LECTURE NOTES

ON

ANALOG COMMUNICATIONS

(AEC005)

B.Tech-ECE-IV semester

Dr.P.Munusamy, Professor, ECE

Ms.G.Ajitha, Assistant Professor, ECE

Ms.P.Saritha, Assistant Professor, ECE

Ms.L.Shruthi, Assistant Professor, ECE

ELECTRONICS AND COMMUNICATION ENGINEERING

INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043, Telangana

ANALOG COMMUNICATIONS

IV Semester: ECE								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
AEC005	**Core**	**L**	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
		3	1	-	4	30	70	100
Contact Classes: 45	**Tutorial Classes: 15**	**Practical Classes: Nil**				**Total Classes: 60**		

OBJECTIVES:

The course should enable the students to:

- **I.** Develop skills for analyzing different types of signals in terms of their properties such as energy, power, correlation and apply for analysis of linear time invariant systems.
- **II.** Analyze various techniques of generation and detection of amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM) signals.
- **III.** Differentiate the performance of AM, FM and PM systems in terms of Power, Bandwidth and SNR (Signal-to-Noise Ratio).
- **IV.** Evaluate Analog Communication system in terms of the complexity of the transmitters and receivers.

UNIT-I	SIGNAL ANALYSIS AND LTI SYSTEMS	Classes: 10

Classification of signals and study of Fourier transforms for standard signals, definition of signal bandwidth; Systems: Definition of system, classification of systems based on properties, linear time invariant system , impulse, step, sinusoidal response of a linear time invariant system, transfer function of a linear time invariant system, distortion less transmission through a linear time invariant system; system bandwidth; Convolution and correlation of signals: Concept of convolution, graphical representation of convolution, properties of convolution; Cross correlation ,auto correlation functions and their properties, comparison between correlation and convolution.

UNIT-II	AMPLITUDE AND DOUBLE SIDE BAND SUPPRESSED CARRIER MODULATION	Classes: 10

Introduction to communication system, need for modulation, frequency division multiplexing; Amplitude modulation, definition; Time domain and frequency domain description, single tone modulation, power relations in amplitude modulation waves; Generation of amplitude modulation wave using ,square law and switching modulators; Detection of amplitude modulation waves using square law and envelope detectors; Double side band modulation: Double side band suppressed carrier time domain and frequency domain description; Generation of double side band suppressed carrier waves using balanced and ring modulators; Coherent detection of double side band suppressed carrier modulated waves; Costas loop; Noise in amplitude modulation, noise in double side band suppressed carrier.

UNIT-III	SINGLE SIDE BAND AND VESTIGIAL SIDE BAND MODULATION	Classes: 08

Frequency domain description, frequency discrimination method for generation of amplitude modulation single side band modulated wave; time domain description; Phase discrimination method for generating amplitude modulation single side band modulated waves; Demodulation of single side band waves.
Noise in single side band suppressed carrier; Vestigial side band modulation: Frequency description, generation of vestigial side band modulated wave; Time domain description; Envelope detection of a vestigial side band modulation wave pulse carrier; Comparison of amplitude modulation techniques; Applications of different amplitude modulation systems.

UNIT-IV	ANGLE MODULATION	Classes: 09
Basic concepts, frequency modulation: Single tone frequency modulation, spectrum analysis of sinusoidal frequency modulation wave, narrow band frequency modulation, wide band frequency modulation, transmission bandwidth of frequency modulation wave, phase modulation, comparison of frequency modulation and phase modulation; Generation of frequency modulation waves, direct frequency modulation and indirect frequency modulation, detection of frequency modulation waves: Balanced frequency discriminator, Foster Seeley discriminator, ratio detector, zero crossing detector, phase locked loop, comparison of frequency modulation and amplitude modulation; Noise in angle modulation system, threshold effect in angle modulation system, pre-emphasis and de-emphasis.		
UNIT-V	**RECEIVERS AND SAMPLING THEORM**	**Classes: 08**
Receivers: Introduction, tuned radio frequency receiver, super heterodyne receiver, radio frequency amplifier, mixer, local oscillator, intermediate frequency amplifier, automatic gain control; Receiver characteristics: Sensitivity, selectivity, image frequency rejection ratio, choice of intermediate frequency, fidelity; Frequency modulation receiver, amplitude limiting, automatic frequency control, comparison with amplitude modulation receiver; Sampling: Sampling theorem, graphical and analytical proof for band limited signals, types of sampling, reconstruction of signal from its samples.		

UNIT I

SIGNAL ANALYSIS AND LTI SYSTEMS

WHAT IS A SIGNAL

We are all immersed in a sea of signals. All of us from the smallest living unit, a cell, to the most complex living organism (humans) are all time receiving signals and are processing them. Survival of any living organism depends upon processing the signals appropriately. What is signal? To define this precisely is a difficult task. Anything which carries information is a signal. In this course we will learn some of the mathematical representations of the signals, which has been found very useful in making information processing systems. Examples of signals are human voice, chirping of birds, smoke signals, gestures (sign language), fragrances of the flowers. Many of our body functions are regulated by chemical signals, blind people use sense of touch. Bees communicate by their dancing pattern. Some examples of modern high speed signals are the voltage charger in a telephone wire, the electromagnetic field emanating from a transmitting antenna, variation of light intensity in an optical fiber. Thus we see that there is an almost endless variety of signals and a large number of ways in which signals are carried from on place to another place. In this course we will adopt the following definition for the signal: A signal is a real (or complex) valued function of one or more real variable(s).When the function depends on a single variable, the signal is said to be one dimensional. A speech signal, daily maximum temperature, annual rainfall at a place, are all examples of a one dimensional signal. When the function depends on two or more variables, the signal is said to be multidimensional. An image is representing the two dimensional signal, vertical and horizontal coordinates representing the two dimensions. Our physical world is four dimensional (three spatial and one temporal).

1.2 CLASSIFICATION OF SIGNALS

As mentioned earlier, we will use the term signal to mean a real or complex valued function of real variable(s). Let us denote the signal by $x(t)$. The variable t is called independent variable and the value x of t as dependent variable. We say a signal is continuous time signal if the independent variable t takes values in an interval. For example $t \epsilon (-\infty, \infty)$, $or\ t\epsilon\ [0, \infty]$ $or\ t\ \epsilon[T0, T1]$.

The independent variable t is referred to as time, even though it may not be actually time. For example in variation if pressure with height t refers above mean sea level. When t takes vales in a countable set the signal is called a discrete time signal. For example $t\ \epsilon\{0, T, 2T, 3T, 4T, ...\}$ or $t\ \epsilon\{...-1, 0, 1, ...\}$ or $t\ \epsilon\{1/2, 3/2, 5/2, 7/2, ...\}$ etc. For convenience of presentation we use the notation x[n] to denote discrete time signal. Let us pause here and clarify the notation a bit. When we write $x(t)$ it has two meanings. One is value of x at time t and the other is the pairs($x(t)$, t) allowable value of t. By signal we mean the second interpretation. To keep this distinction we will use the following notation: $\{x(t)\}$ to denote the continuous time signal. Here $\{x(t)\}$ is short notation for $\{x(t),\ t\ \epsilon\ I\}$ where I is the set in which t takes the value. Similarly for discrete time signal we will use the notation $\{x[n]\}$, where $\{x[n]\}$ is short for $\{x[n], n\_I\}$. Note that in $\{x(t)\}$ and $\{x[n]\}$ are dummy variables i.e. $\{x[n]\}$ and $\{x[t]\}$ refer to the same signal. Some books use the notation $x[\cdot]$ to denote $\{x[n]\}$ and $x[n]$ to denote value of x at time$\cdot$ $x[n]$ refers to the whole waveform, while $x[n]$ refers to a particular value. Most of the books do not make this distinction clean and use $x[n]$ to denote signal and $x[n]$ to denote a particular value.

As with independent variable t, the dependent variable x can take values in a continues set or in a countable set. When both the dependent and independent variable take value in intervals, the signal is called an analog signal. When both the dependent and independent variables take values in countable sets (two sets can be quite different) the signal is called Digital signal. When we use digital computers to do processing we are doing digital signal processing. But most of the theory is for discrete time signal processing where default variable is continuous. This is because of the mathematical simplicity of discrete time signal processing. Also digital signal processing tries to implement this as closely as possible. Thus what we study is mostly discrete time signal processing and what is really implemented is digital signal processing.

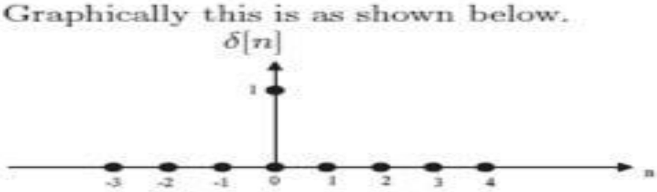
1.3 ELEMENTARY SIGNALS

There are several elementary signals that feature prominently in the study of digital signals and digital signal processing.

(a)Unit sample sequence *δ*[*n*]: Unit sample sequence is defined by

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

Graphically this is as shown below.



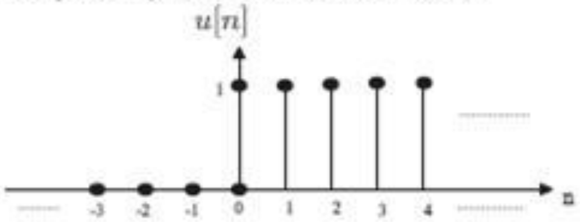
Unit sample sequence is also known as impulse sequence. This plays role akin to the impulse function $\delta(t)$ of continues time. The continues time impulse $\delta(t)$ is purely a mathematical construct while in discrete time we can actually generate the impulse sequence.

(b)Unit step sequence $u[n]$: Unit step sequence is defined by

(b)Unit step sequence $u[n]$: Unit step sequence is defined by

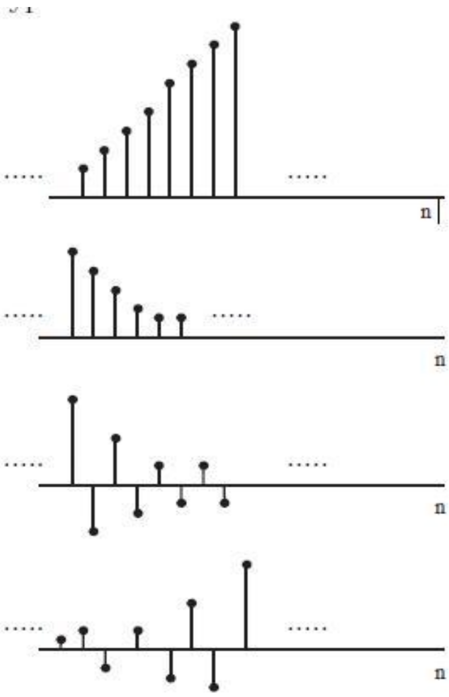
$$u[n] = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

Graphically this is as shown below



Exponential sequence: The complex exponential signal or sequence $x[n]$
is defined by $x[n] = C\,\alpha^n$
where C and α are, in general, complex numbers.

Real exponential signals: If C and α are real, we can have one of the several type of behaviour illustrated below



$\{x[n] = \alpha^n,\ \alpha > 1\}$

$\{x[n] = \alpha^n,\ 0 < \alpha < 1\}$

$\{x[n] = \alpha^n,\ -1 < \alpha < 0\}$

$\{x[n] = \alpha^n,\ \alpha < -1\}$

SIMPLE OPERATIONS AND PROPERTIES OF SEQUENCES

2.1 Simple operations on signals

In analyzing discrete-time systems, operations on sequences occur frequently.

Some operations are discussed below.

2.1.1 Sequence addition:

Let $\{x[n]\}$ and $\{y[n]\}$ be two sequences. The sequence addition is defined as term by term addition. Let $\{z[n]\}$ be the resulting sequence

$\{z[n]\} = \{x[n]\} + \{y[n]\}$, where each term $z[n] = x[n] + y[n]$
We will use the following notation

$\{x[n]\} + \{y[n]\} = \{x[n] + y[n]\}$

2.1.2 Scalar multiplication:

Let a be a scalar. We will take a to be real if we consider only the real valued signals, and take a to be a complex number if we are considering complex valued sequence. Unless otherwise stated we will consider complex valued sequences. Let the resulting sequence be denoted by $w[n]$

$\{w[n]\} = ax[n]$ is defined by $w[n] = ax[n]$, each term is multiplied by a We will use the notation $aw[n] = aw[n]$

Note: If we take the set of sequences and define these two operators as addition and scalar multiplication they satisfy all the properties of a linear vector space.

2.1.3 Sequence multiplication:

Let $\{x[n]\}$ and $\{y[n]\}$ be two sequences, and $\{z[n]\}$ be resulting sequence

$\{z[n]\} = \{x[n]\}\{y[n]\}$, where $z[n] = x[n]y[n]$.
The notation used for this will be $\{x[n]\}\{y[n]\} = \{x[n]y[n]\}$

Now we consider some operations based on independent variable n.

2.1.4 Shifting

This is also known as translation. Let us shift a sequence $\{x[n]\}$ by $n0$ units, and the resulting sequence by $\{y[n]\}$

$\{y[n]\} = z{-}n0(\{x[n]\})$
where $z{-}n0()$is the operation of shifting the sequence right by $n0$ unit.
The terms are defined by $y[n] = x[n{-}n)]$. We will use short notation $\{x[n{-}n0]\}$

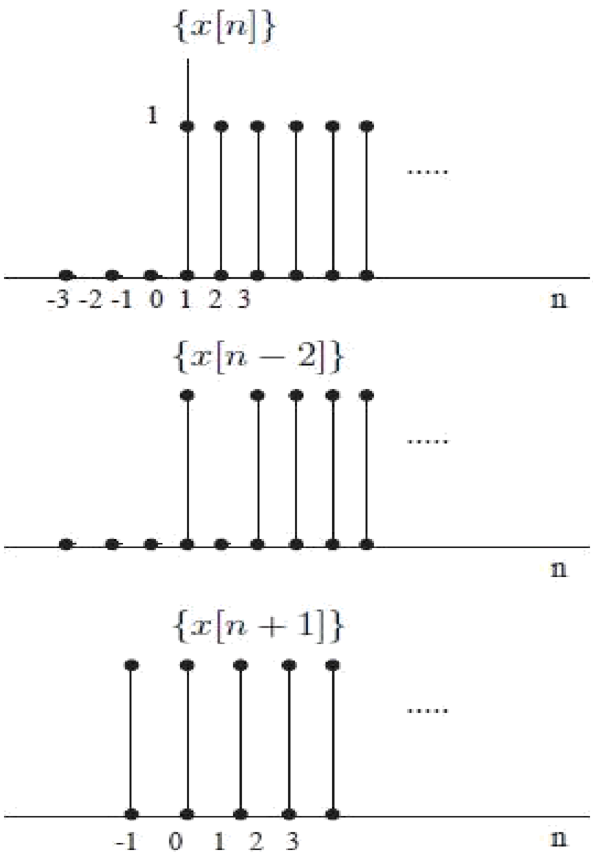
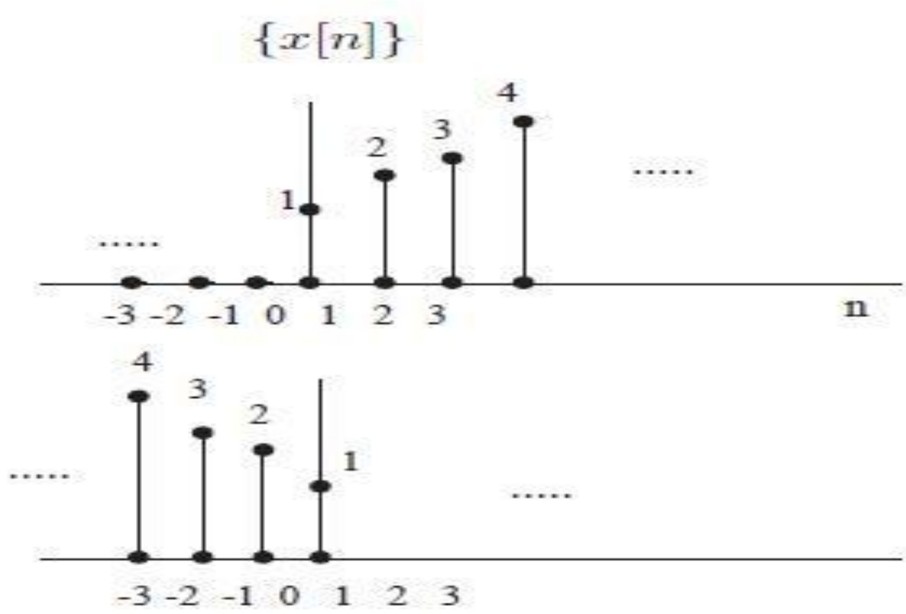


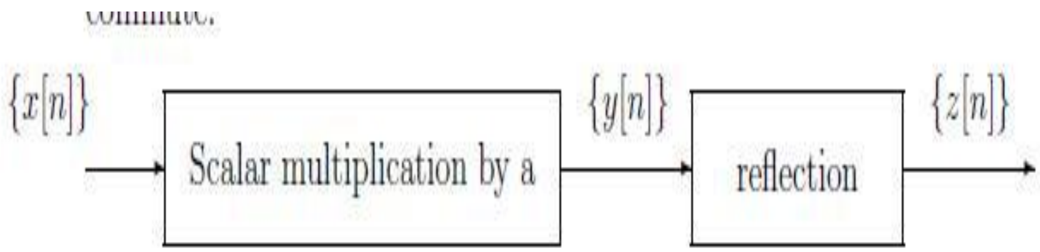
Figure above show some examples of shifting. A negative value of n_0 shift towards right.

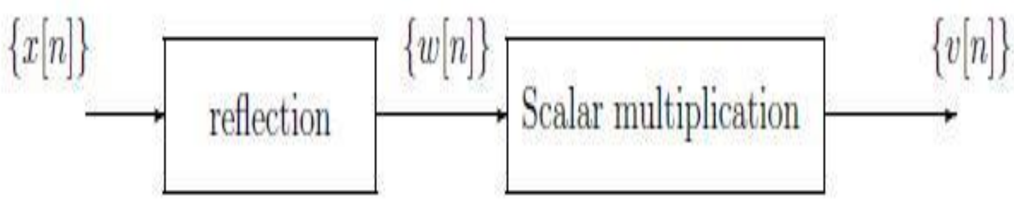
2.1.5 Reflection:

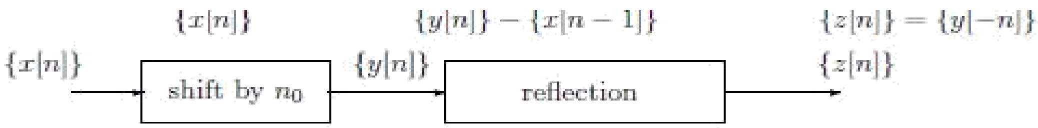
Let *{x[n]}* be the original sequence, and *{y[n]}* be reflected sequence, then *y*[*n*] is defined by $y[n] = x[-n]$

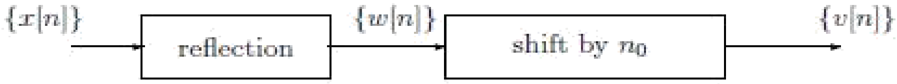


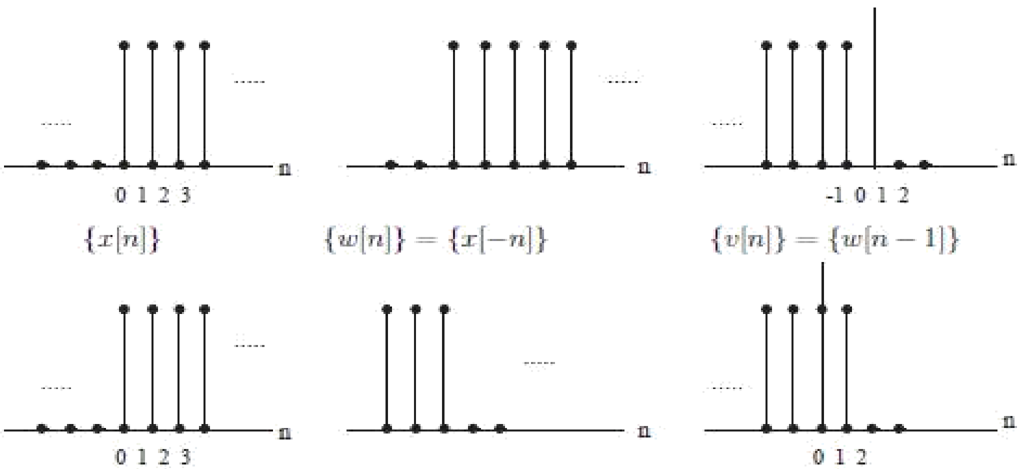
We will denote this by *{x*[*n*]*}*. When we have complex valued signals, sometimes we reflect and do the complex conjugation, ie, $y[n]$ is defined by $y[n]$
S $x * [-n]$, where * denotes complex conjugation. This sequence will be denoted by $\{x * [-n]\}$.
We will learn about more complex operations later on. Some of these operations commute, i.e. if we apply two operations we can interchange their order and some do not commute. For example scalar multiplication and reflection.











We can combine many of these operations in one step, for example $\{y[n]\}$ may be defined as $y[n] = 2x[3 - n]$.

2.2 SOME PROPERTIES OF SIGNALS:

2.2.1 Energy of a Signal:

The total enery of a signal *{x*[*n*]*}* is defined by

$$E_x = \sum_{n=-\infty}^{\infty} |x[n]|^2$$

A signal is reffered to as an energy signal, if and only if the total energy of

the signal *Ex* is finite. An energy signal has a zero power and a power signal has infinite energy. There are signals which are neither energy signals nor power signals. For example *{x*[*n*]*}* defined by *x*[*n*] = *n* does not have finite power or energy

2.2.2 Power of a signal:

If *{x*[*n*]*}* is a signal whose energy is not finite, we define power of the signal

$$y(n) = -\sum_{i=1}^{N} a_i y(n-i) + \sum_{j=0}^{M} b_j x(n-j).$$

$$P_x = \lim_{N\to\infty} \frac{1}{(2N+1)} \sum_{n=-N}^{N} |x[n]|^2$$

A signal is referred to as a power signal if the power P_x satisfies the condition

$$0 < P_x < \infty$$

2.2.3 Periodic Signals:

An important class of signals that we encounter frequently is the class of periodic signals. We say that a signal $\{x[n]\}$ is periodic period N, where N is a positive integer, if the signal is unchanged by the time shift of N ie.,

$$\{x[n]\} = \{x[n+N]\}$$

or $x[n] = x[n+N$ for all n.
Since $\{x[n]\}$ is same as $\{x[n+N]\}$, it is also periodic so we get

$$\{x[n]\} = \{x[n+N]\} = \{x[n+N+N]\} = \{x[n+2N]\}$$

Generalizing this we get $\{x[n]\} = \{x[n+kN]\}$, where k is a positive integer. From this we see that $\{x[n]\}$ is periodic with $2N$, $3N$, ... The fundamental period N_0 is

the smallest positive value N for which the signal is periodic. The signal illustrated below is periodic with fundamental period $N_0 = 4$. $\{x[n]\}$ By change of variable we can write $\{x[n]\} = \{x[n+N]\}$ as $\{x[m-N]\} = \{x[m]\}$ and then we see that

$$\{x[n]\} = \{x[n+kN]\},$$

for all integer values of k, positive, negative or zero. By definition, period of

a signal is always a positive integer n. Except for a all zero signal all periodic signals have infinite energy. They may have finite power. Let $\{x[n]\}$ be periodic with period N, then the power P_x is given by

$$P - x = \lim_{M\to\infty} \frac{1}{(2M+1)} \sum_{n=-M}^{M} |x[n]|^2$$

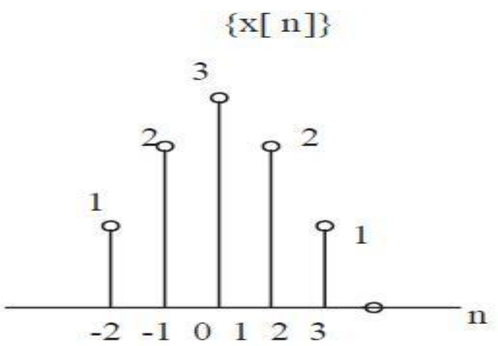
$$= \lim_{M\to\infty} \tfrac{1}{2M+1}[\sum_{n=0}^{N-1} |x[n]|^2 + \sum_{n=N}^{2N-1} |x[n]|^2 + ...$$
$$+ \sum_{n=(k-1)N-1}^{kN-1} |x[n]|^2 + \sum_{n=kN}^{M} |x[n]|^2 + \sum_{n=-N}^{-1} |x[n]|^2 + ...$$
$$+ \sum_{n=-kN}^{-(k-1)N-1} |x[n]|^2 + \sum_{n=-M}^{-kN-1} |x[n]|^2]$$

where k is largest integer such that $kN - 1 \le M$. Since the signal is periodic, sum over one period will be same for all terms. We see that k is approximately equal to M/N(it is integer part of this) and for large M we get $2M/N$ terms and limit $2M/(2M+1)$ as M goes to infinite is one we get

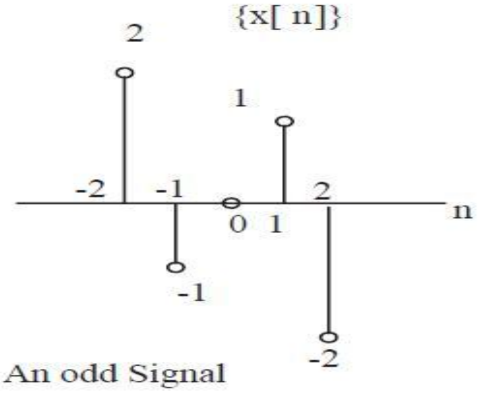
$$P_x = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2$$

2.2.4 Even and odd signals:

A real valued signal $\{x[n]\}$ is referred as an even signal if it is identical to its time reversed counterpart ie, if $\{x[n]\} = \{x[-n]\}$ A real signal is referred to as an odd signal if $\{x[n]\} = \{-x[-n]\}$ An odd signal has value 0 at $n = 0$ as $x[0] = -x[n] = -x[0]$



An even signal



An odd Signal

as a sum of an even signal and an odd signal. Consider the signals

$$Ev(\{x[n]\}) = \{x_e[n]\} = \{1/2(x[n] + x[-n])\}$$

and

$$Od(\{x[n]\}) = \{x_0[n]\} = \{1/2(x[n] - x[-n])\}$$

We can see easily that

$$\{x[n]\} = \{x_e[n]\} + \{x_o[n]\}$$

The signal $\{x[n]\}$ is called the even part of $\{x[n]\}$. We can verify very easily that $\{x_e[n]\}$ is an even signal. Similarly, $\{x_0[n]\}$ is called the odd part of $\{x[n]\}$ and is an odd signal. When we have complex valued signals we use a slightly different terminology. A complex valued signal $\{x[n]\}$ is referred to as a conjugate symmetric signal if $\{x[n]\} = \{x^*[-n]$, where x^*refers to the complex conjugate of x.Here we do reflection and comple conjugation. If $\{x[n]\}$ is real valuedthis is same as an even signal. A complex signal $\{x[n]\}$ is referred to as a conjugate antisymmetric signal if $\{x[n]\} = \{-x^*[-n]\}$. We can express any complex valued signal as sum conjugate symmetric and conjugate antisymmetric signals. We use notation similar to above $Ev(\{x[n]\}) = \{x_e[n]\} =\{1/2(x[n] + x^*[-n])\}$ and $Od(\{x[n]\}) = \{x_0[n]\} = \{1/2(x[n] - x* [-n])\}$ the $\{x[n]\} = \{x_e[n]\} + \{x_o[n]\}$. We can see easily that $\{xe[n]\}$ is conjugate symmetric signal and $\{x_o[n]\}$ is conjugate antisymmetric signal. These definitions reduce to even and odd signals in case signals takes only real values.

2.3 PERIODICITY PROPERTIES OF SINUSOIDAL SIGNALS

Let us consider the signal $\{x[n]\} = \{\cos w_0 n\}$. We see that if we replace $w0$ by $(w_0 + 2\pi)$ we get the same signal. In fact the signal with frequency $w_0\}2\pi$,$w0\}4\pi$ and so on. This situation is quite different from continuous time signal $\{\cos w_0 t, -\infty < t < \infty\}$ where each frequency is different. Thus in discrete time we need to consider frequency interval of length 2π only. As we increase w; 0 to π signal oscillates more and more rapidly. But if we further increase frequency from π to 2π the rate of oscillations decreases. This can be seen easily by plotting signal $\cos w_0 n\}$ for several values of w_0. The signal $\{\cos w_0 n\}$ is not periodic for every value of w_0. For the signal to be periodic with pe
riod $N > 0$, we should have

$$\{\cos w_0 n\} = \{cos w_0(n + N)\}$$

that is $w_0 N$ should be some multiple of 2π.

$$w_0 N = 2\pi m$$

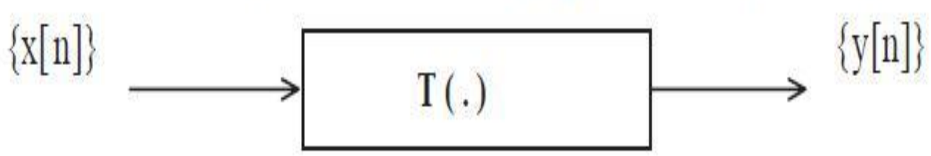
or

$$\frac{w_0}{2\pi} = \frac{m}{N}$$

Thus signal $\{\cos w_0 n\}$ is periodic if and only if $w_0 = 2\pi$ is a rational number. Above observations also hold for complex exponential signal $\{x[n]\} = \{e^{jw_0 n}\}$

2.3.1.Discrete-Time Systems

A discrete-time system can be thought of as a transformation or operator that maps an input sequence $\{x[n]\}$ to an output sequence $\{y[n]\}$



By placing various conditions on $T(\cdot)$ we can define different classes of systems.

3.BASIC SYSTEM PROPERTIES

3.1 Systems with or without memory:

A system is said to be memory less if the out put for each value of the independent variable at a given time n depends only on the input value at time (t) For example system specified by the relationship $y[n] = \cos(x[n]) + z$ is memory less. A particularly simple memory less system is the identity system defined by $y[n] = x[n]$ In general we can write input-output relationship for memory less system as $y[n] = g(x[n])$. Not all systems are memory less. A simple example of system with memory is a delay defined by $y[n] = x[n - 1]$ A system with memory retains or stores information about input values at times other than the current input value.

3.2 Inevitability

A system is said to be invertible if the input signal $\{x[n]\}$ can be recovered from the output signal $\{y[n]\}$. For this to be true two different input signals should produce two different outputs. If some different input signal produce same output signal then by processing output we can not say which input produced the output. Example of an invertible system is

$$y[n] = \sum_{k=-\infty}^{n} x[k]$$

then

$$x[n] = y[n] - y[n-1]$$

Example if a non-invertible system is

$$y[n] = 0$$

That is the system produces an all zero sequence for any input sequence. Since every input sequence gives all zero sequence, we can not find out which input produced the output. The system which produces the sequence $\{x[n]\}$ from sequence $\{y[n]\}$ is called the inverse system. In communication system, decoder is an inverse of the encoder.

3.3 Causality

A system is causal if the output at anytime depends only on values of the input at the present time and in the past. $y[n] = f(x[n], x[n-1], \ldots)$. All memory less systems are causal. An accumulator system defined by

$$y[n] = \sum_{x[h]}$$

is also causal. The system defined by

$$y[n] = \frac{1}{2N+1} \sum_{k=-N}^{N} x[n-k]$$

For real time system where n actually denoted time causalities is important. Causality is not an essential constraint in applications where n is not time, for example, image processing. If we case doing processing on recorded data, then also causality may not be required.

3.4 Stability

There are several definitions for stability. Here we will consider bounded input bonded output(BIBO) stability. A system is said to be BIBO stable if every bounded input produces a bounded output. We say that a signal $\{x[n]\}$ is bounded if

$$|x[n]| < M < \infty \qquad \text{for all } n$$

The moving averafe system

$$y[n] = \frac{1}{2N+1} \sum_{k=-N}^{N} x[n]$$

is stable as $y[n]$ is sum of finite numbers and so it is bounded. The accumulator system defined by

$$y[n] = \sum_{k=-\infty}^{n} x[k]$$

is unstable. If we take $\{x[n]\} = \{u[n]\}$, the unit step then $y[0] = 1$, $y[1] = 2$, $y[2] = 3$, are$y[n] = n + 1$, $n \geq 0$ so $y[n]$ grows without bound.

3.5 Time invariance

A system is said to be time invariant if the behaviour and characteristics of the system do not change with time. Thus a system is said to be time invariant if a time delay or time advance in the input signal leads to identical delay or advance in the output signal. Mathematically if

$$\{y[n]\} = T(\{x[n]\})$$

then

$$\{y[n-n_0]\} = T(\{x[n-n_0]\}) \text{ for any } n_0$$

Let us consider the accumulator system

$$y[n] = \sum_{k=-\infty}^{n} x[k]$$

If the input is now $\{x_1[n]\} = \{x[n-n_0]\}$ then the corresponding output is

$$y_1[n] = \sum_{k=-\infty}^{n} x_1[k]$$

$$= \sum_{k=-\infty} x[k]$$

The shifted output signal is given by

$$y[n-n_0] == \sum_{k=-\infty}^{n-n_0} x[k]$$

The two expression look different, but infact they are equal. Let us change the index of summation by $l = k - n_0$ in the first sum then we see that

$$y_1[n] == \sum_{l=-\infty}^{n-n_0} x[l]$$

$$= y[n-n_0]$$

Hence, $\{y[n]\} = \{y[n-n_0]\}$ and the system is time-invariant. As a second example consider the system defined by

$$y[n] = nx[n]$$

if

$$\{x_1[n]\} = \{x[n-n_0]\}$$

$$y_1[n] = nx_1[n] = nx[n-n_0]$$

while

$$y[n-n_0] = (n-n_0)x[n-n_0]$$

and so the system is not time-invariant. It is time varying. We can also see this by giving a counter example. Suppose input is $\{x[n]\} = \{\delta[n]\}$ then output is allzero sequence. If the input is $\{\delta[n-1]\}$ then output is $\{\delta[n-1]\}$ which is definitely not a shifted version version of all zero sequence.

3.6 Linearity

This is an important property of the system. We will see later that if we have system which is linear and time invariant then it has a very compact representation. A linear system possesses the important property of super position: if an input consists of weighted sum of several signals, the output is also weighted sum of the responses of the system to each of those input signals. Mathematically let $\{y_1[n]\}$ be the response of the system to the input $\{x_1[n]\}$ and let $\{y_2[n]\}$ be the response of the system to the input $\{x_2[n]\}$. Then the system is linear if:

1. **Additivity**: The response to $\{x_1[n]\} + \{x_2[n]\}$ is $\{y_1[n]\} + \{y_2[n]\}$
2. **Homogeneity**: The response to $a\{x_1[n]\}$ is $a\{y_1[n]\}$, where a is any real number if we are considering only real signals and a is any complex number if we are considering complex valued signals.

3. **Continuity:** Let us consider $\{x_1[n]\}, \{x_2[n]\}, ...\{x_k[n]\}...$ be countably infinite number of signals such that lim$\{ x_k[n]\} = \{x[n]\}$ Let the corresponding output signals be denoted by $\{yn[n]\}$ $k\to\infty$ and Lim $\{ yn[n]\} =\{y[n]\}$ We say that system processes the continuity property $k\to\infty$ if the response of the system to the limiting input $\{x[n]\}$ is limit of the responses $\{y[n]\}$.
$T($ lim$\{ xk[n]\})$ = lim $T(\{Xk[n]\})$ $k\to\infty k\to\infty$

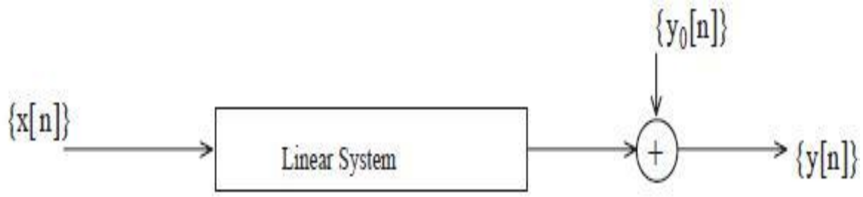
The additive and continuity properties can be replaced by requiring that We say that system posseses the continuity property system is additive for countably infinite number if signals i.e. response to$\{x1[n]\}+\{x2[n]\}+...+\{xn[n]\}+...$ is $\{y1[n]\}+\{y2[n]\}+...+\{yk[n]\}+$....Most of the books do not mention the continuity property. They state only finite additivity and homogeneity. But from finite additivity we can not deduce c....... additivity. This distinction becomes very important in continuous time systems. A system can be linear without being time invariant and it can be time invariant without being linear. If a system is linear, an all zero input sequence will produce a all zero output sequence. $\{0\}$ denote the all zero sequence ,then $\{0\} = 0.\{x[n]\}$.
If $T(\{x[n]\} = \{y[n]\})$ then by homogeneity property $T(0.\{x[n]\}) = 0.\{y[n]\}$

$T(\{0\}) = \{0\}$

Consider the system defined by

$y[n] = 2x[n] + 3$

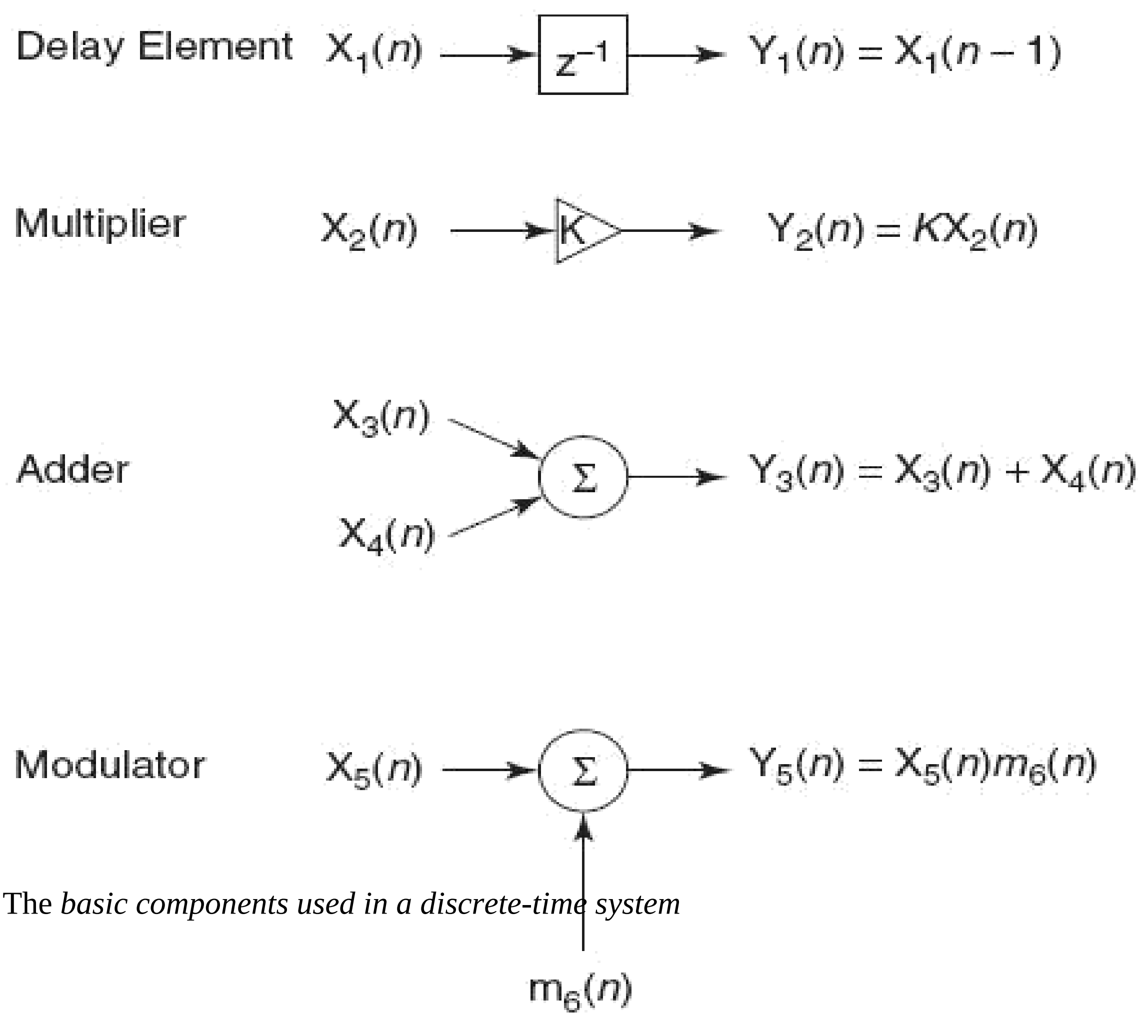
This system is not linear. This can be verified in several ways. If the input is all zero sequence $\{0\}$, the output is not an all zero sequence. Although the defining equation is a linear equation is x and y the system is nonlinear. The output of this system can be represented as sum of a linear system and another signal equal to the zero input response. In this case the linear system is $y[n] = 2x[n]$ and the zero-input response is $y_0[n] = 3$ for all n



systems correspond to the class of incrementally linear system. System is linear in term of difference signal i.e if we define $\{x_d[n]\} = \{x_1[n]\} - \{X_2[n]\}$and $\{y_d[n]\} = \{y_1[n]\} - \{y_2[n]\}$. Then in terms of $\{x_d[n]\}$ and $\{y_d[n]\}$ the system is linear.

4. MODELS OF THE DISCRETE-TIME SYSTEM

First let us consider a discrete-time system as an interconnection of only three basic components: the delay elements, multipliers, and adders. The input– output relationships for these components and their symbols are shown in Figure below. The fourth component is the modulator, which multiplies two or more signals and hence performs a nonlinear operation.



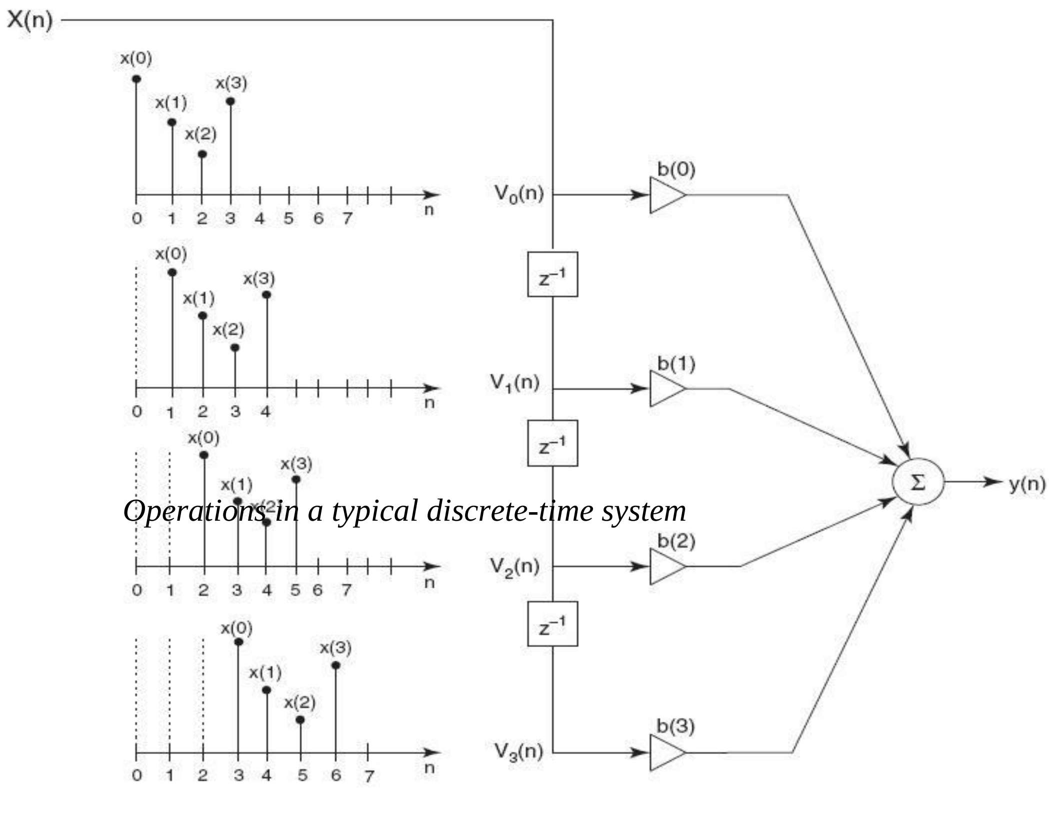
The *basic components used in a discrete-time system*

simple discrete-time system is shown in Figure 5, where input signal $x(n)= \{x(0), x(1), x(2), x(3)\}$ is shown to the left of $v_0(n)= x(n)$. The signal $v_1(n)$shown on the left is the signal $x(n)$delayed by T seconds or one sample, so, $v_1(n)= x(n -1)$. Similarly, $v(2)$and $v(3)$are the signals obtained from $x(n)$when it is delayed by $2T$ and $3T$ seconds: $v_2(n)= x(n - 2)$and $v_3(n)= x(n - 3)$. When we say that the signal $x(n)$is delayed by T, $2T$, or $3T$ seconds, we mean that the samples of the sequence are present T, $2T$, or $3T$ seconds *later*, as shown by the plots of the signals to the left of $v_1(n)$, $v_2(n)$, and $v_3(n)$. But at any given time $t = nT$, the samples in $v_1(n)$, $v_2(n)$, and $v_3(n)$ are the samples of the input signal that occur T,$2T$, and $3T$ seconds *previous* to $t= nT$. For example, at $t = 3T$, the value of the sample in $x(n)$is $x(3)$, and the values present in $v_1(n)$, $v_2(n)$and $v_3(n)$are $x(2)$, $x(1)$, and $x(0)$ respectively.Agood understanding of the operation of the discrete-time system as illustrated in above Figure is essential in analyzing, testing, and debugging the operation of the system when available software is used for the design, simulation, and hardware implementation of the system.
It is easily seen that the output signal in above Figure is

$$y(n) = b(0)v(0) + b(1)v(1) + b(2)v(2) + b(3)v(3)$$

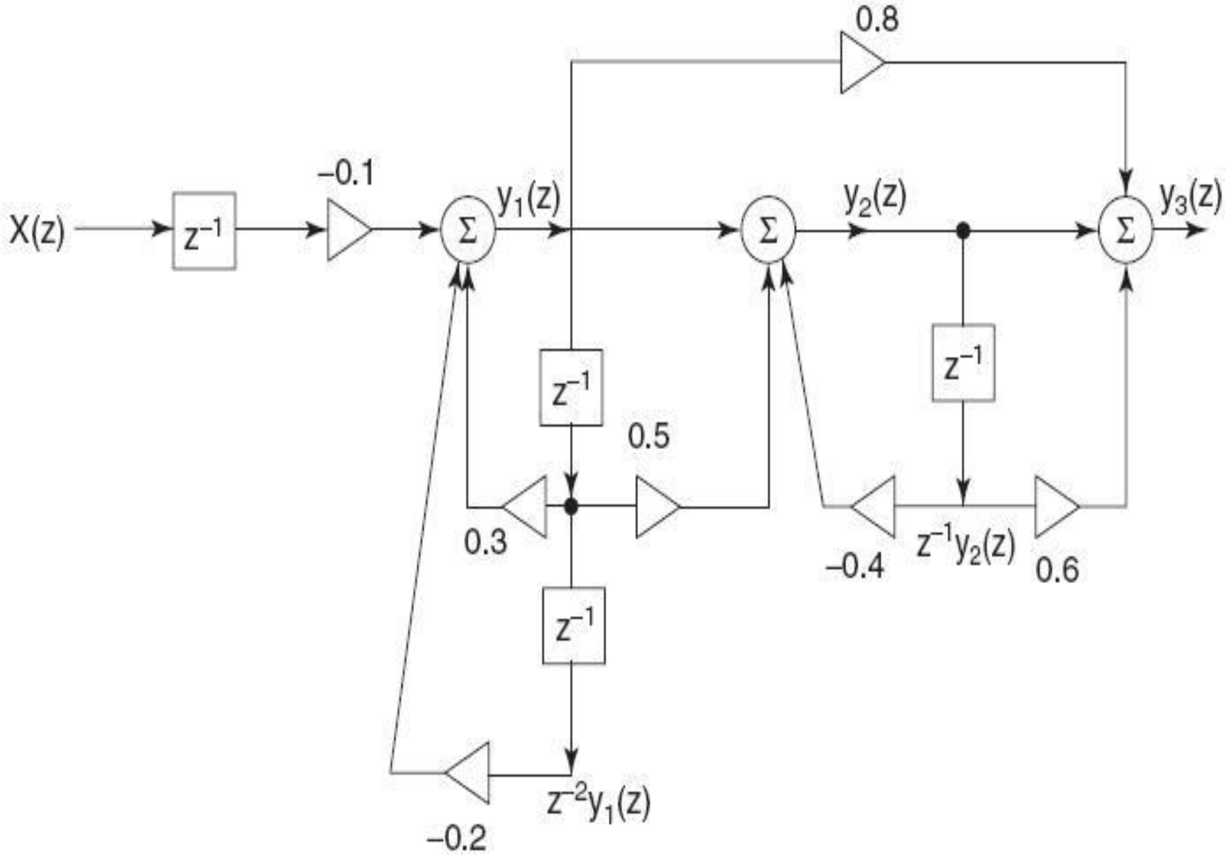
$$= b(0)x(n) + b(1)x(n-1) + b(2)x(n-2) + b(3)x(n-3)$$

where $b(0)$, $b(1)$, $b(2)$, $b(3)$are the gain constants of the multipliers. It is also easy to see from the last expression that the output signal is the weighted sum of the current value and the previous three values of the input signal. So this gives us an input–output relationship for the system shown in below



Operations in a typical discrete-time system

Now we consider another example of a discrete-time system, shown in Figure 5. Note that a fundamental rule is to express the output of the adders and generate as many equations as the number of adders found in this circuit diagram for the discrete-time system. (This step is similar to writing the node equations for an analog electric circuit.) Denoting the outputs of the three adders as $y_1(n)$, $y_2(n)$, and $y_3(n)$, we get



Schematic circuit for a discrete-time system.

$$y_1(n) = 0.3y_1(n-1) - 0.2y_1(n-2) - 0.1x(n-1)$$

$$y_2(n) = y_1(n) + 0.5y_1(n-1) - 0.4y_2(n-1)$$

$$y_3(n) = y_2(n) + 0.6y_2(n-1) + 0.8y_1(n)$$

These three equations give us a mathematical model derived from the model shown in above that is schematic in nature. We can also derive (draw the circuit realization) the model shown in Figure 5 from the same equations given above. After eliminating the internal variables $y_1(n)$and $y_2(n)$; that relationship constitutes the third model for the system. The general form of such an input– output relationship is

Eq(1) $$y(n) = -\sum_{k=1}^{N} a(k)y(n-k) + \sum_{k=0}^{M} b(k)x(n-k)$$

or in another equivalent form

$$\sum_{k=0}^{N} a(k)y(n-k) = \sum_{k=0}^{M} b(k)x(n-k); \qquad a(0) = 1$$

Eq(1) shows that the output $y(n)$is determined by the weighted sum of the previous N values of the output and the weighted sum of the current and previous $M + 1$ values of the input. Very often the coefficient $a(0)$as shown in Eq(2) is normalized to unity.

5. LINEAR TIME-INVARIANT, CAUSAL SYSTEMS

In this section, we study linear time-invariant causal systems and focus on properties such as linearity, time invariance, and causality.

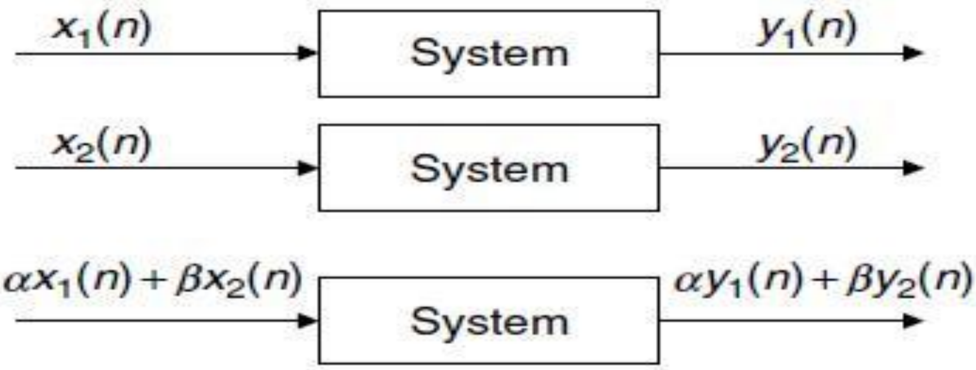
5.1 Linearity:

A linear system is illustrated in below figure, where $y_1(n)$ is the system output using an input $x_1(n)$, and $y_2(n)$ is the system output using an input $x_2(n)$. This Figure illustrates that the system output due to the weighted sum inputs $\alpha x_1(n) + \beta x_2(n)$ is equal to the same weighted sum of the individual outputs obtained from their corresponding inputs, that is

$y(n)=\alpha y_1(n) + \beta y_2(n)$

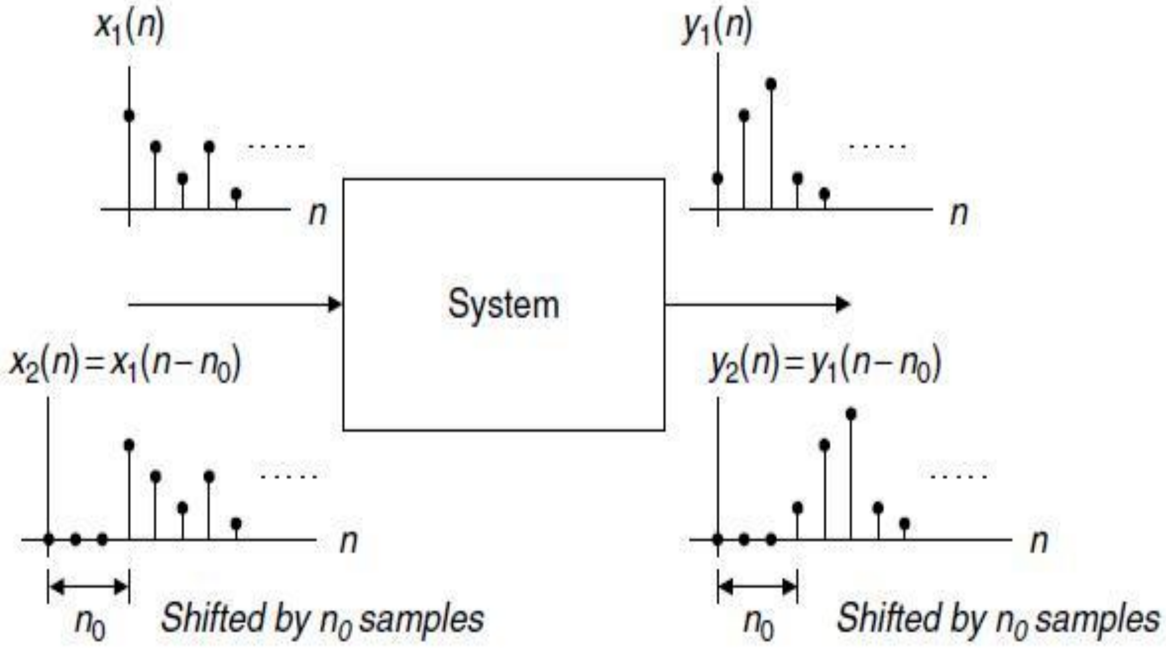
where α and β are constants.

For example, assuming a digital amplifier as $y(n)=10x(n)$, the input is multiplied by 10 to generate the output. The inputs $x_1(n) = u(n)$ and $x_2(n) =\delta(n)$ generate the outputs $y_1(n) =10u(n)$ and $y_2(n) = 10\delta(n)$, respectively. If, as described in below Figure , we apply to the system using the combined input $x(n)$, where the first input is multiplied by a constant 2 while the second input is multiplied by a constant 4, $x(n) = 2x_1(n) + 4x_2(n) =2u(n) + 4\delta(n)$,



5.2 Time Invariance

A time-invariant system is illustrated in Figure below, where $y_1(n)$ is the system output for the input $x_1(n)$. Let $x2(n) = x_1(n - n0)$ be the shifted version of $x_1(n)$



by n_0 samples. The output $y_2(n)$ obtained with the shifted input $x2(n) = x1(n - n_0)$is equivalent to the output $y_2(n)$ acquired by shifting $y_1(n)$ by n_0 samples,$y_2(n) = y_1(n - n_0)$.This can simply be viewed as the following. If the system is time invariant and $y_1(n)$ is the system output due to the input $x_1(n)$,then the shifted system input $x_1(n - n_0)$ will produce a shifted system output $y_1(n - n_0)$by the same amount of time n_0.

5.3 Differential Equations and Impulse Responses:

A causal, linear, time-invariant system can be described by a difference equation having the following general form:

$y(n) + a_1y(n - 1) + . . . + a_Ny(n - N) = b_0x(n) + b_1x(n -1) + . . . + b_Mx(n -M)$ where $a_1, . . . , a_N$ and $b_0, b_1, . . . , b_M$ are the coefficients of the difference equation. It can further be written as
$y(n) = - a_1y(n - 1) -. . . - a_Ny(n - N)+ b_0x(n) + b_1x(n - 1) + . . . + b_Mx(n -M)$

1. FOURIER SERIES COEFFICIENTS OF PERIODIC IN DIGITAL SIGNALS:

Let us look at a process in which we want to estimate the spectrum of a periodic digital signal x(n) sampled at a rate of fs Hz with the fundamental period T_0=NT, as shown in below, where there are N samples within the duration of the fundamental period and T = 1/fs is the sampling period. For the time being, we assume that the periodic digital signal is band limited to have all harmonic frequencies less than the folding frequency fs=2 so that aliasing does not occur. According to Fourier series analysis (Appendix B), the coefficients of the Fourier series expansion of a periodic signal x(t) in a complex form is

$$c_k = \frac{1}{T_0} \int_{T_0} x(t) e^{-jk\omega_0 t} dt \qquad -\infty < k < \infty, \tag{4.1}$$

where k is the number of harmonics corresponding to the harmonic frequency of kf_0 and $\omega_0 = 2\pi/T_0$ and $f_0 = 1/T_0$ are the fundamental frequency in radians per second and the fundamental frequency in Hz, respectively. To apply Equation (4.1), we substitute $T_0 = NT$, $\omega_0 = 2\pi/T_0$ and approximate the integration over one period using a summation by substituting $dt = T$ and $t = nT$. We obtain

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}}, \quad -\infty < k < \infty. \tag{4.2}$$

Since the coefficients c_k are obtained from the Fourier series expansion in the complex form, the resultant spectrum c_k will have two sides. There is an important feature of Equation (4.2) in which the Fourier series coefficient c_k is periodic of N. We can verify this as follows

$$c_{k+N} = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi(k+N)n}{N}} = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi kn}{N}} e^{-j2\pi n}. \tag{4.3}$$

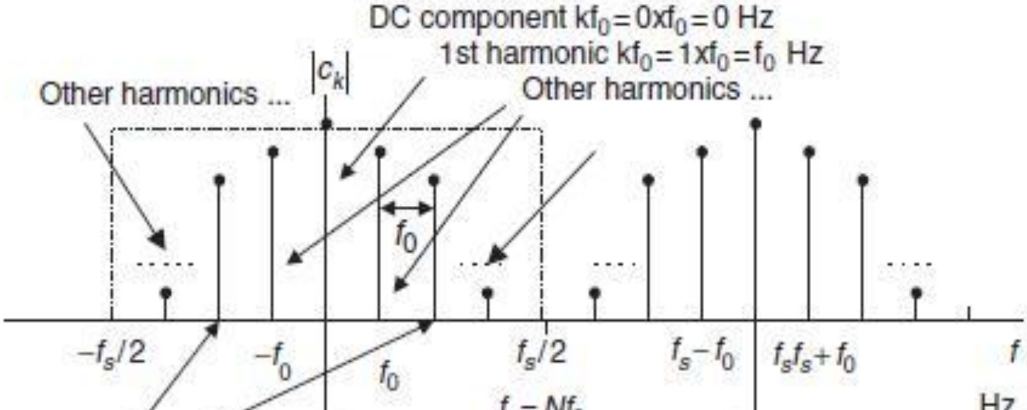
Since $e^{-j2\pi n} = \cos(2\pi n) - j\sin(2\pi n) = 1$, it follows that

$$c_{k+N} = c_k. \tag{4.4}$$

Therefore, the two-sided line amplitude spectrum jckj is periodic, as shown in

Figure 4.3. We note the following points:

a. As displayed in Figure 4.3, only the line spectral portion between the frequency fs=2 and frequency fs=2 (folding frequency) represents the frequency information of the periodic signal.



b. Notice that the spectral portion from fs=2 to fs is a copy of the spectrum in the negative frequency range from _fs=2 to 0 Hz due to the spectrum being periodic for every Nf_0 Hz. Again, the amplitude spectral components indexed from fs=2 to fs can be folded at the folding frequency fs=2 to match the amplitude spectral components indexed from 0 to fs=2 in terms of fs _ f Hz, where f is in the range from fs=2 to fs. For convenience, we compute the spectrum over the range from 0 to fs Hz with nonnegative indices, that is,

$$c_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}}, \; k = 0,1,\ldots, N-1. \qquad (4.5)$$

c. For the kth harmonic, the frequency is f = kf_0 Hz. The frequency spacing between the consecutive spectral lines, called the frequency resolution, is f_0 Hz

7. Discrete Fourier Transform

Now, let us concentrate on development of the DFT. In below Figure shows one way to obtain the DFT formula. First, we assume that the process acquires data samples from digitizing the interested continuous signal for a duration of T seconds. Next, we assume that a periodic signal x(n) is obtained by copying the acquired N data samples with the duration of T to itself repetitively. Note that we assume continuity between the N data sample frames. This is not true in practice. We will tackle this problem in Section 4.3. We determine the Fourier series coefficients using one-period N data samples and Equation (4.5). Then we multiply the Fourier series coefficients by a factor of N to obtain

$$X(k) = Nc_k = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}}, \quad k = 0,\, 1, \ldots,\, N-1,$$

where X(k) constitutes the DFT coefficients. Notice that the factor of N is a constant and does not affect the relative magnitudes of the DFT coefficients X(k). As shown in the last plot, applying DFT with N data samples of x(n) sampled at a rate of fs (sampling period is T = 1/fs) produces N complex DFT

Using periodicity, it follows that

$$c_{-1} = c_3 = j0.5, \text{ and } c_{-2} = c_2 = 0.$$

b. The amplitude spectrum for the digital signal is sketched in Figure 4.5.

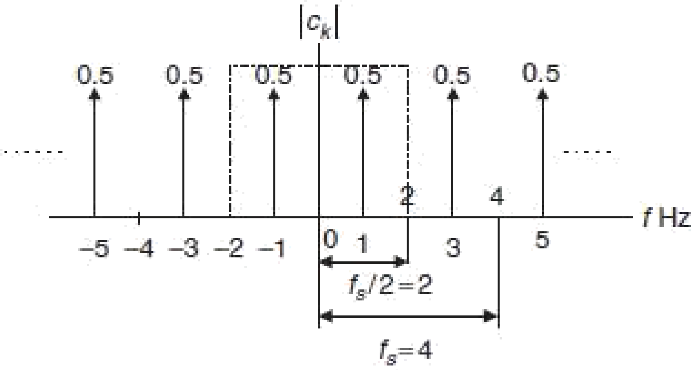


FIGURE 4.5 Two-sided spectrum for the periodic digital signal in Example 4.1.

As we know, the spectrum in the range of -2 to 2 Hz presents the information of the sinusoid with a frequency of 1 Hz and a peak value of 2|c1| = 1, which is converted from two sides to one side by doubling the spectral value. Note that we do not double the direct-current (DC) component.

PROPERTIES OF DISCRETE FOURIER TRANSFORM

As a special case of general Fourier transform, the discrete time transform shares all properties (and their proofs) of the Fourier transform discussed above, except now some of these properties may take different forms. In the following, we always assume $\mathcal{F}[x[m]] = X(e^{j\omega})$ and $\mathcal{F}[y[m]] = Y(e^{j\omega})$.

1. **Linearity** $\mathcal{F}[ax[m] + by[m]] = aX(e^{j\omega}) + bY(e^{j\omega})$

2. **Time Shifting** $\mathcal{F}[x[m - m_0]] = e^{-jm_0\omega}X(e^{j\omega})$

Proof:

$$\mathcal{F}[x[m - m_0]] = \sum_{m=-\infty}^{\infty} x[m - m_0]e^{-j\omega m}$$

If we let $m' = m - m_0$, the above becomes

$$\mathcal{F}[x[-m]] = X(e^{-j\omega})$$

$$\mathcal{F}[x[m - m_0]] = \sum_{m=-\infty} x[m']e \quad) = e^{-j\omega m_0} X(e^{j\omega})$$

= **Time Reversal**

$$\mathcal{F}[x[m]e^{j\omega_0 m}] = X(e^{j(\omega - \omega_0)})$$

= **Frequency Shifting**

= **Differencing**

Differencing is the discrete-time counterpart of differentiation.

$$\mathcal{F}[x[m] - x[m-1]] = (1 - e^{-j\omega})X(e^{j\omega})$$

Proof:

$$\mathcal{F}[x[m] - x[m-1]] = \mathcal{F}[x[m]] - \mathcal{F}[x[m-1]]$$

$$X(e^{j\omega}) - X(e^{j\omega})e^{-j\omega} = (1 - e^{-j\omega})X(e^{j\omega})$$

= **Differentiation in frequency**

$$\mathcal{F}^{-1}[j\frac{d}{d\omega}X(e^{j\omega})] = m\ x[m]$$

proof: Differentiating the definition of discrete Fourier transform with respect to ω, we get

$$\frac{d}{d\omega}X(e^{j\omega}) = \frac{d}{d\omega}\sum_{m=-\infty}^{\infty} x[m]e^{-j\omega m} = \sum_{m=-\infty}^{\infty} x[m]\frac{d}{d\omega}e^{-j\omega m}$$

$$= \sum_{m=-\infty}^{\infty} -jmx[m]e^{-j\omega m}$$

1. **Convolution Theorems**

The convolution theorem states that convolution in time domain corresponds to multiplication in frequency domain and vice versa:

$$\mathcal{F}[x[n] * y[n]] = X(e^{j\omega})\ Y(e^{j\omega}) \qquad (a)$$

$$\mathcal{F}[x[n]\ y[n]] = X(e^{j\omega}) * Y(e^{j\omega}) \qquad (b)$$

Recall that the convolution of periodic signals $x_T(t)$ and $y_T(t)$ is

$$x_T(t) * y_T(t) \triangleq \frac{1}{T}\int_T x_T(\tau)y_T(t-\tau)d\tau$$

Here the convolution of periodic spectra $X(f)$ and $Y(f)$ is similarly defined as

$$X(e^{j\omega})*Y(e^{j\omega}) = \frac{1}{\Omega}\int_{\Omega} X(e^{j\omega'})Y(e^{j(\omega-\omega')})\,d\omega' = \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\omega'})Y(e^{j(\omega-\omega')})\,d\omega'$$

Proof of (a):

$$\begin{aligned}
\mathcal{F}[x[n]*y[n]] &= \sum_{n=-\infty}^{\infty}[\sum_{m=-\infty}^{\infty} x[m]y[n-m]\,]e^{-jn\omega} \\
&= \sum_{m=-\infty}^{\infty} x[m][\sum_{n=-\infty}^{\infty} y[n-m]e^{-j(n-m)\omega}]e^{-jm\omega} \\
&= X(j\omega)\;Y(j\omega)
\end{aligned}$$

roof of (b):

$$\begin{aligned}
\mathcal{F}[x[n]y[n]] &= \sum_{n=-\infty}^{\infty} x[n]y[n]e^{-jn\omega} = \sum_{n=-\infty}^{\infty}[\frac{1}{2\pi}\int_0^{2\pi} X(j\omega')e^{jn\omega'}d\omega']y[n]e^{-jn\omega} \\
&= \frac{1}{2\pi}\int_0^{2\pi} X(j\omega')[\sum_{n=-\infty}^{\infty} e^{jn\omega'}y[n]e^{-jn\omega}]d\omega' \\
&= \frac{1}{2\pi}\int_0^{2\pi} X(j\omega')\sum_{n=-\infty}^{\infty} y[n]e^{-jn(\omega-\omega')}d\omega' \\
&= \frac{1}{2\pi}\int_0^{2\pi} X(j\omega')Y(j(\omega-\omega'))d\omega' = X(j\omega)*Y(j\omega)
\end{aligned}$$

= **Parseval's Relation**

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 = \frac{1}{2\pi} \int_0^{2\pi} |X(e^{j\omega})|^2 d\omega$$

The **circular convolution**, also known as **cyclic convolution**, of two aperiodic functions occurs when one of them is convolved in the normal way with a periodic summation of the other function. That situation arises in the context of the Circular convolution theorem. The identical operation can also be expressed in terms of the periodic summations of both functions, if the infinite integration interval is reduced to just one period. That situation arises in the context of the discrete-time Fourier transform (DTFT) and is also called **periodic convolution**. In particular, the transform (DTFT) of the product of two discrete sequences is the periodic convolution of the transforms of the individual sequences.

For a periodic function x_T, with period T, the convolution with another function, h, is also periodic, and can be expressed in terms of integration over a finite interval as follows**:**

For a periodic function x_T, with period T, the convolution with another function, h, is also periodic, and can be expressed in terms of integration over a finite interval as follows**:**

$$\begin{aligned}(x_T * h)(t) \quad &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} h(\tau) \cdot x_T(t-\tau)\, d\tau \\ &= \int_{t_o}^{t_o+T} h_T(\tau) \cdot x_T(t-\tau)\, d\tau,\end{aligned}$$

[2]

where t_o is an arbitrary parameter, and h_T is a periodic summation of h, defined by**:**

$$h_T(t) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} h(t-kT) = \sum_{k=-\infty}^{\infty} h(t+kT).$$

This operation is a **periodic convolution** of functions x_T and h_T. When x_T is expressed as the periodic summation of another function, x, the same operation may also be referred to as a **circular convolution** of functions h and x.

__Discrete sequences__

Similarly, for discrete sequences and period **N**, we can write the **circular convolution** of functions *h* and *x* as**:**

$$
\begin{aligned}
(x_N * h)[n] &\stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} h[m] \cdot x_N[n-m] \\
&= \sum_{m=-\infty}^{\infty} \left(h[m] \cdot \sum_{k=-\infty}^{\infty} x[n-m-kN] \right).
\end{aligned}
$$

This corresponds to matrix multiplication, and the kernel of the integral transform is a circular matrix

 If a sequence, $x[n]$, represents samples of a continuous function, $x(t)$, with Fourier transform $X(f)$, its DTFT is a periodic summation of $X(f)$.

 Proof**:**

$$
\begin{aligned}
&\int_{-\infty}^{\infty} h(\tau) \cdot x_T(t-\tau)\, d\tau \\
&= \sum_{k=-\infty}^{\infty} \left[\int_{t_o+kT}^{t_o+(k+1)T} h(\tau) \cdot x_T(t-\tau)\ d\tau \right] \\
&\stackrel{\tau \rightarrow \tau + kT}{=} \sum_{k=-\infty}^{\infty} \left[\int_{t_o}^{t_o+T} h(\tau + kT) \cdot x_T(t - \tau - kT)\ d\tau \right] \\
&= \int_{t_o}^{t_o+T} \left[\sum_{k=-\infty}^{\infty} h(\tau + kT) \cdot \underbrace{x_T(t - \tau - kT)}_{X_T(t-\tau),\text{ by periodicity}} \right] d\tau \\
&= \int_{t_o}^{t_o+T} \underbrace{\left[\sum_{k=-\infty}^{\infty} h(\tau + kT) \right]}_{\stackrel{\text{def}}{=}\ h_T(\tau)} \cdot x_T(t-\tau)\ d\tau \qquad (QED)
\end{aligned}
$$

Definition of the Fourier Transform

The Fourier transform (FT) of the function *f .(x)* is the function *F(ω)* where:

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} \, dx$$

and the inverse Fourier transform is

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} \, d\omega$$

Recall that $i = \sqrt{-1}$ and $e^{i\theta} = \cos\theta + i\sin\theta$.

Think of it as a transformation into a different set of basis functions. The Fourier transform uses complex exponentials (sinusoids) of various frequencies as its basis functions.(Other transforms, such as Z, Laplace, Cosine, Wavelet, and Hartley, use different basic functions).

A Fourier transform A Fourier transform pair is often written

$f(x) \leftrightarrow F(\omega)$, or $F(f(x)) = F(\omega)$ where *F* is the Fourier transform operator. If *f .x/* is thought of as a signal (i.e. input data) then we call *F(ω)*the signal‟s *spectrum*. If *f* is thought of as the *impulse response* of a filter (which operates on input data to produce output data) then we call *F* the filter‟s *frequency response*. (Occasionally the line between what‟s signal and what‟s filter becomes blurry).

Example of a Fourier Transform

Suppose we want to create a filter that eliminates high frequencies but retains low frequencies (this is very useful in anti aliasing). In signal processing terminology, this is called an *ideal low pass filter*. So we‟ll specify a box-shaped frequency response with cutoff frequency ω_C

$$F(\omega) = \begin{cases} 1 & |\omega| \le \omega_c \\ 0 & |\omega| > \omega_c \end{cases}$$

What is its impulse response?

We know that the impulse response is the inverse Fourier transform of the frequency response, so taking off our signal processing hat and putting on our mathematics hat, all we need to do is evaluate:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega x}\,d\omega$$

for this particular $F(\omega)$:

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{i\omega x}\,d\omega \\ &= \frac{1}{2\pi}\frac{e^{i\omega x}}{ix}\Big|_{\omega=-\omega_c}^{\omega_c} \\ &= \frac{1}{\pi x}\frac{e^{i\omega_c x} - e^{-i\omega_c x}}{2i} \\ &= \frac{\sin \omega_c x}{\pi x} \qquad\qquad \text{since } \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \\ &= \frac{\omega_c}{\pi}\mathrm{sinc}\left(\frac{\omega_c}{\pi}x\right) \end{aligned}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$. For antialiasing with unit-spaced samples, you want the cutoff frequency to equal the Nyquist frequency, so $\omega_c = \pi$.

Fourier Transform Properties

Rather than write "the Fourier transform of an X function is a Y function", we write the shorthand: $X \leftrightarrow Y$. If z is a complex number and $z = x + iy$ where x and y are its real and imaginary parts, then the complex conjugate of z is $z^* = x - iy$. A function $f(u)$ is *even* if $f(u) = f(-u)$, it is *odd* if $f(u) = -f(-u)$, it is conjugate symmetric if $f(u) = f^*(-u)$, and it is conjugate antisymmetric if $f(u) = -f^*(-u)$.

Convolution Theorem

The Fourier transform of a convolution of two signals is the product of their

Fourier transforms: $f \circledast g \leftrightarrow FG.$. The convolution of two continuous signals *f* and *g* is

$$(f \circledast g)(x) = \int_{-\infty}^{+\infty} f(t)g(x-t)\,dt$$

So $\int_{-\infty}^{+\infty} f(t)g(x-t)\,dt \leftrightarrow F(\omega)G(\omega)$.

The Fourier transform of a product of two signals is the convolution of their Fourier transforms: $fg \leftrightarrow F \circledast G/2\pi$.

Delta Functions

The (Dirac) delta function $\delta(x)$ is defined such that $\delta(x) = 0$ for all $x \neq 0$, $\int_{-\infty}^{+\infty} \delta(t)\,dt = 1$, and for any $f(x)$:

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

The latter is called the *sifting property* of delta functions. Because convolution with a delta is linear shift-invariant filtering, translating the delta by *a* will translate the output by *a*:

$$\Big(f(x) \circledast \delta(x-a)\Big)(x) = f(x-a)$$

Two-point DFT (N=2)

$W_2 = e^{-i\pi} = -1$, and

$$A_k = \sum_{n=0}^{1} (-1)^{kn} a_n = (-1)^{k \cdot 0} a_0 + (-1)^{k \cdot 1} a_1 = a_0 + (-1)^k a_1$$

so

$$\begin{aligned} A_0 &= a_0 + a_1 \\ A_1 &= a_0 - a_1 \end{aligned}$$

Four-point DFT (N=4)

$W_4 = e^{-i\pi/2} = -i$, and

$$A_k = \sum_{n=0}^{3} (-i)^{kn} a_n = a_0 + (-i)^k a_1 + (-i)^{2k} a_2 + (-i)^{3k} a_3 = a_0 + (-i)^k a_1 + (-1)^k a_2 + i^k a_3$$

so

$$\begin{aligned} A_0 &= a_0 + a_1 + a_2 + a_3 \\ A_1 &= a_0 - ia_1 - a_2 + ia_3 \\ A_2 &= a_0 - a_1 + a_2 - a_3 \\ A_3 &= a_0 + ia_1 - a_2 - ia_3 \end{aligned}$$

This can also be written as a matrix multiply:

$$\begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

More on this later.

To compute A quickly, we can pre-compute common subexpressions:

$$\begin{aligned} A_0 &= (a_0 + a_2) + (a_1 + a_3) \\ A_1 &= (a_0 - a_2) - i(a_1 - a_3) \\ A_2 &= (a_0 + a_2) - (a_1 + a_3) \\ A_3 &= (a_0 - a_2) + i(a_1 - a_3) \end{aligned}$$

Convolution

Let's consider two time series, g_i and h_i, where the index i runs from $-\infty$ to ∞. The convolution of these two time series is defined as

$$(g * h)_i = \sum_{j=-\infty}^{\infty} g_{i-j} h_j \qquad (1)$$

This definition is applicable to time series of infinite length. If g and h are finite, they can be extended to infinite length by adding zeros at both ends. After this trick, called zero padding, the definition in Eq. (1) becomes applicable. For example, the sum in Eq. (1) becomes

$$(g * h)_i = \sum_{j=0}^{n-1} g_{i-j} h_j \qquad (2)$$

for the finite time series $h_0, \ldots, h_{n-1}$.

Exercise 1 *Convolution is commutative and associative. Prove that* $g * h = h * g$ *and*

1) $* (g * h) = (f * g) * h$.

Exercise 2 *Convolution is distributive over addition. Prove that* $(g_1 + g_2) * h = g_1 * h + g_2 * h$. *This means that filter ing a signal via convolution is a linear operation.*

Although g and h are treated symmetrically by the convolution, they generally have very different natures. Typically, one is a signal that goes on indefinitely in time. The other is concentrated near time zero, and is called a filter. The output of the convolution is also a signal, a filtered version of the input signal.

Note that filtering a signal via convolution is a linear operation. This is an important property, because it simplifies the mathematics. There are also nonlinear methods of filtering, but they involve more technical difficulties. Because of time limitations, this class will cover linear filters only. Accordingly, we will discuss only neurobiological examples for which linear models work well. But these examples are exceptions to the rule that most everything in biology is nonlinear. Don't jump to the conclusion that linear models are always sufficient.

In Eq. (2), we chose h_i to be zero for all negative i. This is called a *causal filter*, because $g * h$ is affected by h in the present and past, but not in the future. In some contexts, the causality constraint is not important, and one can take $h_{-M}, \ldots, h_M$ to be nonzero, for example.

Formulas are nice and compact, but now let's draw some diagrams to see how convolution works. To take a concrete example, assume a causal filter $(h_0, \ldots, h_{n-1})$. Then the ith component of the convolution $(g * h)_i$ involves aligning g and h this way:

$$\begin{array}{cccccccccccc} \cdots & g_{i-m-1} & g_{i-m} & g_{i-m+1} & \cdots & g_{i-2} & g_{i-1} & g_i & g_{i+1} & g_{i+2} & \cdots \\ \cdots & 0 & 0 & h_{m-1} & \cdots & h_2 & h_1 & h_0 & 0 & 0 & \cdots \end{array}$$

In words, $(g * h)_i$ is computed by looking at the signal g through a window of length

S starting at time i and extending back to time $i - m + 1$. The weighted sum of the signals in the window is taken, using the coefficients given by h.

last is f_{M2+N2}. So discarding the first $|N_1|$ and last N_2 elements of f leaves us with $f_{M1}, \ldots, f_{M2}$. This is time-aligned with the signal $g_{M1}, \ldots, g_{M2}$, and has the same length.

Another motivation for discarding elements at the beginning and end is that they may be corrupted by edge effects. If you are really worried about edge effects, you may have to discard even more elements, which will leave f shorter than g.

Firing rate

Consider a spike train $\rho_1, \ldots, \rho_N$. One estimate of the probability of firing is

$$p = \frac{1}{N} \sum_i \rho_i \qquad (3)$$

This estimate is satisfactory, as long as it makes sense to describe the whole spike train by a single probability that does not vary with time. This is an assumption of statistical *stationarity*.

More commonly, it's a better model to assume that the probability varies slowly with time (is nonstationary). Then it's better to apply something like Eq. (3) to small segments of the spike train, rather than to the whole spike train. For example, the formula

$$p_i = (\rho_{i+1} + \rho_i + \rho_{i-1})/3 \qquad (4)$$

estimates the probability at time i by counting the number of spikes in three time bins, and then dividing by three. In the first problem set, you were instructed to smooth the spike train like this, but to use a much wider window. In general, choosing the size of window involves a tradeoff. A larger window minimizes the effects of statistical sam- pling error (like flipping a coin many times to more accurately determine its probability of coming up heads). But a larger window also reduces the ability to follow more rapid changes in the probability as a function of time.

Note that Eq. (4) isn't to be trusted near the edges of the signal, as the filter operates on the zeros that surround the signal.

There are other methods for estimating probability of firing, many of which can be expressed in the convolutional form,

There are many different ways to choose w, depending on the particulars of the application. Previously we chose w be of length n, with nonzero values equal to $1/n$. This is sometimes called a "boxcar" filter. MATLAB comes with a lot of other filter shapes. Try typing `help bartlett`, and you'll find more information about the Bartlett and other types of windows that are good for smoothing. Depending on the context, you might want a causal or a noncausal filter for estimating probability of firing.

This is causal, but has infinite duration.

Exercise 3 *Prove that the exponential filter is equivalent to*

$p_i = (1 - \gamma)p_{i-1} + \gamma\rho_i$

4 Impulse response

Consider the signal consisting of a single impulse at time zero,The convolution of this signal with a filter h iswhich is just the filter h again. In other words h, is the response of the filter to an impulse, or the *impulse response* function. If the impulse is displaced from time 0 to time i, then the result of the convolution is the filter h, displaced by i time steps.

UNIT II

AMPLITUDE MODULATION AND DOUBLE SIDE BAND SUPPRESSED CARRIER MODULATION

Objective:

The transmission of information-bearing signal over a band pass communication channel, such as telephone line or a satellite channel usually requires a shift of the range of frequencies contained in the signal to another frequency range suitable for transmission. A shift in the signal frequency range is accomplished by modulation. This chapter introduces the definition of modulation, need of modulation, types of modulation- AM, PM and FM, Various types of AM, spectra of AM, bandwidth requirements, Generation of AM & DSB-SC, detection of AM & DSB-SC, and power relations. After studying this chapter student should be familiar with the following

1) Need for modulation

2) Definition of modulation
3) Types of modulation techniques – AM, FM, PM

4) AM definition - Types of AM –Standard AM, DSB, SSB, and VSB
5) Modulation index or depth of modulation and % modulation
SpectrandBandwidth of all types of AM

6) Generation of AM wave using Square law modulator & Switching Modulator

7) Generation of DSB wave using Balanced modulator & Ring modulator

8) Detection of AM wave using Square law detector & Envelope detector

9) Detection of DSB wave using Synchronous detection & Costas loop

10) Power and current relations

11) Problems
12) Frequency Translation

Communication is a process of conveying message at a distance. If the

distance is involved is beyond the direct communication,the communication engineering comes into the picture. The bran engineering which deals with communication systems is known as telecommunication engineering. Telecommunication engineering is classified into two types based on Transmission media. They are:

1)Line communication

2)Radio communication

In Line communication the media of transmission is a pair of conductors called transmission line. In this technique signals are directly transmitted through the transmission lines. The installation and maintenance of a transmission line is not only costly and complex, but also overcrowds the open space.

Modulation: Modulation is defined as the process by which some characteristics (i.e. amplitude, frequency, and phase) of a carrier are varied in accordance with a modulating wave.

Demodulation is the reverse process of modulation, which is used to get back the original message signal. Modulation is performed at the transmitting end whereas demodulation is performed at the receiving end.

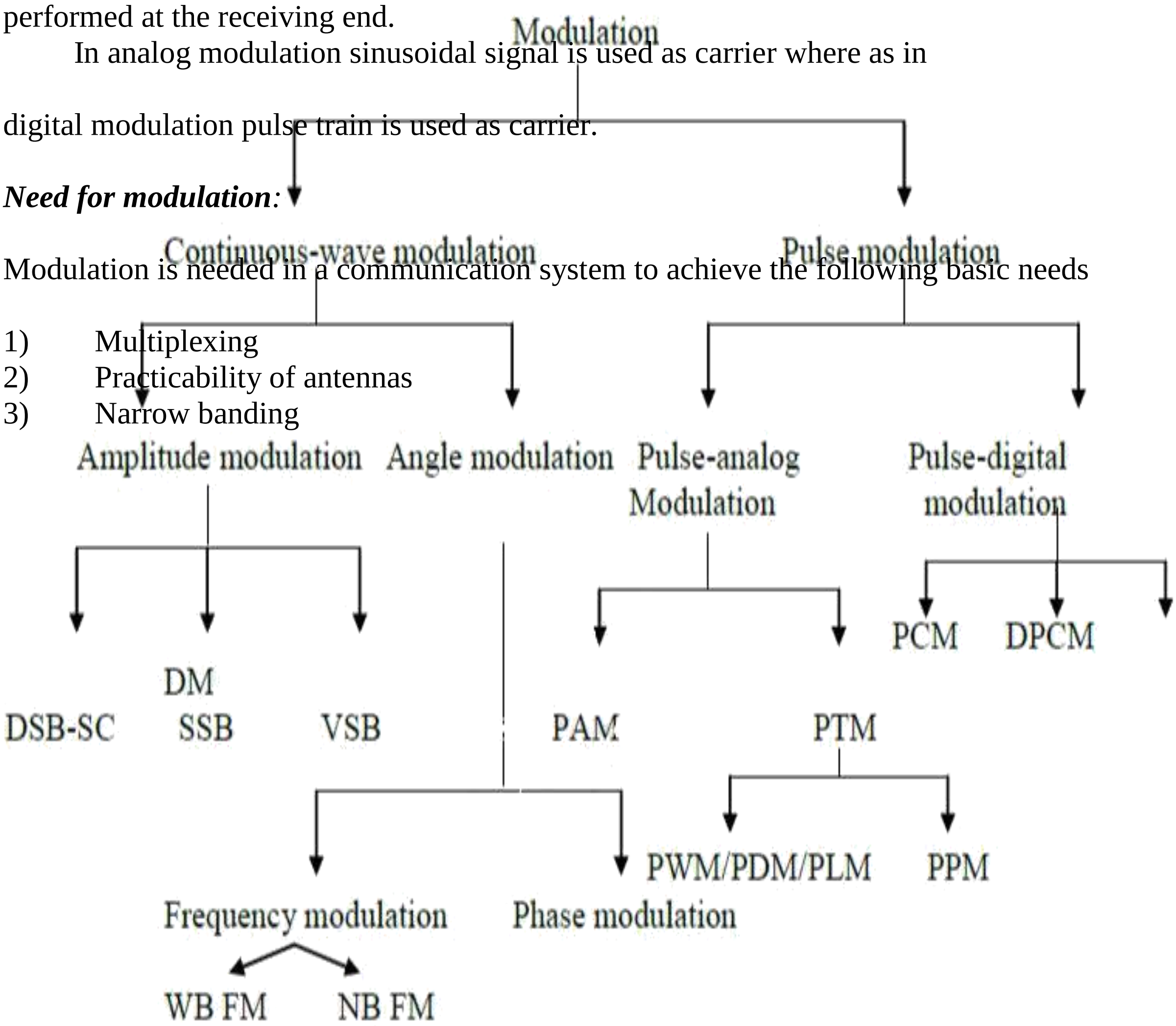
In analog modulation sinusoidal signal is used as carrier where as in

digital modulation pulse train is used as carrier.

Need for modulation:

Modulation is needed in a communication system to achieve the following basic needs

1) Multiplexing
2) Practicability of antennas
3) Narrow banding



Types of modulation:

Continuous wave modulation (CW): When the carrier wave is continuous in nature the modulation process is known as continuous wave modulation.

Pulse modulation: When the carrier wave is a pulse in nature the modulation process is known as continuous wave modulation

Amplitude modulation (AM): A modulation process in which the amplitude of the carrier is varied in accordance with the instantaneous value of the modulating signal.

Amplitude modulation

Amplitude modulation is defined as the process in which the amplitude of the carrier signal is varied in accordance with the modulating signal or message signal.

Consider a sinusoidal carrier signal C (t) is defined as

$$C(t) = A_c \mathrm{Cos}\,(2\pi f_c t + \Theta)\, t$$

For our convenience, assume the phase angle of the carrier signal is zero. An amplitude-modulated (AM) wave S(t) can be described as function of time is given by

$$S(t) = A_c\,[1+k_a m(t)] \cos 2\pi f_c t$$

Where ka = Amplitude sensitivity of the modulator.

The amplitude modulated (AM) signal consists of both modulated carrier signal and un modulated carrier signal.
There are two requirements to maintain the envelope of AM signal is same as the shape of base band signal.

The amplitude of the kam(t) is always less than unity i.e., $|k_a m(t)|<1$ for all 't'.
The carrier signal frequency fc is far greater than the highest frequency component *W* of the message signal m (t) i.e., $f_c >> W$

Assume the message signal m (t) is band limited to the interval –W f W

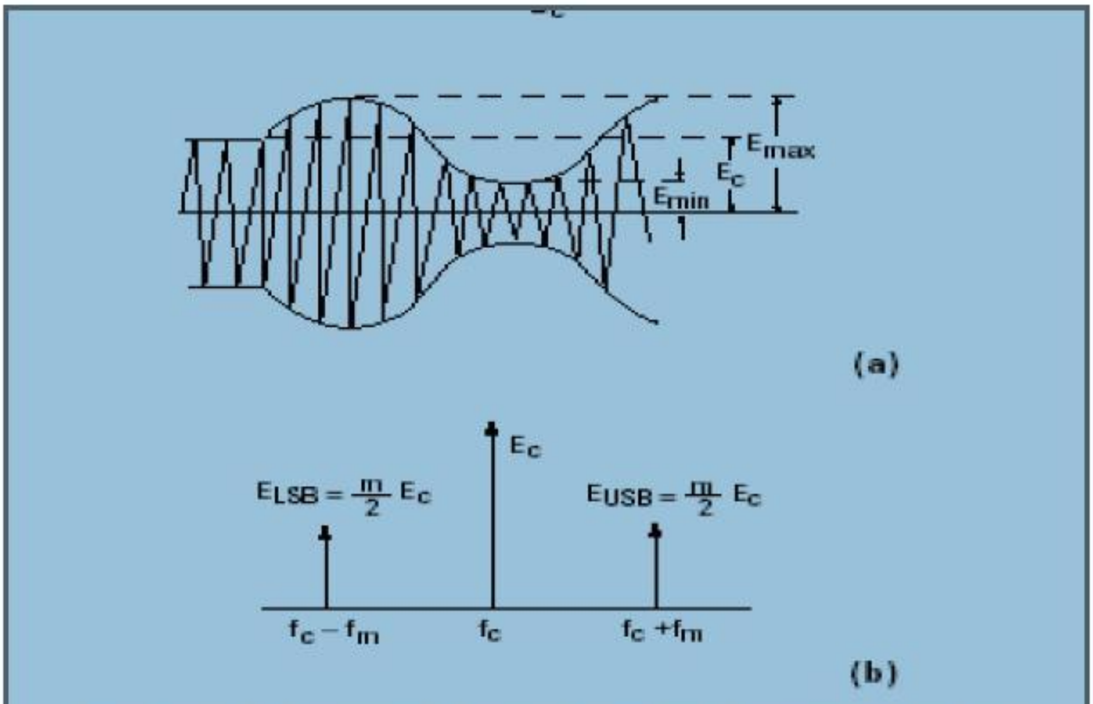


Figure 2(a)(b). Calculation of degree of amplitude modulation from time domain and frequency domain displays

9

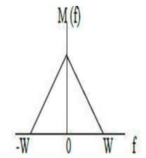


Fig. 1: Spectrum of message signal

The Fourier transform of AM signal S (t) is

$$S(f) = A_c/2\ [\delta(f-f_c)+ \delta(f+f_c)]+k_aA_c/2[M(f-f_c)+ M(f+f_c)]$$

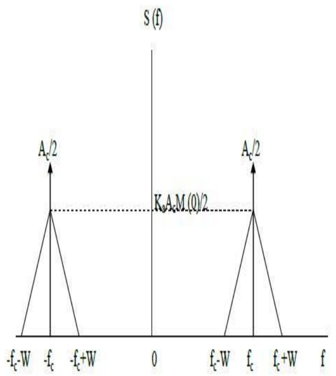


Fig : Spectrum of AM signal

The AM spectrum consists of two impulse functions which are located at fc and -fc and weighted by Ac/2, two USBs, band of frequencies from fc to fc +W and band of frequencies from -fc-W to –fc, and two LSBs, band of frequencies from fc-W to fc and -fc to -fc+W.
The difference between highest frequency component and lowest frequency component is known as transmission bandwidth. i.e.,

BT = 2W
The envelope of AM signal is Ac [1+kam (t)].

Single-tone modulation:

In single-tone modulation modulating signal consists of only one frequency component where as in multi-tone modulation modulating signal consists of more than one frequency component.

S (t) = Ac[1+kam(t)]cos 2πfct ………..(i)

Let m (t) = Amcos 2πfmt

Substitute m (t) in equation (i)

S (t) = Ac [1+ka Amcos 2π fmt] cos 2π fct

Replace the term ka Am by which is known as modulation index or modulation factor.

Modulation index is defined as the ratio of amplitude of message signal to the amplitude of carrier signal. i.e.,

= Am/Ac

(In some books modulation index is designated as ***"m"***)

Which can also be expressed in terms of
Amax and Amin? = (Amax-Amin)/ (Amax+Amin)

Where Amax = maximum amplitude of the modulated carrier signal

Amin = minimum amplitude of the modulated carrier signal

T (t) = Ac cos (2π fct)+Ac /2[cos2 π(fc+fm)t]+ Ac /2[cos2π (fc-fm)t]

Fourier transform of S (t) is

S (f) =Ac/2[(f-fc) + (f+fc)] +Ac /4[(f-fc-fm) + (f+fc+fm)]

+ Ac /4[(f- fc+fm) + (f+fc-fm)]

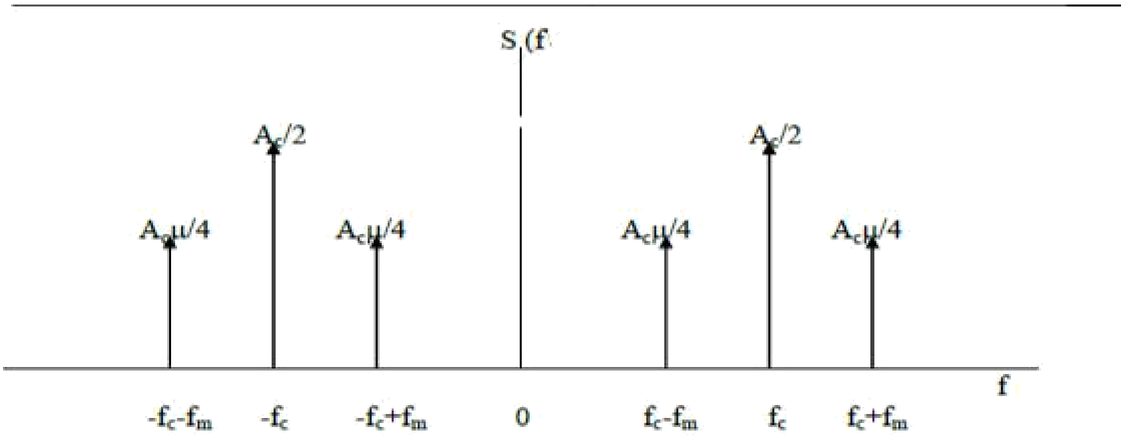


Fig. Spectrum of Single tone AM signal

Power calculations of single-tone AM signal:

The standard time domain equation for single-tone AM signal is given

T (t) = Accos (2 fct) +Ac /2[cos2 (fc+fm) t] + Ac /2[cos2 (fc-fm) t]

Power of any signal is equal to the mean square value of the signal

Carrier power Pc = $Ac^2/2$

Upper Side Band power P_{USB} = $Ac^2\ ^2/8$

Lower Side Band power P_{LSB} = $Ac^2\ ^2/8$

Total power $P_T = Pc + P_{LSB} + P_{USB}$

Total power $P_T = Ac^2/2 + Ac^2\ ^2/8 + Ac^2\ ^2/8$

$P_T = P_c\ [1+\ ^2/2]$

Multi-tone modulation:

In multi-tone modulation modulating signal consists of more than one

frequency component where as in single-tone modulation modulating signal consists of only one frequency component.

S (t) = Ac [1+k_am (t)] cos 2πfct……….. (i)

Let m (t) = Am1cos 2 fm1t + Am2cos 2 fm2t Substitute m (t) in equation (i)

S (t) = Ac [1+ka Am1cos 2 fm1t+ka Am2cos 2 fm2t] cos 2 fct

Replace the term ka Am1 by 1 and Am2 by 2 S (t) = Accos (2 fct) + Ac 1/2[cos2 (fc+fm1) t]+Ac 1/2[cos2 (fc-fm1) t] + Ac 2/2[cos2 (fc+fm2) t] + Ac 2/2[cos2 (fc-fm2) t]

Fourier transform of S (t) is

$$S(f)=A_c/2[\delta(f-f_c)+\delta(f+f_c)]+A_c\mu_1/4[\delta(f-f_c-f_{m1})+\delta(f+f_c+f_{m1})]+$$

$$A_c\mu_1/4[\delta(f-f_c+f_{m1})+\delta(f+f_c-f_{m1})]+A_c\mu_2/4[\delta(f-f_c-f_{m2})+\delta(f+f_c+f_{m2})]+A_c\mu_2/4[\delta(f-f_c+f_{m2})+\delta(f+f_c-f_{m2})]$$

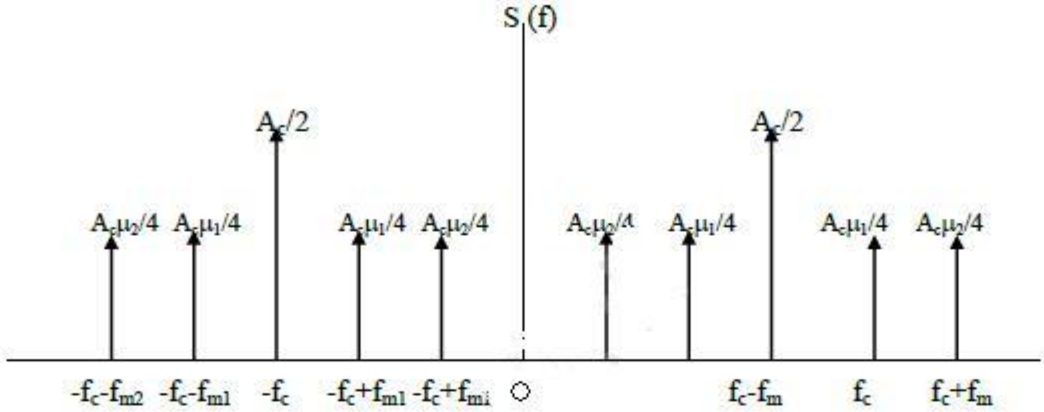


Fig. Spectrum of Multi tone AM signal

Power of Multi-tone AM signal is given by

$$P_T = P_c\,[1+\mu_1^2/2+\mu_2^2 2+ \ldots\ldots\ldots +\mu_n^2 2]$$

$$P_T = P_c\,[1+\mu_t/2]$$

Where $\mu_t=\sqrt{\mu_1^2+\mu_2^2+\ldots\ldots\ldots+\mu_n^2}$

Transmission efficiency (ŋ):-

Transmission efficiency is defined as the ratio of total side band power to the total transmitted power.

i.e., =PSB/PT or ^{2}/ (2+ ^{2})

Advantages of Amplitude modulation:-

Generation and detection of AM signals are very easy

It is very cheap to build, due to this reason it I most commonly used in AM radio broad casting

Disadvantages of Amplitude of modulation:-

Amplitude modulation is wasteful of power

Amplitude modulation is wasteful of band width

Application of Amplitude modulation: - AM Radio Broadcasting

Generation of AM waves

There are two methods to generate AM waves

- Square-law modulator
- Switching modulator

Square-law modulator: -

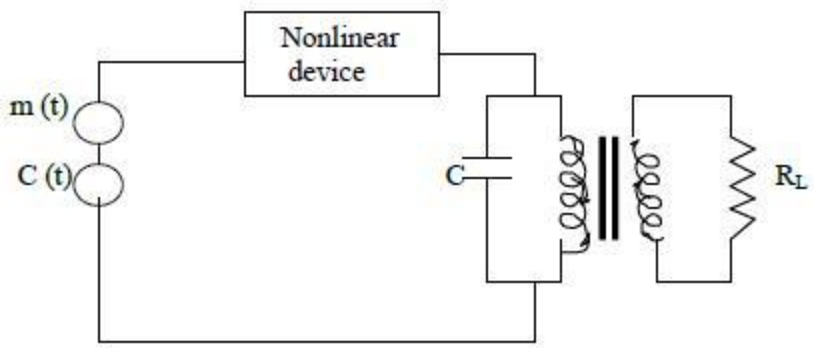


Fig. Square-law Modulator.

A Square-law modulator requires three features: a means of summing the carrier and modulating waves, a nonlinear element, and a band pass filter for extracting the desired modulation products. Semi-conductor diodes and transistors are the most common nonlinear devices used for implementing square law modulators. The filtering requirement is usually satisfied by using a single or double tuned filters.

When a nonlinear element such as a diode is suitably biased and operated in restricted portion of its characteristic curve, that is ,the signal applied to the diode is relatively weak, we find that transfer characteristic of diode-load resistor combination can be represented closely by a square law :

V_0 (t) = a_1V_i (t) + a_2 V_i^2(t) ……………….(i)

Where a1, a2 are constants

Now, the input voltage Vi (t) is the sum of both carrier and message signals i.e.,

Vi (t) =$A_c$cos 2πfct+m (t) ……………. (ii)

Substitute equation (ii) in equation (i) we get

V_0 (t) =a_1A_c [1+k_am (t)] cos2 f_ct +a_1m (t) +a_2A_c2cos22 fct+a_2m^2(t)……….(iii)

Where ka =$2a_2/a_1$

Now design the tuned filter /Band pass filter with center frequency fc and pass band frequency width 2W.We can remove the unwanted terms by passing this output voltage V0(t) through the band pass filter and finally we will get required AM signal.

V0 (t) =a1Ac [1+2a2/a1 m (t)] cos2 fct

Assume the message signal m (t) is band limited to the interval –W f W

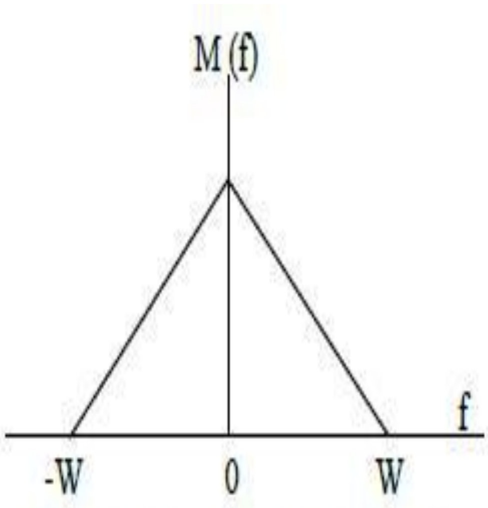


Fig .Spectrum of message signal

The Fourier transform of output voltage V_O (t) is given by

$$V_O(f) = a_1 A_C/2[\delta(f-f_c) + \delta(f+f_c)] + a_2 A_C [M(f-f_c) + M(f+f_c)]$$

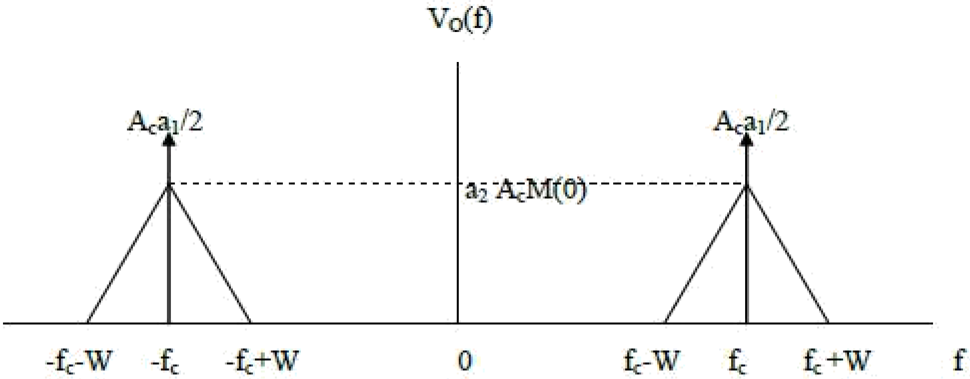
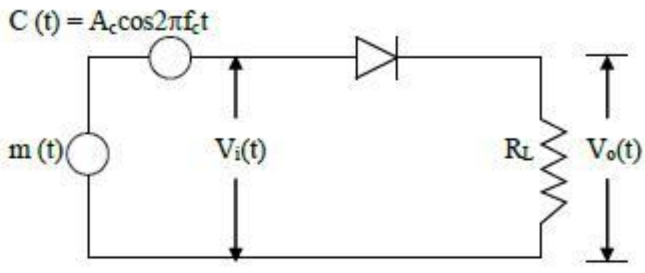


Fig: Spectrum of AM signal

The AM spectrum consists of two impulse functions which are located at fc & -fc and weighted by Aca1/2 & a2Ac/2, two USBs, band of frequencies from fc to fc +W and band of frequencies from -fc-W to –fc, and two LSBs, band of frequencies from fc-W to fc & -fc to -fc+W.

Switching Modulator: -



Assume that carrier wave C (t) applied to the diode is large in amplitude, so that it swings right across the characteristic curve of the diode .we assume that the diode acts as an ideal switch, that is, it presents zero impedance when it is forward-biased and infinite impedance when it is reverse-biased. We may thus approximate the transfer characteristic of the diode-load resistor combination by a piecewise-linear characteristic.

The input voltage applied Vi (t) applied to the diode is the sum of both carrier and message signals.

Vi (t) $=A_c \cos 2\pi f_c t+m(t)$ …………….(i)

During the positive half cycle of the carrier signal i.e. if C (t)>0, the diode is forward biased, and then the diode acts as a closed switch. Now the output voltage Vo (t) is same as the input voltage Vi (t) . During the negative half cycle of the carrier signal i.e. if C(t)<0, the diode is reverse biased, and then the diode acts as a open switch. Now the output voltage V_O (t) is zero i.e. the output voltage varies periodically between the values input voltage Vi (t) and zero at a rate equal to the carrier frequency fc.

i.e., Vo (t) = $[A_c\cos 2\pi f_c t+m(t)]\, P(t)$……….(ii)

Where gp(t) is the periodic pulse train with duty cycle one-half and period Tc=1/fc and which is given by

gP(t)= ½+2/ ∑ [(-1)n-1/(2n-1)]cos [2 fct(2n-1)]…………(iii)

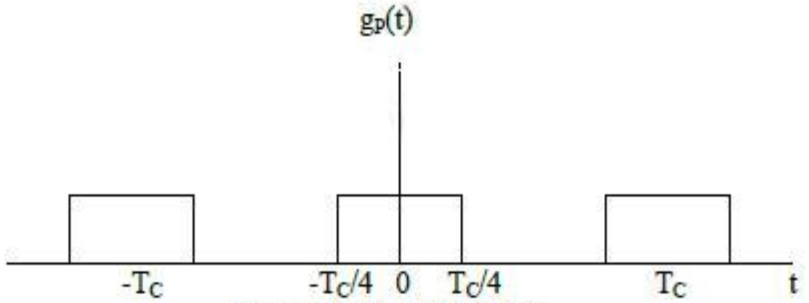


Fig. Periodic pulse train

V0 (t) =Ac/2[1+kam (t)] cos2 fct +m (t)/2+2AC/cos 2πfct……….(iii)

Where ka = 4/ AC

Now design the tuned filter /Band pass filter with center frequency fc and pass band frequency width 2W.We can remove the unwanted terms by passing this output voltage V0(t) through the band pass filter and finally we will get required AM signal.

$V0\ (t) = Ac/2[1+kam\ (t)]\cos 2\pi fct$

Assume the message signal m(t) is band limited to the interval –W f W

The AM spectrum consists of two impulse functions which are located at fc & -fc and weighted by Aca1/2 & a2Ac/2, two USBs, band of frequencies from fc to fc +W and band of frequencies from -fc-W to –fc, and two LSBs, band of frequencies from fc-W to fc & -fc to -fc+W.

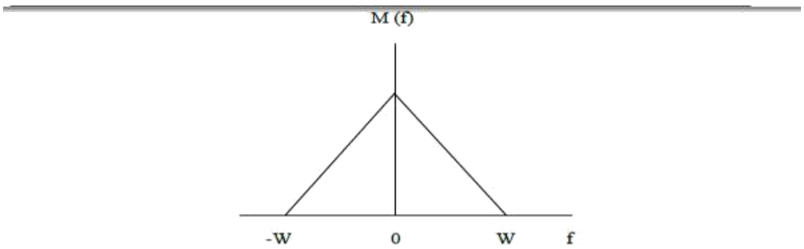


Fig: Spectrum of message signal

The Fourier transform of output voltage V_O (t) is given by

$V_O(f) = A_C/4[\delta(f-f_c) + \delta(f+f_c)] + A_C/\pi[M(f-f_c) + M(f+f_c)]$

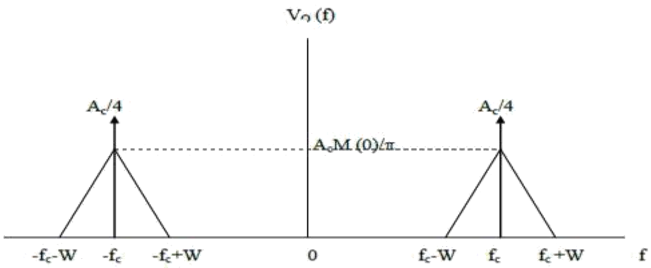


Fig. Spectrum of AM signal

Demodulation of AM waves:

There are two methods to demodulate AM signals.
They are:

- Square-law detector
- Envelope detector

Square-law detector:-

A Square-law modulator requires nonlinear element and a low pass filter for extracting the desired message signal. Semi-conductor diodes and transistors are the most common nonlinear devices used for implementing square law modulators. The filtering requirement is usually satisfied by using a single or double tuned filters. When a nonlinear element such as a diode is suitably biased and operated in a

restricted portion of its characteristic curve, that is ,the signal applied to the diode is relatively weak, we find that transfer characteristic of diode-load resistor combination can be represented closely by a square law :

$V_0 (t) = a_1 V_i (t) + a_2 V_i^2 (t)$ ……………….(i)
Where a_1, a_2 are constants

Now, the input voltage $V_i (t)$ is the sum of both carrier and message signals i.e., $V_i (t) = A_c [1+k_a m (t)] \cos 2 f_c t$…………….(ii)

Substitute equation (ii) in equation (i) we get

$V_0 (t) = a_1 A_c [1+k_a m (t)] \cos 2 f_c t + 1/2\, a_2 A_c^2 [1+2 k_a m (t) + k_a^2 m^2 (t)]$
$[\cos 4 f_c t]$………..(iii)

Now design the low pass filter with cutoff frequency f is equal to the required message signal bandwidth. We can remove the unwanted terms by passing this output voltage V_0 (t) through the low pass filter and finally we will get required message signal.

$V_0 (t) = A_c^2 a_2 m (t)$

The Fourier transform of output voltage $V_O (t)$ is given by $V_O (f) = A_c^2 a_2 M (f)$

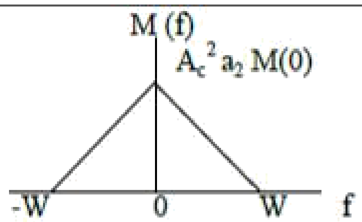


Fig: Spectrum of Output signal

Envelope Detector:

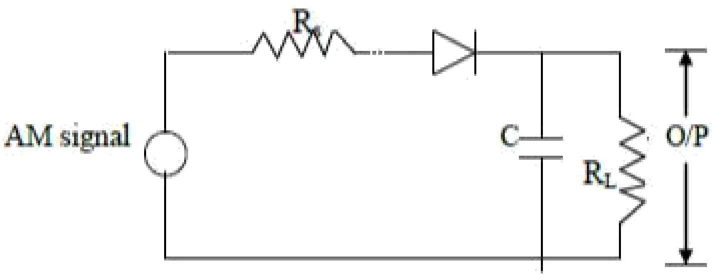


Fig: Envelope detector

Envelope detector is used to detect high level modulated levels, whereas

square-law detector is used to detect low level modulated signals (i.e., below 1v). It is also based on the switching action or switching characteristics of a diode. It consists of a diode and a resistor-capacitor filter. The operation of the envelope detector is as follows. On a positive half cycle of the input signal, the diode is forward biased and the capacitor C charges up rapidly to the peak value of the input signal. When the input signal falls below this value, the diode becomes reverse biased and the capacitor C discharges slowly through the load resistor Rl . The discharging process continues until the next positive half cycle. When the input signal becomes greater than the voltage across the capacitor, the diode conducts again and the process is repeated.

The charging time constant RsC is very small when compared to the

carrier period 1/fc i.e., RsC << 1/fc
Where Rs = internal resistance of the voltage source.
C = capacitor
fc = carrier frequency
i.e., the capacitor C charges rapidly to the peak value of the signal. The discharging time constant RlC is very large when compared to the

charging time constant i.e.,

$1/fc << RlC << 1/W$

Where Rl = load resistance value

W = message signal bandwidth
i.e., the capacitor discharges slowly through the load resistor.

Advantages:

- It is very simple to design
- It is inexpensive
- Efficiency is very high when compared to Square Law detector

Disadvantage:

- Due to large time constant, some distortion occurs which is known

as diagonal clipping i.e., selection of time constant is somewhat difficult

Application:

- It is most commonly used in almost all commercial AM Radio receivers.

Types of Amplitude modulation:-

There are three types of amplitude modulation.
They are:

- Double Sideband-Suppressed Carrier(DSB-SC) modulation
- Single Sideband(SSB) modulation
- Vestigial Sideband(SSB) modulation

DOUBLE SIDEBAND-SUPPRESSED CARRIER (DSBSC) MODULATION

Double sideband-suppressed (DSB-SC) modulation, in which the transmitted wave consists of only the upper and lower sidebands. Transmitted power is saved through the suppression of the carrier wave, but the channel bandwidth requirement is same as in AM (i.e. twice the bandwidth of the message signal). Basically, double sideband-suppressed (DSB-SC) modulation consists of the product of both the message signal m (t) and the carrier signal c(t),as follows:

S (t) =c (t) m (t)

S (t) =Ac cos (2 fct) m (t) The modulated signal s (t) undergoes a phase reversal whenever the message signal m (t) crosses zero. The envelope of a DSB-SC modulated signal is different from the message signal. The transmission bandwidth required by DSB-SC modulation is the same as that for amplitude modulation which is twice the bandwidth of the message signal, 2W. Assume that the message signal is band-limited to the interval $–W \leq f \leq W$

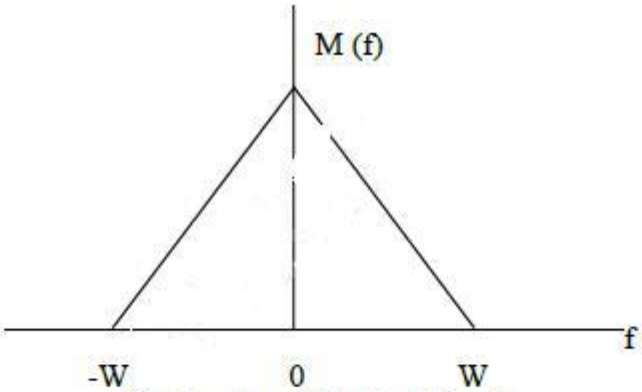


Fig. Spectrum of Baseband signal

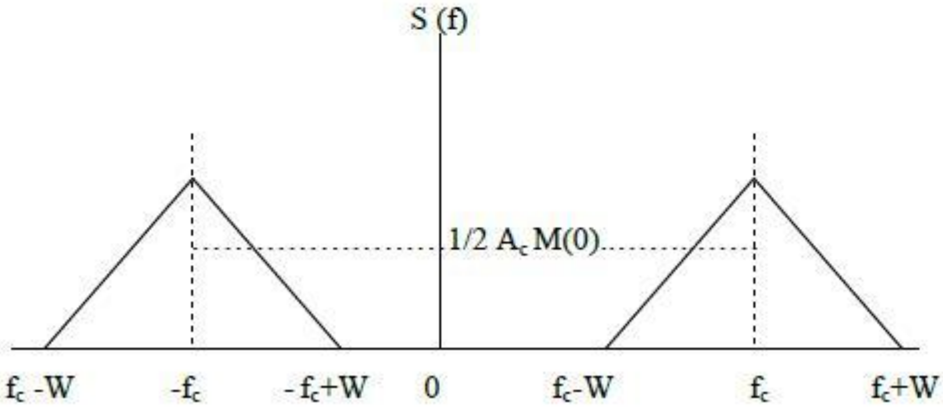


Fig. Spectrum of DSBSC wave

Single-tone modulation:-

In single-tone modulation modulating signal consists of only one frequency component where as in multi-tone modulation modulating signal consists of more than one frequency components. The standard time domain equation for the DSB-SC modulation is given by

S (t) =Ac cos (2π f_ct) m (t)………………… (1)

Assume m (t) =$A_m$cos (2π f_mt)……………….. (2)

Substitute equation (2) in equation (1) we will get

$$S(t) = A_c A_m \cos(2\pi f_c t) \cos(2\pi f_m t)$$

S (t) = Ac Am/2[cos 2π (fc-fm) t + cos 2π (fc+fm)t]…………… (3)

The Fourier transform of s (t) is

S (f) =Ac Am/4[δ (f-fc-fm) + δ (f+fc+fm)] + Ac Am/4[δ fc+fm) + δ(f+fc+fm)]

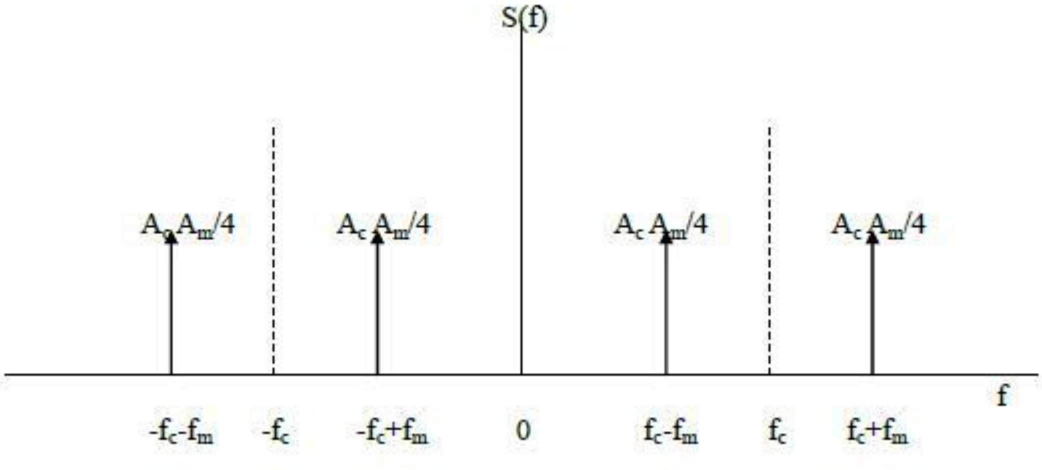


Fig. Spectrum of Single tone DSBSC wave

Power calculations of DSB-SC waves:-

Total power PT = PLSB+PUSB
Total power PT =Ac2Am2/8+Ac2Am2/8
Total power PT =Ac2Am2/4

Generation of DSB-SC waves:-

There are two methods to generate DSB-SC waves. They are:

Balanced modulator
Ring modulator

Balanced Modulator:-

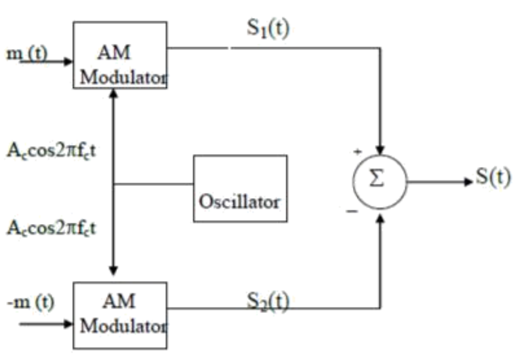


Fig. Balanced Modulator

One possible scheme for generating a DSBSC wave is to use two AM modulators arranged in a balanced configuration so as to suppress the carrier wave, as shown in above fig. Assume that two AM modulators are identical, except for the sign reversal of the modulating signal applied to the input of one of the modulators. Thus the outputs of the two AM modulators can be expressed as follows:

$S1\ (t) = Ac\ [1+kam\ (t)] \cos 2\pi\ fct$

and

$S2\ (t) = Ac\ [1- kam\ (t)] \cos 2\pi\ fct$

Subtracting S2 (t) from S1 (t), we obtain

$S\ (t) = S1\ (t) – S2\ (t)$

$S\ (t) = 2Ac\ kam\ (t) \cos 2\ \pi fct$

Hence, except for the scaling factor 2ka the balanced modulator output is equalto product of the modulating signal and the carrier signal

The Fourier transform of s (t) is

$$S\ (f) =kaAc\ [M\ (f-fc) + M\ (f+fc)]$$

Assume that the message signal is band-limited to the interval $–W \leq f \leq W$

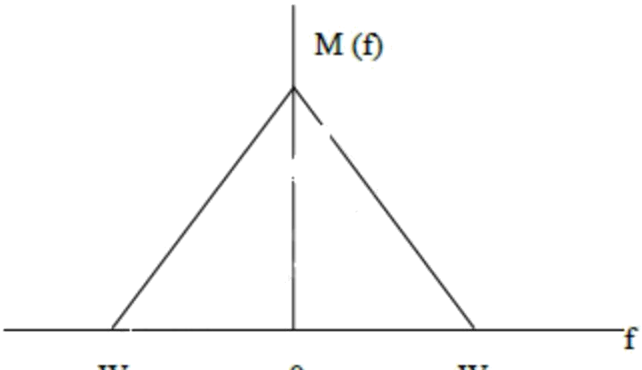
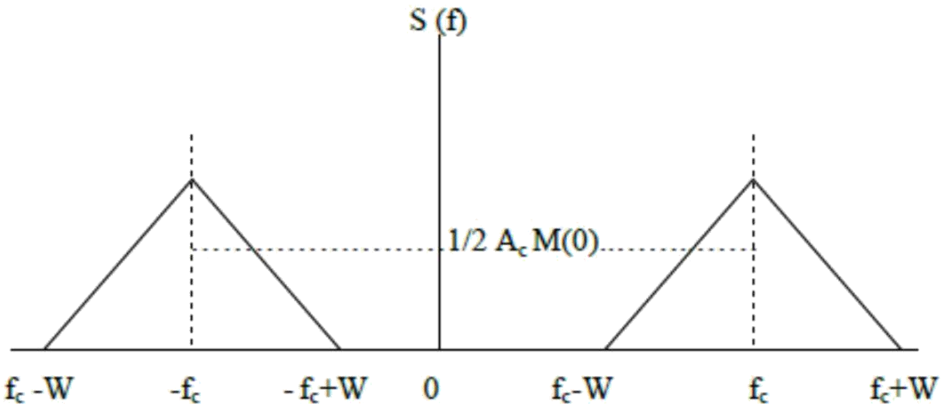
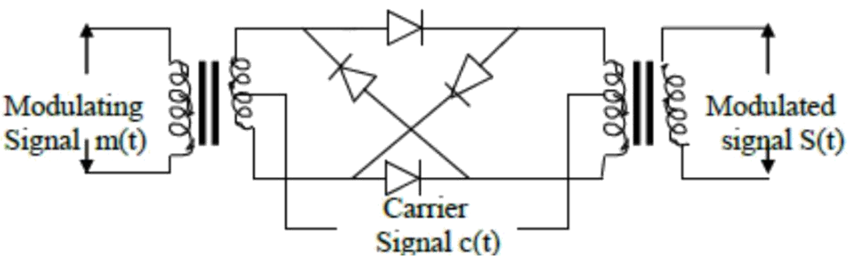


Fig. Spectrum of Baseband signal



Ring modulator:-



One of the most useful product modulator, well suited for generating a DSBSC wave, is the ring modulator shown in above figure. The four diodes form ring in which they all point in the same way-hence the name. The diodes are controlled by a square-wave carrier c (t) of frequency fc, which applied longitudinally by means of to center-tapped transformers. If the transformers are perfectly balanced and the diodes are identical, there is no leakage of the modulation frequency into the modulator output.

On one half-cycle of the carrier, the outer diodes are switched to their forward resistance rf and the inner diodes are switched to their backward resistance rb .on other half-cycle of the carrier wave, the diodes operate in the opposite condition. The square wave carrier c (t) can be represented by a Fourier series as follows:

$$c(t)=4/\pi \sum_{n=1}^{\infty} (-1)^{n-1}/(2n-1) \cos [2\pi f_c t(2n-1)]$$

When the carrier supply is positive, the outer diodes are switched ON

and the inner diodes are switched OFF, so that the modulator multiplies the message signal by +1 When the carrier supply is positive, the outer diodes are switched ON and the inner diodes are switched OFF, so that the modulator multiplies the message signal by +1.when the carrier supply is negative, the outer diodes are switched OFF and the inner diodes are switched ON, so that the modulator multiplies the message signal by -1. Now, the Ring modulator output is the product of both message signal m (t) and carrier signal c (t).

$$S(t) = c(t)\, m(t)$$

$$S(t) = 4/\pi \sum_{n=1}^{\infty} (-1)^{n-1}/(2n-1) \cos [2\pi f_c t (2n-1)]\, m(t)$$

For n=1

$$S(t) = 4/\pi \cos (2\pi f_c t)\, m(t)$$

There is no output from the modulator at the carrier frequency i.e the

modulator output consists of modulation products. The ring modulator is

sometimes referred to as a double-balanced modulator, because it is balanced with respect to both the message signal and the square wave carrier signal. The Fourier transform of s (t) is

$$S(f) = 2/\pi [M(f-f_c) + M(f+f_c)]$$

Assume that the message signal is band-limited to the interval $-W \le f \le W$

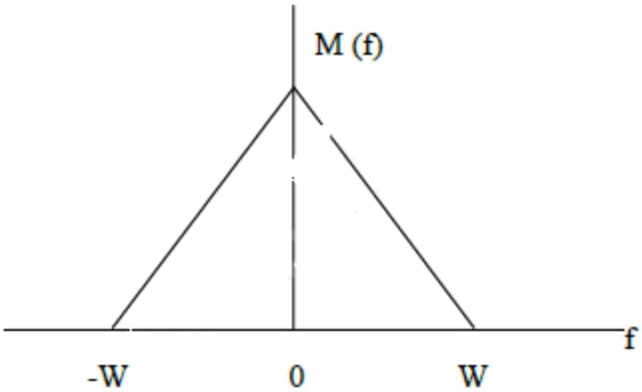
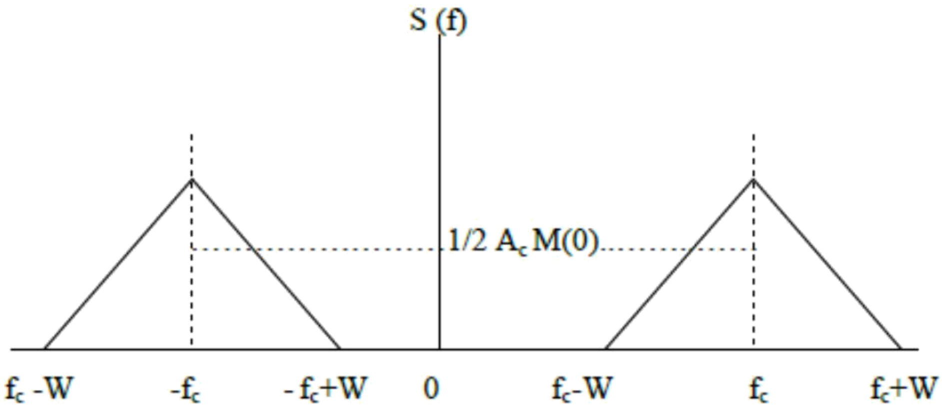


Fig. Spectrum of Baseband signal



<u>Coherent Detection of DSB-SC Waves:-</u>

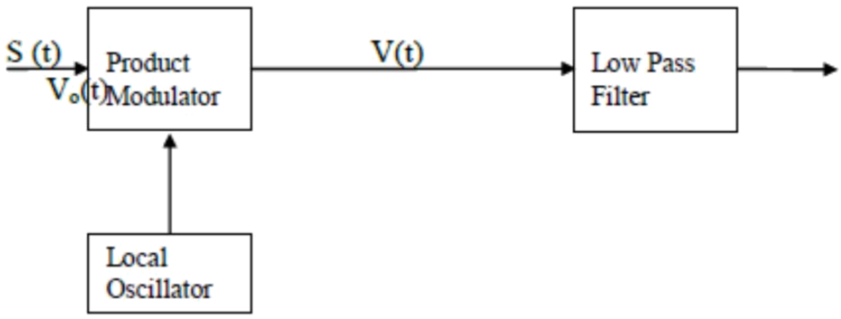


Fig. Coherent detection of DSBSC waves.

The base band signal m (t) can be recovered from a DSB-SC wave s (t) by multiplying s(t) with a locally generated sinusoidal signal and then low pass filtering the product. It is assumed that local oscillator signal is coherent or synchronized, in both frequency and phase ,with the carrier signal c(t) used in the product modulator to generate s(t).this method of demodulation is know as coherent detection or synchronous demodulation. The product modulator produces the product of both input signal and local oscillator and the output of the product modulator v (t) is given by

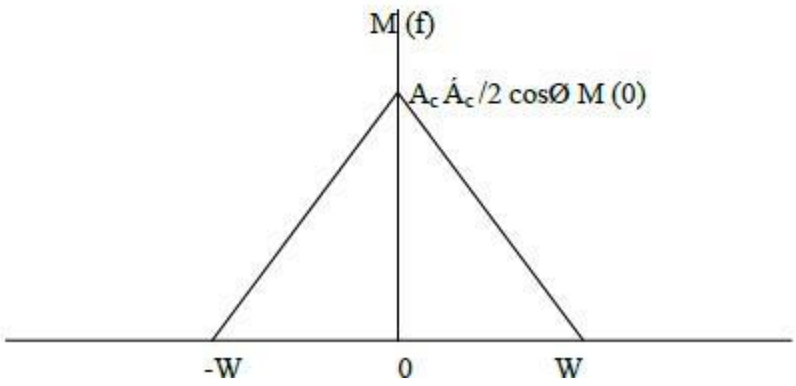
$v(t) = \acute{A}c\cos(2\pi fct+Ø)\, s(t)$

$v(t) = \acute{A}c\cos(2\pi fct+Ø)\, A c\cos 2\pi fct\, m(t)$
$v(t) = Ac\, \acute{A}c/2 \cos(2\pi fct+Ø)\, m(t) + Ac\, \acute{A}c/2 \cos Ø\, m(t)$

The high frequency can be eliminated by passing this output voltage to the Low Pass Filter. Now the Output Voltage at the Low pass Filter is given by $v0(t) = Ac\, \acute{A}c/2 \cos Ø\, m(t)$

The Fourier transform of $v_o(t)$ is

$V_O(f) = A_c \acute{A}_c/2 \cos Ø\, M(f)$

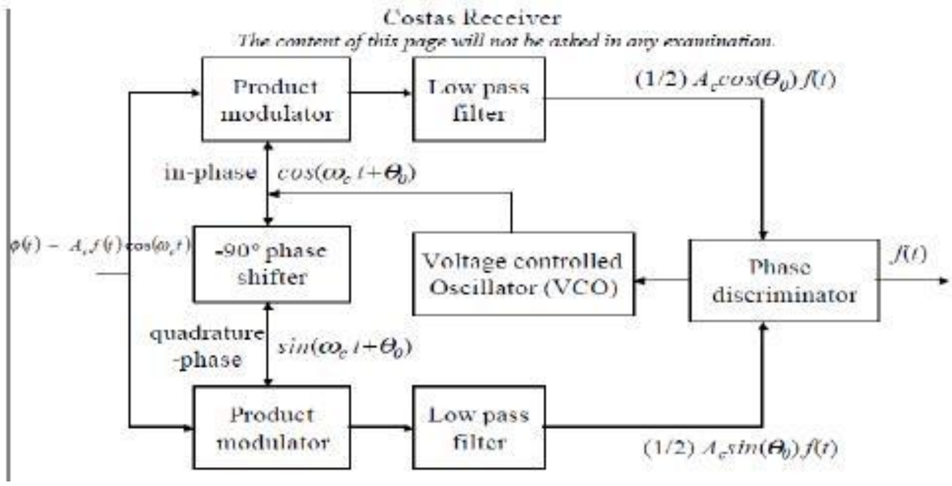


The demodulated signal is proportional to the message signal m (t) when the phase error is constant. The Amplitude of this Demodulated signal is maximum when Ø=0, and it is minimum (zero) when Ø=±π/2 the zero demodulated signal, which occurs for Ø=±π/2 represents quadrature null effect of the coherent detector.

Disadvantages.
First, with conventional AM, carrier power constitutes two thirds or more of the total transmitted power .This is a major drawback because the carrier contains no information.

Conventional AM systems utilize twice as much bandwidth as needed with SSB systems. With SSB transmission, the information contained in the USB is identical the information contained in the LSB. Therefore transmitting both sidebands is redundant.
Consequently, Conventional AM is both power and bandwidth inefficient, which are the two predominant considerations when designing modern electronic communication systems.

COSTA'S Loop



Objective:
Noise is ever present and limits the performance of virtually every system. The presence of noise degrades the performance of the Analog and digital communication systems. This chapter deals with how noise affects different Analog modulation techniques. After studying this chapter the should be familiar with the following

Various performance measures of communication systems SNR calculations for DSB-SC, SSB-SC, Conventional AM, FM (threshold effect, threshold extension, pre-emphasis and

deemphasis)and PM.
Key points:

- The presence of noise degrades the performance of the Analog and digital communication systems
- The extent to which the noise affects the performance of communication system is measured by the output signal-to-noise power ratio or the probability of error.
- The SNR is used to measure the performance of the Analog communication systems, whereas the probability of error is used as a performance measure of digital communication systems
- figure of merit = γ = SNRo/SNRi

- The loss or mutilation of the message at low predetection SNR is called as the threshold effect. The threshold occurs when SNRi is about 10dB or less.

➢ Output SNR :

So= output signal power

Si = input signal power

fM = base band signal frequency range

The input noise is white with spectral density = η/2 You havepreviously studiedideal analog communication systems. Our aim here is to compare the performance of different analog modulation schemes in the presence of noise.The performance will be measured in terms of the signal-to-noise ratio (SNR) at the output of the receiverNote 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.

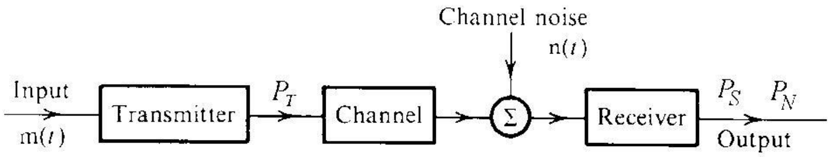


Figure 3.1: Model of an analog communication system. [Lathi, Fig. 12.1]

A model of a typical communication system is shown in Fig. 3.1, where we assume that a modulated signal with power *PT* is transmitted over a channel with additive noise. At the output of the receiver the signal and noise powers are *PS* and *PN* respectively, and hence, the output SNR is SNR*o* = *P S/PN* . This ratio can be increased as much as desired simply by increasing the transmitted power. However, in practice the maximum value of *PT* 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 difference mod- ulation 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 communi- cation 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 is shown in Fig. 3.2(a), where *m*(*t*) is the band-limited mes- sage signal, and *W* is its bandwidth.

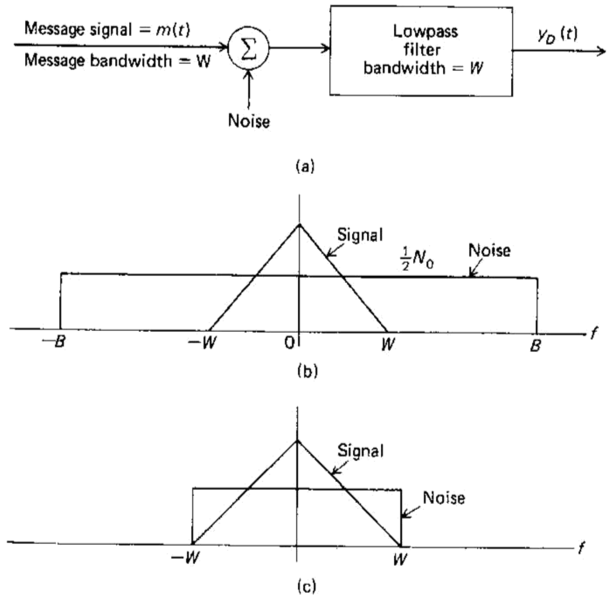


Figure 3.2: Baseband communication system: (a) model, (b) signal spectra at filter input, and (c) signal spectra at filter output. [Ziemer & Tranter, Fig. 6.1]

An example signal PSD is shown in Fig. 3.2(b). 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 $No/2$ over some bandwidth $B > W$, as shown in Fig. 3.2(b). For a basic baseband system, the transmitted power is identical to the message power, i.e., $PT = P$.

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. 3.2(c), and the average noise power is given by

Thus, the SNR at the receiver output is
SNR baseband $=NoW$

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.

Noise in DSB-SC

The *predetection* signal (i.e., just before the multiplier in Fig. 3.3) is

$$x(t) = s(t) + n(t) \qquad (3.8)$$

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 $No/2$ over a bandwidth of $2W$ (centered on the carrier frequency), as shown in Fig. 2.5. Hence, using the bandpass representation

(2.25) the predetection signal is

$$x(t) = [Am(t) + n_c(t)] \cos(2\pi f_c t) - n_s(t) \sin(2\pi f_c t) \qquad (3.9)$$

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

$$y(t) = 2 \cos(2\pi fct)x(t)$$

$$= Am(t)[1 + \cos(4\pi fc\, t)] + n(t)[1 + \cos(4\pi f_c t)]$$

$$-n_s(t)\sin(4\pi f_c t) \qquad (3.10)$$

where we have used (3.6) and

$$2 \cos x \sin x = \sin(2x) \qquad (3.11)$$

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

$$\tilde{y}(t) = Am(t) + n_c(t) \qquad (3.12)$$

The signal power at the receiver output is

$$PS = E\{A^2m^2(t)\} = A^2E\{m^2(t)\} = A^2P \qquad (3.13)$$

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

$$\mathrm{SNR}_o = \frac{A^2 P}{2 N_o W}$$

To make a fair comparison with a baseband system, we need to calculate the transmitted power
We conclude that a DSB-SC 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

After low-pass filtering this becomes

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

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

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

UNIT III

SSB MODULATION AND VESTIGIAL SIDE BAND MODULATION

Generation of SSB waves:

- Filter method
- Phase shift method
- Third method (Weaver's method)

Demodulation of SSB waves:

- Coherent detection: it assumes perfect synchronization between the local carrier and that used in the transmitter both in frequency and phase.

Effects of frequency and phase errors in synchronous detection-DSB-SC, SSB-SC:

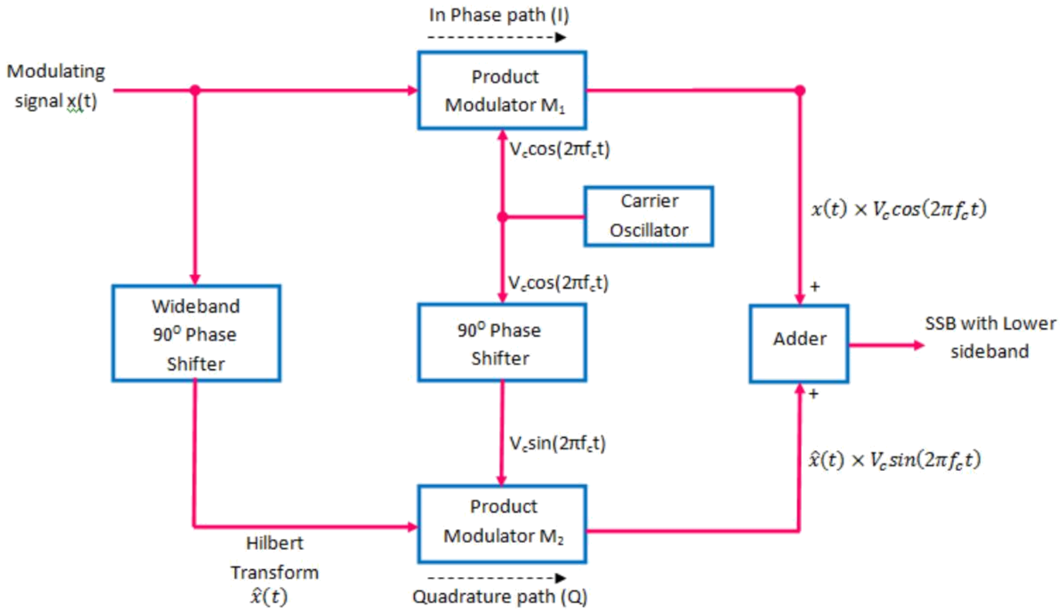
Any error in the frequency or the phase of the local oscillator signal in the receiver, with respect to the carrier wave, gives rise to distortion in the demodulated signal.The type of distortion caused by frequency error in the demodulation process is unique to SSB modulation systems. In order to reduce the effect of frequency error distortion in telephone systems, we have to limit the frequency error to 2-5 Hz.The error in the phase of the local oscillator signal results in phase distortion, where each frequency component of the message signal undergoes a constant phase shift at the demodulator output. This phase distortion is usually not serious with voice communications because the human ear is relatively insensitive to phase distortion; the presence of phase distortion gives rise to a Donald Duck voice effect.

Phase Shift Method for the SSB Generation

Fig. 1 shows the block diagram for the phase shift method of SSB generation .

This system is used for the suppression of lower sideband .

This system uses two balanced modulators M_1 and M_2 and two 90^o phase shifting networks as shown in fig. 1 .



Working Operation

The message signal x(t) is applied to the product modulator M_1 and through a 90^o phase shifter to the product modulator M_2 .

Hence, we get the Hilbert transform

$\hat{x}(t)$

at the output of the wideband 90^o phase shifter .

The output of carrier oscillator is applied as it is to modulator M_1 whereas it s passed through a 90^o phase shifter and applied to the modulator M_2 .

$$Output\ of\ M_1 = x(t) \times V_c cos(2\pi f_c t)$$

$$and\quad Output\ of\ M_2 = \hat{x}(t) \times V_c sin(2\pi f_c t)$$

The outputs of M_1 and M_2 are applied to an adder .

Generation of VSB Modulated wave:

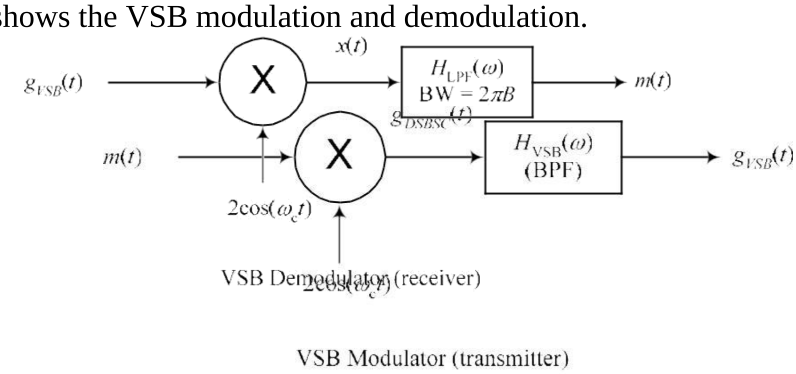
To generate a VSB modulated wave, we pass a DSBSC modulated wave through a sideband-shaping filter.

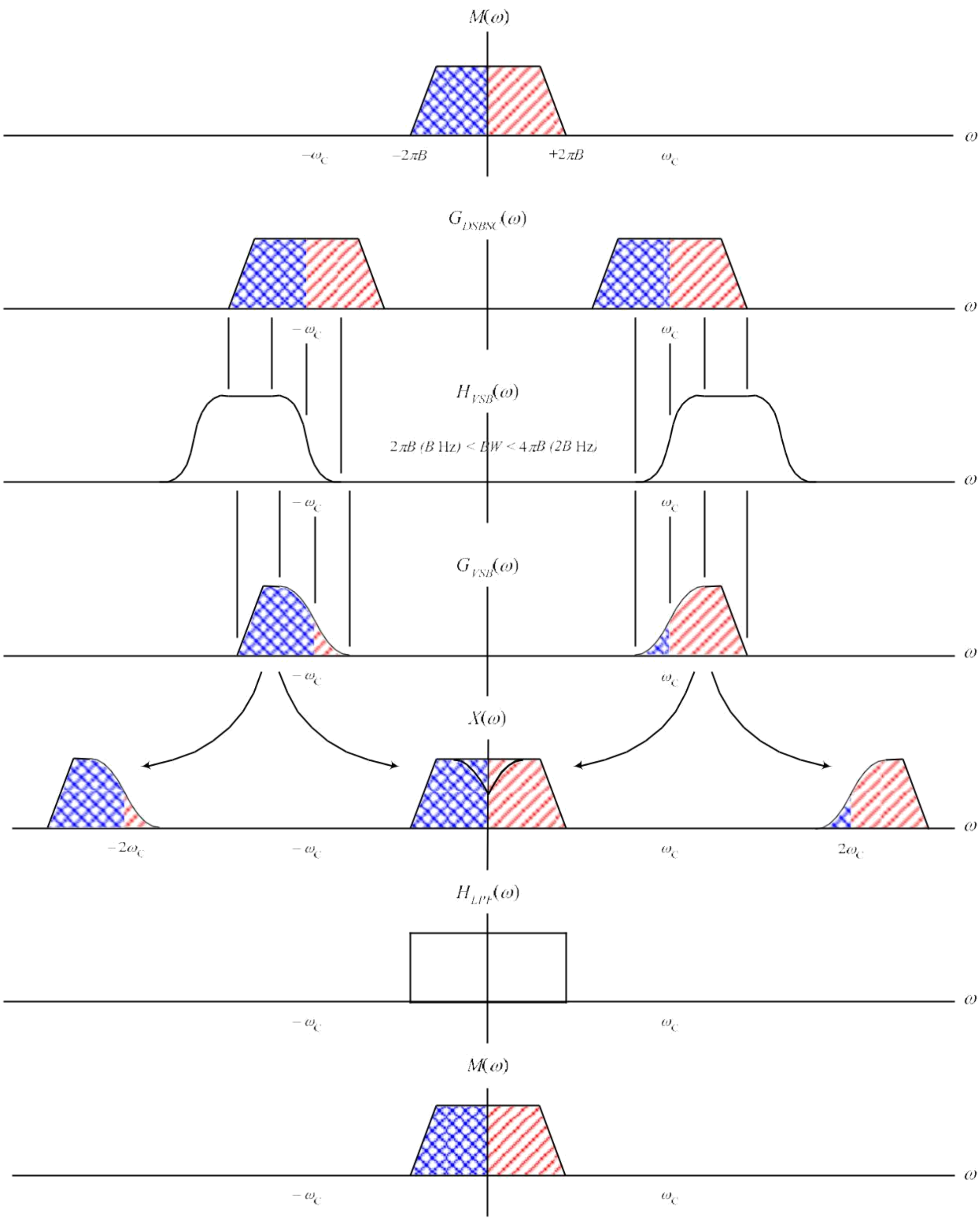
In commercial AM radio broadcast systems standard AM is used in preference to DSBSC or SSB modulation.Suppressed carrier modulation systems require the minimum transmitter power and minimum transmission bandwidth.Suppressed carrier systems are well suited for point –to-point communications.SSB is the preferred method of modulation for long-distance transmission of voice signals over metallic circuits, because it permits longer spacing between the repeaters.VSB modulation requires a transmission bandwidth that isintermediate between that required for SSB or DSBSC.VSB modulation technique is used in TV transmissionDSBSC, SSB, and VSB are examples of linear modulation. In Commercial TV broadcasting, the VSB occupies a width of about 1.25MHz, or about one-quarter of a full sideband.

Vestigial Side Band Modulation

As mentioned last lecture, the two methods for generating SSB modulated signals suffer some problems. The selective–filtering method requires that the two side bands of the DSBSC modulated signal which will be filtered are separated by a guard band that allows the bandpass filters that are used to have non–zero transition band (so it allows for real filters). An ideal Hilbert transform for the phase–shifting method is impossible to build, so only an approximation of that can be used. Therefore, the SSB modulation method is hard, if not impossible, build. A compromise between the DSBSC modulation and the SSB modulation is known as Vestigial Side Band (VSB) modulation. This type of modulation is generated using a

similar system as that of the selective–filtering system for SSB modulation. The following block diagram shows the VSB modulation and demodulation.





$M(\omega)$
ω
$-\omega_C$
$-2\pi B$
$+2\pi B$
ω_C
$G_{DSBSC}(\omega)$
$-\omega_C$
ω_C
$H_{VSB}(\omega)$
$2\pi B$ (B Hz) < BW < $4\pi B$ (2B Hz)
$G_{VSB}(\omega)$
$X(\omega)$
$-2\omega_C$
$2\omega_C$
$H_{LPF}(\omega)$
$M(\omega)$

The above example for generating VSB modulated signals assumes that the VSB filter

($HVSB$( )) that the transition band of the VSB filter is symmetric in a way that adding the part that remains in the filtered signal from the undesired side band to the missing part of the desired side band during the process of demodulation produces an undusted signal at baseband. In fact, this condition is not necessary if the LPF in the demodulator can take care of any distortion that happens when adding the different components of the bandpass components at baseband.

To illustrate this, consider a baseband message signal $m(t)$ that has the FT shown in the following figure.

The DSBSC modulated signal from that assuming that the carrier is 2cos(Ct) (the 2 in the carrier is placed there for convenience) is

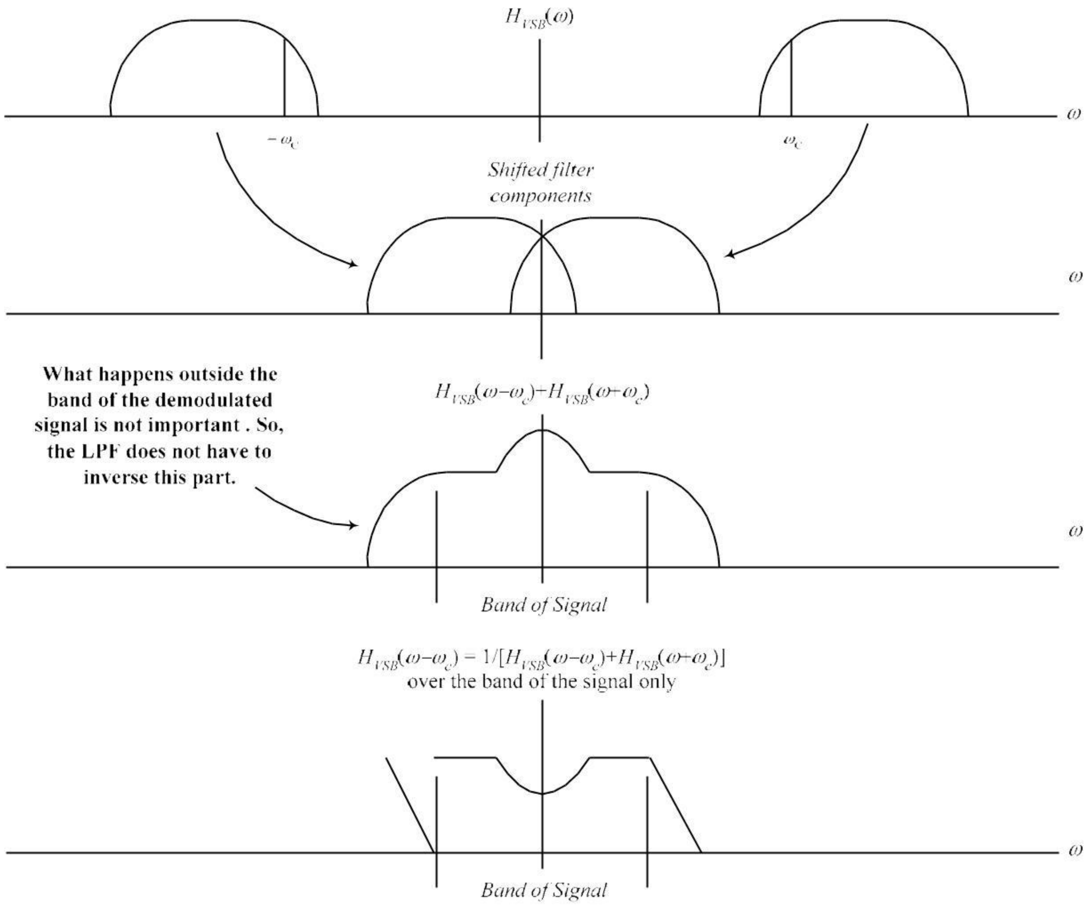
$g_{DSBSC}(t) \qquad m(t)\cos(_{C}t)$

Note that the VSB filter is not an ideal filter with flat transfer function, so it has to appear in the equation defining the VSB signal.

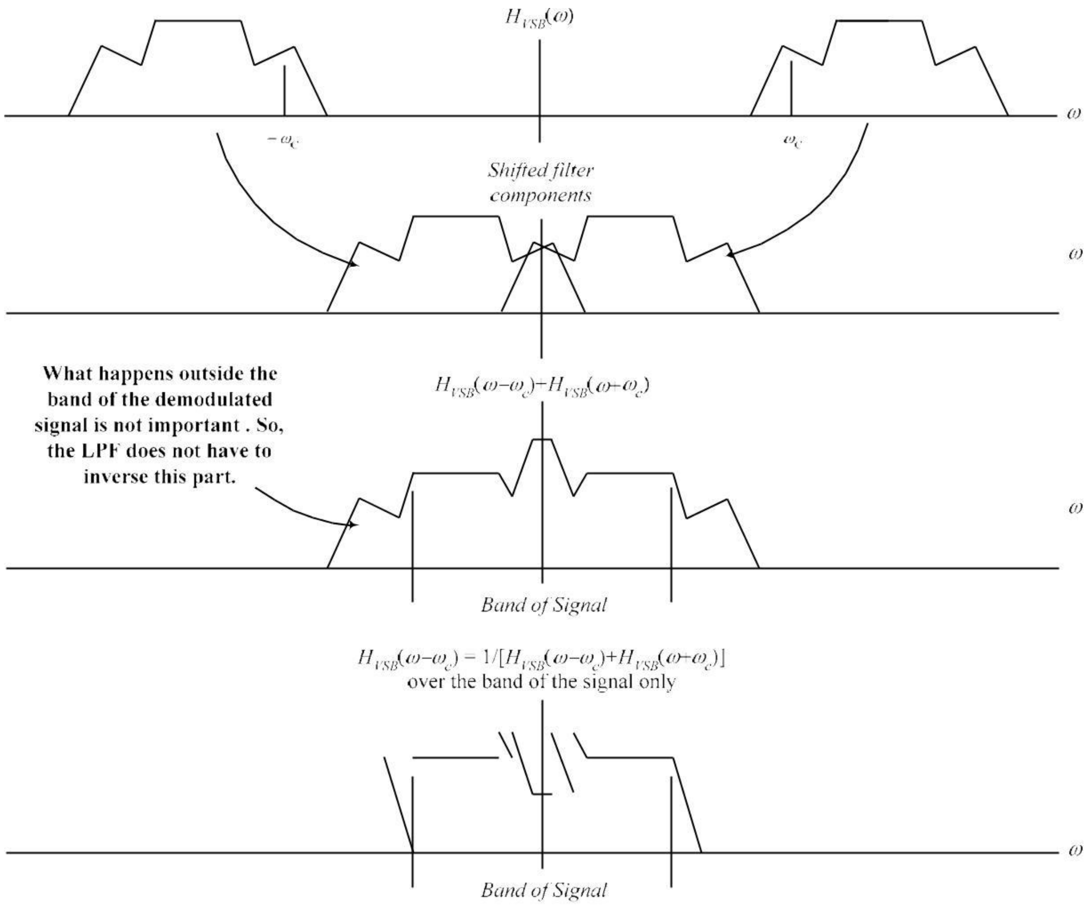
Now, let us demodulate this VSB signal using the demodulator shown above but use a non–ideal filter HLPF( ) (the carrier here is also multiplied by 2 just for convenience)

— —

So, this filter must be a LPF that has a transfer function around 0 frequency that is related to the VSB filter as given above. To illustrate this relationship, consider the following VSB BPF example.



Another example follows.



Multiplexing:

It is a technique whereby a number of independent signals can be combined into a composite signal suitable for transmission over a common channel. There are two types of multiplexing techniques

4. Frequency division multiplexing (FDM) : The technique of separating the signals in frequency is called as FDM

5. Time division multiplexing: The technique of separating the signals in time is called as TDM.

UNIT-IV

ANGLE MODULATION

Objective:

It is another method of modulating a sinusoidal carrier wave, namely, angle Modulation in which either the phase or frequency of the carrier wave is varied according to the message signal. After studying this the student should be familiar with the following

- Definition of Angle Modulation
- Types Angle Modulation- FM & PM
- Relation between PM & FM
- Phase and Frequency deviation
- Spectrum of FM signals for sinusoidal modulation – sideband
- features, power content.
- Narrow band and Wide band FM
- BW considerations-Spectrum of a constant BW FM, Carson's Rule
- Phasor Diagrams for FM signals
- Multiple frequency modulations – Linearity.
- FM with square wave modulation.

This deals with the generation of Frequency modulated wave and detection of original message signal from the Frequency modulated wave. After studying this chapter student should be familiar with the following Generation of FM Signals

i. Direct FM – Parameter Variation Method (Implementation using varactor, FET)

ii. Indirect FM – Armstrong system, Frequency Multiplication.

FM demodulators- Slope detection, Balanced Slope Detection, Phase Discriminator (Foster Seely), Ratio Detector.

Angle modulation: there are two types of Angle modulation techniques namely

2. Phase modulation
3. Frequency modulation

Phase modulation (PM) is that of angle modulation in which the angular argument θ (t) is varied linearly with the message signal m(t), as shown by

$\theta(t) = 2\pi f_c t + k_p m(t)$
where $2\pi f_c t$ represents the angle of the unmodulated carrier
k_p represents the phase sensitivity of the modulator(radians/volt). The phase modulated wave
$s(t)=A_c\cos[2\pi f_c t+k_p m(t)]$

Frequency modulation (FM) is that of angle modulation in which the instantaneous frequency fi(t) is varied linearly with the message signal m(t), as shown by

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

Where fc represents the frequency of the unmodulated carrier

kf represents the frequency sensitivity of the modulator(Hz/volt) The frequency modulated wave
s(t)=Accos[2πfct+2πkf otm(t)dt]

FM wave can be generated by first integrating m(t) and then using the result as the input to a phase modulator

PM wave can be generated by first differentiating m(t) and then using the result as the input to a frequency modulator. Frequency modulation is a Non-linear modulation process. Single tone FM:

Consider $m(t)=A_m\cos(2\pi f_m t)$

The instantaneous frequency of the resulting FM wave
$f_i(t) = f_c + k_f A_m\cos(2\pi f_m t)$

= fc+ f cos(2πfmt)

wheref = kf Am is called as frequency deviation

θ (t) =2π fi(t)dt

=2πfct+ f/fm sin(2πfmt)

= 2πfct+β sin(2πfmt)

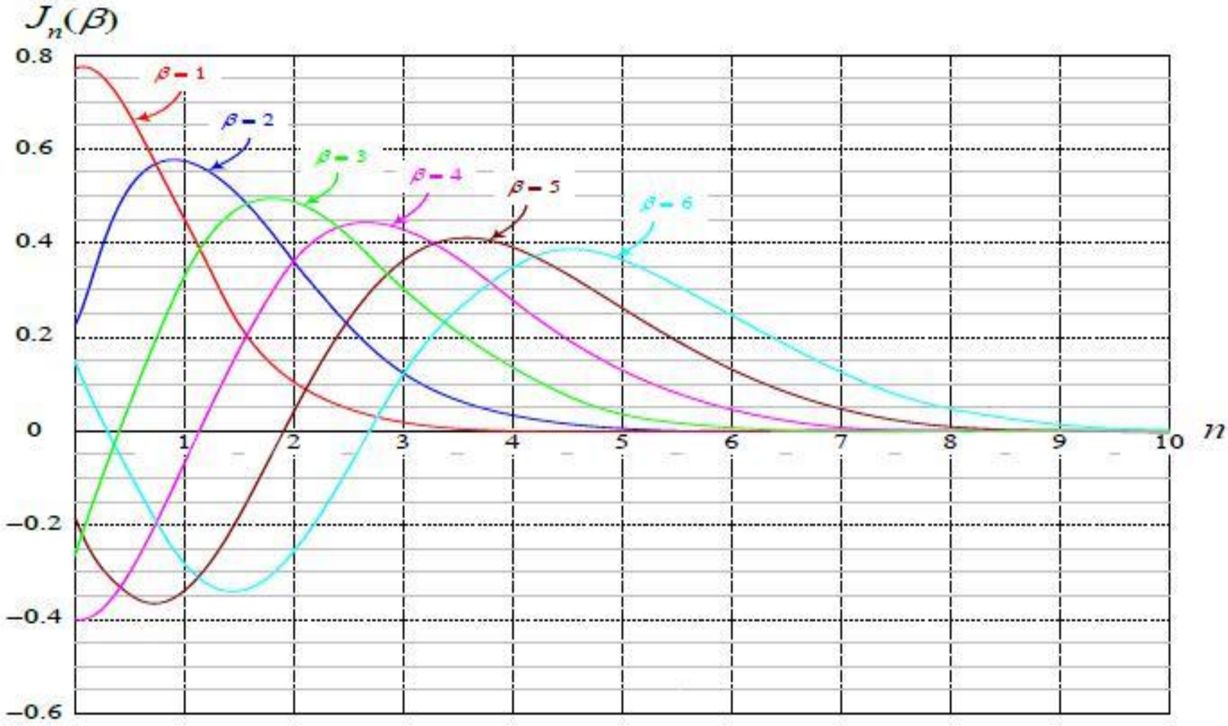
Where β= f/fm= modulation index of the FM wave

When β<<1 radian then it is called as narrowband FM consisting essentially of a carrier, an upper side-frequency component, and a lower side-frequency component.

When β>>1 radian then it is called as wideband FM which contains a carrier and an infinite number of side-frequency components located symmetrically around the carrier.

The envelope of an FM wave is constant, so that the average power of such a wave dissipated in a 1-ohm resistor is also constant.

Plotting the Bessel function of the first kind *J*n( ) for different orders *n* and different values of is shown below.



$J_n(\beta)$	β=1	β=2	β=3	β=4	β=5	β=6
n=0	0.7652	0.2239	-0.2601	-0.3971	-0.1776	0.1506
n=1	0.4401	0.5767	0.3391	-0.0660	-0.3276	-0.2767
n=2	0.1149	0.3528	0.4861	0.3641	0.0466	-0.2429
n=3	0.0196	0.1289	0.3091	0.4302	0.3648	0.1148
n=4	0.0025	0.0340	0.1320	0.2811	0.3912	0.3576
n=5	0.0002	0.0070	0.0430	0.1321	0.2611	0.3621
n=6	0.0000	0.0012	0.0114	0.0491	0.1310	0.2458
n=7	0.0000	0.0002	0.0025	0.0152	0.0534	0.1296
n=8	0.0000	0.0000	0.0005	0.0040	0.0184	0.0565
n=9	0.0000	0.0000	0.0001	0.0009	0.0055	0.0212
n=10	0.0000	0.0000	0.0000	0.0002	0.0015	0.0070

Frequency Spectrum of FM:

The FM modulated signal in the time domain is given by:

$S(t)=A_c\sum^{\infty}_{n=-\infty} J_n(\beta)\mathrm{Cos}[(\acute{\omega}_c+n\ \acute{\omega}_m)t]$

From this equation it can be seen that the frequency spectrum of an FM waveform with a sinusoidal modulating signal is a discrete frequency spectrum made up of components spaced at frequencies of c ± *n* m
.

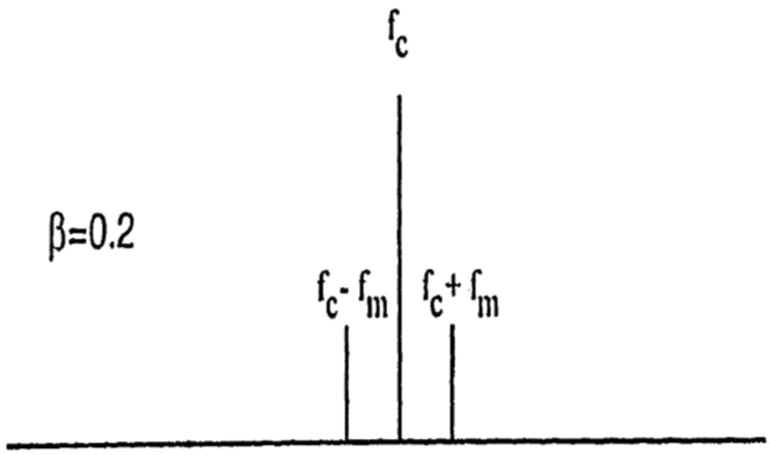
By analogy with AM modulation, these frequency components are called sidebands.

We can see that the expression for s(t) is an infinite series. Therefore the frequency spectrum of an FM signal has an infinite number of sidebands.

The amplitudes of the carrier and sidebands of an FM signal are given by the corresponding Bessel functions, which are themselves functions of the modulation index.

Spectra of an FM Signal with Sinusoidal Modulation

The following spectra show the effect of modulation index, , on the bandwidth of an FM signal, and the relative amplitudes of the carrier and sidebands



Carson's Rule: Bandwidth is twice the sum of the maximum

frequency deviation and the modulating frequency.

BW=2(f+ fm)
The nominal BW 2 f = 2 βfm

Key points:

Generation of FM waves:

1. Indirect FM: This method was first proposed by Armstrong. In this method the modulating wave is first used to produce a narrow-band FM wave, and frequency multiplication is next used to increase the frequency deviation to the desired level.
2. Direct FM: In this method the carrier frequency is directly varied in accordance with the incoming message signal.

Detection of FM waves:
To perform frequency demodulation we require 2-port device that produces an output signal with amplitude directly proportional to the instantaneous frequency of a FM wave used as the input signal.

Fm detectors

1)Slope detector

2)Balanced Slope detector(Travis detector, Triple-tuned-discriminator)

3)Phase discriminator (Foster seeley discriminator or center-tuned discriminator)
4)Ratio detector PLL demodulator and Quadrature detector

- The Slope detector, Balanced Slope detector, Foster seeley discriminator, and Ratio detector are one forms of tuned –circuit frequency discriminators.

- Tuned circuit discriminators convert FM to AM and then demodulate the AM envelope with conventional peak detectors.
- Disadvantages of slope detector – poor linearity, difficulty in tuning, and lack of provisions for limiting.
- A Balanced slope detector is simply two single ended slope detectors connected in parallel and fed 180o out of phase.
- Advantage of Foster-seeley discriminator: output voltage-vs-frequency deviation curve is more linear than that of a slope detector, it is easier to tune.
- Disadvantage of Foster-seeley discriminator: a separate limiter circuit must precede it.
- Advantage of Ratio detector over Foster seeley discriminator: it is relatively immune to amplitude variations in its input signal.

FM DETECTORS:

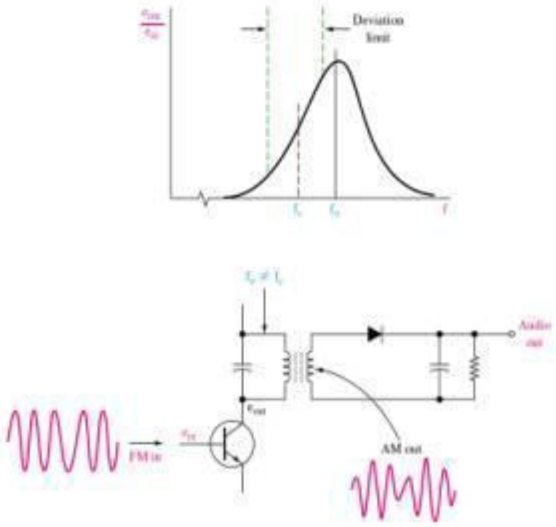
FM detectors convert the frequency variations of the carrier back into a replica of the original modulating signal. There are 5 basic types of FM detectors:

1. Slope detector

2. Foster-Seely Discriminator
3. Ratio Detector
4. Quadrature Detector
5. Phase-Locked Loop (PLL) detector

1. SLOPE DETECTOR

The slope detector is the simplest type of FM detector. A schematic diagram of a slope detector appears below:



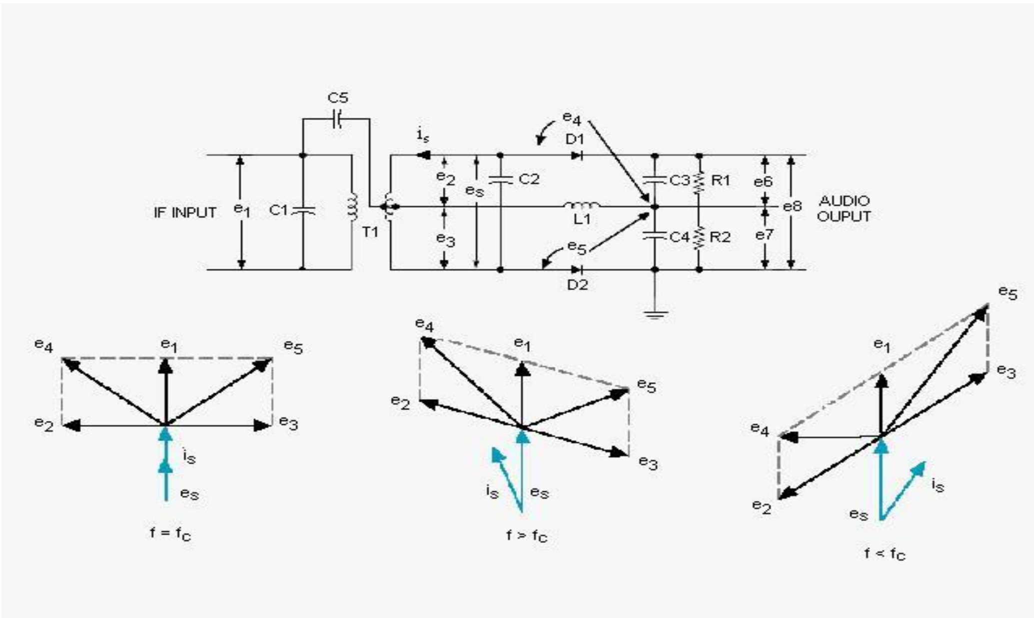
The operation of the slope detector is very simple. The output network of an amplifier is tuned to a frequency that is slightly more than the carrier frequency + peak deviation. As the input signal varies in frequency, the output signal across the LC network will vary in amplitude because of the band pass properties of the tank circuit. The output of this amplifier is AM, which can be detected using a diode detector.

The circuit shown in the diagram above looks very similar to the last IF amplifier and detector of an AM receiver, and it is possible to receive NBFM on an AM receiver by detuning the last IF transformer. If this transformer is tuned to a frequency of approximately 1 KHz above the IF frequency, the last IF amplifier will convert NBFM to AM.
In spite of its simplicity, the slope detector is rarely used because it has poor linearity. To see why this is so, it is necessary to look at the expression for the voltage across the primary of the tuned transformer in the sloped detector

The voltage across the transformer's primary winding is related to the squareof the frequency. Since the frequency deviation of the FM signal is directlyproportional to the modulating signal's amplitude, the output of the slopedetector will be distorted. If the bandwidth of the FM signal is small, it ispossible to approximate the response of the slope detector by a linear function,and a slope detector could be used to demodulate an NBFM signal

2. FOSTER-SEELY DISCRIMINATOR

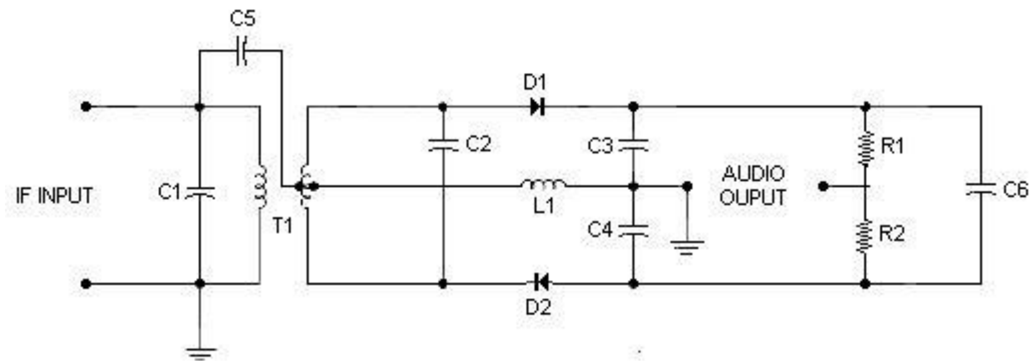
The Foster-Seely Discriminator is a widely used FM detector. The detector consists of a special center-tapped IF transformer feeding two diodes. The schematic looks very much like a full wave DC rectifier circuit. Because the input transformer is tuned to the IF frequency, the output of the discriminator is zero when there is no deviation of the carrier; both halves of the center tapped transformer are balanced. As the FM signal swings in frequency above and below the carrier frequency, the balance between the two halves of the center-tapped secondary are destroyed and there is an output voltage proportional to the frequency deviation.



The discriminator has excellent linearity and is a good detector for WFM and NBFM signals. Its major drawback is that it also responds to AM signals. A good limiter must precede a discriminator to prevent AM noise from appearing in the output.

2.RATIO DETECTOR

The ratio detector is a variant of the discriminator. The circuit is similar to the discriminator, but in a ratio detector, the diodes conduct in opposite directions. Also, the output is not taken across the diodes, but between the sum of the diode voltages and the center tap. The output across the diodes is connected to a large capacitor, which eliminates AM noise in the ratio detector output. The operation of the ratio detector is very similar to the discriminator, but the output is only 50% of the output of a discriminator for the same input signal.



1. SSB-SC:

So/Si =1/4
No= ηfM/4

SNRo= Si/ ηfM
2. DSB-SC:
So/Si =1/2
No= ηfM/2
SNRo= Si/ ηfM
3. DSB-FC:
SNRo= {m2/(2+m2)}Si/ ηfM

- Figure of merit of FM:

γFM = 3/2β2

- Figure of merit of AM & FM :

γFM/ γAM = 9/2β2 = 9/2 (BFM/BAM)2

- The noise power spectral density at the output of the demodulator in PM is flat within the message bandwidth whereas for FM the noise power spectrum has a parabolic shape.

- The modulator filter which emphasizes high frequencies is called the pre-emphasis filter(HPF) and the demodulator filter which is the inverse of the modulator filter is called the de-emphasis filter(LPF).

UNIT-V

RECEIVERS AND SAMPLING THEORM

Introduction

This unit centers around basic principles of the super heterodyne receiver. In The article, we will discuss the reasons for the use of the super heterodyne and various topics which concern its design, such as the choice of intermediate frequency, the use of its RF stage, oscillator tracking, band spread tuning and frequency synthesis. Most of the information is standard text book material, but put together as an introductory article, it can provide somewhere to start if you are contemplating building a receiver, or if you are considering examining specifications with an objective to select a receiver for purchase.

TRF Receiver

Early valve radio receivers were of the Tuned Radio Frequency (TRF) type consisting of one or a number of tuned radio frequency stages with individual tuned circuits which provided the selectivity to separate one received signal from the others. A typical receiver copied from a 1929 issue of "The Listener In" is shown in Figure 1. Tuned circuits are separated by the radio frequency (RF) amplifier stages and the last tuned circuit feeds the AM detector stage. This receiver belongs to an era before the introduction of the screen grid valve and it is interesting to observe the grid-plate capacity neutralisation applied to the triode RF amplifiers to maintain amplifier stability. In these early receivers, the individual tuning capacitors were attached to separate tuning dials, as shown in Figure 2, and each of these dials had to be reset each time a different station was selected. Designs evolved for receivers with only one tuning dial, achieved by various methods of mechanical ganging the tuning capacitors, including the ganged multiple tuning capacitor with a common rotor shaft as used today.

The bandwidth of a tuned circuit of given Q is directly proportional to its operational frequency and hence, as higher and higher operating frequencies came into use, it became more difficult to achieve sufficient selectivity using the TRF

Receiver system.

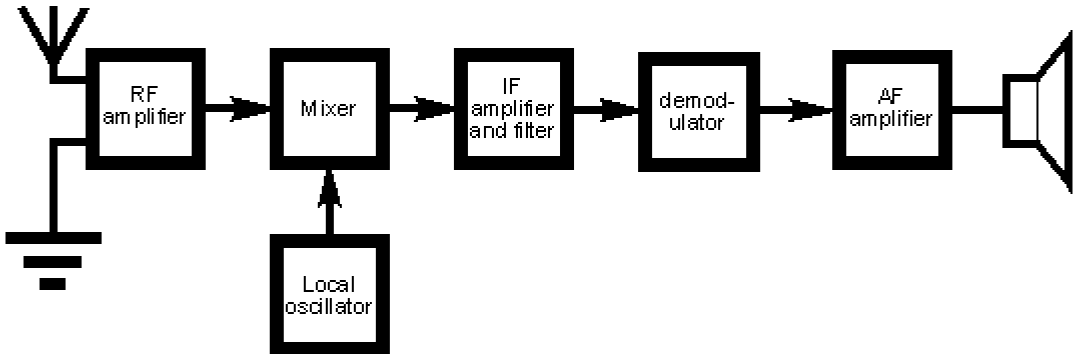


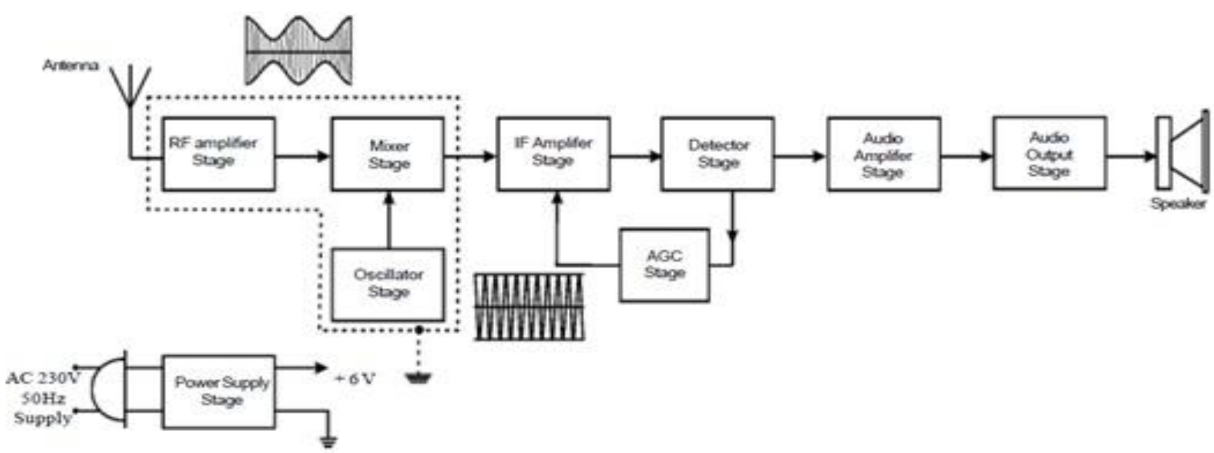
FIGURE: AM RECEIVER

The Super Heterodyne Principle

The super heterodyne (short for supersonic heterodyne) receiver was first evolved by Major Edwin Howard Armstrong, in 1918. It was introduced to the market place in the late 1920s and gradually phased out the TRF receiver during the 1930s.

The principle of operation in the super heterodyne is illustrated by the diagram in Figure 4. In this system, the incoming signal is mixed with a local oscillator to produce sum and difference frequency components. The lower frequency difference component called the intermediate frequency (IF), is separated from the other components by fixed tuned amplifier stages set to the intermediate frequency. The tuning of the local oscillator is mechanically ganged to the tuning of the signal circuit or radio frequency (RF) stages so that the difference intermediate frequency is always the same fixed value. Detection takes place at intermediate frequency instead of at radio frequency as in the TRF receiver

Superheterodyne Receiver.



Use of the fixed lower IF channel gives the following advantages:

1. For a given Q factor in the tuned circuits, the bandwidth is lower making it easier to achieve the required selectivity.
2. At lower frequencies, circuit losses are often lower allowing higher Q factors to be achieved and hence, even greater selectivity and higher gain in the tuned circuits.
3. It is easier to control, or shape, the bandwidth characteristic at one fixed frequency. Filters can be easily designed with a desired band pass characteristic and slope characteristic, an impossible task for circuits which tune over a range of frequencies.
4. Since the receiver selectivity and most of the receiver pre-detection gain, are both controlled by the fixed IF stages, the selectivity and gain of the super heterodyne receiver are more consistent over its tuning range than in the TRF receiver.

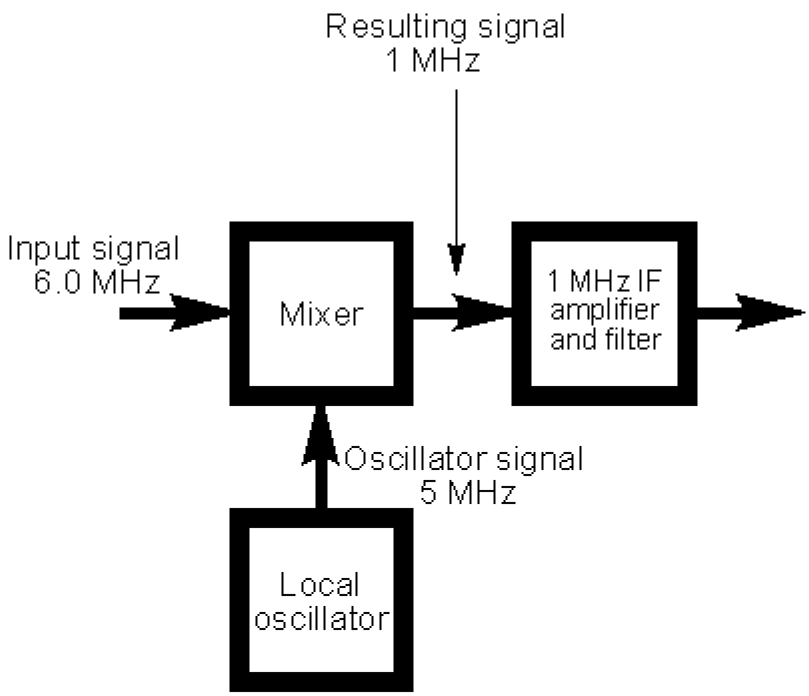


Figure : An illustration of how image frequency provides a second mixing product.

Second Channel or Image frequency

One problem, which has to be contended within the super heterodyne receiver, is its ability to pick up a second or imago frequency removed from the signal frequency by a value equal to twice the intermediate frequency.
To illustrate the point, refer Figure 5. In this example, we have a signal frequency of 1 MHz which mix to produce an IFof 455kHz. A second or image signal, with a frequency equal to 1 MHz plus (2 x 455) kHz or 1.910 MHz, can also mix with the 1.455 MHz to produce the 455 kHz.

Reception of an image signal is obviously undesirable and a function of the RF tuned

circuits (ahead of the mixer), is to provide sufficient selectivity to reduce the image sensitivity of the receiver to tolerable levels.

Choice of intermediate frequency

Choosing a suitable intermediate frequency is a matter of compromise. The lower the IF used, the easier it is to achieve a narrow bandwidth to obtain good selectivity in the receiver and the greater the IF stage gain. On the other hand, the higher the IF, the further removed is the image frequency from the signal frequency and hence the better the image rejection. The choice of IF is also affected by the selectivity of the RF end of the receiver. If the receiver has a number of RF stages, it is better able to reject an image signal close to the signal frequency and hence a lower IF channel can be tolerated.

Another factor to be considered is the maximum operating frequency the receiver. Assuming Q to be reasonably constant, bandwidth of a tuned circuit is directly proportional to its resonant frequency and hence, the receiver has its widest RF bandwidth and poorest image rejection at the highest frequency end of its tuning range.

A number of further factors influence the choice of the intermediate frequency:

1. The frequency should be free from radio interference. Standard intermediate frequencies have been established and these are kept dear of signal channel allocation. If possible, one of these standard frequencies should be used.

2. An intermediate frequency which is close to some part of the tuning range of the receiver is avoided as this leads to instability when the receiver is tuned near thefrequency of the IF channel.

3. Ideally, low order harmonics of the intermediate frequency (particularly second and third order) should not fall within the tuning range of the receiver. This requirement cannot always be achieved resulting in possible heterodyne whistles at certain spots within the tuning range.

4. Sometimes, quite a high intermediate frequency is chosen because the channel must pass very wide band signals such as those modulated by 5 MHz video used in television. In this case the wide bandwidth circuits are difficult to achieve unless quite high frequencies are used.

5. For reasons outlined previously, the intermediate frequency is normally lower than the RF or signal frequency. However, there we some applications, such as in tuning the Low Frequency (LF) band, where this situation could be reversed. In this case, there are difficulties in making the local oscillator track with the signal circuits.

Some modern continuous coverage HF receivers make use of the Wadley Loop or a synthesised VFO to achieve a stable first oscillator source and these have a first intermediate frequency above the highest signal frequency. The reasons for this will be discussed later.

<u>Standard intermediate frequencies</u>

Various Intermediate frequencies have been standardised over the years. In the early days of the superheterodyne, 175 kHz was used for broadcast receivers in the USA and Australia. These receivers were notorious for their heterodyne whistles caused by images of broadcast stations other than the one tuned. The 175 kHz IF was soon overtaken by a 465 kHz allocation which gave better image response. Another compromise of 262kHz between 175 and 465 was also used to a lesser extent. The 465 kHz was eventually changed to 455 kHz, still in use today.

In Europe, long wave broadcasting took place within the band of 150 to 350 kHz and a more suitable IF of 110 kHz was utilised for this band.

The IF of 455 kHz is standard for broadcast receivers including many communication receivers. Generally speaking, it leads to poor image response when used above 10 MHz. The widely used World War 2 Kingsley AR7 receiver used an IF of 455 kHz but it also utillised two RF stages to achieve improved RF selectivity and better image response. One commonly used IF for shortwave receivers is 1.600 MHz and this gives a much improved image response for the HF spectrum.

Amateur band SSB HF transceivers have commonly used 9 MHz as a receiver intermediate frequency in common with its use as a transmitter intermediate frequency. This frequency is a little high for ordinary tuned circuits to achieve the narrow bandwidth needed in speech communication; however, the bandwidth in the amateur transceivers is controlled by specially designed ceramic crystal filter networks in the IF channel.

Some recent amateur transceivers use intermediate frequencies slightly below 9 MHz. A frequency of 8.830 MHz can be found in various Kenwood transceivers and a frequency of 8.987.5 MHz in some Yaesu transceivers. This change could possibly be to avoid the second harmonic of the IF falling too near the edge of the more recently allocated 18 MHz WARC band. (The edge of the band is 18.068 MHz).

General coverage receivers using the Wadley Loop, or a synthesised band set VFO, commonly use first IF channels in the region of 40 to 50 MHz

An IF standard for VHF FM broadcast receivers is 10.7 MHz In this case, the FM deviation used is 75 kHz and audio range is 15 kHz. The higher IF is very suitable as the wide bandwidth is easily obtained with good image rejection. A less common IF is 4.300 MHz believed to have been used in receivers tuning the lower end of the VHF spectrum.

As explained earlier, a very high intermediate frequency is necessary to achieve the wide bandwidth needed for television and the standard in Australia is the frequency segment of 30.500 to 30.6.000 MHz

Multiple Conversion Super Heterodyne Receiver

In receivers tuning the upper HF and the VHF bands, two (or even more) IF channels are commonly used with two (or more) stages of frequency conversion. The lowest frequency IF channel provides the selectivity or bandwidth control that is needed and the highest frequency IF channel is used to achieve good Image rejection. A typical system used in two meter FM amateur transceivers is shown in Figure 6. In this system, IF channels of 10.7 MHz and 455 kHz are used with double conversion. The requirement Is different to that of the wideband FM broadcasting system as frequency deviation is only 5 kHz with an audio frequency spectrum limited to below 2.5 kHz. Channel spacing is 25 kHz and bandwidth is usually limited to less than 15 kHz so that the narrower bandwidth 455 kHz IF channel is suitable. Some modern HF SSB transceivers use a very high frequency IF channel such as 50 MHz. Combined with this, a last IF channel of 455 KHz is used to provide selectivity and bandwidth control. Where there is such a large difference between the first and last intermediate frequency, three stages of conversion and a middle frequency IF channel are needed. This is necessary to prevent on image problem initiating in the 50 MHz IF channel due to insufficient selectivity in that channel. For satisfactory operation, the writer suggests a rule of thumb that the frequency ratio between the RF channel and the first IF channel, or between subsequent IF channels, should not exceed a value of 10.

The RF Amplifier

A good receiver has at least one tuned RF amplifier stage ahead of the first mixer. As discussed earlier, one function of the RF stage is to reduce the image frequency level into the mixer. The RF stage also carries out a number of other useful functions:

1. The noise figure of a receiver is essentially determined by the noise generated in the first stage connected to the aerial system. Mixer stages are inherently more noisy than straight amplifiers and a function of the RF amplifier is to raise the signal level into the mixer so that the signal to noise ratio is determined by the RF amplifier characteristics rather than those of the mixer.

2. There Is generally an optimum signal Input level for mixer stages. If the signal level is increased beyond this optimum point, the levels of inter modulation products steeply increase and these products can cause undesirable effects in the receiver performance. If the signal level is too low, the signal to noise rate will be poor. A function of the RF amplifier is to regulate the signal level into the mixer to maintain a more constant, near optimum, level. To achieve this

regulation, the gain of the RF stage is controlled by an automatic gain control system, or a manual gain control system, or both.

3. Because of its non-linear characteristic, the mixer is more prone to cross modulation

from a strong signal on a different frequency than is the RF
amplifier. The RF tuned circuits, ahead of the mixer, help to reduce the level of the unwanted signal into the mixer input and hence reduce the susceptibility of the mixer to cross-modulation.

4. If, by chance, a signal exists at or near the IF, the RF tuned circuits provide attenuation to that signal.

5. The RF stage provides isolation to prevent signals from the local oscillator reaching the aerial and causing interference by being radiated.

Oscillator Tracking

Whilst the local oscillator circuit tunes over a change in frequency equal to that of the RF circuits, the actual frequency is normally higher to produce the IF frequency difference component and hence less tuning capacity change is needed than in the RF tuned circuits. Where a variable tuning gang capacitor has sections of the same capacitance range used for both RF and oscillator tuning, tracking of the oscillator and RF tuned circuits is achieved by capacitive trimming and padding.

Figure shows a local oscillator tuned circuit (L2,C2) ganged to an RF tuned circuit (Ll,Cl) with Cl and C2 on a common rotor shaft. The values of inductance are set so that at the centre of the tuning range, the oscillator circuit tunes to a frequency equal to RF or signal frequency plus intermediate frequency.
A capacitor called a padder, in series with the oscillator tuned circuit, reduces the maximum capacity in that tuning section so that the circuit tracks with the RF section near the low frequency end of the band.

Small trimming capacitors are connected across both the RF and oscillator tuned circuits to adjust the minimum tuning capacity and affect the high frequency end of the band. The oscillator trimmer is preset with a little more capacity than the RF trimmer so that the oscillator circuit tracks with RF trimmer near the high frequency end of the band.

Curve A is the RF tuning range. The solid curve B shows the ideal tuning range required for the oscillator with a constant difference frequency over the whole tuning range. Curve C shows what would happen if no padding or trimming were applied. Dotted curve B shows the correction applied by padding and trimming. Precise tracking is achieved at three points in the tuning range with a tolerable error between these points.

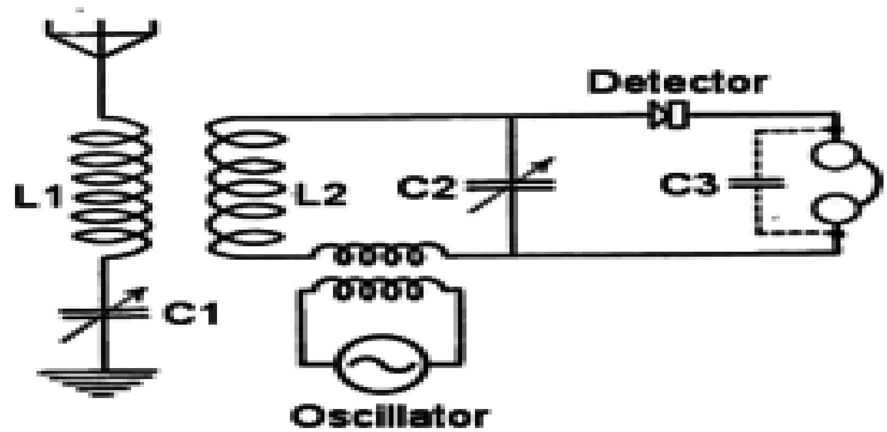


Figure : Tracking Circuit

A capacitor called a padder, in series with the oscillator tuned circuit, reduces the maximum capacity in that tuning section so that the circuit tracks with the RF section near the low frequency end of the band.

Small trimming capacitors are connected across both the RF and oscillator tuned circuits to adjust the minimum tuning capacity and affect the high

frequency end of the band. The oscillator trimmer is preset with a little more capacity than the RF trimmer so that the oscillator circuit tracks with RF trimmer near the high frequency end of the band.

Curve A is the RF tuning range. The solid curve B shows the ideal tuning range required for the oscillator with a constant difference frequency over the whole tuning range. Curve C shows what would happen if no padding or trimming were applied. Dotted curve B shows the correction applied by padding and trimming. Precise tracking is achieved at three points in the tuning range with a tolerable error between these points.

Where more than one band is tuned, not only are separate inductors required for each band, but also separate trimming and padding capacitors, as the degree of capacitance change correction is different for each band.

The need for a padding capacitor can be eliminated one band by using a tuning gang capacitor with a smaller number of plates in the oscillator section than in the RF sections. If tuning more than one band, the correct choice of capacitance for the oscillator section will not be the same for all bands and padding will still be required on other bands.

Alignment of the tuned circuits can be achieved by providing adjustable trimmers and padders. In these days of adjustable magnetic cores in the inductors, the padding capacitor is likely to be fixed with the lower frequency end of the band essentially set by the adjustable cores.

OSCILLATOR STABILITY

The higher the input frequency of a receiver, the higher is the first local oscillator frequency and the greater is the need for oscillator stability. A given percentage frequency drift at higher frequencies amounts to a larger percentage drift in IF at the detector. Good stability is particularly important in a single sideband receiver as a small change in signal frequency is very noticeable as a change in the speech quality, more so than would be noticeable in AM or FM systems.

Frequency stability in an oscillator can be improved by care in the way it is designed and built. Some good notes on how to build a stable variable frequency oscillator were prepared by Draw Diamond VK3XU, and published in Amateur Radio, January 1 1998.

One way to stabilize a receiver tunable oscillator is to use an automatic frequency control (AFC) system. To do this, a frequency discriminator can be operated from the last IF stage and its output fed back via a low pass filter (or long time constant circuit) to a frequency sensitive element in the oscillator. Many of today's receivers and transceivers also make use of phase locked loop techniques to achieve frequency control.

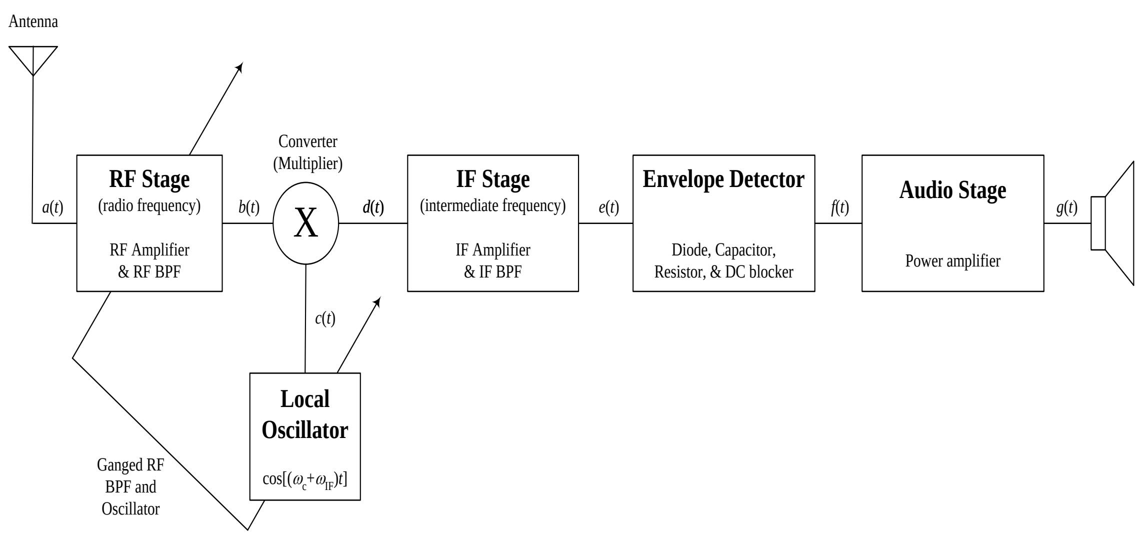
Where there are several stages of frequency conversion and the front end is tuned, the following oscillator stages, associated with later stage conversion, are usually fixed in frequency and can be made stable by quartz crystal control. In this case, receiver frequency stability is set by the first oscillator stability.

One arrangement, which can give better stability, is to crystal lock the first oscillator stage but tune the first IF stage and second oscillator stage as shown in Figure. In this case, the RF tuned circuits are sufficiently broadband to cover a limited tuning range (such as an amateur band) but selective enough to attenuate the image frequency and other possible unwanted signals outside the tuning range. This is the method used when a converter Is added to the front end of a HF receiver to tune say the two meter band.

The RF circuits in the converter are fixed, the converter oscillator is crystal locked and the HF receiver RF and first oscillator circuits become the tunable first IF stage and second tunable oscillator, respectively. Since the HF receiver tunable oscillator is working at a lower frequency than the first oscillator in the converter, the whole system is inherently more stable than if the converter oscillator were tuned. As stated earlier, the system is restricted to a limited tuning range and this leads to a discussion on band spread tuning and other systems incorporating such ideas as the Wadley Loop.

Superheterodyne AM Radio Receiver

Since the inception of the AM radio, it spread widely due to its ease of use and more importantly, it low cost. The low cost of most AM radios sold in the market is due to the use of the full amplitude modulation, which is extremely inefficient in terms of power as we have seen previously. The use of full AM permits the use of the simple and cheap envelope detector in the AM radio demodulator. In fact, the AM demodulator available in the market is slightly more complicated than a simple envelope detector. The block diagram below shows the construction of a typical AM receiver and the plots below show the signals in frequency–domain at the different parts of the radio.



Description of the AM Superheterodyne Radio Receiver

<u>Signal $a(t)$ at the output of the Antenna:</u> The antenna of the AM radio receiver receives the whole band of interest. So it receives signals ranging in frequency from around 530 kHz to 1650 kHz as shown by $a(t)$ in the figure. Each channel in this band occupies around 10 kHz of bandwidth and the different channels have center frequencies of 540, 550, 560, …. , 1640 kHz.

<u>Signal $b(t)$ at the output of the RF (Radio Frequency) Stage:</u> The signal at the output of the antenna is extremely week in terms of amplitude. The radio cannot process this signal as it is, so it must be amplified. The amplification does not amplify the whole spectrum of the AM band and it does not amplify a single channel, but a range of channels is amplified around the desired channel that we would like to receive. The reason for using a BPF in this stage although the desired channel is not completely separated from adjacent channels is to avoid possible interference of some channels later in the demodulation process if the whole band was allowed to pass (assume the absence of this BPF and try demodulating the two channels at the two edges of the AM band, you will see that one of these cannot be demodulated). Also, the reason for not extracting the desired channel alone is that extracting only that channel represents a big challenge since the filter that would have to extract it must have a constant bandwidth of 10 kHz and a center frequency in the range of 530 kHz to 1650 kHz. Such a filter is extremely difficult

to design since it has a high Q–factor (center frequency/bandwidth) let alone the fact that its center frequency is variable. Therefore, the process of extracting only one channel is left for the following stages where a filter with constant center frequency may be used. Note in the block diagram above that the center frequency of the BPF in the RF stage is controlled by a variable capacitor with a value that is modified using a knob in the radio (the tuning knob).

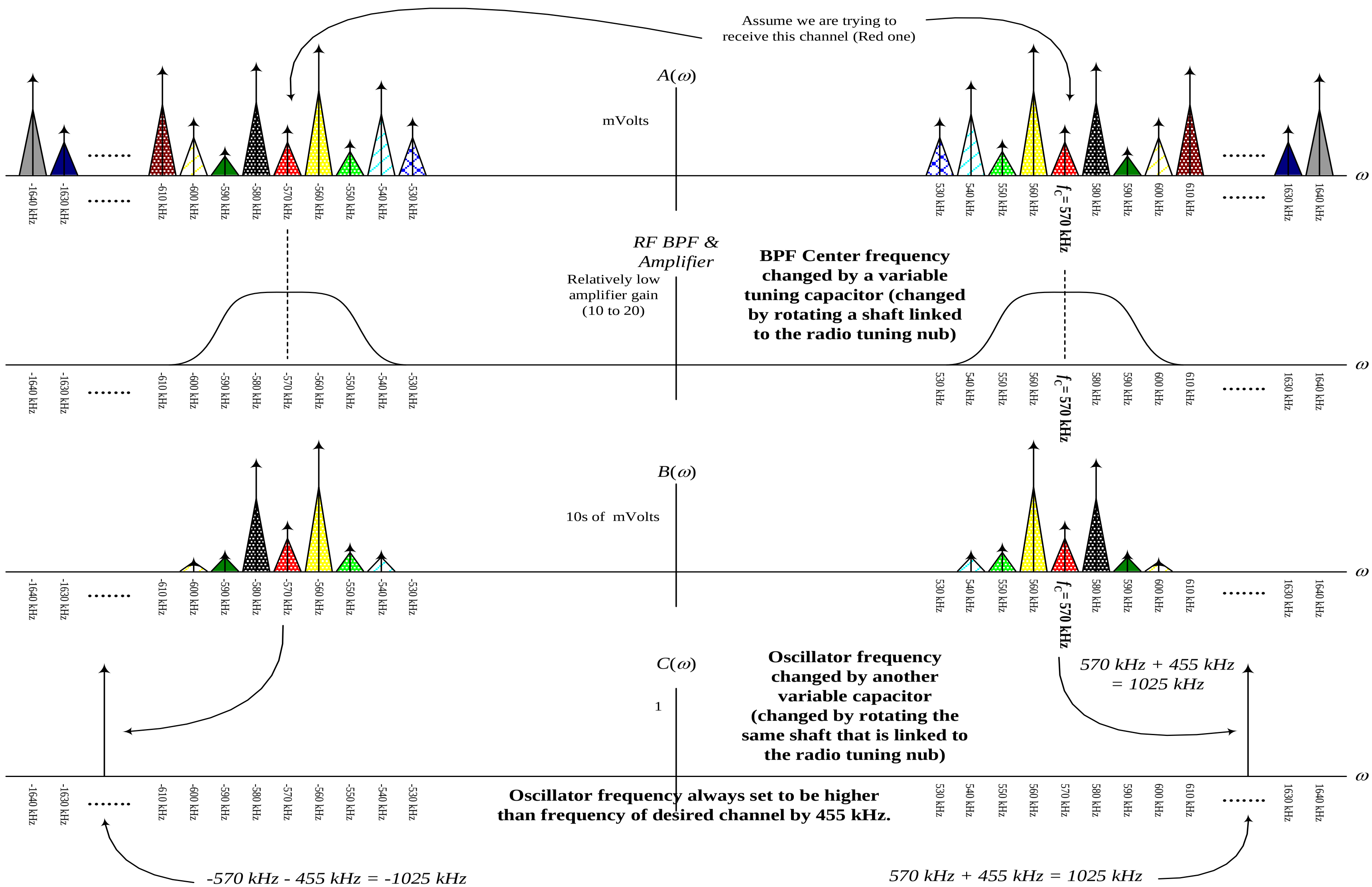
Signal $c(t)$ at the output of the Local Oscillator: This is simply a sinusoid with a variable frequency that is a function of the carrier frequency of the desired channel. The purpose of multiplying the signal $b(t)$ by this sinusoid is to shift the center frequency of $b(t)$ to a constant frequency that is called IF (intermediate frequency). Therefore, assuming that the desired channel (the channel you would like to listen to) has a frequency of f_{RF} and the IF frequency that we would like to move that channel to is f_{IF}, one choice for the frequency of the local oscillator is to be $f_{RF} + f_{IF}$. The frequency of the local oscillator is modified in the radio using a variable capacitor that is also controlled using the same tuning knob as the variable capacitor that controls the center frequency of the BPF filter in the RF stage. The process of controlling the values of two elements such as two variable capacitors using the same knob by placing them on the same shaft is known as GANGING.

Signal $d(t)$ at the output of the Multiplier (Usually called frequency converter or mixer): The signal here should contain the desired channel at the constant frequency f_{IF} regardless of the original frequency of the desired channel. Remember that this signal does not only contain the desired channel but it contains also several adjacent channels and also contain images of these channels at the much higher frequency $2f_{RF} + f_{IF}$ (since multiplying by a cosine shifts the frequency of the signal to the left and to the right). When this type of radios was first invented, a standard was set for the value for the IF frequency to be 455 kHz. There is nothing special about this value. A range of other values can be used.

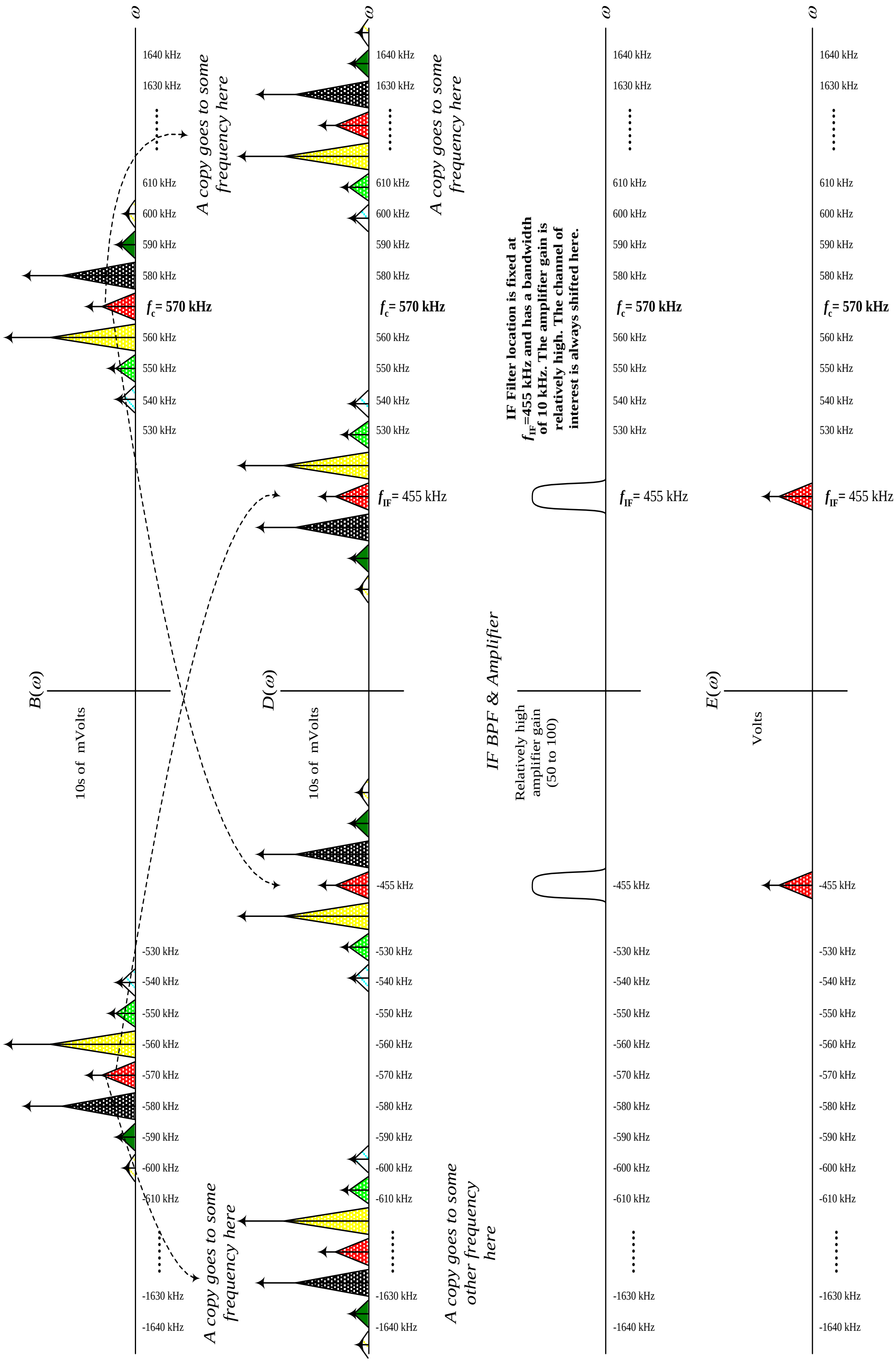
Signal $e(t)$ at the output of the IF Stage: Now that the desired channel is located at the IF frequency, a relatively simple to create BP filter with BW of 10 kHz and center frequency of f_{IF} can be used to extract only the desired channel and reject all adjacent channels. This filter has a constant Q factor of about 455/10 = 45.5 (which is not that difficult to create), but more importantly has a constant center frequency. Therefore the output of this stage is the desired channel alone located at the IF frequency. This stage also contains a filter that amplifies the signal to a level that is sufficient for an envelope detector to operate on.

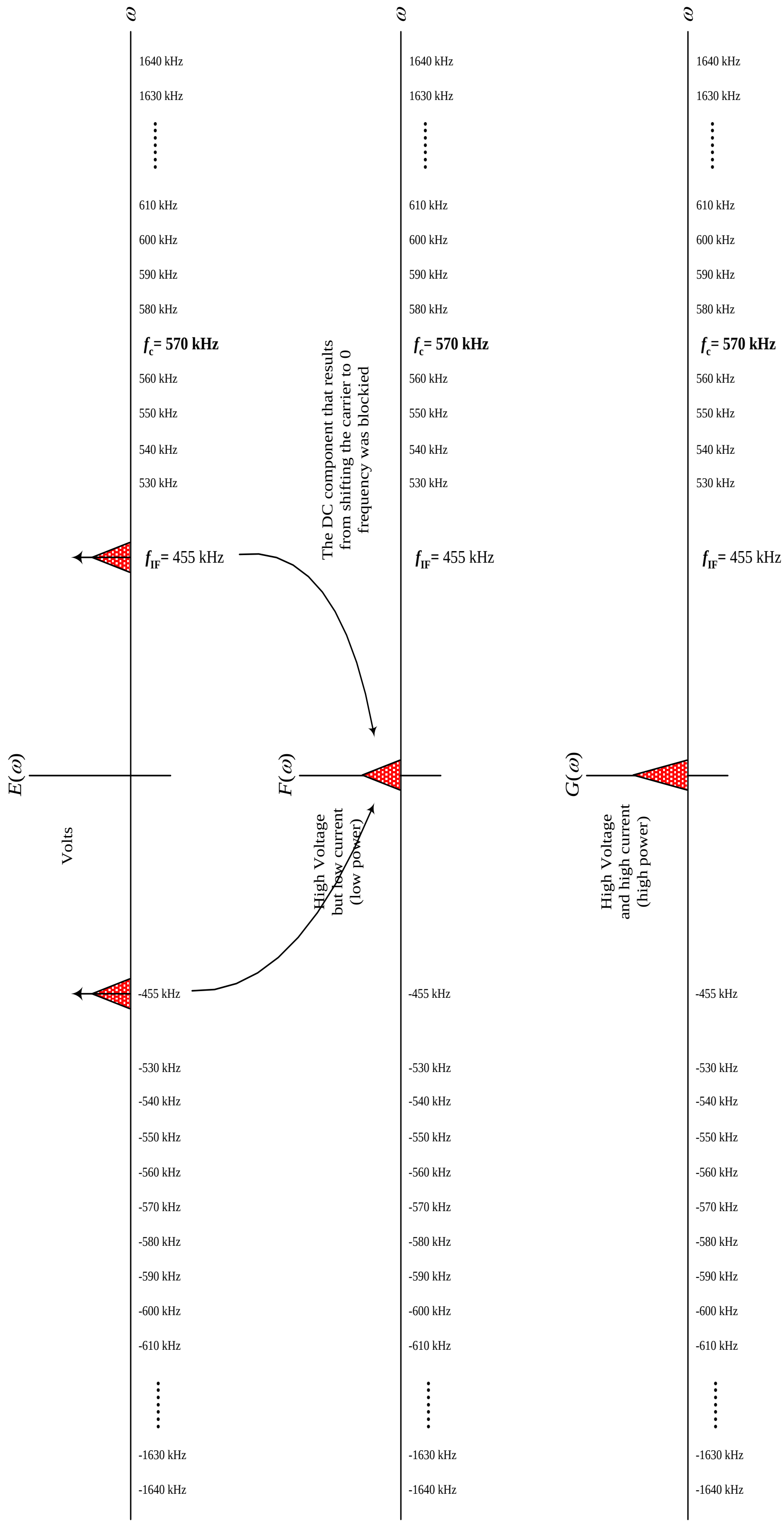
Signal $f(t)$ at the output of the Envelope Detector: The signal above is input to an envelope detector that extracts the original unmodulated signal from the modulated signal and also rejects any DC that is present in that signal. The output of that stage becomes the original signal with relatively low power.

Signal $g(t)$ at the output of the Audio Stage (Power Amplifier): Since the output of the envelope detector is generally weak and is not sufficient to drive a large speaker, the use of an amplifier that increases the power in the signal is necessary. Therefore, the output of that stage is the original audio signal with relatively high power that can directly be input to a speaker.



Assume we are trying to receive this channel (Red one)
A(ω)
mVolts
RF BPF & Amplifier
Relatively low amplifier gain (10 to 20)
BPF Center frequency changed by a variable tuning capacitor (changed by rotating a shaft linked to the radio tuning nub)
B(ω)
10s of mVolts
C(ω)
1
Oscillator frequency changed by another variable capacitor (changed by rotating the same shaft that is linked to the radio tuning nub)
570 kHz + 455 kHz = 1025 kHz
Oscillator frequency always set to be higher than frequency of desired channel by 455 kHz.
-570 kHz - 455 kHz = -1025 kHz
570 kHz + 455 kHz = 1025 kHz
fC = 570 kHz
-1640 kHz
-1630 kHz
-610 kHz
-600 kHz
-590 kHz
-580 kHz
-570 kHz
-560 kHz
-550 kHz
-540 kHz
-530 kHz
530 kHz
540 kHz
550 kHz
560 kHz
570 kHz
580 kHz
590 kHz
600 kHz
610 kHz
1630 kHz
1640 kHz
ω


B(ω)
10s of mVolts
A copy goes to some frequency here
A copy goes to some frequency here
D(ω)
10s of mVolts
A copy goes to some frequency here
A copy goes to some other frequency here
IF BPF & Amplifier
Relatively high amplifier gain (50 to 100)
IF Filter location is fixed at f_IF=455 kHz and has a bandwidth of 10 kHz. The amplifier gain is relatively high. The channel of interest is always shifted here.
E(ω)
Volts
ω
f_c= 570 kHz
f_IF= 455 kHz
-455 kHz
530 kHz
540 kHz
550 kHz
560 kHz
580 kHz
590 kHz
600 kHz
610 kHz
1630 kHz
1640 kHz
-530 kHz
-540 kHz
-550 kHz
-560 kHz
-570 kHz
-580 kHz
-590 kHz
-600 kHz
-610 kHz
-1630 kHz
-1640 kHz


E(ω)
F(ω)
G(ω)
ω
Volts
1640 kHz
1630 kHz
610 kHz
600 kHz
590 kHz
580 kHz
fc = 570 kHz
560 kHz
550 kHz
540 kHz
530 kHz
fIF = 455 kHz
-455 kHz
-530 kHz
-540 kHz
-550 kHz
-560 kHz
-570 kHz
-580 kHz
-590 kHz
-600 kHz
-610 kHz
-1630 kHz
-1640 kHz
The DC component that results from shifting the carrier to 0 frequency was blockied
High Voltage but low current (low power)
High Voltage and high current (high power)

Section 6: Sampling & Reconstruction

This section is concerned with digital signal processing systems capable of operating on analogue signals which must first be sampled and digitised. The resulting digital signals often need to be converted back to analogue form or "reconstructed". Before starting, we review some facts about analogue signals.

6.1. Some analogue signal theory:

1. Given an analogue signal $x_a(t)$, its analogue Fourier Transform $X_a(j\omega)$ is its spectrum where ω is the frequency in radians per second. $X_a(j\omega)$ is complex but we often concentrate on the modulus $|X_a(j\omega)|$.

2. An analogue unit impulse $\delta(t)$ may be visualised as a very high (infinitely high in theory) very narrow (infinitesimally narrow) rectangular pulse, applied starting at time t=0, with area (height in volts times width in seconds) equal to one volt-second. The area is the impulse strength. We can increase the impulse strength by making the pulse we visualise higher for a given narrowness. The "weighted" impulse then becomes $A\delta(t)$ where A is the new impulse strength. We can also delay the weighted impulse by τ seconds to obtain $A\delta(t-\tau)$. An upward arrow labelled with "A" denotes an impulse of strength A.

6.2 Sampling an analogue signal

Given an analogue signal $x_a(t)$ with Fourier Transform $X_a(j\omega)$, consider what happens when we sample $x_a(t)$ at intervals of T seconds to obtain the discrete time signal $\{x[n]\}$ defined as the sequence:

$$\{ \ldots, x[-1],\ \underline{x[0]},\ x[1],\ x[2],\ x[3], \ldots \}$$

with the underlined sample occurring at time t=0. It follows that $x[1] = x_a(T)$, $x[2] = x_a(2T)$, etc. The sampling rate is 1/T Hz or 2π/T radians/second.

Define a new analogue signal $x_S(t)$ as follows:

$$x_s(t) = \sum_{n=-\infty}^{\infty} x[n]\,\delta(t - nT)$$
$$= \text{sample}_T\{x(t)\}$$

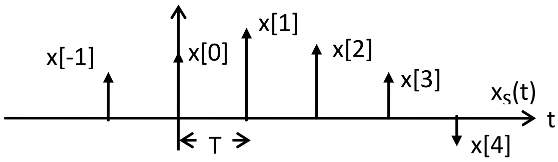


Fig. 1

where $\delta(t)$ denotes an analogue impulse. As illustrated in Figure 1, $x_s(t)$ is a succession of delayed impulses, each impulse being multiplied by a sample value of $\{x[n]\}$. The Fourier transform of $x_s(t)$ is:

$$X_s(j\omega) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x[n]\,\delta(t - nT)\; e^{-j\omega t}\, dt$$

$$= \sum_{n=-\infty}^{\infty} x[n] \int_{-\infty}^{\infty} \delta(t - nT)\; e^{-j\omega t}\, dt$$

$$= \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\omega nT}$$

If we replace Ω by ωT, this expression is ready known to us as the discrete time Fourier transform (DTFT) of $\{x[n]\}$.

6.3. Relative frequency: Remember that Ω is 'relative frequency' in units of "radians per sample". It is related to ordinary frequency, ω in radians/second, as follows:-

$$\Omega = \omega T = \omega/f_s \text{ where } f_s \text{ is the sampling rate in Hertz.}$$

6.4 Discrete time Fourier transform (DTFT) related to Fourier Transform:

The DTFT of $\{x[n]\}$ is therefore identical to the Fourier transform (FT) of the analogue signal $x_S(t)$ with Ω denoting ωT.

6.5. Relating the DTFT of $\{x[n]\}$ to the FT of $x_a(t)$:

What we really want to do is relate the DTFT of $\{x[n]\}$ to the Fourier Transform (FT) of the original analogue signal $x_a(t)$. To achieve this we quote a convenient form of the 'Sampling Theorem':

> Given any signal $x_a(t)$ with Fourier Transform $X_a(j\omega)$, the Fourier Transform of $x_S(t) = \text{sample}_T\{x_a(t)\}$ is $X_S(j\omega) = (1/T)\text{repeat}_{2\pi/T}\{X_a(j\omega)\}$.

- By '$\text{sample}_T\{x_a(t)\}$' we mean a series of impulses at intervals T each weighted by the appropriate value of $x_a(t)$ as seen in fig 1.
- By '$\text{repeat}_{2\pi/T}\{X_a(j\omega)\}$' we mean (loosely speaking) $X_a(j\omega)$ repeated at frequency intervals of $2\pi/T$. This definition will be made a bit more precise later when we consider 'aliasing'.

This theorem states that $X_s(j\omega)$ is equal to the sum of an infinite number of identical copies of $X_a(j\omega)$ each scaled by 1/T and shifted up or down in frequency by a multiple of $2\pi/T$ radians per second, i.e.

$$X_s(j\omega) = \frac{1}{T}X_a(j\omega) + \frac{1}{T}X_a(j(\omega - 2\pi/T)) + \frac{1}{T}X_a(j(\omega + 2\pi/T)) + \cdots\cdots$$

This equation is valid for any analogue signal $x_a(t)$.

6.6: Significance of the Sampling Theorem:

For an analogue signal $x_a(t)$ which is band-limited to a frequency range between $-\pi/T$ and $+\pi/T$ radians/sec ($\pm f_s/2$ Hz) as illustrated in Figure 2, $X_a(j\omega)$ is zero for all values of ω with $|\omega| \geq \pi/T$. It follows that

$$X_s(j\omega) = (1/T)\, X_a(j\omega) \quad \text{for} \quad -\pi/T < \omega < \pi/T$$

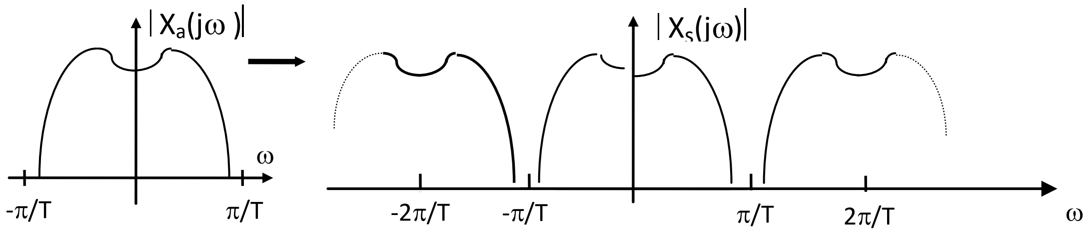


Fig. 2

This is because $X_a(j(\omega - 2\pi/T))$, $X_a(j(\omega + 2\pi/T)$ and $X_a(j\omega)$ do not overlap. Therefore if we take the DTFT of {x[n]} (obtained by sampling $x_a(t)$), set $\omega=\Omega/T$ to obtain $X_S(j\omega)$, and then remove everything outside $\pm\pi/T$ radians/sec and multiply by T, we get back the original spectrum $X_a(j\omega)$ exactly; we have lost nothing in the sampling process. From the spectrum we can get back to $x_a(t)$ by an inverse FT. We can now feel confident when applying DSP to the sequence {x[n]} that it truly represents the original analogue signal without loss of fidelity due to the sampling process. This is a remarkable finding of the "Sampling Theorem".

6.7: Aliasing distortion

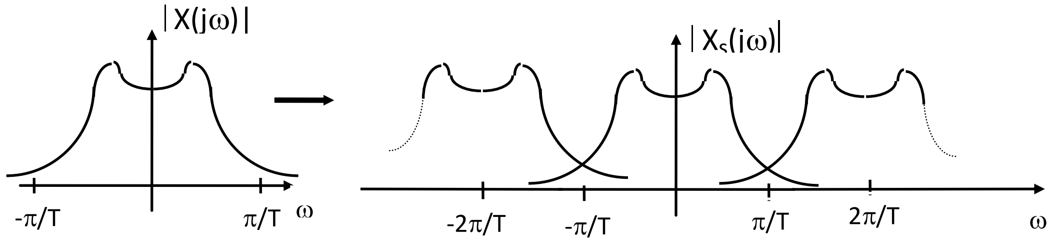


Fig. 3

In Figure 3, where $X_a(j\omega)$ is not band-limited to the frequency range $-\pi/T$ to π/T, overlap occurs between $X_a(j(\omega-2\pi/T))$, $X_a(j\omega)$ and $X_a(j(\omega+2\pi/T))$. Hence if we take $X_s(j\omega)$ to represent $X_a(j\omega)/T$ in this case for

$-\pi/T < \omega < \pi/T$, the representation will not be accurate, and $X_s(j\omega)$ will be a distorted version of $X_a(j\omega)/T$. This type of distortion, due to overlapping in the frequency domain, is referred to as aliasing distortion.

- The precise definition of 'repeat$_{2\pi/T}\{X(j\omega)\}$' is "the sum of an infinite number of identical copies of $X_a(j\omega)$ each scaled by 1/T and shifted up or down in frequency by a multiple of $2\pi/T$ radians per second". It is only when $X_a(j\omega)$ is band-limited between $\pm\pi/T$ that our earlier 'loosely speaking' definition strictly applies. Then there are no 'overlaps' which cause aliasing.

The properties of $X(e^{j\Omega})$, as deduced from those of $X_s(j\omega)$ with $\Omega= \omega T$, are now summarised.

6.8: Properties of DTFT of {x[n]} related to Fourier Transform of $x_a(t)$:

(i) If {x[n]} is obtained by sampling $x_a(t)$ which is bandlimited to $\pm f_s/2$ Hz (i.e. $\pm 2\pi/T$ radians/sec),

at f_s (= 1/T) samples per second then

$X(e^{j\Omega}) = (1/T)\, X_a(j\omega) \quad$ for $\quad -\pi < \Omega \;(=\omega T) < \pi$

Hence $X(e^{j\Omega})$ is closely related to the analogue frequency spectrum of $x_a(t)$ and is referred to

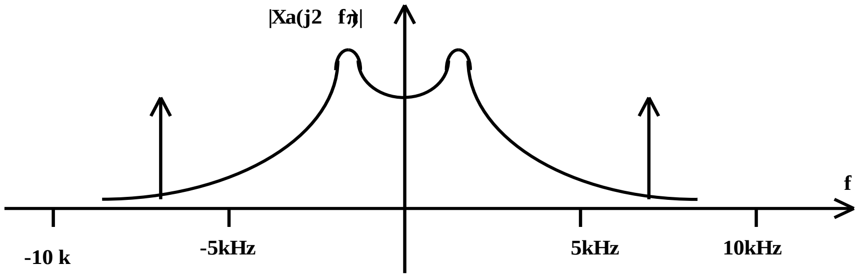
as the "spectrum" of {x[n]}.

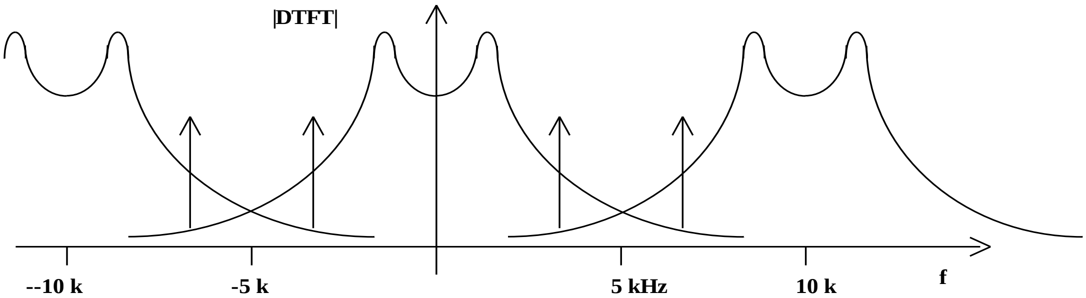
(ii) $X(e^{j\Omega})$ is the Fourier Transform of an analogue signal $x_s(t)$ consisting of a succession of impulses at intervals of $T = 1/f_s$ seconds multiplied by the corresponding elements of {x[n]}.

6.9: Anti-aliasing filter: To avoid aliasing distortion, we have to low-pass filter $x_a(t)$ to band-limit the signal to $\pm f_S/2$ Hz. It then satisfies "Nyquist sampling criterion".

Example: $x_a(t)$ has a strong sinusoidal component at 7 kHz. It is sampled at 10 kHz without an anti-aliasing filter. What happens to the sinusoids?

Solution:





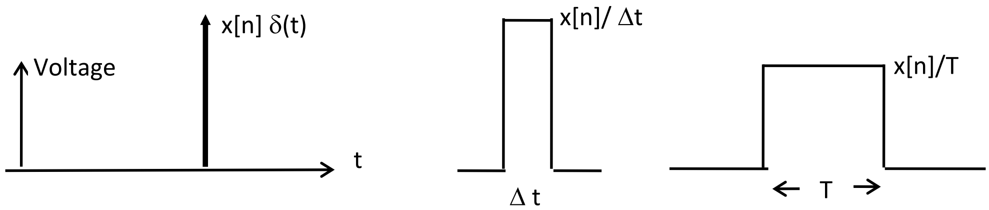
It becomes a 3 kHz (=10 – 7kHz) sine-wave & distorts the signal.

6.10: Reconstruction of $x_a(t)$:

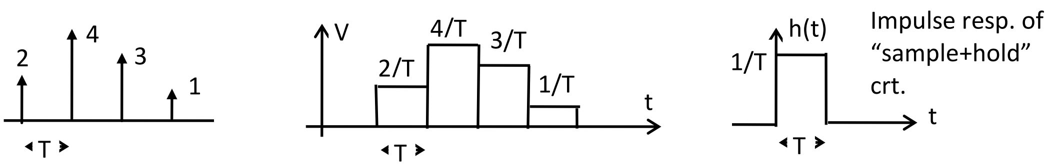
Given a discrete time signal {x[n]}, how can we reconstruct an analogue signal $x_a(t)$, band-limited to $\pm f_s/2$ Hz, whose Fourier transform is identical to the DTFT of {x[n]} for frequencies in the range $-f_s/2$ to $f_s/2$?

Ideal reconstruction: In theory, we must first reconstruct $x_s(t)$ (requires ideal impulses) and then filter using an ideal low-pass filter with cut-off π/T radians/second.

In practice we must use an approximation to $x_s(t)$ where each impulse is approximated by a pulse of finite voltage and non-zero duration:-

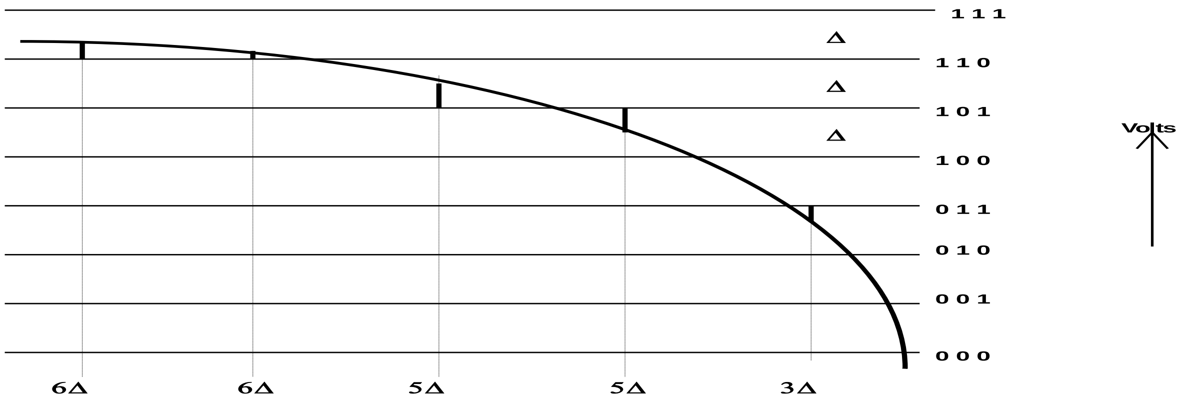


The easiest approach in practice is to use a "sample and hold" (sometimes called “zero order hold”) circuit to produce a voltage proportional to (1/T)x(t) at t = mT, and hold this fixed until the next sample is available at t = (m+1)T. This produces a “staircase” wave-form as illustrated below. The effect of this approximation may be studied by realising that the sample and hold approximation could be produced by passing $x_s(t)$ through a linear circuit, which we can call a sample and hold filter, whose impulse response is as shown below:-



A graph of the gain-response of the sample & hold circuit shows that the gain at ω = 0 is 0 dB, and the gain at π/T is 20 $\log_{10}(2/\pi)$ = -3.92 dB. Hence the reconstruction of $x_a(t)$ using a sample and hold approximation to $x_s(t)$ rather than $x_s(t)$ itself incurs a frequency dependent loss (roll-off) which increases towards about 4 dB as ω increases towards π/T. This is called the 'sample & hold roll-off' effect. In some cases the loss is not too significant and can be disregarded. In other cases a compensation filter may be used to cancel out the loss.

6.11: Quantisation error: The conversion of the sampled voltages of $x_a(t)$ to binary numbers produces a digital signal that can be processed by digital circuits or computers. As a finite number of bits will be available for each sample, an approximation must be made whenever a sampled value falls between two voltages represented by binary numbers or quantisation levels.

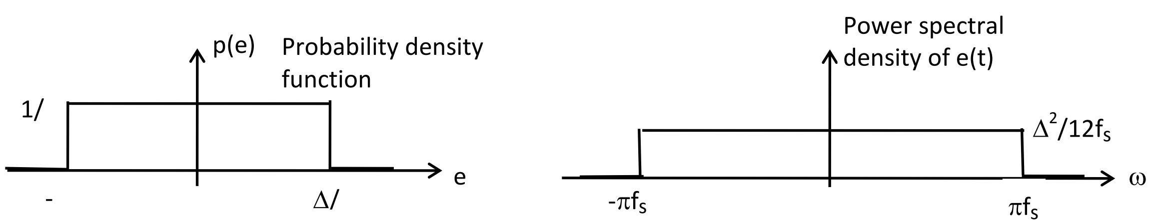


An m bit uniform A/D converter has 2^m quantisation levels, Δ volts apart. Rounding the true samples of {x[n]} to the nearest quantisation level for each sample produces a quantised sequence { $\hat{x}$ [n]} with elements:

$$\hat{x}[n] = x[n] + e[n] \quad \text{for all n.}$$

Normally e[n] lies between -Δ/2 and +Δ/2, except when the amplitude of x[n] is too large for the range of quantisation levels.

Ideal reconstruction (using impulses and an f_s/2 cut-off ideal low-pass filter) from $\hat{x}$ [n] will produce the analogue signal $x_a(t)$ + e(t) instead of $x_a(t)$, where e(t) arises from the quantisation error sequence {e[n]}. Like $x_a(t)$, e(t) is bandlimited to $\pm f_s/2$ by the ideal reconstruction filter. {e[n]} is the sampled version of e(t). Under certain conditions, it is reasonable to assume that if samples of $x_a(t)$ are always rounded to the nearest available quantisation level, the corresponding samples of e(t) will always lie between -Δ/2 and +Δ/2 (where Δ is the difference between successive quantisation levels), and that any voltage in this range is equally likely regardless of $x_a(t)$ or the values of any previous samples of e(t). It follows from this assumption that at each sampling instant t = mT, the value e of e(t) is a random variable with zero mean and uniform probability distribution function. It also follows that the power spectral density of e(t) will show no particular bias to any frequency in the range -f_s/2 to f_s/2 will therefore be flat as shown below:-



In signal processing, the 'power' of an analogue signal is the power that would be dissipated in a 1 Ohm resistor when the signal is applied to it as a voltage. If the signal were converted to sound, the power would tell us how loud the sound would be. It is well known that the power of a sinusoid of amplitude A is $A^2/2$ watts. Also, it may be shown that the power of a random (noise) signal with the probability density function shown above is equal to $\Delta^2/12$ watts. We will assume these two famous results without proof.
A useful way of measuring how badly a signal is affected by quantisation error (noise) is to calculate the following quantity:

$$\text{Signal to quantisation noise ratio (SQNR)} = 10 \log_{10}\left[\frac{\text{signal power}}{\text{quantisation noise power}}\right] \text{ dB.}$$

<u>Example</u>: What is the SQNR if the signal power is (a) twice and (b) 1,000,000 times the quantisation noise power?
<u>Solution:</u> (a) 10 log10(2) = 3 dB. (b) 60 dB.

To make the SQNR as large as possible we must arrange that the signal being digitised is large enough to use all quantisation levels without excessive overflow occurring. This often requires that the input signal is amplified before analogue-to-digital (A/D) conversion.

Consider the analogue-to-digital conversion of a sine-wave whose amplitude has been amplified so that it uses the maximum range of the A/D converter. Let the number of bits of the uniformly quantising A/D converter be m and let the quantisation step size be Δ volts. The range of the A/D converter is from
$-2^{m-1}\Delta$ to $+2^{m-1}\Delta$ volts and therefore the sine-wave amplitude is $2^{m-1}\Delta$ volts.
The power of this sine-wave is $(2^{m-1}\Delta)^2 / 2$ watts and the power of the quantisation noise is $\Delta^2/12$ watts. Hence the SQNR is:-

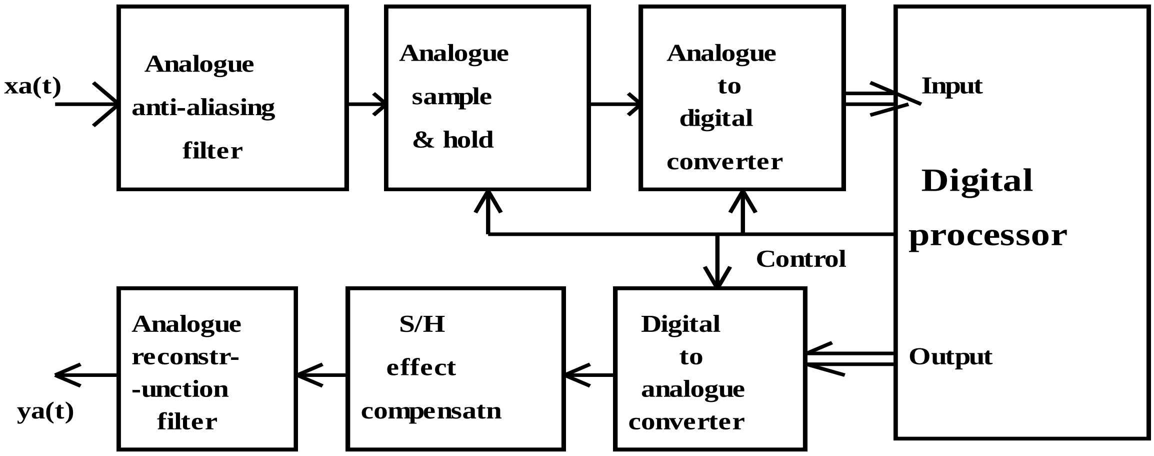
$$10 \log_{10}\left[\frac{2^{2m-2}\Delta^2/2}{\Delta^2/12}\right] = 10 \log_{10}\left[3 \times 2^{2m-1}\right] = 10\,[\log_{10}(3) + (2m-1)\log_{10}(2)]$$

= 1.8 + 6m dB. (i.e. approx 6 dB per bit)

This simple formula is often assumed to apply for a wider range of signals which are approximately sinusoidal.

<u>Example</u>: (a) How many bits are required to achieve a SQNR of 60 dB with sinusoidally shaped signals amplified to occupy the full range of a uniformly quantising A/D converter?
(b) What SQNR is achievable with a 16-bit uniformly quantising A/D converter applied to sinusoidally shaped signals?
<u>Solution:</u> (a) About ten bits. (b) 97.8 dB.

6.12 Block diagram of a DSP system for analogue signal processing



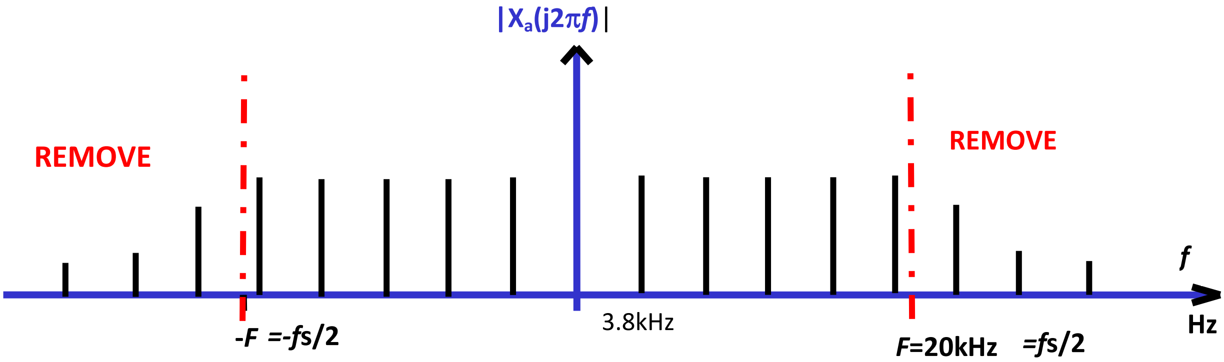
Antialiasing LPF: Analogue low-pass filter with cut-off $f_S/2$ to remove (strictly, to sufficiently attenuate)
any spectral energy which would be aliased into the signal band.
Analogue S/H: Holds input steady while A/D conversion process takes place.
A/D convertr: Converts from analogue voltages to binary numbers of specified word-length.
Quantisation error incurred. Samples taken at f_S Hz.
Digital processor: Controls S/H and ADC to determine f_S fixed by a sampling clock connected via an
input port. Reads samples from ADC when available, processes them & outputs
to DAC. Special-purpose DSP devices (microprocessors) designed specifically for this
type of processing.
D/A convertr: Converts from binary numbers to analogue voltages. "Zero order hold" or "stair-case
like" waveforms normally produced.
S/H compensation: Zero order hold reconstruction multiplies spectrum of output by sinc($\pi f/f_S$)
Drops to about 0.64 at $f_S/2$. Lose up to -4 dB.
S/H filter compensates for this effect by boosting the spectrum as it approaches $f_S/2$.
Can be done digitally before the DAC or by an analogue filter after the DAC.
Reconstruction LPF: Removes "images" of $-fs/2$ to $fs/2$ band produced by S/H reconstruction.
Specification similar to that of input filter.

Example: Why must analogue signals be low-pass filtered before they are sampled?

If {x[n]} is obtained by sampling xa(t) at intervals of T, the DTFT X(ejΩ) of {x[n]} is (1/T)repeat2π/T{Xa(jω)}. This is equal to the FT of $x_S(t)$ = sampleT(xa(t))
If xa(t) is bandlimited between π±/T then Xa(jω) =0 for |ω| > π/T.
It follows that X(ejΩ) = (1/T)Xa(jω) with Ω=ωT. No overlap.
We can reconstruct xa(t) perfectly by producing the series of weighted impulses xS(t) & low-pass filtering. No informatn is lost. In practice using pulses instead of impulses give good approximation.
Where xa(t) is not bandlimited between ±π/T then overlap occurs & XS(jω) will not be identical to Xa(jω) in the frequency range ±*fs*/2 Hz. Lowpass filtering x*S*(t) produces a distorted (aliased) version of xa(t). So before sampling we must lowpass filter xa(t) to make sure that it is bandlimited to π±/T i.e. *fS*/2 Hz

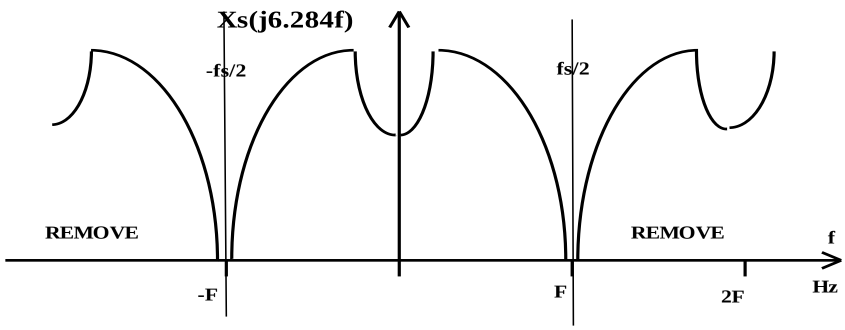
6.13. Choice of sampling rate: Assume we wish to process $x_a(t)$ band-limited to $\pm F$ Hz. F could be 20 kHz, for example.
In theory, we could choose $f_S = 2F$ Hz. e.g. 40 kHz. There are two related problems with this choice.
(1) Need very sharp analogue anti-aliasing filter to remove everything above F Hz.
(2) Need very sharp analogue reconstruction filter to eliminate images (ghosts):

To illustrate problem (1), assume we sample a musical note at 3.8kHz (harmonics at 7.6, 11.4, 15.2, 19, 22.8kHz etc. See figure below. Music bandwidth F=20kHz sampled at $2F$ = 40 kHz. Need to filter off all above F without affecting harmonics below F. Clearly a very sharp ('brick-wall') low-pass filter would be required and this is impractical (or impossible actually).



The consequences of not removing the musical harmonics above 20kHz (i.e. at 22.8, 26.6 and 30.4 kHz) by low-pass filtering would that they would be 'aliased' and become sine-waves at 40-22.8 = 17.2 kHz, 40-26.6 = 13.4 kHz, and 40-30.4 = 9.6 kHz. These aliased frequencies are not harmonics of the fundamental 3.8kHz and will sound 'discordant'. Worse, if the 3.8 kHz note increases for the next note in a piece of music, these aliased tones will decrease to strange and unwanted effect.

To illustrate problem (2) that arises with reconstruction when $f_S = 2F$, consider the graph below.



Again it is clear that a 'brick-wall filter is needed to remove the inmages (ghosts) beyond ±F without affecting the music in the frequency range ±F .

Effect of increasing the sampling rate:

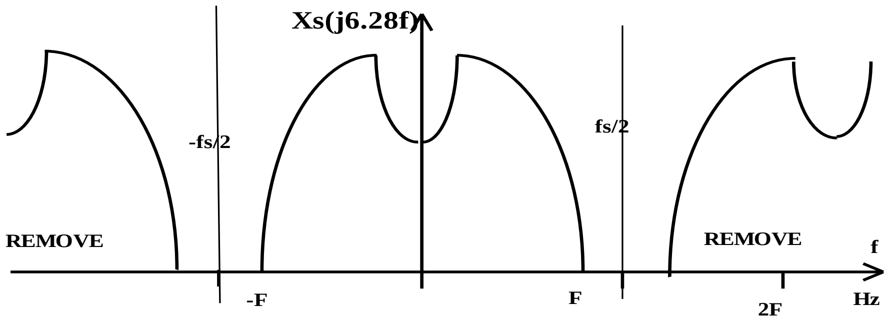
Slightly over-sampling: Consider effect on the input antialiasing filter requirements if the music bandwidth remains at $\pm F = \pm 20$kHz, but instead of sampling at $fS = 40$ kHz we ‘slightly over-sample’ at

44.1kHz. See the diagram below. To avoid input aliasing, we must filter out all signal components above $fS/2 = 22.05$ kHz without affecting the music within ±20kHz.

We have a ‘guard-band’ from 20 to 22.05 kHz to allow the filter’s gain response to ‘roll off’ gradually.

So the analogue input filter need not be ‘brick-wall’.

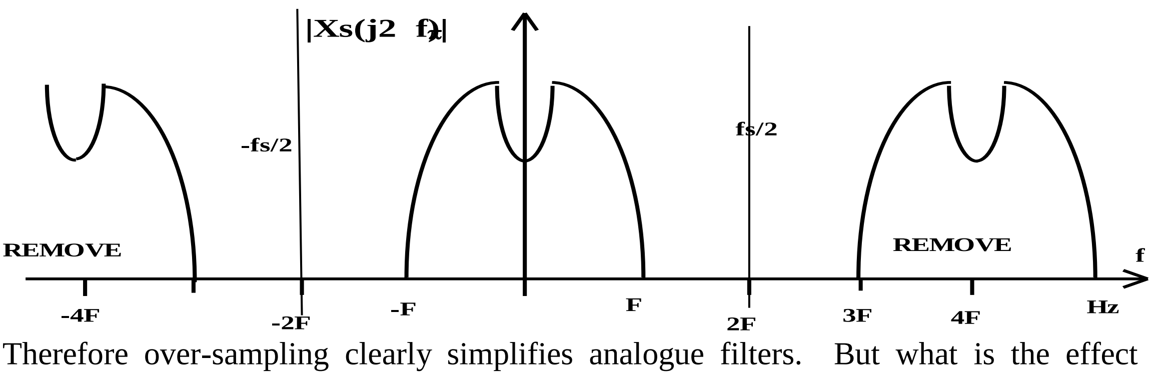
It may be argued that guard-band is 4.1kHz as the spectrum between 22.05 and 24.1kHz gets aliased to the range 20-22.05kHz which is above 20kHz. Once the signal has been digitised without aliasing, further digital filtering may be applied to efficiently remove any signal components between 20 kHz and 22.05 kHz that arise from the use of a simpler analogue input antialiasing filter.



Analogue filtering is now easier. To avoid aliasing, the analog input filter need only remove everything above $f_S - F$ Hz. For HI-FI, it would need to filter out everything above 24.1 kHz without affecting 0 to 20 kHz.

Higher degrees of over-sampling: Assume we wish to digitally process signals bandlimited to $\pm F$ Hz and that instead of taking 2F samples per second we sample at twice this rate, i.e. at $4F$ Hz.

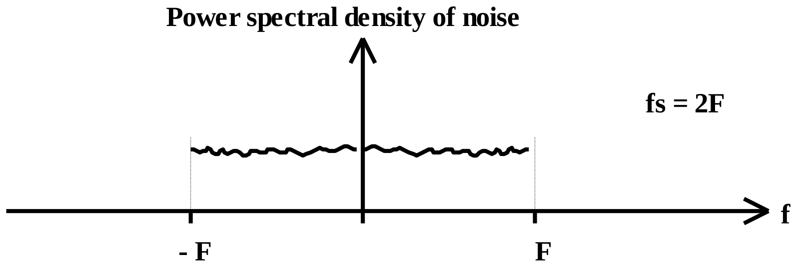
The anti-aliasing input filter now needs to filter out only components above 3F (*not* 2F) without distorting 0 to F. Reconstruction is also greatly simplified as the images start at ± 3F as illustrated below. These are now easier to remove without affecting the signal in the frequency range $\pm F$.

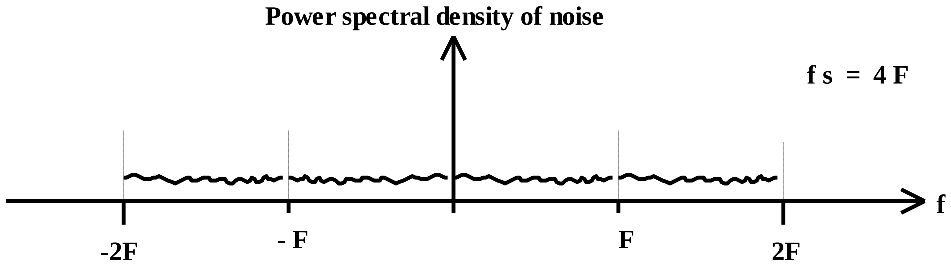


Therefore over-sampling clearly simplifies analogue filters. But what is the effect on the SQNR?

Does the SQNR (a) reduce, (b) remain unchanged or (c) increase ?

As Δ and m remain unchanged the maximum achievable SQNR Hz is unaffected by this increase in sampling rate. However, the quantisation noise power is assumed to be evenly distributed in the frequency range $\pm f_S/2$, and f_S has now been doubled. Therefore the same amount of quantisation noise power is now more thinly spread in the frequency range ±2F Hz rather than ±F Hz.





It would be a mistake to think that the reconstruction low-pass filter must always have a cut-off frequency equal to half the sampling frequency. The cut-off frequency should be determined according to the bandwidth of the signal of interest, which is ±F in this case. The reconstruction filter should therefore have cut-off frequency F Hz. Apart from being easier to design now, this filter will also in principle remove or significantly attenuate the portions of quantisation noise (error) power between F and 2F.
Therefore, assuming the quantisation noise power to be evenly distributed in the frequency domain, setting the cut-off frequency of the reconstruction low-pass filter to F Hz can remove about half the noise power and hence add 3 dB to the maximum achievable SQNR.
Over-sampling can increase the maximum achievable SQNR and also reduces the S/H reconstruction roll-off effect. Four times and even 256 times over-sampling is used. The cost is an increase in the number of bits per second (A/D converter word-length times the sampling rate), increased cost of processing, storage or transmission and the need for a faster A/D converter. Compare the 3 dB gained by doubling the sampling rate in our example with what would have been achieved by doubling the number of ADC bits m. Both result in a doubling of the bit-rate, but doubling m increases the max SQNR by 6m dB i.e. 48 dB if an 8-bit A/D converter is replaced by a (much more expensive) 16-bit A/D converter. The following table illustrates this point further for a signal whose bandwidth is assumed to be 5 kHz.

m	f_S	Max SQNR	bit-rate
10	10 kHz	60 dB	100 k
10	20 kHz	63 dB	200 k
12	10 kHz	72 dB	120 k

Conclusion: Over-sampling simplifies the analogue electronics, but is less economical with digital processing/storage/transmission resources.

6.14. Digital anti-aliasing and reconstruction filters. We can get the best of both worlds by sampling at a higher rate, 4F say, with an analogue filter to remove components beyond ±3F, and then applying a digital filter to the digitised signal to remove components beyond ± F. The digitally filtered signal may now be down-sampled ("decimated") to reduce the sampling rate to 2F. To do this, simply omit alternate samples.

To reconstruct: "Up-sample" by placing zero samples between each sample:

e.g. { …, 1, 2, 3, 4, 5, …} with f_S = 10 kHz

becomes {…, 1, 0, 2, 0, 3, 0, 4, 0, 5, 0, …} at 20 kHz.

This creates images (ghosts) in the DTFT of the digital signal. These images occur from *F* to 2*F* and can be removed by a digital filter prior to the A/D conversion process. We now have a signal of bandwidth F sampled at $f_S = 4F$ rather than $2F$. Reconstruct as normal but with the advantages of a higher sampling rate.

6.15: Compact Disc (CD) format:

Compact discs store high-fidelity (hi-fi) sound of 20 kHz bandwidth sampled at 44.1 kHz. Each of the two stereo channels is quantised to 16-bits per sample, and is given another 16-bits for errors protection. Music therefore stored at 44100 x 32=1.4112Mbytes/s (with FEC).

A recording studio will over-sample & use a simple analogue input filter. Once the signal has been digitised at a sampling rate much higher than the required 44.1 kHz, a digital antialiasing filter may be applied to band-limit the signal to the frequency range ±20 kHz. So if the simple analogue input filter has not quite removed all the energy above 20 kHz, this energy is now removed (i.e. very strongly attenuated) digitally.

So now we have a digitised signal that is definitely well band-limited within ±20 kHz. But it is sampled at a much higher rate than can be stored on the CD. It must be 'down-sampled' (decimated) to 44.1 kHz for storing on the CD. If you think about it, all that is necessary is to omit samples; i.e. if the sampling rate is four times 44.1 kHz (=196.4 kHz) , we just take one sample, discard 3, take 1, discard 3, and so on. You are actually sampling a digital signal and it works because of the sampling theorem. If you are not convinced, consider the 196.4 kHz sampled version being ideally converted back to analogue and then sampling this 20 kHz bandwidth analogue signal at 44.1 kHz. But you don't actually need to do the D to A conversion and resampling as exactly the same result is obtained by omitting samples.

Most CD players "up-sample" the digital signal read from the CD by inserting zeros to obtain a signal sampled at say 88.2, 176.4 kHz or higher. Inserting zeros is not good enough by itself. It creates images (ghosts) in the spectrum represented by the 'up-sampled' digital signal. It is as though you have produced an analogue version of $x_s(t) = \text{sample}_T\{x_a(t)\}$ with spikes at 44.1 kHz, and then resampled this at say 176.4 kHz (4 times up-sampling) without removing images between ±88.2 kHz. Actually you haven't produced this analogue signal, but the effect is the same. So these images created within the digital signal by 'up-sampling' must be removed by digital filtering. The digital filter is a digital 'reconstruction filter' and its requirements are the same as for an analogue reconstruction filter; i.e . it must remove spectral energy outside the range ±20 kHz without affecting the music in the range ±20 kHz. Fortunately this filtering task is much easier to do digitally than with an analogue filter.

After the digital filtering, we have a 20 kHz bandwidth music sampled not at 44.1 kHz, but now at a higher rate, say 176.4 kHz or higher. We apply it in the normal way to a 'digital to analogue converter' (DAC) with staircase reconstruction. The DAC may have to be a bit faster than that needed for a 44.1kHz sampling rate.

The DAC output will have to be low-pass filtered by an analogue reconstruction filter required to remove images (ghosts) without affecting the ±20kHz music. But the ghosts are now much

higher in frequency. In fact with four times over-sampling, the first ghost starts at 192.4 – 20 kHz which is 172.4 kHz. This gives us a considerable 'guard-band' between 20 kHz and 172.4 kHz allowing the analogue low-pass reconstruction filter's response to fall off quite gradually. A simple analogue reconstruction filter is now all that is required.
In conclusion, with up-sampling, the reconstruction filtering is divided between the digital and the analogue processing and simplifies the analalogue processing required at the expense of a faster DAC. An added advantage with over-sampling is that the 'sample & hold' effect (up to 4 dB attenuation) that occurs without up-sampling is now greatly reduced because effectively shorter pulses (closer to impulses) are being used because of the four times faster DAC. Four, 8 or 16 times over-sampling was commonly used with 14-bit and 16 bit D/A converters. For 8-times over-sampling seven zeros are inserted between each sample.

"Bit-stream" converters up- (over-) sample to such a degree (typically 256) that a one-bit ADC is all that is required. This produces high quantisation noise, but the noise is very thinly spread in the frequency-domain. Most of it is filtered off by very simple analogue reconstruction filter. For 256 times over-sampling the gain in SQNR is only 3 x 8 = 24 dB. This is not enough if a 1-bit D/A converter is being used. Some more tricks are needed, for example noise-shaping which distributes the noise energy unevenly in the frequency-domain with greater spectral density well above 20 kHz where the energy will be filtered off.

<u>Example:</u> A DSP system for processing sinusoidal signals in the range 0 Hz to 4 kHz samples at 20 kHz with an 8-bit ADC. If the input signal is always amplified to use the full dynamic range of the ADC, estimate the SQNR in the range 0 to 4 kHz. How would the SQNR be affected by decreasing fS to 10 kHz and replacing the 8-bit ADC by a 10-bit device? Are there any disadvantages in doing this?

<u>Example</u>: We have discussed how to change the sampling rate of a digitised signal by the insertion or removal of samples before or after digital filtering. Consider how you could change the sampling rate from 8 kHz to 10 kHz and also vice-versa. Then consider how you would up-sample from a 16kHz sampling rate to 44.1 kHz. All this must be done without perceived changes in duration or pitch; i.e. you can't just increase or decrease the clock speed without modifying the samples..