Chapter 3

Sampling and Reconstruction

Signals encountered in real-life applications are usually in continuous time. To facilitate digital processing, a continuous-time signal must be converted to a sequence of numbers. The process of converting a continuous-time signal to a sequence of numbers is called *sampling*. A motion picture is a familiar example of sampling. In a motion picture, a continuously varying scene is converted by the camera to a sequence of frames. The frames are taken at regular time intervals, typically 24 per second. We then say that the scene is sampled at 24 frames per second. Sampling is essentially a selection of a finite number of data at any finite time interval as representatives of the infinite amount of data contained in the continuous-time signal in that interval. In the motion picture example, the frames taken at each second are representatives of the continuously varying scene during that second.

When we watch a motion picture, our eyes and brain fill the gaps between the frames and give the illusion of a continuous motion. The operation of filling the gaps in the sampled data is called *reconstruction*. In general, reconstruction is the operation of converting a sampled signal back to a continuous-time signal. Reconstruction provides an infinite (and continuously varying) number of data at any given time interval out of the finite number of data in the sampled signal. In the motion picture example, the reconstructed continuous-time scene exists only in our brain.

Naturally, one would not expect the reconstructed signal to be absolutely faithful to the original signal. Indeed, sampling leads to distortions in general. The fundamental distortion introduced by sampling is called *aliasing*. Aliasing in motion pictures is a familiar phenomenon. Suppose the scene contains a clockwise-rotating wheel. As long as the speed of rotation is lower than half the number of frames per second, our brain perceives the correct speed of rotation. When the speed increases beyond this value, the wheel appears to rotate *counterclockwise* at a reduced speed. Its apparent speed is now the number of frames per second minus its true speed. When the speed is equal to the number of frames per second, it appears to stop rotating. This happens because all frames now sample the wheel in an identical position. When the speed increases further, the wheel appears to rotate clockwise again, but at a reduced speed. In general, the wheel always appears to rotate at a speed not higher than half the number of frames per second, either clockwise or counterclockwise.

Sampling is *the* fundamental operation of digital signal processing, and avoiding (or at least minimizing) aliasing is the most important aspect of sampling. Thorough understanding of sampling is necessary for any practical application of digital signal processing. In most engineering applications, the continuous-time signal is given in an electrical form (i.e., as a voltage waveform), so sampling is performed by an electronic circuit (and the same is true for reconstruction). Physical properties (and limitations) of electronic circuitry lead to further distortions of the sampled signal, and these need to be thoroughly understood as well.

In this chapter we study the mathematical theory of sampling and its practical aspects. We first define sampling in mathematical terms and derive the fundamental result of sampling theory: the sampling theorem of Nyquist, Whittaker, and Shannon. We examine the consequences of the sampling theorem for signals with finite and infinite bandwidths. We then deal with reconstruction of signals from their sampled values. Next we consider physical implementation of sampling and reconstruction, and explain the deviations from ideal behavior due to hardware limitations. Finally, we discuss several special topics related to sampling and reconstruction.

3.1 Two Points of View on Sampling

Let $x(t)$ be a continuous function on the real line. Sampling of the function amounts to picking its values at certain time points. In particular, if the sampling points are nT, $n \in \mathbb{Z}$, it is called *uniform sampling*, and T is called the *sampling interval*. The numbers

$$f_{\text{sam}} = \frac{1}{T}, \quad \omega_{\text{sam}} = \frac{2\pi}{T}$$

are called the *sampling frequency* (or *sampling rate*) and *angular sampling frequency*, respectively. The sampled function is then the sequence

$$x[n] = x(nT), \quad n \in \mathbb{Z}. \tag{3.1}$$

This definition can be extended to discontinuous functions, provided the discontinuities are isolated (no more than a finite number of them on any finite interval), and the limits at each discontinuity point exist from both left and right. In this case it is common to define

$$x[n] = 0.5[x(nT^-) + x(nT^+)] \tag{3.2}$$

if nT is a discontinuity point.

An alternative description of the sampling operation is as follows. Recall the impulse train $p_T(t)$ defined in (2.47) and let

$$x_p(t) = x(t)p_T(t) = \sum_{n=-\infty}^{\infty} x(t)\delta(t-nT) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t-nT). \tag{3.3}$$

Then $x_p(t)$ is, formally, a continuous-time signal, since it is defined for all t. However, it is clear that the information it conveys about $x(t)$ is limited to the values $x(nT)$, $n \in \mathbb{Z}$, since $x_p(t)$ is identically zero at all other points.

We thus have two ways to look at the sampled signal:

1. To consider it as a sequence of numbers $x[n] = x(nT)$, $n \in \mathbb{Z}$, or as a discrete-time signal. We refer to $x[n]$ as *point sampling* of $x(t)$.
2. To consider it as a continuous-time signal $x_p(t)$. We refer to $x_p(t)$ as *impulse sampling* of $x(t)$.

Figure 3.1 illustrates the sampling operation from the two points of view. Part a shows a continuous-time signal, with the sampling points emphasized. Part b shows the discrete-time signal $x[n]$ obtained by point sampling. Part c shows the continuous-time signal $x_p(t)$ obtained by impulse sampling.

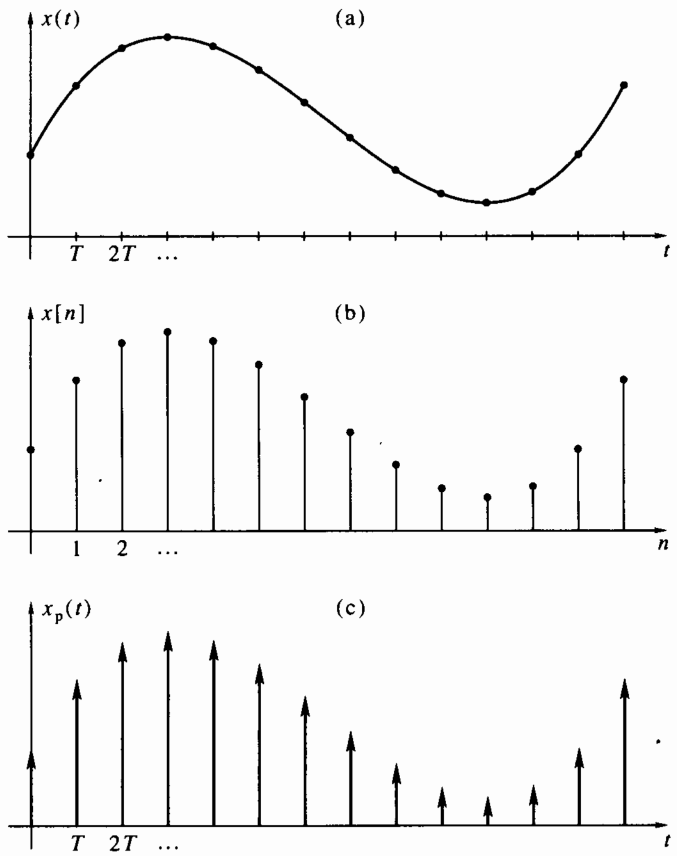

Figure 3.1 The sampling operation: (a) the continuous-time signal $x(t)$; (b) the point-sampled sequence $x[n]$; (c) the impulse-sampled signal $x_p(t)$.

The two points of view on sampling are equivalent. Physically, only point sampling is possible (and even this is an idealization of real-life sampling). Impulse sampling is convenient for mathematical derivations, since known results from continuous-time signal analysis can be used.

Example 3.1 Here are a few examples of sampled signals.

1. Let $x(t)$ be the sinusoidal signal

$$x(t) = A\cos(\omega_0 t + \phi_0).$$

Then,

$$x[n] = A\cos(\omega_0 Tn + \phi_0).$$

The sequence $x[n]$ is a *discrete-time sinusoid*, with a discrete angular frequency $\omega_0 T$. The discrete frequency is measured in radians (or radians per sample, but not radians per second).

2. Let $x(t)$ be the one-sided exponential function

$$x(t) = \begin{cases} 0, & t < 0, \\ Ae^{-\alpha t}, & t \geq 0, \end{cases} \tag{3.4}$$

where $\alpha \geq 0$. Then,

$$x[n] = \begin{cases} 0, & n < 0, \\ Ae^{-\alpha T n}, & n \geq 0. \end{cases} \tag{3.5}$$

We emphasize that, in this case, it is common to define $x[0]$ as $x(0^+)$, rather than as the midpoint between the two discontinuity limits. The signal $x[n]$ is a geometric series, with parameter $e^{-\alpha T}$.

3. A special case of a one-sided exponential occurs when $\alpha = 0$, $A = 1$. The continuous-time and discrete-time signals are then

$$v(t) = \begin{cases} 0, & t < 0, \\ 1, & t \geq 0, \end{cases} \quad \text{and} \quad v[n] = \begin{cases} 0, & n < 0, \\ 1, & n \geq 0. \end{cases} \tag{3.6}$$

These are the continuous-time and discrete-time *unit-step* signals, respectively.

4. The continuous-time delta function $\delta(t)$ *cannot be sampled*, since it does not have a finite value at $t = 0$. It is a gross mistake to assume that its sampled version is the discrete-time delta $\delta[n]$.

5. Let $x(t) = \text{rect}(t/T_0)$. Then the sampled signal is

$$x[n] = \begin{cases} 1, & |n| \leq \left\lfloor \frac{0.5T_0}{T} \right\rfloor, \\ 0, & \text{otherwise.} \end{cases}$$

In particular, if $T_0 < 2T$, then the sampled signal is $\delta[n]$. □

3.2 The Sampling Theorem

Suppose we are given a continuous-time signal $x(t)$ whose Fourier transform is $X^{\text{F}}(\omega)$. Then the impulse-sampled signal $x_{\text{p}}(t)$ has Fourier transform $X_{\text{p}}^{\text{F}}(\omega)$, and the point-sampled signal $x[n]$ has Fourier transform $X^{\text{f}}(\theta)$, defined as

$$X_{\text{p}}^{\text{F}}(\omega) = \int_{-\infty}^{\infty} x_{\text{p}}(t)e^{-j\omega t}dt, \quad X^{\text{f}}(\theta) = \sum_{n=-\infty}^{\infty} x(nT)e^{-j\theta n}. \tag{3.7}$$

The questions we address in this section are (1) How are $X_{\text{p}}^{\text{F}}(\omega)$ and $X^{\text{f}}(\theta)$ related to each other, and (2) how are they related to $X^{\text{F}}(\omega)$? The answers would enable us to understand the consequences of sampling, to assess the nature and amount of distortion introduced by sampling, and to tell under what conditions it is possible to completely avoid distortions due to sampling.

The relationships among the three Fourier transforms are given by the famous *sampling theorem*.[1]

Theorem 3.1 The Fourier transforms $X^{\text{F}}(\omega)$, $X_{\text{p}}^{\text{F}}(\omega)$, and $X^{\text{f}}(\theta)$ satisfy the following relationships:

$$X_{\text{p}}^{\text{F}}(\omega) = \sum_{n=-\infty}^{\infty} x(nT)e^{-jn\omega T} \tag{3.8}$$

$$= \frac{1}{T}\sum_{k=-\infty}^{\infty} X^{\text{F}}\left(\omega - \frac{2\pi k}{T}\right), \tag{3.9}$$

$$X^{\text{f}}(\theta) = X_{\text{p}}^{\text{F}}\left(\frac{\theta}{T}\right) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X^{\text{F}}\left(\frac{\theta - 2\pi k}{T}\right). \tag{3.10}$$

Proof To prove (3.8), substitute the definition (3.3) of $x_{\mathrm{p}}(t)$ in the definition of Fourier transform:

$$X_{\mathrm{p}}^{\mathrm{F}}(\omega) = \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT) \right] e^{-j\omega t} dt$$

$$= \sum_{n=-\infty}^{\infty} x(nT) \int_{-\infty}^{\infty} \delta(t - nT) e^{-j\omega t} dt = \sum_{n=-\infty}^{\infty} x(nT) e^{-jn\omega T}. \tag{3.11}$$

To prove (3.9), recall that (1) a product in the time domain translates to a convolution in the frequency domain; (2) the Fourier transform of an impulse train is an impulse train in the frequency domain; see (2.52). Therefore,

$$X_{\mathrm{p}}^{\mathrm{F}}(\omega) = \frac{1}{2\pi}\{P_T^{\mathrm{F}} * X^{\mathrm{F}}\}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X^{\mathrm{F}}(\lambda) \left[\frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega - \lambda - \frac{2\pi k}{T}\right) \right] d\lambda$$

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}\left(\omega - \frac{2\pi k}{T}\right). \tag{3.12}$$

Finally, (3.10) follows when substituting $\theta = \omega T$ in (3.8), (3.9).

An alternative proof of (3.9) The following proof does not rely on the properties of the impulse train:

$$x(nT) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X^{\mathrm{F}}(\omega) e^{j\omega nT} d\omega = \sum_{k=-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-(2k+1)\pi/T}^{-(2k-1)\pi/T} X^{\mathrm{F}}(\omega) e^{j\omega nT} d\omega \right]$$

$$= \sum_{k=-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} X^{\mathrm{F}}\left(\omega - \frac{2\pi k}{T}\right) e^{j\omega nT} d\omega \right]$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{1}{T} \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}\left(\frac{\theta - 2\pi k}{T}\right) \right] e^{j\theta n} d\theta. \tag{3.13}$$

In passing from the first line of (3.13) to the second, we used the property

$$e^{-j2\pi nk} = 1, \quad n, k \in \mathbb{Z}.$$

Equation (3.13) gives, by the definition of the Fourier transform (2.93) and its inverse formula (2.95),

$$\sum_{n=-\infty}^{\infty} x(nT) e^{-j\omega nT} = \frac{1}{T} \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}\left(\omega - \frac{2\pi k}{T}\right). \tag{3.14}$$

This is the same as (3.9). □

The sampling theorem tells us that the Fourier transform of a discrete-time signal obtained from a continuous-time signal by sampling is related to the Fourier transform of the continuous-time signal by three operations:

1. Transformation of the frequency axis according to the relation $\theta = \omega T$.
2. Multiplication of the amplitude axis by a factor $1/T$.
3. Summation of an infinite number of replicas of the given spectrum, shifted horizontally by integer multiples of the angular sampling frequency ω_{sam}.

As a result of the infinite summation, the Fourier transform of the sampled signal is periodic in θ with period 2π. We therefore say that *sampling in the time domain gives rise to periodicity in the frequency domain.*

Example 3.2 Consider the signal $x(t) = te^{-\alpha t}v(t)$, where $\alpha > 0$. Its Fourier transform is

$$X^{\mathrm{F}}(\omega) = \int_0^\infty te^{-(\alpha+j\omega)t}dt = \frac{1}{(\alpha+j\omega)^2}. \tag{3.15}$$

The sampled signal is $x[n] = nTe^{-\alpha Tn}v[n]$, and its Fourier transform is

$$X^{\mathrm{f}}(\theta) = \sum_{n=0}^{\infty} nTe^{-(\alpha T+j\theta)n} = \frac{Te^{-(\alpha T+j\theta)}}{[1-e^{-(\alpha T+j\theta)}]^2}. \tag{3.16}$$

The sampling theorem then leads us to conclude that

$$\sum_{k=-\infty}^{\infty} \frac{1}{[\alpha T + j(\theta - 2\pi k)]^2} = \frac{e^{-(\alpha T+j\theta)}}{[1-e^{-(\alpha T+j\theta)}]^2}. \tag{3.17}$$

This infinite sum is not straightforward to prove directly, but it follows easily from sampling theory; see Problem 3.6 for another example. □

3.3 The Three Cases of Sampling

A continuous-time signal $x(t)$ is called *band limited* if its Fourier transform vanishes outside a certain frequency range, that is, if there exists $\omega_{\mathrm{m}} > 0$ such that

$$X^{\mathrm{F}}(\omega) = 0 \text{ for } |\omega| \geq \omega_{\mathrm{m}}. \tag{3.18}$$

The angular frequency ω_{m} is called the *bandwidth* of the signal. (When we learn about band-pass signals, in Section 3.6, we shall modify this definition.) Sometimes we shall say that the bandwidth is $\pm\omega_{\mathrm{m}}$, to emphasize that the Fourier transform is nonzero for both positive and negative frequencies. The definition can be extended to allow $X^{\mathrm{F}}(\pm\omega_{\mathrm{m}}) \neq 0$, provided there are no delta functions at $\omega = \pm\omega_{\mathrm{m}}$. The interval $[-\omega_{\mathrm{m}}, \omega_{\mathrm{m}}]$ is called the *frequency support* of the signal. For a signal that is not band limited, the frequency support is $(-\infty, \infty)$. Bandwidth thus defined is measured in radians per second. Expressed in hertz, the bandwidth is

$$f_{\mathrm{m}} = \frac{\omega_{\mathrm{m}}}{2\pi}.$$

Figure 3.2(a) illustrates the Fourier transform of a band-limited signal. Figure 3.2(b) shows, for comparison, the Fourier transform of a signal that is not band limited.

For now, we regard the property of being band limited as purely mathematical and defer the question of physical existence of such signals until later.

Example 3.3 The Fourier transform of the function $\mathrm{sinc}(t)$ is $\mathrm{rect}(\omega/2\pi)$, compare (2.35). The Fourier transform is nonzero only on the interval $|\omega| \leq \pi$, hence $\mathrm{sinc}(t)$ is a band-limited signal, with $\omega_{\mathrm{m}} = \pi$. On the other hand, the signal $x(t)$ in Example 3.2 is not band limited, because, as we see from (3.15), its Fourier transform does not vanish for any ω. □

Suppose we sample a band-limited signal $x(t)$ and we choose the sampling frequency such that $f_{\mathrm{sam}} \geq 2f_{\mathrm{m}}$. The spectra of the continuous-time signal and the sampled signal are shown in Figure 3.3. *The replicas do not overlap in this case.* In particular, the Fourier transform of the sampled signal in the range $\theta \in [-\pi, \pi]$ is given by [cf. (3.10)]

$$X^{\mathrm{f}}(\theta) = \frac{1}{T}X^{\mathrm{F}}\left(\frac{\theta}{T}\right), \quad -\pi \leq \theta \leq \pi. \tag{3.19}$$

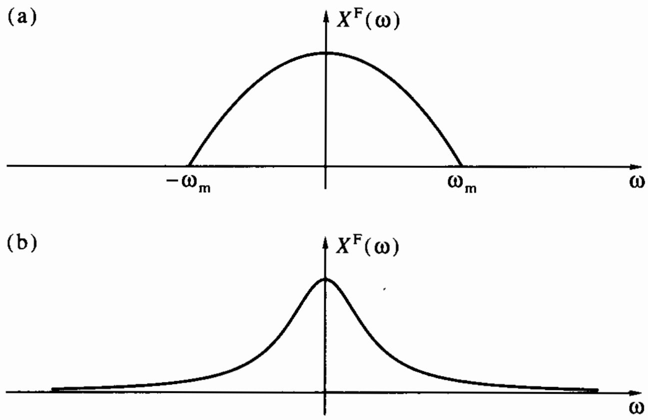

Figure 3.2 Fourier transform: (a) of a band-limited signal; (b) of a signal that is not band limited.

The conclusion is that the Fourier transform of the sampled signal in the principal frequency range $[-\pi, \pi]$ preserves the shape of the Fourier transform of the given signal, except for multiplication of the frequency and amplitude axes by constant factors. The shape at frequencies outside the interval $[-\pi, \pi]$ is obtained by periodic extension, as is always true for Fourier transforms of discrete-time signals.

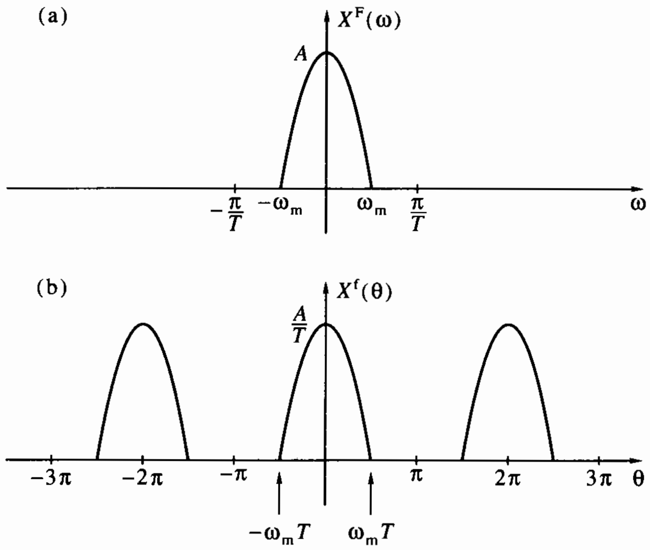

Figure 3.3 Sampling of a band-limited signal above the Nyquist rate: (a) Fourier transform of the continuous-time signal; (b) Fourier transform of the sampled signal.

The lowest sampling rate for which (3.19) holds is

$$f_{\text{sam}} = \frac{1}{T} = 2f_{\text{m}}.$$

This is called the *Nyquist rate*, or the *critical rate* for sampling of a band-limited signal. The first case of sampling can therefore be summarized as follows:

> If a band-limited signal is sampled at a rate equal to or greater than the Nyquist rate $2f_m$, the shape of the Fourier transform of the sampled signal in the range $\theta \in [-\pi, \pi]$ is identical to the shape of the Fourier transform of the given signal, except for multiplication of the frequency axis by a factor T, and multiplication of the amplitude axis by a factor $1/T$.

The fundamental relationship between continuous-time and discrete-time frequencies

$$\boxed{\theta = \omega T = 2\pi f T}$$

is of extreme importance, and should be borne in mind at all times. The quantities ω, f measure the number of radians or cycles, respectively, per second. The quantity θ measures the number of radians per sample. Since each sample extends over T seconds, we get the relationship shown in the box.

Since the Fourier transform of a band-limited signal is not distorted if the sampling rate meets the Nyquist condition, we expect to be able to reconstruct the continuous-time signal $x(t)$ from it samples $x(nT)$ in this case. Reconstruction is discussed in Section 3.4.

Example 3.4 Consider the signal

$$x(t) = \text{sinc}(f_0 t).$$

It follows from (2.35) and (2.7) that

$$X^{\text{F}}(\omega) = \frac{1}{f_0}\text{rect}\left(\frac{\omega}{2\pi f_0}\right).$$

Therefore, the Nyquist rate is $\omega_{\text{sam}} = 2\pi f_0$, or $f_{\text{sam}} = f_0$. The sampled signal is then

$$x[n] = \text{sinc}(n) = \delta[n].$$

The Fourier transform of the sampled signal is

$$X^{\text{f}}(\theta) = 1.$$

This result is perhaps intriguing: Surely, there are many continuous-time signals whose sampling at $T = 1/f_0$ would give the unit-sample signal. It follows from this result that, of all such signals, the sinc is the only one whose bandwidth is limited to $0.5f_0$ Hz, and no such signal can have a smaller bandwidth. This result has profound consequences in digital communication theory. A signal that, when sampled at an interval T, gives a unit impulse, is called a *Nyquist-T signal.*[2] We conclude that a Nyquist-T signal must have a bandwidth not smaller than $(0.5/T)$ Hz, and only the sinc signal achieves this lower bound. □

Example 3.5 Consider the signal

$$y(t) = \text{sinc}^2(f_0 t).$$

We recall that multiplication in the time domain translates to convolution in the frequency domain. The convolution of two identical rect functions is a triangular function; therefore,

$$Y^{\text{F}}(\omega) = \frac{1}{2\pi}\{X^{\text{F}} * X^{\text{F}}\}(\omega) = \begin{cases} \frac{1}{f_0}\left(1 - \frac{|\omega|}{2\pi f_0}\right), & |\omega| \le 2\pi f_0, \\ 0, & |\omega| > 2\pi f_0. \end{cases}$$

The Nyquist rate for $y(t)$ is $\omega_{\text{sam}} = 4\pi f_0$, or $f_{\text{sam}} = 2f_0$. The sampled signal is

$$y[n] = \text{sinc}^2(0.5n).$$

The Fourier transform of the sampled signal is

$$Y^{\text{f}}(\theta) = 2\left(1 - \frac{|\theta|}{\pi}\right), \quad -\pi \le \theta \le \pi.$$

□

We now consider the case of a band-limited signal sampled at a rate *lower* than the Nyquist rate. Figure 3.4 shows what happens in this case, using a sampling rate $f_{\text{sam}} = 3f_{\text{m}}/2$ as an example. Now the shape of the Fourier transform in the range $\theta \in [-2\pi/3, 2\pi/3]$ is preserved, but the shape in the range $|\theta| \in (2\pi/3, \pi]$ is distorted. Distortion occurs because, in this frequency range, two adjacent replicas overlap and their superposition [as expressed in (3.10)] gives rise to the shape shown in the figure. An alternative way to describe this phenomenon is this: The high-frequency contents of the continuous-time signal (in the range $|\omega| \in (3\omega_{\text{m}}/4, \omega_{\text{m}}])$ disguises itself as a low-frequency contents (in the range $|\omega| \in [\omega_{\text{m}}/2, 3\omega_{\text{m}}/4))$ and is added to the original contents in this range.

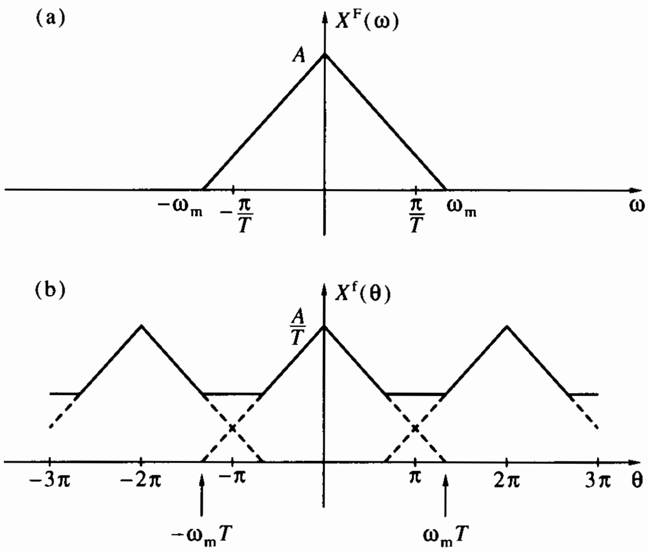

Figure 3.4 Sampling of a band-limited signal below the Nyquist rate: (a) Fourier transform of the continuous-time signal; (b) Fourier transform of the sampled signal.

The phenomenon we have described is called *aliasing*. Aliasing is caused by sampling at a rate lower than the Nyquist rate for the given signal. Since the shape of the Fourier transform in the range $\theta \in [-\pi, \pi]$ is not preserved, we expect to be unable to reconstruct the continuous-time signal from its samples. In conclusion:

> If a band-limited signal is sampled at a rate lower than the Nyquist rate $2f_{\text{m}}$, the shape of the Fourier transform of the sampled signal in the range $\theta \in [-\pi, \pi]$ is distorted relative to the Fourier transform of the given signal. This distortion, which is called aliasing, results from overlapping of the replicas in the sampling formula (3.10).

Example 3.6 Let

$$x_1(t) = \cos(1.2\pi t), \quad x_2(t) = \cos(0.8\pi t).$$

Suppose the two signals are sampled at interval $T = 1$ second. Then

$$x_1(nT) = \cos(1.2\pi n) = \cos(0.8\pi n) = x_2(nT).$$

Therefore, the signals $x_1(t)$ and $x_2(t)$ become indistinguishable when sampled at interval $T = 1$ second. This is illustrated in Figure 3.5. We say that $x_1(t)$ is *aliased* as $x_2(t)$. □

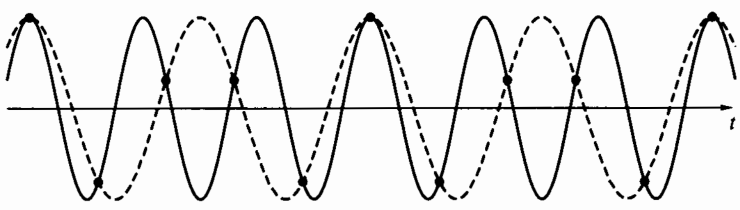

Figure 3.5 Aliasing of a sinusoidal signal. Solid line: signal at 0.6 Hz; dashed line: signal at 0.4 Hz; circles: samples at $T = 1$ second.

Example 3.7 Consider the signal

$$x(t) = a_1 \cos(0.5\pi t) + a_2 \cos(1.75\pi t).$$

Let the sampling interval be $T = 1$ second. The sampled signal is then

$$x[n] = a_1 \cos(0.5\pi n) + a_2 \cos(1.75\pi n) = a_1 \cos(0.5\pi n) + a_2 \cos(0.25\pi n),$$

since $\cos(1.75\pi n) = \cos(0.25\pi n)$ for all n. As we see, the frequency of the first component of the sampled signal is $\theta = 0.5\pi$, which is related to the corresponding frequency of $x(t)$ by the correct relationship $\theta = \omega T$. Since, however, the second component becomes a low frequency signal, it is aliased. □

Example 3.8 Consider the borderline case of sampling the signal

$$x(t) = \cos(2\pi f_0 t - \phi_0)$$

at a rate $f_{\text{sam}} = 2f_0$. Will it be subject to aliasing? If we take $\phi_0 = 0$, we will get

$$x[n] = \cos(\pi n) = (-1)^n,$$

which seems to be alias free. However, if we take $\phi_0 = \pi/2$, we will get

$$x[n] = \sin(\pi n) = 0,$$

so the sampled signal vanishes completely! The answer is therefore that the choice $f_{\text{sam}} = 2f_0$ is *not* permitted in this case. Another argument for this conclusion is: Since the Fourier transform of $x(t)$ contains delta functions at $\omega = \pm 2\pi f_0$, sampling at $f_{\text{sam}} = 2f_0$ violates Nyquist's no-aliasing condition. □

Example 3.9 Consider the signal $x(t)$ whose Fourier transform is

$$X^{\mathrm{F}}(\omega) = \begin{cases} \dfrac{1}{f_0}, & |\omega| \le \omega_1, \\ \dfrac{0.5}{f_0}\left[1 + \cos\left(\dfrac{|\omega| - \omega_1}{2\alpha f_0}\right)\right], & \omega_1 < |\omega| \le \omega_2, \\ 0, & |\omega| > \omega_2, \end{cases} \tag{3.20}$$

where $0 < \alpha < 1$, and

$$\omega_1 = (1-\alpha)\pi f_0, \quad \omega_2 = (1+\alpha)\pi f_0. \tag{3.21}$$

This signal is band limited, its bandwidth being $(1+\alpha)\pi f_0$, which is more than πf_0, but less than $2\pi f_0$. The shape of $X^{\mathrm{F}}(\omega)$ is shown in Figure 3.6(a). This is called an *excess bandwidth, raised-cosine spectrum*, and α is called the bandwidth excess. For example, when $\alpha = 0.4$, we call it a 40 percent raised-cosine spectrum.

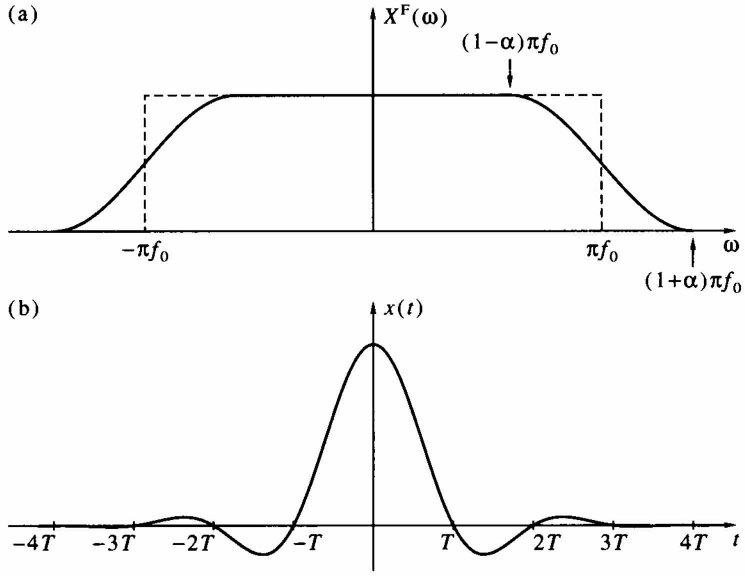

Figure 3.6 A raised-cosine signal: (a) the spectrum; (b) the waveform.

The inverse Fourier transform of $X^{\mathrm{F}}(\omega)$ is given by

$$\begin{aligned} x(t) &= \frac{1}{\pi f_0}\int_0^{(1-\alpha)\pi f_0} \cos(\omega t)d\omega \\ &+ \frac{0.5}{\pi f_0}\int_{(1-\alpha)\pi f_0}^{(1+\alpha)\pi f_0}\left[1+\cos\left(\frac{1}{2\alpha f_0}[\omega-(1-\alpha)\pi f_0]\right)\right]\cos(\omega t)d\omega \\ &= \frac{\sin(\pi f_0 t)\cos(\pi\alpha f_0 t)}{\pi f_0 t[1-(2\alpha f_0 t)^2]}. \end{aligned} \tag{3.22}$$

The signal is shown in Figure 3.6(b). We observe that sampling at a rate $f_{\mathrm{sam}} = f_0$ yields the discrete-time signal

$$x[n] = \delta[n] \quad \Longrightarrow \quad X^{\mathrm{f}}(\theta) = 1.$$

Thus the sampled signal is obviously aliased, since $f_{\mathrm{sam}} = f_0$ is below the Nyquist rate. Moreover, the signal $x(t)$ is Nyquist-T for $T = 1/f_0$. Another way to arrive at this conclusion is to observe that in forming the sum

$$f_0 \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}(\omega - 2\pi k f_0),$$

the cosine shapes exactly add up to 1 at all intervals of overlap, so the result of the infinite sum is identically 1 at all frequencies.

Signals with raised-cosine spectra are useful in digital communication applications, thanks to their Nyquist-T property. They provide an example in which aliasing is not only harmless, but necessary for proper operation! □

The third and final case to be considered is that of a signal $x(t)$ whose bandwidth is not limited, as illustrated in Figure 3.7. In this case, the sum (3.10) includes an infinite number of nonzero terms, so the shape of the Fourier transform of the sampled signal must be distorted. Part b of the figure shows, in a thick solid line, the result of the infinite summation. The other lines (thin solid, two dashed, and two dotted) show five of the terms in the sum, corresponding to $k = -2, -1, 0, 1, 2$. The conclusion is that sampling of an infinite bandwidth signal always gives rise to aliasing, no matter how high the sampling rate.

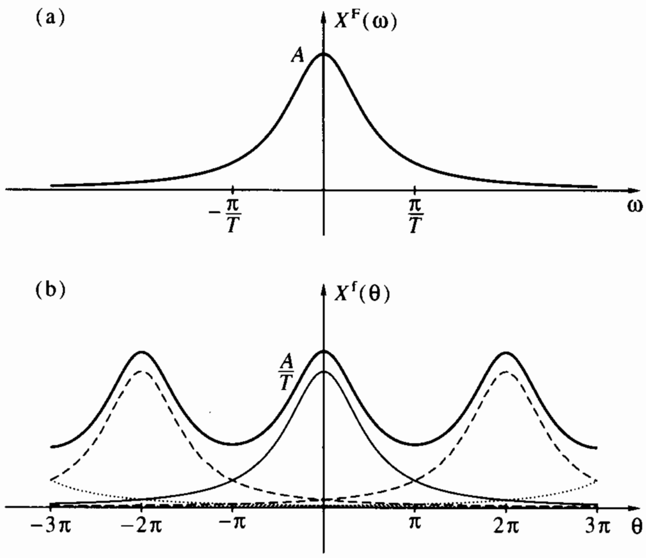

Figure 3.7 Sampling of a signal of an infinite bandwidth: (a) Fourier transform of the continuous-time signal; (b) Fourier transform of the sampled signal.

A famous theorem in the theory of Fourier transforms asserts that a signal of finite duration (i.e., which is identically zero outside a certain time interval) must have an infinite bandwidth.[3] Real-life signals always have finite duration, so we conclude that their bandwidth is always infinite. Sampling theory therefore implies that real-life signals must become aliased when sampled. Nevertheless, the bandwidth of a real-life signal is always *practically finite*, meaning that the percentage of energy outside a certain frequency range is negligibly small. It is therefore permitted, in most practical situations, to sample at a rate equal to twice the practical bandwidth, since then the effect of aliasing will be negligible.

Example 3.10 The practical bandwidth of a speech signal is about 4 kHz. Therefore, it is common to sample speech at 8-10 kHz. The practical bandwidth of a musical audio signal is about 20 kHz. Therefore, compact-disc digital audio uses a sampling rate of 44.1 kHz. □

A common practice in sampling of a continuous-time signal is to filter the signal *before* it is passed to the sampler. The filter used for this purpose is an analog low-pass filter whose cutoff frequency is not larger than *half the sampling rate.* Such a filter is called an *antialiasing* filter.

In summary, the rules of safe sampling are:

- Never sample below the Nyquist rate of the signal. To be on the safe side, use a safety factor (e.g., sample at 10 percent higher than the Nyquist rate).
- In case of doubt, use an antialiasing filter before the sampler.

Example 3.11 This story happened in 1976 (a year after Oppenheim and Schafer's classic *Digital Signal Processing* was published, but sampling and its consequences were not yet common knowledge among engineers); its lesson is as important today as it was then. A complex and expensive electrohydraulic system had to be built as part of a certain large-scale project. The designer of the system constructed a detailed mathematical model of it, and this was given to a programmer whose task was to write a computer simulation program of the system. When the simulation was complete, the system was still under construction. The programmer then reported that, under certain conditions, the system exhibited nonsinusoidal oscillations at a frequency of about 8 Hz, as shown in Figure 3.8. This gave rise to a general concern, since such a behavior was judged intolerable. The designer declared that such oscillations were not possible, although high-frequency oscillations, at about 100 Hz, were possible. Further examination revealed the following: The simulation had been carried out at a rate of 1000 Hz, which was adequate. However, to save disk storage (which was expensive those days) and plotting time (which was slow), the simulation output had been stored and plotted at a rate of 100 Hz. In reality, the oscillations were at 108 Hz, but as a result of the choice of plotting rate, they were aliased and appeared at 8 Hz. When the simulation output was plotted again, this time at 1000 Hz, it showed the oscillations at their true frequency, see Figure 3.9. □

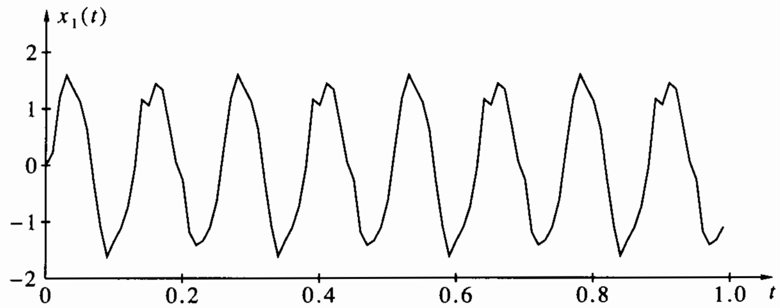

Figure 3.8 The apparent 8 Hz oscillations of the electrohydraulic system in Example 3.11.

3.4 Reconstruction

Suppose we are given a sampled signal $x[n]$ that is known to have been obtained from a band-limited signal $x(t)$ by sampling at the Nyquist rate (or higher). Since the Fourier transform of the sampled signal preserves the shape of the Fourier transform of the continuous-time signal, we should be able to reconstruct $x(t)$ exactly from its samples. How then do we accomplish such reconstruction? The answer is given by Shannon's reconstruction theorem.

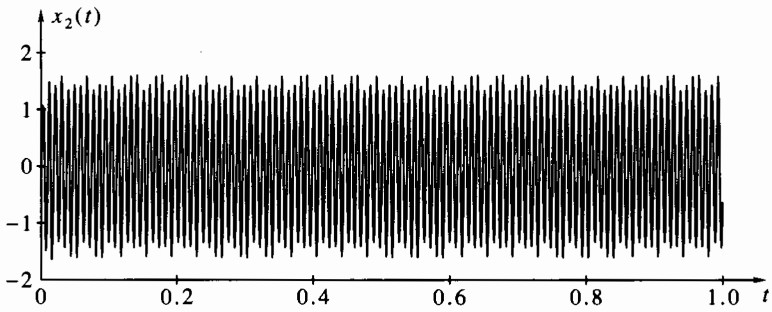

Figure 3.9 The true 108 Hz oscillations of the electrohydraulic system in Example 3.11.

Theorem 3.2 A band-limited signal $x(t)$ whose bandwidth is smaller than $\pm\pi/T$ can be exactly reconstructed from its samples $\{x(nT),\ -\infty < n < \infty\}$, using the formula

$$\boxed{x(t) = \sum_{n=-\infty}^{\infty} x(nT)\operatorname{sinc}\left(\frac{t-nT}{T}\right). \qquad (3.23)}$$

Proof As we did for the sampling theorem, we shall give two proofs of Shannon's theorem, one that is based on the impulse train and one that is not. Consider first $X_{\mathrm{p}}^{\mathrm{F}}(\omega)$, the Fourier transform of $x_{\mathrm{p}}(t)$, and recall from (3.9) that in the nonaliased case it is related to $X^{\mathrm{F}}(\omega)$ through

$$X^{\mathrm{F}}(\omega) = TX_{\mathrm{p}}^{\mathrm{F}}(\omega)H^{\mathrm{F}}(\omega), \tag{3.24}$$

where $H^{\mathrm{F}}(\omega)$ is an ideal low-pass filter, that is,

$$H^{\mathrm{F}}(\omega) = \operatorname{rect}\left(\frac{\omega T}{2\pi}\right) \tag{3.25}$$

(see Figure 3.10). Therefore, $x(t)$ is given by the convolution

$$x(t) = T\int_{-\infty}^{\infty} h(\tau)x_{\mathrm{p}}(t-\tau)d\tau. \tag{3.26}$$

The impulse response of the ideal low-pass filter is obtained from (2.35) as

$$h(t) = \frac{1}{T}\operatorname{sinc}\left(\frac{t}{T}\right). \tag{3.27}$$

Therefore,

$$x(t) = \int_{-\infty}^{\infty} \operatorname{sinc}\left(\frac{\tau}{T}\right)\left[\sum_{n=-\infty}^{\infty} x(nT)\delta(t-\tau-nT)\right]d\tau = \sum_{n=-\infty}^{\infty} x(nT)\operatorname{sinc}\left(\frac{t-nT}{T}\right). \tag{3.28}$$

We now give a direct proof of (3.23). We have from (3.19)

$$\begin{aligned} x(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} X^{\mathrm{F}}(\omega)e^{j\omega t}d\omega = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} X^{\mathrm{f}}(\omega T)e^{j\omega t}d\omega \\ &= \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T}\left[\sum_{n=-\infty}^{\infty} x(nT)e^{-j\omega nT}\right]e^{j\omega t}d\omega \\ &= \frac{T}{2\pi}\sum_{n=-\infty}^{\infty} x(nT)\int_{-\pi/T}^{\pi/T} e^{j\omega(t-nT)}d\omega = \sum_{n=-\infty}^{\infty} x(nT)\operatorname{sinc}\left(\frac{t-nT}{T}\right). \end{aligned} \tag{3.29}$$

□

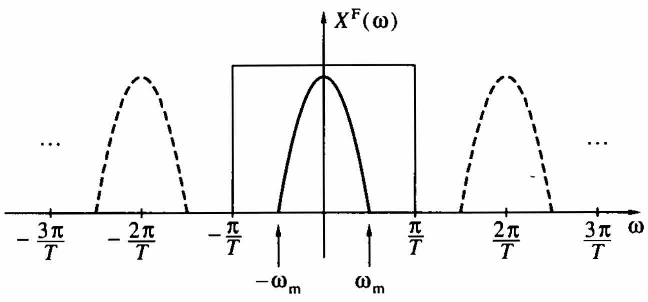

Figure 3.10 Reconstruction of a band-limited signal by an ideal low-pass filter (replicas shown by dashed lines will be eliminated by the filter).

Let us interpret the meaning of the reconstruction formula (3.23). Consider first a time point that coincides with one of the sampling points, say $t = n_0T$. We have, by the properties of the sinc function,

$$\text{sinc}\left(\frac{n_0T - nT}{T}\right) = \delta[n - n_0]. \tag{3.30}$$

The right side of (3.23) is indeed equal to $x(n_0T)$, as it should be. This is true for all n_0, so we conclude that the reconstruction formula is correct at the sampling points. Now choose a point t located between n_0T and $(n_0 + 1)T$. Then none of the values of the sinc function will be zero in general. However, points closer to n_0T will contribute to the sum more than points farther from n_0T, since the sinc is a decaying function (although not monotonically decaying). All these contributions, infinite in number, will add exactly to $x(t)$, as guaranteed by Shannon's formula. What if $x(t)$ has an irregular behavior between the two sampling points n_0T and $(n_0 + 1)T$, which the sinc function doesn't see? The answer is that the bandwidth of such a signal is necessarily higher than the limit π/T imposed by Shannon's theorem. Stated in different words: A signal whose bandwidth is limited to π/T must be fairly smooth—it cannot have more than "half an oscillation" between any two adjacent sampling points.

The Shannon reconstructor (also called *Shannon interpolator* or *sinc interpolator*) is a noncausal system. Reconstruction of $x(t)$ at any time t requires knowledge of the entire sequence $\{x(nT),\ -\infty < n < \infty\}$. Therefore, Shannon's reconstructor is important mainly from a theoretical viewpoint. In real-time signal processing, we must use a causal reconstructor, preferably one that is convenient to implement. Let $h(t)$ be the impulse response of the reconstructor. The reconstruction formula is then

$$\hat{x}(t) = \sum_{n=-\infty}^{\lfloor t/T \rfloor} x(nT)h(t - nT). \tag{3.31}$$

The impulse response $h(t)$ should be chosen such that $\hat{x}(t)$ is a good approximation of $x(t)$. The limits of the sum reflect the causality requirement (the upper limit is the largest integer not larger than t/T).

A simple device that realizes (3.31) is the *zero-order hold* (ZOH) circuit. A zero-order hold is a device that, when given an input sample $x[n]$, maintains a constant output equal to $x[n]$ during the interval $[nT, nT + T)$ and then resets its output to zero. When fed with the sequence $\{x[n],\ -\infty < n < \infty\}$, it yields the output

$$\hat{x}(t) = x[n], \quad nT \le t < nT + T, \quad \text{for all } n \in \mathbb{Z}. \tag{3.32}$$

The impulse response of the ZOH device is

$$h_{\mathrm{zoh}}(t) = \begin{cases} 1, & 0 \le t < T, \\ 0, & \text{otherwise.} \end{cases} \tag{3.33}$$

The impulse response (3.33) is shown in Figure 3.11; the response $\hat{x}(t)$ to the input sequence $x(nT)$ is shown in Figure 3.12. The latter figure also shows, in a dashed line, the ideal waveform $x(t)$, which would be obtained by a Shannon reconstructor.

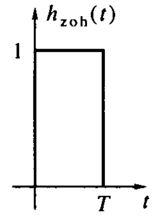

Figure 3.11 Impulse response of a zero-order hold.

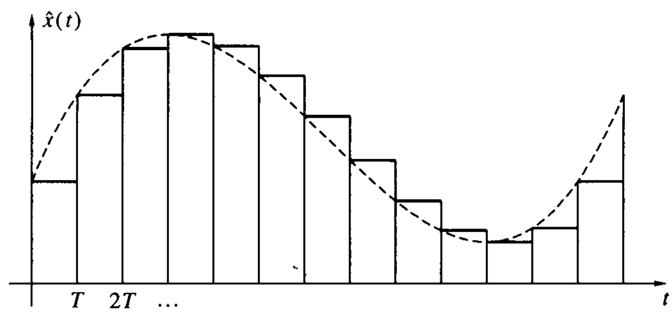

Figure 3.12 Time response of a zero-order hold. Staircase line: actual response $\hat{x}(t)$; dashed line: ideal response $x(t)$.

To understand to what extent $\hat{x}(t)$ approximates $x(t)$, let us compute the frequency response of the ZOH. As we recall, the frequency response of the ideal (Shannon) reconstructor is a perfect rectangle on $-\pi/T \le \omega \le \pi/T$, with zero phase; see (3.25). By analyzing how the frequency response of the ZOH deviates from the perfect rectangle, we will understand the nature of distortions introduced by the ZOH. We have from (3.33)

$$H_{\mathrm{zoh}}^{\mathrm{F}}(\omega) = \int_0^T e^{-j\omega t} dt = \frac{1 - e^{-j\omega T}}{j\omega} = T \operatorname{sinc}\left(\frac{\omega T}{2\pi}\right) e^{-j0.5\omega T}. \tag{3.34}$$

The magnitude and phase responses of $H_{\mathrm{zoh}}^{\mathrm{F}}(\omega)$ are shown in Figure 3.13. We observe the following differences with respect to the ideal low-pass filter:

1. The magnitude response at low frequencies is not flat, but decays gradually. Furthermore, it decreases to zero at $\omega = \pm 2\pi/T$, rather than at $\pm\pi/T$.
2. The magnitude response has nonvanishing ripple at high frequencies, so the reconstructed signal $\hat{x}(t)$ has undesired high-frequency energy. In the time domain, the high-frequency energy is apparent in the staircaselike form of the output created by the hold operation.
3. The phase of the response is not zero, but piecewise linear, with slope $-0.5T$.

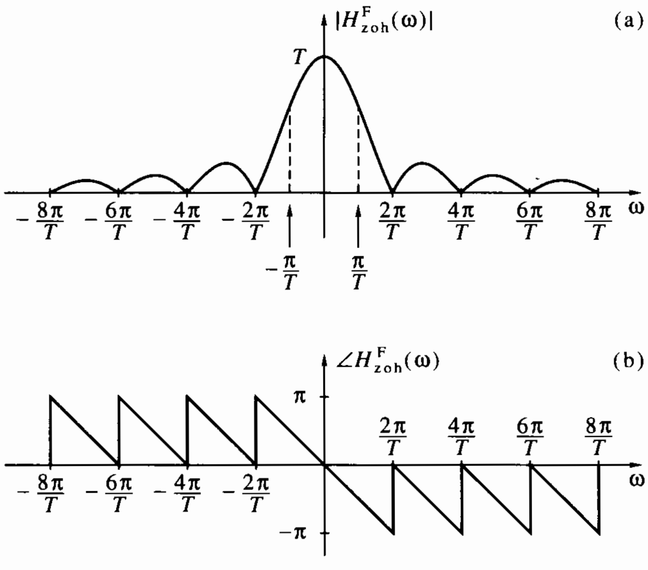

Figure 3.13 Frequency response of a zero-order hold: (a) magnitude; (b) phase.

The undesirable effects of the zero-order hold can be partly overcome by passing its output through a low-pass analog filter. The ideal magnitude response $|G^F(\omega)|$ of such a filter is shown in Figure 3.14. The magnitude response is the inverse of $|H^F_{zoh}(\omega)|$ in the frequency range $[-\pi/T, \pi/T]$ and zero elsewhere. We thus get from (3.34) that

$$|G^F(\omega)| = \left[\operatorname{sinc}\left(\frac{\omega T}{2\pi}\right)\right]^{-1} \operatorname{rect}\left(\frac{\omega T}{2\pi}\right). \tag{3.35}$$

Such a filter is impossible to implement, due its infinitely steep cutoff, so in practice an approximation must be used instead.

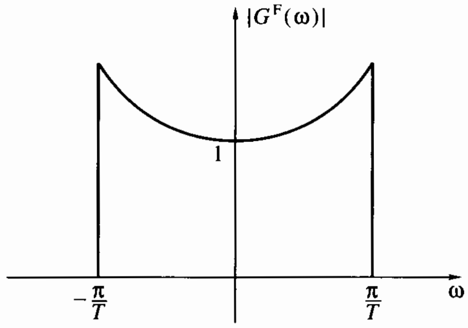

Figure 3.14 Magnitude response of an ideal reconstruction filter at the output of a ZOH.

Other reconstruction devices, more sophisticated than the ZOH, are sometimes used, but are not discussed here (see Problems 3.29 and 3.41 for two alternatives). We remark that non-real-time reconstruction offers much more freedom in choosing the reconstruction filter and allows for a greatly improved frequency response (compared with the one shown in Figure 3.13). Non-real-time signal processing involves sampling and storage of a finite (but potentially long) segment of a physical signal, followed by off-line processing of the stored samples. If such a signal needs to be reconstructed, it is perfectly legitimate, and even advisable, to use a noncausal filter for reconstruction.

We summarize our discussion of reconstruction by showing a complete typical DSP system, as depicted in Figure 3.15. The continuous-time signal $x(t)$ is passed through an antialiasing filter, then fed to the sampler. The resulting discrete-time signal $x[n]$ is then processed digitally as needed in the specific application. The discrete-time signal at the output $y[n]$ is passed to the ZOH and then low-pass filtered to give the final continuous-time signal $\hat{y}(t)$.

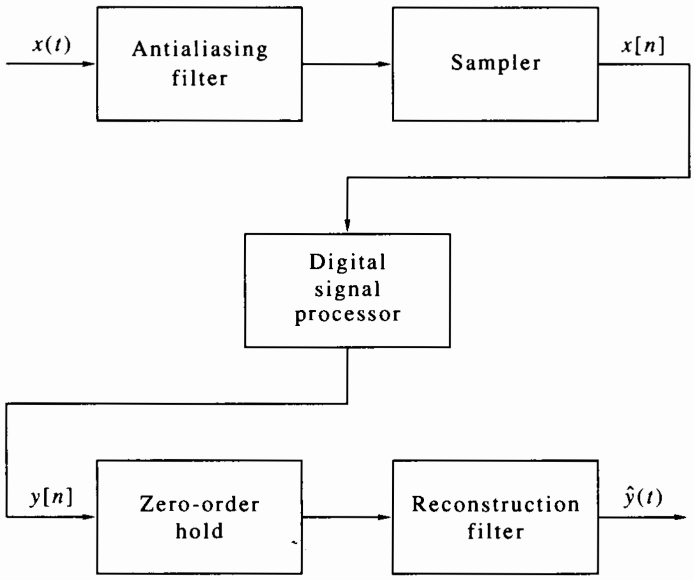

Figure 3.15 A typical digital signal processing system.

3.5 Physical Aspects of Sampling and Reconstruction*

So far we have described sampling and reconstruction as mathematical operations. We now describe electronic circuitry for implementing those two operations. We shall see that hardware limitations introduce imperfections to both sampling and reconstruction. These imperfections are chiefly *nonlinearities*, of which there are three major types:

1. Saturation, since voltages must be confined to a certain range, depending on the specific electronic circuit.
2. Quantization, since only a finite number of bits can be handled by the circuit.
3. Nonlinearities introduced by electronic components (tolerances of resistors, nonlinearity of operational amplifiers, and so on).

Nonlinearities of these types appear in both sampling and reconstruction in a similar manner. Each of the two operations, sampling and reconstruction, also has its own characteristic imperfections. Sampling is prone to signal smearing, since it must be performed over finite time intervals. Switching delays make reconstruction prone to instantaneous deviations of the output signal (known to engineers as *glitches*).

3.5.1 Physical Reconstruction

Physical reconstruction is implemented using a device called *digital-to-analog converter*, or D/A. A D/A converter approximates zero-order hold operation. It accepts a discrete-time signal $x[n]$ in a form of a sequence of binary numbers. Each binary number is held fixed by a data register for a period of T seconds (the sampling interval). The binary number at the register's output is converted to a voltage waveform $\hat{x}(t)$. The voltage is approximately proportional to the present value of $x[n]$, and remains fixed for T seconds. When the next binary number $x[n+1]$ appears at the register's output, $\hat{x}(t)$ changes accordingly. The result is approximately the staircase waveform shown in Figure 3.12. Figure 3.16 depicts this sequence of operations schematically.

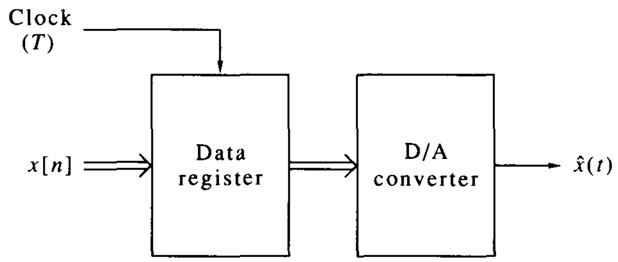

Figure 3.16 Schematic diagram of physical reconstruction by a D/A converter.

A simple electronic circuit for implementing a D/A converter is shown in Figure 3.17. The figure illustrates a 3-bit D/A, but it can be easily extended to any number of bits. This D/A is *unipolar*, that is, the binary number at its input is considered nonnegative and the voltage at its output has only one polarity (always negative in Figure 3.17, assuming that V_{ref} is positive). The signal $x[n]$ is represented by the binary number $b_0 b_1 \dots b_{B-1}$, where B is the number of bits, 3 in this example. The most significant bit (MSB) is b_0, and the least significant bit (LSB) is b_{B-1}. It is convenient to think of this number as a fraction in the range $[0, 1-2^{-B}]$, that is, as representing the numerical value $\sum_{k=0}^{B-1} b_k 2^{-(k+1)}$. The circuit operates as follows:

1. Assuming that the gain of the operational amplifier is high (approximately infinity), its inverting input has nearly zero potential (*virtual ground* is the common term). Therefore, regardless of the positions of the switches, the bottom terminals of the $2R$ resistances of the resistor ladder behave as if they were grounded. Application of rules for parallel-series connection of resistors shows that the total resistance seen by the voltage source V_{ref} is $2R$.
2. Assuming that the voltage source is ideal (i.e., its internal resistance is zero), the current flowing from it is $I = V_{\text{ref}}/2R$. The current splits into two equal parts at the node, so $I_0 = I/2$. Continuing in this manner, we observe that $I_1 = I_0/2 = I/4$ and, in general,
$$I_k = 2^{-(k+1)} I, \quad 0 \le k \le B-1. \tag{3.36}$$
3. Assume that the switches are ideal and that the kth switch is in the right position when $b_k = 1$ and in the left position is when $b_k = 0$. Then the current fed by the kth section to the operational amplifier's input is 0 or I_k, depending on whether b_k is 0 or 1. In other words, this current is $b_k I_k$. Therefore we get, by current balancing at the input of the operational amplifier (assuming it has an infinite

input resistance),

$$\frac{V_{out}}{2R} = -\sum_{k=0}^{B-1} b_k I_k = -I \sum_{k=0}^{B-1} b_k 2^{-(k+1)}, \tag{3.37}$$

$$V_{out} = -V_{ref} \sum_{k=0}^{B-1} b_k 2^{-(k+1)}. \tag{3.38}$$

As we see, the output voltage is proportional to the binary number at the input.[4]

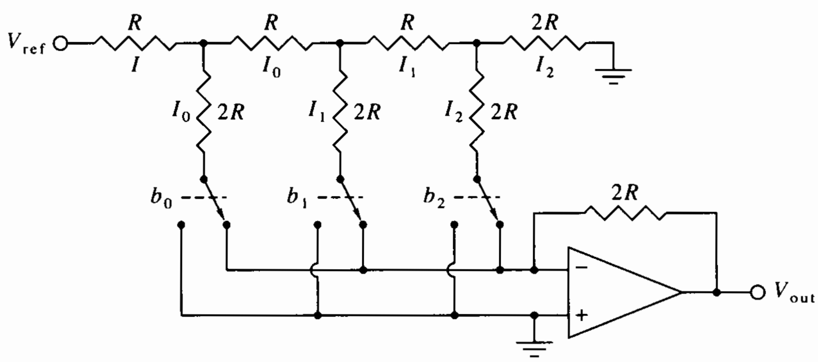

Figure 3.17 Schematic diagram of a digital-to-analog converter.

Practical D/A converters are often *bipolar*; their output voltage can be either positive or negative, depending on the sign of the binary number at the input. Table 3.1 illustrates, in its first two columns, the correspondence between binary numbers and voltages (relative to V_{ref}) for a 3-bit bipolar D/A. The binary word 00...0 corresponds to the smallest (most negative) voltage and the word 11...1 corresponds to the largest (most positive) voltage. This correspondence is called *offset binary*. The possible voltage values are symmetric with respect to 0; it is impossible to get zero voltage; finally, the largest possible absolute voltage is $(1 - 2^{-B})V_{ref}$. The absolute possible voltage represents the *saturation level* of the D/A. The voltage increment, also called the *quantization level*, is $2^{-(B-1)}V_{ref}$.

$8\frac{V_{out}}{V_{ref}}$	offset binary	two's-complement
−7	000	100
−5	001	101
−3	010	110
−1	011	111
1	100	000
3	101	001
5	110	010
7	111	011

Table 3.1 Correspondence between output voltage and binary representations (offset binary and two's-complement) in a bipolar D/A converter.

Numbers in a computer are usually represented in a *two's-complement* form.[5] The two's-complement representations of the eight possible voltages in the 3-bit case are shown in the third column of Table 3.1. As we see, two's-complement representation is obtained from offset binary representation by a simple rule: Invert the most significant bit and leave the other bits unchanged. Figure 3.18 illustrates the correspondence between voltages and binary numbers in the two's-complement case.

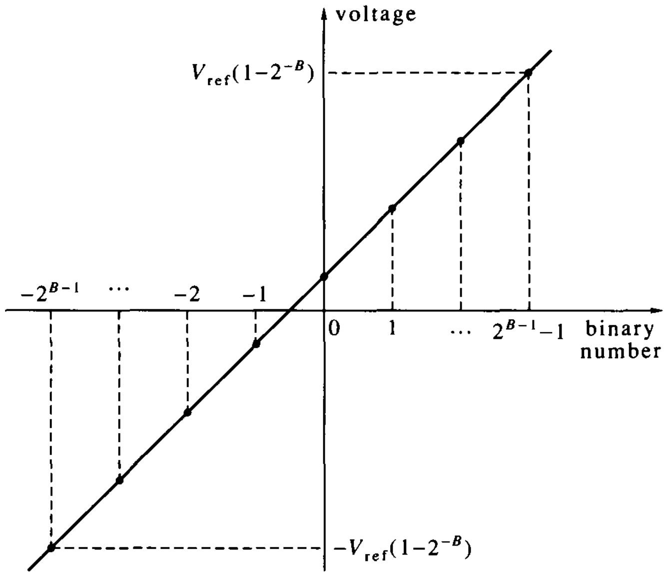

Figure 3.18 Correspondence between two's-complement binary numbers and voltages in a digital-to-analog converter.

The ratio between the maximum possible voltage and half the quantization level is called the *dynamic range* of the D/A. As we have seen, this number is $2^B - 1$, or nearly 2^B. The corresponding parameter in decibels is about $6B$ dB. Thus, a 10-bit D/A has a dynamic range of 60 dB. For example, high-fidelity music usually requires a dynamic range of 90 dB or more, so at least a 16-bit D/A is necessary for this purpose. D/A converters with high dynamic ranges are expensive to manufacture, since they impose tight tolerances on the analog components.

3.5.2 Physical Sampling

Physical sampling is implemented using a device called *analog-to-digital converter*, or A/D. The A/D device approximates point sampling. It accepts a continuous-time signal $x(t)$ in a form of an electrical voltage and produces a sequence of binary numbers $x[n]$, which approximate the corresponding samples $x(nT)$. Often the electrical voltage is not fed to the A/D directly, but through a device called *sample-and-hold*, or S/H; see Figure 3.19. Sample-and-hold is an analog circuit whose function is to measure the input signal value at the clock instant (i.e., at an integer multiple of T) and hold it fixed for a time interval long enough for the A/D operation to complete. Analog-to-digital conversion is potentially a slow operation, and variation of the input voltage during the conversion may disrupt the operation of the converter. The S/H prevents such disruption by keeping the input voltage constant during conversion. When the input voltage variation is slow relative to the speed of the A/D, the S/H is not needed and the input voltage may be fed directly to the A/D.

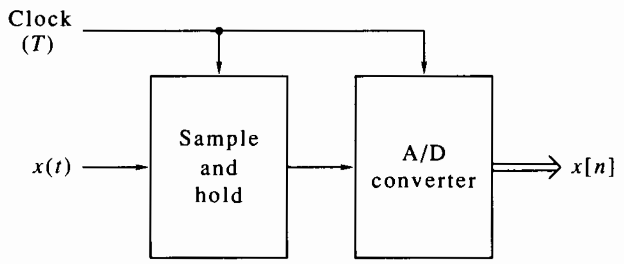

Figure 3.19 Schematic diagram of physical sampling by an A/D converter.

Similarly to D/A converters, A/D converters exhibit saturation and quantization. Figure 3.20 depicts two types of A/D response. Part a shows a *rounding* A/D; its binary output corresponding to an input voltage V_{in} is given by

$$x = \left\lfloor \frac{V_{\text{in}}}{V_{\text{ref}}} 2^{B-1} + 0.5 \right\rfloor, \tag{3.39}$$

where V_{ref} is the reference voltage to the A/D, B is the number of bits, and $\lfloor a \rfloor$ denotes the integer nearest to a from below. As we see from (3.39) and from the figure, the binary number can be up to half the quantization level above or below the ideal output (represented in the figure by a dashed line). The rounding A/D thus has a symmetric response, except near the positive and negative saturation levels. The reason for the asymmetry at the ends is that the number of possible values is necessarily even, being a power of 2. In two's-complement arithmetic, the maximum positive value is $2^{B-1} - 1$ and the maximum negative value is -2^{B-1}. Correspondingly, the positive saturation level is slightly smaller than the negative one.

Figure 3.20(b) shows a *truncating* A/D; its binary output corresponding to an input voltage V_{in} is given by

$$x = \left\lfloor \frac{V_{\text{in}}}{V_{\text{ref}}} 2^{B-1} \right\rfloor. \tag{3.40}$$

As we see from (3.40) and from the figure, the binary number is always less than the ideal output (represented in the figure by a dashed line) by up to one quantization level. The truncating A/D thus has an asymmetric response.

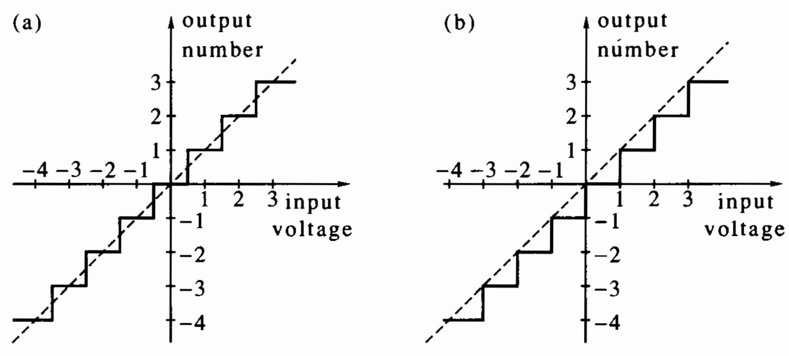

Figure 3.20 Quantization in analog-to-digital converter: (a) rounding; (b) truncation. Staircase lines show the actual responses; dashed lines show the ideal responses.

Figure 3.21 illustrates the result of sampling an analog signal with a rounding A/D. In this example the A/D is 4 bits, or 16 quantization levels (8 for positive values of the input signal). As we see, rounding results in an error that lies in the range plus or minus half the quantization level. In addition, we see how the extreme positive values of the input signal are chopped because of saturation. The error in the sampled values due to quantization (but not due to saturation) is called *quantization noise.*

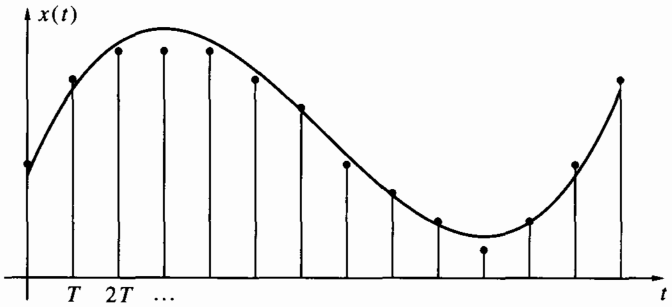

Figure 3.21 Quantization error in a rounding A/D.

Analog-to-digital converters can be implemented in several ways, depending on speed and cost considerations. The speed of an A/D converter is measured by the maximum number of conversions per second. The time for a single conversion is approximately the reciprocal of the speed. Usually, the faster the A/D, the more complex is the hardware and the costlier it is to build. Common A/D implementations include:

1. Successive approximation A/D; see Figure 3.22. This A/D builds the output bits in a feedback loop, one at a time, starting with the most significant bit (MSB). Feedback is provided by a D/A, which converts the output bits of the A/D to an electrical voltage. Initially the shift register MSB is set to 1, and all the other bits to 0. This sets the data register MSB to 1, so the D/A output becomes $0.5V_{\text{ref}}$. The comparator decides whether this is lower or higher than the input voltage V_{in}. If it is lower, the MSB retains the value 1. If it is higher, the MSB is reset to 0. At the next clock cycle, the 1 in the shift register shifts to the bit below the MSB, and sets the corresponding bit of the data register to 1. Again a comparison is made between the D/A output and V_{in}. If $V_{\text{in}} > V_{\text{fb}}$, the bit retains the value 1, otherwise it is reset to 0. This process is repeated B times, until all the data register bits are set to their proper values. Such an A/D is relatively inexpensive, requiring only a D/A, a comparator, a few registers, and simple logic. Its conversion speed is proportional to the number of bits, because of its serial operation. Successive approximation A/D converters are suitable for many applications, but not for ones in which speed is of prime importance.

   When a bipolar A/D converter is required, it is convenient to use offset-binary representation, since then the binary number is a monotone function of the voltage. The representation can be converted to two's-complement by inverting the MSB.

2. Flash A/D; see Figure 3.23 for a 3-bit example. This converter builds all output bits in parallel by directly comparing the input voltage with all possible output values. It requires $2^B - 1$ comparators. As is seen from Figure 3.23, the bottom comparators up to the one corresponding to the input voltage will be set to 1, whereas the ones above it will be set to 0. Therefore, we can determine that the quantized voltage is n quantization levels up from $-V_{\text{ref}}$ if the nth comparator is set to 1 and the $(n + 1)$st comparator is set to 0. This is accomplished by

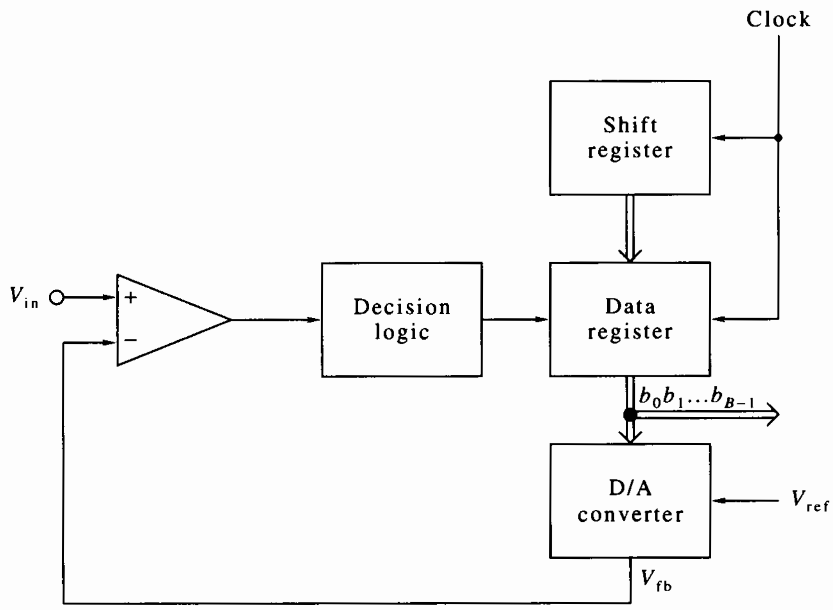

Figure 3.22 Schematic diagram of a successive approximation A/D converter.

the AND gates shown in the figure. The quantized voltage is $-V_{ref}$ if the bottom comparator is set to 0, and is $(1 - 2^{-(B-1)})V_{ref}$ if the top comparator is set to 1. The encoding logic then converts the gate outputs to the appropriate binary word. This scheme implements a truncating A/D; if a rounding A/D is required, it can be achieved by changing the reference voltages V_{ref} and $-V_{ref}$ to $(1 - 2^{-B})V_{ref}$ and $-(1 + 2^{-B})V_{ref}$, respectively. Another possibility is to change the bottom and top resistors to $0.5R$ and $1.5R$, respectively.

Flash A/D converters are the fastest possible, since all bits are obtained simultaneously. On the other hand, their hardware complexity grows exponentially with the number of bits, so they become prohibitively expensive for a large number of bits. Their main application is for conversion of video signals, since these require high speeds (on the order of 10^7 conversions per second), whereas the number of bits is typically moderate.

3. Half-flash A/D; see Figure 3.24. This A/D offers a compromise between speed and complexity. It uses two flash A/D converters, each for half the number of bits. The number of comparators is $2(2^{B/2} - 1)$, which is significantly less than the number of comparators in a flash A/D converter having the same number of bits. The $B/2$ most significant bits are found first, and then converted to analog using a $B/2$-bit D/A. The D/A output is subtracted from the input voltage and used, after being passed through a S/H, to find the $B/2$ least significant bits. The conversion time is about twice that of a full-flash A/D.

4. Sigma-delta A/D converter. This type of converter provides high resolutions (i.e., a large number of bits) with relatively simple analog circuitry. It is limited to applications in which the signal bandwidth is relatively low and speed is not a major factor. The theory of sigma-delta converters relies on concepts and techniques we have not studied yet. We therefore postpone the explanation of such converters to Section 14.7; see page 586.

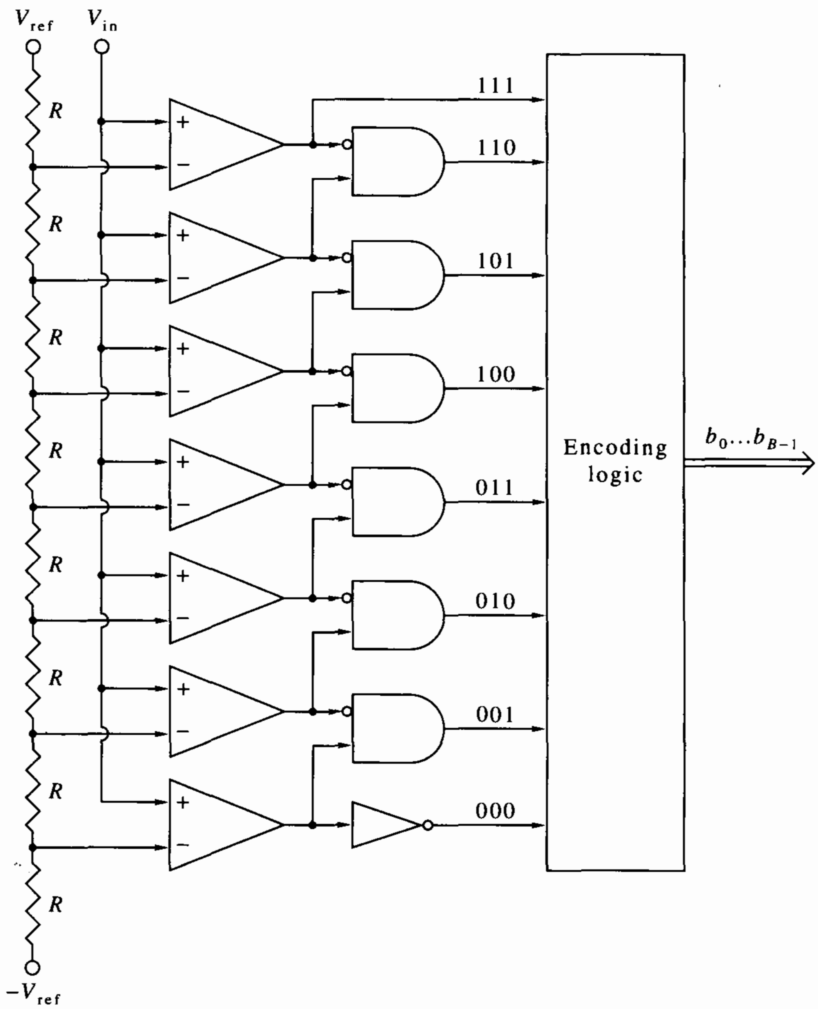

Figure 3.23 Schematic diagram of a flash A/D converter.

Now we discuss the circumstances under which an S/H is needed at the input of the A/D. For safe operation without S/H, the input voltage should not change by more than half the quantization level during the conversion time. Assume that the input voltage is a sinusoid of amplitude V_{ref} and frequency f. The greatest rate of change occurs when the sinusoid crosses zero, its value there being $2\pi f V_{ref}$. Let us denote by f_{ad} the number of conversions per second, so the duration of a single conversion is $1/f_{ad}$. Then the maximum amount of change during the conversion time is $2\pi f V_{ref}/f_{ad}$. This must be less than $2^{-B}V_{ref}$, so we arrive at the condition

$$f \leq \frac{f_{ad}}{\pi 2^{B+1}}. \tag{3.41}$$

Condition (3.41) imposes an upper limit on the frequency of the input signal for which an S/H is not needed. If the frequency of the input signal exceeds the right side of (3.41), an S/H is necessary for safe operation.

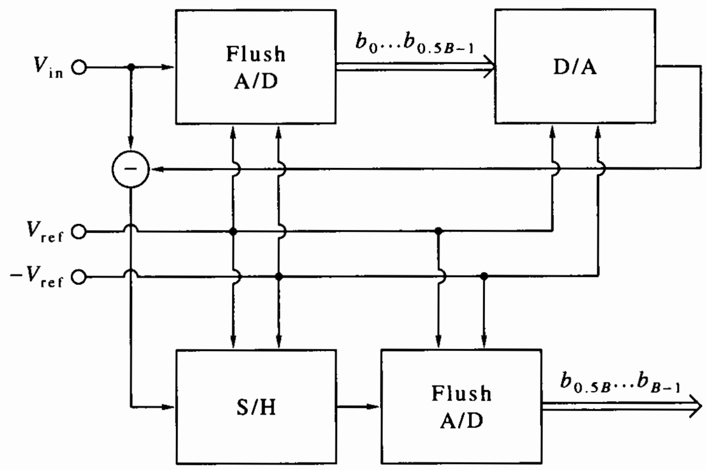

Figure 3.24 Schematic diagram of a half-flash A/D converter.

3.5.3 Averaging in A/D Converters

So far we have assumed ideal operation of the S/H circuit. In reality, S/H performs averaging of the continuous-time signal over a short, but finite period. Thus, a physical S/H can be mathematically described (at least approximately) by the relationship

$$x[n] = \frac{1}{\Delta} \int_{nT-\Delta}^{nT} x(t)dt, \tag{3.42}$$

where Δ is the averaging interval, assumed to be smaller than the sampling interval T (otherwise the A/D will not be able to operate in real time). The operation (3.42) is depicted in Figure 3.25.

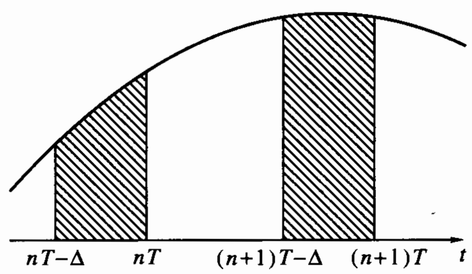

Figure 3.25 Averaging in A/D converter.

Intuitively, if $\Delta \ll T$, then physical sampling must be well approximated by point sampling. However, when Δ is a sizable fraction of T, we expect low-pass filtering effect that is not negligible. We now analyze this effect and express it in precise terms. Define the auxiliary signal $y(t)$ as the indefinite integral of $x(t)$, that is,

$$y(t) = \int^{t} x(\tau)d\tau + C, \tag{3.43}$$

where C is an arbitrary constant. Then we have from (3.42)

$$x[n] = \frac{1}{\Delta}\{y(nT) - y(nT - \Delta)\}. \tag{3.44}$$

The sequence $y(nT)$ is the point sampling of $y(t)$ and $y(nT - \Delta)$ is the point sampling of $y(t-\Delta)$. The Fourier transform of the former is $(j\omega)^{-1}X^{\mathrm{F}}(\omega)$ and that of the latter is $(j\omega)^{-1}X^{\mathrm{F}}(\omega)e^{-j\omega\Delta}$. Therefore we get, by linearity and by the sampling theorem (3.10),

$$X^{\mathrm{f}}(\theta) = \frac{1}{T}\sum_{k=-\infty}^{\infty} \tilde{X}^{\mathrm{F}}\left(\frac{\theta - 2\pi k}{T}\right), \tag{3.45}$$

where we define

$$\tilde{X}^{\mathrm{F}}(\omega) = \frac{1 - e^{-j\omega\Delta}}{j\omega\Delta} X^{\mathrm{F}}(\omega) = e^{-j0.5\omega\Delta}\operatorname{sinc}\left(\frac{\omega\Delta}{2\pi}\right) X^{\mathrm{F}}(\omega). \tag{3.46}$$

As we see, sampling with averaging is equivalent to low-pass filtering followed by point sampling. The transfer function of the low-pass filter is identical to that of a ZOH, except that the hold interval is Δ. When Δ is small with respect to T, the attenuation and phase lag of the low-pass filter are negligible. At the other extreme, when Δ approaches T, we get the same magnitude attenuation and phase delay as described for the ZOH; see Figure 3.13. Recall, however, that if the signal $x(t)$ was passed through an antialiasing filter, its bandwidth does not exceed $\pm\pi/T$. Therefore, only the region $[-\pi/T, \pi/T]$ of Figure 3.13 is of interest in this case.

In summary, ideal point sampling is a good approximation of physical sampling if the averaging interval of the S/H is small with respect to the sampling interval. If not, there is an additional low-pass filtering effect that needs to be taken into account, as expressed by (3.46).

3.6 Sampling of Band-Pass Signals*

A continuous-time signal $x(t)$ is called *band pass* if its Fourier transform vanishes outside a certain frequency range that does not include $\omega = 0$: in other words, if there exist $0 < \omega_1 < \omega_2$ such that

$$X^{\mathrm{F}}(\omega) = 0 \ \text{ for } \ |\omega| \le \omega_1 \ \text{ and for } \ |\omega| \ge \omega_2. \tag{3.47}$$

Signals transmitted electromagnetically—such as radio, radar, or signals transmitted via optical fibers—are band pass. It is common to define the *bandwidth* of a band-pass signal as $\omega_2 - \omega_1$ [or $(\omega_2 - \omega_1)/2\pi$ in hertz]. For example, the bandwidth of commercial AM radio transmission is 10 kHz in the United States and 9 kHz in Europe and many other countries. The bandwidth of commercial FM radio transmission is about 180 kHz.

Band-pass signals are band limited to $\pm\omega_2$, so they can be sampled at a rate ω_2/π without being aliased. Since, however, the frequency support of the signal is only $2(\omega_2 - \omega_1)$, it appears to be wasteful to sample at such a rate if $\omega_2 \gg (\omega_2 - \omega_1)$. We now show that indeed, a band-pass signal can be sampled at a rate not much higher than $(\omega_2 - \omega_1)/\pi$ without being subject to aliasing.

Consider first the special case of ω_2 an integer multiple of the bandwidth, say

$$\omega_2 = L(\omega_2 - \omega_1). \tag{3.48}$$

Let us sample the signal at an interval

$$T = \frac{\pi}{\omega_2 - \omega_1} = \frac{\pi L}{\omega_2}. \tag{3.49}$$

As usual, denote the impulse-sampled signal by $x_{\mathrm{p}}(t)$ and its Fourier transform by $X_{\mathrm{p}}^{\mathrm{F}}(\omega)$. Then, by the sampling theorem,

$$X_{\mathrm{p}}^{\mathrm{F}}(\omega) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X^{\mathrm{F}}\left(\omega - \frac{2\pi k}{T}\right) = \frac{1}{T}\sum_{k=-\infty}^{\infty} X^{\mathrm{F}}(\omega - 2k(\omega_2 - \omega_1)). \tag{3.50}$$

Now, $X^{\mathrm{F}}(\omega - 2k(\omega_2 - \omega_1))$ is nonzero only in the range

$$\omega_1 \le |\omega - 2k(\omega_2 - \omega_1)| \le \omega_2,$$

or

$$(2k - L)(\omega_2 - \omega_1) \le \omega \le (2k + 1 - L)(\omega_2 - \omega_1),$$
$$(2k + L - 1)(\omega_2 - \omega_1) \le \omega \le (2k + L)(\omega_2 - \omega_1).$$

This is illustrated in Figure 3.26 for even L (4 in this case) and odd L (3 in this case). As we see, the replicas corresponding to different values of k *do not overlap*. In particular, the replicas corresponding to $k = \pm L/2$ if L is even, or $k = \pm(L-1)/2$ if L is odd, appear in the interval

$$-(\omega_2 - \omega_1) \le \omega \le \omega_2 - \omega_1,$$

which corresponds to the interval $[-\pi, \pi]$ in the θ domain. The conclusion is that the sampled signal *is not aliased*, despite being sampled at a rate smaller than Nyquist's critical rate.

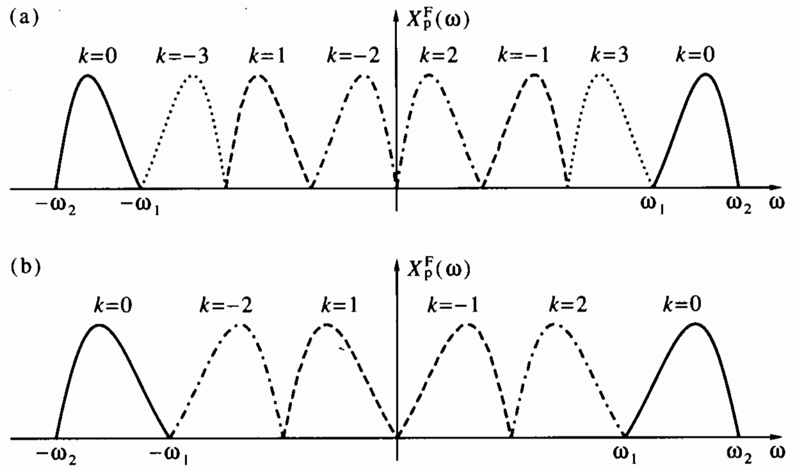

Figure 3.26 Sampling of a band-pass signal below the Nyquist rate when $(\omega_2 - \omega_1)$ is an integer multiple of the bandwidth: (a) $L = 4$; (b) $L = 3$.

Next consider the general case, where ω_2 is not an integer multiple of the bandwidth. The idea in this case is to extend the bandwidth of $x(t)$ artificially to the range

$$\omega_0 < |\omega| < \omega_2,$$

such that (1) $\omega_0 \le \omega_1$ and (2) ω_2 is an integer multiple of $\omega_2 - \omega_0$, say

$$\omega_2 = L(\omega_2 - \omega_0). \tag{3.51}$$

We can then sample the signal at an interval

$$T = \frac{\pi}{(\omega_2 - \omega_0)} = \frac{\pi L}{\omega_2} \tag{3.52}$$

and the sampled signal will be alias free, by the same argument as before. Figure 3.27 illustrates the procedure of extending the bandwidth and sampling at the interval given by (3.52).

The integer factor L is calculated as follows. We have

$$\omega_0 = \frac{L-1}{L}\omega_2 \le \omega_1 \implies L \le \frac{\omega_2}{\omega_2 - \omega_1}. \tag{3.53}$$

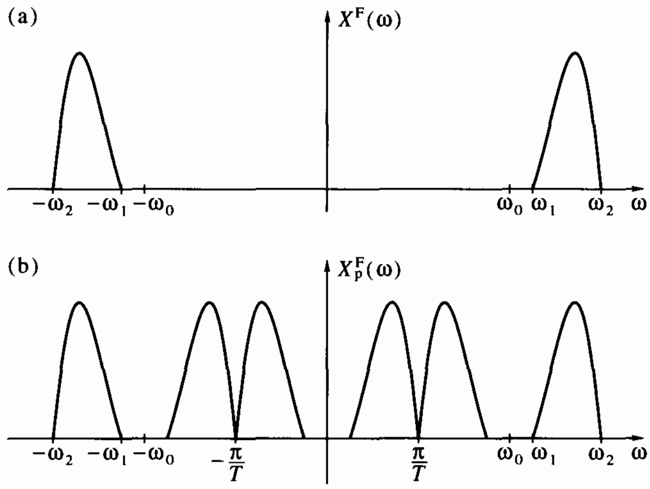

Figure 3.27 Sampling of a band-pass signal below the Nyquist rate in a general case: (a) Fourier transform of the continuous-time signal; (b) Fourier transform of the sampled signal.

But L must be an integer, so

$$L = \left\lfloor \frac{\omega_2}{\omega_2 - \omega_1} \right\rfloor. \tag{3.54}$$

In summary, the signal can be sampled at interval

$$T = \frac{\pi \lfloor \omega_2/(\omega_2 - \omega_1) \rfloor}{\omega_2}. \tag{3.55}$$

This is only slightly different from $\pi/(\omega_2 - \omega_1)$ if $\omega_2 \gg (\omega_2 - \omega_1)$.

When implementing band-pass sampling, the following point is of extreme importance and should not be overlooked: Regardless of the sampling rate, the A/D converter must be speed compatible with the highest signal frequency ω_2, rather than with the sampling rate. Recall from our discussion of physical sampling that A/D conversion involves averaging. Unless the averaging interval Δ is much smaller than one period of ω_2, averaging will disrupt the information and the samples will be grossly distorted. Therefore, A/D converters for band-pass sampling are faster, hence costlier, than A/D converters for low-pass signals.

Example 3.12 Radio receivers, since the invention of the *superheterodyne*, convert their high-frequency input signal to an intermediate frequency, or IF. The great advantage of the superheterodyne is that the IF frequency is constant for the entire range of frequencies of the input signal. This makes it easy to amplify the signal and control its bandwidth; hence most of the amplification is done at the IF stage. The IF signal is, like the high-frequency input signal, modulated by the information signal, which is usually low frequency. Traditionally, the IF output is demodulated and low-pass filtered to provide the information signal.

Modern digital communication receivers often rely on digital processing of the information. At the time of writing, there is a growing interest in *direct IF sampling*, as an alternative to the procedure of demodulation followed by low-pass filtering and sampling. Direct IF sampling has the potential of eliminating costly and space-consuming analog hardware. For example, suppose that the IF frequency is 1 MHz and the information bandwidth is 38.4 kHz. In this case we have

$$L = \lfloor 1019.2/38.4 \rfloor = 26,$$

and the sampling frequency is

$$f_{\text{sam}} = \frac{2 \times 1019.2}{26} = 78.4\,\text{kHz}.$$

The sample-and-hold time (or the conversion time for a flash A/D) should be about 0.1 microsecond or less in this case. □

When sampling a band-pass signal, there is seldom a need to reconstruct it in the pass band (usually either there is no need for reconstruction at all, or reconstruction is needed in the base band). If band-pass reconstruction is needed, it can be performed by passing the sampled signal through an ideal band-pass filter with frequency response

$$H^{\text{F}}(\omega) = \begin{cases} T, & \omega_0 \le |\omega| \le \omega_2, \\ 0, & \text{otherwise,} \end{cases} \tag{3.56}$$

see Figure 3.28. The computation of the corresponding impulse response $h(t)$ is discussed in Problem 3.34.

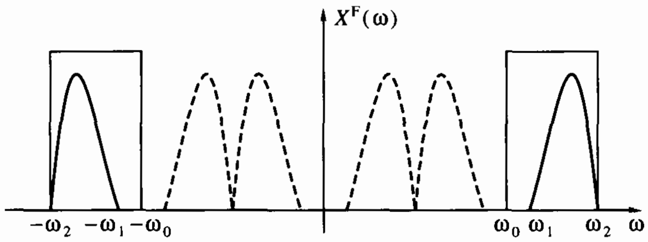

Figure 3.28 Reconstruction of a band-pass signal by an ideal band-pass filter (replicas shown by dashed lines will be eliminated by the filter).

Instead of extending the bandwidth on the left, we can also extend it on the right. You are asked to carry out the procedure and compare the resulting sampling rate with the one given in (3.55) (see Problem 3.32).

3.7 Sampling of Random Signals*

When a continuous-time random signal is sampled, the result is a discrete-time random signal. Our interest here is to learn how the parameters of the signal—mean, variance, covariance function, and PSD—are affected by sampling.

Let $x(t)$ be a continuous-time WSS random signal and $x[n] = x(nT)$ its point sampling. The mean and variance of the sampled signal are not affected by sampling, since

$$E(x[n]) = E(x(nT)) = \mu_x, \tag{3.57a}$$

$$E\{(x[n] - \mu_x)^2\} = E\{(x(nT) - \mu_x)^2\} = \gamma_x. \tag{3.57b}$$

The covariance sequence of $x[n]$ is derived as follows:

$$\begin{aligned} \kappa_x[m] &= E\{(x[n+m] - \mu_x)(x[n] - \mu_x)\} \\ &= E\{[x(nT + mT) - \mu_x][x(nT) - \mu_x]\} = \kappa_x(mT). \end{aligned} \tag{3.58}$$

We conclude that:

1. Sampling of a WSS random signal yields a discrete-time WSS random signal.
2. The covariance sequence of the sampled signal is obtained by sampling the covariance function of the continuous-time signal at the same sampling interval.

The sampling theorem relations (3.10) and (3.58) immediately imply:

Theorem 3.3 The PSD of a sampled WSS signal $x[n]$ is related to that of the continuous-time WSS signal $x(t)$ by

$$K_x^{\mathrm{f}}(\theta) = \frac{1}{T}\sum_{k=-\infty}^{\infty} K_x^{\mathrm{F}}\left(\frac{\theta - 2\pi k}{T}\right). \tag{3.59}$$

□

Formula (3.59) can be regarded as a sampling theorem for random WSS signals. It implies that Nyquist's no-aliasing condition, as well as the aliasing phenomenon in case Nyquist's condition is violated, apply to the power spectra of random signals in the same manner as they apply to the Fourier transforms of nonrandom signals. One difference should be borne in mind, however. Whereas the right side of (3.10) is, in general, an infinite sum of complex-valued functions, the right side of (3.59) is a sum of real-valued nonnegative functions. The sampling theorem for random signals applies to the PSDs of the signals, not to their Fourier transforms.

White noise represents an exception to the sampling formulas (3.58) and (3.59). White noise cannot be point sampled. This is hardly surprising since, as we saw in Section 2.6, continuous-time white noise has infinite variance and does not exist as a physical entity. On the other hand, band-limited white noise has finite variance and its point sampling is well defined. The covariance and spectral characteristics of sampled band-limited white noise depend on the sampling rate relative to the noise bandwidth. An important special case occurs when the signal is sampled exactly at the Nyquist rate. In this case, the sampled signal becomes a discrete-time white noise. This can be seen either from (2.78) in the covariance domain, or from (2.77) in the frequency domain. The variance of the discrete-time white noise is then N_0/T, or $N_0\omega_{\mathrm{m}}/\pi$. This result is of importance when sampling signals accompanied by noise. Often the noise has large bandwidth, sometimes much larger than the signal bandwidth. If the sampling rate is chosen according to the signal bandwidth, the noise will be aliased and its variance may increase considerably after sampling, due to the summation in (3.59). In such cases it expedient to insert an antialiasing filter before the sampler. The antialiasing filter limits the bandwidth of the noise before the sampler. As a result, the sampled signal will be accompanied by discrete-time white noise whose variance is the smallest possible, since aliasing is prevented by the antialiasing filter. The following example illustrates such a case.

Example 3.13 Binary phase-shift keying (BPSK) is one of the simplest methods for transmitting digital information.[6] Suppose we are given a sequence of bits $b[n]$, appearing every T seconds. A non-return to zero (NRZ) signal $x(t)$ for this sequence is defined as

$$x(t) = \begin{cases} 1, & b[n] = 0, \\ -1, & b[n] = 1, \end{cases} \quad \text{for} \quad nT \le t < (n+1)T. \tag{3.60}$$

A typical waveform of an NRZ signal is shown in Figure 3.29(a). When such a waveform is used for modulating a sinusoidal carrier wave, the resulting high-frequency wave has the form

$$m(t) = x(t)\cos(\omega_0 t). \tag{3.61}$$

The signal $m(t)$ is a BPSK signal; it has phase 0° whenever the bit is 0 and phase 180° whenever the bit is 1; hence the name *binary phase-shift keying*. However, here we are interested in the signal $x(t)$, rather than in the modulated signal $m(t)$.

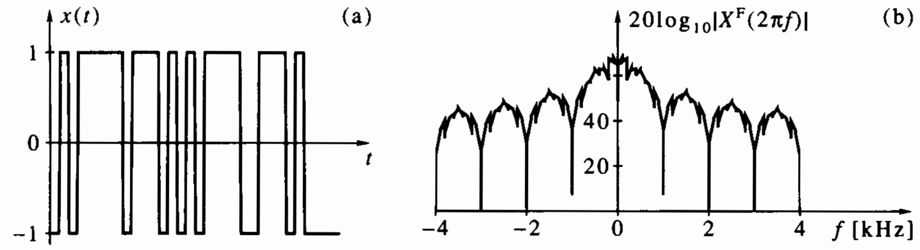

Figure 3.29 An NRZ signal: (a) waveform; (b) magnitude spectrum.

Figure 3.29(b) shows a typical magnitude spectrum of an NRZ signal. Here the bits appear at a rate of 1000 per second and the spectrum is shown in the range ±4 kHz. As we see, the magnitude decays rather slowly as the frequency increases. This behavior of the spectrum is problematic because communication systems usually require narrowing the spectrum as much as possible for a given bit rate.

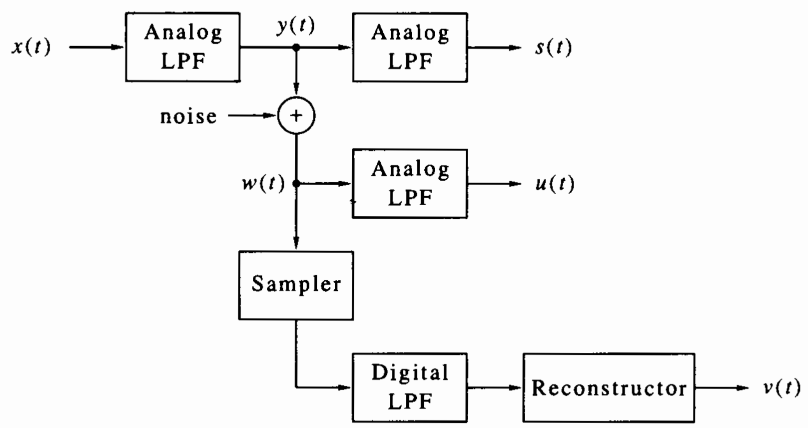

Figure 3.30 Block diagram of the system in Example 3.13.

A common remedy is to pass the NRZ signal through a low-pass filter (LPF) before it is sent to the modulator, as shown in the top part of Figure 3.30. The bandwidth of the filter is on the order of the bit rate, in our case 1 kHz. Figure 3.31(a) shows the result of passing the signal $x(t)$ through such a filter [denoted by $y(t)$], and Figure 3.31(b) shows the corresponding spectrum. As we see, the magnitude now decays much more rapidly as the frequency increases.

The received high-frequency signal is demodulated and then passed through another low-pass filter, as shown in the top part of Figure 3.30. This new filter is often identical or similar to the filter used in the transmitter. The signal at the output of this filter, denoted by $s(t)$, is used for detecting the transmitted bits.

In practice, the communication channel adds noise to the signal, resulting in a new signal $w(t)$, as shown in the middle part of Figure 3.30. When $w(t)$ is passed through the analog low-pass filter, it yields a signal $u(t)$ different from $s(t)$.

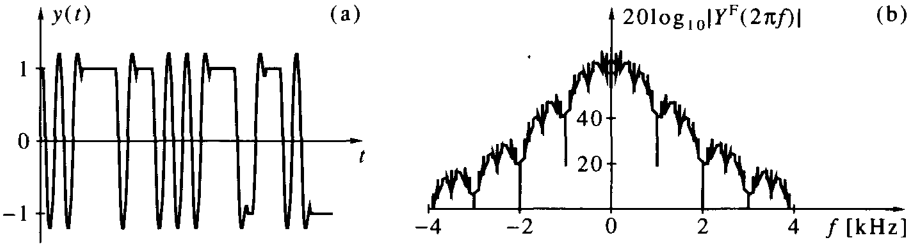

Figure 3.31 A filtered NRZ signal: (a) waveform; (b) magnitude spectrum.

To get to the point we wish to illustrate, let us assume that an engineer who was given the task of implementing the filter at the receiver decided to replace the traditional analog filter by a digital filter, as shown in the bottom part of Figure 3.30. The engineer decided to sample the demodulated signal at a rate of 8 kHz, judged to be more than enough for subsequent digital low-pass filtering to 1 kHz. The engineer also built the traditional analog filter for comparison and tested the two filters on real data provided by the receiver. The waveform $u(t)$ of the analog filter output is shown in Figure 3.32(a); the theoretical signal $s(t)$ is shown for comparison (dotted line). As we see, the real-life waveform at the output of the analog filter is quite similar to the theoretical one. However, the reconstructed output of the digital filter $v(t)$ was found to be distorted, as shown in Figure 3.32(b).

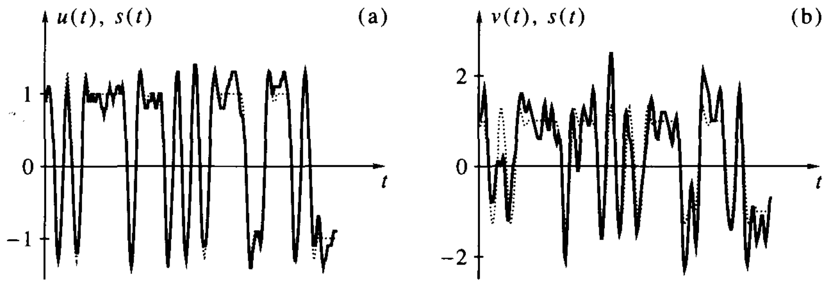

Figure 3.32 A received BPSK signal: (a) analog filtering; (b) digital filtering.

To find the source of the problem, the engineer recorded the waveform and the spectrum of the input signal to the two filters $w(t)$. Figure 3.33 shows the result. Contrary to the simulated signal, which is synthetic and smooth, the real-life signal is noisy. The noise has large bandwidth (about 100 kHz in this example), of which only a small part is shown in the figure. The analog filter attenuates most of this noise, retaining only the noise energy within ± 1 kHz. This is why the signals $u(t)$ and $s(t)$ are similar. On the other hand, the noise energy is aliased in the sampling process, and appears to the digital filter as energy in the range ± 4 kHz. The digital filter removes about 75 percent of this energy (the part outside 1 kHz), but the remaining 25 percent is enough to create the distortion seen in Figure 3.32(b).

The lesson of this example is that an analog antialiasing filter should have been inserted prior to the sampler, to remove the noise at frequencies higher than 4 kHz. With such a filter, the two systems would have performed approximately the same. □

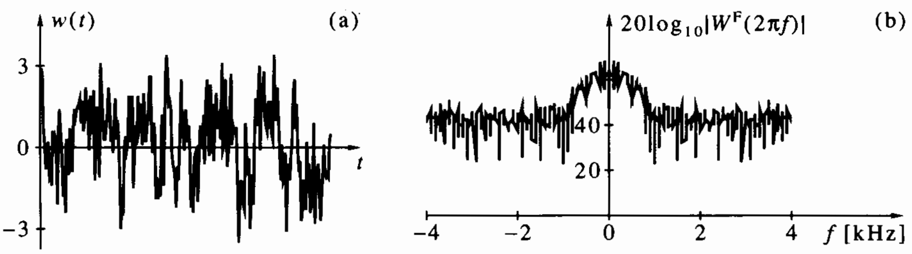

Figure 3.33 A noisy BPSK signal: (a) waveform; (b) magnitude spectrum.

3.8 Sampling in the Frequency Domain*

Suppose we are given a continuous-time signal $x(t)$ with a Fourier transform $X^{\mathrm{F}}(\omega)$, and we wish to sample its Fourier transform, that is, to let

$$X^{\mathrm{F}}[k] = X^{\mathrm{F}}(k\Omega), \quad k \in \mathbb{Z},$$

for some $\Omega > 0$. How is the time signal corresponding to $X^{\mathrm{F}}[k]$ related to $x(t)$, and under what conditions can we reconstruct $X^{\mathrm{F}}(\omega)$ from its samples? We can answer these questions by invoking duality of the Fourier transform. Define $Y^{\mathrm{F}}(\omega)$ as the impulse sampling of $X^{\mathrm{F}}(\omega)$, that is,

$$Y^{\mathrm{F}}(\omega) = \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}(k\Omega)\delta(\omega - k\Omega). \tag{3.62}$$

Then we have, by the dual of the sampling formula (3.9),

$$y(t) = \frac{1}{\Omega}\sum_{n=-\infty}^{\infty} x\left(t - \frac{2\pi n}{\Omega}\right). \tag{3.63}$$

As a result of the infinite summation, the signal corresponding to the sampled Fourier transform is periodic in t with period $2\pi/\Omega$. We therefore say that *sampling in the frequency domain gives rise to periodicity in the time domain.*

Formula (3.63) implies that $y(t)$ is *time aliased*; it is the signal obtained when shifting $x(t)$ by all integer multiples of $2\pi/\Omega$ and summing. In particular, if $x(t)$ is *time limited* to a duration not greater than $2\pi/\Omega$, then it can be recovered unambiguously from $y(t)$ because then the shifted replicas $x(t - 2\pi n/\Omega)$ do not overlap. The support of $x(t)$ does not have to be symmetric around $t = 0$; only its total duration matters.

When $x(t)$ is time limited to a duration $2\pi/\Omega$, we can recover $X^{\mathrm{F}}(\omega)$ from its samples using the dual of Shannon's reconstruction formula

$$X^{\mathrm{F}}(\omega) = \sum_{k=-\infty}^{\infty} X^{\mathrm{F}}(k\Omega)\operatorname{sinc}\left(\frac{\omega - k\Omega}{\Omega}\right). \tag{3.64}$$

Example 3.14 Suppose we wish to measure the impulse response of an electronic amplifier. Impulse response is difficult to measure directly because it is impossible to input an ideal impulse (i.e., a delta function) to a physical device. A convenient alternative is to measure the frequency response of the amplifier, and obtain the impulse response by inverse Fourier transform. Frequency response can be measured by feeding the amplifier with a sinusoidal input, and measuring the gain and phase shift of the output sinusoid. By repeating this process for a finite number of discrete frequencies, we get a sampling of the Fourier transform of the amplifier.

Suppose it is known a priori that the effective duration of the impulse response is no longer than 10 microseconds, and we are interested in a time resolution of 0.1

microsecond. How should we design the test procedure? By what we have said, the frequency increment should be 100 kHz or less (the reciprocal of the maximum duration) to avoid time aliasing. To meet the time resolution requirement, we should use frequencies up to 10 MHz (the reciprocal of the time resolution). In summary, the frequency response should be measured for the frequencies

$$\omega[k] = k\Omega = 2\pi k \cdot 10^5, \quad 0 \le k \le 100. \qquad \square$$

3.9 Summary and Complements

3.9.1 Summary

In this chapter we introduced the mathematical theory of sampling and reconstruction, as well as certain physical aspects of these two operations. The basic concept is point sampling of a continuous-time signal, which amounts to forming a sequence of regularly spaced time points (T seconds apart) and picking the signal values at these time points. An equivalent mathematical description of this operation is impulse sampling, which amounts to multiplying the continuous-time signal by an impulse train.

A fundamental result in sampling theory is the sampling theorem (3.10), which expresses the Fourier transform of a sampled signal as a function of the continuous-time signal. As the theorem shows, sampling leads to periodic replication of the Fourier transform. A major consequence of this replication is *aliasing*: The spectral shape of the sampled signal is distorted by high-frequency components, disguised as low-frequency components. The physical implication of aliasing is that high-frequency information is lost and low-frequency information is distorted.

An exception to the aliasing phenomenon occurs when the continuous-time signal is band limited and the sampling rate is at least twice the bandwidth. In such a case there is no aliasing: The Fourier transforms of the continuous-time signal and the sampled signal are equal up to a proportionality constant. This implies that all the information in the continuous-time signal is preserved in the sampled signal.

To prevent aliasing or to minimize its adverse effects, it is recommended to low-pass filter the continuous-time signal prior to sampling, such that the relative energy at frequencies above half the sampling rate will be zero or negligible. Such a low-pass filter is called an antialiasing filter.

A second fundamental result in sampling theory is the reconstruction theorem (3.23), which expresses the continuous-time signal as a function of the sampled signal values, provided that the signal has not been aliased during sampling. Ideal reconstruction is performed by an ideal low-pass filter whose cutoff frequency is half the sampling rate. Such an operation is physically impossible, so approximate reconstruction schemes are required. The simplest and most common approximate reconstructor is the zero-order hold (3.33). The zero-order hold has certain undesirable effects—nonuniform gain in the low-frequency range and nonzero gain in the high range. Both effects can be mitigated by an appropriate low-pass analog filter at the output of the zero-order hold.

We have devoted part of the chapter to physical circuits for sampling and reconstruction: analog-to-digital (A/D) and digital-to-analog (D/A) converters. Physical devices have certain undesirable effects on the signal, which cannot be completely avoided. The most prominent effect is quantization of the signal value to a finite number of bits. Other effects are smearing (or averaging), delays, and various nonlinearities.

Many applications involve band-pass or modulated signals. Such signals can often be sampled at a rate much lower than twice the highest frequency without being subject to aliasing. The lowest sampling rate at which a band-pass signal can be sampled without aliasing is typically slightly larger than twice its net bandwidth. Reconstruction of a sampled band-pass signal can be performed by an analog band-pass filter.

The last topic presented in this chapter is sampling in the frequency domain. The mathematical properties of sampling in the frequency domain are dual to those obtained for time-domain sampling. Sampling in the frequency domain leads to time aliasing in general. Time aliasing can be avoided if the signal has finite duration and the frequency sampling interval is smaller than the inverse of the duration.

3.9.2 Complements

1. [p. 48] Nyquist [1928] is credited with calling the attention of the engineering community to the lower limit $2f_{\mathrm{m}}$ on the sampling frequency. Shannon [1949] provided a rigorous proof of the sampling theorem and the reconstruction formula (3.23). However, Whittaker [1915] preceded both, although his work was apparently unknown to electrical engineers of his time.
2. [p. 52] A linear time-invariant communication channel whose impulse response is a Nyquist-T signal can be used to transmit digital communication signals at a symbol rate $1/T$ without intersymbol interference. The channel whose impulse response is $h(t) = \mathrm{sinc}(t/T)$ has the minimum bandwidth of all such channels. This property was discovered by Nyquist. The raised-cosine channel, whose impulse response is given by (3.22), also has the Nyquist-T property, but its bandwidth is larger than that of the sinc channel (for the same T). See Haykin [1994, Sec. 7.5] for a detailed discussion of this subject.
3. [p. 56] This property of the Fourier transform follows from the following argument. A finite duration signal $x(t)$ that has a Fourier transform $X^{\mathrm{F}}(\omega)$ also possesses an analytic Laplace transform $X^{\mathrm{L}}(s)$ on the entire complex plane. A theorem in complex function theory states that a function analytic in a domain has a countable number of zeros at most in the domain [Markushevich, 1977, Sec. 82, Theorem 17.3]. In particular, $X^{\mathrm{F}}(\omega)$ has no more than a countable number of zeros on $-\infty < \omega < \infty$. Therefore, $X^{\mathrm{F}}(\omega)$ cannot be band limited.
4. [p. 64] Sometimes a sample-and-hold circuit is inserted at the output of the D/A converter (see Section 3.5.2 for an explanation of the operation of this device). The sample-and-hold device helps overcome transients caused by unequal switching times of the D/A switches; it is particularly useful when the D/A operates at a high speed.
5. [p. 64] Two's-complement *integer* arithmetic is defined as follows: The range of numbers representable by B bits is from -2^{B-1} to $2^{B-1} - 1$. If x is a number in this range, it is represented by the positive number $x \bmod 2^B$, expressed as a B-bit binary number. Note that the MSB of positive numbers is 0 and that of negative numbers is 1. Two's-complement *fractional* arithmetic is defined similarly, except that the range of numbers is from -1 to $1-2^{-(B-1)}$, and a number x in this range is represented by the positive integer $(2^B x) \bmod 2^B$.
6. [p. 75] For further reading on binary phase-shift keying see, for example, Haykin [1994, Sec. 8.11].

3.10 Problems

3.1 We are given the signal

$$x[n] = \begin{cases} (-1)^m, & n = 2m, \\ 0, & n = 2m+1. \end{cases}$$

Specify two possible continuous-time signals $x(t)$ from which $x[n]$ could have been obtained by sampling, one in which $x[n]$ is not aliased, and one in which it is aliased. Specify the sampling interval T in each case.

3.2 The signal

$$x(t) = e^{-0.02t^2}\operatorname{sinc}(t)$$

was sampled at interval T. It was then found that the Fourier transform of the sampled signal is

$$X^{\mathrm{f}}(\theta) = 1.$$

What is the minimum T for which such a result is possible? If this is impossible for any T, explain why.

3.3 The signal $x(t)$ has the Fourier transform

$$X^{\mathrm{F}}(\omega) = \frac{\pi}{\omega_{\mathrm{m}}}\left[1 + \cos\left(\frac{\pi\omega}{\omega_{\mathrm{m}}}\right)\right], \quad |\omega| \le \omega_{\mathrm{m}}.$$

The signal is sampled at interval $T = 2\pi/\omega_{\mathrm{m}}$. Find the sampled signal $x[n]$.

3.4 Let $x(t)$ be a band-limited signal, whose Fourier transform $X^{\mathrm{F}}(\omega)$ is identically zero outside the interval $[-1.5\omega_0, 1.5\omega_0]$. Define $T = 2\pi/\omega_0$ and let

$$y(t) = \sum_{n=-\infty}^{\infty} x(t - nT).$$

(a) Prove that $y(t)$ has the form

$$y(t) = C_0 + C_1\cos(\omega_0 t + \phi_0),$$

where C_0, C_1, ϕ_0 are constants.

(b) Express C_0, C_1, ϕ_0 it terms of $X^{\mathrm{F}}(\omega)$.

3.5 Prove the following properties of a band-limited signal sampled at or above the Nyquist rate:

$$\text{Area conservation:} \quad \int_{-\infty}^{\infty} x(t)dt = T\sum_{n=-\infty}^{\infty} x(nT), \tag{3.65}$$

$$\text{Energy conservation:} \quad \int_{-\infty}^{\infty} |x(t)|^2 dt = T\sum_{n=-\infty}^{\infty} |x(nT)|^2. \tag{3.66}$$

3.6 Repeat Example 3.2 for the signal

$$x(t) = e^{-\alpha|t|}, \quad \alpha > 0, \quad -\infty < t < \infty,$$

and establish an infinite sum formula similar to (3.17) for this case.

3.7 Let $x(t)$ be a periodic signal with period T_0 and let $X^{\mathrm{S}}[k]$ be its Fourier series coefficients. Find a general expression for $X^{\mathrm{f}}(\theta)$, the Fourier transform of the sampled signal $x(nT)$, in terms of the $X^{\mathrm{S}}[k]$. Hint: Remember that $X^{\mathrm{f}}(\theta)$ must be defined first on $\theta \in [-\pi, \pi)$ and then extended periodically.

3.8 Let $x(t)$ be a continuous-time complex periodic signal with period T_0. The signal is band limited, such that its Fourier series coefficients $X^S[k]$ vanish for $|k| > 3$.

(a) The signal is sampled at interval $T = T_0/N$, where N is integer. What is the minimum N that meets Nyquist's condition?

(b) With this value of N, what is the minimum number of samples from which $X^S[k]$ can be computed? Explain how to perform the computation.

(c) Instead of sampling as in part a, we sample at $T = T_0/5.5$. Plot the Fourier transform of the point-sampled signal as a function of θ. Is it possible to compute the $X^S[k]$ in this case? If so, what is the minimum number of samples and how can the computation be performed?

3.9 A continuous-time signal $x(t)$ is passed through a filter with impulse response $h(t)$, and then sampled at interval T; see Figure 3.34(a). The signal is band limited to $\pm\omega_1$, and the frequency response of the filter is band limited to $\pm\omega_2$. We wish to change the order of the operations: Sample the signal first and then pass the sampled signal through a digital filter; see Figure 3.34(b). We require that:

- the impulse response of the digital filter be $Th(nT)$;
- the outputs of the two systems be equal for any input signal $x(t)$ that meets the bandwidth restriction.

What is the condition on the sampling interval T to meet this requirement?

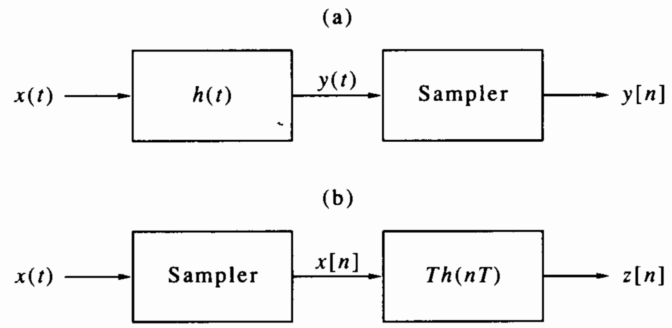

Figure 3.34 Pertaining to Problem 3.9.

3.10 Consider a continuous-time signal whose Fourier transform is

$$Y^F(\omega) = \begin{cases} \dfrac{1}{f_0}, & |\omega| \le \omega_1, \\ \dfrac{1}{f_0}\cos\left(\dfrac{|\omega| - \omega_1}{4\alpha f_0}\right), & \omega_1 < |\omega| \le \omega_2, \\ 0, & |\omega| > \omega_2, \end{cases} \tag{3.67}$$

where ω_1 and ω_2 are defined in (3.21).

(a) Compute the signal $y(t)$. Take care to compute separately for $t = 0$, for $t = \pm 1/(4\alpha f_0)$, and for all other t.

(b) Assume that the signal is sampled at interval $T = 1/f_0$. Is the sampled signal alias free?

(c) With the same sampling interval, is the signal Nyquist-T (as defined in Example 3.9)?

(d) Repeat parts b and c for the signal

$$x(t) = \{y * y\}(t).$$

Hint: This problem is related to Example 3.9.

3.11 The signal $x(t)$ has the Fourier transform

$$X^{\mathrm{F}}(\omega) = \begin{cases} 1 - \frac{|\omega|T}{\pi}, & |\omega| \le \frac{\pi}{T}, \\ 0, & \text{otherwise.} \end{cases}$$

Let $y(t)$ be the impulse sampling of $x(t)$ using the alternating impulse train $q_T(t)$ defined in Problem 2.24, that is,

$$y(t) = x(t)q_T(t).$$

Compute $Y^{\mathrm{F}}(\omega)$ as a function of $X^{\mathrm{F}}(\omega)$. Draw the shape of $Y^{\mathrm{F}}(\omega)$.

3.12 Let $x(t)$ be a periodic square wave with period T_0, that is,

$$x(t) = \begin{cases} 1, & nT_0 \le t < (n+0.5)T_0, \\ -1, & (n+0.5)T_0 \le t < (n+1)T_0, \end{cases} \quad \text{for} \quad -\infty < n < \infty.$$

(a) Show that the sampled signal $x(nT)$ is always aliased, regardless of the choice of T.

(b) We wish to sample $x(t)$ after passing it through an antialiasing filter. Let $y(t)$ be the output of the antialiasing filter. We require that (1) $y(nT)$ be nonaliased with respect to $y(t)$, (2) $y(t)$ contain at least 99 percent of the energy of $x(t)$, and (3) the frequency responses of $y(t)$ and $x(t)$ be identical in the range $\pm\pi/T$. What sampling interval T will meet these requirements, and what is the cutoff frequency of the antialiasing filter? Hint: Use Parseval's relationship for the Fourier series of $x(t)$.

3.13 We are given the continuous-time signal

$$x(t) = a\sin(\omega_0 t) + b\cos(2\omega_0 t).$$

The signal is sampled by impulse sampling at interval $T = 0.5\pi/\omega_0$.

(a) Show that the sampled signal $x_{\mathrm{p}}(t)$ is periodic, find its period and compute its Fourier series coefficients $X_{\mathrm{p}}^{\mathrm{S}}[k]$.

(b) Suppose that a malfunction causes all samples

$$\{x(3mT),\ -\infty < m < \infty\}$$

to be lost and their values to be replaced by zero. Let $y(t)$ denote the resulting impulse-sampled signal. The signal $y(t)$ is passed through an ideal low-pass filter with frequency response

$$H^{\mathrm{F}}(\omega) = T\,\mathrm{rect}\left(\frac{\omega}{3\omega_0}\right).$$

Find the signal $z(t)$ at the output of the filter.

3.14 Let

$$x(t) = \frac{\beta}{\beta^2 + t^2}, \quad \beta > 0, \quad -\infty < t < \infty.$$

The signal is sampled at interval T. Find $X^{\mathrm{f}}(0)$ for the sampled signal $x[n]$. Hint: Find the Fourier transform of the signal $y(t) = 1/(\beta - jt)$ first, by duality to the signal $x_1(t)$ in Problem 2.3.

3.15 Let $x(t)$ be the continuous-time signal

$$x(t) = \begin{cases} 1, & |t| \le (N + 0.5)T, \\ 0, & \text{otherwise}, \end{cases}$$

where N is a positive integer.

(a) Find $X^{\mathrm{f}}(\theta)$, the Fourier transform of the sampled signal $x(nT)$.

(b) Use the result of part a to derive a closed-form expression for the infinite sum

$$\sum_{k=-\infty}^{\infty} \operatorname{sinc}\left[\frac{(N + 0.5)(\omega T - 2\pi k)}{\pi}\right].$$

3.16 We are given the signal $x(t)$ whose Fourier transform is

$$X^{\mathrm{F}}(\omega) = \begin{cases} \cos\left(\frac{\pi\omega}{\omega_0}\right), & |\omega| \le \omega_0, \\ 0, & |\omega| > \omega_0. \end{cases}$$

Testing of the signal has revealed that an unwanted noise has been added to the signal in the frequency range $0.5\omega_0 < |\omega| \le \omega_0$. It was therefore decided to disregard the contents of the signal in this frequency range and to use only the information in the range $|\omega| \le 0.5\omega_0$.

(a) What is the minimum sampling frequency ω_{sam} for which there will be no aliasing in the range $|\omega| \le 0.5\omega_0$? Assume that antialiasing filtering before sampling is not permitted.

(b) After sampling at the rate found in part a, the signal is reconstructed using an ideal low-pass filter with cutoff frequency $0.5\omega_0$. Give an expression for the signal $\hat{x}(t)$ at the output of the reconstructor.

3.17 Figure 3.35 shows one period of a continuous-time periodic signal $x(t)$ with period $T_0 = 6$. The signal is sampled at interval $T = 1$, resulting in the discrete-time signal $x[n]$. The discrete-time signal is then reconstructed using the ideal reconstructor (3.23), resulting in the continuous-time signal $\hat{x}(t)$.

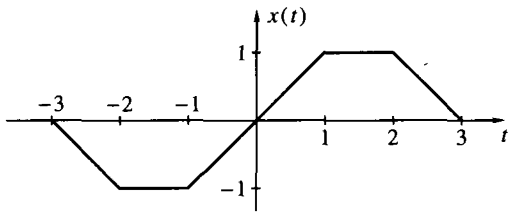

Figure 3.35 Pertaining to Problem 3.17.

(a) Is $\hat{x}(t)$ periodic?

(b) Is $\hat{x}(t)$ equal to $x(t)$? Answer without performing any calculations.

(c) Compute $\hat{x}(t)$.

3.18 The signal $x(t)$ is periodic, with period T_0. We are given that its Fourier series coefficients $X^{\mathrm{S}}[k]$ are zero for $|k| > K$ for some given K. Let $N = 2K + 1$. The signal is sampled at interval $T = T_0/N$. Prove that $x(t)$ can be expressed in terms of its samples over one period, $\{x(nT),\ 0 \le n \le N-1\}$, as follows:

$$x(t) = \frac{1}{N}\sum_{n=0}^{N-1} x(nT) D\left(\frac{t}{T_0} - \frac{n}{N}, N\right), \tag{3.68}$$

where

$$D(a,N) = \frac{\sin(0.5Na)}{\sin(0.5a)}. \tag{3.69}$$

Hint: Use Shannon's interpolation formula (3.23) first and then (2.157).

3.19 The signal $x(t)$ is sampled at the instants

$$t = nT,\ nT + a, \quad n \in \mathbb{Z},$$

where a is a constant, $0 < a < T$. Note that this is a nonuniform sampling. Let $q(t)$ be an impulse train consisting of Dirac delta functions at the sampling instants. Let

$$x_q(t) = x(t)q(t),$$

that is, $x_q(t)$ is the impulse sampling of $x(t)$.

(a) Write an explicit expression for $q(t)$.

(b) Express $X_q^F(\omega)$ as a function of $X^F(\omega)$. Bring the result to the form

$$X_q^F(\omega) = \frac{1}{T}\sum_{k=-\infty}^{\infty} A_k X^F(\omega - k\omega_{sam}).$$

Specify the values of A_k and ω_{sam}.

(c) Let $X^F(\omega)$ be given by

$$X^F(\omega) = \begin{cases} 1 - \dfrac{2|\omega|}{\omega_{sam}}, & -0.5\omega_{sam} \le \omega \le 0.5\omega_{sam}, \\ 0, & \text{otherwise,} \end{cases}$$

where ω_{sam} is as found in part b. Let $a = 0.25T$. Draw, on separate plots, the real and imaginary parts of $X_q^F(\omega)$ in the range $-2.5\omega_{sam} \le \omega \le 2.5\omega_{sam}$.

(d) Under the conditions of part c, what reconstruction filter is needed to reconstruct $x(t)$ from $x_q(t)$?

3.20 We are given a band-limited signal $x(t)$, whose Fourier transform is nonzero only for $|f| \le 8$ kHz. The signal is sampled at interval T, resulting in the discrete-time signal $x[n]$. The signal $x[n]$ is passed through a nonlinear system, resulting in the discrete-time signal

$$y[n] = (x[n])^3 + 0.5x[n].$$

The signal $y[n]$ is reconstructed by an ideal low-pass filter. What is the minimal sampling rate $1/T$ that will ensure that the reconstructed signal will be

$$\hat{y}(t) = [x(t)]^3 + 0.5x(t)$$

up to a scale factor?

3.21 Recall Problem 3.7 and suppose we reconstruct $\hat{x}(t)$ from $x(nT)$ using Shannon's formula (3.23).

(a) Show that if T/T_0 is not a rational number, the reconstructed signal $\hat{x}(t)$ is not periodic.

(b) Show that if T/T_0 is rational, that is,

$$\frac{T}{T_0} = \frac{p}{q},$$

where p, q are coprime integers, then $\hat{x}(t)$ is periodic. Show that, in this case, $\hat{x}(t)$ is a finite sum of sinusoidal signals. Find the period of $\hat{x}(t)$ as a function of T, p, and q.

3.22 Let $x(t)$ be the signal

$$x(t) = 3\cos(100\pi t) + 2\sin(250\pi t).$$

(a) The signal is sampled at an interval $T_1 = 0.0025$ second, yielding the discrete-time signal $x[n]$. Now we reconstruct $\hat{x}(t)$ from $x[n]$ using an ideal reconstructor corresponding to $T_2 = 0.005$ second, that is,

$$\hat{x}(t) = \sum_{n=-\infty}^{\infty} x[n]\mathrm{sinc}\left(\frac{t - nT_2}{T_2}\right).$$

Give an explicit expression for $\hat{x}(t)$.

(b) Now let $T_1 = 0.005$ second and $T_2 = 0.0025$ second, and repeat part a.

3.23 Let $x(t)$ be band limited to $\pm\omega_{\mathrm{m}}$ and let $T = \pi/\omega_{\mathrm{m}}$. Define

$$y(t) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t - 2nT).$$

(a) Find the Fourier transform of $y(t)$ as a function of $X^{\mathrm{F}}(\omega)$. Hint: Write $y(t)$ as an impulse sampling of a signal related to $x(t)$, then use property (2.7) of the Fourier transform.

(b) Is the Fourier transform of $y(t)$ aliased? Explain both mathematically and based on physical reasoning.

3.24 Let $x(t)$ be band limited to $\pm\omega_{\mathrm{m}}$ and let $T = \pi/\omega_{\mathrm{m}}$. The signal $x(t)$ is sampled at interval T and then reconstructed according to the formula

$$z(t) = \sum_{n=-\infty}^{\infty} x(nT)\mathrm{sinc}\left(\frac{t - 2nT}{T}\right)$$

(note the factor 2 in the numerator of the argument of the sinc).

(a) Express $Z^{\mathrm{F}}(\omega)$ in terms of $X^{\mathrm{F}}(\omega)$ and plot the result. Hint: Express $z(t)$ as a convolution with a certain impulse-sampled signal and use the solution to Problem 3.23; then pass to the frequency domain.

(b) Give a closed-form expression for $z(t)$ if

$$x(t) = \cos(0.6\omega_{\mathrm{m}}t).$$

(c) Show that the operation in part a can be also performed as follows:

- Form a new sequence $u[n]$ by inserting zeros between adjacent points of $x(nT)$, that is, define

$$u[n] = \begin{cases} x(0.5nT), & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases}$$

- Reconstruct $z(t)$ from $u[n]$ by Shannon's formula (3.23).

3.25 The discrete-time signal

$$x[n] = \cos\left(\frac{2\pi n}{16}\right)$$

is reconstructed by a zero-order hold with $T = 1$, to give the continuous-time signal $y(t)$. This signal is sampled in the following manner:

$$z[n] = y((2n + 0.5)T).$$

Find an expression for $z[n]$.

3.26 The signal $x(t)$ has the Fourier transform

$$X^{\mathrm{F}}(\omega) = \begin{cases} 1, & 0.25\pi \le |\omega| \le 1.25\pi, \\ 0, & \text{otherwise.} \end{cases}$$

The signal is sampled at interval $T = 1$ to give the discrete-time signal $x[n]$. The signal $x[n]$ is given to a Shannon reconstructor (with $T = 1$), yielding the continuous-time signal $y(t)$. Compute $y(t)$.

3.27 The signal

$$x(t) = \cos^3(0.4\pi t)$$

is sampled at interval $T = 1$ (point sampling) and then reconstructed by a Shannon reconstructor that uses $T = 0.5$. Compute the reconstructed signal.

3.28 The signal

$$x(t) = \cos(0.9\pi t)\cos(0.4\pi t)$$

is sampled at interval $T = 1$ second, yielding the discrete-time signal $x[n]$. The discrete-time signal is passed through an ideal high-pass filter, which eliminates all frequencies below 0.6π. The signal at the output of the filter, $y[n]$, is reconstructed by a Shannon reconstructor having $T = 1$ second. Find an expression for the signal $y(t)$ at the output of the reconstructor.

3.29 A *first-order hold* is a reconstructor that works as follows: At time $t = nT$ it computes the straight line connecting $(nT - T, x[nT - T])$ and $(nT, x[nT])$. During the interval $[nT, nT + T)$ it takes $\hat{x}(t)$ as the ordinate of the point on the straight line whose abscissa is t, see Figure 3.36.

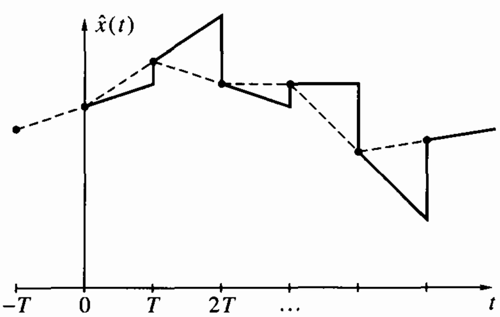

Figure 3.36 Pertaining to Problem 3.29.

(a) Compute and plot the impulse response of the first-order hold.

(b) Compute and plot the frequency response of the first-order hold (magnitude and phase).

(c) Compare the first-order hold with the zero-order hold and draw conclusions.

3.30 Let $x[n]$ be a discrete-time signal. Suppose that, because of a malfunction, all odd-numbered values are replaced by zero, and we get the signal

$$y[n] = \begin{cases} x[n], & n \text{ even}, \\ 0, & n \text{ odd}. \end{cases}$$

It is suggested to reconstruct $x[n]$ by a digital ZOH, which is defined by

$$\hat{x}[2m] = \hat{x}[2m+1] = y[2m], \quad m \in \mathbb{Z}.$$

Show that the digital ZOH operation is equivalent to passing $y[n]$ through a digital linear time-invariant filter. Compute the magnitude and phase responses of this filter.

3.31 Consider the following reconstructor (note that it is not causal):

- Between every two points of the sampled signal, $x(nT)$ and $x(nT+T)$, we compute a midsample by linear interpolation, that is,

$$\hat{x}(nT + 0.5T) = 0.5[x(nT) + x(nT + T)].$$

- We interleave x and $\hat{x}$ in their order of their appearance; that is, we form the discrete-time signal

$$y[n] = \begin{cases} x(mT), & n = 2m, \\ \hat{x}(mT + 0.5T), & n = 2m+1. \end{cases}$$

- We pass $y[n]$ through a ZOH (at interval $0.5T$) to get the reconstructed signal $\hat{y}(t)$.

(a) Find and plot the impulse response of the reconstructor.

(b) Compute the frequency response of the reconstructor.

3.32 Suppose that, instead of choosing ω_0 to the left of the interval $[\omega_1, \omega_2]$, as we did in Section 3.6, we choose it to the right of the interval. Repeat the procedure described there for this case. Show that the resulting sampling interval is always smaller than the one given in (3.55).

3.33 Explain why in practice it is usually advisable, in sampling of band-pass signals, to extend the bandwidth to both left and right.

3.34 Write the impulse response $h(t)$ of the reconstruction filter (3.56).

3.35 Let $x(t)$ be a signal whose Fourier transform $X^{\mathrm{F}}(\omega)$ is nonzero only on $3 \le |\omega| \le 9$. However, only the frequencies $5 \le |\omega| \le 7$ contain useful information whereas the other frequencies contain only noise.

(a) What is the smallest sampling rate that will enable exact reconstruction of the *useful* signal, if we do not perform any filtering on $x(t)$ before sampling?

(b) How will the answer to part a change if it is permitted to pass $x(t)$ through a filter before sampling?

3.36 $x(t)$ is a band-pass signal whose Fourier transform is nonzero only for $3.5 \le \omega \le 4.5$. Let

$$y(t) = [x(t)]^2.$$

What is the smallest rate at which $y(t)$ can be sampled such that it will be possible to reconstruct $y(t)$ exactly from its samples?

3.37 Let $x(t)$ be a complex band-limited signal on $[-\omega_m, \omega_m]$ and assume that $X^F(\omega_m) = X^F(-\omega_m) = 0$. Define

$$y(t) = x(t) \sum_{n=1}^{\infty} \lambda^n \sin(n\omega_0 t),$$

where λ is a given real number in the range $0 < \lambda < 1$.

(a) Express $Y^F(\omega)$ in terms of $X^F(\omega)$.

(b) What is the minimum value of ω_0 that enables exact reconstruction of $x(t)$ from $y(t)$?

(c) Suggest a procedure for reconstructing $x(t)$ from $y(t)$ when the condition on ω_0 is met.

3.38 A digital communication channel is given, capable of transmitting 19,200 bits per second. We wish to use the channel to transmit a band-limited analog signal $x(t)$, by sampling and digitizing. The magnitude of the analog signal is limited to $|x(t)| \le x_{max}$. The error between the digitized signal and $x(t)$ must not exceed $\pm 10^{-4} x_{max}$.

(a) What is the required number of bits of the A/D?

(b) What is the maximum bandwidth of the analog signal for which the channel can be used?

3.39 The signal $x(t) = \cos(\omega_0 t)$ is sampled by a nonideal sampler, as described in Section 3.5.3, with averaging interval Δ. The discrete-time signal $x[n]$ is then reconstructed, using an ideal (Shannon) reconstructor. Assume that the sampling interval T meets the Nyquist condition and that $\Delta < T$. Derive an expression for the reconstructed signal.

3.40 The purpose of this problem is to demonstrate that plots of signals sampled only slightly above the Nyquist rate can be misleading.

(a) Let

$$x(t) = \sin(0.98\pi t).$$

Sample the signal at interval $T = 1$ and plot it for $0 \le n \le 100$. Does the plot look like a sinusoid?

(b) Sample the signal at $T = 1/8$ and plot it for $0 \le n \le 800$. Then examine the signal details by plotting it for $0 \le n \le 100$. Does the plot look like a sinusoid now?

(c) Interpret what you have seen and draw conclusions.

3.41* Ben and Erik are given the continuous-time signal

$$x(t) = \cos(2\pi f_0 t + \phi_0).$$

They are told that $\phi_0 = 0.25\pi$, but they know nothing about f_0. Each is asked to sample the signal at a rate of his choice and report the value of f_0 based on the sampled signal. Ben uses a sampling frequency $f_{sam} = 150$ Hz, and reports that $f_0 = 50$ Hz. Erik uses a sampling frequency $f_{sam} = 240$ Hz, and reports that $f_0 = 20$ Hz.

(a) Is it possible, based on this information, to determine the true value of f_0? If not, what are all the possible values of f_0?

(b) If it is known that $f_0 < 1000\,\text{Hz}$, is it then possible to determine the true value of f_0?

3.42* We are given two band-pass signals: $x_1(t)$ is limited to the frequency range 1000–1350 Hz; $x_2(t)$ is limited to the frequency range 2000–2400 Hz. We are required to transmit *both* signals, in a digital form, over a single channel. The following scheme is proposed for this purpose; see Figure 3.37. Each signal is sampled separately, then the two discrete-time signals $x_1[n]$, $x_2[n]$ are combined to a single sequence $z[n]$ by a packetizer. The packetizer takes N_1 samples from the first signal and N_2 samples from the second signal as they become available, builds a sequence of $N_1 + N_2$ regularly spaced samples, and passes them to the channel. At the receiving end of the channel there is a distributor, which separates $z[n]$ back to the individual sequences $x_1[n]$, $x_2[n]$ and outputs each at regularly spaced intervals to the reconstructors. Finally, the reconstructors rebuild the continuous-time signals.

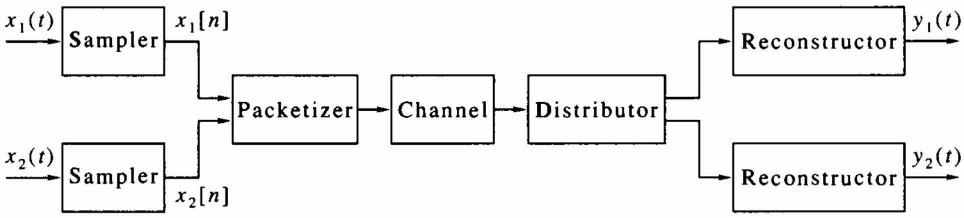

Figure 3.37 Pertaining to Problem 3.42.

(a) What are the minimum sampling rates for the two continuous-time signals?

(b) How many samples per second must the channel be capable of transmitting?

(c) What is the minimum packet size $N_1 + N_2$? How many packets are transmitted every second?

3.43* Let $x(t)$ be a real signal, band limited to $\pm\omega_m$. Consider the double-side-band modulated signal (see Problem 2.15)

$$y(t) = x(t)\cos(\omega_0 t),$$

where $\omega_0 \gg \omega_m$. The signal $y(t)$ is band-pass. DSB is often used in wireless communication. In receiving DSB signals, the most common scheme is to demodulate the signal first by

$$z(t) = 2y(t)\cos(\omega_0 t) = 2x(t)\cos^2(\omega_0 t) = x(t)[1 + \cos(2\omega_0 t)].$$

The signal $z(t)$ is then passed through a low-pass filter, which removes the high-frequency component $x(t)\cos(2\omega_0 t)$. The output of the filter is $x(t)$. This can be sampled at an interval $T \le \pi/\omega_m$ for digital processing.

(a) Show that $x(nT)$ can also be obtained by sampling $y(t)$ directly, without demodulating it first. This requires a careful choice of T. Find the largest T (i.e., the smallest sampling rate) for which this can be done, and explain why it works. Interpret your solution in the frequency domain.

(b) Suppose now that the receiver does not know ω_0 accurately and uses $\omega_0 + \Delta\omega$ instead, where $\Delta\omega$ is a small frequency error. Show that both schemes are prone to failure in this case, and explain exactly why and how they fail.

3.44* You are given the task of designing a digital receiver for radio transmission from a satellite. The transmitted signal is double side-band, as defined in Problem 3.43. The nominal carrier frequency is $f_0 = 160\,\text{MHz}$ and the information bandwidth is $f_m = 19.2\,\text{kHz}$. The satellite's orbit is at altitude $h = 300\,\text{km}$. The frequency generator at the satellite has relative accuracy $\pm 0.5 \times 10^{-6}$, and the one at the receiver has relative accuracy $\pm 3 \times 10^{-6}$. The receiver is required to work from horizon to horizon, that is, at all possible positions of the satellite relative to the receiver.

This problem deals with the front end of the receiver, which consists of a DSB demodulator, a low-pass filter, and a sampler; see Figure 3.38. The three parameters you have to determine are (1) the nominal frequency of the demodulator f_d (see the figure); (2) the sampling rate T; (3) the cutoff frequency of the low-pass filter f_c. Because of the inaccuracies of the various frequencies, you decide to demodulate the signal only partially, such that the Fourier transform remains band pass and its positive and negative parts do not overlap under any circumstances. You leave to the digital processor the task of handling the residual modulation.

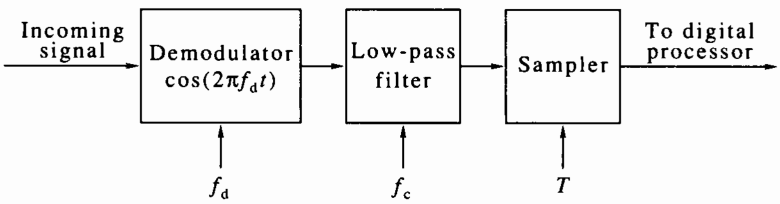

Figure 3.38 Pertaining to Problem 3.44.

Design the three parameters according to the supplied information, using the following hints and guidelines:

- If you have already solved Problem 3.43 you are aware of the inadequacy of the solution proposed there for the situation described here, so do not attempt it.
- You will need a few facts from your college physics education. In particular:
  (a) The Earth radius is $r_0 = 6370\,\text{km}$ and the Earth gravity at sea level is $g(0) = 9.81\,\text{m/s}^2$. The gravity at altitude h, denoted by $g(h)$, is inversely proportional to the square of the distance of a point at that altitude from the center of the Earth.
  (b) The orbital velocity of the satellite is the square root of the product of the gravity at its orbit by the radius of the orbit.
  (c) The Earth rotates once every 24 hours.
  (d) Doppler effect acts to change the frequency of an electromagnetic wave; look at a physics book to remind yourself of the formula.
- It is good engineering practice to insert reasonable safety margins at various places.

3.45* This problem introduces the concept of a T-*parent* of a discrete-time signal.

(a) Prove the following claim:

> Let $x[n]$ be a discrete-time signal whose Fourier transform $X^{\mathrm{f}}(\theta)$ exists. Then, for every positive T, there exists a unique continuous-time signal $x(t)$ such that $X^{\mathrm{F}}(\omega)$ is band limited to $|\omega| \le \pi/T$ and $x[n] = x(nT)$.

The signal $x(t)$ can be called the T-*parent* of $x[n]$.

(b) Find the T-parent of the signal

$$x[n] = \begin{cases} 1, & n = -1, 0, 1, \\ 0, & \text{otherwise.} \end{cases}$$

(c) Generalize part a to any finite duration signal, that is, a signal satisfying

$$x[n] = 0, \quad n < n_1, \; n > n_2.$$

(d) Is the signal

$$x(t) = e^{-t/T}, \; t \ge 0$$

the T-parent of

$$x[n] = e^{-n}, \; n \ge 0?$$

Give reasons.