Table of Contents

Syllabus

Course Catalog

3 Credit hours (3 h lectures). Signal Analysis: equivalent low-pass and band-pass models, Hilbert Transform, power spectral density. Amplitude modulation and demodulation: large carrier and suppressed carrier, single side band, and vestigial side band, coherent and non-coherent detection; Angle modulation and demodulation: FM and PM, wide band and narrow band FM, transmission bandwidth, generation and demodulation of FM. Noise representation and analysis: SNR analysis of AM and FM systems. Pulse modulation techniques: sampling theorem, PAM, PPM, PWM, PCM, Delta Modulation.

Textbook

John G. Proakis and Masoud Salehi (2002), *Communication Systems Engineering*, 2nd ed., Prentice-Hall.

References

1. Simon Haykin (2001). *Communication Systems,* 4th ed. Wiley.
2. Wayne Tomasi (2001). *Electronic Communications Systems, Fundamentals through Advanced*. 4th ed. Prentice Hall.
3. R. E. Ziemer and W. H. Traner (1995). *Principles of Communications*. 4th ed. Wiley.
4. Leon Couch II (2001). *Digital and Analog Communication Systems*. 6th ed. Prentice Hall.
5. A. B. Carlson (1986). *Communication Systems*. 3rd ed. McGraw-Hill.

Instructor

Instructor: Dr. Mohammad M. Banat

Email Address: banat@just.edu.jo

Prerequisites

Prerequisites by topic Signal Analysis, Random Signal Analysis

Prerequisites by course EE 260, EE 360

Prerequisite for EE 551

Topics Covered

Week	Topics	Chepters in Text
1	Introduction	0
2-3	Review of Signal Analysis	Apx 2
4	Lowpass Signal Representation	Apx 2
5-8	Amplitude Modulation	2
9-11	Angle Modulation	2
12	Pulse Modulation	3
13-15	Review of Random Signal Analysis	1
16	Noise in Communication Systems	2

Evaluation

Assessment Tool	**Expected Due Date**	**Weight**
Mid-Term Exam	Sat. 7 August 2021	25%
Class Work		25%
Final Exam		50%

Objectives and Outcomes

Objectives	*Outcomes*
1. Ability to analyze signals and systems in time and frequency domains [1]	1.1. Understanding Fourier transform and Fourier series for signals and systems and its properties [1]. 1.2. Introducing band-pass models and the low-pass equivalent [1]. 1.3. Studying Hilbert Transform [1]. 1.4. Analyzing the transmission of signals through linear systems [1].
2. Understanding and analyzing different types of amplitude modulated signals [1]	2.1. Defining amplitude modulation (AM) and its forms [1]. 2.2. Differentiating between the double sideband-suppressed carrier, single sideband, and vestigial sideband modulations [1]. 2.3. Understanding the generation and detection of AM signals [1]. 2.4. Understanding frequency translation and frequency-division multiplexing [1].
3. Understanding and analyzing different types of angle modulated signals [1]	3.1. Defining angle modulation [1]. 3.2. Explaining different forms of frequency modulation (FM) (i.e., Narrow-band FM and Wide-band FM) [1]. 3.3. Understanding the generation and detection of FM signals [1]. 3.4. Understanding the phase-lock loop [1]. 3.5. Studying nonlinear effects in FM signals [1]. 3.6. Introducing the superheterodyne receiver [1].
4. Understanding sampling theorem and pulse modulation techniques [1]	4.1. Appreciating the need for digitizing analogue signals [1]. 4.2. Understanding the sampling theorem [1]. 4.3. Differentiating between pulse-amplitude modulation and pulse-position modulation [1]. 4.4. Introducing time division multiplexing [1]. 4.5. Introducing the quantization process [1]. 4.6. Appreciating the importance of pulse-code modulation [1]. 4.7. Introducing delta modulation [1].
5. Ability to evaluate the performance of modulated signals in the presence of additive white Gaussian noise [1,2]	5.1. Understanding the effects noise in different AM receivers and FM receivers [1,2]

Contribution of Course to Meeting the Professional Component

The course contributes to building the fundamental basic concepts, applications, and design of communication systems in electrical Engineering.

0-Evaluation

Relationship to Program Outcomes (%)

1	2	3	4	5	6	7
90	10					

I. Introduction to Communications

A communication system is used to transmit information from one or more sending parties to one or more receiving parties.

The purpose of a communication system is to deliver an information message signal from an information source in recognizable form to a user destination, with the source and the user being physically separated from each other.

Figure I.1: Communication system

Transmitter:	***Transmits data***
Channel:	***Transmission Medium***
Receiver:	***Receives Data***

I.1.A. Transmitter

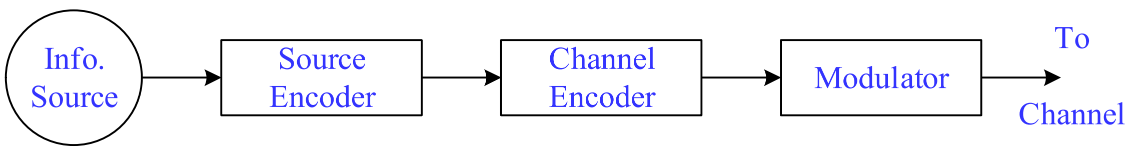


Figure I.2: Transmitter

Information Source:	**Generates data to be transmitted,** ***usually in the form of symbols from a finite alphabet.***
Source Encoder:	**Removes redundancy from the source symbol stream,** ***in order to reduce the required transmission bit rate.***
Channel Encoder:	**Adds controlled redundancy to the source encoder symbol stream,** ***in order to improve system error rate performance.***
Modulator:	**Maps the symbol stream into a finite set of signal waveforms.** ***Each different symbol is assigned a different waveform that is transmitted during the duration of the symbol.***

I.1.B. Receiver

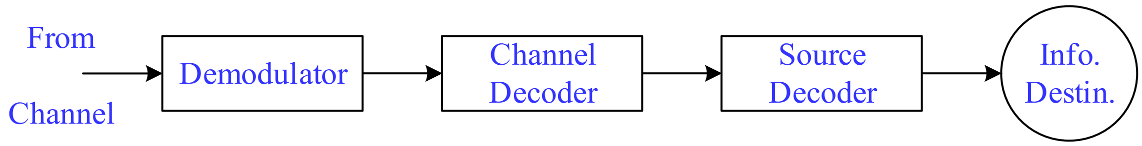


Figure I.3: Receiver

Demodulator:	**Re-maps the received waveforms into an encoded symbol stream.** ***Some symbol errors may occur due to channel impairments. Receiver performance is usually measured by the error rate.***
Channel Decoder:	**Removes the redundancy added by the channel encoder.** ***Some decoding errors may occur due to demodulator errors.***
Source Decoder:	***Recovers part (lossy) or all (lossless) of the redundancy removed by the source encoder.***
Information Destination:	**Restores original form of transmitted information. Restoration is usually imperfect due to demodulation/decoding errors.**

I.1.C. Channel

Channel Types

Linear	***Nonlinear***
Time-Invariant	***Time-Varying***
Distorting	***Non-Distorting***
Stochastic	***Deterministic***
Wideband	***Narrowband***

Channel Impairments

Attenuation	***Noise***
Distortion	***Interference***
Multipath and Fading	***Jamming***
Nonlinearities	

Transmission over Communication Channels

Guided Transmission	***Unguided Transmission***
Wires	***Radio/Wireless***
Cables	***Lightwave***
Waveguides	

0-Objectives and Outcomes

OPTICAL FIBERS

I.2. Communication Resources

A transmitted communication signal consumes power and occupies bandwidth. Using higher transmit power results in better transmission reliability. Allocating more transmission bandwidth results in better transmission quality. A general system design objective is to use these power and bandwidth resources as efficiently as possible.

In most communication channels, one resource may be considered more important than the other. We may therefore classify communication channels as power limited or band limited. For example, the telephone circuit is a typical band-limited channel, whereas a space communication link or satellite channel is typically power limited.

I.2.A. POWER

The transmitted power is the average power of the transmitted signal. Increasing the transmitted signal power generally improves the system performance. However, the cost of power consumption has to be taken into account.

> Some communication systems perform better than other systems using the same average transmitted power.

I.2.B. BANDWIDTH

The channel bandwidth is defined as the band of frequencies allocated for the transmission of the message signal. Channels with wider bandwidths generally carry more information. However, the cost of licensing bandwidth has to be taken into account. Sometimes the demanded bandwidth is not available at any cost, due to bandwidth allocation by local and international organizations.

> Some communication systems perform better than other systems using the same bandwidth.

I.3. Modulation

To achieve the goal of information transmission from the transmitter to the receiver, the transmitter modifies the message signal into a form suitable for transmission over the channel. This modification is achieved by means of a process known as modulation, which involves varying some parameter of a carrier wave in accordance with the message signal. The receiver recreates the original message signal from a degraded version of the transmitted signal after propagation through the channel. This recreation is accomplished by using a process known as demodulation, which is the reverse of the modulation process used in the transmitter.

One important way to make the signal suitable for transmission over the channel is to carry the low-frequency signal on a higher-frequency carrier signal. To see how this can be done, imagine a 4 kHz audio signal to be transmitted over a wireless channel. The transmitting antenna height h is usually a faction, say a quarter, of the signal wavelength λ. The antenna height should then be

$$\begin{aligned} h &= \frac{\lambda}{4} \\ &= \frac{c/f}{4} = \frac{3\times10^{8}/4\times10^{3}}{4} \\ &= 18.75 \text{ km !!!} \end{aligned} \tag{I.1}$$

However, with a 4 MHz carrier signal frequency, the antenna height becomes

$$\begin{aligned} h &= \frac{3\times10^{8}/4\times10^{6}}{4} \\ &= 18.75 \text{ m} \end{aligned} \tag{I.2}$$

Clearly enough, the first antenna height is too far from practical, while the second one is reasonably practical, especially for broadcasting stations.

Owing to the unavoidable presence of noise and distortion in the received signal, we find that the receiver cannot recreate the original message signal exactly. The resulting degradation in overall system performance is influenced by the type of modulation scheme used. Specifically, we find that some modulation schemes are less sensitive to the effects of noise and distortion than others.

We may classify the modulation process into continuous-wave modulation and pulse modulation. In continuous-wave (CW) modulation, a sinusoidal wave is used as the carrier. When the amplitude of the carrier is varied in accordance with the message signal, we have amplitude modulation (AM), and when the angle of the carrier is varied, we have angle modulation. The latter form of CW modulation may be further subdivided into frequency modulation (FM) and phase modulation (PM), in which the instantaneous frequency or phase of the carrier, respectively, is varied in accordance with the message signal.

In pulse modulation, on the other hand, the carrier consists of a periodic sequence of rectangular pulses. Pulse modulation can itself be of an analog or digital type. In analog pulse modulation, the amplitude, duration, or position of a pulse is varied in accordance with sample values of the message signal. In such a case, we speak of pulse-amplitude modulation (PAM), pulse-duration modulation (PDM), and pulse-position modulation (PPM).

The standard digital form of pulse modulation is known as pulse-code modulation (PCM) that has no CW counterpart. PCM starts out essentially as PAM, but with an important modification: The amplitude of each modulated pulse (i.e., sample of the original message signal) is quantized or rounded off to the nearest value in a prescribed set of discrete amplitude levels and then coded into a corresponding sequence of binary symbols. The binary symbols 0 and 1 are themselves represented by pulse signals that are suitably shaped for transmission over the channel. In any event, as a result of the quantization process, some information is always lost and the original message signal cannot therefore be reconstructed exactly. However, provided that the number of quantizing (discrete amplitude) levels is large enough, the distortion produced by the quantization process is not discernible to the human ear in the case of a speech signal or the human eye in the case of a two-dimensional image.

I.4. Multiplexing

In introducing the idea of modulation, we stressed its importance as a process that ensures the transmission of a message signal over a prescribed channel. There is another important benefit, namely, multiplexing, that results from the use of modulation. Multiplexing is the process of combining several message signals for their simultaneous transmission over the same channel. Three commonly used methods of multiplexing are as follows:

- Frequency-division multiplexing (FDM), in which CW modulation is used to translate each message signal to reside in a specific frequency slot inside the passband of the channel by assigning it a distinct carrier frequency; at the receiver, a bank of filters is used to separate the different modulated signals and prepare them individually for demodulation.
- Time-division multiplexing (TDM), in which pulse modulation is used to position samples of the different message signals in non-overlapping time slots.
- Code-division multiplexing (CDM), in which each message signal is identified by a distinctive code.

In FDM the message signals overlap with each other at the channel input; hence the system may suffer from crosstalk (i.e., interaction between message signals) if the channel is nonlinear. In TDM the message signals use the full passband of the channel, but on a timeshared basis. In CDM the message signals are permitted to overlap in both time and frequency across the channel.

Mention should also be made of wavelength-division multiplexing (WDM), which is special to optical fibers. In WDM, wavelength is used as a new degree of freedom by concurrently operating distinct portions of the wavelength spectrum (i.e., distinct colors) that are accessible within the optical fiber. However, recognizing the reciprocal relationship that exists between the wavelength and frequency of an electromagnetic wave, we may say that WDM is a form of FDM.

I.5. Analog and Digital Communications

In an analog communication system, there is no significant effort made by the system designer to tailor the waveform of the transmitted signal to suit the channel at any deeper level. On the other hand, digital communication theory endeavors to find a finite set of waveforms that are closely matched to the characteristics of the channel and which are therefore more tolerant of channel impairments. In so doing, reliable communication is established over the channel. In the selection of good waveforms for digital communication over a noisy channel, the design is influenced solely by the channel characteristics. However, once the appropriate set of waveforms for transmission over the channel has been selected, the source information can be encoded into the channel waveforms, and the efficient transmission of information from the source to the user is thereby ensured. In summary, the use of digital communications provides the capability for information transmission that is both efficient and reliable.

From this discussion, it is apparent that the use of digital communications requires a considerable amount of electronic circuitry, but nowadays electronics are inexpensive, due to the ever-increasing availability of very-large-scale integrated (VLSI) circuits in the form of silicon chips.

Thus although cost considerations used to be a factor in selecting analog communications over digital communications in the past, that is no longer the case.

I.6. Signal Classifications

BASEBAND	**BANDPASS**
PERIODIC	**APERIODIC**
ENERGY	**POWER**

I.6.A. BASEBAND SIGNALS

When the frequency spectrum of the signal is ***centered at zero*** frequency and is of ***finite extent***, the signal is ***baseband***.

In communication systems, the message is usually represented using a baseband waveform. Note that in Figure I.4, the signal bandwidth is equal to W .

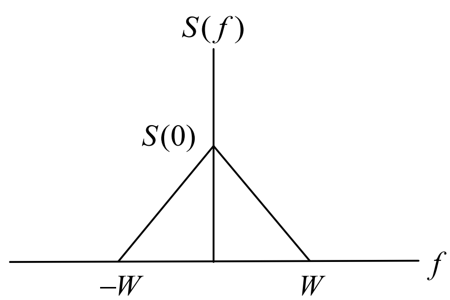


Figure I.4: Spectrum of Baseband Signal

I.6.B. BANDPASS SIGNALS

When the frequency spectrum of the signal is ***centered at*** $f_c \gg 0$ and is of ***finite extent***, the signal is ***bandpass***.

In communication systems, the transmitted signal is usually bandpass. Note that in Figure I.5, the signal bandwidth is equal to $2W$.

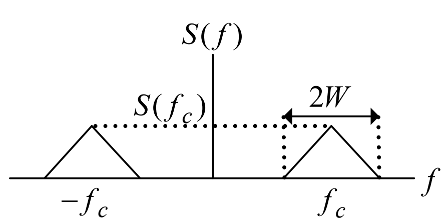


Figure I.5: Spectrum of Bandpass Signal

I.7. Fourier Analysis

Let $g(t)$ be a non-periodic deterministic signal. The Fourier transform is defined as:

$$G(f)=\int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}dt \tag{I.3}$$

The inverse transform is:

$$g(t)=\int_{-\infty}^{\infty} G(f)e^{j2\pi ft}df \tag{I.4}$$

I.7.A. Dirac Delta Function

$$\delta(t)=\begin{cases}0, & t\neq 0\\ \text{undefined}, & t=0\end{cases} \tag{I.5}$$

Note that $\delta(t)$ is an even function of time.

$$\begin{aligned}&\int_{-\infty}^{\infty}\delta(t)dt=1\\ &\int_{1}^{4}\delta(t)dt=0\\ &\int_{-2}^{7}\delta(t-3)dt=1\\ &\int_{-2}^{1}\delta(t-3)dt=0\end{aligned} \tag{I.6}$$

Sifting Property

Give that $g(t)$ is a continuous function of time,

$$\int_{-\infty}^{\infty} g(t)\delta(t-t_0)dt=g(t_0) \tag{I.7}$$

$$\delta(t-t_0)=\delta(t_0-t) \tag{I.8}$$

$$\int_{-\infty}^{\infty} g(t)\delta(t_0-t)dt=g(t_0)\Rightarrow\int_{-\infty}^{\infty} g(\tau)\delta(t-\tau)d\tau=g(t)\Rightarrow g(t)*\delta(t)=g(t) \tag{I.9}$$

I.7.B. FOURIER TRANSFORM OF PERIODIC SIGNALS

Let $g_{T_0}(t)$ be periodic with period T_0. The Fourier series of $g_{T_0}(t)$ is:

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_0 t} \tag{I.10}$$

where c_n is the complex Fourier series coefficient, given by:

$$c_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) e^{-j2\pi n f_0 t} dt \tag{I.11}$$

$f_0 = 1/T_0$ is known as the fundamental frequency.

Given,

$$g(t) = \begin{cases} g_{T_0}(t), & -\frac{T_0}{2} < t \le \frac{T_0}{2} \\ 0, & \text{elsewhere} \end{cases} \tag{I.12}$$

Then,

$$g_{T_0}(t) = \sum_{m=-\infty}^{\infty} g(t - mT_0) \tag{I.13}$$

Using the time shifting property of the Fourier Transform,

$$G_{T_0}(f) = G(f) \sum_{m=-\infty}^{\infty} e^{-j2\pi f m T_0} \tag{I.14}$$

Equation (I.11) can be rewritten using $g(t)$ in the form

$$\begin{aligned} c_n &= f_0 \int_{-\infty}^{\infty} g(t) e^{-j2\pi n f_0 t} dt \\ &= f_0 G(n f_0) \end{aligned} \tag{I.15}$$

Therefore, (I.10) can be rewritten as

$$g_{T_0}(t) = f_0 \sum_{n=-\infty}^{\infty} G(n f_0) e^{j2\pi n f_0 t} \tag{I.16}$$

Poisson's Sum Formula

Using (I.13), the result in (I.16) can be rewritten in the form

$$\sum_{m=-\infty}^{\infty} g(t-mT_0) = f_0 \sum_{n=-\infty}^{\infty} G(nf_0)e^{j2\pi nf_0 t} \tag{I.17}$$

Equation (I.17) is known as the Poisson's sum formula.

I.7.C. Properties of the Fourier Transform

Linearity

$$a_1 g_1(t) + a_2 g_2(t) \rightleftharpoons a_1 G_1(f) + a_2 G_2(f) \tag{I.18}$$

Time Scaling

$$g(at) \rightleftharpoons \frac{1}{|a|} G\left(\frac{f}{a}\right) \tag{I.19}$$

Duality

$$g(t) \rightleftharpoons G(f) \Rightarrow G(t) \rightleftharpoons g(-f) \tag{I.20}$$

Time Shifting

$$g(t-t_0) \rightleftharpoons G(f)e^{-j2\pi f t_0} \tag{I.21}$$

Frequency Shifting

$$g(t)e^{j2\pi f_0 t} \rightleftharpoons G(f-f_0) \tag{I.22}$$

Area in the Time Domain

$$\int_{-\infty}^{\infty} g(t)dt = G(0) \tag{I.23}$$

Area in the Frequency Domain

$$\int_{-\infty}^{\infty} G(f)df = g(0) \tag{I.24}$$

Time Differentiation

$$\frac{d}{dt} g(t) \rightleftharpoons j2\pi f G(f) \tag{I.25}$$

Time Integration

$$\int_{-\infty}^{t} g(\xi)d\xi = \frac{1}{j2\pi f}G(f) + \frac{G(0)}{2}\delta(f) \tag{I.26}$$

Conjugation

$$g^*(t) \rightleftharpoons G^*(-f) \tag{I.27}$$

Time Multiplication

$$g_1(t)g_2(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\xi)G_2(f-\xi)d\xi \tag{I.28}$$

Time Convolution

$$\int_{-\infty}^{\infty} g_1(\xi)g_2(t-\xi)d\xi = G_1(f)G_2(f) \tag{I.29}$$

I.7.D. Fourier Transform Pairs

$$1 \rightleftharpoons \delta(f) \tag{I.30}$$

$$\delta(t) \rightleftharpoons 1 \tag{I.31}$$

$$\delta(t-t_0) \rightleftharpoons e^{-j2\pi f t_0} \tag{I.32}$$

$$e^{j2\pi f_c t} \rightleftharpoons \delta(f-f_c) \tag{I.33}$$

$$\cos(2\pi f_c t) \rightleftharpoons \frac{1}{2}\left[\delta(f-f_c)+\delta(f+f_c)\right] \tag{I.34}$$

$$\sin(2\pi f_c t) \rightleftharpoons \frac{1}{j2}\left[\delta(f-f_c)-\delta(f+f_c)\right] \tag{I.35}$$

$$\text{rect}\left(\frac{t}{T}\right) \rightleftharpoons T\,\text{sinc}(fT) \tag{I.36}$$

$$\text{sinc}(Wt) \rightleftharpoons \frac{1}{W}\text{rect}\left(\frac{f}{W}\right) \tag{I.37}$$

$$e^{-at}u(t) \rightleftharpoons \frac{1}{a+j2\pi f}, \quad a>0 \tag{I.38}$$

$$e^{-a|t|} \rightleftharpoons \frac{2a}{a^2 + (2\pi f)^2} \tag{I.39}$$

$$e^{-\pi t^2} \rightleftharpoons e^{-\pi f^2} \tag{I.40}$$

$$\begin{cases} 1 - \dfrac{|t|}{T}, & |t| < T \\ 0, & |t| \geq T \end{cases} \rightleftharpoons T \operatorname{sinc}^2(fT) \tag{I.41}$$

$$\operatorname{sgn}(t) \rightleftharpoons \frac{1}{j\pi f} \tag{I.42}$$

$$\frac{1}{\pi t} \rightleftharpoons -j \operatorname{sgn}(f) \tag{I.43}$$

$$u(t) \rightleftharpoons \frac{1}{2}\left(\delta(f) + \frac{1}{j\pi f} \right) \tag{I.44}$$

$$\sum_{i=-\infty}^{\infty} \delta(t - iT_0) \rightleftharpoons \frac{1}{T_0} \sum_{i=-\infty}^{\infty} \delta\left(f - \frac{i}{T_0} \right) \tag{I.45}$$

I.8. Transmission of Signals Through Linear Systems

Let the impulse response of an LTI system be $h(t)$. Let the input be $x(t)$. Therefore, the output is given by:

$$\begin{aligned} y(t) &= x(t) * h(t) \\ &= \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau \end{aligned} \tag{I.46}$$

This is called the convolution integral. Equivalently, we may write

$$\begin{aligned} y(t) &= h(t) * x(t) \\ &= \int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau \end{aligned} \tag{I.47}$$

Exercise I.1

Determine the output of an LTI system when the input and impulse response are given, respectively by:

$$x(t) = u(t) - u(t-T)$$

$$h(t) = e^{-\alpha t}u(t)$$

where T and α are finite positive constants.

Consider a linear time-invariant system with impulse response $h(t)$ driven by a complex exponential input of unit amplitude and frequency f, i.e.,

$$x(t) = e^{j2\pi ft} \tag{I.48}$$

The response of the system to this input is obtained using the convolution operation as

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)e^{j2\pi f(t-\tau)}d\tau \\ &= e^{j2\pi ft}\int_{-\infty}^{\infty} h(\tau)e^{-j2\pi f\tau}d\tau \end{aligned} \tag{I.49}$$

Define the frequency response of the system as the Fourier transform of its impulse response (Frequency Response),

as shown by

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}dt \tag{I.50}$$

Hence,

$$\begin{aligned} y(t) &= H(f)e^{j2\pi ft} \\ &= |H(f)|e^{j\beta(f)}e^{j2\pi ft} \end{aligned} \tag{I.51}$$

The response of a linear time-invariant system to a complex exponential function of frequency f is, therefore, the same complex exponential function multiplied by a frequency-dependent coefficient $H(f)$. Note that the value of $H(f)$ depends on the frequency of the input.

The frequency response $H(f)$ is, in general, a complex quantity, so we may express it in the polar form

$$H(f) = |H(f)|e^{j\beta(f)} \tag{I.52}$$

where $|H(f)|$ is called the magnitude response, and $\beta(f)$ is the phase, or phase response. In the special case of a linear system with a real-valued impulse response $h(t)$, the frequency response $H(f)$ exhibits conjugate symmetry, which means that

$$|H(-f)| = |H(f)| \tag{I.53}$$

$$\beta(-f) = -\beta(f) \tag{I.54}$$

Exercise I.2

Determine the output of an LTI system when the input and impulse response are given, respectively by:

$$x(t) = 8\sin\left(20\pi t + \frac{\pi}{3}\right)$$

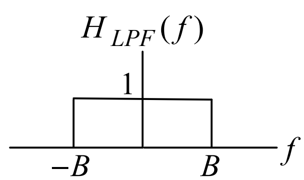
$$h(t) = e^{-3t}u(t)$$

I.8.A. FILTERS

Low-Pass Filters

The ideal lowpass filter is defined by the frequency response

$$\begin{aligned} H_{LPF}(f) &= \begin{cases} 1, & |f| \le B \\ 0, & \text{otherwise} \end{cases} \\ &= \text{rect}\left(\frac{f}{2B}\right) \end{aligned} \tag{I.55}$$



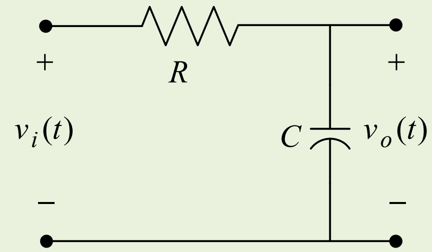
where B is the filter cut off frequency, or bandwidth.

Exercise I.3

Determine the impulse response of $H_{LPF}(f)$.

Exercise I.4

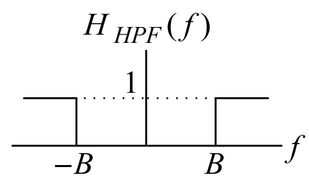
Determine the frequency response and the impulse response of the following circuit



High-Pass Filters

The ideal lowpass filter is defined by the frequency response

$$\begin{aligned} H_{HPF}(f) &= \begin{cases} 1, & |f| \geq B \\ 0, & \text{otherwise} \end{cases} \\ &= 1 - H_{LPF}(f) \end{aligned} \tag{I.56}$$

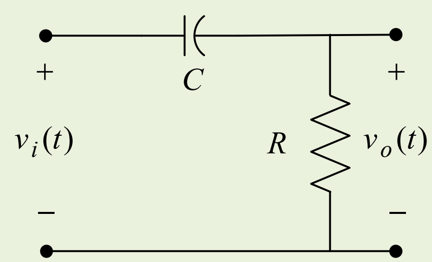


Exercise I.5

Determine the impulse response of $H_{HPF}(f)$.

Exercise I.6

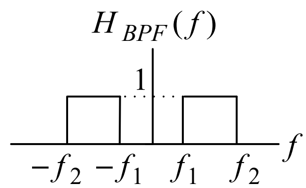
Determine the frequency response and the impulse response of the following circuit



Band-Pass Filters

The ideal lowpass filter is defined by the frequency response

$$H_{BPF}(f) = \begin{cases} 1, & f_1 \le |f| \le f_2 \\ 0, & \text{otherwise} \end{cases}$$
$$= \text{rect}\left(\frac{f - (f_1 + f_2)/2}{f_2 - f_1}\right) + \text{rect}\left(\frac{f + (f_1 + f_2)/2}{f_2 - f_1}\right) \quad \text{(I.57)}$$



Exercise I.7

Determine the impulse response of $H_{BPF}(f)$.

Exercise I.8

Sketch a BPF circuit, and determine its frequency response and impulse response.

I.8.B. BANDWIDTH

The time-domain and frequency-domain descriptions of a signal are inversely related. In particular, we may make the following important statements:

- If the time-domain description of a signal is changed, the frequency-domain description of the signal is changed in an inverse manner, and vice versa. This inverse relationship prevents arbitrary specifications of a signal in both domains. In other words, we may specify an arbitrary function of time or an arbitrary spectrum, but we cannot specify both of them together.
- In general, if a signal is strictly limited in frequency, the time-domain description of the signal will trail on indefinitely, even though its amplitude may assume a progressively smaller value. We say a signal is strictly limited in frequency or strictly band limited if its Fourier transform is exactly zero outside a finite band of frequencies. The sinc pulse is an example. An exception of this rule if the Gaussian waveform, which is infinitely extended in both domains.
- The bandwidth of a signal provides a measure of the extent of significant spectral content of the signal for positive frequencies. When the signal is strictly band limited, the bandwidth is well defined. Consider the pulse

$$x(t) = \text{sinc}(2Wt) = \frac{\sin(2\pi Wt)}{2\pi Wt}$$

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

The bandwidth of this signal is W.

There is no universally accepted definition of bandwidth. Nevertheless, there are some commonly used definitions for bandwidth.

Null-to-null bandwidth

When the spectrum of a signal is symmetric with a main lobe bounded by well-defined nulls (i.e., frequencies at which the spectrum is zero), we may use the main lobe as the basis for defining the bandwidth of the signal.

Specifically, if the signal is low-pass (i.e., its spectral content is centered around the origin), the bandwidth is defined as one half the total width of the main spectral lobe since only one half of this lobe lies inside the positive frequency region. For example, a rectangular pulse of duration T seconds has a main spectral lobe of total width $2/T$ Hertz centered at the origin. Accordingly, we may define the bandwidth of this rectangular pulse as $1/T$ Hertz.

If, on the other hand, the signal is band-pass with main spectral lobes centered around $\pm f_c$ where f_c is large enough, the bandwidth is defined as the width of the main lobe for positive frequencies.

On the basis of the definitions presented here, we may state that shifting the spectral content of a low-pass signal by a sufficiently large frequency has the effect of doubling the bandwidth of the signal; such a frequency translation is attained by using modulation.

3 dB Bandwidth

If the signal is low-pass, the 3-dB bandwidth is defined as the separation between zero frequency, where the magnitude spectrum attains its peak value, and the positive frequency, at which the amplitude spectrum drops to $1/\sqrt{2}$ of its peak value.

Example I.1

Let

$$x(t)=e^{-at}u(t),\text{ where } a>0$$

Then

$$X(f)=\frac{1}{a+j2\pi f}\Rightarrow\left|X(f)\right|^2=\frac{1}{a^2+\left(2\pi f\right)^2}\Rightarrow\left|X(0)\right|^2=\frac{1}{a^2}$$

$$\left|X(f_0)\right|^2=\frac{1}{2}\left|X(0)\right|^2=\frac{1}{2a^2}\text{ when } 2\pi f_0=a\Rightarrow f_0=\frac{a}{2\pi}$$

If the signal is band-pass, centered at $\pm f_c$, the 3-dB bandwidth is defined as the separation (along the positive frequency axis) between the two frequencies at which the magnitude spectrum of the signal drops to $1/\sqrt{2}$ of the peak value at f_c.

RMS Bandwidth

For a lowpass $g(t)$,

$$W_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} f^2 |G(f)|^2 \, df}{\int_{-\infty}^{\infty} |G(f)|^2 \, df} \right)^{1/2} \tag{I.58}$$

I.8.C. Time-Bandwidth Product

For any family of pulse signals that differ by a time-scaling factor, the product of the signal's duration and its bandwidth is always a constant, as shown by

$$\text{Duration} \times \text{Bandwidth} = \text{Constant} \tag{I.59}$$

Example I.2

A rectangular pulse $x(t)$ of duration T seconds has a bandwidth (defined on the basis of the positive-frequency part of the main lobe) equal to $1/T$ hertz, making the time-bandwidth product of the pulse equal unity.

The pulse $x(2t)$ has a duration of $T/2$ and a bandwidth of $2/T$, making the time-bandwidth product of the pulse equal to unity again.

RMS Duration

For a $g(t)$ that is centered about the origin,

$$T_{\text{rms}} = \left(\frac{\int_{-\infty}^{\infty} t^2 |g(t)|^2 \, dt}{\int_{-\infty}^{\infty} |g(t)|^2 \, dt} \right)^{1/2} \tag{I.60}$$

$$T_{\text{rms}} W_{\text{rms}} \geq \frac{1}{4\pi} \tag{I.61}$$

The equality in (I.61) holds when $g(t) = e^{-\pi t^2}$ (gaussian pulse).

I.8.D. Inner Product

The inner product (dot product) of two signals $x(t)$ and $y(t)$ (also called the cross-correlation) is defined as

$$\begin{aligned} R_{xy} &= \langle x(t), y(t) \rangle \\ &= \int_{-\infty}^{\infty} x(t) y^*(t) dt \\ &= \int_{-\infty}^{\infty} X(f) Y^*(f) df \end{aligned} \tag{I.62}$$

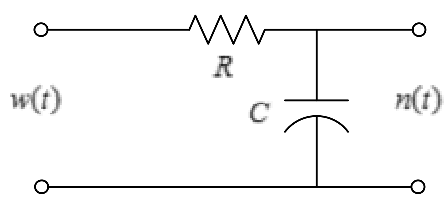
The complex quantity ρ_{xy}, called the cross-correlation coefficient of $x(t)$ and $y(t)$, is defined as

$$\begin{aligned} \rho_{xy} &= \frac{R_{xy}}{\sqrt{\mathcal{E}_x \mathcal{E}_y}} \\ &= \frac{\langle x(t), y(t) \rangle}{\sqrt{\mathcal{E}_x \mathcal{E}_y}} \end{aligned} \tag{I.63}$$

$$\left| \rho_{xy} \right| \le 1 \tag{I.64}$$

I.8.E. Noise Equivalent Bandwidth

Suppose that a white noise source of power spectral density $N_0/2$ is connected to the input of the simple RC low-pass filter



Note that

$$S_W(f) = \frac{N_0}{2} \tag{I.65}$$

$$\begin{aligned} H(f) &= \frac{1/(j2\pi fC)}{R + 1/(j2\pi fC)} \\ &= \frac{1}{1 + j2\pi RCf} \end{aligned} \tag{I.66}$$

$$|H(f)|^2 = \frac{1}{1 + (2\pi RCf)^2} \tag{I.67}$$

$$|H(0)|^2 = 1 \tag{I.68}$$

$$W_{3\text{dB}} = \frac{1}{2\pi RC} \tag{I.69}$$

$$S_N(f) = \frac{N_0}{2} \frac{1}{1 + (2\pi RCf)^2} \tag{I.70}$$

$$\begin{aligned} P_N &= \int_{-\infty}^{\infty} S_N(f)\,df \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} \frac{1}{1 + (2\pi RCf)^2}\,df \\ &= \frac{N_0}{4RC} \end{aligned} \tag{I.71}$$

We find that the average output noise power of the filter is proportional to the filter bandwidth. We may generalize this statement to include all kinds of low-pass filters by defining a noise equivalent bandwidth. Suppose that a zero-mean white noise source of power spectral density $N_0/2$ is connected to the input of a low-pass filter with the transfer function $H(f)$. At the filter output,

$$\begin{aligned} P_N &= \frac{N_0}{2} \int_{-\infty}^{\infty} |H(f)|^2\,df \\ &= N_0 \int_{0}^{\infty} |H(f)|^2\,df \end{aligned} \tag{I.72}$$

Consider next the same source of white noise connected to the input of an ideal lowpass filter of zero-frequency response $H(0)$ and bandwidth B. In this case, the average output noise power is

$$P_N = N_0 B H^2(0) \tag{I.73}$$

Equating P_N in (I.73) with P_N in (I.72) yields

$$B = \frac{\int_0^\infty |H(f)|^2 \, df}{H^2(0)} \tag{I.74}$$

***

II. Lowpass Equivalent Signal Modeling

II.1. Hilbert Transform

II.1.A. Basics

The Hilbert transform is defined by:

$$\begin{aligned}\hat{g}(t) &= g(t) * \frac{1}{\pi t} \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau} d\tau\end{aligned} \qquad \text{(II.1)}$$

$$\frac{1}{\pi t} \rightleftharpoons -j\,\text{sgn}(f) \qquad \text{(II.2)}$$

$$\text{sgn}(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0 \\ -1, & f < 0 \end{cases} \qquad \text{(II.3)}$$

$$\hat{G}(f) = -j\,\text{sgn}(f) G(f) \qquad \text{(II.4)}$$

Hilbert transform shifts positive frequencies by -90°, and negative frequencies by 90°.

II.1.B. Properties of the Hilbert Transform

1. $\left|\hat{G}(f)\right| = \left|G(f)\right|$.
2. $\hat{\hat{g}}(t) = -g(t)$.
3. If $g(t)$ is even, then $\hat{g}(t)$ is odd.
4. If $g(t)$ is odd, then $\hat{g}(t)$ is even.
5. Orthogonality: $\int_{-\infty}^{\infty} g(t)\hat{g}(t)dt = 0$.
6. Energy: $\int_{-\infty}^{\infty} g^2(t)dt = \int_{-\infty}^{\infty} \hat{g}^2(t)dt$

II.1.C. Pre-Envelope

Let $g(t)$ be a real bandpass signal. Note that $G(f)$ is conjugate symmetric, i.e.,

$$G(-f) = G^*(f) \qquad \text{(II.5)}$$

The pre-envelopes of $g(t)$ are defined as follows:

$$g_+(t) = g(t) + j\hat{g}(t) \tag{II.6}$$

$$\begin{aligned} g_-(t) &= g(t) - j\hat{g}(t) \\ &= g_+^*(t) \end{aligned} \tag{II.7}$$

This means that

$$G_-(f) = G_+^*(-f) \tag{II.8}$$

Note that from (II.6) and (II.7),

$$g(t) = \frac{1}{2}\left[g_+(t) + g_-(t)\right] \tag{II.9}$$

This means that

$$\begin{aligned} G(f) &= \frac{1}{2}\left[G_+(f) + G_-(f)\right] \\ &= \frac{1}{2}\left[G_+(f) + G_+^*(-f)\right] \end{aligned} \tag{II.10}$$

From (II.10) we can easily see that $G_+(f)$ is sufficient to reconstruct $G(f)$. From (II.6):

$$\begin{aligned} G_+(f) &= G(f) + \mathrm{sgn}(f)G(f) & G_-(f) &= G(f) - \mathrm{sgn}(f)G(f) \\ &= G(f)\left[1 + \mathrm{sgn}(f)\right] \quad | & & G(f)\left[1 - \mathrm{sgn}(f)\right] \\ &= \begin{cases} 2G(f), & f > 0 \\ G(0), & f = 0 \\ 0, & f < 0 \end{cases} & &= \begin{cases} 2G(f), & f < 0 \\ G(0), & f = 0 \\ 0, & f > 0 \end{cases} \end{aligned} \tag{II.11}$$

Note that when $g(t)$ is bandpass $G_+(f)$ and $G_-(f)$ do not overlap. The energy in $g(t)$ can be found from

$$\begin{aligned} \mathcal{E}_g &= \int_{-\infty}^{\infty} \left|g(t)\right|^2 dt \\ &= \int_{-\infty}^{\infty} \left|G(f)\right|^2 df \end{aligned} \tag{II.12}$$

Substituting for $G(f)$ from (II.10) yields

$$\mathcal{E}_g = \frac{1}{4}\int_{-\infty}^{\infty} \left\| \left[G_+(f) + G_-(f) \right] \right\|^2 df \tag{II.13}$$

Since $G_+(f)$ and $G_-(f)$ do not overlap, (II.13) can be rewritten in the form

$$\begin{aligned} \mathcal{E}_g &= \frac{1}{4}\int_{-\infty}^{\infty} \left(\left| G_+(f) \right|^2 + \left| G_-(f) \right|^2 \right) df \\ &= \frac{1}{2}\mathcal{E}_{g_+} \end{aligned} \tag{II.14}$$

II.1.D. Complex Envelope

Let $g(t)$ be narrowband, with $G(f)$ centered at f_c. Then,

$$\tilde{g}(t) = g_+(t) e^{-j2\pi f_c t} \tag{II.15}$$

Or,

$$g_+(t) = \tilde{g}(t) e^{j2\pi f_c t} \tag{II.16}$$

This means that

$$\tilde{G}(f) = G_+(f + f_c) \tag{II.17}$$

Note that from (II.17),

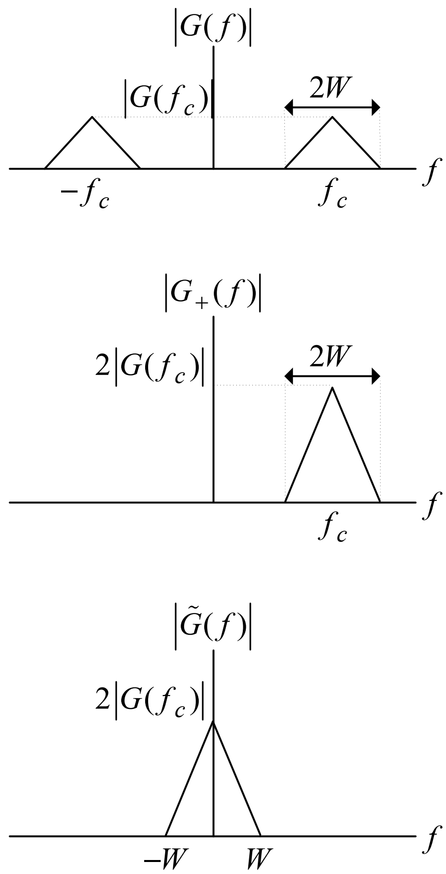
$$\mathcal{E}_{\tilde{g}} = \mathcal{E}_{g_+} \tag{II.18}$$

This means that

$$\mathcal{E}_g = \frac{1}{2}\mathcal{E}_{\tilde{g}} \tag{II.19}$$

Note that

$$\begin{aligned} \mathcal{E}_g &= \int_{-\infty}^{\infty} g(t) g^*(t) dt \\ &= \left\langle g(t), g(t) \right\rangle \end{aligned} \tag{II.20}$$



$$\begin{aligned} g(t) &= \mathrm{Re}\left\{ g_{+}(t) \right\} \\ &= \mathrm{Re}\left\{ \tilde{g}(t) e^{j2\pi f_c t} \right\} \end{aligned} \tag{II.21}$$

$$\tilde{g}(t) = g_I(t) + j g_Q(t) \tag{II.22}$$

$$g(t) = g_I(t)\cos\left(2\pi f_c t\right) - g_Q(t)\sin\left(2\pi f_c t\right) \tag{II.23}$$

$$\begin{aligned} \hat{g}(t) &= \mathrm{Im}\left\{ g_{+}(t) \right\} \\ &= \mathrm{Im}\left\{ \tilde{g}(t) e^{j2\pi f_c t} \right\} \end{aligned} \tag{II.24}$$

$$\hat{g}(t) = g_Q(t)\cos\left(2\pi f_c t\right) + g_I(t)\sin\left(2\pi f_c t\right) \tag{II.25}$$

Note that

$$\langle g_1(t), g_2(t)\rangle = \frac{1}{2}\langle \tilde{g}_1(t), \tilde{g}_2(t)\rangle \tag{II.26}$$

Note also that

$$\rho_{g_1 g_2} = \mathrm{Re}\left\{\rho_{\tilde{g}_1 \tilde{g}_2}\right\} \tag{II.27}$$

$$x(t) = x_I(t)\cos\left(2\pi f_c t\right) - x_Q(t)\sin\left(2\pi f_c t\right) \tag{II.28}$$

Example II.1

Let $m(t)$ be a baseband signal with bandwidth W and energy $\mathcal{E}_m$. Let $x(t) = m(t)\cos\left(2\pi f_c t\right)$ and $y(t) = m(t)\sin\left(2\pi f_c t\right)$, where $f_c > W$.

$$x_I(t) = m(t)$$
$$x_Q(t) = 0$$

$$y_I(t) = 0$$
$$y_Q(t) = -m(t)$$

$$\tilde{x}(t) = m(t)$$
$$\tilde{y}(t) = -jm(t)$$

$$\rho_{\tilde{x}\tilde{y}} = j\mathcal{E}_m$$

$$\rho_{xy} = 0$$

Since their cross-correlation is zero, $x(t)$ and $y(t)$ are said to be orthogonal.

II.1.E. Envelope

The complex envelope $\tilde{g}(t)$ can be written in polar form as follows:

$$\tilde{g}(t) = a(t)e^{j\varphi(t)} \tag{II.29}$$

$$\begin{aligned} a(t) &= \left|\tilde{g}(t)\right| \\ &= \left|g_+(t)e^{-j2\pi f_c t}\right| \\ &= \left|g_+(t)\right| \end{aligned} \tag{II.30}$$

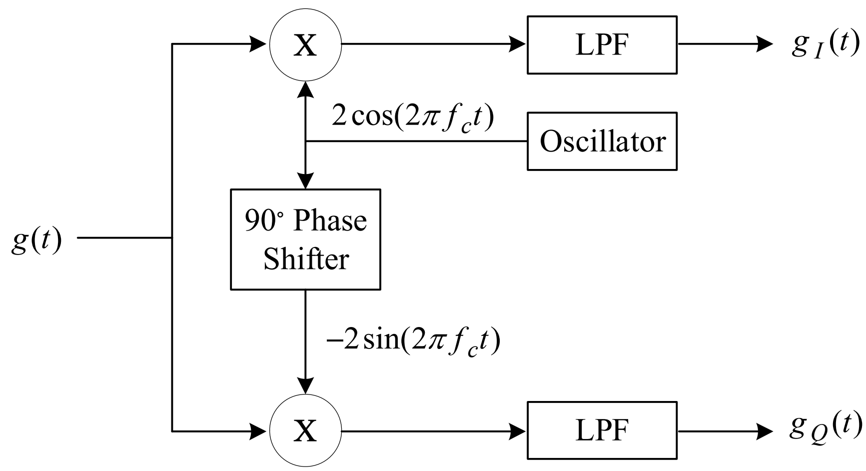
$$a(t) = \sqrt{g_I^2(t) + g_Q^2(t)} \tag{II.31}$$

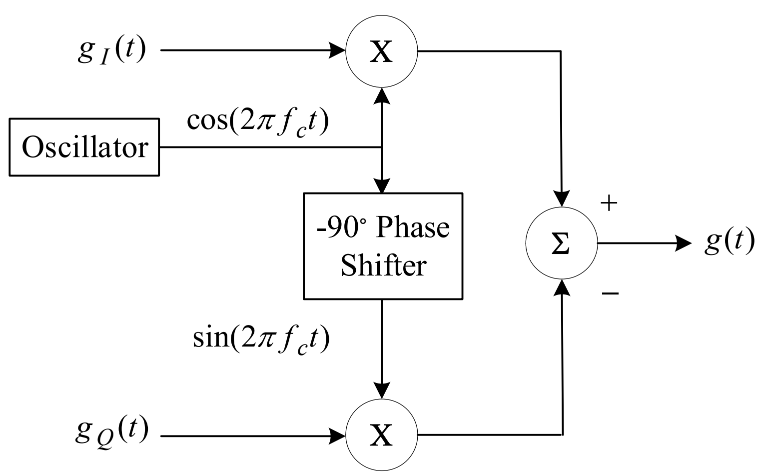
$$\varphi(t) = \tan^{-1}\left(\frac{g_Q(t)}{g_I(t)}\right) \quad \text{(II.32)}$$

$$g(t) = a(t)\cos\left(2\pi f_c t + \varphi(t)\right) \quad \text{(II.33)}$$

$$g_I(t) = a(t)\cos\varphi(t) \quad \text{(II.34)}$$

$$g_Q(t) = a(t)\sin\varphi(t) \quad \text{(II.35)}$$





II.1.F. Band-Pass Systems

Let $x(t)$ be a narrowband signal. Then,

$$x(t) = x_I(t)\cos\left(2\pi f_c t\right) - x_Q(t)\sin\left(2\pi f_c t\right) \quad \text{(II.36)}$$

$$\tilde{x}(t) = x_I(t) + jx_Q(t) \quad \text{(II.37)}$$

Let $x(t)$ be applied to a linear time-invariant band-pass system with impulse response $h(t)$.

$$h(t) = h_I(t)\cos\left(2\pi f_c t\right) - h_Q(t)\sin\left(2\pi f_c t\right) \tag{II.38}$$

$$\tilde{h}(t) = h_I(t) + jh_Q(t) \tag{II.39}$$

$$\begin{aligned} h(t) &= \mathrm{Re}\left\{\tilde{h}(t)e^{j2\pi f_c t}\right\} \\ &= \frac{1}{2}\left(\tilde{h}(t)e^{j2\pi f_c t} + \tilde{h}^*(t)e^{-j2\pi f_c t}\right) \end{aligned} \tag{II.40}$$

$$H(f) = \frac{1}{2}\left(\tilde{H}(f - f_c) + \tilde{H}^*(-f - f_c)\right) \tag{II.41}$$

Note that since $h(t)$ is a real function, then $H(f)$ satisfies

$$H^*(f) = H(-f) \tag{II.42}$$

This can also be deduced from (II.41).

Note that in (II.41), $\tilde{H}(f - f_c)$ is centered at f_c while $\tilde{H}^*(-f - f_c)$ is centered at $-f_c$.

Therefore,

$$\tilde{H}(f - f_c) = 2H(f), \quad f > 0 \tag{II.43}$$

$$y(t) = \mathrm{Re}\left\{\tilde{y}(t)e^{j2\pi f_c t}\right\} \tag{II.44}$$

$$\tilde{y}(t) = \frac{1}{2}\tilde{x}(t) * \tilde{h}(t) \tag{II.45}$$

$$\begin{aligned} \tilde{y}(t) &= \frac{1}{2}\left(x_I(t) + jx_Q(t)\right) * \left(h_I(t) + jh_Q(t)\right) \\ &= \frac{1}{2}\left[\left(x_I(t) * h_I(t) - x_Q(t) * h_Q(t)\right) + j\left(x_I(t) * h_Q(t) + x_Q(t) * h_I(t)\right)\right] \end{aligned} \tag{II.46}$$

Therefore,

$$y_I(t) = \frac{1}{2}\left(x_I(t) * h_I(t) - x_Q(t) * h_Q(t)\right) \tag{II.47}$$

$$y_Q(t) = \frac{1}{2}\left(x_I(t) * h_Q(t) + x_Q(t) * h_I(t)\right) \tag{II.48}$$

***

II: Lowpass Equivalent Signal Modeling

II.1-Hilbert Transform

III. AMPLITUDE MODULATION

The purpose of a communication system is to transmit information-bearing signals through a communication channel separating the transmitter from the receiver. Information bearing signals are generally baseband signals.

The proper use of the communication channel requires a shift of the range of baseband frequencies into bandpass frequency ranges suitable for transmission, and a corresponding shift back to the original frequency range after reception.

A shift of the range of frequencies in a signal is accomplished by using modulation, which is defined as the process by which some characteristic of a carrier is varied in accordance with a modulating wave (message signal).

A common form of the carrier is a sinusoidal wave, in which case we speak of a continuous-wave modulation process.

The baseband signal is referred to as the modulating wave, and the result of the modulation process is referred to as the modulated wave. Modulation is performed at the transmitting end of the communication system. At the receiving end of the system, we usually require the original baseband signal to be restored. This is accomplished by using a process known as demodulation, which is the reverse of the modulation process.

In AM, the amplitude of a sinusoidal carrier is varied linearly in accordance with an incoming message signal.

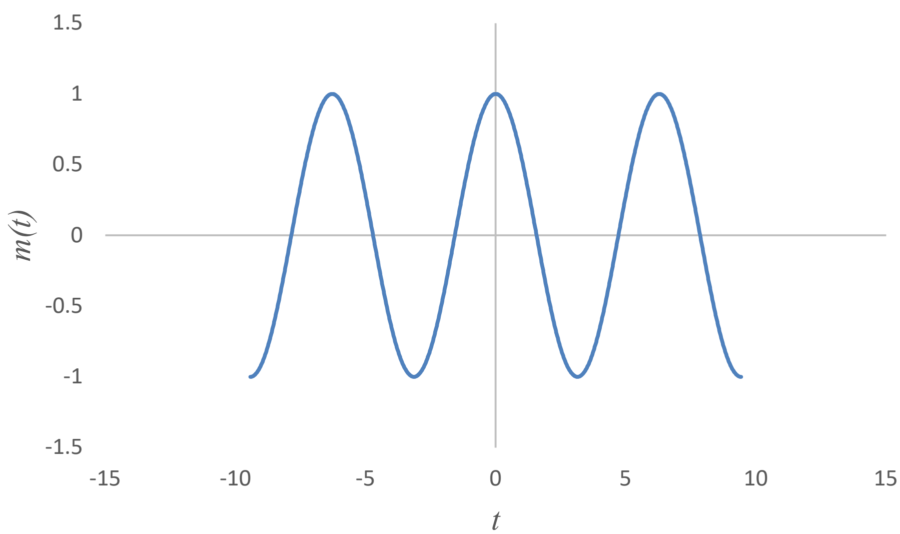


Figure III.1: Message (modulating) signal

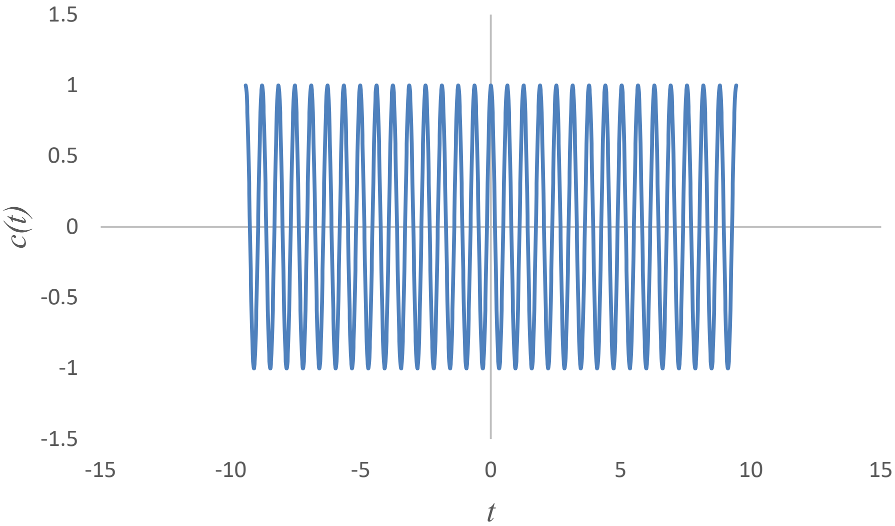


Figure III.2: Carrier signal

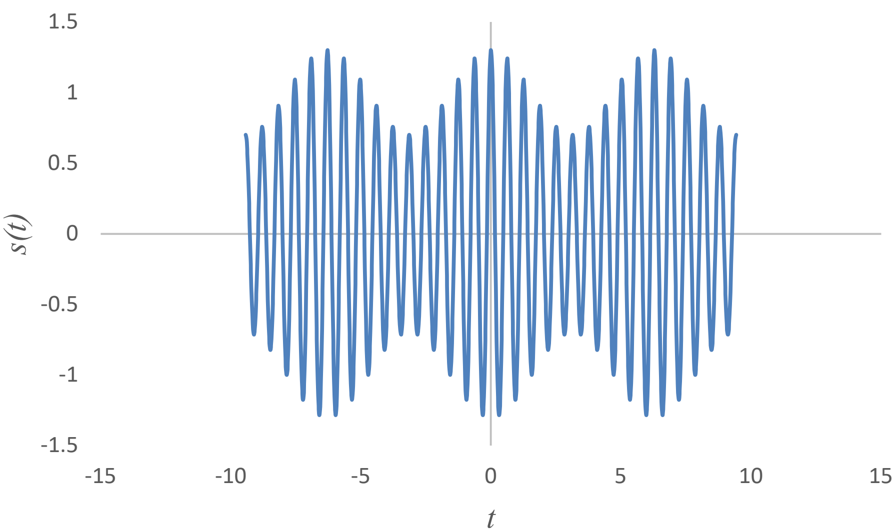


Figure III.3: Modulated signal

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

$$c(t) = A_c \cos\left(2\pi f_c t\right) \tag{III.1}$$

Let $m(t)$ denote the baseband signal that carries the specification of the message. The source of carrier wave $c(t)$ is physically independent of the source responsible for generating $m(t)$.

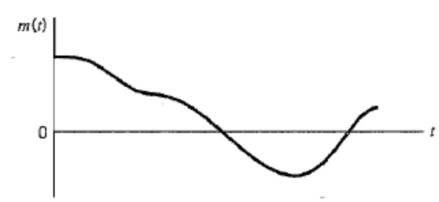
Amplitude modulation (AM) is defined as a process in which the amplitude of the carrier wave $c(t)$ is varied about a mean value, linearly with the baseband signal $m(t)$.

An amplitude-modulated (AM) wave may thus be described, in its most general form, as a function of time as follows:

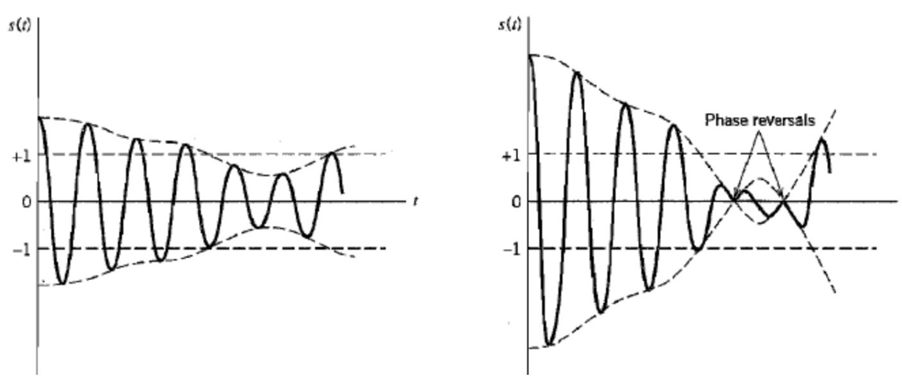
$$\begin{aligned} s(t) &= A_c\left[1+k_a m(t)\right]\cos\left(2\pi f_c t\right) \\ &= \underbrace{A_c\cos\left(2\pi f_c t\right)}_{\text{Unmodulated Carrier}} + \underbrace{A_c k_a m(t)\cos\left(2\pi f_c t\right)}_{\text{Sidebands}} \end{aligned} \qquad \text{(III.2)}$$

$$S(f) = \underbrace{\frac{A_c}{2}\left[\delta(f-f_c)+\delta(f+f_c)\right]}_{\text{Unmodulated Carrier}} + \underbrace{\frac{k_a A_c}{2}\left[M(f-f_c)+M(f+f_c)\right]}_{\text{Sidebands}}$$

where k_a is a constant called the amplitude sensitivity of the modulator responsible for the generation of the modulated signal $s(t)$. Typically, the carrier amplitude A_c and the message signal $m(t)$ are measured in volts, in which case k_a is measured in V^{-1}.



(a)



We observe that the envelope of $s(t)$ has essentially the same shape as the baseband signal $m(t)$ provided that two requirements are satisfied:

1. The amplitude of $k_a m(t)$ is always less than unity, that is,

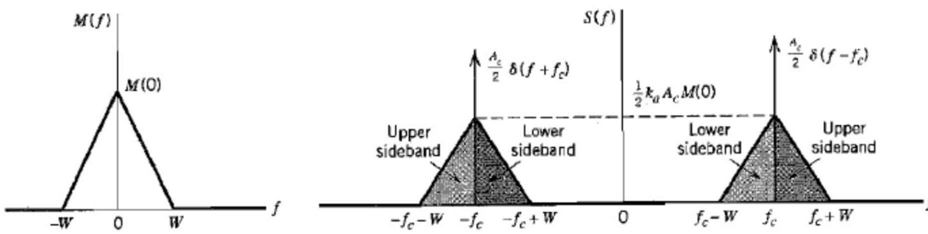
$$\left|k_a m(t)\right| < 1,\ \forall t \qquad \text{(III.3)}$$

$$-1 < k_a m(t) < 1,\ \forall t$$

2. The carrier frequency f_c is much larger than the message signal bandwidth W.

From (III.2),

$$S(f)=\frac{A_c}{2}\left[\delta\left(f-f_c\right)+\delta\left(f+f_c\right)\right]+\frac{k_a A_c}{2}\left[M\left(f-f_c\right)+M\left(f+f_c\right)\right] \qquad \text{(III.4)}$$



$$B_T=2W \qquad \text{(III.5)}$$

III.1. Virtues of Amplitude Modulation

- In the transmitter, amplitude modulation is accomplished using a nonlinear device. For example, in the switching modulator discussed in **Problem 2.3**, the combined sum of the message signal and carrier wave is applied to a diode, with the carrier amplitude being large enough to swing across the characteristic curve of the diode. Fourier analysis of the voltage developed across a resistive load reveals the generation of an AM component, which may be extracted by means of a band-pass filter.
- In the receiver, amplitude demodulation is also accomplished using a nonlinear device. For example, we may use a simple and yet highly effective circuit known as the envelope detector, which is discussed in **Problem 2.5**. The circuit consists of a diode connected in series with the parallel combination of a capacitor and load resistor. Some version of this circuit is found in most commercial AM radio receivers. Provided that the carrier frequency is high enough and the percentage modulation is less than 100 percent, the demodulator output developed across the load resistor is nearly the same as the envelope of the incoming AM wave, hence the name "envelope detector."

III.2. Limitations of Amplitude Modulation

- Amplitude modulation is wasteful of power. The carrier wave is completely independent of the information-bearing signal. The transmission of the carrier wave therefore represents a waste of power, which means that in amplitude modulation only a fraction of the total transmitted power is actually affected by the message signal.
- Amplitude modulation is wasteful of bandwidth. The upper and lower sidebands of an AM wave are uniquely related to each other by virtue of their symmetry about the carrier frequency; hence, given the magnitude and phase spectra of either sideband, we can uniquely determine the other. This means that insofar as the transmission of information is concerned, only one sideband is necessary, and the communication channel therefore needs to provide only the same bandwidth as the baseband signal. In light of this observation,

amplitude modulation is wasteful of bandwidth as it requires a transmission bandwidth equal to twice the message bandwidth.

To overcome these limitations, we must make certain modifications: suppress the carrier and modify the sidebands of the AM wave. These modifications naturally result in increased system complexity. In effect, we trade system complexity for improved use of communication resources. The basis of this trade-off is linear modulation. In a strict sense, full amplitude modulation does not qualify as linear modulation because of the presence of the carrier wave.

III.2.A. EXAMPLE

Let

$$m(t) = A_m \cos\left(2\pi f_m t\right) \text{ (Single Tone)}$$

$$P_m = \frac{A_m^2}{2}$$

$$M(f) = \frac{A_m}{2}\left[\underbrace{\delta(f - f_m)}_{f_m} + \underbrace{\delta(f + f_m)}_{-f_m}\right]$$

$$W = f_m$$

$$c(t) = A_c \cos\left(2\pi f_c t\right)$$

$$P_c = \frac{A_c^2}{2}$$

$$C(f) = \frac{A_c}{2}\left[\underbrace{\delta(f - f_c)}_{f_c} + \underbrace{\delta(f + f_c)}_{-f_c}\right]$$

$$f_m << f_c$$

$$s(t) = A_c\left[1 + k_a A_m \cos(2\pi f_m t)\right]\cos(2\pi f_c t)$$
$$= \underbrace{A_c \cos(2\pi f_c t)}_{\text{Unmodulated Carrier}} + \underbrace{A_c k_a A_m \cos(2\pi f_m t)\cos(2\pi f_c t)}_{\text{Sidebands}}$$
$$= \underbrace{A_c \cos(2\pi f_c t)}_{\text{Unmodulated Carrier}} + \underbrace{\frac{A_c k_a A_m}{2}\cos 2\pi(f_c - f_m)t}_{\text{Lower Sideband}} + \underbrace{\frac{A_c k_a A_m}{2}\cos 2\pi(f_c + f_m)t}_{\text{Upper Sideband}}$$

$$k_a A_m = \mu \text{ (Modulation Index)}$$

$$0 \le \mu \le 1 \text{ (to avoid overmodulation)}$$

$$s(t) = A_c\left[1 + \mu\cos(2\pi f_m t)\right]\cos(2\pi f_c t)$$
$$= \underbrace{A_c \cos(2\pi f_c t)}_{\text{Unmodulated Carrier}} + \underbrace{\mu A_c \cos(2\pi f_m t)\cos(2\pi f_c t)}_{\text{Sidebands}}$$
$$= \underbrace{A_c \cos(2\pi f_c t)}_{\text{Unmodulated Carrier}} + \frac{\mu A_c}{2}\left[\cos 2\pi(f_c - f_m)t + \cos 2\pi(f_c + f_m)t\right]$$

$$P_s = \frac{A_c^2}{2} + \frac{\left(\frac{\mu A_c}{2}\right)^2}{2} + \frac{\left(\frac{\mu A_c}{2}\right)^2}{2}$$
$$= P_c + P_{\text{LSB}} + P_{\text{USB}}$$

$$P_s = P_c + P_c\frac{\mu^2}{4} + P_c\frac{\mu^2}{4} = P_c + P_{\text{Sidebands}}$$
$$= \left(1 + \frac{\mu^2}{2}\right)P_c$$

$$P_c \le P_s \le \frac{3}{2}P_c$$

III.2-Limitations of Amplitude Modulation

$$\frac{P_c}{P_s} = \frac{1}{1 + \frac{\mu^2}{2}}$$

$$\frac{2}{3} P_s \le P_c \le P_s$$

$$0 \le P_{\text{Sidebands}} \le \frac{1}{3} P_s$$

$$0 \le P_{\text{LSB}} = P_{\text{USB}} \le \frac{1}{6} P_s$$

$$S(f) = \underbrace{\frac{A_c}{2}\left[\underbrace{\delta(f - f_c)}_{f_c} + \underbrace{\delta(f + f_c)}_{-f_c}\right]}_{\text{Unmodulated Carrier}}$$

$$+ \underbrace{\frac{\mu A_c}{4}\left[\left(\underbrace{\delta(f - f_c + f_m)}_{f_c - f_m} + \underbrace{\delta(f + f_c - f_m)}_{-(f_c - f_m)}\right) + \left(\underbrace{\delta(f - f_c - f_m)}_{f_c + f_m} + \underbrace{\delta(f + f_c + f_m)}_{-(f_c + f_m)}\right)\right]}_{\text{Sidebands}}$$

III.3. Linear Modulation Schemes

In its most general form, linear modulation is defined by

$$s(t) = s_I(t)\cos\left(2\pi f_c t\right) - s_Q(t)\sin\left(2\pi f_c t\right) \tag{III.6}$$

where $s_I(t)$ is the in-phase component of the modulated wave $s(t)$, and $s_Q(t)$ is its quadrature component.

In linear modulation, both $s_I(t)$ and $s_Q(t)$ are low-pass signals that are linearly related to the message signal $m(t)$.

- Double sideband-suppressed carrier (DSB-SC) modulation, where only the upper and lower sidebands are transmitted. Bandwidth= $2W$.
- Single sideband (SSB) modulation, where only one sideband (the lower sideband or the upper sideband) is transmitted. Bandwidth= W .
- Vestigial sideband (VSB) modulation, where only a vestige (i.e., trace) of one of the sidebands and a correspondingly modified version of the other sideband are transmitted. $W \le \text{Bandwith} \le 2W$.

Type of Modulation	In-Phase Component $s_I(t)$	Quadrature Component $s_Q(t)$	Comments
DSB-SC	$m(t)$	0	$m(t)$ = message signal
SSB:[a]			
(a) Upper sideband transmitted	$\frac{1}{2}m(t)$	$\frac{1}{2}\hat{m}(t)$	$\hat{m}(t)$ = Hilbert transform of $m(t)$
(b) Lower sideband transmitted	$\frac{1}{2}m(t)$	$-\frac{1}{2}\hat{m}(t)$	
VSB:			
(a) Vestige of lower sideband transmitted	$\frac{1}{2}m(t)$	$\frac{1}{2}m'(t)$	$m'(t)$ = output of the filter of frequency response $H_Q(f)$ due to $m(t)$. For the definition of $H_Q(f)$, see Eq. (2.16)
(b) Vestige of upper sideband transmitted	$\frac{1}{2}m(t)$	$-\frac{1}{2}m'(t)$	

The in-phase component $s_I(t)$ is solely dependent on the message signal $m(t)$.

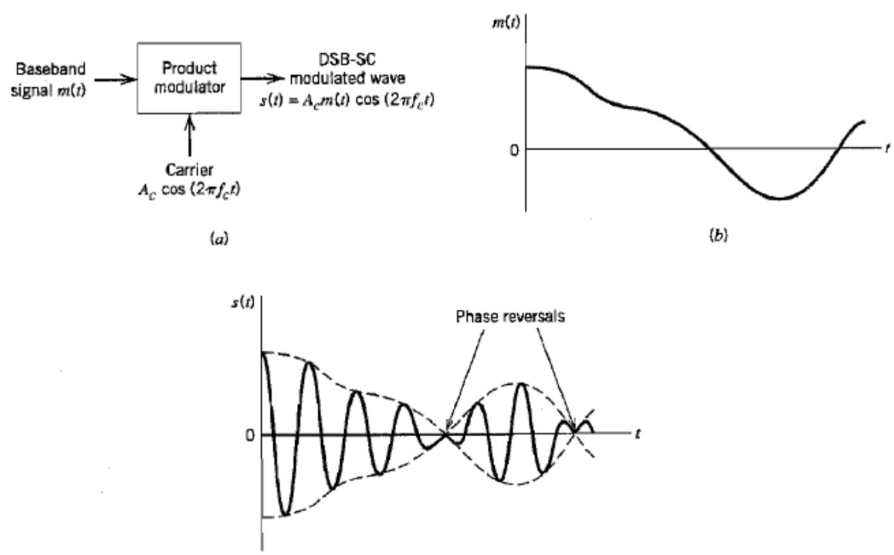
The quadrature component $s_Q(t)$ is a filtered version of $m(t)$. The spectral modification of the modulated wave sit) is solely due to $s_Q(t)$.

To be more specific, the role of the quadrature component (if present) is merely to interfere with the in-phase component, so as to reduce or eliminate power in one of the sidebands of the modulated signal sit), depending on how the quadrature component is defined.

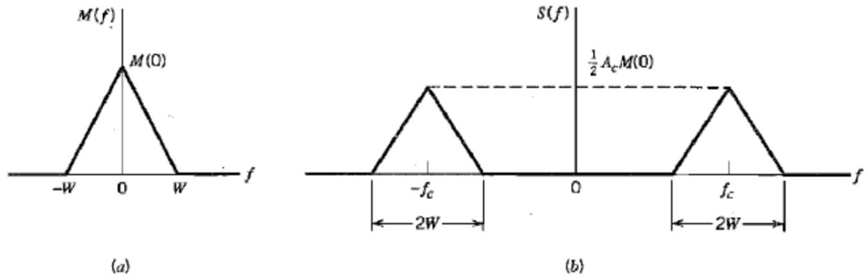
III.3.A. Double Sideband-Suppressed Carrier (DSB-SC) Modulation

This form of linear modulation is generated by using a product modulator that simply multiplies the message signal $m(t)$ by the carrier wave $c(t)$.

$$s(t) = m(t)c(t) = A_c m(t)\cos(2\pi f_c t) \quad \text{(III.7)}$$



(c)

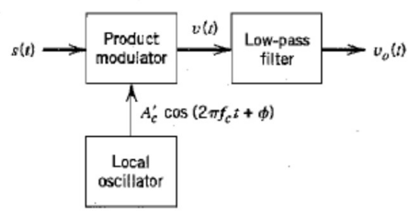


$$S(f)=\frac{A_c}{2}\left[M\left(f-f_c\right)+M\left(f+f_c\right)\right] \quad \text{(III.8)}$$

Of course, the transmission bandwidth required by DSB-SC modulation is the same as that for amplitude modulation, namely, $2W$.

Coherent Detection

The baseband signal $m(t)$ can be uniquely recovered from a DSB-SC wave $s(t)$ by first multiplying $s(t)$ with a locally generated sinusoidal wave and then low-pass filtering the product.



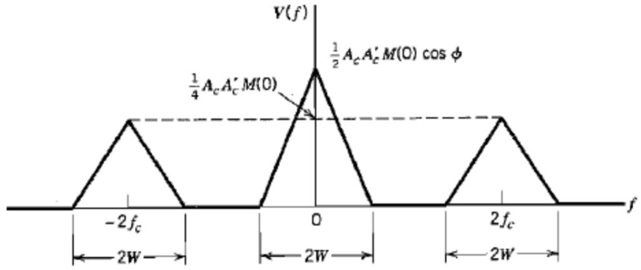
$$s(t)=m(t)c(t)=A_c m(t)\cos\left(2\pi f_c t+\theta\right) \quad \text{(III.9)}$$

It is assumed that the local oscillator signal is exactly coherent or synchronized, in both frequency and phase, with the carrier wave $c(t)$ used in the product modulator to generate $s(t)$. This method of demodulation is known as coherent detection.

It is instructive to derive coherent detection as a special case of the more general demodulation process using a local oscillator signal of the same frequency but arbitrary phase difference ϕ, measured with respect to the carrier wave $c(t)$. Thus, denoting the local oscillator signal by $A_c'\cos\left(2\pi f_c t+\phi\right)$. The product modulator output is

$$v(t)=\frac{1}{2}A_c A_c'\cos\left(2\pi 2 f_c t+\phi\right)m(t)+\frac{1}{2}A_c A_c'\cos\left(\phi\right)m(t) \quad \text{(III.10)}$$

The first term in (III.10) represents a DSB-SC modulated signal with a carrier frequency $2f_c$ whereas the second term is proportional to the baseband signal $m(t)$.



It is apparent that the first term in (III.10) is removed by the low-pass filter, provided that the cut-off frequency of this filter is greater than W but less than $2f_c - W$. This requirement is satisfied by choosing $f_c > W$. At the low-pass filter output we then obtain a signal given by

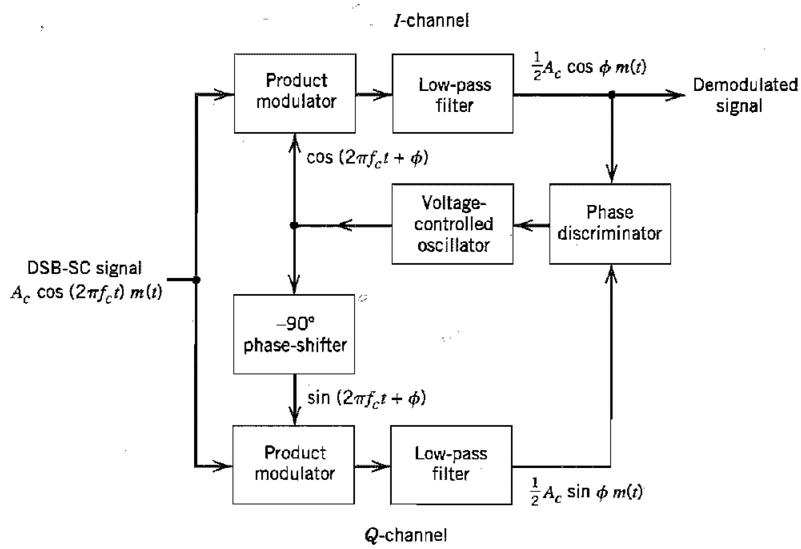
$$v_o(t) = \frac{1}{2} A_c A_c' \cos(\phi) m(t) \qquad \text{(III.11)}$$

The demodulated signal is therefore proportional to $m(t)$ when the phase error is a constant. The amplitude of this demodulated signal is maximum when $\phi = 0$, and it is minimum (zero) when $\phi = \pm\pi/2$. As long as the phase error is constant, the detector provides an undistorted version of the original baseband signal $m(t)$.

In practice, however, we usually find that the phase error varies randomly with time, due to random variations in the communication channel. The result is that at the detector output, the multiplying factor $\cos(\phi)$, also varies randomly with time, which is obviously undesirable. Therefore, provision must be made in the system to maintain the local oscillator in the receiver in perfect synchronism, in both frequency and phase, with the carrier wave used to generate the DSB-SC modulated signal in the transmitter. The resulting system complexity is the price that must be paid for suppressing the carrier wave to save transmitter power.

Costas Receiver

One method of obtaining a practical synchronous receiver system, suitable for demodulating DSB-SC waves, is to use the Costas receiver. This receiver consists of two coherent detectors supplied with the same input signal, namely, the incoming DSB-SC wave $A_c \cos(2\pi f_c t) m(t)$, but with individual local oscillator signals that are in phase quadrature with respect to each other.



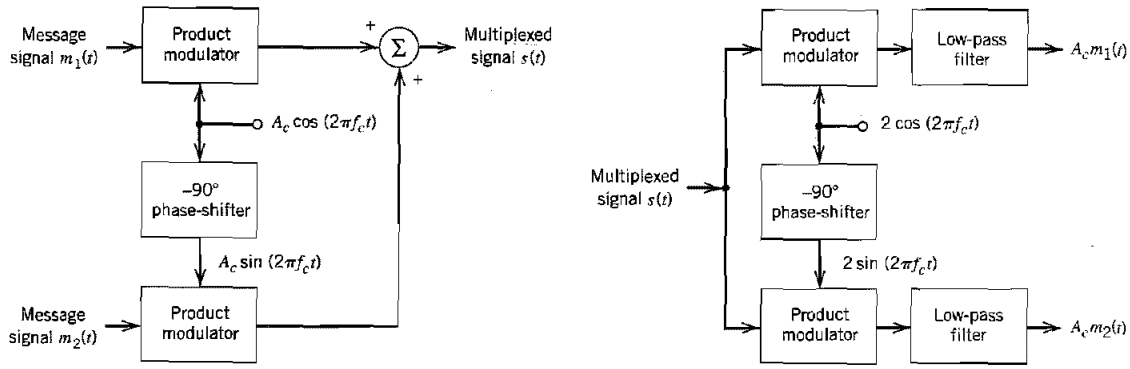
The frequency of the local oscillator is adjusted to be the same as the carrier frequency f_c which is assumed known a priori. The detector in the upper path is referred to as the in-phase coherent detector or I-channel, and that in the lower path is referred to as the quadrature-phase coherent detector or Q-channel. These two detectors are coupled together to form a negative feedback system designed in such a way as to maintain the local oscillator synchronous with the carrier wave.

Suppose that the local oscillator signal is of the same phase as the carrier wave used to generate the incoming DSB-SC wave. Under these conditions, we find that the I-channel output contains the desired demodulated signal $m(t)$, whereas the Q-channel output is zero due to the quadrature null effect of the Q-channel.

Suppose next that the local oscillator phase drifts from its proper value by a small angle ϕ radians. The I-channel output will remain essentially unchanged, but there will now be some signal appearing at the Q-channel output, which is proportional to $\sin(\phi) \approx \phi$ for small ϕ.This Q-channel output will have the same polarity as the I-channel output for one direction of local oscillator phase drift and opposite polarity for the opposite direction of local oscillator phase drift. Thus, by combining the I- and Q-channel outputs in a phase discriminator (which consists of a multiplier followed by a low-pass filter), as shown in the last figure, a DC control signal is obtained that automatically corrects for local phase errors in the voltage-controlled oscillator.

III.3.B. Quadrature Amplitude Multiplexing (QAM)

The quadrature null effect of the coherent detector may also be put to good use in the construction of the so-called quadrature-carrier multiplexing or quadrature-amplitude modulation (QAM). This scheme enables two DSB-SC modulated waves (resulting from the application of two physically independent message signals) to occupy the same channel bandwidth, and yet it allows for the separation of the two message signals at the receiver output. It is therefore a bandwidth-conservation scheme.



The transmitter part of the system involves the use of two separate product modulators that are supplied with two carrier waves of the same frequency but differing in phase by -90 degrees. The transmitted signal consists of the sum of these two product modulator outputs, as shown by

$$s(t) = A_c m_1(t)\cos(2\pi f_c t) + A_c m_2(t)\sin(2\pi f_c t) \tag{III.12}$$

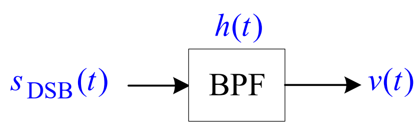
where $m_1(t)$ and $m_2(t)$ denote the two different message signals applied to the product modulators. Thus $s(t)$ occupies a channel bandwidth of $2W$ centered at carrier frequency f_c, where W is the message bandwidth of $m_1(t)$ or $m_2(t)$. We may view $A_c m_1(t)$ as the in-phase component of the multiplexed band-pass signal and $-A_c m_2(t)$ as its quadrature component.

At the receiver, the multiplexed signal $s(t)$ is applied simultaneously to two separate coherent detectors that are supplied with two local carriers of the same frequency but differing in phase by -90 degrees. The output of the top detector is $A_c m_1(t)$, whereas the output ofthe bottom detector is $A_c m_2(t)$. For the system to operate satisfactorily, it is important to maintain the correct phase and frequency relationships between the local oscillators used in the transmitter and receiver parts of the system.

To maintain this synchronization, we may send a pilot signal outside the passband of the modulated signal. In this method, the pilot signal typically consists of a low-power sinusoidal tone whose frequency and phase are related to the carrier wave $c(t)$; at the receiver, the pilot signal is extracted by means of a suitably tuned circuit and then translated to the correct frequency for use in the coherent detector.

III.4. Filtering of Sidebands

Consider the system



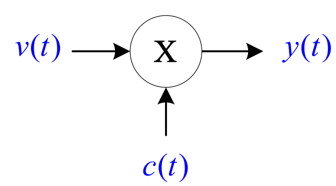
$$\begin{aligned} s_{\text{DSB}}(t) &= m(t)c(t) \\ &= A_c m(t)\cos\left(2\pi f_c t\right) \end{aligned} \tag{III.13}$$

$$S_{\text{DSB}}(f) = \frac{A_c}{2}\left[M\left(f - f_c\right) + M\left(f + f_c\right)\right] \tag{III.14}$$

Given that the bandwidth of the message signal $m(t)$ is equal to W, the bandwidth of $s_{\text{DSB}}(t)$ is equal to $2W$. To reduce the bandwidth, let $s_{\text{DSB}}(t)$ be applied to a BPF with impulse response $h(t)$. In the frequency domain, the filter output is equal to

$$V(f) = S_{\text{DSB}}(f)H(f) \tag{III.15}$$

To recover the message signal at the receiver, we start by applying the received signal $v(t)$ to a product modulator, the other input of which is $c(t)$.



The product modulator output is given by

$$\begin{aligned} y(t) &= v(t)c(t) \\ &= A_c v(t)\cos\left(2\pi f_c t\right) \end{aligned} \tag{III.16}$$

In the frequency domain we have

$$Y(f) = \frac{A_c}{2}\left[V\left(f - f_c\right) + V\left(f + f_c\right)\right] \tag{III.17}$$

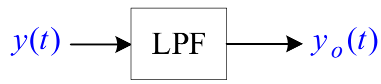
Substituting (III.15) into (III.17) yields

$$Y(f) = \frac{A_c}{2}\left[S_{\text{DSB}}(f - f_c)H(f - f_c) + S_{\text{DSB}}(f + f_c)H(f + f_c)\right] \tag{III.18}$$

Now, substituting (III.14) into (III.18) produces

$$Y(f) = \frac{A_c^2}{4}\begin{pmatrix} M(f)\left[H(f - f_c) + H(f + f_c)\right] \\ +M\left(f - 2f_c\right)H(f - f_c) \\ +M\left(f + 2f_c\right)H(f + f_c) \end{pmatrix} \tag{III.19}$$

The term in the first line in (III.19) is centered at $f = 0$ and has a bandwidth W, the term in the second line is centered at $f = 2f_c$, while the term in the third line is centered at $f = -2f_c$. Now, let's filter $y(t)$ by a LPF with bandwidth W as shown in the system below



The filter output in the frequency domain is equal to

$$Y_o(f) = \frac{A_c^2}{4} M(f)\left[H(f - f_c) + H(f + f_c)\right] \qquad \text{(III.20)}$$

To make sure that $m(t)$ can be recovered from $y_o(t)$, we must have

$$H(f - f_c) + H(f + f_c) = 1, \quad -W \le f \le W \qquad \text{(III.21)}$$

This can be written more conveniently as

$$H(f_c + f) = 1 - H(f_c - f), \quad -W \le f \le W \qquad \text{(III.22)}$$

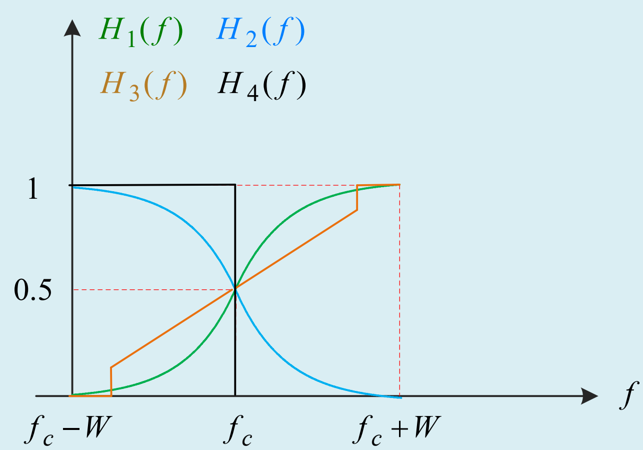
when $f = 0$ we have

$$H(f_c) = 0.5 \qquad \text{(III.23)}$$

This creates an odd symmetry in the function $H(f)$ around the point $(f, H) = (f_c, 0.5)$.

Example III.1

Functions in the below figure satisfy odd symmetry around $(f, H) = (f_c, 0.5)$.



Example III.2

In Example III.1, let $f_c = 100$ kHz, $W = 10$ kHz, $H(92\text{k}) = 0.3$ and $H(105\text{k}) = 0.8$. Then, we should have

$H(100\text{k}) = 0.5$

$H(108\text{k}) = 0.7$

$H(95\text{k}) = 0.2$

Note that from the analysis of bandpass systems we have

$$v_I(t) = \frac{1}{2}\left(s_{\text{DSB},I}(t) * h_I(t) - s_{\text{DSB},Q}(t) * h_Q(t)\right) \tag{III.24}$$

$$v_Q(t) = \frac{1}{2}\left(s_{\text{DSB},I}(t) * h_Q(t) + s_{\text{DSB},Q}(t) * h_I(t)\right) \tag{III.25}$$

where

$$s_{\text{DSB},I}(t) = A_c m(t) \tag{III.26}$$

$$s_{\text{DSB},Q}(t) = 0 \tag{III.27}$$

Substituting (III.26) and (III.27) into (III.24) yields

Type of Modulation	In-Phase Component $s_I(t)$	Quadrature Component $s_Q(t)$	Comments
DSB-SC	$m(t)$	0	$m(t)$ = message signal
SSB:[a]			
(a) Upper sideband transmitted	$\frac{1}{2}m(t)$	$\frac{1}{2}\hat{m}(t)$	$\hat{m}(t)$ = Hilbert transform of $m(t)$
(b) Lower sideband transmitted	$\frac{1}{2}m(t)$	$-\frac{1}{2}\hat{m}(t)$	
VSB:			
(a) Vestige of lower sideband transmitted	$\frac{1}{2}m(t)$	$\frac{1}{2}m'(t)$	$m'(t)$ = output of the filter of frequency response $H_Q(f)$ due to $m(t)$. For the definition of $H_Q(f)$, see Eq. (2.16)
(b) Vestige of upper sideband transmitted	$\frac{1}{2}m(t)$	$-\frac{1}{2}m'(t)$	

$$v_I(t) = \frac{A_c}{2} m(t) * h_I(t) \tag{III.28}$$

Substituting (III.26) and (III.27) into (III.25) yields

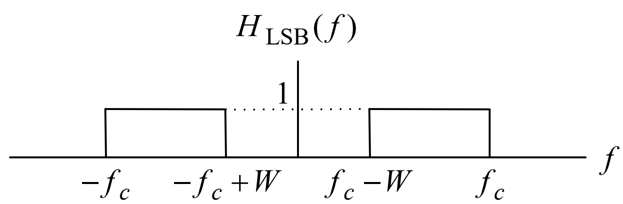
$$v_Q(t) = \frac{A_c}{2} m(t) * h_Q(t) \tag{III.29}$$

III.4.A. SINGLE-SIDEBAND MODULATION

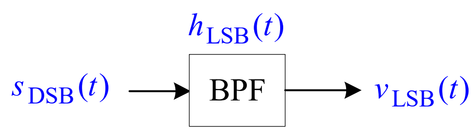
In single-sideband modulation, only the upper or lower sideband is transmitted. We may generate such a modulated wave by using the frequency-discrimination method that consists of two stages:

- The first stage is a product modulator, which generates a DSB-SC modulated wave.
- The second stage is a band-pass filter, which is designed to pass one of the sidebands of this modulated wave and suppress the other.

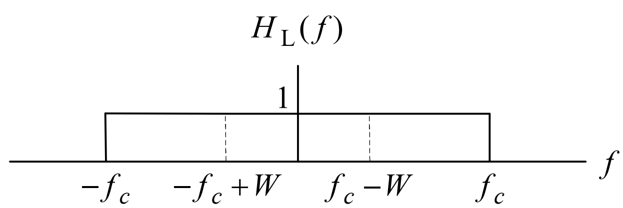
Consider the BPF that generates an LSB signal from a DSB-SC signal, shown below.



Consider the below system, where a DSB-SC signal is filtered by the bandpass filter with the impulse response $h_{\text{LSB}}(t)$.



It can be easily see that the same LSB signal generated by this filter can be generated by the LPF shown below.

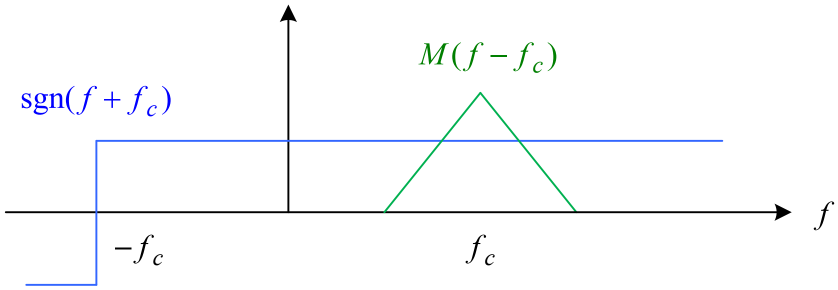


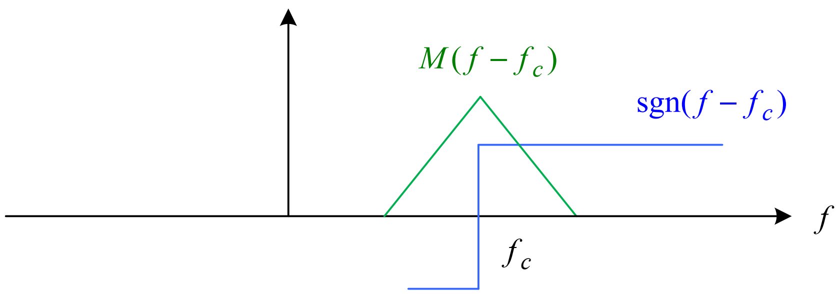
It can be easily seen that

$$\begin{aligned} H_{\text{L}}(f) &= \text{rect}\left(\frac{f}{2f_c}\right) \\ &= \frac{1}{2}\left[\text{sgn}\left(f+f_c\right)-\text{sgn}\left(f-f_c\right)\right] \end{aligned} \tag{III.30}$$

Substituting (III.30) and (III.14) into (III.15) yields

$$V_{\text{LSB}}(f) = \frac{A_c}{4}\left[M\left(f-f_c\right)+M\left(f+f_c\right)\right]\left[\text{sgn}\left(f+f_c\right)-\text{sgn}\left(f-f_c\right)\right] \tag{III.31}$$





$$V_{\text{LSB}}(f) = \frac{A_c}{4}\left[M\left(f - f_c\right) + M\left(f + f_c\right)\right] + \frac{A_c}{4}\left[M\left(f + f_c\right)\text{sgn}\left(f + f_c\right) - M\left(f - f_c\right)\text{sgn}\left(f - f_c\right)\right] \quad \text{(III.32)}$$

To determine $v_{\text{LSB}}(t)$, we first that

$$\frac{A_c}{2} m(t)\cos\left(2\pi f_c t\right) \overset{\mathcal{F}}{\rightleftarrows} \frac{A_c}{4}\left[M\left(f - f_c\right) + M\left(f + f_c\right)\right] \quad \text{(III.33)}$$

Now, since

$$\hat{m}(t) \overset{\mathcal{F}}{\rightleftarrows} -j\,\text{sgn}(f)M(f) \quad \text{(III.34)}$$

Then

$$\hat{m}(t)e^{j2\pi f_c t} \overset{\mathcal{F}}{\rightleftarrows} -j\,\text{sgn}(f - f_c)M(f - f_c) \quad \text{(III.35)}$$

$$\hat{m}(t)e^{-j2\pi f_c t} \overset{\mathcal{F}}{\rightleftarrows} -j\,\text{sgn}(f + f_c)M(f + f_c) \quad \text{(III.36)}$$

Using (III.33), (III.35) and (III.36) to apply the inverse Fourier transform to both sides of (III.32), we get

$$v_{\text{LSB}}(t) = \frac{A_c}{2} m(t)\cos(2\pi f_c t) + \frac{A_c}{j4}\hat{m}(t)\left[e^{j2\pi f_c t} - e^{-j2\pi f_c t}\right]$$
$$= \frac{A_c}{2} m(t)\cos(2\pi f_c t) + \frac{A_c}{2}\hat{m}(t)\sin(2\pi f_c t) \quad \text{(III.37)}$$

Note that

$$v_{\text{LSB},\,I}(t) = \frac{A_c}{2} m(t) \quad \text{(III.38)}$$

$$v_{\text{LSB},\,Q}(t) = -\frac{A_c}{2}\hat{m}(t) \quad \text{(III.39)}$$

The last result implies that

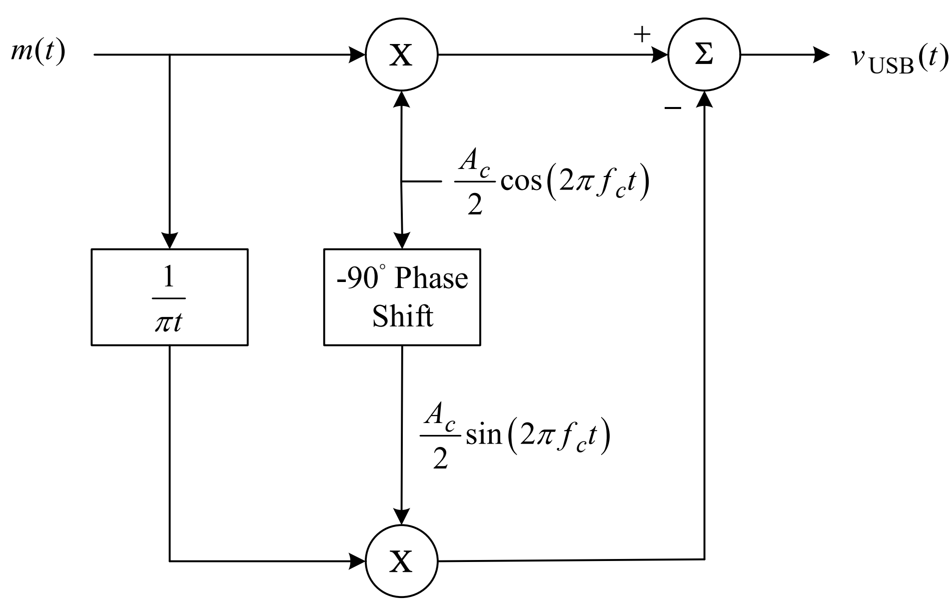
$$h_{\text{LSB},\,Q}(t) = -\frac{1}{\pi t} \quad \text{(III.40)}$$

A similar procedure leads to the following expression of a USB signal.

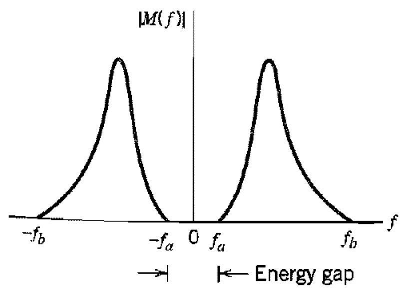
$$v_{\text{USB}}(t) = \frac{A_c}{2} m(t)\cos(2\pi f_c t) - \frac{A_c}{2}\hat{m}(t)\sin(2\pi f_c t) \quad \text{(III.41)}$$

The last result implies that

$$h_{\text{USB},\,Q}(t) = \frac{1}{\pi t} \quad \text{(III.42)}$$



From a practical viewpoint the most severe requirement of SSB generation arises from the unwanted sideband. The nearest frequency component of the unwanted sideband is separated from the desired sideband by twice the lowest frequency component of the message (modulating) signal. The implication here is that for the generation of an SSB modulated signal to be possible, the message spectrum must have an energy gap centered at the origin. This requirement is naturally satisfied by voice signals, whose energy gap is about 600 Hz wide (i.e., it extends from - 300 to + 300 Hz).

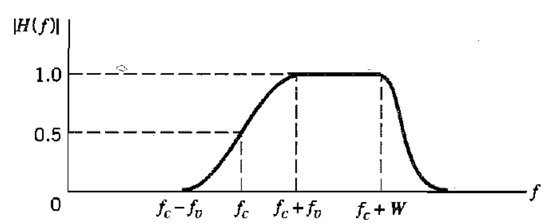


To demodulate a SSB modulated signal $s(t)$, we may use a coherent detector, which multiplies $s(t)$ by a locally generated carrier and then low-pass filters the product. This method of demodulation assumes perfect synchronism between the oscillator in the coherent detector and the oscillator used to supply the carrier wave in the transmitter.

It is inevitable that there will always be some phase error in the local oscillator output with respect to the carrier wave used to generate the incoming SSB modulated wave. The effect of this phase error is to introduce a phase distortion in the demodulated signal, where each frequency component of the original message signal undergoes a constant phase shift. This phase distortion is tolerable in voice communications, because the human ear is relatively insensitive to phase distortion. In the transmission of music and video signals, on the other hand, the presence of this form of waveform distortion is unacceptable.

III.4.B. Vestigial Sideband Modulation

In vestigial sideband (VSB) modulation, one of the sidebands is partially suppressed and a vestige of the other sideband is transmitted to compensate for that suppression. A popular method is to first generate a DSB-SC modulated wave and then pass it through a bandpass filter. It is the special design of the band-pass filter that distinguishes VSB modulation from SSB modulation. Assuming that a vestige of the lower sideband is transmitted, the frequency response $H(f)$ of the band-pass filter takes the form shown below.



This frequency response is normalized, so that at the carrier frequency we have $H(f_c)=1/2$. The important feature to note is that the cutoff portion of the frequency response around the carrier frequency exhibits odd symmetry. That is, inside the transition interval $f_c - f_v \le |f| \le f_c + f_v$, the following two conditions are satisfied:

- The sum of the values of the magnitude response $|H(f)|$ at any two frequencies equally displaced above and below f_c is unity.
- The phase response is linear.

That is, $H(f)$ satisfies the condition for $-W \le f \le W$

$$H(f - f_c) + H(f + f_c) = 1 \tag{III.43}$$

Note also that outside the frequency band of interest (i.e., $|f| > f_c + W$), the frequency response $H(f)$ may have an arbitrary specification. Accordingly, the transmission bandwidth of VSB modulation is

$$B_T = W + f_v \tag{III.44}$$

The VSB modulated wave is described in the time domain as

$$s(t) = \frac{1}{2} A_c m(t)\cos(2\pi f_c t) \pm \frac{1}{2} A_c m'(t)\sin(2\pi f_c t) \tag{III.45}$$

$$m'(t) = m(t) * h_Q(t) \tag{III.46}$$

where the plus sign corresponds to the transmission of a vestige of the upper sideband, and the minus sign corresponds to the transmission of a vestige of the lower sideband. The signal $m'(t)$ in the quadrature component of $s(t)$ is obtained by passing the message signal $m(t)$ through a filter whose frequency response $H_Q(f)$ satisfies the following requirement for $-W \le f \le W$

$$H_Q(f) = j\left[H(f - f_c) - H(f + f_c)\right] \tag{III.47}$$

It is of interest to note that SSB modulation may be viewed as a special case of VSB modulation. Specifically, when the vestigial sideband is reduced to zero (i.e., we set $f_v = 0$), the modulated takes the limiting form of a single sideband modulated wave. Note that this case we also have

$$m'(t) = \hat{m}(t) \tag{III.48}$$

III.5. Frequency Translation

The basic operation involved in single-sideband modulation is in fact a form of frequency translation, which is why single-sideband modulation is sometimes referred to as frequency mixing, or heterodyning.

The idea of frequency translation described herein may be generalized as follows. Suppose that we have a modulated wave $s_1(t)$ whose spectrum is centered on a carrier frequency f_1, and the requirement is to translate it upward in frequency such that its carrier frequency is changed from f_1, to a new value f_2. This requirement may be accomplished using a mixer. The mixer consists of a product modulator followed by a band-pass filter.

Up Conversion

The translated carrier frequency f_2 is greater than the incoming carrier frequency f_1.

Down Conversion

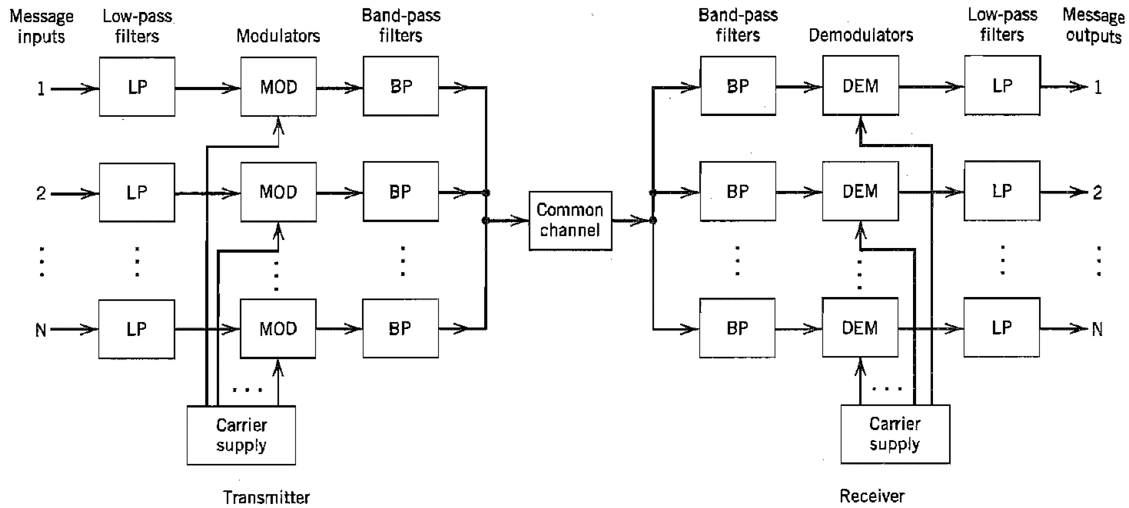
The translated carrier frequency f_2 is smaller than the incoming carrier frequency f_1.

III.6. Frequency-Division Multiplexing (FDM)

Another important signal processing operation is multiplexing, whereby a number of independent signals can be combined into a composite signal suitable for transmission over a common channel. Voice frequencies transmitted over telephone systems, for example, range from 300 to 3100 Hz. To transmit a number of these signals over the same channel, the signals must be kept apart so that they do not interfere with each other, and thus they can be separated at the receiving end. This is accomplished by separating the signals either in frequency or in time. The technique of separating the signals in frequency is referred to as frequency-division multiplexing (FDM).

A block diagram of an FDM system is shown below. The incoming message signals are assumed to be of the low-pass type, but their spectra do not necessarily have nonzero values all the way down to zero frequency. Following each signal input, we have shown a low-pass filter, which is designed to remove high-frequency components that do not contribute significantly to signal representation but are capable of disturbing other message signals that share the common channel. These low-pass filters may be omitted only if the input signals are sufficiently band limited initially. The filtered signals are applied to modulators that shift the frequency ranges of the signals so as to occupy mutually exclusive frequency intervals. The necessary carrier frequencies needed to perform these frequency translations are obtained from a carrier supply.

The most widely used method of modulation in frequency-division multiplexing is single sideband modulation, which, in the case of voice signals, requires a bandwidth that is approximately equal to that of the original voice signal. In practice, each voice input is usually assigned a bandwidth of 4 kHz. The band-pass filters following the modulators are used to restrict the band of each modulated wave to its prescribed range. The resulting bandpass filter outputs are next combined in parallel to form the input to the common channel.



At the receiving terminal, a bank of band-pass filters, with their inputs connected in parallel, is used to separate the message signals on a frequency-occupancy basis. Finally, the original message signals are recovered by individual demodulators. Note that the FDM system shown above operates in only one direction. To provide for two-way transmission, as in telephony, for example, we have to completely duplicate the multiplexing facilities, with the components connected in reverse order and with the signal waves proceeding from right to left.

***

IV. ANGLE MODULATION

In angle modulation the angle of the carrier wave is varied according to the baseband signal. In this method of modulation, the amplitude of the carrier wave is maintained constant. An important feature of angle modulation is that it can provide better discrimination against noise and interference than amplitude modulation. This improvement in performance is achieved at the expense of increased transmission bandwidth; that is, angle modulation provides us with a practical means of exchanging channel bandwidth for improved noise performance. Such a tradeoff is not possible with amplitude modulation, regardless of its form.

Let $\theta_i(t)$ denote the instantaneous angle of a modulated sinusoidal carrier, assumed to be a function of the message signal. We express the resulting angle-modulated wave as

$$s(t) = A_c \cos\left(\theta_i(t)\right) \tag{IV.1}$$

where A_c is the carrier amplitude.

We define the instantaneous frequency of the angle-modulated signal $s(t)$ as follows:

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

In the simple case of an unmodulated carrier, the angle $\theta_i(t)$ is

$$\theta_i(t) = 2\pi f_c t + \phi_c \tag{IV.3}$$

There are an infinite number of ways in which the angle $\theta_i(t)$ may be varied in some manner with the message (baseband) signal. However, we shall consider only two commonly used methods, phase modulation and frequency modulation, defined as follows:

1. Phase modulation (PM) is that form of angle modulation in which the angle $\theta_i(t)$ is varied linearly with the message signal $m(t)$, as shown by

$$\theta_i(t) = 2\pi f_c t + k_p m(t) \tag{IV.4}$$

The term $2\pi f_c t$ represents the angle of the unmodulated carrier; and the constant k_p represents the phase sensitivity of the modulator, expressed in radians per volt on the assumption that $m(t)$ is a voltage waveform. For convenience, we have assumed that the angle of the unmodulated carrier is zero at $t = 0$. The phase-modulated signal $s(t)$ is thus described in the time domain by

$$s(t) = A_c \cos\left(2\pi f_c t + k_p m(t)\right) \tag{IV.5}$$

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

$$f_i(t) = f_c + k_f m(t) \tag{IV.6}$$

The term f_c represents the frequency of the unmodulated carrier, and the constant k_f represents the frequency sensitivity of the modulator, expressed in Hertz per volt on the assumption that $m(t)$ is a voltage waveform.

Integrating (IV.6) with respect to time and multiplying the result by 2π, we get

$$\theta_i(t) = 2\pi f_c t + 2\pi k_f \int m(t)dt \tag{IV.7}$$

where, for convenience, we have assumed that the angle of the unmodulated carrier wave is zero at $t = 0$. The frequency-modulated signal is therefore described in the time domain by

$$s(t) = A_c \cos\left(2\pi f_c t + 2\pi k_f \int m(t)dt\right) \tag{IV.8}$$

A consequence of allowing the angle $\theta_i(t)$ to become dependent on the message signal $m(t)$ as in (IV.4) or on its integral as in Equation (IV.7) is that the zero crossings of a PM signal or FM signal no longer have a perfect regularity in their spacing; zero crossings refer to the instants of time at which a waveform changes from a negative to a positive value or vice versa. This is one important feature that distinguishes both PM and FM signals from an AM signal. Another important difference is that the envelope of a PM or FM signal is constant (equal to the carrier amplitude), whereas the envelope of an AM signal is dependent on the message signal.

Comparing (IV.5) with (IV.8) reveals that an FM signal may be regarded as a PM signal in which the modulating wave is $\int m(t)dt$ in place of $m(t)$. This means that an FM signal can be generated by first integrating $m(t)$ and then using the result as the input to a phase modulator. Conversely, a PM signal can be generated by first differentiating $m(t)$ and then using the result as the input to a frequency modulator. We may thus deduce all the properties of PM signals from those of FM signals and vice versa. Henceforth, we concentrate our attention on FM signals.

IV.1. Frequency Modulation

The FM signal $s(t)$ defined by (IV.8) is a nonlinear function of the modulating signal $m(t)$, which makes frequency modulation a nonlinear modulation process. Consequently, unlike amplitude modulation, the spectrum of an FM signal is not related in a simple manner to that of the modulating signal; rather, its analysis is much more difficult than that of an AM signal.

- We consider the simplest case possible, namely, that of a single-tone modulation that produces a narrowband FM signal.
- We next consider the more general case also involving a single-tone modulation, but this time the FM signal is wideband.

Consider a sinusoidal modulating signal defined by

$$m(t) = A_m \cos\left(2\pi f_m t\right) \quad \text{(IV.9)}$$

The instantaneous frequency of the resulting FM signal equals

$$\begin{aligned} f_i(t) &= f_c + k_f A_m \cos\left(2\pi f_m t\right) \\ &= f_c + \Delta f \cos\left(2\pi f_m t\right) \end{aligned} \quad \text{(IV.10)}$$

where

$$\Delta f = k_f A_m \quad \text{(IV.11)}$$

The quantity Δf is called the frequency deviation, representing the maximum departure of the instantaneous frequency of the FM signal from the carrier frequency f_c. A fundamental characteristic of an FM signal is that the frequency deviation Δf is proportional to the amplitude of the modulating signal and is independent of the modulation frequency.

The angle $\theta_i(t)$ of the FM signal is obtained as

$$\begin{aligned} \theta_i(t) &= 2\pi \int f_i(t)dt \\ &= 2\pi f_c t + \frac{\Delta f}{f_m} \sin\left(2\pi f_m t\right) \end{aligned} \quad \text{(IV.12)}$$

The ratio of the frequency deviation Δf to the modulation frequency f_m is commonly called the modulation index of the FM signal. We denote it by β:

$$\begin{aligned} \beta &= \frac{\Delta f}{f_m} \\ &= \frac{k_f A_m}{f_m} \end{aligned} \quad \text{(IV.13)}$$

This leads to

$$\theta_i(t) = 2\pi f_c t + \beta \sin\left(2\pi f_m t\right) \quad \text{(IV.14)}$$

The parameter β represents the phase deviation of the FM signal, that is, the maximum departure of the angle $\theta_i(t)$ from the angle $2\pi f_c t$ of the unmodulated carrier; hence, β is measured in radians.

The FM signal itself is given by

$$s(t) = A_c \cos\left(2\pi f_c t + \beta \sin\left(2\pi f_m t\right)\right) \quad \text{(IV.15)}$$

Depending on the value of the modulation index β, we may distinguish two cases of frequency modulation:

- Narrowband FM, for which β is small compared to one radian.
- Wideband FM, for which β is large compared to one radian.

IV.1.A. NARROWBAND FREQUENCY MODULATION

Note that

$$s(t) = A_c \cos\left(2\pi f_c t\right)\cos\left(\beta \sin\left(2\pi f_m t\right)\right) - A_c \sin\left(2\pi f_c t\right)\sin\left(\beta \sin\left(2\pi f_m t\right)\right) \quad \text{(IV.16)}$$

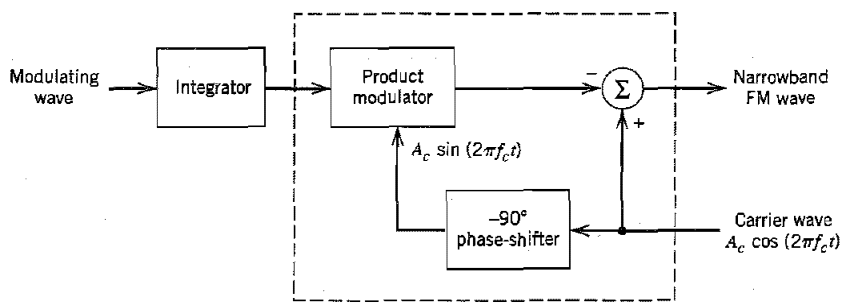
Assuming that the modulation index is small compared to one radian, we may use the following approximations:

$$\cos\left(\beta \sin\left(2\pi f_m t\right)\right) \approx 1 \quad \text{(IV.17)}$$

$$\sin\left(\beta \sin\left(2\pi f_m t\right)\right) \approx \beta \sin\left(2\pi f_m t\right) \quad \text{(IV.18)}$$

Hence, (IV.16) simplifies to

$$s(t) \approx A_c \cos\left(2\pi f_c t\right) - \beta A_c \sin\left(2\pi f_m t\right)\sin\left(2\pi f_c t\right) \quad \text{(IV.19)}$$



Ideally, an FM signal has a constant envelope and, for the case of a sinusoidal modulating signal of frequency f_m the angle $\theta_i(t)$ is also sinusoidal with the same frequency.

Equation (IV.19) may be expanded as follows:

$$s(t) \approx A_c \cos\left(2\pi f_c t\right) + \frac{1}{2}\beta A_c \left[\cos\left(2\pi\left(f_c + f_m\right)t\right) - \cos\left(2\pi\left(f_c - f_m\right)t\right)\right] \quad \text{(IV.20)}$$

This expression is somewhat similar to the corresponding one defining an AM signal. The basic difference between an AM signal and a narrowband FM signal is that the algebraic sign of the lower side frequency in the narrowband FM is reversed. Thus, a narrowband FM signal requires essentially the same transmission bandwidth (i.e., $2f_m$) as the AM signal.

IV.1.B. WIDEBAND FREQUENCY MODULATION

In general, an FM signal produced by a sinusoidal modulating signal, is in itself nonperiodic unless the carrier frequency f_c is an integral multiple of the modulation frequency f_m. However, we may simplify matters by using the complex representation of bandpass signals. Specifically, we assume that the carrier frequency is large enough (compared to the bandwidth of the FM signal) to justify rewriting this equation in the form

$$\begin{aligned} s(t) &= \mathrm{Re}\left\{A_c e^{j2\pi f_c t + j\beta \sin(2\pi f_m t)}\right\} \\ &= \mathrm{Re}\left\{\tilde{s}(t) e^{j2\pi f_c t}\right\} \end{aligned} \quad \text{(IV.21)}$$

where

$$\tilde{s}(t) = A_c e^{j\beta \sin(2\pi f_m t)} \quad \text{(IV.22)}$$

Thus, unlike the original FM signal $s(t)$, the complex envelope $\tilde{s}(t)$ is a periodic function of time with a fundamental frequency equal to the modulation frequency f_m. We may therefore expand $\tilde{s}(t)$ in the form of a complex Fourier series as follows:

$$\tilde{s}(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_m t} \quad \text{(IV.23)}$$

where the complex Fourier coefficients are defined by

$$\begin{aligned} c_n &= f_m \int_{-1/2f_m}^{1/2f_m} \tilde{s}(t) e^{-j2\pi n f_m t} dt \\ &= f_m A_c \int_{-1/2f_m}^{1/2f_m} e^{\left[j\beta \sin(2\pi f_m t) - j2\pi n f_m t\right]} dt \end{aligned} \quad \text{(IV.24)}$$

Define a new variable:

$$x = 2\pi f_m t \quad \text{(IV.25)}$$

Hence, we may rewrite (IV.24) in the new form

$$c_n = \frac{A_c}{2\pi} \int_{-\pi}^{\pi} e^{j[\beta \sin(x) - nx]} dx \quad \text{(IV.26)}$$

The integral on the right-hand side of (IV.26), except for a scaling factor, is recognized as the n^{th} order Bessel function of the first kind. This function is commonly denoted $J_n(\beta)$, as shown by

$$J_n(\beta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j[\beta \sin(x) - nx]}dx \tag{IV.27}$$

Accordingly, we may reduce (IV.26) to

$$c_n = A_c J_n(\beta) \tag{IV.28}$$

Using (IV.28), the complex envelope of the FM signal can be written in the form

$$\tilde{s}(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi n f_m t} \tag{IV.29}$$

Therefore, the FM signal becomes

$$s(t) = A_c \operatorname{Re}\left\{ \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi(f_c + n f_m)t} \right\} \tag{IV.30}$$

Interchanging the order of summation and evaluation of the real part in the right-hand side of (IV.30), we finally get

$$s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos\left(2\pi\left(f_c + n f_m\right)t\right) \tag{IV.31}$$

$$\begin{aligned} P_s &= \frac{A_c^2}{2} \sum_{n=-\infty}^{\infty} J_n^2(\beta) \\ &= \frac{A_c^2}{2} \end{aligned}$$

This is the desired form for the Fourier series representation of the single-tone FM signal. The discrete spectrum of $s(t)$ is obtained by taking the Fourier transforms of both sides of (IV.31); we thus have

$$S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta)\left[\delta\left(f - f_c - n f_m\right) + \delta\left(f + f_c + n f_m\right)\right] \tag{IV.32}$$

The Bessel function $J_n(\beta)$ has the following properties

$$J_n(\beta) = (-1)^n J_{-n}(\beta) \tag{IV.33}$$

For small values of the modulation index β, we have the following three approximations

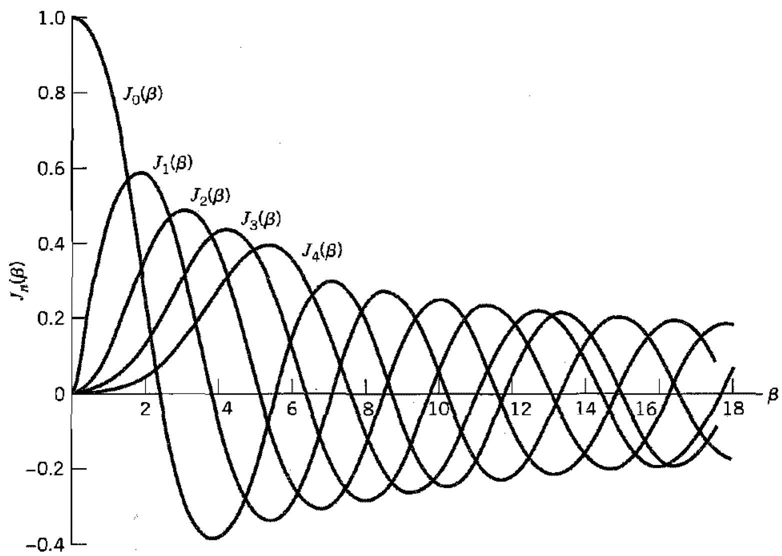
$$J_0(\beta) \approx 1 \tag{IV.34}$$

$$J_1(\beta) \approx \frac{\beta}{2} \qquad \text{(IV.35)}$$

$$J_n(\beta) \approx 0, \quad n \geq 2 \qquad \text{(IV.36)}$$

The third property is

$$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1 \qquad \text{(IV.37)}$$



From the figure above, it is very important to note that

$$J_0(2.4) \approx 0 \qquad \text{(IV.38)}$$

Note that the carrier component amplitude in (IV.31) is $A_c J_0(\beta)$.

IV.1.C. Properties of the Wideband FM Signal

- The spectrum of an FM signal contains a carrier component and an infinite set of side frequencies located symmetrically on either side of the carrier at frequency separations of $f_m, 2f_m, 3f_m, \cdots$. In this respect, the result is unlike that which prevails in an AM system, since in an AM system a sinusoidal modulating signal gives rise to only one pair of side frequencies.
- For the special case of β small compared with unity, only the Bessel coefficients $J_0(\beta)$ and $J_1(\beta)$ have significant values, so that the FM signal is effectively composed of a carrier and a single pair of side frequencies at $f_c \pm f_m$. This situation corresponds to the special case of narrowband FM that was considered earlier.

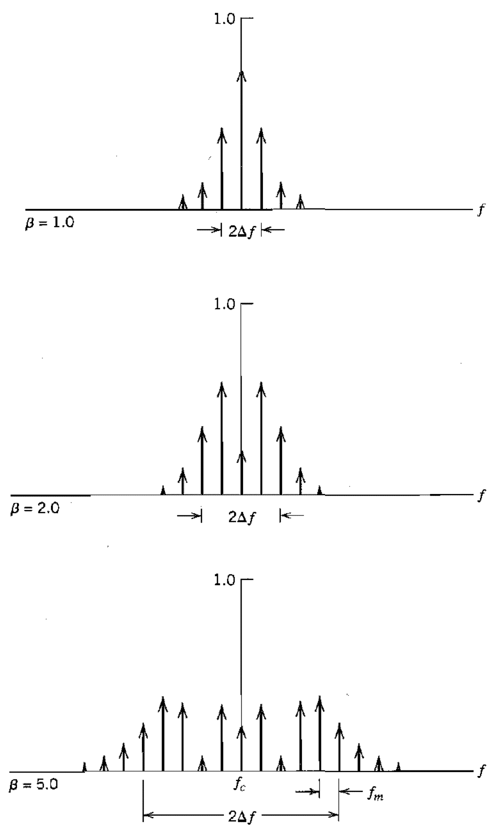
- The amplitude of the carrier component varies with β according to $J_0(\beta)$. That is, unlike an AM signal, the amplitude of the carrier component of an FM signal is dependent on the modulation index. The physical explanation for this property is that the envelope of an FM signal is constant, so that the average power of such a signal developed across a 1-ohm resistor is also constant, as shown by

$$P = \frac{A_c^2}{2} \tag{IV.39}$$

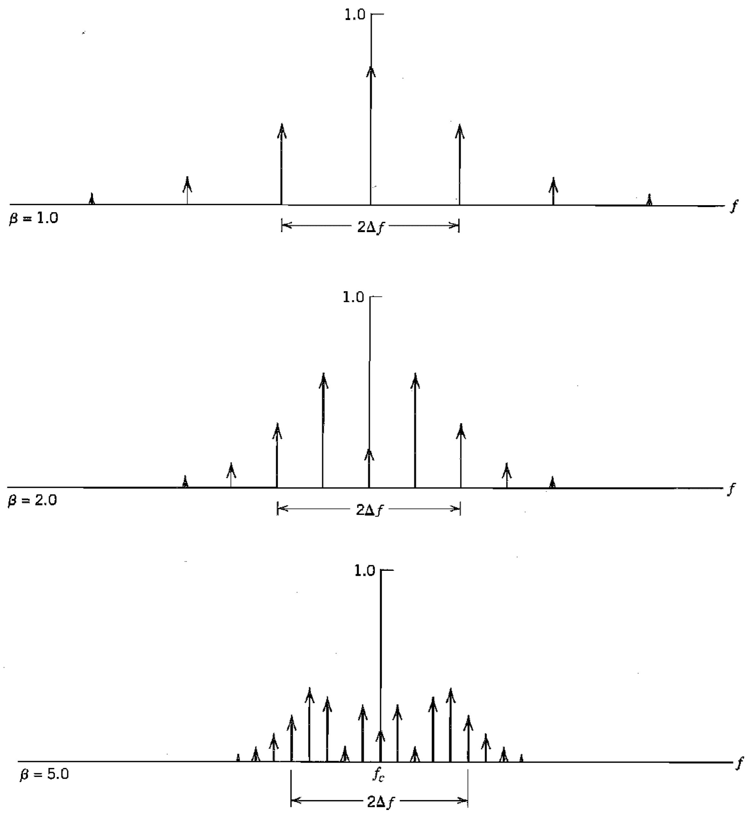
When the carrier is modulated to generate the FM signal, the power in the side frequencies may appear only at the expense of the power originally in the carrier, thereby making the amplitude of the carrier component dependent on β. Note that the average power of an FM signal may also be determined from Equation (IV.31), obtaining

$$\begin{aligned} P &= \frac{A_c^2}{2} \sum_{n=-\infty}^{\infty} J_n^2(\beta) \\ &= \frac{A_c^2}{2}\left[J_0^2(\beta) + 2\sum_{n=1}^{\infty} J_n^2(\beta) \right] \end{aligned} \tag{IV.40}$$

Effect of Modulating Signal Amplitude on Spectrum



Effect of Modulating Signal Frequency on Spectrum



IV.1.D. Transmission Bandwidth of FM Signals

In theory, an FM signal contains an infinite number of side frequencies so that the bandwidth required to transmit such a signal is similarly infinite in extent. In practice, however, we find that the FM signal is effectively limited to a finite number of significant side frequencies compatible with a specified amount of distortion. We may therefore specify an effective bandwidth required for the transmission of an FM signal. Consider first the case of an FM signal generated by a single-tone modulating wave of frequency f_m. In such an FM signal, the side frequencies that are separated from the carrier frequency f_c by an amount greater than the frequency deviation Δf decrease rapidly toward zero, so that the bandwidth always exceeds the total frequency excursion, but nevertheless is limited. Specifically, for large values of the modulation index, the bandwidth approaches, and is only slightly greater than, the total frequency excursion $2\Delta f$. On the other hand, for small values of the modulation index, the spectrum of the FM signal is effectively limited to the carrier frequency f_c and one pair of side frequencies at $f_c \pm f_m$ so that the bandwidth approaches $2f_m$. We may thus define an approximate rule for the transmission bandwidth of an FM signal generated by a single-tone modulating signal of frequency f_m as follows:

$$\begin{aligned} B_T &\approx 2\Delta f + 2 f_m \\ &= 2\Delta f\left(1+\frac{1}{\beta}\right) \end{aligned} \tag{IV.41}$$

This empirical relation is known as **Carson's rule**.

Consider next the more general case of an arbitrary modulating signal $m(t)$ with its highest frequency component denoted by W. The bandwidth required to transmit an FM signal generated by this modulating signal is estimated by using a worst-case tone modulation analysis. Specifically, we first determine the so-called deviation ratio D, defined as the ratio of the frequency deviation Δf, which corresponds to the maximum possible amplitude of the modulation signal $m(t)$, to the highest modulation frequency W; these conditions represent the extreme cases possible.

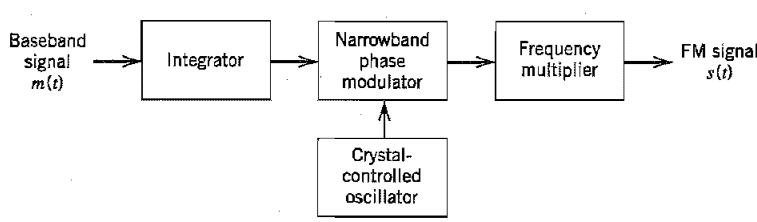
$$D = \frac{\Delta f}{A_{\max}} \tag{IV.42}$$

The deviation ratio D plays the same role for non-sinusoidal modulation that the modulation index β plays for the case of sinusoidal modulation. Then, replacing β by D and replacing f_m with W, we may use Carson's rule to obtain a value for the transmission bandwidth of the FM signal. From a practical viewpoint, Carson's rule somewhat underestimates the bandwidth requirement of an FM signal.

IV.2. Generation of FM Signals

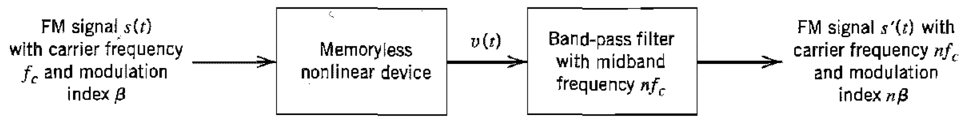
There are essentially two basic methods of generating frequency-modulated signals, namely, direct FM and indirect FM. In the direct method the carrier frequency is directly varied in accordance with the input baseband signal, which is readily accomplished using a voltage-controlled oscillator. In the indirect method, the modulating signal is first used to produce a narrowband FM signal, and frequency multiplication is next used to increase the frequency deviation to the desired level. The indirect method is the preferred choice for frequency modulation when the stability of carrier frequency is of major concern as in commercial radio broadcasting.

IV.2.A. Indirect FM



The message (baseband) signal is first integrated and then used to phase-modulate a crystal-controlled oscillator; the use of crystal control provides frequency stability. To minimize the distortion inherent in the phase modulator, the maximum phase deviation or modulation index is kept small, thereby resulting in a narrowband FM signal. The narrowband FM signal is next multiplied in frequency by means of a frequency multiplier so as to produce the desired wide band FM signal.

A frequency multiplier consists of a nonlinear device followed by a band-pass filter.



The implication of the nonlinear device being memoryless is that it has no energy-storage elements. The input-output relation of such a device may be expressed in the general form

$$v(t) = a_1 s(t) + a_2 s^2(t) + \cdots + a_n s^n(t) + \cdots \tag{IV.43}$$

The input $s(t)$ is an FM signal defined by

$$s(t) = A_c \cos\left(2\pi f_c t + 2\pi k_f \int m(t)dt\right) \tag{IV.44}$$

whose instantaneous frequency is

$$f_i(t) = f_c + k_f m(t) \tag{IV.45}$$

The mid-band frequency of the band-pass filter is set equal to nf_c. Moreover, the band-pass filter is designed to have a bandwidth equal to n times the transmission bandwidth of $s(t)$. After band-pass filtering of the nonlinear device's output $v(t)$, we have a new FM signal defined by

$$s'(t) = A'_c \cos\left(2\pi n f_c t + 2\pi n k_f \int m(t)dt\right) \tag{IV.46}$$

whose instantaneous frequency is

$$f'_i(t) = nf_c + nk_f m(t) \tag{IV.47}$$

Note that

$$\beta' = n\beta \tag{IV.48}$$

IV.2.B. Direct FM

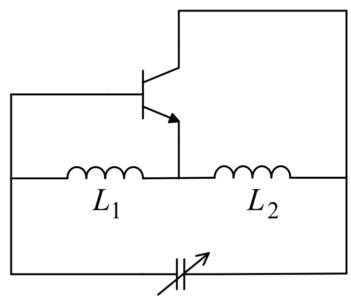
Hartley Oscillator

The circuit below shows a voltage-controlled Hartley oscillator. The capacitance $C(t)$ is linearly proportional to the message signal $m(t)$. This capacitance is most often called a varactor or a varicap.

For single-tone modulation, we can assume that $C(t)$ and $m(t)$ are related as follows:

$$C(t) = C_0 + \Delta C \cos\left(2\pi f_m t\right) \tag{IV.49}$$

where C_0 is the value of the capacitance when $m(t)$ is not present or equal to zero, and ΔC is a constant that is proportional to A_m.



The oscillator instantaneous frequency is given by:

$$f_i(t) = \frac{1}{2\pi\sqrt{\left(L_1 + L_2\right)C(t)}} \tag{IV.50}$$

This can be rewritten in the form

$$f_i(t) = f_0 \frac{1}{\sqrt{1 + \frac{\Delta C}{C_0}\cos\left(2\pi f_m t\right)}} \tag{IV.51}$$

where

$$f_0 = \frac{1}{2\pi\sqrt{\left(L_1 + L_2\right)C_0}} \tag{IV.52}$$

where f_0 is the free-running frequency of the oscillator. When $\Delta C << C_0$, we can use the approximation

$$\begin{aligned} &\frac{1}{\sqrt{1+x}} \approx 1 - \frac{x}{2}, \text{ for } x << 1 \\ &\sqrt{1+x} \approx 1 + \frac{x}{2} \end{aligned} \tag{IV.53}$$

Therefore,

$$f_i(t) = f_0\left(1 - \frac{\Delta C}{2C_0}\cos\left(2\pi f_m t\right)\right) \tag{IV.54}$$

Let's compare this formula with the general equation for the FM instantaneous frequency, while replacing f_0 with f_c.

$$f_i(t) = f_c + k_f m(t) \tag{IV.55}$$

The result is

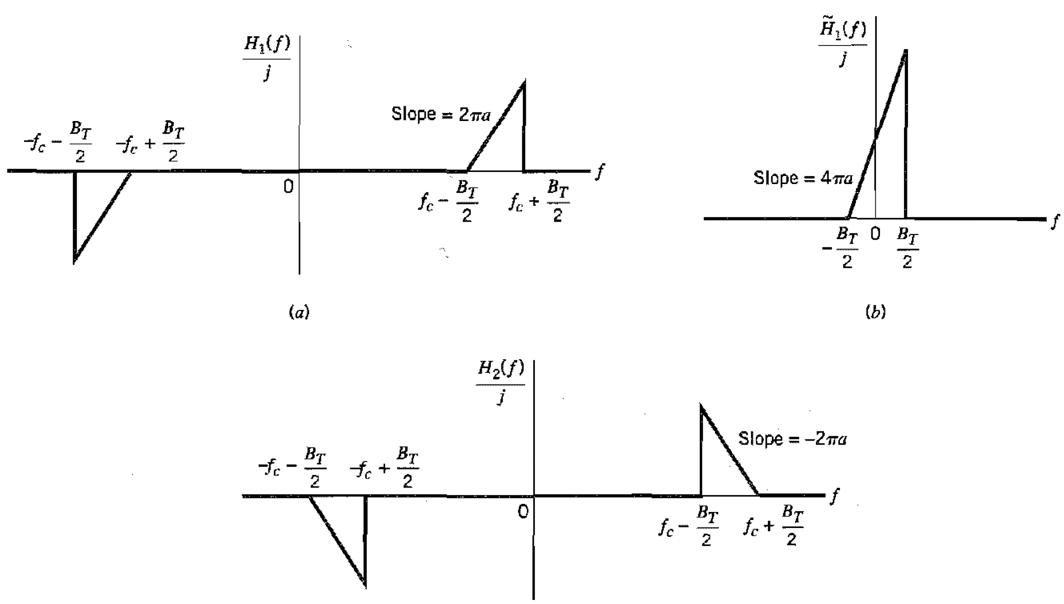
$$k_f = -\frac{\Delta C}{2C_0} f_0 \tag{IV.56}$$

IV.3. FM Demodulation

The objective is to produce a transfer characteristic that is the inverse of that of the frequency modulator, which can be realized directly or indirectly. Here we describe a direct method of frequency demodulation involving the use of a popular device known as a frequency discriminator, whose instantaneous output amplitude is directly proportional to the instantaneous frequency of the input FM signal.

Frequency Discriminator

Basically, the frequency discriminator consists of a slope circuit followed by an envelope detector. An ideal slope circuit is characterized by a frequency response that is purely imaginary, varying linearly with frequency inside a prescribed frequency interval.



(c)

Consider the frequency response

$$H_1(f)=\begin{cases} j2\pi a\left(f-f_c+\dfrac{B_T}{2}\right), & f_c-\dfrac{B_T}{2}\le f\le f_c+\dfrac{B_T}{2} \\ j2\pi a\left(f+f_c-\dfrac{B_T}{2}\right), & -\left(f_c+\dfrac{B_T}{2}\right)\le f\le -\left(f_c-\dfrac{B_T}{2}\right) \\ 0, & \text{elsewhere} \end{cases} \quad \text{(IV.57)}$$

Note that

$$\tilde{H}_1\left(f-f_c\right)=2H_1(f), \quad f>0 \quad \text{(IV.58)}$$

Therefore,

$$\tilde{H}_1(f)=\begin{cases} j4\pi a\left(f+\dfrac{B_T}{2}\right), & -\dfrac{B_T}{2}\le f\le \dfrac{B_T}{2} \\ 0, & \text{elsewhere} \end{cases} \quad \text{(IV.59)}$$

The incoming FM signal is given by:

$$s(t)=A_c\cos\left[2\pi f_c t+2\pi k_f\int m(t)dt\right] \quad \text{(IV.60)}$$

Given that the carrier frequency is high compared to the transmission bandwidth of the FM signal, the complex envelope of $s(t)$ is

$$\tilde{s}(t)=A_c\exp\left[j2\pi k_f\int m(t)dt\right] \quad \text{(IV.61)}$$

Let $\tilde{s}_1(t)$ denote the complex envelope of the response of the slope circuit due to $s(t)$. Then,

$$\begin{aligned}\tilde{S}_1(f)&=\frac{1}{2}\tilde{S}(f)\tilde{H}_1(f)\\ &=\begin{cases} j2\pi a\left(f+\dfrac{B_T}{2}\right)\tilde{S}(f), & -\dfrac{B_T}{2}\le f\le \dfrac{B_T}{2} \\ 0, & \text{elsewhere} \end{cases}\end{aligned} \quad \text{(IV.62)}$$

Hence,

$$\tilde{s}_1(t)=a\left[\frac{d\tilde{s}(t)}{dt}+j\pi B_T\tilde{s}(t)\right] \quad \text{(IV.63)}$$

Substituting for $\tilde{s}(t)$ in (IV.63),

$$\tilde{s}_1(t) = j\pi B_T a A_c \left[1 + \frac{2k_f}{B_T} m(t)\right] \exp\left[j2\pi k_f \int m(t)dt\right] \tag{IV.64}$$

The desired response of the slope circuit is therefore

$$s_1(t) = \pi B_T a A_c \left[1 + \frac{2k_f}{B_T} m(t)\right] \cos\left[2\pi f_c t + 2\pi k_f \int m(t)dt + \frac{\pi}{2}\right] \tag{IV.65}$$

The signal $s_1(t)$ is a hybrid-modulated signal in which both amplitude and frequency of the carrier wave vary with the message signal $m(t)$. However, provided that we choose

$$\left|\frac{2k_f}{B_T} m(t)\right| < 1 \tag{IV.66}$$

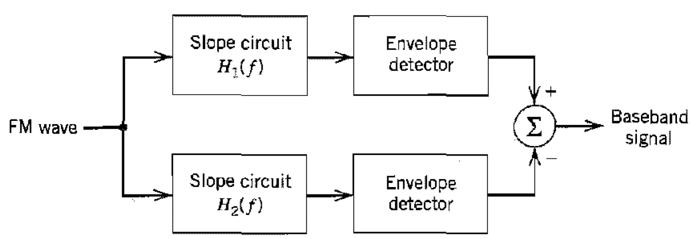
then we may use an envelope detector to recover the amplitude variations and thus, except for a bias term, obtain the original message signal. The resulting envelope-detector output is therefore

$$|\tilde{s}_1(t)| = \pi B_T a A_c \left[1 + \frac{2k_f}{B_T} m(t)\right] \tag{IV.67}$$

The bias term $\pi B_T a A_c$ in the right-hand side of Equation (IV.67) is proportional to the slope a of the transfer function of the slope circuit. This suggests that the bias may be removed by subtracting from the envelope-detector output $|\tilde{s}_1(t)|$ the output of a second envelope detector preceded by the complementary slope circuit with a frequency response $H_2(f)$ as described above.

The two slope circuits are related by

$$\tilde{H}_2(f) = \tilde{H}_1(-f) \tag{IV.68}$$



$$|\tilde{s}_2(t)| = \pi B_T a A_c \left[1 - \frac{2k_f}{B_T} m(t)\right] \tag{IV.69}$$

The difference between the two envelopes in Equations (IV.67) and (IV.69) is

$$\begin{aligned} s_o(t) &= \left|\tilde{s}_1(t)\right| - \left|\tilde{s}_2(t)\right| \\ &= 4\pi k_f a A_c m(t) \end{aligned} \quad \text{(IV.70)}$$

which is a scaled version of the original message signal $m(t)$ and free from bias.

***

V. RANDOM PROCESSES

A signal is said to be deterministic if there is no uncertainty about its time-dependent behavior at any instant of time. However, in many real-world problems the use of a deterministic signal model is inappropriate because the physical phenomenon of interest involves too many unknown factors. Nevertheless, it may be possible to consider a signal model described in probabilistic terms in that we speak of the probability of a future value lying between two specified limits. In such a case, the signal is said to be stochastic or random.

Consider, for example, a radio communication system. The received signal in such a system usually consists of an information-bearing signal component, a random interference component, and random channel noise. The information-bearing signal component may represent, for example, a voice signal that, typically, consists of randomly spaced bursts of energy of random duration. The interference component may represent signals produced by other communication systems operating in the vicinity of the radio receiver. A major source of channel noise is thermal noise, which is caused by the random motion of the electrons in conductors and devices at the front end of the receiver. We thus find that the received signal is random in nature.

Although it is not possible to predict the exact value of the signal in advance, it is possible to describe the signal in terms of statistical parameters such as the mean value, the average power and the power spectral density.

V.1. Mathematical Definition of a Random Process

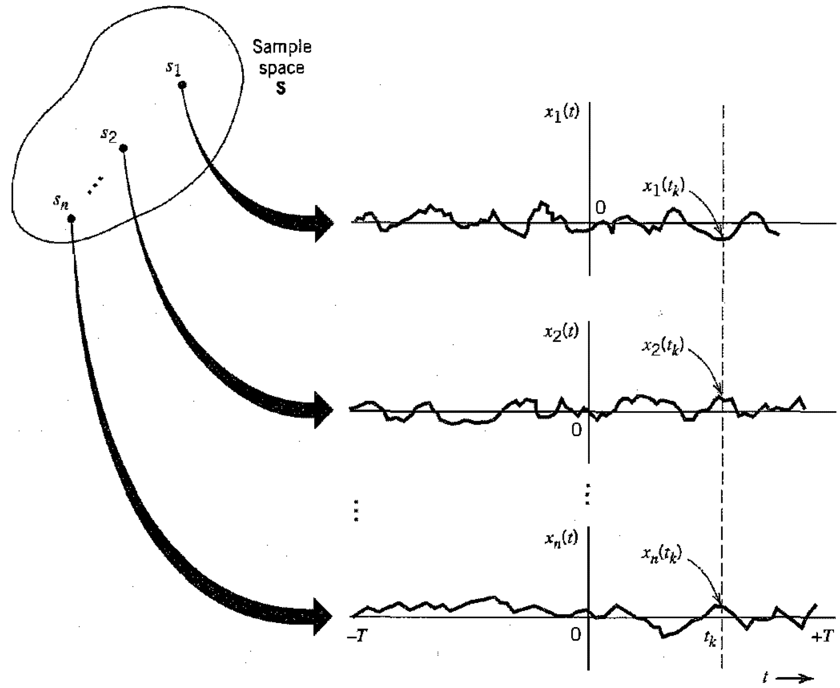
Random processes have two properties:

- First, they are functions of time.
- Second, they are random in the sense that before conducting an experiment, it is not possible to exactly define the waveforms that will be observed in the future.

Elements of the sample space are sample functions.

- For a random variable, the outcome of a random experiment is mapped into a number.
- For a random process, the outcome of a random experiment is mapped into a waveform that is a function of time.

Consider a random process $X(t)$ that is initiated at $t=-\infty$. Let $X(t_1), X(t_2), \cdots, X(t_k)$ denote the random variables obtained by observing the random process $X(t)$ at times $t_1, t_2, \cdots, t_k$, respectively. The joint distribution function of this set of random variables is $F_{X(t_1),X(t_2),\cdots,X(t_k)}(x_1, x_2, \cdots, x_k)$. Suppose next that we shift all the observation times by a fixed amount τ, thereby obtaining a new set of random variables $X(t_1+\tau), X(t_2+\tau), \cdots, X(t_k+\tau)$.



The random process $X(t)$ is said to be stationary in the strict sense or strictly stationary if the following condition holds:

$$F_{X(t_1+\tau),X(t_2+\tau),\cdots,X(t_k+\tau)}(x_1,x_2,\cdots,x_k) = F_{X(t_1),X(t_2),\cdots,X(t_k)}(x_1,x_2,\cdots,x_k) \qquad \text{(V.1)}$$

for all time shifts τ, all k, and all possible choices of observation times $t_1, t_2, \cdots, t_k$.

In other words, a random process $X(t)$, initiated at time $t = -\infty$, is strictly stationary if the joint distribution of any set of random variables obtained by observing the random process $X(t)$ is invariant with respect to the location of the origin $t = 0$.

Note that the finite-dimensional distributions in (V.1) depend on the relative time separation between random variables but not on their absolute time. That is, the random process has the same probabilistic behavior through all time.

Similarly, we may say that two random processes $X(t)$ and $Y(t)$ are jointly strictly stationary if the joint finite-dimensional distributions of the two sets of random variables $X(t_1), X(t_2), \cdots, X(t_k)$ and $Y(\xi_1), X(\xi_2), \cdots, X(\xi_l)$ are invariant with respect to the origin $t = 0$ for all k and l and all choices of observation times $t_1, t_2, \cdots, t_k$ and $\xi_1, \xi_2, \cdots, \xi_l$.

First Order Stationarity

Let $k = 1$, then from (V.1):

$$F_{X(t_1+\tau)}(x_1) = F_{X(t_1)}(x_1) \qquad \text{(V.2)}$$

That is, the first-order distribution function of a stationary random process is independent of time.

Second Order Stationarity

Let $k=2$ and $\tau=-t_1$, then from (V.1):

$$F_{X(t_1),X(t_2)}(x_1,x_2)=F_{X(0),X(t_2-t_1)}(x_1,x_2), \quad \forall t_1,t_2 \tag{V.3}$$

That is, the second-order distribution function of a stationary random process depends only on the time difference between the observation times and not on the particular times at which the random process is observed.

V.2. Wide-Sense Cyclostationary Processes

A continuous-time random signal $X(t)$ is wide-sense cyclostationary (WSC) with period T if

$$\mu_X(t+T)=\mu_X(t) \tag{V.4}$$

and

$$R_X(t+T,\tau)=R_X(t,\tau) \tag{V.5}$$

$$R_X(t,\tau)=\mathrm{E}\left[X(t)X(t+\tau)\right]=\text{autocorrelation function}$$

Note that periodicity is in the time (t) variable. Note also that, in this case, the statistical power, is given by

$$\begin{aligned} P_X(t)&=\mathrm{E}\left[|X(t)|^2\right] \\ &=R_X(t,0) \end{aligned} \tag{V.6}$$

This quantity is also periodic with period T, so that it is reasonable to define the average power as:

$$\begin{aligned} \overline{P}_X&=\frac{1}{T}\int_0^T P_X(t)dt \\ &=\frac{1}{T}\int_0^T R_X(t,0)dt \end{aligned} \tag{V.7}$$

The average autocorrelation function $\overline{R}_X(\tau)$ of $X(t)$ is evaluated by averaging $R_X(t,\tau)$ over a period (with respect to the variable t), that is:

$$\overline{R}_X(\tau)=\frac{1}{T}\int_0^T R_X(t,\tau)dt \tag{V.8}$$

Then the average power spectral density (PSD) is computed as

$$\overline{S}_X(f) = \int_{-\infty}^{\infty} \overline{R}_X(\tau)\, e^{-j2\pi f\tau} d\tau \tag{V.9}$$

V.3. Mean, Correlation, and Covariance Functions

Consider a strictly stationary random process $X(t)$. The mean of this process is given by:

$$\begin{aligned}\mu_X(t) &= \mathrm{E}\left[X(t)\right] \\ &= \int_{-\infty}^{\infty} x f_{X(t)}(x) dx\end{aligned} \tag{V.10}$$

For a strictly stationary random process, $f_{X(t)}(x)$ is independent of time. Consequently, the mean of a strictly stationary process is a constant, as shown by

$$\mu_X(t) = \mu_X, \quad \text{for all } t \tag{V.11}$$

The autocorrelation function of $X(t)$ is given by:

$$\begin{aligned}R_X(t_1,t_2) &= \mathrm{E}\left[X(t_1)X(t_2)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(t_1),X(t_2)}(x_1,x_2) dx_1 dx_2\end{aligned} \tag{V.12}$$

For a strictly stationary random process, $f_{X(t_1),X(t_2)}(x_1,x_2)$ depends only on the difference between the observation times t_1 and t_2. This, in turn, implies that the autocorrelation function of a strictly stationary process depends only on the time difference $t_2 - t_1$.

$$R_X(t_1,t_2) = R_X(t_2 - t_1), \quad \text{for all } t_1 \text{ and } t_2 \tag{V.13}$$

The auto covariance function of a strictly stationary process $X(t)$ is written as

$$\begin{aligned}C_X(t_1,t_2) &= \mathrm{E}\left[\left(X(t_1) - \mu_X\right)\left(X(t_2) - \mu_X\right)\right] \\ &= R_X(t_2 - t_1) - \mu_X^2\end{aligned} \tag{V.14}$$

This implies that the auto covariance function of a strictly stationary process depends only on the time difference $t_2 - t_1$.

This equation also shows that if we know the mean and autocorrelation function of the process, we can uniquely determine the auto covariance function. The mean and autocorrelation function are therefore sufficient to describe the first two moments of the process.

1. The mean and autocorrelation function only provide a partial description of the distribution of a random process $X(t)$.
2. The conditions of Equations (V.11) and (V.13), involving the mean and autocorrelation function, respectively, are not sufficient to guarantee that the random process $X(t)$ is strictly stationary.

The class of random processes that satisfy (V.11) and (V.13) are known as wide-sense stationary.

Example

Let $X(t) = A\cos\left(2\pi f_c t + \Theta\right)$. Let Θ be uniformly distributed over $\left[-\pi, \pi\right]$.

$$\mu_X(t) = \mathrm{E}\left[X(t)\right]$$

$$f_\Theta(\theta) = \frac{1}{2\pi}, \left|\theta\right| \le \pi$$

$$\begin{aligned}\mu_X(t) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} X(t)d\theta \\ &= \frac{A}{2\pi}\int_{-\pi}^{\pi}\cos\left(2\pi f_c t + \theta\right)d\theta \\ &= \frac{A}{2\pi}\sin\left(2\pi f_c t + \theta\right)\Big|_{-\pi}^{\pi}\end{aligned}$$

$$\mu_X(t) = 0$$

$$\begin{aligned}R_X(t, t+\tau) &= \mathrm{E}\left[X(t)X(t+\tau)\right] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} X(t)X(t+\tau)d\theta \\ &= \frac{A^2}{2\pi}\int_{-\pi}^{\pi}\cos\left(2\pi f_c t + \theta\right)\cos\left(2\pi f_c t + 2\pi f_c \tau + \theta\right)d\theta \\ &= \frac{A^2}{4\pi}\int_{-\pi}^{\pi}\cos\left(2\pi f_c \tau\right)d\theta + \frac{A^2}{4\pi}\int_{-\pi}^{\pi}\cos\left(2\pi 2 f_c t + 2\pi f_c \tau + 2\theta\right)d\theta = \frac{A^2}{2}\cos\left(2\pi f_c \tau\right)\end{aligned}$$

$X(t) = A\cos\left(2\pi f_c t + \Theta\right)$ is WSS.

Example

Let $X(t) = A\cos\left(2\pi f_c t + \Theta\right)$. Let Θ be uniformly distributed over $\left[0, \pi\right]$.

$$\mu_X(t) = \mathrm{E}\left[X(t)\right]$$

$$f_\Theta(\theta) = \frac{1}{\pi},\ 0 \le \theta \le \pi$$

$$\begin{aligned}\mu_X(t) &= \frac{1}{\pi}\int_0^\pi X(t)d\theta \\ &= \frac{A}{\pi}\int_0^\pi \cos\left(2\pi f_c t + \theta\right)d\theta \\ &= \frac{A}{\pi}\sin\left(2\pi f_c t + \theta\right)\Big|_0^\pi\end{aligned}$$

$$\mu_X(t) = -\frac{2A}{\pi}\sin\left(2\pi f_c t\right)$$

$X(t) = A\cos\left(2\pi f_c t + \Theta\right)$ is **not** WSS.

Example

Let $X(t) = A\cos\left(2\pi f_c t + \Theta\right)$. Let A be uniformly distributed over $\left[-A_m, A_m\right]$.

$$\mu_X(t) = \mathrm{E}\left[X(t)\right]$$

$$f_A(a) = \frac{1}{2A_m},\ |a| \le A_m$$

$$\mu_X(t) = \frac{1}{2A_m} \int_{-A_m} X(t)da$$

$$= \frac{1}{2A_m} \int_{-A_m}^{A_m} a\cos\left(2\pi f_c t + \theta\right) da$$

$$= \frac{1}{2A_m} \cos\left(2\pi f_c t + \theta\right) \int_{-A_m}^{A_m} ada$$

$$= \frac{a^2}{4A_m} \cos\left(2\pi f_c t + \theta\right) \Big|_{-A_m}^{A_m}$$

$$\mu_X(t) = 0$$

$$R_X(t, t+\tau) = \mathrm{E}\left[X(t)X(t+\tau)\right]$$

$$= \frac{1}{2A_m} \int_{-A_m}^{A_m} X(t)X(t+\tau)da$$

$$= \frac{1}{2A_m} \int_{-A_m}^{A_m} a^2 \cos\left(2\pi f_c t + \theta\right)\cos\left(2\pi f_c t + 2\pi f_c \tau + \theta\right) da$$

$$= \frac{1}{2A_m} \cos\left(2\pi f_c t + \theta\right)\cos\left(2\pi f_c t + 2\pi f_c \tau + \theta\right) \int_{-A_m}^{A_m} a^2 da$$

$$= \frac{A_m^2}{3} \cos\left(2\pi f_c t + \theta\right)\cos\left(2\pi f_c t + 2\pi f_c \tau + \theta\right)$$

$X(t) = A\cos\left(2\pi f_c t + \Theta\right)$ is **not** WSS.

V.3.A. Properties of the Autocorrelation Function

For a stationary process,

$$R_X(\tau) = \mathrm{E}\left[X(t)X(t+\tau)\right] \tag{V.15}$$

Mean-Square Value (Total Power)

$$R_X(0) = \mathrm{E}\left[X^2(t)\right] \tag{V.16}$$

Symmetry

$$R_X(\tau) = R_X(-\tau) \tag{V.17}$$

Maximum Correlation

$$\left|R_X(\tau)\right| \le \left|R_X(0)\right| \tag{V.18}$$

Exercise:
Is $2|\tau|+1$ a possible autocorrelation function? **NO**.

V.3.B. Cross-Correlation Functions

$$R_{XY}(t,u) = \mathrm{E}\left[X(t)Y(u)\right] \tag{V.19}$$

$$R_{YX}(t,u) = \mathrm{E}\left[Y(t)X(u)\right] \tag{V.20}$$

Correlation Matrix

$$R(t,u) = \begin{bmatrix} R_X(t,u) & R_{XY}(t,u) \\ R_{YX}(t,u) & R_Y(t,u) \end{bmatrix} \tag{V.21}$$

If the random processes $X(t)$ and $Y(t)$ are each stationary and, in addition, they are jointly stationary, then the correlation matrix can be written as

$$R(\tau) = \begin{bmatrix} R_X(\tau) & R_{XY}(\tau) \\ R_{YX}(\tau) & R_Y(\tau) \end{bmatrix} \tag{V.22}$$

The cross-correlation function is not generally an even function of τ as was true for the autocorrelation function, nor does it have a maximum at the origin. However, it does obey a certain symmetry relationship as follows

$$R_{XY}(\tau) = R_{YX}(-\tau) \tag{V.23}$$

V.4. Ergodic Processes

The expectations or ensemble averages of a random process $X(t)$ are averages "across the process". For example, the mean of a random process $X(t)$ at some fixed time t_k is the expectation of the random variable $X(t_k)$ that describes all possible values of the sample functions of the process observed at time $t = t_k$.

Naturally, we may also define long-term sample averages, or time averages that are averages "along the process".

We are therefore interested in relating ensemble averages to time averages, for time averages represent a practical means available to us for the estimation of ensemble averages of a random process. The key question, of course, is: When can we substitute time averages for ensemble averages?

Time Average

Consider the sample function $x(t)$ of a stationary process $X(t)$, with the observation interval defined as $-T \le t \le T$. The DC value of $x(t)$ is defined by the time average

$$\mu_x(T) = \frac{1}{2T}\int_{-T}^{T} x(t)dt \tag{V.24}$$

Clearly, the time average $\mu_x(T)$ is a random variable, as its value depends on the observation interval and which particular sample function of the random process $X(t)$ is picked. Since the process $X(t)$ is assumed to be stationary, the mean of the time average $\mu_x(T)$ is given by (after interchanging the operation of expectation and integration):

$$\begin{aligned} \mathrm{E}\left[\mu_x(T)\right] &= \frac{1}{2T}\int_{-T}^{T} \mathrm{E}\left[x(t)\right]dt \\ &= \frac{1}{2T}\int_{-T}^{T} \mu_X dt \\ &= \mu_X \end{aligned} \tag{V.25}$$

The time average $\mu_x(T)$ represents an unbiased estimate of the ensemble-averaged mean μ_X.

The process $X(t)$ is ergodic in the mean if two conditions are satisfied:

- The time average $\mu_x(T)$ approaches μ_X as T approaches infinity;

$$\lim_{T\to\infty} \mu_x(t) = \mu_X \tag{V.26}$$

- The variance of $\mu_x(T)$ approaches zero as T approaches infinity;

$$\lim_{T\to\infty} \mathrm{var}\left[\mu_x(T)\right] = 0 \tag{V.27}$$

Time-Averaged Autocorrelation

$$R_x(\tau,T) = \frac{1}{2T}\int_{-T}^{T} x(t+\tau)x(t)dt \tag{V.28}$$

This is a random variable.

The process $X(t)$ is ergodic in the autocorrelation function if

$$\lim_{T\to\infty} R_x(\tau,T) = R_X(\tau) \quad \text{(V.29)}$$

$$\lim_{T\to\infty} \operatorname{var}\left[R_x(\tau,T)\right] = 0 \quad \text{(V.30)}$$

V.5. Power Spectral Density (PSD)

$$R_X(\tau) = \frac{A^2}{2}\cos\left(2\pi f_c \tau\right)$$

$$S_X(f) = \int_{-\infty}^{\infty} \frac{A^2}{2}\cos\left(2\pi f_c \tau\right) e^{-j2\pi f\tau} d\tau$$

$$S_X(f) = \int_{-\infty}^{\infty} \frac{A^2}{2}\cos\left(2\pi f_c t\right) e^{-j2\pi ft} dt$$

$$R_X(\xi) = \frac{A^2}{2}\cos\left(2\pi f_c \xi\right)$$

Einstein-Wiener-Khintchine Relations

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau$$ (**Fourier Transform of the autocorrelation**) (V.31)

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{j2\pi f\tau} df$$ (**Inverse Fourier Transform of the PSD**) (V.32)

V.5.A. Properties of the PSD

$$S_X(0) = \int_{-\infty}^{\infty} R_X(\tau) d\tau \quad \text{(V.33)}$$

$$\begin{aligned} \mathrm{E}\left[X^2(t)\right] &= R_X(0) \\ &= \int_{-\infty}^{\infty} S_X(f) df \end{aligned} \quad \text{(V.34)}$$

$$S_X(f) \geq 0 \quad \text{(V.35)}$$

$$S_X(-f) = S_X(f) \quad \text{(V.36)}$$

Example V.1

Sinusoidal Wave with Random Phase

Consider the random process $X(t) = A\cos(2\pi f_c t + \Theta)$, where Θ is a uniformly distributed random variable over the interval $[-\pi, \pi]$. Then,

$$R_X(\tau) = \frac{A^2}{2}\cos(2\pi f_c \tau)$$

$$S_X(f) = \frac{A^2}{4}\left[\delta(f - f_c) + \delta(f + f_c)\right]$$

Example V.2

Mixing of a Random Process with a Sinusoidal Process

Let $X(t)$ be a stationary random process. Consider the random process $Y(t) = X(t)\cos(2\pi f_c t + \Theta)$, where Θ is a uniformly distributed random variable over the interval $[0, 2\pi]$. Then,

$$R_Y(\tau) = \frac{1}{2}R_X(\tau)\cos(2\pi f_c \tau)$$

$$S_Y(f) = \frac{1}{4}\left[S_X(f - f_c) + S_X(f + f_c)\right]$$

V.5.B. Cross Power Spectral Densities

$$S_{XY}(f) = \int_{-\infty}^{\infty} R_{XY}(\tau) e^{-j2\pi f\tau} d\tau \quad \text{(V.37)}$$

$$S_{YX}(f) = \int_{-\infty}^{\infty} R_{YX}(\tau) e^{-j2\pi f\tau} d\tau \quad \text{(V.38)}$$

$$R_{XY}(\tau) = R_{YX}(-\tau) \quad \Rightarrow \quad S_{XY}(f) = S_{YX}(-f) = S_{YX}^*(f) \quad \text{(V.39)}$$

Let $Z(t) = X(t) + Y(t)$. If $X(t)$ and $Y(t)$ are uncorrelated, then

$$S_Z(f) = S_X(f) + S_Y(f) \quad \text{(V.40)}$$

V.6. Gaussian Random Processes

Let us suppose that we observe a random process $X(t)$ for an interval that starts at time $t=0$ and lasts until $t=T$. Suppose also that we weight the random process $X(t)$ by some function $g(t)$ and then integrate the product $g(t)X(t)$ over this observation interval, thereby obtaining a random variable Y defined by

$$Y = \int_0^T g(t)X(t)dt \tag{V.41}$$

$X(t)$ is a Gaussian random process if Y is Gaussian with a finite mean square value. The PDF of Y can be written in the form

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \tag{V.42}$$

V.6.A. Central Limit Theorem

The central limit theorem provides the mathematical justification for using a Gaussian process as a model for a large number of different physical phenomena in which the observed random variable, at a particular instant of time, is the result of a large number of individual random events.

Let $X_1(t), \cdots, X_N(t)$ be a set of random variables that satisfies the following requirements:

1. $\{X_n(t)\}_{n=1}^N$ are statistically independent.
2. $\{X_n(t)\}_{n=1}^N$ are identically distributed, i.e., they all have the same PDF with the same mean μ_X and the same variance σ_X^2.

Let

$$Y_n = \frac{X_n - \mu_X}{\sigma_X} \tag{V.43}$$

Then,

$$\begin{aligned} \mu_{Y_n} &= \mathrm{E}\left[Y_n(t)\right] \\ &= 0 \\ &= \mu_Y \end{aligned} \tag{V.44}$$

$$\begin{aligned}\sigma_{Y_n}^2 &= 1 \\ &= \sigma_Y^2\end{aligned} \quad \text{(V.45)}$$

Let

$$V_N = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} Y_n \quad \text{(V.46)}$$

The central limit theorem states that the probability distribution of V_N approaches a normalized Gaussian distribution $\mathcal{N}(0,1)$ in the limit as the number of random variables N approaches infinity.

V.6.B. Properties of the Gaussian Process

- If a Gaussian process $X(t)$ is applied to a stable linear filter, then the random process $Y(t)$ developed at the output of the filter is also Gaussian.
- Consider the set of random variables or samples $X(t_1),\cdots,X(t_n)$ obtained by observing a random process $X(t)$ at times $t_1,\cdots,t_n$. If the process $X(t)$ is Gaussian, then this set of random variables is jointly Gaussian for any n, with their n-fold joint probability density function being completely determined by specifying the set of means $\mu_{X_i}(t)$ and the set of covariance functions $C_X(t_k,t_l)$ for $i,k,l = 1,\cdots,n$. The joint PDF is given by:

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}\Delta^{1/2}} \exp\left(-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T \Sigma^{-1}\left(\underline{x}-\underline{\mu}\right)\right) \quad \text{(V.47)}$$

where

$$\underline{X} = \left[X(t_1) \quad \cdots \quad X(t_n)\right]^T \quad \text{(V.48)}$$

$$\underline{x} = \left[x_1 \quad \cdots \quad x_n\right]^T \quad \text{(V.49)}$$

$$\underline{\mu} = \left[\mu_1 \quad \cdots \quad \mu_n\right]^T \quad \text{(V.50)}$$

$$\Sigma = \begin{bmatrix} C_X(t_1,t_1) & \cdots & C_X(t_1,t_n) \\ \vdots & \ddots & \vdots \\ C_X(t_n,t_1) & \cdots & C_X(t_n,t_n) \end{bmatrix} \quad \text{(V.51)}$$

$$\Delta = \det\left(\Sigma\right) \quad \text{(V.52)}$$

- If a Gaussian process is stationary, then the process is also strictly stationary.

- If the random variables $X(t_1),\cdots,X(t_n)$, obtained by sampling a Gaussian process $X(t)$ at times $t_1,\cdots,t_n$, are uncorrelated, that is,

$$\mathrm{E}\left[\left(X(t_k)-\mu_{X(t_k)}\right)\left(\left(X(t_i)-\mu_{X(t_i)}\right)\right)\right]=0 \qquad \text{(V.53)}$$

for $i \neq k$, then these random variables are statistically independent.

V.7. Transmission of a Random Process Through a Linear Time-Invariant Filter

Consider an LTI system with impulse response $h(t)$. Let the system input be a stationary random process $X(t)$. The output process is given by

$$\begin{aligned} Y(t) &= X(t) * h(t) \\ &= \int_{-\infty}^{\infty} h(\xi) X(t-\xi) d\xi \end{aligned} \qquad \text{(V.54)}$$

$$\begin{aligned} \mu_Y(t) &= \mathrm{E}\left[Y(t)\right] \\ &= \mathrm{E}\left[\int_{-\infty}^{\infty} h(\xi) X(t-\xi) d\xi\right] \\ &= \int_{-\infty}^{\infty} h(\xi) \mathrm{E}\left[X(t-\xi)\right] d\xi \\ &= \int_{-\infty}^{\infty} h(\xi) \mu_X(t-\xi) d\xi \\ &= \mu_X(t) * h(t) \end{aligned} \qquad \text{(V.55)}$$

When the input random process $X(t)$ is stationary, the mean $\mu_X(t)$ is a constant μ_X, so that

$$\begin{aligned} \mu_Y(t) &= \mu_X \int_{-\infty}^{\infty} h(\xi) d\xi \\ &= \mu_X H(0) \end{aligned} \qquad \text{(V.56)}$$

$$\begin{aligned} R_Y(t,u) &= \mathrm{E}\left[Y(t)Y(u)\right] \\ &= \mathrm{E}\left[\int_{-\infty}^{\infty} h(\xi) X(t-\xi) d\xi \int_{-\infty}^{\infty} h(\gamma) X(u-\gamma) d\gamma\right] \end{aligned} \qquad \text{(V.57)}$$

$$\begin{aligned} R_Y(t,u) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\xi)h(\gamma)\,\mathrm{E}\left[X(t-\xi)X(u-\gamma)\right]d\xi d\gamma \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\xi)h(\gamma)R_X(t-\xi,u-\gamma)d\xi d\gamma \end{aligned} \tag{V.58}$$

If $X(t)$ is stationary,

$$\begin{aligned} R_Y(\tau) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\xi)h(\gamma)R_X(\tau-\xi+\gamma)d\xi d\gamma \\ &= \int_{-\infty}^{\infty} h(\gamma)\int_{-\infty}^{\infty} h(\xi)R_X(\tau+\gamma-\xi)d\xi d\gamma \\ &= \int_{-\infty}^{\infty} h(\gamma)q(\tau+\gamma)d\gamma \\ &= q(\tau)*h(-\tau) \\ &= R_X(\tau)*h(\tau)*h(-\tau) \end{aligned} \tag{V.59}$$

$$S_Y(f) = \left|H(f)\right|^2 S_X(f) \tag{V.60}$$