

Consider a signal $x(t)$ which is periodic

$$x_T(t) = \begin{cases} x(t), & -T/2 < t < T/2 \\ 0, & \text{elsewhere} \end{cases} = x(t)\cdot\Pi(\frac{t}{T})$$

Using definition of power signal we obtain

$$P = \lim_{T\to\infty}\frac{1}{T}\int_{t=-\frac{T}{2}}^{+\frac{T}{2}} x^2(t)dt = \lim_{T\to\infty}\frac{1}{T}\int_{t=-\infty}^{+\infty} x^2(t)dt$$

If $x(t)$ is a power signal, then its power is finite, and the truncated signal $x_T(t)$ is an energy signal as long as T is finite. If $x_T(t) \circ\!\!- X_T(\omega)$, then according to Parseval's theorem

$$E_{x_T} = \int_{t=-\infty}^{+\infty} x_T^2(t)dt = \frac{1}{2\pi}\int_{\omega:-\infty}^{+\infty} |X_T(\omega)|^2 d\omega$$

Hence. P_x, the power of $x(t)$ is given by

$$P_x = \lim_{T\to\infty}\frac{E_{x_T}}{T} = \lim_{T\to\infty}\frac{1}{T}\left[\frac{1}{2\pi}\int_{\omega=-\infty}^{+\infty} |X_T(\omega)|^2 d\omega\right]$$

As T increases, the duration of $x_T(t)$ increases, and its energy E_{xT} also increases proportionately. This means $|X_T(\omega)|^2$ also increases with T and as $T \to \infty$ $|X_T(\omega)|^2 \to \infty$. Since $x(t)$ is a power signal, the integral on the right side must converge, convergence permits us to interchange the order of the limiting process

$$P_x = \int_{\omega=-\infty}^{+\infty} \lim_{T\to\infty}\frac{|X_T(\omega)|^2}{T} d\omega$$

We define **Power Spectral Density** $S_x(\omega)$ as

$$S_{xx}(\omega) = \lim_{T\to\infty}\frac{|X_T(\omega)|^2}{T}$$

Observe that PSD is the time average of ESD, PSD is always a positive, real. If the $x(t)$ is a voltage signal, the units of PSD are volts squared per hertz.

Time Auto-correlation Function of Power Signals

The time autocorrelation function $r_{xx}(\tau)$ of a real power signal $x(t)$ is

$$r_{xx}(\tau) = \frac{1}{T}\int_{t=-T/2}^{+T/2} x^*(t)x(t+\tau)dt$$

From prior discussion we know that

$$r_{xx}(\tau) = \frac{1}{T}\int_{t=-T/2}^{+T/2} x(t)x^*(t-\tau)dt$$

which implies

$$r_{xx}(\tau) = r_{xx}(-\tau)$$

Interpretation of Power Spectral Density

Because the PSD is a time average of the ESD of $x(t)$. We can readily show that PSD $S_{xx}(\omega)$ represents the power per unit bandwidth (in hertz) of the spectral component at frequency ω. The spectral component within the band ω_1 to ω_2 is given by

$$\Delta P_x = \frac{1}{\pi} \int_{\omega_1}^{\omega_2} S_{xx}(\omega) d\omega$$

Input & Output Power Spectral Densities

Because the PSD is a time average of the ESD. If $x(t)$ and $y(t)$ are in the input and output signals of LTI system with transfer function $H(\omega)$ then

$$S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega)$$

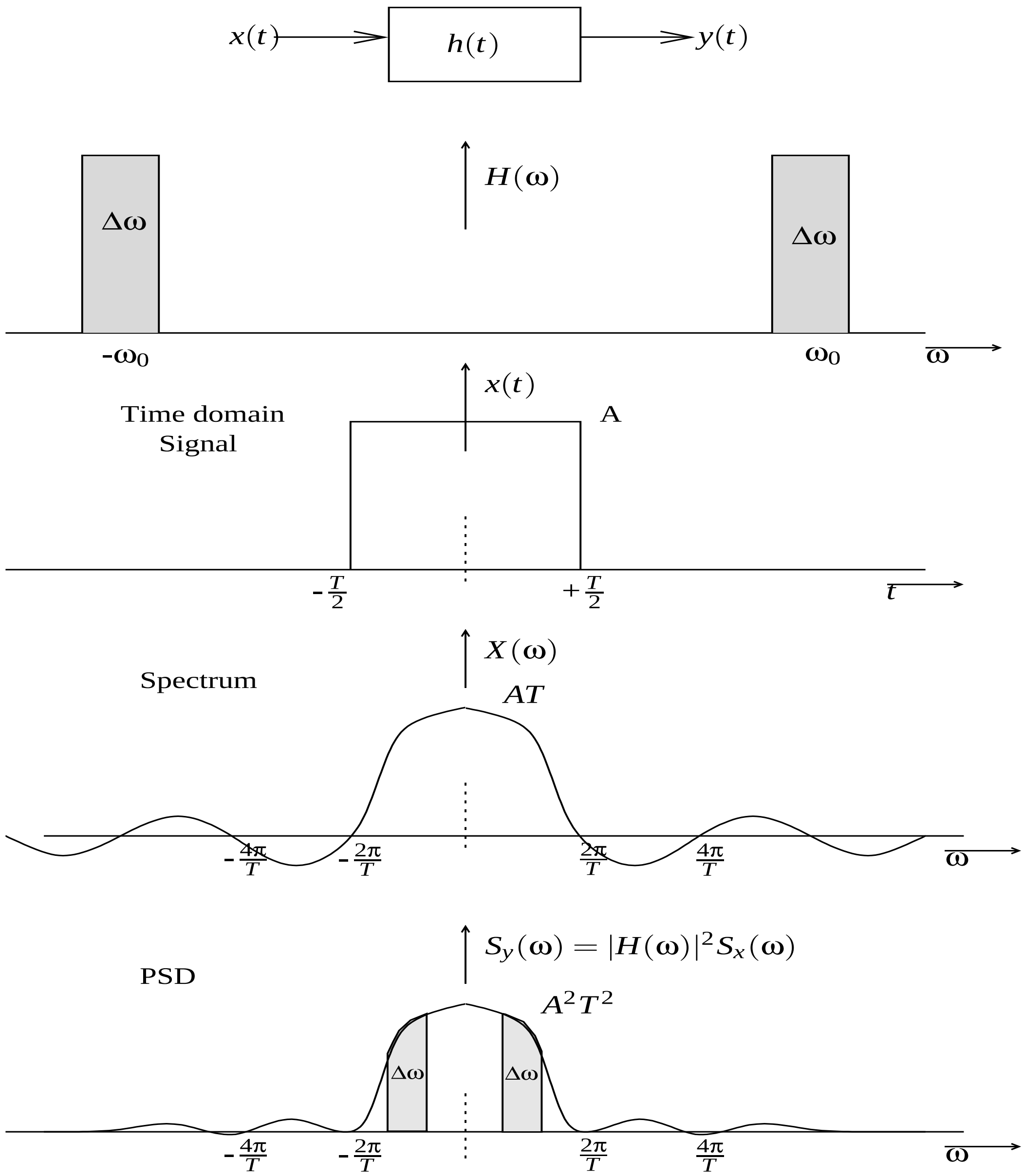


Figure 1.14: The concentration of power in unit spectrum

Power spectral density of periodic waveform

For a periodic waveform, the power spectral density (PSD) is given by

$$S_{xx}(\omega) = \sum_{n=-\infty}^{+\infty} |c_n|^2 \delta_0(\omega - n\omega_0)$$

Let x(t) be defined as

$$\begin{aligned}
x(t) &= \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_0 t} \\
r_{xx}(\tau) &= \langle x^*(t)x(t+\tau)\rangle \\
&= \langle \sum_{n=-\infty}^{+\infty} c_n^* e^{-jn\omega_0 t} \sum_{m=-\infty}^{+\infty} c_m e^{jm\omega_0 (t+\tau)}\rangle \\
&= \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} c_n^* c_m e^{-jn\omega_0 t} e^{jm\omega_0 (t+\tau)}
\end{aligned}$$

But $e^{j\omega_0 (n-m)t} = \delta_{mn}$, this $r_{xx}(\tau)$ reduces to

$$\begin{aligned}
r_{xx}(\tau) &= \sum_{n=-\infty}^{+\infty} |c_n|^2 e^{jn\omega_0 \tau} \\
S_{xx}(\omega) &= \mathfrak{F}[r_{xx}(\tau)] = \mathfrak{F}\left[\sum_{n:-\infty}^{+\infty} |c_n|^2 e^{jn\omega_0 \tau}\right] \\
&= \sum_{n=-\infty}^{+\infty} |c_n|^2 \mathfrak{F}[e^{jn\omega_0 \tau}] = \sum_{n=-\infty}^{+\infty} |c_n|^2 \delta(\omega - n\omega_0)
\end{aligned}$$

Chapter 2

Analog Communication

Outline

In this chapter the most commonly used analog modulation schemes are discussed. The objective of this study is to introduce the fundamental concepts of analog modulations schemes. A detailed discussion on hardware design is beyond the scope of this course. Interested students are referred to text books like Lyon Couch and Simon Haykin. The objective of this chapter is familiarize students with the following

- Amplitude modulation and demodulation.
- Phase modulation and demodulation.
- Frequency modulation and demodulation.
- Analog systems.

Basic Terminology

- **Modulation** is defined as a process by which some characteristic of a carrier is varied in accordance with the modulating wave.
- The message signal $m(t)$ is known as the **Modulating** signal.
- The carrier signal $c(t)$ is known as the **Modulated** signal.
- At the receiver the original signal is recovered through **demodulation** which is the inverse of the modulation process.

In this section we briefly touch up the analog schemes used for modulation of analog message.

2.1 Analog modulation schemes

2.1.1 Amplitude Modulation

Consider a sinusoidal carrier wave $c(t)$ defined as

$$c(t) = A_c \cos(2\pi f_c t) \tag{2.1}$$

Extra Reading

For further detail refer Couch section 5.1-3 Haykin section 2.2.

where the peak value A_c is the carrier amplitude and f_c is the carrier frequency. The phase of the carrier is not considered thus far. Let $m(t)$ denote the baseband message signal. Amplitude modulation is defined as a process in which amplitude of the carrier wave $c(t)$ is varied linearly with the message signal $m(t)$. The time domain representation of an amplitude modulated signal is defined as

$$s(t) = \underbrace{A_c\Big[1 + k_a m(t)\Big]}\cos(2\pi f_c t) \tag{2.2}$$

where k_a is a constant called amplitude sensitivity/modulation index of the modulator. The amplitude of the time function multiplying $\cos(2\pi f_c t)$ is called the envelope of the AM wave $s(t)$. This gain can be expressed equivalently as

$$a(t) = A_c|1 + k_a m(t)| \tag{2.3}$$

Two cases of particular interest arise depending on the magnitude of $k_a m(t)$, compared to unity. For case 1, we have

$$|k_a m(t)| \leq 1, \qquad \text{for all } t \tag{2.4}$$

in this condition, the term $1 + k_a m(t)$ is always non-negative. The expression can be expressed as

$$a(t) = A_c[1 + k_a m(t)] \qquad \text{for all } t \tag{2.5}$$

for the other case

$$|k_a m(t)| > 1, \text{for some } t \tag{2.6}$$

in this situation (2.3) must be used to determine envelope of modulated signal. The term $k_a m(t)$ can be conveniently termed as the modulation gain. In spectral representation the Fourier transform of the modulated signal $s(t)$ can be expressed in terms of modulating signal $m(t)$ as

$$\begin{aligned} S(f) = &\frac{A_c}{2}\{\delta(f - f_c) + \delta(f + f_c)\} \\ &+ \frac{K_a A_c}{2}\{M(f - f_c) + M(f - f_c)\} \end{aligned} \tag{2.7}$$

where $S(f)$ and $M(f)$ are the spectra of the modulated and modulating signal respectively. The spectrum of modulated signal is illustrated in Fig.2.1.

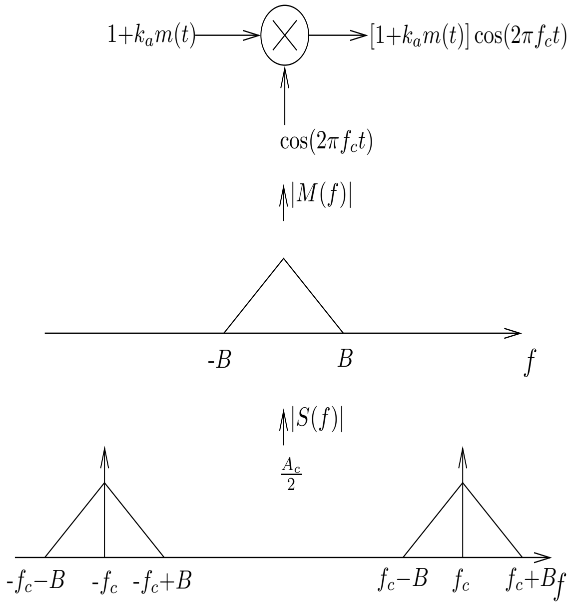
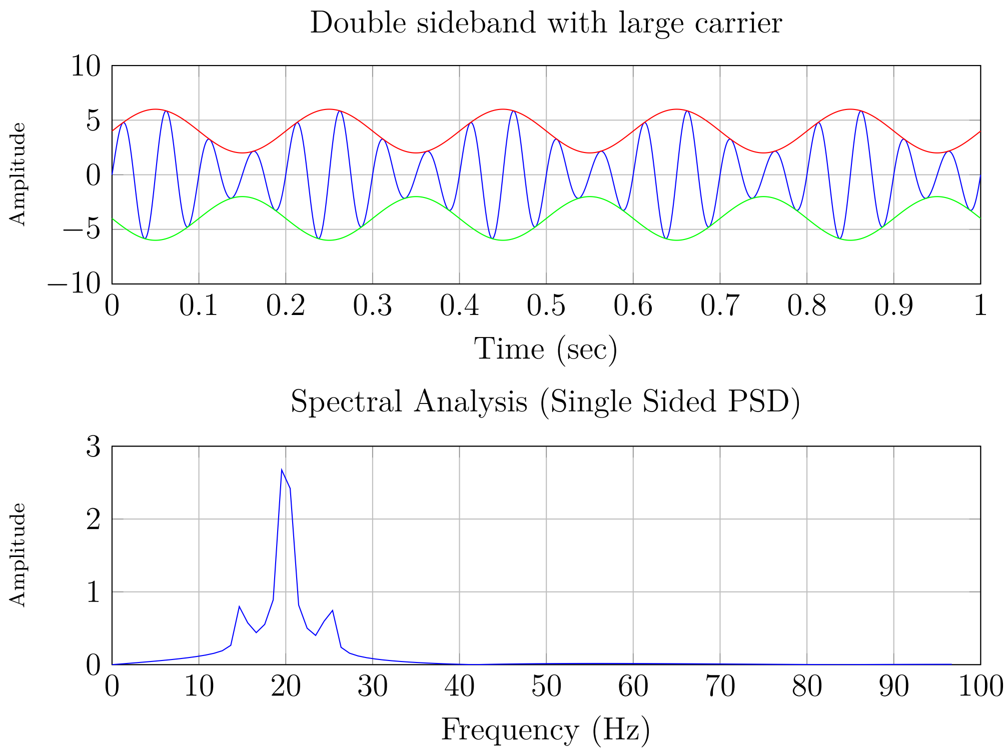
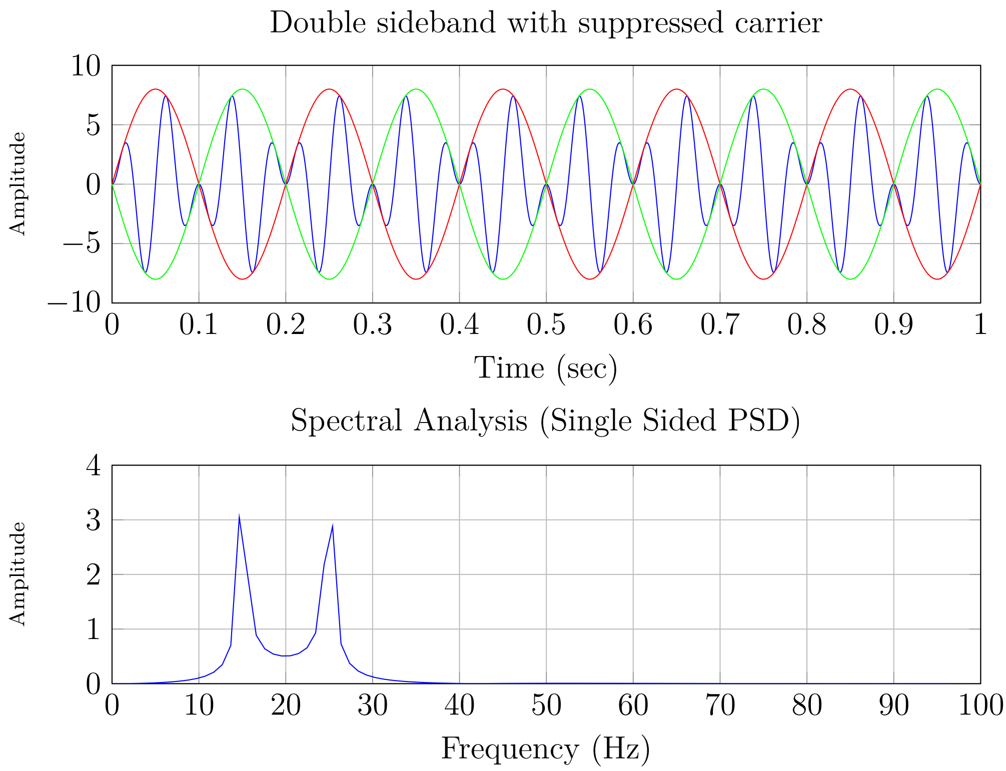


Figure 2.1: Spectral representation of AM modulated signal.

The graphical illustration of time domain AM modulated signal is presented in Fig.2.2



(a) Time and spectral domain representation of double sideband system with large carrier.



(b) Time and spectral domain representation of carrier suppressed system.

Figure 2.2: Time and spectral representation of DSB and CS amplitude modulation scheme.

An important parameter used to define the level of modulation is commonly known as '*modulation index*'. The modulation index may be defined as

$$k_a = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}} \tag{2.8}$$

1. The percentage modulation is less than 100%, so as to avoid envelope distortion.

2. The message bandwidth, W, is small compared to the carrier frequency f_c, so that the envelope $a(t)$ may be visualized satisfactorily.

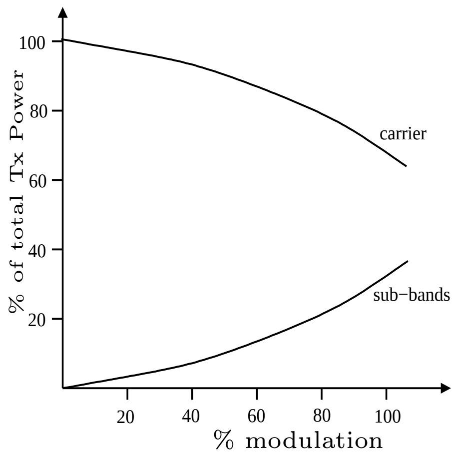


Figure 2.3: Power distribution between subcarrier and sideband with DSBAM.

Variants of AM

There exist several variants of amplitude modulation which offer bandwidth and power efficiency of transmission system.

Double-sideband suppressed-carrier(DSBSC)

The mathematical expression for DSBSC is defined as

$$\begin{aligned} s(t) &= c(t)m(t) \\ &= A_c \cos(2\pi f_c t)m(t) \end{aligned} \tag{2.9}$$

The modulated signal undergoes a phase reversal whenever message signal crosses zero. The envelope of the DSBSC is different from the actual message signal as illustrated in Fig. 2.2.

Single sideband(SSB)

Standard amplitude modulation and DSBSC modulation are wasteful of bandwidth because they both require bandwidth equal to twice the message bandwidth. However using the fact that upper and lower sidebands are uniquely related to each other by virtue of symmetry about carrier frequency.

$$
\begin{aligned}
s_1(t) =& A_c[1 + k_a m(t)] \cos(2\pi f_c t) \\
s_2(t) =& A_c[1 - k_a m(t)] \cos(2\pi f_c t)
\end{aligned}
$$

subtracting $s_2(t)$ from $s_1(t)$ we have

$$
\begin{aligned}
s(t) =& s_1(t) - s_2(t) \\
& 2k_a A_c \cos(2\pi f_c t) m(t) \qquad (2.10)
\end{aligned}
$$

The detailed discussion on generation of these special AM signals is not pursued here. The graphical illustration of these signals is presented in Fig. 2.4.

Note:
TV broadcasts use another variant of AM known as Vestigial sideband AM

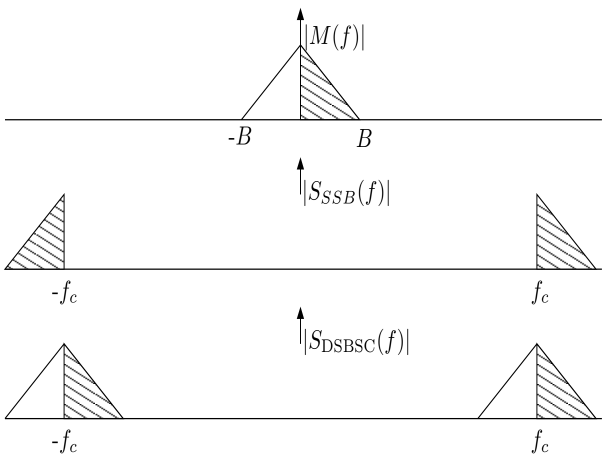


Figure 2.4: Spectral representation of DSB-SC and SSB amplitude modulated signals.

Generation of AM signal

There exist a number of possible devices which can AM modulate the incoming signal for example square law and switching modulator. In the following we just discuss square law modulator briefly.

This modulator consists of a summing device adding up the carrier and the modulating waves, a nonlinear element and a bandpass filter.

The sum of the modulating and modulated signals is applied to a nonlinear device such as a (suitably biased) diode. The outcome of the diode can be defined as

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t) \tag{2.11}$$

where a_1 and a_2 are constant. The input voltage signal $v_1(t)$ is the sum of the carrier wave and the modulating signal

$$v_1(t) = A_c \cos(2\pi f_c t) + m(t) \tag{2.12}$$

Therefore rewriting $v_2(t)$ in terms of $v_1(t)$ the output signal is

$$v_2(t) = \underbrace{a_1 A_c \left[1 + \frac{2a_2}{a_1} m(t)\right] \cos(2\pi f_c t)}_{}$$

$$+ a_1 m(t) + a_2 m^2(t) + a_2 A_c^2 cos^2(2\pi f_c t) \tag{2.13}$$

Demodulation of AM signal

An envelope detector is a simple and yet highly efficient device. This device is suitable when modulation index is less than 100%. Ideally an envelope detector produces an output signal that follows the envelope of the input waveform exactly. The circuit diagram of an envelope detector is illustrated in Fig.2.5.

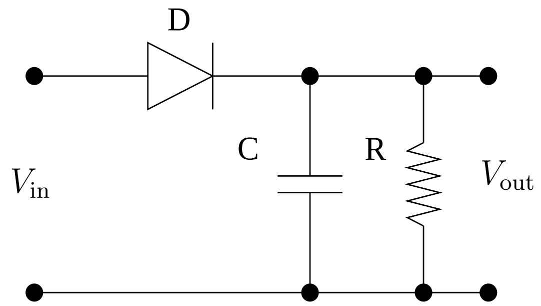


Figure 2.5: The electrical schematic of a simple amplitude demodulator (envelope detector).

The signal can be demodulated using the square law modulator technique. Once again consider the transfer characteristics of nonlinear device

$$v_2(t) = a_1 v_1(t) + a_2 v_1^2(t) \tag{2.14}$$

where $v_1(t)$ and $v_2(t)$ are the input and output voltages, respectively and a_1 and a_2. When such device is used to demodulation of an AM signal, at its input we have

$$v_1(t) = A_c[1 + k_a m(t)]\cos(2\pi f_c t) \tag{2.15}$$

substituting (2.15) in (2.14) we get

$$\begin{aligned} v_2(t) = {} & a_1 A_c[1 + k_a m(t)]\cos(2\pi f_c t) \\ & + \frac{1}{2} a_2 A_c^2 [1 + 2k_a m(t) + k_a^2 m^2(t)][1 + \cos(4\pi f_c t)] \end{aligned} \tag{2.16}$$

where desired signal is $a_2 A_c^2 k_a m(t)$ can be obtained through a low pass filter, however another interfering term $\frac{1}{2} a_2 A_c^2 k_a^2 m^2(t)$ will give rise to inband distortion.

Representation of FM and PM signals

Phase modulation (PM) and frequency modulation (FM) are special cases of *angle modulated* signalling. In this kind of signalling the complex envelope is

$$g(t) = A_c e^{j\theta(t)} \tag{2.17}$$

Here the real envelope $|g(t)|=A_c$ is a constant and the phase $\theta(t)$ is a linear function of the modulating signal $m(t)$. However, $g(t)$ is nonlinear function of the modulation. The resulting angle-modulated signal is defined as

$$s(t) = A_c \cos[\omega_c t + \theta(t)] \tag{2.18}$$

For PM, the phase is directly proportional to the modulating signal that is

$$\theta(t) = D_p m(t) \tag{2.19}$$

where the proportionality constant D_p is the phase sensitivity of the phase modulator, having units of radians per volt. For FM, the phase is proportional to the integral of $m(t)$ so that

Note:
m_p and m_f is the modulating message for phase and frequency modulation respectively.

$$\theta(t) = D_f \int_{\tau=-\infty}^{t} m(\tau) d\tau \tag{2.20}$$

where the frequency deviation constant D_f has units of radians/volt-second. We see that if we have PM modulated signal $m_p(t)$, there is also FM on the signal, corresponding to a different modulation waveshape that is given by

$$m_f(t) = \frac{D_p}{D_f}\left[\frac{dm_p(t)}{dt}\right] \tag{2.21}$$

where subscripts f and p denote frequency and phase, respectively. Similarly, for an FM signal modulated $m_f(t)$, the corresponding phase modulated signal is

$$m_p(t) = \frac{D_f}{D_p}\int_{-\infty}^{t} m_f(\sigma) d\sigma \tag{2.22}$$

PM circuit may be used the synthesize an FM circuit by inserting an integrator in cascade with the phase modulator input. Direct PM circuit is realized by passing an unmodulated sinusoidal signal through a time-varying circuit which introduces a phase shift that varies with the applied modulating voltage.

DEFINITION: If a bandpass signal is represented by

$$s(t) = R(t)\cos(\psi(t)) \tag{2.23}$$

where $\psi(t)=\omega_c t+\theta(t)$, then the instantaneous frequency of $s(t)$ is

$$\begin{aligned} f_i(t) &= \frac{1}{2\pi}\omega_i(t) = \frac{1}{2\pi}\left[\frac{d\psi(t)}{dt}\right] \\ &= f_c + \frac{1}{2\pi}\left[\frac{d\theta(t)}{dt}\right] \\ &= f_c + \frac{1}{2\pi}D_f m(t) \end{aligned}$$

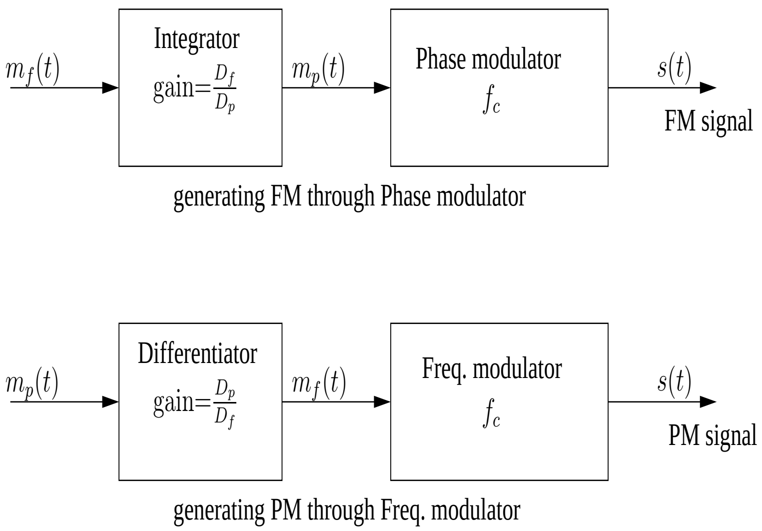


Figure 2.6: Interconnection of FM and PM signal generation processes.

2.1.2 Phase Modulation

Extra Reading
For further detail refer
Couch section 5.5
Haykin section 2.6.

A phase modulated signal may be expressed as

$$s(t) = A_c\cos[\theta(t)] \tag{2.24}$$

A_c is a constant carrier amplitude and $\theta(t)$ is varied by message signal $m(t)$. The mathematical form of this variation is determined by the type of angle modulation of the interest. The change in frequency over certain duration Δt depends on change in phase

$$f_{\Delta t}(t) = \frac{\theta(t + \Delta t) - \theta(t)}{2\pi \Delta t} \tag{2.25}$$

Instantaneous frequency of the angle-modulated wave $s(t)$

$$\begin{aligned} f_i(t) &= \lim_{\Delta t \to 0} f_{\Delta t}(t) \\ &= \lim_{\Delta t \to 0} \left[\frac{\theta(t + \Delta) - \theta(t)}{2\pi \Delta t} \right] \\ &= \frac{1}{2\pi} \frac{d\theta(t)}{dt} \end{aligned} \tag{2.26}$$

The phase modulated signal may be conceived as a rotating phasor of length A_c and angle $\theta(t)$. In the case of unmodulated carrier the angle $\theta(t)$ is

$$\theta(t) = 2\pi f_c t + \phi \tag{2.27}$$

the phasor rotates with a constant angular velocity equal to $2\pi f_c$.

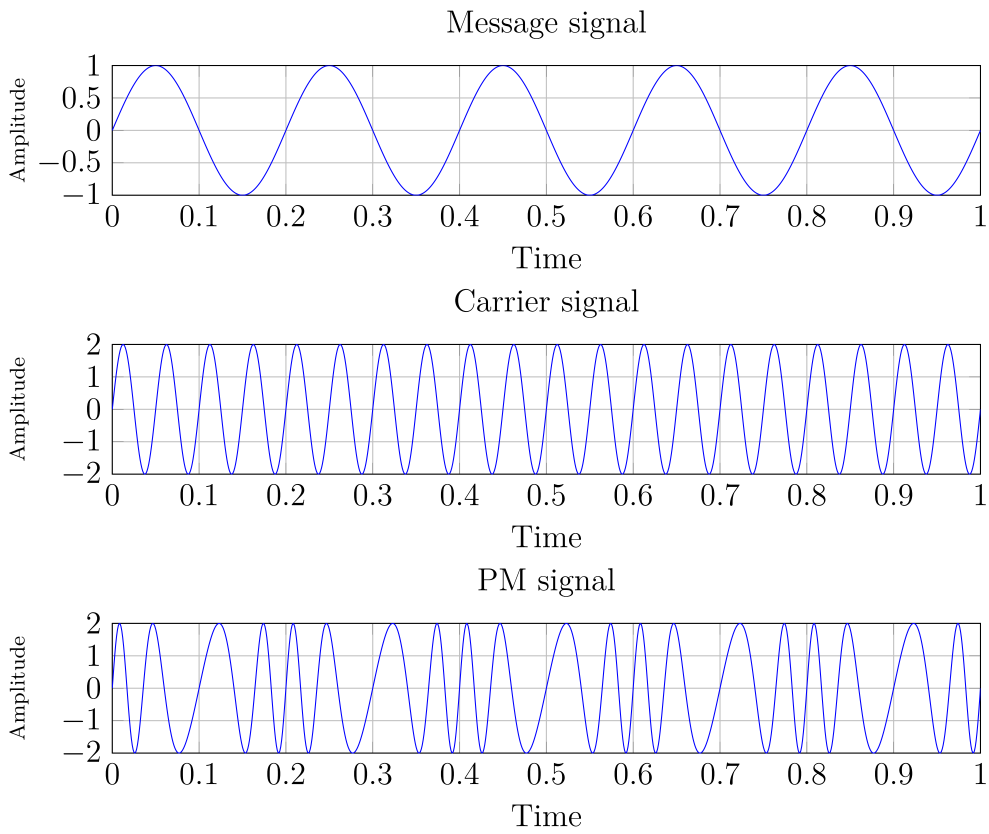


Figure 2.7: Illustration of phase modulated signal.

2.1.3 Frequency Modulation

Extra Reading

For further detail refer

Couch section 5.6

Haykin section 2.7.

The FM wave $s(t)$ is defined by

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(t)dt\right] \tag{2.28}$$

is a nonlinear function of the modulating signal $m(t)$. Hence it is a non-linear modulation scheme. Consequently there does not exist a simple relation between FM wave and its spectrum. For analysis it is convenient to assume that $m(t)$ is a simple sinusoidal signal

$$m(t) = A_m \cos(2\pi f_m t) \tag{2.29}$$

The instantaneous frequency of the resulting FM wave equals

$$\begin{aligned} f_i(t) &= f_c + k_f A_m \cos(2\pi f_m t) \\ &= f_c + \Delta f \cos(2\pi f_m t) \end{aligned} \tag{2.30}$$

where Δf is the frequency deviation, representing the maximal departure from the instantaneous frequency of the FM wave from the carrier frequency f_c. The deviation of the frequency is proportional to the amplitude of the modulating wave.

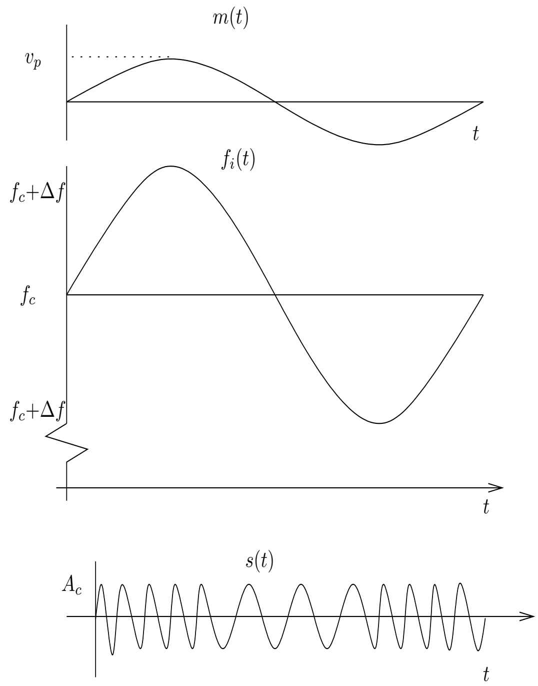


Figure 2.8: Illustration of frequency modulation of a sinusoidal signal.

Redefining the argument $\theta(t)$ of the FM wave i.e.

$$\begin{aligned} \theta(t) &= 2\pi \int_0^t f_i(t) dt \\ &= 2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t) \end{aligned} \tag{2.31}$$

The ratio of the frequency deviation Δf to the modulation frequency f_m is commonly known as the modulation index of the FM wave.

$$\beta = \frac{\Delta f}{f_m} \tag{2.32}$$

The spectrum representation of the FM modulated signal is presented in Fig. 2.9

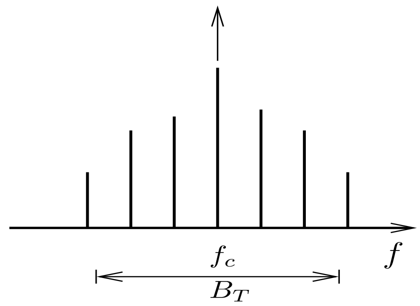


Figure 2.9: Spectrum of an FM modulated signal.

The spectral bandwidth of the modulated signal is directly proportional to the modulation index.

Salient Features

- The envelope of the FM wave is a constant, so the average power of such modulated signals is also constant.
- FM modulated signals are more immune to additive noise since the message is not modulated on the amplitude of the carrier.

Generation of FM signal

FM signal is generated through either direct or indirect modulation of the carrier signal. In direct modulation scheme the carrier frequency is controlled directly by the modulating signal where as in the indirect method of producing frequency modulation, the modulating waves are first used to produce a narrow band FM wave and frequency multiplication is next used to increase the frequency deviation to a desired level.

$$s_1(t) = A_1 \cos[2\pi f_1 t + \phi_1(t)] \tag{2.33}$$

where f_1 is the carrier frequency and A_1 is the carrier amplitude. The angular argument $\phi_1(t)$ is related to $m(t)$ as

$$\phi_1(t) = 2\pi k_1 \int_0^t m(t)dt \tag{2.34}$$

with k_1 being frequency sensitivity index, assume the frequency variations are small we have

$$\begin{aligned} \cos[\phi(t)] &= 1 \\ \sin[\phi(t)] &\approx \phi(t) \end{aligned} \tag{2.35}$$

using this approximation the $s_1(t)$ can be expressed

$$\begin{aligned} s_1(t) &\approx A_1\cos(2\pi f_1 t) - A_1\sin(2\pi f_1 t)\phi_1(t) \\ &= A_1\cos(2\pi f_1 t) - 2\pi k_1 A_1\sin(2\pi f_1 t) \int_0^t m(t)dt \end{aligned} \tag{2.36}$$

which defines a narrowband FM wave. The next step in generating in-direct FM signal generation is frequency multiplication. A frequency multiplier is a non-linear device (such as diode or transistor) followed by a bandpass filter. The non-linear device is assumed to be memoryless. The input-output relation of a memoryless nonlinear device is defined as

$$s_2(t) = a_1 s_1(t) + a_2 s_1^2(t) + \cdots + a_n s_1^n(t) \tag{2.37}$$

Performing the expansion in (2.37) we find that $s_2(t)$ has a dc component and n different frequency components $(f_1, 2f_1, \cdots, nf_1)$ and frequency deviations (Δf_1,$2\Delta f_1$, $\cdots$, $n\Delta f_1$). The values of Δf_1 is determined by frequency sensitivity k_1. The FM signal can be obtained by

1. Passing the FM wave centred at the carrier frequency nf_1 and with frequency deviation $n\Delta f_1$.

2. Suppressing all other FM spectra.

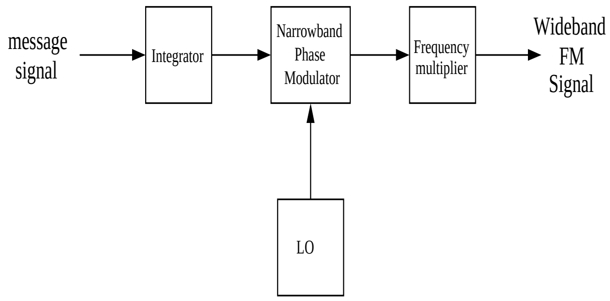


Figure 2.10: The process of indirect FM signal generation.

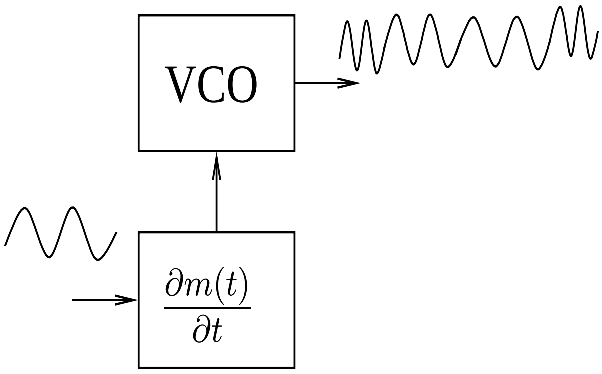


Figure 2.11: The simplified design of an FM modulator.

2.2 Phase Lock Loop

Extra Reading
For further detail refer
Couch section 4.4
Haykin section 2.14

The phase lock loop is an essential component of almost all modern communication systems. It is basically a negative feedback loop which consists of *multiplier*, *a low pass filter* and a *voltage controlled oscillator* (VCO). A block diagram structure of PLL is illustrated in the Fig. 2.12.

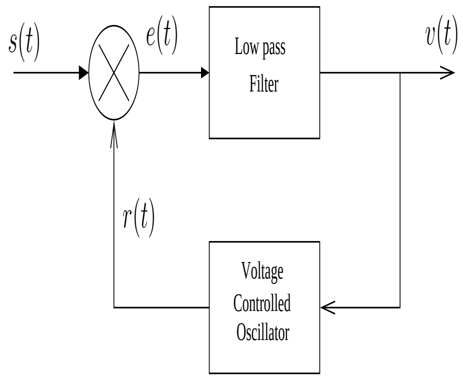


Figure 2.12: The block diagram of PLL.

Operation: Assuming the VCO is tuned such that when control voltage is zero, two conditions are satisfied (1) frequency of VCO is precisely set as unmodulated carrier f_c and (2) VCO output has 90° phase-shift with respect to the unmodulated carrier.

If FM modulated signal applied to PLL is defined as

$$s(t) = A_c \sin[2\pi f_c t + \phi_1(t)] \tag{2.38}$$

with modulating signal defined as

$$\phi_1(t) = 2\pi k_f \int_0^t m(t)dt \tag{2.39}$$

where k_f is the frequency sensitivity of the frequency modulator. Let the VCO output be defined as

$$r(t) = A_v \cos[2\pi f_c t + \phi_2(t)] \tag{2.40}$$

with a control voltage $v(t)$ applied to VCO we have

$$\phi_2(t) = 2\pi k_v \int_0^t v(t)dt \tag{2.41}$$

where k_v is the sensitivity of the VCO measure in *hertz per volt*. The incomming FM signal $s(t)$ and the VCO output $r(t)$ are applied to the multiplier producing the following

- A high frequency component represented by

$$k_m A_c A_v \sin[4\pi f_c t + \phi_1(t) + \phi_2(t)]$$

- a low frequency component

$$k_m A_c A_v \sin[\phi_1(t) - \phi_2(t)]$$

with k_m is the *multiplier gain* measure in volt^{-1}. The high frequency component can be eliminated by low pass filtering, thus the only component left is

$$e(t) = \underbrace{k_m k_v A_c A_v}_{k_0} \sin[\phi_e(t)] \tag{2.42}$$

where $\sin(\phi_e(t)) \approx \phi_e(t)$ which accurate to within 4% for $\phi_e(t)$ less than 0.5 rad, k_m are the multiplier gain, k_v is the frequency sensitivity of VCO and A_v and A_c, are amplitudes of VCO and carrier signal respectively, $\phi_e(t)$ is defined as

$$\begin{aligned} \phi_e(t) &= \phi_1(t) - \phi_2(t) \\ &= \phi_1(t) - 2\pi k_v \int_0^t v(t) dt \end{aligned} \tag{2.43}$$

with the output of the low pass filter $v(t)$ defined as

$$v(t) = \int_{-\infty}^{+\infty} \phi_e(\tau) h(t-\tau) d\tau \tag{2.44}$$

with $h(t)$ being the impulse response of the low pass filter. The phase error $\phi_e(t)$ is related to input phase $\phi_1(t)$ by integro-differential equation.

$$\begin{aligned} \frac{d\phi_e(t)}{dt} &= \frac{d\phi_1(t)}{dt} - 2\pi k_0 \int_{\tau=-\infty}^{+\infty} \sin[\phi_e(\tau)] h(t-\tau) d\tau \\ \frac{d\phi_1(t)}{dt} &= \frac{d\phi_e(t)}{dt} + 2\pi k_0 \int_{\tau=-\infty}^{+\infty} \sin[\phi_e(\tau)] h(t-\tau) d\tau \end{aligned} \tag{2.45}$$

Using Laplace transform properties for simple derivation we can write the transfer function of the system as

$$\begin{aligned} s\phi_1(s) &= s\phi_e(s) + 2\pi k_0 \phi_e(s) H(s) \\ \phi_e(s) &= \frac{\phi_1(s)}{1 + \underbrace{2\pi k_0 \frac{H(s)}{s}}_{L(s)}} \end{aligned} \tag{2.46}$$

The output of the control system $v(t)$ can be expressed in Laplace domain as

$$V(s) = \frac{2\pi k_0}{2\pi k_v} \phi_e(s) H(s). \tag{2.47}$$

By using the definition of $\phi_e(s)$ from (2.46) we can express output signal as

$$\begin{aligned} V(s) &= \frac{k_0}{k_v} \frac{\phi_1(s)}{1 + \underbrace{2\pi k_0 \frac{H(s)}{s}}_{L(s)}} H(s) \\ &= 2\pi k_0 s\phi_1(s) \frac{L(s)}{1 + L(s)} \end{aligned} \tag{2.48}$$

where $L(s)$ is the open loop gain. Now using the fact that the gain of LPF is large in the frequency band of our interest therefore the approximation

$$\frac{L(s)}{1 + L(s)} \approx 1 \tag{2.49}$$

is valid and the output signal can be finally expressed as

$$V(s) = \frac{1}{2\pi k_v} s\phi_1(s) \tag{2.50}$$

and finally corresponding time-domain relation is

$$v(t) \approx \frac{1}{2\pi k_v} \frac{d\phi_1(t)}{dt} \tag{2.51}$$

and $\phi_1(t)$ is related to the modulating wave $m(t)$

$$v(t) = \frac{k_f}{k_v} m(t) \tag{2.52}$$

That is the output $v(t)$ of the phase-locked loop is approximately the same except for the scale factor as original message signal $m(t)$ and the frequency demodulation is accomplished.

2.3 Applications of PLL

The following diagrams illustrate various applications of PLL in general communication systems

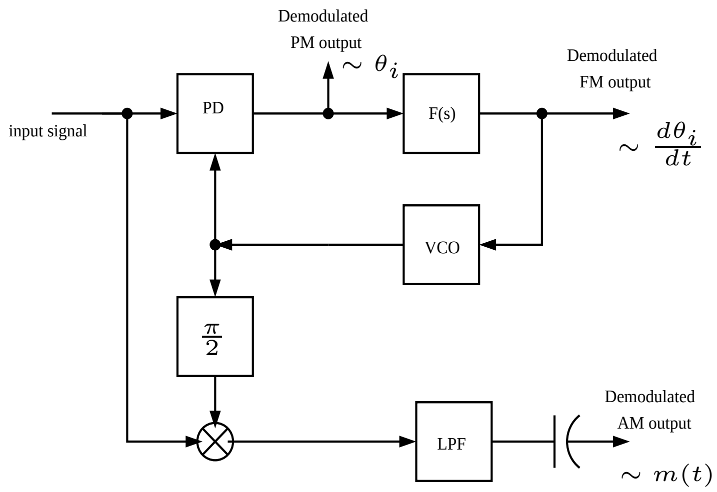


Figure 2.13: Application of PLL as FM,PM and AM demodulator.

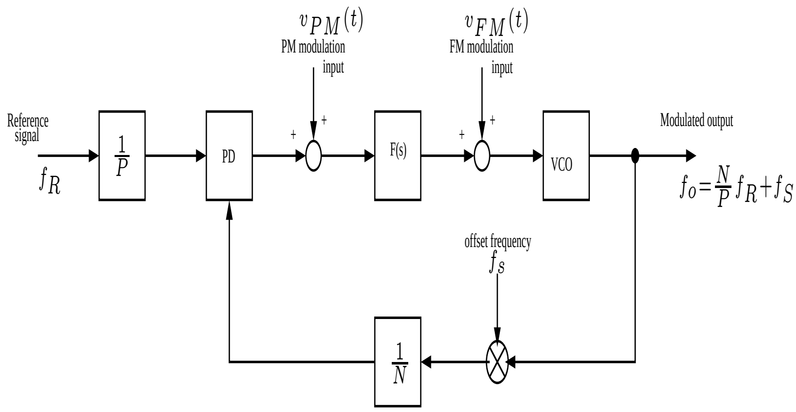


Figure 2.14: Fractional frequency synthesizer with FM and PM modulation.

Example

Consider that we have a base LO operating at 1 MHz. Design a frequency synthesizer to generate a reference clock of 133.5 MHz.

Costas Loop

Extra Reading

For further detail refer

Couch section 5.4

Proakis section 6.4.

It is well known technique for demodulating double sideband amplitude modulation with suppressed carrier. The schematic diagram is illustrated in fig. 2.15. Let the received RF signal be defined as

$$r(t) = A_c m(t)\cos(2\pi f_c t + \phi) + n(t) \tag{2.53}$$

is multiplied by $\cos(2\pi f_c t + \bar{\phi})$ and $\sin(2\pi f_c t + \bar{\phi})$ which is derived from the output of the VCO. The In-phase and Quadrature-phase products are

$$\begin{aligned} y_I(t) &= [A_c m(t)\cos(2\pi f_c t + \phi) + n_I(t)\cos(2\pi f_c t) - n_Q \sin(2\pi f_c t)] \cdot \cos(2\pi f_c t + \bar{\phi}) \\ &= \frac{A_c}{2} m(t)\cos\Delta\phi + \frac{1}{2}[n_I \cos\bar{\phi} + n_Q \sin\bar{\phi}] \\ &\qquad + \text{double frequency components} \end{aligned} \tag{2.54}$$

$$\begin{aligned} y_Q(t) &= [A_c m(t)\cos(2\pi f_c t + \phi) + n_I(t)\cos(2\pi f_c t) - n_Q \sin(2\pi f_c t)] \cdot \sin(2\pi f_c t + \bar{\phi}) \\ &= \frac{A_c}{2} m(t)\sin\Delta\phi + \frac{1}{2}[n_I \cos\bar{\phi} - n_Q \sin\bar{\phi}] \\ &\qquad + \text{double frequency components} \end{aligned} \tag{2.55}$$

Note:
Most common application Costas loop carrier recovery in DSB SC demodulation.

where $\Delta\phi = \phi - \bar{\phi}$. The double frequency terms are eliminated by the LPFs following the mixer.

An error signal is generated by multiplying the In-phase and Quadrature-phase outputs after LPFs.

$$\begin{aligned}
e(t) &= y_I(t) \times y_Q(t) \\
&= \frac{A_c^2}{4} m^2(t) \sin 2\Delta\phi \\
&\quad + \frac{A_c}{4}[n_I(t)\cos\bar{\phi} + n_Q(t)\sin\bar{\phi}]\sin\Delta\phi \\
&\quad + \frac{A_c}{4}[n_I(t)\sin\bar{\phi} - n_Q(t)\cos\bar{\phi}]\cos\Delta\phi \\
&\quad + \frac{1}{4}[n_I(t)\cos\bar{\phi} + n_Q(t)\sin\bar{\phi}][[n_I(t)\sin\bar{\phi} - n_Q(t)\cos\bar{\phi}]]
\end{aligned}$$

Note:
The factor $\sin(2\phi_e)$ implies more robust error tracking.

This error signal is filtered by the loop filter whose output is the control voltage driving the VCO.

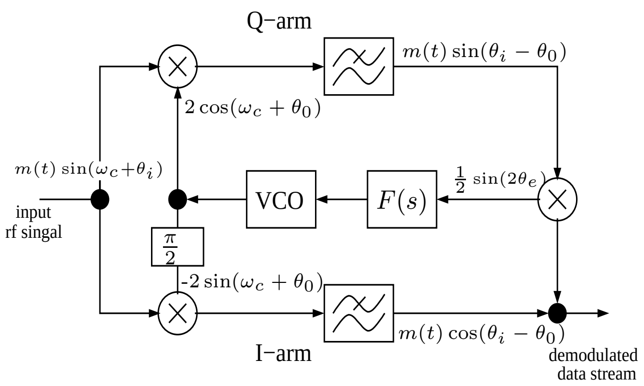


Figure 2.15: Block diagram of Costas Loop.

2.3.1 Mathematical modeling of Bandpass RF signals

Real world transmitted signals are all real valued, but in literature we often find the reference to Quadrature modulation techniques which exploits the orthogonality feature of Quadrature modulation.

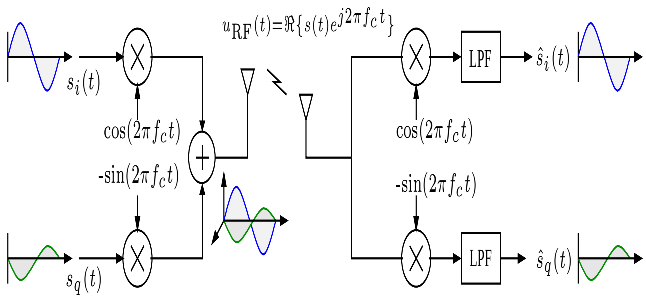


Figure 2.16: Block diagram Tx Rx RF chain for Quadrature transmission system.

By Quadrature modulation we mean that two independent signal are combined into one **complex** signal as *real* and *imaginary* parts. After transmission through RF link it is possible to separate the two signals easily.

Extra Reading

For further detail refer Haykin section 6.4.

The idea of quadrature modulation and demodulation is illustrated in the fig.2.16. The complex continuous signal $s(t)$ which consists of two real and imaginary components also commonly known as *inphase* and *quadrature* components respectively. The real component of the signal is multiplied with $\cos \omega_c t$ while the imaginary component is multiplied with a $-\sin \omega_c t$. The carrier modulated output is defined as

$$u_{\mathrm{RF}}(t) = s_I(t)\cos\omega_c t - s_Q(t)\sin\omega_c t \Rightarrow \Re\{s(t)e^{j\omega_c t}\} \quad \text{(please verify)} \tag{2.56}$$

The structure of the quadrature receiver is also illustrated in the fig. 2.16. In short the received baseband signal can be defined as

$$r(t) = \mathrm{LPF}\{u_{\mathrm{RF}(t)}e^{-j\omega_c t}\} \tag{2.57}$$

Simple derivation is provided here

$$\begin{aligned} r(t) &= (s_I(t)\cos\omega_c t - s_Q(t)\sin\omega_c t) \times (\cos\omega_c t - j\sin\omega_c t) \\ &= S_I(t)\cos\omega_c t\cos\omega_c t - js_I(t)\cos\omega_c t\sin\omega_c t \\ &\quad - s_Q(t)\sin\omega_c t\cos\omega_c t + js_Q(t)\sin\omega_c t\sin\omega_c t \end{aligned} \tag{2.58}$$

Now using the following trigonometric identities

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha+\beta)+\cos(\alpha-\beta)] \tag{2.59a}$$
$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)] \tag{2.59b}$$
$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)] \tag{2.59c}$$
$$\cos\alpha\sin\beta = \frac{1}{2}[\sin(\alpha+\beta)-\sin(\alpha-\beta)] \tag{2.59d}$$

using (2.59) in (2.58) we obtain

$$\begin{aligned} r(t) = s_I(t)(\cos 0\omega_c t + \cos 2\omega_c t) - js_I(t)(\sin 0\omega_c t - \sin 0\omega_c t) \\ - s_Q(t)(\sin 2\omega_c t + \sin 0\omega_c t) + js_Q(t)(\cos 0\omega_c t + \cos 2\omega_c t) \end{aligned} \tag{2.60}$$

After LPF filter residue left at the output of receiver is

$$r(t) = s_I(t) + js_Q(t) \tag{2.61}$$

This chapter is limited to a very superficial review of the analog communication systems. You are referred to

1. An introduction to Analog and Digital Communication by Simon Haykin
2. Digital and Analog Communication Systems by Leon W. Couch

for further detail.

An extensive detail of Analog schemes is beyond the scope of this course.

Chapter 3

Fundamentals of Digital Communication

Outline

The objective of this chapter is to study a fundamental communication system. Digital communication system is a broad topic which consists of several fundamental concepts and due to time constraints we aim to study only a selective set of components in this chapter. The fundamental concepts covered in this chapter are

Ⓐ Data formatting	Ⓔ Matched filter
Ⓑ Information & Entropy	Ⓕ Channel equalization
Ⓒ Pulse shaping	Ⓖ Channel Coding
Ⓓ Channel modeling	Ⓗ Bit error analysis

Now having explained the baseband transmission model we can discuss the effects of a communication system namely ISI, and relevant criteria and then comes the question what is the optimal receiver and (sub)optimal equalizer device.
The Fig. 3.1 illustrates the simplistic block diagram of a general communication system. The block illustrates several essential and non-essential components of a typical transmission link. In this chapter our objective is to selectively study some

of the fundamental concepts of digital communication system and lay a strong mathematical foundation for advanced studies. Before we even begin to explore the basic elements of communication system we would like to briefly discuss the information theoretic aspects of data source.

Extra Reading

For Further detail refer

Haykin section 9.2

Proakis section 12.2

3.1 Information and Entropy

The total information content for a symbol m is inversely proportional to its probability $p(m)$.

$$I_m = \log \frac{1}{p(m)} = -\log p(m) \tag{3.1}$$

Please note that total information contained by two messages is

$$I_1 + I_2 = -\log_2 p(m_1) - \log_2 p(m_2) = -\log_2 [p(m_1)p(m_2)] \qquad \text{(bits)} \tag{3.2}$$

The average amount of uncertainty contained by an event/a message is defined as its **entropy**.

$$H = \sum_{m=1}^{M} p(m) \log_2 \frac{1}{p(m)} \tag{3.3}$$

Entropy of Binary Source:

For the case when $M = 2$ (a system with 2 symbols i.e. (0,1)) if $p(1) = \rho$ and $p(0) = 1 - \rho$ then

$$H = \rho \log_2 \frac{1}{\rho} + (1-\rho) \log_2 \frac{1}{1-\rho} \tag{3.4}$$

For ASCII alphabet, there exist 128 equally likely symbols therefore entropy of a symbol is $H = -\log_2(1/128) = 7$ symbols. However in practice all of the symbols are

Note:

Typical units of entropy 'bit'

entropy is always positive!

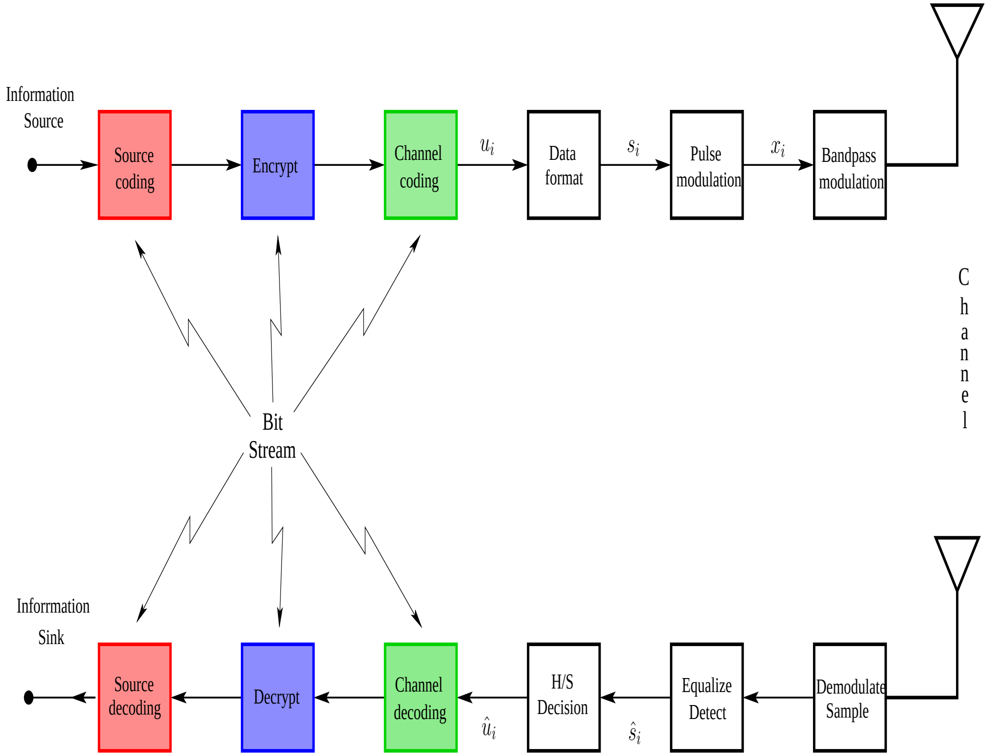


Figure 3.1: Simplified structure of a general transmission systems.

equally likely and statistically independent. Therefore

$$H = \sum_{m=1}^{128} p(m) \log_2 \frac{1}{p(m)} < 7 \text{ bits/symbol} \tag{3.5}$$

Entropy thus represents the minimum number of binary digits required per symbol (averaged over a long sequence of symbols). The maximum possible entropy of any source is defined as

$$H_{\max} = \log_2 M (\text{bit/symbol}) \tag{3.6}$$

The results of these examples can be generalized as

Vocab. Size	no. binary digits	Symbol probability
2	1	1/2
4	2	1/4
8	3	1/8
16	4	1/16
32	5	1/32
64	6	1/64
128	7	1/128

3.1.1 Shannon-Hartley Capacity Theorem

Extra Reading
For further detail refer Sklar section 9.4

Shannon in 1949 showed that the system capacity C of a channel perturbed by additive white Gaussian noise (AWGN) is a function of average received signal power S, the average noise power N, and bandwidth W. The capacity relationship (Shannon-Hartley Theorem) can be stated as

$$C = W \log_2 \left(1 + \frac{S}{N}\right) \tag{3.7}$$

Note:
Spectral efficiency is defined as $\frac{R}{W}$(bits/s/Hz) for arbitrarily small BER efficiency $\rightarrow \frac{C}{W}$.

where W is in hertz and logarithm is taken to the base 2, and channel capacity is measured in bits/s. It is theoretically possible to transmit information over such a channel at any rate R, where $R \leq C$, with an arbitrarily small error probability using a sufficiently complicated coding scheme. For an information rate $R > C$, it is not possible to find a code that can achieve an arbitrarily small error probability.

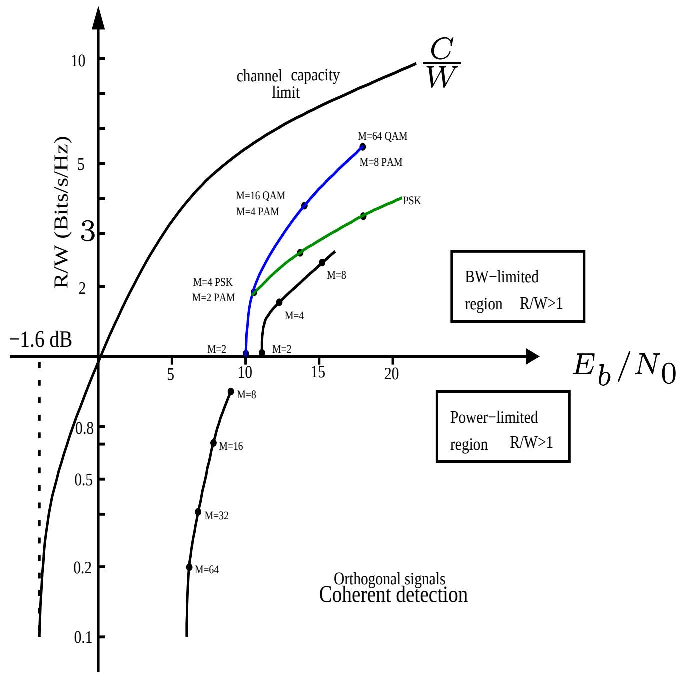


Figure 3.2: Graphical illustration of equation (3.7).

Shannon's work showed that the values of S, N and W set a limit on transmission rate not on error probability. Shannon used (3.7) to graphically illustrate a bound of achievable performance of practical systems. This plot gives normalized channel capacity C/W in bits/s/Hz as a function of the channel SNR. Noise power is proportional to the bandwidth

$$N = N_0 W \tag{3.8}$$

A related Fig. 3.3 indicates the normalized channel bandwidth W/C in Hz/bits/s as a function of SNR in the channel. Fig.3.2 is sometimes used to illustrate the power-bandwidth trade-off inherent in the ideal channel.

$$\frac{C}{W} = \log_2\left(1 + \frac{S}{N_0 W}\right) \tag{3.9}$$

For the case where transmission bit rate is equal to channel capacity R=C, using the following identity

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{W}{R}\right) \tag{3.10}$$

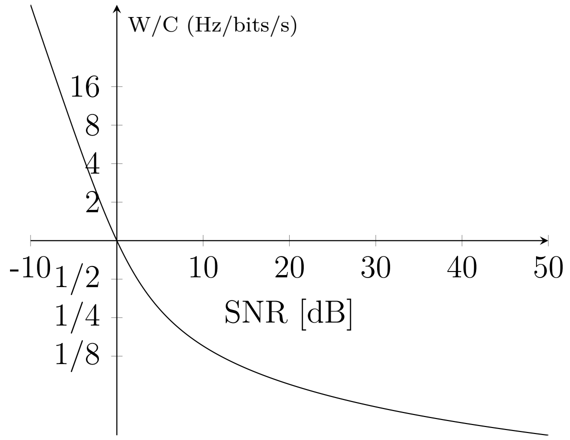


Figure 3.3: Graphical illustration of equation (3.9).

we can rewrite the expression (3.7)

$$\begin{aligned} \frac{C}{W} &= \log_2\left[1+\left(\frac{S}{N_0W}\right)\right] \\ 2^{\mathrm{C/W}} &= 1+\frac{E_b}{N_0}\left(\frac{C}{W}\right) \\ \frac{E_b}{N_0} &= \frac{W}{C}(2^{\mathrm{C/W}}-1) \end{aligned} \tag{3.11}$$

Fig. 3.3 is a plot W/C versus E_b/N_0 in accordance with (3.11). The asymptotic behaviour of this curve as $C/W \to 0$ (or $W/C \to \infty$). There exists a limiting value of E_b/N_0 below which there can be no error-free communication at any information rate. Using the identity

$$\lim_{x\to 0}(1+x)^{1/x} = e \tag{3.12}$$

we can calculate the limiting value of E_b/N_0 as follows:Let

$$x = \frac{E_b}{N_0}\left(\frac{C}{W}\right) \tag{3.13}$$

then from above

$$\frac{C}{W} = x\log_2(1+x)^{1/x} \tag{3.14}$$

In the limit, as $C/W \to 0$, we get

$$\frac{E_b}{N_0} = \frac{1}{\log_2 e} = 0.693 \tag{3.15}$$

or in decibels

$$\frac{E_b}{N_0} = -1.6dB \tag{3.16}$$

Important Points

- For infinite bandwidth i.e $W \to \infty$ the ratio E_b/N_0 approaches $\log 2 = 0.693$, this value is called Shannon's limit for an AWGN channel assuming code rate of zero. This corresponds to the limiting value of channel capacity (i.e. at $W \to \infty$).
- The capacity boundary defined by the curve for critical bit rate R=C, separates the combination of the system parameters that have the potential for supporting error free transmission ($R < C$), for those which error free transmission is not possible ($R > C$).
- The diagram highlights potential trade-off among E_b/N_0, R/W and the probability of symbol error P_e. In particular the movement of the operating point along a horizontal line as trading P_e versus E_b/N_0 for a fixed R/W or the movement of the operating point along vertical line as trading P_e versus R/W for a fixed E_b/N_0.

In the subsequent sections certain blocks of the transmission system are discussed in detail.

3.2 Data Format

Data formatting is an important facet of digital communication. Several binary bits can be mapped into one symbol. There exist several mapping schemes. In general there exist only $M = 2^k$ possible symbols, each symbol represents k info bits. There exist several data mapping possibilities. In the following only a few well known schemes are discussed.

Pulse Amplitude Modulation

The waveform of pulse amplitude modulation (PAM) may be represented as

$$s_m(t) = A_m g(t) \qquad 1 \leq m \leq M \tag{3.17}$$

where $g(t)$ is a pulse of duration T, and A_m $1 \leq m \leq M$ denotes set of M possible amplitudes corresponding to $M = 2^k$ (each symbol represents k info. bits).
The average energy of PAM transmitted symbols

$$\begin{aligned}\mathcal{E}_{\text{avg}} &= \frac{\mathcal{E}_g}{M}\sum_{m=1}^{M} A_m^2 \\ &= \frac{2\mathcal{E}_g}{M}(1^2 + 3^2 + \cdots + (M-1)^2)\end{aligned}$$

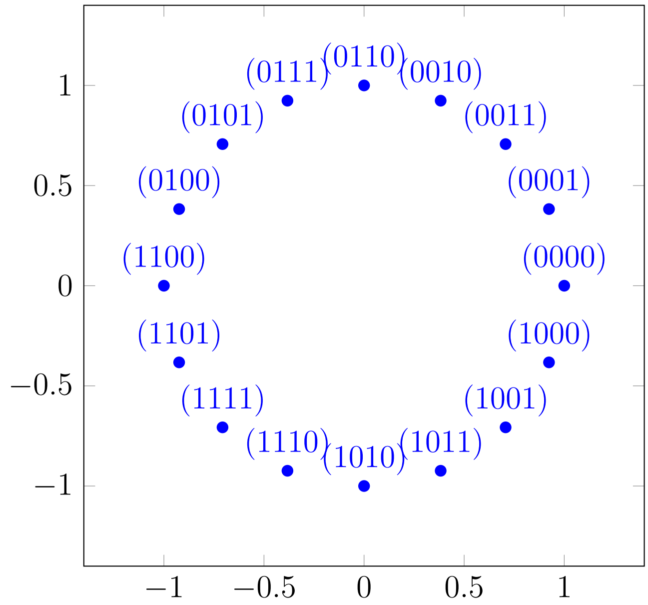
which simplifies into

$$\begin{aligned}&= \frac{2\mathcal{E}_g}{M} \times \frac{M(M^2-1)}{6} \\ &= \frac{(M^2-1)}{3}\mathcal{E}_g. \end{aligned}\tag{3.18}$$

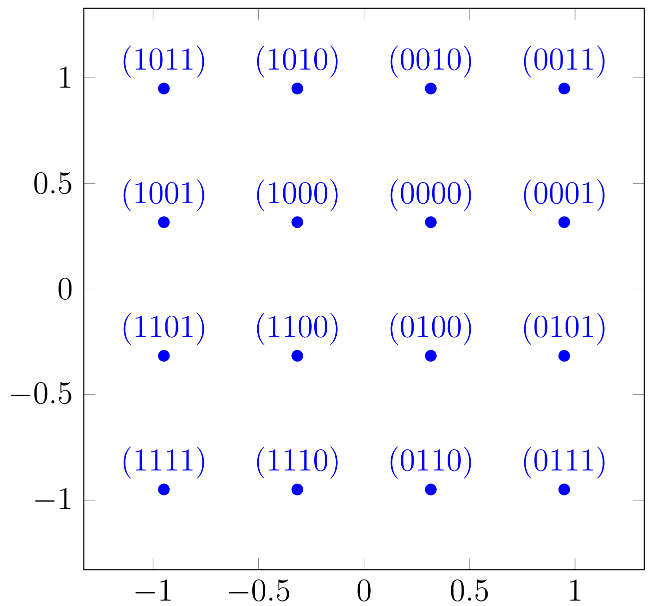
Phase Modulation

In phase modulation, the M-signal waveform is represented as

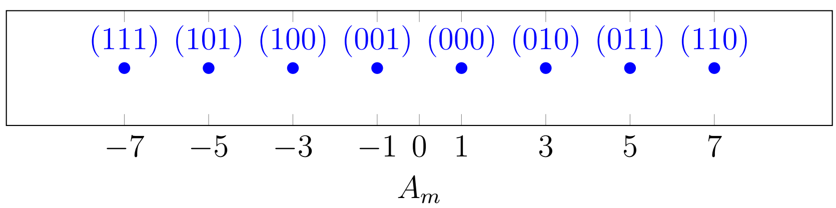
$$\begin{aligned}s_m(t) &= \Re\left\{g(t)e^{j\frac{2\pi(m-1)}{M}}e^{j2\pi f_c t}\right\}, \quad m = 1, \ldots, M \\ &= g(t)\cos\left[2\pi f_c t + \frac{2\pi}{M}(m-1)\right] \\ &= g(t)\cos\left(\frac{2\pi}{M}(m-1)\right)\cos(2\pi f_c t) - g(t)\sin\left(\frac{2\pi}{M}(m-1)\right)\sin(2\pi f_c t)\end{aligned}$$



(a) 16 PSK Constellation



(b) 16 QAM Constellation



(c) 8 PAM Constellation

Figure 3.4: The constellation diagrams of PAM, QAM and PSK.

where $g(t)$ is the pulse shape and $\theta_m=2\pi(m-1)/M$ with $m=1,\ldots,M$ being possible values of the carrier to convey the transmitted information. Digital phase modulation is called phase shift keying. The average energy of the transmitted signal can be defined as

$$\mathcal{E}_{\text{avg}} = \frac{1}{2}\mathcal{E}_g. \tag{3.19}$$

The euclidean distance between possible signal points is defined as

$$\begin{aligned} d_{mn} &= \sqrt{\|\mathbf{s}_m - \mathbf{s}_n\|^2} \\ &= \sqrt{\mathcal{E}_g\left[1 - \cos\left(\frac{2\pi}{M}(m-n)\right)\right]} \end{aligned} \tag{3.20}$$

and the minimum distance corresponding to $|m-n| = 1$ is

$$d_{min} = \sqrt{\mathcal{E}_g\Big(1 - \cos\frac{2\pi}{M}\Big)} = \sqrt{2\mathcal{E}_g \sin^2\frac{\pi}{M}} \tag{3.21}$$

Quadrature Amplitude Modulation

This scheme maps information bits on $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$. The resulting mapping scheme is called quadrature amplitude modulation (QAM). The corresponding signal may be expressed as

$$\begin{aligned} s_m(t) &= \Re\Big\{(A_{mi} + jA_{mq})g(t)e^{j2\pi f t}\Big\} \\ &= A_{mi}g(t)\cos(2\pi f t) - A_{mq}g(t)\sin(2\pi f t) \qquad m = 1, 2, \cdots, M \end{aligned} \tag{3.22}$$

Euclidean distance between any pair of symbols is defined

$$\begin{aligned} d_{mn} &= \sqrt{\|\mathbf{s}_m - \mathbf{s}_n\|^2} \\ &= \sqrt{\frac{\mathcal{E}_g}{2}[(A_{mi} - A_{ni})^2 + (A_{mq} - A_{nq})^2]} \end{aligned} \tag{3.23}$$

thus the minimal distance between to adjacent symbols can be defined as

$$d_{\min} = \sqrt{2\mathcal{E}_g} \tag{3.24}$$

The average energy of the symbols (for equally likely symbols) may be defined as

$$\begin{aligned}
\mathcal{E}_{\text{avg}} &= \frac{1}{M}\frac{\mathcal{E}_g}{2}\sum_{m=1}^{\sqrt{M}}\sum_{n=1}^{\sqrt{M}}(A_i^2 - A_q^2) \\
&= \frac{\mathcal{E}_g}{2M} \times \frac{2M(M-1)}{3} \\
&= \frac{M-1}{3}\mathcal{E}_g
\end{aligned} \tag{3.25}$$

Gray coded data

By carefully choosing the assignment of data bits to symbols, the bit error can be reduced significantly. The example of gray codes is illustrated in Fig. 3.4.

Important definitions

- Symbol: A member of source alphabets- may or may not be binary.
- Baud: Rate of symbol transmission. i.e. 100 bauds = 100 symbols/sec
- Bit: Quantity of information carried by a symbol with selection probability $p = 0.5$.
- Bit rate: Rate of information transmission (bits/sec)
- SNR: The ratio of total signal power to the total noise power.
- E_b/N_0: The ratio of energy available per bit to the noise power.

3.3 Line coding formats

Extra Reading
For further details refer Sklar section 2.8

In the following section only a few simple line coding techniques are presented. The line codes determine the spectral and electrical characteristics of any given transmission system. Careful choice of line codes can reduce the power, bandwidth requirement and lower the desipitation of power in the communication link. Var-

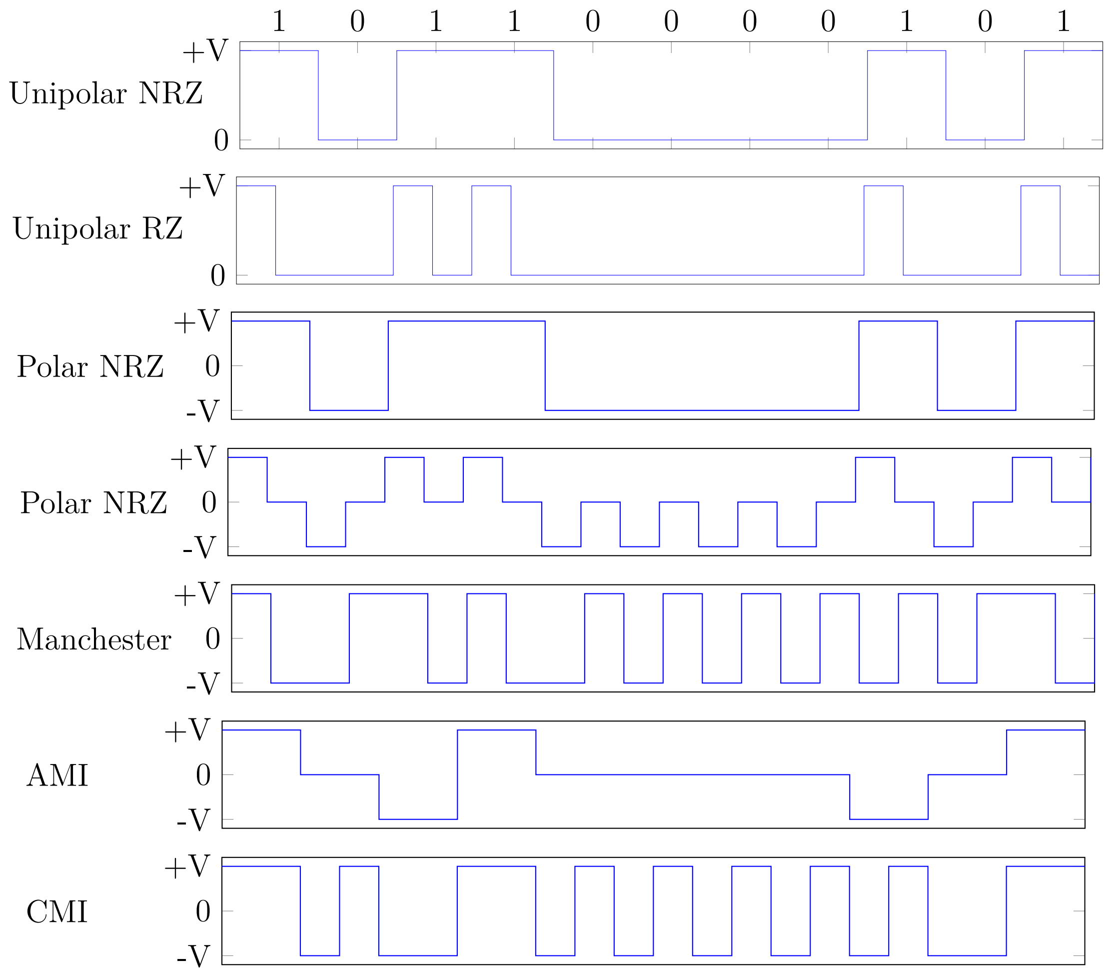


Figure 3.5: Time domain waveform of some common line codes found in literature.

ious characteristics of the popular pulse formats is illustrated in the table 3.1. The spectral characteristics of pulse shapes have very important implications for example possibility of carrier recovery from transmitted signal, required transmit bandwidth, distribution of power in spectrum and DC power desipitation etc. The supplementary slide sheds light on these features.

	Timing Extraction	Error correction	First null bandwidth	AC coupled	Transparent
Unipolar (NRZ)	Difficult	No	f_0	No	No
Unipolar (RZ)	Simple	No	$2f_0$	No	No
Polar (NRZ)	Difficult	No	f_0	No	No
Polar (RZ)	Difficult	No	$2f_0$	No	No
Manchester	Difficult	No	$2f_0$	Yes	No
AMI	Rectify	Yes	f_0	Yes	No
CMI	Simple	Yes	$2f_0$	Yes	Yes

Table 3.1: Comparison of common line codes available in the literature.

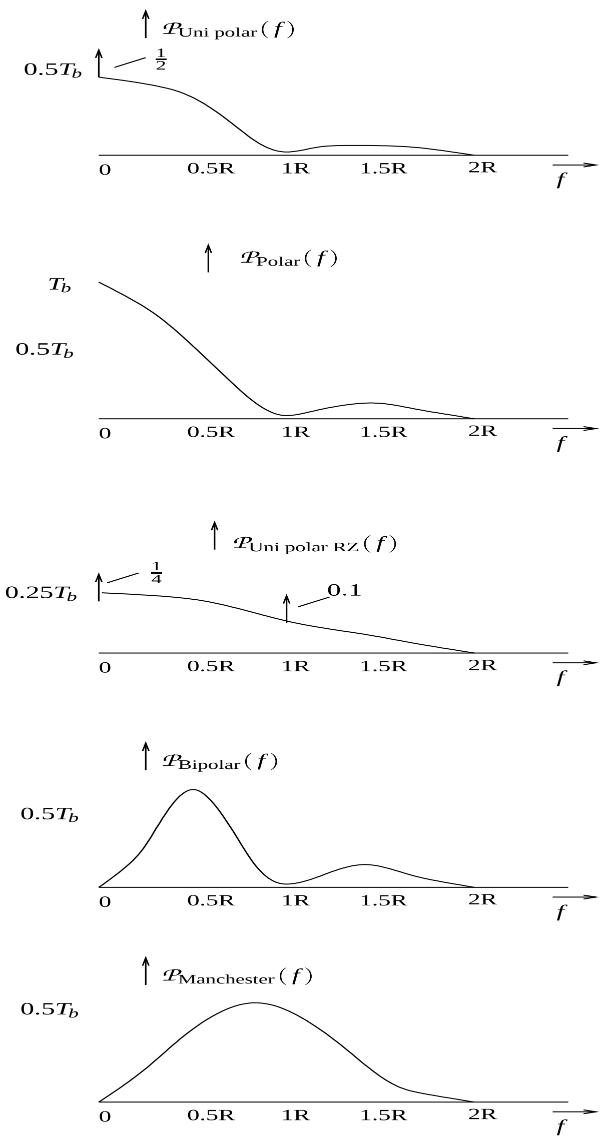


Figure 3.6: Spectral characteristics of common line coding techniques.

The generic PSD of a conventional line code can be evaluated through the following approach. The transmitted signal $s(t)$ can be defined as

$$s(t) = \sum_{n=-\infty}^{+\infty} a_n g(t - nT_s) \tag{3.26}$$

From the defination of power spectral density we know that

$$\mathcal{P}_s(f) = \lim_{T_s \to \infty} \frac{|S_{T_s}(f)|^2}{T_s} \tag{3.27}$$

The auto-correlation function

$$R_s(t+lT_s) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} E\{a_n a_m^*\} g(t + lT_s - nT_s) g^*(t - mT_s) \tag{3.28}$$

Extra Reading
For further detail refer Haykin section 3.8.

Here $E\{\cdot\}$ represents a statistical expectation operator, this expectation operator comes to play because a_n is not a deterministic signal (this fact is not discussed here further, but it means that its sufficient to take a look at certain combination of $a_n a_m^*$ to draw conclusion about the overall autocorrelation function). The Fourier integral of this correlation function can found using the definition (3.27)

$$\begin{aligned}
S(f) \cdot S^*(f) &= \left(\int_{t=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a_n g(t - nT_s) e^{-j2\pi f t} dt \right) \\
&\quad \cdot \left(\int_{t=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} a_{m+l} g(t - lT_s - mT_s) e^{-j2\pi f t} dt \right)^* \\
&= |G(f)|^2 \left(\sum_{n=-\infty}^{+\infty} a_n e^{-j2\pi f n T_s} \right) \left(\sum_{m=-\infty}^{+\infty} a_{m+l} e^{-j2\pi f (m+l) Ts} \right)^* \\
&= |G(f)|^2 \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} E\{a_{n+l} a_m^*\} e^{-j2\pi f (n-m-l) T_s}
\end{aligned} \tag{3.29}$$

The double summation term in (3.29) can be simplified into a single summation since indices n and m in $\sum\sum e^{j2\pi f(n-l-m)}$ just create offset which ranges from $-\infty$ to $+\infty$. Therefore we can rewrite the summation index and

$$\mathcal{F}\{R_s(t - lT_s)\} = |G(f)|^2 \sum_{n=-\infty}^{+\infty} E\{a_n a_{n+l}^*\} e^{-j2\pi f(n)} \tag{3.30}$$

Using the expression for Fourier transform of auto-correlation in (3.27) we can write the generic expression for PSD as follows

$$\mathcal{P}_s(f) = \mathcal{F}\{R_s(t - lT_s)\} = \frac{|G(f)|^2}{T_s} \sum_{n=-\infty}^{+\infty} E\{a_n a_{n+l}^*\} e^{-j2\pi f(n)} \tag{3.31}$$

Now lets apply this definition to calculate the PSD of simple NRZ unipolar transmission using simple rectangle pulses. The signal levels can be defined as

$$a_n = \begin{cases} A & \text{when} \quad u_n = 1 \\ 0 & \text{when} \quad u_n = 0 \end{cases} \tag{3.32}$$

For now we assume the common rectangular pulse as the pulse shape which is defined as

$$g(t) = \Pi\left(\frac{t}{T_b}\right) \tag{3.33}$$

Firstly the auto-correlation function for transmitted data symbols is defined as

$$R_s(l) = E\{a_n a_{n+l}\} = \sum_{i=1}^{4} (a_n a_{n+l}) p_i \tag{3.34}$$

This expression can be written in two terms as

$$\begin{aligned} R_s(0T_s) &= \frac{1}{2}(0)^2 + \frac{1}{2}(A)^2 \\ R_s(lT_s) &= \frac{1}{4}A^2 + \frac{1}{4}A^2 - \frac{1}{2}A^2 - \frac{1}{4}(A)^2 \text{with } l \neq 0 \end{aligned} \tag{3.35}$$

A plot of the above auto-correlation function would be.

Using the linearity properties of the Fourier transform the spectrum of this expression can be easily obtain. Hence, this all can be summed up into

$$\mathcal{P}_s(f) = \frac{A^2}{4}\frac{|G(f)|^2}{T_s}\sum_{k=-\infty}^{\infty}\left(1+\delta(f-\frac{k}{T_s})\right) \tag{3.36}$$

using the standard Fourier transform of $\Pi(\frac{t}{T})$ defined in (1.4) the final expression for *Unipolar NRZ* pulse shaping can be expressed as

$$\mathcal{P}_{\text{unipolar NRZ}}(f) = T_s\frac{A^2}{4}\left(\frac{\sin(\pi f T_b)}{\pi f T_b}\right)^2\sum_{k=-\infty}^{\infty}\left(1+\delta(f-\frac{k}{T_s})\right) \tag{3.37}$$

Extra Reading

For further detail refer Couch section 3.5

Please try to obtain the expression of PSD for other common line codes and critically analyze the power distribution and bandwidth utilization of these codes.

3.4 Pulse Shapes

In the last section we have consider rectangular pulses; From practical point of view rectangular pulse shape is undesirable because it requires infinite bandwidth which is not available. The design criteria for pulse shaping was formulated by *Nyquist*.

$$P(f) = \begin{cases} \frac{1}{2W}, & 0 \le |f| \le f_1 \\ 0, & |f| \ge W \end{cases}$$

which is

$$P(f) = \frac{1}{2W}\text{rect}\left(\frac{f}{2W}\right) \tag{3.38}$$

The most important characteristic of $P(f)$ is that its inverse Fourier transform $p(t)$ satisfies Nyquist condition that

$$p(t) = \begin{cases} 1, & t = 0T \\ 0, & t = \pm nT \quad n = 1, 2, \ldots \end{cases}$$

The pulse may have non-zero value between multiples of T but they do not effect the sampled signal. This pulse is physically unrealizable because sharp transition in $P(f)$ means that the impulse response decays at a very slow rate and this impulse response need to be very long.
This limitation can be encountered by extending the bandwidth of the $P(f)$ from W to $2W$.

3.4.1 Raised Cosine Pulse

Let $P(f)$ denote the overall frequency response composed of three components

$$P(f) = \begin{cases} \frac{1}{2W}, & 0 \leq |f| \leq f_1 \\ \frac{1}{4W}\left[1 + \cos\left(\frac{\pi}{2W\rho}(|f| - W(1-\rho))\right)\right], & f_1 \leq |f| \leq 2W - f_1 \\ 0 & |f| \geq 2W - f_1 \end{cases}$$

with $\rho = 1 - f_1/W$ is the roll-off factor represents excess bandwidth and $\rho \in (0, 1)$. The frequency response of the *raised cosine* pulse is illustrated in the fig. 3.8. The time response $p(t)$ is the inverse Fourier transform of the frequency response $P(f)$ i.e.

$$p(t) = \mathrm{sinc}\Big(2Wt\Big)\left(\frac{\cos 2\pi\rho W t}{1 - 16\rho^2 W^2 t^2}\right) \tag{3.39}$$

In the impulse response for different values of ρ is plotted in fig.3.8.

3.4.2 Root Raised-Cosine Pulse Shaping

A more sophisticated form of pulse shaping uses the *root raised cosine* instead of raised cosine spectrum. Thus using the trigonometric identities we write

$$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta) \tag{3.40}$$

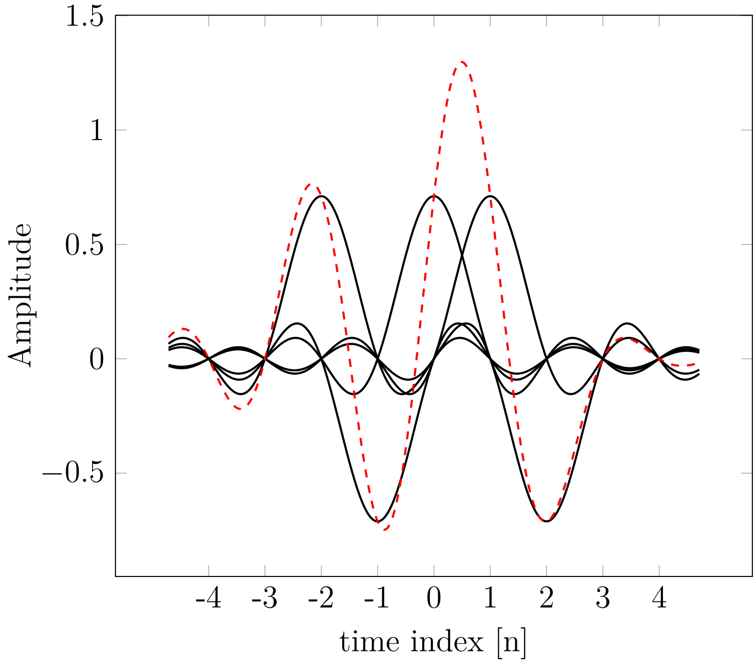


Figure 3.7: ISI free transmission of data with sinc pulse.

In our case $\theta = \pi/2W\rho(|f| - W(1-\rho))$, the spectrum of the pulse shape can be defined as

$$P(f) = \begin{cases} \frac{1}{\sqrt{2W}}, & 0 \leq |f| \leq f_1 \\ \frac{1}{\sqrt{2W}} \cos\left(\frac{\pi}{4W\rho}(|f| - W(1-\rho))\right), & f_1 \leq |f| < 2W - f_1 \\ 0, & |f| \geq 2W - f_1 \end{cases}$$

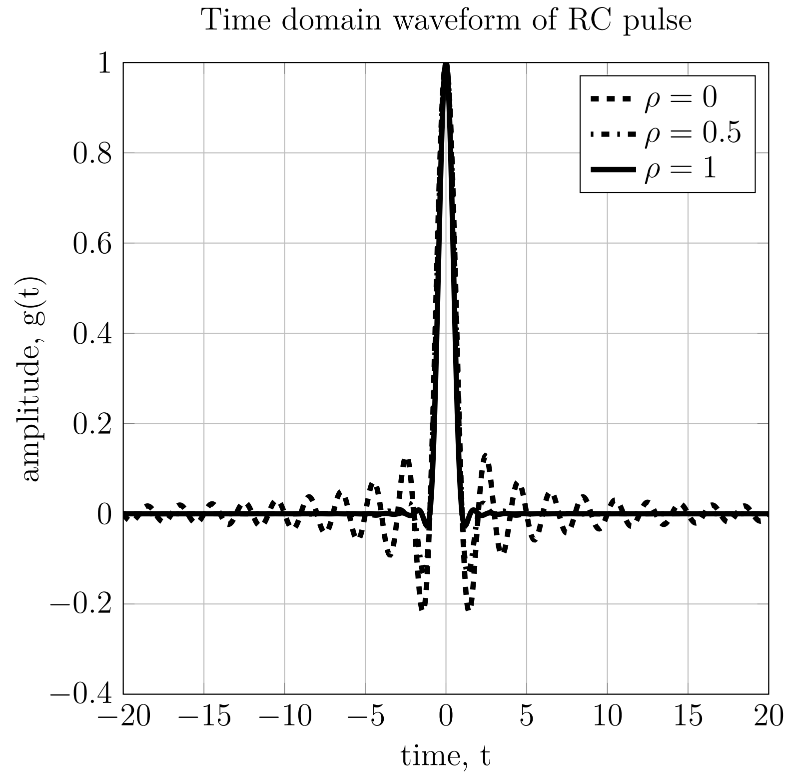
The corresponding time domain function of the impulse is defined as

$$p(t) = \frac{\sqrt{2W}}{(1 - 8\rho W t)^2} \left(\frac{\sin(2\pi W(1-\rho)t)}{2\pi W t} + \frac{4\rho}{\pi} \cos(2\pi W(1+\rho)t) \right) \tag{3.41}$$

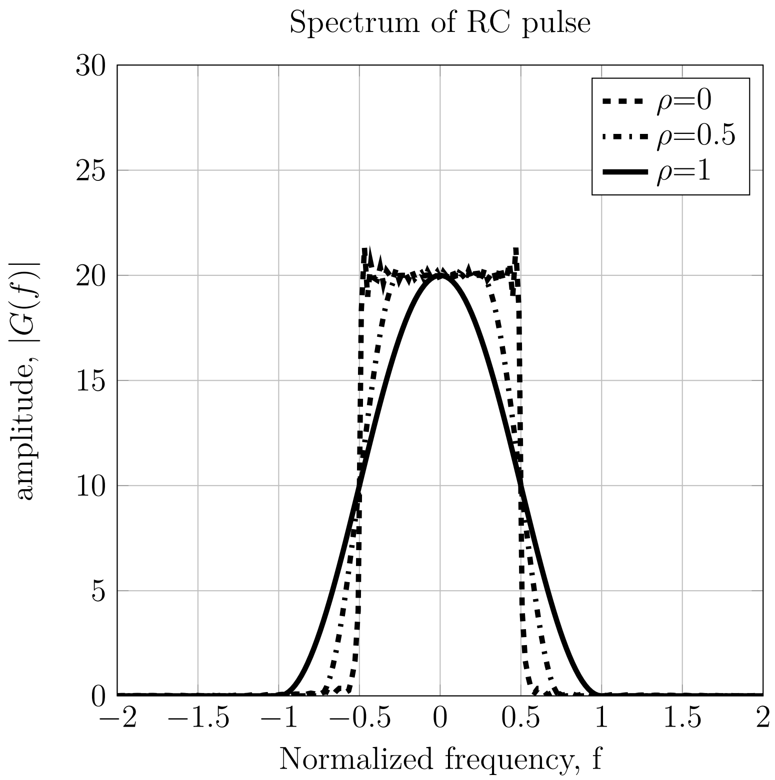
This pulse satisfies the orthogonality condition which is essential for ISI free transmission

$$\int_{-\infty}^{+\infty} p(t)p(t - nT)dt = 0 \quad \text{for } n = \pm 1, \pm 2, \ldots \tag{3.42}$$

Its is worth mentioning that despite of being orthogonal, this pulse does not satisfy (3.39).



(a) Impulse response of raised cosine pulse shape



(b) Frequency response of raised cosine pulse shape

Figure 3.8: Time and frequency response of the raised cosine pulse shape

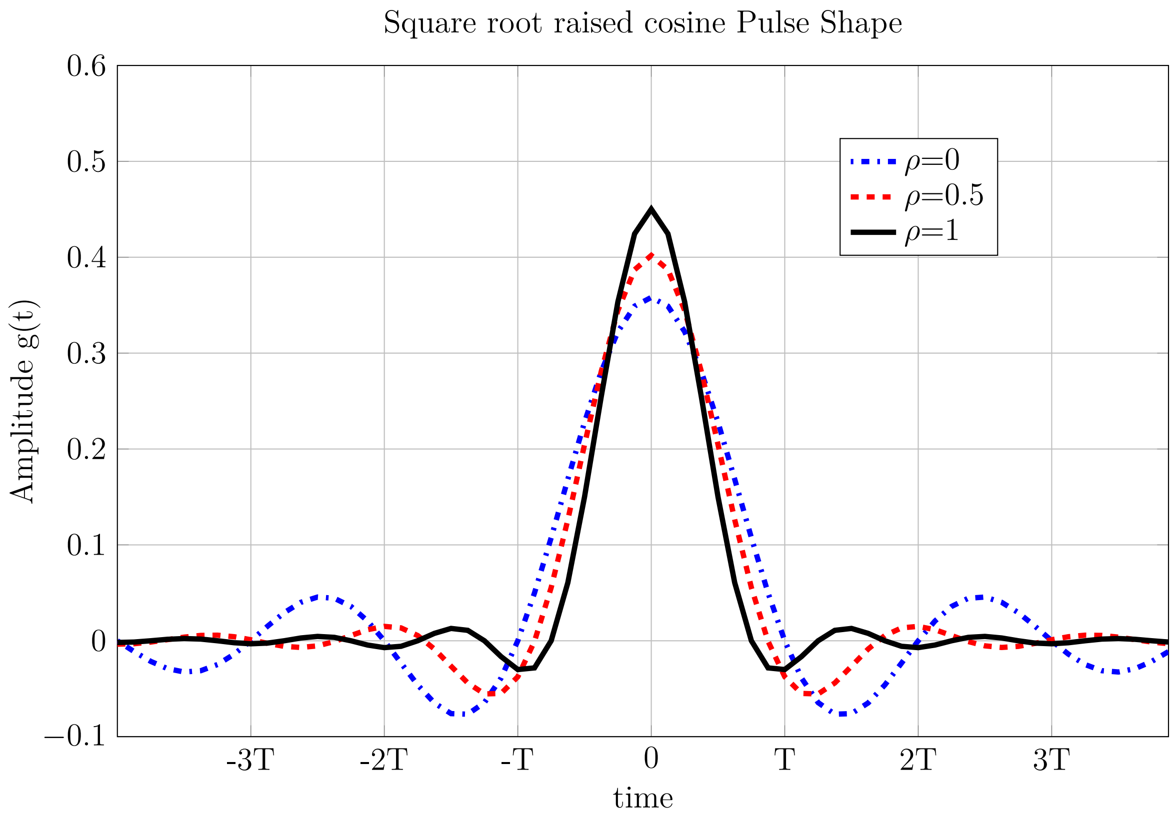


Figure 3.9: Impulse response of square root raised cosine pulse shape

3.5 Digital Bandpass Modulation Techniques

Bandpass modulation (either analog or digital) is the process by which an information signal is converted to a sinusoidal waveform; for digital modulation, such a sinusoid of duration T is referred to as a digital symbol. The information bearing signal can be expressed as

$$s(t) = A(t)\cos[\omega_0 t + \phi(t)] \tag{3.43}$$

where ω_0 is the radian frequency of the carrier and $\phi(t)$ is the phase. Fundamental the digital modulation schemes can be classified in synchronous and asynchronous systems. A brief list of the schemes is tabulated below, and simplistic description is provided there after.

Coherent	Non-coherent
Phase shift Keying	Differential phase shift keying
Frequency shift Keying	Frequency shift keying
Amplitude shift Keying	Amplitude shift keying
Continuous phase shift keying	Continuous phase shift keying
variants thereoff	variants thereoff

Table 3.2: classification of coherent and non-coherent transmission techniques.

Phase Shift Keying

Phase shift keying (PSK) is widely used in military and commercial communication systems. The general analytical expression for PSK is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_0 + \phi_i(t)) \qquad 0 \leq t \leq T$$
$$i = 1, \dots, M$$

where $\phi_i(t)$ will have M discrete values i.e.

$$\phi_i(t) = \frac{2\pi}{M} i, \qquad i{=}1, \dots, M \tag{3.44}$$

The waveform of PSK signal is illustrated in Fig. 3.10.

Frequency Shift Keying

The general analytical expression for FSK modulation is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \qquad 0 \leq t \leq T$$
$$i = 1, \dots, M$$

where frequency term ω_i has M discrete values and an arbitrary constant phase. The waveform of the above signal is illustrated in Fig. 3.10.

Amplitude Shift Keying

The general analytical expression for ASK modulation is

$$s_i(t) = \sqrt{\frac{2E_i}{T}} \cos(\omega_0 t + \phi) \qquad 0 \leq t \leq T$$
$$i = 1, \ldots, M$$

where the gain term E_i has M discrete values and an arbitrary constant phase.

Amplitude Phase Keying

The general analytical expression for APSK modulation is

$$s_i(t) = \sqrt{\frac{2E_i}{T}} \cos(\omega_0 t + \phi_i) \qquad 0 \leq t \leq T$$
$$i = 1, \ldots, M$$

where combination of amplitude and phase terms E_i and ϕ_i have M discrete values.

Differential Phase Keying

The problem with coherent transmission system is that the demodulator requires exact estimate of carrier phase. However in practice carrier phase is extract from the received signal is generally not accurate.

One way to overcome this problem is to encode information in the phase differences between successive signal transmission as opposed to absolute phase encoding. For BPSK scheme information bit 1 may be transmitted by shifting the phase of the carrier by 180° relative to previous phase while bit 0 is transmitted by a zero phase shift relative to previous signalling phase. This scheme can be extended to higher order of constellation.

Example of differential encoding(decoding) process for BPSK system.

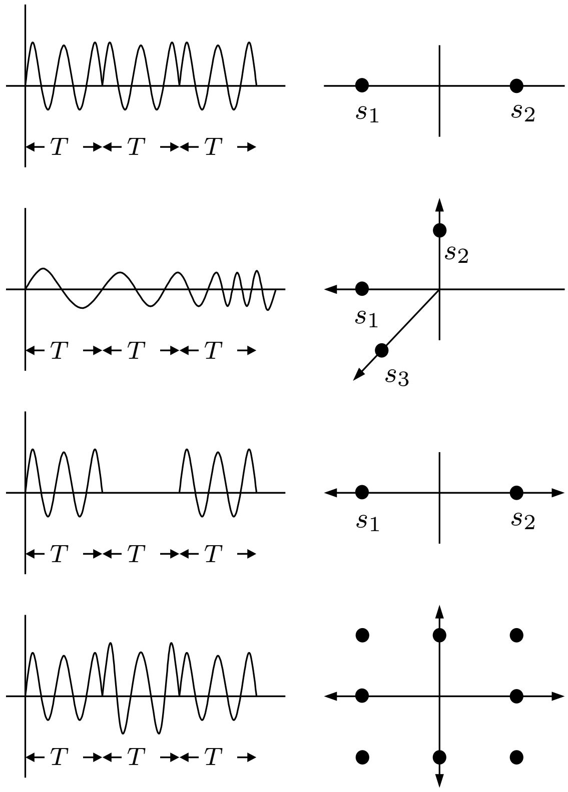


Figure 3.10: Graphical illustration of different shift keying schemes.

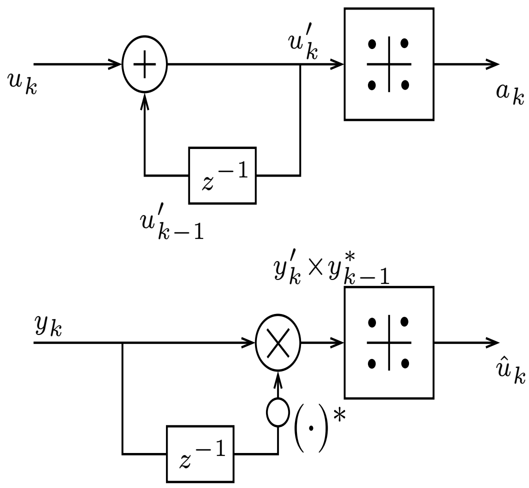


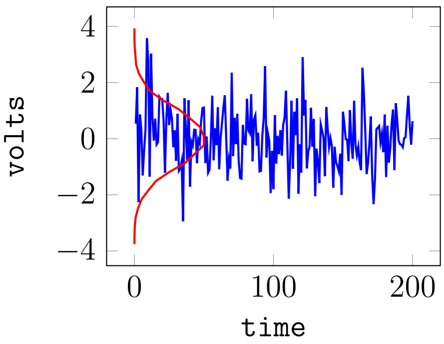
Figure 3.11: Encoder and decoder for differential phase shift keying.

3.6 Communication Channels

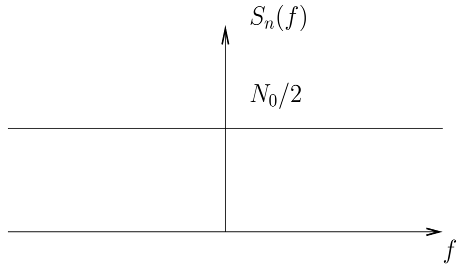
In this section, we consider a few typical channel models commonly found in literature.

k	0	1	2	3	4	5
u_k	1	1	0	1	1	1
u_{k-1}						
u'_k						
a_k=y_k						
y^*_{k-1}						
$\hat{u}_k$						

Table 3.3: where a_k corresponds to mapping binary values into bipolar symbols i.e. ‘$0'$ $\rightarrow 1$ and ‘$1'$ $\rightarrow -1$.



(a) A realization of AWGN process.



(b) PSD of AWGN process.

3.6.1 Noisy channel

The noise analysis of a communication system based on idealized form of noise process called white noise. The adjective white refers to the fact that white light contains equal amount of all frequencies within visible band. The auto-correlation and power spectral density of noise process are illustrated in Fig. **??**.

The PSD of white noise process is defined as

$$S_n(f) = \frac{N_0}{2} = \frac{kT}{2} \tag{3.45}$$

where k is the *Boltzmann's constant*, T is the ambient temperature measured in kelvins and factor of 1/2 has been included to indicate that half power is associated with +ve and -ve frequencies respectively. Some important characteristics of white noise are

- $n(t)$ is a stationary process.
- $n(t)$ is zero mean process. i.e. $E\{n(t)\} = 0$
- $n(t)$ has a variance σ_n^2. i.e. $E\{n^2(t)\} = \sigma_n^2$
- $n(t)$ has autocorrelation defined as $r_{nn}(\tau) = \delta_0(\tau)$
- $n(t)$ is Gaussian process.

$$p(n) = \frac{1}{\sigma_n\sqrt{2\pi}}\exp\Big(-\frac{n^2}{2\sigma_n^2}\Big) \tag{3.46}$$

Practically the extent to which a system is exposed to white noise depends on bandwidth of the receiver. For a noise bandwidth of B_N the average noise power is defined as

$$P_N = N_0 B_N H^2(0) \tag{3.47}$$

where $H^2(0)$ is the channel PSD at DC.

3.6.2 Distortionless Transmission System

In any communication system we require the output waveform to be a replica of the input waveform. In such cases we need to minimize the distortion caused by the amplifier to the communication channel. It is of interest to determine the characteristics of *distortionless transmission*.

$$y(t) = kx(t - t_d) \tag{3.48}$$

Fourier transform of this equation yields

$$Y(\omega) = kX(\omega)e^{-j\omega t_d} \tag{3.49}$$

but

$$Y(\omega) = X(\omega)H(\omega)$$

therefore

$$H(\omega) = ke^{-j\omega t_d} \tag{3.50}$$

This function is required for distortionless transmission. From this equation it follows that

$$\begin{aligned} |H(\omega)| &= k \\ \theta_h(\omega) &= -\omega t_d \end{aligned} \tag{3.51}$$

This implies that for distortionless transmission, the amplitude response $|H(\omega)|$ should be constant, and the phase response $\theta_h(\omega)$ must be linear function of ω. The slope of $\theta_h(\omega)$ with respect to ω is $-t_d$.
The time delay resulting from signal transmission through a system is negative of the slope of the system phase response θ_h, that is

$$t_d(\omega) = -\frac{d\theta_h}{d\omega}$$

If the slope of θ_h is constant (i.e. θ_h is linear with ω), all the components are delayed

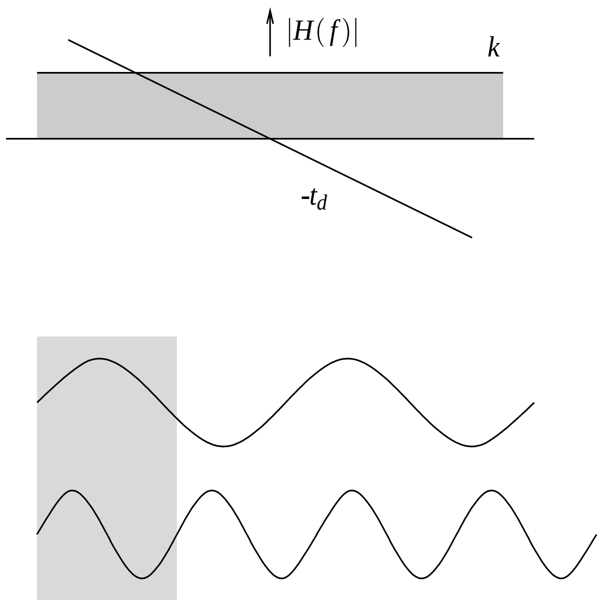


Figure 3.12: magnitude & phase response of distortionless channel.

by the same time interval, but if slope is not constant t_d, the time delay varies with frequency. This means that different frequencies go different amount of time delay and consequently output waveform will not be a replica of the input waveform.$\rightarrow$*if frequency components are delayed equally they will have different phase angle offset in time* t_d. The above is listed as follows

- Amplitude distortion, which occurs when

$$|H(f)| \neq |k|.$$

- Delay distortion which occurs when

$$\theta_h(f) \neq 2\pi t_d f \pm m180^\circ.$$

3.6.3 Non-linear Channel

Nonlinear distortion, which occurs when the system includes nonlinear elements such as diodes, transistor and amplifiers. The distortions described in last section are designated as linear distortions, since they can be described in terms of the transfer function of a linear system. A system having nonlinear distortions can not be described by a transfer function. Instead instantaneous values of input and output are related by a curve of transfer function $y(t) = T[x(t)]$ commonly called *transfer characteristic*, the flattening of output for large input signal is known as *saturation-and-cutoff* effect of transistors. Under small signal inputs, it may be possible to linearize the transfer characteristics in piecewise fashion. The more general approach is a polynomial approximation to curve of the form

$$y(t) = a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + \cdots \tag{3.52}$$

the higher powers of this equation give rise to nonlinear distortion. Even though we have no transfer function, the output spectrum can be found atleast in principle by convolution theorem

$$y(f) = a_1 X(f) + a_2 X(f) * X(f) + a_3 X(f) * X(f) * X(f) + \cdots \tag{3.53}$$

Common reason for nonlinear distortions as amplitude limiter, amplifier non-linearity

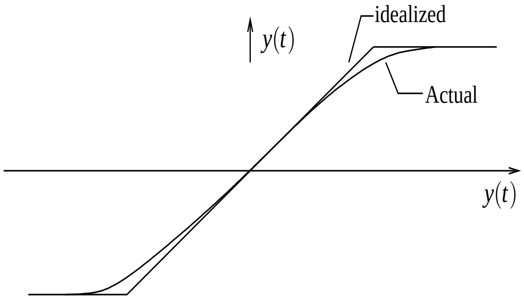


Figure 3.13: transfer characteristic of transistor / amplifier.

and laser diodes; As a consequence of this, the $x(t)$ which was bandlimited to W would contain frequencies beyond W. The components which are outside W can be eliminated via filtering, but the ones inside the band of W can not be removed.

3.6.4 Discrete-time model for channel with memory

Extra Reading
For further detail refer Proakis section 6.4

Fig. 3.14 illustrates the transmission chain of a generic communication system. The transmitter sends data at the rate of $1/T_s$ symbols per second. The input bits $u[n]$ are mapped onto symbols which are then applied to a suitable analog transmit pulse shaping filter $g_{\text{Tx}}(t)$ and is transmitted over a communication channel $c(t)$ (which may be with or without memory), on the receiver we have receive filter $g_{\text{Rx}}(t)$ which is matched to the product of communication channel $c(t)$ and $g_{\text{Tx}}(t)$, output of this filter followed by buad rate the sampling can be equivalently expressed as a discrete-time transversal filter $h[l]$. The input $a[n]$ applied to this equivalent filter $h[l]$ is the sequence of information symbols and its output $y[n]$ is also a sampled discrete-time sequence.

For simplicity of mathematical model in literature the overall cascade of Tx/Rx and bandpass RF channel are combined together

$$h(t) = g_{\text{Tx}}(t) * c(t) * g_{\text{Rx}}(t) \tag{3.54}$$

Note that $h(t)$ is also the pulse shape that will appear at the out of the receiver filter when single flat top pulse is fed into the transmitting filter (i.e. channel impulse

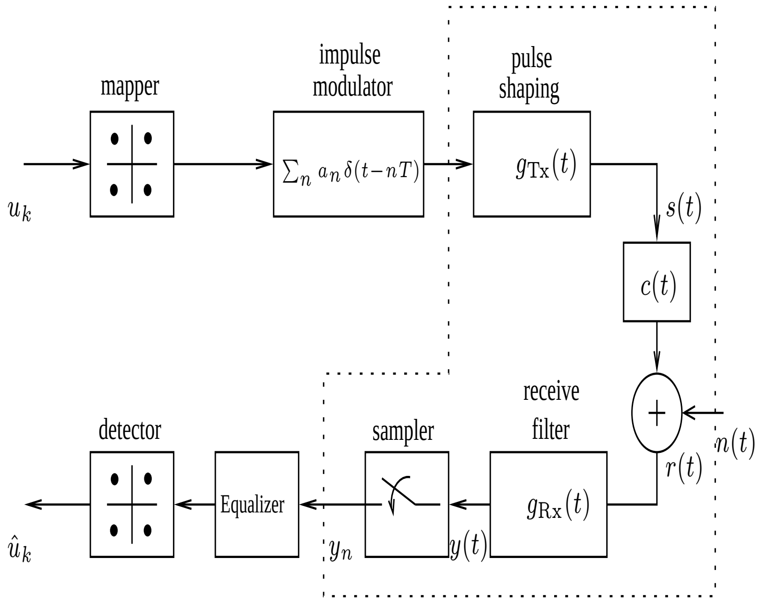


Figure 3.14: Illustration of equivalent discrete time channel model

response).

Since the output of the signal is obtained after the baud-rate sampling therefore the received signal is defined as

$$y[n] = \sum_{l=0}^{L_h-1} h[l]a[n-l] \tag{3.55}$$

The equivalent system transfer function is

$$H(f) = G_{\mathrm{Tx}}(f) \cdot C(f) \cdot G_{\mathrm{Rx}}(f) \tag{3.56}$$

Here $H(f)$ is the combined channel frequency response of the Tx/Rx and RF band-pass filters. Ideally this channel frequency response should be

$$H(f) = \mathfrak{F}\left[T_s\left(\frac{sin(\pi f_s t)}{\pi f_s t}\right)\right] = \frac{1}{f_s}\Pi\left(\frac{f}{f_s}\right) \tag{3.57}$$

For this requirement the $G_{\mathrm{Rx}}(f)$ can be determine as

$$G_{\mathrm{Rx}}(f) = \frac{H(f)}{G_{\mathrm{Tx}}(f) \cdot C(f)} \tag{3.58}$$

When $G_{\mathrm{Rx}}(f)$ is obtained from (3.58) is called the *equalizing filter* since this Rx filter

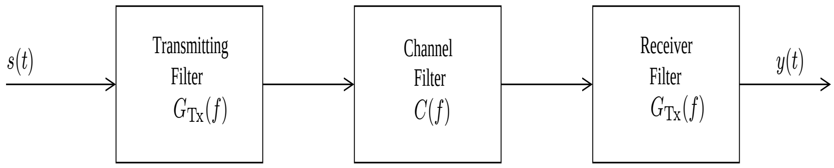


Figure 3.15: Equivalent transmission system

will help minimize the effects of inter symbol interference (ISI). The characteristics of equalizing filter depend on $C(f)$, the RF channel frequency response, as well as on the $H(f)$. If the channel varies in time (as is the case in wireless communication system) the equalizing filter needs to be an adaptive filter. The object of these equalizing filters is to *maximize the eye opening.*

The effects of ISI can be mitigated using certain techniques which are commonly referred to as Nyquist criterion for ISI free transmission. In the following subsections we shed light on these requirements.

This model is used commonly to describe a communication system in the presence of channel with memory. The characteristics of the channel filter depends on the nature of the medium.

3.7 Intersymbol Interference

Extra Reading

For further detail refer

Hayking section 4.4

Sklar section 3.3

If the transmitted signal pulses are filtered improperly as they pass through a communication system , they will spread in time. and pulse for each symbol may be smeared into adjacent time slots and cause ISI as illustrated in the Fig. 3.16(b). This problem was first studied by Nyquist. He discovered three methods for pulse shaping that could be used to eliminate ISI which will be discussed subsequently.
If the transmitted signal pulses are filtered improperly as they pass through a communication system , they will spread in time. and pulse for each symbol may be smeared into adjacent time slots and cause ISI as illustrated in the Fig. 3.16(b). This problem was first studied by Nyquist. He discovered three methods for pulse

shaping that could be used to eliminate ISI which will be discussed subsequently.

$$y[n] = \sum_{l=0}^{L-1} h[l]a[n-l] \tag{3.59}$$

where a_n may take on any of the allowed M multilevels ($M = 2$ for binary signaling) and $h[l]$ is the discrete-time channel impulse response which may be spread over several symbols intervals. The symbol rate is $R = 1/T_s$. It must be said that the equivalent channel $h(t)$ is continuous, but due to baud rate sampling, the response of channel at values other than multiples of T_s does not the effect the output. This fact is illustrated in the fig. 3.17.

3.7.1 Nyquist's First Criterion

Extra Reading
For further detail refer Couch section 3.6

Nyquist's first criterion for eliminating ISI is to use an equivalent transfer function $H_e(f)$, such that the impulse response satisfies the condition

$$h_e(t - kT_s) = \begin{cases} C, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{3.60}$$

where k is an integer, T_s is the symbol interval and C is a nonzero constant. That is, for a single flat-top pulse of level a present at the input of the transmitting filter at $t = 0$, the received pulse would be $ah_e(t)$. It would have a value of aC at $t = kT_S$, but would not cause interference at other sampling times because $h_e(t - kT_s)$=0 for $k \neq 0$.

If we choose that a $\sin(x)/x$ function for $h_e(t)$, then the impulse response of the equivalent channel would be

$$h_e(t) = \frac{sin(\pi f_s t)}{\pi f_s t} \tag{3.61}$$

where f_s is the baud rate, the transmit spectrum looks like

$$H_e(f) = \frac{1}{f_s} \Pi \left(\frac{f}{f_s} \right) \tag{3.62}$$

There will be no ISI. Furthermore, the absolute bandwidth of this transfer function is $B = f_s/2$.

- The overall amplitude transfer function of the equivalent channel $H_e(f)$ is flat over $-B < f < B$ and zero elsewhere. This is physically unrealizable (i.e. the impulse would be non-causal and of infinite duration). $H_e(f)$ is difficult to approximate because sharp transitions on the edges.

- The synchronization of the clock in the decoding sampling circuit has be to almost perfect, since the $sin(x)/x$ pulse decays only as $1/x$ and is zero in adjacent time slots.

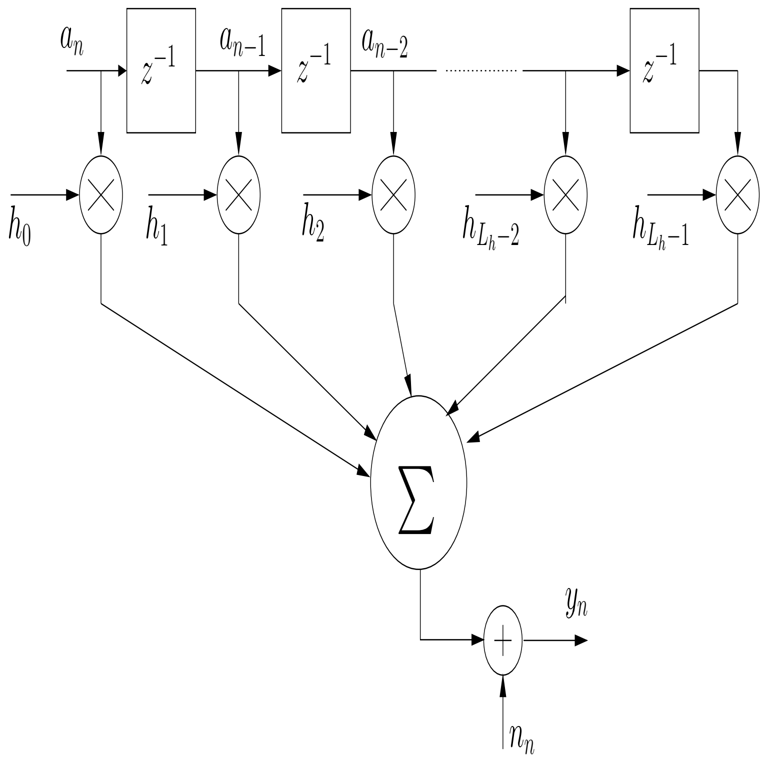
3.7.2 Nyquist's Second & Third Criterion

Nyquist's second scheme for ISI control allows some ISI to be introduced in a controlled way so that it can be cancelled out at the receiver and the data can be recovered without error if no noise is present.

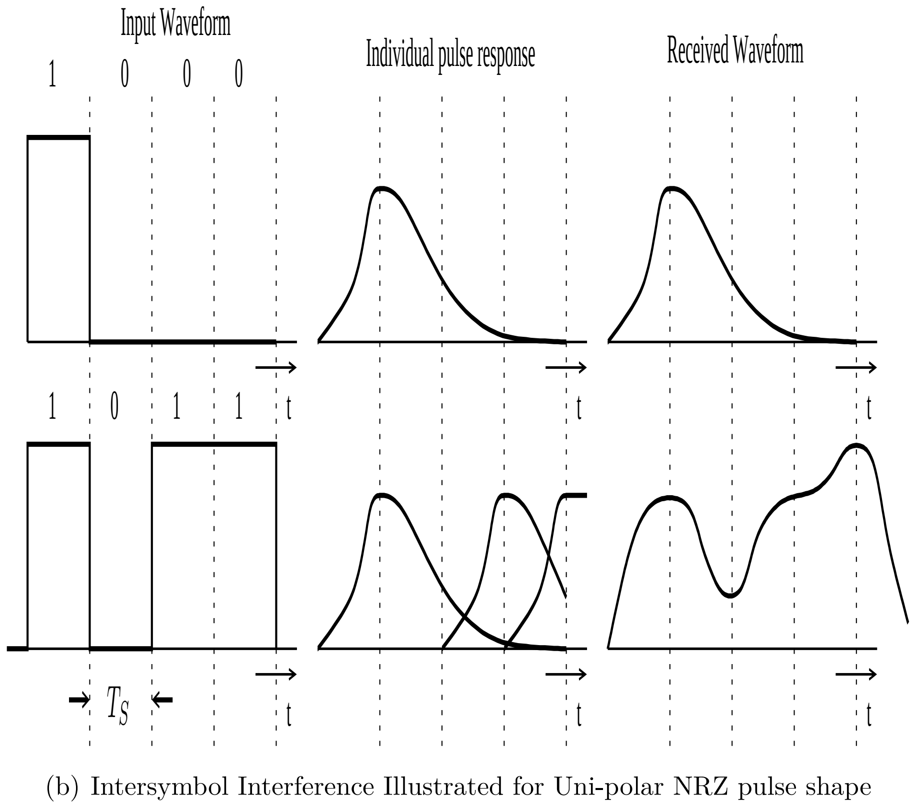
Extra Reading

For further detail refer Couch Section 3.6

In Nyquist's third method ISI is eliminated by choosing $h_e(t)$ so that the area under $h_e(t)$ pulse within the desired symbol interval is not zero but the area under $h_e(t)$ in adjacent symbol interval is zero. For data detection the receiver evaluate the integral under $h_e(t)$ over each T_s interval.



(a) The equivalent channel model.



(b) Intersymbol Interference Illustrated for Uni-polar NRZ pulse shape

Figure 3.16: The equivalent discrete time model of channel with memory and illustration of its effect on rectangular pulses.

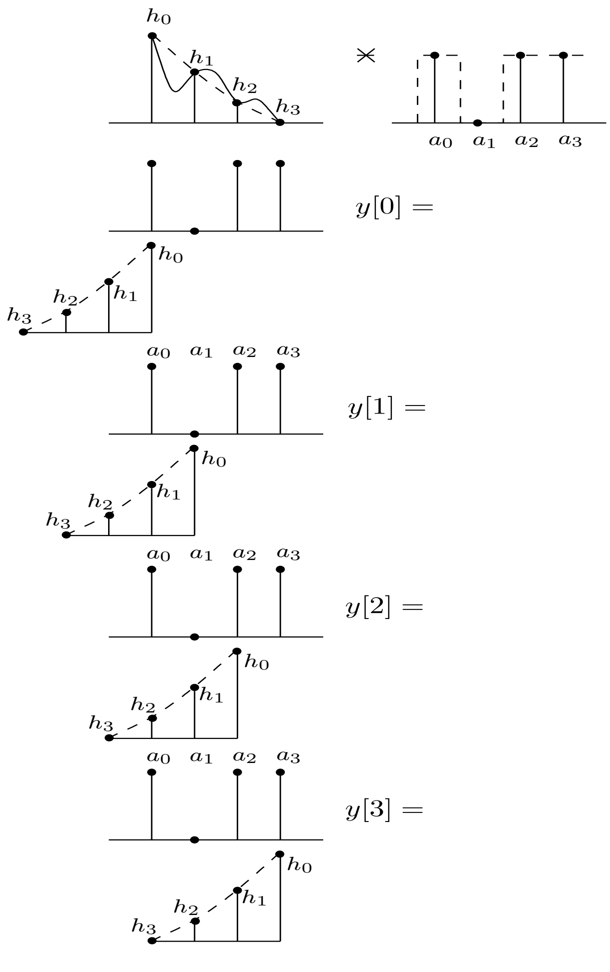


Figure 3.17: Sampled output of a equivalent channel $h(t)$.

3.8 Eye Diagram

The effect of channel filtering and channel noise can be seen by observing the received line code on an analog oscilloscope. The left side of the fig. 3.18 illustrates the received corrupted NRZ waveform for the cases of (a) ideal channel filtering, (b) filtering that produces intersysmbol interference (ISI) and (c) noise plus ISI. On the right hand side of the figure, corresponding oscilloscope presentations of the corrupted signal with multiple sweeps, where each sweep is triggered by a clock signal and sweep width is slightly longer than T_b. These displays on the right are called the *eye patterns*. The eye diagram presents a good insight into the overall

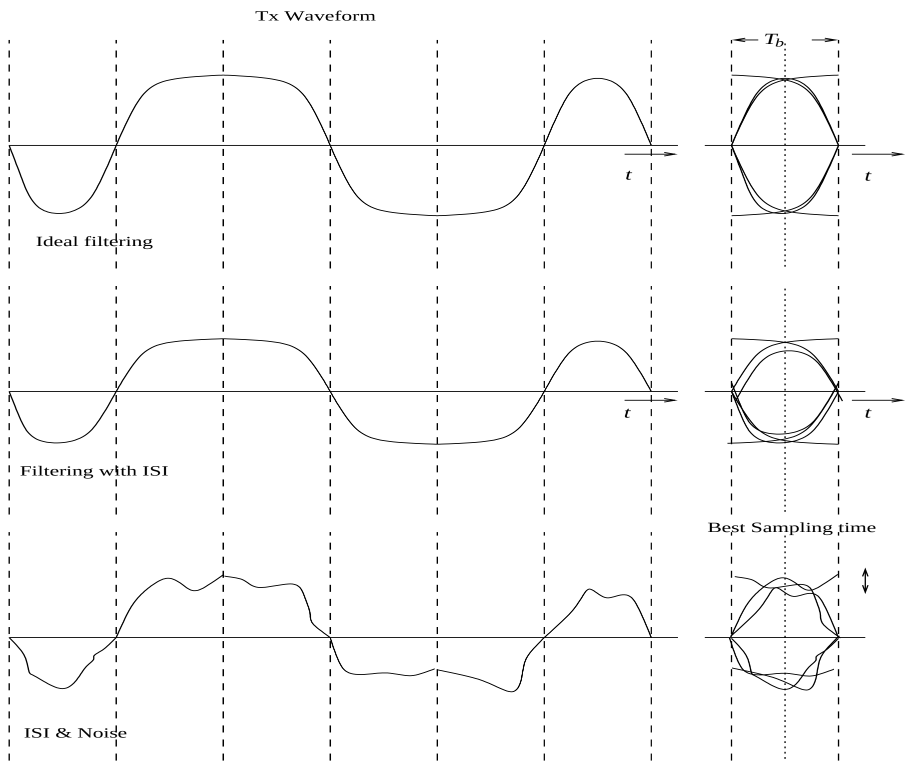


Figure 3.18: Eye diagram pattern

performance a transmission system for example SNR, BER, optimal sampling point, clock synchronization. The illustration of this process is illustrated in the fig. 3.19. Here the maximum level difference between top and bottom lines reference to the term *vertical eye-opening*. The point in symbol duration, where this takes place is the optimal point for baud rate sampling. The *horizontal eye opening* is the duration

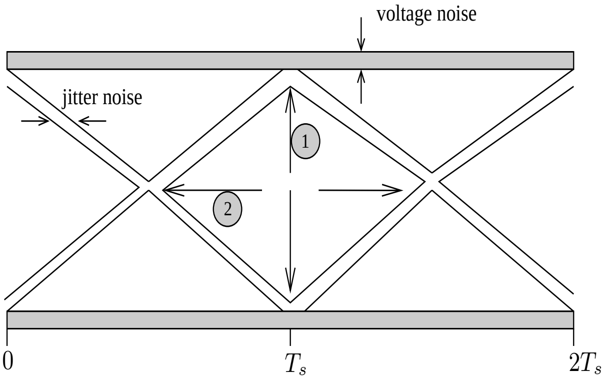


Figure 3.19: Useful features of an eye diagram.

between two successive level crossings. The extent of horizontal eye opening depends on the choice of the shaping pulse used in transmission system. Another example of the eye diagram calculation is illustrated in the figure below

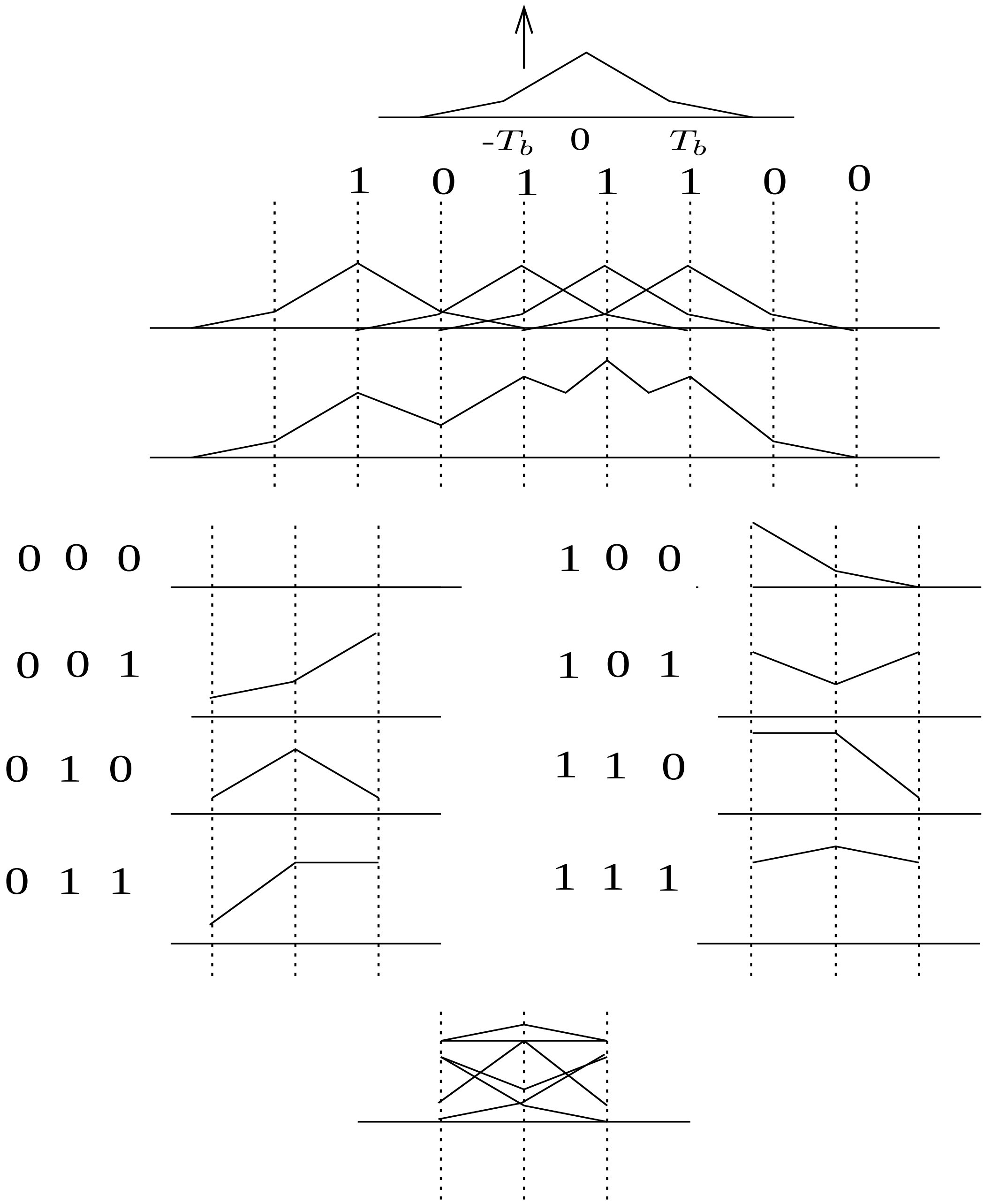


Figure 3.20: Plot of eye-diagram pattern for a pulse span multiple symbol intervals.

3.9 Matched filter

A basic problem that often arises in the study of communication system is that of detecting a pulse transmitted over a channel that is corrupted by channel noise. For the moment we assume that additive noise is the only source of disturbance present in the system.

Consider the receiver model of fig.3.22 involving a linear time-invariant channel impulse response $h(t)$. The output of this filter $y(t)$ consists of pulse signal $g(t)$ corrupted by additive noise $w(t)$

$$r(t) = ag(t) + n(t) \qquad 0 \le t \le T \tag{3.63}$$

where T is an arbitrary observation interval. The pulse signal $g(t)$ may represent a binary symbol 1 or 0 in digital communication system. $n(t)$ is the noise sample $n \sim \mathcal{N}(0, N_0/2)$ effecting the received signal. The function of the receiver is to detect the transmitted symbol in an optimal manner given the received signal $r(t)$. We have to optimize the design of receive filter so as to minimize the effect of noise at the receiver output in some statistical sense. Since the the receive filter is linear, the resulting output $y(t)$ may be expressed as

$$\begin{aligned} y(t) &= h(t) * r(t) = a \cdot h(t) * g(t) + h(t) * n(t) \\ &= a \cdot g_0(t) + \tilde{n}(t) \end{aligned} \tag{3.64}$$

where $g_0(t)$ and $\tilde{n}(t)$ are produced by the signal and noise components present in $r(t)$. A simple way to describe the requirement of output signal component $g_0(t)$ be considerably greater than the output noise component $\tilde{n}(t)$ is to have a filter which makes the instantaneous power of desired signal $g_0(t)$ measured at time t=T_s as large as possible compared with the average power of the output noise $\tilde{n}(t)$. This is equivalent to maximizing the peak pulse SNR, defined as

$$\text{SNR} = \frac{|g_0(t)|^2}{E\{\tilde{n}^2(t)\}} \tag{3.65}$$

where $|g_0(t)|^2$ is the instantaneous power in the output signal and $E\{\tilde{n}^2(t)\}$ is the variance of noise power. The requirement is to find impulse response $h(t)$ of the filter such that output signal SNR is maximized.

Extra Reading

For further detail refer

Proakis section 5.1

Haykin section 4.2

Let $G(f)$ denote Fourier transform of the known signal $g(t)$ and $H(f)$ denote the frequency response of the receiver filter $h(t)$. The Fourier transform of the received signal $g_0(t)$ is $G(f)H(f)$ and $g_0(t)$ is given in terms of inverse Fourier transforms as

$$g_0(t) = \int_{-\infty}^{+\infty} a \cdot H(f)G(f)e^{j2\pi ft}df \tag{3.66}$$

hence, when the filter output is sampled at time $t = T_s$

$$|g_0(t)|^2 = \left|a \cdot \int_{-\infty}^{+\infty} H(f)G(f)e^{j2\pi ft}df\right|^2 \tag{3.67}$$

Now consider the effect on the filter output due to noise alone. The PSD of output noise $n(t)$ is equal to the PSD of the input noise signal $n(t)$ times the squared magnitude response of the filter $|H(f)|^2$. Since $n(t)$ is white with constant PSD $N_0/2$, we have

$$S_N(f) = \frac{N_0}{2}|H(f)|^2 \tag{3.68}$$

The average power of the output noise $n(t)$ is

$$E\{\tilde{n}^2(t)\} = \frac{N_0}{2}\int_{-\infty}^{+\infty} |H(f)|^2 df \tag{3.69}$$

Through substitution in previous equations

$$\text{SNR} = \frac{\left|a \cdot \int_{-\infty}^{+\infty} H(f)G(f)e^{j2\pi ft}df\right|^2}{\frac{N_0}{2}\int_{-\infty}^{+\infty} |H(f)|^2 df} \tag{3.70}$$

The numerator term can be decomposed using *Schwarz's inequality.* One form of the inequality can be stated as

$$\left|\int_{-\infty}^{+\infty} f_1(x)f_2(x)dx\right|^2 \leq \int_{-\infty}^{+\infty} |f_1(x)|^2 dx \int_{-\infty}^{+\infty} |f_2(x)|^2 dx \tag{3.71}$$

The inequality holds if $f_1(x)$=$kf_2^*(x)$, where k is an arbitrary constant. If we identify $H(f)$ with $f_1(x)$ and $S(f)e^{2\pi fT}$ with $f_2(x)$, inserting (3.71) in (3.70) we obtain

$$\text{SNR} = \frac{a^2 \cdot \left|\int_{-\infty}^{+\infty} H(f)df\right|^2 \int_{-\infty}^{+\infty} G(f)e^{j2\pi ft}df\Big|^2}{\frac{N_0}{2}\int_{-\infty}^{+\infty} |H(f)|^2 df} \tag{3.72}$$

which translates into

$$\text{SNR} = \frac{2\mathcal{E}_s}{N_0} \cdot a^2 \tag{3.73}$$

The graphical illustration for a causal impulse response its corresponding matched filter is illustrated in the fig. 3.21. The practical implementation of a matched filter

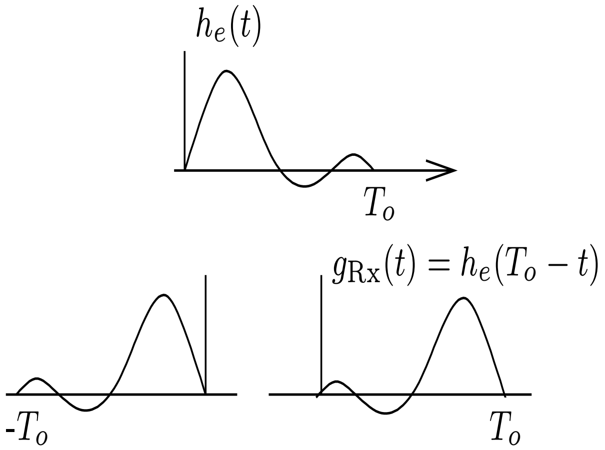


Figure 3.21: Sample illustration of channel impulse response and it's optimal matched filter according to (3.72).

receiver is presented in fig. 3.22.

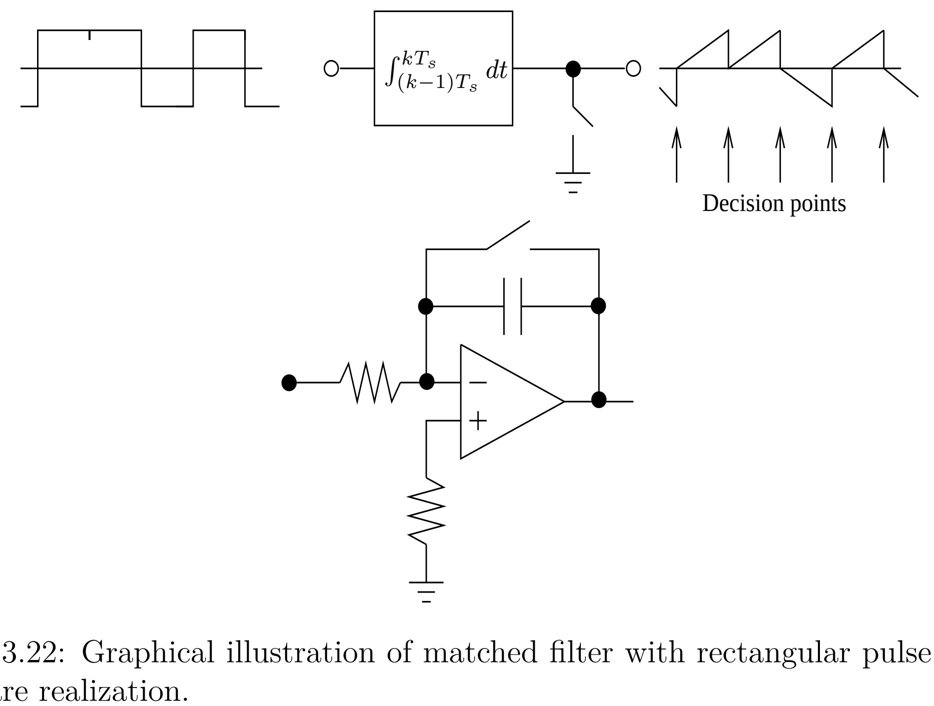


Figure 3.22: Graphical illustration of matched filter with rectangular pulse and its hardware realization.

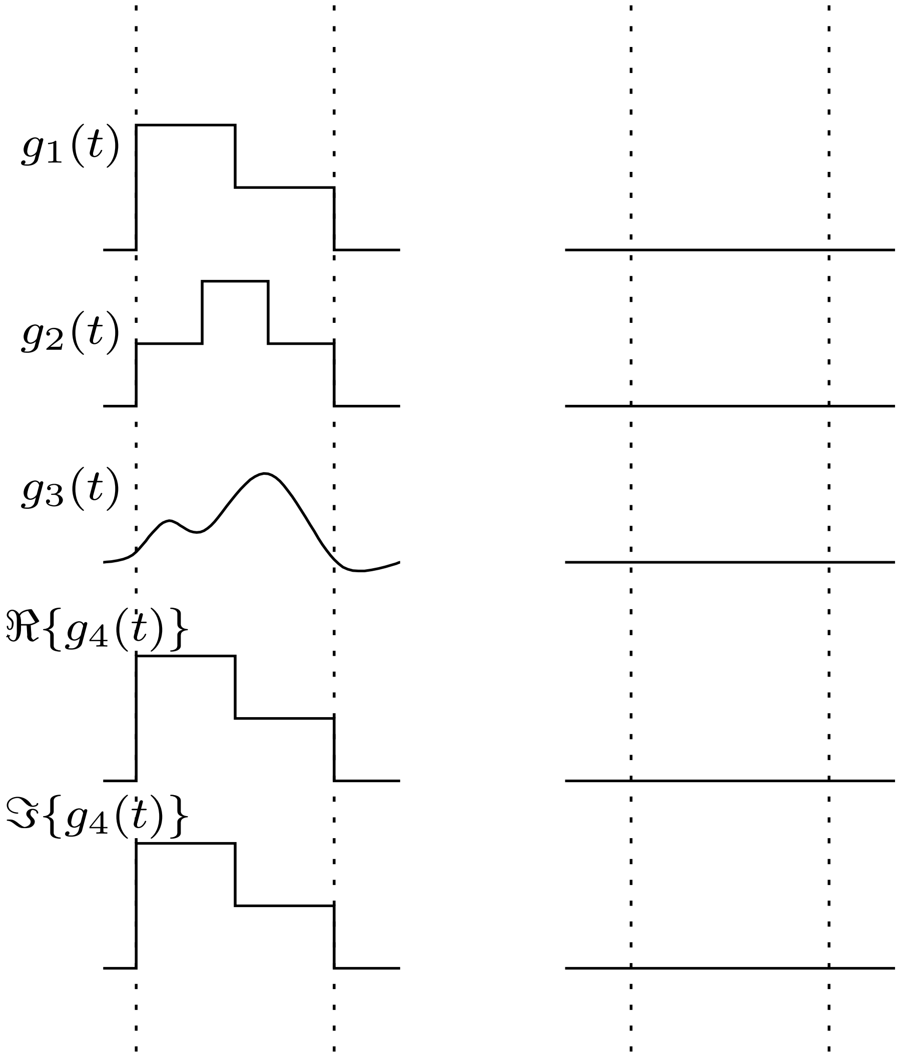


Figure 3.23: Matched filter examples.

3.10 Equalization

To compensate for linear distortion an system may be placed in ISI is a direct consequence of band-limited channels. At any given time a received data symbol $x[k]$ may be influence by L preceding data symbols. In literature there exist several techniques to mitigate the effects of ISI. In the following we discuss certain schemes highlighting possible approaches to solve this problem.

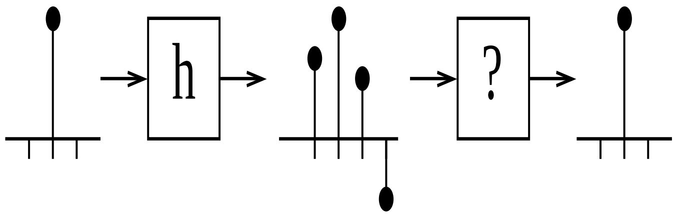


Figure 3.24: A typical transmission system in the presence of equalization.

3.10.1 Transversal filter

One approach to compensate for ISI is to use linear transversal filter. The complexity of such transversal filters grow exponentially with the length of channel memory as evident from (3.76). Graphical illustration of the transveral filter is illustrated in fig. 3.25 below.

$$\hat{s}[k] = \sum_{i=-K}^{K} c[i]y[k-i] \tag{3.74}$$

where $c[i]$ are the $2k{+}1$ complex-values coefficients of the filter. The estimates $\hat{s}[k]$ is quantized to the nearest information symbol. Considerable research has been made on the criterion to optimize the filter coefficients $c[k]$. Since the ultimate objective of a digital communication system is to minimize the bit error rate, it is desireable to choose the coefficients which minimize this performance index. However the BER performance is a highly non-linear function of $c[k]$ therefore optimizing the coefficients under this performance index is not feasible.

The system model can be expressed in matrix notation as

$$\hat{\mathbf{s}} = \mathbf{Y}\mathbf{c} \tag{3.75}$$

where $\mathbf{Y}$ is the diagonal band matrix obtained from the transmitted vector $\mathbf{y}$ at the input of the equalizer (when only one impulse is transmitted through the time invariant filter). $\mathbf{s}$ is the equalized output and $\mathbf{c}$ is the unknown equalizer coefficient which we would like to estimate. A simple least square (LS) solution can be found

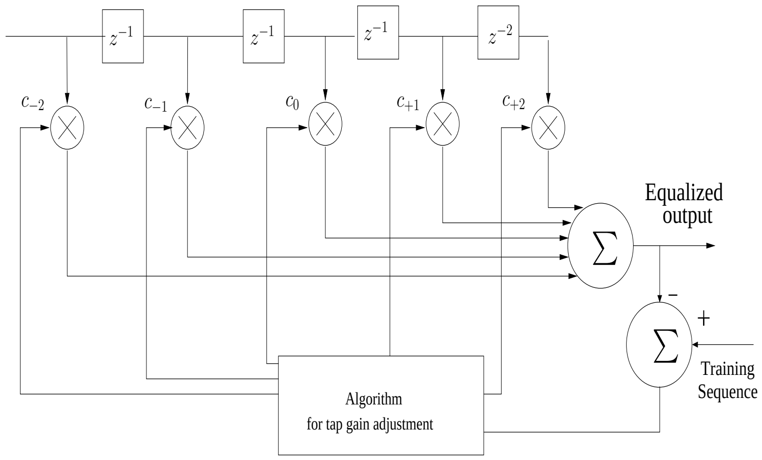


Figure 3.25: The graphical illustration of adaptive transversal filter for zero-forcing equalization.

through

$$\begin{aligned}\mathbf{Y}^T\mathbf{s} &= \mathbf{Y}^T\mathbf{Y}\mathbf{c} \\ \hat{\mathbf{c}} &= \left(\mathbf{Y}^T\mathbf{Y}\right)^{-1}\mathbf{Y}^T\mathbf{s} \qquad (3.76)\end{aligned}$$

The process of ZF equalization for a simple example is illustrated in the following example. Let the received (observed) set of samples $y[k]$=$\{0.0108, -0.0558, 0.1617, 1.00, -0.1$

we can use a LS solution using

$$
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
=
\begin{bmatrix}
0.0110 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.227 & 0.0110 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.1749 & 0.227 & 0.0110 & 0 & 0 & 0 & 0 & 0 & 0 \\
1.000 & -0.1749 & 0.227 & 0.0110 & 0 & 0 & 0 & 0 & 0 \\
0.1617 & 1.000 & -0.1749 & 0.227 & 0.0110 & 0 & 0 & 0 & 0 \\
-0.0558 & 0.1617 & 1.000 & -0.1749 & 0.227 & 0.0110 & 0 & 0 & 0 \\
0.0108 & -0.0558 & 0.1617 & 1.000 & -0.1749 & 0.227 & 0.0110 & 0 & 0 \\
0 & 0.0108 & -0.0558 & 0.1617 & 1.000 & -0.1749 & 0.227 & 0.0110 & 0 \\
0 & 0 & 0.0108 & -0.0558 & 0.1617 & 1.000 & -0.1749 & 0.227 & 0.0110 \\
0 & 0 & 0 & 0.0108 & -0.0558 & 0.1617 & 1.000 & -0.1749 & 0.227 \\
0 & 0 & 0 & 0 & 0.0108 & -0.0558 & 0.1617 & 1.000 & -0.1749 \\
0 & 0 & 0 & 0 & 0 & 0.0108 & -0.0558 & 0.1617 & 1.000 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.0108 & -0.0558 & 0.1617 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0108 & -0.0558 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0108
\end{bmatrix}
\cdot
\begin{bmatrix} c[-3] \\ c[-2] \\ c[-1] \\ c[0] \\ c[1] \\ c[2] \\ c[3] \end{bmatrix}
$$

Which means complete elimination of ISI requires the use of an inverse filter of $F(z)$. The performance of the these equalizers needs to be discussed further.

The matrix representation of the ZFE (for the case when channel and equalizer length of 3 and 4 respectively) can be expressed as

$$
\begin{bmatrix}
y[0] & & & & \\
y[1] & y[0] & & 0 & \\
y[2] & y[1] & y[0] & & \\
y[3] & y[2] & y[1] & y[0] & \\
 & y[3] & y[2] & y[1] & y[0] \\
 & & y[3] & y[2] & y[1] & y[0] \\
 & & & y[3] & y[2] & y[1] \\
 & & 0 & & y[3] & y[2] \\
 & & & & & y[3]
\end{bmatrix}
\cdot
\begin{bmatrix} c[0] \\ c[1] \\ c[2] \\ c[3] \\ c[4] \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
+
\begin{bmatrix} e[0] \\ e[1] \\ e[2] \\ e[3] \\ e[4] \\ e[5] \\ e[6] \\ e[7] \end{bmatrix}
$$

$$\mathbf{Yc} = \mathbf{s}_d + \mathbf{e} \tag{3.77}$$

This equalizer can be expanded to consider fractional spaced equalizers (not considered here). The foremost problem of this type of equalizers is that their complexity increases exponentially with the number of channel coefficients. This problem can be solved by taking the adaptive approach e.g. least mean squared (LMS) or recursive least squared (RLS) methods. These schemes are not considered in this course.

3.11 Channel coding Techniques

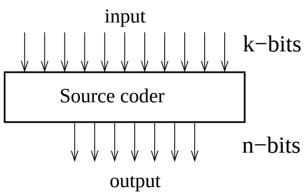
Channel coding techniques are special type of techniques/algorithms which can improve the communication performance by enabling the transmitted signals to tolerate the adverse effects of the channel such as noise, interference and fading. The

Extra Reading

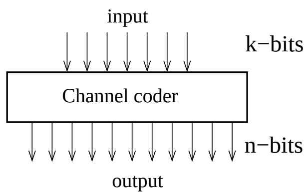
For further detail refer

Sklar section 6.4

Haykin section 10.3



(a) Structure of a block encoder.



(b) Structure of a block decoder.

Note:

Channel Coding is also known as forward error correcting code.

coding techniques can be broadly classified into two classes

1. Waveform coding: Mainly deals with formatting and shaping of data to enhance the performance of the system.

2. Structure sequences: Introduces a structured redundancy to the transmitted sequence which ensures better recovery on the receiver side.

In the case of structured codes, the source data are segmented into blocks of k data bits also called information or message bits; each can represent any one of 2^k distinct messages. The encoder transforms each k-bit data block into a larger block of n bits. The $(n - k)$ bits, which the encoder adds to each data block, are called redundant

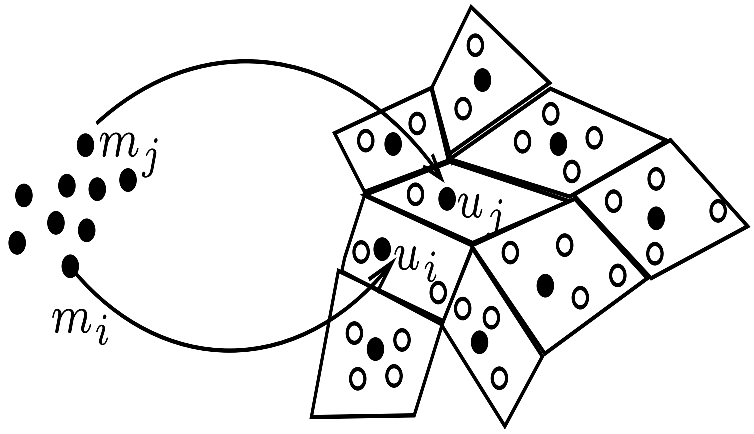


Figure 3.26: Mapping of message blocks to code blocks.

bits, parity bits or check bits; they carry no new information. The code is referred to as an (n, k) code. The ratio of data bits to total bits k/n is called code rate. Code is the portion of information in coded message.

Single Parity Check Codes

Parity check codes use linear sum of the information bits, called parity symbols or parity bits, for error detection or correction. A single parity check code is constructed by adding a single parity bit to a block of bits. The parity bit takes on the value of one or zero. The summation is performed using modulo-2 arithmetic (exclusive-or logic). This code can not automatically correct the bit errors.

(4,3) Parity codes

Configure a $(4, 3)$ even-parity error-detection code such that the parity symbol appears as the leftmost symbol of the codeword. Which error patterns can the code detect?

Generator Matrix

If k is large, a lookup implementation of the encoder become prohibitive. For a $(127, 92)$ code there are 2^{92} possible data vectors. The encoding of these vector may not be suitable through simple lookup table. However complexity can be reduced by generating the codes instead of sorting them.

Message	Parity	Codeword
000	0	0 000
001	1	1 001
010	1	1 010
011	0	0 011
100	1	1 100
101	0	0 101
110	0	0 110
111	1	1 111

Table 3.4: A simple single parity check code.

The generator matrix is the smallest linearly independent set that spans the subspace is called a basis of the subspace, and the number of vectors in this basis set is the dimension of the subspace. Any basis set of k linearly independent n-tuples $\mathbf{v}_1$, $\mathbf{v}_2$, $\cdots$, $\mathbf{v}_k$.

$$\mathbf{G} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & & & \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{bmatrix} \tag{3.78}$$

The generation of codeword $\mathbf{u}$ can be written in matrix notation as the product of $\mathbf{m}$ and generator matrix $\mathbf{G}$

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & \cdots & m_k \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & & & \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{bmatrix} \tag{3.79}$$

Example: A code word assignment for (6,3) code is illustrated for example, there are 2^k=2^3 = 8 messages and there for 8 codewords. The table 3.5 forms a subspace of V_6 (the all zero vector is present, and the sum of any two codewords yields another codeword member in the subspace). Therefore, these codewords represent a linear block code. An important question that arise at this point in time is how was the code word to message assignment chosen for this (6,3) code? A unique assignment for a particular (n, k) code does not exist. A generator matrix is defined as

Message	Codeword
000	000000
100	110100
010	011010
110	101110
001	101001
101	011101
011	110011
111	000111

Table 3.5: An example code for (6,3) parity check code.

$$\mathbf{G} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{3.80}$$

where $\mathbf{v}_1$,$\mathbf{v}_2$ and $\mathbf{v}_3$ are three linearly independent vectors (a subset of the eight code vectors) that can generate all the code vectors. Notice that the sum of any two generating vectors.

$$\begin{aligned} \mathbf{u} &= \begin{bmatrix} m_1 & m_2 & m_3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \underbrace{\overbrace{m_1 \oplus m_3}^{p_1}}_{u_1} & \underbrace{\overbrace{m_1 \oplus m_2}^{p_2}}_{u_2} & \underbrace{\overbrace{m_2 \oplus m_3}^{p_3}}_{u_3} & m_1 & m_2 & m_3 \end{bmatrix} \end{aligned} \tag{3.81}$$

An example code is generated here

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} = 1 \cdot \mathbf{v}_1 + 1 \cdot \mathbf{v}_2 + 0 \cdot \mathbf{V}_3$$

$$= 1 \;\; 1 \;\; 0 \;\; 1 \;\; 0 \;\; 0 + 0 \;\; 1 \;\; 1 \;\; 0 \;\; 1 \;\; 0 + 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0$$

$$= \_\_ \;\; \_\_ \;\; \_\_ \;\; \_\_ \;\; \_\_ \;\; \_\_$$

Thus, the code vector corresponding to the vector is a linear combination of the rows of $\mathbf{G}$, the encoder needs to only store the k rows of G instead of 2^k vectors of the code.

Systematic Linear Block Codes

A systematic (n, k) linear block code is a mapping from a k-dimensional message vector to a n-dimensional codeword in such a way that part of the sequence generated coincides with k messages digits. The remaining $(n-k)$ digits are parity digits. A systematic linear block code will have a generator matrix of the form

$$\begin{aligned}\mathbf{G} &= \begin{bmatrix} \\ \mathbf{P} & \mathbf{I}_k \\ \\ \end{bmatrix} \\ &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1(n-k)} & \vdots & 1 & 0 & \cdots & 0 \\ p_{21} & p_{22} & \cdots & p_{2(n-k)} & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & 0 & 0 & \cdots & 0 \\ p_{k1} & p_{k2} & \cdots & p_{k(n-k)} & \vdots & 0 & 0 & \cdots & 1 \end{bmatrix}\end{aligned} \tag{3.82}$$

where $\mathbf{P}$ is the parity array portion of the the generator matrix p_{ij}=(0 or 1) and $\mathbf{I}_k$ is the $k \times k$ identity matrix. The systematic generator further reduces the encoding complexity since storing of identity matrix is not required.

$$u_1 \quad u_2 \quad \cdots \quad u_n = \begin{bmatrix} m_1 & m_2 & \cdots & m_k \end{bmatrix} \times \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1(n-k)} & \vdots & 1 & 0 & \cdots & 0 \\ p_{21} & p_{22} & \cdots & p_{2(n-k)} & \vdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & 0 & 0 & \cdots & 0 \\ p_{k1} & p_{k2} & \cdots & p_{k(n-k)} & \vdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{3.83}$$

where

$$\begin{aligned} u_i =& m_1p_{1i} + m_2p_{2i} + \cdots + m_kp_{ki} \qquad && \text{for } i = 1, \cdots, (n-k) \\ & m_{i-n+k} && \text{for } i = (n-k+1), \cdots, n \end{aligned} \tag{3.84}$$

and a general systematic code can be expressed as

$$\mathbf{u} = [p_1, p_2, \ldots, p_{n-k}, m_1, m_2, \ldots, m_k] \tag{3.85}$$

where

$$\begin{aligned} p_1 &= m_1p_{11} + m_2p_{21} + \cdots + m_kp_{k1} \\ p_2 &= m_1p_{12} + m_2p_{22} + \cdots + m_kp_{k2} \\ p_{n-k} &= m_1p_{1,n-k} + m_2p_{2,n-k} + \cdots + m_kp_{k,n-k} \end{aligned} \tag{3.86}$$

It must be said that the information bearing vector may be located at any position in the vector. The first parity bit is the sum of the first and third message bits; the second parity bits is the sum of the first and second message bits; and the third parity bit is the sum of the second and third message bits. This system of equations provides insight into the structure of linear block codes.

Parity-Check Matrix

Let us define a matrix $\mathbf{H}$ called the parity-check matrix, that will enable us to decode the received vector. For each $k \times n$ generator matrix $\mathbf{G}$ there exist $(n-k) \times n$ matrix $\mathbf{H}$ such that rows of $\mathbf{G}$ are orthogonal to the rows of $\mathbf{H}$; that is $\mathbf{GH}^T$=$\mathbf{0}$ where $\mathbf{H}^T$ is the transpose of $\mathbf{H}$, and $\mathbf{0}$ is a $k \times (n-k)$ all-zero matrix. $\mathbf{H}^T$ is an $n \times (n-k)$ matrix whose rows are the columns are the rows of $\mathbf{H}$. To full fill the orthogonality requirements for a systematic code, the components of the $\mathbf{H}$ matrix for a systematic code is written as

$$\mathbf{H} = \left[\mathbf{I}_{n-k}|\mathbf{P}^T\right] \tag{3.87}$$

Hence, the $\mathbf{H}^T$ matrix is written as

$$\mathbf{H}^T = \begin{bmatrix} \mathbf{I}_{n-k} \\ \mathbf{P} \end{bmatrix} \tag{3.88}$$

Using defination of parity matrix from (3.82) and (3.86) parity check matrix is defined as

$$\mathbf{H}^T = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ p_{11} & p_{11} & \cdots & p_{11} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & & \vdots & 0 \\ p_{k1} & p_{k2} & \cdots & p_{k,n-k} \end{bmatrix} \tag{3.89}$$

It is easy to verify that the product $\mathbf{uH}^T$ of each codeword $\mathbf{u}$ generated by $\mathbf{G}$ and $\mathbf{H}$ matrix yields

$$\mathbf{uH}^T = p_1 + p_1 \quad p_2 + p_2 \quad \cdots \quad p_{n-k} + p_{n-k} = \mathbf{0} \tag{3.90}$$

where the parity bits $p_1, p_2, \ldots, p_{n-k}$ are defined as above.

Example: From the generator matrix formulated in (3.82) we can formulate the parity check matrix and decode the transmitted message as

$$\mathbf{H}^T = \begin{bmatrix} \mathbf{I}_{n-k} \\ \mathbf{P} \end{bmatrix} \tag{3.91}$$

A detailed elaborate description of the system is defined as

$$\mathbf{uH}^T =$$

$$= \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \_\_\_ & \_\_\_ & \_\_\_ \end{bmatrix} \tag{3.92}$$

Syndrome Testing

Let $\mathbf{r} = [r_1\ r_2\ \dots\ r_n]^T$ be received vector one of 2^n resulting from the transmission of $\mathbf{u} = [u_1\ u_2\ \dots\ u_n]^T$ (one of 2^k n-tuples). We can therefore describe $\mathbf{r}$ as

$$\mathbf{r} = \mathbf{u} + \mathbf{e} \tag{3.93}$$

where $\mathbf{e} = [e_1\ e_2\ \dots\ e_n]^T$ is an error vector or error pattern introduced by the channel. There are total of $2^n - 1$ potential nonzero error patterns in the space of 2^n n-tuples. The syndrome of $\mathbf{r}$ is defined

$$\mathbf{s} = \mathbf{rH}^T \tag{3.94}$$

The syndrome is the result of a parity check performed on $\mathbf{r}$ to determine whether $\mathbf{r}$ is a valid member of the codeword set. If in fact $\mathbf{r}$ is a member, the syndrome $\mathbf{S}$ has a value $\mathbf{0}$. If $\mathbf{r}$ contains detectable errors, the syndrome has some non-zero value, the syndrome can be defined as

$$\begin{aligned} \mathbf{s} &= (\mathbf{u}+\mathbf{e})\mathbf{H}^T \\ &= \mathbf{uH}^T + \mathbf{eH}^T \\ &= \mathbf{eH}^T \end{aligned} \tag{3.95}$$

from (3.93) and (3.95) it is evident that syndrome test whether performed on a corrupted code or error sequence yields similar results and there exists one to one mapping between correctable codes and syndrome. $\rightarrow$ Syndrome code can be used to estimate the error.

Some further observations are

- No rows of $\mathbf{H}$ can be an all zero vector, otherwise an error in the corresponding codeword position would not be detectable.
- All columns of $\mathbf{H}$ must be unique. If two columns are same their errors will be indistinguishable.

Example:Syndrome test

Lets us assume that $\mathbf{u} = [101110]$ from the (6,3) code mentioned earlier. With single bit error in the left-most bit so $\mathbf{r} = [001110]$. From (3.93) and (3.95) we can verify that syndrome of corrupted code vector and error vector is the same.

Error vector	Syndrome
000000	000
000001	101
000010	011
000100	110
001000	001
010000	010
100000	100
010001	111

$$\begin{aligned} \mathbf{s} &= \mathbf{r}\mathbf{H}^T \\ &= \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \mathbf{H}^T \\ &\begin{bmatrix} 1 & 1+1 & 1+1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{3.96}$$

From the above lookup table it is possible to estimate the $\hat{\mathbf{u}}$=$\mathbf{r}$+$\hat{\mathbf{e}}$. Error detection and correction capability depends on the error correction capability of the codes. Typically the error correction capability of code is defined under the following framework.

The Hamming weight $w(\mathbf{u})$ of a codeword $\mathbf{u}$ is defined as the number of ones present in a codeword (e.g. $\mathbf{u} = [101101]$ then $w(\mathbf{u}) = 4$). Furthermore Hamming distance between two codewords $\mathbf{u}_1 = [1011010]$ and $\mathbf{u}_2 = [0111001]$, the hamming distance $d(u_1, u_2) = 4$.

Consider the set of distance between all pairs of codewords in the space U_n. The smallest member of the set is the minimum distance of the code and is denoted by $d_{\min}$. The minimum distance between two codes is also defined as the minimum hamming distance between two valid codes. So a valid code $\mathbf{u}_1$ is differentiable from another valid code as long as the error vector $\mathbf{e}$ has hamming weight less than $d_{\min}$. This idea is illustrated in the fig. 3.27. The minimum distance is like the weakest link in the chain and provides a measure of the codes minimum capability and characterizes codes strength. If $\mathbf{u}_1$ and $\mathbf{u_2}$ are valid codes then $\mathbf{w} = \mathbf{u}_1 + \mathbf{u_2}$ is also a codeword.

In general the error correcting capability is depends on $d_{\min}$ of the channel code,

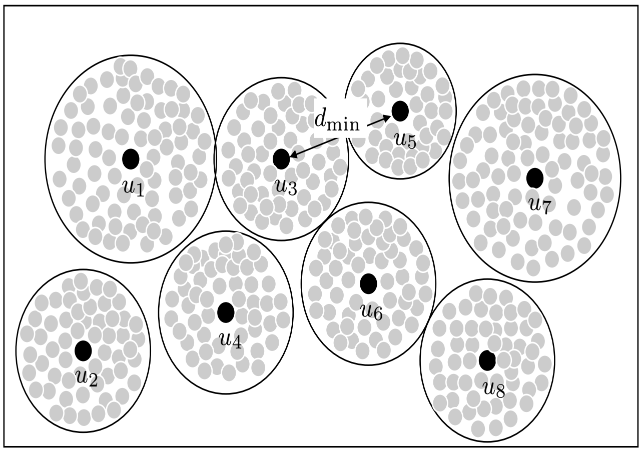


Figure 3.27: Illustration of code space, minimum distance, error detection and correction capabilities.

the error-correcting capability t of a code is defined as the maximum number of guaranteed correctable errors per codeword as written

$$t = \lfloor \frac{d_{\min} - 1}{2} \rfloor \tag{3.97}$$

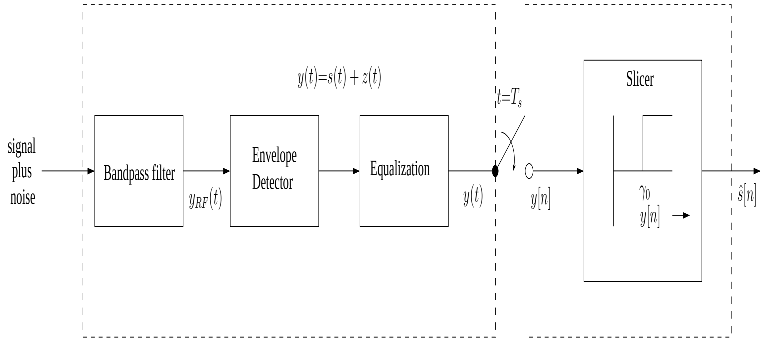
Often a code that corrects all possible sequences of t or fewer errors can also correct certain sequences of $t+1$ errors. However the error detecting capability of the code is typically defined as $e = d_{\min} - 1$.

3.12 BER analysis

The receiver filter and the equalizing filter are shown as two separate blocks in order to emphasize their separate functions. The figure demodulation/detection process

a Signal to sample transformation, where signal waveform is sampled $y[nT_s]$, this is often called test statistics.

b A decision is made regarding the digital meaning of the signal.



The receiver filter and equalizing filter are shown at two separate blocks in order to emphasize their separate functions. The figure highlights two step demodulation / detection process step 1. signal to sample transformation, where signal waveform is sampled $y(nTs)$, this is often called as test statistics. At step 2, a decision is made regarding the digital meaning of the signal. We assume that the input noise is a Gaussian random process and that the receiver filter in the demodulator is linear. A linear operation performed on a Gaussian process will produce a second Gaussian random process. Thus, the filter output noise is Gaussian. The output of the step 1 yields the test statistics

$$y(nTs) = s_i(nTs) + z(nTs) \quad i = 1,2 \tag{3.98}$$

where $s_i(nTs)$ is the desired signal component, and $z(nTs)$ is the noise component. To simplify the notation of (3.98) in the form of

$$y = s_i + z \tag{3.99}$$

The noise component z is a zero mean Gaussian random variable, and thus y is Gaussian random variable with mean of either s_1 or s_2. The probability density function of the noise process $z(t)$ is defined as

$$p(z) = \frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\Big[-\frac{1}{2}(\frac{z^2}{\sigma_z^2})\Big] \tag{3.100}$$

where σ_z^2 is the noise variance. Thus it follows from (3.98) and (3.99) that conditional pdfs $p(y|s_1)$ and $p(y|s_2)$ can be expressed as

$$p(y|s_1) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\Big[-\frac{1}{2}\frac{(y-s_1)^2}{\sigma_z^2}\Big] \tag{3.101}$$

$$p(y|s_2) = \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\Big[-\frac{1}{2}\frac{(y-s_2)^2}{\sigma_z^2}\Big] \tag{3.102}$$

These conditional pdfs are illustrated in the figure below, the rightmost conditional pdf, $p(y|s_1)$, called the likelihood of s_1, illustrates the probability density function of the random variable y(nTs) given that symbol s1 was transmitted. Similarly,

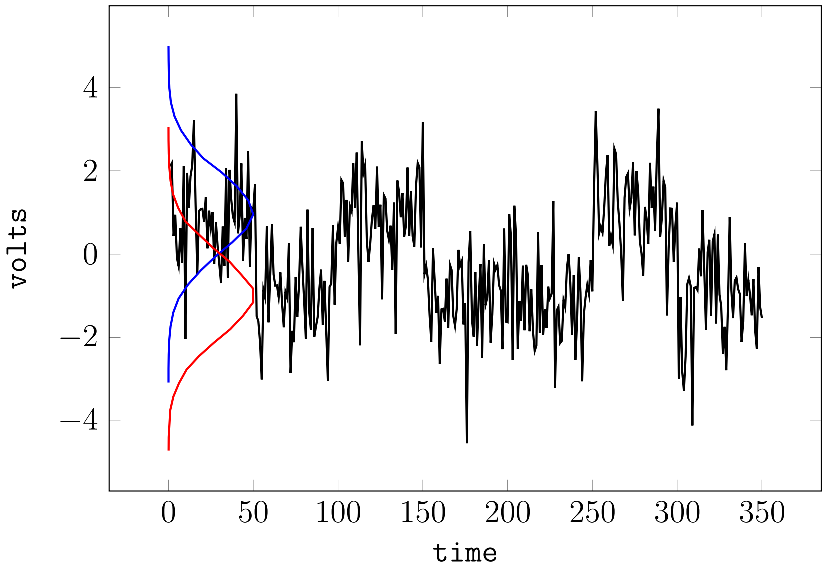
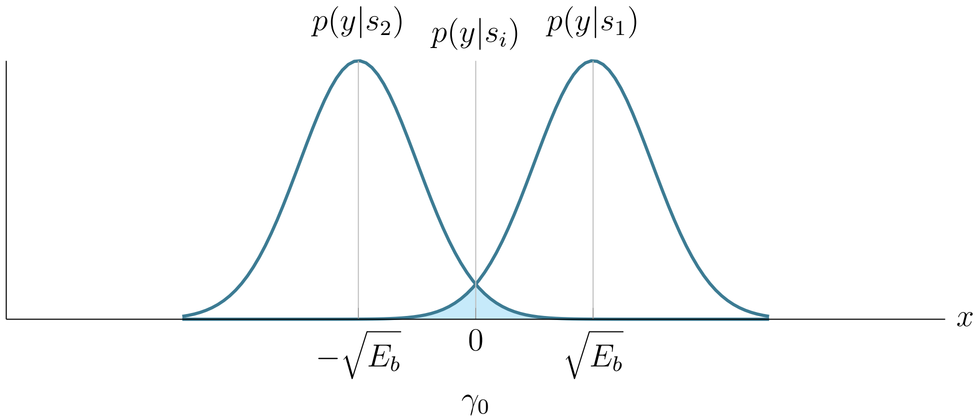


Figure 3.28: A binary NRZ transmission in the presence of AWGN noise SNR 0 dB.

the leftmost conditional pdf $p(y|s_2)$, called the likelihood of s_2, illustrates the pdf $y(nTs)$ that symbol s_2 was transmitted. The received signal energy is important parameter in the detection process. This is why the detection analysis for baseband signals is the same as that for bandpass signal. Since $y(nTs)$ is a voltage signal that is proportional to the energy of the received symbol.

$$p(e|s_1) = P(H_2|s_1) = \int_{-\infty}^{\gamma_0} p(y|s_1)dy \tag{3.103}$$

This is illustrated in the shaded area to the left of 0. Similarly an error occurs when $s_2(t)$ is sent and the channel noise $y(t)$ being greater than 0. The probability of this



occurrence is

$$p(e|s_2) = P(H_1|s_2) = \int_{\gamma_0}^{\infty} p(y|s_2)dy \tag{3.104}$$

The probability of an error is the sum of the probabilities of all the ways that an error can occur. For the binary case, we can express the bit error probability as

$$\begin{aligned} p_b &= \sum_{i=1}^{2} p(e, s_i) = \sum_{i=1}^{2} p(e|s_i)p(s_i) \\ &= p(e|s_1)p(s_1) + p(e|s_2)p(s_2) \end{aligned} \tag{3.105}$$

given the fact that signal $s_1(t)$ was transmitted, an error occurs if hypothesis H_2 is chosen (in other words s_2 is detected at the receiver while $s_1(t)$ was transmitted) ; or given that signal $s_2(t)$ was transmitted an error results if hypothesis $H1$ is chosen. For the case where the apriori probabilities are equal that is $p(s_1) = p(s_2) = 1/2$

$$p_b = \frac{1}{2}p(e|s_1)p(s_1) + \frac{1}{2}p(e|s_2)p(s_2) \tag{3.106}$$

because of the symmetry of the probability density functions

$$p_e = p(H_2|s_1) = p(H_1|s_2) \tag{3.107}$$

The probability of error is numerically equal to the area under the <u>tail</u> of either likelihood function $p(y|s_1)$ or $p(y|s_2)$ falling on the incorrect side of the threshold. We can therefore compute p_b by integrating $p(y|s_1)$ between the limits âĹŠ∞ and

γ_0 or by integrating $p(y|s_2)$ between the limits γ_0 and ∞

$$p_e = \int_{\gamma_0=(s1+s2)/2}^{\infty} (y|s_2)dy \tag{3.108}$$

Here, $\gamma_0 = (s1+s2)/2$ is the optimum threshold, replacing in

$$p_e = \int_{\gamma_0=(s1+s2)/2}^{\infty} \exp\Big[-\frac{1}{2}\Big(\frac{(y-s_2)}{\sigma_z}\Big)^2\Big]dy \tag{3.109}$$

where σ_z^2 is the variance of the noise out of the correlator. Let $u = (y-s_2)/\sigma_z$, then $\sigma_z du = dy$ and

$$p_b = \frac{1}{2}\int_{-\infty}^{\gamma_0} \frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\Big[-\frac{(y-s_1)^2}{2\sigma_z^2}\Big]dy + \frac{1}{2}\int_{\gamma_0}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_z^2}}\exp\Big[-\frac{(y-s_2)^2}{2\sigma_z^2}\Big]dy \tag{3.110}$$

Using two different substitution in both integrals we can write

$$u = -\Big(\frac{y-s_1}{\sigma_z}\Big) \quad \frac{du}{dy} = -\frac{1}{\sigma_z^2} \rightarrow du = -\frac{1}{\sigma_z}dy$$
$$y \rightarrow -\infty \Rightarrow u \rightarrow \infty$$
$$y = \gamma_0 \rightarrow -\Big(\frac{\gamma_0 - s_1}{\sigma_z}\Big)$$

$$u = \Big(\frac{y-s_2}{\sigma_z}\Big) \qquad \frac{du}{dy} = \frac{1}{\sigma_z} \rightarrow du = \frac{1}{\sigma_z}dy$$
$$y \rightarrow \gamma_0 \rightarrow \frac{(\gamma_0 - s_2)}{\sigma_z}$$
$$y \rightarrow \infty \Rightarrow u \rightarrow \infty$$

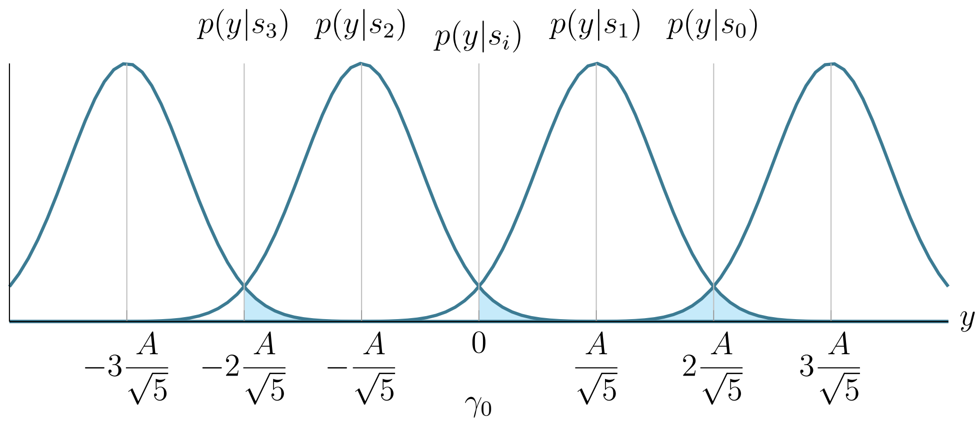
Using the above substitutions we can rewrite the expression for both conditional probability distributions as

$$\begin{aligned}
p_b &= \frac{1}{2}\int_{u=-\frac{\gamma_0 - s1}{\sigma_z}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\Big[-\frac{1}{2}u^2\Big]du + \frac{1}{2}\int_{u=\frac{\gamma_0 - s_2}{\sigma_z}} \frac{1}{\sqrt{2\pi}}\exp\Big[-\frac{1}{2}u^2\Big]du \\
&= \frac{1}{2}Q\Big(\frac{\sqrt{E_b}}{\sigma_z}\Big) + \frac{1}{2}Q\Big(\frac{\sqrt{E_b}}{\sigma_z}\Big) = Q\Big(\frac{\sqrt{E_b}}{\sigma_z}\Big) \\
&= Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big)
\end{aligned} \tag{3.111}$$

where $Q(x)$ is called complementary error function or co-error function is commonly used symbol for probability under tail of Gaussian pdf. Now extending the model to consider a 4 PAM transmission system, the symbol error rate performance can be calculated similarly as The average energy of the constellation can be calculated as

$$\begin{aligned}
E_{\text{avg.}} &= \frac{1}{4}[(-3A)^2 + (-A)^2 + (A)^2 + (3A)^2] \\
E_{\text{avg.}} &= 5A^2 \\
A &= \sqrt{\frac{E_{\text{avg.}}}{5}}
\end{aligned} \tag{3.112}$$

It is a practice to normalize the average energy of the transmitted symbols to one (1). The probability of symbol error for the above set of conditional probability distribution can be calculated by using symmetry of error which is shaded.



$$\begin{aligned}
p_e = p(s_0)&\Big[p(y < \frac{2A}{\sqrt{5}}|s_0)\Big] \\
&+ p(s_1)\Big[p(y > \frac{2A}{\sqrt{5}}|s_1) + p(y < 0|s_1)\Big] \\
&+ p(s_2)\Big[p(y > 0|s_2) + p(y < -\frac{2A}{\sqrt{5}}|s_2)\Big] \\
&+ p(s_3)\Big[p(y > -\frac{2A}{\sqrt{5}}|s_3)\Big] \\
&= \frac{3}{2}Q\Big(\frac{A/\sqrt{5}}{\sigma_z}\Big) \\
&=
\end{aligned} \tag{3.113}$$

BER performances for different modulation schemes are illustrated in the following figures. As we have discussed in the lecture , the performance of transmission system is tied with the operating SNR, similarly the BER performance has a direct connection with the system throughput.
The students are expected to be able to read these graphs, determine required SNR levels are a certain BER performance. Further detail are not given here but must be referred from the lecture.

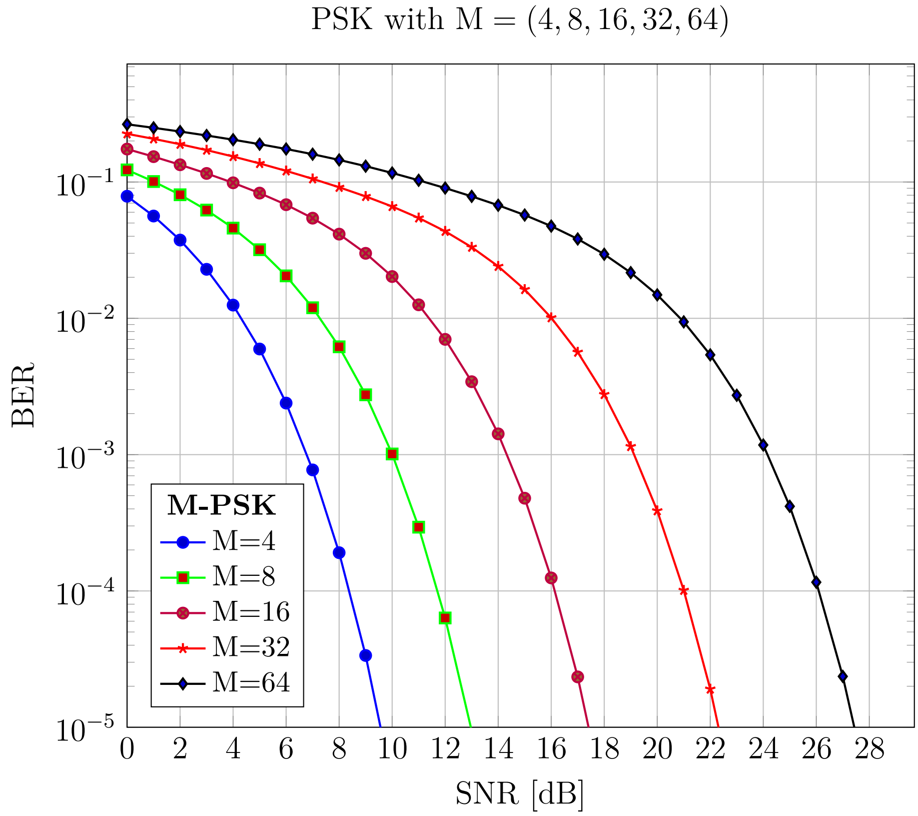


Figure 3.29: BER performance of M-PSK transmission system.

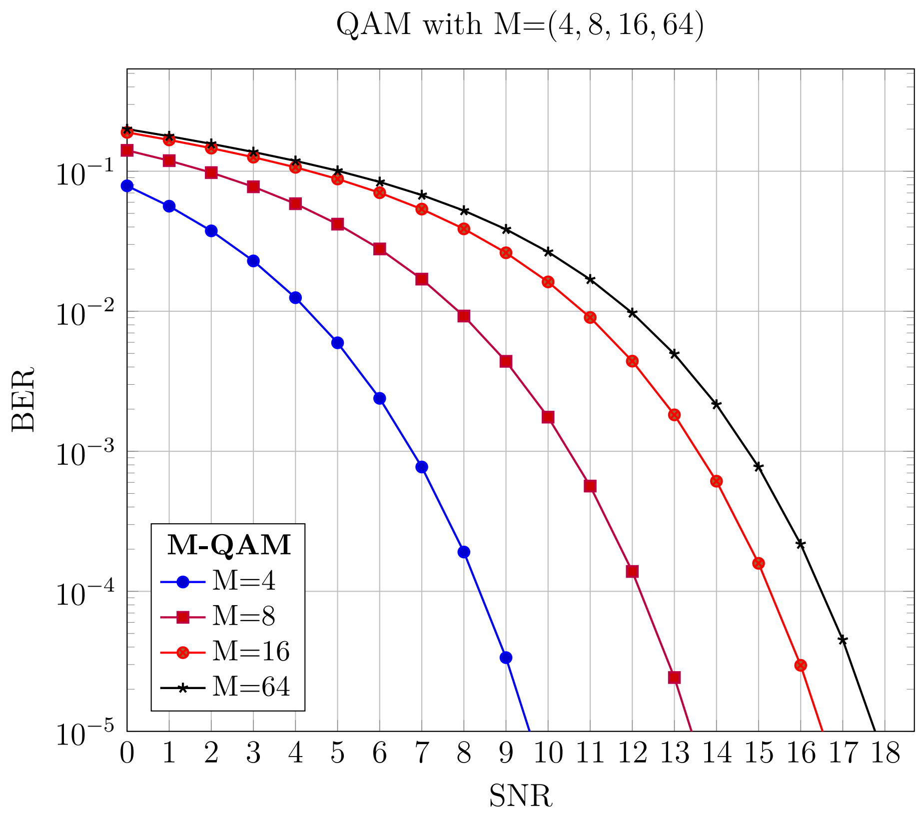


Figure 3.30: BER performance of M-QAM transmission system.

Dear Student, this handout is not a comprehensive document but it mearely serves as a guideline/reference material. You are strongly encouraged to refer these topics further in the following textbooks

1. Digital Communication by Jhon G. Proakis
2. Digital Communication by Bruce Karlson
3. Digital Communication by Grant and Glover

for further detail.

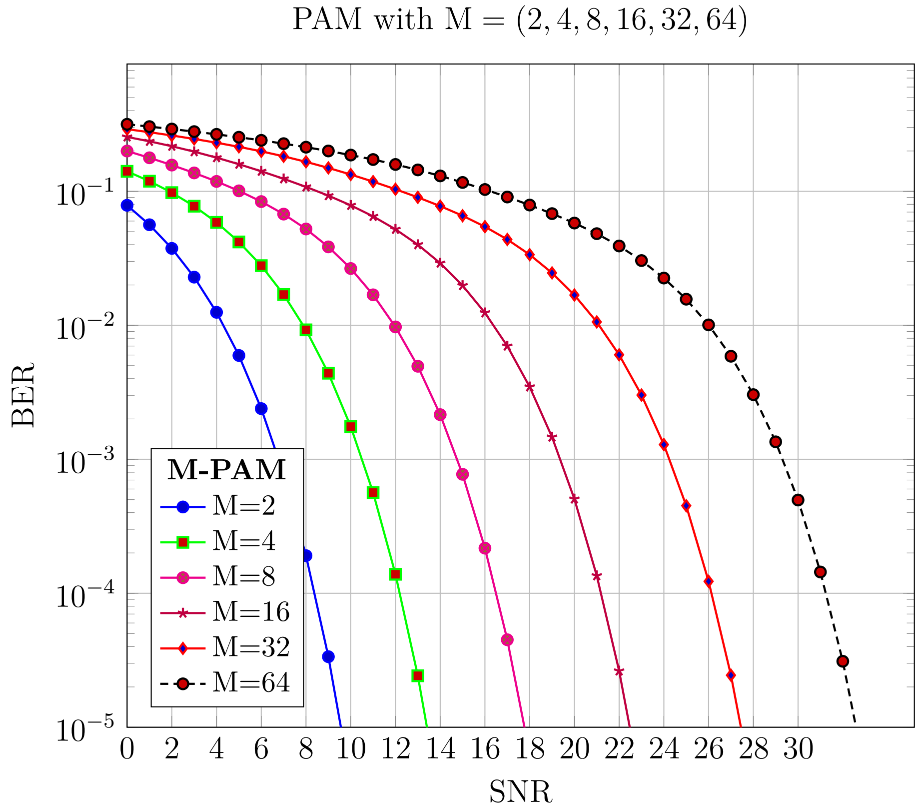


Figure 3.31: BER performance of M-PAM transmission system.

Just because I don't care doesn't mean I don't understand.