6 SAMPLING AND ANALOG-TO-DIGITAL CONVERSION

As briefly discussed in Chapter 1, analog signals can be digitized through sampling and quantization. This analog-to-digital (A/D) conversion sets the foundation of modern digital communication systems. In the A/D converter, the sampling rate must be large enough to permit the analog signal to be reconstructed from the samples with sufficient accuracy. The **sampling theorem**, which is the basis for determining the proper (lossless) sampling rate for a given signal, has played a huge role in signal processing, communication theory, and A/D circuit design.

6.1 SAMPLING THEOREM

We first show that a signal $g(t)$ whose spectrum is band-limited to B Hz, that is,

$$G(f) = 0 \quad \text{for } |f| > B$$

can be reconstructed exactly (without any error) from its discrete time samples taken uniformly at a rate of R samples per second. The condition is that $R > 2B$. In other words, the minimum sampling frequency for perfect signal recovery is $f_s = 2B$ Hz.

To prove the sampling theorem, consider a signal $g(t)$ (Fig. 6.1a) whose spectrum is band-limited to B Hz (Fig. 6.1b).* For convenience, spectra are shown as functions of f as well as of ω. Sampling $g(t)$ at a rate of f_s Hz means that we take f_s uniform samples per second. This uniform sampling can be accomplished by multiplying $g(t)$ by an impulse train $\delta_{T_s}(t)$ of Fig. 6.1c, consisting of unit impulses repeating periodically every T_s seconds, where $T_s = 1/f_s$. This results in the sampled signal $\overline{g}(t)$ shown in Fig. 6.1d. The sampled signal consists of impulses spaced every T_s seconds (the sampling interval). The nth impulse, located at $t = nT_s$, has a strength $g(nT_s)$ which is the value of $g(t)$ at $t = nT_s$. Thus, the relationship between the

* The spectrum $G(f)$ in Fig. 6.1b is shown as real, for convenience. Our arguments are valid for complex $G(f)$.

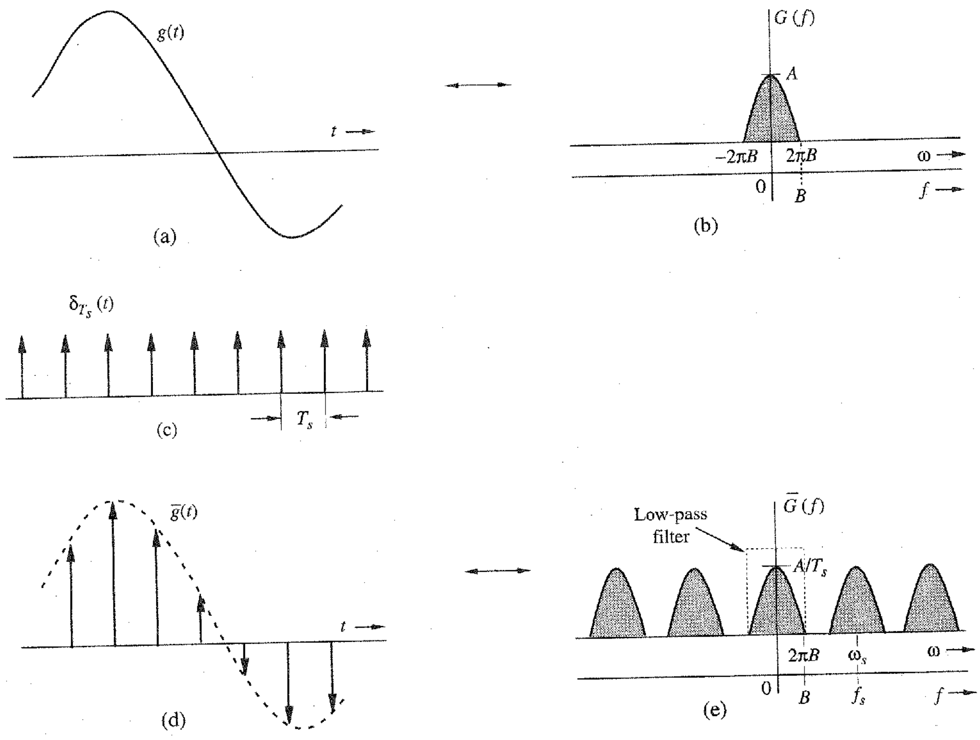

Figure 6.1 Sampled signal and its Fourier spectra.

sampled signal $\overline{g}(t)$ and the original analog signal $g(t)$ is

$$\overline{g}(t) = g(t)\delta_{T_s}(t) = \sum_n g(nT_s)\delta(t - nT_s) \tag{6.1}$$

Because the impulse train $\delta_{T_s}(t)$ is a periodic signal of period T_s, it can be expressed as an exponential Fourier series, already found in Example 3.11 as

$$\delta_{T_s}(t) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} e^{jn\omega_s t} \qquad \omega_s = \frac{2\pi}{T_s} = 2\pi f_s \tag{6.2}$$

Therefore,

$$\begin{aligned}\overline{g}(t) &= g(t)\delta_{T_s}(t) \\ &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} g(t)e^{jn2\pi f_s t}\end{aligned} \tag{6.3}$$

To find $\overline{G}(f)$, the Fourier transform of $\overline{g}(t)$, we take the Fourier transform of the summation in Eq. (6.3). Based on the frequency-shifting property, the transform of the *nth* term is shifted by nf_s. Therefore,

$$\overline{G}(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} G(f - nf_s) \tag{6.4}$$

This means that the spectrum $\overline{G}(f)$ consists of $G(f)$, scaled by a constant $1/T_s$, repeating periodically with period $f_s = 1/T_s$ Hz, as shown in Fig. 6.1e.

After uniform sampling that generates a set of signal samples $\{g(kT_s)\}$, the vital question becomes: **Can $g(t)$ be reconstructed from $\overline{g}(t)$ without any loss or distortion?** If we are to reconstruct $g(t)$ from $\overline{g}(t)$, equivalently in the frequency domain we should be able to recover $G(f)$ from $\overline{G}(f)$. Graphically from Fig. 6.1, perfect recovery is possible if there is no overlap among the replicas in $\overline{G}(f)$. Figure 6.1e clearly shows that this requires

$$f_s > 2B \tag{6.5}$$

Also, the sampling interval $T_s = 1/f_s$. Therefore,

$$T_s < \frac{1}{2B} \tag{6.6}$$

Thus, as long as the sampling frequency f_s is greater than twice the signal bandwidth B (in hertz), $\overline{G}(f)$ will consist of nonoverlapping repetitions of $G(f)$. When this is true, Fig. 6.1e shows that $g(t)$ can be recovered from its samples $\overline{g}(t)$ by passing the sampled signal $\overline{g}(t)$ through an ideal low-pass filter of bandwidth B Hz. The minimum sampling rate $f_s = 2B$ required to recover $g(t)$ from its samples $\overline{g}(t)$ is called the **Nyquist rate** for $g(t)$, and the corresponding sampling interval $T_s = 1/2B$ is called the **Nyquist interval** for the low-pass signal $g(t)$.*

We need to stress one important point regarding the possibility of $f_s = 2B$ and a particular class of low-pass signals. For a general signal spectrum, we have proved that the sampling rate $f_s > 2B$. However, if the spectrum $G(f)$ has no impulse (or its derivatives) at the highest frequency B, then the overlap is still zero as long as the sampling rate is greater than or equal to the Nyquist rate, that is,

$$f_s \geq 2B$$

If, on the other hand, $G(f)$ contains an impulse at the highest frequency $\pm B$, then the equality must be removed or else overlap will occur. In such case, the sampling rate f_s must be greater than $2B$ Hz. A well-known example is a sinusoid $g(t) = \sin 2\pi B(t - t_0)$. This signal is band-limited to B Hz, but all its samples are zero when uniformly taken at a rate $f_s = 2B$ (starting at $t = t_0$), and $g(t)$ cannot be recovered from its Nyquist samples. Thus, for sinusoids, the condition of $f_s > 2B$ must be satisfied.

6.1.1 Signal Reconstruction from Uniform Samples

The process of reconstructing a continuous time signal $g(t)$ from its samples is also known as **interpolation**. In Fig. 6.1, we used a constructive proof to show that a signal $g(t)$ band-limited

* The theorem stated here (and proved subsequently) applies to low-pass signals. A bandpass signal whose spectrum exists over a frequency band $f_c - B/2 < |f| < f_c + B/2$ has a bandwidth B Hz. Such a signal is also uniquely determined by samples taken at above the Nyquist frequency $2B$. The sampling theorem is generally more complex in such case. It uses two interlaced uniform sampling trains, each at half the overall sampling rate $R_s > B$. See, for example, the Refs. 1 and 2.

to B Hz can be reconstructed (interpolated) exactly from its samples. This means not only that uniform sampling at above the Nyquist rate preserves all the signal information, but also that simply passing the sampled signal through an ideal low-pass filter of bandwidth B Hz will reconstruct the original message. As seen from Eq. (6.3), the sampled signal contains a component $(1/T_s)g(t)$, and to recover $g(t)$ [or $G(f)$], the sampled signal

$$\overline{g}(t) = \sum g(nT_s)\delta(t - nT_s)$$

must be sent through an ideal low-pass filter of bandwidth B Hz and gain T_s. Such an ideal filter response has the transfer function

$$H(f) = T_s \,\Pi\left(\frac{\omega}{4\pi B}\right) = T_s \,\Pi\left(\frac{f}{2B}\right) \tag{6.7}$$

Ideal Reconstruction

To recover the analog signal from its uniform samples, the ideal interpolation filter transfer function found in Eq. (6.7) is shown in Fig. 6.2a. The impulse response of this filter, the inverse Fourier transform of $H(f)$, is

$$h(t) = 2BT_s \operatorname{sinc}(2\pi Bt) \tag{6.8}$$

Assuming the use of Nyquist sampling rate, that is, $2BT_s = 1$, then

$$h(t) = \operatorname{sinc}(2\pi Bt) \tag{6.9}$$

This $h(t)$ is shown in Fig. 6.2b. Observe the very interesting fact that $h(t) = 0$ at all Nyquist sampling instants ($t = \pm n/2B$) except $t = 0$. When the sampled signal $\overline{g}(t)$ is applied at the input of this filter, the output is $g(t)$. Each sample in $\overline{g}(t)$, being an impulse, generates a sinc pulse of height equal to the strength of the sample, as shown in Fig. 6.2c. The process is

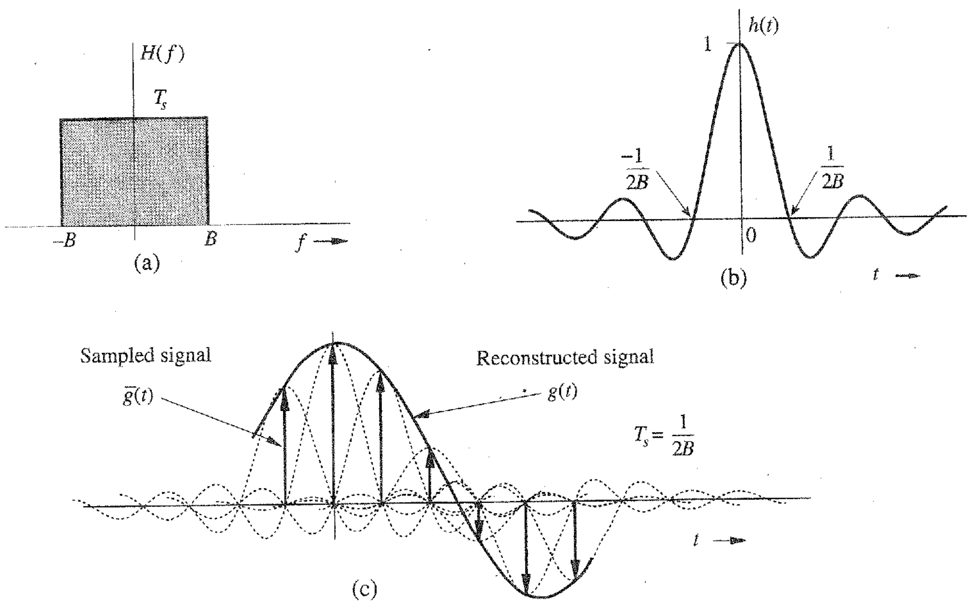

Figure 6.2 Ideal interpolation.

identical to that shown in Fig. 6.6, except that $h(t)$ is a sinc pulse instead of a rectangular pulse. Addition of the sinc pulses generated by all the samples results in $g(t)$. The kth sample of the input $\overline{g}(t)$ is the impulse $g(kT_s)\delta(t-kT_s)$; the filter output of this impulse is $g(kT_s)h(t-kT_s)$. Hence, the filter output to $\overline{g}(t)$, which is $g(t)$, can now be expressed as a sum,

$$g(t) = \sum_k g(kT_s)h(t-kT_s)$$

$$= \sum_k g(kT_s)\,\text{sinc}\,[2\pi B(t-kT_s)] \tag{6.10a}$$

$$= \sum_k g(kT_s)\,\text{sinc}\,(2\pi Bt - k\pi) \tag{6.10b}$$

Equation (6.10) is the **interpolation formula,** which yields values of $g(t)$ between samples as a weighted sum of all the sample values.

Example 6.1 Find a signal $g(t)$ that is band-limited to B Hz and whose samples are

$$g(0) = 1 \quad \text{and} \quad g(\pm T_s) = g(\pm 2T_s) = g(\pm 3T_s) = \cdots = 0$$

where the sampling interval T_s is the Nyquist interval for $g(t)$, that is, $T_s = 1/2B$.

We use the interpolation formula (6.10b) to construct $g(t)$ from its samples. Since all but one of the Nyquist samples are zero, only one term (corresponding to $k = 0$) in the summation on the right-hand side of Eq. (6.10b) survives. Thus,

$$g(t) = \text{sinc}\,(2\pi Bt) \tag{6.11}$$

This signal is shown in Fig. 6.3. Observe that this is the only signal that has a bandwidth B Hz and sample values $g(0) = 1$ and $g(nT_s) = 0$ $(n \neq 0)$. No other signal satisfies these conditions.

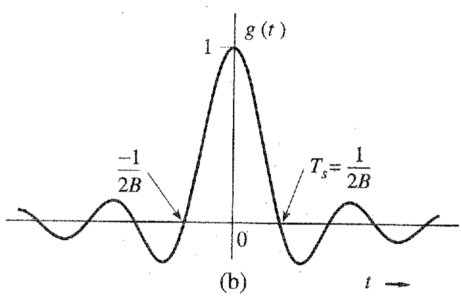

Figure 6.3 Signal reconstructed from the Nyquist samples in Example 6.1.

Practical Signal Reconstruction (Interpolation)

We established in Sec. 3.5 that the ideal low-pass filter is noncausal and unrealizable. This can be equivalently seen from the infinitely long nature of the sinc reconstruction pulse used in the ideal reconstruction of Eq. (6.10). For practical application of signal reconstruction (e.g., a

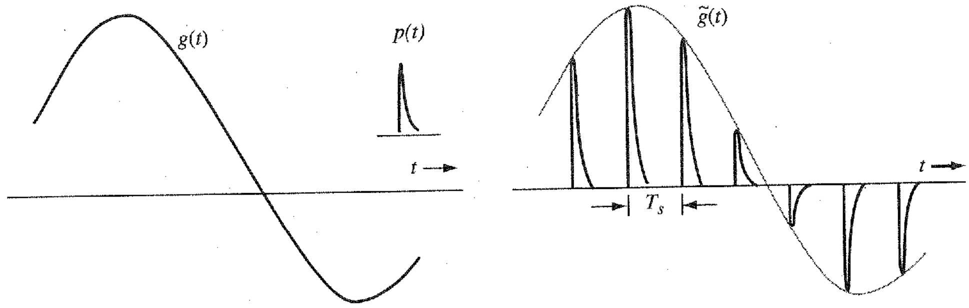

Figure 6.4 Practical reconstruction (interpolation) pulse.

CD player), we need to implement realizable signal reconstruction systems from the uniform signal samples.

For practical implementation, this reconstruction pulse $p(t)$ must be easy to generate. For example, we may apply the reconstruction pulse $p(t)$ as shown in Fig. 6.4. However, we must first use the nonideal interpolation pulse $p(t)$ to analyze the accuracy of the reconstructed signal. Let us denote the new signal from reconstruction as

$$\tilde{g}(t) \triangleq \sum_n g(nT_s)p(t - nT_s) \tag{6.12}$$

To determine its relation to the original analog signal $g(t)$, we can see from the properties of convolution and Eq.(6.1) that

$$\begin{aligned}\tilde{g}(t) &= \sum_n g(nT_s)p(t - nT_s) = p(t) * \left[\sum_n g(nT_s)\delta(t - nT_s)\right] \\ &= p(t) * \overline{g}(t)\end{aligned} \tag{6.13a}$$

In the frequency domain, the relationship between the reconstruction and the original analog signal can rely on Eq. (6.4)

$$\tilde{G}(f) = P(f)\frac{1}{T_s}\sum_n G(f - nf_s) \tag{6.13b}$$

This means that the reconstructed signal $\tilde{g}(t)$ using pulse $p(t)$ consists of multiple replicas of $G(f)$ shifted to the frequency center nf_s and filtered by $P(f)$. To fully recover $g(t)$, further filtering of $\tilde{g}(t)$ becomes necessary. Such filters are often referred to as equalizers.

Denote the equalizer transfer function as $E(f)$. Distortionless reconstruction requires that

$$\begin{aligned}G(f) &= E(f)\tilde{G}(f) \\ &= E(f)P(f)\frac{1}{T_s}\sum_n G(f - nf_s)\end{aligned}$$

This relationship clearly illustrates that the equalizer must remove all the shifted replicas $G(f - nf_s)$ in the summation except for the low-pass term with $n = 0$, that is,

$$E(f)P(f) = 0 \quad |f| > f_s - B \tag{6.14a}$$

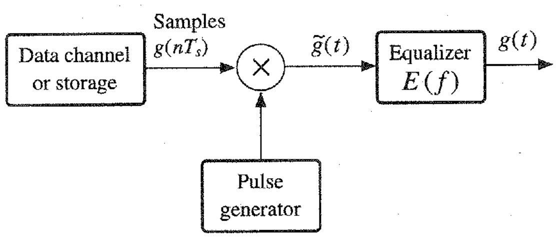

Figure 6.5 Practical signal reconstruction.

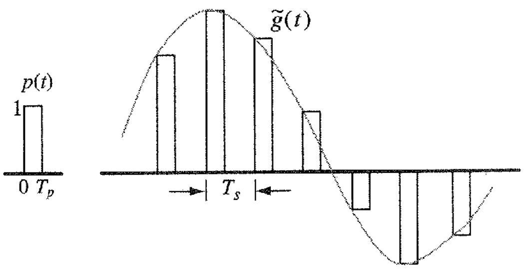

Figure 6.6 Simple interpolation by means of simple rectangular pulses.

Additionally, distortionless reconstruction requires that

$$E(f)P(f) = T_s \quad |f| < B \tag{6.14b}$$

The equalizer filter $E(f)$ must be low-pass in nature to stop all frequency content above $f_s - B$ Hz, and it should be the inverse of $P(f)$ within the signal bandwidth of B Hz. Figure 6.5 demonstrates the diagram of a practical signal reconstruction system utilizing such an equalizer.

Let us now consider a very simple interpolating pulse generator that generates short (zero-order hold) pulses. As shown in Fig. 6.6,

$$p(t) = \Pi\left(\frac{t - 0.5T_p}{T_p}\right)$$

This is a gate pulse of unit height with pulse duration T_p. The reconstruction will first generate

$$\tilde{g}(t) = \sum_n g(nT_s)\,\Pi\left(\frac{t - nT_s - 0.5T_p}{T_p}\right)$$

The transfer function of filter $P(f)$ is the Fourier transform of $\Pi(t/T_p)$ shifted by $0.5T_p$:

$$P(f) = T_p \operatorname{sinc}\left(\pi f T_p\right) e^{-j\pi f T_p} \tag{6.15}$$

As a result, the equalizer frequency response should satisfy

$$E(f) = \begin{cases} T_s/P(f) & |f| \le B \\ \text{Flexible} & B < |f| < (1/T_s - B) \\ 0 & |f| \ge (1/T_s - B) \end{cases}$$

It is important for us to ascertain that the equalizer passband response is realizable. First of all, we can add another time delay to the reconstruction such that

$$E(f) = T_s \cdot \frac{\pi f}{\sin(\pi f T_p)} e^{-j2\pi f t_0} \qquad |f| \leq B \tag{6.16}$$

For the passband gain of $E(f)$ to be well defined, it is imperative for us to choose a short pulse width T_p such that

$$\frac{\sin(\pi f T_p)}{\pi f} \neq 0 \quad |f| \leq B$$

This means that the equalizer $E(f)$ does not need to achieve infinite gain. Otherwise the equalizer would become unrealizable. Equivalently, this requires that

$$T_p < 1/B$$

Hence, as long as the rectangular reconstruction pulse width is shorter than $1/B$, it may be possible to design an analog equalizer filter to recover the original analog signal $g(t)$ from the nonideal reconstruction pulse train. Of course, this is a requirement for a rectangular reconstruction pulse generator. In practice, T_p can be chosen very small, to yield the following equalizer passband response:

$$E(f) = T_s \cdot \frac{\pi f}{\sin(\pi f T_p)} \approx \frac{T_s}{T_p} \qquad |f| \leq B \tag{6.17}$$

This means that very little distortion remains when very short rectangular pulses are used in signal reconstruction. Such cases make the design of the equalizer either unnecessary or very simple. An illustrative example is given as a MATLAB exercise in Sec. 6.9.

We can improve on the zero-order-hold filter by using the **first-order-hold** filter, which results in a linear interpolation instead of the staircase interpolation. The linear interpolator, whose impulse response is a triangle pulse $\Delta(t/2T_s)$, results in an interpolation in which successive sample tops are connected by straight-line segments (Prob. 6.1-7).

6.1.2 Practical Issues in Signal Sampling and Reconstruction

Realizability of Reconstruction Filters

If a signal is sampled at the Nyquist rate $f_s = 2B$ Hz, the spectrum $\overline{G}(f)$ consists of repetitions of $G(f)$ without any gap between successive cycles, as shown in Fig. 6.7a. To recover $g(t)$ from $\overline{g}(t)$, we need to pass the sampled signal $\overline{g}(t)$ through an ideal low-pass filter (dotted area in Fig. 6.7a). As seen in Sec. 3.5, such a filter is unrealizable in practice; it can be closely approximated only with infinite time delay in the response. This means that we can recover the signal $g(t)$ from its samples with infinite time delay.

A practical solution to this problem is to sample the signal at a rate higher than the Nyquist rate ($f_s > 2B$ or $\omega_s > 4\pi B$). This yields $\overline{G}(f)$, consisting of repetitions of $G(f)$ with a finite band gap between successive cycles, as shown in Fig. 6.7b. We can now recover $G(f)$ from $\overline{G}(f)$ [or from $\tilde{G}(f)$] by using a low-pass filter with a gradual cutoff characteristic (dotted area in Fig. 6.7b). But even in this case, the filter gain is required to be zero beyond the first cycle

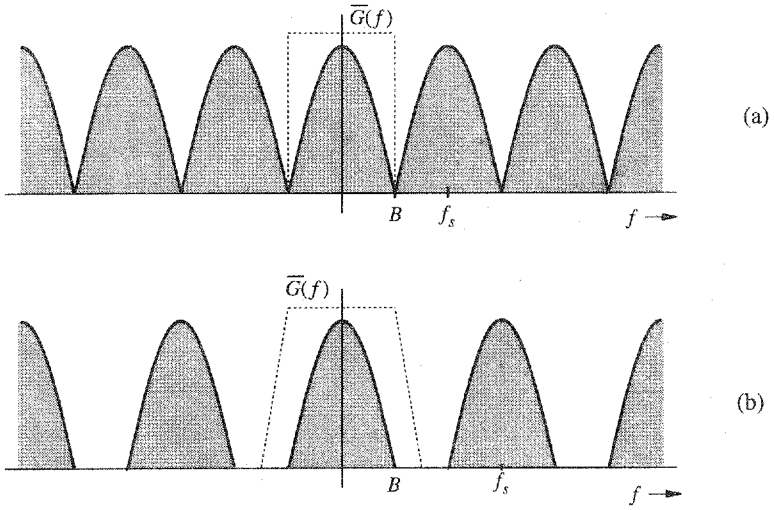

Figure 6.7 Spectra of a sampled signal: (a) at the Nyquist rate; (b) above the Nyquist rate.

of $G(f)$ (Fig. 6.7b). According to the Paley-Wiener criterion, it is impossible to realize even this filter. The only advantage in this case is that the required filter can be better approximated with a smaller time delay. This shows that it is impossible in practice to recover a band-limited signal $g(t)$ exactly from its samples, even if the sampling rate is higher than the Nyquist rate. However, as the sampling rate increases, the recovered signal approaches the desired signal more closely.

The Treachery of Aliasing

There is another fundamental practical difficulty in reconstructing a signal from its samples. The sampling theorem was proved on the assumption that the signal $g(t)$ is band-limited. *All practical signals are time-limited;* that is, they are of finite duration or width. We can demonstrate (Prob. 6.1-8) that a signal cannot be time-limited and band-limited simultaneously. A time-limited signal cannot be band-limited, and vice versa (but a signal can be simultaneously non-time-limited and non-band-limited). Clearly, all practical signals, which are necessarily time-limited, are non-band-limited, as shown in Fig. 6.8a; they have infinite bandwidth, and the spectrum $\bar{G}(f)$ consists of overlapping cycles of $G(f)$ repeating every f_s Hz (the sampling frequency), as illustrated in Fig. 6.8b. Because of the infinite bandwidth in this case, the spectral overlap is unavoidable, regardless of the sampling rate. Sampling at a higher rate reduces but does not eliminate overlapping between repeating spectral cycles. Because of the overlapping tails, $\bar{G}(f)$ no longer has complete information about $G(f)$, and it is no longer possible, even theoretically, to recover $g(t)$ exactly from the sampled signal $\bar{g}(t)$. If the sampled signal is passed through an ideal low-pass filter of cutoff frequency $f_s/2$ Hz, the output is not $G(f)$ but $G_a(f)$ (Fig. 6.8c), which is a version of $G(f)$ distorted as a result of two separate causes:

1. The loss of the tail of $G(f)$ beyond $|f| > f_s/2$ Hz.
2. The reappearance of this tail inverted or folded back onto the spectrum.

Note that the spectra cross at frequency $f_s/2 = 1/2T$ Hz, which is called the *folding* frequency. The spectrum may be viewed as if the lost tail is folding back onto itself at the folding frequency. For instance, a component of frequency $(f_s/2)+f_z$ shows up as, or "impersonates," a component of lower frequency $(f_s/2)-f_z$ in the reconstructed signal. Thus, the components of frequencies above $f_s/2$ reappear as components of frequencies below $f_s/2$. This tail inversion,

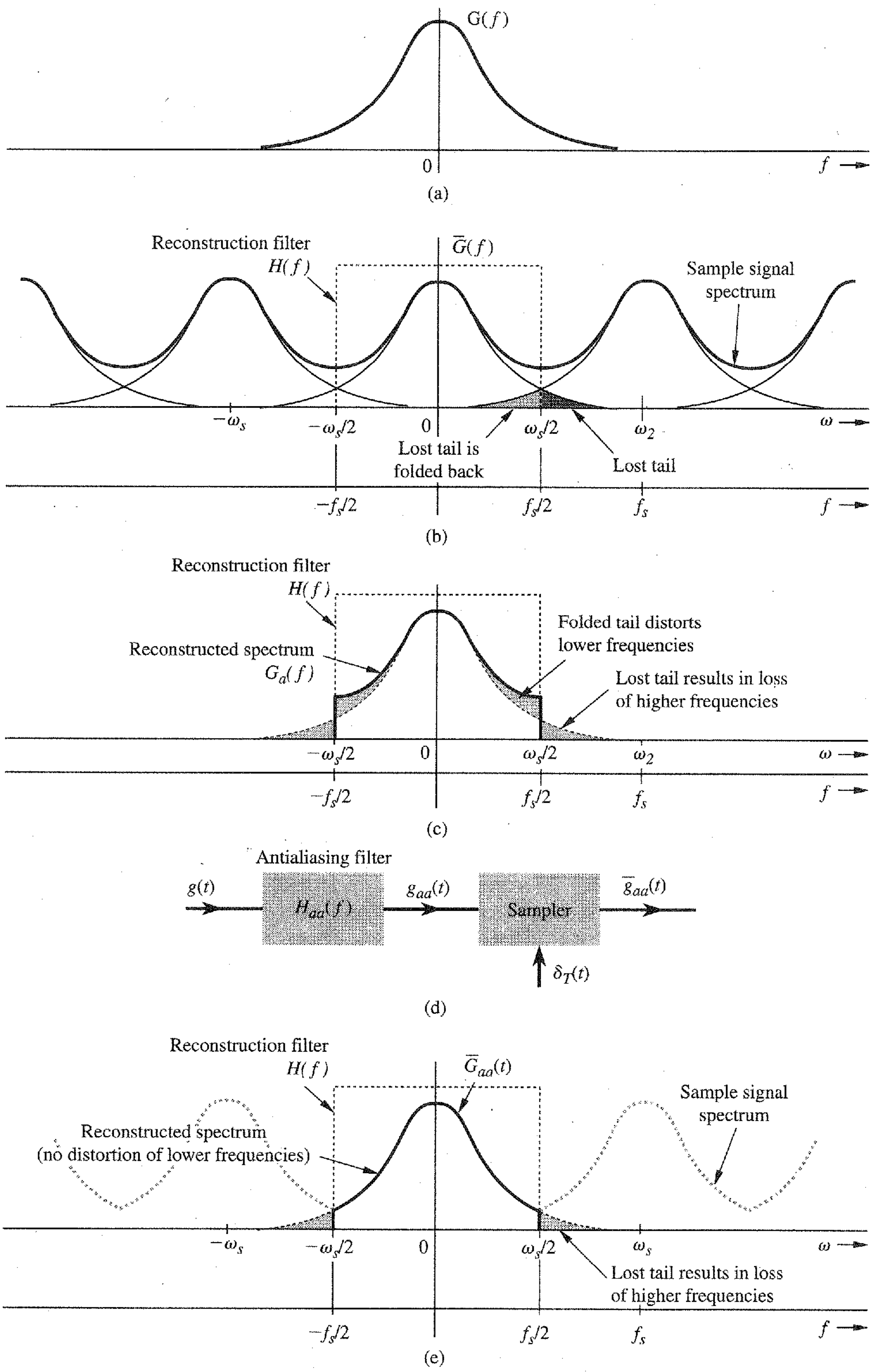

Figure 6.8 Aliasing effect. (a) Spectrum of a practical signal $g(t)$. (b) Spectrum of sampled $g(t)$. (c) Reconstructed signal spectrum. (d) Sampling scheme using antialiasing filter. (e) Sampled signal spectrum (dotted) and the reconstructed signal spectrum (solid) when antialiasing filter is used.

known as *spectral folding* or *aliasing*, is shown shaded in Fig. 6.8b and also in Fig. 6.8c. In the process of aliasing, not only are we losing all the components of frequencies above the folding frequency $f_s/2$ Hz, but these very components reappear (aliased) as lower frequency components in Fig. 6.8b or c. Such aliasing destroys the integrity of the frequency components below the folding frequency $f_s/2$, as depicted in Fig. 6.8c.

The problem of aliasing is analogous to that of an army when a certain platoon has secretly defected to the enemy side but remains nominally loyal to their army. The army is in double jeopardy. First, it has lost the defecting platoon as an effective fighting force. In addition, during actual fighting, the army will have to contend with sabotage caused by the defectors and will have to use loyal platoon to neutralize the defectors. Thus, the army has lost two platoons to nonproductive activity.

Defectors Eliminated: The Antialiasing Filter

If you were the commander of the betrayed army, the solution to the problem would be obvious. As soon as you got wind of the defection, you would incapacitate, by whatever means, the defecting platoon. By taking this action *before the fighting begins*, you lose only one (the defecting)* platoon. This is a partial solution to the double jeopardy of betrayal and sabotage, a solution that partly rectifies the problem and cuts the losses in half.

We follow exactly the same procedure. The potential defectors are all the frequency components beyond the folding frequency $f_s/2 = 1/2T$ Hz. We should eliminate (suppress) these components from $g(t)$ *before sampling* $g(t)$. Such suppression of higher frequencies can be accomplished by an ideal low-pass filter of cutoff $f_s/2$ Hz, as shown in Fig. 6.8d. This is called the *antialiasing filter*. Figure 6.8d also shows that antialiasing filtering is performed before sampling. Figure 6.8e shows the sampled signal spectrum and the reconstructed signal $G_{aa}(f)$ when the antialiasing scheme is used. An antialiasing filter essentially band-limits the signal $g(t)$ to $f_s/2$ Hz. This way, we lose only the components beyond the folding frequency $f_s/2$ Hz. These suppressed components now cannot reappear, corrupting the components of frequencies below the folding frequency. Clearly, use of an antialiasing filter results in the reconstructed signal spectrum $G_{aa}(f) = G(f)$ for $|f| < f_s/2$. Thus, although we lost the spectrum beyond $f_s/2$ Hz, the spectrum for all the frequencies below $f_s/2$ remains intact. The effective aliasing distortion is cut in half owing to elimination of folding. We stress again that the antialiasing operation must be performed *before the signal is sampled.*

An antialiasing filter also helps to reduce noise. Noise, generally, has a wideband spectrum, and without antialiasing, the aliasing phenomenon itself will cause the noise components outside the desired signal band to appear in the signal band. Antialiasing suppresses the entire noise spectrum beyond frequency $f_s/2$.

The antialiasing filter, being an ideal filter, is unrealizable. In practice we use a steep-cutoff filter, which leaves a sharply attenuated residual spectrum beyond the folding frequency $f_s/2$.

Sampling Forces Non-Band-Limited Signals to Appear Band-Limited

Figure 6.8b shows the spectrum of a signal $\overline{g}(t)$ consists of overlapping cycles of $G(f)$. This means that $\overline{g}(t)$ are sub-Nyquist samples of $g(t)$. However, we may also view the spectrum in Fig. 6.8b as the spectrum $G_a(f)$ (Fig. 6.8c), repeating periodically every f_s Hz without overlap. The spectrum $G_a(f)$ is band-limited to $f_s/2$ Hz. Hence, these (sub-Nyquist) samples of $g(t)$

* Figure 6.8b shows that from the infinite number of repeating cycles, only the neighboring spectral cycles overlap. This is a somewhat simplified picture. In reality, all the cycles overlap and interact with every other cycle because of the infinite width of all practical signal spectra. Fortunately, all practical spectra also must decay at higher frequencies. This results in an insignificant amount of interference from cycles other than the immediate neighbors. When such an assumption is not justified, aliasing computations become little more involved.

are actually the Nyquist samples for signal $g_a(t)$. In conclusion, sampling a non-band-limited signal $g(t)$ at a rate f_s Hz makes the samples appear to be the Nyquist samples of some signal $g_a(t)$, band-limited to $f_s/2$ Hz. In other words, sampling makes a non-band-limited signal appear to be a band-limited signal $g_a(t)$ with bandwidth $f_s/2$ Hz. A similar conclusion applies if $g(t)$ is band-limited but sampled at a sub-Nyquist rate.

6.1.3 Maximum Information Rate: Two Pieces of Information per Second per Hertz

A knowledge of the maximum rate at which information can be transmitted over a channel of bandwidth B Hz is of fundamental importance in digital communication. We now derive one of the basic relationships in communication, which states that *a maximum of 2B independent pieces of information per second can be transmitted, error free, over a noiseless channel of bandwidth B Hz*. The result follows from the sampling theorem.

First, the sampling theorem shows that a low-pass signal of bandwidth B Hz can be fully recovered from samples uniformly taken at the rate of $2B$ samples per second. Conversely, we need to show that any sequence of independent data at the rate of $2B$ Hz can come from uniform samples of a low-pass signal with bandwidth B. Moreover, we can construct this low-pass signal from the independent data sequence.

Suppose a sequence of independent data samples is denoted as $\{g_n\}$. Its rate is $2B$ samples per second. Then there always exists a (not necessarily band-limited) signal $g(t)$ such that

$$g_n = g(nT_s) \qquad T_s = \frac{1}{2B}$$

In Figure 6.9a we illustrate again the effect of sampling the non-band-limited signal $g(t)$ at sampling rate $f_s = 2B$ Hz. Because of aliasing, the ideal sampled signal

$$\begin{aligned}\overline{g}(t) &= \sum_n g(nT_s)\delta(t - nT_s) \\ &= \sum_n g_a(nT_s)\delta(t - nT_s)\end{aligned}$$

where $g_a(t)$ is the aliased low-pass signal whose samples $g_a(nT_s)$ equal to the samples of $g(nT_s)$. In other words, sub-Nyquist sampling of a signal $g(t)$ generates samples that can be equally well obtained by Nyquist sampling of a band-limited signal $g_a(t)$. Thus, through Figure 6.9, we demonstrate that sampling $g(t)$ and $g_a(t)$ at the rate of $2B$ Hz will generate the same independent information sequence $\{g_n\}$:

$$g_n = g(nT_s) = g_a(nT_s) \qquad T_s = \frac{1}{2B} \tag{6.18}$$

Also from the sampling theorem, a low-pass signal $g_a(t)$ with bandwidth B can be reconstructed from its uniform samples [Eq. (6.10)]

$$g_a(t) = \sum_n g_n \operatorname{sinc}(2\pi Bt - k\pi)$$

Assuming no noise, this signal can be transmitted over a distortionless channel of bandwidth B Hz, error free. At the receiver, the data sequence $\{g_n\}$ can be recovered from the Nyquist samples of the distortionless channel output $g_a(t)$ as the desired information data.

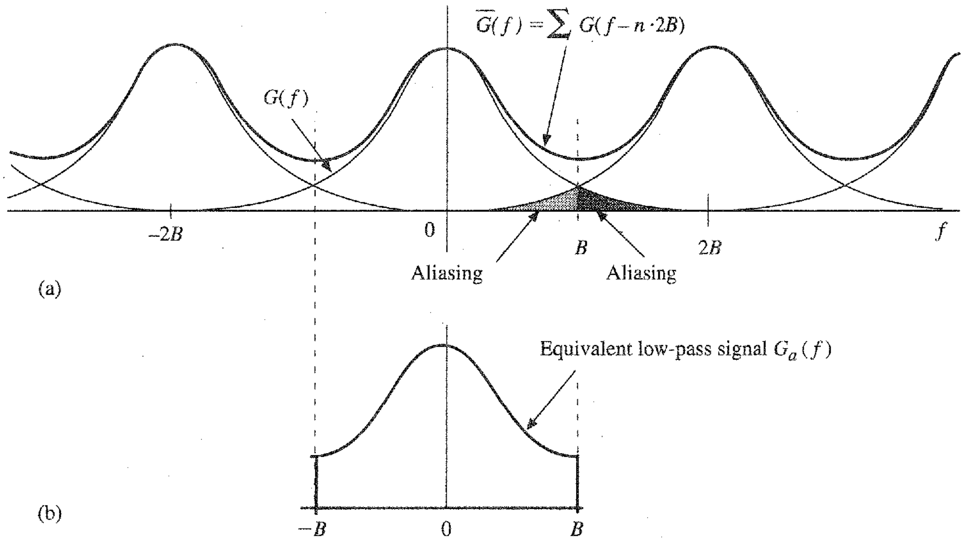

Figure 6.9 (a) Non-band-limited signal spectrum and its sampled spectrum $\bar{G}(f)$. (b) Equivalent low-pass signal spectrum $G_a(f)$ constructed from uniform samples of $g(t)$ at sampling rate $2B$.

This theoretical rate of communication assumes a noise-free channel. In practice, channel noise is unavoidable, and consequently, this rate will cause some detection errors. In Chapter 14, we shall present the Shannon capacity which determines the theoretical error-free communication rate in the presence of noise.

6.1.4 Nonideal Practical Sampling Analysis

Thus far, we have mainly focused on ideal uniform sampling that can use an ideal impulse sampling pulse train to precisely extract the signal value $g(kT_s)$ at the precise instant of $t = kT_s$. In practice, no physical device can carry out such a task. Consequently, we need to consider the more practical implementation of sampling. This analysis is important to the better understanding of errors that typically occur during practical A/D conversion and their effects on signal reconstruction.

Practical samplers take each signal sample over a short time interval T_p around $t = kT_s$. In other words, every T_s seconds, the sampling device takes a short snapshot of duration T_p from the signal $g(t)$ being sampled. This is just like taking a sequence of still photographs of a sprinter during an 100-meter Olympic race. Much like a regular camera that generates a still picture by averaging the picture scene over the window T_p, the practical sampler would generate a sample value at $t = kT_s$ by averaging the values of signal $g(t)$ over the window T_p, that is,

$$g_1(kT_s) = \frac{1}{T_p} \int_{-T_p/2}^{T_p/2} g(kT_s + t)\, dt \tag{6.19a}$$

Depending on the actual device, this averaging may be weighted by a device-dependent averaging function $q(t)$ such that

$$g_1(kT_s) = \frac{1}{T_p} \int_{-T_p/2}^{T_p/2} q(t) g(kT_s + t)\, dt \tag{6.19b}$$

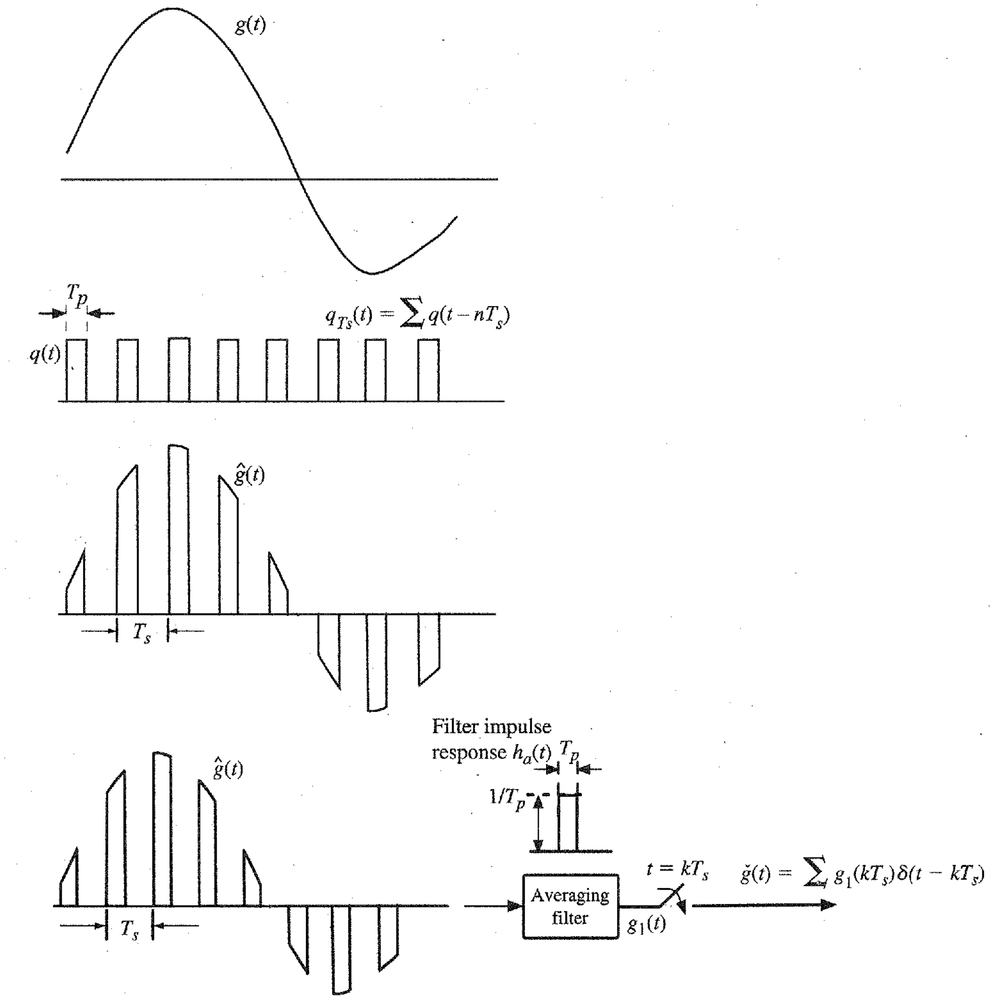

Figure 6.10 Illustration of practical sampling.

Thus we have used the camera analogy to establish that practical samplers in fact generate sampled signal of the form

$$\breve{g}(t) = \sum g_1(kT_s)\delta(t - kT_s) \tag{6.20}$$

We will now show the relationship between the practically sampled signal $\breve{g}(t)$ and the original low-pass analog signal $g(t)$ in the frequency domain.

We will use Fig. 6.10 to illustrate the relationship between $\breve{g}(t)$ and $g(t)$ for the special case of uniform weighting. This means that

$$q(t) = \begin{cases} 1 & |t| \le 0.5T_p \\ 0 & |t| > 0.5T_p \end{cases}$$

As shown in Fig. 6.10, $g_1(t)$ can be equivalently obtained by first using "natural gating" to generate the signal *snapshots*

$$\widehat{g}(t) = g(t) \cdot q_{T_s}(t) \tag{6.21}$$

where

$$q_{T_s}(t) = \sum_{n=-\infty}^{\infty} q(t - nT_s)$$

Figure 6.10b illustrates the snapshot signal $\widehat{g}(t)$. We can then define an averaging filter with impulse response

$$h_a(t) = \begin{cases} \dfrac{1}{T_p} & -\dfrac{T_p}{2} \leq t < \dfrac{T_p}{2} \\ 0 & \text{elsewhere} \end{cases}$$

or transfer function

$$H_a(f) = \text{sinc}\left(\pi f T_p\right)$$

Sending the naturally gated snapshot signal $\widehat{g}(t)$ into the averaging filter generates the output signal

$$g_1(t) = h_a(t) * \widehat{g}(t)$$

As illustrated in Fig. 6.10c, the practical sampler generate a sampled signal $\check{g}(t)$ by sampling the averaging filter output $g_1(kT_s)$. Thus we have used Fig. 6.10c to establish the equivalent process of taking snapshots, averaging, and sampling in generating practical samples of $g(t)$. Now we can examine the frequency domain relationships to analyze the distortion generated by practical samplers.

In the following analysis, we will consider a general weighting function $q(t)$ whose only constraint is that

$$q(t) = 0, \qquad t \notin (-0.5T_p, 0.5T_p)$$

To begin, note that $q_{T_s}(t)$ is periodic. Therefore, its Fourier series can be written as

$$q_{T_s}(t) = \sum_{n=-\infty}^{\infty} Q_n e^{jn\omega_s t}$$

where

$$Q_n = \frac{1}{T_s} \int_{-0.5T_p}^{0.5T_p} q(t) e^{-jn\omega_s t} dt$$

Thus, the averaging filter output signal is

$$\begin{aligned} g_1(t) &= h_a(t) * \left[g(t) q_{T_s}(t)\right] \\ &= h_a(t) * \sum_{n=-\infty}^{\infty} Q_n g(t) e^{jn\omega_s t} \end{aligned} \tag{6.22}$$

In the frequency domain, we have

$$G_1(f) = H(f) \sum_{n=-\infty}^{\infty} Q_n G(f - nf_s)$$
$$= \text{sinc}\left(\pi f T_p\right) \sum_{n=-\infty}^{\infty} Q_n G(f - nf_s) \tag{6.23}$$

Because

$$\breve{g}(t) = \sum_k g_1(kT_s)\delta(t - kT_s)$$

we can apply the sampling theorem to show that

$$\breve{G}(f) = \frac{1}{T_s}\sum_m G_1(f + mf_s)$$
$$= \frac{1}{T_s}\sum_m \text{sinc}\left[\frac{(2\pi f + m2\pi f_s)T_p}{2}\right]\sum_n Q_n G(f + mf_s - nf_s)$$
$$= \sum_\ell \left(\frac{1}{T_s}\sum_n Q_n \,\text{sinc}\left[(\pi f + (n+\ell)\pi f_s)T_p\right]\right) G(f + \ell f_s) \tag{6.24}$$

The last equality came from the change of the summation index $\ell = m - n$.

We can define frequency responses

$$F_\ell(f) = \frac{1}{T_s}\sum_n Q_n \,\text{sinc}\left[(\pi f + (n+\ell)\pi f_s)T_p\right]$$

This definition allows us to conveniently write

$$\breve{G}(f) = \sum_\ell F_\ell(f) G_1(f + \ell f_s) \tag{6.25}$$

For the low-pass signal $G(f)$ with bandwidth B Hz, applying an ideal low-pass (interpolation) filter will generate a distorted signal

$$F_0(f)G(f) \tag{6.26a}$$

in which

$$F_0(f) = \frac{1}{T_s}\sum_n Q_n \,\text{sinc}\left[\pi(f + nf_s)T_p\right] \tag{6.26b}$$

It can be seen from Eqs. (6.25) and (6.26) that the practically sampled signal already contains a known distortion $F_0(f)$.

Moreover, the use of a practical reconstruction pulse $p(t)$ as in Eq. (6.12) will generate additional distortion. Let us reconstruct $g(t)$ by using the practical samples to generate

$$\check{g}(t) = \sum_n g_1(nT_s)p(t - nT_s)$$

Then from Eq. (6.13) we obtain the relationship between the spectra of the reconstruction and the original message $G(f)$ as

$$\check{G}(f) = P(f)\sum_n F_n(f)G(f + nf_s) \tag{6.27}$$

Since $G(f)$ has bandwidth B Hz, we will need to design a new equalizer with transfer function $E(f)$ such that the reconstruction is distortionless within the bandwidth B, that is,

$$E(f)P(f)F_0(f) = \begin{cases} 1 & |f| < B \\ \text{Flexible} & B < |f| < f_s - B \\ 0 & |f| > f_s - B \end{cases} \tag{6.28}$$

This single equalizer can be designed to compensate for two sources of distortion: nonideal sampling effect in $F_0(f)$ and nonideal reconstruction effect in $P(f)$. The equalizer design is made practically possible because both distortions are known in advance.

6.1.5 Some Applications of the Sampling Theorem

The sampling theorem is very important in signal analysis, processing, and transmission because it allows us to replace a continuous time signal by a discrete sequence of numbers. Processing a continuous time signal is therefore equivalent to processing a discrete sequence of numbers. This leads us directly into the area of digital filtering. In the field of communication, the transmission of a continuous time message reduces to the transmission of a sequence of numbers. This opens doors to many new techniques of communicating continuous time signals by pulse trains. The continuous time signal $g(t)$ is sampled, and sample values are used to modify certain parameters of a periodic pulse train. We may vary the amplitudes (Fig. 6.11b), widths (Fig. 6.11c), or positions (Fig. 6.11d) of the pulses in proportion to the sample values of the signal $g(t)$. Accordingly, we can have **pulse amplitude modulation** (PAM), **pulse width modulation** (PWM), or **pulse position modulation** (PPM). The most important form of pulse modulation today is **pulse code modulation** (PCM), introduced in Sec. 1.2. In all these cases, instead of transmitting $g(t)$, we transmit the corresponding pulse-modulated signal. At the receiver, we read the information of the pulse-modulated signal and reconstruct the analog signal $g(t)$.

One advantage of using pulse modulation is that it permits the simultaneous transmission of several signals on a time-sharing basis—**time division multiplexing (TDM)**. Because a pulse-modulated signal occupies only a part of the channel time, we can transmit several pulse-modulated signals on the same channel by interweaving them. Figure 6.12 shows the TDM of two PAM signals. In this manner we can multiplex several signals on the same channel by reducing pulse widths.

Another method of transmitting several baseband signals simultaneously is frequency division multiplexing (FDM), briefly discussed in Chapter 4. In FDM, various signals are multiplexed by sharing the channel bandwidth. The spectrum of each message is shifted to a specific band not occupied by any other signal. The information of various signals is located in nonoverlapping frequency bands of the channel. In a way, TDM and FDM are duals of each other.

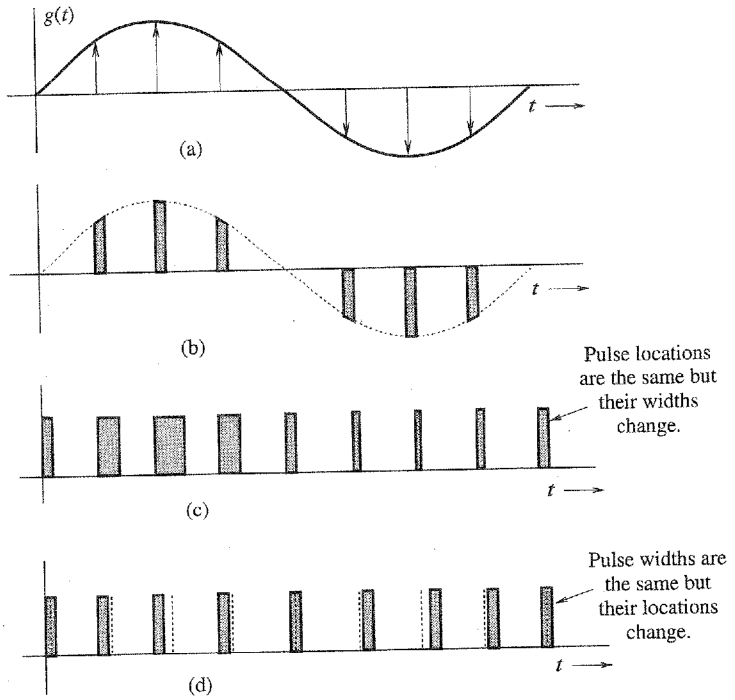

Figure 6.11 Pulse-modulated signals. (a) The unmodulated signal. (b) The PAM signal. (c) The PWM (PDM) signal. (d) The PPM signal.

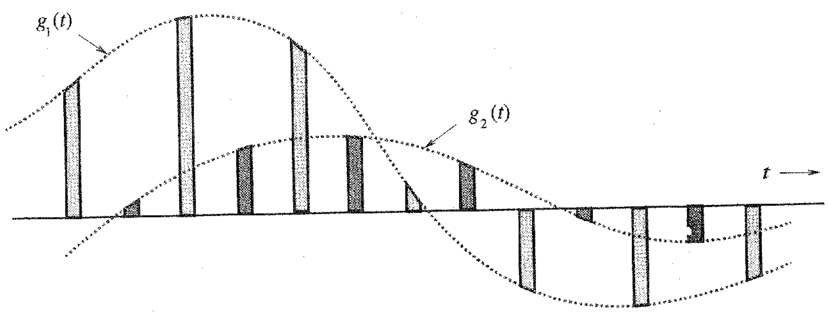

Figure 6.12 Time division multiplexing of two signals.

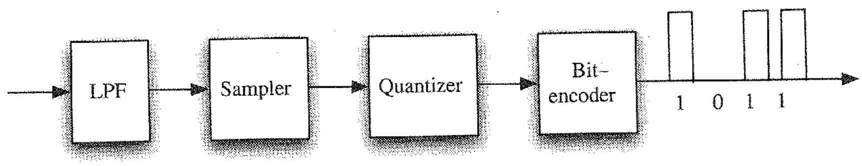

Figure 6.13 PCM system diagram.

6.2 PULSE CODE MODULATION (PCM)

PCM is the most useful and widely used of all the pulse modulations mentioned. As shown in Fig. 6.13, PCM basically is a tool for converting an analog signal into a digital signal (A/D conversion). An **analog** signal is characterized by an amplitude that can take on any value over a continuous range. This means that it can take on an infinite number of values. On the other hand, **digital** signal amplitude can take on only a finite number of values. An analog signal can

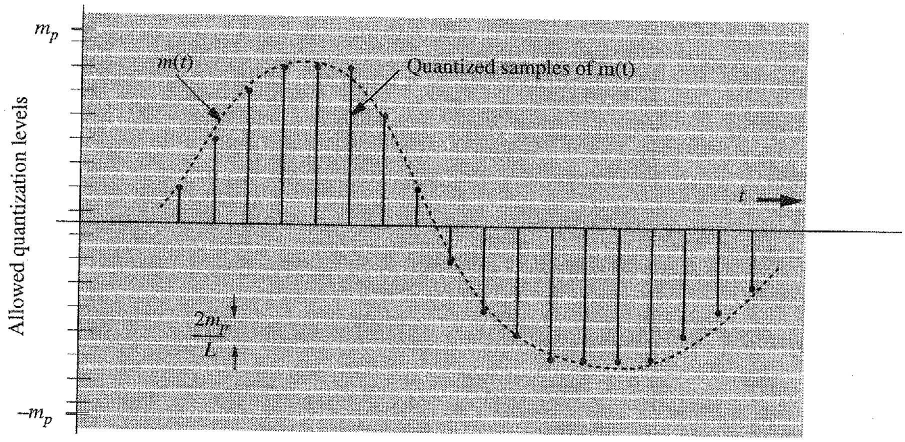

Figure 6.14 Quantization of a sampled analog signal.

be converted into a digital signal by means of sampling and **quantizing**, that is, rounding off its value to one of the closest permissible numbers (or **quantized levels**), as shown in Fig. 6.14. The amplitudes of the analog signal $m(t)$ lie in the range $(-m_p, m_p)$, which is partitioned into L subintervals, each of magnitude $\Delta v = 2m_p/L$. Next, each sample amplitude is approximated by the midpoint value of the subinterval in which the sample falls (see Fig. 6.14 for $L = 16$). Each sample is now approximated to one of the L numbers. Thus, the signal is digitized, with quantized samples taking on any one of the L values. Such a signal is known as an ***L*-ary digital signal**.

From practical viewpoint, a binary digital signal (a signal that can take on only two values) is very desirable because of its simplicity, economy, and ease of engineering. We can convert an L-ary signal into a binary signal by using pulse coding. Such a coding for the case of $L = 16$ was shown in Fig. 1.5. This code, formed by binary representation of the 16 decimal digits from 0 to 15, is known as the **natural binary code (NBC)**. Other possible ways of assigning a binary code will be discussed later. Each of the 16 levels to be transmitted is assigned one binary code of four digits. The analog signal $m(t)$ is now converted to a (binary) digital signal. A **bin**ary dig**it** is called a **bit** for convenience. This contraction of "binary digit" to "bit" has become an industry standard abbreviation and is used throughout the book.

Thus, each sample in this example is encoded by four bits. To transmit this binary data, we need to assign a distinct pulse shape to each of the two bits. One possible way is to assign a negative pulse to a binary **0** and a positive pulse to a binary **1** (Fig. 1.5) so that each sample is now transmitted by a group of four binary pulses (pulse code). The resulting signal is a binary signal.

The audio signal bandwidth is about 15 kHz. However, for speech, subjective tests show that signal articulation (intelligibility) is not affected if all the components above 3400 Hz are suppressed.*[,3] Since the objective in telephone communication is intelligibility rather than high fidelity, the components above 3400 Hz are eliminated by a low-pass filter. The resulting signal is then sampled at a rate of 8000 samples per second (8 kHz). This rate is intentionally kept higher than the Nyquist sampling rate of 6.8 kHz so that realizable filters can be applied for signal reconstruction. Each sample is finally quantized into 256 levels ($L = 256$), which requires a group of eight binary pulses to encode each sample ($2^8 = 256$). Thus, a telephone signal requires $8 \times 8000 = 64,000$ binary pulses per second.

* Components below 300 Hz may also be suppressed without affecting the articulation.

The compact disc (CD) is a more recent application of PCM. This is a high-fidelity situation requiring the audio signal bandwidth to be 20 kHz. Although the Nyquist sampling rate is only 40 kHz, the actual sampling rate of 44.1 kHz is used for the reason mentioned earlier. The signal is quantized into a rather large number ($L = 65,536$) of quantization levels, each of which is represented by 16 bits to reduce the quantizing error. The binary-coded samples (1.4 million bit/s) are then recorded on the compact disc.

6.2.1 Advantages of Digital Communication

Here are some of the advantages of digital communication over analog communication.

1. Digital communication, which can withstand channel noise and distortion much better than analog as long as the noise and the distortion are within limits, is more rugged than analog communication. With analog messages, on the other hand, any distortion or noise, no matter how small, will distort the received signal.

2. The greatest advantage of digital communication over analog communication, however, is the viability of regenerative repeaters in the former. In an analog communication system, a message signal becomes progressively weaker as it travels along the channel, whereas the cumulative channel noise and the signal distortion grow progressively stronger. Ultimately the signal is overwhelmed by noise and distortion. Amplification offers little help because it enhances the signal and the noise by the same proportion. Consequently, the distance over which an analog message can be transmitted is limited by the initial transmission power. For digital communications, a long transmission path may also lead to overwhelming noise and interferences. The trick, however, is to set up repeater stations along the transmission path at distances short enough to be able to detect signal pulses before the noise and distortion have a chance to accumulate sufficiently. At each repeater station the pulses are detected, and new, clean pulses are transmitted to the next repeater station, which, in turn, duplicates the same process. If the noise and distortion are within limits (which is possible because of the closely spaced repeaters), pulses can be detected correctly.* This way the digital messages can be transmitted over longer distances with greater reliability. The most significant error in PCM comes from quantizing. This error can be reduced as much as desired by increasing the number of quantizing levels, the price of which is paid in an increased bandwidth of the transmission medium (channel).

3. Digital hardware implementation is flexible and permits the use of microprocessors, digital switching, and large-scale integrated circuits.

4. Digital signals can be coded to yield extremely low error rates and high fidelity as well as for privacy.

5. It is easier and more efficient to multiplex several digital signals.

6. Digital communication is inherently more efficient than analog in exchanging SNR for bandwidth.

7. Digital signal storage is relatively easy and inexpensive. It also has the ability to search and select information from distant electronic database.

8. Reproduction with digital messages can be extremely reliable without deterioration. Analog messages such as photocopies and films, for example, lose quality at each successive stage of reproduction and must be transported physically from one distant place to another, often at relatively high cost.

* The error in pulse detection can be made negligible.

9. The cost of digital hardware continues to halve every two or three years, while performance or capacity doubles over the same time period. And there is no end in sight yet to this breathtaking and relentless exponential progress in digital technology. As a result, digital technologies today dominate in any given area of communication or storage technologies.

A Historical Note

The ancient Indian writer Pingala applied what turns out to be advanced mathematical concepts for describing prosody, and in doing so presented the first known description of a binary numeral system, possibly as early as the eighth century BCE.[6] Others, like R. Hall in Mathematics of Poetry place him later, circa 200 BCE. Gottfried Wilhelm Leibniz (1646–1716) was the first mathematician in the West to work out systematically the binary representation (using 1s and 0s) for any number. He felt a spiritual significance in this discovery, believing that **1**, representing unity, was clearly a symbol for God, while **0** represented nothingness. He reasoned that if all numbers can be represented merely by the use of **1** and **0**, this surely proves that God created the universe out of nothing!

6.2.2 Quantizing

As mentioned earlier, digital signals come from a variety of sources. Some sources such as computers are inherently digital. Some sources are analog, but are converted into digital form by a variety of techniques such as PCM and delta modulation (DM), which will now be analyzed. The rest of this section provides quantitative discussion of PCM and its various aspects, such as quantizing, encoding, synchronizing, the required transmission bandwidth and SNR.

For quantization, we limit the amplitude of the message signal $m(t)$ to the range $(-m_p, m_p)$, as shown in Fig. 6.14. Note that m_p is not necessarily the peak amplitude of $m(t)$. The amplitudes of $m(t)$ beyond $\pm m_p$ are simply chopped off. Thus, m_p is not a parameter of the signal $m(t)$; rather, it is the limit of the quantizer. The amplitude range $(-m_p, m_p)$ is divided into L uniformly spaced intervals, each of width $\Delta v = 2m_p/L$. A sample value is approximated by the midpoint of the interval in which it lies (Fig. 6.14). The quantized samples are coded and transmitted as binary pulses. At the receiver some pulses may be detected incorrectly. Hence, there are two sources of error in this scheme: *quantization error* and *pulse detection error*. In almost all practical schemes, the pulse detection error is quite small compared to the quantization error and can be ignored. In the present analysis, therefore, we shall assume that the error in the received signal is caused exclusively by quantization.

If $m(kT_s)$ is the kth sample of the signal $m(t)$, and if $\hat{m}(kT_s)$ is the corresponding quantized sample, then from the interpolation formula in Eq. (6.10),

$$m(t) = \sum_k m(kT_s) \text{ sinc } (2\pi Bt - k\pi)$$

and

$$\hat{m}(t) = \sum_k \hat{m}(kT_s) \text{ sinc } (2\pi Bt - k\pi)$$

where $\hat{m}(t)$ is the signal reconstructed from quantized samples. The distortion component $q(t)$ in the reconstructed signal is $q(t) = \hat{m}(t) - m(t)$. Thus,

$$\begin{aligned} q(t) &= \sum_k [\hat{m}(kT_s) - m(kT_s)] \operatorname{sinc}(2\pi Bt - k\pi) \\ &= \sum_k q(kT_s) \operatorname{sinc}(2\pi Bt - k\pi) \end{aligned}$$

where $q(kT_s)$ is the quantization error in the kth sample. The signal $q(t)$ is the undesired signal, and, hence, acts as noise, known as **quantization noise**. To calculate the power, or the mean square value of $q(t)$, we have

$$\begin{aligned} \widetilde{q^2(t)} &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} q^2(t)\, dt \\ &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left[\sum_k q(kT_s) \operatorname{sinc}(2\pi Bt - k\pi)\right]^2 dt \end{aligned} \tag{6.29a}$$

We can show that (see Prob. 3.7-4) the signals $\operatorname{sinc}(2\pi Bt - m\pi)$ and $\operatorname{sinc}(2\pi Bt - n\pi)$ are orthogonal, that is,

$$\int_{-\infty}^{\infty} \operatorname{sinc}(2\pi Bt - m\pi) \operatorname{sinc}(2\pi Bt - n\pi)\, dt = \begin{cases} 0 & m \neq n \\ \dfrac{1}{2B} & m = n \end{cases} \tag{6.29b}$$

Because of this result, the integrals of the cross-product terms on the right-hand side of Eq. (6.29a) vanish, and we obtain

$$\begin{aligned} \widetilde{q^2(t)} &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \sum_k q^2(kT_s) \operatorname{sinc}^2(2\pi Bt - k\pi)\, dt \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_k q^2(kT_s) \int_{-T/2}^{T/2} \operatorname{sinc}^2(2\pi Bt - k\pi)\, dt \end{aligned}$$

From the orthogonality relationship (6.29b), it follows that

$$\widetilde{q^2(t)} = \lim_{T\to\infty} \frac{1}{2BT} \sum_k q^2(kT_s) \tag{6.30}$$

Because the sampling rate is $2B$, the total number of samples over the averaging interval T is $2BT$. Hence, the right-hand side of Eq. (6.30) represents the average, or the mean of the square of the quantization error. The quantum levels are separated by $\Delta v = 2m_p/L$. Since a sample value is approximated by the midpoint of the subinterval (of height Δv) in which the sample falls, the maximum quantization error is $\pm\Delta v/2$. Thus, the quantization error lies in the range $(-\Delta v/2,\ \Delta v/2)$, where

$$\Delta v = \frac{2m_p}{L} \tag{6.31}$$

Assuming that the error is equally likely to lie anywhere in the range $(-\Delta v/2,\ \Delta v/2)$, the mean square quantizing error $\widetilde{q^2}$ is given by*

$$\widetilde{q^2} = \frac{1}{\Delta v}\int_{-\Delta v/2}^{\Delta v/2} q^2\, dq$$

$$= \frac{(\Delta v)^2}{12} \tag{6.32}$$

$$= \frac{m_p^2}{3L^2} \tag{6.33}$$

Because $\widetilde{q^2(t)}$ is the mean square value or power of the quantization noise, we shall denote it by N_q,

$$N_q = \widetilde{q^2(t)} = \frac{m_p^2}{3L^2}$$

Assuming that the pulse detection error at the receiver is negligible, the reconstructed signal $\hat{m}(t)$ at the receiver output is

$$\hat{m}(t) = m(t) + q(t)$$

The desired signal at the output is $m(t)$, and the (quantization) noise is $q(t)$. Since the power of the message signal $m(t)$ is $\widetilde{m^2(t)}$, then

$$S_o = \widetilde{m^2(t)}$$

$$N_o = N_q = \frac{m_p^2}{3L^2}$$

and

$$\frac{S_o}{N_o} = 3L^2 \frac{\widetilde{m^2(t)}}{m_p^2} \tag{6.34}$$

In this equation, m_p is the peak amplitude value that a quantizer can accept, and is therefore a parameter of the quantizer. This means S_o/N_o, the SNR, is a linear function of the message signal power $\widetilde{m^2(t)}$ (see Fig. 6.18 with $\mu = 0$).

---

* Those who are familiar with the theory of probability can derive this result directly by noting that the probability density of the quantization error q is $1/(2m_p/L) = L/2m_p$ over the range $|q| \le m_p/L$ and is zero elsewhere. Hence,

$$\widetilde{q^2} = \int_{-m_p/L}^{m_p/L} q^2 p(q)\, dq = \int_{-m_p/L}^{m_p/L} \frac{L}{2m_p} q^2\, dq = \frac{m_p^2}{3L^2}$$

6.2.3 Principle of Progressive Taxation: Nonuniform Quantization

Recall that S_o/N_o, the SNR, is an indication of the quality of the received signal. Ideally we would like to have a constant SNR (the same quality) for all values of the message signal power $\widetilde{m^2}(t)$. Unfortunately, the SNR is directly proportional to the signal power $\widetilde{m^2}(t)$, which varies from speaker to speaker by as much as 40 dB (a power ratio of 10^4). The signal power can also vary because of the different lengths of the connecting circuits. This indicates that the SNR in Eq. (6.34) can vary widely, depending on the speaker and the length of the circuit. Even for the same speaker, the quality of the received signal will deteriorate markedly when the person speaks softly. Statistically, it is found that smaller amplitudes predominate in speech and larger amplitudes are much less frequent. This means the SNR will be low most of the time.

The root of this difficulty lies in the fact that the quantizing steps are of uniform value $\Delta v = 2m_p/L$. The quantization noise $N_q = (\Delta v)^2/12$ [Eq. (6.32)] is directly proportional to the square of the step size. The problem can be solved by using smaller steps for smaller amplitudes (nonuniform quantizing), as shown in Fig. 6.15a. The same result is obtained by first compressing signal samples and then using a uniform quantization. The input-output characteristics of a compressor are shown in Fig. 6.15b. The horizontal axis is the normalized input signal (i.e., the input signal amplitude m divided by the signal peak value m_p). The vertical axis is the output signal y. The compressor maps input signal increments Δm into larger increments Δy for small input signals, and vice versa for large input signals. Hence, a given interval Δm contains a larger number of steps (or smaller step size) when m is small. The quantization noise is lower for smaller input signal power. An approximately logarithmic compression characteristic yields a quantization noise nearly proportional to the signal power $\widetilde{m^2}(t)$, thus making the SNR practically independent of the input signal power over a large dynamic range[5] (see later Fig. 6.18). This approach of equalizing the SNR appears similar to the use of progressive income tax to equalize incomes. The loud talkers and stronger signals are penalized with higher noise steps Δv to compensate the soft talkers and weaker signals.

Among several choices, two compression laws have been accepted as desirable standards by the ITU-T:[6] the μ-law used in North America and Japan, and the A-law used in Europe and the rest of the world and on international routes. Both the μ-law and the A-law curves have odd symmetry about the vertical axis. The μ-law (for positive amplitudes) is given by

$$y = \frac{1}{\ln(1+\mu)} \ln\left(1 + \frac{\mu m}{m_p}\right) \qquad 0 \le \frac{m}{m_p} \le 1 \tag{6.35a}$$

The A-law (for positive amplitudes) is

$$y = \begin{cases} \dfrac{A}{1+\ln A}\left(\dfrac{m}{m_p}\right) & 0 \le \dfrac{m}{m_p} \le \dfrac{1}{A} \\[2ex] \dfrac{1}{1+\ln A}\left(1 + \ln \dfrac{Am}{m_p}\right) & \dfrac{1}{A} \le \dfrac{m}{m_p} \le 1 \end{cases} \tag{6.35b}$$

These characteristics are shown in Fig. 6.16.

The compression parameter μ (or A) determines the degree of compression. To obtain a nearly constant S_o/N_o over a dynamic range of for input signal power 40 dB, μ should be greater than 100. Early North American channel banks and other digital terminals used a value of $\mu = 100$, which yielded the best results for 7-bit (128-level) encoding. An optimum value

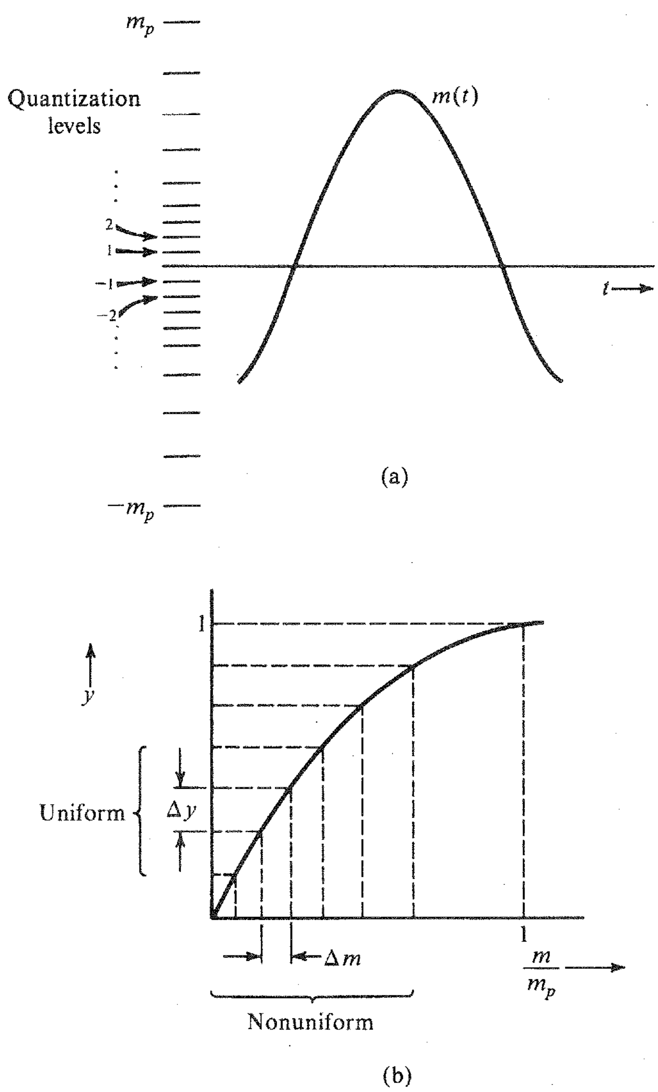

Figure 6.15 Nonuniform quantization.

of $\mu = 255$ has been used for all North American 8-bit (256-level) digital terminals, and the earlier value of μ is now almost extinct. For the A-law, a value of $A = 87.6$ gives comparable results and has been standardized by the ITU-T.[6]

The compressed samples must be restored to their original values at the receiver by using an expander with a characteristic complementary to that of the compressor. The compressor and the expander together are called the **compandor**. Figure 6.17 describes the use of compressor and expander along with a uniform quantizer to achieve nonuniform quantization.

Generally speaking, time compression of a signal increases its bandwidth. But in PCM, we are compressing not the signal $m(t)$ in time but its sample values. Because neither the time scale not the number of samples changes, the problem of bandwidth increase does not arise here. It happens that when a μ-law compandor is used, the output SNR is

$$\frac{S_o}{N_o} \simeq \frac{3L^2}{[\ln(1+\mu)]^2} \qquad \mu^2 \gg \frac{m_p^2}{\widetilde{m^2(t)}} \tag{6.36}$$

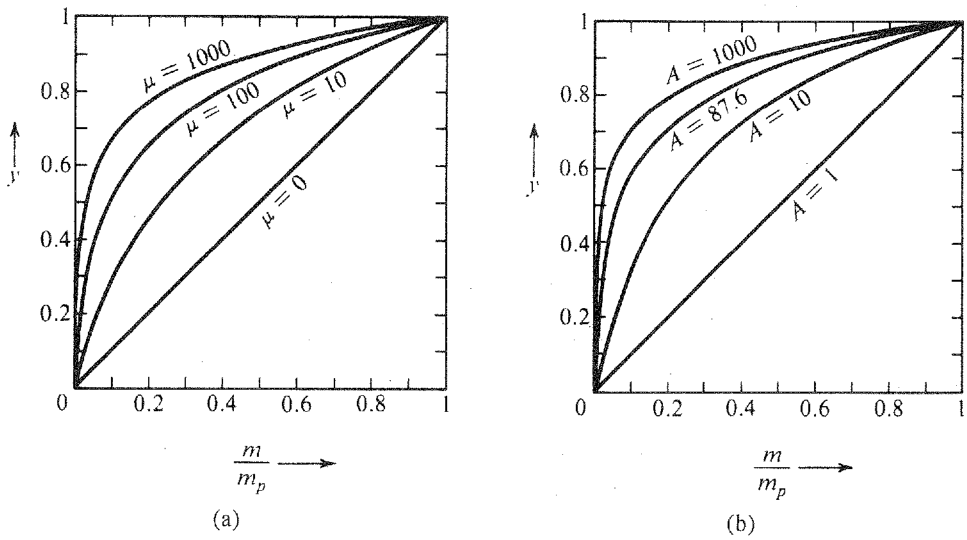

Figure 6.16 (a) μ-Law characteristic. (b) A-Law characteristic.

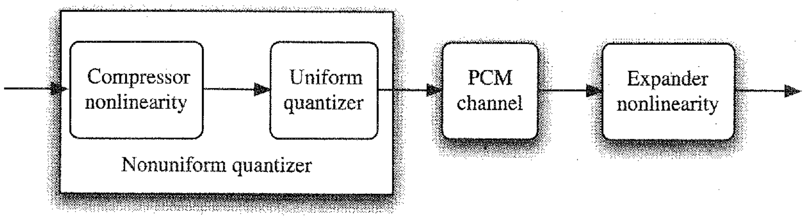

Figure 6.17 Utilization of compressor and expander for nonuniform quantization.

The output SNR for the cases of $\mu = 255$ and $\mu = 0$ (uniform quantization) as a function of $\overline{m^2(t)}$ (the message signal power) is shown in Fig. 6.18.

The Compandor

A logarithmic compressor can be realized by a semiconductor diode, because the V–I characteristic of such a diode is of the desired form in the first quadrant:

$$V = \frac{KT}{q} \ln\left(1 + \frac{I}{I_s}\right)$$

Two matched diodes in parallel with opposite polarity provide the approximate characteristic in the first and third quadrants (ignoring the saturation current). In practice, adjustable resistors are placed in series with each diode and a third variable resistor is added in parallel. By adjusting various resistors, the resulting characteristic is made to fit a finite number of points (usually seven) on the ideal characteristics.

An alternative approach is to use a piecewise linear approximation to the logarithmic characteristics. A 15-segmented approximation (Fig. 6.19) to the eighth bit ($L = 256$) with $\mu = 255$ law is widely used in the D2 channel bank that is used in conjunction with the T1 carrier system. The segmented approximation is only marginally inferior in terms of SNR.[8] The piecewise linear approximation has almost universally replaced earlier logarithmic approximations to the true $\mu = 255$ characteristic and is the method of choice in North American standards.

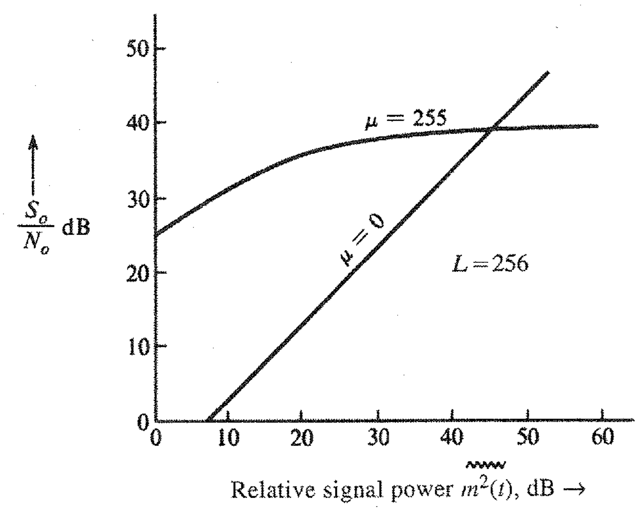

Figure 6.18 Ratio of signal to quantization noise in PCM with and without compression.

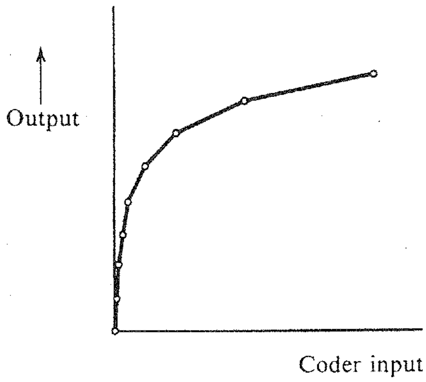

Figure 6.19 Piecewise linear compressor characteristic.

Though a true $\mu = 255$ compressor working with a $\mu = 255$ expander will be superior to similar piecewise linear devices, a digital terminal device exhibiting the true characteristic in today's network must work end-to-end against other network elements that use the piecewise linear approximation. Such a combination of differing characteristics is inferior to either of the characteristics obtained when the compressor and the expander operate using the same compression law.

In the standard audio file format used by Sun, Unix and Java, the audio in "au" files can be pulse-code-modulated or compressed with the ITU-T G.711 standard through either the μ-law or the A-law.[6] The μ-law compressor ($\mu = 255$) converts 14-bit signed linear PCM samples to logarithmic 8-bit samples, leading to storage saving. The A-law compressor ($A = 87.6$) converts 13-bit signed linear PCM samples to logarithmic 8-bit samples. In both cases, sampling at the rate of 8000 Hz, a G.77 encoder thus creates from audio signals bit streams at 64 kilobits per second (kbit/s). Since the A-law and the μ-law are mutually compatible, audio recoded into "au" files can be decoded in either format. It should be noted that the Microsoft WAV audio format also has compression options that use μ-law and A-law.

The PCM Encoder

The multiplexed PAM output is applied at the input of the encoder, which quantizes and encodes each sample into a group of n binary digits. A variety of encoders is available.[7, 10] We shall discuss here the **digit-at-a-time** encoder, which makes n sequential comparisons to generate an n-bit codeword. The sample is compared with a voltage obtained by a combination of reference voltages proportional to 2^7, 2^6, $2^5, \dots,$ 2^0. The reference voltages are conveniently generated by a bank of resistors R, $2R$, 2^2R, $\dots$, 2^7R.

The encoding involves answering successive questions, beginning with whether the sample is in the upper or lower half of the allowed range. The first code digit **1** or **0** is generated, depending on whether the sample is in the upper or the lower half of the range. In the second step, another digit **1** or **0** is generated, depending on whether the sample is in the upper or the lower half of the subinterval in which it has been located. This process continues until the last binary digit in the code has been generated.

Decoding is the inverse of encoding. In this case, each of the n digits is applied to a resistor of different value. The kth digit is applied to a resistor 2^kR. The currents in all the resistors are added. The sum is proportional to the quantized sample value. For example, a binary code word **10010110** will give a current proportional to $2^7 + 0 + 0 + 2^4 + 0 + 2^2 + 2^1 + 0 = 150$. This completes the D/A conversion.

6.2.4 Transmission Bandwidth and the Output SNR

For a binary PCM, we assign a distinct group of n binary digits (bits) to each of the L quantization levels. Because a sequence of n binary digits can be arranged in 2^n distinct patterns,

$$L = 2^n \qquad \text{or} \qquad n = \log_2 L \tag{6.37}$$

each quantized sample is, thus, encoded into n bits. Because a signal $m(t)$ band-limited to B Hz requires a minimum of $2B$ samples per second, we require a total of $2nB$ bit/s, that is, $2nB$ pieces of information per second. Because a unit bandwidth (1 Hz) can transmit a maximum of two pieces of information per second (Sec. 6.1.3), we require a minimum channel of bandwidth B_T Hz, given by

$$B_T = nB \text{ Hz} \tag{6.38}$$

This is the theoretical minimum transmission bandwidth required to transmit the PCM signal. In Secs. 7.2 and 7.3, we shall see that for practical reasons we may use a transmission bandwidth higher than this minimum.

Example 6.2 A signal $m(t)$ band-limited to 3 kHz is sampled at a rate $33\frac{1}{3}\%$ higher than the Nyquist rate. The maximum acceptable error in the sample amplitude (the maximum quantization error) is 0.5% of the peak amplitude m_p. The quantized samples are binary coded. Find the minimum bandwidth of a channel required to transmit the encoded binary signal. If 24 such signals are time-division-multiplexed, determine the minimum transmission bandwidth required to transmit the multiplexed signal.

The Nyquist sampling rate is $R_N = 2 \times 3000 = 6000$ Hz (samples per second). The actual sampling rate is $R_A = 6000 \times (1\frac{1}{3}) = 8000$ Hz.

The quantization step is Δv, and the maximum quantization error is $\pm\Delta v/2$.

Therefore, from Eq. (6.31),

$$\frac{\Delta v}{2} = \frac{m_p}{L} = \frac{0.5}{100} m_p \Longrightarrow L = 200$$

For binary coding, L must be a power of 2. Hence, the next higher value of L that is a power of 2 is $L = 256$.

From Eq. (6.37), we need $n = \log_2 256 = 8$ bits per sample. We require to transmit a total of $C = 8 \times 8000 = 64,000$ bit/s. Because we can transmit up to 2 bit/s per hertz of bandwidth, we require a minimum transmission bandwidth $B_T = C/2 = 32$ kHz.

The multiplexed signal has a total of $C_M = 24 \times 64,000 = 1.536$ Mbit/s, which requires a minimum of $1.536/2 = 0.768$ MHz of transmission bandwidth.

Exponential Increase of the Output SNR

From Eq. (6.37), $L^2 = 2^{2n}$, and the output SNR in Eq. (6.34) or Eq. (6.36) can be expressed as

$$\frac{S_o}{N_o} = c(2)^{2n} \tag{6.39}$$

where

$$c = \begin{cases} \dfrac{3\,\widetilde{m^2(t)}}{m_p^2} & \text{[uncompressed case, in Eq. (6.34)]} \\[2ex] \dfrac{3}{[\ln(1+\mu)]^2} & \text{[compressed case, in Eq. (6.36)]} \end{cases}$$

Substitution of Eq. (6.38) into Eq. (6.39) yields

$$\frac{S_o}{N_o} = c(2)^{2B_T/B} \tag{6.40}$$

From Eq. (6.40) we observe that the SNR increases exponentially with the transmission bandwidth B_T. This trade of SNR for bandwidth is attractive and comes close to the upper theoretical limit. A small increase in bandwidth yields a large benefit in terms of SNR. This relationship is clearly seen by using the decibel scale to rewrite Eq. (6.39) as

$$\begin{aligned} \left(\frac{S_o}{N_o}\right)_{\text{dB}} &= 10 \log_{10}\left(\frac{S_o}{N_o}\right) \\ &= 10 \log_{10}[c(2)^{2n}] \\ &= 10 \log_{10} c + 2n \log_{10} 2 \\ &= (\alpha + 6n) \quad \text{dB} \end{aligned} \tag{6.41}$$

where $\alpha = 10 \log_{10} c$. This shows that increasing n by 1 (increasing one bit in the codeword) quadruples the output SNR (a 6 dB increase). Thus, if we increase n from 8 to 9, the SNR quadruples, but the transmission bandwidth increases only from 32 kHz to 36 kHz (an increase of only 12.5%). This shows that in PCM, SNR can be controlled by transmission bandwidth. We shall see later that frequency and phase modulation also do this. But it requires a doubling of the bandwidth to quadruple the SNR. In this respect, PCM is strikingly superior to FM or PM.

Example 6.3 A signal $m(t)$ of bandwidth $B = 4$ kHz is transmitted using a binary companded PCM with $\mu = 100$. Compare the case of $L = 64$ with the case of $L = 256$ from the point of view of transmission bandwidth and the output SNR.

For $L = 64$, $n = 6$, and the transmission bandwidth is $nB = 24$ kHz,

$$\frac{S_o}{N_o} = (\alpha + 36) \text{ dB}$$

$$\alpha = 10 \log \frac{3}{[\ln (101)]^2} = -8.51$$

Hence,

$$\frac{S_o}{N_o} = 27.49 \text{ dB}$$

For $L = 256$, $n = 8$, and the transmission bandwidth is 32 kHz,

$$\frac{S_o}{N_o} = \alpha + 6n = 39.49 \text{ dB}$$

The difference between the two SNRs is 12 dB, which is a ratio of 16. Thus, the SNR for $L = 256$ is 16 times the SNR for $L = 64$. The former requires just about 33% more bandwidth compared to the latter.

Comments on Logarithmic Units

Logarithmic units and logarithmic scales are very convenient when a variable has a large dynamic range. Such is the case with frequency variables or SNRs. A logarithmic unit for the power ratio is the decibel (dB), defined as $10 \log_{10}$ (power ratio). Thus, an SNR is x dB, where

$$x = 10 \log_{10} \frac{S}{N}$$

We use the same unit to express power gain or loss over a certain transmission medium. For instance, if over a certain cable the signal power is attenuated by a factor of 15, the cable gain is

$$G = 10 \log_{10} \frac{1}{15} = -11.76 \text{ dB}$$

or the cable attenuation (loss) is 11.76 dB.

Although the decibel is a measure of power ratios, it is often used as a measure of power itself. For instance, "100 watt" may be considered to be a power ratio of 100 with respect to 1-watt power, and is expressed in units of dBW as

$$P_{\text{dBW}} = 10 \log_{10} 100 = 20 \text{ dBW}$$

Thus, 100-watt power is 20 dBW. Similarly, power measured with respect to 1 mW power is dBm. For instance, 100-watt power is

$$P_{\text{dBm}} = 10 \log \frac{100\text{ W}}{1\text{ mW}} = 50 \text{ dBm}$$

6.3 DIGITAL TELEPHONY: PCM IN T1 CARRIER SYSTEMS

A Historical Note

Because of the unavailability of suitable switching devices, more than 20 years elapsed between the invention of PCM and its implementation. Vacuum tubes, used before the invention of the transistor, were not only bulky, but they were poor switches and dissipated a lot of heat. Systems having vacuum tubes as switches were large, rather unreliable, and tended to overheat. PCM was just waiting for the invention of the transistor, which happens to be a small device that consumes little power and is a nearly ideal switch.

Coincidentally, at about the time the transistor was invented, the demand for telephone service had become so heavy that the existing system was overloaded, particularly in large cities. It was not easy to install new underground cables because space available under the streets in many cities was already occupied by other services (water, gas, sewer, etc.). Moreover, digging up streets and causing many dislocations was not very attractive. An attempt was made on a limited scale to increase the capacity by frequency-division-multiplexing several voice channels through amplitude modulation. Unfortunately, the cables were primarily designed for the audio voice range (0–4 kHz) and suffered severely from noise. Furthermore, cross talk between pairs of channels on the same cable was unacceptable at high frequencies. Ironically, PCM—requiring a bandwidth several times larger than that required for FDM signals—offered the solution. This is because digital systems with closely spaced regenerative repeaters can work satisfactorily on noisy lines that give poor high-frequency performance.[9] The repeaters, spaced approximately 6000 feet apart, clean up the signal and regenerate new pulses before the pulses get too distorted and noisy. This is the history of the Bell System's T1 carrier system.[3, 10] A pair of wires that used to transmit one audio signal of bandwidth 4 kHz is now used to transmit 24 time-division-multiplexed PCM telephone signals with a total bandwidth of 1.544 MHz.

T1 Time Division Multiplexing

A schematic of a T1 carrier system is shown in Fig. 6.20a. All 24 channels are sampled in a sequence. The sampler output represents a time-division-multiplexed PAM signal. The multiplexed PAM signal is now applied to the input of an encoder that quantizes each sample and encodes it into eight binary pulses—a binary codeword* (see Fig. 6.20b). The signal, now converted to digital form, is sent over the transmission medium. Regenerative repeaters spaced approximately 6000 feet apart detect the pulses and retransmit new pulses. At the receiver, the decoder converts the binary pulses into samples (decoding). The samples are then demultiplexed (i.e., distributed to each of the 24 channels). The desired audio signal is reconstructed by passing the samples through a low-pass filter in each channel.

* In an earlier version, each sample was encoded by seven bits. An additional bit was added for signaling.