10.1 Signal Transmission in Communication Systems

The main role of a communication system is to transmit signals (information) from the source of information (system input) to the user, destination (system output). The transmission is done over a *communication channel* using a *transmitter* and a *receiver*. A simplified basic communication system is presented in Figure 10.1.

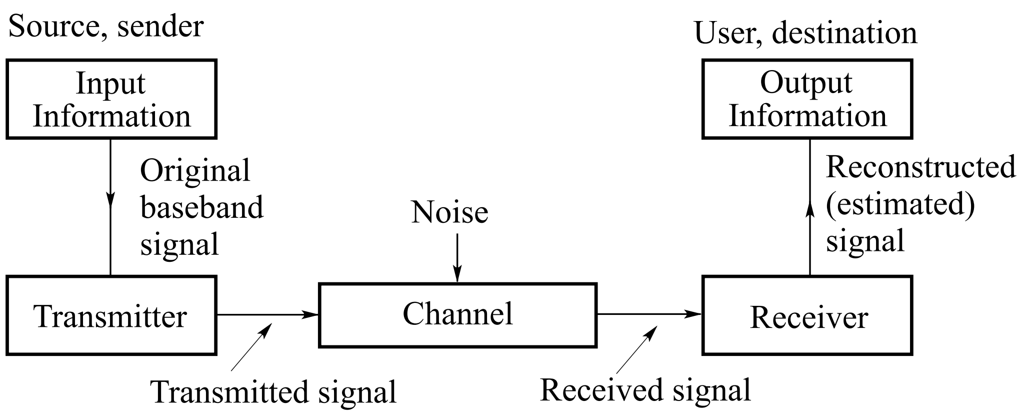

Figure 10.1: Basic communication system

The original signal, usually called, the *baseband* signal (this name will be justified after we explain the modulation concept) is first transformed into the signal convenient for transmission (called the *transmitted signal*) using the transmitter. The transmitter sends such a signal as an electrical or optical (electromagnetic) signal over a communication channel, which represents a physical medium convenient for propagation of electromagnetic waves (low signal attenuation and distortion). Communication channels can be guided media (such as copper wire or optical fiber cable channels) or free-space channels (such as satellite or wireless (radio) channels). The role of the receiver is to convert received signals, theoretically, into baseband signals and pass them to the user. Due to channel attenuation, distortion, and noise, the receiver produces a signal that is only similar but not identical to the baseband signal. Such a signal is called estimated or reconstructed signal. The *estimated signal* can be slightly different than the originally sent signal (baseband signal) especially for voice and video transmissions, since the human eye and ear are unable to detect small errors. However, in the case when we transmit data, the signal transmission must be error free.

Modulation

In a standard communication system, the transmitter is a modulator, and the receiver is a demodulator. The modulator and demodulator together are called modem. We have already introduced the modulation concept within the properties of the Fourier transform. The modulation property of the Fourier transform says the following: Let the signal $x(t)$ have the Fourier transform equal to $X(j\omega)$, where $\omega = 2\pi f$. Then, the Fourier transform of the *modulated signal*, defined as $x(t)\cos(\omega_c t)$, $\omega_c = 2\pi f_c$, is given by

$$\mathcal{F}\{x(t)\cos(\omega_c t)\} = \frac{1}{2}X(j(\omega + \omega_c)) + \frac{1}{2}X(j(\omega - \omega_c))$$

Since $|X(j\omega)|$ denotes the signal spectrum, it can be seen that the *spectrum of the modulated signal is shifted left and right by* ω_c, as represented in Figure 10.2. The frequency ω_c is called the *carrier frequency*. The original signal spectrum is the baseband signal spectrum, and the other two spectra in Figure 10.2 are the modulated signal spectra. This justifies the name the *baseband signal*. Due to the magnitude spectrum symmetry, the positive frequencies carry all information contained in the given signal. We can make two observations from Figure 10.2.

(1) The spectrum of the modulated signal is doubled comparing to the spectrum of the baseband signal. It contains the *upper frequency sideband* and the *lower frequency sideband*, each having the bandwidth equal to the bandwidth of the baseband signal.

(2) Due to frequency translation, the negative frequencies come into the picture, and they form the lower frequency sideband.

Hence, the amplitude modulation procedure presented requires doubling in the spectrum requirements (waste of the frequency band). In Section 10.5, we will study a technique that remedies this problem.

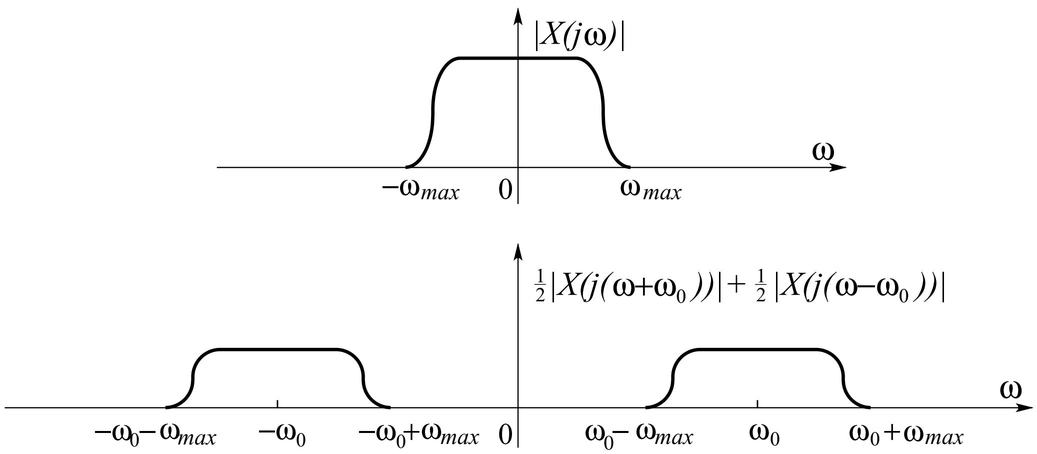

Figure 10.2: The spectrum of the original and modulated signals

The modulation concept indicates one extraordinary possibility that the same channel can be used to simultaneously transmit several signals by appropriately shifting their spectra such that they do not overlap in the frequency domain. Note that the signals may overlap in the time domain. In Figure 3.8, we have considered a telephone network that transmits many telephone signals (calls) simultaneously.

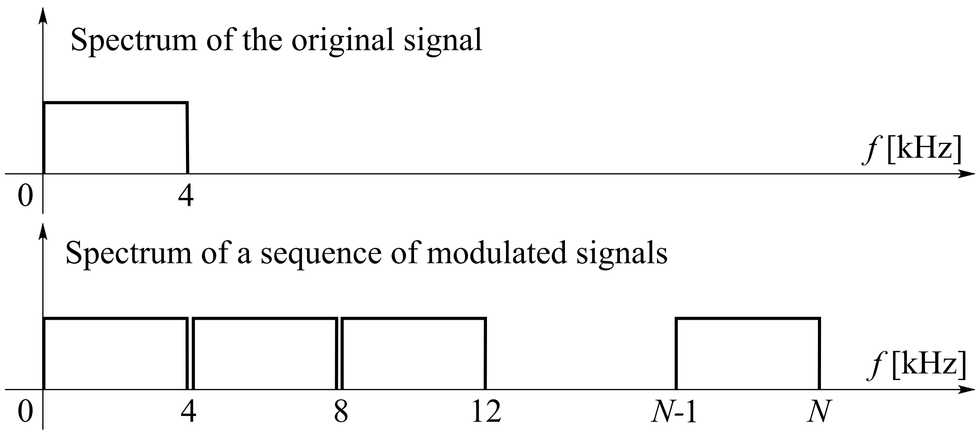

Figure 3.8: Transmission of N telephone modulated signals over the same channel

It can be observed from Figure 3.8 that the users share the frequency band. If we assume that the channel frequency bandwidth is equal to ω_{BW} and that the channel must serve N users (baseband signals), then we see that each user has reserved all the times a part of the channel frequency band equal to ω_{BW}/N. Such a channel sharing is called *frequency division multiplexing* (FDM).

Another channel sharing technique used in communication system practice is the *time division multiplexing* (TDM), a technique in which each user gets the whole frequency band of the channel, but only during a limited period of time. In such a case the users are switched on and off according to the given time schedule. For example, each user uses the whole channel frequency band during the time period of Δt, and they rotate so that each gets a turn after $N\Delta t$ time units (fair sharing of the channel). Note that there is no single criteria by which to judge that one of the channel sharing techniques is better, due to the very simple fact that a *channel with a larger frequency bandwidth has a higher capacity* (it can transmit more units of information per unit of time, it is a faster channel). Hence, there is an interplay between transmitting at high speeds during short periods of time (TDM) and transmitting at low speeds all times (FDM).

Demodulation

The demodulation process is reciprocal to the modulation process. Demodulation is an operation that reconstructs the original baseband signal from its modulated signal. Technically speaking, the demodulator has to cut out (filter out) the frequency band that corresponds to the given baseband signal.

Demodulation can be performed by modulating again the modulated signal

$$\mathcal{F}\{[x(t)\cos(\omega_c t)]\cos(\omega_c t)\} = \mathcal{F}\left\{\frac{1}{2}x(t) + \frac{1}{2}x(t)\cos(2\omega_c t)\right\}$$
$$= \frac{1}{2}X(j\omega) + \frac{1}{4}X(j(\omega + 2\omega_c)) + \frac{1}{4}X(j(\omega - 2\omega_c))$$

By passing this signal through a low-pass filter we can recover the original signal multiplied by 0.5, that is $0.5x(t)$. In Section 10.5 we will say more about both the modulation and demodulation procedures. In the remaining part of this section we will introduced some notions frequently used in signal transmission.

Signal to Noise Ratio

As mentioned earlier, channel noise is most often random in nature. Despite the fact that we will not study channels from the stochastic point of view, we can define a simple quantity that tells us how much, in average, a given channel is noisy. Let P_s denote the average signal power and let P_n denote the average noise power. The *signal to noise ratio in decibels* [dB] is defined by

$$SNR\,[\mathrm{dB}] = 10\log_{10}\left(\frac{P_s}{P_n}\right)$$

Apparently, the higher SNR the better channel.

Channel Capacity

It can be experimentally observed that the channel capacity is directly proportional to its frequency bandwidth. It has also been observed that the higher SNR implies the higher channel capacity.

An exact formula that relates the channel capacity in bits per second, channel frequency bandwidth in Hz, and the channel signal power to noise power ratio was derived by Shannon (also known as Shannon-Harteley's formula). It is given by

$$C\left[\frac{\mathrm{b}}{\mathrm{s}}\right] = w_{BW}\log_2\left(1+\frac{P_s}{P_n}\right)$$

The formula is valid for channels with Guassian noise (noise statistics is completely described by the first and second order moments). In the case of non Gaussian noise, the above formula gives only an approximate lower bound.

Optical Fiber Cable

As the *waveguide medium of the future*, the optical fiber cable has a huge frequency bandwidth that theoretically can reach several hundreds of THz (1 terahertz is equal to 10^{12} Hz). It has also very low signal attenuation of only $0.2\,\mathrm{dB/km}$, which means that

$$10 \log_{10} \left(\frac{P^{inp}}{P^{out}} \right) = 0.2 \left[\frac{\text{dB}}{\text{km}} \right]$$

where P^{inp} and P^{out} represent, respectively, the input and output signal powers. In addition, the optical fiber cable has very low signal distortion. Note that such a cable is made of silica glass (dielectric) and that it transports light signals, also called optical signals. Similarly to the frequency division multiplexing, in optical communication systems, *wavelength division multiplexing* (WDM) is used to transmit simultaneously many signals (eighty or even more) over the same optical fiber channel. The optical wavelength is defined by $\lambda = v/f$, where v is the light speed (equal to $c = 3 \times 10^8\,\text{m}$ in vacuum and $c/\sqrt{\epsilon_r \mu_r} \approx 0.6c$ in a guided media, ϵ_r and μ_r are respectively the medium permittivity and permeability constants) and f the frequency of the corresponding light signal. Note that

DWDM stands for *dense wavelength frequency division multiplexing* that has optical wavelength channels densely spaced every $\mathbf{10^{-9}\,m = 1\,nm}$ (1 *nanometar*).

It is interesting to point out that a channel represents a dynamic system that can be either linear or nonlinear, time invariant or time varying, deterministic or stochastic (see system classification in Section 1.4). For example, telephone channels are linear systems in most cases, wireless channels can be considered as time varying linear systems, fiber optics channels are nonlinear time invariant systems that are often linearized (see section on linearization of nonlinear systems, Section 8.6), satellite channels are nonlinear. Another classification of channels distinguishes between *bandlimited channels* such as telephone networks and *power limited channels* such as optical fiber and satellite channels.

10.2 Signal Correlation, Energy and Power Spectra

In addition to the system frequency bandwidth, the signal power represents another important quantity that engineers are particularly concerned with while transmitting signals. We have already defined signal energy and power in the time domain in Section 2.3. Here, we present their representations in the frequency domain and relate them to the quantity known as the signal correlation function.

Devices called signal correlators are used to measure power of incoming signals in many communication (and signal processing) systems. For example, in wireless communication systems, correlators at the base station measure at all times the signal power of all mobiles in the base station area (cell). Those powers are periodically adjusted such that each mobile has sufficient signal power for a good quality transmission, but not so much signal power as to cause unnecessary interference to the other mobiles that use the same frequency band.

Continuous-Time Signal Correlation

The analytical expression for signal correlation is very similar to the convolution integral, even though signal correlation and signal convolution have completely different physical meanings.

Correlation of two continuous-time signals $x_1(t)$ and $x_2(t)$ is defined by

$$R_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t)x_2(t+\tau)dt$$

where τ is a parameter, $-\infty < \tau < \infty$. More precisely, $R_{12}(\tau)$ is called the *cross-correlation function*. Assuming that the signals $x_1(t)$ and $x_2(t)$ have Fourier transforms respectively given by $X_1(j\omega)$ and $X_2(j\omega)$, that is

$$x_1(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1(j\omega)e^{j\omega t}d\omega, \quad x_2(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X_2(j\omega)e^{j\omega t}d\omega$$

then, we have

$$\begin{aligned} R_{12}(\tau) &= \int_{-\infty}^{\infty} x_1(t)\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} X_2(j\omega)e^{j\omega(t+\tau)}d\omega\right]dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x_1(t)e^{j\omega t}dt\right]X_2(j\omega)e^{j\omega\tau}d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X_1^*(j\omega)X_2(j\omega)e^{j\omega\tau}d\omega \end{aligned}$$

Note that $X_1^*(j\omega) = X_1(-j\omega)$. The last formula indicates that $R_{12}(\tau)$ and $X_1^*(j\omega)X_2(j\omega)$ form the Fourier transform pair, that is

$$R_{12}(\tau) \leftrightarrow X_1^*(j\omega)X_2(j\omega)$$

In the case when $x_1(t) = x_2(t) = x(t)$, we have the definition of the *autocorrelation function* as

$$R(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$$

In this case, we have

$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X^*(j\omega)X(j\omega)e^{j\omega\tau}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(j\omega)|^2 e^{j\omega\tau}d\omega$$

that is, the autocorrelation function and $|X(j\omega)|^2$ form the Fourier transform pair

$$R(\tau) \leftrightarrow |X(j\omega)|^2$$

It can be shown that the autocorrelation function has the following properties:

1) The autocorelation function is even, that is $R(\tau) = R(-\tau)$.

2) $R(0) = E_\infty$, where E_∞ stands for the total signal energy.

3) $R(0) \geq R(\tau), \ \forall \tau$.

4) $R(\tau)$ is continuous in time (like convolution).

The quantity $|X(j\omega)|^2$ defines the signal power at the given frequency ω so that $|X(j\omega)|^2$ is called the *power spectrum*. $|X(j\omega)|^2$ is also called the *energy density spectrum* for the reason to be clear soon. Introducing notation

$$|X(j\omega)|^2 = S(\omega)$$

we have

$$S(\omega) = \int_{-\infty}^{\infty} R(t)e^{-j\omega t}dt, \quad R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}d\omega$$

Note that $S(\omega)$ is a real positive and even function, that is $S(\omega) = S(-\omega) \geq 0$.

It follows from $R(0) = E_\infty$ that

$$E_\infty = \frac{1}{2\pi}\int_{\omega=-\infty}^{\omega=\infty} S(\omega)d\omega$$

It is clear that $S(\omega)$ represents the *energy density in the frequency domain*, which justifies the density energy spectrum name used for $|X(j\omega)|^2 = S(\omega)$.

If one intends to find the signal energy in the frequency domain in any frequency range, say (ω_1, ω_2), then the knowledge of the signal density energy $S(\omega)$ gives the following formula

$$W(\omega_1, \omega_2) = \frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S(\omega)d\omega + \frac{1}{2\pi}\int_{-\omega_2}^{-\omega_1} S(\omega)d\omega = \frac{1}{\pi}\int_{\omega_1}^{\omega_2} S(\omega)d\omega$$

The last expression follows from the fact that $S(\omega)$ is a positive and symmetric function of frequency. This formula determines the distribution of the signal energy in the frequency domain. Here, we see how the "negative" frequencies come into the picture and how the signal energy can be completely expressed in terms of positive frequencies, which reflects physical reality.

Example 10.1: In this example we will find the frequency range that contains the given percentage (50%) of the signal energy. Consider the signal frequency spectrum presented in Figure 10.3.

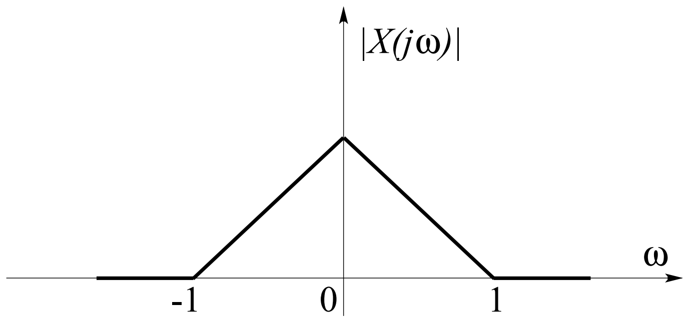

Figure 10.3: The frequency spectrum of a signal

We are looking for the frequency ω_1 such that

$$W(0,\omega_1)=\frac{1}{2}W=\frac{1}{2}\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)d\omega=\frac{1}{4\pi}\int_{-1}^{1}|X(j\omega)|^2d\omega$$

$$=\frac{1}{2\pi}\int_{0}^{1}(1-\omega)^2d\omega=\frac{1}{6\pi}$$

Hence, we have the equality

$$W(0,\omega_1) = \frac{1}{\pi}\int_0^{\omega_1} \left(1-\omega^2\right) d\omega = \frac{1}{2}W = \frac{1}{6\pi}$$

Using the following change of variables $1-\omega = \mu$, the above integral can be easily calculated, which leads to $\omega_1 = 1 - 1/\sqrt[3]{2}$. Note that in this problem we have tacitly taken into account the contribution of "negative" frequencies to the signal energy.

As a measure of similarity of two signals, the so-called correlation coefficient can be defined as

$$-1 \leq c_{12}(\tau) = \frac{R_{12}(\tau)}{\sqrt{R_{11}(0)R_{22}(0)}} \leq 1$$

When the correlation coefficient is close to one then the signals are similar.

Correlation of Periodic Signals

In the case when signals are periodic, the correlation functions can be obtained using the Fourier series. Let $x_1(t) = x_1(t+T)$ and $x_2(t) = x_2(t+T)$, $T < \infty$, then the *cross-correlation function for periodic signals* is defined by

$$R_{12}(\tau) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_1(t)x_2(t+\tau)dt$$

Using the fact that periodic functions can be expressed using the Fourier series

$$x_1(t) = \sum_{n=-\infty}^{n=\infty} X_1(jn\omega_0)e^{jn\omega_0 t}$$

$$x_2(t) = \sum_{n=-\infty}^{n=\infty} X_2(jn\omega_0)e^{jn\omega_0 t}, \quad \omega_0 = \frac{2\pi}{T}$$

we have

$$R_{12}(\tau) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_1(t)\left[\sum_{n=-\infty}^{n=\infty} X_2(jn\omega_0)e^{jn\omega_0(t+\tau)}\right]dt$$

$$= \sum_{n=-\infty}^{n=\infty} X_2(jn\omega_0)\left[\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_1(t)e^{jn\omega_0 t}dt\right]e^{jn\omega_0\tau}$$

$$= \sum_{n=-\infty}^{n=\infty} X_1^*(jn\omega_0)X_2(jn\omega_0)e^{jn\omega_0\tau}$$

Note that $R_{12}(\tau)$ and $X_1^*(jn\omega_0)X_2(jn\omega_0)$ are the corresponding Fourier series pair, which implies that

$$X_1^*(jn\omega_0)X_2(jn\omega_0) = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} R_{12}(\tau)e^{-jn\omega_0\tau}d\tau$$

In the case when $x_1(t) = x_2(t) = x(t)$, we can define the *autocorrelation function for periodic signals* as

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

which leads to

$$R(\tau) = \sum_{n=-\infty}^{n=\infty} |X(jn\omega_0)|^2 e^{jn\omega_0\tau}$$

Introducing the notion of the power spectrum $S(n\omega_0) = |X(jn\omega_0)|^2$ of a periodic signal, we have the corresponding Fourier series pair

$$R(\tau) = \sum_{n=-\infty}^{n=\infty} S(n\omega_0)e^{jn\omega_0\tau}, \quad S(n\omega_0) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} R(\tau)e^{-jn\omega_0\tau}d\tau$$

Note that for periodic signals the autocorrelation function is also an even function. The corresponding spectrum is an even and positive function. In addition, $R(0)$ defines the signal energy during one time period, that is

$$R(0) = \sum_{n=-\infty}^{n=\infty} S(n\omega_0) = \sum_{n=-\infty}^{n=\infty} |X(jn\omega_0)|^2 = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t)dt = W_T$$

This relation also represents *Parseval's theorem for periodic signals*.

10.3 Hilbert Transform

The Hilbert transform plays an important role in communication systems. It can be easily derived using knowledge from Chapter 3 about the Fourier transform. There are two forms of the Hilbert transform. The first form is valid for causal signals and the second form holds for real signals. The first form of the Hilbert transform has applications in linear electrical circuits and electric power systems, and the second form of the Hilbert transform is used in communications systems.

Hilbert Transform for Causal Signals

In the following we will show that in the case of causal signals, the Hilbert transform, in fact, relates the real and imaginary parts of the corresponding Fourier transform. Such a relationship holds for any *causal* real or complex signal (function) $x(t)$. Recall that causal signals are equal to zero for $t < 0$.

Due to causality, we have

$$\mathcal{F}(x(t)) = X(j\omega) = X_{Re}(\omega) + jX_{Im}(\omega) = \int_0^\infty x(t)e^{-j\omega t}dt$$

Causality implies also $x(t) = x(t)u(t)$, where $u(t)$ is the unit step function. The application of the Fourier transform produces

$$\mathcal{F}(x(t)) = X(j\omega) = \mathcal{F}(x(t)u(t)) = \frac{1}{2\pi}X(j\omega) * U(j\omega)$$

Using the expression for the Fourier transform of the unit step function, we obtain

$$X(j\omega) = X_{Re}(\omega) + jX_{Im}(\omega)$$

$$= \frac{1}{2\pi}[X_{Re}(\omega) + jX_{Im}(\omega)] * \left[\pi\delta(\omega) + \frac{1}{j\omega}\right]$$

It is known that the convolution of any signal with the delta impulse signal produces that signal. Using this fact, the above equation is simplified into

$$X_{Re}(\omega) + jX_{Im}(\omega) = \frac{1}{2}X_{Re}(\omega) + j\frac{1}{2}X_{Im}(\omega)$$
$$-\frac{1}{2\pi}X_{Re}(\omega) * \frac{j}{\omega} + \frac{1}{2\pi}X_{Im}(\omega) * \frac{1}{\omega}$$

Equating the real and imaginary parts in the last equation, we have

$$X_{Re}(\omega) = X_{Im}(\omega) * \frac{1}{\pi\omega}$$

and

$$X_{Im}(\omega) = -X_{Re}(\omega) * \frac{1}{\pi\omega}$$

These formulas relate the real and imaginary parts of the Fourier transform of the causal signal $x(t)$ and define the Hilbert transform.

Using the definition of the frequency domain convolution, the last two formulas can be written in the following form

$$X_{Re}(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} X_{Im}(\nu)\frac{1}{\omega - \nu}d\nu$$

and

$$X_{Im}(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} X_{Re}(\nu)\frac{1}{\omega - \nu}d\nu$$

Example 10.2: The unit step signal $u_h(t)$ is a causal signal whose Fourier transform has both the real and imaginary parts, that is

$$\mathcal{F}\{u_h(t)\} = \pi\delta(\omega) + \frac{1}{j\omega} = \pi\delta(\omega) - j\frac{1}{\omega}$$

The imaginary part of the given Fourier transform is related to its real part through the Hilbert transform, that is

$$-\frac{1}{\pi}\int_{-\infty}^{\infty} \pi\delta(\nu)\frac{1}{\omega-\nu}d\nu = -\frac{1}{\omega}$$

This form of the Hilbert transform has applications in linear electrical circuits and electric power systems in order to find the imaginary part of the Fourier transform when its real part is known (obtained experimentally) and vice versa to find the real part from the imaginary part of the Fourier transform. For completeness, it is presented in this section, together with the second form of the Hilbert transform, which has a particular importance for the modulation process in communication systems.

Hilbert Transform for Real Signals

The second form of the Hilbert transform is derived for real Fourier transformable signals. Let $x(t) \leftrightarrow X(j\omega)$, that is

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

Consider the signal $x_+(t)$ whose spectrum is zero for negative frequencies and equal to $2X(j\omega)$ for positive frequencies, that is

$$x_+(t) = \frac{1}{\pi} \int_{0}^{\infty} X(j\omega) e^{j\omega t} d\omega$$

The following relationship exists between the signals $x(t)$ and $x_+(t)$

$$x_+(t) = x(t) + j\hat{x}(t)$$

where $\hat{x}(t)$ is the Hilbert transform of $x(t)$ defined by

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

This result can be shown as follows. From the expression for $x(t)$ and $\hat{x}(t)$, we have

$$X_+(j\omega) = 2U_h(j\omega)X(j\omega)$$

where $U_h(j\omega)$ represents the unit step function in the frequency domain. We know that $u_h(t) \leftrightarrow (1/j\omega) + \pi\delta(\omega)$. Using the duality property, we have

$$\frac{1}{j(-t)} + \pi\delta(t) \leftrightarrow 2\pi U_h(j\omega)$$

$$\Leftrightarrow \frac{1}{j2\pi(-t)} + \frac{1}{2}\delta(t) = \frac{1}{2}\delta(t) + \frac{j}{2\pi t} \leftrightarrow U_h(j\omega)$$

Since the product in the frequency domain corresponds to the convolution in the time domain, we have

$$x_{+}(t) = x(t) * \left[\delta(t) + \frac{j}{\pi t}\right] = x(t) + j\hat{x}(t)$$

Also

$$\hat{X}(j\omega) = \mathcal{F}\left(\frac{1}{\pi t}\right)X(j\omega)$$

Since $\mathrm{sgn}(t) \leftrightarrow 2/j\omega$ then by the duality property, we have

$$\frac{2}{jt} \leftrightarrow 2\pi\mathrm{sgn}(-\omega) \quad \Leftrightarrow \quad \frac{1}{t} \leftrightarrow j\pi\mathrm{sgn}(-\omega) = -j\pi\mathrm{sgn}(\omega)$$

so that

$$\hat{X}(j\omega) = -j\mathrm{sgn}(\omega)X(j\omega)$$

Finding analytically this form of the Hilbert transform requires that the signal Fourier transform is multiplied by $-j\mathrm{sgn}(\omega)$ and that the inverse Fourier transform is applied to the result obtained. This procedure is demonstrated in the next example.

Example 10.3: The Hilbert transform of the sine function is obtained as

$$-j\mathrm{sgn}(\omega)\mathcal{F}\{\sin(\omega_0 t)\} = j\mathrm{sgn}(\omega) \times j\pi[\delta(\omega+\omega_0) - \delta(\omega-\omega_0)]$$

$$= -\pi[\delta(\omega+\omega_0) + \delta(\omega-\omega_0)] = -\mathcal{F}\{\cos(\omega_0 t)\}$$

Hence, the Hilbert transform of the signal $\sin(\omega_0 t)$ is equal to $-\cos(\omega_0 t)$.

Using the definition of the sign function, we have

$$\hat{X}(j\omega) = -j\mathrm{sgn}(\omega)X(j\omega) = \begin{cases} -jX(j\omega) = e^{-j\frac{\pi}{2}}X(j\omega), & \omega > 0 \\ jX(j\omega) = e^{j\frac{\pi}{2}}X(j\omega), & \omega < 0 \end{cases}$$

It can be seen that *for real signals the Hilbert transform introduces the phase shift of* $-90^\circ = -\pi/2$ *for positive frequencies and the phase shift of* $90^\circ = \pi/2$ *for negative frequencies.*

The signal $x_+(t)$ is called the *positive frequency pre-envelope signal* of $x(t)$. Its main feature is given by its spectrum formula

$$X_+(j\omega) = \begin{cases} 2X(j\omega), & \omega > 0 \\ X(0), & \omega = 0 \\ 0, & \omega < 0 \end{cases}$$

The middle relation follows from the property of the frequency domain unit step function, that is $X_+(0) = 2U_h(0)X(0) = X(0)$. Similarly, we can define the *negative frequency pre-envelope signal* of $x(t)$ by $x_-(t) = x(t) - j\hat{x}(t)$. Its spectrum is

$$X_-(j\omega) = \begin{cases} 0, & \omega > 0 \\ X(0), & \omega = 0 \\ 2X(j\omega), & \omega < 0 \end{cases}$$

Applications of this form of the Hilbert transform in communication systems will be discussed in Section 10.5 within the single sideband modulation technique.

10.4 Ideal Filter

Signal filtering plays a very important role in communication systems. Filters can extract from a given frequency spectrum either low frequency components (*low-pass filtering*) or high frequency components (*high-pass filtering*) or signal components that belong to a certain frequency range (*band-pass filtering*). A filter can also eliminate certain components from the signal frequency spectrum (*band-stop filtering*).

It is important to know that an ideal filter that exactly passes the given range of frequency components and exactly suppresses the frequency components outside of that range is not physically realizable. However, the ideal filter has theoretical importance in understanding the interplay between the time and frequencies domains. Moreover, with sight modifications we can construct realizable filters starting with the frequency characteristics of ideal filters.

The frequency characteristics of such an ideal low-pass filter is presented in Figure 10.4. The frequency ω_0 is called the *filter cut-off frequency*. According to Figure 10.4 the ideal low-pass filter transfer function is given by

$$H(j\omega) = \begin{cases} 1 \times e^{-j\omega t_d}, & |\omega| \leq \omega_0 \\ 0, & \text{otherwise} \end{cases}$$

Note that we have assumed that the phase of the ideal filter changes linearly in frequency, which corresponds to the time shift of the filter input signals by t_d (time shifting property of the Fourier transform).

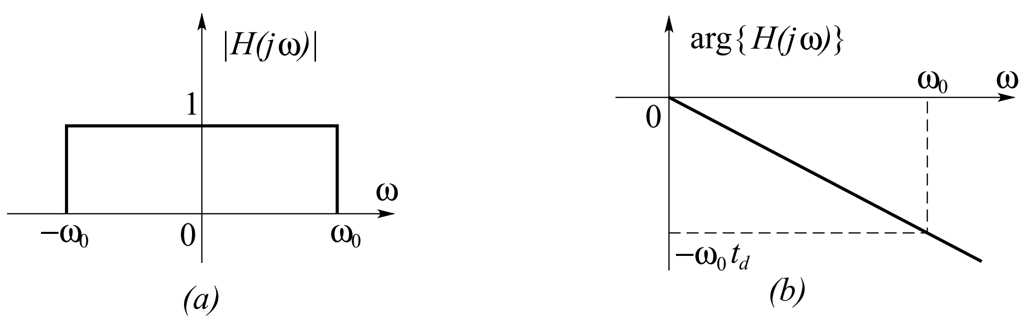

Figure 10.4: The frequency spectra of an ideal low−pass filter: (a) magnitude and (b) phase

In the following, we derive the impulse response of the ideal low-pass filter and show that such an impulse response does not correspond to the impulse response of a causal (real physical) system. We know that a rectangular frequency domain pulse has the following time domain Fourier equivalent (note that in this case $\tau = 2\omega_0$), see Example 3.16

$$p_{2\omega_0}(\omega) \leftrightarrow \frac{2\omega_0}{2\pi}\mathrm{sinc}\left(\frac{2\omega_0}{2\pi}t\right)$$

Using the time shift property of the Fourier transform, we have

$$H(j\omega) = p_{2\omega_0}(\omega)e^{-j\omega_0 t_d} \leftrightarrow \frac{\omega_0}{\pi}\mathrm{sinc}\left(\frac{\omega_0}{\pi}(t - t_d)\right) = h(t)$$

We have obtained the shifted sinc signal whose maximum, at $t = t_d$, is equal to ω_0/t_d. The waveform is present both left and right from the point $t = t_d$, having infinite duration in both directions, see Figure 10.5, where we use MATLAB to plot the corresponding impulse response for $\omega_0 = 5\,\mathbf{rad/s}$ and $t_d = 2\,\mathrm{s}$.

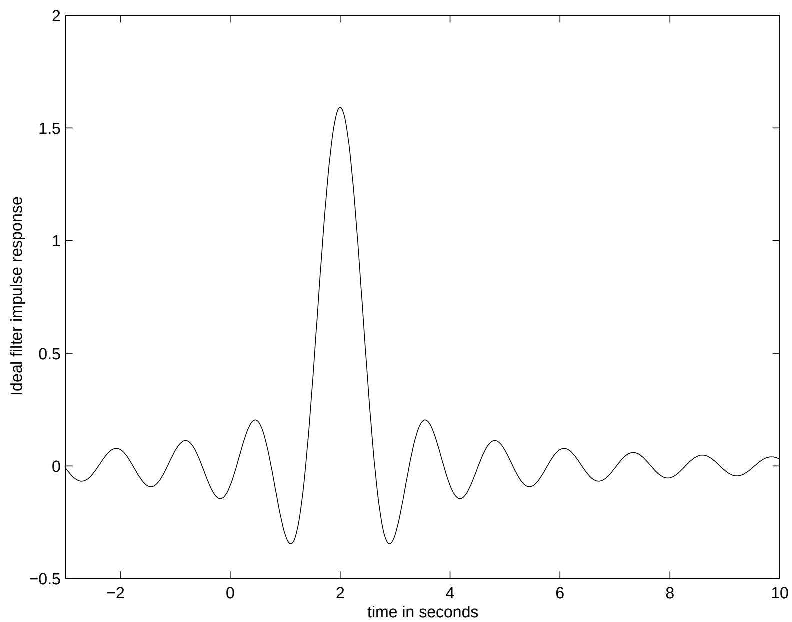

Figure 10.5: MATLAB plot of the impulse response of an ideal low−pass filter

It can be concluded that the ideal low-pass filter impulse response produces a waveform different from zero even for the times ($t < 0$) before the delta impulse input is applied to the filter. That violates the causality of the filter, that is, its physical realizability.

Problem 10.15: Direct derivations of the ideal filter impulse response

$$h(t) = \mathcal{F}^{-1}(H(j\omega)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} p_{2\omega_0}(\omega)e^{-j\omega t_d}e^{j\omega t}d\omega$$

$$= \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{j\omega(t-t_d)}d\omega = \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} \cos\left(\omega(t-t_d)\right)d\omega$$

$$= \frac{1}{2\pi}\frac{\sin\left(\omega(t-t_d)\right)}{t-t_d}\Big|_{\omega=-\omega_0}^{\omega=\omega_0} = \frac{\sin\left(\omega_0(t-t_d)\right)}{\pi(t-t_d)} = \frac{\omega_0}{\pi}\mathrm{sinc}(\omega_0(t-t_d))$$

10.5 Modulation and Demodulation

Signal modulation has been the cornerstone for the development of modern communication theory and its applications. More precisely, the formula

$$\mathcal{F}\{x(t)A_c\cos(\omega_c t)\} = 0.5A_cX(j(\omega+\omega_c)) + 0.5A_cX(j(\omega-\omega_c))$$

defines *amplitude modulation*. The signal $A_c\cos(\omega_c t)$ is the carrier signal with carrier frequency ω_c and carrier amplitude A_c. The modulated signal is $x(t)A_c\cos(\omega_c t)$. The baseband signal $x(t)$ is also called the *message signal*, *modulating signal*, or *original signal*. The spectra of the original (baseband) and modulated signals are presented in Figure 10.2.

There are several types of modulation techniques. In addition to amplitude modulation, we have *frequency modulation* and *phase modulation* techniques, in which, respectively, the carrier signal frequency and the carrier signal phase are affected by the baseband signal. Hence, in those cases the carrier frequency and the carrier phase carry information about the original signal $x(t)$. Frequency and phase amplitude modulation are outside the scope of this introductory chapter on communication systems.

In addition to the sinusoidal carrier, the *train of pulses* is used as the carrier signal. In the case of the train of pulses, we have again *amplitude modulation* (the pulse magnitude is proportional to the original signal magnitude at the given time instant), *pulse duration modulation* (the pulse width is proportional to the magnitude of the original signal), and *pulse position modulation* (the pulse position with respect to the reference position is determined by the magnitude of the original signal). The above pulse modulation techniques are specific for continuous-time (analog) signals. Due to space limitation and introductory nature of this chapter, continuous-time pulse modulation techniques will not be discussed.

For digital signals, we have the *pulse code modulation* technique, in which digital signals are binary encoded and the bits carrying information about signal magnitude are transmitted. We will say more about this modulation technique in the next section, where we present the essence of digital communication systems.

Amplitude Modulation

It can be seen from the previous analysis that the carrier signal amplitude is equal to A_c. Being multiplied by $x(t)$, the carrier signal changes its magnitude according to $x(t)A_c$, which is the way the carrier signal carries information about the signal $x(t)$. There are several variants of the amplitude modulation technique.

Let us demonstrate on a simple example that the envelope of the modulated signal can carry information about the original signal. At the same time we will establish conditions required for such a signal transmission.

Example 10.4: Consider a simple signal $x(t) = te^{-t}u_h(t)$. Its Fourier transform is given by $\mathcal{F}(x(t)) = 1/(1+j\omega)^2 = X(j\omega)$. The modulated signal $x(t)A_c \cos(\omega_c t),\ A_c = 1,$ and the original signal $x(t)$ are presented in Figures 10.6 and 10.7, respectively for $\omega_c = 20\,\mathrm{rad/s}$ and $\omega_c = 2\,\mathrm{rad/s}$.

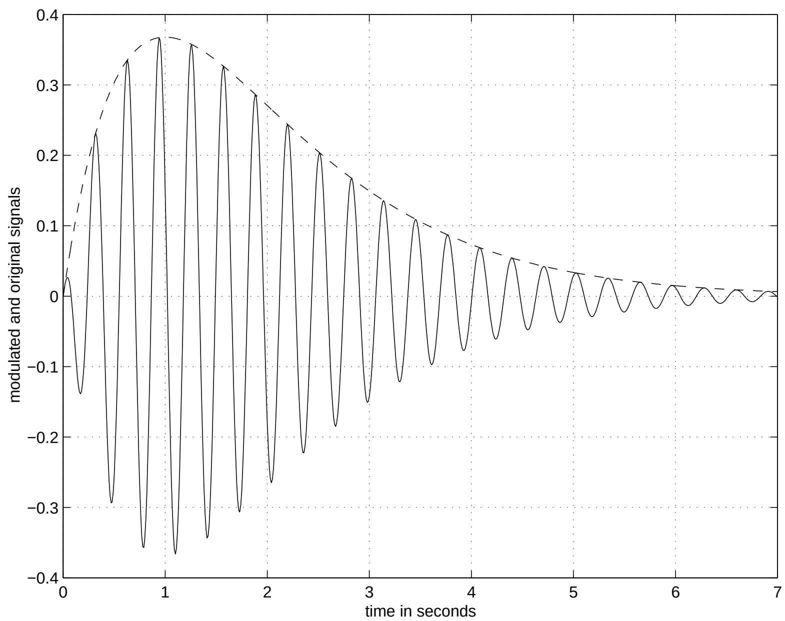

Figure 10.6: Modulated (solid line) and original (dashed line) signals for $\omega_c = 20\,\mathrm{rad/s}$

It can be seen from Figure 10.6 that the modulated signal in its envelope basically carries information about the original signal. It is natural to expect that such information is sufficient for recovery of the original signal. However, it follows from Figure 10.7 that in this case, the recovery process of the original signal from the modulated signal is very difficult if possible at all.

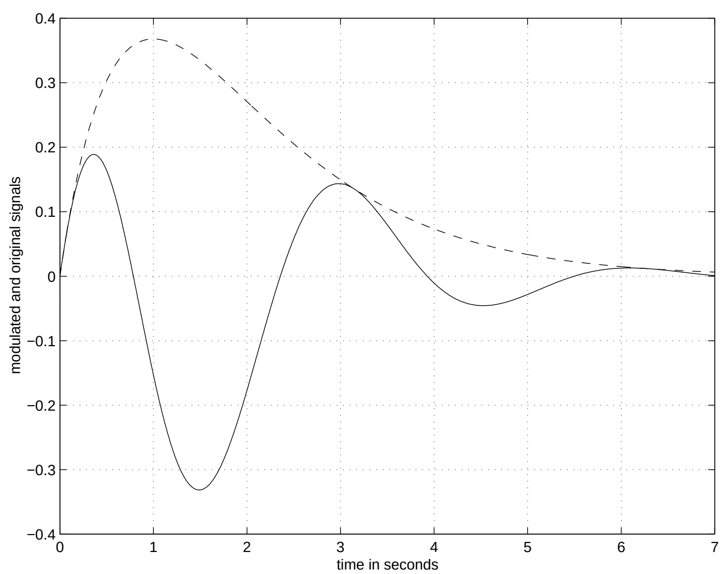

Figure 10.7: Modulated (solid line) and original (dashed line) signals for $\omega_c = 2\,\mathbf{rad/s}$

In Figure 10.8, we have presented the magnitude spectrum of the original signal. It can be seen from this figure that the signal has a significant frequency component at $\boldsymbol{\omega_c = 2\,\mathrm{rad/s}}$ and almost negligible frequency component at $\boldsymbol{\omega_c = 20\,\mathrm{rad/s}}$. We can draw a conclusion that for an easy and accurate signal recovery from the modulated signal the *carrier frequency must be much higher than the frequency of any significant spectral component of the signal*.

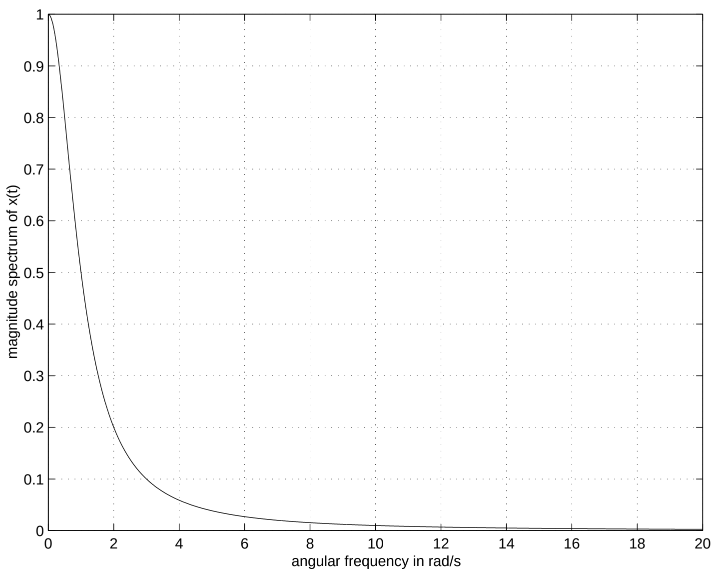

Figure 10.8: The magnitude spectrum of the original signal

Amplitude Modulation with a Transmitted Carrier

Note that in Example 10.4 the signal $x(t) = te^{-t}u_h(t)$ is positive for all t. If $x(t)$ changes its sign for some t, then the modulated signal $x(t)A_c\cos(\omega_c t)$ will change the phase at that time, such that its envelope will be distorted and it will no longer preserve the shape of the original signal. To prevent this problem, we can define the modulated signal using a slightly different modulation formula

$$(1 + k_a x(t))A_c\cos(\omega_c t) \leftrightarrow A_c\pi\delta(\omega + \omega_c) + A_c\pi\delta(\omega - \omega_c)$$

$$+0.5k_a A_c X(j(\omega + \omega_c)) + 0.5k_a A_c X(j(\omega - \omega_c))$$

where k_a is an arbitrary constant called either the *amplitude sensitivity* or *index of modulation*. By choosing this constant such that

$$1 + k_a x(t) > 0 \Rightarrow |k_a x(t)| < 1, \quad \forall t$$

the envelope of the modulated signal will have the shape of the original signal and hence carry information about the original signal at all times. This property will facilitate the use of simple modulators for signal modulation and simple demodulators (envelope detectors) for signal reconstruction.

The frequency domain price for such a time domain convenience is the presence of two additional delta impulses in the frequency spectrum of the modulated signal. Since A_c may have a large value, such as in the case of the modulator known as the switching modulator, a considerable amount of power is wasted in this kind of modulation known as *double sideband with transmitted carrier modulation* (DSB-TC). The originally considered modulation technique ($x(t)A_c\cos(\omega_c t)$) does not require an independent carrier transmission. It is known as *double sideband with suppressed carrier modulation* (DSB-SC).

In both DSB-SC and DSB-TC, the lower and upper signal frequency sidebands are transmitted. Since the signal information is completely contained in either the upper or lower frequency sideband, we conclude that these two modulation techniques waste a significant amount of the channel's frequency band. Exactly half of the frequency band can be saved by transmitting only the lower or upper signal frequency sideband. This can be facilitated by the modulation technique known as *single sideband (SSB) modulation*. Theoretical foundations for SSB modulation lie in the Hilbert transform considered in Section 10.3.

Switching Modulator and Envelope Detector (Demodulator)

Amplitude modulation with the transmitted carrier and corresponding demodulation are easily performed by using pretty simple electrical devices known as the *switching modulator* and *envelope detector (demodulator)*. They are presented in Figures 10.9 and 10.10.

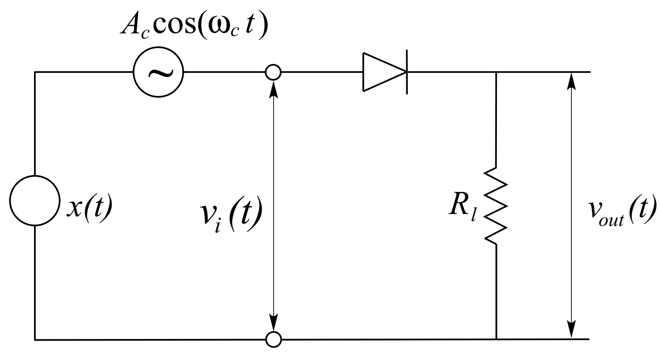

Figure 10.9: Switching modulator

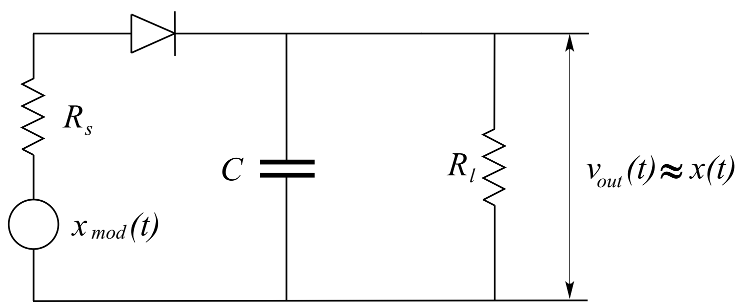

Figure 10.10: Envelope detector

Amplitude modulation for DSB-SC signals requires the use of more complex modulators. The most common of which is called the ring modulator. As the corresponding demodulator, the Costas receiver is mostly recommended. It is beyond the scope of this textbook to go into detail about these devices.

Note that MATLAB has the modulation function `modulate`, which can be used for any of the above three modulation techniques. Its general form is `xmod=modulate(x,fc,fs,'method',parameter)`, where `x` represents samples of the original continuous-time signal sampled with the frequency `fs`. `fc` is the carrier frequency ($f_c = \omega_c/2\pi$). `method` is either `amdsb-sc` or `amdsb-tc` or `amssb`, denoting respectively the modulation method used DSB-SC or DSB-TC or SSB. The choice of the `parameter` should be such that the modulating signal is positive with the minimum equal to zero. The `parameter` is set to zero for DSB-SC and SSB. It can also be omitted since its default value is zero. Similarly, the MATLAB function `demod` performs demodulation, which can be achieved by using the following MATLAB statement `x=demod(xmod,fc,fs,'method')`.

Problem 10.22

In this problem we use MATLAB to find the Fourier transform (spectrum) of the signal presented in Figure 2.8, find its DSB-SC and DSB-TC amplitude modulated signals and plot their spectra. Note that the chosen carrier frequency is $\boldsymbol{f_c = 0.1 f_s = 10\,\mathrm{Hz}}$, which implies that $\boldsymbol{\omega_c = 2\pi f_c = 62.8\,\mathrm{rad/s}}$.

```
Ts=0.01; tf=3; t=0:Ts:tf; tt=0:Ts:1;
xs=-tt+1; x=[zeros(1,1/Ts) xs zeros(1,1/Ts)];
figure (1); subplot(221); plot(t-1,x);
fs=1/Ts; fc=0.1*fs;
xmodSC=modulate(x,fc,fs,'amdsb-sc');
xmodTC=modulate(x,fc,fs,'amdsb-tc',0.1);
subplot(222); plot(t-1,xmodSC);
subplot(224); plot(t-1,xmodTC);
```

```
N=length(x)-1; X=Ts*fft(x,N);
XmodSC=Ts*fft(xmodSC,N);
XmodTC=Ts*fft(xmodTC,N);
k=0:1:N/2-1; w=(2*pi*k/N)/Ts
subplot(223); plot(w,abs(X(1:N/2)));
figure (2)
subplot(211); plot(w,abs(XmodSC(1:N/2)));
subplot(212); plot(w,abs(XmodTC(1:N/2)));
```

The results obtained are presented in FIGURES 10.7 and 10.8. Note that the signal spectra (Fourier transforms) are evaluated using FFT and the formula

$$X(j\omega) \approx T_s X(j\Omega)$$

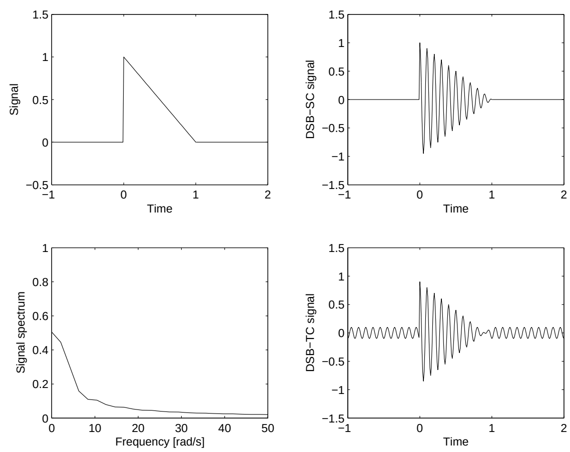

FIGURE 10.7 (Solutions Manual)

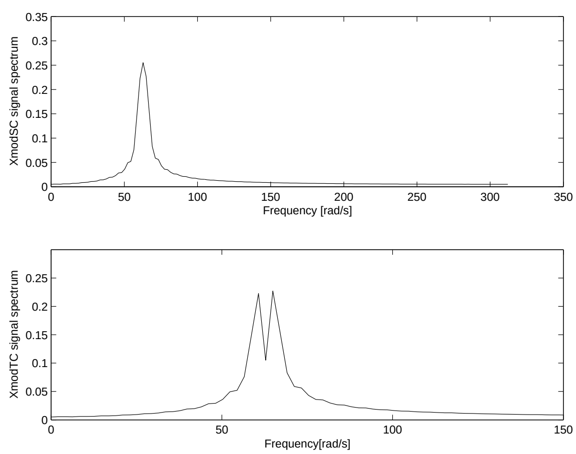

FIGURE 10.8 (Solutions Manual)

Single Sideband Amplitude Modulation

Theoretical foundations for the development of single sideband amplitude modulation lie in the Hilbert transform. The single sideband amplitude modulated signal can be obtained by using the Hilbert transform as follows. Consider the cosine modulated original signal, that is

$$s_{mod}^{cos}(t) = x(t)\cos(\omega_0 t) \leftrightarrow \frac{1}{2}X(j(\omega - \omega_0)) + \frac{1}{2}X(j(\omega + \omega_0))$$

The Hilbert transform of $x(t)$, denoted by $\hat{x}(t)$, modulated by the sine signal is

$$\hat{s}_{mod}^{sin}(t) = \hat{x}(t)\sin(\omega_0 t) \leftrightarrow \frac{j}{2}\hat{X}(j(\omega + \omega_0)) - \frac{j}{2}\hat{X}(j(\omega - \omega_0))$$

The signals $x(t)$ and $\hat{x}(t)$ are related through the Hilbert transform so that

$$\hat{X}(j\omega) = -j\mathrm{sgn}(\omega)X(j\omega)$$

which implies

$$\hat{s}_{mod}^{sin}(t) = \hat{x}(t)\sin(\omega_0 t)$$

$$\leftrightarrow \frac{1}{2}\mathrm{sgn}(\omega+\omega_0)X(j(\omega+\omega_0)) - \frac{1}{2}\mathrm{sgn}(\omega-\omega_0)X(j(\omega-\omega_0))$$

If we form now the new modulated signal as

$$s_{mod}(t) = \frac{1}{2}s_{mod}^{cos}(t) - \frac{1}{2}\hat{s}_{mod}^{sin}(t)$$

its frequency spectrum will be given by

$$\frac{1}{4}[1+\mathrm{sgn}(\omega-\omega_0)]X(j(\omega-\omega_0)) + \frac{1}{4}[1+\mathrm{sgn}(\omega-\omega_0)]X(j(\omega+\omega_0))$$

Having in mind the expression for the signum function, we see that the spectrum of the above modulated signal has the following form

$$
\mathcal{F}(s_{mood}(t)) = \mathcal{F}\left(\frac{1}{2}s^{cos}_{mod}(t) - \frac{1}{2}\hat{s}^{sin}_{mod}(t)\right)
$$

$$
= \begin{cases} 0.5X(j(\omega - \omega_0)), & \omega \geq \omega_0 \\ 0.5X(j(\omega + \omega_0)), & \omega \leq -\omega_0 \\ 0, \ \ \text{otherwise} \end{cases}
$$

This spectrum is presented in Figure 10.11.

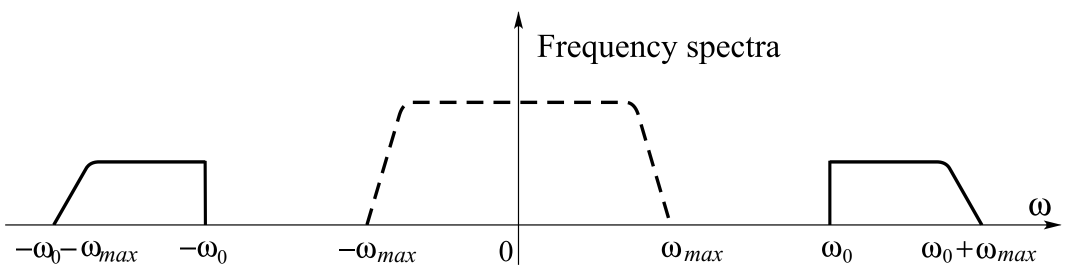

Figure 10.11: The frequency spectra of the original (dashed line) and single sideband modulated signal (solid line)

Similarly, it can be shown that the frequency magnitude spectrum of the signal

$$\frac{1}{2} s_{mod}^{cos}(t) + \frac{1}{2} \hat{s}_{mod}^{sin}(t)$$

contains only the lower frequency sidebands, that is

$$\mathcal{F}\left(\frac{1}{2} s_{mod}^{cos}(t) + \frac{1}{2} \hat{s}_{mod}^{sin}(t)\right) = \begin{cases} 0.5X(j(\omega - \omega_0)), & 0 \leq \omega \leq \omega_0 \\ 0.5X(j(\omega + \omega_0)), & 0 \geq \omega \geq -\omega_0 \\ 0, \ \ \text{otherwise} \end{cases}$$

Demodulation of SSB Signals

The original signal can be extracted from a single sideband amplitude modulated signal by modulating the modulated signal again using the signal of the same frequency and phase, that is

$$\mathcal{F}\{[x(t)\cos(\omega_c t) - \hat{x}(t)\sin(\omega_c t)]\cos(\omega_c t)\}$$

$$= \mathcal{F}\left\{\frac{1}{2}x(t) + \frac{1}{2}x(t)\cos(2\omega_c t) + \frac{1}{2}\hat{x}(t)\sin(2\omega_c t)\right\}$$

$$= \frac{1}{2}X(j\omega) + \frac{1}{4}X(j(\omega + 2\omega_c)) + \frac{1}{4}X(j(\omega - 2\omega_c))$$

$$+\frac{j}{4}\hat{X}(j(\omega + 2\omega_c)) - \frac{j}{4}\hat{X}(j(\omega - 2\omega_c))$$

The original signal can be easily extracted by using a lower pass filter since

This demodulation technique is called *coherent demodulation*

10.6 Digital Communication Systems

Nowadays signal transmission in communication systems is mostly done digitally. The advantage of digital signal transmission techniques is their improved tolerance to noise. Noise is unavoidably present in all communication channels. The rapid development of digital computer networks, digital signal processing, fast electronic and photonic switching devices during the last ten years has facilitated powerful signal transmission techniques that can make digital communication systems more efficient than corresponding analog communication systems.

In the introductory Section 1.1.1, we have introduced the concept of discretization of continuous-time signals with the given sampling period, which leads to the formation of discrete-time signals. The device that performs the signal discretization (sampling) is called the *sampler*. In addition of being discretized, in digital communication systems, signals are also quantized (discretized with respect to the magnitude). The device that performs such a magnitude quantization is called the *quantizer*. Such an obtained discretized and quantized signal is called the digital signal. Finally, the digital signal obtained is encoded into a stream of bits. This process composed of sampling, quantization, and encoding, is known as the *pulse code modulation* (PCM) technique. It is symbolically presented in Figure 10.12.

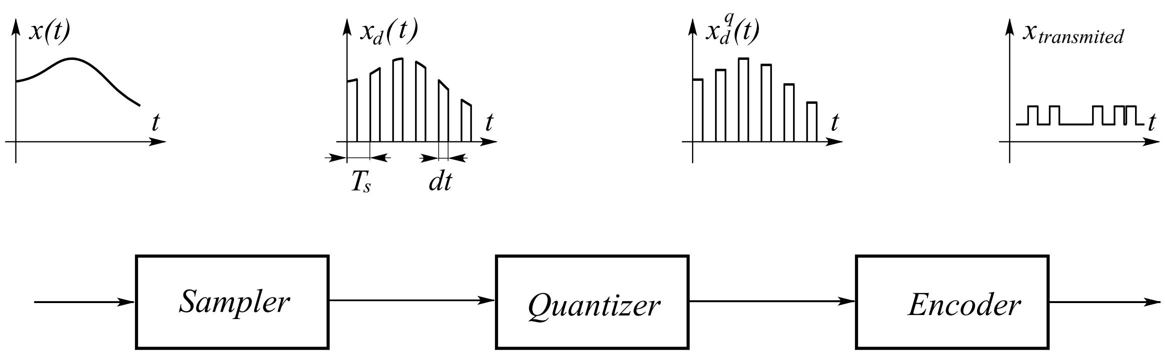

Figure 10.12: Pulse code modulation technique

The transmitter in a digital communication system performs pulse code modulation on an incoming signal and forms the encoded binary signal. The encoded binary signal is then sent over a communication channel as a stream of bits.

In this section, we have presented only the essential idea of digital communications. Further study of digital communications is beyond the scope of this chapter.

Example 10.5 PCM for Speech Signals

Speech (telephone) signals are sampled every $125\,\mu\mathrm{s}$, which generates 8000 samples per second. Quantization of speech signals is performed at $128 = 2^7$ levels, with each quantized sample being encoded using 8 bits (one bit for the sign). This generates $8000 \times 8 = 64000$ bits per second, commonly denoted as $64\,\mathrm{kbps}$ (kilo bits per second). Hence, while talking on the telephone, each user (speaker) generates $64\,\mathrm{kb}$ every second. Owing to recent advances in digital communication networks that use optical fiber channels such a heavy bit stream can be easily handled.