

ANALOG COMMUNICATION TECHNIQUES

PCEC 4302

ELECTRONICS & TELECOMMUNICATION ENGINEERING

5th SEMESTER (B.Tech)

Ashima Rout

ashimarout@gmail.com

Jyotirekha Das

jyotirekha43@gmail.com

Rashmita Sethy

sethyrasmita@gmail.com

Disclaimer:

This document does not claim any originality and cannot be used as a substitute for prescribed textbooks. The information presented here is merely a collection by the committee faculty members for their respective teaching assignments as an additional tool for the teaching-learning process. Various sources as mentioned at the reference of the document as well as freely available material from internet were consulted for preparing this document. The ownership of the information lies with the respective authors or institutions. Further, this document is not intended to be used for commercial purpose and the committee faculty members are not accountable for any issues, legal or otherwise, arising out of use of this document. The committee faculty members make no representations or warranties with respect to the accuracy or completeness of the contents of this document and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose

Syllabus with name of books

ANALOG COMMUNICATION TECHNIQUES (3-1-0)

Module-I : (12 Hours)

SIGNALS AND SPECTRA: An Overview of Electronic Communication Systems, Signal and its Properties, Fourier Series Expansion and its Use, The Fourier Transform, Orthogonal Representation of Signal.

RANDOM VARIABLES AND PROCESSES: Probability, Random variables, Useful Probability Density functions, Useful Properties and Certain Application Issues.

AMPLITUDE MODULATION SYSTEMS: Need for Frequency translation, Amplitude Modulation (*Double Side Band with Carrier DSB-C*), Single Sideband Modulation (SSB) Other AM Techniques and Frequency Division Multiplexing, Radio Transmitter and Receiver.

Module-II : (12 Hours)

ANGLE MODULATION: Angle Modulation, Tone Modulated FM Signal, Arbitrary Modulated FM signal, FM Modulators and Demodulators, Approximately Compatible SSB Systems.

PULSE MODULATION AND DIGITAL TRANSMISSION OF ANALOG SIGNAL:

Analog to Digital (*Noisy Channel and Role of Repeater*), Pulse Amplitude Modulation and Concept of Time division multiplexing, Pulse Width Modulation and Pulse Position Modulation, Digital Representation of Analog Signal.

Module-III : (14 Hours)

MATHEMATICAL REPRESENTATION OF NOISE: Some Sources of Noise, Frequency-domain Representation of Noise, Superposition of Noises, Linear Filtering of Noise.

NOISE IN AMPLITUDE MODULATION SYSTEM : Framework for Amplitude Demodulation, Single Sideband Suppressed Carrier (SSB-SC), Double Sideband Suppressed Carrier (DSB-SC), Double Sideband With Carrier (DSB-C).

NOISE IN FREQUENCY MODULATION SYSTEM : An FM Receiving System, Calculation of Signal to Noise Ratio, Comparison of FM and AM, Preemphasis and Deemphasis and SNR Improvement, Noise in Phase Modulation and Multiplexing Issues, Threshold in Frequency Modulation, Calculation of Threshold in an FM Discriminator, The FM Demodulator using Feedback (FMFB).

Essential Reading:

1. H. Taub, D. L Schilling, G. Saha; *Principles of Communication System, 3rd Edition; 2008, Tata McGraw Hill, India*; ISBN: 0070648115. (**Selected portions from chapters:** Chapter-1, Chapter-2, Chapter-3, Chapter-4, Chapter-5, Chapter-7, Chapter-8, Chapter-9)

Supplementary Reading:

1. Communication System Engineering, Second Edition by Masoud Salehi, John G. Proakis, ISBN: 0130950076 (paperback)
2. Analog Communication by Chandra Sekar, Oxford University Press.
3. Modern Digital and Analog Communication Systems, by B.P. Lathi, Oxford

Module I:

Signal & Spectra

Definition of a signal:

"A signal is a detectable physical quantity or impulse (as a voltage, current, or magnetic field strength) that varies with respect to time, space, temperature or any other independent variable or can be defined as a function $x(t)$ of independent variable 't' by which messages or information about behaviour of a natural or artificial system can be conveyed"

Electrical signals - time varying voltages and currents - in many cases have important properties that are necessary to be measured. Sometimes it is also justified to make any of these kinds of measurements.

- Power in an audio signal - as one can test an audio amplifier's output ability
- Frequency - as one can use an AC tachometer to measure a motor's rpm.
- Amplitude – measurement of signal strength in a communication system.

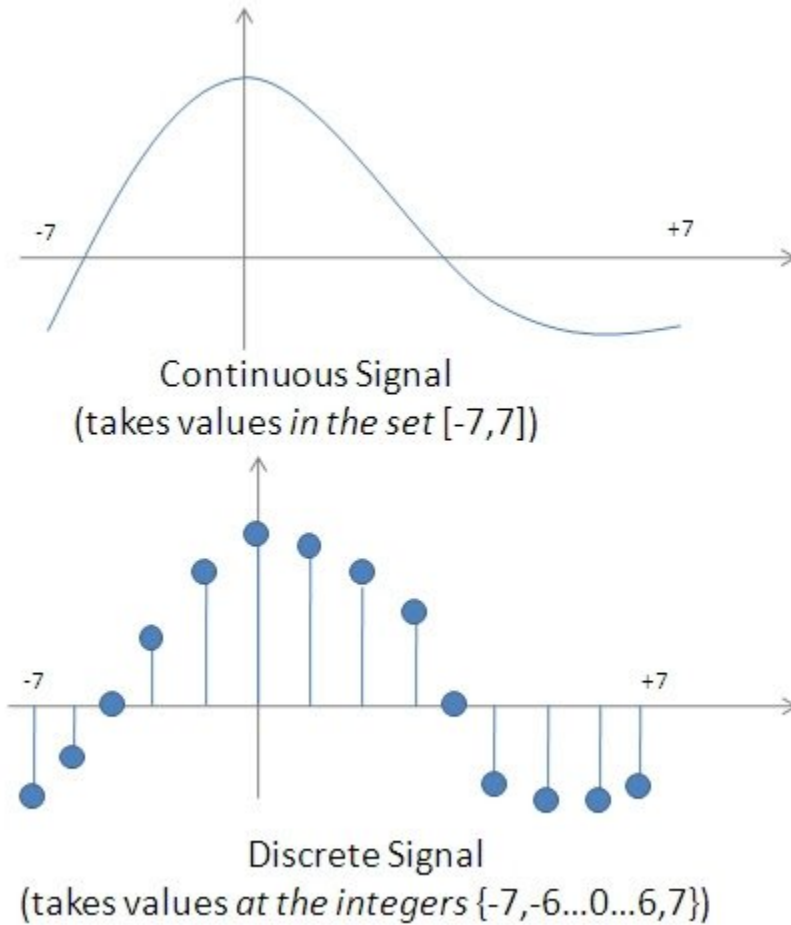
This includes the basic definition of signals. To easily understand signals & systems, we would visualize signals as simple mathematical functions.

Classification of signals:

1. Continuous & Discrete Signals

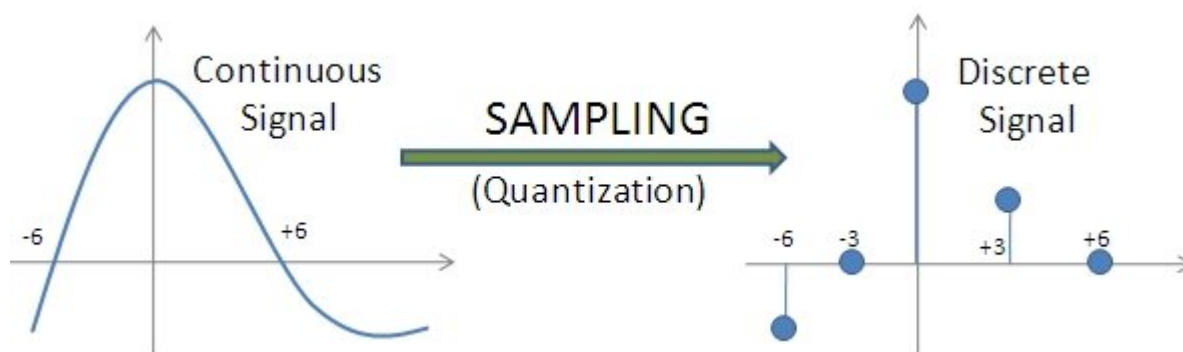
- Continuous signals & those defined over a set of real numbers(\mathbb{R}) & discrete signals are those defined for discrete integers(\mathbb{I}).
For instance, a signal (a function) having the domain $[0,10]$ is continuous & one having

the domain $\{1,2,3\ldots 10\}$ is discrete.



A Continuous Signal can be converted to a Discrete Signal using an Analog-to-Digital Converter (ADC). The conversion consists of a process called **sampling**.

The sampling process simply samples out values of the signal at certain points separated by an equal interval called the sampling period.

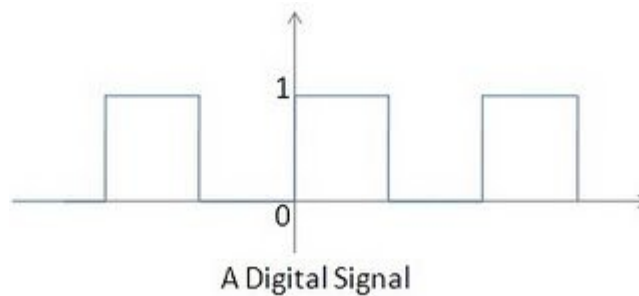


A common application of the above process is a Compact Disc (CD) which is simply a signal sampled at 44.1kHz & Quantized at 16 bits/2 bytes.

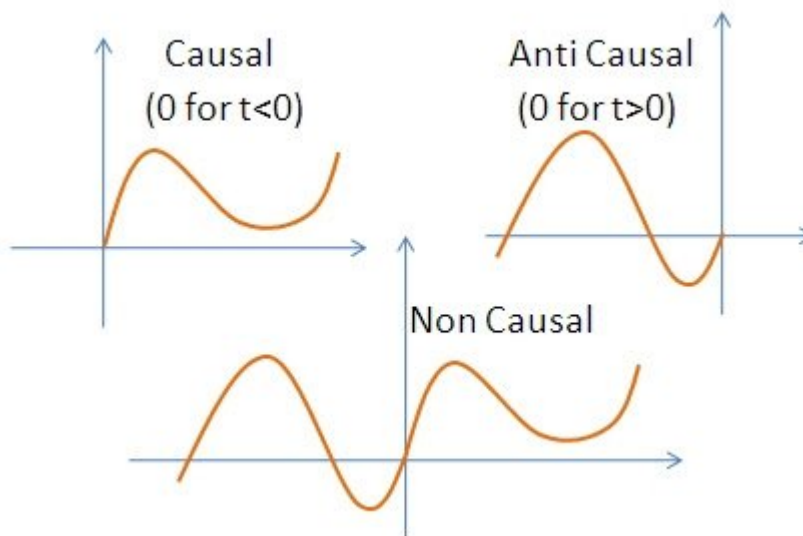
2. Analog & Digital Signals:

Analog signals are continuous electric signals which arise from non-electric signals. The variable of the converted signal is analogous to the non-electric time varying signal & hence, they are called analog signals. A good example is an audio (speech) signal.

Digital signal signals, take only two values-HIGH or LOW, ON or OFF, 0 or 1, TRUE or FALSE, etc. All computers & other gadgets use digital signals to store information. (The term sometimes also refers to discrete time signals which can also take discrete values other than 0's & 1's)

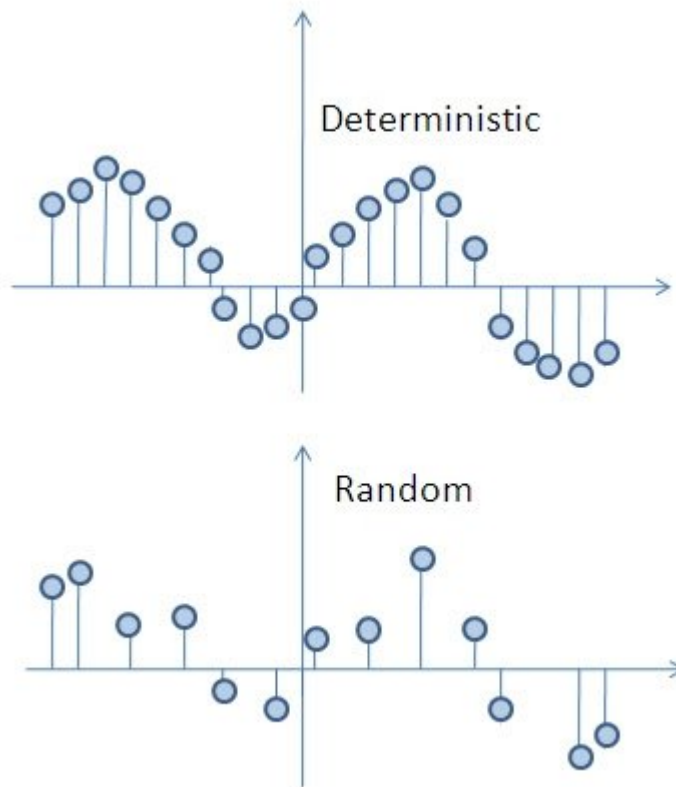


Signals are also classified as Causal and Anti causal signals: ...



3. Deterministic & Random Signals.

A **random signal** takes random values & at a point on the signal, we cannot determine its value just before it or just after it. However, these values can be easily determined for a **deterministic**



signal.

Even and Odd Signal

One of characteristics of signal is symmetry that may be useful for signal analysis. Even signals are symmetric around vertical axis, and Odd signals are symmetric about origin

Even Signal:

A signal is referred to as an even if it is identical to its time - reversed counterparts;

$$x(t) = x(-t)$$

Odd Signal:

A signal is odd if $x(t) = -x(-t)$

An odd signal must be 0 at $t=0$, in other words, odd signal passes the origin.

Using the definition of even and odd signal, any signal may be decomposed into a sum of its even part, and its odd part,

Periodic Signal If the transformed signal $x(t)$ is same as $x(t+nT)$, then the signal is periodic. where T is fundamental period (the smallest period) of signal $x(t)$ In discrete -time, the periodic signal is;

Aperiodic signal: Signal those do not satisfy the above condition and is not repeated over a particular interval of time.

System:

A **System** is any physical set of components that takes a signal, and produces a signal. In terms of engineering, the input is generally some electrical signal X , and the output is another electrical signal(response) Y . However, this may not always be the case. Consider a household thermostat, which takes input in the form of a knob or a switch, and in turn outputs electrical control signals for the furnace.

A main purpose of this book is to try and lay some of the theoretical foundation for future dealings with electrical signals. Systems will be discussed in a theoretical sense only.

Random Variable & Processes:

In probability and statistics a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance. A random variable can take on a set of possible different values.

A random variable's possible values might represent the possible outcomes of a past experiment whose already-existing value is uncertain (for example, as a result of incomplete information or imprecise measurements). They may also conceptually represent either the results of an "objectively" random process (such as rolling a die) or the "subjective" randomness that results from incomplete knowledge of a quantity. The meaning of the probabilities assigned to the potential values of a random variable is not part of probability theory itself but is instead related to some arguments over the interpretation of probability.

The mathematical function describing the possible values of a random variable and their associated probabilities is known as a probability distribution can be discrete which , takes on any of a specified finite or countable list of values, capable with a probability mass function, characteristic of a probability distribution; or continuous, taking any numerical value in an interval or collection of intervals,

The formal mathematical treatment of random variables is a topic in probability theory. In that context, a random variable is understood as a function defined on a sample space whose outputs are numerical values.

A random variable $X: \Omega \rightarrow E$ is a measurable function from the set of possible outcomes Ω to some set E . Usually, $E = \mathbb{R}$. The technical axiomatic definition requires both Ω and E to be measurable spaces.

As a real-valued function, X often describes some numerical quantity of a given event,

e.g., the number of heads after a certain number of coin flips; the heights of different people.

When the range of X is finite or countable infinite, the random variable is called a discrete random variable and its distribution can be described by a probability mass function which assigns a probability to each value in the image of X . If the image is uncountable and infinite then X is called a continuous random variable. In the special case that it is absolutely continuous, its distribution can be described by a probability density function, which assigns probabilities to intervals; in particular, each individual point must necessarily have probability zero for an absolutely continuous random variable. Not all continuous random variables are absolutely continuous

All random variables can be described by their cumulative distribution function, which describes the probability that the random variable will be less than or equal to a certain value.

II. Amplitude modulation (AM) is a modulation technique used for transmitting information via a radio carrier wave. AM works by varying the strength (amplitude) of the carrier in proportion to the waveform being sent. That waveform may, for instance, correspond to the sounds to be reproduced by a loudspeaker, or the light intensity of television pixels. This contrasts with frequency modulation, in which the frequency of the carrier signal is varied, and phase modulation, in which its phase is varied, by the modulating signal.

Need for Modulation:

Ease transmission that necessitates the decision of antenna length, Allows multiplexing of signals, Bandwidth trade-off with SNR

AM was the earliest modulation method used to transmit voice by radio for example it is used in portable two way radio, VHF aircraft radio and in computer modems. "AM" is often used to refer to medium wave AM radio broadcasting.

One disadvantage of all amplitude modulation techniques (not only standard AM) is that the receiver amplifies and detects noise and electromagnetic interference in equal proportion to the signal. Increasing the received signal to noise ratio, say, by a factor of 10 (a 10 decibel improvement), thus would require increasing the transmitter power by a factor of 10. This is in

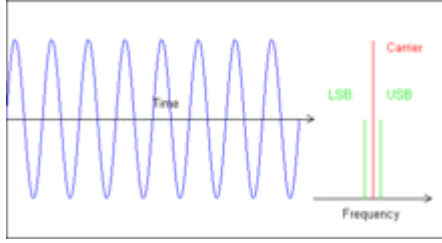
contrast to frequency modulation (FM) and digital where the effect of such noise following demodulation is strongly reduced so long as the received signal is well above the threshold for reception. For this reason AM broadcast is not favored for music and highly fidelity broadcasting, but rather for voice communications and broadcasts (sports, news, talk radio etc.).

Another disadvantage of AM is that it is inefficient in power usage; at least two-thirds of the power is concentrated in the carrier signal. The carrier signal contains none of the original information being transmitted (voice, video, data, etc.). However its presence provides a simple means of demodulation using envelop detection, providing a frequency and phase reference to extract the modulation from the sidebands. In some modulation systems based on AM, a lower transmitter power is required through partial or total elimination of the carrier component; however receivers for these signals are more complex and costly. The receiver may regenerate a copy of the carrier frequency (usually as shifted to the intermediate frequency) from a greatly reduced "pilot" carrier (in reduced carrier- carrier or DSB-RC) to use in the demodulation process. Even with the carrier totally eliminated in double –sideband suppressed-carrier transmission, carrier regeneration is possible using a phase-locked loop. This doesn't work however for single-sideband suppressed-carrier (SSB-SC), Single sideband is nevertheless used widely in amateur radio and other voice communications both due to its power efficiency and bandwidth efficiency (cutting the RF bandwidth in half compared to standard AM). On the other hand, in MW and SW broadcasting, standard AM with the full carrier allows for reception using inexpensive receivers. The broadcaster absorbs the extra power cost to greatly increase potential audience.

An additional function provided by the carrier in standard AM, but which is lost in either single or double-sideband suppressed-carrier transmission, is that it provides an amplitude reference. In the receiver, the automatic gain control (AGC) responds to the carrier so that the reproduced audio level stays in a fixed proportion to the original modulation. On the other hand, with suppressed-carrier transmissions there is *no* transmitted power during pauses in the modulation, so the AGC must respond to peaks of the transmitted power during peaks in the modulation. This typically involves a so-called fast attack, slow decay circuit which holds the AGC level for a second or more following such peaks, in between syllables or short pauses in the program. This is very acceptable for communications radios, where compression of the audio aids intelligibility. However it is absolutely undesired for music or normal broadcast programming, where a faithful reproduction of the original program, including its varying modulation levels, is expected.

A trivial form of AM which can be used for transmitting binary data is on-off keying, the simplest form of amplitude shift keying, in which ones and zeros are represented by the presence or absence of a carrier. On-off keying is likewise used by radio amateurs to transmit Morse code where it is known as continuous wave (CW) operation, even though the transmission is not strictly "continuous."

Simplified analysis of standard AM



Left part: Modulating signal. Right part: Frequency spectrum of the resulting amplitude modulated carrier

Consider a carrier wave (sine wave) of frequency f_c and amplitude A given by:

$$c(t) = A \cdot \sin(2\pi f_c t)$$

Let $m(t)$ represent the modulation waveform. For this example we shall take the modulation to be simply a sine wave of a frequency f_m , a much lower frequency (such as an audio frequency) than f_c :

$$m(t) = M \cdot \cos(2\pi f_m t + \phi)$$

where M is the amplitude of the modulation. We shall insist that $M < 1$ so that $(1+m(t))$ is always positive. Amplitude modulation results when the carrier $c(t)$ is multiplied by the positive quantity $(1+m(t))$:

$$\begin{aligned} y(t) &= [1 + m(t)] \cdot c(t) \\ &= [1 + M \cdot \cos(2\pi f_m t + \phi)] \cdot A \cdot \sin(2\pi f_c t) \end{aligned}$$

In this simple case M is identical to the modulation index, discussed below. With $M=0.5$ the amplitude modulated signal $y(t)$ thus corresponds to the top graph (labelled "50% Modulation") in Figure 4.

Using prosthaphaeresis identities, $y(t)$ can be shown to be the sum of three sine waves:

$$y(t) = A \cdot \sin(2\pi f_c t) + \frac{AM}{2} [\sin(2\pi(f_c + f_m)t + \phi) + \sin(2\pi(f_c - f_m)t - \phi)]$$

Therefore, the modulated signal has three components: the carrier wave $c(t)$ which is unchanged, and two pure sine waves (known as sidebands) with frequencies slightly above and below the carrier frequency f_c .

Modulation Index

The AM modulation index is a measure based on the ratio of the modulation excursions of the RF signal to the level of the unmodulated carrier. It is thus defined as:

$$h = \frac{\text{peak value of } m(t)}{A} = \frac{M}{A}$$

where M and A are the modulation amplitude and carrier amplitude, respectively; the modulation amplitude is the peak (positive or negative) change in the RF amplitude from its unmodulated value. Modulation index is normally expressed as a percentage, and may be displayed on a meter connected to an AM transmitter.

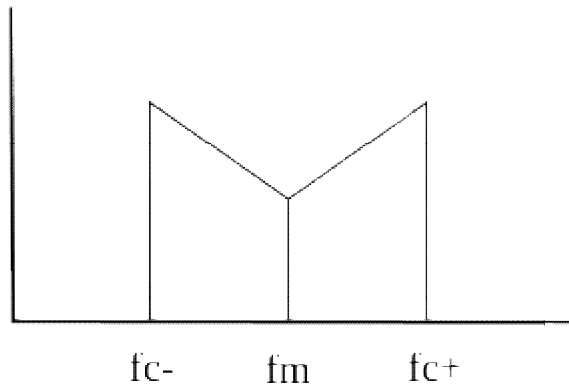
Double-sideband suppressed-carrier transmission (DSB-SC) is transmission in which frequencies produced by amplitude modulation (AM) are symmetrically spaced above and below the carrier frequency and the carrier level is reduced to the lowest practical level, ideally being completely suppressed.

In the DSB-SC modulation, unlike in AM, the wave carrier is not transmitted; thus, much of the power is distributed between the sidebands, which implies an increase of the cover in DSB-SC, compared to AM, for the same power used.

DSB-SC transmission is a special case of double-sideband reduced carrier transmission. It is used for radio data systems.

Spectrum

DSB-SC is basically an amplitude modulation wave without the carrier, therefore reducing power waste, giving it a 50% efficiency. This is an increase compared to normal AM transmission (DSB), which has a maximum efficiency of 33.333%, since 2/3 of the power is in the carrier which carries no intelligence, and each sideband carries the same information. Single Side Band (SSB) Suppressed Carrier is 100% efficient.

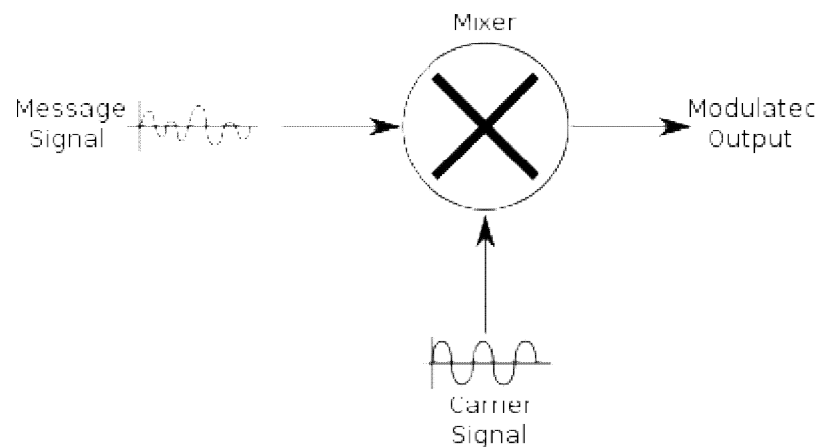


Spectrum plot of an DSB-SC signal: f_m $f_c - f_m$ $f_c + f_m$ f_m

Generation

DSB-SC is generated by a mixer. This consists of a message signal multiplied by a carrier signal. The mathematical representation of this process is shown below, where the product-to-sum trigonometric identity is used.

$$\underbrace{V_m \cos(\omega_m t)}_{\text{Message}} \times \underbrace{V_c \cos(\omega_c t)}_{\text{Carrier}} = \underbrace{\frac{V_m V_c}{2} [\cos((\omega_m + \omega_c) t) + \cos((\omega_m - \omega_c) t)]}_{\text{Modulated Signal}}$$



Demodulation

Demodulation is done by multiplying the DSB-SC signal with the carrier signal just like the modulation process. This resultant signal is then passed through a low pass filter to produce a scaled version of original message signal. DSB-SC can be demodulated if modulation index is less than unity.

$$\begin{aligned}
& \overbrace{\frac{V_m V_c}{2} [\cos((\omega_m + \omega_c)t) + \cos((\omega_m - \omega_c)t)]}^{\text{Modulated Signal}} \times \overbrace{V'_c \cos(\omega_c t)}^{\text{Carrier}} \\
&= \left(\frac{1}{2} V_c V'_c \right) \underbrace{V_m \cos(\omega_m t)}_{\text{original message}} + \frac{1}{2} V_c V'_c V_m [\cos((\omega_m + 2\omega_c)t) + \cos((\omega_m - 2\omega_c)t)]
\end{aligned}$$

The equation above shows that by multiplying the modulated signal by the carrier signal, the result is a scaled version of the original message signal plus a second term. Since $\omega_c \gg \omega_m$, this second term is much higher in frequency than the original message. Once this signal passes through a low pass filter, the higher frequency component is removed, leaving just the original message.

Distortion and Attenuation :

For demodulation, the demodulation oscillator's frequency and phase must be exactly the same as modulation oscillator's, otherwise, distortion and/or attenuation will occur.

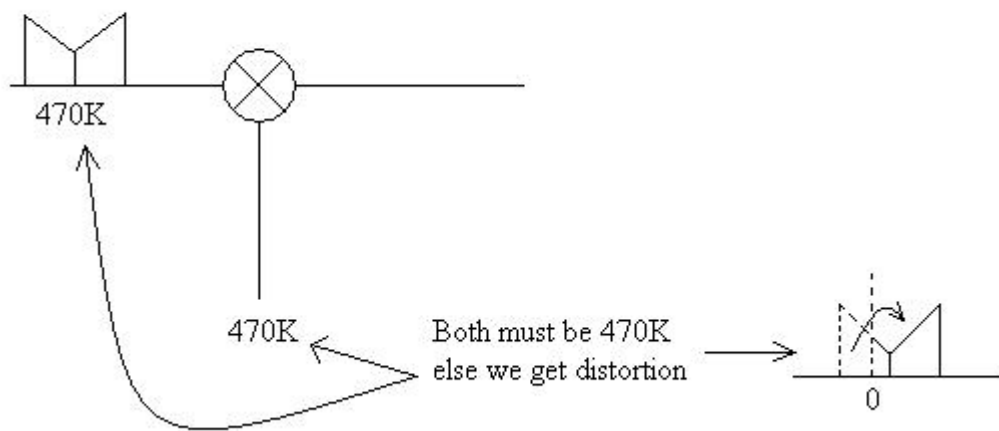
To see this effect, take the following conditions:

- Message signal to be transmitted: $f(t)$
- Modulation (carrier) signal: $V_c \cos(\omega_c t)$
- Demodulation signal (with small frequency and phase deviations from the modulation signal): $V'_c \cos[(\omega_c + \Delta\omega)t + \theta]$

The resultant signal can then be given by

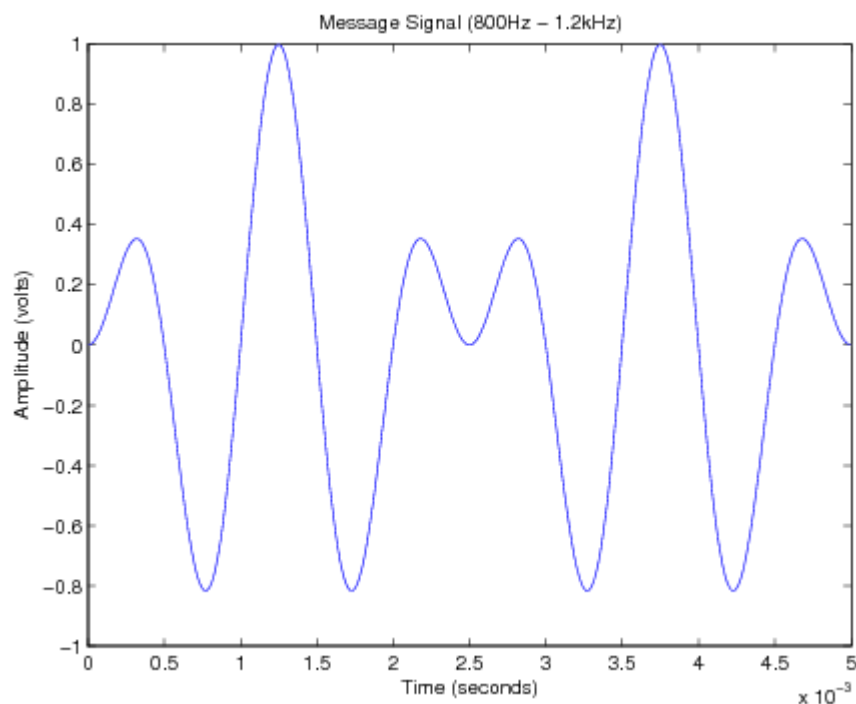
$$\begin{aligned}
& f(t) \times V_c \cos(\omega_c t) \times V'_c \cos[(\omega_c + \Delta\omega)t + \theta] \\
&= \frac{1}{2} V_c V'_c f(t) \cos(\Delta\omega \cdot t + \theta) + \frac{1}{2} V_c V'_c f(t) \cos[(2\omega_c + \Delta\omega)t + \theta] \\
&\xrightarrow{\text{After low pass filter}} \frac{1}{2} V_c V'_c f(t) \cos(\Delta\omega \cdot t + \theta)
\end{aligned}$$

The $\cos(\Delta\omega \cdot t + \theta)$ terms results in distortion and attenuation of the original message signal. In particular, $\Delta\omega \cdot t$ contributes to distortion while θ adds to the attenuation.



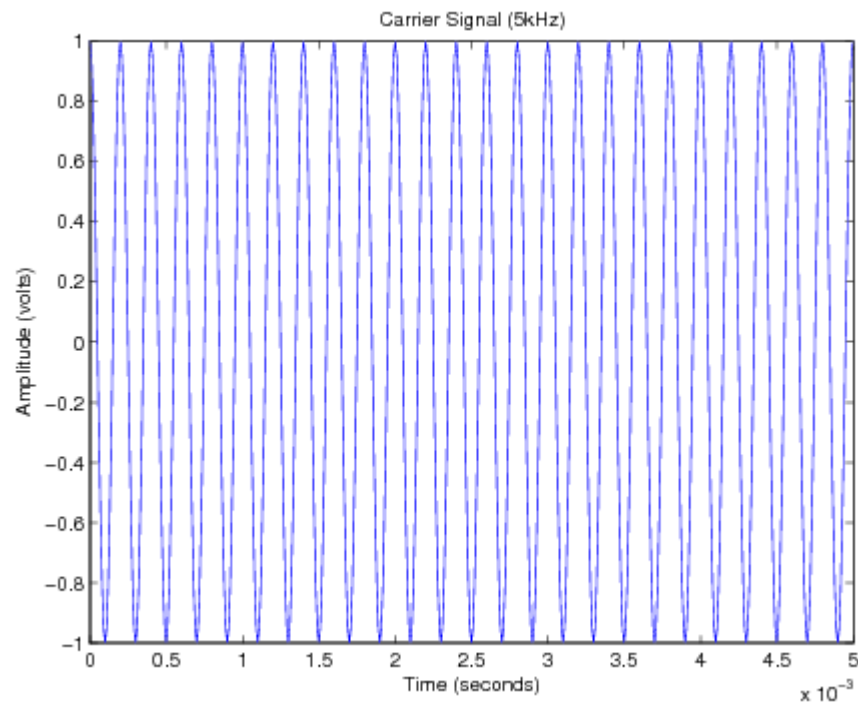
It's working:

This is best shown graphically. Below is a message signal that one may wish to modulate onto a carrier, consisting of a couple of sinusoidal components.

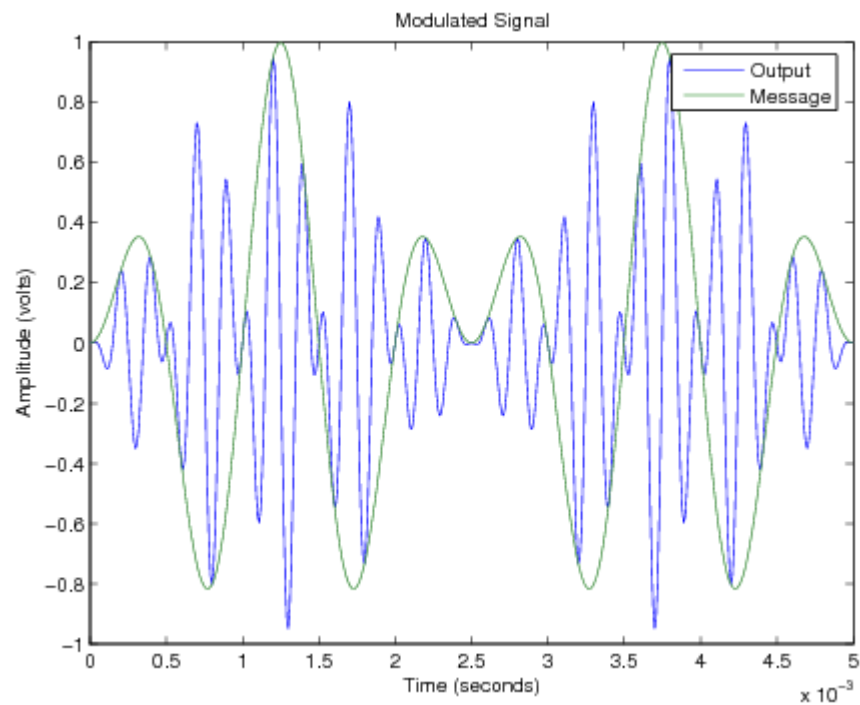


The equation for this message signal is $s(t) = \frac{1}{2} \cos(2\pi 800t) - \frac{1}{2} \cos(2\pi 1200t)$.

The carrier, in this case, is a plain 5 kHz ($c(t) = \cos(2\pi 5000t)$) sinusoid—pictured below.

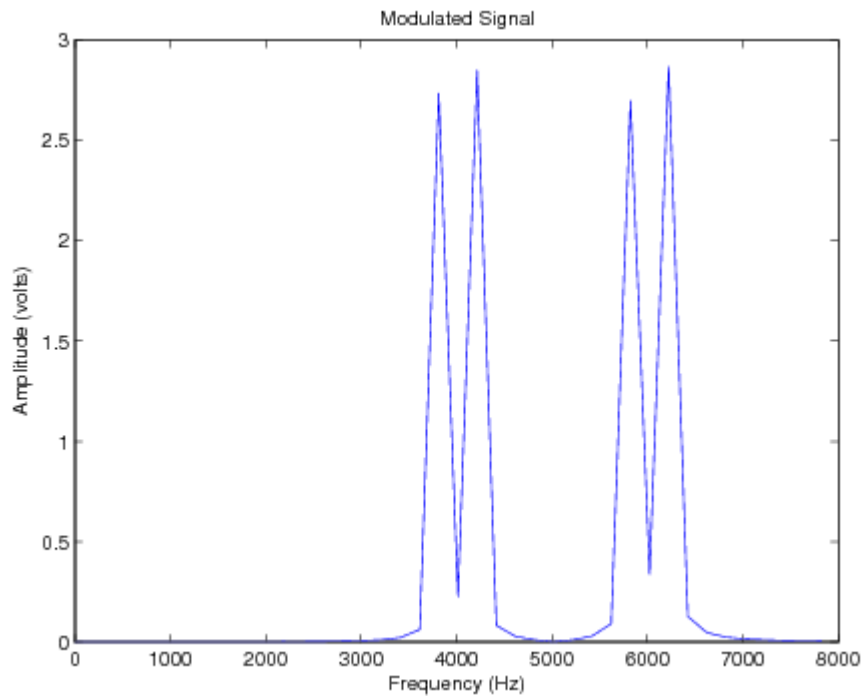


The modulation is performed by multiplication in the time domain, which yields a 5 kHz carrier signal, whose amplitude varies in the same manner as the message signal.



$$x(t) = \underbrace{\cos(2\pi 5000t)}_{\text{Carrier}} \times \underbrace{\left[\frac{1}{2} \cos(2\pi 800t) - \frac{1}{2} \cos(2\pi 1200t) \right]}_{\text{Message Signal}}$$

The name "suppressed carrier" comes about because the carrier signal component is suppressed—it does not appear in the output signal. This is apparent when the spectrum of the output signal is viewed:



MODULE –II

ANGLE MODULATION:

In this type of modulation the modulator output is of constant in amplitude, and the message signal is superimposed on the carrier by varying its frequency and phase together is called angle [1].

Angle modulation encompasses phase modulation (PM) and frequency modulation (FM). The phase angle of a sinusoidal carrier signal is varied according to the modulating signal. In angle modulation, the spectral components of the modulated signal are not related in a simple fashion to the spectrum of the modulating signal [1]. Superposition does not apply and the bandwidth of the modulated signal is usually much greater than the modulating signal bandwidth. Mathematically

Consider a sinusoid $A_c \cos(2\pi f_c t + \phi_0)$, where A_c is the (constant) amplitude, f_c is the (constant) frequency in Hz and ϕ_0 is the initial phase angle. Let the sinusoid be written as $A_c \cos[\theta(t)]$ where $\theta(t) = 2\pi f_c t + \phi_0$. the case where A_c is a constant but, instead of being equal to $2\pi f_c t + \phi_0$, is a function of $m(t)$. This leads to what is known as the angle modulated signal [2]. Two important cases of angle modulation are Frequency Modulation (FM) and Phase modulation (PM).

An important feature of FM and PM is that they can provide much better protection to the message against the channel noise as compared to the linear (amplitude) modulation schemes. Also, because of their constant amplitude nature, they can withstand nonlinear distortion and amplitude fading. The price paid to achieve these benefits is the increased bandwidth requirement; that is, the transmission bandwidth of the FM or PM signal with constant amplitude and which can provide noise immunity is much larger than $2W$, where W is the highest frequency component present in the message spectrum [2].

Now let us define PM and FM. Consider a signal $s(t)$ given by $s(t) = A_c \cos[\theta_i(t)]$ where $\theta_i(t)$, the instantaneous angle quantity, is a function of $m(t)$. We define the instantaneous frequency of the angle modulated wave $s(t)$, as

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} \quad (5.1)$$

(The subscript i in $\theta_i(t)$ or $f_i(t)$ is indicative of our interest in the instantaneous behavior of these quantities). If $\theta_i(t) = 2\pi f_c t + \phi_0$, then $f_i(t)$ reduces to the constant f_c , which is in perfect agreement with our established notion of frequency of a sinusoid. This is illustrated in Fig. 5.1.

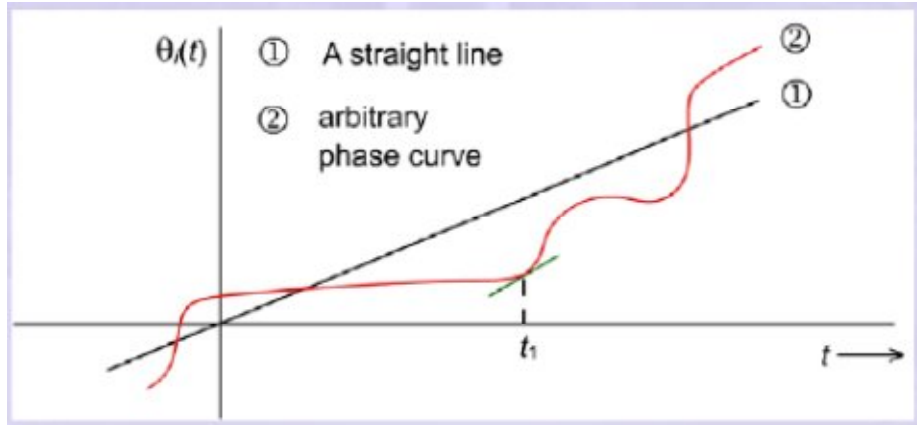


Fig.5.1 Illustration of instantaneous phase and frequency

Curve 1 in Fig. 5.1 depicts the phase behavior of a constant frequency sinusoid with $\phi_0 = 0$. Hence, its phase, as a function of time is a straight line; that is $\theta_1(t) = 2\pi f_c t$. Slope of this line is a constant and is equal to the frequency of the sinusoid. Curve 2 depicts an arbitrary phase behavior; its slope changes with time. The instantaneous frequency (in radians per second) of this signal at $t = t_1$ is given by the slope of the tangent (green line) at that time [2].

a) Phase modulation:

For PM, $\theta_i(t)$ is given by

$$\theta_i(t) = 2\pi f_c t + k_p m(t) \quad (5.2)$$

The term $2\pi f_c t$ is the angle of the un-modulated carrier and the constant k_p is the phase sensitivity of the modulator with the units, radians per volt. (For convenience, the initial phase angle of the un-modulated carrier is assumed to be zero) [1].

Using Eq. 5.2, the phase modulated wave $s(t)$ can be written as

$$[s(t)]_{PM} = A_c \cos[2\pi f_c t + k_p m(t)] \quad (5.3)$$

From Eq. 5.2 and 5.3, it is evident that for PM, the phase deviation of $s(t)$ from that of the un modulated carrier phase is a linear function of the base-band message signal, $m(t)$. The instantaneous frequency of a phase modulated signal depends on

$$\frac{dm(t)}{dt} = m'(t) \text{ because } \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} = f_c + \frac{k_p}{2\pi} m'(t).$$

b) Frequency Modulation:

Let us now consider the case where $f_i(t)$ is a function of $m(t)$; that is,

$$f_i(t) = f_c + k_f m(t) \quad (5.4)$$

$$\theta_i(t) = 2\pi \int_{-\infty}^t f_i(\tau) d\tau \quad (5.5)$$

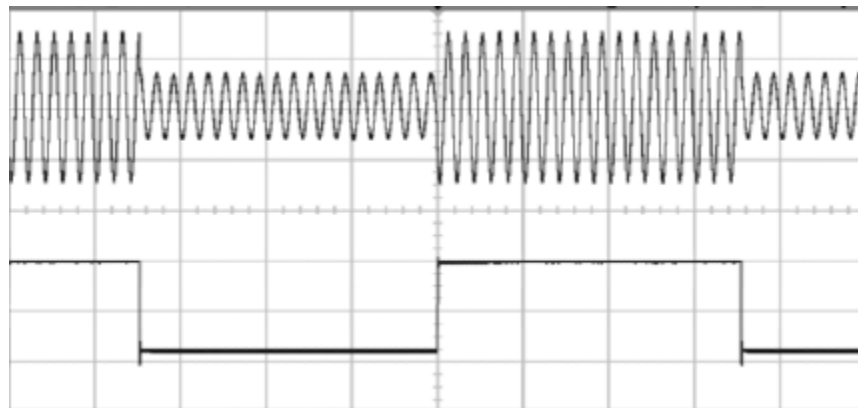
$$= 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \quad (5.6)$$

k_f is a constant, which we will identify shortly. A frequency modulated signal $s(t)$ is described in the time domain by

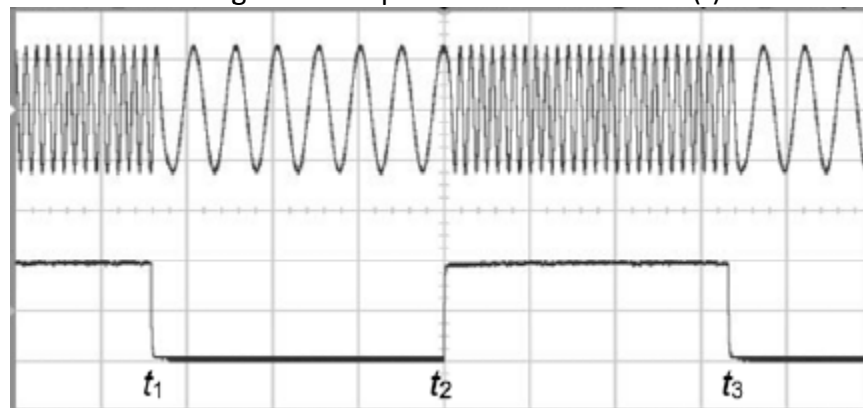
$$[s(t)]_{FM} = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right] \quad (5.7)$$

k_f is termed as the *frequency sensitivity* of the modulator with the units Hz/volt.

From Eq. 5.4 we infer that for an FM signal, the instantaneous frequency



AM signal with square wave shown as $m(t)$



FM signal with square wave shown as $m(t)$

As both PM and FM have constant amplitude A_c , the average power of a PM or FM signal is,

$$P_{av} = \frac{A_c^2}{2},$$

regardless of the value of k_p or k_f .

TONE MODULATED FM SIGNAL:

An angle modulated signal is known as tone modulated if its amplitude arbitrarily set at unity, also referred to as an exponentially modulated signal [1], has the form

$$S_m(t) = A \cos[\omega t + \theta(t)] = \text{Re}\{A \exp[j\phi(t)]\} \quad (1)$$

Where the instantaneous phase $\phi_i(t)$ is defined as

$$\phi_i(t) = \omega t + \theta(t) \quad (2)$$

and the instantaneous frequency of the modulated signal is defined as

$$\omega_i(t) = \frac{d}{dt}[\omega t + \theta(t)] = \omega + \frac{d(\theta(t))}{dt} \quad (3)$$

The functions $\theta(t)$ and $\frac{d(\theta(t))}{dt}$ are referred to as the instantaneous phase and frequency deviations, respectively.

The phase deviation of the carrier $\phi(t)$ is related to the baseband message signal $s(t)$. Depending on the nature of the relationship between $\phi(t)$ and $s(t)$ we have different forms of angle modulation [1].

$$\frac{d(\theta(t))}{dt} = k_f s(t) \quad (4)$$

$$\phi(t) = k_f \int_{t_0}^t s(\lambda) d\lambda + \omega t \quad (5)$$

Where K_f is a frequency deviation constant, (expressed in (radian/sec)/volt).

It is usually assumed that $t_0 = -\infty$ and $\phi(-\infty) = 0$.

Combining Equations-4 and 5 with Equation-1, we can express the frequency modulated signal as

$$S_m(t) = A \cos[\omega t + k_f \int_{-\infty}^t s(\tau) d\tau] \quad (6)$$

Fig. 1 shows a single tone ($s(t)$ message signal), frequency modulated a carrier frequency, represented in time domain.

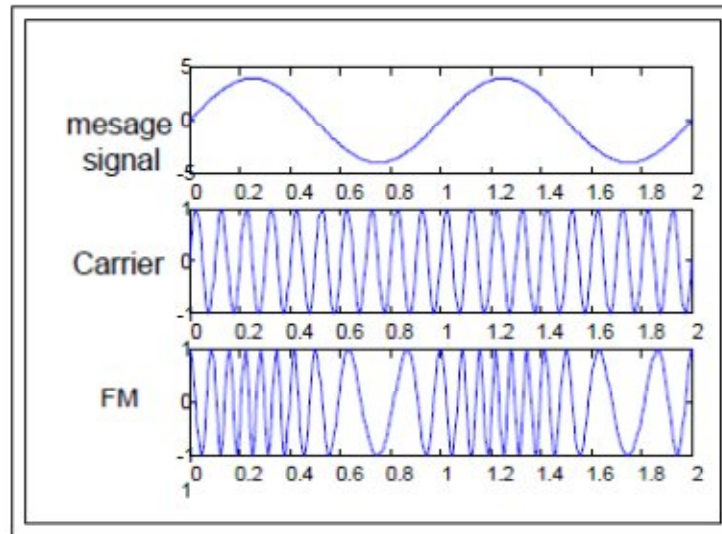


Fig 1: modulation signal (upper), carrier signal (middle), and modulated FM signal(lower)

Bessel function:

Bessel function of the first kind, is a solution of the differential equation

$$\beta^2 \frac{d^2 y}{d\beta^2} + \beta \frac{dy}{d\beta} + (\beta^2 - n^2)y(\beta) = 0$$

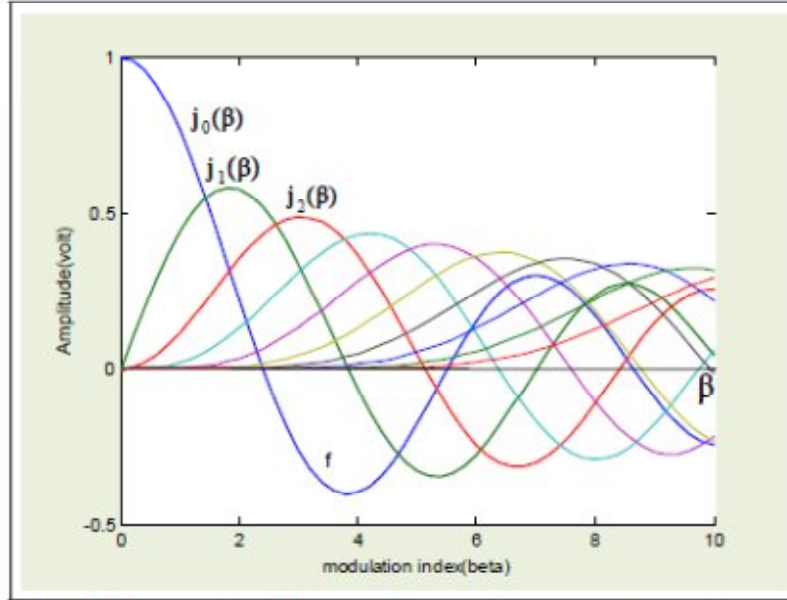


Fig. 2 Bessel function, of kind 1, and order 1 to 10

Bessel function defined for negative and positive real integers. It can be shown that for integer values of n

$$j_{-n}(\beta) = (-1)^n j_n(\beta) \quad (7)$$

$$j_{n-1}(\beta) + j_{n+1}(\beta) = \frac{2n}{\beta} j_n(\beta) \quad (8)$$

$$\sum_{n=-\infty}^{\infty} j_n^2(\beta) = 1 \quad (9)$$

A short listing of Bessel function of first kind of order n and discrete value of argument β , is shown in Table-1, and graph of the function, is shown in Fig. 2

Note that for very small β , value $j_0(\beta)$ approaches unity, while $j_1(\beta)$ to $j_n(\beta)$ approach zero.

$n \backslash \beta$	0	0.2	0.5	1	2	5	8	10
0	<u>1.00</u>	0.99	0.938	0.765	0.224	-0.178	0.172	-0.246
1	0	<u>0.1</u>	<u>0.242</u>	0.440	0.577	-0.328	0.235	0.043
2		0.005	0.031	<u>0.115</u>	0.353	0.047	-0.113	0.255
3				0.02	<u>0.129</u>	0.365	-0.291	0.058
4				0.002	0.034	0.391	-0.105	-0.22
5					0.007	0.261	0.186	-0.234
6						<u>0.131</u>	0.338	-0.14
7						0.053	0.321	0.217
8						0.018	0.223	0.318
9							<u>0.126</u>	0.292
10							0.061	0.208
11							0.026	<u>0.123</u>

Table-1 Bessel function $j_n(\beta)$

Properties of Bessel's Function:

1. Eq. -1.9 indicates that the phase relationship between the sideband components is such that the odd-order lower sidebands are reversed in phase .
2. The number of significant spectral components is a function of argument β (see Table-1). When $\beta \ll 1$, only J_0 , and J_1 , are significant so that the spectrum will consist of carrier plus two sideband components, just like an *AM* spectrum with the exception of the phase reversal of the lower sideband component.
3. A large value of β implies a large bandwidth since there will be many significant sideband components.
4. Transmission bandwidth of 98% of power always occur after $n = \beta + 1$, we note it in table-1 with underline.
5. Carrier and sidebands null many times at special values of β see table-2

Order	0	1	2	3	4	5	6
β for 1st zero	2.40	3.83	5.14	6.38	7.59	8.77	9.93
β for 2nd zero	5.52	7.02	8.42	9.76	11.06	12.34	13.59
β for 3rd zero	8.65	10.17	11.62	13.02	14.37	15.70	17.00
β for 4th zero	11.79	13.32	14.80	16.22	17.62	18.98	20.32
β for 5th zero	14.93	16.47	17.96	19.41	20.83	22.21	23.59
β for 6th zero	18.07	19.61	21.12	22.58	24.02	25.43	26.82
Table-2 Zeroes of Bessel function: Values for β when $j_n(\beta) = 0$							

Power and Bandwidth of Tone modulated signal:

In the previous section we saw that a single tone modulated *FM* signal has an infinite number of sideband components and hence the *FM* spectrum seems to have infinite spectrum. Fortunately, it turns out that for any β a large portion of the power is contained in finite bandwidth. . Hence the determination of *FM* transmission bandwidth depends to the question of how many significant sidebands need to be included for transmission, if the distortion is to be within certain limits.

To determine *FM* transmission bandwidth, let us analyze the power ratio S_n , which is the fraction of the power contained in the carrier plus n sidebands, to the total power of FM signal. We search a value of number of sidebands n , for power ratio $S_n \geq 0.98$.

$$S_n = \frac{\frac{1}{2}A \sum_{k=-n}^n j_k^2(\beta)}{\frac{1}{2}A \sum_{k=-\infty}^{\infty} j_k^2(\beta)} \quad (17)$$

$$S_n \geq 0.98 \quad (18)$$

Using the value properties of Bessel function,, and Table 1, we can show that the bandwidth of FM signal B_T , depends on the number of sidebands n , and FM modulation index β . which can be expressed as

$$B_T \approx 2(\beta + 1)f_m \quad (19)$$

ARBITRARY MODULATED FM SIGNAL:

- Narrow Band FM:

Narrowband *FM* is in many ways similar to *DSB* or *AM* signals. By way of illustration let us consider the *NBFM* signal

$$\begin{aligned} S_m(t) &= A \cos[\omega t + \phi(t)] = A \cos \omega t \cos \phi(t) - A \sin \omega t \sin \phi(t) \\ &\approx A \cos \omega t - A \phi(t) \sin \omega t \end{aligned} \quad (20)$$

Using the approximations $\cos \phi = 1$ and $\sin \phi \approx \phi$, when ϕ is very small. Equation-26 shows that a *NBFM* signal contains a carrier component and a quadrature carrier linearly modulated by (a function of) the baseband signal. Since $s(t)$ is assumed to be bandlimited to f_m therefore $\phi(t)$ is also bandlimited to f_m . Hence, the bandwidth of *NBFM* is $2f_m$, and the *NBFM* signal has the same bandwidth as an *AM* signal.

Narrow band FM modulator:

According to Equation-1.20, it is possible to generate *NBFM* signal using a system such as the one shown in Fig-4 . The signal is integrated prior to modulation and a DSB modulator is used to generate the quadrature component of the *NBFM* signal. The carrier is added to the quadrature component to generate an approximation to a true *NBFM* signal.

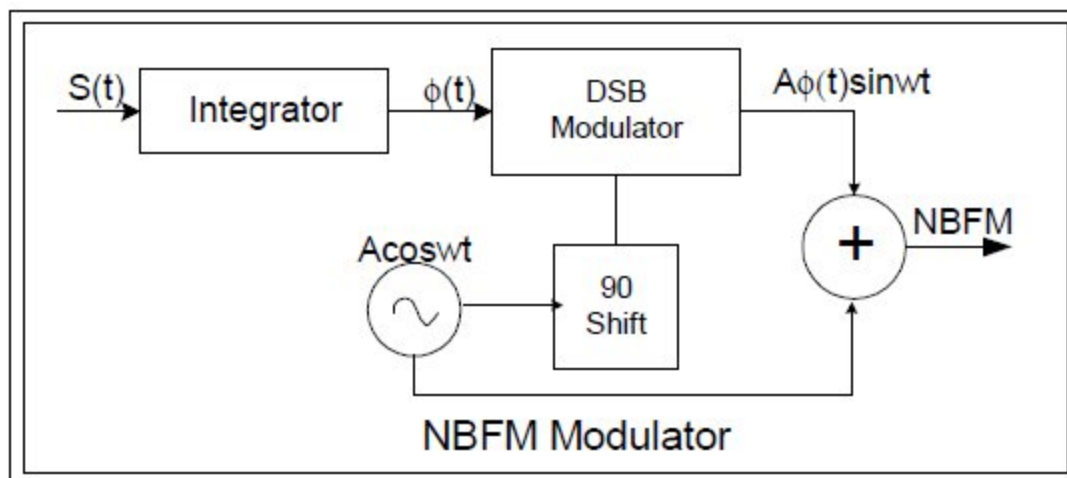


Fig 4 NBFM Modulator

Wideband FM modulator:

There are two basic methods for generating *FM* signals known as direct and indirect methods. The direct method makes use of a device called voltage controlled oscillator (*VCO*) whose oscillation frequency depends linearly on the modulation voltage.

A system that can be used for generating an *FM* signal is shown in Figure-5.

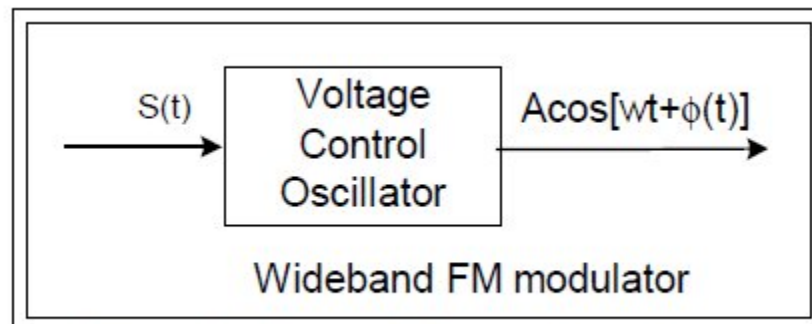


Fig. 5 VCO as wideband FM modulator

The combination of message differentiation that drive a *VCO* produces a *PM* signal. The physical device that generates the *FM* signal is the *VCO* whose output frequency depends directly on the applied control voltage of the message signal. *VCO*'s are easily implemented up to microwave frequencies.

FM MODULATORS AND DEMODULATORS:

- FM MODULATOR:

The other common form of analog data transmission that is familiar to most people is called frequency modulation (FM). In the previous section, the message signal was multiplied with the carrier, modulating its amplitude in order to be transmitted (this amplitude modulation) [1]. With FM, the frequency of the carrier is modulated (varied) as the message changes. FM is what is used to transmit the majority of radio broadcasts. It is generally preferred over AM because it is less sensitive to noise, and it is in fact possible to trade bandwidth for noise performance [2]. The advantages of FM come at the expense of increased complexity in the transmitter and in the receiver. This having been said, we will see a simple FM receiver is actually no more complicated (to build) than its AM counterpart [3].

FM Generation by parameter variation method:

Figure 6.6.5 shows the block diagram of an FM transmitter. The modulator circuit uses parameter-variation method. A pre-emphasis circuit is used to reduce the effect of noise at higher audio frequencies for threshold improvement. The details about this circuit are discussed in Sec. 6.10. Like the AM transmitter, the carrier oscillator generates sub-harmonic of final carrier frequency to achieve frequency stability. A stable oscillation frequency at a lower radio frequency (say 4 MHz) is generated by an oscillator and then it is raised to the final carrier (say 96 MHz) by frequency multipliers. It should be kept in mind that more frequency stability can be obtained if the carrier oscillator operates at low radio frequency. The multiplying circuit not only increases the carrier frequency, but also the frequency deviation by the same factor. This fact is proved in Ex. 6.6.3.

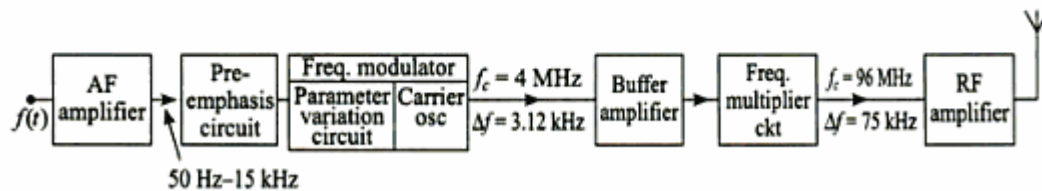


Fig. 6.6.5 Block Diagram of an FM Transmitter using Direct Modulation

We have seen that the direct method has the disadvantages of frequency instability. Variations in environmental conditions (temperature, humidity), supply voltage, aging of active devices etc. can also cause frequency instability. Therefore, FM transmitter must incorporate some auxiliary means for frequency stabilization. A stabilization scheme utilizing feedback principle is shown in Fig. 6.6.6. A stable crystal oscillator provides the reference frequency. The output of the crystal oscillator and the frequency modulator is fed into a mixture, and the difference frequency term is extracted at the output of the mixture. The mixer output then, is fed into a frequency discriminator. The frequency discriminator circuit provides an error voltage whose instantaneous value is proportional to the instantaneous frequency of its input. When the FM wave has frequency exactly equal to the assigned carrier frequency (no drift), the error signal is zero. But, whenever the transmitter carrier frequency drifts from the assigned value, a d.c. error signal of proper polarity is generated. The amplified error voltage of proper polarity is applied to a VCO in order to correct the transmitter frequency to the assigned value.

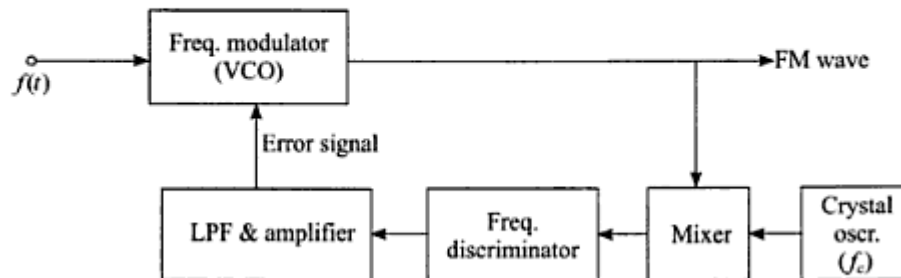


Fig. 6.6.6 A Frequency Stabilization Scheme

FM Generation by Armstrong's indirect method:

In the Armstrong method, frequency stability of a higher order can be obtained because the crystal oscillator can be used as a carrier generator. The basic principle of this method is to generate a narrowband FM (NBFM) indirectly by using the phase-modulation technique, and then converting this NBFM to a wideband FM (WBFM), as shown in Fig. 6.6.7. The distortion is low in NBFM as the modulation index is small. The phase modulation is preferred because of its easy generation schemes. The generation of NBFM is illustrated in Fig. 6.2.1. The multiplier circuit, apart from multiplying the carrier frequency, also increases the frequency deviation (refer Ex. 6.6.3), and thus the NBFM (small deviation) is converted into WBFM (large deviation).

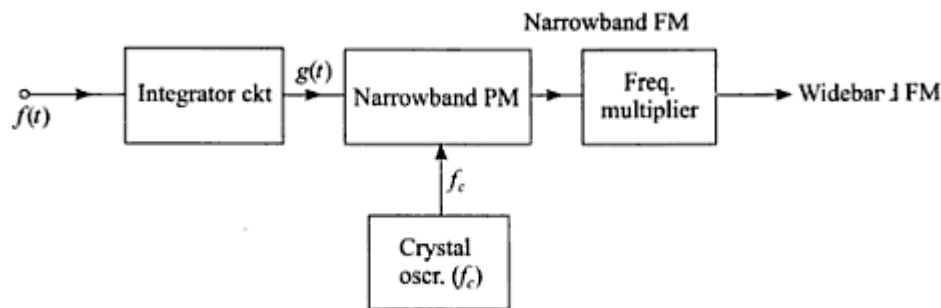


Fig. 6.6.7 Armstrong Method for FM Generation

- FM DEMODULATOR:

An FM demodulator is required to produce an output voltage that is linearly proportional to the input frequency variation. One way to realize the requirement is to use discriminators—devices which distinguish one frequency from another, by converting frequency variations into amplitude variations. The resulting amplitude changes are detected by an envelope detector, just as done by AM detector [4].

$$S_m(t) = A \cos[\omega t + k_f \int_{-\infty}^t s(\tau) d\tau]$$

the discriminator output will be

$$y_d(t) = k_d k_f s(t)$$

Where K_d is the discriminator constant. The characteristics of an ideal discriminator are shown in Fig. 6. Discriminator can be realized by using a filter in the stop band region, in a linear range, assuming that the filter is differentiation in frequency domain [4].

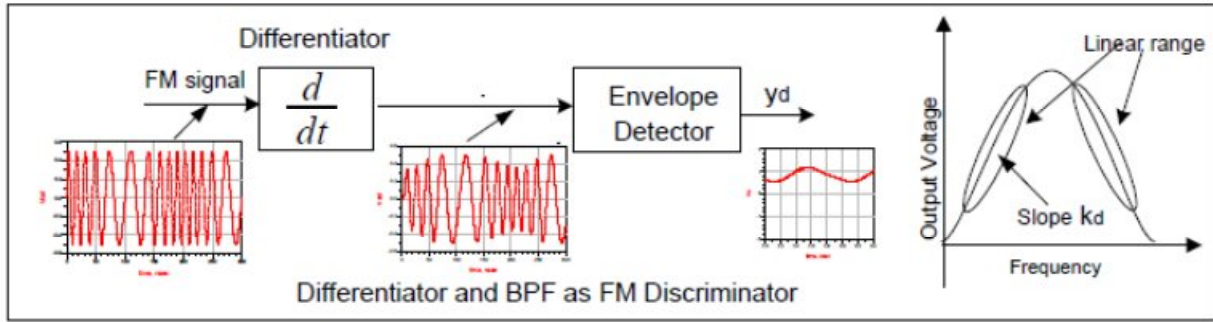


Fig 6 Ideal and real frequency demodulator

An approximation to the ideal discriminator characteristics can be obtained by the use of a differentiation followed by an envelope detector (see Figure-6) . If the input to the differentiator is $S_m(t)$, then the output of the differentiator is

$$S'_m(t) = -A[\omega + k_f s(t)] \sin[\omega t + \phi(t)] \quad (21)$$

With the exception of the phase deviation $\phi(t)$, the output of the differentiator is both amplitude and frequency modulated. Hence envelope detection can be used to recover the message signal. The baseband signal is recovered without any distortion if $\text{Max}\{k_f s(t)\} = 2\pi\Delta f < \omega$, which is easily achieved in most practical systems.

APPROXIMATELY COMPATIBLE SSB (CSSB) SYSTEMS:

- SSB-AM:

- i) Single sideband transmission has several advantages over the conventional amplitude modulation (AM) transmission. It needs less average transmitted power and its transmission band width is only half of that AM [1]. Smaller bandwidth implies conservation of spectrum as well as less susceptibility to frequency selective fading.
- ii) The drawback with SSB is the complexity and cost of the receiver and this makes unsuitable for broadcasting [1]. For conventional AM the broadcasting needs shortwave bands and mediums at its receiver.
- iii) AM receiver which employ simple envelope detector is known as compatible SSB signals. Although an exactly compatible SSB is not feasible hence an approximately compatible is produced i.e. its envelope may not be an exact replica of the modulating signal but its spectrum may differ slightly from the conventional SSB signal [2].
- iv) In this case when one of a side frequency components of a single tone modulated AM signal is suppressed the resultant signal is modulated both in amplitude and phase.
- v) If this approximately compatible system is used for speech or music as AM, then the low frequency fundamental components of the audio modulating signal having good number of harmonic components will fall within the allowed spectrum of the SSB signal.

- vi) The higher harmonic which falls outside the allowed spectrum may not be able to cause much of adjacent channel interference. Since the energy content in the higher harmonic is very small which is not very important.
- vii) Hence the performance of the compatible SSB system with ordinary AM receivers being used for reception which is much more superior than conventional AM [4].

- SSB-FM:

- i) A compatible SSB-FM signal has frequency components extending either above or below the carrier frequency. It can be demodulated by using limiter-discriminator [1].
- ii) The compatible FM signal is constructed by adding amplitude modulation to the frequency modulated signal.
- iii) The limiter removes any and all amplitude modulation which does not affect the recovery of the modulation discriminator by properly adjusting the amplitude modulation either the upper and lower sideband can be removed [1].

PULSE MODULATION AND DIGITAL TRANSMISSION OF ANALOG SIGNAL:

ANALOG TO DIGITAL

- NOISY CHANNEL:

- i) When a signal is transmitted over a long distance communication channel like radio, and when the signal arrives at the receiver it will greatly attenuated and combined with noise and electrical disturbances are also present in the channel receiver. As a result the received signal may not be distinguishable against its background of noise [4],
- ii) If a signal is transmitted over a small distance communication channel like in wires or coaxial cable, which does not provide complete freedom from crosstalk disturbances and electrical noise.
- iii) Hence the receiver noise is often noise source of largest power.

To resolve this problem is simply by raising the strength of the signal level at the transmitting end so high a level that in spite of the attenuation the received signal substantially overrides the noise [4]. This problem can be solved by a repeater (amplifier in a communication channel).

- ROLE OF REPEATER

- i) An amplifier at the receiver will not help the above situation, since at this point both signal and noise levels will be increased together. But a repeater (repeater is the term used for an amplifier in a communication channel) is located at the mid point of the long communications path [4]. The repeater will raise the signal level, and it will raise the level of only the noise introduced in the first half of the communication path.

- ii) Hence such a mid way repeater as contrasted with an amplifier at the receiver has the advantage of improving signal to noise ratio. The mid way repeater will receive burden imposed on transmitter and cable due to higher power requirements when the repeater is not used.
- iii) If the repeaters are connected at each point in between in a communication channel (cascaded connection) which serves to lower the maximum power level encountered on the communication link and improves the signal to noise level at the receiver [3].
- iv) Due to the role of repeater at the receiving end not only the gain of an amplifier increases and increase the signal to noise ratio but also increase the noise level at the receiving end.

Hence for remove this noise at the receiver we have to transmit the digital signal (through sampling) at the transmitter end and also connecting filters to remove the noise at the receiving end.

PULSE AMPLITUDE MODULATION AND CONCEPT OF TIME DIVISION MULTIPLEXING

A technique is called Time division multiplexing by which we may take advantage of sampling principle. Sampling of signals is needed for applications such as time sharing of a signal transmission facility, i.e. time division multiplexing (TDM) [2]. The signal samples may be used in analogue form, e.g. using pulse amplitude modulation (PAM), or they may be quantized and encoded into digital form, i.e. pulse code modulation (PCM).

Since the purpose of TDM is to exploit better the available performance of a signal transmission facility by squeezing as many channels into it as practically possible, this objective will be best achieved by using as few samples as possible per unit time per channel. For this reason it is important to know the minimum sample rate needed to avoid excessive distortion, and also to avoid placing unrealistic or costly demands on the performance of the functional elements of the system (e.g. transition bandwidth of anti aliasing and interpolation filters) [2]. These considerations apply not only to TDM systems, but in fact to all sampled data signal processing systems, which includes all digital processing of analogue data, e.g. digital filtering, digital spectrum analysis. The sampling theorem provides a theoretical basis for estimating a minimum sampling rate in these situations [4].

If two messages are sampled at the same rate but at slightly different times, then two of trains of samples can be added without mutual interaction. In Figure-1.a the signal $x(t)$ and the corresponding PAM signal are depicted. Also in Figure-1.b, TDM of two signals is shown:

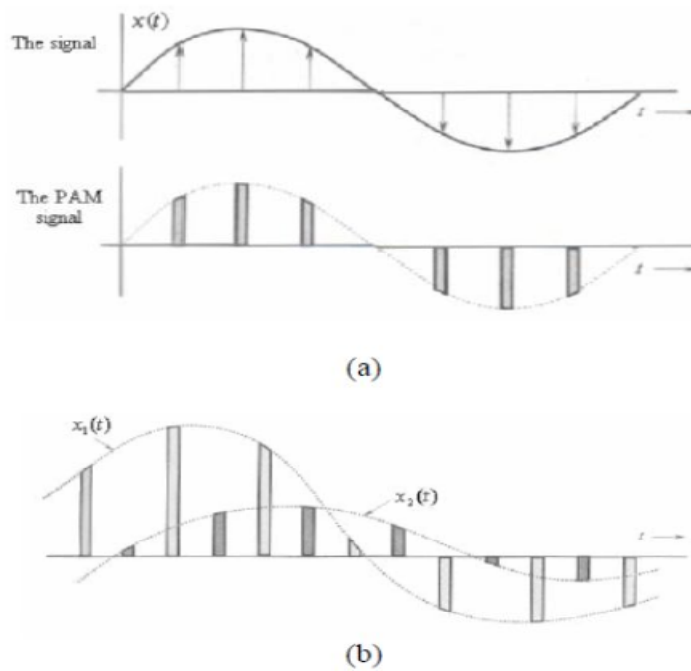
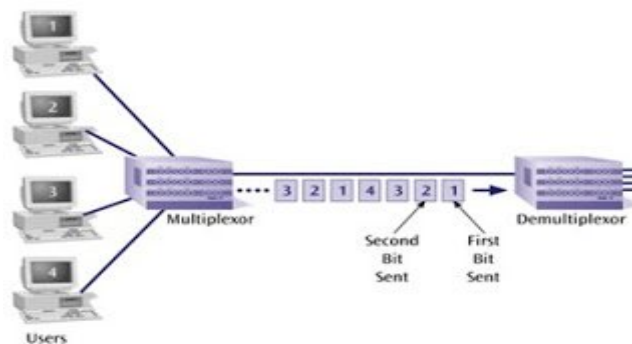


Figure 1. (a) The signal $x(t)$ and its PAM signal
(b) TDM of two signals

Concept of Time Division Multiplexing-TDM:

In TDM, sharing is accomplished by dividing available “transmission time” on a medium/channel among users. Each user of the channel is allotted a small time interval during which he transmits a message [3]. Total time available in the channel is divided, and each user is allocated a time slice. In TDM, users send message sequentially one after another. Each user can use the full channel bandwidth during the period he has control over the channel [4].



The block diagram of PAM/TDM signal production is given in Figure 2 below:

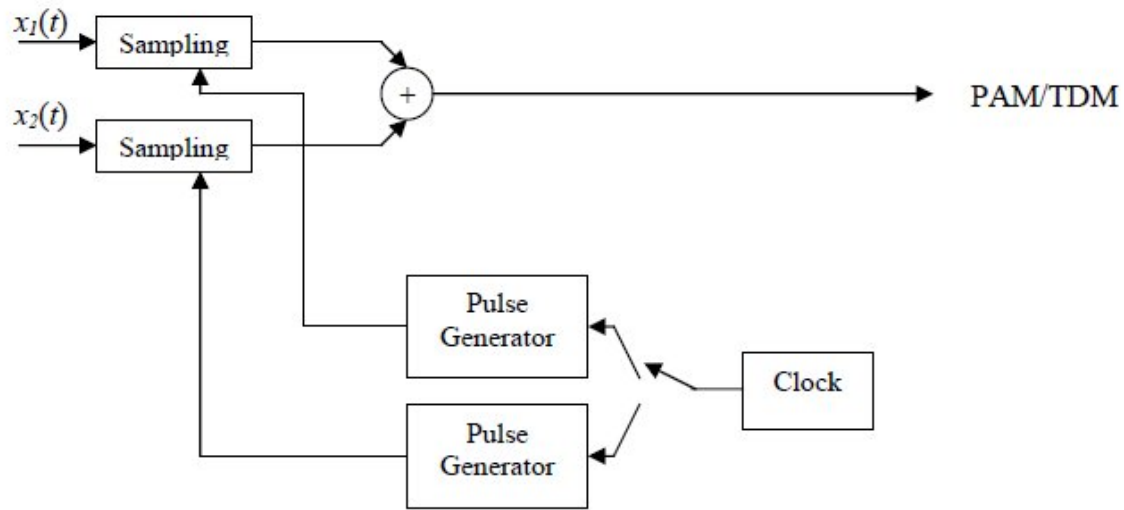


Figure 2. Production of TDM Signal

In Figure 2, if the band-widths of both signals ($x_1(t)$ and $x_2(t)$) is 3 kHz, according to the sampling theorem each signal should be sampled with the frequency of 6 kHz. But in this case the clock frequency should be 12 kHz. The distance between the pulses is $T_n = T_s/n$. Here, n indicates the number of the input signals, and T_s denotes the sampling period required for one signal. The obtained TDM wave is shown in Figure 3.

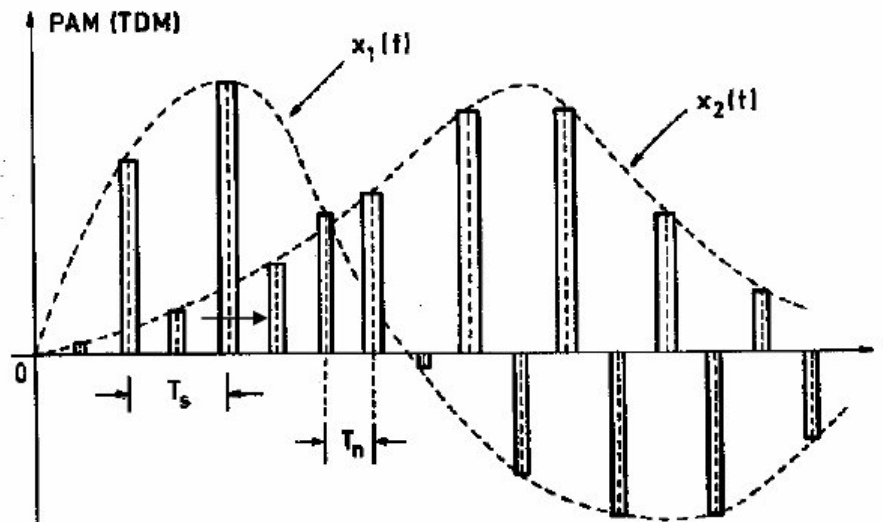
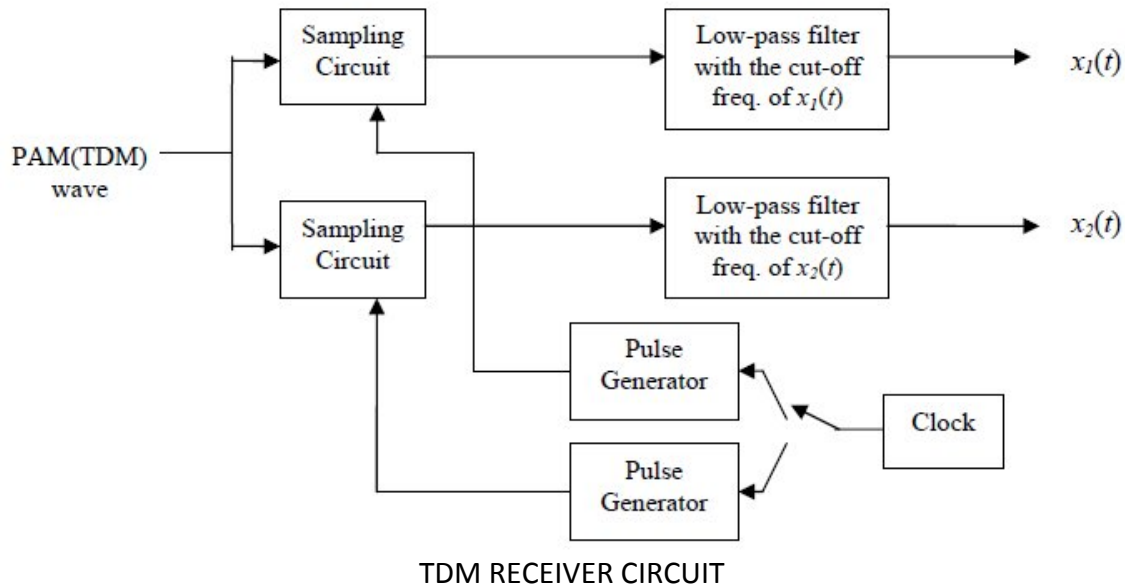


Figure 3. PAM(TDM) Wave

As considered before, if two different signals have the frequency of 3 kHz, the sampling period of each signal will be $T_s = 1/6000 = 166.7 \mu\text{sec}$. Since number of input signals $n=2$, from $T_n = T_s/n$;

the distance between samples becomes $T_n = T_s/2 = 83.3 \mu\text{sec}$. Wave minimum bandwidth to transmit these samples by TDM should be; $B \geq 166.7 \times 10^3 = 6 \text{ kHz}$.

A TDM receiver block diagram is shown in Figure 4 below. The most significant issue in recovery of input signals from TDM signals is the requirement for the proper synchronization between TDM transmitter and receiver. Therefore, the clock signal in the transmitter should be passed to the receiver correctly [4].



TYPES OF SAMPLING:

Sampling is the process of converting an analog signal to a discrete signal [1].

Generally there are two types of sampling are possible.

- i) Natural sampling
- ii) Flat-top sampling.

In Pulse-Amplitude Modulation (PAM) a pulse signal is used to sample an analog signal. The result is a train of constant-width pulses. The amplitude of each pulse is proportional to the amplitude of the message signal at the time of sampling [1]. Figure 2-2 and Figure 2-3 show the time domain appearance of a PAM signal for a triangle wave message signal. In the figures you can see that the PAM signal is made up of small segments (samples) of the message signal.

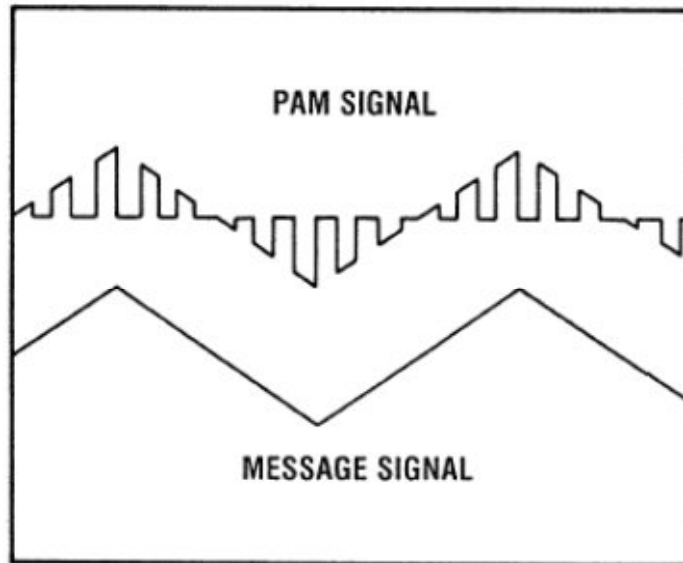


Figure 2-2. Natural sampling – time domain appearance of a PAM signal.

For a PAM signal produced with natural sampling, the sampled signal follows the waveform of the input signal during the time that each sample is taken. Flat-top sampling, however, produces pulses whose amplitude remains fixed during the sampling time. The amplitude value of the pulse depends on the amplitude of the input signal at the time of sampling [1].

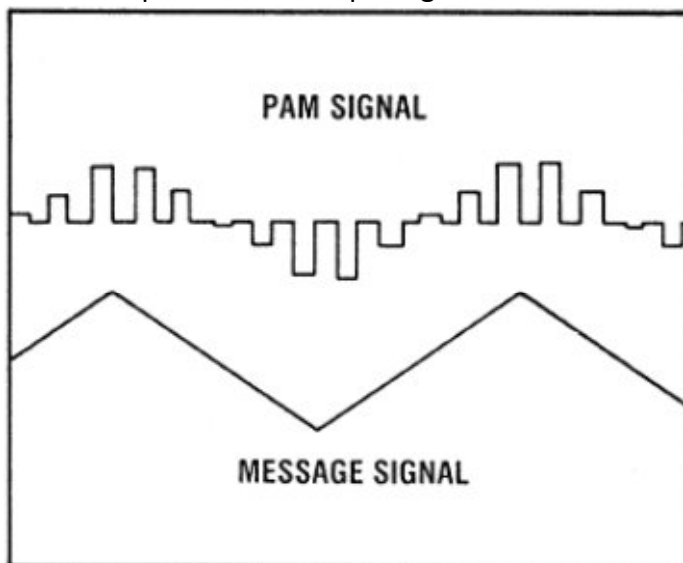


Figure 2-3. Flat-top sampling – time domain appearance of a PAM signal.

PAM signal generation:

A PAM signal is generated by using a pulse train, called the sampling signal (or clock signal) to operate an electronic switch or "chopper". This produces samples of the analog message signal [3], as shown in Figure 2-4.

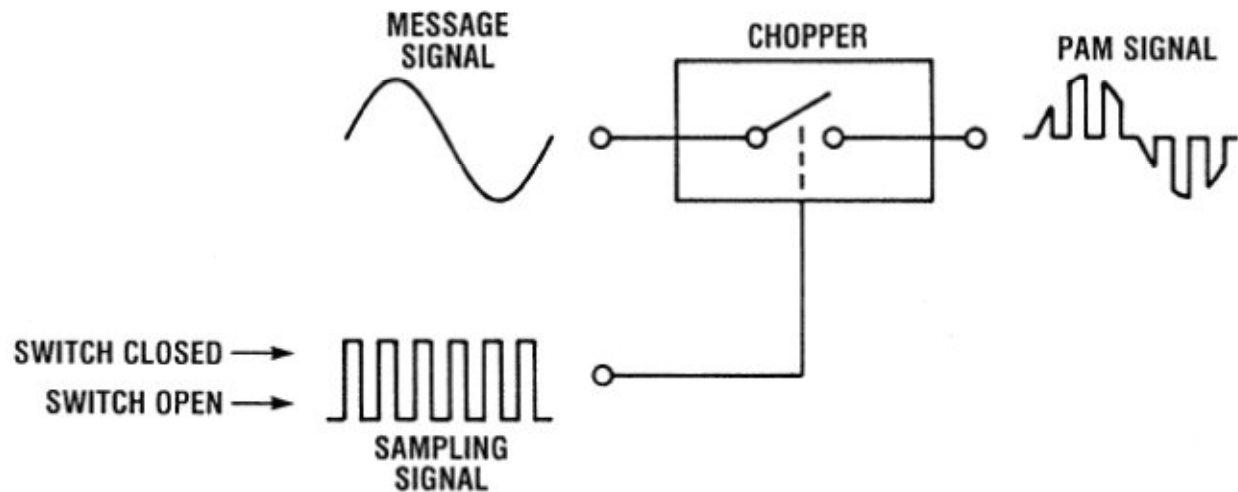


Figure 2-4. Natural sampling – generation of PAM signals.

The switch is closed for the duration of each pulse allowing the message signal at that sampling time to become part of the output. The switch is open for the remainder of each sampling period making the output zero. This type of sampling is called natural sampling [3].

For flat-top sampling, a sample-and-hold circuit is used in conjunction with the chopper to hold the amplitude of each pulse at a constant level during the sampling time, as shown in Figure 2-5.

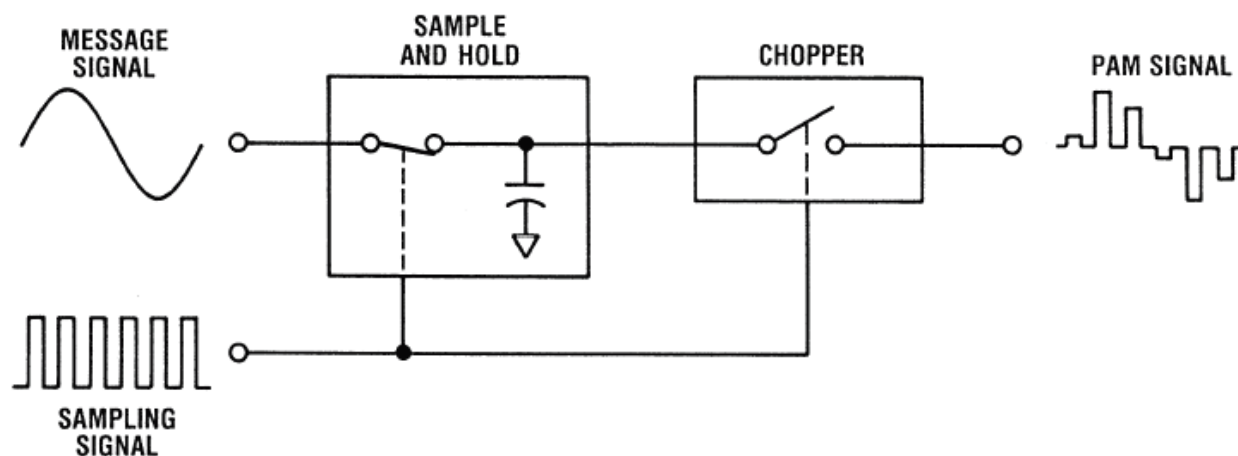


Figure 2-5. Flat-top sampling – generation of PAM signals.

Figure 2-6 shows the relationship between the message signal, the sampling signal, and the resulting PAM signal using natural sampling.

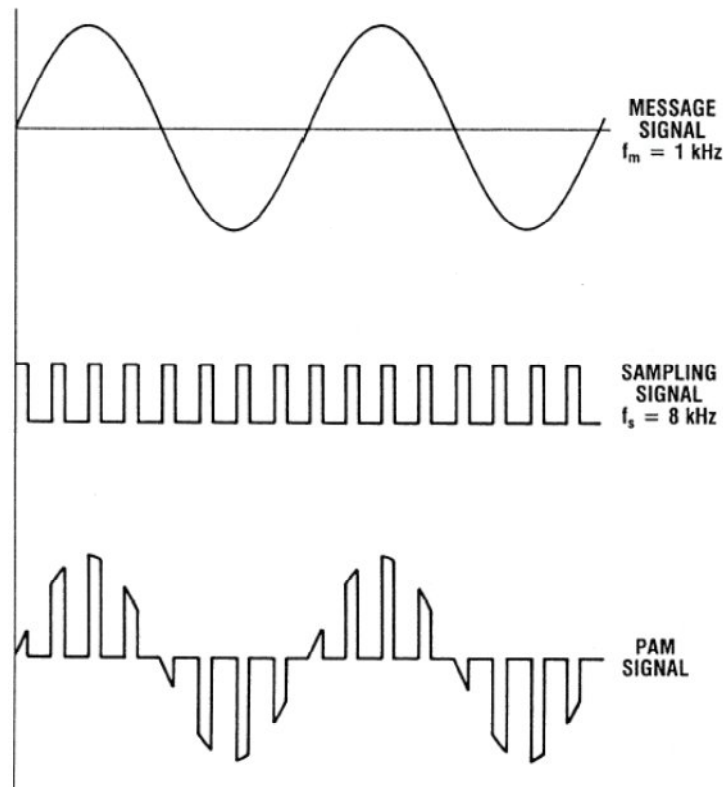


Figure 2-6. Natural sampling.

PULSE TIME MODULATION:

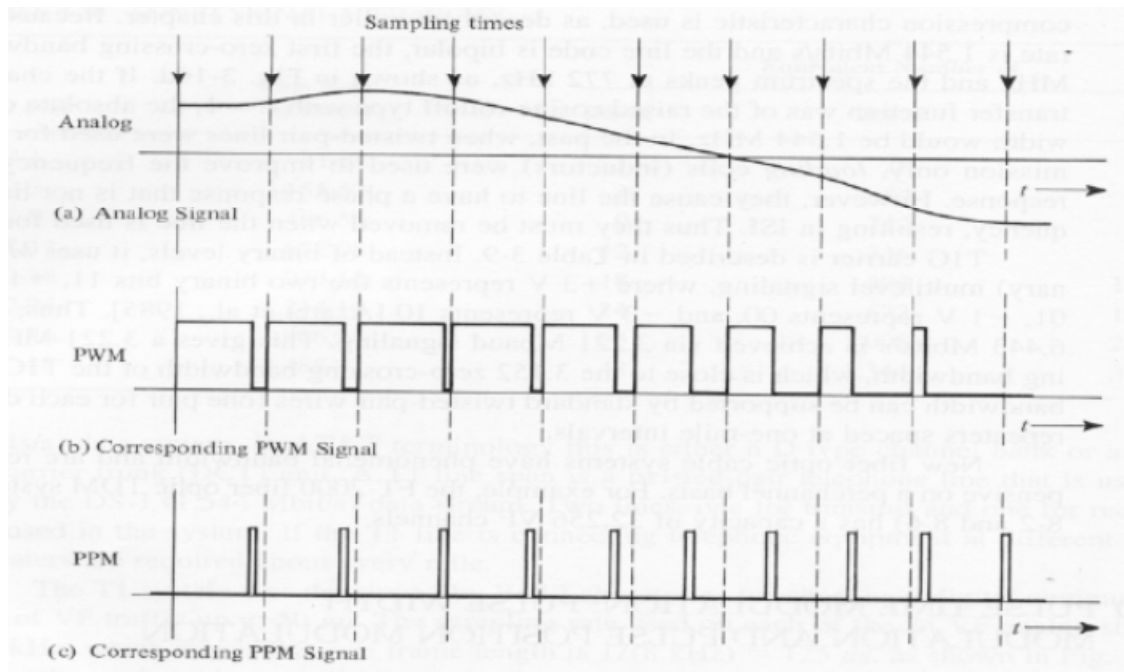
Pulse Time Modulation (PTM) is a class of signaling technique that encodes the sample values of an analog signal onto the time axis of a digital signal [3].

The two main types of pulse time modulation are:

1. Pulse Width Modulation (PWM)
2. Pulse Position Modulation (PPM)

In PWM the sample values of the analog waveform are used to determine the width of the pulse signal. Either instantaneous or natural sampling can be used [2].

In PPM the analog sample values determine the position of a narrow pulse relative to the clocking time [2]. It is possible to obtain PPM from PWM by using a mono-stable multivibrator circuit.



Pulse time modulation signaling

Pulse width Modulation:

This is also known as pulse duration modulation (PDM). Three variations of pulse width modulation are possible.

- i) In first variation the leading edge of the pulse is held constant and change in pulse width with signal is measured with respect to the leading edge [3].
- ii) In second variation the tail edge is held constant and the pulse width is measured with respect to it [3].
- iii) In third variation centre of the pulse is held constant and pulse width changes on either side of the centre of the pulse [3].

The modulating signal is at its positive peak at point (A) and its negative peak at (B). In this case figure 1 (a) the leading edge of the pulse is kept constant and pulse width is measured from the lead edge. From the above figure pulse width is maximum corresponding to point (A) while it is minimum at point (B).

In figure 1(b) the tail edge of the pulse is kept constant and pulse width is measured from the tail end of the pulse. Pulse width is maximum corresponding to positive peak of the modulating signal and minimum at the negative peak.

In figure 1(c) the centre of the pulse is kept constant and pulse extends on either side of the center of the pulse depending upon the modulating signal.

Generation of PWM /PPM Signal

It is basically a monostable multivibrator with a modulating input signal applied at the control voltage input. The control voltage is adjusted internally to $2/3 V_{cc}$ [1]. Externally applied modulating signal changes the control voltage and hence the threshold voltage level. The time period required to charge the capacitor up to threshold voltage level changes giving pulse modulated signal at the output as shown in figure.

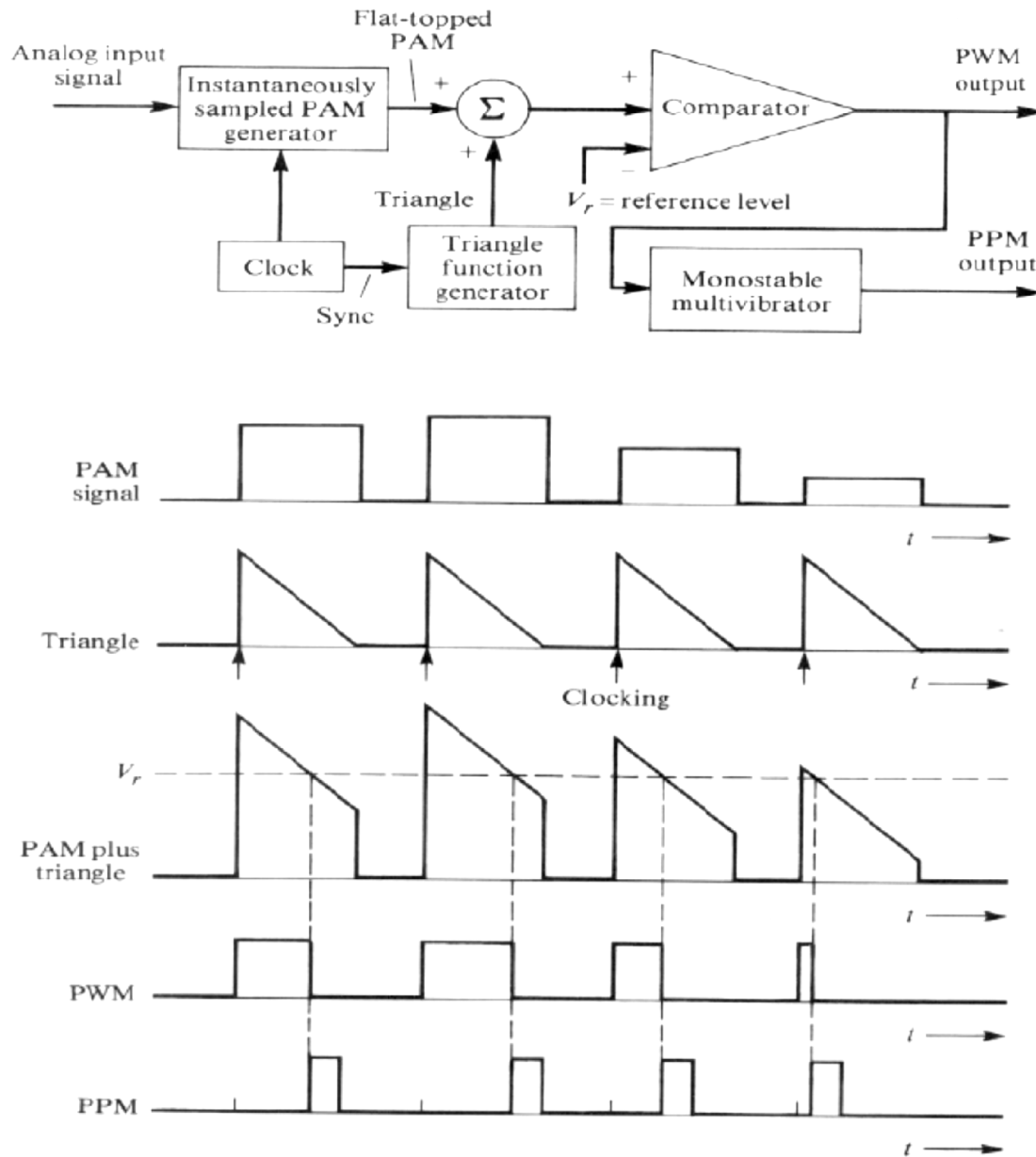


Figure 3-44 Technique for generating instantaneously sampled PTM signals.

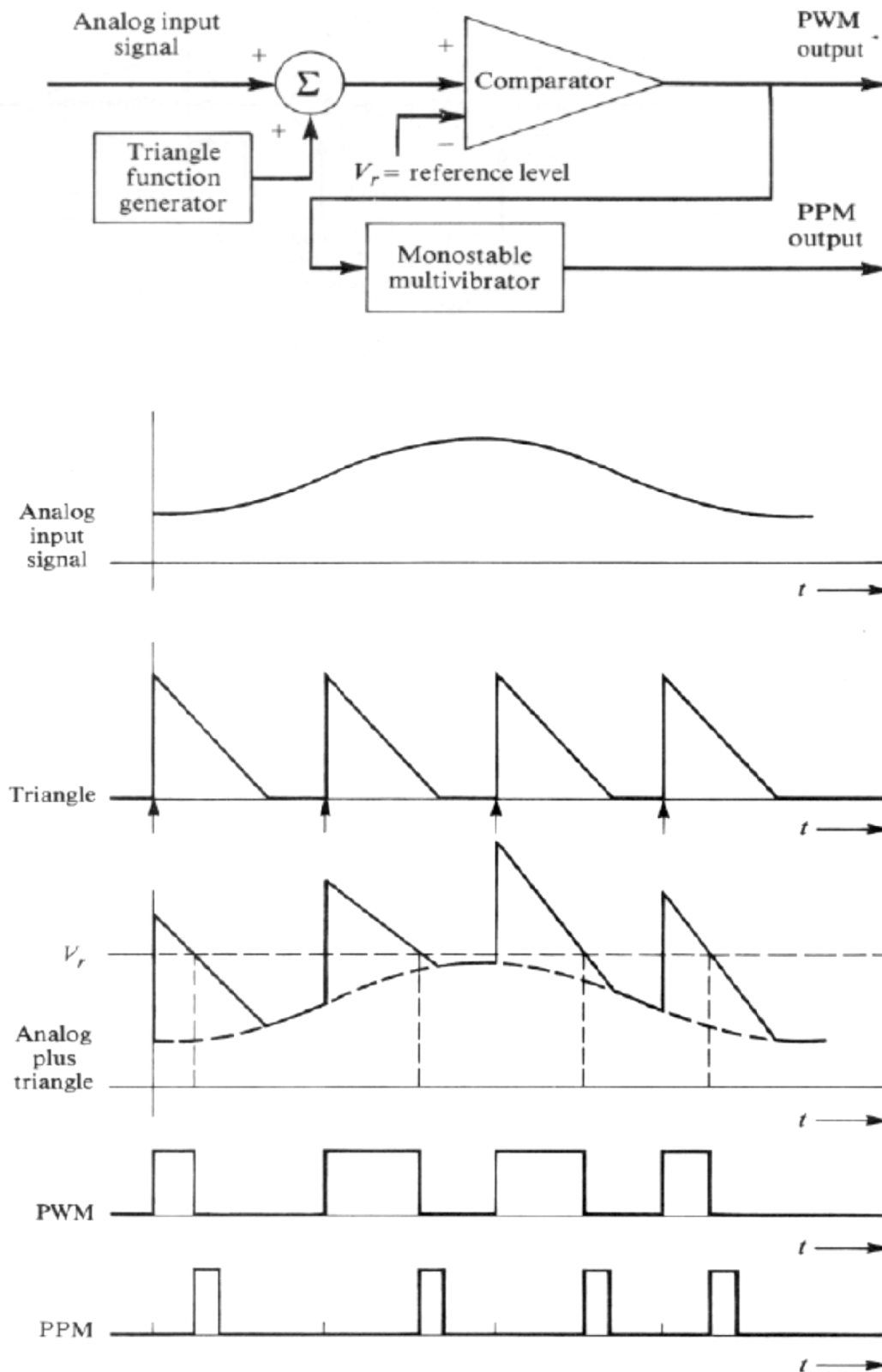


Figure 3-45 Technique for generating naturally sampled PTM signals.

The PWM or PPM signals may be converted back to the corresponding analog signal by a receiving system as shown in Fig. 3-46.

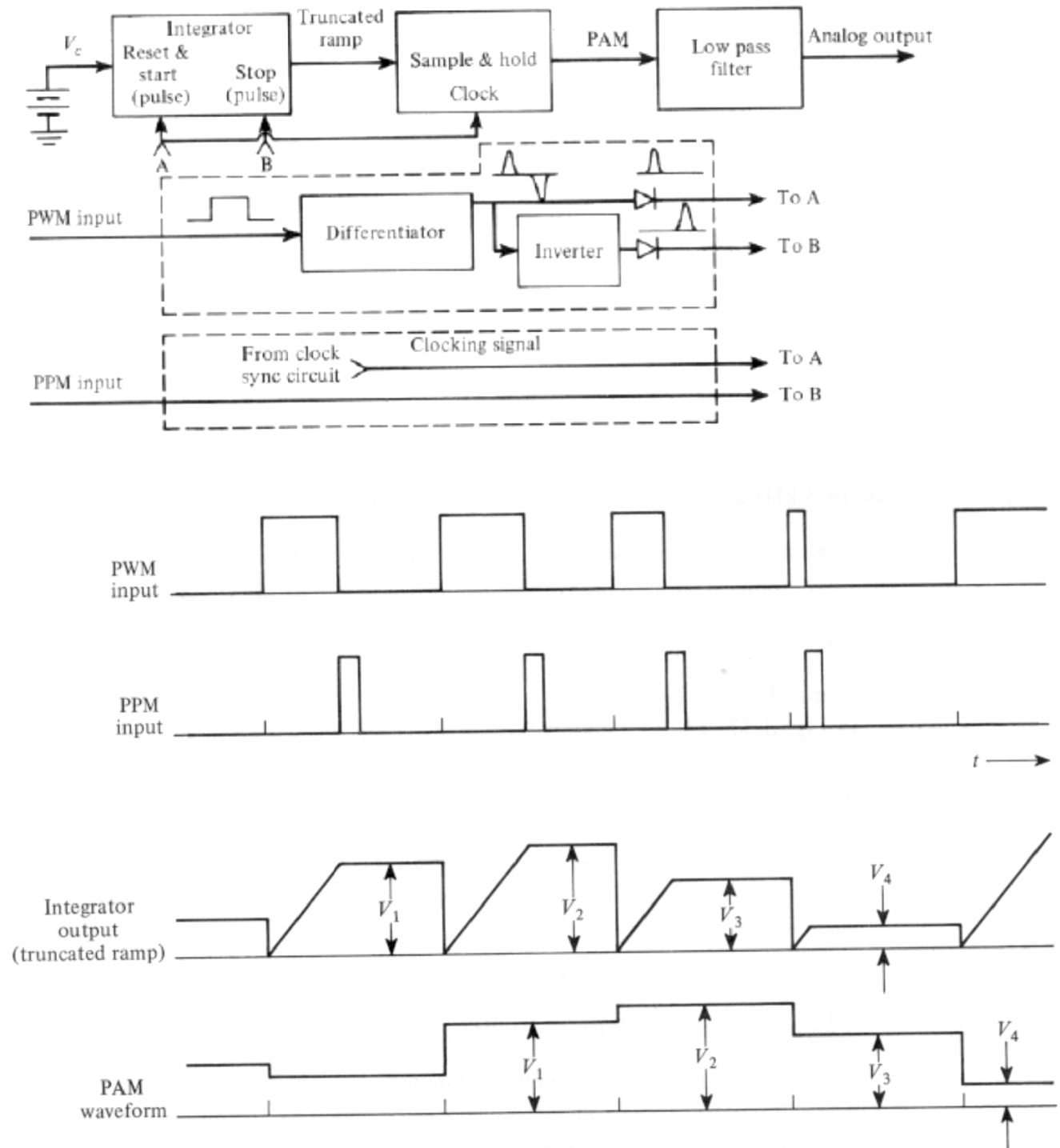


Figure 3-46 Detection of PWM and PPM signals.

For PWM detection the PWM signal is used to start and stop the integration of the integrator. After reset integrator starts to integrate during the duration of the pulse and will continue to do so till the pulse goes low. If integrator has a DC voltage connected as input, the output will be a truncated ramp. After the PWM signal goes low, the amplitude of the truncated ramp will be equal to the corresponding PAM sample value. Then it goes to zero with reset of the integrator [3].

DIGITAL REPRESENTATION OF ANALOG SIGNAL:

Digital representation of analog signal means to discretize the amplitude axis and then represent that signal in terms of binary digits. The process of converting an analog signal to digital signal is called quantization [1].

Quantization of signals:

Quantization is the process of mapping a large set of input values to a (countable) smaller set – such as rounding values to some unit of precision. A device or algorithmic function that performs quantization is called a quantizer [1].

Thus, quantization is the process of approximating a continuous range of values (or a very large set of possible discrete values) by a relatively- small set of discrete symbols or integer values. A common use of quantization is in the conversion of a discrete signal (a sampled continuous signal) into a digital signal by quantizing. Both of these steps (sampling and quantizing) are performed in analog-to-digital converters with the quantization level specified in bits [4]. A specific example would be compact disc (CD) audio which is sampled at 44,100 Hz and quantized with 16 bits (2 bytes) which can be one of 65,536 (i.e. 2^{16}) possible values per sample.

Mathematical description:

The simplest and best- known form of quantization is referred to as scalar quantization, since it operates on scalar (as opposed to multi- dimensional vector) input data [2]. In general, a scalar quantization operator can be represented as

$$Q(x) = g(\lfloor f(x) \rfloor)$$

where x is a real number to be quantized, $\lfloor \cdot \rfloor$ is the floor function, yielding an integer result $i = \lfloor f(x) \rfloor$ that is sometimes referred to as the *quantization index*, $f(x)$ and $g(i)$ are arbitrary real-valued functions.

The integer-valued quantization index i is the representation that is typically stored or transmitted, and then the final interpretation is constructed using $g(i)$ when the data is later interpreted.

In computer audio and most other applications, a method known as *uniform quantization* is the most common. If x is a real-valued number between -1 and 1, a mid-rise uniform quantization operator that uses M bits of precision to represent each quantization index can be expressed as

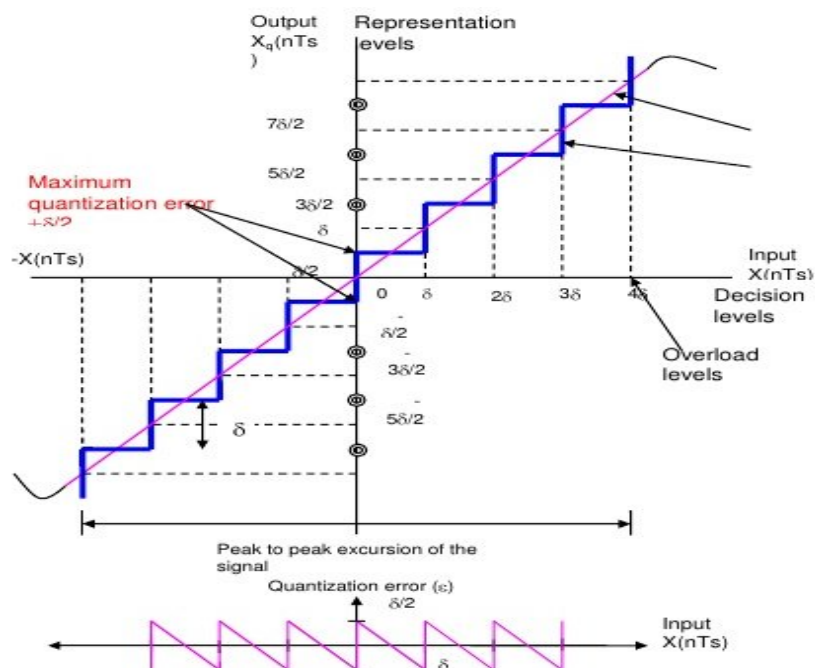
$$Q(x) = \frac{\lfloor 2^{M-1}x \rfloor + 0.5}{2^{M-1}}$$

The value $2^{-(M-1)}$ is often referred to as the *quantization step size*. Using this quantization law and assuming that quantization noise is approximately uniformly distributed over the quantization step size (an assumption typically accurate for rapidly varying x or high M) and further assuming that the input signal x to be quantized is

approximately uniformly distributed over the entire interval from -1 to 1, the signal to noise ratio (SNR) of the quantization can be computed as

$$\frac{S}{N_q} \approx 20 \log_{10}(2^M) = 6.0206M \text{ dB}$$

From this equation, it is often said that the SNR is approximately 6 dB per bit.

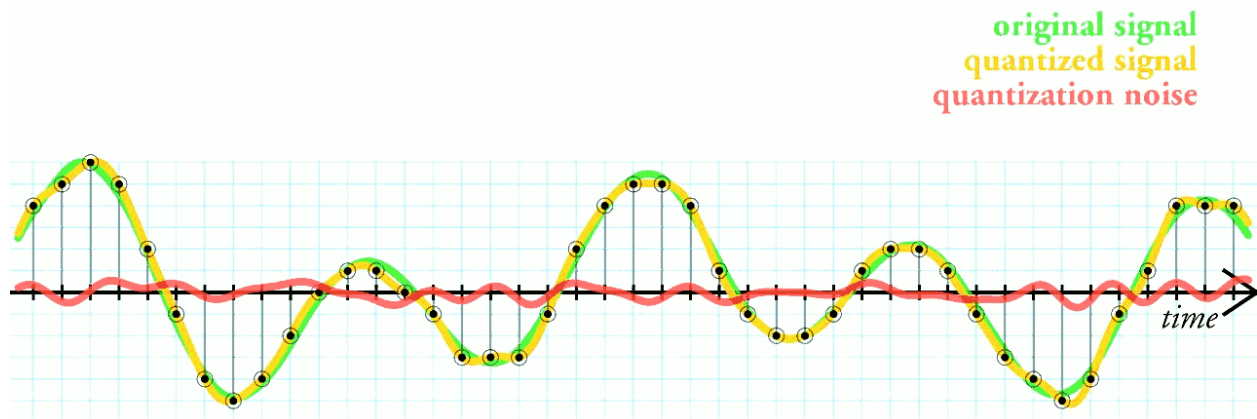


There are two substantially different classes of applications where quantization is used:

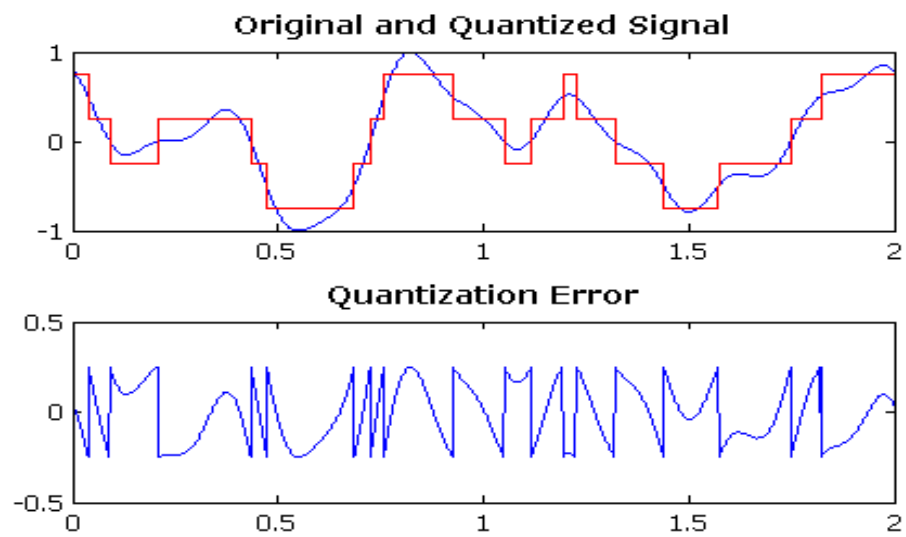
- The first type, which may simply be called rounding quantization, is the one employed for many applications, to enable the use of a simple approximate representation for some quantity that is to be measured and used in other calculations. This category includes the simple rounding approximations used in everyday arithmetic [3].
- The second type, which can be called rate–distortion optimized quantization, is encountered in source coding for "lossy" data compression algorithms, where the purpose is to manage distortion within the limits of the bit rate supported by a communication channel or storage medium [2].

Quantization Error:

In analog-to-digital conversion, the difference between the actual analog value and quantized digital value is called quantization error or quantization distortion. This error is either due to rounding or truncation [1]. The error signal is sometimes modeled as an additional random signal called quantization noise because of its stochastic behavior. Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form ordinarily involves rounding [2].



In the typical case, the original signal is much larger than one least significant bit (LSB). When this is the case, the quantization error is not significantly correlated with the signal, and has an approximately uniform distribution. In the rounding case, the quantization error has a mean of zero and the RMS value is the standard deviation of this distribution, given by $\frac{1}{\sqrt{12}}\text{LSB} \approx 0.289\text{LSB}$. In the truncation case the error has a non-zero mean of $\frac{1}{2}\text{LSB}$ and the RMS value is $\frac{1}{\sqrt{3}}\text{LSB}$. In either case, the standard deviation, as a percentage of the full signal range, changes by a factor of 2 for each 1-bit change in the number of quantize bits. The potential signal-to-quantization-noise power ratio therefore changes by 4, or $10 \cdot \log_{10}(4) = 6.02$ decibels per bit [3].



Module III

Noise:

Background

The performance of any communication system, is ultimately limited by two factors: (i) the transmission bandwidth, and (ii) the noise[3]. Bandwidth is a resource that must be conserved as much as possible, since only a finite electromagnetic spectrum is allocated for any transmission medium. Whatever the physical medium of the channel, the transmitted signal is corrupted in a random manner by a variety of possible mechanisms as it propagates through the system[4]. The term *noise* is used to denote the unwanted waves that disturb the transmission of signals, and over which we have incomplete control. As we will see throughout this course, bandwidth and noise are intimately linked. In this chapter our aim is to develop a model for the noise, with the ultimate goal of using this model in later chapters to assess the performance of various modulation schemes when used in the presence of noise.

A Model of Noise

Sources of noise

In a practical communication system, there are many sources of noise. These source may be external to the system (e.g., atmospheric, galactic, and synthetic noise) or internal to the system.[4]

Internal noise arises due to spontaneous fluctuation of current or voltage in electrical circuits, and consists of both *shot noise* and *thermal noise*.

Shot noise arises in electronic devices and is caused by the random arrival of electrons at the output of semiconductor devices. Because the electrons are discrete and are not moving in a continuous steady flow, the current is randomly fluctuating. The important characteristic of shot noise is that it is *Gaussian* distributed with zero mean (i.e, it has the Gaussian probability densityfunction shown in Figure). This follows from the *central limit theorem*, which states that the sum of η independent random variables approaches a Gaussian distribution as $\eta \rightarrow \infty$. In practice, engineers and statisticians usually accept that the theorem holds when $\eta \geq 6$.

Thermal noise is associated with the rapid and random motion of electrons within a conductor due to thermal agitation[4]. It is also referred to as Johnson noise, since it was first studied experimentally in 1928 by Johnson, who showed that the average power in a conductor due to thermal noise is

$$P_{\text{thermal}} = kTB$$

where k is Boltzman's constant (1.38×10^{-23}), T is the absolute temperature in Kelvin, and B is the bandwidth in hertz. Again, because the number of electrons in a conductor is very large, and their random motions are statistically independent, the central limit theorem indicates that thermal noise is Gaussian distributed with zero mean.

The noise power from a source (not necessarily a thermal source) can be specified by a number called the *effective noise temperature*:

$$T_e = \frac{P}{kB}$$

Effective noise temperature can be interpreted as the temperature of a fictitious thermal noise source at the *input*, that would be required to produce the same noise power at the *output*. Note that if the noise source is not thermal noise, then T_e may have nothing to do with the physical temperature of the device[4].

The important thing to note from this section is that *noise is inevitable*.

The additive noise channel

The simplest model for the effect of noise in a communication system is the additive noise channel, shown in Fig. Using this model the transmitted signal $s(t)$ is corrupted by the addition of

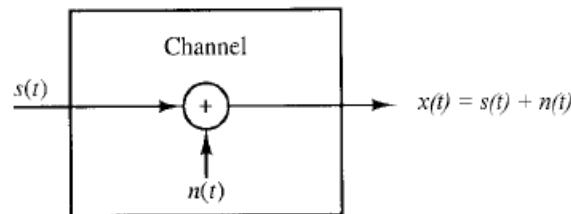


Figure: Additive noise Channel

a random noise signal $n(t)$. If this noise is introduced primarily by electronic components and amplifiers at the receiver, then we have seen that it can be characterized statistically as a Gaussian process. It turns out that the noise introduced in most physical channels is (at least approximately) Gaussian, and thus, this simple model is the predominant one used in communication system analysis and design[3].

A Statistical Description of Noise

As we have already hinted at, noise is completely random in nature. The noise signal $n(t)$ is a time-varying waveform, however, and just like any other signal it must be affected by the system through which it passes. We therefore need a model for the noise that allows us to answer questions such as: How does one quantitatively determine the effect of systems on noise? What happens when noise is picked up at the receiver and passed through a demodulator? And what effect does this have on the original message signal? Here we seek a representation that will enable us to answer such questions in the following chapters.

Background on Probability

Before developing a mathematical model for the noise, we need to define a few terms[3].

Random Variable

Consider a *random experiment*, that is, an experiment whose outcome cannot be predicted precisely.

The collection of all possible separately identifiable outcomes of a random experiment is

called the *sample space*, \mathcal{S} . A *random variable* is a rule or relationship (denoted by x) that assigns a real number x_i to the i^{th} sample point in the sample space. In other words, the random variable x can take on values $x_i \in \mathcal{S}$. The *probability* of the random variable x taking on the value x_i is denoted $P_x(x_i)$.

Cumulative and Probability Density Functions

The *cumulative density function* (cdf) of a random variable x is

$$F_x(x) = P_x(X \leq x)$$

The *probability density function* (pdf) of a random variable x is

$$P_x(x) = \frac{d}{dx} F_x(X)$$

Note that we use upper case P to denote probability, and lower case p to denote a pdf. We also have that

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} P_1(X) dx$$

One specific pdf that we will be particularly interested in is the *Gaussian* pdf, defined as

$$P_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

where m is the *mean* of the distribution and σ^2 is the *variance*. This pdf is shown in Figure.

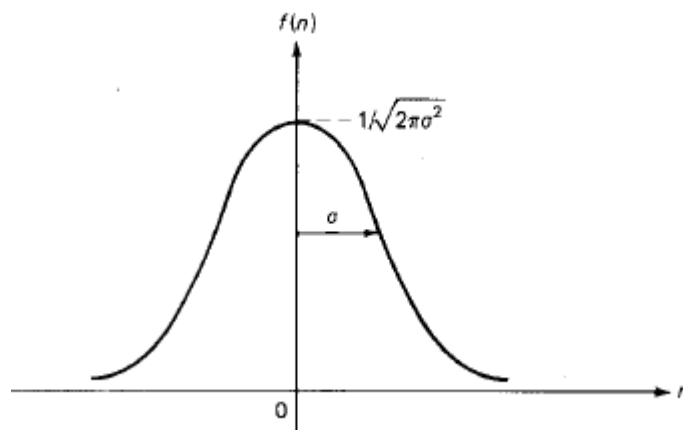


Figure: Gaussian Probability Density Function

Statistical Averages

One is often interested in calculating averages of a random variable. Denote the *expected value* (or mean value) of a random variable x as $E\{x\}$. If the random variable x has a pdf $p_x(x)$, then the expected value is defined as

$$E\{x\} = \int_{-\infty}^{\infty} x p_x(x) dx$$

where $E\{\cdot\}$ denotes the *expectation operator*.

It is also often necessary to find the mean value of a function of a random variable, for example, the mean square amplitude of a random signal. Suppose we want to find $E\{y\}$ where y is a random variable given by

$$y = g(x);$$

where x is a random variable with a known pdf, and $g(\cdot)$ is an arbitrary function. Then,

$$E\{y\} = E\{g(x)\} = \int_{-\infty}^{\infty} g(x) p_x(x) dx$$

The *variance* of a random variable x is defined as

$$\begin{aligned}\sigma_x^2 &= E\{(x - E\{x\})^2\} \\ &= E\{x^2 - E^2\{x\}\}\end{aligned}$$

Note that for a zero-mean random variable, the variance is equal to the mean square.

Random Processes

A random time-varying function is called a *random process*[4]. Let $n(t)$ be such a process. A sample n of $n(t)$ taken at any time t is a random variable with some probability density function. Here we will only consider *stationary* processes, where a stationary process is defined as one for which a sample taken at time t is a random variable whose pdf is independent of t .

Recall that one can view a random variable as an outcome of a random experiment, where the outcome of each trial is a number. Similarly, a random process can be viewed as a random experiment where the outcome of each trial is a waveform that is a function of time. The collection of all possible waveforms is the *ensemble* of the random process.

Two averages exist for random processes. Consider the ensemble of noise processes shown in Figure. Each waveform in this figure represents a different outcome of a random experiment

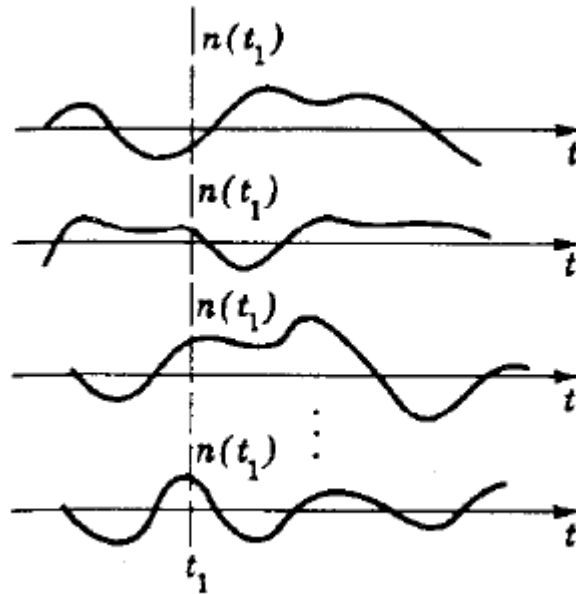


Figure. Ensemble Averages

Time Average

For a specific waveform, one could find the *time average*, defined as

$$\langle n(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} n(t) dt$$

where $\langle . \rangle$ denotes the time average operator. Note that average power (1.2) is just the time average of the magnitude-square of the waveform[3].

Ensemble Average

Alternatively, one could pick a specific time and average across all sample functions of the process at that time. This would give the *ensemble average*

$$E\{n(t)\} = \int_{-\infty}^{\infty} n P_n(n) dn$$

Comments

Note that the mean value $E\{n(t)\}$ locates the center of gravity of the area under the pdf. Random processes for which the time-average and the ensemble-average are identical are called *ergodic* processes. All the random processes encountered in this course will be assumed to be ergodic. The importance of this is that time averages can be easily measured, whereas ensemble averages cannot.

We can now interpret some engineering measures of the noise in terms of statistical quantities:

$$\begin{aligned} \text{DC component: } E\{n(t)\} &= \langle n(t) \rangle \\ \text{Average power: } E\{n^2(t)\} &= \langle n^2(t) \rangle \end{aligned}$$

Notice that for a zero-mean process, the variance is equivalent to the average power, i.e., $\sigma^2 = E\{n^2(t)\}$. This could be measured in the lab using a power metre[3].

Autocorrelation and Power Spectral Density

To understand bandwidth issues relating to random signals, we must now find a reasonable spectral representation of $n(t)$. In particular, we are interested in a frequency representation that reflects an ensemble average of all possible random processes[3].

Autocorrelation

The frequency content of a process depends on how rapidly the amplitude changes as a function of time. This can be measured by correlating the amplitudes at times t_1 and t_2 . Define the *autocorrelation* of a real random process as

$$R_x(t_1, t_2) = E\{x(t_1)x(t_2)\}$$

For a stationary process, the autocorrelation depends only on the time difference, so

$$R_x(\tau) = E\{x(t)x(t+\tau)\}$$

Recall that the average power of a waveform is the mean square. Hence,

$$P = E\{x^2(t)\} = R_x(0)$$

Power Spectral Density

Power spectral density (PSD) is a function that measures the distribution of power of a random signal with frequency. To illustrate the idea, consider a power meter tuned to a frequency f_0 that measures the power in a very narrow bandwidth around f_0 ; the output of this metre would give a good approximation to the PSD at the frequency f_0 . PSD is only defined for stationary signals.

Theorem (Wiener-Khinchine Theorem)

The power spectral density of a random process is defined as the Fourier transform of the autocorrelation[3]:

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-2\pi f \tau} d\tau$$

Since the autocorrelation is thus given by the inverse Fourier transform of the PSD, it follows from the equation that the average power of a random process can be found by integrating the PSD overall frequencies:

$$P = R_x(0) = \int_{-\infty}^{\infty} S_x(f) df$$

One particular example of a PSD that plays an extremely important role in communications and signal processing is one in which the PSD is constant over all frequencies, i.e.,

$$S(f) = \frac{N_0}{2}$$

Noise having such a PSD is referred to as *white noise*, and is used in the same sense as white light which contains equal amounts of all frequencies within the visible band of electromagnetic radiation. Note that the factor $1/2$ is included to indicate that half the power is associated with

positive frequency and half with negative.

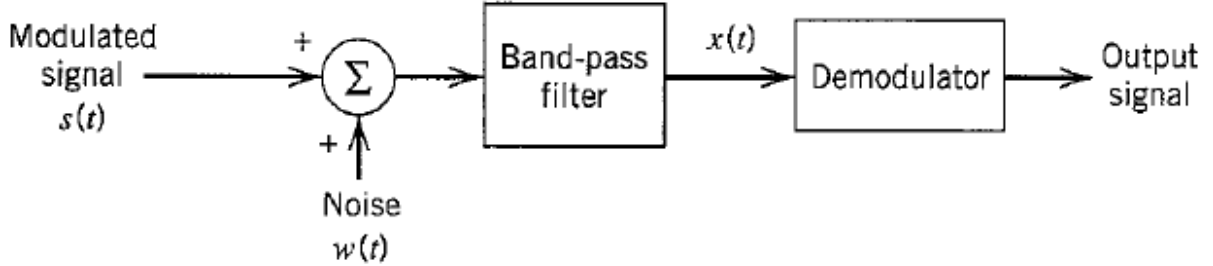


Figure. Receiver Model

Representation of Bandlimited Noise

Development

Any communication system that uses carrier modulation will typically have a bandpass filter at the front-end of the receiver see Fig . This filter will have a bandwidth wide enough to pass the modulated signal of interest, and is designed to restrict out-of-band noise from entering the receiver[4]. Any noise that does enter the receiver will therefore be bandpass in nature, i.e., its spectral magnitude is non-zero only for some band concentrated around the carrier frequency f_c . For example, if white noise have a PSD of $N_0/2$ is passed through such a filter, then the PSD of the noise that enters the receiver is given by

$$S(x) = \begin{cases} \frac{N_0}{2}, & f_c - w \leq |f| \leq f_c + w \\ 0, & \text{otherwise} \end{cases}$$

and is shown in Fig.. We are now in a position to develop a representation specifically for such bandpass noise. To achieve this, we will use a simple artifice, namely, to visualize the noise as being composed of the sum of many closely spaced randomly-phased sine waves.

Consider the bandpass noise signal $n(t)$, whose PSD is given by equation and is shown in Fig. The average noise power in the frequency slices f at frequencies f_k and $-f_k$, is

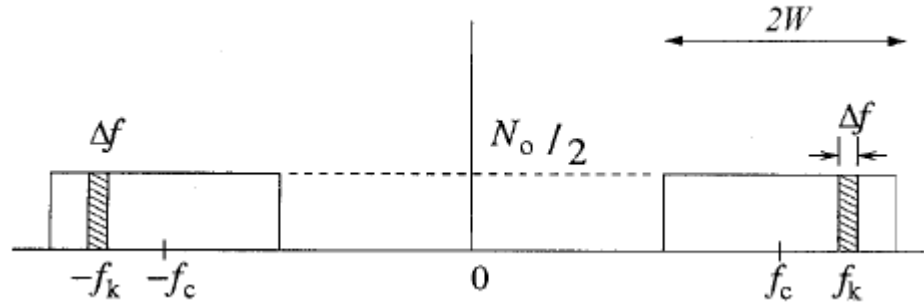


Figure. Power Spectral Density of band limited white noise process $n(t)$

found to be

$$P_k = 2 \frac{N_0}{2} \Delta f$$

where the factor of 2 is present because we are summing the slices at negative and positive frequencies.

For Δf small, the component associated with this frequency interval can be written

$$n_k(t) = a_k \cos(2\pi f_k t + \theta_k)$$

where θ_k is a random phase assumed independent and uniformly distributed in the range $[0; 2\pi)$, and a_k is a random amplitude. It can be shown that the average power of the randomly-phased sinusoid is

$$P_k = E\left\{\frac{a_k^2}{2}\right\}$$

The complete bandpass noise waveform $n(t)$ can be constructed by summing up such sinusoids over the entire band, i.e.,

$$\begin{aligned} n(t) &= \sum_k n_k(t) \\ &= \sum_k a_k \cos(2\pi f_k t + \theta) \end{aligned}$$

where

$$f_k = f_c + k\Delta f$$

Now, let $f_k = (f_c + f_c) + f_c$, and using the identity for the $\cos(\cdot)$ of a sum^k we obtain the required result.

Result

$$n(t) = n_c(t) \cos 2\pi f_c t - n_s(t) \sin 2\pi f_c t$$

where

$$\begin{aligned} n_c(t) &= \sum_k a_k \cos(2\pi(f_k - f_c)t + \theta_k) \\ n_s(t) &= \sum_k a_k \sin(2\pi(f_k - f_c)t + \theta_k) \end{aligned}$$

From equation we see that $f_k - f_c = k\Delta f$. Hence, $n_c(t)$ and $n_s(t)$ are *baseband* signals. The representation for $n(t)$ given by equation is the representation we seek, and is referred to as the *bandpass representation*. Although we have derived it for the specific case of a bandlimited white noise process, it is actually a very general representation that can be used for *any* bandpass signal.

Average power and power spectral density

If this representation of bandpass noise is to be of use in our later analyses, we must find suitable

statistical descriptions[3]. Hence, we will now derive the average power in $n(t)$, together with the average power and PSD for both $n_s(t)$ and $n_c(t)$.

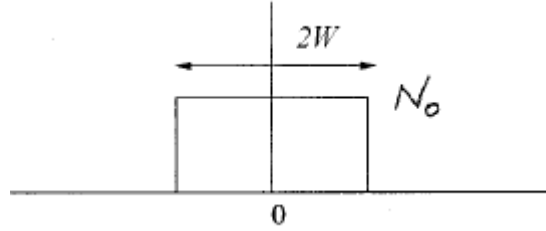


Figure: Power spectral density of each of the baseband noise processes $n_c(t)$ and $n_s(t)$

The average power in $n(t)$ is $P_n = E\{n^2(t)\}$. Recall that for a zero-mean Gaussian process the average power is equal to the variance σ^2 . Substituting in equation yields

$$\begin{aligned} P_n &= E\{n^2(t)\} \\ &= E\left\{\sum_k a_k \cos(2\pi f_k t + \theta_k) \sum_l a_l (\cos 2\pi f_l t + \theta_l)\right\} \\ &= \sum_k \sum_l E\{a_k a_l \cos(2\pi f_k t + \theta_k) \cos 2\pi f_l t + \theta_l\} \end{aligned}$$

Since we have assumed the phase terms are independent, it follows that

$$E\{\cos(2\pi f_k t + \theta_k) \cos(2\pi f_l t + \theta_l)\} = 0 \text{ for } k \neq l$$

and

$$E\{\cos(2\pi f_k t + \theta_k) \cos(2\pi f_l t + \theta_l)\} = E\{\cos^2(2\pi f_k t + \theta_k)\} = \frac{1}{2} \text{ for } k = l$$

Hence,

$$P = E\{n(t)^2\} = \sum_k \frac{E\{a_k^2\}}{2} = \sigma^2$$

This is what you should intuitively expect to obtain, given in equation. A similar derivation for each of $n_c(t)$ and $n_s(t)$ reveals that

$$P_c = E\{n_c(t)^2\} = \sum_k \frac{E\{a_k^2\}}{2} = \sigma^2$$

and

$$P_s = E\{n_s(t)^2\} = \sum_k \frac{E\{a_k^2\}}{2} = \sigma^2$$

Thus, the average power in *each* of the baseband waveforms $n_c(t)$ and $n_s(t)$ is *identical* to the average power in the bandpass noise waveform $n(t)$.

Now, considering the PSD of $n_c(t)$ and $n_s(t)$, we note from above equations that each of these waveforms consists of a sum of closely spaced baseband sinusoids. Thus, each baseband noise waveform will have the same PSD, which is shown in Fig.. Since the average power in each of the baseband waveforms is the same as the average power in the bandpass waveform, it follows that the area under the PSD in Fig. must equal the area under the PSD in Fig. The PSD of $n_c(t)$ and $n_s(t)$ is therefore given by

$$S_c(f) = S_s(f) = \begin{cases} N_0, & |f| < w \\ 0, & \text{otherwise} \end{cases}$$

A phasor interpretation

Finally, we will interpret the bandpass representation in another way. Notice that equation can be written

$$n(t) = \text{Re}\{g(t)e^{j2\pi f_c t}\}$$

where

$$g(t) = n_c(t) + jn_s(t)$$

and $\text{Re}\{.\}$ denotes the real part. We can also write $g(t)$ in terms of magnitude and phase as

$$g(t) = r(t)e^{j\phi(t)}$$

where

$$r(t) = \sqrt{n_c(t)^2 + n_s(t)^2}$$

is the envelope and $\phi(t) = \tan^{-1}[n_s(t)/n_c(t)]$ is the phase of

the noise. The phasor diagram representation[2] is shown in Fig. Because of this represen

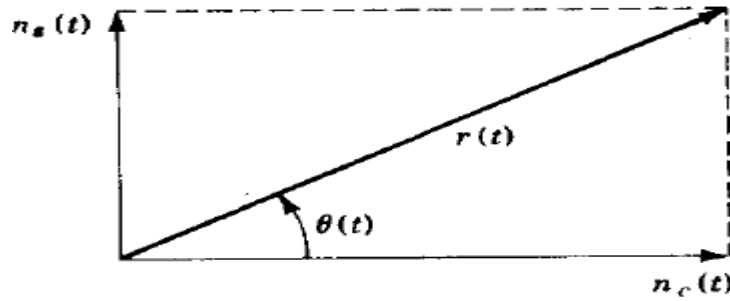


Figure. Phasor representation of bandlimited noise

tation, $n_c(t)$ is often referred to as the *in-phase* component, and $n_s(t)$ as the *quadrature-phase* component. Substituting the magnitude-phase representation for $g(t)$ gives

$$n(t) = r(t) \cos[2\pi f_c t + \phi(t)]$$

This is an intuitively satisfying result. Since the bandpass noise $n(t)$ is narrow band in the vicinity of f_c , one would expect it to be oscillating on the average at f_c . It can be shown that if $n_c(t)$ and $n_s(t)$ are Gaussian-distributed, then the magnitude $r(t)$ has a Rayleigh distribution, and the phase $\phi(t)$ is uniformly distributed.

Noise in Analog Communication Systems

Background

You have previously studied ideal analog communication systems. Our aim here is to compare the performance of different analog modulation schemes in the presence of noise[4]. The performance will be measured in terms of the signal-to-noise ratio (SNR) at the output of the receiver, defined as

$$\text{SNR}_o = \frac{\text{average power of message signal at reciver output}}{\text{average power of noise atthe reciver output}}$$

Note that this measure is unambiguous if the message and noise are additive at the receiver output;
we will see that in some cases this is not so, and we need to resort to approximation methods to obtain a result.

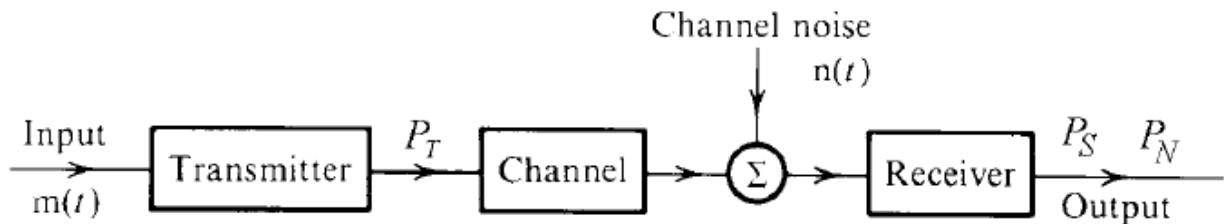


Figure. Model of an analog communication system.

A model of a typical communication system is shown in Fig. 3.1, where we assume that a modulated signal with power P_T is transmitted over a channel with additive noise. At the output of the receiver the signal and noise powers are P_S and P_N respectively, and hence, the output SNR is $\text{SNR}_O = P_S/P_N$. This ratio can be increased as much as desired simply by increasing the transmitted power. However, in practice the maximum value of P_T is limited by considerations such as transmitter cost, channel capability, interference with other channels, etc. In order to make a fair comparison between different modulation schemes, we will compare systems having the same transmitted power. Also, we need a common measurement criterion against which to compare the different modulation schemes. For this, we will use the *baseband* SNR. Recall that all modulation schemes are bandpass (i.e., the modulated signal is centered around a carrier frequency). A baseband communication system is one that does not use modulation. Such a scheme is suitable for transmission over wires, say, but is not terribly practical. As we will see, however, it does allow a direct performance comparison of different schemes.

Baseband Communication System

A baseband communication system[4] is shown in Fig., where $m(t)$ is the band-limited message signal, and W is its bandwidth

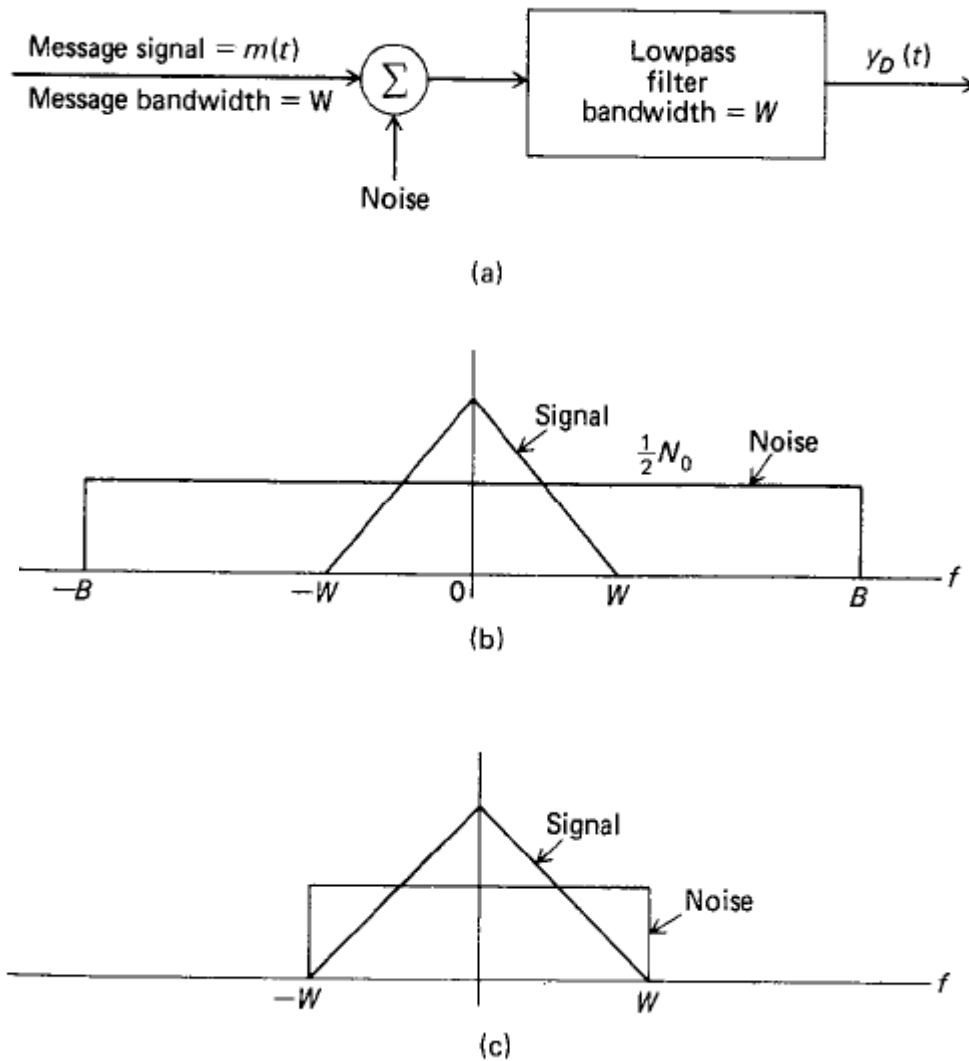


Figure. Baseband Communication System: (a)Model (b) signal spectra at filter input, and (c) signal spectra at filter output.

An example signal PSD is shown in Fig. The average signal power is given by the area under the triangular curve marked "Signal", and we will denote it by P . We assume that the additive noise has a double-sided white PSD of $N_0/2$ over some bandwidth $B > W$, as shown in Fig. For a basic baseband system, the transmitted power is identical to the message power, i.e., $P_T = P$.

The receiver consists of a low-pass filter with a bandwidth W , whose purpose is to enhance the SNR by cutting out as much of the noise as possible. The PSD of the noise at the output of the LPF is shown in Fig. , and the average noise power is given by

$$\int_{-W}^W \frac{N_0}{2} df = N_0 W$$

Thus, the SNR at the receiver output is

$$\text{SNR}_{\text{baseband}} = \frac{P_T}{N_0 W}$$

Notice that for a baseband system we can improve the SNR by: (a) increasing the transmitted power, (b) restricting the message bandwidth, or (c) making the receiver less noisy.

Amplitude Modulation

Review

In amplitude modulation[3], the *amplitude* of a sinusoidal carrier wave is varied linearly with the message signal. The general form of an AM signal is

$$s(t)_{\text{AM}} = [A + m(t)] \cos(2\pi f_c t)$$

where A is the amplitude of the carrier, f_c is the carrier frequency, and $m(t)$ is the message signal. The *modulation index*, μ is defined as

$$\mu = \frac{m_p}{A}$$

where m_p is the peak amplitude of $m(t)$, i.e., $m_p = \max|m(t)|$. Recall that if $\mu = 1$, (i.e., $A \geq m_p$), then the envelope of $s(t)$ will have the same shape as the message $m(t)$, and thus, a simple envelope detector can be used to demodulate the AM wave. The availability of a particularly simple receiver is the major advantage of AM, since as we will see, its noise performance is not great.

If an envelope detector cannot be used, another form of detection known as *synchronous detection* can be used. The block diagram of a synchronous detector is shown in Fig.

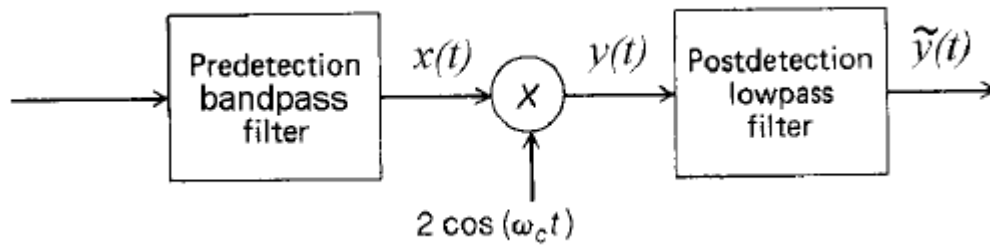


Figure. Synchronous demodulator

process involves multiplying the waveform at the receiver by a local carrier of the same frequency (and phase) as the carrier used at the transmitter. This basically replaces the $\cos(\varphi)$ term by a $\cos 2(\varphi)$ term. From the identity

$$2 \cos^2(x) = 1 + \cos(2x)$$

the result is a frequency translation of the message signal, down to baseband (i.e., $f = 0$) and up to twice the carrier frequency. The low-pass filter is then used to recover the baseband message signal. As one might expect, the main disadvantage with this scheme is that it requires generation of a local carrier signal that is perfectly synchronized with the transmitted carrier.

Notice in equation that the AM signal consists of two components, the carrier $A \cos(2\pi f_c t)$ and the sidebands $m(t) \cos(2\pi f_c t)$. Since transmitting the carrier term is wasteful, another variation of AM that is of interest is one in which the carrier is suppressed. This is referred to as *doubleside band suppressed carrier* (DSB-SC), and is given by

$$s(t)_{\text{DSB-SC}} = Am(t) \cos(2\pi f_c t)$$

In this case the envelope of the signal looks nothing like the original message signal, and a synchronous detector must be used for demodulation.

Noise in DSB-SC

The *predetection* signal (i.e., just before the multiplier[3] in Fig. is

$$x(t) = s(t) + n(t)$$

The purpose of the predetection filter is to pass only the frequencies around the carrier frequency, and thus reduce the effect of out-of-band noise. The noise signal $n(t)$ after the predetection filter is bandpass with a double-sided white PSD of $N_o = \frac{1}{2}$ over a bandwidth of $2W$ (centered on the carrier frequency), as shown in Fig. Hence, using the bandpass representation the predetection signal is

$$x(t) = [Am(t) + n_c(t)] \cos(2\pi f_c t) - n_s(t) \sin(2\pi f_c t)$$

After multiplying by $2 \cos(2\pi f_c t)$, this becomes

$$\begin{aligned} y(t) &= 2 \cos(2\pi f_c t) x(t) \\ &= Am(t)[1 + \cos(4\pi f_c t)] + n_c(t)[1 + \cos(4\pi f_c t)] - n_s(t) \sin(4\pi f_c t) \end{aligned}$$

where we have used and

$$2 \cos x \sin x = \sin(2x)$$

Low-pass filtering will remove all of the $2f_c$ frequency terms, leaving

$$y(t) = Am(t) + n_c(t)$$

The signal power at the receiver output is

$$P_s = E\{A^2 m^2(t)\} = A^2 E\{m^2(t)\} = \frac{A^2}{2}$$

where, recall, P is the power in the message signal $m(t)$. The power in the noise signal $n_c(t)$ is

$$P_N = \int_{-W}^W \frac{N_o}{2} df = 2N_o W$$

since from (2.34) the PSD of $n_c(t)$ is N_o and the bandwidth of the LPF is W . Thus, for the DSB-SC synchronous demodulator, the SNR at the receiver output is

$$\text{SNR}_O = \frac{A^2 P}{2N_o W}$$

To make a fair comparison with a baseband system, we need to calculate the transmitted power

$$P_T = E\{Am(t) \cos(2\pi f_c t)\} = \frac{A^2 P}{2}$$

and substitution gives

$$\text{SNR}_O = \frac{P_T}{N_o W}$$

Comparison gives

$$\text{SNR}_{\text{DSB-SC}} = \text{SNR}_{\text{baseband}}$$

We conclude that a DSB-SC[3] system provides no SNR performance gain over a baseband system.

It turns out that an SSB system also has the same SNR performance as a baseband system.

Noise in AM, Synchronous Detection

For an AM waveform, the predetection signal[3] is

$$x(t) = [A + m(t) + n_c(t)] \cos(2\pi f_c t) - j n_s(t) \sin(2\pi f_c t)$$

After multiplication by $2 \cos(2\pi f_c t)$, this becomes

$$y(t) = A[1 + \cos(4\pi f_c t)] + m(t)[1 + \cos(4\pi f_c t)] + n_c(t)[1 + \cos(4\pi f_c t)] - n_s(t) \sin(2\pi f_c t)$$

After low-pass filtering this becomes

$$\hat{y}(t) = A + m(t) + n_c(t)$$

Note that the DC term A can be easily removed with a DC block (i.e., a capacitor), and most AM demodulators are not DC-coupled.

The signal power at the receiver output is

$$P_s = E\{m^2(t)\}g = P$$

and the noise power is

$$P_N = 2N_0W$$

The SNR at the receiver output is therefore

$$\text{SNR}_o = \frac{P}{2N_0W}$$

The transmitted power for an AM waveform is

$$P_T = \frac{A^2}{2} + \frac{P}{2}$$

and substituting this into the baseband SNR, we find that for a baseband system with the same transmitted power

$$\text{SNR}_{\text{baseband}} = \frac{A^2 + P}{2N_0W}$$

Thus, for an AM waveform using a synchronous demodulator we have

$$\text{SNR}_{\text{AM}} = \frac{P}{A^2 + P} \text{SNR}_{\text{baseband}}$$

In other words, the performance of AM is always worse than that of a baseband system. This is because of the wasted power which results from transmitting the carrier explicitly in the AM waveform.

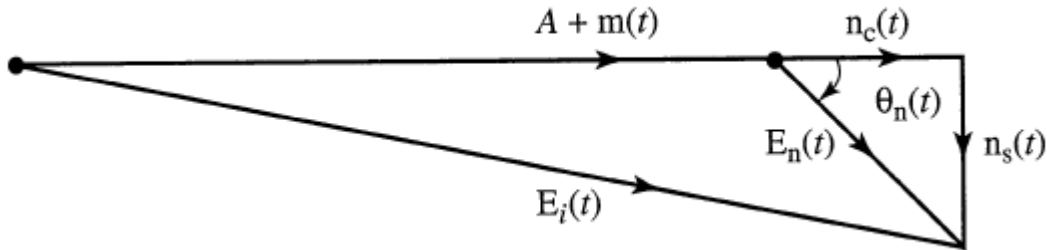


Figure. Phasor diagram of the signals present at an AM receiver

Noise in AM, Envelope Detection

Recall that an envelope detector[4] can only be used if $\mu < 1$. An envelope detector works by detecting the *envelope* of the received signal. To get an appreciation of the effect of this, we will represent the received signal by phasors, as shown in Fig. The receiver output, denoted by “ $E_r(t)$ ” in the figure, will be given by

$$y(t) = \text{envelope of } x(t) \\ = \sqrt{[A + m(t) + n_c(t)]^2 + n_s^2(t)}$$

This expression is somewhat more complicated than the others we have looked at, and it is not immediately obvious how we will find the SNR at the receiver output. What we would like is an approximation to $y(t)$ in which the message and the noise are additive.

(a) Small Noise Case

The receiver output can be simplified if we assume that for almost all t the noise power is small[2], i.e., $n(t) \ll [A + m(t)]$. Hence

$$|A + m(t) + n_c(t)| \gg |n_s(t)|$$

Then, most of the time,

$$y(t) \approx A + m(t) + n_c(t)$$

which is identical to the post-detection signal in the case of synchronous detection. Thus, (ignoring the DC term A again) the output SNR is

$$\text{SNR}_O = \frac{A^2 P}{2N_O W}$$

which can be written in terms of baseband SNR as

$$\text{SNR}_{\text{env}} = \frac{A^2}{2} + \frac{P}{2} \\ \text{SNR}_{\text{baseband}} = \frac{A^2 + P}{2N_O W}$$

Note that whereas $\text{SNR}_{\text{AM}} = \frac{P}{A^2 + P}$ $\text{SNR}_{\text{baseband}}$ is valid always, the expression for SNR_{env} is *only* valid for small noise power.

Large Noise Case

Now consider the case where noise power is large[2], so that for almost all t we have $n(t) \gg [A + m(t)]$. Rewrite as $y^2(t) = [A + m(t) + n_c(t)]^2 + n_s^2(t)$

$$= [A + m(t) + n_c(t)]^2 + 2Am(t) + 2An_c(t) + 2m(t)n_c(t) + n_s^2(t)$$

For $n_c(t) \gg [A + m(t)]$, this reduces to

$$y^2(t) \approx n_c^2(t) + n_s^2(t) + 2[A + m(t)]n_c(t) \\ = E_n^2(t) \left(1 + \frac{2[A + m(t)]n_c(t)}{E_n^2(t)}\right)$$

Where

$$E_n(t) = \sqrt{n_c^2(t) + n_s^2(t)}$$

$E_n(t)$ is the envelope of the noise (as described). But from the phasor diagram in Fig. , we have $n_c(t) = E_n(t) \cos \theta_n(t)$, giving

$$y(t) \approx E_n(t) \sqrt{1 + \frac{2[A+m(t)]\cos \theta_n(t)}{E_n(t)}}$$

Further,

$\sqrt{1+x} \approx 1 + \frac{x}{2}$ for $x \ll 1$, so this reduces to

$$\begin{aligned} y(t) &\approx E_n(t) \left(1 + \frac{[A+m(t)\cos \theta_n(t)]}{E_n(t)}\right) \\ &= E_n(t) + [A + m(t)\cos \theta_n(t)] \end{aligned}$$

The main thing to note is that the output of the envelope detector contains no term that is proportional to the message $m(t)$. The term $m(t) \cos \theta_n(t)$ is the message multiplied by a noise term $\cos \theta_n(t)$, and is no use in recovering $m(t)$. This multiplicative effect corrupts the message to a far greater extent than the additive noise in our previous analysis; the result is that there is a complete loss of information at the receiver. This produces a *threshold effect*, in that below some carrier power level, the performance of the detector deteriorates very rapidly. Despite this threshold effect, we find that in practice it does not matter terribly. This is because the quality of a signal with an output SNR less than about 25 dB is so poor, that no-one would really want to listen to it anyway. And for such a high output SNR, we are well past the threshold level and we find that equation holds. From a practical point of view, the threshold effect is seldom of importance for envelope detectors.

Frequency Modulation

Having studied the effect of additive noise on amplitude modulation systems[1], we will now look at the SNR performance on frequency modulation systems. There is a fundamental difference between these two. In AM, the message information is contained within the amplitude of the signal, and since the noise is additive it adds directly to the modulated signal. For FM, however, it is the frequency of the modulated signal that contains the message information. Since the frequency of a signal can be described by its zero crossings, the effect of noise on an FM signal is determined by the extent to which it changes the zero crossing of the modulated signal. This suggests that the effect of noise on an FM signal will be less than that for an AM system, and we will see in this section that this is in fact the case.

Review

Consider the following general representation of a carrier waveform[1]

$$s(t) = A \cos[\theta_i(t)]$$

where $\theta_i(t)$ is the *instantaneous phase angle*. Comparing this with the generic waveform $A \cos(2\pi ft)$, where f is the frequency, we can define the *instantaneous frequency* as

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt}$$

For an FM system, the instantaneous frequency of the carrier is varied linearly with the message, i.e.,

$$f_i(t) = f_c + k_f m(t)$$

where k_f is the *frequency sensitivity* of the modulator. Hence, the instantaneous phase is

$$\theta_i(t) = 2\pi \int_{-\infty}^t f_i(\tau) d\tau = 2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau$$

and the modulated signal is

$$s(t) = A \cos[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau]$$

There are two things to note about the FM signal: (a) the envelope is constant, and (b) the signal $s(t)$ is a non-linear function of the message signal $m(t)$.

Bandwidth of FM

Let the peak message amplitude be $m_p = \max|m(t)|$, so that the instantaneous frequency[2] will vary between $f_c - k_f m_p$ and $f_c + k_f m_p$. Denote the deviation of the instantaneous frequency from the carrier frequency as the *frequency deviation*

$$\Delta f = k_f m_p$$

Define the *deviation ratio* (also called the FM modulation index in the special case of tone modulated

FM) as

$$\beta = \frac{\Delta f}{W}$$

where

W is the message bandwidth.

Unlike AM, the bandwidth of FM is not dependent simply on the message bandwidth. For Small β the FM bandwidth is approximately twice the message bandwidth (referred to as narrowband FM). But for large β (referred to as wide-band FM) the bandwidth can be much larger than this. A useful rule-of-thumb for determining the transmission bandwidth of an FM signal is *Carson's rule*:

$$B_T = 2W(\beta + 1) = 2(\Delta f + W)$$

Observe that for $\beta \ll 1$, $B_T \approx 2W$ (as is the case in AM). At the other extreme, for $\beta \gg 1$, $B_T \approx 2\Delta f$, which is independent of W .

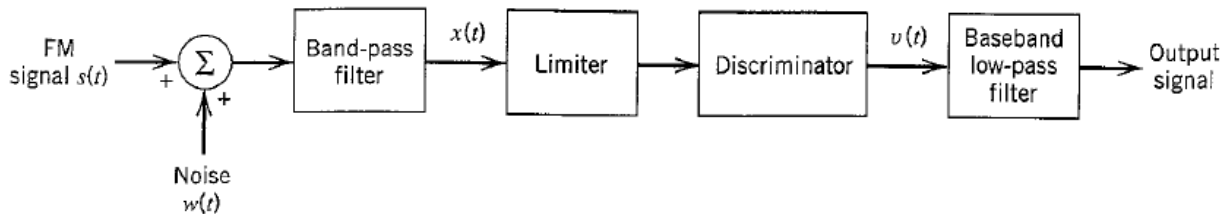


Figure: Model of an FM receiver.

Noise in FM

The model of an FM receiver[1] is shown in Fig., where $s(t)$ is the FM signal, and $w(t)$ is white Gaussian noise with power spectral density $N_0/2$. The bandpass filter is used to remove any signals outside the bandwidth of $f_c \pm B_T/2$, and thus, the predetection noise at the receiver is bandpass with a bandwidth of B_T . Since an FM signal has a constant envelope, the limiter is used to remove any amplitude variations. The discriminator is a device whose output is proportional to the deviation in the instantaneous frequency (i.e., it recovers the message signal), and the final baseband low-pass filter has a bandwidth of W and thus passes the message signal and removes out-of-band noise.

The predetection signal is

$$x(t) = A \cos \left[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\tau) d\tau \right] + n_c(t) \cos(2\pi f_c t) - n_s(t) \sin(2\pi f_c t)$$

First, let us consider the signal power at the receiver output. When the predetection SNR is high, it can be shown that the noise does not affect the power of the signal at the output. Thus, ignoring the noise, the instantaneous frequency of the input signal is

$$f_i = f_c + k_f m(t)$$

and the output of the discriminator (which is designed to simply return the deviation of the instantaneous frequency away from the carrier frequency) is $k_f m(t)$

The output signal power is therefore

$$P_s = k_f^2 P$$

where

P is the average power of the message signal.

Now, to calculate the noise power at the receiver output, it turns out that for high predetection SNR the noise output is approximately independent of the message signal. In this case, we only have the carrier and noise signals present. Thus,

$$\tilde{x}(t) = A \cos(2\pi f_c t) + n_c(t) \cos(2\pi f_c t) - n_s(t) \sin(2\pi f_c t)$$

The phasor diagram of this is shown in Fig. From this diagram, we see that the instantaneous phase is

$$\theta_i(t) = \tan^{-1} \frac{n_s(t)}{A + n_c(t)}$$

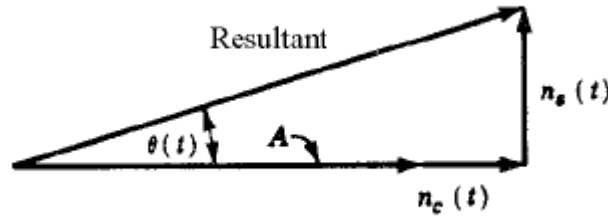


Figure: Phasor diagram of the FM carrier and noise signals.

For large carrier power, then most of the time

$$\theta_i(t) = \tan^{-1} \frac{n_s(t)}{A} \approx \frac{n_s(t)}{A}$$

where the last line follows from $\tan \epsilon \approx \epsilon$ for small ϵ . But, the discriminator output is the instantaneous frequency, given by

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} = \frac{1}{2\pi A} \frac{dn_s(t)}{dt}$$

We know the PSD of $n_s(t)$ shown in Fig., but what is the PSD of $dn_s(t)/dt$?

Fourier theory tells us that:

if $x(t) \leftrightarrow X(f)$

then

$$\frac{dx(t)}{dt} = j2\pi f x(f)$$

In other words, differentiation with respect to time is the same as passing the signal through a system having a transfer function of $H(f) = j2\pi f$. It can be shown that if a signal with PSD $S_i(f)$ is passed through a linear system with transfer function $H(f)$, then the PSD at the output of the system is $S_o(f) = |H(f)|^2 S_i(f)$

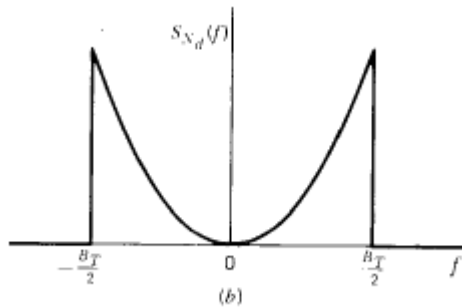
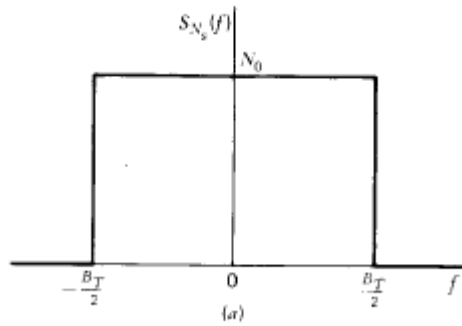
If the PSD of $n_s(t)$ has a value of N_0 within the band $\pm B_T/2$ as shown in Fig, then

$dn_s(t)/dt$ has a PSD of $|j2\pi f|^2 N_0$. The PSD of $dn_s(t)/dt$ before and after the baseband LPF is shown in Fig. (b) and (c) respectively.

Returning to equation, that the PSD of $dn_s(t)/dt$ is known, we can calculate the average noise power at the receiver output. It is given by

$$P_N = \int_{-W}^W S_D(f) df$$

where $S_D(f)$ is the PSD of the noise component at the discriminator output (i.e., the PSD of $f_i(t)$ in equation); the limits of integration are taken between $-W$ and W to reflect the fact that the output



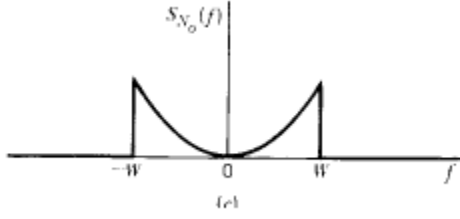


Figure: Power spectral densities for FM noise analysis: (a) $n_s(t)$ (b) $dn_s(t)/dt$ and (c) noise at the receiver output,

signal is low-pass filtered. Thus,

$$P_N = \int_{-W}^W \left(\frac{1}{2\pi A} \right)^2 (2\pi f)^2 N_o df = \int_{-W}^W \frac{N_o}{A^2} (f)^2 df = \frac{2N_o W^3}{3A^2}$$

This expression is quite important, since it shows that the average noise power at the output of a FM receiver is inversely proportional to the carrier power $A^2/2$. Hence, increasing the carrier power has a *noise quieting* effect. This is one of the major advantages of FM systems.

Finally, we have that at the output the SNR is

$$SNR_O = \frac{3A^2 K_f^2 P}{2N_o W^3}$$

Since the transmitted power of an FM waveform is

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

substitution into earlier equation gives

$$SNR_{FM} = \frac{3K_f^2 P}{W^2} SNR_{baseband} = 3\beta^2 \frac{P}{m_p^2} SNR_{baseband}$$

The SNR expression[2] is based on the assumption that the carrier power is large compared to the noise power. It is found that, like an AM envelope detector, the FM detector exhibits a

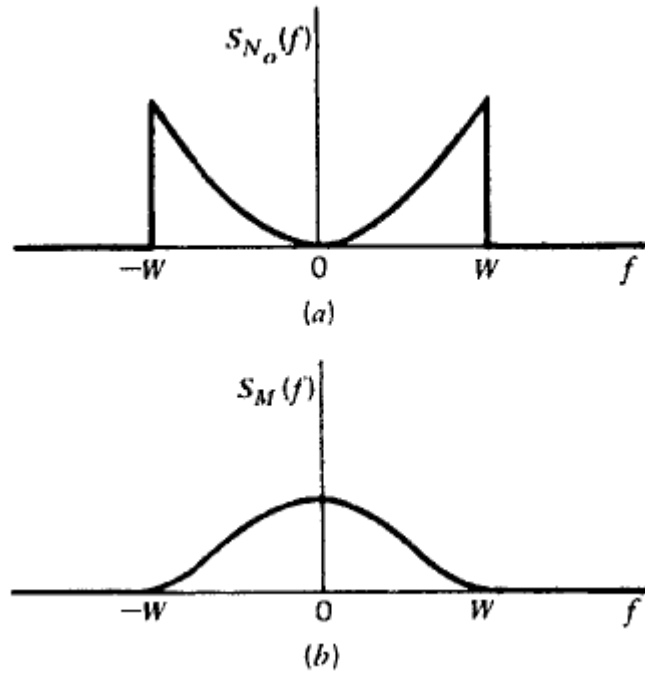


Figure : Power spectral densities of: (a) noise at the output of FM receiver, and (b) a typical message signal

threshold effect. As the carrier power decreases, the FM receiver breaks, as Haykin describes: “At first, individual clicks are heard in the receiver output, and as the carrier-to-noise ratio decreases still further, the clicks rapidly merge into a crackling or sputtering sound”. Experimental studies indicate that this noise mutilation is negligible in most cases if the predetection SNR (i.e., just after the receiver bandpass filter) is above 10. In other words, the threshold point occurs around

$$\frac{A^2}{2N_0\beta_T} = 10$$

Where, recall, $B_T = 2W(\beta+1)$. For predetection SNRs above this value, the output SNR is given by equation

One should note that whereas equation suggests that output SNR for an FM system can be increased arbitrarily by increasing β while keeping the signal power fixed, inspection of equation shows this not to be strictly true. The reason is that if β increases too far, the condition of equation that we are above threshold may no longer be true, meaning that equation no longer provides an expression for the true SNR.

3.4.3 Pre-emphasis and De-emphasis

There is another way in which the SNR of an FM system may be increased [2][3]. We saw in the previous subsection that the PSD of the noise at the detector output has a square-law dependence on frequency. On the other hand, the PSD of a typical message source is not uniform, and typically rolls off at around 6 dB per decade (see Fig.). We note that at high frequencies the

relative message power is quite low, whereas the noise power is quite high (and is rapidly increasing). It is possible that this situation could be improved by reducing the bandwidth of the transmitted message (and the corresponding cutoff frequency of the baseband LPF in the receiver), thus rejecting a large amount of the out-of-band noise. In practice, however, the distortion introduced by low-pass filtering the message signal is unsatisfactory [2].

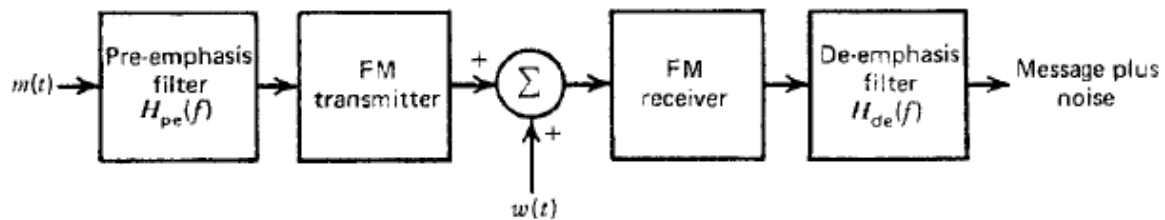
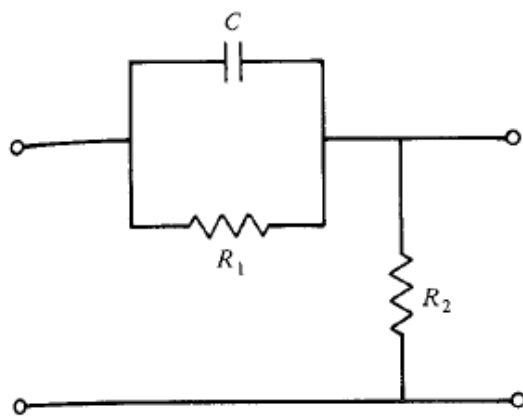
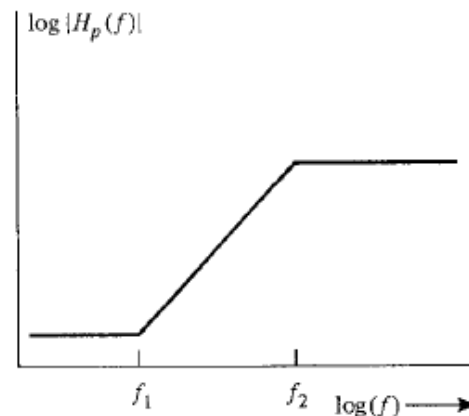


Figure: Pre-emphasis and de-emphasis in an FM system.



$$H_p(f) = K \frac{1 + j(f/f_1)}{1 + j(f/f_2)} \quad \text{where } f_1 = \frac{1}{2\pi\tau_1} = \frac{1}{2\pi R_1 C}, f_2 = \frac{1}{2\pi\tau_2} = \frac{R_1 + R_2}{2\pi R_1 R_2 C}$$

(a) Preemphasis Filter



(b) Bode Plot of Preemphasis Frequency Response

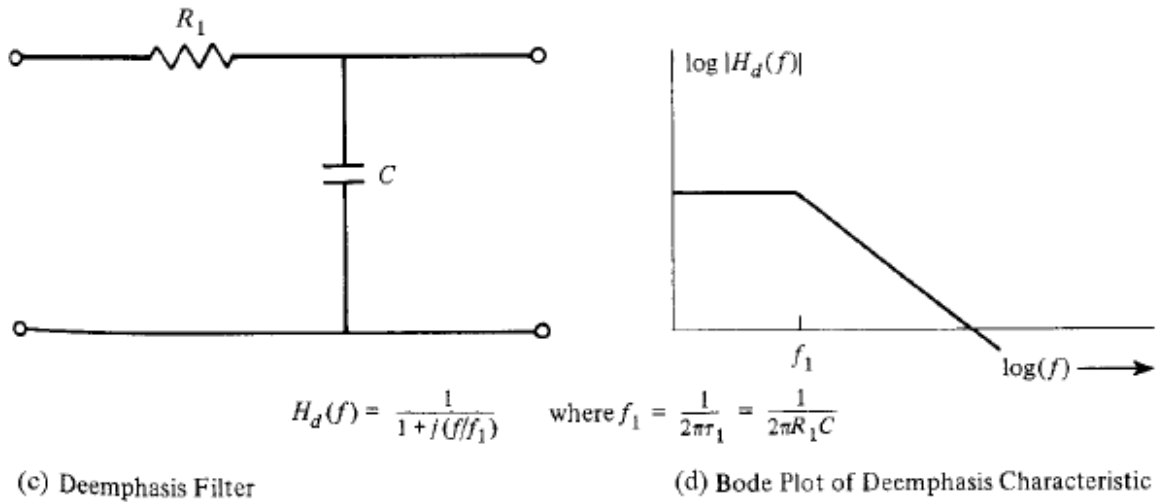


Figure: Simple linear pre-emphasis and de-emphasis circuits.

A better solution is obtained by using the *pre-emphasis* and *de-emphasis*[1] stages shown in Fig. . The intention of this scheme is that $H_{pe}(f)$ is used to artificially emphasize the high frequency components of the message prior to modulation, and hence, before noise is introduced. This serves to effectively equalize the low- and high-frequency portions of the message PSD such that the message more fully utilizes the bandwidth available to it. At the receiver, $H_{de}(f)$ performs the inverse operation by de-emphasizing the high frequency components, thereby restoring the original PSD of the message signal.

Simple circuits that perform pre- and de-emphasis are shown in Fig. , along with their respective frequency responses. Haykin shows that these circuits can improve the output SNR by around 13 dB. In closing this section, we also note that Dolby noise reduction uses an analogous pre-emphasis technique to reduce the effects of noise.

REFERENCES:

[1] H. Taub, D. L Schilling, G. Saha; “Principles of Communication System,” 3rd Edition; 2008, Tata McGraw Hill, India; ISBN: 0070648115.

[2] Modern Digital and Analogue Communication Systems by “B.P. Lathi and Z Ding,” 4th Edition 2010, Oxford University Press, New Delhi.

[3] Communication Systems by “Simon Haykin,” 4th Edition, John Wiley & Sons, Inc.

[4] Communication systems Analog and Digital by sanjay Sharma ,4th edition, katson publications