Principles of Communication

- The communication process:

  Sources of information, communication channels, modulation process, and communication networks

- Representation of signals and systems:

  Signals, Continuous Fourier transform, Sampling theorem, sequences, z-transform, convolution and correlation.

- Stochastic processes:

  Probability theory, random processes, power spectral density, Gaussian process.

- Modulation and encoding:

Basic modulation techniques and binary data transmission:AM, FM, Pulse Modulation, PCM, DPCM, Delta Modulation

- Information theory:

Information, entropy, source coding theorem, mutual information, channel coding theorem, channel capacity, rate-distortion theory.

- Error control coding:

linear bloc codes, cyclic codes, convolution codes

Course Material

1. Text: Simon Haykin, Communication systems, 4th edition, John Wiley & Sons, Inc (2001)
2. References
   (a) B.P. Lathi, Modern Digital and Analog Communcations Systems, Oxford University Press (1998)
   (b) Alan V. Oppenheim and Ronald W. Schafer, Discrete-Time signal processing, Prentice-Hall of India (1989)
   (c) Andrew Tanenbaum, Computer Networks, 3rd edition, Prentice Hall(1998).
   (d) Simon Haykin, "Digital Communication Systems," John Wiley & Sons, Inc.

Course Schedule

Duration: 14 Weeks

- Week 1:* Source of information; communication channels, modulation process and Communication Networks
- Week 2-3:* Signals, Continuous Fourier transform, Sampling theorem
- Week 4-5:* sequences, z-transform, convolution, correlation
- Week 6:* Probability theory - basics of probability theory, random processes
- Week 7:* Power spectral density, Gaussian process
- Week 8:* Modulation: amplitude, phase and frequency
- Week 9:* Encoding of binary data, NRZ, NRZI, Manchester, 4B/5B

- Week 10:* Characteristics of a link, half-duplex, full-duplex, Time division multiplexing, frequency division multiplexing
- Week 11:* Information, entropy, source coding theorem, mutual information
- Week 12:* channel coding theorem, channel capacity, rate-distortion theory
- Week 13:* Coding: linear block codes, cyclic codes, convolution codes
- Week 14:* Revision

Overview of the Course

Target Audience: Computer Science Undergraduates who have not taken any course on Communication

- Communication between a source and a destination requires a channel.
- A signal (voice/video/facsimile) is transmitted on a channel: Basics of Signals and Systems
  - This requires a basic understanding of signals
    - Representation of signals
  - Each signal transmitted is characterised by power.
  - The power required by a signal is best understood by frequency characteristics or bandwidth of the signal:
    - Representation of the signal in the frequency domain - Continuous Fourier transform

- A signal trasmitted can be either analog or digital
  - A signal is converted to a digital signal by first discretising the signal - Sampling theorem - Discrete-time Fourier transform
  - Frequency domain interpretation of the signal is easier in terms of the Z-transform
  - Signals are modified by Communication media, the communication media are characterised as Systems
  - The output to input relationship is characterised by a Transfer Function

- Signal in communcation are characterised by Random variables
  - Basics of Probability
  - Random Variables and Random Processes
  - Expectation, Autocorrelation, Autocovariance, Power Spectral Density

- Analog Modulation Schemes
  - AM, DSB-SC, SSB-SC, VSB-SC, SSB+C, VSB+C
  - Frequency Division Muliplexing
  - Power required in each of the above
- Digital Modulation Schemes
  - PAM, PPM, PDM (just mention last two)
  - Quantisation
  - PCM, DPCM, DM
  - Encoding of bits: NRZ, NRZI, Manchester
  - Power required for each of the encoding schemes
- Information Theory
  - Uncertainty, Entropy, Information
  - Mutual information, Differential entropy
  - Shannon's source and channel coding theorems

- Shannon's information capacity theorem - Analysis of Gaussian channels

- Coding
  - Repetition code
  - Hamming codes
  - Error detection codes: CRC

Analogy between Signal Spaces and Vector Spaces

Consider two vectors V_1 and V_2 as shown in Fig. 1. If V_1 is to be represented in terms of V_2

$$V_1 = C_{12}V_2 + V_e \tag{1}$$

where V_e is the error.

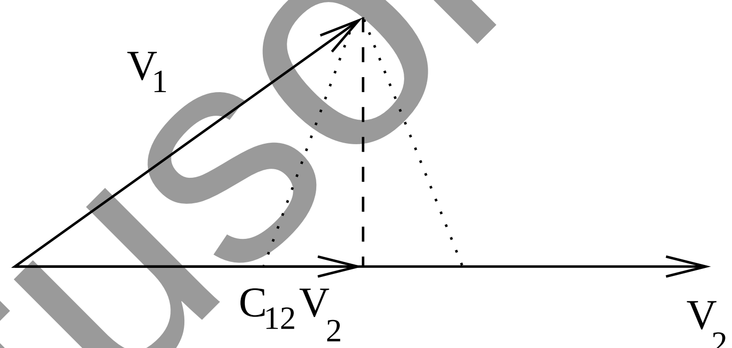


Figure 1: Representation in vector space

The error is minimum when V_1 is projected perpendicularly onto V_2. In this case, C_{12} is computed using dot product between V_1 and V_2.

Component of V_1 along V_2 is

$$= \frac{V_1.V_2}{\|V_2\|} \tag{2}$$

Similarly, component of V_2 along V_1 is

$$= \frac{V_1.V_2}{\|V_1\|} \tag{3}$$

Using the above discussion, analogy can be drawn to signal spaces also.

Let $f_1(t)$ and $f_2(t)$ be two real signals. Approximation of $f_1(t)$ by $f_2(t)$ over a time interval $t_1 < t < t_2$ can be given by

$$f_e(t) = f_1(t) - C_{12} f_2(t) \tag{4}$$

where $f_e(t)$ is the error function.

The goal is to find C_{12} such that $f_e(t)$ is minimum over the interval considered. The energy of the error signal ε given by

$$\varepsilon = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [f_1(t) - C_{12} f_2(t)]^2 \, dt \tag{5}$$

To find C_{12},

$$\frac{\partial \varepsilon}{\partial C_{12}} = 0 \tag{6}$$

Solving the above equation we get

$$C_{12} = \frac{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f_1(t).f_2(t)\, dt}{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f_2^2(t)\, dt} \tag{7}$$

The denominator is the energy of the signal $f_2(t)$.

When $f_1(t)$ and $f_2(t)$ are orthogonal to each other $C_{12} = 0$.

Example: $\sin n\omega_0 t$ and $\sin m\omega_0 t$ be two signals where m and n are integers. When $m \neq n$

$$\int_{\frac{-\pi}{\omega_0}}^{\frac{\pi}{\omega_0}} \sin n\omega_0 t . \sin m\omega_0 t \, dt = 0 \qquad (8)$$

Clearly $\sin n\omega_0 t$ and $\sin m\omega_0 t$ are orthogonal to each other.

Representation of Signals by a set of Mutually Orthogonal Real Functions

Let $g_1(t), g_2(t), ..., g_n(t)$ be n real functions that are orthogonal to each other over an interval t_1, t_2:

$$\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} g_i(t)g_j(t)dt = 0, \quad i \neq j \tag{1}$$

Let

$$\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} g_j(t)g_j(t)dt = K_j \tag{2}$$

$$f(t) = C_1 g_1(t) + C_2 g_2(t) + ... + C_n g_n(t) \tag{3}$$

$$f(t) = \sum_{r=1}^{n} C_r g_r(t) \tag{4}$$

$$f_e(t) = f_(t) - \sum_{r=1}^{n} C_r g_r(t) \tag{5}$$

$$\varepsilon = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [f(t) - \sum_{r=1}^{n} C_r g_r(t)]^2 \, dt \tag{6}$$

To find C_r,

$$\frac{\partial \varepsilon}{\partial C_1} = \frac{\partial \varepsilon}{\partial C_2} = ... = \frac{\partial \varepsilon}{\partial C_r} = 0 \tag{7}$$

When ε is expanded we have

$$\varepsilon = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f^(t) - 2f(t) \sum_{i=1}^{n} C_r g_r(t) + \sum_{r=1}^{n} C_r g_r(t) \sum_{k=1}^{n} C_k g_k(t)\, dt \tag{8}$$

Now all cross terms disappear

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} C_i g_i(t) C_j g_j(t) dt = 0, \quad i \neq j \tag{9}$$

since $g_i(t)$ and $g_j(t)$ are orthogonal to each other.

Solving the above equation we get

$$C_j = \frac{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f(t).g_j(t)\, dt}{\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} g_j^2(t)\, dt} \tag{10}$$

Analogy to Vector Spaces: Projection of $f(t)$ along the signal $g_j(t) = C_j$

Representation of Signals by a set of Mutually Orthogonal Complex Functions

When the basis functions are complex. [a]

$$E_x = \int_{t_1}^{t_2} |x(t)|^2 dt \tag{11}$$

represents the energy of a signal.

Suppose $g(t)$ is represented by the complex signal $x(t)$

[a] $|u+v|^2 = (u+v)(u^* + v^*) = |u|^2 + |v|^2 + u^*v + uv^*$

$$E_e = \int_{t_1}^{t_2} |g(t) - cx(t)|^2 dt \tag{12}$$

$$= \int_{t_1}^{t_2} |g(t)|^2 dt - \left| \frac{1}{\sqrt{E_x}} \int_{t_1}^{t_2} g(t)x^*(t)dt \right|^2 + \tag{13}$$

$$\left| c\sqrt{E_x} - \frac{1}{\sqrt{E_x}} \int_{t_1}^{t_2} g(t)x^*(t)dt \right|^2 \tag{14}$$

Minimising the second term yields

$$c = \frac{1}{E_x} \int_{t_1}^{t_2} g(t)x^*(t)dt \tag{15}$$

Thus the coefficients can be determined by projection $g(t)$ along $x^*(t)$.

Fourier Representation of continuous time signals

Any periodic signal $f(t)$ can be represented with a set of complex exponentials as shown below.

$$
\begin{aligned}
f(t) &= F_0 + F_1 e^{j\omega_0 t} + F_2 e^{j2\omega_0 t} + \cdots + F_n e^{jn\omega_n t} + & (1) \\
& \cdots + F_{-1} e^{-j\omega_0 t} + F_{-2} e^{-j2\omega_0 t} + F_{-n} e^{-jn\omega_n t} + \cdots & (2)
\end{aligned}
$$

The exponential terms are orthogonal to each other because

$$\int_{-\infty}^{+\infty} (e^{jn\omega t})(e^{jm\omega t})^* \, dt = 0, \; m \neq n$$

The energy of these signals is unity since

$$\int_{-\infty}^{+\infty}(e^{jn\omega t})(e^{jm\omega t})^*\,dt = 1,\ m = n$$

Representing a signal in terms of its exponential Fourier series components is called *Fourier Analysis.*
The weights of the exponentials are calculated as

$$F_n = \frac{\displaystyle\int_{t_0}^{t_0+T} f(t).(e^{jn\omega_0 t})^*\,dt}{\displaystyle\int_{t_0}^{t_0+T}(e^{jn\omega_0}t).(e^{jn\omega_0}t)^*\,dt}$$

$$= \frac{1}{T}\int_{t_0}^{t_0+T} f(t).(e^{jn\omega_0 t})^*\,dt$$

Extending this representation to aperiodic signals:
When $T \longrightarrow \infty$ and $\omega_0 \longrightarrow 0$, the sum becomes an *integral* and ω_0 becomes continuous.
The resulting represention is termed as the Fourier Transform ($F(\omega)$) and is given by

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

The signal $f(t)$ can recovered from $F(\omega)$ as

$$f(t) = \int_{-\infty}^{+\infty} F(\omega)e^{j\omega t}\, d\omega$$ [b]

[a] Analysis equation
[b] Synthesis equation

Some Important Functions

Delta function is a very important signal in signal analysis. It is defined as

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

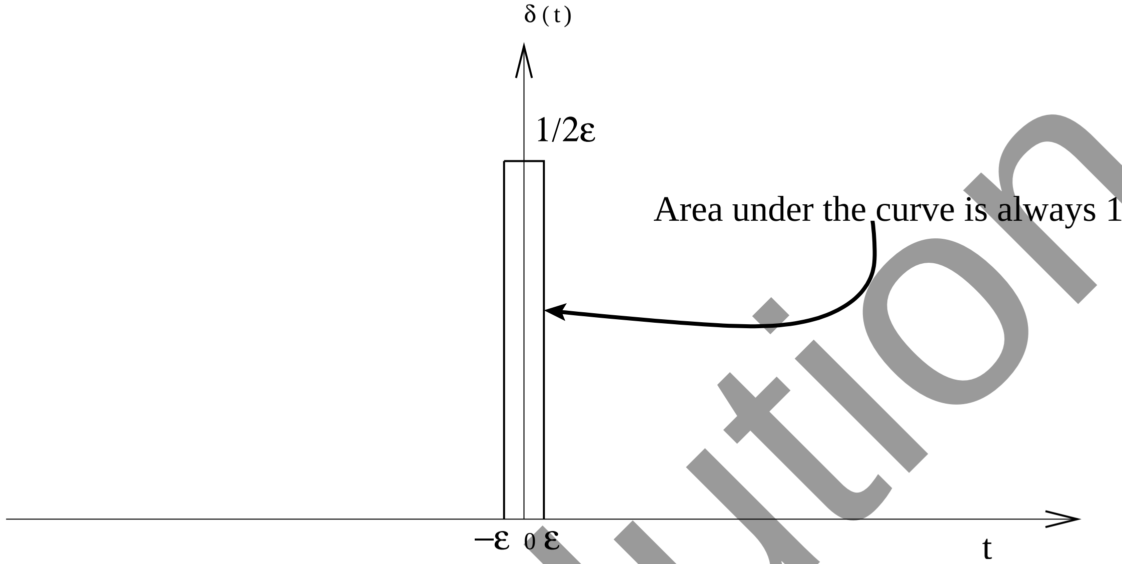


Figure 1: The Dirac delta function

The Dirac delta function is also called the Impulse function. This function can be represented as the limiting function of a number of sampling functions:

1. Gaussian Pulse

$$\delta(t) = \lim_{T \to 0} \frac{1}{T} e^{-\frac{\pi t^2}{T^2}}$$

2. Triangular Pulse

$$\delta(t) = \lim_{T\to 0} \frac{1}{T}\left[1 - \frac{|t|}{T}\right], |t| \le T \quad (3)$$

$$= 0, \; |t| > T \quad (4)$$

3. Exponential Pulse

$$\delta(t) = \lim_{T\to 0} \frac{1}{2T} e^{-\frac{|t|}{T}}$$

4. Sampling Function

$$\int_{-\infty}^{\infty} \frac{k}{\pi} Sa(kt) dt = 1$$

$$\delta(t) = \lim_{k\to\infty} \frac{k}{\pi} Sa(kt)$$

5. Sampling Square function

$$\delta(t) = \lim_{k \to \infty} \frac{k}{\pi} Sa^2(kt)$$

The unit step function is another important function signal processing. It is defined by

$$\begin{aligned} u(t) &= 1, \ t > 0 \\ &= \frac{1}{2}, \ t = 0 \\ &= 0, \ t < 0 \end{aligned}$$

The Fourier transform of the *unit step* can be found only in the limit. Some common Fourier transforms will be discussed.

Fourier Representation of continuous time signals

Properties of Fourier Transform[a]

- Translation Shifting a signal in time domain introduces linear phase in the frequency domain.

$$f(t) \longleftrightarrow F(\omega)$$

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

Proof:

[a] $\mathcal{F}$ *and* F^{-1} *correspond to the Forward and Inverse Fourier transforms*

$$F(\omega) = \int_{-\infty}^{+\infty} f(t - t_0)e^{-j\omega t}\, dt$$

Put $\tau = t - t_0$

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{+\infty} f(\tau)e^{-j\omega(\tau + t_0)}\, dt \\ &= e^{-j\omega t_0}\int_{-\infty}^{+\infty} f(\tau)e^{-j\omega\tau}\, d\tau \qquad (1) \\ &= F(\omega)e^{-j\omega t_0} \qquad (2) \end{aligned}$$

- Modulation A linear phase shift introduced in time domain signals results in a frequency domain.

$$f(t) \longleftrightarrow F(\omega)$$

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

Proof:

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{+\infty} f(t) e^{j\omega_0 t} e^{-j\omega t}\, dt \\ &= \int_{-\infty}^{+\infty} f(t) e^{-j(\omega-\omega_0)t}\, dt \quad (3) \\ &= F(\omega - \omega_0) \quad (4) \end{aligned}$$

- Scaling Compression of a signal in the time domain results in an expansion in frequency domain and vice-versa.

$$f(t) \quad \longleftrightarrow \quad F(\omega)$$
$$f(at) \quad \longleftrightarrow \quad \frac{1}{|a|}F(\frac{\omega}{a})$$

Proof:

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

Put $\tau = at$

If $a > 0$

$$\mathcal{F}(f(at)) \quad = \quad \int_{-\infty}^{+\infty} f(\tau)e^{-j\frac{\omega}{a}\tau}\,d\tau$$
$$= \quad \frac{1}{a}F(\frac{\omega}{a})$$

If $a < 0$

$$\mathcal{F}(f(at)) = -\int_{-\infty}^{+\infty} f(\tau)e^{-j\frac{\omega}{a}\tau}\,d\tau$$
$$= \frac{1}{a}F(\frac{\omega}{a})$$

$$Therefore$$

$$\mathcal{F}(f(at)) = \frac{1}{|a|}F(\frac{\omega}{a})$$

- Duality

$$f(t) \longleftrightarrow F(\omega)$$
$$F(t) \longleftrightarrow 2\pi f(-\omega)$$

Replace t with ω and ω with t in

$$
\begin{aligned}
F(\omega) &= \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}\,dt \\
F(t) &= \int_{-\infty}^{+\infty} f(\omega)e^{-jt\omega}\,d\omega
\end{aligned}
$$

But the inverse Fourier transform of a given FT $f(\omega)$ is

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

Therefore

$$
F(t) = 2\pi\mathcal{F}^{-1}(f(-\omega))
$$

or

$$F(t) \longleftrightarrow 2\pi f(-\omega)$$

Example:

$$\begin{aligned} \delta(t) &\longleftrightarrow 1 \\ 1 &\longleftrightarrow 2\pi\delta(-\omega) \\ &= 2\pi\delta(\omega)^{\text{b}} \end{aligned}$$

- Convolution Convolution of two signals in the time domain results in multiplication of their Fourier transforms.

$$f_1(t) * f_2(t) \longleftrightarrow F_1(\omega)F_2(\omega)$$

$$g(t) = f_1(t) * f_2(t) = \int -\infty^{+\infty} f_1(\tau) f_2(t-\tau) d\tau$$

Proof:

$$\begin{aligned}
\mathcal{F}(g(t)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(\tau) f_2(t-\tau) d\tau \, e^{-j\omega t} dt \\
&= \int_{-\infty}^{+\infty} f_1(t) \int_{-\infty}^{+\infty} f_2(t) e^{-j\omega t} dt d\tau \\
&= \int_{-\infty}^{+\infty} f_1(\tau) F_2(\omega) e^{-j\omega\tau} d\tau \\
&= F_1(\omega) F_2(\omega)
\end{aligned}$$

- Multiplication Multiplication of two signas in the time domain results in convolution of their Fourier transforms

$$f_1(t)f_2(t) \longleftrightarrow \frac{1}{2\pi} F_1(\omega) * F_2(\omega)$$

This can be easily proved using the **Duality Property**

- Differentiation in time

$$\frac{d}{dt} f(t) \longleftrightarrow j\omega F(\omega)$$

Proof:

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

Differentiating both sides w.r.t t yields the result.

- Differentiation in Frequency

$$(-jt)^n f(t) \longleftrightarrow \frac{d^n F(\omega)}{d\omega}$$

This follows from the duality property.

- Integration in time

$$\int_{-\infty}^{t} f(t)dt \longleftrightarrow \frac{1}{j\omega}F(\omega)$$

Some Example Continuous Fourier transforms

- $\mathcal{F}(\delta(t))$

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

Given that

$$\begin{aligned} \int_{-\infty}^{+\infty} f(t)\delta(t)\, dt &= f(0) \int_{-\infty}^{+\infty} \delta(t)\, dt \\ &= f(0) \\ \int_{-\infty}^{+\infty} f(t)\delta(t - t_0)\, dt &= f(t_0) \end{aligned}$$

Therefore $\mathcal{F}(\delta(t)) = 1$

- Linearity of the Fourier transform

$$\mathcal{F}(a_1 f_1(t) + a_2 f_2(t)) = a_1 F_1(\omega) + a_2 F_2(\omega)$$

- $\mathcal{F}(A)$, where A is constant Using the duality property and the linearity property of the Fourier transform

$$\mathcal{F}(A) = 2\pi A \delta(\omega)$$

- Fourier transform of $e^{-at}, t > 0$ (see Figure 1)

$$f(t) = e^{-at},\ t > 0$$

$$F(\omega) = = \int_0^{\infty} e^{-(a+j\omega)t} dt$$

$$= \frac{1}{a + j\omega}$$

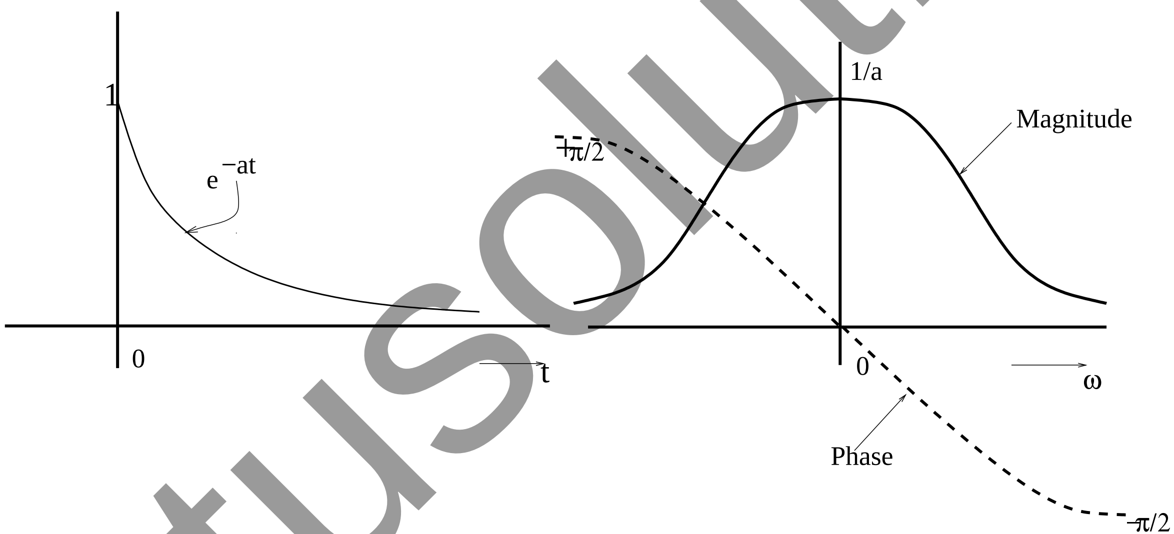


Figure 1: The exponential function and its Fourier transform

- Fourier transform of the unit step function

  The Fourier transform of the unit step function can be obtained only in the limit

$$\begin{aligned}\mathcal{F}(u(t)) &= \lim_{a \to 0} \mathcal{F}(e^{-at}) \\ &= \frac{1}{j\omega}\end{aligned}$$

- Fourier transform of $e^{-a|t|}$(see Figure 2)

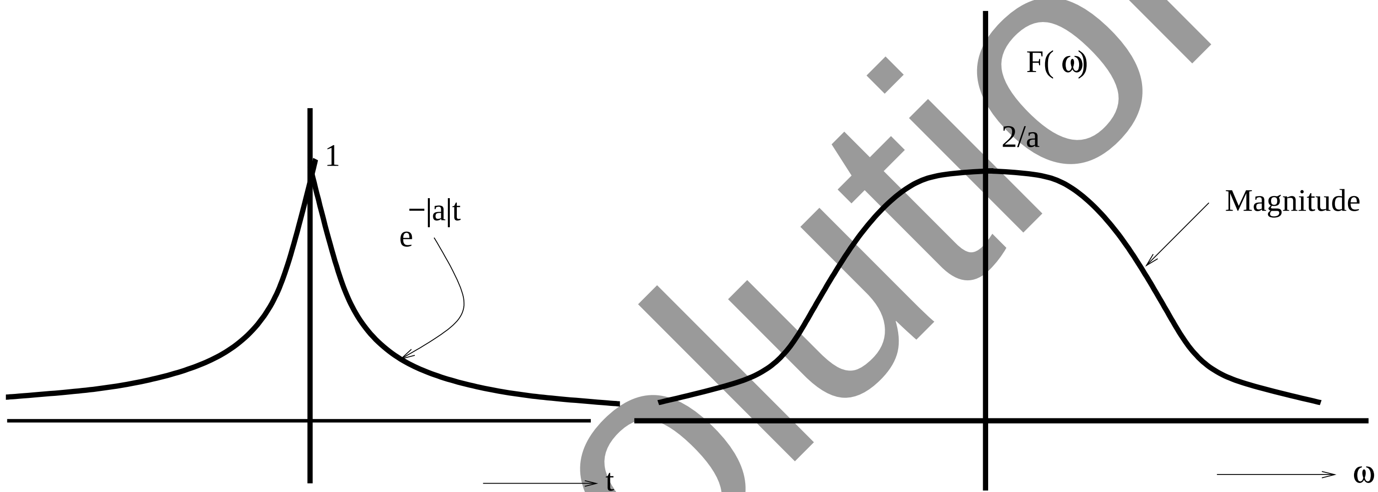


Figure 2: $e^{-|a|t}$ and its Fourier transform

$$
\begin{aligned}
f(t) &= e^{-at},\ t > 0 \\
f(t) &= 1,\ t = 0 \\
&= e^{at},\ t < 0 \\
F(\omega) &= = \int_0^{\infty} e^{-(a+j\omega)t} dt \int_0^{-\infty} e^{-(a-j\omega)t} dt \\
&= \frac{1}{a+j\omega} + \frac{1}{a-j\omega}
\end{aligned}
$$

- Fourier transform of the rectangular function

$$
\begin{aligned}
f(t) &= A, -T/2 \leq t \leq T/2 \\
&= 0,\ otherwise
\end{aligned}
$$

$$
\begin{aligned}
F(\omega) &= \int_{-\frac{T}{2}}^{+\frac{T}{2}} Ae^{-j\omega t}\,dt \\
&= A\frac{e^{-j\omega T/2} - e^{+j\omega T/2}}{-j\omega} \\
&= AT\frac{\sin\frac{\omega T}{2}}{\frac{\omega T}{2}} \\
&= sinc(\frac{\omega T}{2})
\end{aligned}
$$

The rectangular function $rect(t)$ and its Fourier transform $F(\omega)$ are shown in Figure 3

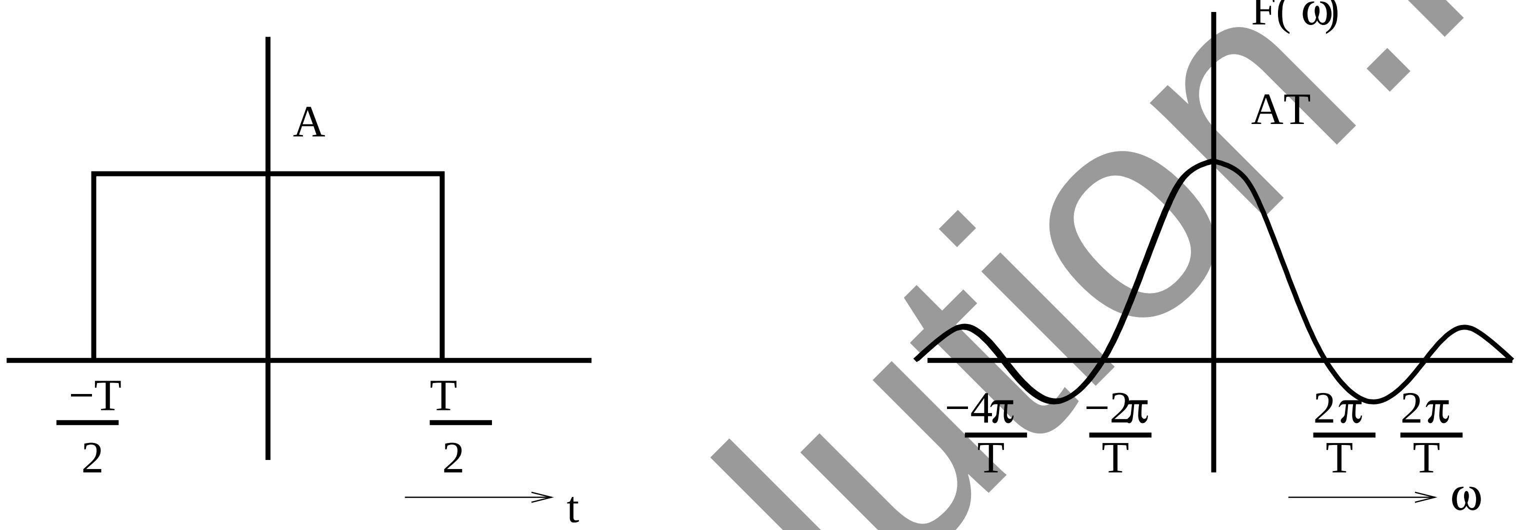


Figure 3: rect(t) and its Fourier transform

- Fourier transform of the *sinc* function
  - Using the duality property, the Fourier transform of the *sinc* function can be determined (see Figure 4).

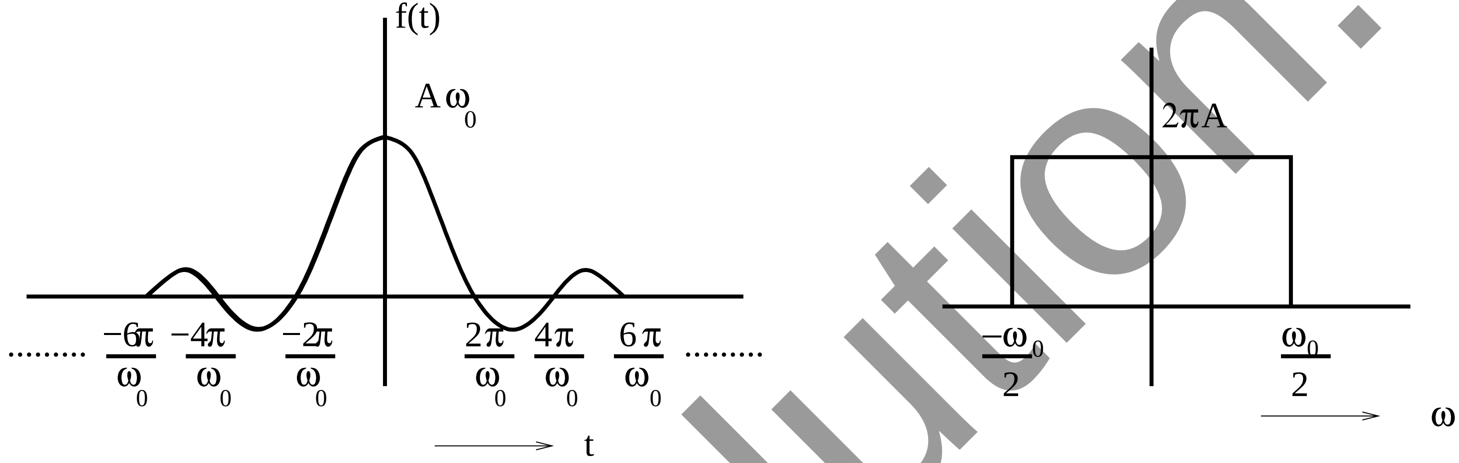


Figure 4: sinc(t) and its Fourier transform

– An important point is that a signal that is "bandlimited" is not "time-limited" while a signal that is "time-limited" is not "bandlimited"

Continuous Fourier transforms of Periodic Functions

- Fourier transform of $e^{jn\omega_0 t}$ Using the frequency shifting property of the Fourier transform

$$\begin{aligned} e^{jn\omega_0 t} &= 1.e^{jn\omega_0 t} \\ \mathcal{F}(e^{jn\omega_0 t}) &= \mathcal{F}(1)\ shifted\ by\ \omega_0 \\ &= 2\pi\delta(\omega - n\omega_0) \end{aligned}$$

- Fourier transform of $\cos\omega_0 t$

$$
\begin{aligned}
\cos \omega_0 t &= \frac{e^{j\omega_0 t} + e^{-j\omega_0 t}}{2} \\
\mathcal{F}(e^{j\omega_0 t}) &= \mathcal{F}(1)\ shifted\ by\ \omega_0 \\
&= 2\pi\delta(\omega - \omega_0) \\
\mathcal{F}(\cos \omega_0 t) &= \pi\delta(\omega - \omega_0) + \pi\delta(\omega + \omega_0)
\end{aligned}
$$

- Fourier transform of a periodic function $f(t)$
  - The periodic function is not absolutely summable.
  - The Fourier transform can be represented by a Fourier series.
  - The Fourier transform of the Fourier series representation of the periodic function (period T) can be computed

$$
\begin{aligned}
f(t) &= \sum_{n=-\infty}^{\infty} F_n e^{jn\omega_0 t}, \omega_0 = \frac{2\pi}{T} \\
\mathcal{F}(f(t)) &= \mathcal{F}(\sum_{n=-\infty}^{\infty} F_n e^{jn\omega_0 t}) \\
&= \sum_{i=-\infty}^{\infty} F_n \mathcal{F}(e^{jn\omega_0 t}) \\
&= 2\pi \sum_{n=-\infty}^{\infty} F_n \delta(\omega - n\omega_0)
\end{aligned}
$$

Note: The Fourier transform is made up of components at discrete frequencies.

- Fourier transform of a periodic function $f(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$ (a periodic train of impulses)

$$f(t) = \sum_{n=-\infty}^{\infty} F_n e^{jn\omega_0 t}, \omega_0 = \frac{2\pi}{T}$$

$$F_n = \frac{1}{T}$$

$$\begin{aligned}
\mathcal{F}(f(t)) &= \frac{1}{T}\mathcal{F}(\sum_{n=-\infty}^{\infty} e^{jn\omega_0 t}) \\
&= \sum_{i=-\infty}^{\infty} F_n \mathcal{F}(e^{jn\omega_0 t}) \\
&= 2\pi\frac{1}{T}\sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) \\
&= \omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0)
\end{aligned}$$

Note: A periodic train of impulses results in a Fourier transform which is also a periodic train of impulses (see Figure 1) .

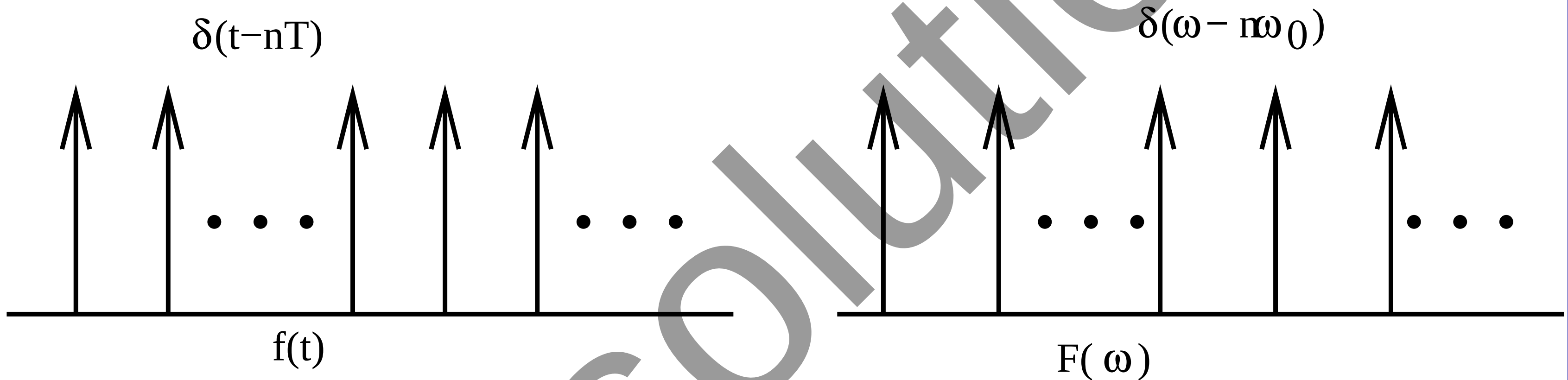


Figure 1: The periodic pulse train and its Fourier transform

Sampling Theorem and its Importance

- Sampling Theorem:
  "A bandlimited signal can be reconstructed exactly if it is sampled at a rate atleast twice the maximum frequency component in it."
  Figure 1 shows a signal $g(t)$ that is bandlimited.

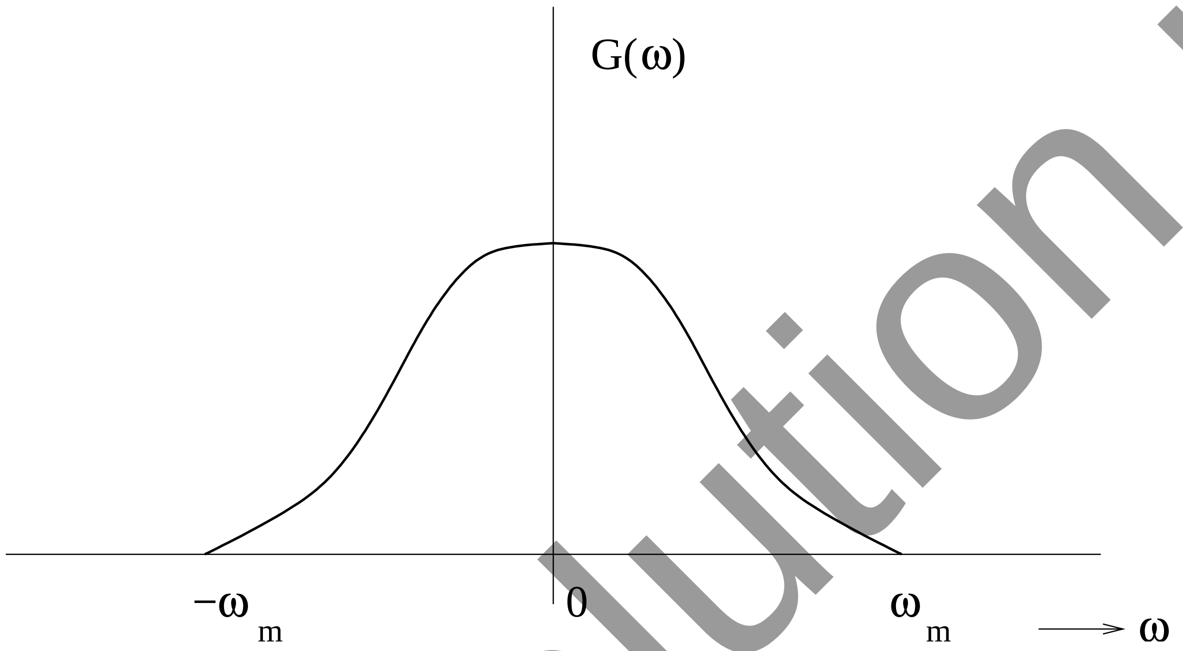


Figure 1: Spectrum of bandlimited signal $g(t)$

- The maximum frequency component of $g(t)$ is f_m. To recover the signal $g(t)$ exactly from its samples it has to be sampled at a rate $f_s \geq 2f_m$.
- The minimum required sampling rate $f_s = 2f_m$ is called

Nyquist rate.

Proof: Let $g(t)$ be a bandlimited signal whose bandwidth is f_m ($\omega_m = 2\pi f_m$).

Figure 2: (a) Original signal $g(t)$ (b) Spectrum $G(\omega)$

$\delta_T(t)$ is the sampling signal with $f_s = 1/T > 2f_m$.

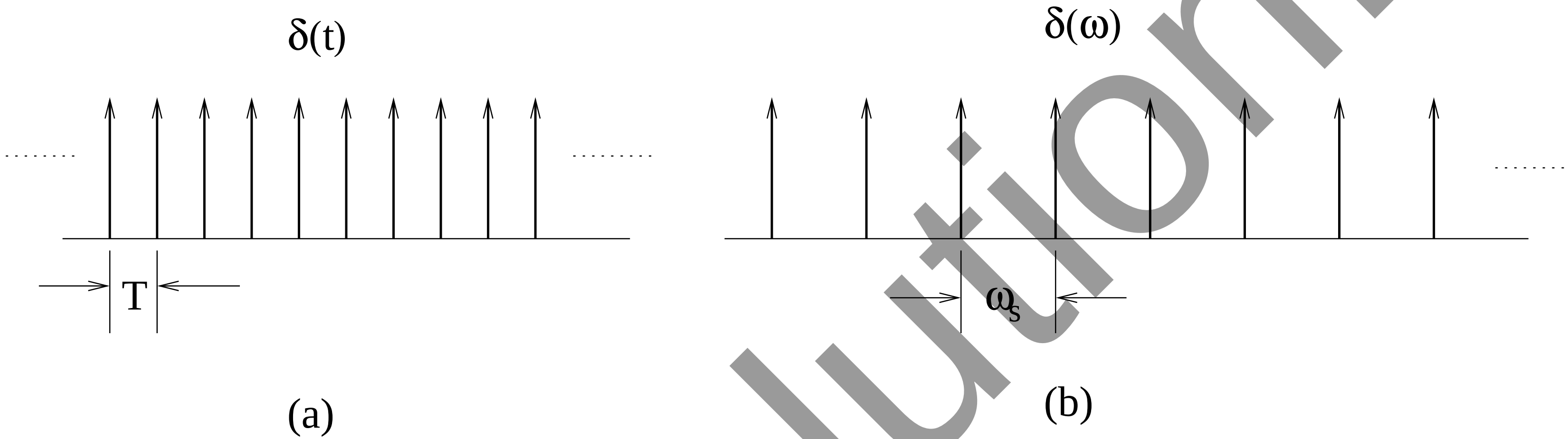


Figure 3: (a) sampling signal $\delta_T(t)$ (b) Spectrum $\delta_T(\omega)$

- Let $g_s(t)$ be the sampled signal. Its Fourier Transform $G_s(\omega)$ is given by

$$\begin{aligned}
\mathcal{F}(g_s(t)) &= \mathcal{F}\left[g(t)\delta_T(t)\right] \\
&= \mathcal{F}\left[g(t)\sum_{n=-\infty}^{+\infty}\delta(t-nT)\right] \\
&= \frac{1}{2\pi}\left[G(\omega) * \omega_0 \sum_{n=-\infty}^{+\infty}\delta(\omega-n\omega_0)\right] \\
G_s(\omega) &= \frac{1}{T}\sum_{n=-\infty}^{+\infty} G(\omega) * \delta(\omega-n\omega_0) \\
G_s(\omega) &= \mathcal{F}\left[g(t) + 2g(t)\cos(\omega_0 t) + 2g(t)\cos(2\omega_0 t) + \cdots\right] \\
G_s(\omega) &= \frac{1}{T}\sum_{n=-\infty}^{+\infty} G(\omega-n\omega_0)
\end{aligned}$$

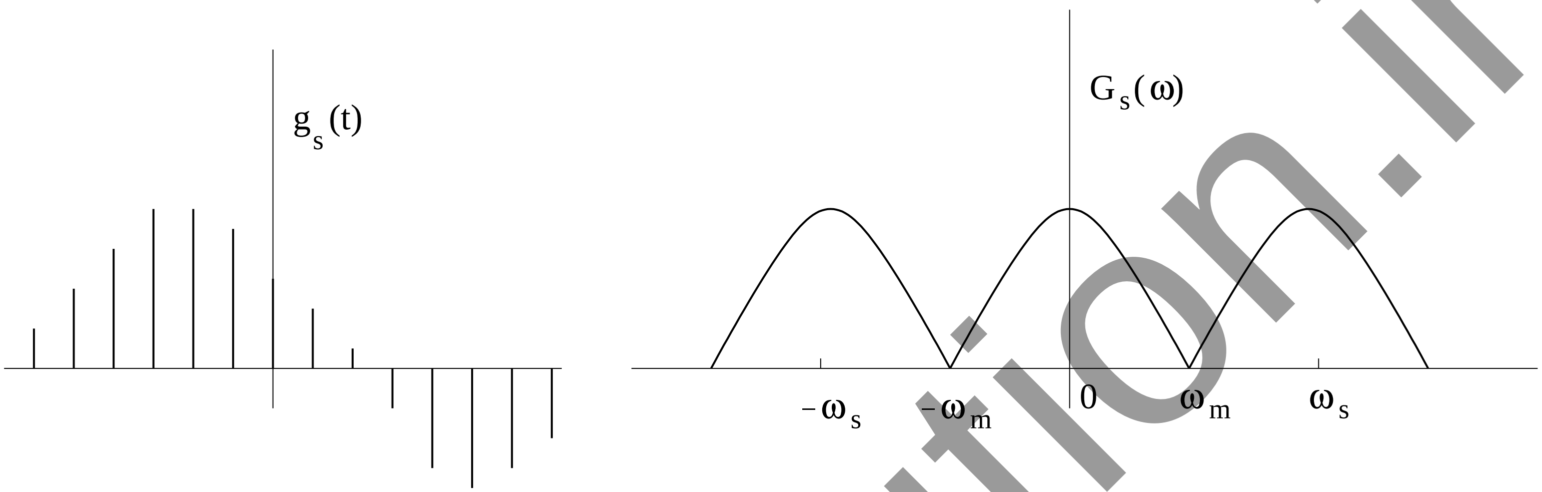


Figure 4: (a) sampled signal $g_s(t)$ (b) Spectrum $G_s(\omega)$

- If $\omega_s = 2\omega_m$, i.e., $T = 1/2f_m$. Therefore, $G_s(\omega)$ is given by

$$G_s(\omega) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} G(\omega - n\omega_m)$$

- To recover the original signal $G(\omega)$:

1. Filter with a Gate function, $H_{2\omega_m}(\omega)$ of width $2\omega_m$.

2. Scale it by T.

$$G(\omega) = TG_s(\omega)H_{2\omega_m}(\omega).$$

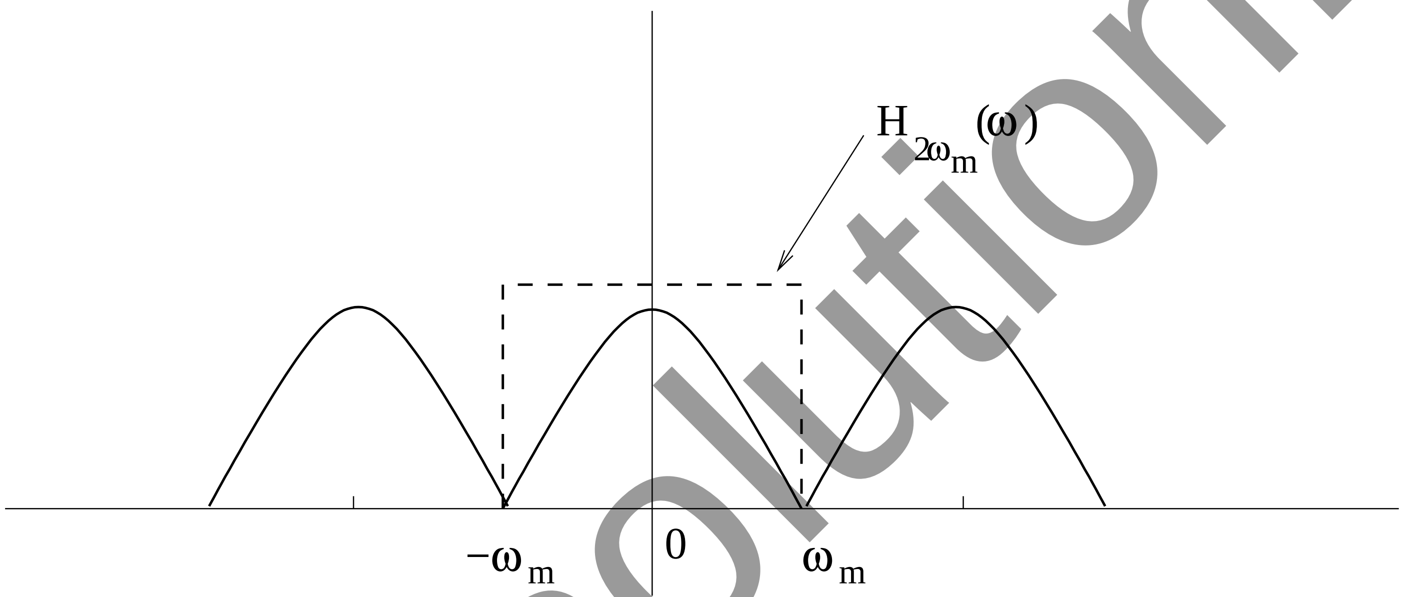


Figure 5: Recovery of signal by filtering with a filter of width $2\omega_m$

- Aliasing
  - Aliasing is a phenomenon where the high frequency components of the sampled signal interfere with each other

because of inadequate sampling $\omega_s < 2\omega_m$.

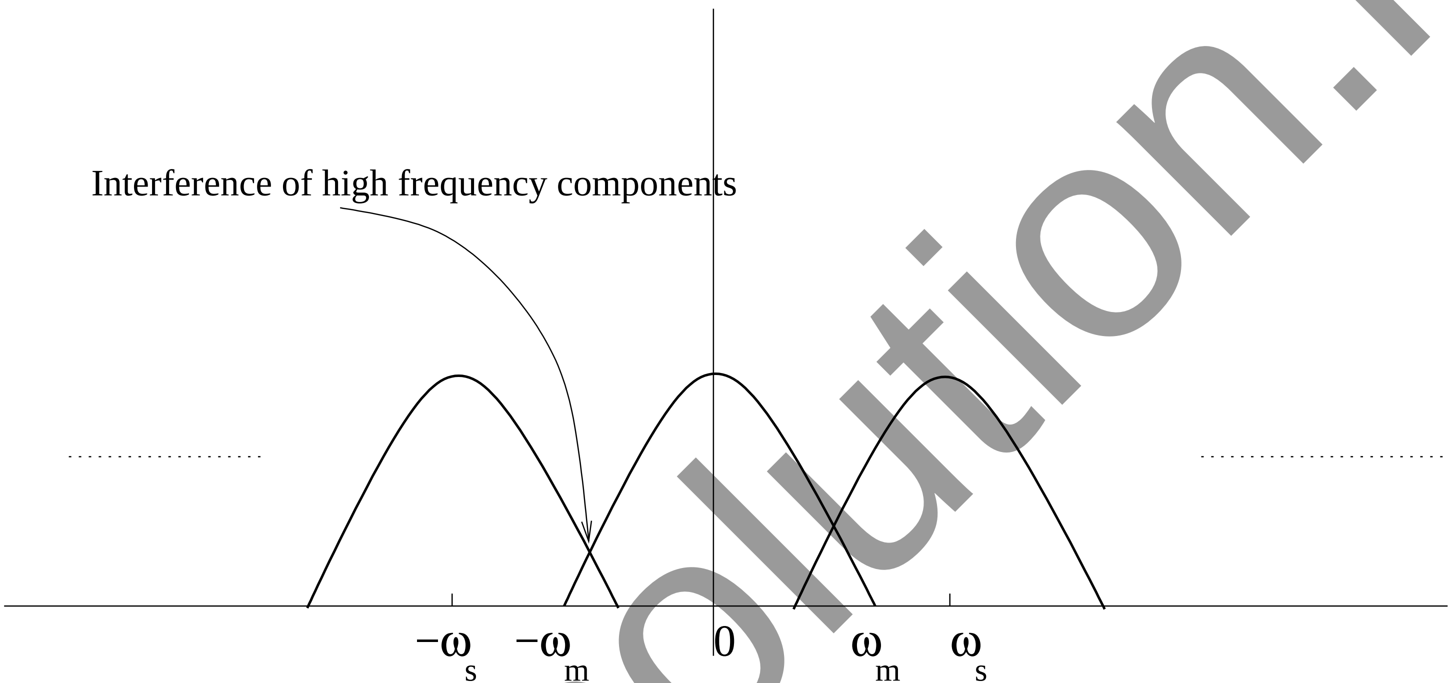


Figure 6: Aliasing due to inadequate sampling

Aliasing leads to distortion in recovered signal. This is the reason why sampling frequency should be atleast twice the bandwidth of the signal.

- Oversampling
  - In practice signal are oversampled, where f_s is significantly higher than Nyquist rate to avoid aliasing.

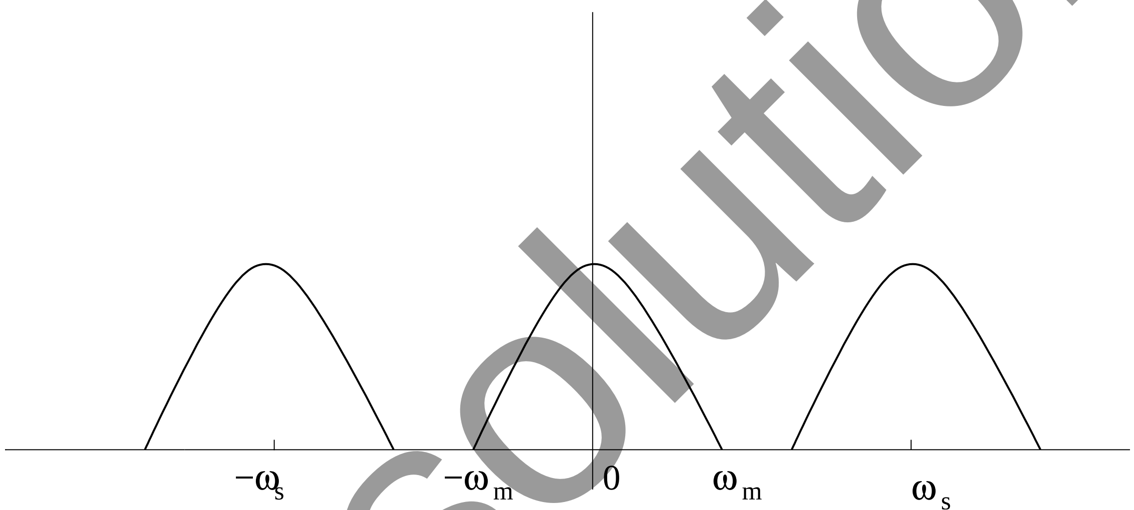


Figure 7: Oversampled signal-avoids aliasing

Problem: Define the frequency domain equivalent of the *Sampling Theorem* and prove it.

Discrete-Time Signals and their Fourier Transforms

Generation of Discrete-time signals

- Discrete time signals are obtained by sampling a continuous time signal.
- The continuous time signal is sampled with an impulse train with sampling period T
- which is usually taken greaterthan or equal to Nyquist Rate to avoid Aliasing.

The Discrete-Time Fourier Transform (DTFT) of a discrete time signal $g(nT)$is represented by[a]

$$G(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} g(nT).e^{-j\omega nT}$$

[a] $G(e^{j\omega})$ represents the DTFT. It signifies the periodicity of the DTFT

In practice, it is assumed that signals are adequately sampled and hence T is dropped to yield:

$$G(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} g(n).e^{-j\omega n}$$

The inverse DTFT is given by:

$$g(n) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} G(\omega).e^{j\omega n} d\omega$$

Some important Discrete-Time Signals

- Discrete time impulse or unit sample function Unit sample function is similar to impulse function in continuous time (see Figure 1. It is defined as follows

$$\delta(n) = \begin{cases} 1, & \text{for } n = 0 \\ 0, & \text{elsewhere} \end{cases}$$

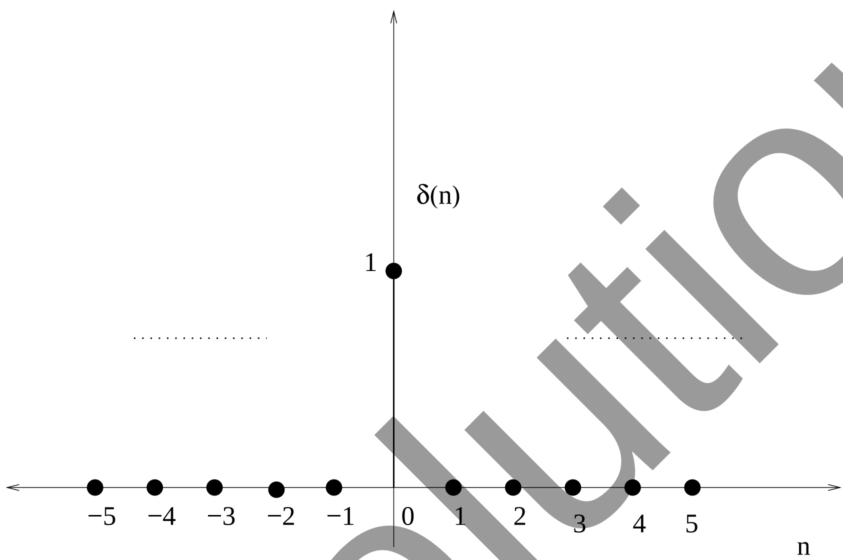


Figure 1: The unit sample function

- Unit step function This is similar to unit step function in continuous time domain (see Figure 2) and is defined as follows

$$u(n) = \begin{cases} 1, & \text{for } n \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

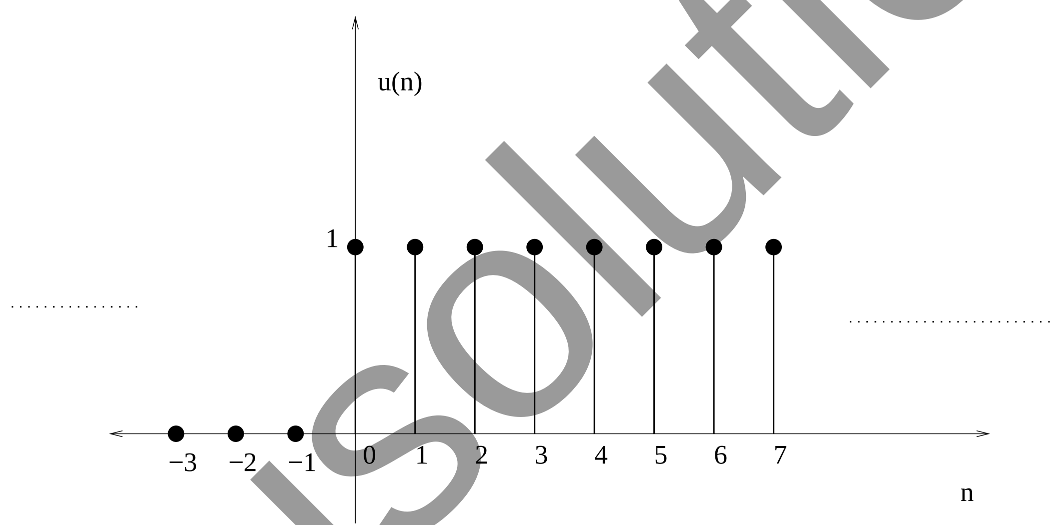


Figure 2: The unit step function

Properties of the Discrete-time Fourier transform

- Time shift property [a]

$$\mathcal{F}(f(n-n_0)) \longleftrightarrow F(e^{j\omega})e^{-j\omega n_0}$$

- Modulation property

$$\mathcal{F}(f(n)e^{-j\omega_0 n}) \longleftrightarrow F(e^{j(\omega-\omega_0)})$$

- Differentiation in the frequency domain

$$\mathcal{F}(nf(n)) \longleftrightarrow \frac{dF(e^{j\omega})}{d\omega}$$

- Convolution in the time domain

$$\mathcal{F}(f(n) * g(n)) \longleftrightarrow \frac{1}{2\pi}F(e^{j\omega})G(e^{j\omega})$$

- Prove that the forward and inverse DTFTs form a pair

[a] A tutorial on this would be appropriate

$$
\begin{aligned}
F(e^{j\omega}) &= \sum_{n=-\infty}^{+\infty} f(n).e^{-j\omega n} \\
f(n) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} F(\omega).e^{j\omega n} d\omega \\
f(n) &= \frac{1}{2\pi}\int_{-\pi}^{+\pi} \sum_{l=-\infty}^{+\infty} f(l).e^{-j\omega l} e^{j\omega n} d\omega \\
&= \sum_{l=-\infty}^{+\infty} f(l).\frac{1}{2\pi}\int_{-\pi}^{+\pi})e^{-j(\omega l-\omega n)} d\omega \\
f(n) &= \sum_{l=-\infty}^{+\infty} f(l)\delta(l-n) \\
&= f(n)
\end{aligned}
$$

Z-transforms

Computation of the Z-transform for discrete-time signals:

- Enables analysis of the signal in the frequency domain.
- Z - Transform takes the form of a polynomial.
- Enables interpretation of the signal in terms of the roots of the polynomial.
- z^{-1} corresponds to a delay of one unit in the signal.

The Z - Transform of a discrete time signal $x[n]$ is defined as

$$X(z) = \sum_{n=-\infty}^{+\infty} x[n].z^{-n} \tag{1}$$

where $z = r.e^{j\omega}$

The discrete-time Fourier Transform (DTFT) is obtained by evaluating Z-Transform at $z = e^{j\omega}$.

or

The DTFT is obtained by evaluating the Z-transform on the unit circle in the z-plane.

The Z-transform converges if the sum in equation 1 converges

Region of Convergence(RoC)

Region of Convergence for a discrete time signal $x[n]$ is defined as a continuous region in z plane where the Z-Transform converges.
In order to determine RoC, it is convenient to represent the Z-Transform as:[a]

$$X(z) = \frac{P(z)}{Q(z)}$$

- The roots of the equation $P(z) = 0$ correspond to the 'zeros' of $X(z)$
- The roots of the equation $Q(z) = 0$ correspond to the 'poles' of $X(z)$
- The RoC of the Z-transform depends on the convergence of the

[a] Here we assume that the Z-transform is rational

polynomials $P(z)$ and $Q(z)$,

- Right-handed Z-Transform
  - Let $x[n]$ be causal signal given by

$$x[n] = a^n u[n]$$

  - The Z - Transform of $x[n]$ is given by

$$
\begin{aligned}
X(z) &= \sum_{n=-\infty}^{+\infty} x[n]z^{-n} \\
&= \sum_{n=-\infty}^{+\infty} a^n u[n] z^{-n} \\
&= \sum_{n=0}^{+\infty} a^n z^{-n} \\
&= \sum_{n=0}^{+\infty} (az^{-1})^n \\
&= \frac{1}{1-az^{-1}} \\
&= \frac{z}{z-a}
\end{aligned}
$$

– The ROC is defined by $|az^{-1}| < 1$ or $|z| > |a|$.

– The RoC for $x[n]$ is the entire region outside the circle $z = ae^{j\omega}$ as shown in Figure 1.

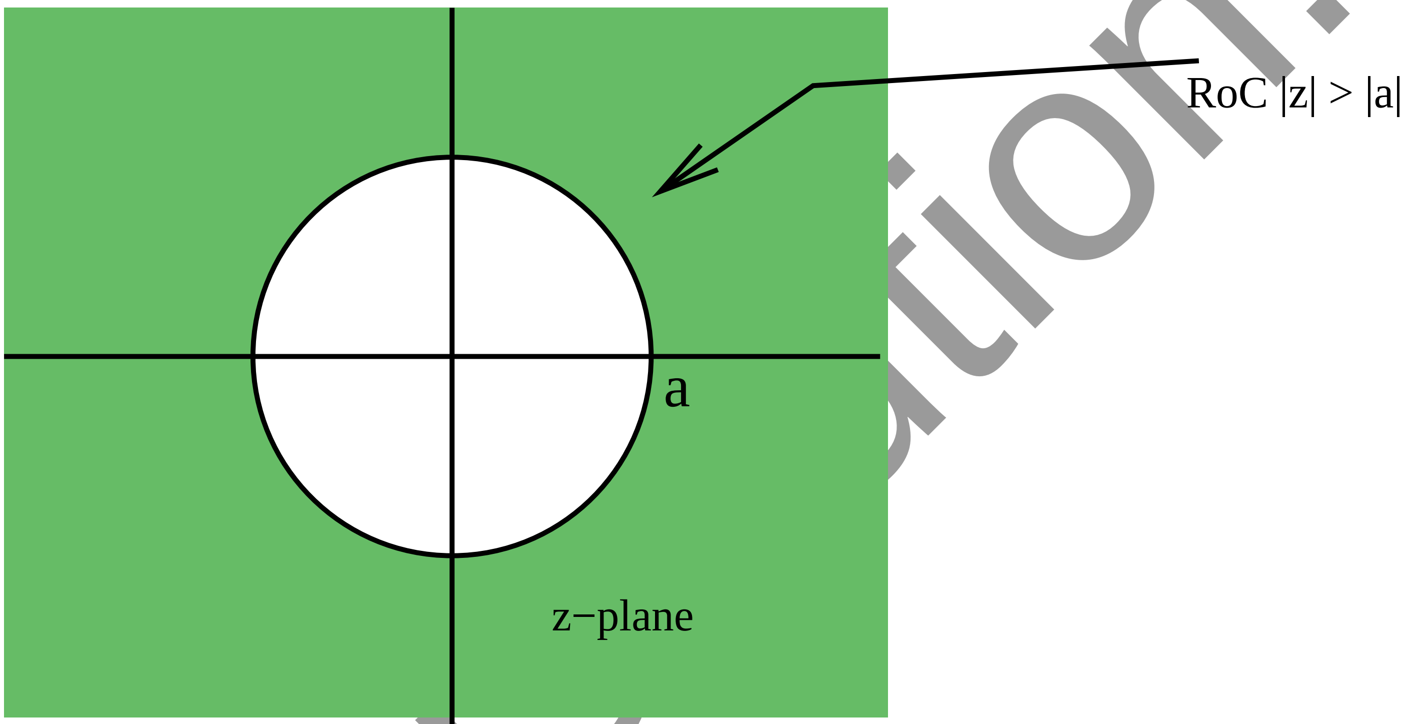


Figure 1: RoC(green region) for a causal signal

- Left-handed Z-Transform

– Let $x[n]$ be an anti-causal signal given by

$$y[n] = -b^n u[-n-1]$$

– The Z - Transform of $y[n]$ is given by

$$
\begin{aligned}
Y(z) &= \sum_{n=-\infty}^{+\infty} y[n]z^{-n} \\
&= \sum_{n=-\infty}^{+\infty} -b^n u[-n-1]z^{-n} \\
&= \sum_{n=-\infty}^{-1} -b^n z^{-n} \\
&= \sum_{n=0}^{+\infty} -(b^{-1}z)^n + 1 \\
&= \frac{1}{1-\frac{z}{b}} + 1 \\
&= \frac{z}{z-b}
\end{aligned}
$$

– $Y(z)$ converges when $|b^{-1}z < 1$ or $|z| < |b|$.
– The RoC for y[n] is the entire region inside the circle $z = be^{j\omega}$ as shown in Figure 2

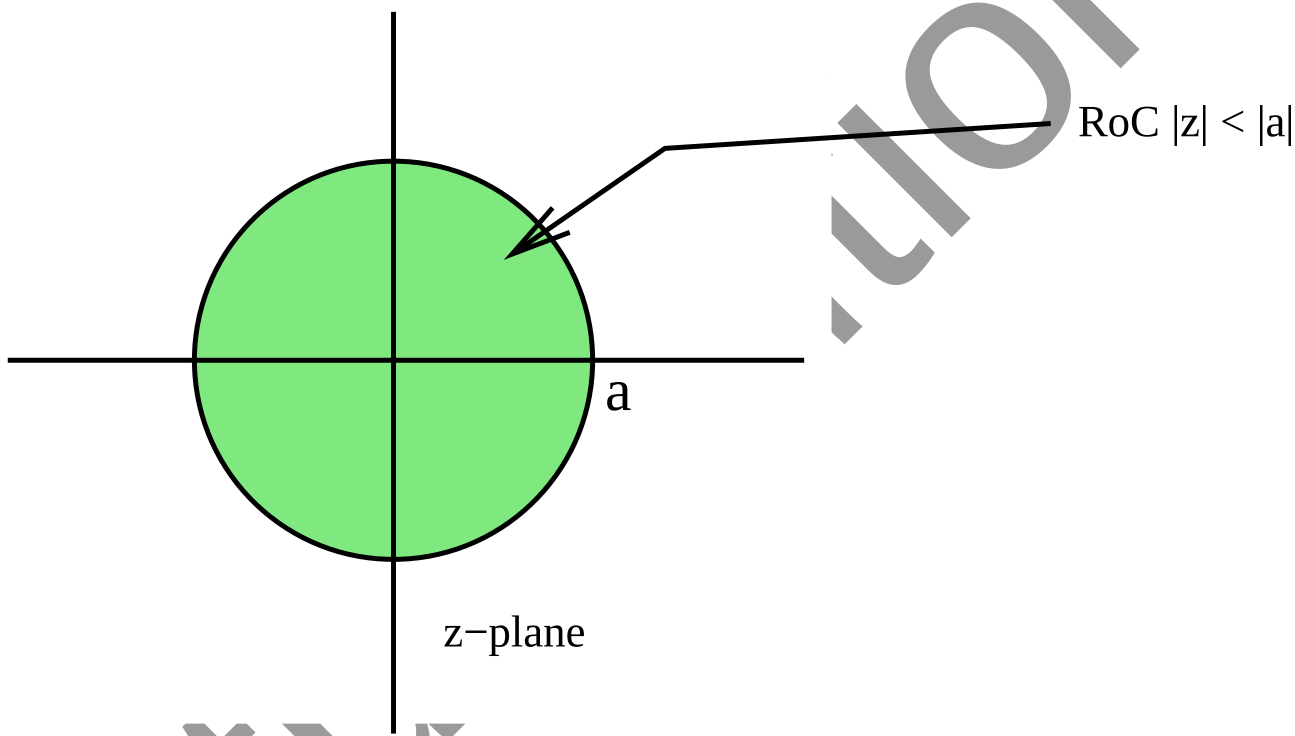


Figure 2: RoC(green region) for an anti-causal signal

- Two-sided Z-Transform
  - Let $y[n]$ be a two sided signal given by

$$y[n] = a^n u[n] - b^n u[-n-1]$$

where, $b > a$

  - The Z - Transform of $y[n]$ is given by

$$
\begin{aligned}
Y(z) &= \sum_{n=-\infty}^{+\infty} y[n] z^{-n} \\
&= \sum_{n=-\infty}^{+\infty} (a^n u[n] - b^n u[-n-1]) z^{-n} \\
&= \sum_{n=0}^{+\infty} a^n z^{-n} - \sum_{n=-\infty}^{-1} b^n z^{-n} \\
&= \sum_{n=0}^{+\infty} (az{-}1)^n - \sum_{n=1}^{+\infty} (b^{-1} z)^n \\
&= \frac{1}{1 - az^{-1}} . \frac{1}{1 - \frac{z}{b}} + 1 \\
&= \frac{z}{z-a} . \frac{z}{z-b}
\end{aligned}
$$

– $Y(z)$ converges for $|b^{-1}z| < 1$ and $|az^{-1}| < 1$ or $|z| < |b|$ and $|z| > |a|$. Hence, for the signal

– The ROC for $y[n]$ is the intersection of the circle $z = be^{j\omega}$ and the circle $z = ae^{j\omega}$ as shown in Figure 3

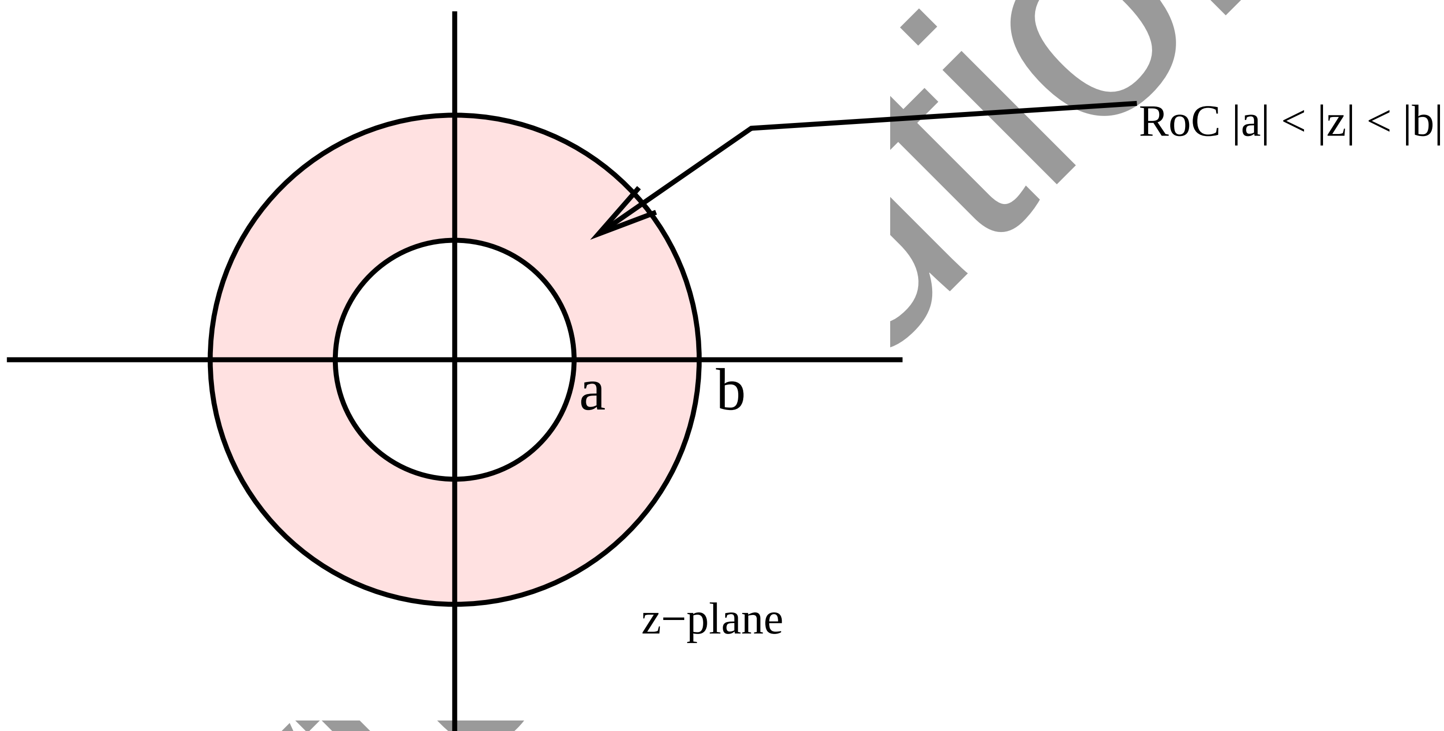


Figure 3: RoC(pink region) for a two sided Z Transform

- Transfer function $H(z)$
  - Consider the system shown in Figure 4.

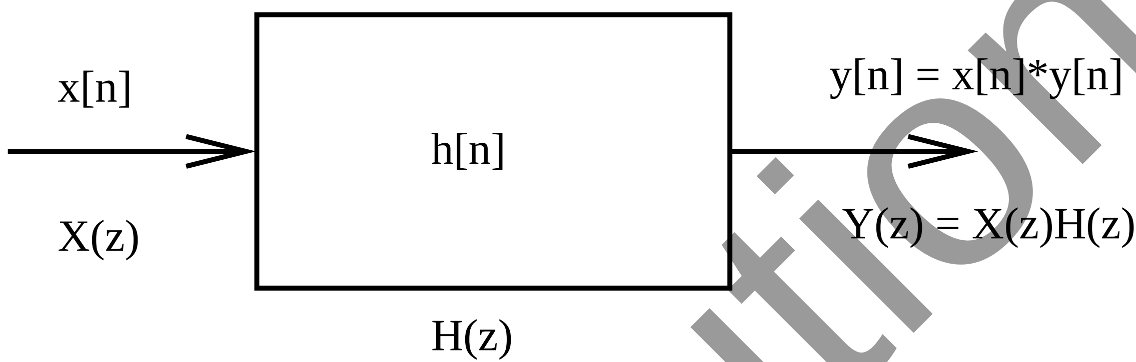


Figure 4: signal - system representation

  - $x[n]$ is the input and $y[n]$ is the output
  - h[n] is the impulse response of the system. Mathematically, this signal-system interaction can be represented as follows

$$y[n] = x[n] * h[n]$$

– In frequency domain this relation can be written as

$$Y(z) = X(z).H(z)$$

or

$$H(z) = \frac{Y(z)}{X(z)}$$

$H(z)$ is called 'Transfer function' of the given system.
In the time domain if $x[n] = \delta[n]$ then $y[n] = h[n]$,
$h[n]$ is called the 'impulse response' of the system.
Hence, we can say that

$$h[n] \longleftrightarrow H(z)$$

Some Examples: Z-transforms

- Delta function

$$\mathcal{Z}(\delta[n]) = 1$$
$$\mathcal{Z}(\delta[n - n_0]) = z^{-n_0}$$

- Unit Step function

$$x[n] = 1, n \geq 0$$
$$x[n] = 0, otherwise$$
$$X(z) = \frac{1}{1 - z^{-1}}, |z| > 1$$

The Z-transform has a real pole at the $z = 1$.

- Finite length sequence

$$
\begin{aligned}
x[n] &= 1,\, 0 \leq n \geq N \\
x[n] &= 0,\, otherwise \\
X(z) &= \frac{1 - z^{-N}}{1 - z^{-1}} \\
&= z^{N-1}\frac{z^{N} - 1}{z - 1}, |z| > 1
\end{aligned}
$$

The roots of the numerator polynomial are given by:

$$z = 0, N zeros\ at\ the\ origin$$

and the n^{th} roots of unity:

$$z = e^{\frac{j2\pi k}{N}}, k = 0, 1, 2, \cdots, N - 1 \qquad (2)$$

- Causal sequences

$$x[n] = (\frac{1}{3})^n u[n] - (\frac{1}{2})^n u[n-1]$$

$$X(z) = \frac{1}{1-\frac{1}{3}z^{-1}} - z^{-1}\frac{1}{1-\frac{1}{2}z^{-1}}, |z| > \frac{1}{3}$$

The Discrete time Fourier transform can be obtained by setting $z = e^{j\omega}$ Figure 5 shows the Discrete Fourier transform for the rectangular function.

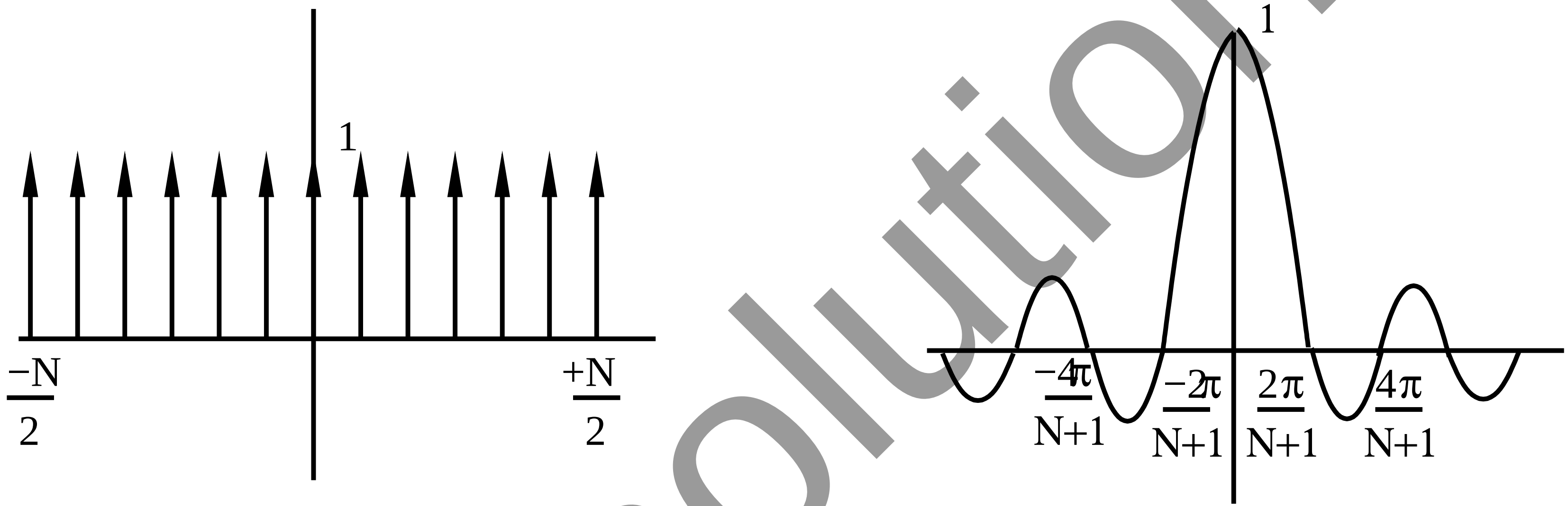


Figure 5: Discrete Fourier transform for the rectangular function

Some Problems

Find the Z-transform (assume causal sequences):

1. $1, \frac{a}{1!}, \frac{a^2}{2!}, \frac{a^3}{3!}, \cdots$
2. $0, a, 0, -\frac{a^3}{3!}, 0, \frac{a^5}{5!}, 0, -\frac{a^7}{7!}, \cdots$
3. $0, a, 0, -\frac{a^2}{2!}, 0, \frac{a^4}{4!}, 0, -\frac{a^6}{6!}, \cdots$

Hint: Observe that the series is similar to that of the exponential series.

Properties of the Z-transform

1. RoC is generally a disk on the z-plane.

$$0 \leq r_R \leq |z| \leq r_L \leq \infty$$

2. Fourier Transform of $x[n]$ converges when RoC includes the unit circle.
3. RoC does not contain any poles.
4. If $x[n]$ is finite duration, RoC contains entire z - plane except for $z = 0$ and $z = \infty$.
5. For a left handed sequence, RoC is bounded by $|z| < min(|a|, |b|)$.
6. For a right handed sequence, RoC is bounded by $|z| > max(|a|; |b|)$.

Inverse Z-transform

To determine the inverse Z-transform, it is necessary to know the RoC.

- RoC decides whether a given signal is causal (exists for positive time), anticausal (exists for negative time) or both causal and anticausal (exists forboth positive and negative time)

Different approaches to compute the inverse Z-transform

- Long division method When Z-Transform is rational, i.e. it can be expressed as the ratio of two polynomials $P(z)$ and $Q(z)$

$$X(z) = \frac{P(z)}{Q(z)}$$

Then, inverse Z-transform can be obtained using long division:

– Divide $P(z)$ by $Q(z)$. Let this be:

$$X(z) = \sum_{i=-\infty}^{\infty} a_i z^{-i} \qquad (1)$$

– The coefficients of the RHS of equation (1) correspond to the time sequencei.e. the coefficients of the quotient of the long division gives the sequence.

- Partial Fraction method
  – the Z-Transform is decomposed into partial fractions
  – the inverse Z-transform of each fraction is obtained independently
  – the inverse sequences are then added

The method of adding inverse Z-transform is illustrated below. Let,

$$X(z) = \frac{\sum_{k=0}^{M} b^k . z^{-k}}{\sum_{k=0}^{N} a^k . z^{-k}}, \qquad M < N$$

$$= \frac{\prod_{k=1}^{M} (1 - c_k . z^{-1})}{\prod_{k=1}^{N} (1 - d_k . z^{-1})}$$

$$= \sum_{k=1}^{N} \frac{A_k}{(1 - d_k . z^{-1})}$$

where,

$$A_k = (1 - d_k.z^{-1})X(z)|_{z=d_k}$$

For s multiple poles at $z = d_i$

$$X(z) = \sum_{k=0}^{M-N} B_r.z^{-1} + \sum_{k=1,k\neq i}^{N} \frac{A_k}{(1 - d_k.z^{-1})} + \sum_{m=1}^{s} \frac{C_m}{(1 - d_i.z^{-1})^m}$$

Properties of the Z-Transform

- Linearity:

$$a_1 x_1[n] + a_2 x_2[n] \longleftrightarrow a_1 X_1(z) + a_2 X_2(z), RoC = R_{x_1} \cap R_{x_2}$$

- Time Shifting Property:

$$\begin{aligned} x[n-n_0] \quad \longleftrightarrow \quad & z^{-n_0} X(z), \\ & RoC = R_x \ (except\ possible\ addition/deletion \\ & of\ \ z = 0\ or\ z = \infty) \end{aligned}$$

- Exponential Weighting:

$$z_0^n x[n] \longleftrightarrow X(z_0^{-1} z), RoC = |z_0| R_x$$

  – The poles of the Z-transform are scaled by $|z_0|$

- Linear Weighting

$$nx(n) \longleftrightarrow -z\frac{dX(z)}{dz},$$

$$RoC = R_x \ (except\ possible\ addition/deletion$$

$$of\ z = 0\ or\ z = \infty)$$

- Time Reversal

$$x[-n] \longleftrightarrow X(z^{-1}), RoC = \frac{1}{R_x}$$

- Convolution

$$x[n] * y[n] \longleftrightarrow X(z)Y(z), RoC = R_x \cap R_y$$

- Multiplication

$$x[n]w[n] \longleftrightarrow \frac{1}{2\pi j}\int X(v)w(\frac{z}{v})v^{-1}dv$$

Inverse Z-Transform Examples

- Using long division: Causal sequence

$$\frac{1}{1-az^{-1}}, RoC = |z| > |a| \quad = \quad 1 + az^{-1} + az^{-2} + az^{-3} + \cdots$$

$$IZT(1 + az^{-1} + a^2z^{-2} + a^3z^{-3} + \cdots) = a^n u[n]$$

- Using long division: Noncausal sequence

$$\frac{1}{1-az^{-1}}, RoC = |z| < |a|$$

Here the IZT is computed as follows:

$$IZT\left(\frac{1}{1-az^{-1}}\right) = IZT\left(\frac{z}{-a+z}\right)$$

This results in:

$$IZT(-a^{-1}z + a^{-2}z^2 + a^{-3}z^3 + \cdots) = -a^n u[-n-1]$$

- Inverse Z-transform - using Power series expansion

$$X(z) = log(1 + az^{-1}), |z| > |a|$$

Using the Power Series expansion for $\log(1 + x)$, $|x| < 1$, we have

$$X(z) = \sum_{1}^{\infty} \frac{(-1)^{n+1} a^n z^{-n}}{n}$$

The IZT is given by

$$\begin{aligned} x[n] &= \frac{(-1)^{n+1} a^n}{n}, n \geq 1 \\ &= 0, n \leq 0 \end{aligned}$$

- Inverse Z-transform - Inspection method

$$a^n u[n] \longleftrightarrow \frac{1}{1 - az^{-1}}, |z| > |a|$$

$$Given X(z) = \frac{1}{1 - \frac{1}{2}z^{-1}}, |z| > |\frac{1}{2}|$$

$$\Longrightarrow x[n] = (\frac{1}{2})^n u[n]$$

- Inverse Z-transform - Partial fraction method
  - Example 1: All-Pole system

$$X(z) = \frac{1}{(1 - \frac{1}{3}z^{-1})(1 - \frac{1}{6}z^{-1})}, |z| > \frac{1}{3}$$

Using partial fraction method, we have:

$$X(z) = \frac{A_1}{1 - \frac{1}{3}z^{-1}} + \frac{A_2}{1 - \frac{1}{6}z^{-1}},$$

$$|z| > \frac{1}{3}$$

$$A_1 = (1 - \frac{1}{3}z^{-1})X(z)|_{z=\frac{1}{3}}$$

$$A_2 = (1 - \frac{1}{6}z^{-1})X(z)|_{z=\frac{1}{6}}$$

$$A_1 = 2$$

$$A_2 = -1$$

$$x(n) = 2(\frac{1}{3})^n u[n] - 1(\frac{1}{6})^n u[n]$$

– Example 2: Pole-Zero system

$$\begin{aligned} X(z) &= \frac{1+2z^{-1}+z^{-2}}{1-\frac{3}{2}z^{-1}+\frac{1}{2}z^{-2}}, |z|>1 \\ &= \frac{(1+z^{-1})^2}{(1-\frac{1}{2}z^{-1})(1-z^{-1})} \\ &= 2+\frac{-1+5z^{-1}}{(1-\frac{1}{2}z^{-1})(1-z^{-1})} \\ &= 2-\frac{9}{1-\frac{1}{2}z^{-1}}+\frac{8}{1-z^{-1}} \\ x[n] &= 2\delta[n]-9(\frac{1}{2})^n u[n]+8u[n] \end{aligned}$$

– Example 3: Finite length sequences

$$X(z) = z^2(1-\frac{1}{2}z^{-1})(1+z^{-1})(1-z^{-1})$$

$$= z^2 - frac12z - 1 + \frac{1}{2}z^{-1}$$

$$= \delta[n+2] - \frac{1}{2}\delta[n+1] - \delta[n] + \frac{1}{2}\delta[n-1]$$

Inverse Z-Transform Problem

1. Given $X(z) = \frac{z}{z-1} - \frac{z}{z-2} + \frac{z}{z-3}$, determine all the possible sequences for $x[n]$.

Hint: Remember that the RoC must be a continuous region

Basics of Probability Theory and Random Processes

Basics of probability theory [a]

- Probability of an event E represented by $P(E)$ and is given by

$$P(E) = \frac{N_E}{N_S} \quad (1)$$

where, N_S is the number of times the experiment is performed and N_E is number of times the event E occured.

Equation 1 is only an approximation. For this to represent the exact probability $N_S \to \infty$.

The above estimate is therefore referred to as *Relative Probability*

Clearly, $0 \leq P(E) \leq 1$.

[a] required for understanding communication systems

- Mutually Exclusive Events

  Let $\mathcal{S}$ be the sample space having N events $E_1, E_2, E_3, \cdots, E_N$. Two events are said to be mutually exclusive or statistically independent if $A_i \cap A_j = \phi$ and $\bigcup_{i=1}^{N} A_i = \mathcal{S}$ for all i and j.

- Joint Probability

  Joint probability of two events A and B represented by $P(A \cap B)$ and is defined as the probability of the occurence of both the events A and B is given by

$$P(A \cap B) = \frac{N_{A \cap B}}{N_S}$$

- Conditional Probability

  Conditional probability of two events A and B represented as

$P(A|B)$ and defined as the probability of the occurence of event A after the occurence of B.

$$P(A|B) = \frac{N_{A \cap B}}{N_B}$$

Similarly,

$$P(B|A) = \frac{N_{A \cap B}}{N_B}$$

This implies,

$$P(B|A)P(A) = P(A|B)P(B) = P(A \cap B)$$

- Chain Rule

  Let us consider a chain of events A_1, A_2, A_3, $\cdots$, A_N which are dependent on each other. Then the probability of occurence of the sequence

$$\begin{aligned} P(A_N, A_{N-1}, A_{N-2}, \\ \cdots, A_2, A_1) &= P(A_N|A_{N-1}, A_{N-2}, \cdots, A_1). \\ & P(A_{N-1}|A_{N-2}, A_{N-3}, \cdots, A_1). \\ & \cdots .P(A_2|A_1).P(A_1) \end{aligned}$$

Bayes Rule

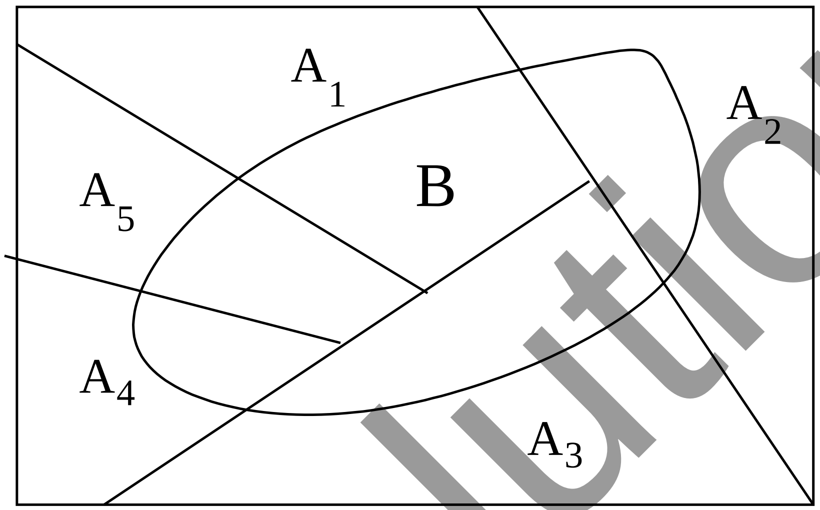


Figure 1: The partition space

In the above figure, if A_1, A_2, A_3, A_4, A_5 partition the sample space $\mathcal{S}$, then $(A_1 \cap B), (A_2 \cap B), (A_3 \cap B), (A_4 \cap B)$, and $(A_5 \cap B)$ partition B.
Therefore,

$$
\begin{aligned}
P(B) &= \sum_{i=1}^{n} P(A_i \cap B) \\
&= \sum_{i=1}^{n} P(B|A_i).P(A_i)
\end{aligned}
$$

In the example figure here, $n = 5$.

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\
&= \frac{P(B|A_i).P(A_i)}{\sum_{i=1}^{n} P(B|A_i).P(A_i)}
\end{aligned}
$$

In the above equation, $P(A_i|B)$ is called posterior probability,

$P(B|A_i)$ is called likelihood, $P(A_i)$ is called prior probability and $\sum_{i=1}^{n} P(B|A_i).P(A_i)$ is called evidence.

Random Variables

Random variable is a function whose domain is the sample space and whose range is the set of real numbers Probabilistic description of a random variable

- Cummulative Probability Distribution:

  It is represented as $F_X(x)$ and defined as

$$F_X(x) = P(X \leq x)$$

  If $x_1 < x_2$, then $F_X(x_1) < F_X(x_2)$ and $0 \leq F_X(x) \leq 1$.

- Probability Density Function:

  It is represented as $f_X(x)$ and defined as

$$f_X(x) = \frac{dF_X(x)}{dx}$$

This implies,

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= \int_{x_1}^{x_2} f_X(x)\,dx \\ f_X(x) &= \frac{dF_X(x)}{dx} \end{aligned}$$

Random Process

A random process is defined as the ensemble(collection) of time functions together with a probability rule (see Figure 2)

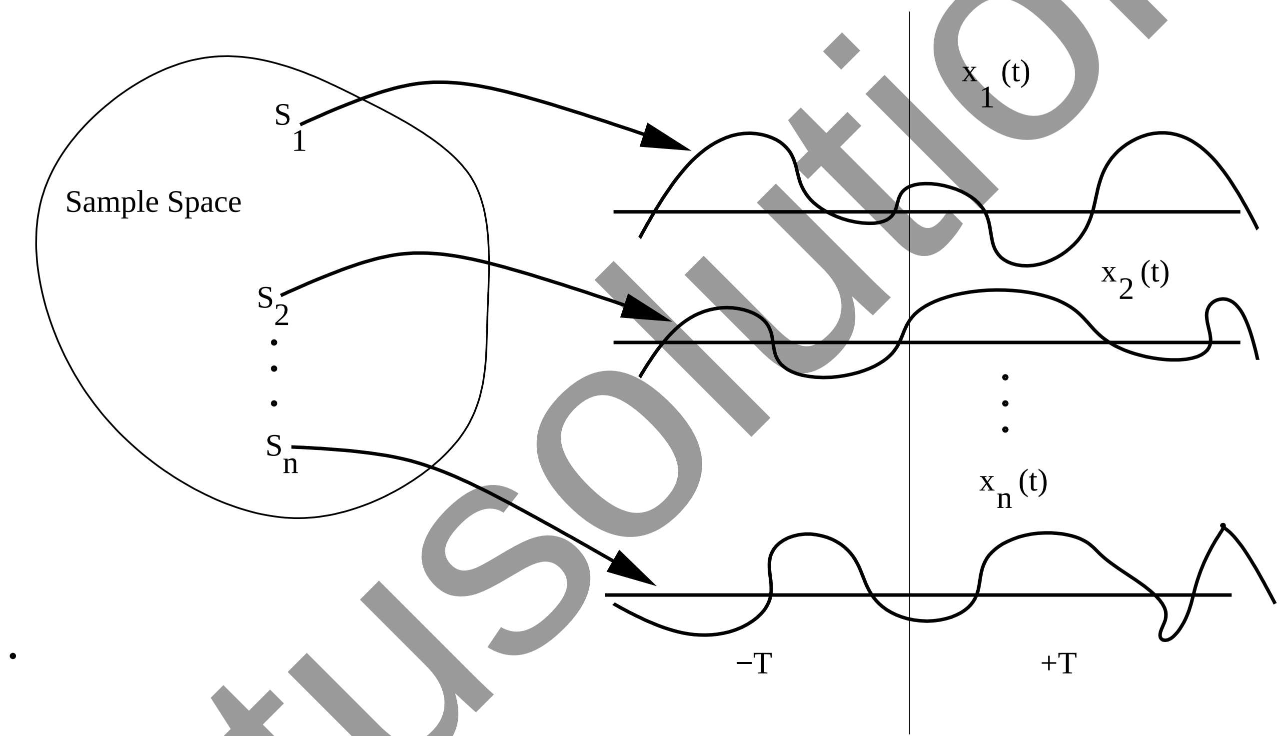


Figure 1: Random Processes and Random Variables

$$x_1(t) \text{ is an outcome of experiment } 1$$
$$x_2(t) \text{ is the outcome of experiment } 2$$
$$\vdots$$
$$x_n(t) \text{ is the outcome of experiment } n$$

- Each sample point in S is associated with a sample function x(t)
- $X(t, s)$ is a random process
  - is an ensemble of all time functions together with a probability rule
  - $X(t, s_j)$ is a realisation or sample function of the random process
  - Probability rules assign probability to any meaningful event associated with an observation An observation is a sample function of the random process
- A random variable:

$\{x_1(t_k), x_2(t_k), ..., x_n(t_k)\} = \{X(t_k, s_1), X(t_k, s_2), ..., X(t_k, s_n)\}$
$X(t_k, s_j)$ constitutes a random variable.

- Outcome of an experiment mapped to a real number
  - An oscillator with a frequency ω_0 with a tolerance of 1%
  - The oscillator can take values between $\omega_0(1 \pm 0.01)$
  - Each realisation of the oscillator can take any value between $(\omega_0)(0.99)$ to $(\omega_0)(1.01)$
  - The frequency of the oscillator can thus be characterised by a random variable
- Stationary random process A random process is said to be stationary if its statistical characterization is independent of the observation interval over which the process was initiated. Mathematically,

$$F_{X(t_1+T)\cdots X(t_k+T)} = F_{X(t_1)\cdots X(t_k)}$$

- Mean, Correlation and Covariance Mean of a stationary random process is independent of the time of observation.

$$\mu_X(t) = E[X(t)] = \mu_x$$

Autocorrelation of a random process is given by:

$$R_X(t_1,t_2) = E[X(t_1)X(t_2)]$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1.x_2 f_{X(t_1)X(t_2)}(x_1,x_2)\,dx_1\,dx_2$$

For a stationary process the autocorrelation is dependent only on the time shiftand not on the time of observation.

Autocovariance of a stationary process is given by

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

- Properties of Autocorrelation

1. $R_X(\tau) = E[X(t+\tau)X(t)]$
2. $R_X(0) = E[X^2(t)]$
3. The autocorrelation function is an even function i.e, $R_X(\tau) = R_X(-\tau)$.
4. The autocorrelation value is maximum for zero shift i.e, $R_X(\tau) \leq R_X(0)$.

Proof:

$$E[(X(t+\tau) \pm X(t))^2] \geq 0$$

$$\implies E[X^2(t+\tau)] + E[X^2(t)] \pm 2E[X(t+\tau)X(t)] \geq 0$$

$$\implies R_X(0) + R_X(0) + 2R_X(\tau) \geq 0$$

$$\implies -R_X(0) \leq R_X(\tau) \leq R_X(0)$$

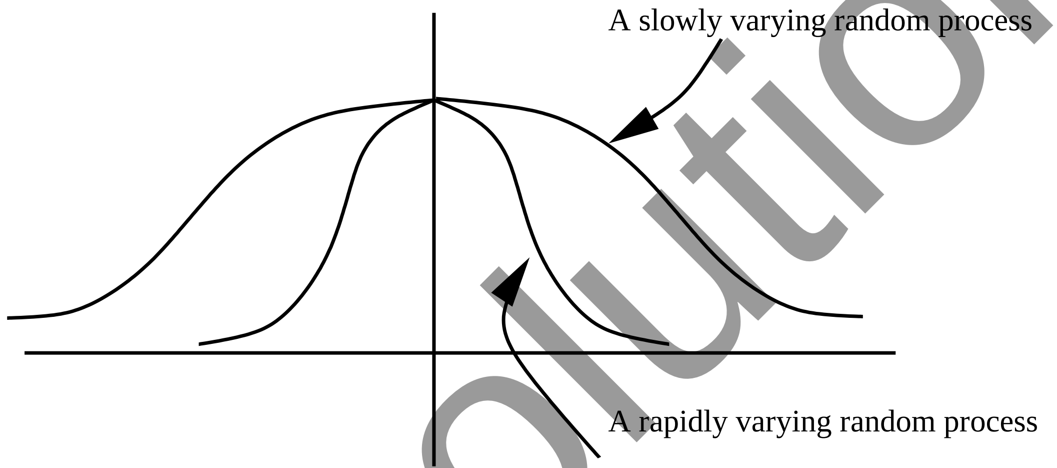


Figure 2: Autocorrelation function of a random process

Random Process: Some Examples

- A sinusoid with random phase

  Consider a sinusoidal signal with random phase, defined by

$$X(t) = a\sin(\omega_0 t + \Theta)$$

where ω_0 and a are constants, and Θ is a random variable that is uniformly distributed over a range of 0 to 2π (see Figure 1)

$$\begin{aligned} f_\Theta(\theta) &= \frac{1}{2\pi}, 0 \leq \theta \leq 2\pi \\ &= 0, elsewhere \end{aligned}$$

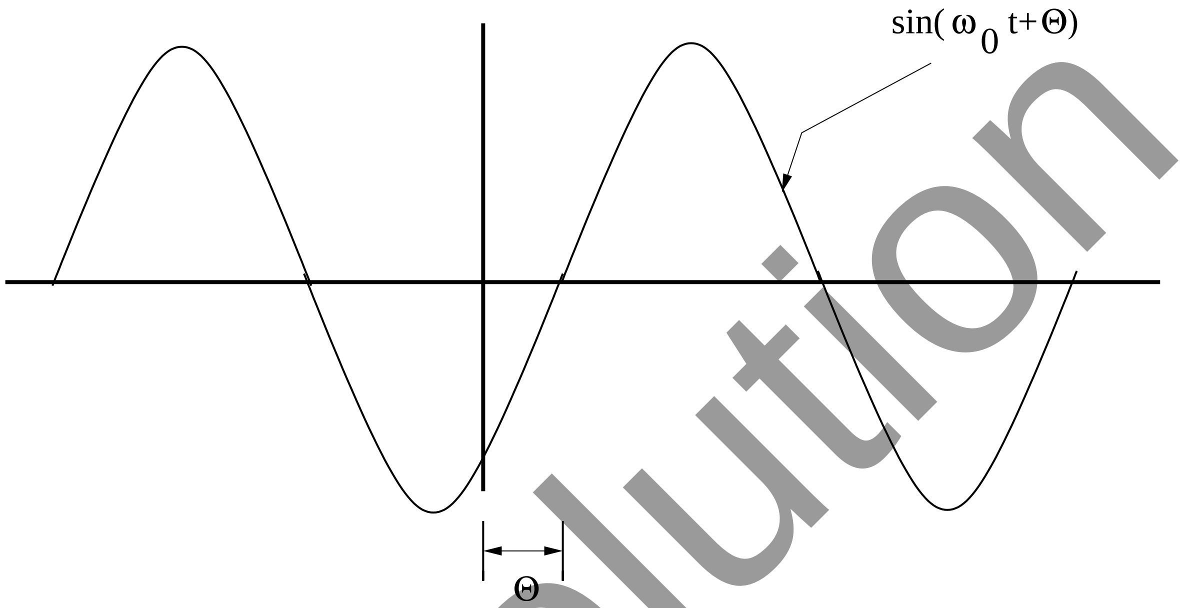


Figure 1: A sinusoid with random phase

This means that the random variable Θ is equally likely to have any value in the range 0 to 2π. The autocorrelation function of $X(t)$ is

$$
\begin{aligned}
R_X(t) &= E[X(t+\tau)X(t)] \\
&= E[\sin(\omega_0 t + \omega_0 \tau) + \Theta)\sin(\omega_0 t + \Theta)] \\
&= \frac{1}{2}E[\sin(2\omega_0 t + \omega_0 \tau + 2\Theta)] + \frac{1}{2}E[\sin(\omega_0 \tau)] \\
&= \frac{1}{2}\int_0^{2\pi} \frac{1}{2\pi}\cos(4\pi f_c t + \omega_0 \tau + 2\Theta)] - \cos(\omega_0 \tau)\, d\Theta
\end{aligned}
$$

The first term intergrates to zero, and so we get

$$
R_X(\tau) = \frac{1}{2}\cos(\omega_0 \tau)
$$

The autocorrelation function is plotted in Figure 2.

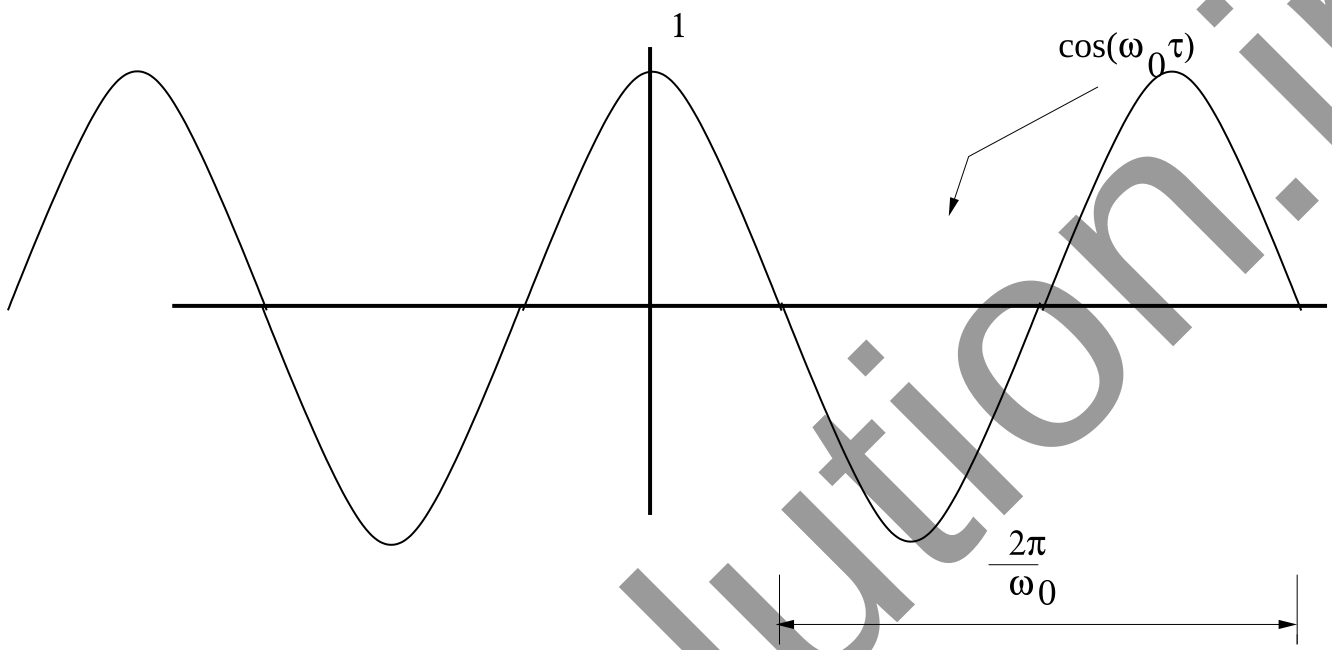


Figure 2: Autocorrelation function of a sinusoid with random phase

The autocorrelation function of a sinusoidal wave with random phase is another sinusoid at the same frequency in the τ - domain

- Random binary wave

Figure 3 shows the sample function $x(t)$ of a random process

$X(t)$ consisting of a random sequence of binary symbols 1 and 0.

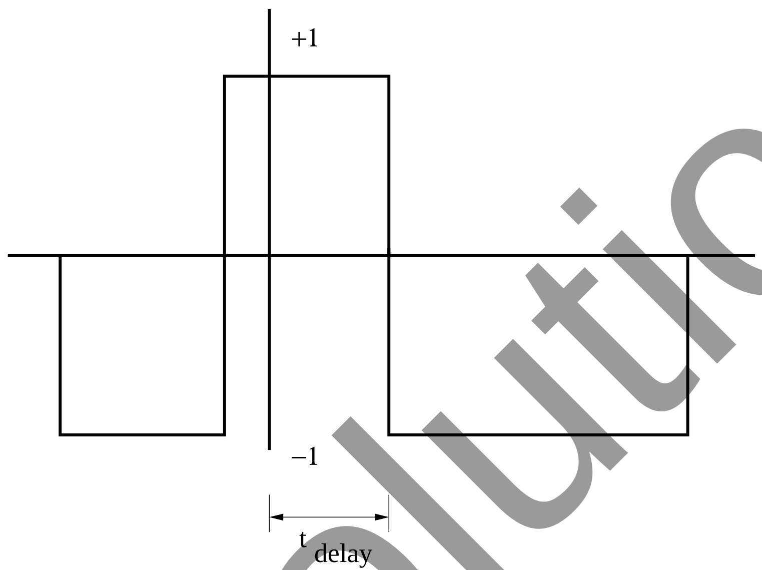


Figure 3: A random binary wave

1. The symbols 1 and 0 are represented by pulses of amplitude $+1$ and -1 volts, respectively and duration T seconds.

2. The starting time of the first pulse, t_{delay}, is equally likely

to lie anywhere between zero and T seconds

3. t_{delay} is the sample value of a uniformly distributed random variable T_{delay} with a probability density function

$$f_{T_{delay}}(t_{delay}) = \frac{1}{T}, 0 \leq t_{delay} \leq T$$
$$= 0, elsewhere$$

4. In any time interval $(n-1)T < t - t_{delay} < nT$, where n is an interger, a 1 or a 0 is determined randomly (for example by tossing a coin: *heads* $\implies$ 1, *tails* $\implies$ 0

– $E[X(t)] = 0$, for all t since 1 and 0 are equally likely.

– Autocorrelation function $R_X(t_k, t_l)$ is given by $E[X(t_k)X(t_l)]$, where $X(t_k)$ and $X(t_l)$ are random variables

Case 1: when $|t_k - t_l| > T$. $X(t_k)$ and $X(t_l)$ occur in different pulse intervals and are therefore independent:

$$E[X(t_k)X(t_l)] = E[X(t_k)]E[X(t_l)] = 0, for|t_k - t_l| > T$$

Case 2: when $|t_k - t_l| < T$, with $t_k = 0$ and $t_l < t_k$. $X(t_k)$ and $X(t_l)$ occur in the same pulse interval provided t_{delay} satisfies the condition $t_{delay} < T - |t_k - t_l|$.

$$\begin{aligned} E[X(t_k)X(t_l)|t_{delay}] &= 1, t_{delay} < T - |t_k - t_l| \\ &= 0, elsewhere \end{aligned}$$

Averaging this result over all possible values of t_{delay}, we get

$$
\begin{aligned}
E[X(t_k)X(t_l)] &= \int_0^{T-|t_k-t_l|} f_{T_{delay}}(t_{delay})\, dt_{delay} \\
&= \int_0^{T-|t_k-t_l|} \frac{1}{T} dt_{delay} \\
&= (1 - \frac{|t_k - t_l|}{T}), |t_k - t_l| < T
\end{aligned}
$$

The autocorrelation function is given by

$$
\begin{aligned}
R_X(\tau) &= (1 - \frac{|\tau|}{T}), |\tau| < T \\
&= 0, |\tau| > T
\end{aligned}
$$

This result is shown in Figure 4

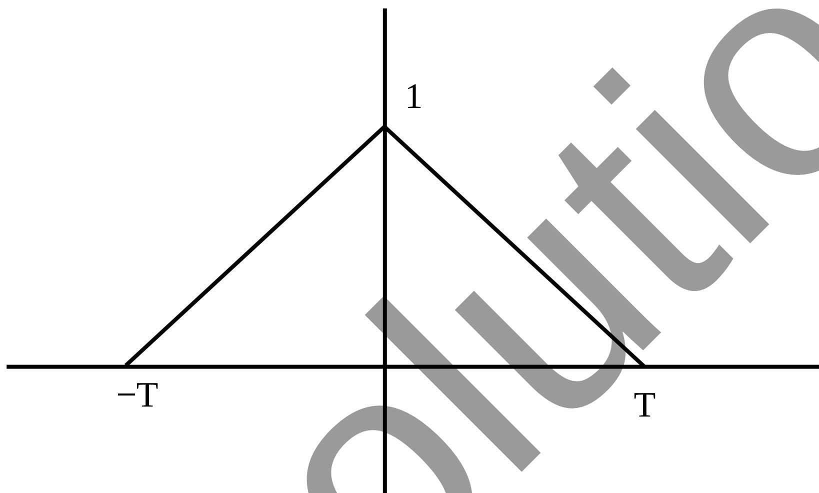


Figure 4: Autocorrelation of a random binary wave

Random Process: Some Examples

- Quadrature Modulation Process Given two random variables $X_1(t)$ and $X_2(t)$

$$X_1(t) = X(t)\cos(2\omega_0 t + \Theta)$$
$$X_2(t) = X(t)\sin(2\omega_0 t + \Theta)$$

where ω_0 is a constant, and Θ is a random variable that is uniformly distributed over a range of 0 to 2π, that is,

$$f_\Theta(\theta) = \frac{1}{2\pi}, 0 \leq \theta \leq 2\pi$$
$$= 0, elsewhere$$

The correlation function of $X_1(t)$ and $X_2(t)$ is

$$
\begin{aligned}
R_{12}(\tau) &= E[X_1(t)X_2(t+\tau)] \\
&= E[X(t)\cos(\omega_0 t + \Theta)X(t-\tau)\sin(2\omega_0(t-\tau)+\Theta)] \\
&= \int_0^{2\pi} \frac{1}{2\pi} X(t)X(t-\tau)\cos(\omega_0 t+\Theta)\sin(\omega_0(t-\tau)+\Theta)\, d\Theta \\
&= \frac{1}{2} R_X(\tau)\sin(\omega_0\tau)
\end{aligned}
$$

Random Process: Time vs. Ensemble Averages

Ensemble averages

- Difficult to generate a number of realisations of a random process
- $\implies$ use time averages
- Mean

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

- Autocorrelation

$$R_x(\tau, T) = \frac{1}{2T}\int_{-T}^{+T} x(t)x(t+\tau)\,dt$$

- Ergodicity A random process is called ergodic if

1. it is ergodic in mean:

$$\lim_{T\to+\infty} \mu_x(T) = \mu_X$$
$$\lim_{T\to+\infty} var[\mu_x(T)] = 0$$

2. it is ergodic in autocorrelation:

$$\lim_{T\to+\infty} R_x(\tau, T) = R_X(\tau)$$
$$\lim_{T\to+\infty} var[R_x(\tau, T)] = 0$$

where μ_X and $R_X(\tau)$ are the ensemble averages of the same random process.

Random Processes and Linear Shift Invariant Systems(LSI)

- The communication channel can be thought of as a system
- The signal that is transmitted through the channel is a realisation of the random process
- It is necessary to understand the behaviour of a signal that is input to a system.
- For analysis purposes it is assumed that a system is LSI
- Linear Shift Invariant(LSI) Systems

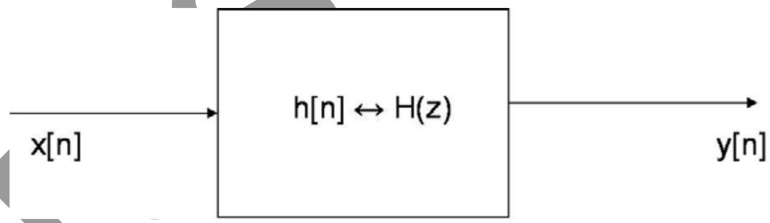


Figure 1: An LSI system

In Figure 1 , $h[n]$ is an LSI system if it satisfies the following properties

– Linearity The system is called linear, if the following equation holds for all signals $x_1[n]$ and $x_2[n]$ and any a and b:

$$\begin{aligned} x_1[n] &\rightarrow y_1[n] \\ x_2[n] &\rightarrow y_2[n] \\ \Longrightarrow a.x_1[n] + b.x_2[n] &\rightarrow a.y_1[n] + b.y_2[n] \end{aligned}$$

– Shift Invariance The system is called Shift Invariant, if the following equation holds for any signal $x[n]$

$$\begin{aligned} x[n] &\rightarrow y[n] \\ \Longrightarrow x[n-n_0] &\rightarrow y[n-n_0] \end{aligned}$$

- The assumption is that the output of the system is linear, in that if the input scaled, the output is scaled by the same factor.
- The system supports superposition
  - When two signals are added in the time domain, the output is equal to the sum of the individual responses
- If the input to the system is delayed by n_0, the output is also delayed by n_0.

Random Process through a linear filter

- A random process $X(t)$ is applied as input to a linear time-invariant filter of impulse response $h(t)$,
- It produces a random process $Y(t)$ at the filter output as shown in Figure 1

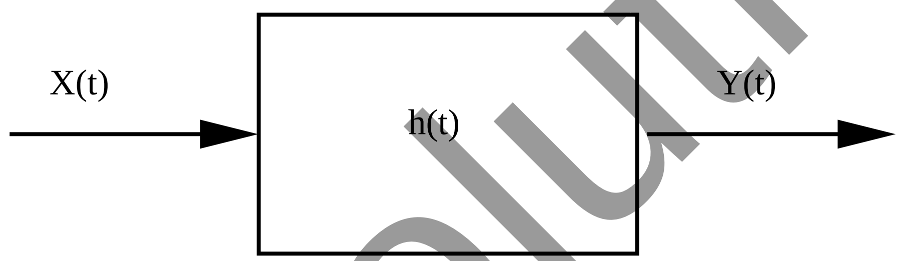


Figure 1: Transmission of a random process through a linear filter

- Difficult to describe the probability distribution of the output random process $Y(t)$, even when the probability distribution of the input random process $X(t)$ is completely specified for $-\infty \leq t \leq +\infty$.

- Estimate characteristics like mean and autocorrelation of the output and try to analyse its behaviour.
- Mean The input to the above system $X(t)$ is assumed stationary. The mean of the output random process $Y(t)$ can be calculated

$$
\begin{aligned}
m_Y(t) &= E[Y(t)] = E\left[\int_{-\infty}^{+\infty} h(\tau)X(t-\tau)\,d\tau\right] \\
&= \int_{-\infty}^{+\infty} h(\tau)E[X(t-\tau)]\,d\tau \\
&= \int_{-\infty}^{+\infty} h(\tau)m_X(t-\tau)\,d\tau \\
&= m_X\int_{-\infty}^{+\infty} h(\tau)\,d\tau \\
&= m_X H(0)
\end{aligned}
$$

where $H(0)$ is the zero frequency response of the system.

- Autocorrelation The autocorrelation function of the output random process $Y(t)$. By definition, we have

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

where t and u denote the time instants at which the process is observed. We may therefore use the convolution integral to write

$$\begin{aligned} R_Y(t,u) &= E\left[\int_{-\infty}^{+\infty} h(\tau_1)X(t-\tau_1)\,d\tau_1 \int_{-\infty}^{+\infty} h(\tau_2)X(t-\tau_2)\,d\tau_2\right] \\ &= \int_{-\infty}^{+\infty} h(\tau_1)\,d\tau_1 \int_{-\infty}^{+\infty} h(\tau_2)E\left[X(t-\tau_1)X(t-\tau_2)\right]\,d\tau_2 \end{aligned}$$

- When the input $X(t)$ is a wide-stationary random process,
  - The autocorrelation function of $X(t)$ is only a function of the difference between the observation times $t-\tau_1$ and

$u - \tau_2$.

– Putting $\tau = t - u$, we get

$$R_Y(\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau_1)h(\tau_2)R_X(\tau - \tau_1 + \tau_2)\,d\tau_1\,d\tau_2$$

– $R_Y(0) = E[Y^2(t)]$

– The mean square value of the output random process $Y(t)$ is obtained by putting $\tau = 0$ in the above equation.

$$\begin{aligned} E[Y^2(t)] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(\tau_1)h(\tau_2)R_X(\tau_2 - \tau_1)\,d\tau_1\,d\tau_2 \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} H(\omega)\exp(j\omega\tau_1)\,d\omega\right] h(\tau_2)R_X(\tau_2 - \tau_1)\,d\tau_1\,d\tau_2 \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} H(\omega)\,d\omega \int_{-\infty}^{+\infty} h(\tau_2)\,d\tau_2 \int_{-\infty}^{+\infty} R_X(\tau_2 - \tau_1)\exp(j2\omega\tau_1)\,d\tau_1 \end{aligned}$$

– Putting $\tau = \tau_2 - \tau_1$

$$E[Y^2(t)] = \frac{1}{2\pi}\int_{-\infty}^{+\infty} H(\omega)\,d\omega \int_{-\infty}^{+\infty} h(\tau_2)\exp(j\omega\tau_2)\,d\tau_2 \int_{-\infty}^{+\infty} R_X(\tau)\exp(-j2\omega\tau)\,d\tau$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} H(\omega)\,d\omega \int_{-\infty}^{+\infty} H^*(\omega)\,d\omega \int_{-\infty}^{+\infty} R_X(\tau)\exp(-j\omega\tau)\,d\tau$$

– This is simply the Fourier Transform of the autocorrelation function $R_X(t)$ of the input random process $X(t)$. Let this transform be denoted by $S_X(f)$.

$$S_X(\omega) = \int_{-\infty}^{+\infty} R_X(\tau)\exp(-j\omega\tau)\,d\tau$$

– $S_X(\omega)$ is called the *power spectral density* or *power spectrum* of the wide-sense stationary random process $X(t)$.

$$E[Y^2(t)] = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(\omega)|^2 S_X(\omega)\,df$$

– *The mean square value of the output of a stable linear time-invariant filter in response to a wide-sense stationary random process is equal to the integral over all frequencies of the power spectral density of the input random process multiplied by the squared magnitude of the transfer function of the filter.*

Definition of Bandwidth

- *"Bandwidth is defined as a band containing all frequencies between upper cut-off and lower cut-off frequencies."* (see Figure 1)

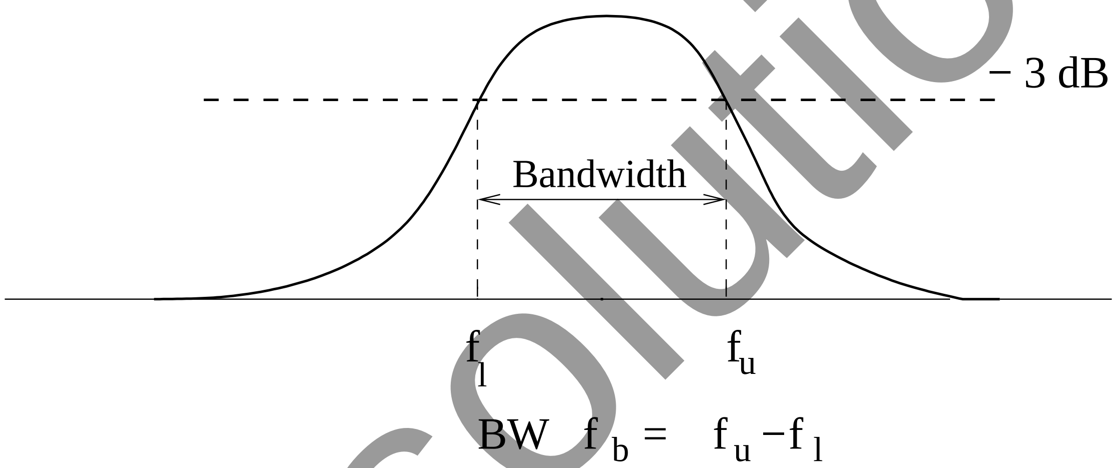


Figure 1: Bandwidth of a signal

Upper and lower cut-off (or $3dB$) frequencies corresponds to the frequencies where the magnitude of signal's Fourier Transform is reduced to half ($3dB$ less than) its maximum value.

- Importance of Bandwidth
  - Bandwidth enables computation of the power required to transmit a signal.
    - Signals that are band-limited are not time-limited
    - Energy of a signal is defined as:

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

    - Energy of a signal that is not time-limited can be computed using Parsevals Theorem[a]:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega$$

[a] The power of a signal is the energy dissipated in a one ohm resistor

- ∗ Multiple signals (of finite bandwidth) can be multiplexed on to a single channel.

- For all pulse signals, duration of the pulse and its bandwidth satisfies inverse relation between frequency and time.

$$duration \times bandwidth = constant$$

  - wider the pulse $\Longrightarrow$ smaller is the bandwidth required
  - wider the pulse $\Longrightarrow$ *Inter-Symbol-Interference* is an issue
  - ideally a pulse that is narrow and has a small bandwidth is required

- Noise equivalent Bandwidth:
  Let, a white noise signal with power $N_0/2$ be fed to an arbitrary Low Pass Filter (LPF) wwith transfer function $H(\omega)$.
  Then, output noise power is given by

$$N_{out} = \frac{1}{2\pi}\frac{N_0}{2}\int_{-\infty}^{+\infty} |H(\omega)|^2 \; d\omega$$
$$= \frac{N_0}{2\pi}\int_{0}^{+\infty} |H(\omega)|^2 \; d\omega$$

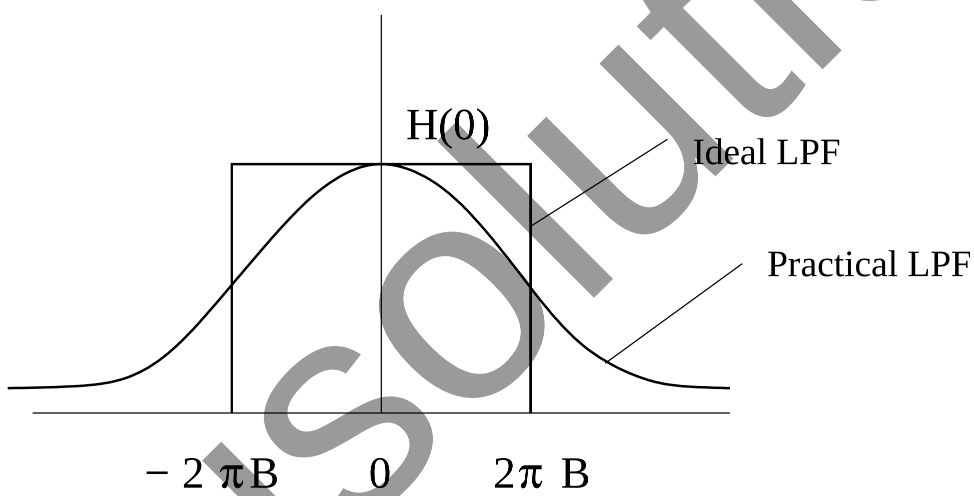


Figure 2: Ideal and Practical LPFs

For an ideal LPF,

$$N_{out} = N_0 B H^2(0)$$

$$\implies B = \frac{\frac{1}{2\pi}\int_0^{+\infty} |H(\omega)|^2 \, d\omega}{H^2(0)}$$

Modulation

Modulation is a process that causes a shift in the range of frequencies in a signal.

- Signals that occupy the same range of frequencies can be separated
- Modulation helps in noise immunity, attentuation - depends on the physical medium

Figure 1 shows the different kinds of analog modulation schemes that are available

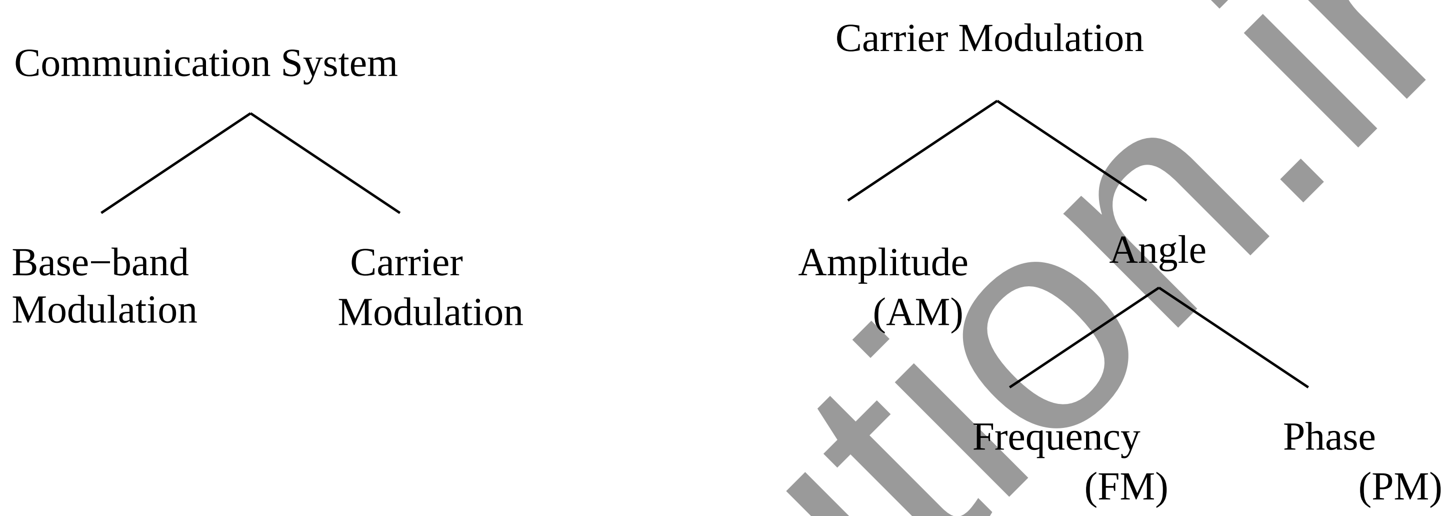


Figure 1: A broad view of communication system

- Amplitude Modulation It is the process where, the amplitude of the carrier is varied proportional to that of the message signal.
  - Amplitude Modulation with carrier
    Let $m(t)$ be the base-band signal, $m(t) \longleftrightarrow M(\omega)$ and $c(t)$ be the carrier, $c(t) = A_c \cos(\omega_c t)$. f_c is chosen such that $f_c >> W$, where W is the maximum frequency component

of $m(t)$.

The amplitude modulated signal is given by

$$s(t) = A_c\left[1 + k_a m(t)\right]\cos(\omega_c t)$$

$$\begin{aligned} S(\omega) &= \pi\frac{A_c}{2}\left(\delta(\omega-\omega_c)+\delta(\omega+\omega_c)\right) + \\ &\quad \frac{k_a A_c}{2}\left(M(\omega-\omega_c)+M(\omega+\omega_c)\right) \end{aligned}$$

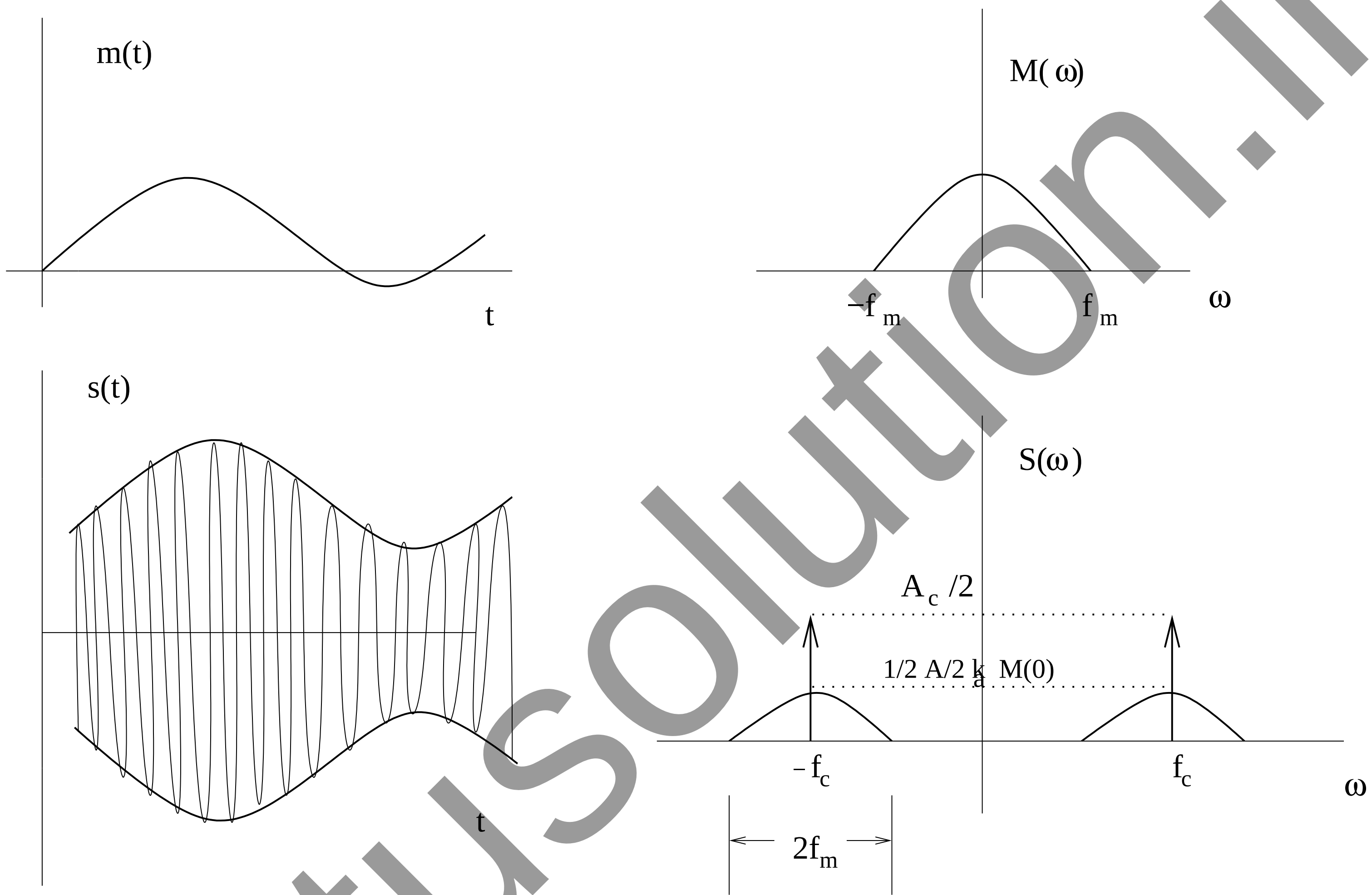


Figure 2: Amplitude modulation

Figure 2 shows the spectrum of the Amplitude Modulated signal.

– k_a is a constant called *amplitude sensitivity*. $k_a m(t) < 1$ and it indicates percentage modulation.

– Modulation in AM: A product modulator is used for generating the modulated signal as shown in Figure 3.

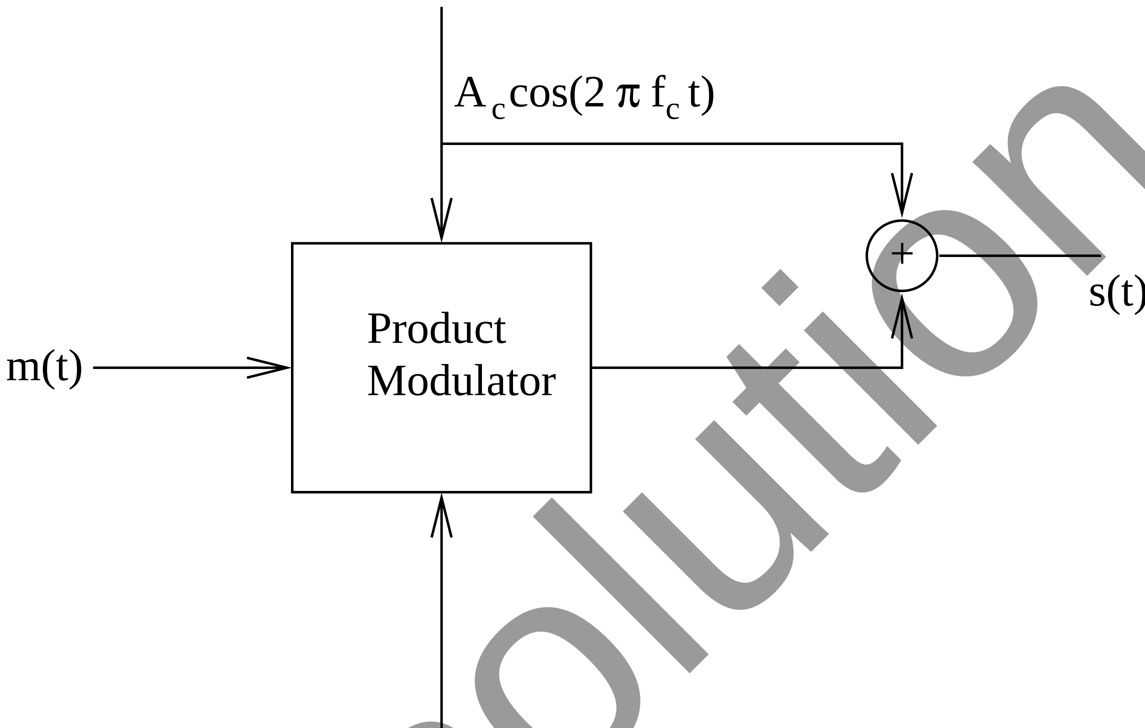


Figure 3: Modulation using product modulator

– Demodulation in AM: An envelope detector is used to get the demodulated signal (see Figure 4).

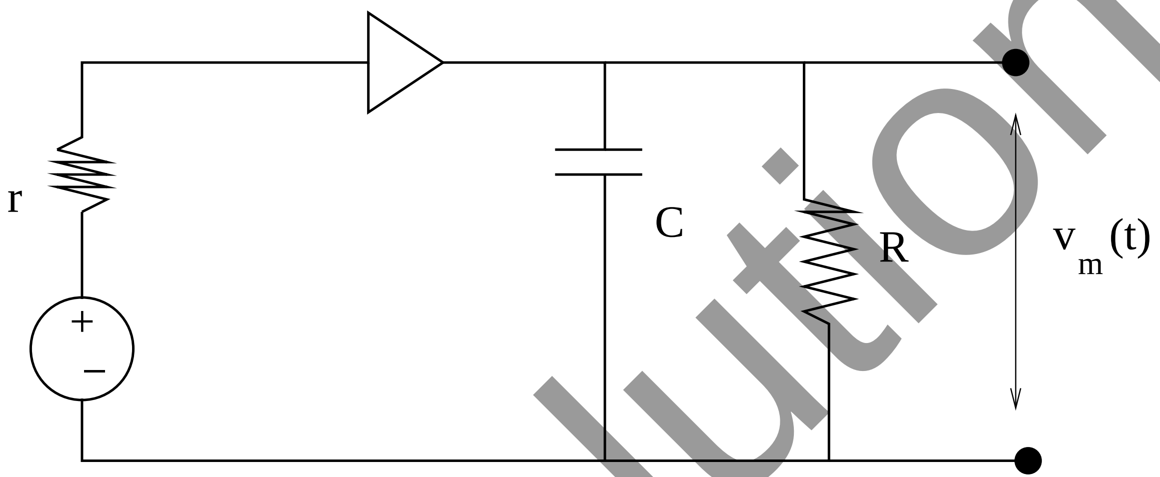


Figure 4: Demodulation using Envelope detector

– The voltage $v_m(t)$ across the resistor R gives the message signal $m(t)$

Double Side Band - Suppressed Carrier (DSB-SC) Modulation

- In AM modulation, transmission of carrier consumes lot of power. Since, only the side bands contain the information about the message, carrier is suppressed. This results in a DSB-SC wave.
- A DSB-SC wave $s(t)$ is given by

$$\begin{aligned} s(t) &= m(t)A_c\cos(\omega_c t) \\ S(\omega) &= \pi\frac{A_c}{2}\left(M(\omega-\omega_c)+M(\omega+\omega_c)\right) \end{aligned}$$

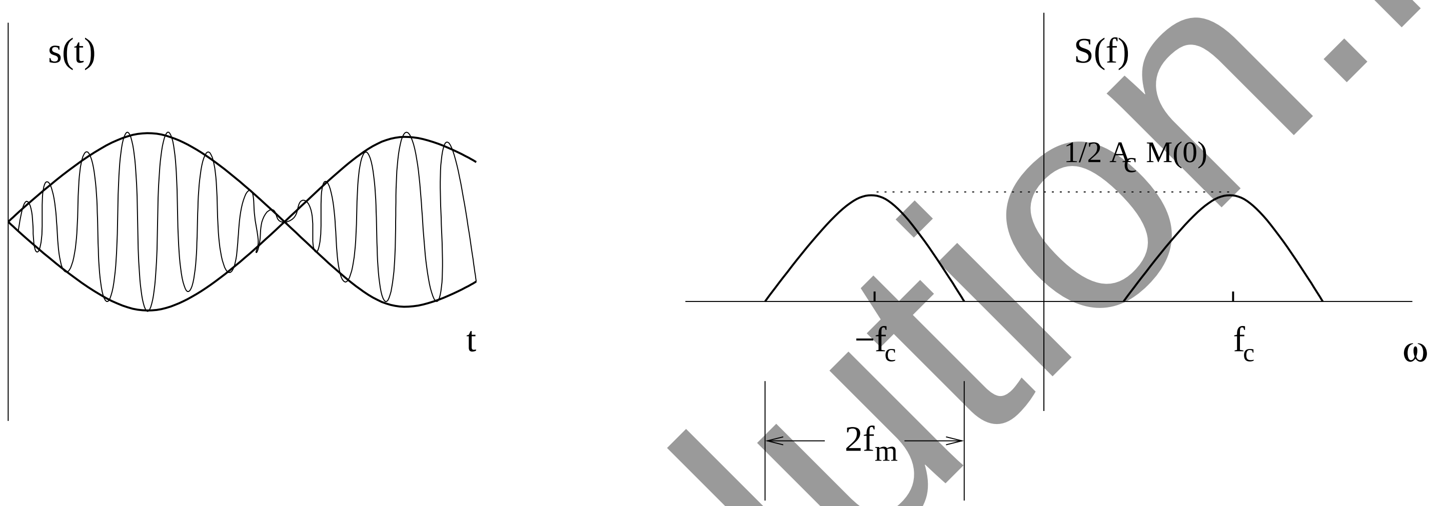


Figure 5: DSB-SC modulation

- Modulation in DSB-SC: Here also product modulator is used as shown in Figure 3, but the carrier is not added. Figure 6 shows the spectrum of the DSB-SC signal.

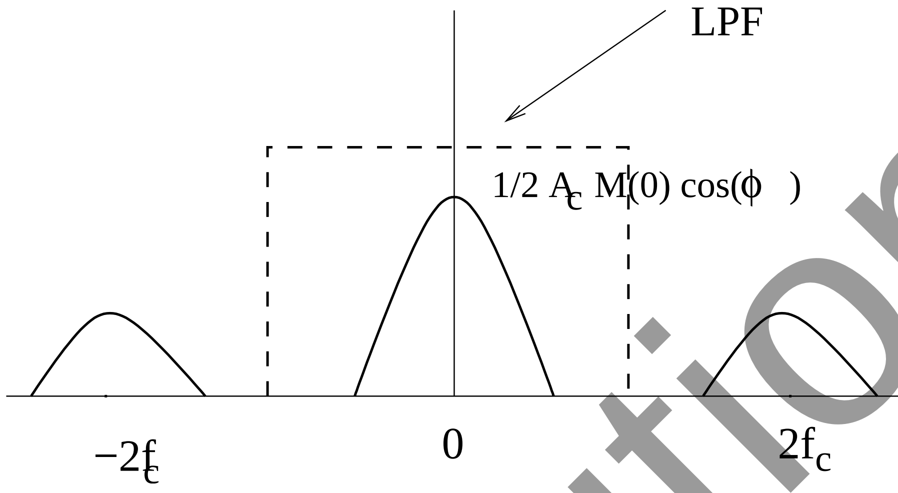


Figure 6: Spectrum of Demodulated DSB-SC signal

- Demodulation in DSB-SC: A coherent demodulator is used. The local oscillator present in the demodulator generates a carrier which has same frequency and phase(i.e. $\phi = 0$ in Figure 7) [a] as that of the carrier in the modulated signal (see Figure 7)

[a] Clearly the design of the demodulator for DSB-SC is more complex than that vanilla AM

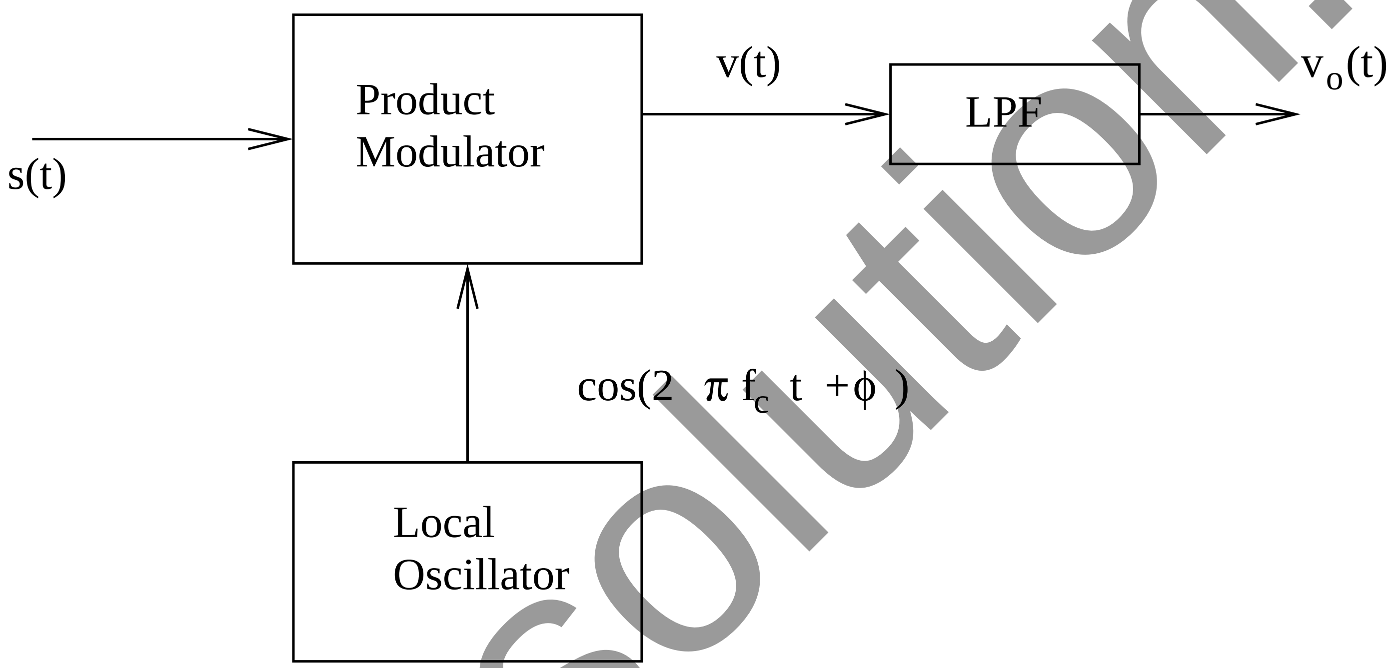


Figure 7: Coherent detector

$$
\begin{aligned}
v(t) &= s(t).\cos(\omega_c t + \phi) \\
&= m(t) A_c \cos(\omega_c t) \cos(\omega_c t + \phi) \\
&= \frac{m(t)}{2} A_c \left[\cos(2\omega_c t + \phi) + \cos(\phi)\right]
\end{aligned}
$$

- If, the demodulator (Figure 7) has constant phase, the original signal is reconstructed by passing $v(t)$ through an LPF.

Single Side Band (SSB) Modulation

- In DSB-SC it is observed that there is symmetry in the bandstructure. So, even if one half is transmitted, the other half can be recovered at the received. By doing so, the bandwidth and power of transmission is reduced by half.

  Depending on which half of DSB-SC signal is transmitted, there are two types of SSB modulation

  1. Lower Side Band (LSB) Modulation
  2. Upper Side Band (USB) Modulation

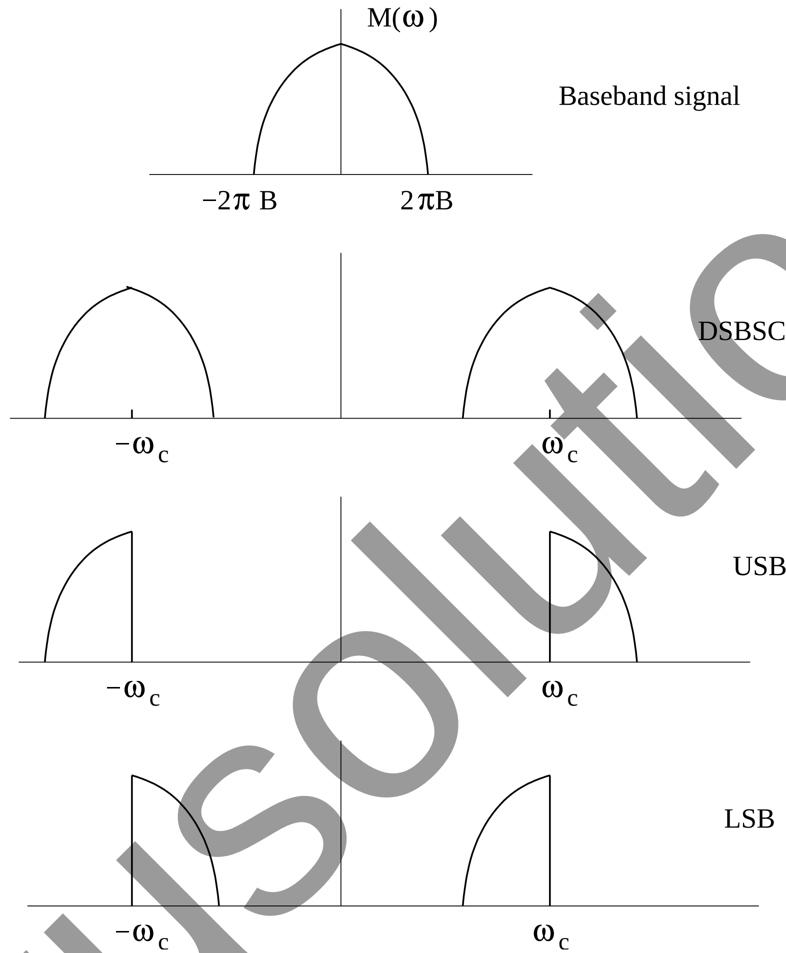


Figure 1: SSB signals from orignal signal

- Mathematical Analysis of SSB modulation

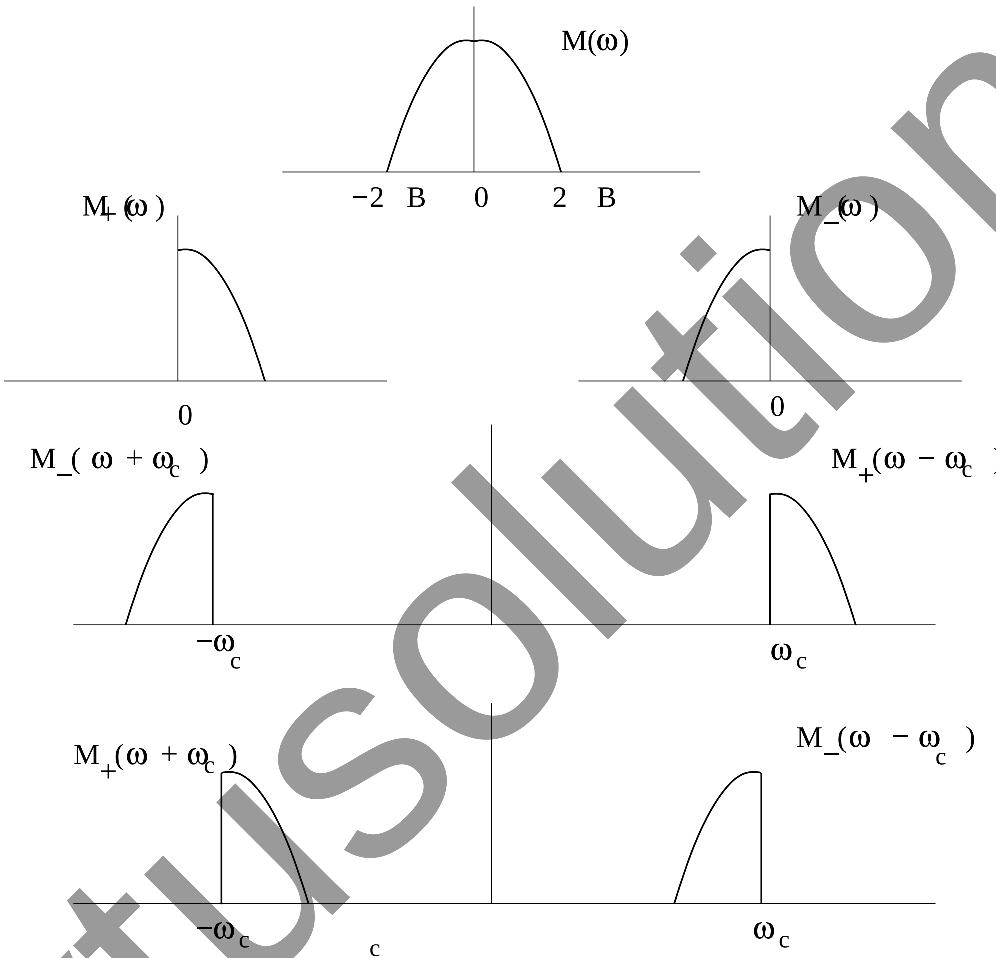


Figure 2: Frequency analysis of SSB signals

- From Figure 2 and the concept of the Hilbert Transform,

$$\begin{aligned}\Phi_{USB}(\omega) &= M_+(\omega-\omega_c)+M_-(\omega+\omega_c)\\ \phi_{USB}(t) &= m_+(t)e^{j\omega_c t}+m_-(t)e^{-j\omega_c t}\end{aligned}$$

But, from complex representation of signals,

$$\begin{aligned}m_+(t) &= m(t)+j\hat{m}(t)\\ m_-(t) &= m(t)-j\hat{m}(t)\end{aligned}$$

So,

$$\phi_{USB}(t) = m(t)\cos(\omega_c t) - \hat{m}(t)\sin(\omega_c t)$$

Similarly,

$$\phi_{LSB}(t) = m(t)\cos(\omega_c t) + \hat{m}(t)\sin(\omega_c t)$$

- Generation of SSB signals A SSB signal is represented by:

$$\phi_{SSB}(t) = m(t)\cos(\omega_c t) \pm \hat{m}(t)\sin(\omega_c t)$$

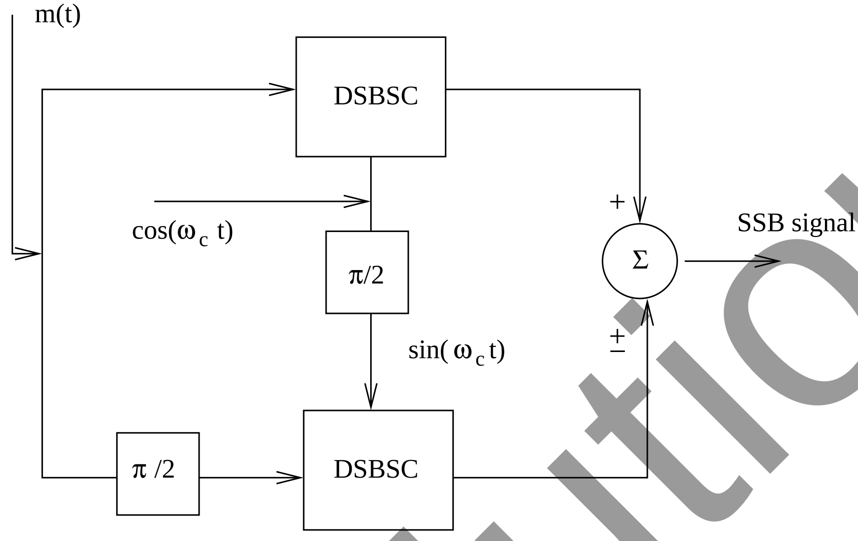


Figure 3: Generation of SSB signals

As shown in Figure 3, a DSB-SC modulator is used for SSB signal generation.

- Coherent Demodulation of SSB signals $\Phi_{SSB}(t)$ is multiplied with $\cos(\omega_c t)$ and passed through low pass filter to get back the orignal signal.

$$\phi_{SSB}(t)\cos(\omega_c t) = \frac{1}{2}m(t)\left[1+\cos(2\omega_c t)\right] \pm \frac{1}{2}\hat{m}(t)\sin(2\omega_c t)$$
$$= \frac{1}{2}m(t) + \frac{1}{2}\cos(2\omega_c t) \pm \frac{1}{2}\hat{m}(t)\sin(2\omega_c t)$$

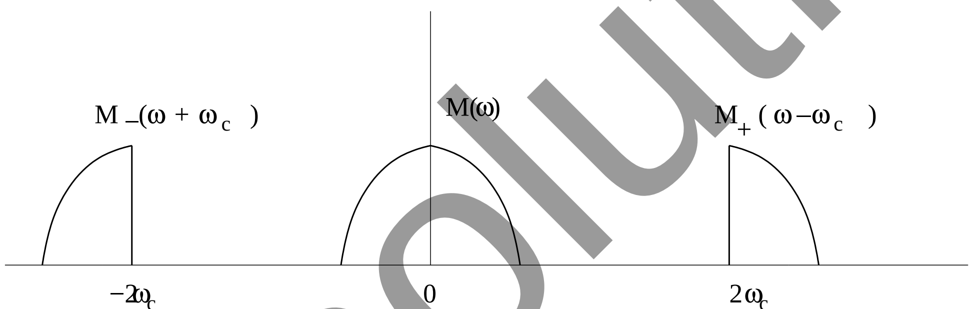


Figure 4: Demodulated SSB signal

The demodulated signal is passed through an LPF to remove unwanted SSB terms.

Vestigial Side Band (VSB) Modulation

- The following are the drawbacks of SSB signal generation:
  1. Generation of an SSB signal is difficult.
  2. Selective filtering is to be done to get the original signal back.
  3. Phase shifter should be exactly tuned to 90^0.
- To overcome these drawbacks, VSB modulation is used. It can viewed as a compromise between SSB and DSB-SC. Figure 5 shows all the three modulation schemes.

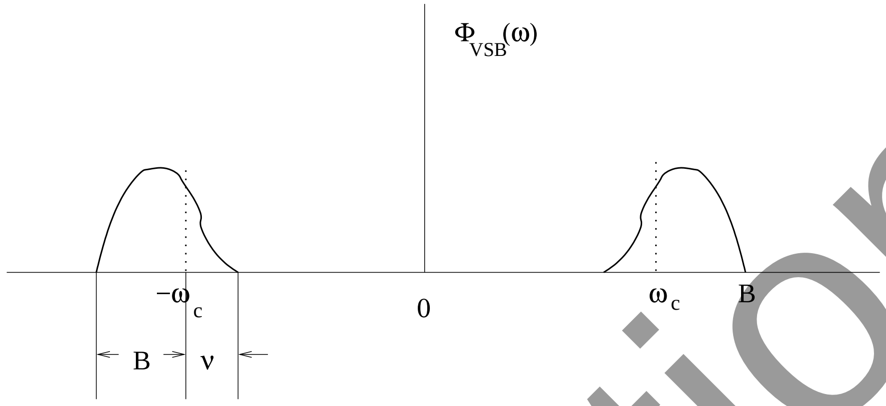


Figure 5: VSB Modulation

- In VSB
  1. One sideband is not rejected fully.
  2. One sideband is transmitted fully and a small part (vestige) of the other sideband is transmitted.
- The transmission BW is $BW_v = B + v$. where, v is the vestigial frequency band. The generation of VSB signal is shown in Figure 6

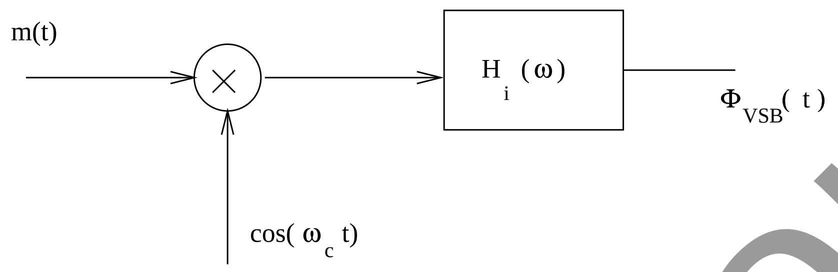


Figure 6: Block Diagram - Generation of VSB signal

Here, $H_i(\omega)$ is a filter which shapes the other sideband.

$$\Phi_{VSB}(\omega) = \left[M(\omega - \omega_c) + M(\omega + \omega_c)\right].H_i(\omega)$$

To recover the original signal from the VSB signal, the VSB signal is multiplied with $\cos(\omega_c t)$ and passed through an LPF such that original signal is recovered.

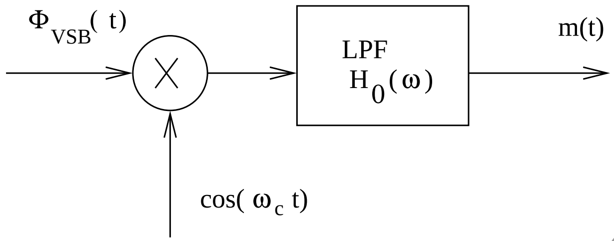


Figure 7: Block Diagram - Demodulation of VSB signal

From Figure 6 and Figure 7, the criterion to choose LPF is:

$$\begin{aligned} M(\omega) &= [\Phi_{VSB}(\omega+\omega_c)+\Phi_{VSB}(\omega-\omega_c)]\,.H_0(\omega) \\ &= [H_i(\omega+\omega_c)+H_i(\omega-\omega_c)]\,.M(\omega).H_0(\omega) \\ \Longrightarrow\ H_0(\omega) &= \frac{1}{H_i(\omega+\omega_c)+H_i(\omega-\omega_c)} \end{aligned}$$

Appendix: The Hilbert Transform

The Hilbert Transform on a signal changes its phase by $\pm 90^0$. The Hilbert transform of a signal $g(t)$ is represented as $\hat{g}(t)$.

$$\hat{g}(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{g(\tau)}{t-\tau}d\tau$$

$$g(t) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\hat{g}(\tau)}{t-\tau}d\tau$$

We, say $g(t)$ and $\hat{g}(t)$ constitute a Hilbert Transform pair. If we observe the above equations, it is evident that Hilbert transform is nothing but the convolution of $g(t)$ with $\frac{1}{\pi t}$.

The Fourier Transform of $\hat{g}(t)$ is computed from signum function $sgn(t)$.

$$sgn(t) \longleftrightarrow \frac{2}{j\omega}$$

$$\implies \frac{1}{\pi t} \longleftrightarrow -jsgn(\omega)$$

Where,

$$sgn(\omega) = \begin{cases} 1, & \omega > 0 \\ 0, & \omega = 0 \\ -1, & \omega < 0 \end{cases}$$

Since, $\hat{g}(t) = g(t) * \frac{1}{\pi t}$,

$$\hat{G}(\omega) = G(\omega) \times -jsgn(\omega)$$

Properties of Hilbert Transform

1. $g(t)$ and $\hat{g}(t)$ have the same magnitude spectrum.
2. If $\hat{g}(t)$ is HT of $g(t)$ then HT of $\hat{g}(t)$ is $-g(t)$.
3. $g(t)$ and $\hat{g}(t)$ are orthogonal over the entire interval $-\infty$ to $+\infty$.

$$\int_{-\infty}^{+\infty} g(t)\hat{g}(t)\,dt = 0$$

Complex representation of signals

If $g(t)$ is a real valued signal, then its complex representation $g_+(t)$ is given by

$$
\begin{aligned}
g_+(t) &= g(t) + j\hat{g}(t) \\
G_+(\omega) &= G(\omega) + sgn(\omega)G(\omega)
\end{aligned}
$$

Therefore,

$$
G_+(\omega) = \begin{cases} 2G(\omega), & \omega > 0 \\ G(0), & \omega = 0 \\ 0, & \omega < 0 \end{cases}
$$

$g_+(t)$ is called *pre-evelope* and exists only for positive frequencies. For negative frequencies $g_-(t)$ is defined as follows:

$$
\begin{aligned}
g_-(t) &= g(t) - j\hat{g}(t) \\
G_-(\omega) &= G(\omega) - sgn(\omega)G(\omega)
\end{aligned}
$$

Therefore,

$$G_-(\omega) = \begin{cases} 2G(\omega), & \omega < 0 \\ G(0), & \omega = 0 \\ 0, & \omega > 0 \end{cases}$$

- Essentially the *pre-envelope* of a signal enables the suppression of one of the sidebands in signal transmission.
- The *pre-envelope* is used in the generation of the SSB-signal.

Angle Modulation

In this type of modulation, the frequency or phase of carrier is varied in proportion to the amplitude of the modulating signal.

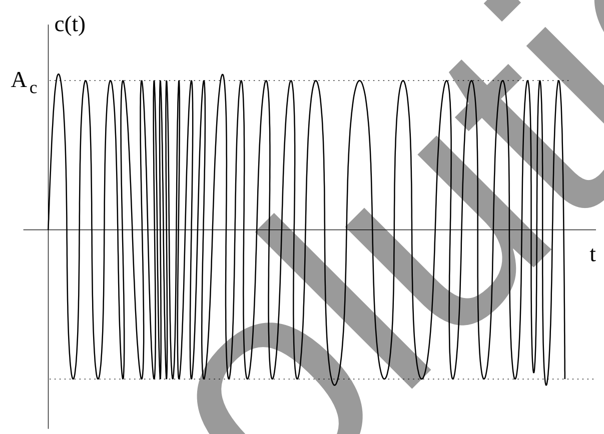


Figure 1: An angle modulated signal

If $s(t) = A_c \cos(\theta_i(t))$ is an angle modulated signal, then

1. Phase modulation:

$$\theta_i(t) = \omega_c t + k_p m(t)$$

where $\omega_c = 2\pi f_c$.

2. Frequency Modulation:

$$\begin{aligned} \omega_i(t) &= \omega_c + k_f m(t) \\ \theta_i(t) &= \int_0^t \omega_i(t)\,dt \\ &= 2\pi \int_0^t f_i(t)\,dt + \int_0^t k_f m(t)\,dt \end{aligned}$$

- Phase Modulation If $m(t) = A_m \cos(2\pi f_m t)$ is the message signal, then the phase modulated signal is given by

$$s(t) = A_c \cos(\omega_c t + k_p m(t))$$

Here, k_p is phase sensitivity or phase modulation index.

- Frequency Modulation If $m(t) = A_m \cos(2\pi f_m t)$ is the message signal, then the Frequency modulated signal is given by

$$2\pi f_i(t) = \omega_c + k_f A_m \cos(2\pi f_m t)$$

$$\theta_i(t) = \omega_c t + \frac{k_f A_m}{2\pi f_m} \sin(2\pi f_m t)$$

here, $\frac{k_f A_m}{2\pi}$ is called *frequency deviation* (Δf) and $\frac{\Delta f}{f_m}$ is called *modulation index* (β). The Frequency modulated signal is given by

$$s(t) = A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t))$$

Depending on how small β is FM is either *Narrowband FM*($\beta << 1$) or *Wideband FM*($\beta \approx 1$).

– Narrow-Band FM (NBFM)

In NBFM $\beta << 1$, therefor $s(t)$ reduces as follows:

$$\begin{aligned} s(t) &= A_c \cos(2\pi f_c t + \beta \sin(2\pi f_m t)) \\ &= A_c \cos(2\pi f_c t) \cos(\beta \sin(2\pi f_m t)) - \\ &\quad A_c \sin(2\pi f_c t) \sin(\beta \sin(2\pi f_m t)) \end{aligned}$$

Since, β is very small, the above equation reduces to

$$s(t) = A_c \cos(2\pi f_c t) - A_c \beta \sin(2\pi f_m t) \sin(2\pi f_c t)$$

The above equation is similar to AM. Hence, for NBFM the bandwidth is same as that of AM i.e., $2 \times \textit{message bandwidth}(2 \times B)$.

A NBFM signal is generated as shown in Figure **??**.

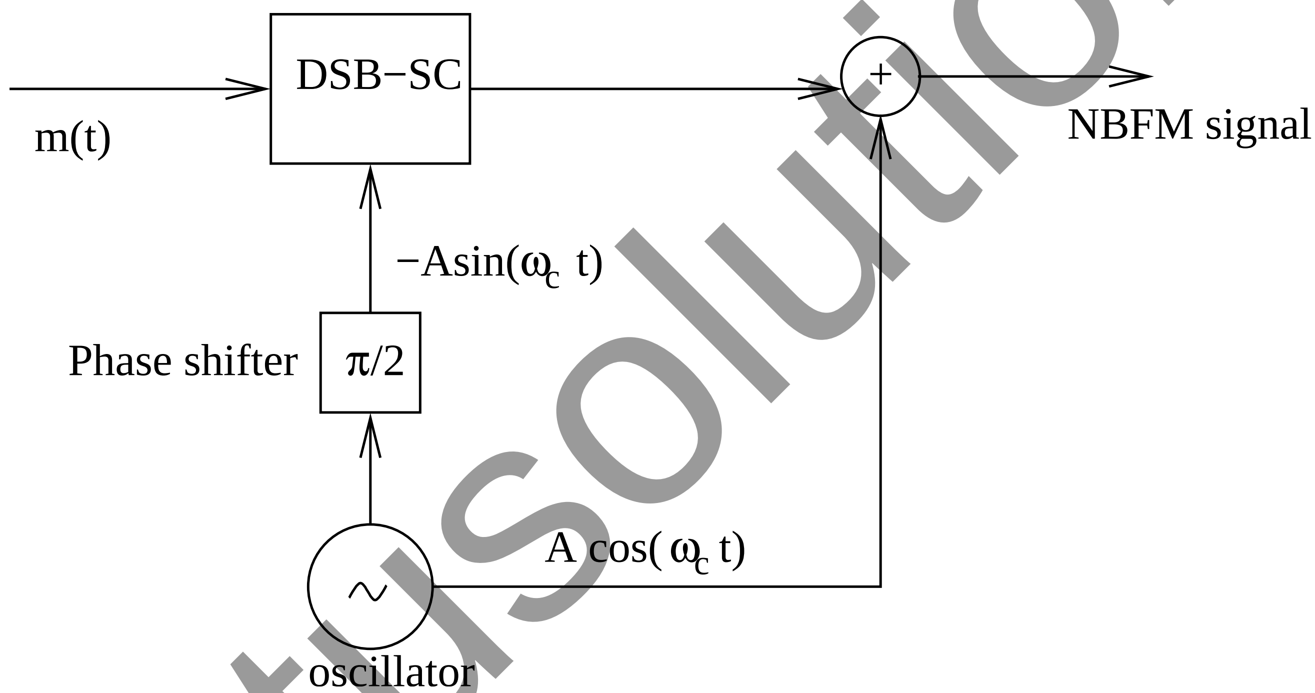


Figure 2: Generation of NBFM signal

– Wide-Band FM (WBFM)

A WBFM signal has theoritically infinite bandwidth. Spectrum calculation of WBFM signal is a tedious process. For, practical applications however the Bandwidth of a WBFM signal is calculated as follows:

Let $m(t)$ be bandlimited to BHz and sampled adequately at $2B$Hz. If time period $T = 1/2B$ is too small, the signal can be approximated by sequence of pulses as shown in Figure **??**

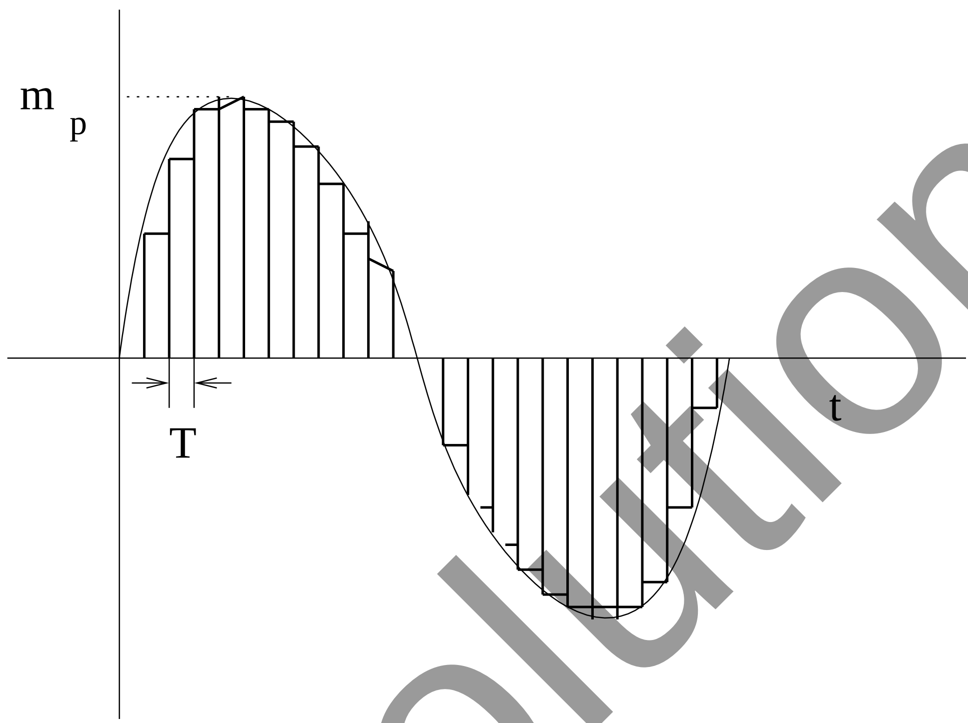


Figure 3: Approximation of message signal

If tone modulation is considered, and the peak amplitude of the sinusoid is m_p, the minimum and maximum frequency deviations will be $\omega_c - k_f m_p$ and $\omega_c + k_f m_p$ respectively. The spread of pulses in frequency domain will be $\frac{2\pi}{T} = 4\pi B$

as shown in Figure **??**

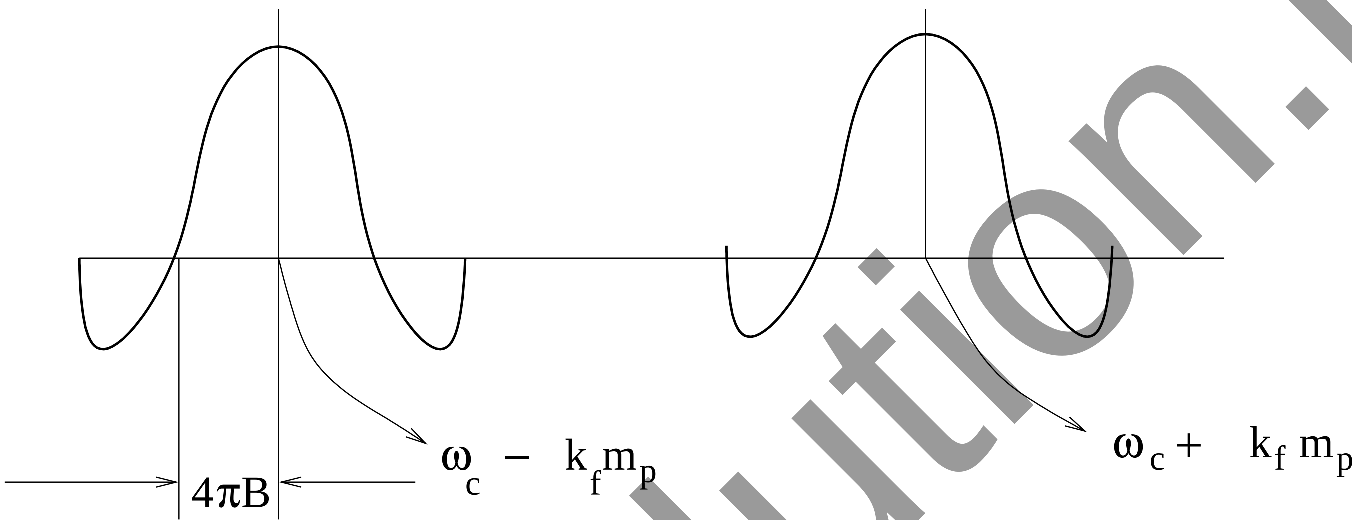


Figure 4: Bandwidth calculation of WBFM signal

Therefore, total BW is $2k_f m_p + 8\pi B$ and if frequency deviation is considered

$$BW_{fm} = \frac{1}{2\pi}(2k_f m_p + 8\pi B)$$

$$BW_{fm} = 2(\Delta f + 2B)$$

* The bandwidth obtained is higher than the actual value. This is due to the staircase approximation of $m(t)$.
* The bandwidth needs to be readjusted. For NBFM, k_f is very small an d hence Δf is very small compared to B. This implies

$$B_{fm} \approx 4B$$

  But the bandwidth for NBFM is the same as that of AM which is $2B$
* A better bandwidth estimate is therefore:

$$\begin{aligned} BW_{fm} &= 2(\Delta f + B) \\ BW_{fm} &= 2(\frac{k_f m_p}{2\pi} + B) \end{aligned}$$

  This is also called *Carson's Rule*

– Demodulation of FM signals

Let $\Phi_{fm}(t)$ be an FM signal.

$$\Phi_{fm}(t) = \left[A\cos(\omega_c t + k_f \int_0^t m(\alpha)\, d\alpha)\right]$$

This signal is passed through a differentiator to get

$$\dot{\Phi}_{fm}(t) = A\left(\omega_c + k_f m(t)\right)\sin\left(\omega_c t + k_f \int_0^t m(\alpha)\, d\alpha\right)$$

If we observe the above equation carefully, it is both amplitude and frequency modulated.

Hence, to recover the original signal back an envelope detector can be used. The envelope takes the form (see Figure **??**):

$$Envelope = A\,(\omega_c + k_f m(t))$$

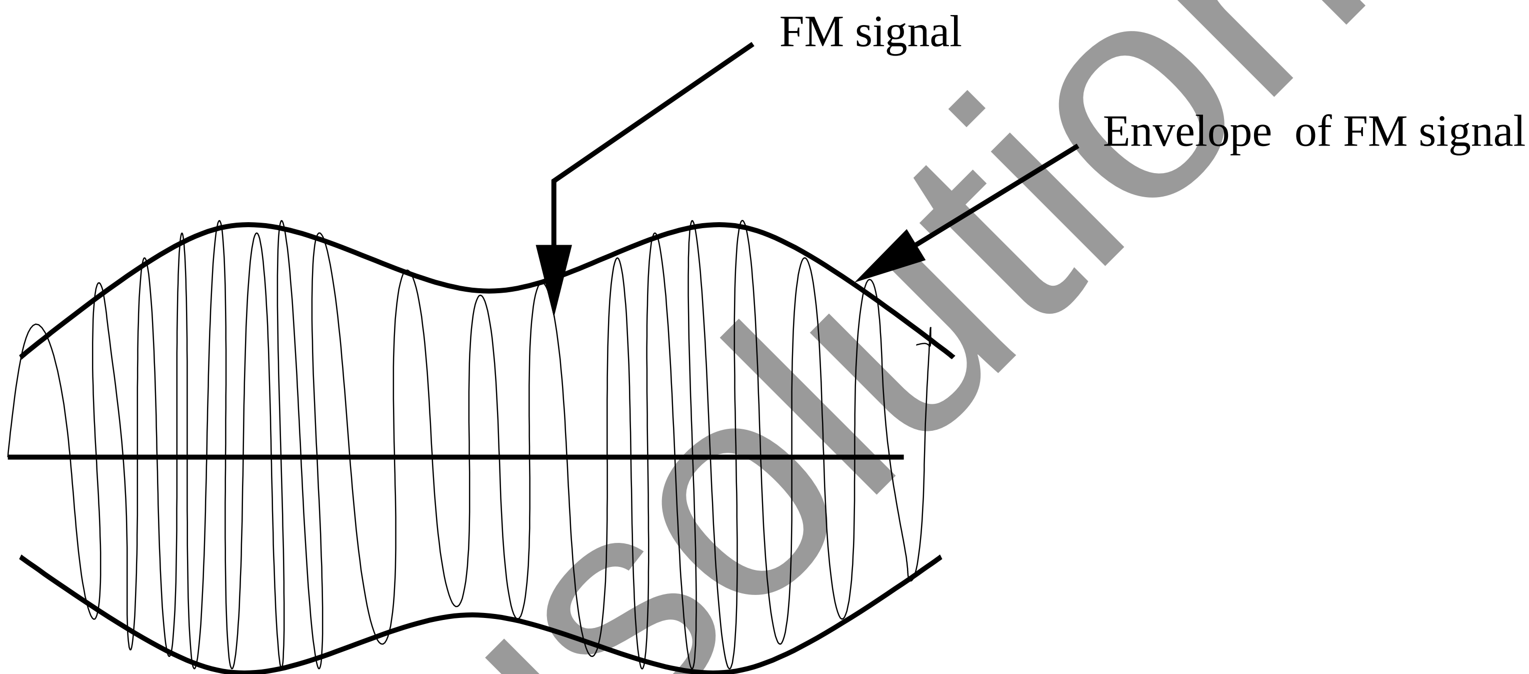


Figure 5: FM signal - both Amplitude and Frequency Modulation

The block diagram of the demodulator is shown in Figure **??**

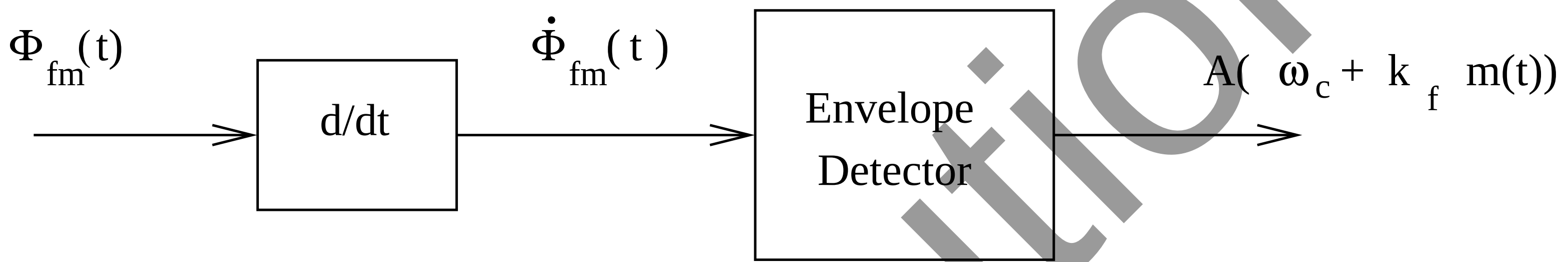


Figure 6: Demodulation of an FM signal

- The analysis for *Phase Modulation* is identical.
  - Analysis of bandwidth in PM

$$\begin{aligned}
\omega_i &= \omega_c + k_p m'(t) \\
m'_p &= [m'(t)]_{max} \\
\Delta\omega &= k_p m_p \\
BW_{pm} &= 2(\Delta f + B) \\
BW_{pm} &= 2(\frac{k_p m'_p}{2\pi} + B)
\end{aligned}$$

– The difference between FM and PM is that the bandwidth is independent of signal bandwidth in FM while it is strongly dependent on signal bandwidth in PM. [a]

[a]owing to the bandwidth being dependent on the peak of the *derivative of m(t)* rather than $m(t)$ itself

Angle Modulation: An Example

- An angle-modulated signal with carrier frequency $\omega_c = 2\pi \times 10^6$ is described by the equation:

$$\phi_{EM}(t) = 12\cos(\omega_c t + 5\sin 1500t + 10\sin 2000\pi t)$$

1. Determine the power of the modulating signal.
2. What is Δf?
3. What is β?
4. Determine $\Delta\phi$, the phase deviation.
5. Estimate the bandwidth of $\phi_E M(t)$?

1. $P = 12^2/2 = 72\ units$
2. Frequency deviation Δf, we need to estimate the instantaneous frequency:

$$\omega_i = \frac{d}{dt}\theta(t) = \omega_c + 7,500\cos 1500t + 20,000\pi t$$

The deviation of the carrier is $\Delta\omega = 7,500\cos 1500t + 20,000\pi t$. When the two sinusoids add in phase, the maximum value will be $7,500 + 20,000\pi$ Hence $\Delta f = \frac{\Delta\omega}{2\pi} = 11,193.66Hz$

3. $\beta = \frac{\Delta f}{B} = \frac{11,193.66}{1000} = 11.193$
4. The angle $\theta(t) = \omega_c t + 5\sin 1500t + 10\sin 2000\pi t)$. The maximum angle deviation is 15, which is the phase deviation.
5. $B_{EM} = 2(\Delta f + B) = 24,387.32\ Hz$

Noise Analysis - AM, FM

The following assumptions are made:

- Channel model
  - distortionless
  - Additive White Gaussian Noise (AWGN)
- Receiver Model (see Figure 1)
  - ideal bandpass filter
  - ideal demodulator

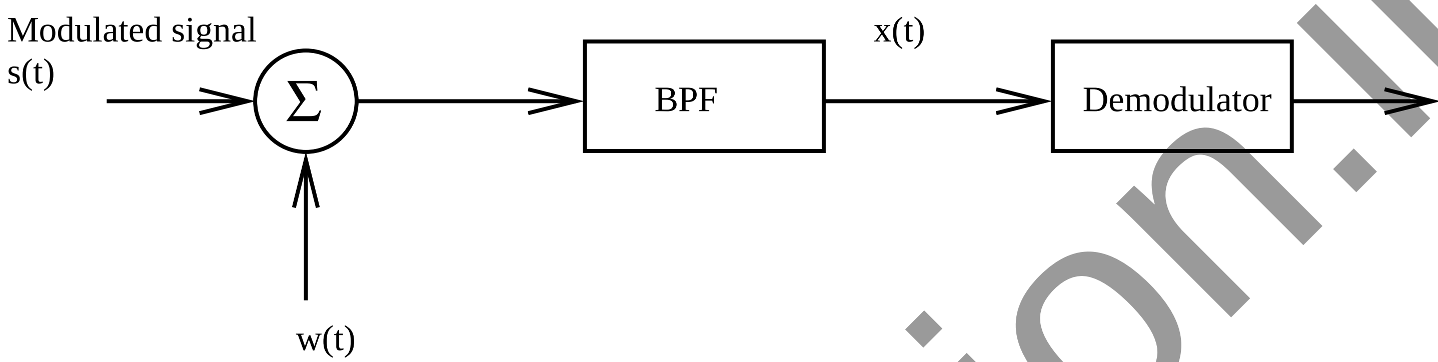


Figure 1: The Receiver Model

- BPF (Bandpass filter) - bandwidth is equal to the message bandwidth B
- midband frequency is ω_c.

Power Spectral Density of Noise

- $\frac{N_0}{2}$, and is defined for both positive and negative frequency (see Figure 2).
- N_0 is the average power/(unit BW) at the front-end of the

receiver in AM and DSB-SC.

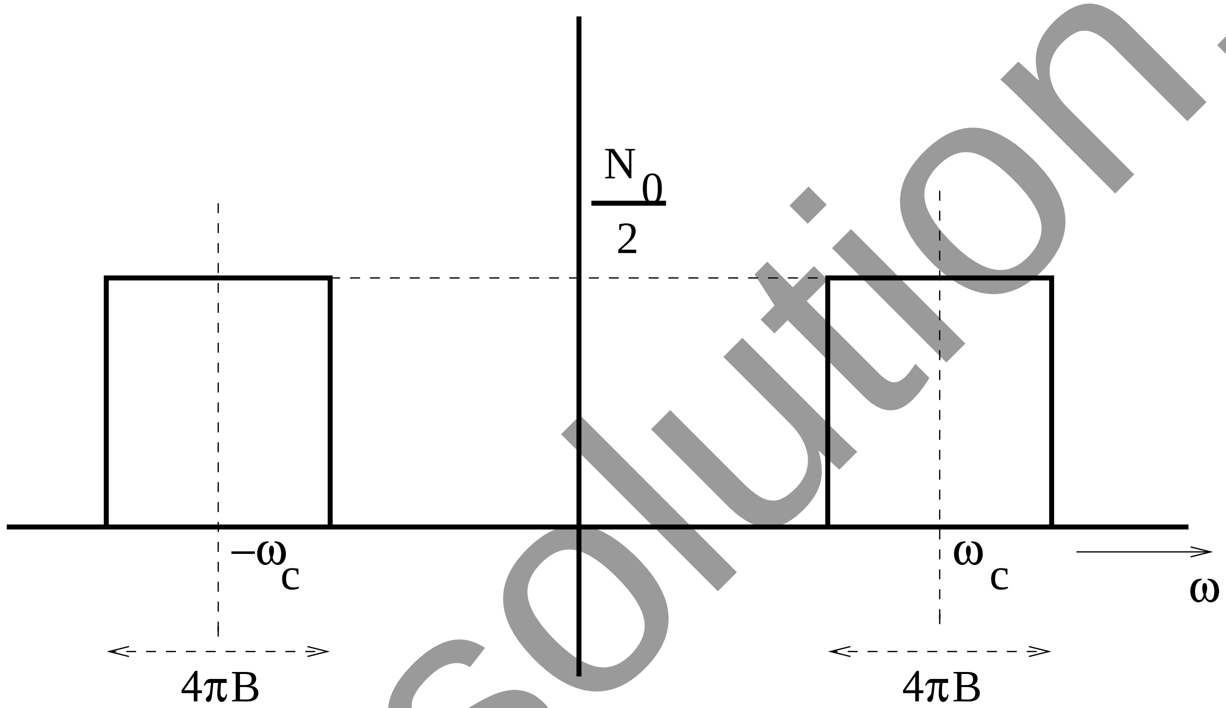


Figure 2: Bandlimited noise spectrum

The filtered signal available for demodulation is given by:

$$\begin{aligned} x(t) &= s(t) + n(t) \\ n(t) &= n_I(t)\cos\omega_c t \\ &\quad -n_Q(t)\sin\omega_c t \end{aligned}$$

$n_I(t)\cos\omega_c t$ is the in-phase component and

$n_Q(t)\sin\omega_c t$ is the quadrature component.

$n(t)$ is the representation for narrowband noise.

There are different measures that are used to define the Figure of Merit of different modulators:

- Input SNR:

$$(SNR)_I = \frac{Average\ power\ of\ modulated\ signal\ s(t)}{Average\ power\ of\ noise}$$

- Output SNR:

$$(SNR)_O = \frac{Average\ power\ of\ demodulated\ signal\ s(t)}{Average\ power\ of\ noise}$$

The Output SNR is measured at the receiver.

- Channel SNR:

$$(SNR)_C = \frac{Average\ power\ of\ modulated\ signal\ s(t)}{Average\ power\ of\ noise\ in\ message\ bandwidth}$$

- Figure of Merit (FoM) of Receiver:

$$FoM = \frac{(SNR)_O}{(SNR)_C}$$

To compare across different modulators, we assume that (see Figure 3):

- The modulated signal $s(t)$ of each system has the same average power
- Channel noise $w(t)$ has the same average power in the message bandwidth B.

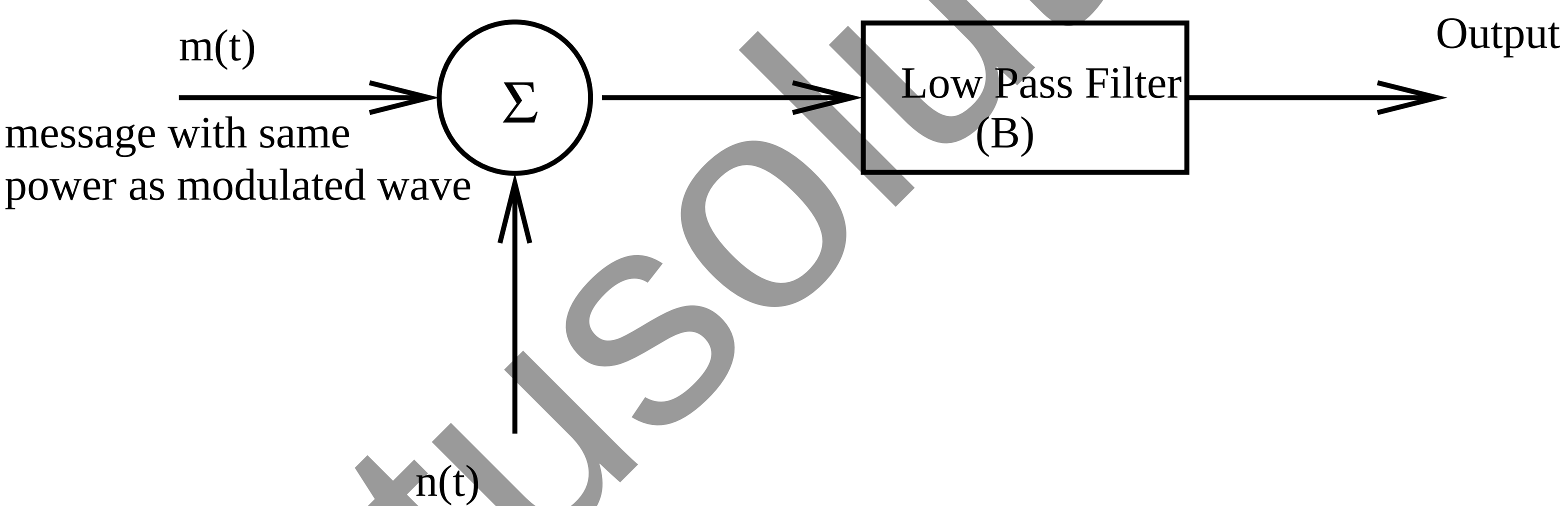


Figure 3: Basic Channel Model

Figure of Merit (FoM) Analysis

- DSB-SC (see Figure 4)

$$\begin{aligned} s(t) &= CA_c \cos(\omega_c t) m(t) \\ (SNR)_C &= \frac{A_c^2 C^2 P}{2BN_0} \\ P &= \int_{-2\pi B}^{+2\pi B} S_M(\omega) d\omega \\ x(t) &= s(t) + n(t) \\ & CA_c \cos(\omega_c t) m(t) \\ & + n_I(t) \cos \omega_c t + n_Q(t) \sin \omega_c t \end{aligned}$$

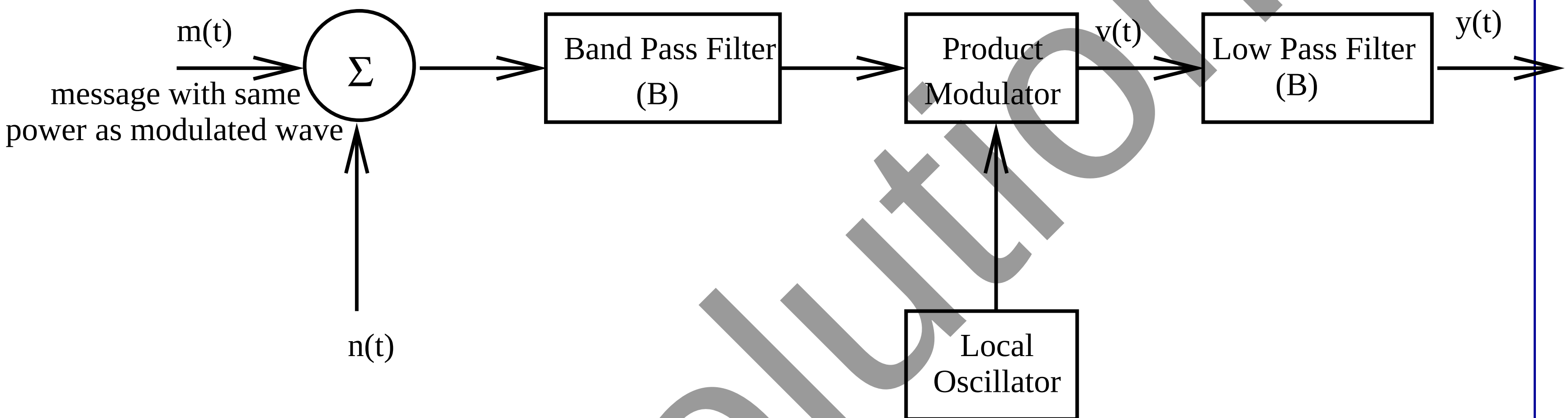


Figure 4: Analysis of DSB-SC System in Noise

The output of the product modulator is

$$
\begin{aligned}
v(t) &= x(t)\cos(\omega_c t) \\
&= \frac{1}{2}A_c m(t) + \frac{1}{2}n_I(t) \\
&\quad + \frac{1}{2}[CA_c m(t) + n_I(t)]\cos 2\omega_c t \\
&\quad - \frac{1}{2}n_Q(t)\sin 2\omega_c t
\end{aligned}
$$

The Low pass filter output is:

$$
= \frac{1}{2}A_c m(t) + \frac{1}{2}n_I(t)
$$

- $\Longrightarrow$ ONLY inphase component of noise $n_I(t)$ at the output
- $\Longrightarrow$ Quadrature component of noise $n_Q(t)$ is filtered at the output
- Band pass filter width $= 2B$

Receiver output is $\frac{n_I(t)}{2}$
Average power of $n_I(t)$ same as that $n(t)$

$$\begin{aligned} Average\ noise\ power &= (\frac{1}{2})^2 2BN_0 \\ &= \frac{1}{2}BN_0 \\ (SNR)_{O,DSB-SC} &= \frac{C^2A_c^2P/4}{BN_0/2} \\ &= \frac{C^2A_c^2P}{2BN_0} \\ FoM_{DSB-SC} &= \left(\frac{(SNR)_O}{(SNR)_C}\right)|_{DSB-SC} = 1 \end{aligned}$$

- Amplitude Modulation
  - The receiver model is as shown in Figure 5

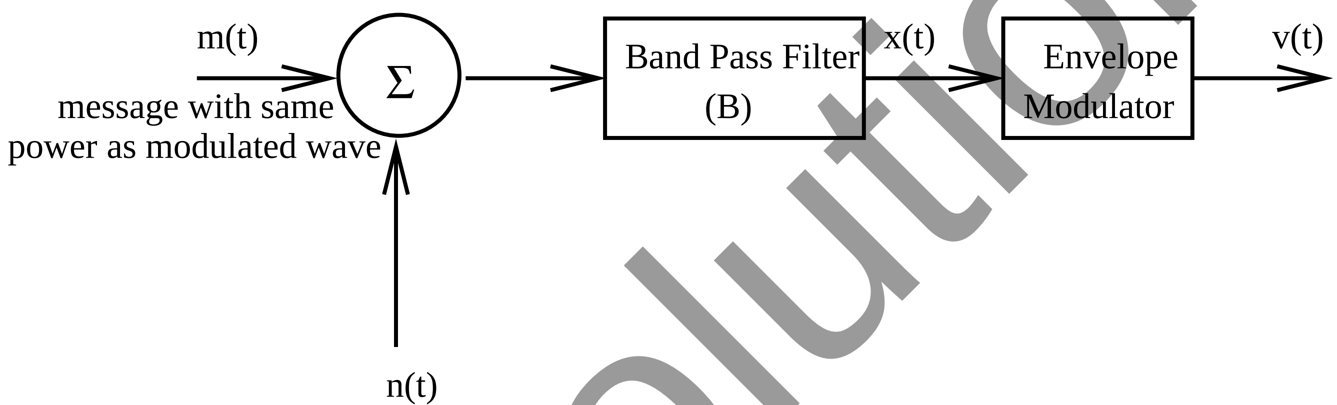


Figure 5: Analysis of AM System in Noise

$$
\begin{aligned}
s(t) &= A_c[1 + k_a m(t)] \cos \omega_c t \\
(SNR)_{C,AM} &= \frac{A_c^2(1 + k_a^2 P)}{2BN_0} \\
x(t) &= s(t) + n(t) \\
&= [A_c + A_c k_a m(t) + n_I(t)] \cos \omega_c t \\
&\quad -n_Q(t) \sin \omega_c t \\
y(t) &= \textit{envelope of } x(t) \\
&= \left[[A_c + A_c k_a m(t) + n_I(t)]^2 + n_Q^2(t)\right]^{\frac{1}{2}} \\
&\approx A_c + A_c k_a m(t) + n_I(t) \\
(SNR)_{O,AM} &\approx \frac{A_c^2 k_a^2 P}{2BN_0} \\
FoM_{AM} &= \left(\frac{(SNR)_O}{(SNR)_C}\right)|_{AM} = \frac{k_a^2 P}{1 + k_a^2 P}
\end{aligned}
$$

Thus the FoM_{AM} is always inferior to FoM_{DSB-SC}

– Frequency Modulation

* The analysis for FM is rather complex
* The receiver model is as shown in Figure 6

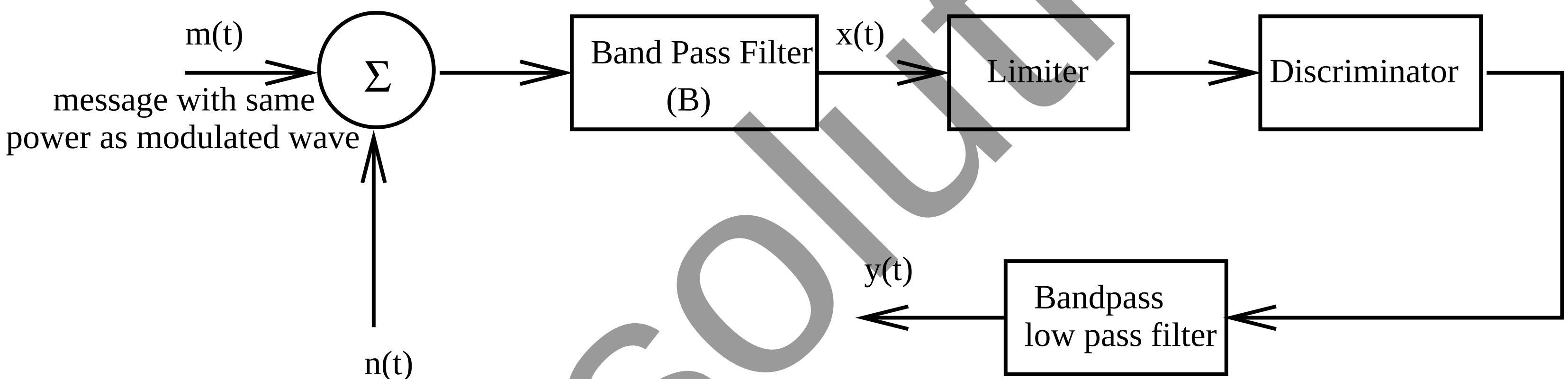


Figure 6: Analysis of FM System in Noise

$$(SNR)_{O,FM} = \frac{3A_c^2 k_f^2 P}{2N_0 B^3}$$

$$(SNR)_{C,FM} = \frac{A_c^2}{2BN_0}$$

$$FoM_{FM} = \left(\frac{(SNR)_O}{(SNR)_C}\right)|_{FM} = \frac{3k_f^2 P}{B^2}$$

The significance of this is that *when the carrier SNR is high, an increase in transmission bandwidth B_T provides a corresponding quadratic increase in output SNR or FoM_{FM}*

Digital Modulation

- Continuous-wave(CW) modulation (recap):
  - A parameter of a sinusoidal carrier wave is varied continuously in accordance with the message signal.
    - Amplitude
    - Frequency
    - Phase
- Digital Modulation:
  - Pulse Modulation: Analog pulse modulation: A periodic pulse train isused as a carrier. The following parameters of the pulse are modified in accordance with the message signal. Signal is transmitted at discrete intervals of time.
    - Pulse amplitude
    - Pulse width
    - Pulse duration

- Pulse Modulation: Digital pulse modulation: Message signal represented in a form that is discrete in both amplitude and time.
  - The signal is transmitted as a sequence of coded pulses
  - No continuous wave in this form of transmission

Analog Pulse Modulation

- Pulse Amplitude Modulation(PAM)
  - Amplitudes of regularly spaced pulses varied in proportion to the corresponding sampled values of a continuous message signal.
  - Pulses can be of a rectangular form or some other appropriate shape.
  - Pulse-amplitude modulation is similar to natural sampling, where the message signal is multiplied by a periodic train of rectangular pulses.
  - In natural sampling the top of each modulated rectangular pulse varies with the message signal, whereas in PAM it is maintained flat. The PAM signal is shown in Figure 1.

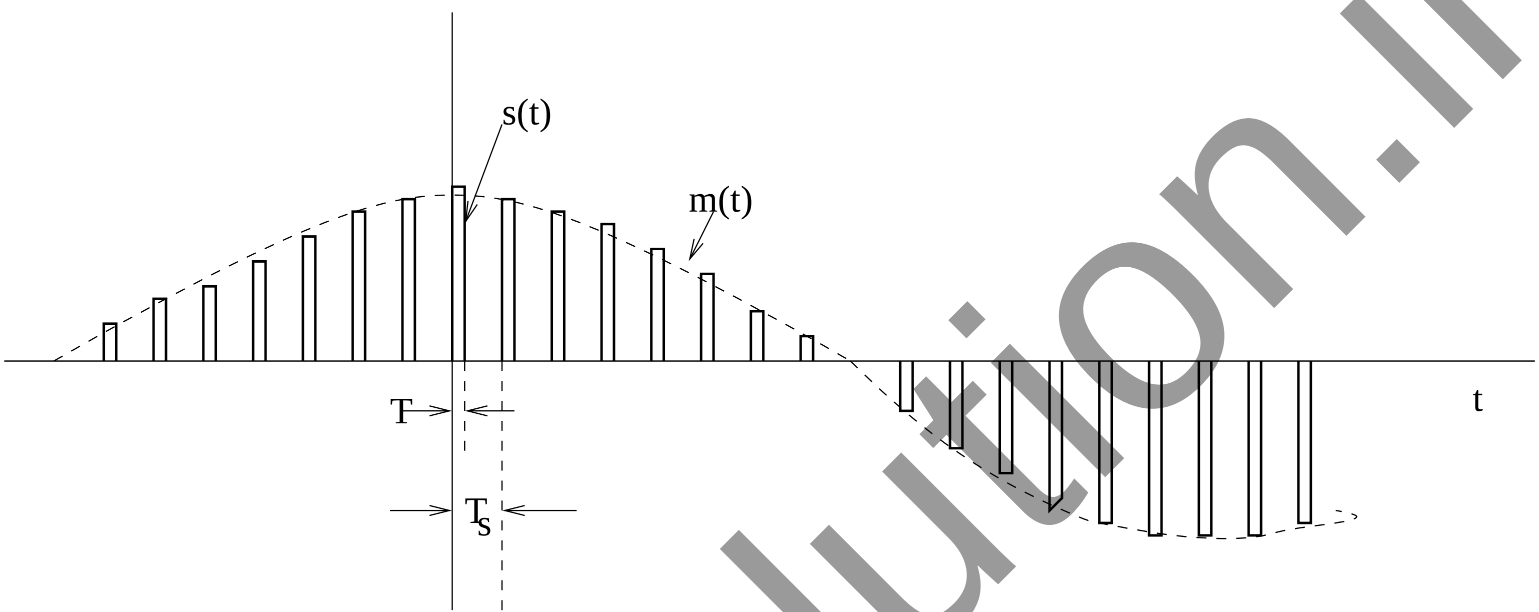


Figure 1: PAM signal

- Mathematical Analysis of PAM signals
  - Let $s(t)$ denote the sequence of the flat-top pulses generated in the manner described Figure 1. We may express the PAM signal as

$$s(t) = \sum_{n=-\infty}^{+\infty} m(nT_s)h(t - nT_s)$$

– $h(T_s)$ is a standard rectangular pulse of unit amplitude and duration T, defined as follows

$$h(t) = \begin{cases} = 1, & 0 \leq t \leq T \\ = \frac{1}{2}, & t = 0, t = T \\ = 0, \text{otherwise} \end{cases}$$

– The instantaneously sampled version of $m(t)$ is given by

$$m_\delta(t) = \sum_{n=-\infty}^{+\infty} m(nT_s)\delta(t - nT_s)$$

– Therefore, we get

$$m_\delta(t) * h(t) = \int_{-\infty}^{+\infty} m_\delta(\tau) h(t-\tau)\, d\tau$$

$$= \int_{-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} m(nT_s)\delta(\tau - nT_s) h(t-\tau)\, d\tau$$

$$= \sum_{n=-\infty}^{+\infty} m(nT_s) \int_{-\infty}^{+\infty} \delta(\tau - nT_s) h(t-\tau)\, d\tau$$

using the shifting property of the delta function, we obtain

$$s(t) = m_\delta(t) * h(t) = \sum_{n=-\infty}^{+\infty} m(nT_s) h(t - nT_s)$$

$$S(\omega) = M_\delta(\omega) * H(\omega)$$
$$= \frac{\omega_s}{2\pi} \sum_{k=-\infty}^{\infty} M(\omega - k\omega_s) H(\omega)$$

– Since, we use flat top samples, $H(\omega) = T sinc\left(\omega \frac{T}{2}\right) e^{-j\omega\frac{T}{2}}$. This results in distortion and a delay of $\frac{T}{2}$. To correct this the magnitude of the equalizer is chosen as $\dfrac{1}{T sinc\left(\omega \frac{T}{2}\right)}$.

– The message signal $m(t)$ can be recovered from the PAM signal $s(t)$ as shown in Figure 2.

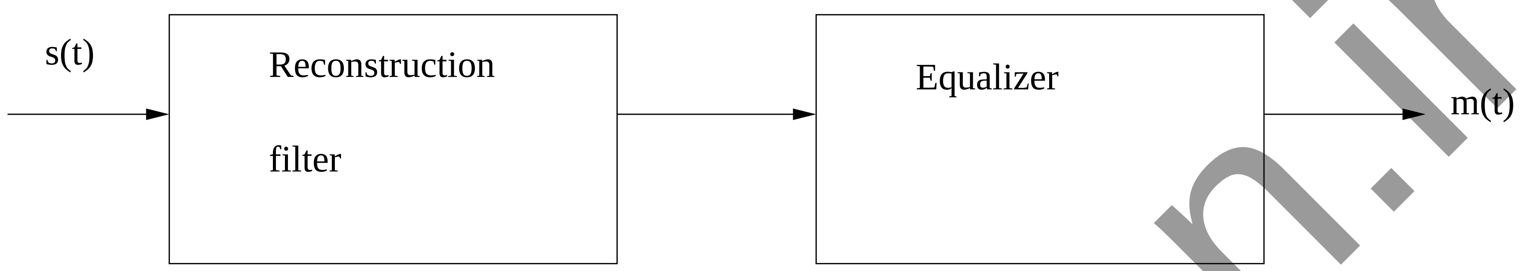


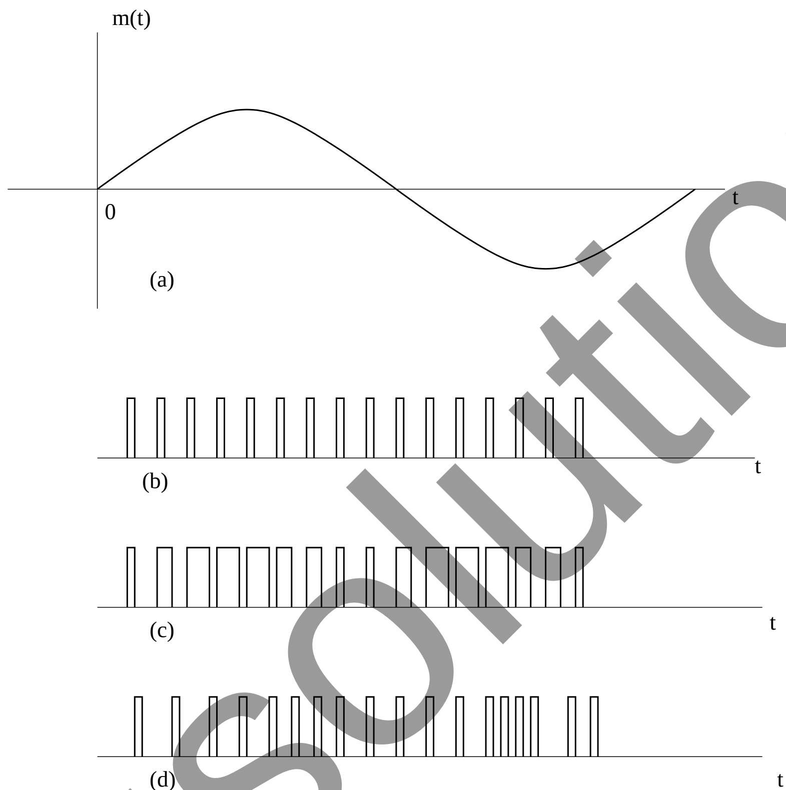
Figure 2: recovering message signal from PAM signal

- Other forms of Pulse Modulation

1. *Pulse-duration modulation*(PDM), also referred to as *Pulse-width modulation*, where samples of the message signal are used to vary the duration of the individual pulses in the carrier.
2. *Pulse-position modulation*(PPM), where the position of a pulse relative to its unmodulated time of occurence is varied in accordance with the message signal. It is similar to FM.

The other two types of modulation schemes are shown in

Figure 3.



(a) modulating wave
(b) Pulse carrier
(c) PDM wave
(d) PPM wave

Figure 3: other pulse modulation schemes

Pulse Digital Modulation

Pulse Code Modulation

- Discretisation of both time and amplitude.
- Discretisation of amplitude is called *quantisation.*
  - Quantisation involves conversion of an analog signal amplitude to discrete amplitude.
  - Quantisation process is memoryless and instantaneous.
  - A quantiser can be of uniform or non uniform.
    - Uniform quantiser: the representation levels are uniformly spaced
    - Nonuniform quantiser: the representation levels are non-uniformly spaced.

Uniform Quantisers

- Quantiser type: The quantiser characteristic can be of midtread or midrise quantizer. These two types are shown in Figure 4.

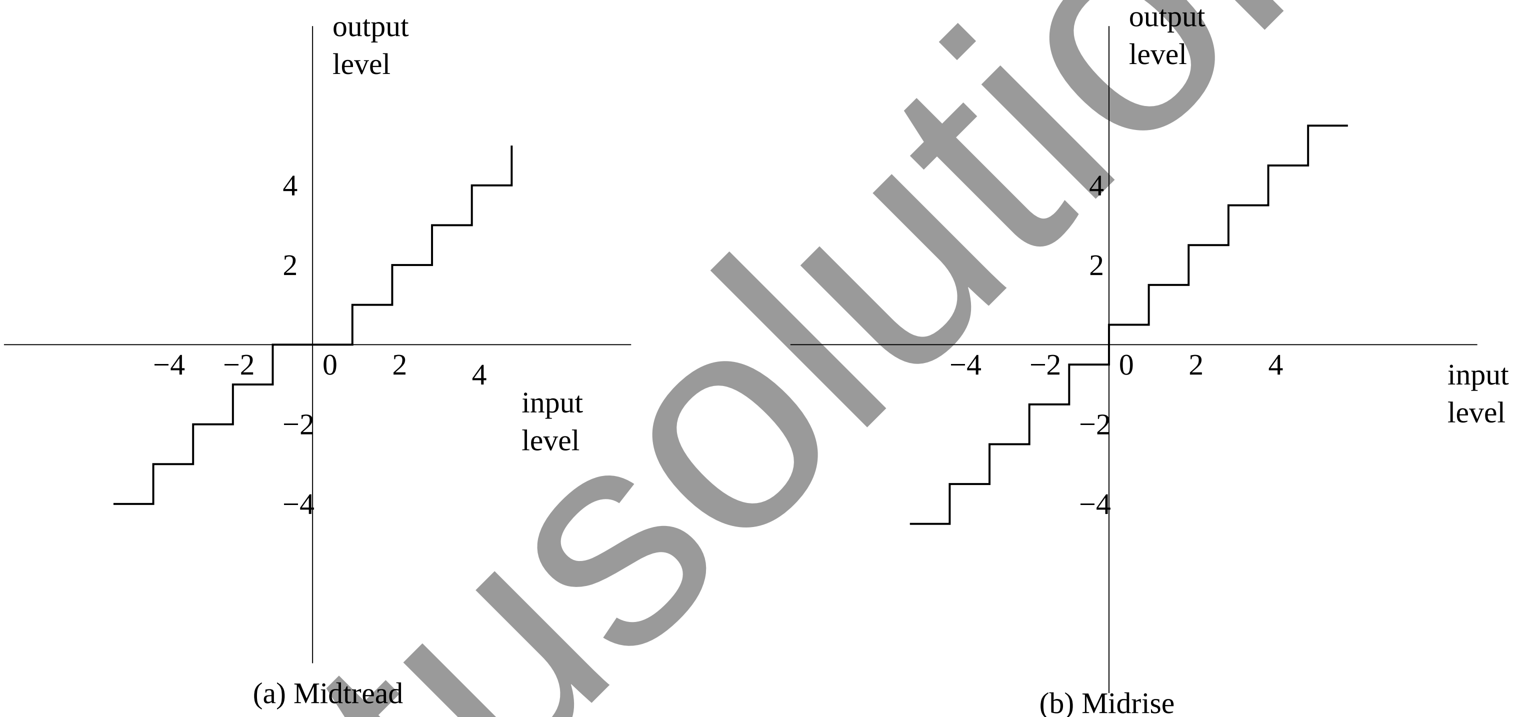


Figure 4: Different types of uniform quantisers

- Quantization Noise: Quantisation introduces an error defined as the difference between the input signal m and the output signal v. The error is called *Quantisation Noise.*
- We may therefore write

$$Q = M - V$$

- Analysis of error in uniform quantisation:
  - The input M has zero mean, and the quantiser assumed to be symmetric $\implies$ Quantiser output V and therefore the quantization error Q also will have a zero mean.
  - Consider that input m has continuous amplitude in the range $(-m_{max}, m_{max})$. Assuming a uniform quantiser of midrise type, we get the step size of the quantiser given by

$$\Delta = \frac{2m_{max}}{L}$$

where, L is the total number of representation levels.

- The quantization error Q will have its sample values bounded by $\frac{-\Delta}{2} \leq q \leq \frac{\Delta}{2}$.
- For small Δ, assume that the quantisation error Q is a uniformly distributed random variable. Thus,
- The probability density function of quantisation error Q can be given like this

$$f_Q(q) = \begin{cases} \frac{1}{\Delta}, & \frac{-\Delta}{2} \leq q \leq \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases}$$

- Since we assume mean of Q is zero, the variance is same as the mean square value.

$$
\begin{aligned}
\sigma_Q^2 &= E[Q^2] \\
&= \int_{-\frac{\Delta}{2}}^{+\frac{\Delta}{2}} q^2 f_Q(q)\, dq \\
&= \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{+\frac{\Delta}{2}} q^2\, dq \\
&= \frac{\Delta^2}{12}
\end{aligned}
$$

- Typically, the L-ary number k, denotes the kth representation level of the quantiser,
- It is transmitted to the receiver in binary form.
- Let R denote the number of bits per sample used in the construction of the binary code. We may then write

$$L = 2^R$$

$$R = log_2(L)$$

Hence, we get

$$\Delta = \frac{2m_{max}}{2^R}$$

also,

$$\sigma_Q^2 = \frac{1}{3}m_{max}^2 2^{-2R}$$

– Let P denote the average power of the message signal $m(t)$. We may then express the output signal to noise ratio of a uniform quantiser as

$$(SNR)_O = \frac{P}{\sigma_Q^2}$$

$$= \frac{3P}{m_{max}^2}.2^{2R}$$

- Example: Sinusoidal Modulating Signal

– Let the amplitude be A_m.

– Power of the signal, $P = \dfrac{A_m^2}{2}$.

– Range of the singal : $-A_m, A_m$.

– Hence, $\sigma_Q^2 = \frac{1}{3}A_m^2 2^{-2R}$.

– $\Longrightarrow (SNR)_O = \dfrac{\frac{A_m^2}{2}}{\frac{1}{3}A_m^2 2^{-2R}}$

– Thus, $(SNR)_O = \frac{3}{2}2^{2R}$ or

– $10\log_{10}(SNR)_O = 1.8 + 6RdB$.

- The major issue in using a uniform quantiser is that no effort is made to reduce quantisation power for a presumed set of levels.

Non Uniform Quantisers

The main advantages of using non-uniform quantizer are:
colorblack

1. Protection of weak passages over loud passages.
2. Enable uniform precision over the entire range of the voice signal.
3. Fewer steps are required in comparison with uniform quantizer.

A nonuniform quantiser is equivalent to passing the baseband signal through a compressor and then applying the compressed signal to a uniform quantizer. A particular form of compression law that is used in practice is μ - law, which is defined as follows

$$|v| = \frac{\log(1+\mu|m|)}{\log(1+\mu)}$$

where, m and v are the normalized input and output voltages and

μ is a positive constant.

Another compression law that is used in practice is the so called A - law defined by

$$|v| = \begin{cases} \frac{A|m|}{1+\log A}, & 0 \leq |m| \leq \frac{1}{A} \\ \frac{1+\log(A|m|)}{1+\log A}, & \frac{1}{A} \leq |m| \leq 1 \end{cases}$$

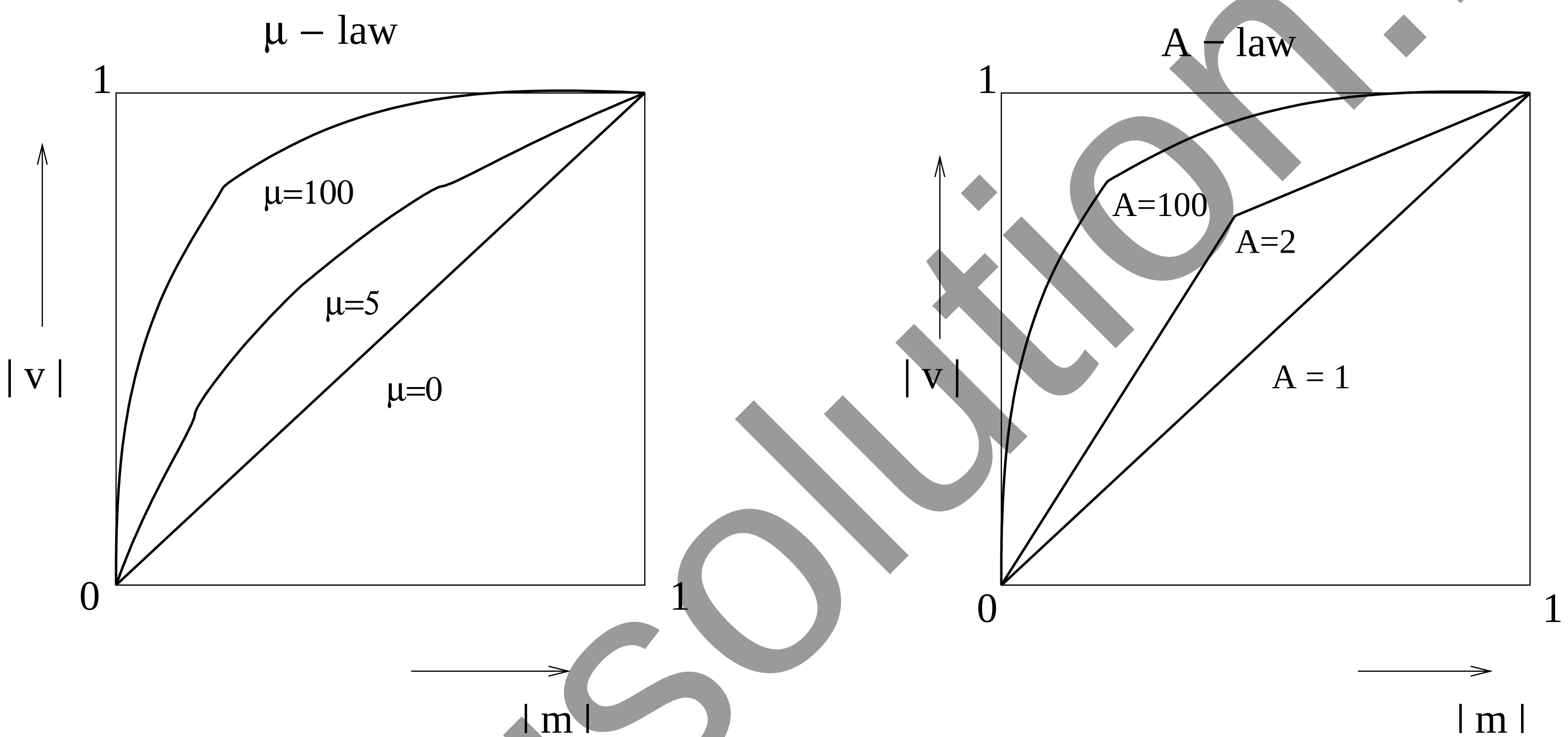


Figure 5: Compressions in non-uniform quantization

Differential Pulse code Modulation(DPCM)

- In DPCM, we transmit not the present sample $m[k]$, but $d[k]$ which is the difference between $m[k]$ and its predicted value $\hat{m}[k]$.
- At the receiver, we generate $\hat{m}[k]$ from the past sample values to which the received $d[k]$ is added to generate $m[k]$. Figure 6 shows the DPCM transmitter.

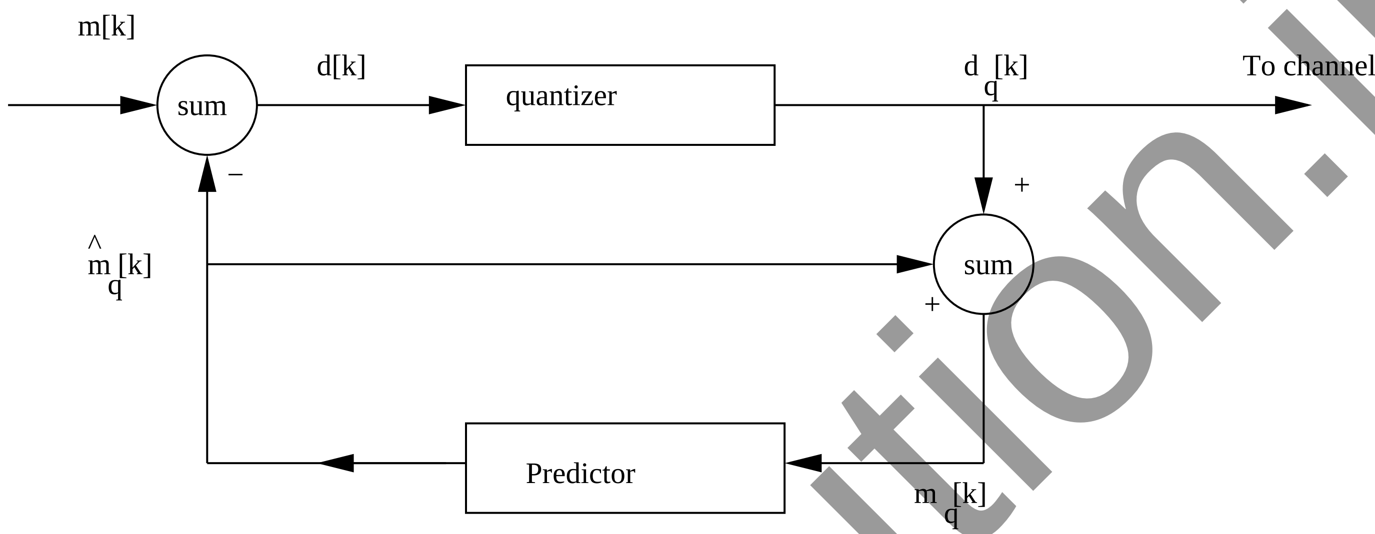


Figure 6: DPCM Transmitter

- Analysis of DPCM: If we take the quantised version, the predicted value as $\hat{m}_q[k]$ of $m_q[k]$. The difference

$$d[k] = m[k] - \hat{m}_q[k]$$

which is quantized to yield

$$d_q[k] = d[k] + q[k]$$

The predictor input $m_q[k]$ is

$$\hat{m}_q[k] = \hat{m}_q[k] + d_q[k]$$
$$= m[k] - d[k] + d_q[k]$$
$$= m[k] + q[k]$$

- The quantized signal $d_q[k]$ is now transmitted over the channel. At the receiver, $\hat{m}[k]$ is predicted from the previous samples and $d[k]$ is added to it to get $m[k]$. A DPCM receiver is shown in Figure 7

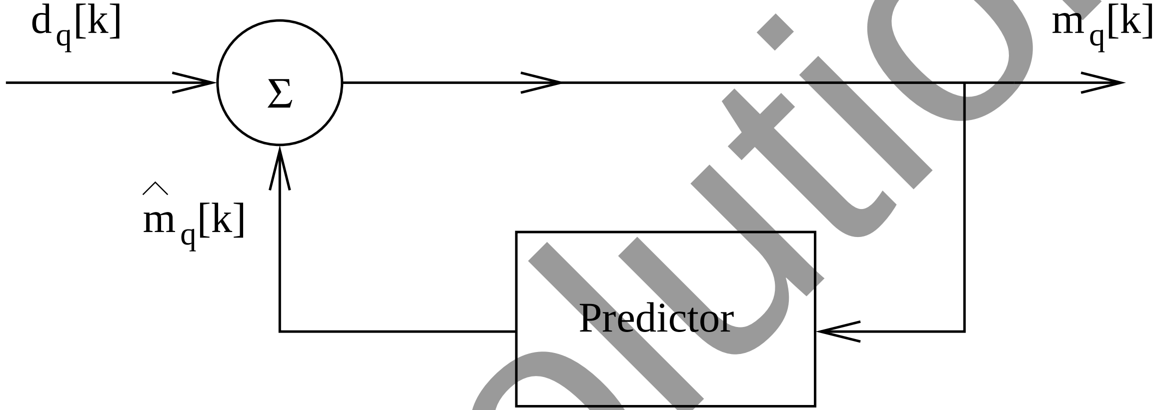


Figure 7: DPCM Receiver

Delta Modulation

- Delta Modulation uses a first order predictor.
- It is a one bit DPCM.
- DM quantizer uses only two levels($L = 2$)
- The signal is oversampled(atleast 4 times the Nyquist rate) to ensure better correlation among the adjacent samples.

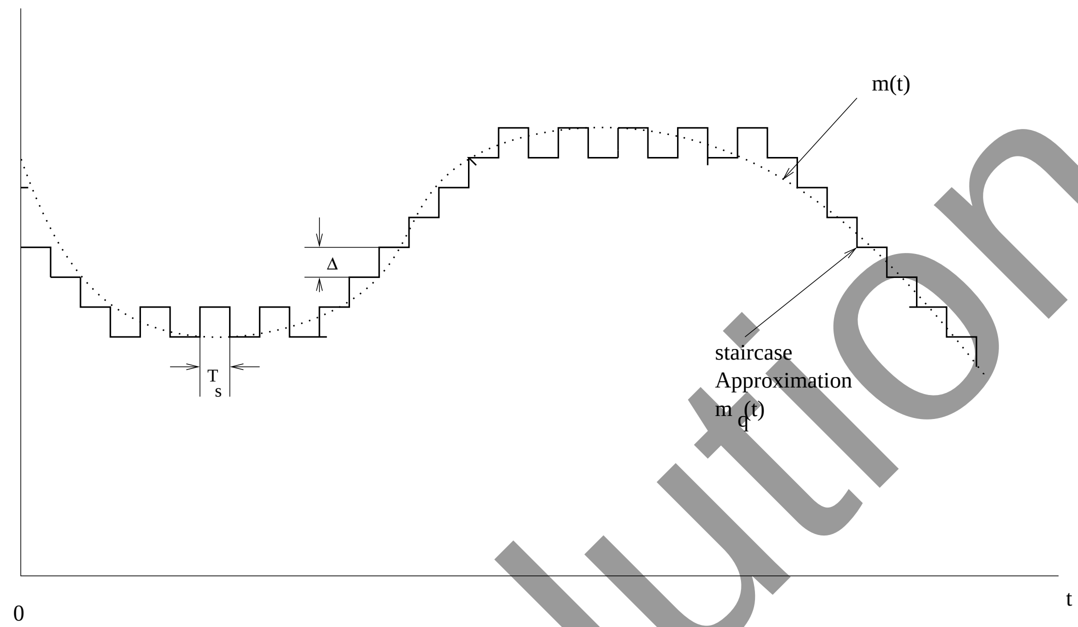


Figure 8: Delta Modulation

- DM provides a staircase approximation to the oversampled version of the message signal.
- The difference between the input and the approximation is quantised into only two levels, namely, $\pm\Delta$, corresponding to

positive and negative differences.

- For this approximation to work, the signal should *not change rapidly.*
- Analysis of Delta Modulation:
  - Let $m(t)$ denote the input signal, and $m_q(t)$ denote its staircase approximation. For convenience of presentation, we adopt the following notation that is commonly used in the digital signal processing literature:

$$m[n] = m(nT_s), n = 0, \pm 1, \pm 2, \cdots$$

  T_s is the sampling period and $m(nT_s)$ is a sample of the signal $m(t)$ taken at time $t = nT_s$, and likewise for the samples of other continuous-time signals.
  - The basic principles of delta modulation is the following set

of discrete-time relations:

$$e[n] = m[n] - m_q[n-1]$$
$$e_q = \Delta sgn(e[n])$$
$$m_q[n] = m_q[n-1] + e_q[n]$$

– $e[n]$ is an error signal representing the difference between the present sample $m[n]$ of the input signal and the latest approximation $m_q[n-1]$ to it.

– $e_q[n]$ is the quantized version of $e[n]$,

– $sgn(.)$ is the signum function.

– The quantiser output $m_q[n]$ is coded to produce the DM signal.

– The transmitter and receiver of the DM is shown in Figure

9.

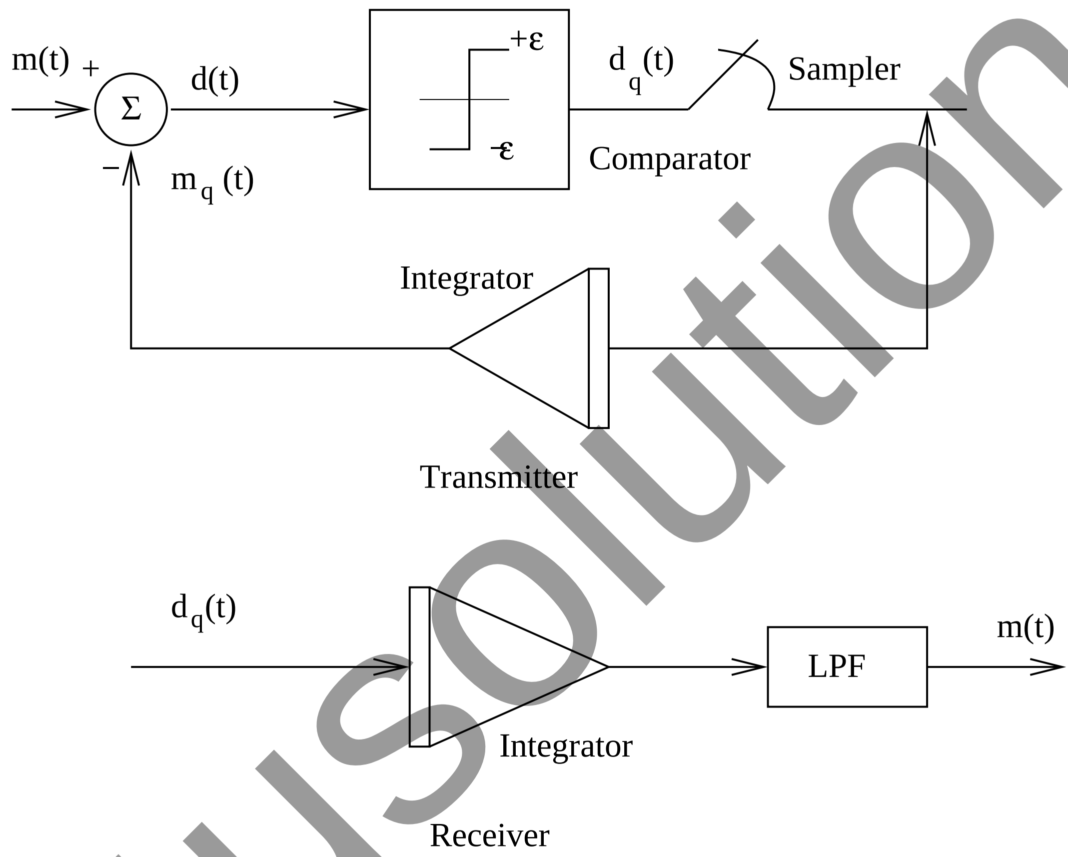


Figure 9: Transmitter and Receiver for Delta Modulation

Uncertainity, Information, and Entropy

Probabilistic experiment involves the observation of the output emitted by a discrete source during every unit of time. The source output is modeled as a discrete random variable, S,which takes on symbols form a fixed finite alphabet.

$$\mathcal{S} = s_0, s_1, s_2, \cdots, s_{k-1}$$

with probabilities

$$P(S = s_k) = p_k, k = 0, 1, \cdots, K - 1$$

We assume that the symbols emitted by teh source during successive signaling intervals are statistically independent. A source having the jproperties jist described is called *discrete memoryless source*, memoryless in the sense that the symbol emitted at any time is independent of previous choices.

We can define the amount of information contained in each symbols.

$$I(s_k) = log(\frac{1}{p_k})$$

Here, generally use log_2 since in digital communications we will be talking about bits. The above expression also tells us that when there is more uncertainity(less probability) of the symbol being occured then it conveys more information. Some properties of information are summarized here:

1. for certain event i.e, $p_k = 1$ the information it conveys is zero, $I(s_k) = 0$.
2. for the events $0 \leq p_k \leq 1$ the information is always $I(s_k) \geq 0$.
3. If for two events $p_k > p_i$, the information content is always $I(s_k) < I(s_i)$.
4. $I(s_k s_i) = I(s_k) + I(s_i)$ if s_k and s_i are statistically independent.

The amount of information $I(s_k)$ produced by the source during an arbitrary signalling interval depends on the symbol s_k emitted by the source at that time. Indeed, $I(s_k)$ is a discrete random variable that takes on the values $I(s_0), I(s_1), \cdots, I(s_{K-1})$ with probabilites $p_0, p_1, \cdots, p_{K-1}$ respectively. The mean of $I(s_k)$ over teh source alphabet $\mathcal{S}$ is given by

$$\begin{aligned} H(\mathcal{S}) &= E[I(s_k)] \\ &= \sum_{k=0}^{K_1} p_k I(s_k) \\ &= \sum_{k=0}^{K_1} p_k log_2(\frac{1}{p_k}) \end{aligned}$$

This important quantity is called entropy of a discrete memoryless source with source alphabet $\mathcal{S}$. It is a measure of average information content per source symbol.

Some properties of Entropy

The entropy $H(\mathcal{S})$ of a discrete memoryless source is bounded as follows:

$$0 \leq H(\mathcal{S}) \leq log_2(K)$$

where K is the radix of the alphabet $\mathcal{S}$ of the source. Furthermore, we may make two statements:

1. $H(\mathcal{S}) = 0$, if and only if the probability $p_k = 1$ for some k, and the remaining probabilities in teh set are all zero; this lower bound on entropy corresponds to no uncertainity.
2. $H(\mathcal{S}) = log_2(K)$, if and only if $p_k = \frac{1}{K}$ for all k; this upper bound on entropy corresponds to maximum uncertainity.

Shannon Source Coding Theorem

An important problem in communication is the efficient representation of data generated by a discrete source. The process by which this representation is accomplished is called source encoding.

Our primary interest is in the development of an efficient source encoder that satisfies two functional requirements:

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

We define the average code word length, $\bar{L}$, of the source encoder as

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

In physical terms, the parameter $\bar{L}$ represents the average number of bits per source symbol used in the source encoding process. Let $L_m in$ denote the minimum possible value of $\bar{L}$. We then define the coding efficency of the source encoder as

$$\eta = \frac{L_m in}{\bar{L}}$$

The source encoder is said to be efficient when η approaches unity. According to the source-coding theorem, the entropy $H(\mathcal{S})$ represents a fundamental limit on the average number of bits per source symbol necessary to represent a discrete memoryless source in that it can be made as small as, but no smaller than, the entropy $H(\mathcal{S})$. Thus with $L_m in = H(\mathcal{S})$, we may rewrite the efficiency of a source encoder in terms of the entropy $H(\mathcal{S})$ as

$$\eta = \frac{H(S)}{\bar{L}}$$

Information Theory and Source Coding

- Scope of Information Theory
  1. Determine the irreducible limit below which a signal cannot be compressed.
  2. Deduce the ultimate transmission rate for reliable communication over a noisy channel.
  3. Define Channel Capacity - the intrinsic ability of a channel to convey information.

  The basic setup in Information Theory has:

  – a source,

  – a channel and

  – destination.

  The output from source is conveyed through the channel and received at the destination. The *source* is a random variable S, which takes symbols from a finite alphabet i.e.,

$$S = \{s_0, s_1, s_2, \cdots, s_{k-1}\}$$

with probabilities

$$P(S = s_k) = p_k \ where \ k = 0, 1, 2, \cdots, k-1$$

and

$$\sum_{k=0}^{k-1} p_k = 1$$

The following assumptions are made about the source

1. Source generates symbols that are statistically independent.
2. Source is memoryless i.e., the choice of present symbol does not depend on the previous choices.

- Information: Information is defined as

$$\mathcal{I}(s_k) = \log_{base}\left(\frac{1}{p_k}\right)$$

- In digital commmunication, data is binary, the '*base*' is always 2.

Properties of Information

1. Information conveyed by a deterministic event is nothing i.e.,

$$\mathcal{I}(s_k) = 0, \; for \; p_k = 1^{a}.$$

2. Information is always positive i.e.,

$$\mathcal{I}(s_k) \geq 0 \; for \; p_k \leq 1$$

[a] Give examples about deterministic events, sun rising in the east as opposed to that of the occurence of the Tsunami

3. Information is never lost i.e.,

$$\mathcal{I}(s_k) \ is \ never \ negative$$

4. More information is conveyed by a *less* probable event than a *more* probable event

$$\mathcal{I}(s_k) > \mathcal{I}(s_i), \ for \ p_k < p_i$$ [a]

[a] a program that runs *without faults* as opposed to a program that gives a *segmentation fault*

Properties of Information

- Entropy: The Entropy($\mathcal{H}(s)$) of a source is defined as:

  the average information generated by a discrete memoryless source.

  Entropy defines the uncertainity in the source.

$$\mathcal{H}(s) = \sum_{k=0}^{K-1} p_k \mathcal{I}(s_k)$$

- Properties of Entropy:

  1. If K is the number of symbols generated by a source, then its entropy is bounded by

$$0 \leq \mathcal{H}(s) \leq \log_2 K$$

2. If the source has an inevitable symbol s_k for which $p_k = 1$ then

$$\mathcal{H}(s) = 0$$

3. If the source generates all the symbols equiprobably i.e., $p_k = \frac{1}{K}, \forall\ k$, then

$$\mathcal{H}(s) = \log_2 K$$

- Proof:
  –

$$\mathcal{H}(s) \geq 0$$

If the source has one symbol s_k which has $P(s = s_k) = p_k = 1$, then according to the properties, all other probabilities should be zero i.e., $p_i = 0, \forall\ i, i \neq k$. Therefore, from the definition of entropy

$$\mathcal{H}(s) = \sum_{k=0}^{K-1} p_k \log\left(\frac{1}{p_k}\right)$$

$$= p_k \log\left(\frac{1}{p_k}\right)$$

$$= 1 \log(1)$$

$$= 0$$

Therefore $\mathcal{H}(s) \geq 0$

–

$$\mathcal{H}(s) \leq \log_2 K$$

Consider a discrete memoryless source which has two different probability distributions

$$p_0, p_1, \cdots, p_k$$

and

$$q_0, q_1, \cdots, q_k$$

on the alphabet

$$S = s_0, s_1, \cdots, s_k$$

We begin with the equation

$$\sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right) = \frac{1}{\log_e 2} \sum_{k=0}^{K-1} p_k \log_e\left(\frac{q_k}{p_k}\right) \qquad (1)$$

To analyze Eq. 1, we consider the following inequality.

$$\log_e x \leq x - 1 \qquad (2)$$

Now, using Eq.2 in Eq. 1,

$$\begin{aligned}\sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right) &\leq \frac{1}{\log_e 2}\sum_{k=0}^{K-1} p_k\left(\frac{q_k}{p_k}-1\right)\\ &\leq \frac{1}{\log_e 2}\left(\sum_{k=0}^{K-1} q_k - \sum_{k=0}^{K-1} p_k\right)\\ &\leq \frac{1}{\log_e 2}(1-1) = 0\end{aligned} \quad (3)$$

$$\implies \sum_{k=0}^{K-1} p_k \log\left(\frac{q_k}{p_k}\right) \leq 0$$

Suppose, $q_k = \frac{1}{K}, \forall\ k$ then, $\sum_{k=0}^{K-1} q_k \log\left(\frac{1}{q_k}\right) = \log_2 K$ So, in Eq. 3 becomes

$$\sum_{k=0}^{K-1} p_k \log q_k - \sum_{k=0}^{K-1} p_k \log p_k \leq 0$$

$$\Longrightarrow \ -\sum_{k=0}^{K-1} p_k \log K \leq \sum_{k=0}^{K-1} p_k \log p_k$$

$$\Longrightarrow \ \sum_{k=0}^{K-1} p_k \log K \geq \sum_{k=0}^{K-1} p_k \log \left(\frac{1}{p_k}\right)$$

$$\Longrightarrow \ \log K \sum_{k=0}^{K-1} p_k \geq \sum_{k=0}^{K-1} p_k \log \left(\frac{1}{p_k}\right)$$

$$\Longrightarrow \sum_{k=0}^{K-1} p_k \log \left(\frac{1}{p_k}\right) \leq \log_2 K$$

Therefore $\mathcal{H}(s) \leq \log_2 K$

Entropy of a binary memoryless channel

Consider a discrete memoryless binary source shown defined on the alphabet $S = \{0, 1\}$. Let the probabilities of symbols 0 and 1 be p_0 and $1 - p_0$ respectively.

The entropy of this channel is given by

$$\mathcal{H}(s) = -p_0 \log_2 p_0 - (1 - p_0) \log_2 (1 - p_0)$$

According to the properties of Entropy, $\mathcal{H}(s)$ has a maximum value when $p_0 = \frac{1}{2}$ as demonstrated in Figure 1.

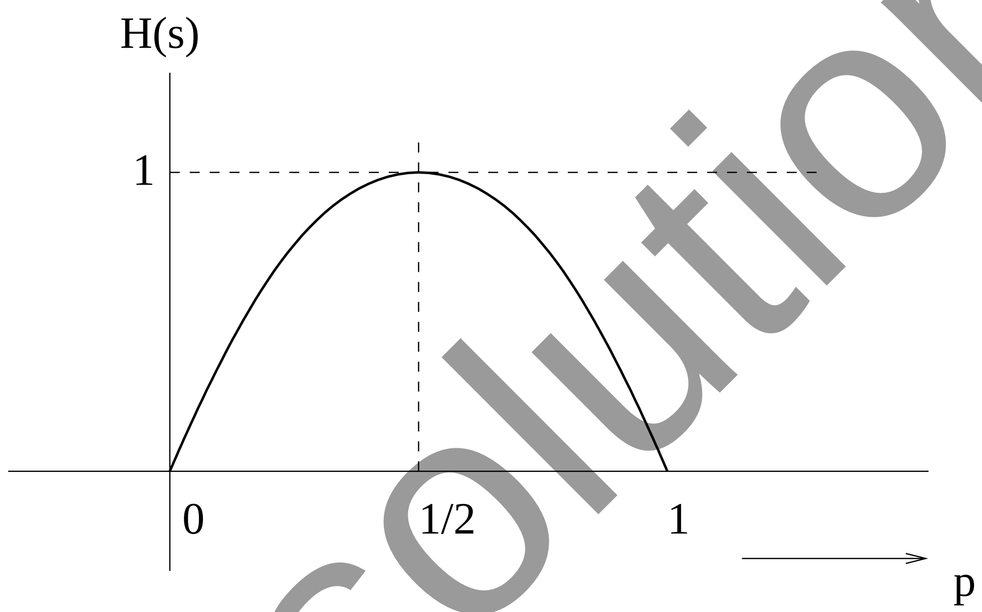


Figure 1: Entropy of a discrete memoryless binary source

Channel Coding

Channel Coding is done to ensure that the signal transmitted is recovered with very low probability of error at the destination.

Let X and Y be the random variables of symbols at the source and destination respectively. The description of the channel is shown in the Figure. 1

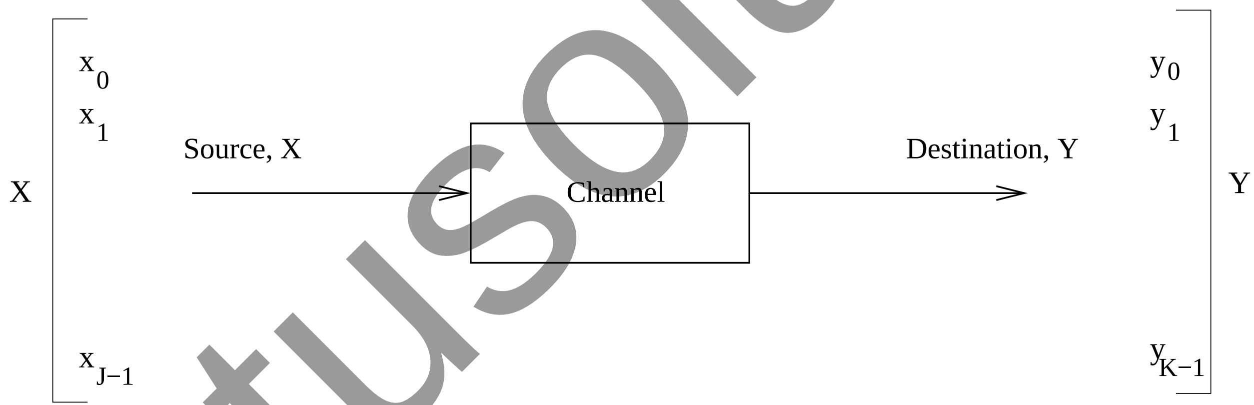


Figure 1: Description of a channel

The channel is described by a set of transition probabilities

$$P(Y = y_k | X = x_j) = p(y_k | x_j), \forall j, k$$

such that

$$\sum_{k=0}^{K-1} p(y_k | x_j) = 1, \forall j.$$

The joint probability is now given by

$$\begin{aligned} p(x_j, y_k) &= P(X = x_j, Y = y_k) \\ &= P(Y = y_k | X = x_j) P(X = x_j) \\ &= p(y_k | x_j) p(x_j) \end{aligned}$$

- Binary Symmetric Channel
  - A discrete memoryless channel with $J = K = 2$.
  - The Channel has two input symbols($x_0 = 0, x_1 = 1$) and two output symbols($y_0 = 0, y_1 = 1$).
  - The channel is symmetric because the probability of receiving a 1 if a 0 is sent is same as the probability of receiving a 0 if a 1 is sent.
  - The conditional probability of error is denoted by p. A binary symmetric channel is shown in Figure. 2 and its transition probability matrix is given by

$$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

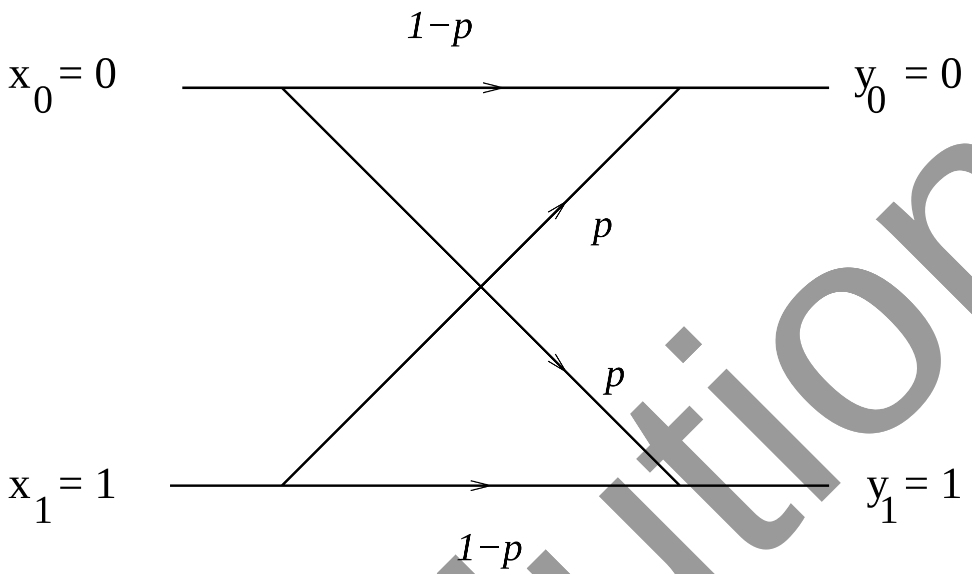


Figure 2: Binary Symmetric Channel

- Mutual Information

If the output Y is the noisy version of the channel input X and $H(\mathcal{X})$ is the uncertainity associated with X, then the uncertainity about X after observing Y, $H(\mathcal{X}|\mathcal{Y})$ is given by

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

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j|y_k)p(y_k)log_2\left[\frac{1}{p(x_j|y_k)}\right] \quad (2)$$

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k)log_2\left[\frac{1}{p(x_j|y_k)}\right] \quad (3)$$

The quantity $H(\mathcal{X}|\mathcal{Y})$ is called *Conditional Entropy*. It is the amount of uncertainity about the channel input after the channel output is observed. Since $H(\mathcal{X})$ is the uncertainity in channel input before observing the output, $H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y})$ represents the uncertainity in channel input that is resolved by observing the channel output. This uncertainity measure is termed as *Mutual Information* of the channel and is denoted by

$I(\mathcal{X};\mathcal{Y})$.

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \tag{4}$$

Similarly,

$$I(\mathcal{Y};\mathcal{X}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) \tag{5}$$

- Properties of Mutual Information
  - Property 1:

    The mutual information of a channel is symmetric, that is

$$I(\mathcal{X};\mathcal{Y}) = I(\mathcal{Y};\mathcal{X}) \tag{6}$$

Proof:

$$H(\mathcal{X}) = \sum_{j=0}^{J-1} p(x_j) \log_2 \left[\frac{1}{p(x_j)} \right] \quad (7)$$

$$= \sum_{j=0}^{J-1} p(x_j) \log_2 \left[\frac{1}{p(x_j)} \right] \sum_{k=0}^{K-1} p(y_k|x_j) \quad (8)$$

$$= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(y_k|x_j) p(x_j) \log_2 \left[\frac{1}{p(x_j)} \right] \quad (9)$$

$$= \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left[\frac{1}{p(x_j)} \right] \quad (10)$$

Substituting Eq.3 and Eq.10 in Eq.4 and then combining, we obtain

$$I(\mathcal{X};\mathcal{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j, y_k)\log_2\left[\frac{p(x_j|y_k)}{p(x_j)}\right] \quad (11)$$

From Bayes' rule for conditional probabilities, we have

$$\frac{p(x_j|y_k)}{p(x_j)} = \frac{p(y_k|x_j)}{p(y_k)} \quad (12)$$

Hence, from Eq.11 and Eq.12

$$I(\mathcal{X};\mathcal{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j, y_k)\log_2\left[\frac{p(y_k|x_j)}{p(y_k)}\right] = I(\mathcal{Y};\mathcal{X}) \quad (13)$$

– Property 2:

The mutual is always non-negative, that is

$$I(\mathcal{X};\mathcal{Y}) \geq 0$$

Proof:

We know,

$$p(x_j|y_k) = \frac{p(x_j, y_k)}{p(y_k)} \tag{14}$$

Substituting Eq. 14 in Eq. 13, we get

$$I(\mathcal{X};\mathcal{Y}) = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left[\frac{p(x_j, y_k)}{p(x_j)p(y_k)}\right] = I(\mathcal{Y};\mathcal{X}) \tag{15}$$

Using the following fundamental inequality which we derived discussing the properties of Entropy,

$$\sum_{k=0}^{K-1} p_k \log_2 \left(\frac{q_k}{p_k} \right) \leq 0$$

Drawing the similarities between the right hand side of the above inequality and the left hand side of Eq. 13, we can conclude that

$$I(\mathcal{X};\mathcal{Y}) \geq 0$$

– Property 3:

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

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X},\mathcal{Y})$$

where, the joint entropy $H(\mathcal{X}, \mathcal{Y})$ is defined as

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

Proof:

$$H(\mathcal{X}, \mathcal{Y}) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{p(x_j)p(y_k)}{p(x_j, y_k)} \right) + \quad (16)$$

$$\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)p(y_k)} \right) \quad (17)$$

$$= I(\mathcal{X}; \mathcal{Y}) + \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)p(y_k)} \right) \quad (18)$$

But,

$$\sum_{j=0}^{J-1}\sum_{k=0}^{K-1} p(x_j, y_k) \log_2 \left(\frac{1}{p(x_j)p(y_k)} \right) \tag{19}$$

$$= \sum_{j=0}^{J-1} \log_2 \left(\frac{1}{p(x_j)} \right) \sum_{k=0}^{K-1} p(x_j, y_k) + \tag{20}$$

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

$$= \sum_{j=0}^{J-1} p(x_j) \log_2 \left(\frac{1}{p(x_j)} \right) + \sum_{k=0}^{K-1} p(y_k) \log_2 \left(\frac{1}{p(y_k)} \right) \tag{22}$$

$$= H(\mathcal{X}) + H(\mathcal{Y}) \tag{23}$$

Therefore, from Eq. 18 and Eq. 23, we have

$$H(\mathcal{X}, \mathcal{Y}) = -I(\mathcal{X}; \mathcal{Y}) + H(\mathcal{X}) + H(\mathcal{Y})$$

Problems

The following problems may be given as exercises.

1. Show that the mutual information is zero for a deterministic channel.
2. Prove that $I(\mathcal{X};\mathcal{Y}) = min(H(\mathcal{Y}), H(\mathcal{X}))$
3. Prove that $I(\mathcal{X};\mathcal{Y}) = min(\log(|\mathcal{Y}|), \log(|\mathcal{X}|))$

Source Coding

1. Source symbols encoded in binary
2. The average codelength must be reduced
3. Remove redundancy ⇒ reduces bit-rate

Consider a discrete memoryless source on the alphabet

$$S = \{s_0, s_1, \cdots, s_k\}$$

Let the corresponding probabilities be

$$\{p_0, p_1, \cdots, p_k\}$$

and codelengths be

$$\{l_0, l_1, \cdots, l_k\}.$$

Then, the average codelength(average number of bits per symbol) of the source is defined as

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

If L_{min} is the minimum possible value of $\bar{L}$, then the coding efficiency of the source is given by η.

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

For an efficient code η approaches unity.

The question: What is smallest average codelength that is possible?

The Answer: Shannon's source coding theorem

Given a discrete memoryless source of entropy $\mathcal{H}(s)$, the average codeword length $\bar{L}$ for any distortionless source encoding scheme is bounded by

$$\bar{L} \geq \mathcal{H}(s)$$

Since, $\mathcal{H}(s)$ is the fundamental limit on the average number of bits/symbol, we can say

$$L_{min} \geq \mathcal{H}(s)$$

$$\Longrightarrow \eta = \frac{\mathcal{H}(s)}{\bar{L}}$$

Data Compaction:

1. Removal of redundant information prior to transmission.
2. Lossless data compaction – no information is lost.
3. A source code which represents the output of a discrete memoryless source should be uniquely decodable.

Source Coding Schemes for Data Compaction

Prefix Coding

1. The *Prefix Code* is variable length source coding scheme where no code is the prefix of any other code.
2. The prefix code is a uniquely decodable code.
3. But, the converse is not true i.e., all uniquely decodable codes may not be prefix codes.

Table 1: Illustrating the definition of prefix code

Symbol	Prob.of Occurrence	Code I	Code II	Code III
s_0	0.5	0	0	0
s_1	0.25	1	10	01
s_2	0.125	00	110	011
s_3	0.125	11	111	0111

Table 2: Table is reproduced from S.Haykin's book on Communication Systems

From 1 we see that Code I is not a prefix code. Code II is a prefix code. Code III is also uniquely decodable but not a prefix code.

Prefix codes also satisfies Kraft-McMillan inequality which is given by

$$\sum_{k=0}^{K-1} 2^{-l_k} \leq 1$$

Kraft-McMillan inequality maps codewords to a binary tree as shown is Figure 1.

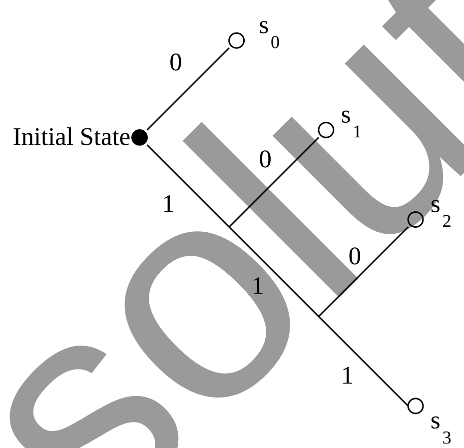


Figure 1: Decision tree for Code II

Given a discrete memoryless source of entropy $\mathcal{H}(s)$, a prefix code can be constructed with an average code-word length $\bar{l}$, which is bounded as follows:

$$\mathcal{H}(s) \leq \bar{(}L) < \mathcal{H}(s) + 1 \tag{1}$$

The left hand side of the above equation, the equality is satisfied owing to the condition that, any symbol s_k is emitted with the probability

$$p_k = 2^{-l_k} \tag{2}$$

where, l_k is the length of the codeword assigned to the symbol s_k. Hence, from Eq. 2, we have

$$\sum_{k=0}^{K-1} 2^{-l_k} = \sum_{k=0}^{K-1} p_k = 1 \tag{3}$$

With this condition, the Kraft-McMillan inequality tells that a prefix code can be constructed such that the length of the codeword assigned to source symbol s_k is $-\log_2 p_k$. Therefore, the average codeword length is given by

$$\bar{L} = \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}} \tag{4}$$

and the corresponding entropy is given by

$$\mathcal{H}(s) = \sum_{k=0}^{K-1} \left(\frac{1}{2^{l_k}}\right) \log_2(2^{l_k}) = \sum_{k=0}^{K-1} \frac{l_k}{2^{l_k}} \tag{5}$$

Hence, from Eq. 5, the equality condition on the leftside of Eq. 1, $\bar{L} = \mathcal{H}(s)$ is satisfied.

To prove the inequality condition we will proceed as follows:

Let $\bar{L_n}$ denote the average codeword length of the extended prefix code. For a uniquely decodable code, $\bar{L_n}$ is the smallest possible.

Huffman Coding

1. Huffman code is a prefix code
2. The length of codeword for each symbol is roughly equal to the amount of information conveyed.
3. The code need not be unique (see Figure 3)

A Huffman tree is constructed as shown in Figure. 3, (a) and (b) represents two forms of Huffman trees. We see that both schemes have same average length but different variances.

Variance is a measure of the variability in codeword lengths of a source code. It is defined as follows:

$$\sigma^2 = \sum_{k=0}^{K-1} p_k (l_k - \bar{L})^2 \tag{6}$$

where, p_k is the probability of kth symbol. l_k is the codeword length of kth symbol and $\bar{L}$ is the average codeword length. It is reasonable to choose the huffman tree which gives greater variace.

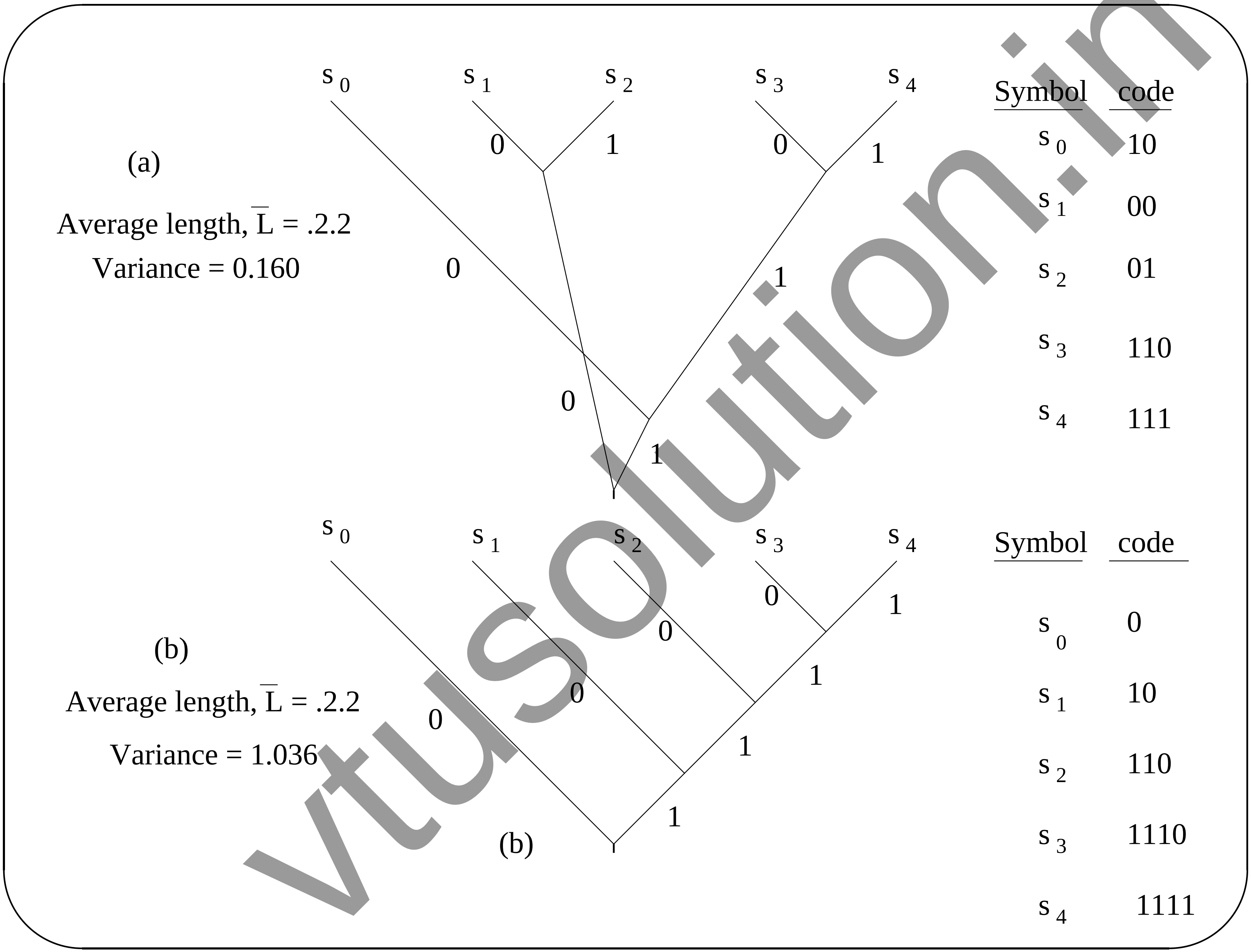


Figure 2: Huffman tree

Drawbacks:

1. Requires proper statistics.
2. Cannot exploit relationships between words, phrases etc.,
3. Does not consider redundancy of the language.

Lempel-Ziv Coding

1. Overcomes the drawbacks of Huffman coding
2. It is an adaptive and simple encoding scheme.
3. When applied to English text it achieves 55% in contrast to Huffman coding which achieves only 43%.
4. Encodes patterns in the text

This algorithm is accomplished by parsing the source data stream into segments that are the shortest subsequences not encountered previously. (see Figure 3 – the example is reproduced from S.Haykin's book on "Communication Systems.")

Let the input sequence be

000101110010100101.........

We assume that 0 and 1 are known and stored in codebook

subsequences stored : 0, 1
Data to be parsed: 000101110010100101.........

The shortest subsequence of the data stream encountered for the first time and not seen before is 00
subsequences stored: 0, 1, 00
Data to be parsed: 0101110010100101.........

The second shortest subsequence not seen before is 01; accordingly, we go on to write

Subsequences stored: 0, 1, 00, 01
Data to be parsed: 01110010100101.........

We continue in the manner described here until the given data stream has been completely parsed. The code book is shown below:

Numerical positions:	1	2	3	4	5	6	7	8	9
subsequences:	0	1	00	01	011	10	010	100	101
Numerical Representations:	–		11	12	42	21	41	61	62
Binary encoded blocks:			0010	0011	1001	0100	1000	1100	1101

Figure 3: Lempel-Ziv Encoding

Channel Coding

- Channel Capacity

  Channel Capacity, C is defined as

  'the maximum mutual information $I(\mathcal{X};\mathcal{Y})$ in any single use of the channel, where the maximization is over all possible input probability distributions $\{p(x_j)\}$ on X"

$$C = \max_{p(x_j)} I(\mathcal{X};\mathcal{Y}) \tag{1}$$

C is measured in bits/channel-use, or bits/transmission.

Example:

For, the binary symmetric channel discussed previously, $I(\mathcal{X};\mathcal{Y}$ will be maximum when $p(x_0) = p(x_1) = \frac{1}{2}$. So, we have

$$C = I(\mathcal{X};\mathcal{Y}|_{p(x_0)=p(x_1)=\frac{1}{2}} \tag{2}$$

Since, we know

$$p(y_0|x_1) = p(y_1|x_0) = p \tag{3}$$

$$p(y_0|x_0) = p(y_1|x_1) = 1 - p \tag{4}$$

Using the probability values in Eq. 3 and Eq. 4 in evaluating Eq. 2, we get

$$\begin{aligned} C &= 1 + p\log_2 p + (1-p)\log_2(1-p) \\ &\implies 1 - H(p) \end{aligned} \tag{5}$$

- Channel Coding Theorem:

  Goal: Design of channel coding to increase resistance of a digital communication system to channel noise.

  The channel coding theorem is defined as

  1. Let a discrete memoryless source
     - with an alphabet S
     - with an entropy $H(S)$
     - produce symbols once every T_s seconds
  2. Let a discrete memoryless channel
     - have capacity C
     - be used once every T_c seconds.
  3. Then if,

$$\frac{H(S)}{T_s} \le \frac{C}{T_c} \tag{6}$$

There exists a *coding scheme* for which the source output can be *transmitted over the channel* and be *reconstructed* with an *arbitrarily small probability of error.* The parameter $\frac{C}{T_c}$ is called *critical rate.*

4. Conversely, if

$$\frac{H(S)}{T_s} > \frac{C}{T_c} \tag{7}$$

it is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error.

Example:

Considering the case of a binary symmetric channel, the source entropy $H(S)$ is 1. Hence, from Eq. 6, we have

$$\frac{1}{T_s} \leq \frac{C}{T_c} \tag{8}$$

But the ratio $\frac{T_c}{T_s}$ equals the code rate, r of the channel encoder. Hence, for a binary symmetric channel, if $r \leq C$, then there exists a code capable of achieving an arbitrarily low probability of error.

Information Capacity Theorem:

The Information Capacity Theorem is defined as

'The information capacity of a continuous channel of bandwidth B hertz, perturbed by additive white Gaussian noise of power spectral density $\frac{N_0}{2}$ and limited in bandwidth to B, is given by

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right) \tag{9}$$

where P is the average transmitted power. **Proof:**

Assumptions:

1. band-limited, power-limited Gaussian channels.
2. A zero-mean stationary process $X(t)$ that is band-limited to B

hertz, sampled at Nyquist rate of $2B$ samples per second

3. These samples are transmitted in T seconds over a noisy channel, also band-limited to B hertz.

The number of samples, K is given by

$$K = 2BT \tag{10}$$

We refer to X_k as a sample of the transmitted signal. The channel output is mixed with additive white Gaussian noise(AWGN) of zero mean and power spectral density $N_0/2$. The noise is band-limited to B hertz. Let the continuous random variables Y_k, $k = 1, 2, \cdots, K$ denote samples of the received signal, as shown by

$$Y_k = X_k + N_k \tag{11}$$

The noise sample N_k is Gaussian with zero mean and variance given by

$$\sigma^2 = N_0 B \tag{12}$$

The transmitter power is limited; it is therefore

$$E[X_k^2] = P \tag{13}$$

Now, let $I(X_k; Y_k)$ denote the mutual information between X_k and Y_k. The capacity of the channel is given by

$$C = \max_{f_{X_k}(x)} I(X_k; Y_k) : E[X_k^2] = P \tag{14}$$

The mutual information $I(X_k; Y_k)$ can be expressed as

$$I(X_k; Y_k) = h(Y_k) - h(Y_k|X_k) \tag{15}$$

This takes the form

$$I(X_k; Y_k) = h(Y_k) - h(N_k) \tag{16}$$

When a symbol is transmitted from the source, noise is added to it. So, the total power is $P + \sigma^2$.
For the evaluation of the information capacity C, we proceed in three stages:

1. The variance of sample Y_k of the received signal equals $P + \sigma^2$. Hence, the differential entropy of Y_k is

$$h(Y_k) = \frac{1}{2}log_2[2\pi e(P + \sigma^2)] \tag{17}$$

2. The variance of the noise sample N_k equals σ^2. Hence, the differential entropy of N_k is given by

$$h(N_k) = \frac{1}{2}log_2(2\pi e\sigma^2) \tag{18}$$

3. Now, substituting above two equations into $I(X_k; Y_k) = h(Y_k) - h(N_k)$ yeilds

$$C = \frac{1}{2}log_2(1 + \frac{P}{\sigma^2})bits\ per\ transmission \tag{19}$$

With the channel used K times for the transmission of K samples of the process $X(t)$ in T seconds, we find that the information capacity per unit time is K/T times the result given above for C.

The number K equals $2BT$. Accordingly, we may express the information capacity in the equivalent form:

$$C = Blog_2(1 + \frac{P}{N_0B})\,bits\ per\ second \tag{20}$$

Error Control Coding

- Channel is noisy
- Channel output prone to error
- $\Rightarrow$ we need measure to ensure correctness of the bit stream transmitted

Error control coding aims at developing methods for coding to check the correctness of the bit stream transmitted.

The bit stream representation of a symbol is called the codeword of that symbol.

Different error control mechanisms:

- Linear Block Codes
- Repetition Codes
- Convolution Codes

Linear Block Codes

A code is linear if two codes are added using modulo-2 arithmetic produces a third codeword in the code.

Consider a (n, k) linear block code. Here,

1. n represents the codeword length
2. k is the number of message bit
3. $n - k$ bits are error control bits or parity check bits generated from message using an appropriate rule.

We may therefore represent the codeword as

$$c_i = \begin{cases} b_i, & i = 0, 1, \cdots, n-k-1 \\ m_{i+k-n}, & i = n-k, n-k+1, \cdots, n-1 \end{cases} \quad (1)$$

The $(n-k)$ parity bits are linear sums of the k message bits.

$$b_i = p_{0i}m_0 + p_{1i}m_1 + \cdots + p_{k-1,i}m_{k-1} \tag{2}$$

where the coefficients are

$$p_{ij} = \begin{cases} 1, & \text{if} b_i \text{depends on} m_j \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

We define the 1-by-k message vector, or information vector, **m**, the 1-by-$(n-k)$ parity vector **b**, and the 1-by-n codevector **c** as follows:

$$\mathbf{m} = [m_0, m_1, \cdots, m_{k-1}]$$
$$\mathbf{b} = [b_0, b_1, \cdots, b_{n-k-1}]$$
$$\mathbf{n} = [n_0, n_1, \cdots, n_{n-1}]$$

We may thus write simultaneous equations in matrix equation form as

$$\mathbf{b} = \mathbf{mP}$$

where $\mathbf{P}$ is a k by $n-k$ matrix defined by

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0,n-k-1} \\ p_{10} & p_{11} & \cdots & p_{1,n-k-1} \\ \vdots & \vdots & \vdots & \vdots \\ p_{k-1,0} & p_{k-1,1} & \cdots & p_{k-1,n-k-1} \end{bmatrix}$$

$\mathbf{c}$ matrix can be expressed as a partitioned row vector in terms of the vectors $\mathbf{m}$ and $\mathbf{b}$ as follows

$$c = [b : m]$$
$$\implies c = m[P : I_k]$$

where I_k is a k-by-k identity matrix. Now, we define k-by-n generator matrix as

$$\mathbf{G} = [\mathbf{P} : I_k]$$
$$\implies \mathbf{c} = \mathbf{mG}$$

Closure property of linear block codes:
Consider a pair of code vectors c_i and c_j corresponding to a pair of message vectors m_i and m_j respectively.

$$c_i + c_j = m_i G + m_j G$$
$$= (m_i + m_j)G$$

The modulo-2 sum of m_i and m_j represents a new message vector. Correspondingly, the modulo-2 sum of c_i and c_j represents a new code vector.

Suppose,

$$H = [I_{n-k}|P^T]$$

$$HG^T = [I_{n-k} : P^T]\left[\frac{P^T}{I_k}\right]$$

$$= P^T + P^T$$

$$= 0$$

postmultiplying the basic equation $c = mG$ with H^T, then using the above result we get

$$cH^T = mGH^T$$
$$= 0$$

- A valid codeword yields the above result.
- The matrix H is called the parity-check matrix of the code.
- The set of equations specified in the above equation are called parity-check-equations.

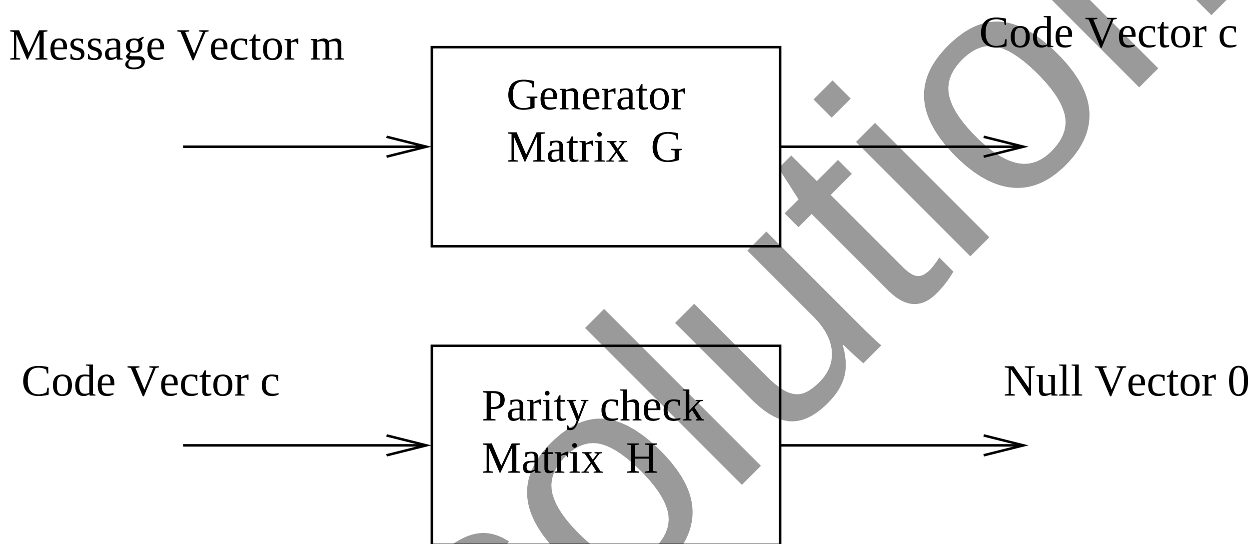


Figure 1: Generator equation and parity check equation

Repetition Codes

This is the simplest of linear block codes. Here, a single message bit is encoded into a block of n identical bits, producing an $(n, 1)$ block code. This code allows variable amount of redundancy. It has only two codewords - all-zero codeword and all-one codeword.

Example

Consider a linear block code which is also a repetition code. Let $k = 1$ and $n = 5$. From the analysis done in linear block codes

$$G = [1111 : 1] \tag{4}$$

The parity check matrix takes the form

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & :1 \\ 0 & 1 & 0 & 0 & :1 \\ 0 & 0 & 1 & 0 & :1 \\ 0 & 0 & 0 & 1 & :1 \end{bmatrix}$$

Syndrome Decoding

Let c be the original codeword sent and $r = c + e$ be the received codeword, where e is the error vector.

$$e_i = \begin{cases} 1, \text{if an error has occured in the ith position} 0, \text{otherwise} \end{cases} \tag{5}$$

Decoding of r

A $1 \times (n - k)$ matrix called error syndrome matrix is calculated

$$s = rH^T \tag{6}$$

Properties of Syndrome:

1. Syndrome depends only on error.

$$s = (c + e)H^T s = cH^T + eH^T \quad (7)$$

Since, cH^T is zero,

$$s = eH^T \quad (8)$$

2. All error patterns that differ by a codeword have the same syndrome.
If there are k message bits, then 2^k distinct codewords can be formed. For any error pattern, e there are 2^k distinct vectors e_i i.e., $e_i = e + c_i, fori = 0, 1, \cdots, 2^k - 1$. Now, multiplying with parity check matrix,

$$\begin{aligned} s_i &= (c_i + e)H^T \\ e_i H^T &= eH^T + c_i H^T \\ &= eH^T \end{aligned} \tag{9}$$

Thus, the syndrome is independent of i.

Hamming Distance

- *Hamming weight*, $w(c)$ is defined as the number of nonzero elements in a codevector.
- *Hamming distance*, $d(c_1, c_2)$ between two codewords c_1 and c_2 is defined as the number of bits in which they differ.
- *Minimum distance*, d_{min} is the minimum hamming distance between two codewords.

It is of importance in error control coding because, error detection and correction is possible if

$$t \leq \frac{1}{2}(d_{min} - 1) \tag{10}$$

where, t is the Hamming weight.

Syndrome based coding scheme for linear block codes

- Let $c_1, c_2, \cdots, c_k$ be the 2^k code vectors of an (n, k) linear block code.
- Let r denote the received vector, which may have one of 2^n possible values.
- These 2^n vectors are partitioned into 2^k disjoint subsets $\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_{2_k}$ in such a way that the ith subset $\mathcal{D}_i$ corresponding to code vector c_i for $1 \leq i \leq 2^k$.
- The received code vector r is decoded into c_i if it is in the ith subset. A standard array of the linear block code is prepared as follows.

$$\begin{bmatrix} c_1 = 0 & c_2 & c_3 & \cdots & c_i & \cdots & c_{2^k} \\ e_2 & c_2 + e_2 & c_3 + e_2 & \cdots & c_i + e_2 & \cdots & c_{2^k} + e_2 \\ e_3 = 0 & c_2 + e_3 & c_3 + e_3 & \cdots & c_i + e_3 & \cdots & c_{2^k} + e_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ e_j = 0 & c_2 + e_j & c_3 + e_j & \cdots & c_i + e_j & \cdots & c_{2^k} + e_j \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ e_{2^{n-k}} = 0 & c_2 + e_{2^{n-k}} & c_3 + e_{2^{n-k}} & \cdots & c_i + e_{2^{n-k}} & \cdots & c_{2^k} + e_{2^{n-k}} \end{bmatrix}$$

The 2^{n-k} rows of the arrays represents the cosets of the code, and their first elements $e_2, \cdots, e_{2^{n-k}}$ are called coset leaders. The probability of decoding is minimized by choosing the most likely occuring error pattern as coset leader. For, binary symmetric channel, the smaller the Hamming weight of an error pattern the more likely is to occur.

The decoding procedure is defined in following three steps.

1. Let r be the received vector, and its syndrome is calculate $s = rH^T$.
2. Within each coset characterized by the syndrome s, identify the coset leader, e_0.
3. Compute $c = r + e_0$ which is the decoded version of r.

Cyclic Codes

Cyclic property: Any cyclic shift of a codeword in the code is also a codeword.

Cyclic codes are well suited for error detection. Certain set of polynomials are chosen for this purpose.

Properties of polynomials:

1. Any polynomial $B(x)$ can be divided by a divisor polynomial $C(x)$ if $B(x)$ is of higher degree than $C(x)$.
2. Any polynomial $B(x)$ can be divided once by $C(x)$ if both are of same degree.
3. subtraction uses exclusive OR.

Cyclic Redundancy Check Codes

Let M(x) be the original message polynomial of kth degree. The following steps are followed in getting the codeword in CRC.

1. Multiply the message polynomial $M(x)$ by x^{n-k}.
2. Divide $x^{n-k}M(x)$ by the generator polynomial $G(x)$, obtaining the remainder $B(x)$.
3. Add $B(x)$ to $x^{n-k}M(x)$, obtaining the code polynomial $C(x)$.

Example:(5,3) cyclic redundancy code with generative polynomial $x^3 + x^2 + 1(1101)$.
Let the message polynomial be $x^5 + x^4 + x(110010)$. Message code after multiplying with x^3 be 110010000. Now, the encoding and decoding of cyclic redundancy codes is shown in the figure below.

Let the message bit stream be 110010.

Hence the message polynomial is $x^5 + x^4 + x$

Let the generating polynomial be $x^3 + x^2 + 1$

Considering a 3−bit CRC, The following process is carried out at encoder and decoder

At the encoder

```
          1001
1101 | 110010000
       1101
       ---------
          1100
          1101
          -------
           100   <---- Remainder
```

Transmitted bit stream with 3−bit CRC added. − 110010100

At the Decoder

If no error.

```
          1001
1101 | 110010100
       1101
       ---------
          1101
          1101
          ------
           000   <---- remainder is zero
```

one bit error

```
          100101
1101 | 110011100
       1101
       ---------
          1111
          1101
          ------
           1000
           1101
           -----
            101
             ^
             |
   Remainder is nonzero
```

Figure 2: example of CRC