J. C. Candy

Chapter 1

An Overview of Basic Concepts *

1.1 INTRODUCTION

This chapter reviews the main properties of oversampling techniques that are useful for converting signals between analog and digital formats. Oversampling has become popular in recent years because it avoids many of the difficulties encountered with conventional methods for analog-to-digital and digital-to-analog (A/D, D/A) conversion, especially for those applications that call for high-resolution representation of relatively low-frequency signals.

Conventional converters, illustrated in Figure 1.1, are often difficult to implement in fine-line very large scale integration (VLSI) technology. These difficulties arise because conventional methods need precise analog components in their filters and conversion circuits and because their circuits can be very vulnerable to noise and interference. The virtue of the conventional methods is their use of a low sampling frequency, usually the Nyquist rate of the signal (i.e., twice the signal bandwidth).

A low-pass filter at the input to the encoder of Figure 1.1 attenuates high-frequency noise and out-of-band components of the signal that alias into the signal when sampled at the Nyquist rate. Properties of this filter are usually specified for each application. The A/D circuit can take a number of different forms, such as flash converters for fast operation, successive-approximation converters for moderate rates, and ramp converters for slow ones. At the decoder a filter smooths the sampled output of the D/A circuit; the amount of smoothing required is usually part of the specification of the system. The circuits of these conventional converters require high-accuracy analog components in order to achieve high overall resolution.

*This chapter is a rewrite of material from reference [1].

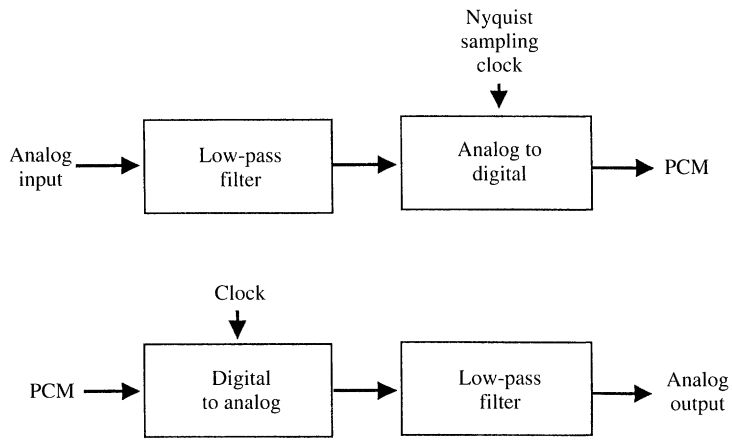

Figure 1.1 Conventional pulse code modulation (PCM), including analog filters for curtailing the aliasing noise in the encoder and for smoothing the output from the decoder.

Oversampling converters, illustrated in Figure 1.2, can use simple and relatively high-tolerance analog components to achieve high resolution, but they require fast and complex digital signal processing stages. These converters modulate the analog signal into a simple code, usually single-bit words, at a frequency much higher than the Nyquist rate.

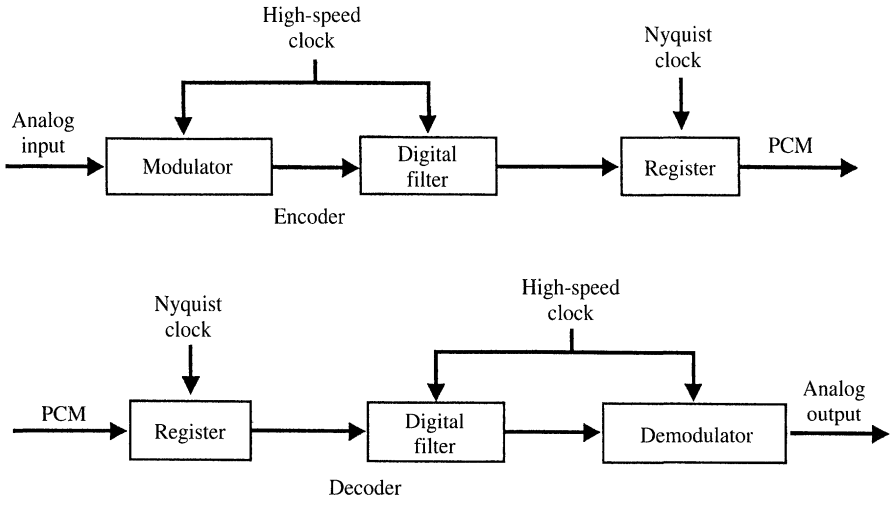

Figure 1.2 Oversampling pulse code modulation. The modulation and demodulation occur at sufficiently high sampling rate that digital filters can provide most for the antialiasing and smoothing functions.

We shall show that the design of the modulator can trade resolution in time for resolution in amplitude in such a way that imprecise analog circuits can be tolerated. The use of high-frequency modulation and demodulation eliminates the need for abrupt cutoffs in the analog antialiasing filter at the input to the A/D converter, as well as in the filters that smooth the analog output of the D/A converter. Digital filters are used instead as illustrated in Figure 1.2. A digital filter smooths the output of the modulator, attenuating noise, interference, and high-frequency components of the signal before they can alias into the signal band when the code is resampled at the Nyquist rate. Another digital filter interpolates the code in the decoder to a high word rate before it is demodulated to analog form.

Oversampling converters make extensive use of digital signal processing, taking advantage of the fact that fine-line VLSI is better suited for providing fast digital circuits than for providing precise analog circuits. Because their sampling rate usually needs to be several orders of magnitude higher than the Nyquist rate, oversampling methods are best suited for relatively low-frequency signals. They have found use in such applications as digital audio, digital telephony, and instrumentation. Future applications in video and radar systems are imminent as faster technologies become available.

An important difference between conventional converters and oversampling ones involve testing and specifying their performance. With conventional converters there is a one-to-one correspondence between input and output sample values, and hence one can describe their accuracy by comparing the values of corresponding input and output samples. In contrast there is no similar correspondence in oversampling converters because they inherently include digital low-pass filters, and hence each input sample value contributes to a whole train of output samples. Consequently, it has been useful to borrow techniques from communication technology to describe the performance of oversampling converters. Thus we measure their root-mean-square (rms) noise under various conditions, the distortion they introduce into sinusoidal signals, and their frequency responses. An important task in designing an oversampling converter is therefore the calculation of rms values of modulation noise and its spectral density. Examples of such calculations will be given in following sections.

This chapter is organized into four main sections. Following this introduction, Section 1.2 describes some basic properties of the quantization noise. It then introduces delta-sigma modulation as a technique for shaping the spectrum of quantization noise, moving most of the noise power to high frequencies, well outside the band of the signal, where it is removed by digital filtering. A number of other modulators are also described. Section 1.3 discusses the design of digital filters that decimate the modulated signal, converting it from a sequence of short digital words occurring at a high rate into long words occurring at the Nyquist rate. Section 1.4 describes oversampling D/A converters.

1.2 DIGITAL MODULATION

1.2.1 Quantization

Quantization of amplitude and sampling in time are at the heart of all digital modulators. Periodic sampling at rates more than twice the signal bandwidth need not introduce distortion, but quantization does, and our primary objective in designing modulators is to limit this distortion. We begin our discussion by describing some basic properties of

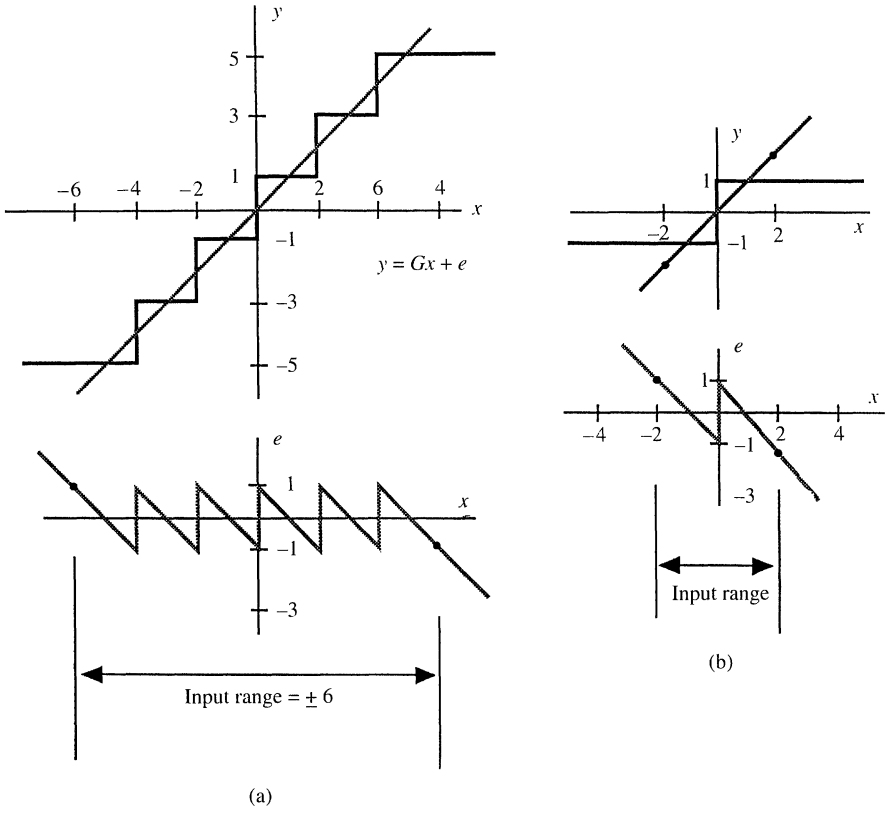

Figure 1.3 (a) An example of a uniform multilevel quantization characteristic that is represented by linear gain G and an error e. (b) For two-level quantization the gain G is arbitrary.

quantization that will be useful for specifying the noise from modulators. Figure 1.3(a) shows a uniform quantization that rounds off a continuous amplitude signal x to odd integers in the range ±5. In this example the level spacing Δ is 2. We will find it useful to represent the quantized signal y by a linear function Gx with an error e: that is,

$$y = Gx + e \tag{1.1}$$

The gain G is the slope of the straight line that passes through the center of the quantization characteristic so that, when the quantizer does not saturate (i.e., when $-6 \leq x \leq 6$), the error is bounded by $\pm\Delta/2$. Notice that the above consideration remains applicable to a two-level (single-bit) quantizer, as illustrated in Figure 1.3(b), but in this case the choice of gain G is arbitrary.

The error is completely defined by the input, but if the input changes randomly between samples by amounts comparable with or greater than the threshold spacing, without causing saturation, then the error is largely uncorrelated from sample to sample and has equal probability of lying anywhere in the range $\pm\Delta/2$. If we further assume that the error has statistical properties that are independent of the signal, then we can represent

it by a noise, and some important properties of modulators can be determined. In many cases experiments have confirmed these properties, but there are two important instances where they may not apply: when the input is constant, and when it changes regularly by multiples or submultiples of the step size between sample times, as can happen in feedback circuits.

When we treat the quantization error e as having equal probability of lying anywhere in the range $\pm\Delta/2$, its mean square value is given by

$$e_{\text{rms}}^2 = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} e^2\, de = \frac{\Delta^2}{12} \tag{1.2}$$

For the ensuing discussion of spectral densities of the noise, we shall employ a one-sided representation of frequencies: that is, we assume that all the power is in the positive range of frequencies. When a quantized signal is sampled at frequency $f_s = 1/T$, all of its power folds into the frequency band $0 \le f < f_s/2$. Then, if the quantization noise is white, the spectral density of the sampled noise is given by

$$E(f) = e_{\text{rms}}\sqrt{\frac{2}{f_s}} = e_{\text{rms}}\sqrt{2T} \tag{1.3}$$

We can use this result to analyze examples of oversampling modulators. Consider first ordinary pulse code modulation (PCM). A signal lying in the frequency band $0 \le f < f_0$, to which a dither signal contained in the band $f_0 \le f < f_s/2$ is added, is pulse code modulated at f_s. The oversampling ratio (OSR), defined as the ratio of the sampling frequency f_s to the Nyquist frequency $2f_0$, is given by the integer

$$\text{OSR} = \frac{f_s}{2f_0} = \frac{1}{2f_0T} \tag{1.4}$$

If the dither is sufficiently large and busy to whiten and decorrelate the quantization error, the noise power that falls into the signal band will be given by

$$n_0^2 = \int_0^{f_0} e^2(f)\, df = e_{\text{rms}}^2(2f_0T) = \frac{e_{\text{rms}}^2}{\text{OSR}} \tag{1.5}$$

Thus we have the well-known result that oversampling reduces the in-band rms noise from ordinary quantization by the square root of the oversampling ratio. Therefore each doubling of the sampling frequency decreases the in-band noise by 3 dB, increasing the resolution by only half a bit.

1.2.2 Delta–Sigma Modulation

1.2.2.1 First–Order Feedback Quantizer. A more efficient oversampling quantizer is the delta–sigma ($\Delta\Sigma$) modulator shown in Figure 1.4(a). Although $\Delta\Sigma$ modulators usually employ two-level quantization, we commence our discussion by assuming the modulator contains a multilevel, uniform quantizer with unity gain $G = 1$. The input

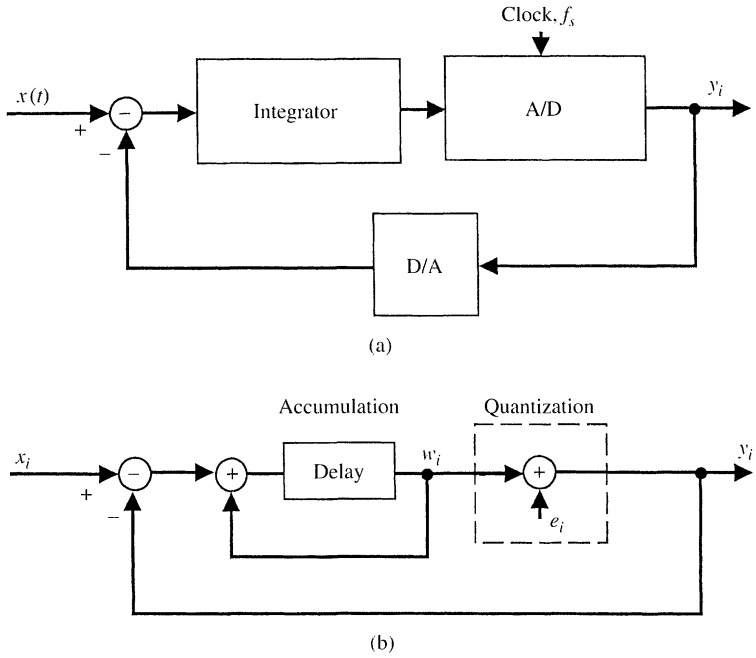

Figure 1.4 A block diagram of a ΔΣ quantizer and its sampled-data equivalent circuit.

to the circuit feeds to the quantizer via an integrator, and the quantized output feeds back to subtract from the input signal. This feedback forces the average value of the quantized signal to track the average input. Any persistent difference between them accumulates in the integrator and eventually corrects itself. Figure 1.5 illustrates the response of the circuit to a ramp input; it shows how the quantized signal oscillates between two levels that are adjacent to the input value in such a manner that its local average equals the average input value [2].

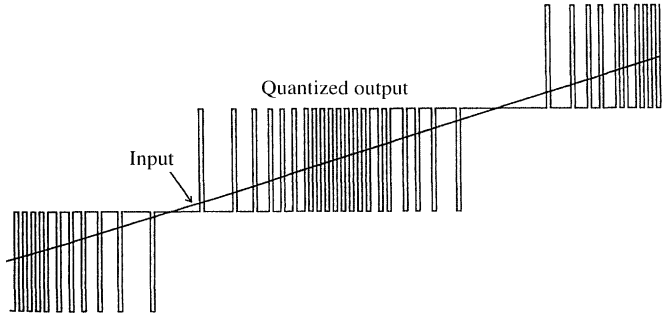

Figure 1.5 The response of a multilevel ΔΣ quantizer to a ramp input. A two-level response is obtained by curtailing input amplitude to a range of values that lies between two adjacent quantization levels.

1.2.2.2 Modulation Noise in Busy Signals. We analyze the modulator by means of the equivalent circuit shown in Figure 1.4(b). Here an added signal e represents the quantization error in accordance with Eq. (1.1) and the quantization gain G set to unity. Because this is a sampled-data circuit, we represent the integration by accumulation, also with unity gain. It can easily be shown that the output of the accumulator is

$$w_i = x_{i-1} - e_{i-1} \tag{1.6}$$

and the quantized signal is

$$y_i = x_{i-1} + (e_i - e_{i-1}) \tag{1.7}$$

Thus this circuit differentiates the quantization error, making the modulation error the first difference of the quantization error while leaving the signal unchanged, except for a delay.

To calculate the effective resolution of the $\Delta\Sigma$ modulator, we now assume that the input signal is sufficiently busy that the error e behaves as white noise that is uncorrelated with the signal. The spectral density of the modulation noise

$$n_i = e_i - e_{i-1} \tag{1.8}$$

may then be expressed as

$$N(f) = E(f)\left|1 - \varepsilon^{-j\omega T}\right| = 2e_{\text{rms}}\sqrt{2T}\sin\left(\frac{\omega T}{2}\right) \tag{1.9}$$

where $\omega = 2\pi f$.

Figure 1.6 compares this spectral density with that of the quantization noise when the oversampling ratio is 16. Clearly, feedback around the quantizer reduces the noise at low

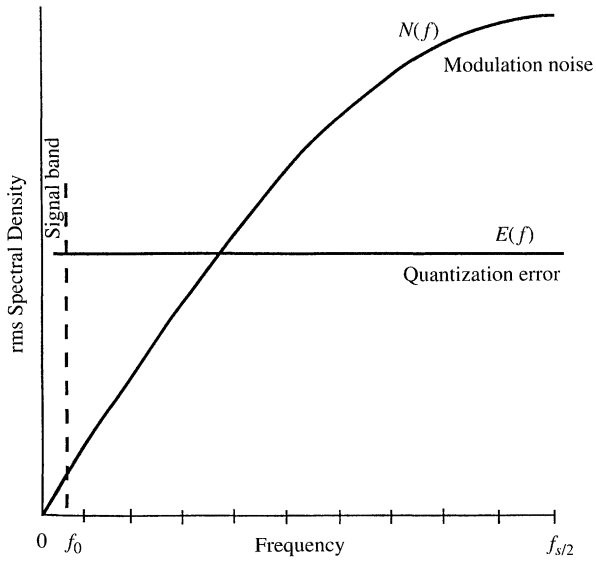

Figure 1.6 The spectral density of the noise $N(f)$ from $\Delta\Sigma$ quantization compared with that of ordinary quantization $E(f)$.

frequencies but increases it at high frequencies. The total noise power in the signal band is

$$n_0^2 = \int_0^{f_0} |N(f)|^2 \, df \approx e_{\text{rms}}^2 \frac{\pi^2}{3} (2f_0T)^3 \qquad f_s^2 >> f_0^2 \tag{1.10}$$

and its rms value is

$$n_0 \approx e_{\text{rms}} \frac{\pi}{\sqrt{3}} (2f_0T)^{3/2} = e_{\text{rms}} \frac{\pi}{\sqrt{3}} (\text{OSR})^{-3/2} \tag{1.11}$$

Each doubling of the oversampling ratio of this circuit reduces the noise by 9 dB and provides 1.5 bits of extra resolution. The improvement in resolution requires that the modulated signal be decimated to the Nyquist rate with a sharply selective digital filter. Otherwise, the high-frequency components of the noise will spoil the resolution when it is sampled at the Nyquist rate. Some early oversampling converters employed primitive decimation. One merely averaged the output samples of the modulator over each Nyquist interval to get a PCM signal. References [2] and [3] show that the rms noise in this PCM can be expressed as $\sqrt{2}e_{\text{rms}}(2f_0T)$. They also show that taking a triangularly weighted sum over each Nyquist interval gives an rms noise $4e_{\text{rms}}(2f_0T)^{1.5}$. An optimization of these techniques for attenuating the high-frequency noise is given in reference [4]. These decimators permit more noise to alias into the signal band than do the ones that employ filters having impulse responses that are longer than one Nyquist interval, but the techniques have been useful because their circuit implementation can be very simple.

This derivation of the average properties of modulation noise depends on representing the quantization error as white uncorrelated noise. But the analysis in Chapter 2, which does not depend on this assumption, shows that Eq. (1.11) may apply even when the error is not white. Moreover, it also shows that the quantization error is rarely truly white.

1.2.2.3 Pattern Noise from ΔΣ Modulation with dc Inputs. When the input to the modulator is a dc signal, the quantized signal bounces between two levels, keeping its mean value equal to the input. Figure 1.7 demonstrates that the oscillation may be repetitive; it returns to its starting condition after seven clock periods. The frequency of repetition depends on the input level; in this example the input is $3\Delta/7$ away from a level, and this results in a pattern that repeats every seven periods. When the repetition frequency lies in the signal band, the modulation is noisy, but when it does not, the modulation is quiet.

Figure 1.8 shows how the in-band rms modulation noise depends on the dc input level, for a ΔΣ modulator having quantization levels at ±1 and an oversampling ratio of 16. The decimating filter that processes the modulation is the one described in Section 1.3. There are peaks of noise adjacent to integer divisions of the space between levels; elsewhere the noise is small. This structure of the quantization noise is called *pattern noise.* The largest peaks can exceed the expected noise level [Eq. (1.11)], which is at −41 dB in this example.

Surprisingly, it is also quite easy to get a mathematical expression [5, 6] for the noise from ΣΔ modulation with dc input. Let x be the input level to the modulator and Y' the

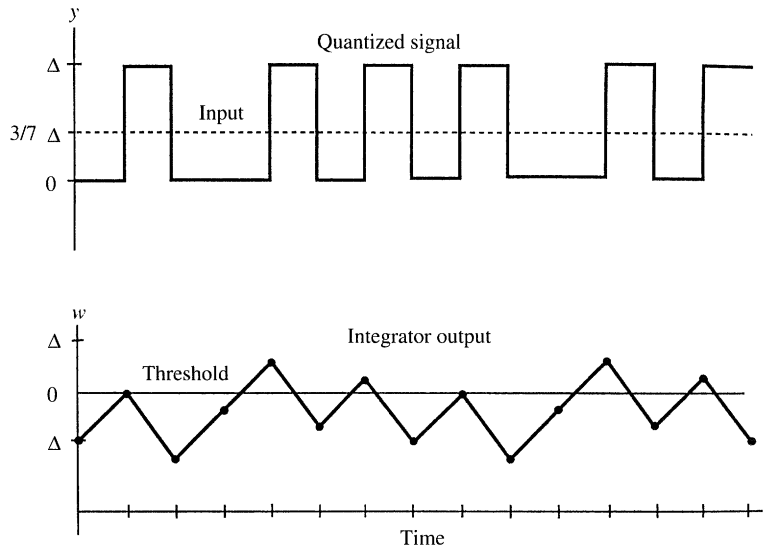

Figure 1.7 Waveforms in a ΔΣ circuit for a constant input situated $\frac{3}{7}\Delta$ above a quantization level.

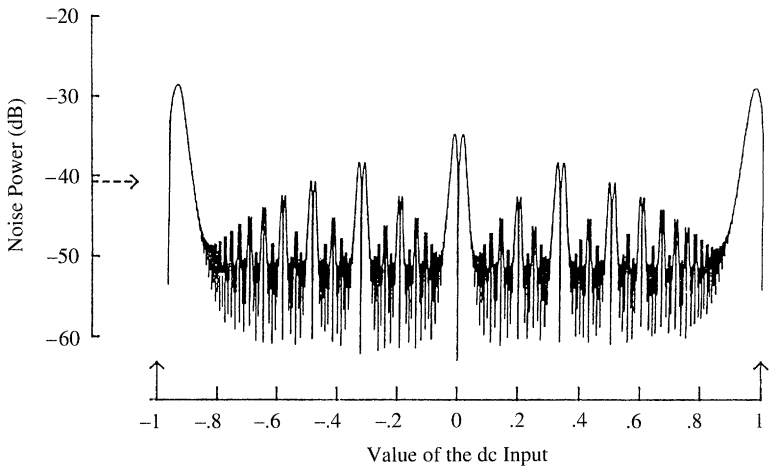

Figure 1.8 Noise from ΔΣ modulation for dc inputs. Quantization levels are at ±1, and the noise is plotted for dc inputs lying between these levels. Peaks of the noise occur adjacent to integer divisions of the level spacing.

adjacent quantization level. The modulator output can then be expressed as

$$y(t) = Y' + \sum_l \sum_k \frac{\sin(\pi l x')}{\pi l} \exp\left(j\pi + \frac{lx' + k}{T} t\right) \tag{1.12}$$

where $x' = (x - Y')/\Delta$. Thus, $y(t)$ has a dc component equal to the input x, accompanied by tones of frequency $(lx' + k)f_s$. The tones that lie in the signal band represent the inherent noise of the conversion. A sum of the powers of these tones, taking account of the response of the decimation filter, gives a good description of the pattern noise [5]. The following properties of pattern noise are noteworthy:

- The *height* of each peak is inversely proportional to the oversampling ratio.
- The *width* of each peak is inversely proportional to the oversampling ratio.
- The *power* in each peak is inversely proportional to the oversampling ratio cubed.
- The *height* and *width* of each peak are inversely proportional to the denominator of the reduced fraction that describes the position of the peak within the quantization interval relative to the level spacing. This fraction is $\frac{3}{7}$ in Figure 1.7.
- The average noise represented by the graph is given in Eq. (1.11).
- About half of the total power is in the end peaks and $\frac{1}{16}$ in the center ones.

The noise pattern in Figure 1.8 can be integrated against time as a function of a slowly changing input to get a measure of the noise introduced into a signal. Figure 1.9 shows the experimentally measured signal-to-noise ratio for two input sine waves, plotted against their amplitudes. Curve (a) corresponds to sine waves centered between two levels, and (b) corresponds to sine waves offset from center. A 0-dB input corresponds to a peak amplitude of $\Delta/2$. For comparison, the dotted line shows the values predicted from Eq. (1.11). The resolution is better than predicted when the larger peaks are not included, but it is sometimes worse when they are. The dependence of noise on signal values and the fact that the noise is composed of tones are reasons why this modulation is rarely used. When it is, dithering is usually applied to randomize the quantization noise and destroy tones that would be disturbing in audio applications, [7]. Dither will be described in Chapter 3.

1.2.2.4 Dead Zones in $\Delta\Sigma$ Modulation. The second graph in Figure 1.7 shows the output of the integrator for a steady input equal to $3\Delta/7$. Notice that this particular waveform may be raised as much as $\Delta/7$, with respect to the quantizer threshold level, without changing the sequence of decisions. Such a change of level at the output of the integrator corresponds to an impulse at its input; consequently, small fast changes of input may be ignored by the modulator, under certain conditions. It can be shown that the location and extent of the transient dead zones correspond in position and size with the peaks of noise in Figure 1.8. For most applications, the pattern noise is more noticeable than are the dead zones. However, when the integrator is a leaky one, having low dc gain, the dead zones can be significant.

We next describe some practical properties of the $\Delta\Sigma$ circuit because this simple modulator is useful for illustrating feedback quantization. Knowledge of its properties will help us to explain improved modulators.

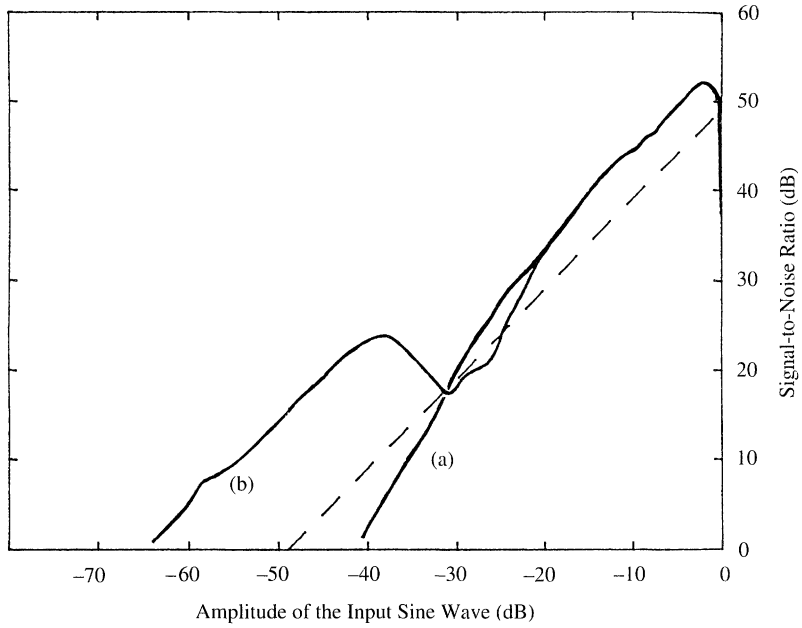

Figure 1.9 Graph of modulation noise plotted against the amplitude of applied sine waves; 0 dB corresponds to an amplitude of $\Delta/2$. Curve (a) is for sine waves centered midway between levels; curve (b) is for sine waves biased $\Delta/64$ away from center. The dashed line is the calculated noise.

1.2.2.5 Influence of Circuit Parameters on ΔΣ Modulation

NET GAIN IN THE FEEDBACK LOOP. Our discussion has so far assumed unity sample gain in every component of the modulator. Figure 1.10 shows a modulator that includes a constant gain G in the forward path of the feedback. Small deviations of this from unity have little effect on the overall properties, provided the net gain in the feedback loop is large. The gain of the accumulator is

$$H(f) = \frac{z^{-1}}{1 - z^{-1}} \approx (j\omega T)^{-1} = [j\pi(2fT)]^{-1} \qquad fT << 1 \tag{1.13}$$

where $z = \exp(j\omega T)$. In the signal band this gain has a modulus greater than one quarter of the oversampling ratio [Eq. (1.4)], which is usually sufficiently large. Measurements on real modulators and simulations [2, 8, 9] have demonstrated that with small gains (i.e., $G < 0.7$), the circuit responds sluggishly to changing inputs. With gains greater than 1.3, the quantized signal bounces by more than two levels and eventually goes unstable when the gain exceeds 2, as can be predicted from the linearized model of Figure 1.4(b). For most applications 10% gain accuracy is tolerable for this circuit.

POSITIONING THE QUANTIZATION THRESHOLDS. Because of the need to have short delay, the quantizer in a multilevel ΔΣ modulator usually takes the form of a flash A/D.

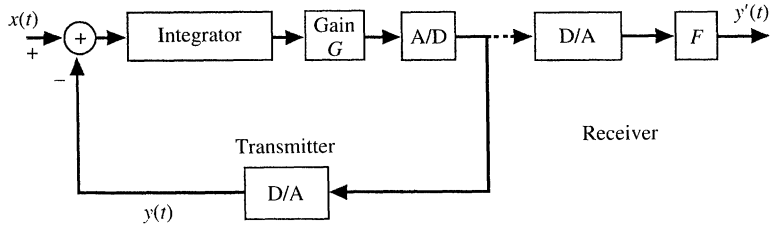

Figure 1.10 Block diagram of a ΔΣ modulator including a gain G in the feedforward path and a nonlinearity F in the representative decoder.

The gain of the quantizer, defined in Eq. (1.1), is the level spacing divided by the threshold spacing, therefore misplaced thresholds may be regarded as a nonlinearity of gain. Such nonlinearity following the high gain of the integration in the forward path of the feedback loop has little effect on baseband properties of the overall modulator [2]. Misplacing the thresholds by as much as a quarter of the spacing is sometimes tolerable.

POSITIONING THE D/A QUANTIZATION LEVELS. Misplaced levels of the D/A in the feedback path of Figure 1.4(a) are more serious than misplaced thresholds because they introduce nonlinearity directly into the signal [2, 10, 11]. The feedback action forces the average value of the quantized amplitude, y, to track the input, even when levels are misplaced. If the effective D/A converter at the receiver in Figure 1.10 matches the one in the transmitter, the output y' will track y. When it does not, the mismatch can be represented as nonlinearity F at the receiver. Such nonlinearity usually must be very small, and this calls for highly accurate D/A converters [10].

TWO-LEVEL QUANTIZATION. Using two-level quantization avoids the need for matched level spacing. A misplacement of one level, as illustrated in Figure 1.11, introduces a change of quantization range and a dc offset, neither of which need be critical.

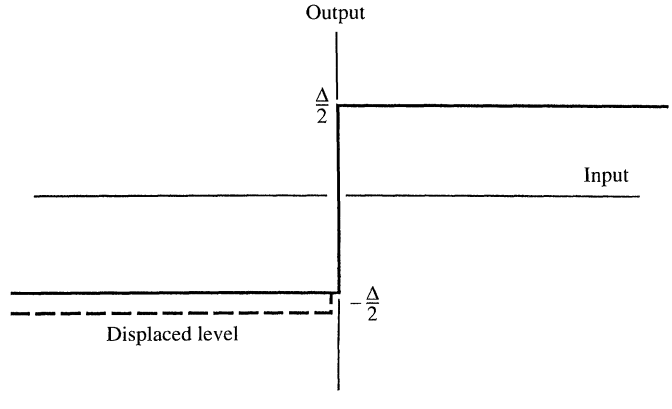

Figure 1.11 Diagram illustrating a misplaced level in two-level quantization.

Two-level quantization requires only one threshold so the concept of gain G in Eq. (1.1) is now arbitrary. Nevertheless, analysis in Chapter 2 shows that results (1.9) and (1.11) can apply.

Two-level modulators can have very robust circuits: The threshold need not be accurately positioned because it is preceded by the high dc gain of the integrator. The sample gain of the integrator is not critical because it drives a single threshold stage. The quantization levels need be positioned only to accommodate the range of input signals.

LEAKAGE IN THE INTEGRATOR. When the integrator in Figure 1.4(a) includes leakage (α), its transfer function is given by

$$H(z) = \frac{z^{-1}}{1 - \alpha z^{-1}} \tag{1.14}$$

and its dc gain by

$$H_0 = H(1) = \frac{1}{1 - \alpha} \tag{1.15}$$

The output of the modulator can be expressed as

$$Y = \frac{z^{-1}X}{1 + (1 - \alpha)z^{-1}} + \frac{(1 - \alpha z^{-1})E}{1 + (1 - \alpha)z^{-1}} \tag{1.16}$$

There is increased noise at low frequency, as illustrated schematically in Figure 1.12. If the dc gain of the integrator is at least equal to the oversampling ratio, the increase in baseband noise is less than 0.3 dB. This condition also ensures that the dead zone described in Section 1.2.2.4 is not troublesome; it could be with smaller gains.

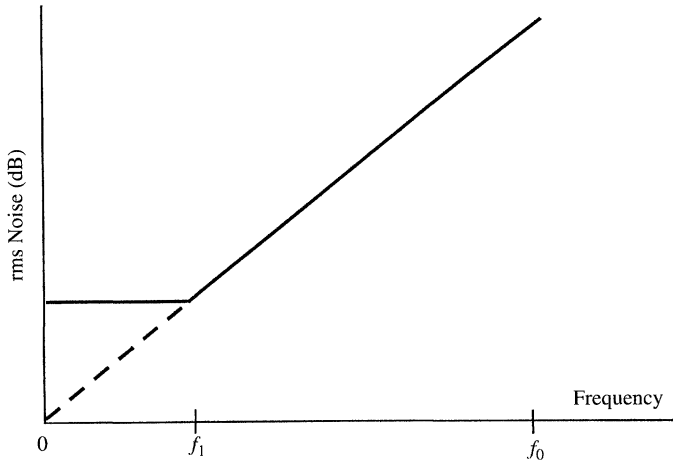

Figure 1.12 Illustration of the effects of leakage in the integrator on the spectral density of modulation noise.

1.2.3 High-Order Modulation

1.2.3.1 Predicting In-Band Values of Quantization Error. In a ΔΣ circuit, feedback via an integrator shapes the spectrum of the modulation noise, placing most of its energy outside the signal band. In general, the characteristics of the filter included in the feedback loop determine the shape of the noise spectrum [12]. In this section we discuss a number of filters and circuit structures that are improvements on ordinary ΔΣ circuits.

The objective of using improved noise shaping filters is to reduce the net noise in the signal band. To do this well, we need to subtract from the quantization error a quantity whose in-band component is a good prediction of the in-band error. Ordinary ΔΣ modulation subtracts the previous error [Eq. (1.7)]. Higher order prediction should give better results than this first-order prediction.

1.2.3.2 Noise in High-Order ΔΣ Modulation. We will see that there are several circuit arrangements that give second-order predictions of the quantization error and that the one shown in Figure 1.13 is easy to build and is tolerant of circuit imperfection. It is an iteration of ΔΣ feedback loops. The output of this modulator can be expressed as

$$y_i = x_{i-1} + (e_i - 2e_{i-1} + e_{i-2}) \tag{1.17}$$

so that the modulation noise is now the second difference of the quantization error. The spectral density of this noise is

$$N(f) = E(f)\left(1 - \varepsilon^{-j\omega T}\right)^2 \tag{1.18}$$

For busy signals

$$|N(f)| = 4e_{\text{rms}}\sqrt{2T}\sin^2\left(\frac{\omega T}{2}\right) \tag{1.19}$$

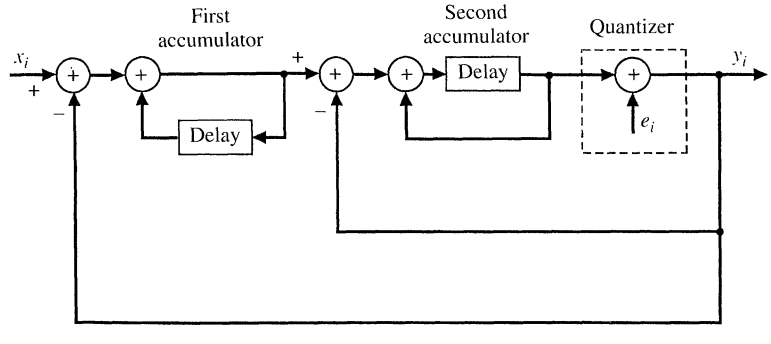

Figure 1.13 Second-order ΔΣ quantizer.

and the rms noise in the signal band is given by

$$n_0 \approx e_{\text{rms}} \frac{\pi^2}{\sqrt{5}} (2f_0 T)^{5/2} = e_{\text{rms}} \frac{\pi^2}{\sqrt{5}} \text{OSR}^{-5/2} \qquad f_s^2 >> f_0^2 \tag{1.20}$$

This noise falls by 15 dB for every doubling of the sampling frequency, providing 2.5 extra bits of resolution [8, 13].

The technique can be extended to provide higher order predictions by adding more feedback loops to the circuit [8]. In general, when a modulator has L loops and is not overloaded, it can be shown that the spectral density of the modulation noise is

$$|N_L(f)| = e_{\text{rms}} \sqrt{2T} \left[2 \sin\left(\frac{\omega T}{2} \right) \right]^L \tag{1.21}$$

For oversampling ratios greater than 2, the rms noise in the signal band is given approximately by

$$n_0 = e_{\text{rms}} \frac{\pi^L}{\sqrt{2L+1}} (2f_0 T)^{L+1/2} \tag{1.22}$$

This noise falls $3(2L-1)$ decibels for every doubling of the sampling rate, providing $(L - \frac{1}{2})$ extra bits of resolution, but we shall see that there are difficulties in implementing circuits containing more than two integrators. Figure 1.14 plots the in-band noise against

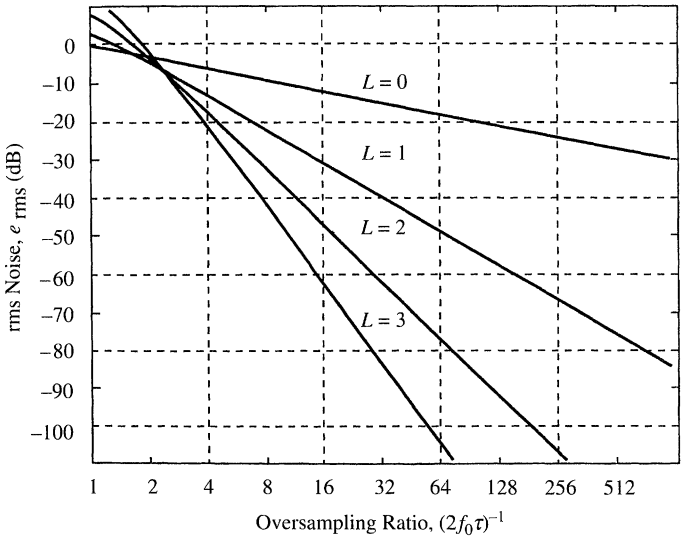

Figure 1.14 The rms noise that enters the signal band for oversampling ratios in the range 1 through 512, assuming busy input signals. Graphs are plotted for ordinary quantization without feedback $L = 0$, and first-, second-, and third-order $\Delta\Sigma$ quantization. 0 dB of noise corresponds to that of PCM sampled at the Nyquist rate. A common level spacing is used in all the quantizers.

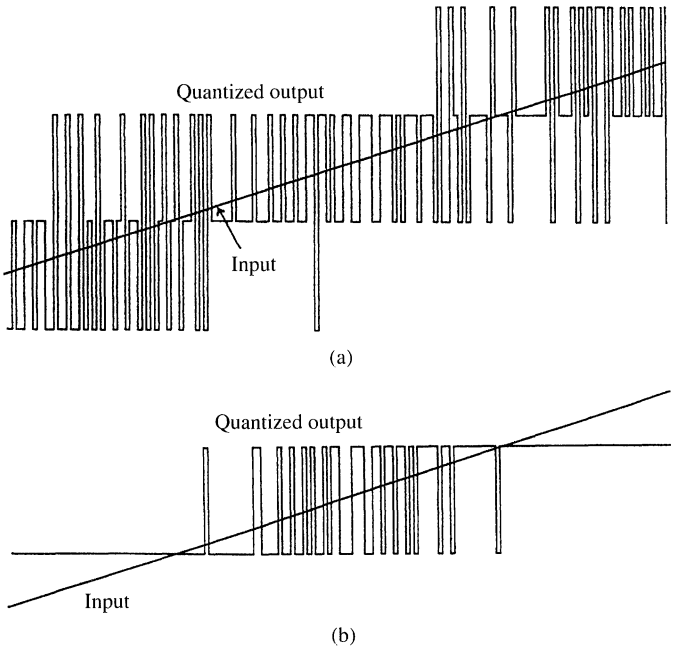

Figure 1.15 Response of a second-order $\Delta\Sigma$ quantizer to a ramp input for both multilevel and two-level quantization.

the oversampling ratio for examples of PCM, and modulators with one, two, and three feedback loops. These graphs are derived from result (1.21), which assumes white, uncorrelated quantization error, and this may not be valid unless the signal is sufficiently busy to randomize the error or unless sufficient dithering is included. It is fortunate that the noise in second- and higher order circuits is more random than it is in the first-order ones [8]. The randomizing influence is provided by the retention of noise in the integrators and depends on the circuits having long-term memory.

Figure 1.5 illustrates the output of the first-order modulator oscillating between two levels adjacent to the input value. The noise ranges in amplitude between $\pm\Delta$, which is consistent with Eq. (1.8). The output of the second-order modulator oscillates predominantly between three levels, but occasionally reaches a fourth, which is consistent with the expression for the noise in Eq. (1.17). Figure 1.15(a) shows the output of a second-order modulator having quantization levels at integer values, when responding to a ramp, and Table 1.1 lists its output for a steady input of 1.3. The output includes levels 0, 1, 2, and sometimes 3 in seemingly random order, but keeping its average close to the input value. These measurements commenced with arbitrary initial values in the integrators. Starting with integer values results in repetitive patterns in the output.

1.2.3.3 Dynamic Range of the Modulators. The oscillation of the signal uses up some of the dynamic range of the circuit, and if the quantizer is not to overload the

TABLE 1.1 A STRING OF 200 OUTPUT SAMPLES FROM A SECOND-ORDER DELTA-SIGMA MODULATOR THAT QUANTIZES ANALOG VALUES INTO INTEGERS*

0	2	1	2	1	1	2	1	1	2	0	2	2	0	2	2	0	2	1	2	1	1	1	2	1	1	2	0	3	0	2	1	1	2
1	1	2	1	1	1	2	1	2	0	2	2	0	2	1	2	0	3	0	2	1	1	1	2	2	0	2	1	1	2	1	1	2	1
1	1	2	1	2	0	2	1	2	1	1	2	0	2	2	0	2	2	0	2	1	2	1	1	1	2	1	1	2	1	1	1	2	1
1	2	1	1	2	1	1	1	2	1	2	0	2	2	1	2	1	2	1	1	1	2	1	1	1	2	2	0	2	1	1	2	1	1
2	1	1	1	2	1	2	0	2	2	0	2	1	2	0	2	2	1	1	1	1	2	1	2	1	1	1	2	1	1	2	1	1	1
2	1	1	2	1	1	2	1	1	1	2	1	2	0	2	2	0	2	1	2	1	1	1	2	1	1	2	0	3	0				

*The input to the modulator is a constant 1.3 value, and the measurement commences with arbitrary initial conditions.

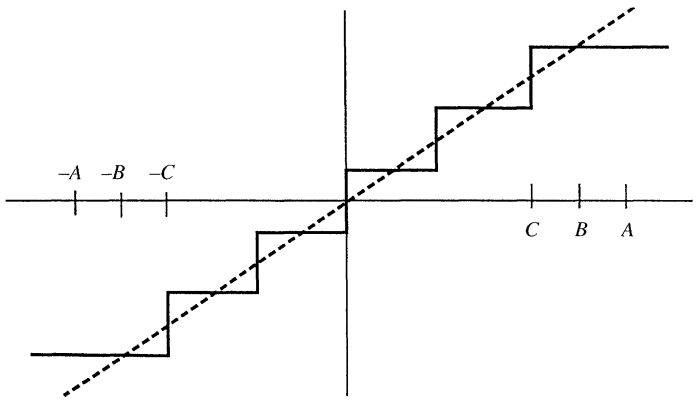

Figure 1.16 Range of amplitudes that can be accommodated by multilevel quantizers. Ordinary quantization accommodates input in the range $\pm A$, and first-order $\Delta\Sigma$ quantization accommodates $\pm B$. Second-order $\Delta\Sigma$ accommodates $\pm C$ with small probability of overloading.

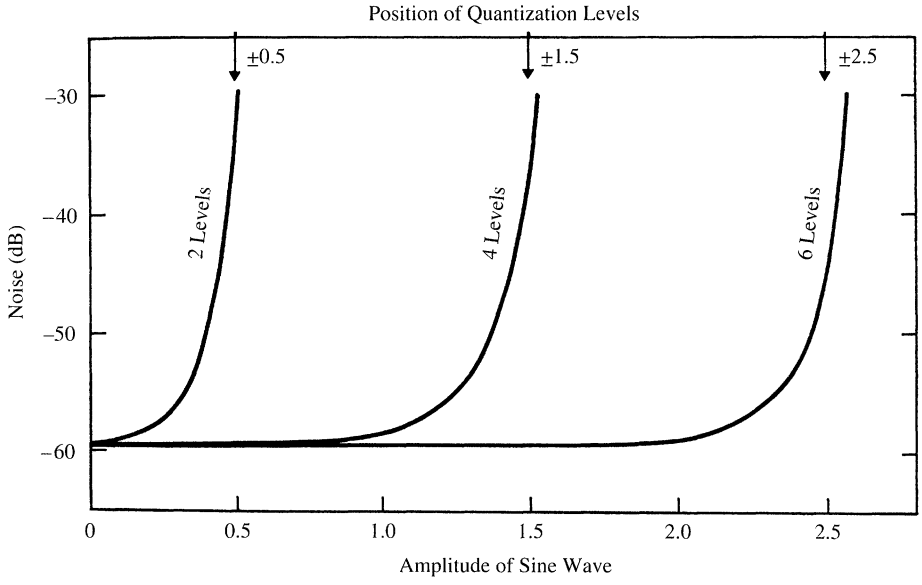

Figure 1.17 Noise introduced into sine waves of various amplitudes by second-order $\Delta\Sigma$ quantization with either 2, 4, or 6 quantization.

input amplitude to the modulator needs to be limited. Figure 1.16 shows the range of inputs that can be accommodated in several modulators, each employing six-level quantization. Ordinary PCM requires that its input be restricted to $\pm A$ in order that the quantization error lie in the range $\pm\Delta/2$. Inputs to a corresponding first-order $\Delta\Sigma$ modulator need be restricted to $\pm B$ for them to be interpolated by oscillation between two levels. Inputs to a second-order modulator need be restricted to $\pm C$ to prevent frequent overloading of its quantizer. This permits the output to oscillate between three levels, but overload can occur occasionally when the oscillation attempts to step outside this three-level range.

When large inputs cause the quantizer to overload, the modulation noise increases as illustrated in Figure 1.17. This plots the noise introduced into sine waves of various amplitude, during second-order modulation. Graphs are drawn for three cases: two-, four-, and six-level quantization. The levels are positioned at ±0.5, ±1.5, and ±2.5. The results demonstrate that the *excess noise* due to overloading the quantizer increases quite slowly with increasing amplitude. Even two-level quantization has a useful range, despite the fact that theoretically it is overloaded for all conditions, except zero input with zero initial conditions. The response of a second-order modulator having two-level quantization to a ramp input is shown in Figure 1.15(b). Examples of such modulators are described in later chapters.

For small input amplitudes, the noise in these modulators agrees with Eq. (1.20). Simulations show that the excess noise introduced into larger inputs appears as odd-order harmonic distortion of centrally biased sine waves and includes a minute increase in the gain of the fundamental. The excess noise decreases with increased oversampling ratio but increases with the frequency of the applied sine wave. A complete characterization of the excess noise is not yet available, but attempts have been made to analyze it [14].

1.2.3.4 Influence of Circuit Parameters on Second-Order Modulators

CIRCUIT TOLERANCES. The second-order modulator in Figure 1.13, like the first-order one in Figure 1.4, is very tolerant of circuit imperfections, especially when two-level quantization is employed. Compared with first-order systems, the second-order system has one more design parameter available; it is the ratio of the gains of the two feedback paths. The outer path dominates in determining the low-frequency properties of the circuit, while the inner path serves to stabilize the system, and determines high-frequency properties. Matching their relative gains to within ±5% is usually satisfactory. Sometimes, the gain of the inner loop is deliberately increased to compensate for delay in the outer loop [15].

RANGE OF INTEGRATION. An important parameter is the range of signal amplitudes that must be accommodated at the outputs of the integrators. Simple theory gives an adequate description of these signals for multilevel quantization that does not saturate. The output of the first integrator is given by $x_i - e_i + e_{i-1}$, the second by $x_{i-1} - 2e_{i-1} + e_{i-2}$, with e and x bounded by $\pm\Delta/2$. These results are inadequate for describing two-level modulators, and we resort to simulations. Figure 1.18 plots the maximum signal level at the output of the integrators as a function of a dc input level. The signal in the first integrator remains well bounded, but the second one becomes very large as the input exceeds 0.4Δ. Clipping this signal at about three times the range of the input signal has little effect on overall performance of the modulator [8, 15].

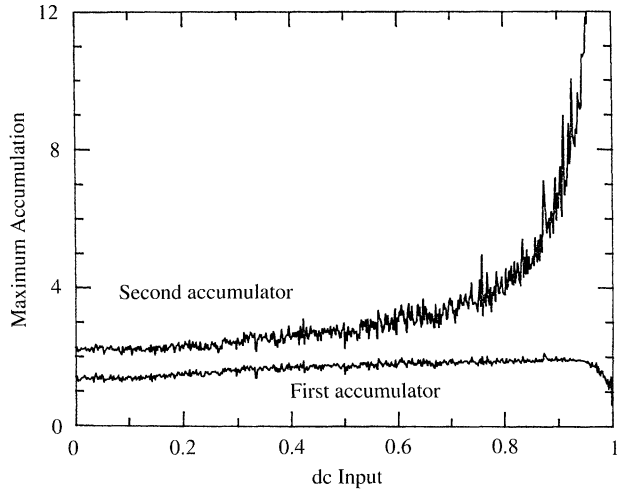

Figure 1.18 Maximum amplitude of signals in the accumulators of a second-order $\Delta\Sigma$ modulator. The two quantization levels are at ± 1. In practice the first accumulation is often clipped at ± 2 and the second effectively at ± 4.

LEAKAGE IN THE INTEGRATORS. First-order modulators need integrators with dc gains H_0 that are greater than the oversampling ratio, in order to have low noise. Calculations of noise in second-order modulators indicate that somewhat lower gains could be tolerated because the gains of two integrator amplifiers are cascaded in the outer loop. But there is another consideration: Leakage can permit the oscillation of the quantized signal to settle into regular patterns when there is insufficient long-term memory to randomize it. This is most noticeable at the center of the range where the output can settle into a $+1, -1, +1, -1$ pattern. The effect is illustrated by Figure 1.19, which shows the filtered output of a modulator responding to a very slowly changing ramp. The full range of the output signal is ± 1 V; Figure 1.19 has such an expanded scale that the noise is apparent. At the center of the range the output locks into the pattern and the input is ignored in the range ± 0.2 mV. It may be shown that the width of the dead zone is given approximately by $1.5\Delta H_0^{-2}$, and for this to be less than twice the rms noise requires that the dc gain of each integrator satisfy

$$H_0 \geq (2f_0T)^{-5/4} \tag{1.23}$$

The dead zone is seldom noticeable because it is present only in very slowly changing signals: It takes time for the oscillations to settle into a pattern. Such a dead zone could actually be useful for audio applications, which need a very quiet idle state.

1.2.3.5 Limit Cycles in Third-Order $\Delta\Sigma$ Modulators. Simple linear theory predicts that the third-order modulator shown in Figure 1.20 has an rms noise given by Eq. (1.22) with $L = 3$. This can be realized in practice with a multilevel quantizer that does not overload [8]; but the circuit is much more sensitive to circuit values than the first-

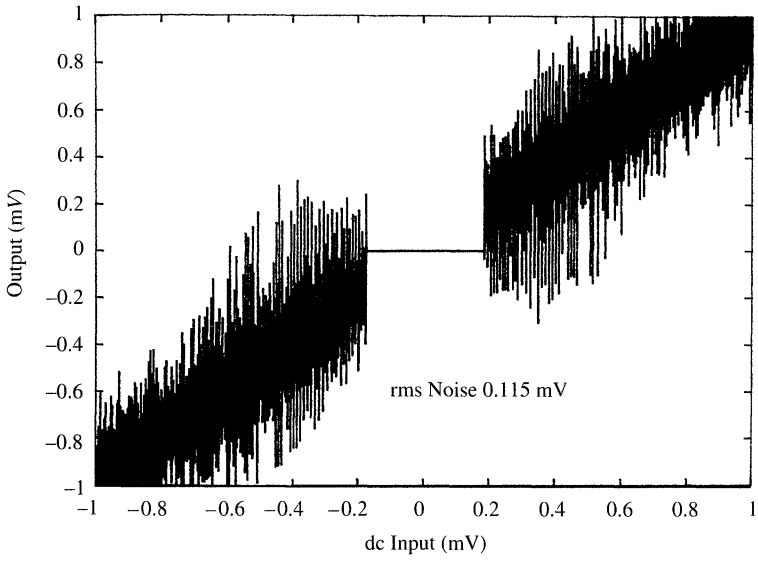

Figure 1.19 Illustration of the dead zone caused by leakage in the accumulators of second-order ΔΣ quantization. The dc gain of each accumulator is 64, and the oversampling ratio is also 64. The range of input and output amplitudes that can be accommodated is ±1 V, and the noise is less than that of 12-bit PCM.

and second-order ones. For example, the equivalent linear circuit of this modulator becomes unstable with quantizer gains G in excess of 1.15 compared with 2.0 and 1.33 for first- and second-order modulators.

More seriously, the third-order circuit is also unstable when its quantizer gain falls below 0.3. When the quantizer saturates, its effective gain falls, and this usually results in

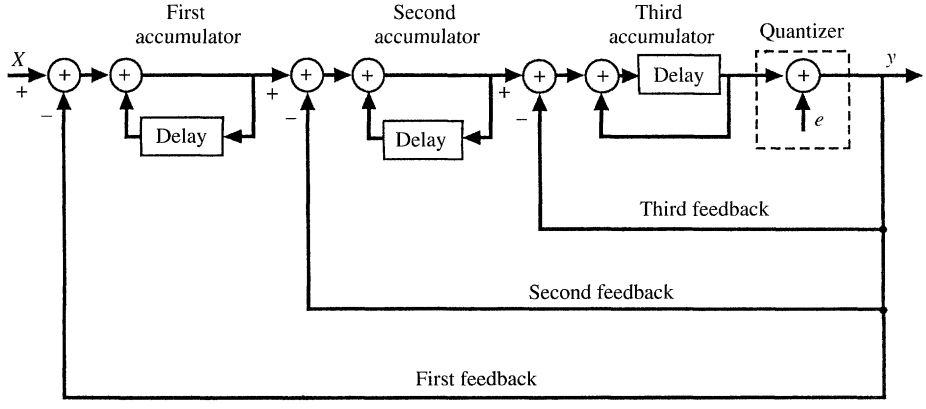

Figure 1.20 Third-order ΔΣ quantizer.

an instability in which the circuit settles into a large-amplitude low-frequency limit cycle. In this state the clipped signals fed back via the two inner feedback paths are small compared with the signals emerging from the integrators. Properties of the outer loop dominate; it contains three integrators and a delay, a strong basis for instability. Unembellished, two-level, third-order ΔΣ modulators cannot escape from this condition. Their circuits can be made stable by clipping the outputs of the integrators or including other nonlinearites that make the inner feedback effective when the quantizer saturates. The noise performance of these modified modulators is considerably worse than Eq. (1.21) predicts. Better performance is obtained by redesigning the filter used in the feedback loop.

1.2.3.6 Noise Shaping Using Filters with Nonmonotonic Transfer Functions. The ΔΣ modulators described so far contain filters in their feedback loop that have multiple poles at dc and zeros at high frequency to stabilize the circuits. The frequency responses of these filters fall monotonically through the range 0 to $f_s/2$.

The noise at the output of the modulator is shaped approximately as the inverse of the filter characteristic. Later chapters describe modulators that replace these filters with a more rectangular high-pass filter. The poles are distributed through the signal band in order to lower the in-band noise. The zeros are chosen to flatten the filter response at high frequency in order to reduce the high-frequency noise and prevent it from using up dynamic range. A noise spectrum [16] obtained from such a modulator is given in Figure 1.21.

Modulators with two-level quantization and fourth- and fifth-order filters have been successfully built with this technique. Their circuits, illustrated in Figure 1.22, are based

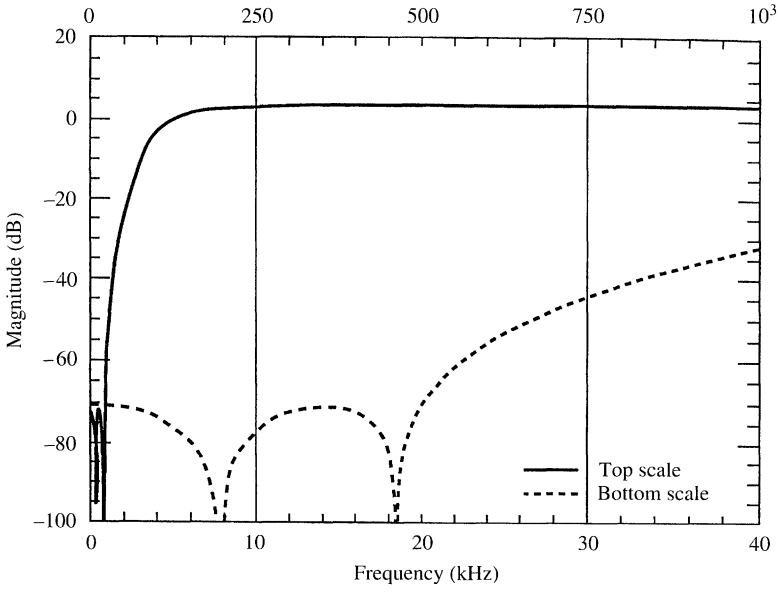

Figure 1.21 Spectral densities of the noise from a generalized feedback of the type shown in Figure 1.22.

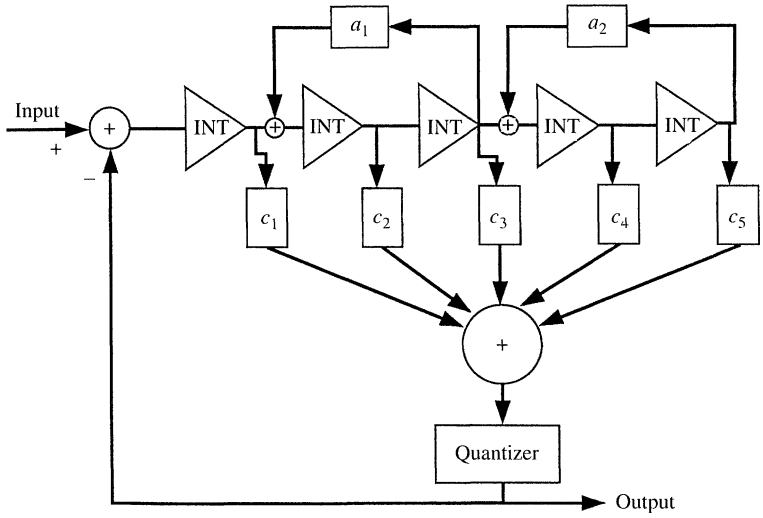

Figure 1.22 Fifth-order feedback quantizer. The loop gain includes two complex and one real pole with zeros positioned to ensure stability. The quantization is two level.

on cascaded integrators [16] with feedback branches dimensioned to position the poles and feedforward ones dimensioned to position the zeros. The danger of these circuits locking into high-frequency limit cycles is avoided by allowing the integrators to clip at quite small amplitudes. The input amplitudes are limited to less than $\pm\Delta/4$ to avoid distorting the signal. Alternative structures are described in other chapters of this book.

The noise performances of these modulators are poorer than anticipated by Eq. (1.21), but better than that obtained from second-order $\Delta\Sigma$ modulators. For example, 16-bit encoding of 20-kHz signals has been obtained by using a fourth-order modulation at 3 MHz.

1.2.4 Some Alternative Modulator Structures

1.2.4.1 Error Feedback. Noise-shaping quantization was first introduced using the structure shown in Figure 1.23. In this circuit, the difference between the input

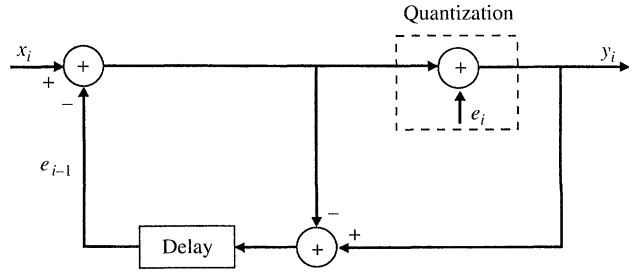

Figure 1.23 Quantizer with error feedback.

and the output of the quantizer is a measure of the quantization error, which is fed back and subtracted from the next input sample. The circuit is algebraically equivalent to the $\Delta\Sigma$ circuit in Figure 1.4, but it has the serious practical disadvantage that inaccuracies in the analog subtractors have a strong impact on the modulator's properties. We shall see, however, that the circuit can be used as a demodulator because there the processing is performed digitally. The circuit can be generalized by replacing the delay with a prediction filter, design methods for which are given in [12].

1.2.4.2 Cascaded Modulators. The performance of a modulator can be improved by taking a measure of its noise, digitizing that measure in a second modulator, and combining the output of the two modulators in a way that cancels the noise of the first modulator. This technique was proposed for use with two Δ modulators and has since been widely applied to $\Delta\Sigma$ modulators [17]. Figure 1.24 shows a method for cascading two first-order $\Delta\Sigma$ modulators. The output of the integrator in the first modulator is fed to the second modulator. Its output is digitally differentiated and subtracted from the output of the first modulator to provide the net output of the circuit. We have used e' to denote the quantization error in the first modulator and e that of the second one. When scaling factors are ignored, it can be shown that the net output of the circuit may be expressed in the form

$$y_i = x_{i-2} + (1-g)(e'_{i-2} - e'_{i-3}) + (e_i - 2e_{i-1} + e_{i-2}) \tag{1.24}$$

where g is a measure of the accuracy of the error cancellation. It depends on a number of parameters, including the precision of component values and the low-frequency gain of the first integrator. Ideally, g is unity; then the noise of the first modulator does not contribute to the output. The remaining noise is the second difference of the quantization error

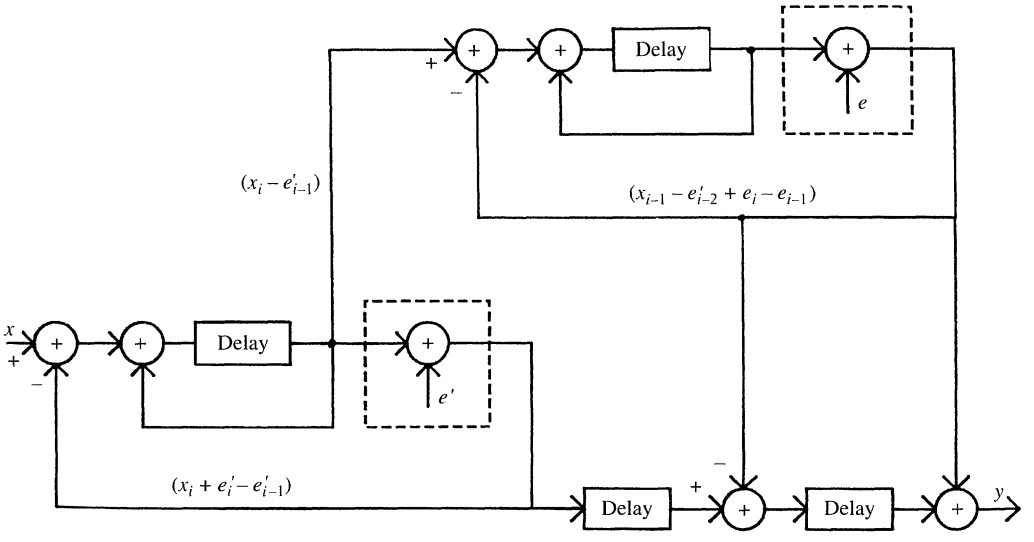

Figure 1.24 Cascade of two first-order $\Delta\Sigma$ modulators. The second modulator serves to digitally encode the quantization error e' of the first modulator so that e' may be cancelled from the net output.

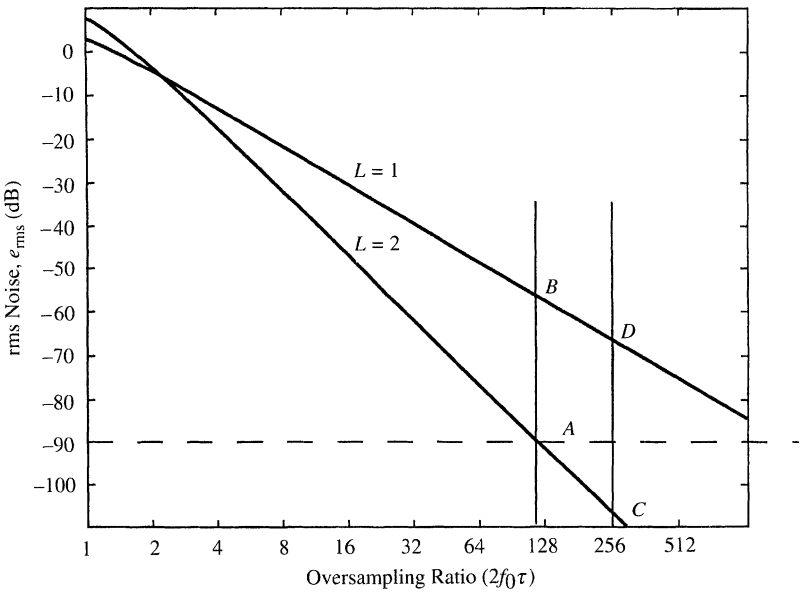

Figure 1.25 A graphical comparison of noise sources in the cascaded modulators illustrated by Figure 1.24.

from the second modulator: It is in the same form as the noise of a second-order ΔΣ modulator given in Eq. (1.17).

The following question is a first consideration in designing cascaded modulators: How close to unity must the factor g be? This can be determined from Figure 1.25, which was derived from Figure 1.14. As an example, suppose we are designing a modulator similar to the one in Figure 1.24 to provide resolution equivalent to 15-bit PCM, which corresponds to a noise of –90 dB on the ordinate. An ideal second-order modulator oversampling by a factor of 120, represented by point A on the graph would meet the requirement. At this sampling rate the noise from the first-order modulation is given by the ordinate of point B. It is at –57 dB. We therefore need sufficient precision to make the term $(1 - g)$ much less than $57 - 90 = -33$ dB; that is, g needs to be well within 2% of unity. In practice this requirement is tightened by needs to scaling signal amplitudes in practical circuits. We can be looking for component tolerances below 0.1% and amplifier gains in excess of 10,000.

The need for such precision is alleviated by raising the sampling rate: for example, at an oversampling ratio of 256, corresponding to points C and D on the graph. The second-order noise is now at –108 dB, well below the requirement. The first-order noise is –68 dB, and we need to reduce this by only 22 dB (i.e., keep g within 8% of unity) to achieve a net noise of –90 dB.

Because of the difficulty in obtaining adequate precision, the noise from these cascaded circuits is often dominated by the noise from the first stage, which in our example is first order and will include peaks of pattern noise. This difficulty is avoided by using a second-order modulator for the first stage [18, 19]. When the performance is constrained

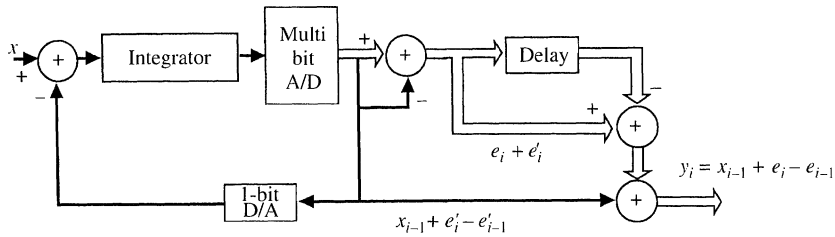

Figure 1.26 A delta modulator and its sampled-data equivalent circuit.

by circuit imperfection, there is little advantage in using higher than first-order modulation as the second stage or adding a third stage.

When circuits have sufficient precision to eliminate the noise of the first stage from the output, the cascade modulator has several attractive features. For example, when $(1 - g) = 0$, the circuit in Figure 1.24 provides second-order modulation yet the feedback loops are first order with two-level quantization. The output can oscillate between four levels, illustrated in Figure 1.15(a), and there need be no excess noise caused by quantizer overload provided the first-order input stage does not overload. Adding a third stage to the cascade can provide third-order modulation without incurring the dangers of instability associated with third-order feedback quantization.

An ingenious circuit [20] that can be interpreted as a cascade of a two-level first-order $\Sigma\Delta$ modulator with multilevel PCM is shown in Figure 1.26. Only the sign bit of the A/D drives the feedback. The complete digital word from the A/D is reduced by subtracting the value of its sign bit (this is equivalent to inverting the sign bit of a 2's complement code). The first difference of the resulting code adds to the original sign bit to provide the output. This technique can obviously be extended to higher order modulators.

1.2.4.3 Delta Modulation. Most early work on oversampling was concerned with Δ modulation. Later work turned its attention to $\Delta\Sigma$ modulation because its circuits are more robust. The main difference between the techniques is that $\Delta\Sigma$ modulators, and other noise-shaping modulators, change the spectrum of the noise but leave the signal unchanged. By contrast, Δ modulation and other signal-predicting modulators shape the spectrum of the modulated signal but leave the quantization noise unchanged at the receiver. It is the need for a filter at the receiver to restore the signal that makes signal-predicting modulators vulnerable to analog circuit inaccuracy. These filters usually have high gain in the signal band and thus magnify distortion introduced in the channel or the D/A.

Figure 1.27 is a diagram of a Δ modulator and demodulator. It transmits the first difference of the signal; consequently an integrator is needed at the receiver. The output is contaminated with the quantization error itself, and it saturates by clipping the derivative of the signal (slope overloading). In contrast the output of a $\Delta\Sigma$ modulator is contaminated by the first difference of the quantization error and saturates by clipping amplitudes. To compare these modulators, we now calculate their signal-to-noise ratios for sinusoidal inputs. The largest sine wave that the $\Delta\Sigma$ modulator can accommodate without saturating has peak value $\Delta/2$. Its rms noise is given by Eq. (1.11), and therefore the maximum rms

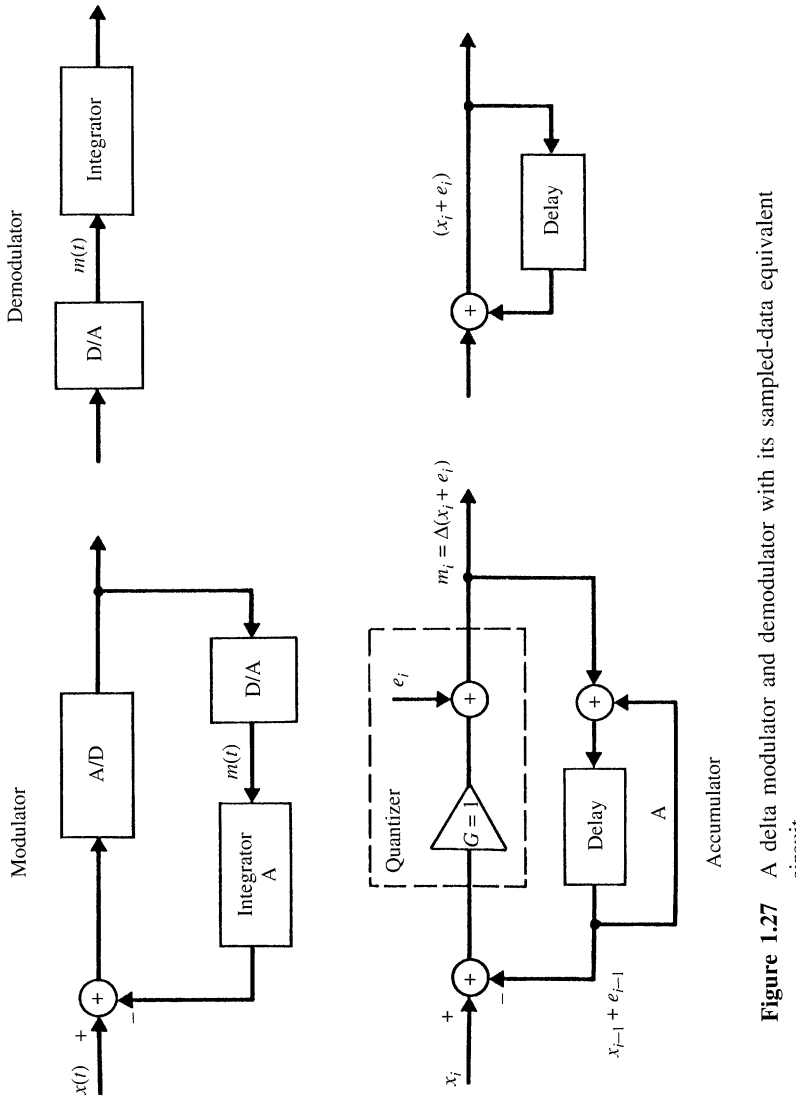

Figure 1.27 A delta modulator and demodulator with its sampled-data equivalent circuit.

signal-to-noise ratio (SNR) can be expressed as

$$\mathrm{SNR}_{\mathrm{dsm}} = \sqrt{\frac{3}{8}}\frac{\Delta}{\pi e_{\mathrm{rms}}}(2f_0T)^{-3/2} = \frac{\sqrt{4.5}}{\pi}(2f_0T)^{-3/2} \tag{1.25}$$

The peak value of the largest sine wave that a delta modulator will accommodate is slope limited and is given by $\Delta/\omega T$, where ω is the angular frequency of the sinusoid signal. If f' is the frequency at which the signal source delivers its steepest slope, the amplitude of sinusoidal inputs need be constrained to be less than $\Delta/(2\pi f'T)$. The rms inband noise introduced into the signal is given by (3). Therefore the signal-to-noise ratio is given by

$$\mathrm{SNR}_{\mathrm{dm}} = \frac{\sqrt{6}}{\pi}\left(\frac{f_0}{f'}\right)(2f_0T)^{-3/2} \tag{1.26}$$

For general signals $f' = f_0$ then Δ modulation has only a 1-dB advantage over $\Delta\Sigma$ modulation. For some speech signals $f' = f_0/3$. Then Δ modulation has a 10-dB advantage. The Δ modulation and signal predictive modulators can be useful for applications where f' is much less than f_0 and when clipping slopes is more tolerable than clipping amplitudes, as in some audio and video application.

1.3 DECIMATING THE MODULATED SIGNAL

1.3.1 Multistage Decimation

The output of the modulator represents the input signal together with its out-of-band components, modulation noise, circuit noise, and interference. The digital filter shown in the encoder of Figure 1.2 serves to attenuate all of the out-of-band energy of this signal so that it may be resampled at the Nyquist rate without incurring significant noise penalty because of aliasing.

A fairly simple filter would suffice to remove the modulation noise alone because its spectrum rises slowly; for example, the noise increases 12 dB per octave for second-order $\Delta\Sigma$ modulation. However, abrupt low-pass filters are often needed to remove out-of-band components of the signal. And such filters are expensive to build at the elevated sampling rates of the modulator. In practice, it nearly always pays to perform the decimation in more than one stage [21]. This is illustrated in Figure 1.28, using the example of 4-kHz

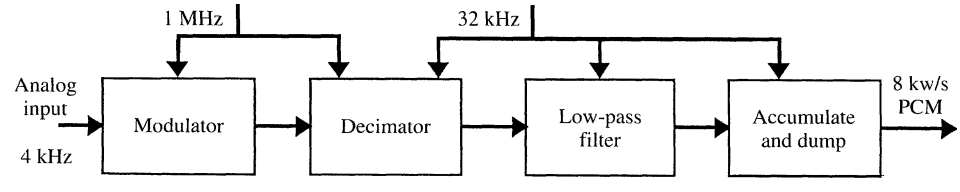

Figure 1.28 Decimating the output of the modulator in two stages, from 1 MHz to 32 kHz and then to 8 kHz.

telephone signals that have been modulated at 1 MHz. The first stage of decimation lowers the word rate from 1 MHz to 32kHz, an intermediate decimation frequency which is four times the Nyquist rate. The filter in this stage is designed primarily to remove modulation noise, because that noise dominate at high frequency. Out-of-band components of the signal dominate at lower frequency, and these are attenuated by the abrupt low-pass filter in the final stage of decimation. As the signal propagates through the filters and resampling stages, the word length increases from 1 to 16 bits in order to preserve the resolution as the word rate decreases. We will describe the design of these decimating circuits individually and explain why an intermediate frequency of four times the Nyquist rate was selected.

1.3.2 Design of the First-Stage Decimator

Figure 1.29 illustrates the action of the decimator. Frequencies below f_0 form the signal band, f_D is the intermediate decimation frequency, and f_s is the modulation frequency. The raised-cosine curve represents the spectral density of the quantization noise arising from second-order modulation. When this noise is sampled at f_D, its components in the vicinity of f_D and harmonics of f_D fold into the signal band. Consequently, it is sensible to place zeros of the decimation filter at these frequencies. There is no need for an abrupt cutoff at f_0 because noise in the range f_0 to $f_D - f_0$ folds on itself without entering the signal band. A small droop in the response over the signal band can easily be compensated in the filters of the next stage.

A convenient filter for this decimation has a frequency response based on sampled $\text{sinc}(\pi f/f_D)$ functions. An example is shown in Figure 1.29. The simplest of these decimators is the accumulate-and-dump circuit. If its input samples are x_i occurring at rate

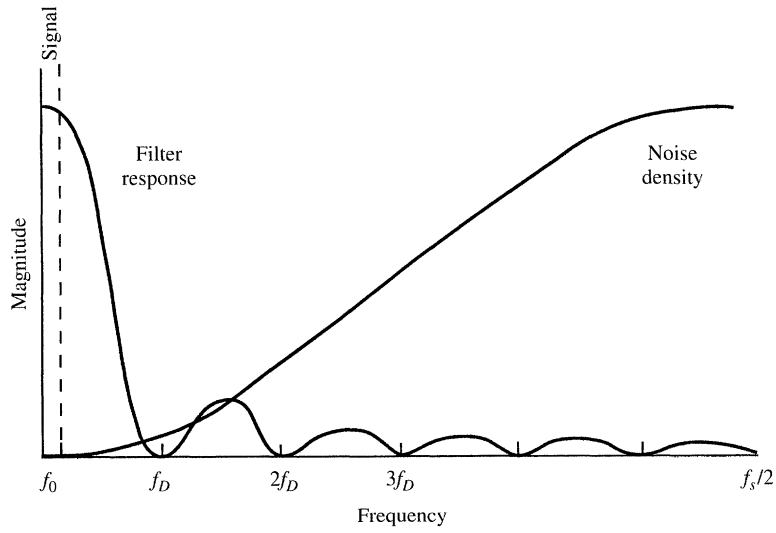

Figure 1.29 Decimating with a filter having sinc^2 frequency response; f_s is the modulation rate, f_D the intermediate decimation frequency, and $0 \le f < f_0$ the signal band.

f_s and output samples are y_k occurring at f_D, then

$$y_k = \frac{1}{N} \sum_{i=N(k-1)}^{Nk-1} x_i \tag{1.27}$$

where the decimation ratio N is the integer ratio of the input frequency to the output frequency,

$$N = \frac{f_s}{f_D} \tag{1.28}$$

The transfer function of the filter is

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{N} \sum_{i=0}^{N-1} z^{-i} = \frac{1}{N} \frac{1 - z^{-N}}{1 - z^{-1}} \tag{1.29}$$

and its frequency response is given for $z = \varepsilon^{j\omega T}$ by

$$H(f) = \frac{\text{sinc}(\pi f N T)}{\text{sinc}(\pi f T)} \tag{1.30}$$

This has zeros at f_D and all of its harmonics in the range $f_0 \leq f < f_s$. This simple filter was used for decimation in oversampling A/D converters [2] at a time when it was important to have simple digital processing circuits. Much better performance can now be obtained by using a filter that is represented by a product of sinc functions [22, 23]. It has been shown

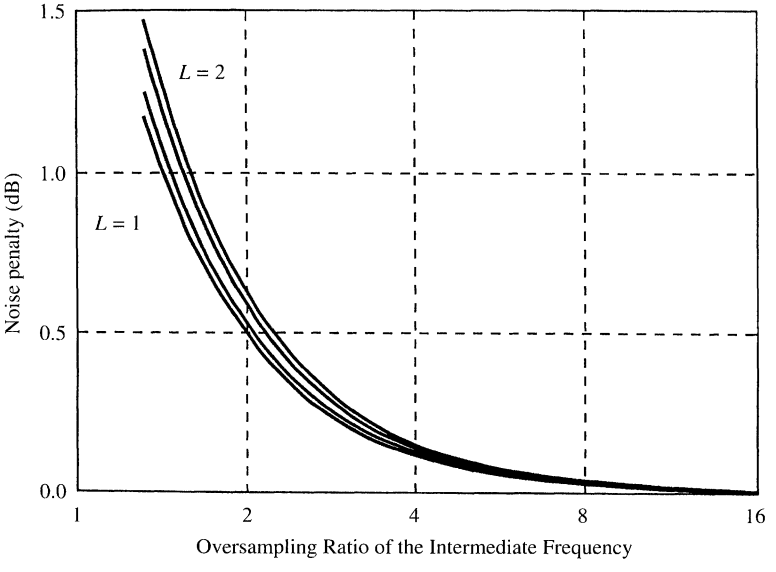

Figure 1.30 Increase in noise caused by decimation with sinc^{L+1} filters. $N = f_s/f_D$ is the decimation ratio. Results are plotted for first-order $\Delta\Sigma$ modulation $L = 1$ and second order $L = 2$.

that a filter function $[\text{sinc}(\pi fNT)/\text{sinc}(\pi fT)]^{(L+1)}$ is close to being optimum for decimating the signal from $\Delta\Sigma$ modulation of order L, which have noise spectral densities given by Eq. (1.12). The order of the filter should be one more than the order of the modulation. The penalty for using this class of decimation is typically less than a 0.5-dB increase in noise. Figure 1.30 plots the penalty against the oversampling ratio $(2f_0NT)^{-1}$ of the intermediate frequency signal.

When the intermediate frequency is four times the Nyquist rate, the penalty is about 0.14 dB, but it increases as the intermediate frequency is lowered. Another factor that influences the choice of intermediate frequency is the droop in the frequency response of the filter at the edge of the signal band f_0. This is plotted in Figure 1.31 against the intermediate oversampling ratio for filters of various orders K and decimation ratios N. With third-order decimation and an intermediate oversampling ratio of 4, the droop is about 2.75 dB, but it increases rapidly if the intermediate oversampling ratio is lowered. It usually is inconvenient to compensate for more than 3 dB of droop.

Besides attenuating the modulation noise, the filter must also provide sufficient attenuation of the high-frequency components of the signal that alias into the signal when resampled at the intermediate frequency. We can see in Figure 1.29 that this attenuation is least at the frequency $f_D - f_0$. The attenuation at this frequency is plotted in Figure 1.32 for various conditions; it is about 50 dB for an intermediate oversampling ratio of 4, third-order decimation, and decimation ratios N greater than 15.

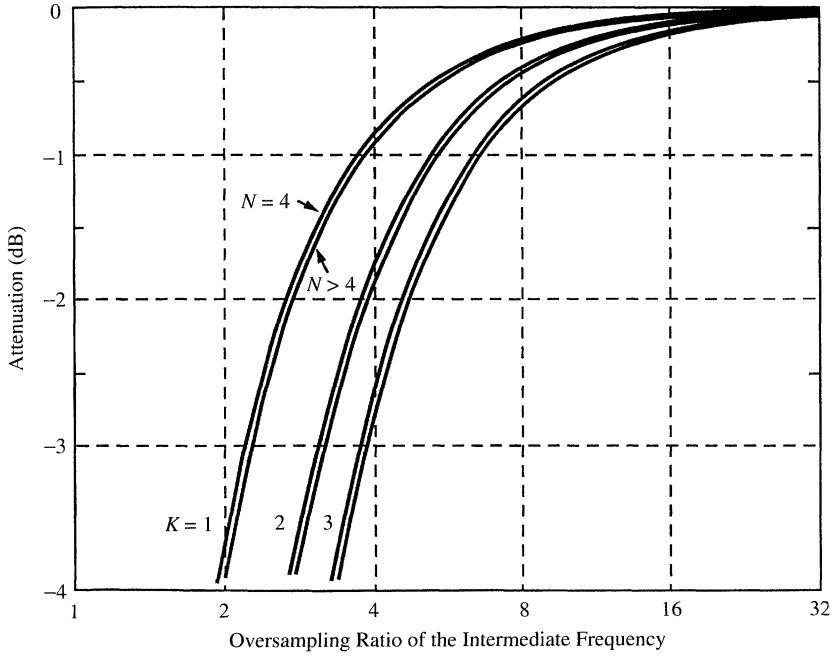

Figure 1.31 Attenuation of sinc^k decimation filters at the edge of the signal band, f_0. This amount of droop in the signal needs to be compensated.

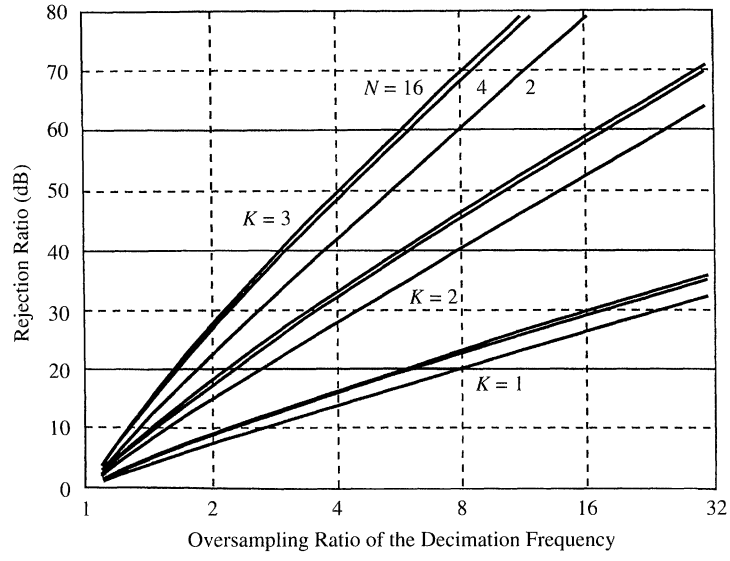

Figure 1.32 A graph of

$$\frac{\text{sinc}^k\{\pi(f_D - f_0)NT\}}{\text{sinc}^k\{\pi(f_D - f_0)T\}}$$

It is attenuation of out-of-band components of the signal at frequency $f_D - f_0$ for sinc^k decimation; N is the decimation ratio $N = f_s/f_D$. This attenuation should meet the antialiasing requirement of the application. In Section 1.4.2 we show how this graph can be used to measure the attenuation of an interpolator.

An intermediate oversampling ratio of about 4 and sinc^k decimation has favorable characteristic for use with ΔΣ modulation in many applications. Using ratios less than 4 results in rapidly deteriorating characteristics, higher ratios give less favorable design requirements for the low-pass filter in the next decimating stage. The results in Figure 1.30 do not apply to modulators that have sharply rising noise spectral densities, such as described in Section 1.2.3.6. One of the penalties of using these modulators is the fact that they need more complex decimation filters than do the ordinary ΔΣ modulators that have spectral densities [Eq. (1.21)] that rise more slowly with frequency. The next section shows that sinc^k decimators can have very simple implementations.

1.3.3 Implementing sinc Decimators

Good designs of decimators place the resampling within the filter so that the sample values that are not needed for the output are not calculated [21]. The input section of the filter processes the bits of short words in parallel at the high word rate. After resampling, the later stages of the filter can process the bits of the longer word serially at the lower word rate.

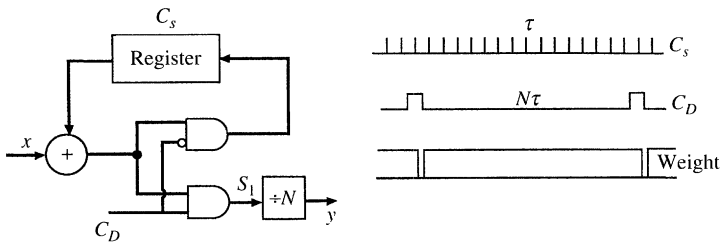

Figure 1.33 Accumulate-and-dump circuit. C_s is the input clock and C_D the output clock.

Figure 1.33 shows an implementation of the accumulate-and-dump function given by Eq. (1.29). Input words are added to the contents of the register. The sum is placed back in the register, except at the time of every Nth input word, when the sum S_1 goes to the output, while the register is cleared.

Figure 1.34 shows a related implementation of the sinc^2 decimation [22]. It comprises a cascade of two accumulate-and-dump circuits. Their outputs S_1 and S_2 are separately converted from parallel to serial words and combined to give the next output. The output of the first accumulator, S_1, is $NH(z)X(z)$, where $H(z)$ is given by Eq. (1.29), and the output of the second is given by

$$\frac{S_2}{X} = \sum_{i=0}^{N-1} \sum_{k=0}^{i} z^{-k} = \sum_{i=0}^{N-1} (i+1)z^{-i} = \frac{1-z^{-N}}{\left(1-z^{-1}\right)^2} - \frac{Nz^{-N}}{1-z^{-1}} \tag{1.31}$$

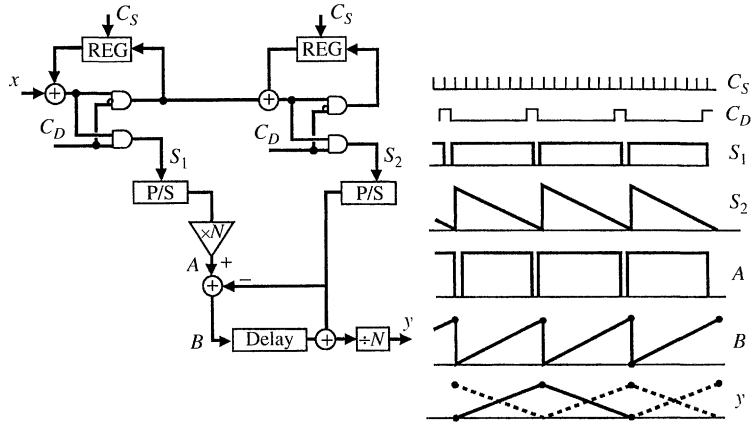

Figure 1.34 A sinc^2 decimating circuit, with a diagram of weighting factors in the summations of input samples at various points in the circuit; P/S denotes a change from parallel to serial format.

An expression for the signal filtered by the second-order sinc function can then be constructed as

$$\frac{1}{N^2}\left(\frac{1-z^{-N}}{1-z^{-1}}\right)^2 X = \frac{1}{N^2}\left[(NS_1 - S_2)z^{-N} + S_2\right] \tag{1.32}$$

and this is implemented at the low word rate because the expression does not include the short delay z^{-1}, associated with the fast clock.

These decimating circuits are relatively simple to implement because the first accumulator needs to hold no more than $\log(N) + b$ bits and the second one $\log_2[N(N+1)] + b$ bits, b being the number of bits in the input word. When N is a power of 2, multiplication and division by N in Figures 1.33 and 1.34 are mere changes in the significance of the bits.

A third-order decimator can be designed by appending a third accumulator generating a sum S_3. This is combined with the other accumulations according to the relationship

$$\left(\frac{1-z^{-N}}{1-z^{-1}}\right)^3 X = \left(1-z^{-N}\right)^2 S_3 + Nz^{-N}\left(1-z^{-N}\right)\left(S_2 + \frac{S_1}{2}\right) + \frac{N^2 z^{-N}}{2}\left(1-z^{-N}\right)S_1 \tag{1.33}$$

An alternative method for designing decimators that gives simpler circuits for third- and higher-order filters [24] is illustrated by Figure 1.35. The input signal feeds to a cascade of K accumulators, which in normal operation are not reset. This provides the filter action $(1-z^{-1})^{-K}$. The signal is then resampled at rate f_D and feeds to a cascade of K differentiators to generate the decimating function $[(1-z^{-N})/(1-z^{-1})]^K$ of the input.

An apparent objection to this method is the fact that the accumulators need to be very large if sufficient space is provided to prevent their overflowing. The difficulty is avoided by employing modulo arithmetic [24]. Each accumulator and differentiator stage holds only sufficient bits to accommodate the length of the output word; no more than

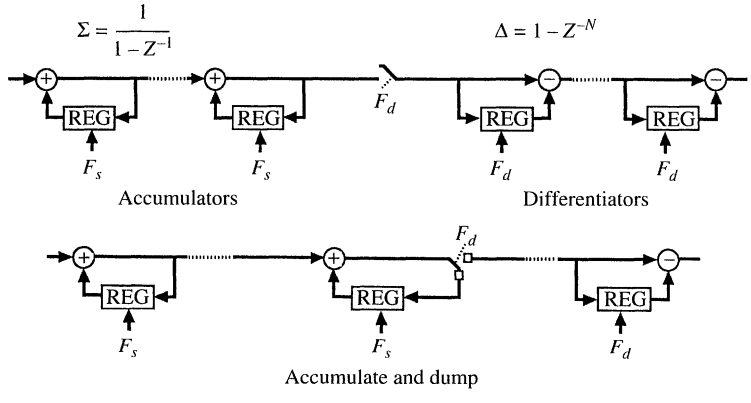

Figure 1.35 A sinc^K decimating circuit that comprises K accumulators, followed by resampling and K differentiators. All additions are performed modulo 2^b, b being the number of bits required in the output word. The second circuit resamples using an accumulate-and-dump circuit.

$K \log_2(N) + b$ bits. The circuits are allowed to overflow naturally. It can be shown that this overflow does not affect the net output of the decimator. The circuit can be simplified by replacing one accumulator stage and one differentiating stage, together with the resampling switch by an accumulate-and-dump stage, as shown in the second circuit in Figure 1.35.

1.3.4 The Low-Pass Filter

The low-pass filter in the final stage of the decimator in Figure 1.28 is designed to meet the antialiasing requirements of the input signal. Its circuit can usually be very simple [22] and because the word rate f_D is low, the circuit may process bits in serial rather than in parallel. Moreover, the word length of the coefficients can be so short that dedicated multiplier circuits can be used rather than shared ones. The accumulate-and-dump stage performs the final resampling to the Nyquist rate. Its frequency response [Eq. (1.30)] contributes useful zeros to the low-pass filter function. The positioning of these zeros depends on the intermediate oversampling ratio; there sometimes is advantage in using a nonbinary ratio such as 5 in order to place zeros off the axes.

Figure 1.36 shows the frequency response of a low-pass filter that is intended for decimating 4 kHz telephone signals from an intermediate sampling frequency of 40 kHz to the 8-kHz Nyquist rate. Zeros in the response near 4.5 and 6 kHz are included in two recursive sections; zeros at 10 and 20 kHz lie on the imaginary and negative axis of the z

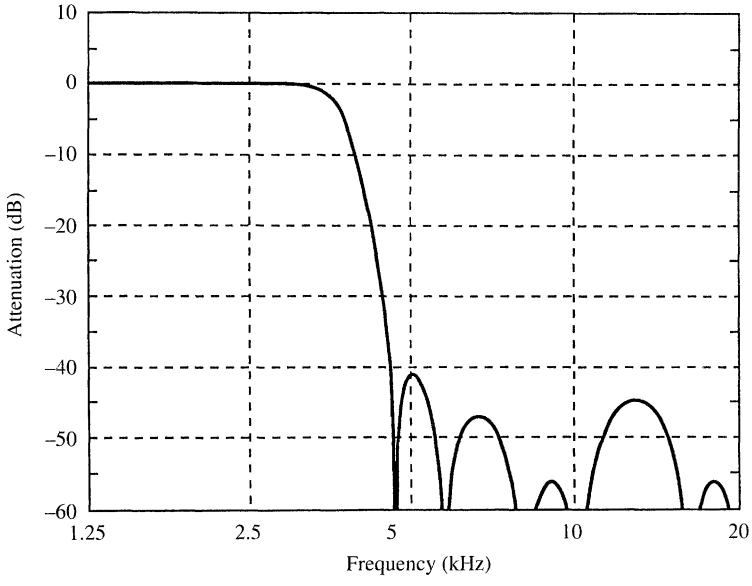

Figure 1.36 Frequency response of a low-pass filter used for decimating 4 kHz telephone signals from 40 to 8 kilowords per second (kw/s). Its z transform is given by

$$\left(\frac{1-\frac{3}{2}z^{-1}+z^{-2}}{1-\frac{11}{8}z^{-1}+\frac{5}{8}z^{-2}}\right)\left(\frac{1-\frac{5}{4}z^{-1}+z^{-2}}{1-\frac{101}{64}z^{-1}+\frac{7}{8}z^{-2}}\right)(1+z^{-1})(1+z^{-2})\left(\frac{1-z^{-5}}{1-z^{-1}}\right)$$

plane, and zeros at 8 and 16 kHz are provided by the accumulate-and-dump stage. The poles of the recursive stages are chosen to flatten the in-band response. These circuits are easy to build because the word lengths of the coefficients are short [22].

1.4 OVERSAMPLING D/A CONVERTERS

1.4.1 Demodulating Signals at Elevated Word Rates

The lower part of Figure 1.2 shows an outline of an oversampling D/A converter. In this circuit a digital filter interpolates sample values of the input signal in order to raise the word rate well above the Nyquist rate [21]. A demodulator then truncates the words and converts them to analog form at the high sample rate. In most applications it is advantageous to raise the word rate of the signal in stages, in much the same way as it was decimated in stages at the encoder. We illustrate the details of the oversampling method for D/A conversion by an example that processes 4-kHz telephone signals encoded into 16-bit words at 8 kHz. Figure 1.37 shows an outline of this oversampling D/A converter [22]. The input words enter a register from which they feed into a low-pass filter at 32 kHz: Each word repeats four times. The output of the filter resembles a PCM encoding of the signal at 32 kHz. The next stage is a linear interpolation that inserts three new values between each adjacent pair of 32-kHz samples, raising the word rate to 128 kHz. The words enter a register from which they feed the demodulator at 1 MHz; each word repeats eight times. The demodulator rounds off the code to single-bit words, converts them to analog levels, and smooths these with an analog filter. The single-bit quantization occurs in a feedback circuit that shapes the spectrum of the quantization noise, moving most of the power far above the signal band. The 1 MHz demodulation rate is sufficiently high that a very simple analog filter will smooth the noise.

The filtering actions [21] that are inherent in the interpolation that raises the word rate from 8 kHz to 1 MHz smooth out sampling images of the signal, leaving only those adjacent to the new sampling rate, 1 MHz, and its harmonics. Figure 1.38 illustrates this action: (a) represents the spectral density of the baseband signal, (b) is spectral density when sampled at the Nyquist rate, (c) is the frequency response of the low-pass filter including the sinc response of the holding register, both (d) and (e) represent the output spectrum of the low-pass filter, (f) is the sinc^2 response of linear interpolation, (g) is the result of this interpolation, (h) is the frequency response of the final holding register, and (i) is the spectral density of its output.

The filter requirements for attenuating sampling images of the signal at the decoder are usually less stringent than are the requirements for preventing aliasing at the encoder.

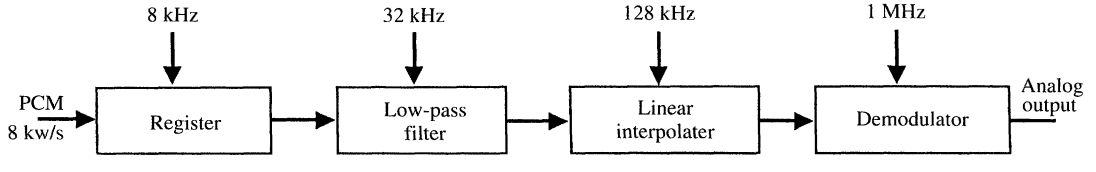

Figure 1.37 Oversampling decoder for 4 kHz signals.

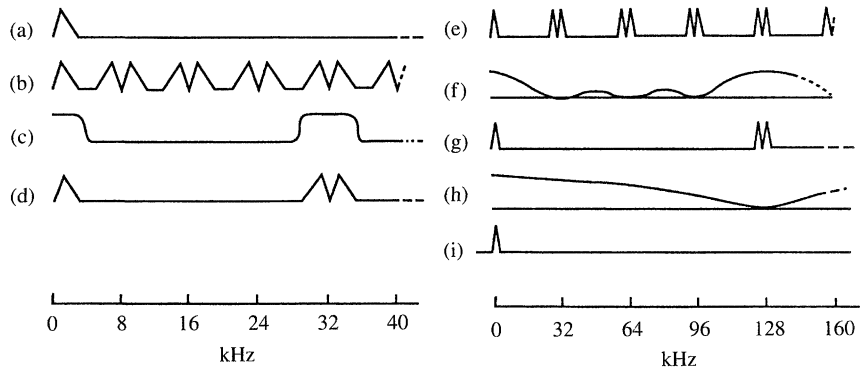

Figure 1.38 Spectral densities of signals, and the frequency response of filters used for interpolating sample values: (a) spectral density of the signal; (b) spectral density of the sampled signal; (c) low-pass filter characteristic; (d, e) spectral density of the filter output on different frequency scales; (f) frequency response of linear interpolation; (g) spectral density of the interpolated signal; (h) frequency response of the holding register; (i) spectral density of the held signal.

Consequently, a copy of the low-pass filter structure used in the encoder of Figure 1.28 is used as the low-pass filter in the decoder of Figure 1.27. The zeros in the filter response, which were provided by the accumulate-and-dump at the output of the A/D converter, are now provided by the holding register at the input of the D/A converter.

1.4.2 Interpolating with sinc^K-Shaped Filter Functions

The frequency responses associated with the linear interpolation circuit and the holding register are sampled sinc^K functions. The amount by which they attenuate sampling images of the signal needs to be calculated in order to determine good values for the intermediate sampling frequencies used between stages of the circuit. For this purpose, let f_I be the rate at which digital words are applied to an interpolating stage, and Nf_I the rate at which words emerge at its output. Here, N is defined as the interpolation ratio. The frequency response of this class of interpolation stage can be expressed as

$$I(f) = \frac{\text{sinc}^K(\pi f/f_I)}{\text{sinc}^K(\pi f/Nf_I)} \tag{1.34}$$

where K is the order of the interpolation. Images of the signal will be situated adjacent to f_I and all of its harmonics (i.e., in the frequency range $kf_I \pm f_0$), as illustrated in Figure 1.38. The attenuation of the unwanted images is least at the frequency $f_I - f_0$. The attenuation at this frequency may be read directly from the graphs in Figure 1.32, which was provided originally for evaluating decimators. The relevant value of the abscissa is now the oversampling ratio of f_I at the input to the interpolator (i.e., $f_I/2f_0$).

As an example, some telephone applications require that sampling images be attenuated by at least 28 dB. For interpolation ratios N greater than 3 this is achieved by using a holding register ($K = 1$) and an oversampling ratio of at least 16 at the input to the stage. Linear interpolation ($K = 2$) is satisfactory for oversampling ratios down to 4. These results are somewhat conservative because the digital low-pass filter attenuates the signal at the upper edge of its passband, and the analog filter at the output contributes to the smoothing of images.

1.4.3 Demodulator Stage

1.4.3.1 Quantizing the Digital Signal. The final stage of demodulation rounds off the long digital words to short ones, preferably all the way to single-bit words that will conveniently convert to analog form. Circuit structures of these quantizers resemble those of the modulators described in Section 1.2. The main difference is that the signals processed in the demodulator are digital instead of analog; hence, there is little trouble in achieving high precision. Properties of quantizing noise derived in Section 1.2 apply the noise of demodulation and may be used to determine the rates required to ensure adequately low demodulation noise. The need for this quantization distinguishes oversampling D/A converters from their conventional counterparts. Conventional converters contain no such intentional source of error, but they are much more sensitive to analog circuit imprecision.

Just as there are several forms of oversampling modulators, so also there are corresponding forms of demodulators. In general, there is no clear general advantage of one demodulator structure over others; the best choice depends on requirements of the application and on properties of the technology. Because much of the signal processing is digital in demodulators and analog in modulators, the trade-offs in their designs differ. We next discuss the design of several demodulator structures.

1.4.3.2 Quantization with Error Feedback. The error feedback circuit in Figure 1.23 is unsuitable for use as a modulator because of its sensitivity to inaccuracy in the analog subtractors, but this sensitivity is not a major concern in design of a demodulator. Figure 1.39 shows a digital implementation of an error feedback quantizer [25]. It uses digital codes that represent only positive values. The sum generated by the adder is quantized by using only its most significant bits as output. The remaining bits then represent the negative of the quantization error. These are delayed and added to the next input sample for error correction. In the extreme case of single-bit quantization, only the carry bit from the adder constitutes the output, and all the sum bits feed back.

The noise introduced into the signal by this quantization is the same form as first-order modulation noise [Eq. (1.8)]. For busy signals its spectral density is given by Eq. (1.9), and the noise power in the signal band by Eq. (1.11). For slowly changing inputs, the noise has a pattern structure as illustrated by Figure 1.8. For this and other reasons, first-order demodulation does not find wide application today; high-order demodulators are usually preferred. The circuit was useful at times when it was important to have very simple circuits.

Higher order noise shaping is achieved by replacing the delay in the feedback path by a prediction filter [12]. Figure 1.40 shows a noise-shaping demodulator of this kind. The

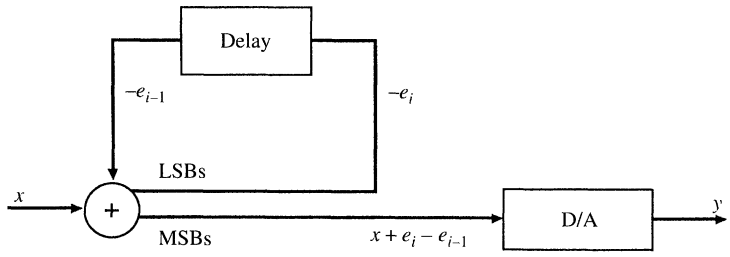

Figure 1.39 Digital quantization with error feedback (LSB, least significant bit; MSB, most significant bit).

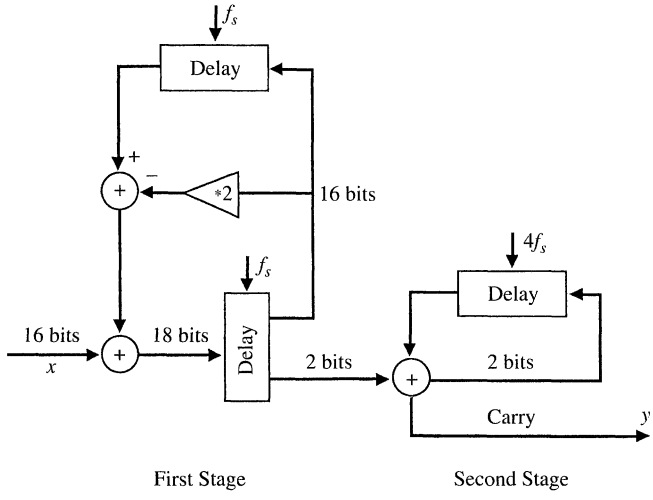

Figure 1.40 Second-order error-feedback quantizer with an auxiliary first-order quantization to 1-bit code.

first stage of the circuit is a second-order quantizer. This circuit requires that its output words comprise of at least two bits if clipping of signal amplitudes in the feedback loop is to be avoided. Including a clipping circuit in order to get single-bit outputs causes significant increase in noise, no matter where the clipping occurs. To avoid this penalty, the second stage of the circuit in Figure 1.40 is included to perform a first-order quantization of the two-bit words to single-bit ones [26]. This second stage is clocked at least four times faster than the first stage. The technique is possible when this higher frequency is achievable, but then, one also has the option of avoiding the need for the second stage altogether, by clocking the first stage fast enough to allow for the increased noise caused by clipping the feedback signal.

1.4.3.3 Cascaded Demodulators. Demodulator circuits can be cascaded [27] in the same way that modulators are cascaded in Figure 1.24. Figure 1.41 shows a cascade of two first-order quantizers in which the derivative of the second-stage output is added to the output of the first stage. This results in a cancellation of the first-stage noise from the net output. In contrast to cascaded analog quantizers, these digital quantizers can provide perfect cancellation because there is no inherent reason for error in the digital circuits besides the two quantization errors.

Even when the outputs of the individual stages are single-bit words, the net output contains two-bit words. This is an advantage of cascaded modulators because it allows the output to oscillate between four levels and avoids introducing excess noise into large signals. However, for demodulation this advantage is outweighed by the need for high precision in the four-level D/A at the output. This difficulty has been overcome to some extent by providing a separate D/A converter at the output of each stage. Then the signal passes only through the first-stage D/A, which should be two level to avoid distorting the signal. Only noise passes through the second D/A, and its imperfections result in imperfect noise cancellation, not signal distortion. Reference [27] describes a demodulator where the differentiation of the second-stage output is performed in the analog circuit; then both D/A converters can be single-bit ones. The precision required in matching the analog circuits is similar to the accuracy requirement of parameter g in Eq. (1.24).

1.4.3.4 Circuit Design for ΔΣ Demodulation. A digital implementation of an ordinary ΔΣ demodulator can serve to quantize the signal in a demodulator. There are many ways of designing these circuits, and Figure 1.42 shows one example of a second-order quantizer. This circuit introduces noise that is given by Eqs. (1.19) and (1.20). Although it is not troubled by leakage in the accumulators, tones may be present in the

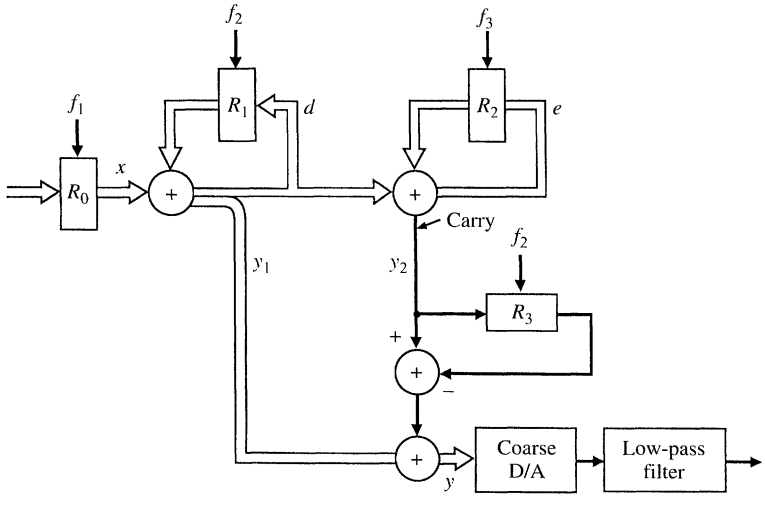

Figure 1.41 Cascade of two first-order digital quantizers with digital combining of their outputs. Single-bit words travel on the bold single-line paths. Multibit words travel on the wide paths.

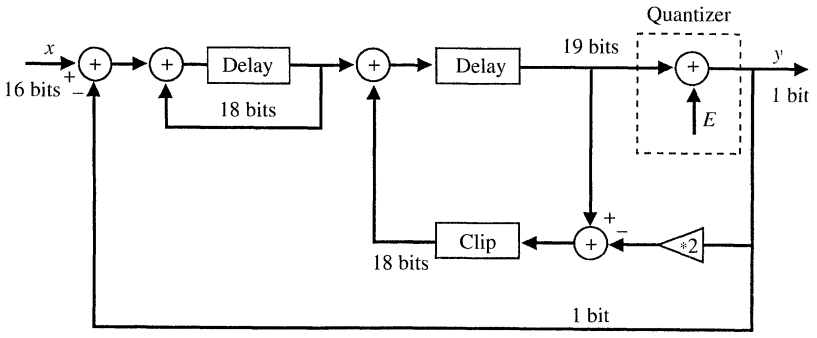

Figure 1.42 Equivalent circuit of a digital second-order ΔΣ quantizer.

noise. The randomness in the oscillation pattern of second-order modulators depends on avoiding signals that are rational multiples of the level spacing in the accumulators. This is not possible in digital implementations, where the signal values are always rational. One method of ensuring randomness is to inject a relatively large dither signal. Another is to add a random bit pattern to the least significant bit of the input, to mimic the effect of having irrational values stored in the first accumulator.

Structures used for higher order feedback quantizers of the type described in Section 1.2.3.6 may also be used in demodulators. It may also be possible to use higher order digital ΣΔ quantizers with digital control circuits that prevent them from going into saturating limit cycles.

1.5 CONCLUSION

Oversampling methods can provide very high resolutions even when relatively inaccurate analog components are used. For example, 20-bit resolution has been reported for 20-kHz audio applications. The ever-increasing speed capabilities of new VLSI technology will allow larger oversampling ratios and possibly higher resolutions, but this will soon be limited by circuit noise.

Designers of oversampling converters can select from a wide variety of architectures for modulators and demodulators each with its own advantages and disadvantages. They can make trade-offs between oversampling ratios, resolution, circuit complexity, and circuit tolerances and choose from numerous designs of digital decimation and interpolation filters. The later chapters of this book will describe many of these options in more detail than has been attempted here. The second chapter will lay a rigorous foundation for the design equations that have been introduced here.

REFERENCES

[1] J. C. Candy and G. C. Temes, "Oversampling methods for A/D and D/A conversion," *Oversampling Delta-Sigma Data Converters,* IEEE Press, New York, 1992, pp. 1–275.

[2] J. C. Candy, "A use of limit cycle oscillations to obtain robust analog-to-digital converters," *IEEE Trans. Commun.,* vol. COM-22, pp. 298–305, March 1974.

[3] J. C. Candy, Y. C. Ching, and D. S. Alexander, "Using triangularly weighted interpolation to get 13-bit PCM from a sigma-delta modulator," *IEEE Trans. Commun.,* vol. COM-29, pp. 815–830, June 1981.

[4] A. N. Netravali, "Optimum filters for interpolative A/D converters," *Bell Sys. Tech. J.,* vol. 56, pp.1629–1641, Nov. 1977.

[5] J. C. Candy and O. J. Benjamin, "The structure of quantization noise from sigma-delta modulation," *IEEE Trans. Commun.,* vol. COM-29, pp. 1316–1323, Sept. 1981.

[6] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory,* vol. IT-36, pp. 1220–1244, Nov. 1990.

[7] B. H. Leung, R. Neff, P. R. Gray, and R. W. Brodersen, "Area-efficient multichannel oversampled PCM voice-band coder," *IEEE J. Solid-State Circuits,* vol. SC-23, pp. 1351–1357, Dec. 1988.

[8] J. C. Candy, "A use of double integration in sigma-delta modulation," *IEEE Trans. Commun.,* vol. COM-33, pp. 249–258, March 1985

[9] M. W. Hauser and R. W. Brodersen, "Circuit and technology considerations for MOS delta-sigma A/D converters," *IEEE Proc. ISCAS '86,* pp.1310–1315, May 1986.

[10] T. Cataltepe, G. C. Temes, and L. E. Larson, "Digitally corrected multi-bit Σ-Δ data converters," *IEEE Proc. ISCAS '89,* pp. 647–650, May 1989.

[11] R. W. Adams, "Companded predictive delta modulation; a low cost conversion technique for digital recording," *J. Audio Eng. Soc.*, vol. 32, pp. 659–672, Sept. 1984.

[12] H. A. Spang III and P. M. Schultheiss, "Reduction of quantizing noise by use of feedback," *IRE Trans. Commun. Sys.,* pp. 373–380, Dec. 1962.

[13] B. E. Brandt, D. E. Wingard, and B. A. Wooley, "Second-order sigma-delta signal acquisition," *IEEE J. Solid-State Circuits,* vol. SC-26, pp. 618–627, April 1991.

[14] S. H. Ardalan and J. J. Paulos, "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Trans. Circuits Sys.,* vol. CAS-34, pp. 593–603, June 1987.

[15] B. E. Boser and B. A. Wooley, "The design of sigma-delta modulation analog-to-digital converters," *IEEE J. Solid-State Circuits,* vol. SC-23, pp. 1298–1308, Dec. 1988.

[16] K. C. H. Chao, S. Nedeem, W. L. Lee, and C. G. Sodini, "A higher order topology for interpolative modulators for oversampling A/D conversion," *IEEE Trans. Circuits Sys.,* vol. Cas-37, pp. 309–318, March 1990.

[17] Y. Matsuya, K. Uchimura, A. Iwata, et al., "A 16-bit oversampling A-to-D conversion technology using triple-integration noise shaping," *IEEE J. Solid-State Circuits,* vol. SC-22, pp. 921–929, Dec. 1987.

[18] L. Logo and M. Copeland, "A 13 bit ISDN-band oversampled ADC using two-stage third order noise shaping," *IEEE Proc. Custom IC Conf.,* pp. 21.2.1–21.2.4, Jan. 1988.

[19] L. A. Williams III and B. A. Wooley, "Third-order cascade sigma-delta modulators," *IEEE Trans. Circuits Sys.,* vol. CAS-38, pp. 489–498, May 1991.

[20] T. C. Leslie and B. Singh, "An improved sigma-delta modulator architure," *IEEE Proc. ISCAS '90*, pp. 372–375, May 1990.

[21] R. E. Crochiere and L. R. Rabiner, "Interpolation and decimation of digital signals—a tutorial review," *Proc. IEEE,* vol. 69, pp. 300–331, March 1981.

[22] J. C. Candy, B. A. Wooley, and O. J. Benjamin, "A voiceband codec with digital filtering," *IEEE Trans. Commun.,* vol. COM-29, pp. 815–830, June 1981.

[23] J. C. Candy, "Decimation for sigma delta modulation," *IEEE Trans. Commun.,* vol. CAS-31, pp. 913–924, Nov. 1984.

[24] E. B. Hogenaur, "An economical class of digital filters for decimation and interpolation," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-29, April 1981.

[25] G. R. Ritchie, J. C. Candy, and W. H. Ninke, "Interpolative digital to analog converters," *IEEE Trans. Commun.,* vol. COM-22, pp. 1797–1806, Nov. 1974.

[26] H. G. Musmann and W. Korte, "Generalized interpolative method for digital/analog conversion of PCM signals," U.S. Patent, No. 4,467,316, 1984 (filed 1981).

[27] J. C. Candy and An-Ni Huynh, "Double interpolation for digital-to-analog conversions," *IEEE Trans. Commun.,* vol COM-34, pp. 77–81, Jan. 1986.