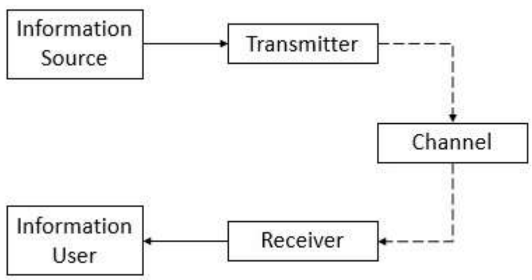
Subject Introduction

A typical communication system starts with a source of information of interest that we want to send somewhere. Normally, it is required to be converted to electrical form using a transducer. After its electrical conversion, we use an equipment called transmitter, the work of which is to prepare the signal for the physical medium which it is to be sent over. In essence, the job of the transmitter is to match the properties of he information signal to the properties of the physical medium over which it is to be sent so that efficient communication is possible. An abstraction of all kinds of physical mediums through which transmission can take place is termed as a channel. At the other end or destination, there is a receiver the job of which is to interpret the received information in a way that is usable to the user. Hence, in a communication system, there are five major blocks connected as shown below.



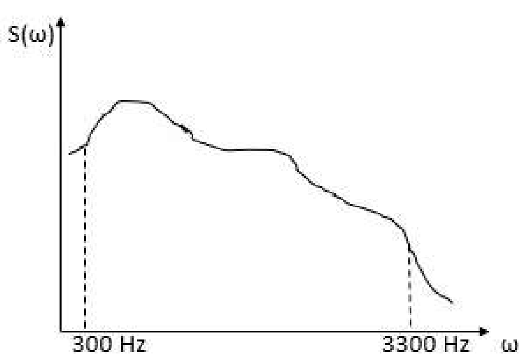
Communication System Block diagram

Information Sources

Communication can use a wide range of sources of information based upon the type of data being transmitted. Voice / Speech signals make up the bulk of communication and is easily the most common form of information. TV and Fax involve communication using moving and still pictures respe3ctively. Data signals are commonly used in the communication of computers. There might be many other signals or information sources, but these three broad categories roughly cover all useful class of sources.

One of the most basic frameworks needed in the discussion of deterministic signals used as information sources is the spectrum of the signal. In its basic sense, the spectrum can be understood as the Fourier Transform of the signal. As is known, the FT of a real signal is (normally) a complex function, meaning that it has both magnitude as well as phase components varying with respect to frequency, called Magnitude and Phase spectrums respectively. Following is a brief discussion on the afore mentioned types of signals in context of their spectrums.

Voice/Speech signals



Frequency Spectrum of typical Speech signal

A typical spectrum of human speech signals can be drawn as shown here. As can be seen in the graph, there might be some peaks, but the maximum amount of the intelligent data used in the speech/ voice communication can be seen to lie in the frequency range of 300Hz to 3300Hz. Including some guard band, we may infer that the bandwidth (maximum possible relevant frequency in the spectrum) of a speech signal is 4KHz. It is noteworthy though, that the actual spectrum of the voice signal may go up to 7 – 8 KHZ. Thus, for studio quality voice processing, the bandwidth must be kept of the same order. In telephonic communication, however, the above mentioned 4KHz bandwidth is considered adequate.

Note: Voice and speech are two forms of the most studied and researched forms of information. It is observed that the data that is generated from a voice signal (binary data in the form of bits) is much smaller in comparison to the bandwidth it occupies. Thus, scientists have tried and succeeded in generating methods of transmitting voice signals over considerably less bandwidth.

Picture signals

As discussed earlier, the picture signal can be of two forms.

a) Moving Pictures/ Television/ Video transmission
While transmitting picture signals with fast changing characteristics, interpreted as moving pictures, care must be taken that we don't loose the essence of motion itself. We first need to convert the special picture signal (2D in nature) into and electrical form (1D time-varying signal). This is done using a picture tube and the output is scanned (horizontally) line wise. This scan must be fast enough to retain the essence of motion of the pictures. It can hence be interpreted, that this fast variation required would mean larger bandwidth requirement. Typically, the TV signals are seen to have a bandwidth of 4.5MHz to 5MHz.

b) Still Pictures/Fax/Photo Transmission
In a fax signal or a signal transmitting a single image, speed of transmission is not a concern. The image in the form of electrical signals can be scanned at any speed convenient to us based upon the available hardware. Thus, Fax signals are possible to transit over general telephone lines which are designed to operate on a bandwidth much smaller that that of video signals.

Data signals

This is a very different type of information in the sense that it is digital in nature in comparison to the other sources discussed so far, which were analog. It must be noted though that the channel or the physical medium through which transmission takes place is still analog in nature, hence in the context of effective transmission, it doesn't make much difference. What is interesting to note though is that the bandwidth of transmission of data signals will depend upon the rate at which data transmission is needed (data rate in bits/s). Thus, a fast data rate system requires larger bandwidth in comparison to a slow data rate system that requires low bandwidth. Consider the example of mobile networks evolving from 2G to 3G to 4G and so on. The idea is to increase the bandwidth allocation per user to increase the rate at which transmission can take place.

Transmitter

Basic function of the transmitter is to match the signal characteristics of the output of the information source with the characteristics of the channel. Let us try to understand this using an example. Say the transmitter needs to transmit data on to free space. For such transmission, one requires an appropriate antenna, the design of which depends upon the frequency of operation. It is a known fact in electromagnetics, that the height of an antenna transmitting EM waves must of the order of the wavelength of the wave. Considering a voice signal, the maximum frequency of transmission is kept at 4KHz. Thus, the wavelength of such a signal is nearly 75km! Obviously, an antenna of comparable height will be impossible to design efficiently. Hence, the information source output is not matched to the channel of transmission. In this case, the job of the transmitter would be to embed the information signal on to another signal for which the antenna height requirement is much smaller. It can easily be deduced that such signals will be of very high frequency. As these signals carry the information signal in some form, these are known as carrier signals. Typically, the carrier wave is a pure sinusoid of high frequency. The process of embedding the information on to a carrier is called modulation. Thus, ease of radiation is one of the reasons why modulation is needed. There are some other needs which may also be studied as advantages and will be discussed a little later.

What this example establishes is that the role of the transmitter is what makes it possible for information exchange over any physical medium to take place. Based upon the output of the source and the requirement of the channel, the design of this transmitter will vary. An obvious observation here is that the receiver will have to be equipped with an appropriate demodulator to reverse the effect of the transmitter on the information signal.

Channel

A channel is basically a physical medium. Some examples are cables, optical fibre, free space, etc. Another way of looking at a channel is as an abstraction. This means that channel is to be modelled in an abstract fashion. Thus, all the effects of the transmitter, receiver and physical medium that are undesirable or unexpected are modelled as a block called channel. These undesired effects can be simply called noise. Hence, to say that noise exists in a communication system simply means that noise exists in the channel.

Noise in Communication Systems

Broadly, noise sources can be classified into two categories: Internal noise and External Noise.

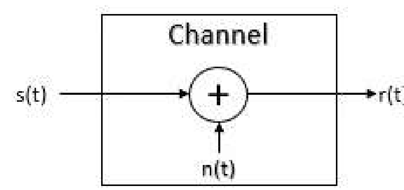
External noises are attributed to sources outside the circuit. Natural sources may include lightening, atmospheric noises, cosmic EM waves, etc. Man-made sources may include EM distortions generated by electrical lines, commutator switches in motors, ignition noises in automobiles, etc. Along with this, other Radio frequency interferences have become more prominent with the advent of multiuser communication over similar frequency spectrums. Multipath fading is also a common issue in wireless communication. An in-depth discussion of different forms of noises will be done later in this course.

Mathematical Models of Communication Channels

In the design of communication systems for transmitting information through physical channels, we find it convenient to construct mathematical models that reflect the most important characteristics of the transmission medium. Then, the mathematical model for the channel is used in the design of the channel encoder and modulator at the transmitter and the demodulator and channel decoder at the receiver. Mentioned below is a brief description of the channel models that are frequently used to characterize many of the physical channels that we encounter in practice.

a) The Additional Noise Channel

The simplest mathematical model for a communication channel is the additive noise channel. In this model the transmitted signal s(t) is corrupted by an additive random noise process n(t). Physically, the additive noise process may arise from electronic components and amplifiers at the receiver of the communication system or from interference encountered in transmission as in the case of radio signal transmission.



$$r(t) = as(t) + n(t)$$

If the noise is introduced by electronic components and amplifiers at the receiver, it may be characterized as thermal noise. This type of noise is characterized statistically as a Gaussian noise process. Hence, the resulting mathematical model for the channel is usually called the additive Gaussian noise channel. Because this channel model applies to a broad class of physical communication channels and because of its mathematical tractability, this is the predominant channel model used in our communication system analysis and design. Channel attenuation is easily incorporated into the model. When the signal undergoes attenuation in transmission through the channel, the received signal is

$$r(t) = as(t) + n(t)$$

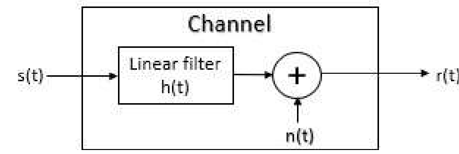
where a represents the attenuation factor.

b) The Linear Filter Channel

In some physical channel such as wireline telephone channels, filters are used to ensure that the transmitted signals do not exceed specified bandwidth limitations and, thus, do not interfere with one another. Such channels are generally characterized mathematically as linear filter channels with additive noise. Hence, if the channel input is the signal s(t), the channel output is the signal

$$r(t) = s(t) * h(t) + n(t)$$
$$= \int_{-\infty}^{\infty} h(\tau)s(t-\tau)d\tau + n(t)$$

where h(t) is the impulse response of the linear filter.

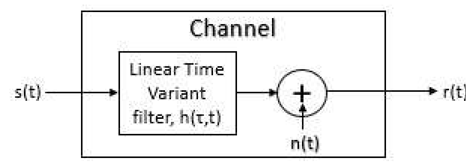


$$r(t) = s(t) * h(t) + n(t)$$

c) The Linear Time-Variant Filter Channel

Physical channels such as underwater acoustic channels and ionospheric radio channels which result in time-variant multipath propagation of the transmitted signal may be characterized mathematically as time-variant linear filters. Such linear filters are characterized by time-variant channel impulse response h(τ,t) which is the response of the channel at time τ, due to an impulse applied at time τ-t. Thus, t represents the 'age' variable. For an input signal s(t), the channel output signal is

$$r(t) = s(t) * h(\tau, t) + n(t)$$
$$= \int_{-\infty}^{\infty} h(\tau, t)s(t-\tau)d\tau + n(t)$$



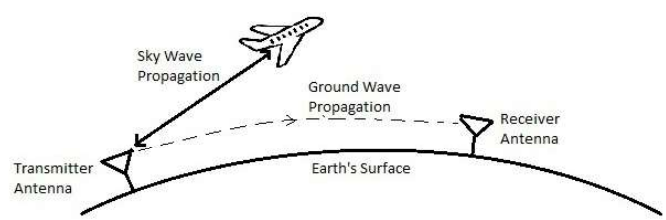
$$r(t) = s(t) * h(\tau, t) + n(t)$$

Propagation in a Communication Channel

The various methods of propagation of signals in a communication channel depends upon the type of channel itself. Based on this criterion, the classification can be done as follows:

a) EM wave propagation channel – Free space channel
b) Guided EM wave propagation channel
c) Optical Channels

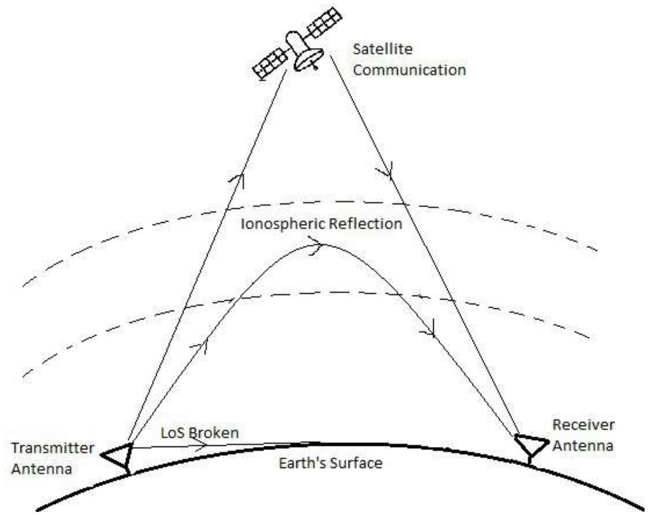
As such, pure free space communication cannot take place in the earth's atmosphere. But still wireless communication can be considered free space communication under specific assumptions.



As is known, a medium bounded on one side by a conductor gives rise to guided waves. In the case shown here, transmitter antenna is sending an EM wave to a nearby receiver. In such a case, the Earth's surface would act as a conductor and the curvature of Earth guides the wave. This is called the ground wave propagation and classified as a guided transmission (type b). The limitation of such a system is that it can only be used for low frequency transmission, since for higher frequencies, the attenuation due to Earth's surface increases significantly. Therefore, ground wave propagation is well suited for lower frequencies.

Consider now, that the communication needs to take place between a transmitter at the Earth's surface and an Aeroplane. Such communication can only take place in Line of Sight (LoS). Since this involved transmission outwards from the Earth's surface, for all practical purposes it can be termed as free space communication, although a more suited name is Sky Wave Communication. This transmission can work for higher frequencies as well.

LoS communication is also possible between stations on the Earth's surface. This is, though, limited by the distance between the stations. Due to the curvature of the Earth, there will be a point beyond which the signal will be obstructed by the surface itself.



The way out of this problem is to transmit the wave skywards and somehow have it reflected to the destination. One way of doing so is making use of a natural phenomenon. In the Earth's atmosphere, there is a layer, called Ionosphere, which is made up of charged ions and acts as a passive reflector. The limitation of this is that the ionosphere is capable of reflecting frequencies up to 30MHz. Higher frequencies will escape the ionosphere and move into free space.

For higher frequencies (>>30MHz), we need to have a reflecting satellite in an orbit around the planet. If that satellite only reflects the wave, it is called a passive satellite. In case it received and retransmits the wave, it is called an active satellite.

Among the guided wave channels, some more examples that can be quoted are:

a) Cables and Wire pairs: Depending upon the lower frequencies or higher frequencies, a pair of wires may be modelled as a lumped circuit or transmission lines respectively. In each case, they are used as guided media to transmit EM waves.
b) Waveguides: They are essentially pipes (cylindrical or rectangular) used to transfer EM waves from transmitter circuit to the antenna.

Optical Channels

Optical communication is possible using Fibre Optics as well as Free Space Optics. Fibre optics is a cylindrical waveguide used for transmission of optical signals, which in turn are essentially EM waves. Free Space Optics is typically like a Free Space EM communication. The main difference is that the information is transmitted in the form of highly focussed beam of light to enable transmission from point A to B. Mainly, Free Space Optics is used in sky wave communication, but it is also finding many terrestrial applications as well.

Note: For further reading on the topic, please refer to Chapter 1 of "Fundamentals of Communication Systems" by John G. Proakis and Masoud Salehi.

Review of relevant mathematical concepts

Some mathematical formulas are relevant to the study of this course.

Trigonometric relations:

1) $\cos(A+B) = \cos A \cos B - \sin A \sin B$
2) $\cos(A-B) = \cos A \cos B + \sin A \sin B$
3) $\sin(A+B) = \sin A \cos B + \cos A \sin B$
4) $\sin(A-B) = \sin A \cos B - \cos A \sin B$
5) $\sin 2A = 2 \sin A \cos A$
6) $\cos 2A = \cos^2 A - \sin^2 A$
7) $\cos^2 A = \frac{1+\cos 2A}{2}$
8) $\sin^2 A = \frac{1-\cos 2A}{2}$
9) $2\cos A \cos B = \cos(A+B) + \cos(A-B)$
10) $2 \sin A \sin B = \cos(A-B) - \cos(A+B)$

Fourier Transform relations

Analysis equation:

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt$$

Synthesis equation:

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

Symmetry relations for Fourier Transform:

$$x(t)\ real \qquad \begin{aligned} X(j\omega) &= X^*(-j\omega) \\ Re\{X(j\omega)\} &= Re\{X(-j\omega)\} \\ Im\{X(j\omega)\} &= -Im\{X(-j\omega)\} \\ |X(j\omega)| &= |X(-j\omega)| \\ \angle X(j\omega) &= -\angle X(-j\omega) \end{aligned}$$

If $x(t)$ *real and even* then $X(j\omega)$ *real and even*

If $x(t)$ *real and odd* then $X(j\omega)$ *imaginary and odd*

Other important properties of Fourier Transform:

$$x(t-t_0) \leftrightarrow e^{-j\omega t_0} X(j\omega)$$

$$x(t) * y(t) \leftrightarrow X(j\omega)\, Y(j\omega)$$

$$e^{j\omega_0 t} x(t) \leftrightarrow X\big(j(\omega-\omega_0)\big)$$

$$x(t)\, y(t) \leftrightarrow \frac{1}{2\pi} X(j\omega) * Y(j\omega)$$

$$\frac{dx(t)}{dt} \leftrightarrow j\omega X(j\omega)$$

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(j\omega)|^2\, d\omega$$

Common Fourier Transform pairs:

$$\sum_{k=-\infty}^{\infty} a_k e^{-jk\omega_0 t} \leftrightarrow 2\pi \sum_{k=-\infty}^{\infty} a_k \delta(\omega-\omega_0)$$

$$\cos \omega_0 t \leftrightarrow \pi[\delta(\omega+\omega_0) + \delta(\omega-\omega_0)]$$

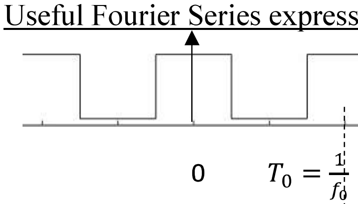
$$\sin \omega_0 t \leftrightarrow \frac{\pi}{j}[\delta(\omega-\omega_0) - \delta(\omega+\omega_0)]$$

$$x(t) = \begin{cases} 1 & |t| < T \\ 0 & |t| > T \end{cases} \leftrightarrow \frac{2\sin \omega T}{\omega}$$

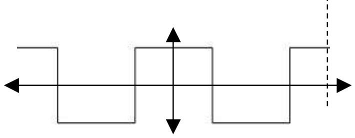
$$1 \leftrightarrow 2\pi\delta(\omega)$$

$$\delta(t) \leftrightarrow 1$$

Useful Fourier Series expressions:



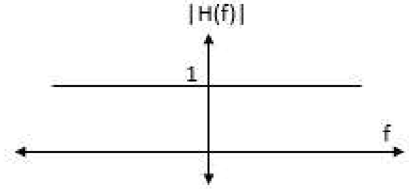
$$s(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos[2\pi f_0 t(2n-1)]$$



$$s(t) = \frac{4}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos\,[2\pi f_0 t(2n-1)]$$

Hilbert Transformation

Consider and LTI system with a transfer function

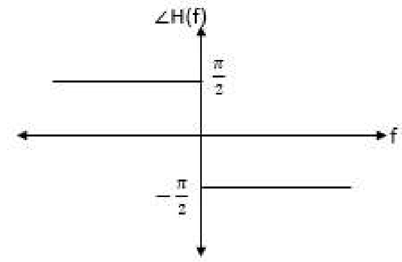


$$H(f) = -j\,sgn(f)$$

$$where \quad sgn(f) = \begin{cases} 1 & f > 0 \\ 0 & f = 0 \\ -1 & f > 0 \end{cases}$$

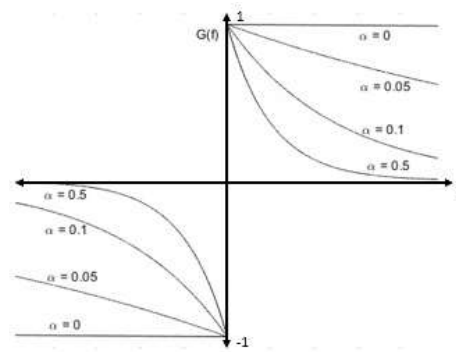
$$hence \quad |H(f)| = 1 \quad \forall f$$

$$and \quad \angle H(f) = \begin{cases} -\frac{\pi}{2} & f > 0 \\ \frac{\pi}{2} & f < 0 \end{cases}$$



Such a system is called a Hilbert Transformer. This implies that the Hilbert transform introduces a phase shift of $-\frac{\pi}{2}$ at all frequencies.

To figure out how the output will be affected by the Hilbert Transformer in the time domain, Fourier inverse of the given function can be evaluated. To this, we make use of the following assessment.



Consider a function:

$$G(f) = \begin{cases} e^{-\alpha f} & \forall\, f > 0 \\ -e^{\alpha f} & \forall\, f < 0 \end{cases}$$

It should be evident that as $\alpha \to 0, G(f) \to sgn(f)$. Hence, evaluating the Fourier Inverse of $G(f)$ and applying the limit would result in the Fourier Inverse of $sgn(f)$.

$$\mathcal{F}^{-1}\{G(f)\} = g(t) = \int_{-\infty}^{0} -e^{\alpha f} e^{j2\pi f t} df + \int_{0}^{\infty} e^{-\alpha f} e^{j2\pi f t} df$$

$$= \int_{-\infty}^{0} -e^{(\alpha + j2\pi\ )f} df + \int_{0}^{\infty} e^{(-\alpha + j2\pi t)f} df$$

$$= -\frac{1}{\alpha + j2\pi t} + \frac{1}{\alpha - j2\pi t}$$

Applying the limit $\alpha \to 0$, we get the Fourier inverse of a signum function as

$$\mathcal{F}^{-1}\{sgn(f)\} = \lim_{\alpha \to 0} g(t)$$

$$= -\frac{1}{j2\pi t} + \frac{1}{-j2\pi t} = \frac{j}{\pi t}$$

Hence, the Fourier Inverse of Hilbert Transform can be written as:

$$h(t) = \mathcal{F}^{-1}\{-j\,sgn(f)\} = \frac{1}{\pi t}$$

Therefore, for any signal $x(t)$, the Hilbert Transform, represented by $\hat{x}(t)$, can be written (using convolution property) as $\hat{x}(t) = x(t) * h(t)$

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

Hilbert Transform Properties

1. $\hat{\hat{x}}(t) = -x(t)$ since the whole signal goes through phase shift of π.
2. Energy and power of $x(t)$ and that of $\hat{x}(t)$ are the same; since $|X(f)| = |\hat{X}(f)|$
   Proof:
$$\begin{aligned}
&\hat{X}(f) = -j\, sgn(f) X(f) \\
&\Rightarrow \left|\hat{X}(f)\right| = |X(f)| \\
&\Rightarrow \int \left|\hat{X}(f)\right|^2 df = \int |X(f)|^2 df \\
&\Rightarrow \int \hat{x}^2(t) dt = \int x^2(t) dt
\end{aligned}$$
3. $x(t)$ and $\hat{x}(t)$ are orthogonal.
4. For a low pass signal ($m(t)$) and a high pass signal ($c(t)$) with non-overlapping spectra, the Hilbert Transform of their product is given by,
$$x(t) = m(t)c(t) \qquad \Rightarrow \qquad \hat{x}(t) = m(t)\,\hat{c}(t)$$
5. Analytic representation of any real valued signal $x(t)$ can be written as
$$x_p(t) = x(t) + j\,\hat{x}(t)$$
   where Analytic representation means the portion of signal with only the positive frequency components.

Consider an example.

Since $\mathcal{H}\{\cos 2\pi f_0 t\} = \sin 2\pi f_0 t$ and $\mathcal{H}\{\sin 2\pi f_0 t\} = -\cos 2\pi f_0 t$

hence, if $x(t) = m(t)\cos\omega_0 t$ *then* $\hat{x}(t) = m(t)\sin\omega_0 t$

and if $x(t) = m(t)\sin\omega_0 t$ *then* $\hat{x}(t) = -m(t)\cos\omega_0 t$

Therefore, the Analytic representation of $x(t)$ may be written as

$$x_p(t) = m(t)\cos\omega_0 t + j\, m(t)\sin\omega_0 t$$

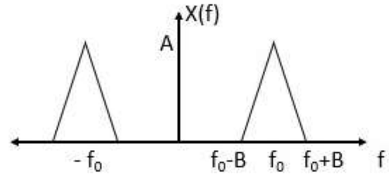
It must be noted here that

$$\begin{aligned}
\left|x_p(t)\right| &= \sqrt{x^2(t) + \hat{x}^2(t)} = Magnitude\ envelop\ of\ x(t) \\
&= \sqrt{m^2(t)\cos^2\omega_0 t + m^2(t)\sin^2\omega_0 t} \\
&= |m(t)|
\end{aligned}$$

Hence in this context, $x_p(t)$ is termed as the complex envelop of $x(t)$.

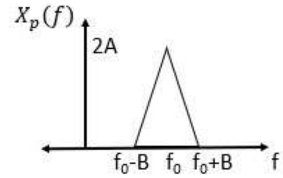
Analytic Representation of Band Pass Signals

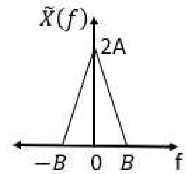
The multiplication of low pass and high pass signals normally results in band pass signals. Analytic (complex) representation of band pass signals is also an important tool.



To understand this, let's have a look at the spectrum.

$$X_p(f) = X(f) + j\hat{X}(f)$$
$$= X(f) + sgn(f)X(f)$$
$$= \begin{cases} 2X(f) & f > 0 \\ 0 & f < 0 \end{cases}$$





Let us assume that the signal $x(t)$ has a band pass spectrum as shown in the figure. Therefore, the spectrum of $x_p(t)$ will have only the positive sided component. This spectrum can in turn be observed as a translated version of a spectrum about 0. Such a translation is possible in frequency domain only when there is a complex exponential multiplied in the time domain.

$$x_p(t) = \tilde{x}(t)e^{j2\pi f_0 t}$$

What we may deduce from this is that $|x_p(t)| = |\tilde{x}(t)|$.

Further, $x(t) = Re\{x_p(t)\}$. We can imagine $\tilde{x}(t)$ to be a complex valued signal such that

$$\tilde{x}(t) = x_{Re}(t) + jx_{Im}(t)$$

Hence,

$$x(t) = x_{Re}(t)\cos 2\pi f_0 t - x_{Im}(t)\sin 2\pi f_0 t$$

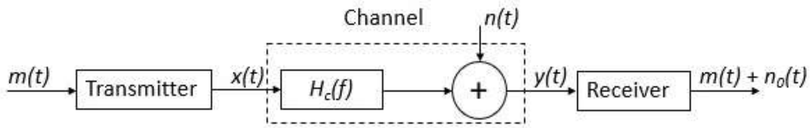
which again re-iterates the fact that any bandpass signal can be represented as a product of a low pass (representative) term with a high frequency carrier. The value of $x(t)$ has two components, one associated with $\cos 2\pi f_0 t$ and another with $\sin\ 2\pi f_0 t$, thus one of them is in phase with the carrier and the other is at a quadrature phase difference with the carrier. For this reason, the component $x_{Re}(t)$ is also called the in-phase component $x_I(t)$ and the component $x_{Im}(t)$ is also called the quadrature component $x_Q(t)$. This representation, called the quadrature representation, makes the analysis of band pass signals very much convenient. Even the other representation

$$x(t) = Re\{\tilde{x}(t)e^{j2\pi f_0 t}\}$$

is equally graceful. This representation is called the complex envelop representation.

Fundamentals of Analog Signal Transmission

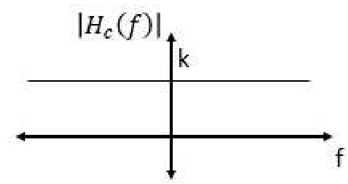
Let us consider a simple communication system block diagram.

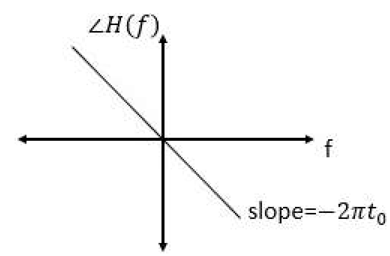


Here, $m(t)$ is a duly formatted message signal which is enough to drive the transmission circuit. Transmitter block contains everything required to prepare the signal for transmission over the channel. The effect of the physical channel on the transmitted signal $x(t)$ can be modelled in 2 parts. The first part is the description of the distortion introduced on the signal which can be modelled as a linear filter. This is basically because the channel typically introduces a frequency or phase error on to the signal which can be nicely modelled as a filter. The model used here is a filter with transfer function $H_c(f)$. The second part of the channel will account for the addition of noise (external) into the transmitted signal. This is thus represented as an adder which adds the filtered input and the noise $n(t)$. the resultant signal $y(t)$ is then fed to the receiver. The receiver tries to remove as much noise as possible and thus generates a replica of the message with some minimum irremovable noise components $n_0(t)$.

Distortion-less Transmission

The filtering effect of the channel modelled as $H_c(f)$ in the discussion above is commonly called as the channel distortion. Ignoring the effect of external noise, the ideal response of the channel would be to produce a replica of the transmitted signal at the receiver input. Such a condition is called distortion-less transmission. Let us define here what modifications in the signal are tolerable to us and what would be counted as distortion.





Firstly, amplitude scaling effect on the signal, (i.e. the channel introduces only a constant amplitude change), is not counted as distortion as it can be easily overcome. Secondly, constant time delays in the reception of the signal as an exact replica will also not be considered as distortion. To sum up, mathematically we can say that

$$y(t) = kx(t - t_0)$$

may be called the description for an ideal distortion-less channel. Same thing can be represented in the frequency domain as

$$Y(f) = ke^{-j2\pi f t_0} X(f)$$

Therefore, the $H_c(f)$ channel function should ideally be

$$H_c(f) = ke^{-j2\pi f\;{}_0}$$

Therefore, for ideal distortion-less transmission, the channel transfer function will be flat in magnitude response and linearly frequency dependent phase response.

This type of response may be too much to expect due to its restricted behaviour over infinite frequency range. It may therefore be sufficient, for the channel to have the given characteristics in the required frequency range (or operational bandwidth). What happens outside that range may not be of any academic significance. Hence, if the signal is defined to be operationally relevant in the range $-B \le f < -B$, then

$$|H_c(f)| = k \qquad \forall\ |f| \le B \qquad \angle H_c(f) = -2\pi t_0 \quad \forall\ |f| \le B$$

or simply, $$H_c(f) = ke^{-j2\pi f\;{}_0} \quad \forall\ |f| \le B$$

In practice, even these relaxed conditions may be unattainable and as a result, unignorable changes occur in the signal, which are termed as **distortion.**

Types of Distortion

Distortions in the signals can be classified broadly into two categories: Linear Distortion and Non-Linear Distortion.

a) Linear Distortion: Distortion caused by the non-ideal behaviour of a linear filter, these distortions, whether in magnitude or in phase responses, can be termed as linear distortion.

   a. Amplitude Distortion

   In simple terms, this would mean that $|H_c(f)| \neq k \quad \forall \quad |f| < B$

   This means, the magnitude value of $H_c(f)$ is not constant over the bandwidth of interest, the result of which is that the channel is introducing amplitude distortion. Hence, the channel is not introducing the same level of attenuation at all frequencies. (Time domain effects of the same may be a little more complicated to determine and would be dependent on the type of signal being used). Tolerable value of amplitude distortion is considered as 1dB. Any larger variations in the magnitude are un-ignorable and have to be taken into account.

   b. Phase Distortion/ Delay Distortion

   This distortion arises when the phase transfer function that behaves non-ideally,

   i.e. $\angle H_c(f) \neq -2\pi t_0 f \pm m\pi; \qquad |f| < B$

   Note: The factor of $\pm m\pi$ introduces a multiplication constant of ± 1 which can be included in the magnitude factor k.

   If this happens then different frequency components get delayed by different amount of time, causing improper reception at the receiver end. Fortunately, in Analog communication, typically for speech signals, delay distortion is of no consequence. The reason for this is that human ear is insensitive to delay distortion. This is not valid for other information signals though. Images are affected by phase distortion and it must be considered as human eye is not insensitive to phase distortion. Similarly, for data communication, phase distortion will cause havoc.

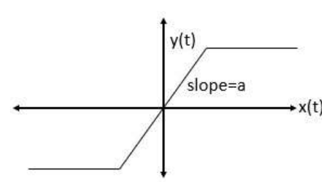
To sum up, linear distortion implies that different frequency components are treated differently, whether in magnitude or phase or both, by the channel. One convenient point in this regard in that, at least theoretically, its compensation is quite simple. We can simply put a filter with transfer function that is inverse of the channel transfer function at the receiver which would simply negate the effect of the channel. This process is called equalization.

This means that if we assume $H_c(f)H_{eq}(f) = ke^{-j2\pi f\,{}_0}$; $|f| < B$, then in theory, channel effect can be completely nullified. In practice, though, there are a couple of reasons why this isn't so simple. Firstly, we are assuming that we know the channel transfer function. This is a possibility and channel characteristics are very rarely possible to model. But even if we are aware of channel characteristics, there is another difficulty which we have conveniently ignored. This problem in the existence of external noise components. In the presence of external noise, equalization may even result in the amplification of the noise component as well.

b) Non-Linear Distortion

As the name suggests, this distortion arises due to non-linear behaviour of devices. It typically arises due to non linear transfer characteristics of amplifiers, mixers, etc.

For example, the practical characteristics of typical amplifiers can be drawn as shown on the side. Ideally, $y(t) = ax(t)$ should be true for all $x(t)$ for this system to be an ideal linear operation. But practically, all amplifiers face a specific non linearity called the saturation non linearity. This arises due to the fixed value of the power supply for the amplifier.



Thus, if the amplifier works in the linear range, the output is distortion free. But if the input further increases to a value beyond the linear range, the output is no longer linear, hence causing distortion. Mathematically, such non-linear functions are often modelled using polynomial representation. A typical model is

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

where for a given or appropriate degree of this polynomial, it is theoretically, possible to find the coefficients $a_1, a_2, a_3, \ldots$

An important way in which non-linear distortion differs from linear distortion is that non-linear distortion can create certain frequency components which may not even exist in the given message signal whereas linear distortion only affects the frequency components present in the message signal.

Take a typical non-linear model for example.

$$y(t) = a_1 x(t) + a_2 x^2(t)$$

Here in case x(t) is a sinusoidal function, then

a) If x(t) only has frequency component at f_1, then y(t) will have frequency components at f_1 and $2f_1$
b) If x(t) has frequency components at f_1 and f_2, then y(t) will have frequency components at f_1, f_2, $f_1 \pm f_2$, $2f_1$ and $2f_2$

Such frequency components are called inter-modulation components or harmonic components. Due to this reason, non-linear distortion is also known as **Harmonic Distortion.**

This is an extremely big issue in communication as the channel spreads the message much beyond the allotted message bandwidth B. Thus, if a different frequency range was allotted to some receiver, they would not only recover their intended message, but also the extra frequencies which are so generated causing crosstalk or what is known as co-channel interference.

Analog Modulation of Carriers

Let us now discuss the transmission of band pass signals. This is imperative because using baseband signals, there is not much we may achieve due to a couple of reasons.

1. Baseband signals have confined bandwidth, which is low pass in nature, i.e. the frequency band of the signal is symmetrical about and centred at the origin. These multiple signals cannot be transmitted over the same wired channel.
2. If we are to transmit a signal wirelessly, the wavelength of the signal must be comparable to the height of the antenna used for transmission. As the baseband signals have practically low frequency values, their wavelengths are very large. For ex. A signal with frequency 4kHz has wavelength of 75km. Thus, an antenna of height comparable to this impractical.

It is thus prudent to embed the information to be transmitted into a high frequency periodic wave called the carrier by varying one or more of its attributes, hence converting a baseband signal into a bandpass signal. This process is called **Modulation**. The need of modulation can thus be summed up with the following points.

a) Ease of Radiation: Due to frequency scaling transmission of higher frequency takes place resulting in reduced bandwidth and reduced height of the antenna for radiation.
b) Multiplexing: Multiple baseband signals centred on different carrier frequencies can be transmitted over the same line simultaneously.
c) Convenient Signal Processing: Different kinds of applications may find it more convenient to process signals in a specific frequency band in comparison to others. Thus, modulation may help is scaling the information to a frequency for which the technology is available.
d) Noise Reduction: Different frequency bands may have different characteristics of noise. Hence it is more prudent to use frequencies which offer better noise characteristics.

Typically, we can discuss various forms of Analog Modulations. Three main types are Amplitude modulation, Frequency modulation and Phase modulation. Out of these, Amplitude modulation is possible to carry out in many ways. Collectively these different types are termed as Linear Modulation Schemes.

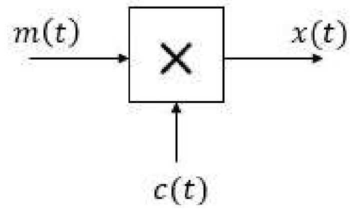
Before discussing details of different modulation types, we must define the nomenclature to be used in the discussion henceforth.

$m(t)$: Modulating Signal $\quad$ $c(t) = A_c\cos\,\omega_c t$: Carrier Signal $\quad$ $x_{<Mod.\ type}\ (t)$: Modulated Signal

$r(t)$: Intermediate Received Signal $\quad$ $y(t)$: Demodulated Signal

Linear Modulation

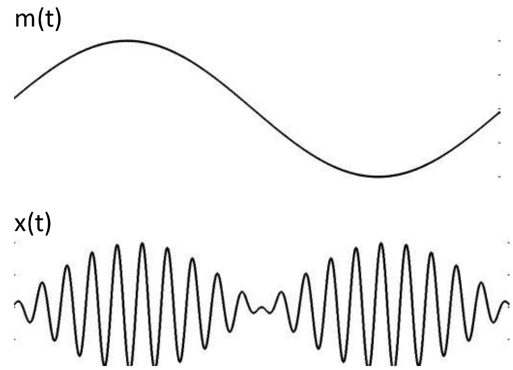
In the very basic form, modulation or frequency translation is the multiplication of the message waveform and carrier waveform. This process is called **frequency mixing** or **Heterodyning**. This mixing operation forms the basis of the study of all types of linear amplitude modulations. Hence, all linear amplitude modulations employ variations of the multiplier operations.



Types of Linear Amplitude Modulations

1) Double Side Band – Suppressed Carrier (DSB-SC)
   This system employs the simple multiplication process as mentioned earlier. The output here may thus be written as

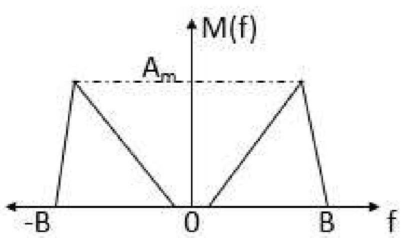
$$x(t) = m(t)c(t)$$
$$= A_c m(t) \cos \omega_c t$$

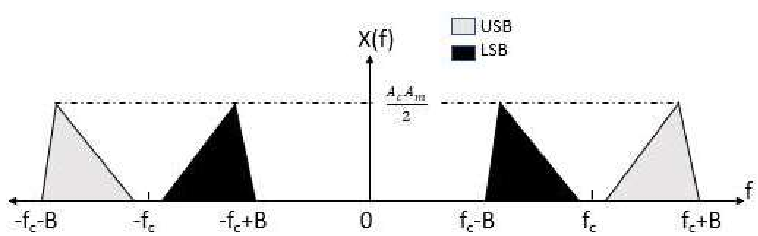


   Let us try to examine what this waveform would physically look like.
   For the sake of discussion, assume that $m(t)$ is a sinusoidal signal. At any given instant the amplitude of the modulated signal can be seen to have the value $A_c m(t)$.
   But as $A_c m(t)$ is a varying function of time, the high frequency carrier is thus scaled in amplitude within an envelope created by the function $A_c m(t)$. Note here that when the message signal $m(t)$ crosses the zero value to go from positive to negative, the modulated signal $x(t)$ faces a phase change of 180^o.

   Spectrum of DSB-SC Signal
   Assume that $m(t)$ has a spectrum (real valued) represented by $M(f)$ as shown here. Then by the modulation property

$$X_{DSBSC}(f) = \frac{A_c}{2}[M(f+f_c) + M(f-f_c)]$$





Here is what can be deducted from the spectrum of $X(f)$.

a) Bandwidth of modulated signal is 2B
   The original signal m(t) is real-valued and hence has an even symmetric magnitude spectrum. Therefore, the same information is sent in the band $0 \to B$ and in $-B \to 0$. In DSB-SC, both these bands are scaled to a higher frequency resulting in duplicity of data. As a result, the frequency bands $f_c \to f_c + B, -f_c - B \to -f_c$ and $f_c - B \to f_c, -f_c \to -f_c + B$ have the same information and both are transmitted simultaneously. Therefore, the bandwidth of the DSB-SC signal is twice that of the message signal, i.e. 2B.
   Here, the bands $\{f_c \to f_c + B\}$ & $\{-f_c - B \to -f_c\}$ are called Upper Side Bands (USB) since they represent frequency components higher than the carrier frequency. Similarly, the frequency bands $\{f_c - B \to f_c\}$ and $\{-f_c \to -f_c + B\}$ are called Lower Side Band (LSB).

   For this reason, this system is called Double Side-Band, i.e. DSB.

b) There is no explicit carrier component in the signal. This is evident as no impulse signals characterising the presence of a sinusoidal carrier at f_c and $-f_c$ are observed in the spectrum. This is the reason why this system is known as Suppressed Carrier system. The term suppressed is attributed to the way the modulation circuit suppresses the carrier component.

For this reason, this type of modulation is called Double Side Band – Suppressed Carrier.

Transmitted Power of DSB – SC

Total transmitted power would be the power of the modulated signal. Therefore,

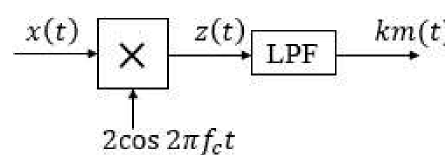
$$\begin{aligned} S_{DSBSC} &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} A_c^2 m^2(t)\cos^2 2\pi f_c t\, dt \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{A_c^2 m^2(t)}{2}(1+\cos 4\pi f_c t)dt \\ &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{A_c^2 m^2(t)}{2}dt + \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{A_c^2 m^2(t)}{2}\cos 4\pi f_c t\, dt \end{aligned}$$

In comparison to f_c, the frequencies in $m(t)$ will be very low. Therefore, as $m(t)$ is a very slowly varying signal in comparison to $\cos 2\pi f_c t$, the second part of the above-mentioned expression can be calculated assuming $m^2(t)$ as a constant with respect to $\cos 4\pi f_c t$. Hence, the value of this integration approaches 0, since the integration of a sinusoid over integral multiples of time period is always 0.

$$\begin{aligned} \Rightarrow \quad S_{DSBSC} &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{A_c^2 m^2(t)}{2}dt \\ &= \frac{A_c^2}{2}\left[\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} m^2(t)dt\right] \\ &= S_c S_m \\ where \quad S_c &= \frac{A_c^2}{2} \quad and \quad S_m = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} m^2(t)dt \end{aligned}$$

Recovery of Baseband Signals from DSB-SC / Demodulation of DSB-SC/ Coherent Demodulation

We can see intuitively, that the modulated waveform needs to be scaled back to the values of frequency centred around 0. What we can do is scale the waveform $x(t)$ again with the same carrier frequency. This would result in generation of frequency components centred around 0 and $2f_c$.



To remove the high frequency component, we may use a low pass filter. Thus, we will be able to retrieve the original message frequency scaled in amplitude by some amount.

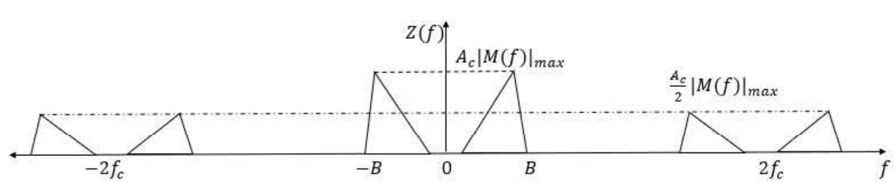
Let us mathematically analyse this system.

$$z(t) = [A_c m(t)\cos\omega_c t]\,[2\cos\omega_c t]$$

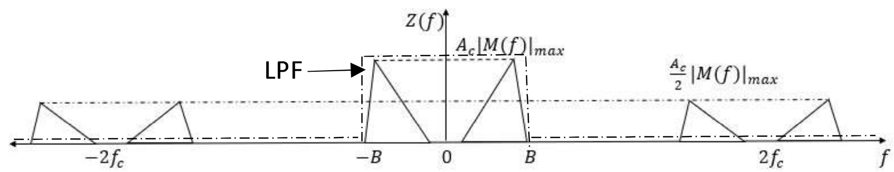
$$\begin{aligned} &= A_c m(t)(2\cos^2\omega_c t) = A_c m(t)(1+\cos 2\omega_c t) \\ &= A_c m(t) + A_c m(t)\cos 2\omega_c t \end{aligned}$$

By simple Fourier Transform, the spectrum of $Z(f)$ can be written as

$$Z(f) = A_c M(f) + \frac{A_c}{2}[M(f-2f_c) + M(f+f_c)]$$



It can be easily observed now, that the high frequency component is completely separated from the baseband spectrum. The bandwidth of the message spectrum is known to be B which is much less than $2f_c$. Therefore, a low pass filter with bandwidth slightly greater than B can be used for message recovery.



One limitation of this system is that the exact knowledge of carrier frequency and phase is needed at the receiver. Only then will the local oscillator at the receiver be able to generate the replica of the carrier waveform for demodulation. If these waveforms (original carrier and locally generated oscillation at receiver) are not identical in frequency and phase, exact demodulation will not be possible. Hence this type of demodulation is called synchronous demodulation. A more commonly used terminology, however, is **Coherent Demodulation**.

Effect of Loss of Coherence

Assume that the local carrier is generating a waveform

$$c_{LO}(t) = 2\cos[(\omega_c + \delta\omega)t + \theta]$$

In such a case, the recovered signal $y(t)$ can be written as

$$\begin{aligned} y(t) &= [A\, m(t) \cos \omega_c t][2\cos[(\omega_c + \delta\omega)t + \theta]] \\ &= A\, m(t)[2\cos(\omega_c t)\cos[(\omega_c + \delta\omega)t + \theta]] \\ &= A\, m(t)[\cos(\delta\omega\, t + \theta) + \cos[(2\omega_c + \delta\omega)t + \theta]] \end{aligned}$$

After Low Pass Filtering,

$$y(t) = A\, m(t)\cos(\delta\omega\, t + \theta)$$

Let's see the effect of this.

a) Assuming $\delta\omega = 0$ $\qquad$ $y(t) = A\, m(t)\cos\theta$
At certain conditions where, $\theta = \frac{\pi}{2}$, $y(t) = 0$. Hence $\cos\theta$ becomes an attenuation factor. Thus, phase synchronization is a must for coherent demodulation.
b) Assume $\theta = 0$ $\qquad$ $y(t) = A\, m(t)\cos(\delta\omega\, t)$
This instead of receiving only $m(t)$, we will receive a waveform which is a product of two functions of time. As $\delta\omega$ is typically a very small value, we will observe $m(t)$ to vary in an envelop defined by cosine function. Such a varying distortion is known as **Warbling Effect.**

We therefore require a carrier recovery circuit at the receiver to figure out the exact frequency and phase of the carrier which makes the demodulation design complicated and more expensive. In practice, DSB-SC is used in the following applications:

1. P2P two-way radio communication
2. Citizen Band radio
3. Air Traffic Control Radios
4. Key-less Garage Door Opener remotes
5. Colour information in Analog TV broadcast
6. Stereo information in FM radio broadcast, etc.

Special case: $m(t) = A_m \cos \omega_m t$

$$B = f_m; \quad Bandwidth = 2f_m; \quad S_m = \frac{A_m^2}{2}; \quad S_{DSBSC} = \frac{A_m^2 A_c^2}{4}; \quad S_{LSB} = S_{USB} = \frac{A_m^2 A_c^2}{8}$$

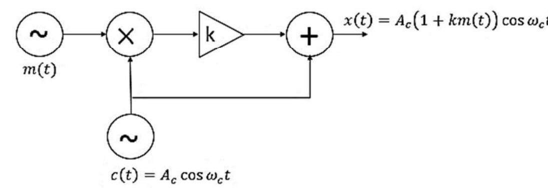
$$X(f) = \frac{A_m A_c}{4}[\delta(f + f_c + f_m) + \delta(f + f_c - f_m) + \delta(f - f_c + f_m) + \delta(f - f_c - f_m)]$$

2) Double Side Band – Full Carrier (DSB-FC)

In transmission of voice or music signals over broadcast channels, the receiving demodulators need to be as simple as possible. One of the major reasons why coherent demodulation is needed in DSB-SC is because the modulated waveform faces a 180^o phase change every time the modulating signal changes its polarity, i.e. goes from positive to negative or vice versa. This is because the multiplication process leads to the amplitude of the carrier wave to be proportional to $|m(t)|$ (refer to Hilbert Transform section).

Therefore, if we can modify the message signal to always have magnitude greater than or equal to 0, then there will be no phase shift and hence, the envelop will be the same as the modified message signal. This simplifies the demodulation process greatly.

The process of generation can be mathematically understood with the block diagram shown here.



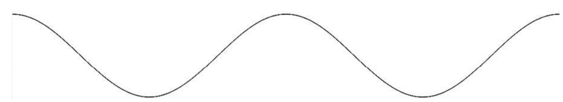
Block Diagram for DSB-FC Generation

A simple analysis of the block diagram gives the expression:

$$x_{DSBFC}(t) = A_c\big(1 + km(t)\big) \cos\omega_c t$$

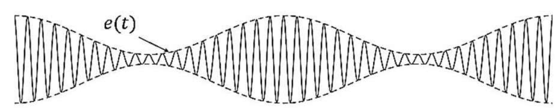
This may be written as

$$x_{DSBFC}(t) = e(t) \cos\omega_c t$$



Message Signal

where $e(t) = A_c\big(1 + km(t)\big)$ can be called the envelop function of the modulated waveform.



Modulated Wave

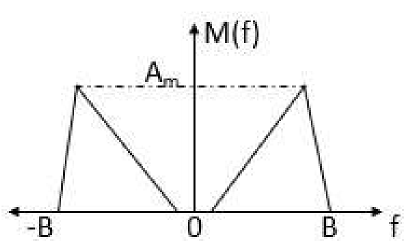
The scaling factor k and addition of carrier ensures that $e(t) \geq 0, \forall t$.

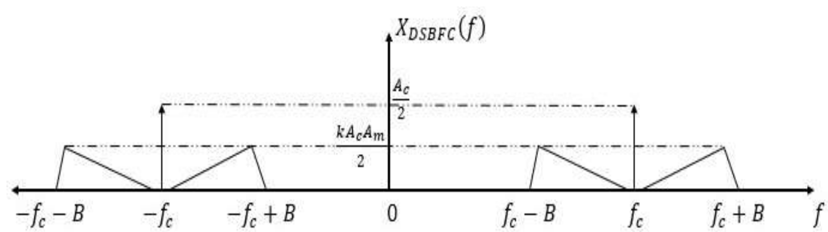
The spectrum of DSB-FC ca be analysed as follows. Re-writing the equation of modulated wave as

$$x_{DSBFC}(t) = A_c \cos\omega_c t + A_c k[m(t) \cos\omega_c t]$$

By simple Fourier Transform of the equation, the frequency domain equation ca ne be written as:

$$X_{DSBFC}(f) = \frac{A_c}{2}[\delta(f + f_c) + \delta(f - f_c)] + \frac{A_c k}{2}[M(f + f_c) + M(f - f_c)]$$





It is clearly visible that the bandwidth requirement of this system is same as DSB-SC, i.e. 2B. Furthermore, the modulated wave transmission requires some extra power since the carrier is also being sent. Since this extra power is not carrying any relevant message information, hence DSB-FC is **Power Inefficient.**

To summarise:

1. Envelop $e(t) \propto m(t)$ and $e(t) > 0$. This makes the recovery of $m(t)$ very simple.
2. Bandwidth requirement is the same as DSB-SC, i.e. 2B or double the message bandwidth.
3. DSB-FC is power inefficient as the carrier transmission requires extra power which is not being used for communication of any useful information.
4. Even though the system is bot bandwidth and power inefficient, still the easy recover and cheaper demodulation circuits reduce the cost of receiver significantly.

Discussing the attributes of the waveform further, we define a new parameter μ, which, in accordance with the value of the constant k and the message peak amplitude (say) A_m, decides the permissible extent of message scaling and hence ensures $e(t) \geq 0$, so that easier recovery is possible. This parameter μ is called **Modulation Index**, and can be defined as

$$\mu = \frac{[e(t)]_{max} - [e(t)]_{min}}{[e(t)]_{max} + [e(t)]_{min}}$$

Further, it can easily be deduced that

$$\mu = \frac{A_c(1 + km(t)_{max}) - A_c(1 + km(t)_{min})}{A_c(1 + km(t)_{max}) + A_c(1 + km(t)_{min})}$$

$$\mu = \frac{k[(m(t)_{max}) - (m(t)_{min})]}{2 + k[(m(t)_{max}) + (m(t)_{min})]}$$

Note: $m(t)$ represents the original unscaled version of the message signal. If $m(t) \nless 0$ for any t, k would simply be 1.

It is important to understand here the importance physical relevance of this parameter.

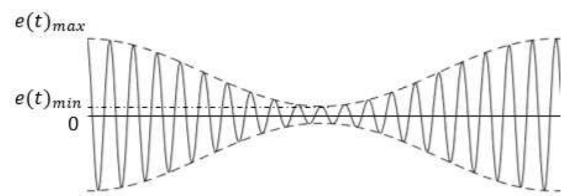
It is quite evident that the parameter k is of great importance in the design of DSB-FC. To maintain the condition of $e(t) \geq 0$, k must be chosen small. However, this also affects the power contained in the sideband which is in turn dependent on k^2. We must, therefore, use a value of k just enough to maintain the envelop condition. Ideally, we want

$$e(t) = \big(1 + km(t)\big) \geq 0 \quad \forall\, t$$
$$\Rightarrow e(t)_{min} = \big(1 + k(m(t)_{min})\big) \geq 0$$
$$\Rightarrow \quad k \leq \frac{-1}{m(t)_{min}} \quad (inequality\ reversal\ since\ m(t)_{min}\ is\ supposed\ to\ be\ negative)$$

Using this value, it can be shown that

a) $For\ k < \frac{-1}{m(t)_{min}} \quad ; \quad \mu < 1$

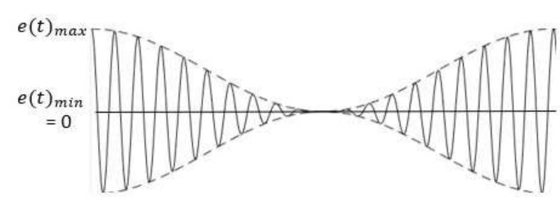
This case has a valid DSB-FC but has comparatively smaller power in the sidebands that carry the information, reducing the power efficiency of the system. This case is called Under Modulation.



$\mu < 1$

b) $For\ k = \frac{-1}{m(t)_{min}} \quad ; \quad \mu = 1$

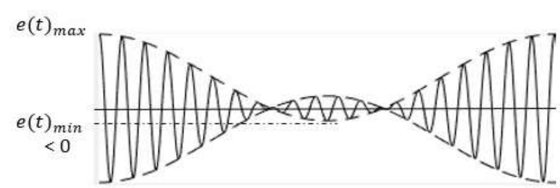
This case is the perfect DSB-FC which is theoretically still recoverable, maintaining maximum permissible power in the sidebands. This is therefore called Fully modulated or Perfectly modulated DSB-FC waveform.



$\mu = 1$

c) $For\ k > \frac{-1}{m(t)_{min}} \quad ; \quad \mu > 1$

This case invalidates the envelop condition making $e(t) \propto |m(t)|$ instead of simply $m(t)$ making simpler recovery impossible. This case is called Over Modulation.



$\mu > 1$

Since the parameter μ is now understood as a clear indicator of the extent of modulation, we may define an entity known as **Depth of Modulation** written in percentage form as $(\mu \times 100)\%$.

Transmitted Power of DSB-FC

Let's calculate the power of the final modulated waveform.

$$
\begin{aligned}
S_{DSBFC} &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \left[A_c\big(1+km(t)\big)\cos\omega_c t\right]^2 dt \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} A_c^2\left(1+k^2m^2(t)+2km(t)\right)\cos^2(\omega_c t)\ \ dt \\
&= \lim_{T\to\infty} \frac{1}{2T}\left[\int_{-T}^{T} A_c^2\cos^2(\omega_c t)\,dt + \int_{-T}^{T} A_c^2k^2m^2(t)\cos^2(\omega_c t)\,dt + \int_{-T}^{T} 2A_c km(t)\cos^2(\omega_c t)\,dt\right] \\
&= \frac{A_c^2}{2} + \frac{A_c^2}{2}\lim_{T\to\infty}\frac{1}{2T}k^2\int_{-T}^{T} m^2(t)dt
\end{aligned}
$$

if the mean of $m(t)$ is 0.

$$
\begin{aligned}
&\Rightarrow & S_{DSBFC} &= S_c + S_cS_m \\
&where & S_c &= \frac{A_c^2}{2} \\
&and & S_m &= \lim_{T\to\infty}\frac{1}{2T}k^2\int_{-T}^{T} m^2(t)\,dt
\end{aligned}
$$

The useful power out of this is only the sideband power which has the relevant information of the message signal. Hence, the power efficiency of the DSB-SC modulated wave may be calculated as

$$
\begin{aligned}
\eta &= \frac{S_cS_m}{S_c+S_cS_m} \\
&= \frac{\frac{A_c^2}{2}S_m}{\frac{A_c^2}{2}+\frac{A_c^2}{2}S_m} \\
&= \frac{S_m}{1+S_m}
\end{aligned}
$$

Special Case:

a) $\boldsymbol{m(t) = A_m\cos\omega_m t}$; Monotonic Input

$$
m(t)_{min} = -A_m \qquad ; \qquad m(t)_{max} = A_m
$$

$$
\begin{aligned}
\mu &= \frac{k\big(A_m-(-A_m)\big)}{2+k\big(A_m+(-A_m)\big)} \\
&= kA_m \\
\Rightarrow\ \ x_{DSBFC}(t) &= A_c(1+kA_m\cos\omega_m t)\cos\omega_c t \\
&= A_c(1+\mu\cos\omega_m t)\cos\omega_c t
\end{aligned}
$$

$$
\begin{aligned}
X_{DSBFC}(f) &= \frac{A_c}{2}[\delta(f+f_c)+\delta(f-f_c)] \\
&\quad + \frac{A_c\mu}{4}[\delta(f+f_c+f_m)+\delta(f+f_c-f_m)+\delta(f-f_c+f_m)+\delta(f-f_c-f_m)]
\end{aligned}
$$

$$
S_m = \frac{k^2A_m^2}{2} = \frac{\mu^2}{2} \qquad ; \qquad S_{DSBFC} = \frac{A_c^2}{2}+\frac{A_c^2}{2}\left(\frac{k^2A_m^2}{2}\right) = \frac{A_c^2}{2}+\frac{A_c^2\mu^2}{4}
$$

$$
S_{LSB} = S_{USB} = \frac{A_c^2\mu^2}{8} \qquad\qquad \eta = \frac{\frac{\mu^2}{2}}{1+\frac{\mu^2}{2}} = \frac{\mu^2}{\mu^2+2}
$$

Hence, with $\mu = 1$, the best possible power efficiency in case of a monotonic input is only 33.33%

b) $\boldsymbol{m(t) = \sum_{n=1}^{N} A_{m_n} \cos \omega_{m_n} t}$**; Multi-tone Input**

$$x_{DSBFC}(t) = A_c\left(1 + k\left(\sum_{n=1}^{N} A_{m_n} \cos \omega_{m_n} t\right)\right)\cos \omega_c t = A_c\left(1 + \sum_{n=1}^{N} \mu_n \cos \omega_{m_n} t\right)$$

If we can ensure that

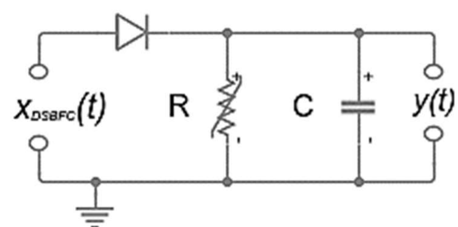
$$k\sum_{n=1}^{N} A_{m_n} = \sum_{n=1}^{N} \mu_n \le 1$$

then, the resultant wave will be a valid DSB-FC wave. In that scenario

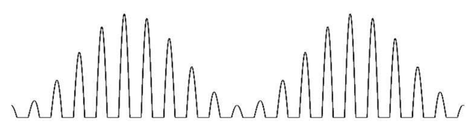
$$\begin{aligned} S_{DSBFC} &= S_c + S_c\left(S_{m_1} + S_{m_2} + \cdots + S_{m_n}\right) \\ &= \frac{A_c^2}{2}\left(1 + \frac{\mu_1^2}{2} + \frac{\mu_2^2}{2} + \cdots + \frac{\mu_n^2}{2}\right) \\ &= \frac{A_c^2}{2}\left(1 + \frac{\mu_T^2}{2}\right) \\ where \quad & \mu_T^2 = \mu_1^2 + \mu_2^2 + \cdots + \mu_N^2 \end{aligned}$$

Recovery of baseband signal from DSB-FC/ Envelop detector/ Diode Detector

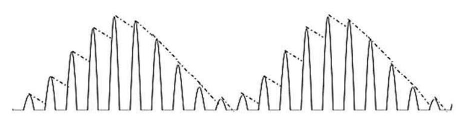
Consider the simple circuit given here. The diode acts as a rectifier and hence, only the positive half cycle of the modulated wave will be able to conduct the diode. The output of the diode will hence be a rectified waveform as shown here. Assuming the value of the resistance is large, whole of the rectified output (almost) will pass through the capacitor C. Charging of the capacitor will happen quickly and the output will almost trace back the incoming positive voltage. When the pulse begins to fall, the capacitor starts to discharge through R, but that process is slow. As charging starts again quite quickly, the output waveform roughly traces the envelop waveform as shown. This output can further be smoothened using an appropriate LPF. However, it is critical to have the correct RC discharging time constant for appropriate recovery of the signal. Let us now examine this condition.



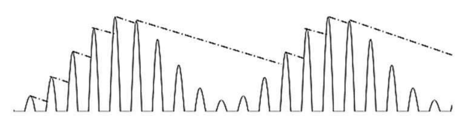
Diode Detector Circuit



Rectified Waveform



Output of the RC circuit



Peak Clipping effect due to improper choice of RC time constant

Firstly, let us examine the charging path of the capacitor. Assume that the source coupling will have a resistance R_s and the forward biased diode presents a resistance r_f. Then the charging path will be a series combination of R_s & r_f. The discharging path will be through the load resistance R. Thus, for appropriate working of this circuit, the charging time constant most be extremely small in comparison to the discharging time constant.

$$\begin{aligned} \Rightarrow \quad & RC \gg \left(R_s + r_f\right)C \\ or \quad & R \gg R_s + r_f \end{aligned}$$

Another way to look at it is that charging should occur much faster in comparison to the time period of the carrier waveform.

$$\left(R_s + r_f\right)C \ll \frac{1}{f_c}$$

Thus, charging starts as soon as positive slope in encountered.

Similarly, we can observe that discharging time constant must be very large in comparison to carrier time period.

$$RC \gg \frac{1}{f_c}$$

But what may occur here is that RC might be so large that some smaller peaks may be missed while discharging as shown before. This results in the peak clipping distortion which is knows as clipping distortion. Thus, RC must be larger that carrier time period, but must also be smaller than the inverse of message bandwidth, ensuring that the discharging will be capable of following the most sudden of changes in the message signal. Hence, the condition for appropriate detection of message signal is

$$\frac{1}{f_c} \ll RC \ll \frac{1}{B}$$

It is evident, thus, that for the recovery, an empirical knowledge of the carrier frequency is enough. In case of variable tuning for different carrier frequencies, the resistance R can be used as a potentiometer/ variable resistor. DSBF-FC is the oldest form of modulation that was invented in the late 1800's. With the advancements in technology, the scheme has lost some of its relevance. In practice, DSB-FC is still used in the following applications.

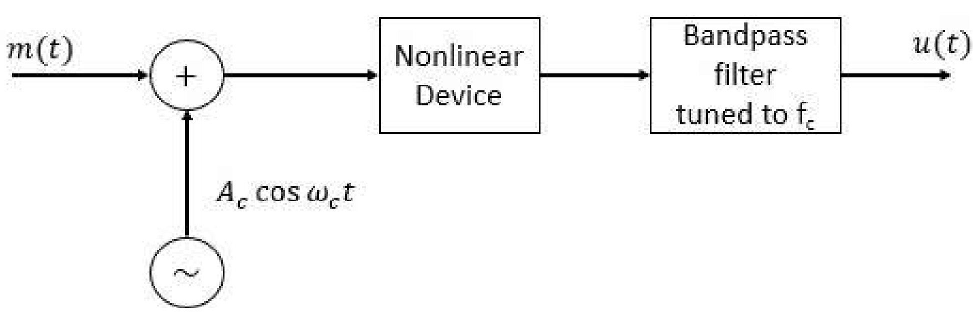
1. Broadcast Transmission: AM is still widely used for broadcasting either long, medium or short-wave bands. The received signal is simple to break down into the baseband signal and hence the equipment cost to the user is very little and it is easy to manufacture. All India Radio AM transmission is one of the examples.
2. Air band radio: The use of AM in the aerospace industry is widespread. The VHF transmissions made by airborne equipment still use AM. The radio contact between ground to air and ground to ground use AM signals.
3. Quadrature Amplitude Modulation: AM is still used in the transmission of data in pretty much everything, from short range transmission such as Wi-Fi to cellular communications, etc. QAM is formed by mixing two carriers out of phase by 90^0.

Note: Coherent Demodulation method will also work for DSB-FC under any value of μ.

Practical Design of Modulators

There are different methods of designing practical circuits for generating DSB-SC and DSB-FC waveforms. Let us discuss some of the methods commonly used in practice.

a) Power Law Modulation:
Consider a device such as PN junction diode that have non-linear input output characteristics. Suppose that the input to such a device is the sum of message signal $m(t)$ and the carrier $A_c \cos \omega_c t$. The non-linearity will generate a product of the message $m(t)$ with the carrier, plus additional terms. The desired modulated signal can be filtered out by passing the output of the nonlinear device through a band pass filter.



To elaborate, suppose that the non-linear device has an input output characteristic of the form

$$v_o(t) = a_1 v_i(t) + a_2 v_i^2(t)$$

where $v_i(t)$ is the input signal, $v_o(t)$ is the output signal and the parameters a_1, a_2 are constants. Then, if the input to the non-linear device is

$$v_i(t) = m(t) + A_c \cos \omega_c t$$

and its output is

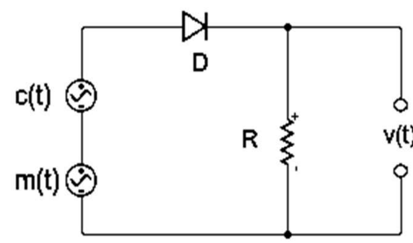
$$\begin{aligned} v_o(t) &= a_1[m(t) + A_c \cos \omega_c t] + a_c[m(t) + A_c \cos \omega_c t]^2 \\ &= a_1 m(t) + a_2 m^2(t) + a_2 A_c^2 \cos^2 \omega_c t + A_c a_1 \left[1 + \frac{2a_2}{a_1} m(t)\right] \cos \omega_c t \end{aligned}$$

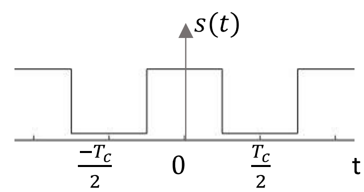
The output of the bandpass filter with bandwidth 2B centred at $f = f_c$ yields

$$u(t) = A_c a_1 \left[1 + \frac{2a_2}{a_1} m(t)\right] \cos \omega_c t$$

where $\frac{2a_2|m(t)|}{a_1} < 1$ must be true by design. Thus, the signal generated by this method will be a conventional DSB-FC signal.

b) Switching Modulator:
Another method of generating a DSB-FC modulated signal is by means of a switching modulator. Such a modulator can be implemented as shown here. The net input voltage $v_i(t)$ will be the sum of $c(t)$ and $m(t)$, i.e. carrier and message signals. With the assumption that $A_c \gg m(t) \quad \forall t$, the diode will conduct only for the positive cycles of the carrier and will be cut-off during the negative cycles. The output across the load resistor would thus be





$$v_0(t) = \begin{cases} v_i(t), & c(t) > 0 \\ 0, & c(t) < 0 \end{cases}$$

This switching operation may be viewed mathematically as a multiplication pf the input $v_i(t)$ with the switching function $s(t)$, i.e.

$$v_o(t) = [m(t) + A_c \cos \omega_c t] s(t)$$

where $s(t)$ is shown in the figure.
Since $s(t)$ is a periodic function, it is represented in the Fourier series as

$$s(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{2n-1} \cos[2\pi f_0 t(2n-1)]$$

Hence, we can re-write $v_o(t)$ as

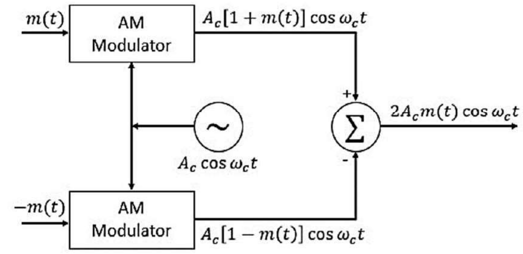
$$\begin{aligned} v_o(t) &= [m(t) + A_c \cos \omega_c t] \; s(t) \\ &= \frac{A_c}{2}\left[1 + \frac{4}{\pi A_c} m(t)\right] \cos \omega_c t + \text{other terms} \end{aligned}$$

The desired DSB-FC modulated signal is obtained by passing $v_o(t)$ through a bandpass filter with the centre frequency $f = f_c$ and bandwidth 2B. At its output, we have the desired conventional DSB-FC signal

$$u(t) = \frac{A_c}{2}\left[1 + \frac{4}{\pi A_c} m(t)\right] \cos \omega_c t$$

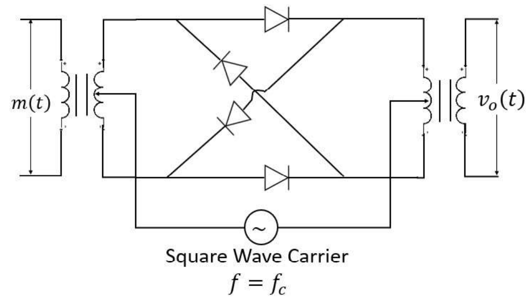
c) Balanced Modulator

A relatively simple method for generating a DSB-SC signal is to use two conventional DSB-FC modulators arranged in the configuration illustrated here. For example, we may use two square law modulators as previously described. Care must be taken to select modulators with approximately identical characteristics so that the carrier component cancels out at the summing junction.



Balanced Modulator Configuration

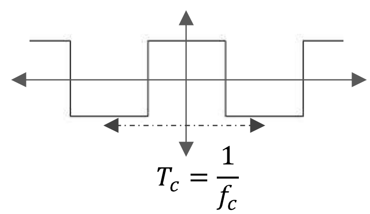
A modified arrangement of the same logic is developed as a circuit known as the **Ring modulator**



Ring Modulator Circuit

In this circuit the switching of the diodes is controlled by a square wave of frequency f_C, denoted as c(t), which is applied to the centre taps of the two transformers. When $c(t) > 0$, the top and the bottom diodes conduct, while the two diodes in the crossarms are off. In this case, the message signal $m(t)$ is multiplied by +1. When $c(t) < 0$, the diodes in the crossarms of the ring conduct, while the two diodes are switched off. In this case, the message signal $m(t)$ will be multiplied by –1. Consequently, the operation of a ring modulator can be described mathematically as a multiplier of $m(t)$ by the square wave carrier $c(t)$, i.e.

$$v_o(t) = m(t)c(t)$$



Square Carrier with f_C frequency

Since $c(t)$ is a periodic function, it is represented by Fourier Series

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

Hence, the desired DSB-SC AM signal $u(t)$ is obtained by passing $v_o(t)$ through a bandpass filter with the centre frequency f_C and the bandwidth 2B.

From this discussion, it is evident that a balanced modulator or in this case a ring modulator, basically perform the operation of multiplication of the message $m(t)$ with $A_c \cos \omega_c t$. Since this is nothing but the mixing operation required in many modulators, the balanced modulator is simply understood as a *Mixer*.

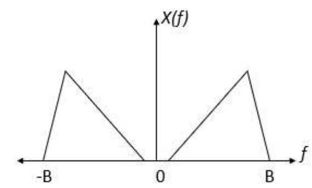
Also, if the carrier is being fed to the circuit is not through centre tapping, then due to mismatch of the inductive amplification, the carrier will not be cancelled off, hence the final output will be DSB-FC instead.

<u>Note</u>: DSB-FC is the traditionally accepted modulation scheme for Commercial Broadcast AM Radio. The carrier frequency band for this transmission scheme is a global standard and is kept in the range of 540kHz to 1600kHz.

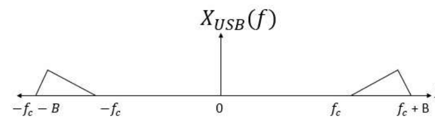
3) <u>Single Side Band – Suppressed Carrier (SSB-SC)</u>

In all the systems discussed till now, we have observed that both bandwidth and power are improperly utilized and thus none of the systems is observed to be optimally power or bandwidth efficient. All the necessary information that needs to be transmitted is found to be sufficiently carried by a single side band of the modulated signal. Despite that, we transmit both side bands to keep the design of the transmitter relatively simple and sometimes even the carrier itself to simplify the design of the receiver. However, if we are able willing to allow the design complexity (hence added cost) and can design a system that transmits only one side band of modulated signal, we will be able to achieve both better power efficiency as well as bandwidth efficiency. Such a system is often called suppressed side band modulation or Single Side Band – Suppressed Carrier modulation.

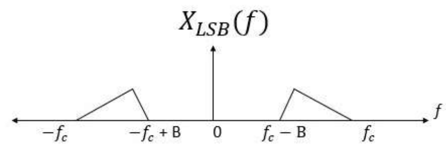
SSB-SC modulation can be achieved using 2 popularly known methods.



<u>Original Message Spectrum</u>



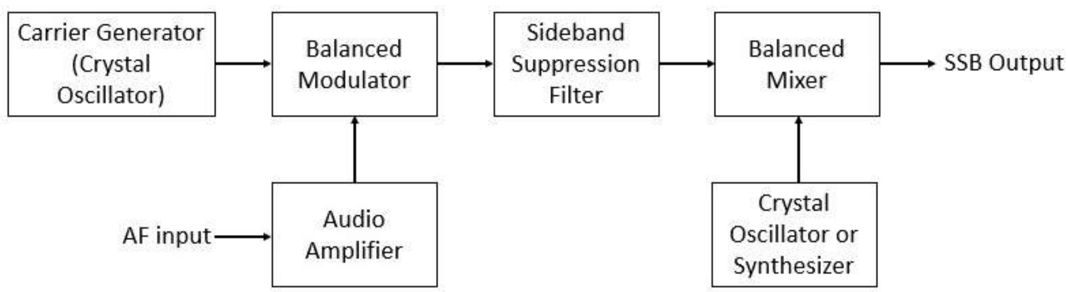
<u>Upper Sideband Spectrum</u>



<u>Lower Sideband Spectrum</u>

a) <u>The Filter Method</u>

The filter system is the simplest system of the three – after the balanced modulator, the unwanted sideband is removed (heavily attenuated) by a filter. The block diagram of a SSB transmitter employing this system is shown below.



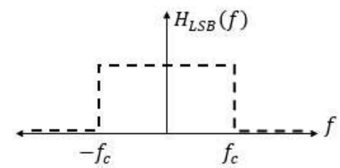
<u>Filter Method of Sideband Suppression</u>

The key circuits in this transmitter are the balanced modulator and the sideband-suppression filter. Let us examine the special considerations involving the sideband suppression filter. Such a filter must have a flat bandpass and extremely high attenuation outside the Pass band (the higher the better). Typically, in radio communication systems, the frequency range used for voice is 300Hz to 3000Hz in most cases. If it is required to suppress the lower sideband and if the transmitting frequency is f, then the lowest frequency that this filter must pass without the attenuation is $f + 300$ Hz, whereas the highest frequency that must be fully attenuated is $f - 300$ Hz. In other words, the filter's response must change from zero to attenuation to full attenuation over a range of only 600 Hz. If the transmitting frequency much above 10 MHz, this is virtually impossible. The situation becomes even worse if lower modulating frequencies are employed, such as the 50 Hz minimum in AM broadcasting (Music transmission). In order to obtain a filter response as suggested above, the Q of the tuned circuits used must be very high. As the transmitting frequency is raised, so must the Q be raised, until a situation is reached where the necessary Q is so high that there is no practicable method of achieving it.

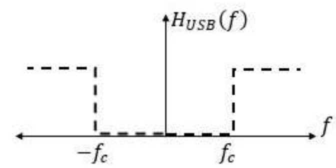
All possible designs of such sideband filters have the same disadvantage – the maximum operating frequency is below the usual transmitting frequencies. This is a reason for the balanced mixer as shown in the figure above. It is very much like a balanced modulator, except that the sum frequency is much farther from the crystal oscillator frequency than the USB was from the carrier, so that it can be selected with a tuned circuit. In this mixer, the frequency of the crystal oscillator or synthesizer is added to the SSB signal from the filter, the frequency thus being raised to the value desired for transmission. Such an arrangement also allows the transmitter to be tuneable. If the transmitting frequency is much higher than the operating frequency of the sideband filter, then two stages of mixing will be required.

b) Phase Shift Method

Because of the inherent disadvantages of the physical filters, it is prudent to find a way to design the sideband filter with required characteristics. Intuitively, we can assume that the sideband filter will take the shape of an ideal Low pass filter with cut off frequency f_c, in case we want the Lower Sideband Spectrum as shown before. Similarly, for upper sideband transmission, the sideband filter would take the shape of an ideal High pass filter.



LSB Filter Response



USB Filter Response

Note: Practically, LPF and HPF of such high frequency are not generally used. Instead, we may use a Bandpass Filter tuned to the LSB or the USB as per requirement. Secondly, we require filters with highly idealistic response as the other sideband must be completely removed (heavily attenuated).

We shall now deduce how such filter shapes can be achieved. Let us discuss the LSB Case. The LPF sideband filter can be designed easily by using signum functions.

$$H_{LSB}(f) = \frac{1}{2}\left(sgn(f + f_c) - sgn(f - f_c)\right)$$

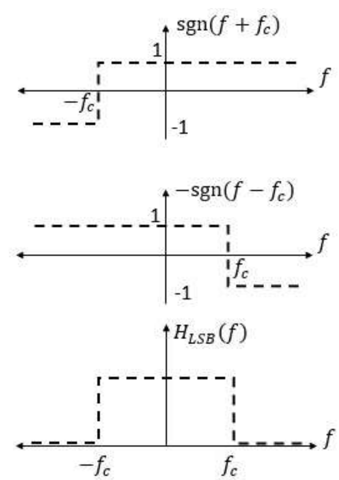
This representation gives us a way of defining how the SSB signal would behave in time domain and thus we may even define a way of better implementation of SSB modulation.

The overall expression of SSB spectrum can be represented as a product of the DSB-SC spectrum and $H_{LSB}(f)$.

$$X_{DSBSC}(f) = \frac{1}{2}A_c[M(f + f_c) + M(f - f_c)]$$

$$H_{LSB}(f) = \frac{1}{2}[sgn(f + f_c) - sgn(f - f_c)]$$

$$X_{LSB}(f) = X_{DSBSC}(f) \cdot H_{LSB}(f)$$



LPF design Method

$$X_{LSB}(f) = \frac{1}{4}A_c[M(f + f_c)sgn(f + f_c) + M(f - f_c)sgn(f + f_c)$$
$$-M(f + f_c)sgn(f - f_c) - M(f - f_c)sgn(f - f_c)]$$

Note (from the figure) that $sgn(f + f_c)$ is 1 for the case of positive part spectrum and $-sgn(f - f_c)$ is 1 for the negative part of the spectrum. Hence, the 2nd and 3rd terms in the above equation simply reduce to $M(f - f_c)$ and $M(f + f_c)$ respectively. The overall equation can be rewritten as

$$X_{LSB}(f) = \frac{1}{4}A_c[M(f - f_c) + M(f + f_c)] + \frac{1}{4}A_c[M(f + f_c)sgn(f + f_c) - M(f - f_c)sgn(f - f_c)]$$

$$\Rightarrow X_{LSB}(f) = \frac{1}{2}X_{DSBSC}(f) + \frac{1}{4}A_c[M(f + f_c)sgn(f + f_c) - M(f - f_c)sgn(f - f_c)]$$

To understand the behaviour of this spectrum in time domain, we first need to understand how to take the Fourier Inverse transform for the second part. The first part is already known to have Fourier inverse as $\frac{1}{2}m(t) \cdot c(t)$.

To understand the second term in the spectrum equation, let us use the knowledge of Hilbert transform.

$$\widehat{m}(t) \leftrightarrow -jsgn(f)M(f)$$

Also, recall the frequency translation theorem of Fourier Transform.

$$m(t)e^{\pm j2\pi f_c t} \leftrightarrow M(f \mp f_c)$$

Therefore,

$$\widehat{m}(t)e^{\pm j2\pi f_c t} \leftrightarrow -jM(f \mp f_c)sgn(f \mp f_c)$$

Hence, the Fourier Inverse of the second term can be written as

$$\mathcal{F}^{-1}\left\{\frac{1}{4}A_c[M(f+f_c)sgn(f+f_c) - M(f-f_c)sgn(f-f_c)]\right\}$$

$$= -A_c\frac{1}{4j}\widehat{m}(t)e^{-j2\pi f_c t} + A_c\frac{1}{4j}\widehat{m}(t)e^{j2\pi f_c t}$$

$$= \frac{A_c\widehat{m}(t)}{2}\left[\frac{e^{j2\pi f_c t} - e^{-j2\pi f_c t}}{2j}\right]$$

$$= \frac{1}{2}\widehat{m}(t)A_c\sin(2\pi f_c t)$$

$$= \frac{1}{2}\widehat{m}(t) \cdot \hat{c}(t)$$

Therefore, the LSB modulated equation can be written as

$$x_{LSB}(t) = \frac{1}{2}[m(t) \cdot c(t) + \widehat{m}(t) \cdot \hat{c}(t)]$$

Similarly, HPF filter for the Upper sideband filtering can be written as

$$H_{USB}(f) = 1 - \frac{1}{2}[sgn(f+f_c) - sgn(f-f_c)]$$

Using the analysis done before, the spectrum for USB can be written as:

$$X_{USB}(f) = X_{DSBSC}(f) - \frac{1}{2}\left\{X_{DSBSC}(f) + \frac{1}{2}A_c[M(f+f_c)sgn(f+f_c) - M(f-f_c)sgn(f-f_c)]\right\}$$

$$\Rightarrow X_{USB}(f) = \frac{1}{2}X_{DSBSC}(f) - \frac{1}{4}A_c[M(f+f_c)sgn(f+f_c) - M(f-f_c)sgn(f-f_c)]$$

In time domain, the USB modulated equation can be written as

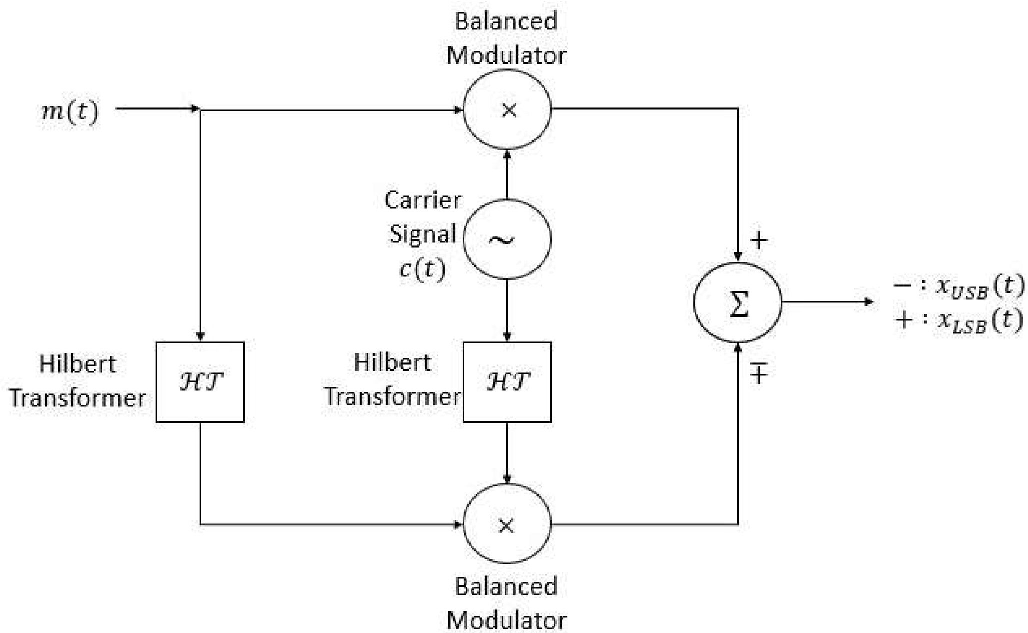
$$x_{USB}(t) = \frac{1}{2}[m(t) \cdot c(t) - \widehat{m}(t) \cdot \hat{c}(t)]$$

These equations give us an alternate scheme of SSB implementation known as the **Phase Shift Method.** The difficulty here has been passed on from the design of a high-quality band-pass filter to the design of an ideal Hilbert Transformer that will give a $-90°$ phase shift to all the frequencies in the signal bandwidth. Although such a design is still quite difficult to implement, this is much simpler than designing a high Q band pass filter. In practice, both designs are used.

Note: In general, SSB may be difficult to visualize for an arbitrary m(t) but it may be done for certain type of signals, for e.g. m(t) being a cosine wave. In that case

$$x_{LSB}(t) = \frac{1}{2}[A_m A_c \cos\{(\omega_c - \omega_m)t\}] \qquad and \qquad x_{USB}(t) = \frac{1}{2}[A_m A_c \cos\{(\omega_c + \omega_m)t\}]$$

This, obviously, is a single sinusoid without any amplitude variations. It is prudent to note that for different message signals this behaviour changes and we observe amplitude variations w.r.t to the message signal.



Block Diagram for Phase Shift Method

Recovery of baseband signal from SSB-SC

There are two methods of SSB-SC demodulation.

a) Coherent Demodulation

As discussed earlier, coherent demodulation method can be directly used for demodulation. It can be illustrated as follows.

$$\begin{aligned} r(t) &= x(t) \cdot \cos \omega_c t \\ &= \frac{A_c}{2}[m(t) \cos \omega_c t \mp \hat{m}(t) \sin \omega_c t] \cdot \cos \omega_c t \\ &= \frac{A_c}{2}[m(t) \cos^2(\omega_c t) \mp \hat{m}(t) \sin(\omega_c t) \cos(\omega_c t)] \\ &= \frac{A_c}{4}[m(t)\{1 + \cos(2\omega_c t)\} \mp \hat{m}(t) \sin 2\omega_c t] \\ &= \frac{A_c}{4} m(t) + \frac{A_c}{4}[m(t) \cos 2\omega_c t \mp \hat{m}(t) \sin 2\omega_c t] \end{aligned}$$

After the LPF blocks the higher frequency components, the resultant wave would simply be

$$y(t) = \frac{A_c}{4} m(t)$$

Consider now that there is a phase incoherence, i.e. local carrier generates $\cos\big(\omega_c t + \theta(t)\big)$ where $\theta(t)$ is a time varying phase function that can affect both phase and frequency performance. In that case

$$\begin{aligned} r(t) &= x(t) \cos\big(\omega_c t + \theta(t)\big) \\ &= \frac{1}{2} A_c\big[m(t) \cos(\omega_c t) \cos\big(\omega_c t + \theta(t)\big) \mp \sin(\omega_c t) \cos\big(\omega_c t + \theta(t)\big)\big] \\ &= \frac{1}{2} A_c m(t) \left[\frac{\cos\big(2\omega_c t + \theta(t)\big) + \cos \theta(t)}{2}\right] \mp \frac{1}{2} A_c \hat{m}(t) \left[\frac{\sin\big(2\omega_c t + \theta(t)\big) - \sin \theta(t)}{2}\right] \end{aligned}$$

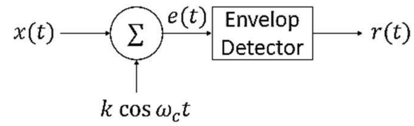
Again, LPF blocks the high frequency signals giving us

$$y(t) = \frac{A_c}{4}[m(t) \cos \theta(t) \pm \hat{m}(t) \sin \theta(t)]$$

Therefore, when $\theta(t) = 0$ then $y(t) = \frac{A_c}{4} m(t)$ i.e. perfect replica. For any other value of $\theta \neq 0$, this will be a highly distorted signal. Thus, SSB will be even more sensitive to phase or frequency incoherence that DSB case.

b) Carrier Reinsertion Method of Demodulation

In this method, we use the local carrier generator and add this carrier wave to the received signal. Under certain conditions, this sum can result in the output having an envelop of the shape of $m(t)$. Then we can use simply an envelop detector to get the desired output.



Carrier Reinsertion Block Diagram

As shown in the figure,

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

This representation can be written in the form

$$e(t) = R(t) \cos\big(\omega_c t + \theta(t)\big)$$

where,

$$R(t) = \sqrt{\left(\frac{1}{2}A_c m(t) + k\right)^2 + \left(\frac{1}{2}A_c \hat{m}(t)\right)^2}$$

and

$$\theta(t) = \pm \tan^{-1}\left[\frac{\frac{1}{2}A_c \hat{m}(t)}{\frac{1}{2}A_c m(t) + k}\right]$$

If we want R(t) to roughly follows m(t), then we'll have to assume that

$$\frac{1}{2}A_c m(t) + k \gg \frac{1}{2}A_c \hat{m}(t)$$

which can be done by assuming a large k. In that case,

$$R(t) \cong \frac{1}{2}A_c m(t) + k$$

which will be the required envelop proportional to m(t).

It is important to note that even in this case, frequency and phase coherence in important. But it is essentially more convenient in certain conditions to implement or handle summing operation than mixing operation. For e.g. speech signals have the convenience which makes the use of this demodulator convenient to get the right quality of m(t). Note that SSB may be used to transmit speech but never to transmit music.

Transmission Power off SSB-SC signal:

Total transmitted power of the modulated signal would be

$$S_{SSBSC} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \left[\frac{1}{2}A_c\{m(t)\cdot \cos \omega_c t \pm \hat{m}(t)\cdot \sin \omega_c t\}\right]^2 dt$$

$$= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{1}{4}A_c^2[m^2(t) \cos^2 \omega_c t + \hat{m}^2(t) \sin^2 \omega_c t \pm m(t)\hat{m}(t) \sin 2\omega_c t]dt$$

$$= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{1}{4}A_c^2 \left[m^2(t)\left\{\frac{(1 + \cos 2\omega_c t)}{2}\right\} + \hat{m}^2(t)\left\{\frac{1 - \cos 2\omega_c t}{2}\right\} \pm m(t)\hat{m}(t) \sin 2\omega_c t\right] dt$$

$$= \frac{1}{4}A_c^2 \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \frac{1}{2}[m^2(t) + \hat{m}^2(t)]\, dt$$

since $m^2(t), \hat{m}^2(t)$ *and* $m(t)\hat{m}(t)$ are all too slow varying and $\cos 2\omega_c t$ & $\sin 2\omega_c t$ integrated over $-T$ *to* T reduce to 0. Also, by property of $m(t)$ & $\hat{m}(t)$ have the same energies.

$$\Rightarrow S_{SSBSC} = \frac{1}{4}A_c^2 \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} m^2(t)dt$$

$$\Rightarrow S_{SSBSC} = \frac{1}{2} S_c S_m$$

$$where \quad S_c = \frac{1}{2} A_c^2 \quad and \quad S_m = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} m^2(t)dt$$

∴ The net power content of the SSB-SC signal is half of the power content of the DSB-SC.

Special case: $m(t) = A_c \cos \omega_c t$

$$m(t) = A_m \cos \omega_m t \quad , \quad c(t) = A_c \cos \omega_c t \quad and \quad \hat{m}(t) = A_m \sin \omega_m t \quad , \quad \hat{c}(t) = A_c \sin \omega_c t$$

$$x_{LSB}(t) = \frac{A_m A_c}{2} [\cos \omega_m t \cdot \cos \omega_c t + \sin \omega_m t \cdot \sin \omega_c t] = \frac{A_m A_c}{2} \cos[(\omega_c - \omega_m)t]$$

$$x_{USB}(t) = \frac{A_m A_c}{2} [\cos \omega_m t \cdot \cos \omega_c t - \sin \omega_m t \cdot \sin \omega_c t] = \frac{A_m A_c}{2} \cos[(\omega_c + \omega_m)t]$$

$$S_{SSBSC} = S_{LS} = S_{USB} = \frac{A_m^2 A_c^2}{8}$$

$$X_{LSB}(f) = \frac{A_m A_c}{4} [\delta(\omega + \omega_c - \omega_m) + \delta(\omega - \omega_c + \omega_m)]$$

$$X_{USB}(f) = \frac{A_m A_c}{4} [\delta(\omega + \omega_c + \omega_m) + \delta(\omega - \omega_c - \omega_m)]$$

In practice the SSB-SC modulation is used in the applications where power saving and low bandwidth is needed. For e.g.:

1. Wireless communications like Amateur Radio used for military applications
2. Radar Communication
3. Point to Point telephonic communication systems

4) Vestigial Side Band

We may now turn to the question of transmitting the video signals required for the proper reception of television, noting that the bandwidth occupied by such signals is at least 4MHz. If we use DSB schemes, the required transmission bandwidth would be around 9-10MHz which is not practical. The use of a form of SSB is clearly indicated here to ensure spectrum conservation. To simplify video demodulation in the receiver, the carrier is, in practice, sent undiminished. Because of the phase response of filters, near the edges of the flat bandpass, would have a harmful effect on the received video signals in a TV receiver, a portion of the unwanted (lower) sideband must also be transmitted. The result is vestigial sideband transmission.

Essentially, the modulation system is the same as the Filter method of SSB generation. The SSB filter is simply replaced with a VSB filter. It is important to note that we would like for the coherent demodulation to work for this modulation scheme as well. Hence, that can give us a condition on the VSB filter design.

We know that

$$\begin{aligned} X_{VSB}(f) &= X_{DSBSC}(f) \cdot H_{VSB}(f) \\ &= \frac{A_c}{2} [M(f + f_c) + M(f - f_c)] \cdot H_{VSB}(f) \end{aligned}$$

Also, at the coherent demodulator

$$R(f) = \frac{1}{2} [X_{VSB}(f + f_c) + X_{VSB}(f - f_c)]$$

From these two equations, we can rewrite the equation of $R(f)$ as

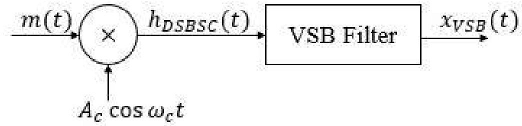
$$R(f) = \frac{A_c}{4} [\{M(f - 2f_c) + M(f)\}H(f - f_c) + \{M(f + 2f_c) + M(f)\}H(f + f_c)]$$

$$\Rightarrow R(f) = \frac{A_c}{4} M(f)[H(f - f_c) + H(f + f_c)] + \frac{A_c}{4} [M(f - 2f_c)H(f - f_c) + M(f + 2f_c)H(f + f_c)]$$

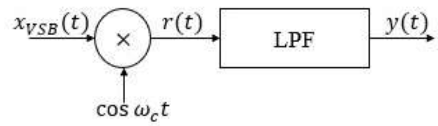
The second half of the equation would be centred around $2f_c$ and would be simply filtered out by the LPF centred around origin.

Hence, the overall received signal $Y(f)$, would be

$$Y(f) = \frac{A_c}{4} M(f)[H(f - f_c) + H(f + f_c)]$$



VSB Modulator



VSB Demodulator

For the final output to be the same as or, at best, a scaled version of the message signal, we get a condition on the sideband filter transfer function.

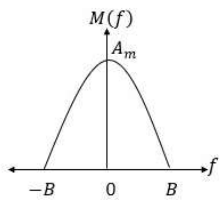
$$H(f - f_c) + H(f + f_c) = c$$

where ‘c’ is a constant. For convenience, we can assume it to be 1 making the equation

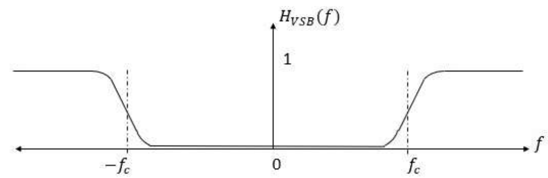
$$H(f - f_c) + H(f + f_c) = 1$$

This condition is called **Vestigial Symmetry**.

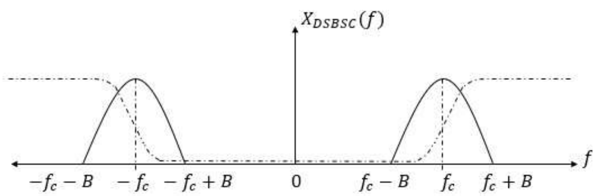
The same can be highlighted by the diagrammatic treatment of the same derivation as shown below.



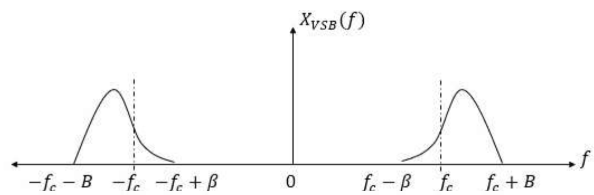
a) Message Spectrum



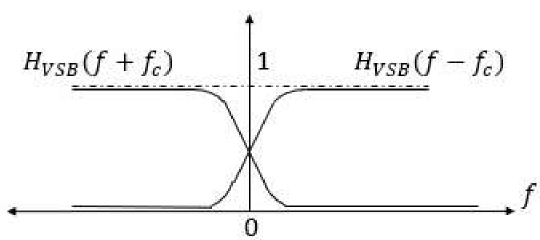
b) Vestigial Sideband filter



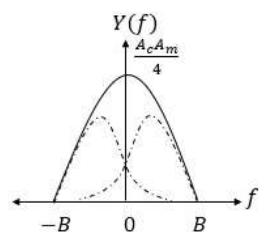
c) Application of VSB Filter on DSBSC spectrum



d) Vestigial Sideband Spectrum



e) Vestigial Symmetry



f) Received Signal Spectrum

As is evident, the bandwidth of the modulated signal would be $B + \beta$.

Representation of VSB Signal

VSB signal, in general, is also a band pass signal just like an SSB signal. Thus, it will also have an in-phase and a quadrature phase component.

$$x_{VSB}(t) = s_I(t) \cos 2\pi f_c t - s_Q(t) \sin 2\pi f_c t$$

So now, we need to figure out $s_I(t)$ and $s_Q(t)$.

Let us try to figure out how $s_I(t)$ can be recovered. Simple coherent detection w. r. t. $\cos 2\pi f_c t$ will result in the detection of $s_I(t)$ completely. In such a case, the spectrum of $s_I(t)$ can be written as

$$S_I(f) = \begin{cases} X(f - f_c) + X(f + f_c) & \forall \ \ |f_c| < B \\ 0 & \forall \ \ |f_c| > B \end{cases}$$

Again, recall that we had written $X(f)$ as

$$X(f) = \frac{A_c}{2} [M(f - f_c) + M(f + f_c)] H_{VSB}(f)$$

Hence, $S_I(f)$ can be rewritten as

$$S_I(f) = \begin{cases} \frac{A_c}{2} M(f)[H(f - f_c) + H(f + f_c)] & \forall\, |f_c| < B \\ 0 & \forall\, |f_c| > B \end{cases}$$

Also, knowing that,

$$H(f - f_c) + H(f + f_c) = 1$$
$$\Rightarrow \quad S_I(f) = \frac{A_c}{2} M(f)$$
$$or \quad s_I(t) = \frac{A_c}{2} m(t)$$

Hence, from our evaluation as done above and our observations on equations of SSB, it can simply be said that the in-phase component of the signal remains the same for both the cases of sideband suppression techniques.

Let us now try to figure out $s_Q(t)$. For this we need to do coherent demodulation w. r. t. $\sin 2\pi f_c t$. In that case we may observe that

$$S_Q(f) = \begin{cases} j[X(f - f_c) - X(f + f_c)] & |f_c| < B \\ 0 & |f_c| > B \end{cases}$$

$$S_Q(f) = LPF\left[\begin{cases} j\frac{A_c}{2}[\{M(f - 2f_c) + M(f)\}H(f - f_c) - \{M(f + 2f_c) + M(f)\}H(f + f_c)] & \forall |f| < B \\ 0 & \forall |f| > B \end{cases}\right]$$

$$S_Q(f) = j\frac{A_c}{2} M(f)[H(f - f_c) - H(f + f_c)]$$

Interpretation of this result is that the quadrature component can be recovered using a sideband filter with transfer function

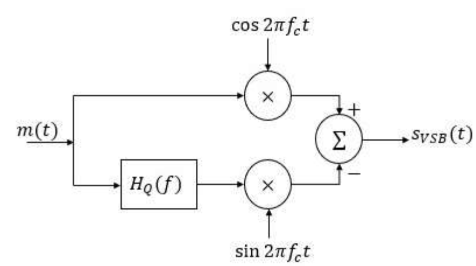
$$H_Q(f) = j[H(f - f_c) - H(f + f_c)]$$

We may define $m'(t)$ as

$$m_Q(t) = m(t) \circledast \mathcal{F}^{-1}[H_Q(f)]$$

Therefore,

$$s_Q(t) = \frac{A_c}{2}[m(t)\cos 2\pi f_c t - m_Q(t)\sin 2\pi f_c t]$$



VSB Modulation (Alternate Method)

Hence, the alternate implementation of the VSB system may be done as shown in the figure above.

Note: As $\beta \to 0$, $H_{VSB}(f) \to Hilbert\ Transform$, $m_Q(t) \to \hat{m}(t)$ and $VSB \to SSB$