# I

# PRELIMINARIES

# 1

# INTRODUCTION

In this chapter we provide a brief and concise review of foundational topics that are of broad interest and usefulness in wireless communication engineering technologies. The notation used throughout is introduced in Section 1.1, and the basics of electrical circuits and signals are reviewed in Section 1.2, including fundamentals of circuit analysis, voltage or current as signals, alternating current, phasors, impedance, and matched loads. This provides a basis for our review of signals and systems in Section 1.3, which includes properties of linear time-invariant systems, Fourier analysis and frequency-domain concepts, representations of bandpass signals, and modeling of random signals. Then in Section 1.4, we focus on signals and systems concepts specifically for communications systems. The reader is expected to have come across much of the material in this chapter in a typical undergraduate electrical engineering program. Therefore, this chapter is written in review form; it is not meant for a student who is encountering all this material for the first time.

Similarly, reviews of foundational topics are provided in Chapters 2, 6, and 10 for the following areas:

- *Chapter 2:* review of selected topics in electromagnetics, transmission lines, and testing, as a foundation for radio frequency (RF), antennas, and propagation
- *Chapter 6:* review of selected topics in digital signal processing, digital communcations over wireless links, the cellular concept, spread spectrum, and othogonal frequency-division multiflexing (OFDM), as a foundation for wireless access technologies

*Fundamentals of Wireless Communication Engineering Technologies*, First Edition. K. Daniel Wong.

- *Chapter 10:* review of selected topics in fundamental networking concepts, Internet protocol (IP) networking, and teletraffic analysis, as a foundation for network and service architectures

Compared to the present chapter, the topics in Chapters 2, 6, and 10 are generally more specific to particular areas. Also, we selectively develop some of the topics in those chapters in more detail than we do in this chapter.

## 1.1 NOTATION

In this section we discuss the conventions we use in this book for mathematical notation. A list of symbols is provided in Appendix D.

$\mathcal{R}$ and $\mathcal{C}$ represent the real and complex numbers, respectively. Membership in a set is represented by $\in$ (e.g., $x \in \mathcal{R}$ means that $x$ is a real number). For $x \in \mathcal{C}$, we write $\Re\{x\}$ and $\Im\{x\}$ for the real and imaginary parts of $x$, respectively.

log represents base-10 logarithms unless otherwise indicated (e.g., $\log_2$ for base-2 logarithms), or where an expression is valid for all bases.

Scalars, which may be real or even complex valued, are generally represented by italic type (e.g., $x$, $y$), whereas vectors and matrices will be represented by bold type (e.g., **G**, **H**). We represent a complex conjugate of a complex number, say an impedance $Z$, by $Z^*$. We represent the magnitude of a complex number $x$ by $|x|$. Thus, $|x|^2 = xx^*$.

For $x \in \mathcal{R}$, $\lfloor x \rfloor$ is the largest integer $n$ such that $n < x$. For example, $\lfloor 5.67 \rfloor = 5$ and $\lfloor -1.2 \rfloor = -2$.

If $\mathbf{G}$ is a matrix, $\mathbf{G}^\mathbf{T}$ represents its transpose.

When we refer to a matrix, vector, or polynomial as being *over* something (e.g., *over the integers*), we mean that the components (or coefficients, in the case of polynomials) are numbers or objects of that sort.

If $x(t)$ is a random signal, we use $< x(t) >$ to refer to the time average and $\overline{x(t)}$ to refer to the ensemble average.

## 1.2 FOUNDATIONS

Interconnections of electrical elements (resistors, capacitors, inductors, switches, voltage and current sources) are often called a *circuit*. Sometimes, the term *network* is used if we want "circuit" to apply only to the more specific case of where there is a closed loop for current flow. In Section 1.2.1 we review briefly this type of electrical network or circuit. Note that this use of "network" should not be confused with the very popular usage in the fields of computer science and telecommunications, where we refer to computer networks and telecommunications networks (see Chapters 9 to 12 for further discussion). In Chapter 2 we will see how transmission lines (Section 2.3.3) can be modeled as circuit elements and so can be part of electrical networks and circuits.

In *electronic* networks and circuits, we also have components with *gain* and/or *directionality*, such as semiconductor devices, which are known as *active* components (as opposed to *passive* components, which have neither gain nor directionality). These are outside the scope of this book, except for our discussion on RF engineering in Chapter 3. Even there, we don't discuss the physics of the devices or compare different device technologies. Instead, we take a "signals and systems" perspective on RF, and consider effects such as noise and the implications of nonlinearities in the active components.

### 1.2.1 Basic Circuits

Charge, $Q$, is quantified in coulombs. Current is charge in motion:

$$I = \frac{dQ}{dt} \qquad \text{amperes} \tag{1.1}$$

The direction of current flow can be indicated by an arrow next to a wire. For convenience, $I$ can take a negative value if current is flowing in the direction opposite from that indicated by the arrow.

Voltage is the difference in electric potential:

$$V = RI \qquad \text{volts} \tag{1.2}$$

Like current, there is a direction associated with voltage. It is typically denoted by $+$ and $-$. $+$ is at higher potential than $-$, and voltage drops going from $+$ to $-$. For convenience, $V$ can take a negative value if a voltage drop is in the direction opposite from that indicated by $+$ and $-$

- Power:

$$P = \frac{V^2}{R}, \qquad P = I^2 R \qquad \text{watts} \tag{1.3}$$

- Resistors in series:

$$R = R_1 + R_2 + \cdots + R_n \tag{1.4}$$

- Resistors in parallel:

$$R = \frac{R_1 R_2 \cdots R_n}{R_1 + R_2 + \cdots + R_n} \tag{1.5}$$

### 1.2.2 Capacitors and Inductors

A capacitor may be conceived of in the form of two parallel plates. For a capacitor with capacitance $C$ farads, a voltage $V$ applied across its plates results in charges $+Q$ and $-Q$ accumulating on the two plates.

$$Q = CV \tag{1.6}$$

$$I = \frac{dQ}{dt} = C\frac{dV}{dt} \tag{1.7}$$

A capacitor acts as an open circuit under direct-current (dc) conditions.

- Capacitors in series:

$$C = \frac{C_1 C_2 \cdots C_n}{C_1 + C_2 + \cdots + C_n} \tag{1.8}$$

- Capacitors in parallel:

$$C = C_1 + C_2 + \cdots + C_n \tag{1.9}$$

An inductor is often in the form of a coil of wire. For an inductor with inductance $L$ henries, a change in current of $dI/dt$ induces a voltage $V$ across the inductor:

$$V = L\frac{dI}{dt} \tag{1.10}$$

An inductor acts as a short circuit under dc conditions.

- Inductors in series:

$$L = L_1 + L_2 + \cdots + L_n \tag{1.11}$$

- Inductors in parallel:

$$L = \frac{L_1 L_2 \cdots L_n}{L_1 + L_2 + \cdots + L_n} \tag{1.12}$$

As hinted at by (1.3), an ideal capacitor or ideal inductor has no resistance and does not dissipate any power as heat. However, a practical model for a real inductor has an ideal resistor in series with an ideal inductor, and they are both in parallel with an ideal capacitor.

### 1.2.3 Circuit Analysis Fundamentals

A *node* in a circuit is any place where two or more circuit elements are connected. A *complete loop* or *closed path* is a continuous path through a circuit that begins and ends at the same node.

*Kirchhoff's Current Law.* *The sum of all the currents entering is zero.* This requires at least one current to have a negative sign if one or more of the others is positive. Alternatively, we say that the sum of all the current entering a node is equal to the sum of all the current leaving a node.

*Kirchhoff's Voltage Law.* *The sum of all the voltage drops around any complete loop (or closed path) is zero.* This requires at least one voltage drop to have a negative sign if one or more of the others is positive.

**1.2.3.1 *Equivalent Circuits*** Often, a subcircuit is connected to the rest of the circuit through a pair of terminals, and we are interested to know what the voltage and current are across these terminals, not how the subcircuit is actually implemented. Norton and Thévenin equivalent circuits can be used for this purpose, for any circuit comprising linear elements. A *Thévenin equivalent circuit* comprises a single voltage source, $V_T$, in series with a single resistor, $R_T$. A *Norton equivalent circuit* comprises a single current source, $I_N$, in parallel with a single resistor, $R_N$. A Thévenin equivalent circuit can be converted to a Norton equivalent circuit, or vice versa, by a simple source transformation.

### 1.2.4 Voltage or Current as Signals

A voltage or current can be interpreted as a signal (e.g., for communications purposes). We usually write $t$ explicitly to emphasize that it is a function of $t$ [e.g., $v(t)$ or $i(t)$ for a voltage signal or current signal, respectively].

If $x(t)$ is a signal, we say that $x(t)$ is

- An *energy signal* if

$$0 < \int_{-\infty}^{\infty} x^2(t)\, dt < \infty \tag{1.13}$$

- A *power signal* if

$$0 < \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} x^2(t)\, dt < \infty \tag{1.14}$$

A *periodic* signal is a signal for which a $T \in \mathcal{R}$ can be found such that

$$x(t) = x(t+T) \qquad \text{for } -\infty < t < \infty \tag{1.15}$$

and the smallest such $T$ is called the *period* of the signal.

The duration of a signal is the time interval from when its begins to be nonnegligible to when its stops being nonnegligible.† Thus, a signal can be of finite duration or of infinite duration.

*Sinusoidal Signals.* Any sinusoid that is a function of a single variable (say, the time variable, $t$; later, in Section 2.1.1.4, we see sinusoids that are functions of both

† We say nonnegligible rather than nonzero to exclude trivial blips outside the duration of the signal.

temporal and spatial variables) can be written as

$$A\cos(\omega t+\phi)=A\cos(2\pi ft+\phi)=A\sin(2\pi ft+\phi+\pi/2)=A\angle\phi \tag{1.16}$$

where $A$ is amplitude ($A \in \mathcal{R}$), $\omega$ is *angular frequency* (radians/second), $f$ is *frequency* (cycles/second, i.e., hertz or $s^{-1}$), $\phi$ is *phase angle*, and where the last equality shows that the shorthand notation $A\angle\phi$ can be used when $f$ and the sinusoidal reference time are known implicitly. The period $T$ is

$$T=\frac{1}{f}=\frac{2\pi}{\omega} \tag{1.17}$$

*Continuous-Wave Modulation Signals.* A continuous-wave modulation signal is a sinusoidal signal that is *modulated* (changed) in a certain way based on the information being communicated. Most communications signals are based on continuous-wave modulation, and we expand on this important topic in Section 1.4.

*Special Signals.* A fundamental building block in continuous-time representation of digital signals is the rectangular pulse signal, a rectangular function given by

$$\Pi(t)=\begin{cases}1 & \text{for } |t|\le 1/2\\ 0 & \text{for } |t|>1/2\end{cases} \tag{1.18}$$

The triangle signal is also commonly used, but not as frequently. It is denoted by

$$\Lambda(t)=\begin{cases}1-|t| & \text{for } |t|\le 1\\ 0 & \text{for } |t|>1\end{cases} \tag{1.19}$$

$\Pi(t)$ and $\Lambda(t)$ are shown in Figure 1.1.

The sinc signal is given by

$$\text{sinc}(t)=\begin{cases}(\sin\pi t)/\pi t & \text{for } |t|\ne 0\\ 1 & \text{for } t=0\end{cases} \tag{1.20}$$

Although it may be described informally as $(\sin\pi t)/\pi t$, $(\sin\pi t)/\pi t$ is actually undefined at $t=0$, whereas sinc$(t)$ is 1 at $t=0$. The sinc function is commonly seen in communications because it is the Fourier transform of the rectangular pulse signal. Note that in some fields (e.g., mathematics), sinc$(t)$ may be defined as $(\sin t)/t$, but

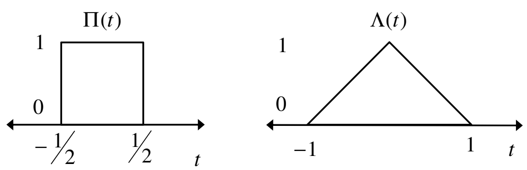


**FIGURE 1.1** $\Pi(t)$ and $\Lambda(t)$ functions.

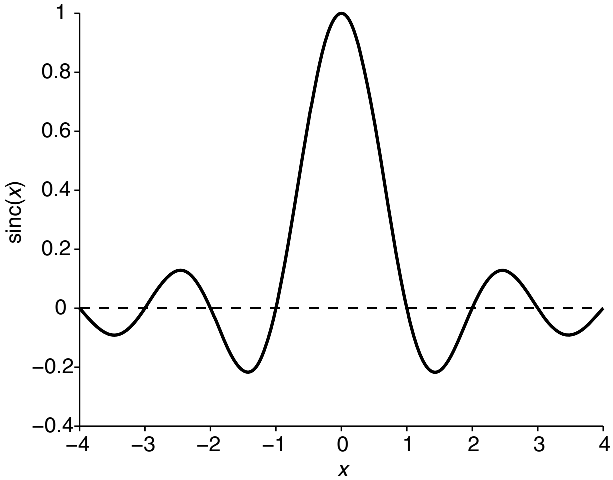


**FIGURE 1.2** Sinc function.

here we stick with our definition, which is standard for communications and signal processing. The sinc function is shown in Figure 1.2.

*Decibels.* It is sometimes convenient to use a log scale when the range of amplitudes can vary by many orders of magnitude, such as in communications systems where the signals have amplitudes and powers that can vary by many orders of magnitude. The standard way to use a log scale in this case is by the use of decibels, defined for any signal voltage or current signal $x(t)$ as

$$10 \log x^2(t) = 20 \log x(t) \tag{1.21}$$

If the signal $s(t)$ is known to be a power rather than a voltage or current, we don't have to convert it to a power, so we just take $10 \log s(t)$. If the power quantity is in watts, it is sometimes written as dBW, whereas if it is in milliwatts, it is written as dBm. This can avoid ambiguity in cases where we just specify a dimensionless quantity $A$, in decibels, as $10 \log A$.

### 1.2.5 Alternating Current

With alternating current (ac) the voltage sources or current sources generate time-varying signals. Then (1.3) refers only to the *instantaneous power*, which depends on the instantaneous value of the signal. It is often also helpful, perhaps more so, to consider the *average power*. Let $v(t) = V_0 \cos 2\pi ft$, where $V_0$ is the maximum voltage (and correspondingly, let $I_0$ be the maximum current), then the average power $P_{av}$ is

$$P_{av} = \frac{V_0^2}{2R}, \qquad P_{av} = \frac{I_0^2 R}{2} \tag{1.22}$$

Equation (1.22) can be obtained either by averaging the instantaneous power directly over one cycle, or through the concept of *rms voltage* and *rms current*. The rms voltage is defined for any periodic signal (not just sinusoidally periodic) as

$$V_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T v^2(t)\,dt} \tag{1.23}$$

Then we have (again, for any periodic signal, not just sinusoidally periodic)

$$P_{\text{av}} = \frac{V_{\text{rms}}^2}{R}, \qquad P_{\text{av}} = I_{\text{rms}}^2 R \tag{1.24}$$

which looks similar to (1.3). For sinusoidally time-varying signals, we have further,

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}}, \qquad I_{\text{rms}} = \frac{I_0}{\sqrt{2}} \tag{1.25}$$

### 1.2.6 Phasors

When working with sinusoidal signals, it is often convenient to work with the *phasor representation* of the signals. Of the three quantities amplitude, phase, and frequency, the phasor representation includes only the amplitude and phase; the frequency is implicit.

Starting from our sinusoid in (1.16) and applying Euler's identity (A.1), we obtain

$$A\cos(2\pi ft + \phi) = A\Re\left\{e^{j(2\pi ft+\phi)}\right\} = \Re\left\{Ae^{j(2\pi ft+\phi)}\right\} \tag{1.26}$$

We just drop the $e^{j2\pi ft}$ and omit mentioning that we need to take the real part, and we have a phasor,

$$Ae^{j\phi} \tag{1.27}$$

Alternatively, we can write the equivalent,

$$A(\cos\phi + j\sin\phi) \tag{1.28}$$

which is also called a phasor. In either case, we see that a phasor is a complex number representation of the original sinusoid, and that it is easy to get back the original sinusoid by multiplying by $e^{j2\pi ft}$ and taking the real part. A hint of the power and convenience of working with phasor representations can be seen by considering differentiation and integration of phasors. Differentiation and integration with respect to $t$ are easily seen to be simple multiplication and division, respectively, by $j2\pi f$.

*Rotating Phasors.* Sometimes it helps to think of a phasor not just as a static point in the complex plane, but as a rotating entity, where the rotation is at frequency $f$ revolutions (around the complex plane) per second, or $w$ radians per second. This is consistent with the $e^{j2\pi ft}$ term that is implicit in phasors. The direction of rotation is as illustrated in Figure 1.3.

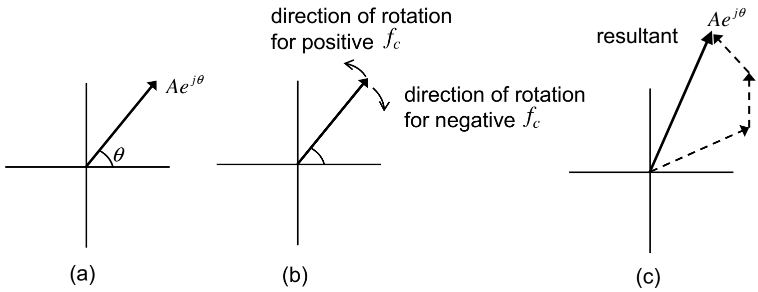


**FIGURE 1.3** (a) Phasor in the complex plane; (b) rotating phasors and their direction of rotation; (c) vector addition of phasors.

*Expressing Familiar Relationships in Terms of Phasors.* Returning to familiar relationships such as (1.2) or (1.3), we find no difference if $v(t)$, $i(t)$ are in phasor representation; however, for capacitors and inductors we have

$$I = j2\pi fCV \quad \text{and} \quad V = j2\pi fLI \tag{1.29}$$

Thus, if we think in terms of rotating phasors, then from (1.29) we see that with a capacitor, $I$ rotates 90° ahead of $V$, so it *leads* $V$ (and $V$ *lags* $I$), whereas with an inductor, $V$ leads $I$ ($I$ lags $V$).

Meanwhile, Kirchhoff's laws take the same form for phasors as they do for non-phasors, so they can continue to be used. Thévenin and Norton equivalent circuits can also be used, generalized to work with impedance, a concept that we discuss next.

### 1.2.7 Impedance

From (1.29) it can be seen that in phasor representation, resistance, inductance, and capacitance all have the same form:

$$V = ZI \tag{1.30}$$

Thus, the concept of *impedance*, $Z$, emerges, where $Z$ is $R$ for resistance, $j2\pi fL$ for inductance, and $1/j2\pi fC$ for capacitance, and $Z$ is considered to be in ohms. The complex part of $Z$ is also known as *reactance*.

Impedance is a very useful concept. For example, Thévenin's and Norton's equivalent circuits work in the same way with phasors, except that impedance is substituted for resistance.

### 1.2.8 Matched Loads

For a linear circuit represented by a Thévenin equivalent voltage $V_T$ and Thévenin equivalent impedance $Z_T$, the maximum power is delivered to a load $Z_L$ when

$$Z_L = Z_T^* \tag{1.31}$$

(NB: It is the complex conjugate of $Z_T$, not $Z_T$ itself, in the equation.) This result can be obtained by writing the expression for power in terms of $Z_L$ and $Z_T$, taking partial derivatives with respect to the load resistance and load reactance, and setting both to 0.

## 1.3 SIGNALS AND SYSTEMS

Similarly, suppose that we have a system (e.g., a circuit) that takes an input $x(t)$ and produces an output $y(t)$. Let $\longrightarrow$ represent the operation of the system [e.g., $x(t) \longrightarrow y(t)$]. Suppose that we have two different inputs, $x_1(t)$ and $x_2(t)$, such that $x_1(t) \longrightarrow y_1(t)$ and $x_2(t) \longrightarrow y_2(t)$. Let $a_1$ and $a_2$ be any two scalars. The system is *linear* if and only if

$$a_1 x_1(t) + a_2 x_2(t) \longrightarrow a_1 y_1(t) + a_2 y_2(t) \tag{1.32}$$

The phenomenon represented by (1.32) can be interpreted as the *superposition* property of linear systems. For example, given knowledge of the response of the system to various sinusoidal inputs, we then know the response of the system to any linear combination of sinusoidal signals. This makes Fourier analysis (Section 1.3.2) very useful.

A system is *time-invariant* if and only if

$$x(t - t_0) \longrightarrow y(t - t_0) \tag{1.33}$$

Systems that are both linear and time invariant are known as *LTI* (linear time-invariant) *systems*.

A system is *stable* if bounded input signals result in bounded output signals.

A system is *causal* if any output does not come before the corresponding input.

### 1.3.1 Impulse Response, Convolution, and Filtering

An impulse (or unit impulse) signal is defined as

$$\delta(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{1.34}$$

and also where

$$\int_{-\infty}^{\infty} \delta(t) = 1 \tag{1.35}$$

Strictly speaking, $\delta(t)$ is not a function, but to be mathematically rigorous requires measure theory or the theory of generalized functions. $\delta(t)$ could also be thought of as

$$\lim_{T \to \infty} T\Pi\,(tT) \tag{1.36}$$

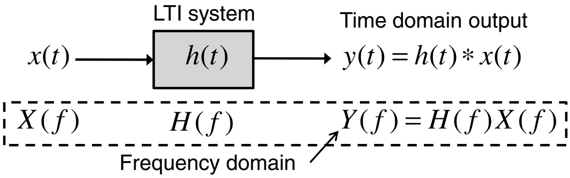


**FIGURE 1.4** Mathematical model of an LTI system.

Thus, we often view it as the limiting case of a narrower and narrower pulse whose area is 1.

All LTI systems can be characterized by their *impulse response*. The impulse response, $h(t)$, is the output when the input is an impulse signal; that is,

$$\delta(t) \longrightarrow h(t) \tag{1.37}$$

Convolution: The output of an LTI system with impulse response $h(t)$, given an input $x(t)$, is

$$y(t) = h(t) * x(t) = \int_{\tau=-\infty}^{\tau=\infty} x(\tau)h(t-\tau)\,d\tau = \int_{\tau=-\infty}^{\tau=\infty} h(\tau)x(t-\tau)\,d\tau \tag{1.38}$$

This is shown as the output of the LTI system in Figure 1.4.

With (1.38) in mind, whenever we put a signal $x(t)$ into an LTI system, we can think in terms of the system as *filtering* the input to produce the output $y(t)$, and $h(t)$ may be described as the impulse response of the filter. Although the term *filter* is used in the RF and baseband parts of wireless transmitters and receivers, $h(t)$ can equally well represent the impulse response of a communications channel (e.g., a wire, or wireless link), in which case we may then call it the *channel response* or simply the *channel*.

***1.3.1.1 Autocorrelation*** It is sometimes useful to quantify the similarity of a signal at one point in time with itself at some other point in time. Autocorrelation is a way to do this. If $x(t)$ is a complex-valued energy signal (a real-valued signal is a special case of a complex-valued signal, where the imaginary part is identically zero, and the complex conjugate of the signal is equal to the signal itself), we define the autocorrelation function, $R_{xx}(\tau)$, as

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x^*(t+\tau)\,dt \qquad \text{for } -\infty < \tau < \infty \tag{1.39}$$

For a complex-valued periodic power signal with period $T_0$,

$$R_{xx}(\tau) = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} x(t)x^*(t+\tau)\,dt \qquad \text{for } -\infty < \tau < \infty \tag{1.40}$$

whereas for a complex-valued power signal, in general,

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x^*(t+\tau)\,dt \qquad \text{for } -\infty < \tau < \infty \tag{1.41}$$

### 1.3.2 Fourier Analysis

*Fourier analysis* refers to a collection of related techniques where:

- A signal can be broken down into sinusoidal components *(analysis)*
- A signal can be constructed from constituent sinusoidal components *(synthesis)*

This is very useful in the study of linear systems because the effects of such a system on a large class of signals can be studied by considering the effects of the system on sinusoidal inputs using the superposition principle. (NB: The term *analysis* here can be used to refer either to just the breaking down of a signal into sinusoidal components, or in the larger sense to refer to the entire collection of these related techniques.)

Various Fourier *transforms* are used in analysis, and *inverse transforms* are used in synthesis, depending on the types of signals involved. For most practical purposes, there is a one-to-one relationship between a time-domain signal and its Fourier transform, and thus we can think of the Fourier transform of a signal as being a different *representation* of the signal. We usually think of there being two domains, the *time domain* and the *frequency domain*. The (forward) transform typically transforms a *time-domain representation* of a signal into a *frequency-domain representation*, whereas the inverse transform transforms a frequency-domain representation of a signal into a time-domain representation.

***1.3.2.1 (Continuous) Fourier Transform*** The (continuous) Fourier transform of a signal $x(t)$ is given by

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt \tag{1.42}$$

where $j = \sqrt{-1}$, and the inverse Fourier transform is given by

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df \tag{1.43}$$

Table 1.1 gives some basic Fourier transforms.

***1.3.2.2 Fourier Series*** For periodic signals $x(t)$ with period $T$, the Fourier series (exponential form) coefficients are the set $\{c_n\}$, where $n$ ranges over all the integers,

**TABLE 1.1 Fourier Transform Pairs**[a]

| Time Domain, $x(t)$ | Frequency Domain, $X(f)$ |
|---|---|
| $\delta(t)$ | $1$ |
| $1$ | $\delta(f)$ |
| $\delta(t - t_0)$ | $e^{-j2\pi f t_0}$ |
| $e^{\pm j2\pi f_0 t}$ | $\delta(f \mp f_0)$ |
| $\cos 2\pi f_0 t$ | $\frac{1}{2}[\delta(f - f_0) + \delta(f + f_0)]$ |
| $\sin 2\pi f_0 t$ | $\frac{1}{2j}[\delta(f - f_0) - \delta(f + f_0)]$ |
| $u(t) = \begin{cases} 1 & \text{for } t > 0 \\ 0 & \text{for } t < 0 \end{cases}$ | $\frac{1}{2}\delta(f) + \frac{1}{j2\pi f}$ |
| $e^{-at}u(t),\ a > 0$ | $\frac{1}{a + j2\pi f}$ |
| $te^{-at}u(t),\ a > 0$ | $\frac{1}{(a + j2\pi f)^2}$ |
| $e^{-a\|t\|},\ a > 0$ | $\frac{2a}{a^2 + (2\pi f)^2}$ |
| $\Pi\left(\frac{t}{T}\right)$ | $T \text{ sinc } fT$ |
| $B \text{ sinc } Bt$ | $\Pi\left(\frac{f}{B}\right)$ |
| $\Lambda\left(\frac{t}{T}\right)$ | $T \text{ sinc}^2 fT$ |
| $\sum_{k=-\infty}^{\infty} \delta(t - kT)$ | $\frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right)$ |

[a] $\Pi(t)$ and $\Lambda(t)$ are the rectangle and triangle functions defined in Section 1.2.4. $\sum_{k=-\infty}^{\infty} \delta(t - kT)$ is also known as an impulse train.

and $c_n$ is given by

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-j2\pi f_0 n t}\, dt \tag{1.44}$$

where $f_0 = 1/T$, and the Fourier series representation of $x(t)$ is given by

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{j2\pi f_0 n t} \tag{1.45}$$

***1.3.2.3 Relationships Between the Transforms*** The (continuous) Fourier transform can be viewed as a limiting case of Fourier series as the period $T$ goes to $\infty$, and the signal thus becomes aperiodic. Since $f_0 = 1/T$, let $f = nf_0 = n/T$. Using (1.44), then

$$\begin{aligned}\lim_{T\to\infty} c_n T &= \lim_{T\to\infty} \int_{-T/2}^{T/2} x(t)e^{-j2\pi nt/T}\,dt \\ &= \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt \\ &= X(f) \end{aligned} \tag{1.46}$$

Since $1/T$ goes to zero in the limit, we can write $1/T$ as $\Delta f$. $\Delta f \to 0$ as $T \to \infty$. Then (1.45) can be written as

$$\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} T\frac{1}{T} c_n e^{j2\pi f_0 nt} \\ &= \sum_{n=-\infty}^{\infty} (c_n T) e^{j2\pi n f_0 t}\frac{1}{T} \\ &= \sum_{n=-\infty}^{\infty} (c_n T) e^{j2\pi n(\Delta f)t}\Delta f \end{aligned} \tag{1.47}$$

$$\lim_{\Delta f\to 0} x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df \tag{1.48}$$

where we used the substitution from (1.46) in the last step.

***1.3.2.4 Properties of the Fourier Transform*** Table 1.2 lists some useful properties of Fourier transforms. Combining properties from the table with known Fourier transform pairs from Table 1.1 lets us compute many Fourier transforms and inverse transforms without needing to perform the integrals (1.42) or (1.43).

**TABLE 1.2 Properties of the Fourier Transform**

| Concept | Time Domain, $x(t)$ | Frequency Domain, $X(f)$ |
|---|---|---|
| Scaling | $x(at)$ | $\frac{1}{\lvert a\rvert} X\left(\frac{f}{a}\right)$ |
| Time shifting | $x(t - t_0)$ | $X(f)e^{-j2\pi f t_0}$ |
| Frequency shifting | $x(t)e^{j2\pi f_0 t}$ | $X(f - f_0)$ |
| Modulation | $x(t)\cos(j2\pi f_0 t + \phi)$ | $\frac{1}{2}\left(X(f - f_0)e^{j\phi} + X(f + f_0)e^{-j\phi}\right)$ |
| Differentiation | $\frac{d^n x}{dt^n}$ | $(j2\pi f)^n X(f)$ |
| Convolution | $x(t) * y(t)$ | $X(f)Y(f)$ |
| Multiplication | $x(t)y(t)$ | $X(f) * Y(f)$ |
| Conjugation | $x^*(t)$ | $X^*(-f)$ |

### 1.3.3 Frequency-Domain Concepts

Some frequency-domain concepts are fundamental for understanding communications systems. A miscellany of comments on the frequency domain:

- In the rotating phasor viewpoint, $e^{j2\pi f_0 t}$ is a phasor rotating at $f_0$ cycles per cycle. But $\mathcal{F}[e^{j2\pi f_0 t}] = \delta(f - f_0)$. Thus, frequency-domain components of the form $\delta(f - f_0)$ for any $f_0$ can be viewed as rotating phasors.
- Negative frequencies can be viewed as rotating phasors rotating clockwise, whereas positive frequencies rotate counterclockwise.
- For LTI systems, $Y(f) = X(f)H(f)$, where $Y(f)$, $X(f)$, and $H(f)$ are the Fourier transforms of the output signal, input signal, and impulse response, respectively. See Figure 1.4.

***1.3.3.1 Power Spectral Density*** *Power spectral density* (PSD) is a way to see how the signal power is distributed in the frequency domain. We have seen that a periodic signal can be written in terms of Fourier series [as in (1.45)]. Similarly, the PSD $S_x(f)$ of periodic signals can be expressed in terms of Fourier series:

$$S_x(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} |c_n|^2 \delta\left(t - \frac{n}{T}\right) \tag{1.49}$$

where $c_n$ are the Fourier series coefficients as given by (1.44).

For nonperiodic power signals $x(t)$, let $x_T(t)$ be derived from $x(t)$ by

$$x_T(t) = x(t)\Pi(t/T) \tag{1.50}$$

Then $x_T(t)$ is an energy signal with a Fourier transform $X_T(f)$ and an energy spectral density $|X_T(f)|^2$. Then the power spectral density of $x(t)$ can be defined as

$$S_x(f) = \lim_{T\to\infty} \frac{1}{T} |X_T(f)|^2 \tag{1.51}$$

Alternatively, we can apply the *Wiener–Kinchine theorem*, which states that

$$S_x(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f\tau}\, d\tau \tag{1.52}$$

In other words, the PSD is simply the Fourier transform of the autocorrelation function. It can be shown that (1.51) and (1.52) and equivalent. Either one can be used to define the PSD and the other can be shown to be equivalent. Whereas (1.51) highlights the connection with the Fourier transform of the signal, (1.52) highlights the connection with its autocorrelation function.

Note that the Wiener–Kinchine theorem applies whether or not $x(t)$ is periodic. Thus, in the case that $x(t)$ is periodic with period $T$, clearly also $R_{xx}(\tau)$ is periodic with the same period. Let $R'_{xx}(t)$ be equal to $R_{xx}(t)$ within one period, $0 \le t \le T$, and zero elsewhere, and let $S'_x(f)$ be the power spectrum of $R'_{xx}(t)$. Note that

$$\begin{aligned} R_{xx}(t) &= \sum_{k=-\infty}^{\infty} R'_{xx}(t-kT) \\ &= \sum_{k=-\infty}^{\infty} R'_{xx}(t) * \delta(t-kT) \\ &= R'_{xx}(t) * \sum_{k=-\infty}^{\infty} \delta(t-kT) \end{aligned} \tag{1.53}$$

Then

$$\begin{aligned} S_x(f) &= \mathcal{F}(R_{xx}(\tau)) \\ &= \mathcal{F}\left(R'_{xx}(t) * \sum_{k=-\infty}^{\infty} \delta(t-kT)\right) \\ &= \mathcal{F}\left(R'_{xx}(t)\right) \mathcal{F}\left(\sum_{k=-\infty}^{\infty} \delta(t-kT)\right) \\ &= S'_x(f)\frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f-\frac{n}{T}\right) \end{aligned} \tag{1.54}$$

***One-Sided vs. Two-Sided PSD.*** The PSD that we have been discussing so far is the *two-sided PSD*, which has both positive and negative frequencies. It reflects the fact the a real sinusoid (e.g., a cosine wave) is the sum of two complex sinusoids rotating in opposite directions at the same frequency (thus, at a positive and a negative frequency). The *one-sided PSD* is a variation that has no negative frequency components and whose positive frequency components are exactly twice those of the two-sided PSD. The one-sided PSD is useful in some cases: for example, for calculations of noise power.

***1.3.3.2 Signal Bandwidth*** Just as in the time domain, we have a notion of duration of a signal (Section 1.2.4), in the frequency domain we have an analogous notion of *bandwidth*. A first-attempt definition of bandwidth might be the interval or range of frequencies from when the signal begins to be nonnegligible to when it stops being nonnegligible (as we sweep from lower to higher frequencies). This is imprecise but can be quantified in various ways, such as:

- *3-dB bandwidth* or *half-power bandwidth*
- *Noise-equivalent bandwidth* (see Section 3.2.3.2)

Often, it is not so much a question of finding the *correct* way of defining bandwidth but of finding a useful way of defining bandwidth for a particular situation.

Bandwidth is fundamentally related to channel capacity in the following celebrated formula:

$$C = B \log\left(1 + \frac{S}{N}\right) \tag{1.55}$$

The base of the logarithm determines the units of capacity. In particular, for capacity in bits/second,

$$C = B \log_2\left(1 + \frac{S}{N}\right) \tag{1.56}$$

To obtain capacity in bits/second, we use (1.56) with $B$ in hertz and $S/N$ on a linear scale (not decibels).

This concept of capacity is known as *Shannon capacity.* Later (e.g., in Section 6.3.2) we will see other concepts of capacity.

### 1.3.4 Bandpass Signals and Related Notions

Because bandpass signals have most of their spectral content around a carrier frequency, say $f_c$, they can be written in an envelope-and-phase representation:

$$x_b(t) = A(t)\cos[2\pi f_c t + \phi(t)] \tag{1.57}$$

where $A(t)$ and $\phi(t)$ are a slowly varying envelope and phase, respectively.

Most communications signals while in the communications medium are continuous-wave modulation signals, which tend to be *bandpass* in nature.

***1.3.4.1 In-phase/Quadrature Description*** A bandpass signal $x_b(t)$ can be written in envelope-and-phase form, as we have just seen. We can expand the cosine term using (A.8), and we have

$$\begin{aligned} x_b(t) &= A(t)\left[\cos(2\pi f_c t)\cos\phi(t) - \sin(2\pi f_c t)\sin\phi(t)\right] \\ &= x_i(t)\cos(2\pi f_c t) - x_q(t)\sin(2\pi f_c t) \end{aligned} \tag{1.58}$$

where $x_i(t) = A(t)\cos\phi(t)$ is the *in-phase* component, and $x_q(t) = A(t)\sin\phi(t)$ is the *quadrature* component. Later, in Section 6.1.8.1, we prove that the in-phase and quadrature components are orthogonal, so can be used to transmit independent bits without interfering with each other.

If we let $X_i(f) = \mathcal{F}[x_i(t)]$, $X_q(f) = \mathcal{F}[x_q(t)]$, and $X_b(f) = \mathcal{F}[x_b(t)]$, then

$$X_b(f) = \frac{1}{2}\left[X_i(f + f_c) + X_i(f - f_c)\right] - \frac{j}{2}\left[X_q(f + f_c) - X_q(f - f_c)\right] \tag{1.59}$$

***1.3.4.2 Lowpass Equivalents*** There is another useful representation of bandpass signals, known as the *lowpass equivalent* or *complex envelope* representation. Going from the envelope-and-phase representation to lowpass equivalent is analogous to going from a rotating phasor to a (nonrotating) phasor; thus we have

$$x_{\mathrm{lp}}(t) = A(t)e^{j\phi(t)} \tag{1.60}$$

which is analogous to (1.27). An alternative definition given in some other books is

$$x_{\mathrm{lp}}(t) = \frac{1}{2}A(t)e^{j\phi(t)} \tag{1.61}$$

which differs by a factor of 1/2. [This is just a matter of convention, and we will stick with (1.60).]

The lowpass equivalent signal is related to the in-phase and quadrature representation by

$$x_{\mathrm{lp}}(t) = x_i(t) + jx_q(t) \tag{1.62}$$

and we also have

$$x_b(t) = \Re\left[x_{\mathrm{lp}}(t)e^{j2\pi f_c t}\right] \tag{1.63}$$

In the frequency domain, the lowpass equivalent is the positive-frequency part of the bandpass signal, translated down to dc (zero frequency):

$$\begin{aligned} X_{\mathrm{lp}}(f) &= \left[X_i(f) + jX_q(f)\right] \\ &= 2X_b(f + f_c)u(f + f_c) \end{aligned} \tag{1.64}$$

where $u(f)$ is the step function (0 for $f < 0$, and 1 for $f \geq 0$).

Interestingly, we can represent filters or transfer functions with lowpass equivalents, too, so we have

$$Y_{\mathrm{lp}}(f) = H_{\mathrm{lp}}(f)X_{\mathrm{lp}}(f) \tag{1.65}$$

where

$$H_{\mathrm{lp}}(f) = H_b(f + f_c)u(f + f_c) \tag{1.66}$$

### 1.3.5 Random Signals

In well-designed communications systems, the signals arriving at a receiver appear random. Thus, it is important to have the tools to analyze random signals. We assume that the reader has knowledge of basic probability theory, including probability distribution or density, cumulative distribution function, and expectations [4].

Then a *random variable* can be defined as mapping from a sample space into a range of possible values. A sample space can be thought of as the set of all outcomes of an experiment. We denote the sample space by $\Omega$ and let $\omega$ be a variable that can represent each possible outcome in the sample space. For example, we consider a coin-flipping experiment with outcome either heads or tails, and we define a random variable by

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{heads} \\ 2 & \text{if } \omega = \text{tails} \end{cases} \tag{1.67}$$

where the domain of $\omega$ is the set {heads, tails}. If $P(\text{heads}) = 2/3$ and $P(\text{tails}) = 1/3$, then $P(X = 1) = 2/3$ and $P(X = 2) = 1/3$. The *average* (also called *mean*, or *expected value*) of $X$ is $(2/3)(1) + (1/3)(2) = 4/3$. Note that when we write just $X$, we have omitted the $\omega$ for notational simplicity.

***1.3.5.1 Stochastic Processes*** Now we consider cases where instead of just mapping each point in the sample space, $\omega$, to a value, we map each $\omega$ to a function. To emphasize that the mapping is to a function, and that this is therefore not the same as a normal random variable, it is called a *stochastic process* or *random process*. It could also be called a *random function*, but that could be confused with *random variable*, so it may be best to stick with *random variable* in general and *stochastic process* in cases where the mapping is to a function. Depending on the application, we may think of a stochastic process as a *random signal*.

For example, a stochastic process could be defined by a sinusoid with a random phase (e.g., a phase that is uniformly distributed between 0 and $2\pi$):

$$x(t, \omega) = \cos(2\pi ft + \phi) \tag{1.68}$$

where $\phi(\omega)$ is a random variable distributed uniformly between 0 and $2\pi$ (and where we usually omit writing the $\omega$, for convenience). Stochastic processes in wireless communications usually involve a time variable, $t$, and/or one or more spatial variables (e.g., $x$, $y$, $z$), so we can write $f(x, y, z, t, \omega)$ or just $f(x, y, z, t)$ if it is understood to represent a stochastic process.

The entire set of functions, as $\omega$ varies over the entire sample space, is called an *ensemble*. For any particular outcome, $\omega = \omega_i$, $x(t)$ is a specific *realization* (also known as *sample*) of the random process. For any given fixed $t = t_0$, $x(t_0)$ is a random variable, $X_0$, that represents the ensemble at that point in time (and hence a stochastic process can be viewed as an uncountably infinite set of random variables). Each of these random variables has a density function $f_{X_0}(x_0)$ from which its *first-order statistics* can be obtained. For example, we can obtain the mean $\int x f_{X_0}(x)\, dx$, the variance, and so on. The relationship between random variables associated with two different times $t_0$ and $t_1$ is often of interest. For example, let their joint distribution be written as $f_{X_0,X_1}(x_0, x_1)$; then, if

$$f_{X_0,X_1}(x_0, x_1) = f_{X_0}(x_0) f_{X_1}(x_1) \tag{1.69}$$

the two random variables are said to be *independent* or *uncorrelated*. The *second-order statistics* may be obtained from the joint distribution. This can be extended to the joint distribution of three or more points in time, so we have the *n*th-*order statistics*.

As an example of these ideas, assume that at a radio receiver we have a signal $r(t)$ that consists of a deterministic signal $s(t)$ in the presence of additive white Gaussian noise (AWGN), $n(t)$. If we model the AWGN in the usual way, $r(t)$ is a stochastic process:

$$r(t) = s(t) + n(t) \tag{1.70}$$

Because of the nature of AWGN, $n(t_1)$ and $n(t_2)$ are uncorrelated for any $t_1 \neq t_2$. Furthermore, since AWGN is Gaussian distributed, the first-order statistics depend on only two parameters (i.e., the mean and variance). Since $\overline{n(t)} = 0$ for all $t$, we just need to know the variance, $\sigma^2(t_1)$, $\sigma^2(t_2)$, and so on. Must we have $\sigma^2(t_1) = \sigma^2(t_2)$ for $t_1 \neq t_2$? We discuss this in Section 1.3.5.4. Here, we have just seen that a deterministic communications signal that is corrupted by AWGN can be modeled as a stochastic process.

***1.3.5.2 Time Averaging vs. Ensemble Averaging*** Averages are still useful for many applications, but since in this case we now have multiple variables over which an average may be taken, it often helps to specify to which average we are referring. If we are working with a specific realization of the random signal, we can take the *time average*. For a periodic signal (in time, $t$) with period $T_0$,

$$< x(t) >= \frac{1}{T_0} \int_0^{T_0} x(t)\, dt \tag{1.71}$$

If it is not a periodic signal, we may still consider a time average as given by

$$< x(t) >= \lim_{T \to \infty} \int_{-T/2}^{T/2} \frac{x(t)}{T}\, dt \tag{1.72}$$

Besides the time average, we also have the *ensemble average*, over the entire ensemble, resulting in a function (unlike the time average, which results in a value). For a discrete probability distribution this may be written as

$$\overline{x(t)} = \sum p_{x,t} x \tag{1.73}$$

where $p_{x,t}$ is the probability of event $x(t)$ at time $t$. The ensemble average for a continuous probability distribution can be written as

$$\overline{x(t)} = \int f_{X_t}(x) x\, dx \tag{1.74}$$

In this book we generally use $< \cdot >$ to denote time averaging or spatial averaging, and $\bar{\cdot}$ to denote ensemble averaging.

***1.3.5.3 Autocorrelation*** As we saw in Section 1.3.1.1, for deterministic signals the autocorrelation is a measure of the similarity of a signal with itself.

The autocorrelation function of a stochastic process $x(t)$ is

$$R_{xx}(t_1, t_2) = \overline{x(t_1)x(t_2)} \tag{1.75}$$

Unlike the case of deterministic signals, this is an ensemble average and in general is a function of two variables representing two moments of time rather than just a time difference. In general, it requires knowledge of the joint distribution of $x(t_1)$ and $x(t_2)$. Soon we will see that these differences go away when $x(t)$ is an ergodic process.

***1.3.5.4 Stationarity, Ergodicity, and Other Properties*** Going back to example (1.70), we saw that $n(t)$ was uncorrelated at any two different times. However, do the mean and variance have to be constant for all time? Clearly, they do not. In that radio receiver example, suppose that the temperature is rising. To make things simple, we suppose that the temperature is rising monotonically as $t$ increases. Then, as we will see in Section 3.2, Johnson–Nyquist noise in the receiver is increasing monotonically with time. Thus,

$$\sigma^2(t_1) < \sigma^2(t_2) \qquad \text{for } t_1 < t_2$$

If, instead,

$$\sigma^2(t_1) = \sigma^2(t_2) \qquad \text{for all } t_1 \neq t_2$$

there is a sense in which the stochastic process $n(t)$ is stationary—its variance doesn't depend on time.

The concept of stationarity has to do with questions of how the statistics of the signal change with time. For example, consider a random signal at $m$ time instances, $t_1, t_2, \ldots, t_m$. Suppose that we consider the joint distribution $f_{X_{t_1}, X_{t_2}, \ldots, X_{t_m}}(x_1, x_2, \ldots, x_m)$. Then a stochastic process is considered *strict-sense stationary* (SSS) if it is invariant to time translations for all sets $t_1, t_2, \ldots, t_m$, that is,

$$f_{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_m+\tau}}(x_1, x_2, \ldots, x_m) = f_{X_{t_1}, X_{t_2}, \ldots, X_{t_m}}(x_1, x_2, \ldots, x_m) \tag{1.76}$$

A weaker sense of stationarity is often seen in communications applications. A stochastic process is *weak-sense stationary* (WSS) if

1. The mean value is independent of time.
2. The autocorrelation depends only on the time difference $t_2 - t_1$ (i.e., it is a function of $\tau = t_2 - t_1$), so it may be written as $R_{xx}(\tau)$ [or $R_x(\tau)$ or simply $R(\tau)$] to keep this property explicit.

The class of WSS processes is larger than and includes the complete class of SSS processes. Similarly, there is another property, ergodicity, such that the class of SSS processes includes the complete class of ergodic processes. A random process is *ergodic* if it is SSS and if all ensemble averages are equal to the corresponding time averages. In other words, for ergodic processes, time averaging and ensemble averaging are equivalent.

*Autocorrelation Revisited.* For random processes that are WSS (including SSS and ergodic processes), the autocorrelation becomes $R(\tau)$, where $\tau$ is the time difference. Thus, (1.75) becomes

$$R_{xx}(\tau) = \overline{x(t)x(t+\tau)} \tag{1.77}$$

which is similar to (1.39).

Furthermore, for ergodic processes, we can even do a time average, so the autocorrelation then converges to the case of the autocorrelation of a deterministic signal (in the case of the ergodic process, we just pick any sample function and obtain the autocorrelation from it as though it were a deterministic function).

#### 1.3.5.5 Worked Example: Random Binary Signal

Consider a random binary wave, $x(t)$, where every symbol lasts for $T_s$ seconds, and independently of all other symbols, it takes the values $A$ or $-A$ with equal probability. Let the first symbol transition after $t = 0$ be at $T_{\text{trans}}$. Clearly, $0 < T_{\text{trans}} < T_s$. We let $T_{\text{trans}}$ be distributed uniformly between 0 and $T_s$.

The mean at any point in time $t$ is

$$E[x(t)] = A(0.5) + (-A)(0.5) = 0 \tag{1.78}$$

The variance at any point in time $t$ is

$$\sigma^2 = E[x^2(t)] - (E[x(t)])^2 = A^2 - 0 = A^2 \tag{1.79}$$

To figure out if it is WSS, we still need to see if the autocorrelation is dependent only on $\tau = t_2 - t_1$. We analyze the two autocorrelation cases:

- If $|t_2 - t_1| > T_s$, then $R_{xx}(t_1, t_2) = 0$ by the independence of each symbol from every other symbol.
- If $|t_2 - t_1| < T_s$, it depends on whether $t_1$ and $t_2$ lie in the same symbol (in which case we get $\sigma^2$) or in adjacent symbols (in which case we get zero).

What is the probability, $P_a$, that $t_1$ and $t_2$ lie in adjacent symbols? Let $t_1' = t_1 - kT_s$ and $t_2' = t_2 - kT_s$, where $k$ is the unique integer such that we get both $0 \le t_1' < T_s$ and $0 \le t_2' < T_s$. Then, $P_a = P(T_{\text{trans}}$ lies between $t_1'$ and $t_2') = |t_2 - t_1|/T_s$.

$$E[x(t_1)x(t_2)] = A^2(1 - P_a) = A^2\left(1 - \frac{|t_2 - t_1|}{T_s}\right) = A^2\left(1 - \frac{|\tau|}{T_s}\right) \tag{1.80}$$

Hence, it is WSS. And using the triangle function notation, we can write the complete autocorrelation function compactly as

$$R_{xx}(\tau) = A^2 \Lambda(\tau/T_s) \tag{1.81}$$

This is shown in Figure 1.5.

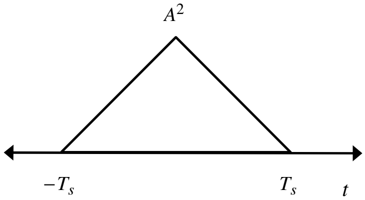


**FIGURE 1.5** Autocorrelation function of the random binary signal.

***1.3.5.6 Power Spectral Density of Random Signals*** For a random signal to have a meaningful power spectral density, it should be wide-sense stationary.

Each realization of the random signal would have its own power spectral density, different from other realizations of the same random process. It turns out that the (ensemble) average of the power spectral densities of each of the realizations, loosely speaking, is the most useful analog to the power spectral density of a deterministic signal. To be precise, the following procedure can be used on a random signal, $x(t)$, to estimate its PSD, $S_x(f)$. Let us denote the estimate by $\widetilde{S}_x(f)$.

1. Observe $x(t)$ over a period of time, say, 0 to $T$; let $x_T(t)$ be the truncated version of $x(t)$, as specified in (1.50), and let $X_T(f)$ be the Fourier transform of $x_T(t)$. Then its energy spectral density may be computed as $|X_T(f)|^2$.
2. Observe many samples $x_T(t)$ repeatedly, and compute their corresponding Fourier transforms $X_T(f)$ and energy spectral densities, $|X_T(f)|^2$.
3. Compute $\widetilde{S}_x(f)$ by computing the ensemble average $\overline{(1/T)\,|X_T(f)|^2}$.

One may wonder how to do step 2 in practice. Assuming that $x(t)$ is ergodic, then $\overline{(1/T)\,|X_T(f)|^2}$ is equivalent to time averaging, so we get a better and better estimate $\widetilde{S}_x(f)$ by obtaining $x_T(t)$ over many intervals of $T$ from the same sample function, and then computing

$$\widetilde{S}_x(f) = \left\langle \frac{1}{T}\,|X_T(f)|^2 \right\rangle \tag{1.82}$$

This procedure is based on the following definition of the PSD for random signals:

$$S_x(f) = \lim_{T\to\infty} \frac{1}{T}\,\overline{|X_T(f)|^2} \tag{1.83}$$

which is analogous to (1.51).

Also, as with deterministic signals, the Wiener–Kinchine theorem applies, so

$$S_x(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-j2\pi f\tau}\,d\tau \tag{1.84}$$

which can be shown to be equivalent to (1.83).

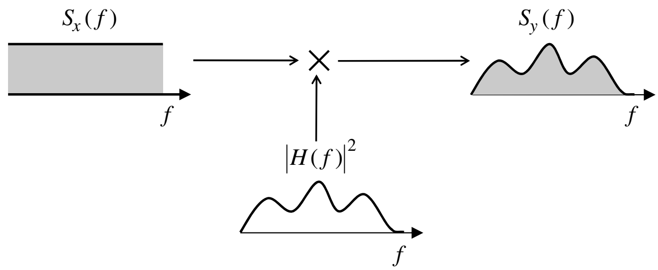


**FIGURE 1.6** Filtering and the PSD.

***1.3.5.7 Worked Example: PSD of a Random Binary Signal*** Consider the random binary signal from Section 1.3.5.5. What is the power spectral density of the signal? What happens as $T_s$ approaches zero?

We use the autocorrelation function, as in (1.81), and take the Fourier transform to obtain

$$S_x(f) = A^2 T_s \operatorname{sinc}^2(fT_s) \tag{1.85}$$

As $T_s$ gets smaller and smaller, the autocorrelation function approaches an impulse function. At the same time, the first lobe of the PSD is between $-1/T_s$ and $1/T_s$, so the it becomes very broad and flat, giving it the appearance of the classic "white noise."

***1.3.5.8 LTI Filtering of WSS Random Signals*** Once we can show that a random signal is WSS, the PSD behaves "like" the PSD of a deterministic signal in some ways; for example, when passing through a filter we have (Figure 1.6)

$$S_y(f) = |H(f)|^2 S_x(f) \tag{1.86}$$

where $S_x(f)$ and $S_y(f)$ are the PSDs of the input and output signals, respectively, and $H(f)$ is the LTI system/channel that filters the input signal.

For example, if $S_x(f)$ is flat (as with white noise), $S_y(f)$ takes on the shape of $H(f)$. In communications, a canonical signal might be a "random" signal around a carrier frequency $f_c$, with additive white Gaussian noise (AWGN) but with interfering signals at other frequencies, so we pass through a filter (e.g., an RF filter in an RF receiver) to reduce the magnitude of the interferers.

***1.3.5.9 Gaussian Processes*** A *Gaussian process* is one where the distribution $f_{X_t}(x)$ is Gaussian and all the distributions $f_{X_{t_1},X_{t_2},\ldots,X_{t_m}}(x_1, x_2, \ldots, x_m)$ for all sets $t_1, t_2, \ldots, t_m$ are joint Gaussian distributions.

For a Gaussian process, if it is WSS, it is also SSS.

***1.3.5.10 Optimal Detection in Receivers*** An important example of the use of random signals to model communications signals is the model of the signal received

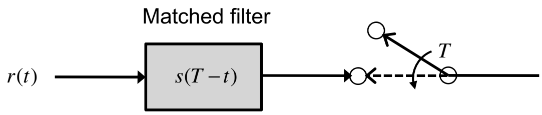


**FIGURE 1.7** Matched filter followed by symbol rate filtering.

at a digital communications receiver. We give examples of modulation schemes used in digital (and analog) systems in Section 1.4. But here we review some fundamental results on optimal detection.

*Matched Filters.* We consider the part of a demodulator after the frequency down-translation, such that the signal is at baseband. Here we have a *receiving filter* followed by a sampler, and we want to optimize the receiving filter. For the case of an AWGN channel, we can use facts about random signals [such as (1.86)] to prove that the optimal filter is the *matched filter*. By *optimal* we are referring to the ability of the filter to provide the largest signal-to-noise ratio at the output of the sampler at time $t = T$, where the signal waveform is from $t = 0$ to $T$.

**If the signal waveform is $s(t)$, the matched filter is $s(T - t)$ [or more generally, a scalar multiple of $s(T - t)$].**

The proof is outside the scope of this book but can be found in textbooks on digital communications. The matched filter is shown in Figure 1.7, where $r(t)$ is the received signal, and the sampling after the matched filtering is at the symbol rate, to decide each symbol transmitted.

*Correlation Receivers.* Also known as *correlators*, correlation receivers provide the same decision statistic that matched filters provide (Exercise 1.5 asks you to show this). If $r(t)$ is the received signal and the transmitted waveform is $s(t)$, the correlation receiver obtains

$$\int_0^T r(t)s(t)\,dt \tag{1.87}$$

## 1.4 SIGNALING IN COMMUNICATIONS SYSTEMS

Most communications systems use continuous-wave modulation as a fundamental building block. An exception is certain types of ultrawideband systems, discussed in Section 17.4.2. In continuous-wave modulation, a sinusoid is modified in certain ways to convey information. The unmodulated sinusoid is also known as the *carrier*. The earliest communications systems used analog modulation of the carrier.

These days, with source data so often in digital form (e.g., from a computer), it makes sense to communicate digitally also. Besides, digital communication has advantages over analog communication in how it allows error correction, encryption, and other processing to be performed. In dealing with noise and other channel impairments, digital signals can be recovered (with bit error rates on the order of $10^{-3}$ to $10^{-6}$, depending on the channel and system design), whereas analog signals are only degraded.

Generally, we would like digital communications with:

- Low bandwidth signals—so that it takes less "space" in the frequency spectrum, allowing more room for other signals
- Low-complexity devices—to reduce costs, power consumption, and so on.
- Low probability of errors

The trade-offs between these goals is the focus of much continuing research and development.

If we denote the carrier frequency by $f_c$ and the bandwidth of the signal by $B$, the design constraints of antennas and amplifiers are such that they work best if $B \ll f_c$, so this is usually what we find in communications systems. Furthermore, $f_c$ needs to be within the allocated frequency band(s) (as allocated by regulators such as Federal Communications Commission in the United States; see Section 17.4) for the particular communication system. The signals at these high frequencies are often called *RF* (radio-frequency) *signals* and must be handled with care with special RF circuits; this is called *RF engineering* (more on this in Chapter 3).

### 1.4.1 Analog Modulation

*Amplitude modulation* (AM) is given by

$$A_c(1 + \mu x(t)) \cos 2\pi f_c t \tag{1.88}$$

where the information signal $x(t)$ is normalized to $|x(t)| \leq 1$ and $\mu$ is the *modulation index*. To avoid signal distortion from *overmodulation*, $\mu$ is often set as $\mu < 1$. When $\mu < 1$, a simple *envelope detector* can be used to recover $x(t)$. AM is easy to detect, but has two drawbacks: (1) The unmodulated carrier portion of the signal, $A_c$, represents wasted power that doesn't convey the signal; and (2) Letting $B_b$ and $B_t$ be the baseband and transmitted bandwidths, respectively, then for AM, $B_t = 2B_b$, so there is wasted bandwidth in a sense. Schemes such as DSB and SSB attempt to reduce wasted power and/or wasted bandwidth.

*Double-sideband modulation* (DSB), also known as *double-sideband suppressed-carrier modulation* to contrast it with AM, is AM where the unmodulated carrier is not transmitted, so we just have

$$A_c x(t) \cos 2\pi f_c t \tag{1.89}$$

Although DSB is more power efficient than AM, simple envelope detection unfortunately cannot be used with DSB. As in AM, $B_t = 2B_b$.

*Single-sideband modulation* (SSB) achieves $B_t = B_b$ by removing either the upper or lower sideband of the transmitted signal. Like DSB, it suppresses the carrier to avoid wasting power. Denote the *Hilbert transform* of $x(t)$ by $\tilde{x}(t)$; then

$$\tilde{x}(t) = x(t) * \frac{1}{\pi t}$$

and we can write an SSB signal as

$$A_c \left[ x(t) \cos \omega_c t \pm \tilde{x}(t) \sin \omega_c t \right] \tag{1.90}$$

where the plus or minus sign depends on whether we want the lower sideband or upper sideband.

*Frequency modulation* (FM), unlike linear modulation schemes such as AM, is a nonlinear modulation scheme in which the frequency of the carrier is modulated by the message.

### 1.4.2 Digital Modulation

To transmit digital information, the basic modulation schemes transmit blocks of $k = \log_2 M$ bits at a time. Thus, there are $M = 2^k$ different finite-energy waveforms used to represent the $M$ possible combinations of the bits. Generally, we want these waveforms to be as "far apart" from each other as possible within certain energy constraints. The *symbol rate* or *signaling rate* is the rate at which new symbols are transmitted, and it is denoted $R$. The data rate is often denoted by $R_b$ bits/second (also written bps), and it is also called the *baud rate*. Clearly, $R_b = kR$. The symbol period $T_s$ is the inverse of the symbol rate, and is the time spent transmitting each symbol before it is time for the next symbol.

A bandlimited channel with bandwidth $B$ can support only up to the *Nyquist rate* of signaling, $R_{\text{Nyquist}} = 2B$. Thus, the signaling rate is constrained by

$$R \leq R_{\text{Nyquist}} = 2B \tag{1.91}$$

Digital modulation schemes, especially when the modulation is of the phase or frequency of the carrier, are often referred to as *shift keying* [e.g., amplitude shift keying (ASK), phase shift keying (PSK), and frequency shift keying (FSK)]. Use of the word *keying* in this context may have come from the concept of Morse code keys for telegraph but is useful for distinguishing digital modulation from analog modulation (e.g., FSK refers to a frequency-modulated digital signal, whereas FM refers to the traditional analog modulation signal that goes by that name). Nevertheless, the distinction is not always retained [e.g., a popular family of digital modulation schemes often goes by the name QAM (rather than QASK)].

***1.4.2.1 Pulse Shaping*** A digital modulator takes a simple continuous-time representation of our digital signal and outputs a continuous-time version of our signal, as will be seen in Section 1.4.2.2. How do we prepare our discrete-time digital data to enter a digital modulator? One way of converting from discrete time to continuous time is to let our data be represented by different baseband pulses for different values. For example, using a basic "rectangle" function, a 1 might be represented as

$p(t) = \pi(t/T_s)$ and a 0 by $-p(t) = -\pi(t/T_s)$ going into the digital modulator; this type of signaling, where one pulse is the exact negative of the other, is called *binary antipodal signaling*.

A problem with using a simple rectangle function in this way is that the spectral occupancy of the signals coming out of the digital modulator would be high—the Fourier transform of the rectangle function is the sinc function, which has relatively large spectral sidebands. Thus, it would be inefficient for use in bandwidth-critical systems such as wireless systems. Thus, it is important to use other *pulse-shaping functions*, $p(t)$, that can shape the spectral characteristics to use available spectrum more efficiently. However, not just any $p(t)$ can be used, because it also needs to be chosen to avoid adding *intersymbol interference* unnecessarily between nearby symbols. For example, if we (foolishly) used $p(t) = \pi(t/2T_s)$, every symbol would "spill over" into the preceding and/or subsequent symbol (in time) and interfere with them. There is a *Nyquist criterion* for $p(t)$ to avoid intersymbol interference that can be found in digital communications textbooks. Within the constraints of this criterion, the *raised cosine pulse*, illustrated in Figure 1.8, has emerged as a popular choice for $p(t)$. The frequency and time domains are shown in the subplots at the top and bottom of the figure, respectively. In the frequency domain we see the raised cosine shape from which the function gets its name. The *roll-off factor*, $\alpha$, is a parameter that determines how sudden or gradual the "roll-off" of the pulse is. In one extreme, $\alpha = 0$, we have a "brick wall" shape in frequency and the familiar sinc function in time (the light solid line on the plots). At the other extreme, $\alpha = 1$, we have the most roll-off, so, the bandwidth expands to twice as much as the $\alpha = 0$ case, as can be seen in the top subplot, with the thick solid line. The case of $\alpha = 0.5$ is also plotted in dashed lines in both subplots, and it can be seen to be between the two extremes. For smaller $\alpha$, the signal occupies less bandwidth, but the time sidelobes are higher, potentially resulting in more intersymbol interference and errors in practical receivers. For larger $\alpha$, the signal occupies more bandwidth but has smaller time sidelobes. In practice, to achieve the raised cosine transfer function, a matching pair of *square-root raised cosine filters* are used in the transmitter and receiver, since the receiver would have a matched filter (Section 1.3.5.10). The product of the two square-root raised cosine filters (in the frequency domain) gives the raised cosine shape at the output of the matched filter in the receiver.

#### 1.4.2.2 Digital Modulation Schemes

We show examples of common digital modulation schemes. We write examples of these waveforms using lowpass equivalent representation (Section 1.3.4.2) for convenience. In all cases, $p(t)$ is the *pulse-shaping function*.

*Pulse amplitude modulation* (PAM) uses waveforms of the form

$$A_m p(t) \qquad \text{for } m = 1, 2, \ldots, M \tag{1.92}$$

For optimal spacing, the $A_m$ are arranged in a line with equal spacing between consecutive points.

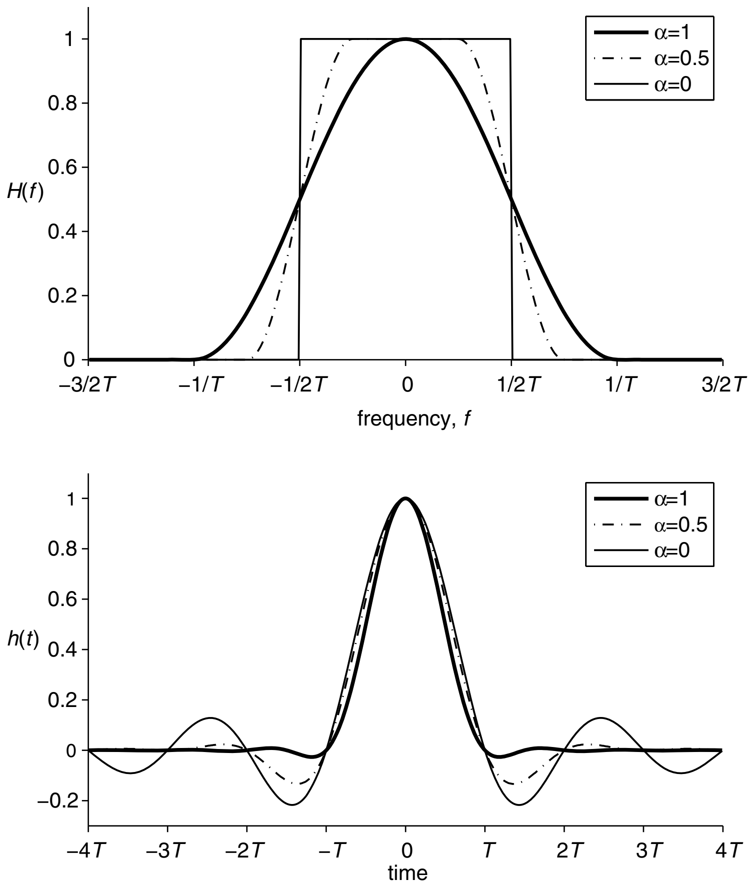


**FIGURE 1.8** Family of raised cosine pulses.

To conserve bandwidth, *SSB PAM* may be used:

$$A_m[p(t) + j\tilde{p}(t)] \qquad \text{for } m = 1, 2, \ldots, M \tag{1.93}$$

*Quadrature amplitude modulation* (QAM), where different bits are put in the in-phase ($A_{i,m}$) and quadrature ($A_{q,m}$) streams, can be written

$$(A_{i,m} + jA_{q,m})p(t) \qquad \text{for } m = 1, 2, \ldots, M/2 \tag{1.94}$$

Normally, wireless systems would use a form of QAM [e.g., 4-QAM (often just called QAM for short), 16-QAM, 32-QAM, 64-QAM] rather than PAM. Between QAM and PAM, QAM is more efficient because PAM does not exploit the quadrature dimension to transmit information. (For a review of the in-phase and quadrature concept, and to see why different bits can be put in in-phase and quadrature, refer to Sections 1.3.4.1 and 6.1.8.1.) The values $A_{i,m}$ and $A_{q,m}$ for $m = 1, 2, \ldots, M/2$ are chosen to be as

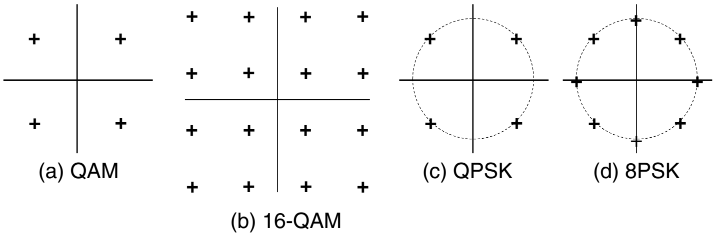


**FIGURE 1.9** Signal constellations for various digital modulation schemes.

far apart from one another (in signal space) as they can be, given an average power constraint. This is because the farther apart they are, the lower the bit error rates. Examples of 4-QAM and 16-QAM are shown in Figure 1.9.

*Phase shift keying* (PSK) uses waveforms of different phases to represent the different bit combinations:

$$e^{j\theta_m} p(t) \qquad \text{for } m = 1, 2, \ldots, M \tag{1.95}$$

*Binary PSK* (BPSK) is PSK with $m = 1$, *quadrature PSK* (QPSK) is PSK with $m = 2$, and *8-PSK* is PSK with $m = 3$. QPSK is very popular in wireless systems because it is more efficient than BPSK. 8-PSK is seen in EDGE (Section 8.1.3), for example. QPSK and 8-PSK are shown in Figure 1.9.

***1.4.2.3 Signal Constellations*** A good way to visualize the waveforms in a digital modulation scheme is through the *signal constellation* diagram. We have seen that the (lowpass equivalent of the) $M$ possible waveforms in general (except for modulation schemes like PAM) lie in the complex plane. We can therefore plot all the points in the complex plane, and the result is known as the *signal constellation*, some examples of which are shown in Figure 1.9. Notice that the signal constellation of 4-QAM happens to be the same as that of QPSK.

When we discuss wireless access technologies, we elaborate on selected aspects of digital modulation (Section 6.2), especially those having to do with design choices typically encountered in wireless systems.

### 1.4.3 Synchronization

In a digital receiver, two main types of synchronization are needed at the physical layer (there may also be other types of synchronization at higher layers, e.g., frame synchronization, multimedia synchronization, etc.):

- Carrier phase synchronization
- Symbol timing synchronization and recovery

*Carrier phase synchronization* is about figuring out, and recovering, a carrier signal frequency and phase. *Symbol timing synchronization and recovery* is about figuring out the locations (in time) of the temporal boundaries between symbols. It is also known as *clock recovery*.

## EXERCISES

**1.1** The form of the Fourier series given in Section 1.3.2 is the exponential form. Show how this is equivalent to the trigonometric form

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos 2\pi f_0 nt + b_n \sin 2\pi f_0 nt \tag{1.96}$$

Express $c_n$ in terms of $a_n$ and $b_n$.

**1.2** Instead of the random binary waveform we saw in Section 1.3.5.5, we have a random digital waveform. So it takes not just two values, 1 and −1, but a range of values over a distribution: say, a Gaussian distribution with mean 0 and variance $\sigma^2$. Find the autocorrelation function of the random digital waveform. How does it compare with the autocorrelation function of the random binary waveform given by (1.81)?

**1.3** Suppose we have a signal $x(t)$ that is multiplied by a sinusoid, resulting in the signal $y(t) = x(t)\cos 2\pi ft$. Assume that $x(t)$ is independent of the sinusoid but could otherwise be a (deterministic or random) signal with autocorrelation function $R_{xx}(\tau)$. Show that the autocorrelation of $y(t)$ is given by

$$R_{yy}(\tau) = R_{xx}(\tau)\left(\frac{1}{2}\cos 2\pi f t\right) \tag{1.97}$$

**1.4** Continuing from Exercise 1.3, what is the effect on the power spectral density of multiplication by a sinusoid? In other words, express the power spectral density of $y(t)$ in terms of the power spectral density of $x(t)$. This is a fundamental and useful result, since it means that we can up-convert and down-convert signals to and from carrier frequencies, and the autocorrelation function and power spectral density behave in this predictable way.

**1.5** Show that a matched filter followed by sampling at $t = T$ produces the same output as a correlation receiver.

## REFERENCES

1. A. B. Carlson. *Communication Systems*, 3rd ed. McGraw-Hill, New York, 1986.
2. L. Couch. *Digital and Analog Communication Systems*, 7th ed. Prentice Hall, Upper Saddle River, NJ, 2007.

3. J. W. Nilsson. *Electric Circuits*, 3rd ed. Addison-Wesley, Reading MA, 1990.
4. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
5. B. Sklar. *Digital Communications: Fundamentals and Applications*, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 2001.