An Overview of Sigma-Delta Converters

How a 1-bit ADC achieves more than 16-bit resolution

PERVEZ M. AZIZ,
HENRIK V. SORENSEN, and
JAN VAN DER SPIEGEL

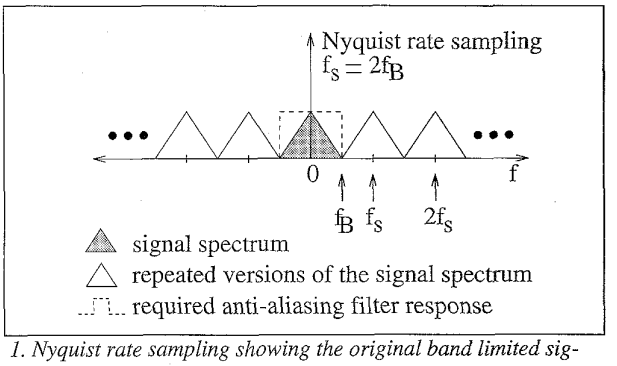
Although real world signals are analog, it is often desirable to convert them into the digital domain using an analog to digital converter (ADC). Motivating designers to apply this conversion is the efficient transmission and storage of digital signals. Digital representation of an audio signal, for example, allows CD players to achieve virtually error free storage using optical disks [1]. Intricate processing of the signal may also necessitate analog to digital (A/D) conversion, since such processing is only feasible in the digital domain using either conventional digital computers or special purpose digital signal processors (DSPs). Signal processing in the digital domain is also extremely useful in such areas as biomedical applications, providing the needed accuracy for tasks such as ultrasound imaging.

One technique, sigma-delta modulation, has become quite popular for achieving high resolution. One significant advantage of the method is that analog signals are converted using only a 1-bit ADC and analog signal processing circuits having a precision that is usually much less than the resolution of the overall converter.

Although sigma-delta concepts have existed since the middle of the century, it is only in the last two decades that this method has become more attractive [2]. One reason is that recent advances in VLSI technology, focused towards realizing high speed densely packed digital circuits, have made feasible the adequate digital processing of the bit stream. Using sigma-delta A/D methods, high resolution can be obtained for only low to medium signal bandwidths.

This article briefly describes conventional A/D conversion, as well as its performance modeling. We then look at the technique of *oversampling*, which can be used to improve the resolution of classical A/D methods. We discuss how sigma-delta converters use the technique of *noise shaping* in addition to oversampling to allow high resolution conversion of relatively low bandwidth signals. Next, we examine the use of sigma-delta converters to convert narrowband bandpass signals with high resolution. Several parallel sigma-delta converters, which offer the potential of extending high resolution conversion to signals with higher bandwidths, are also described.

1053-5888/96/$4.00©1996IEEE



1. Nyquist rate sampling showing the original band limited signal spectrum, periodically repeated versions of the signal spectrum due to sampling, and the anti-aliasing filter response needed to band limit the signal.

PCM A/D Conversion

Nyquist Rate Conversion

Sampling

Analog to digital conversion of a signal is traditionally described in terms of two separate operations: uniform sampling in time, and quantization in amplitude. In the sampling process, a continuous time signal is sampled at uniformly spaced time intervals, T_s. The samples, $x[n]$, of the continuous time signal, $x(t)$ can be represented as $x[n] = x(nT_s)$. The effect, in the frequency domain, of the sampling process is to create periodically repeated versions of the signal spectrum at multiples of the sampling frequency $f_s = 1/T_s$ [3, pp. 80-87]. This relationship is written in Eq. 1, where $X_s(f)$ represents the spectrum of the sampled signal, and $X(f)$ is the spectrum of the original continuous time signal.

$$X_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X(f - kf_s) \tag{1}$$

The sampling process is shown graphically in Fig. 1 for the case where $f_s = 2f_B$, and f_B is the bandwidth of the signal. In general, the signal can be reconstructed back to continuous time if the repeated versions of the signal spectrum do not overlap. Thus, the signal must be band limited to half the sampling rate, i.e., a signal with bandwidth f_B must be sampled at a rate greater than twice the bandwidth, $f_s \geq 2\ f_B$. Interference between the repeated versions of the signal spectrum is known as aliasing and it prevents reconstruction of the signal.

Even if a signal is nominally band limited to $f_s/2$, an anti-aliasing filter is often used to ensure that the signal is indeed band limited. For example, speech has a nominal bandwidth of 4 kHz and so in principle can be sampled at 8 kHz. However, there is some residual signal energy above 4 kHz, which results in aliasing if a 8 kHz sampling rate is used. The anti-aliasing filter is a continuous time analog filter preceding the sampler.

The case where $f_s = 2f_B$ is known as Nyquist rate sampling, and clearly the anti-aliasing filter here must have a very sharp cutoff at frequency $f_B = f_s/2$ as shown in Fig 1. Later, we will discuss how the sharp cutoff requirement on this filter can be relaxed.

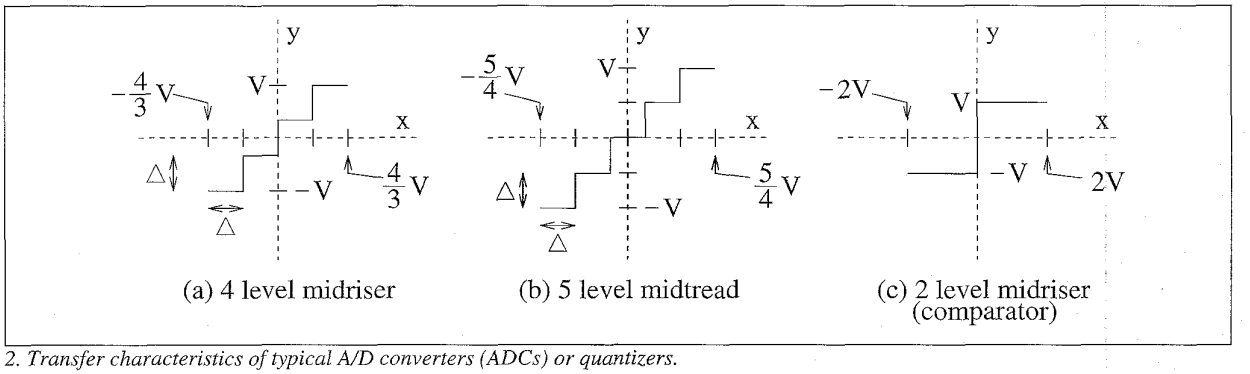
The discretization or quantization in time as a result of the sampling is an invertible operation, since no signal information is lost and the original continuous time signal can be perfectly reconstructed [3, pp. 80-91]. Note that although Fig. 1 shows the sampling process for the case where the signal is a baseband signal, i.e., the spectrum has a bandwidth centered at DC frequency, Eq. 1 still describes the sampled spectrum even if the signal spectrum is centered at some higher frequency f_c. In this case, for a signal bandwidth f_B, the signal spectrum occupies the region $[f_c - f_B/2, f_c + f_B/2]$, and it will still be possible to avoid aliasing and reconstruct the signal provided that $f_s \geq 2\ f_B$.

Quantization

Once sampled, the signal samples must also be quantized in amplitude to a finite set of output values. Typical transfer characteristics of quantizers or A/D converters with an input signal sample, $x[n]$, and an output, $y[n]$, are shown in Fig. 2.

Quantization is a non-invertible process, since an infinite number of input amplitude values are mapped to a finite number of output amplitude values. The quantized output amplitudes are usually represented by a digital code word composed of a finite number of bits. For example, for the 1 bit A/D converter of Fig. 2c, the output levels V and $-V$ can be mapped to digital codes "1" and "0." The digital code words are often said to be in pulse code modulation (PCM) format.

Another way of looking at this would be to plot the digital



2. Transfer characteristics of typical A/D converters (ADCs) or quantizers.

code words instead of the quantized amplitude values for y in Fig 2. The quantized output amplitude values can also be considered the output of an ideal digital to analog converter (DAC) whose inputs are the corresponding digital code words.

An ADC or quantizer with Q output levels is said to have N bits of resolution where $N = \log_2(Q)$. As should be clear from Fig 2, for an ADC with Q quantization levels, only input values separated by at least $\Delta = 2V/(Q-1)$ can be distinguished or resolved to different output levels. N digital bits are needed to encode the Q codewords corresponding with each output level. The difference between the binary digital codes for two adjacent output levels is one least significant bit (LSB) of the overall N bit codeword. Consequently, a difference in input amplitudes corresponds to a one LSB difference in the digital output code words.

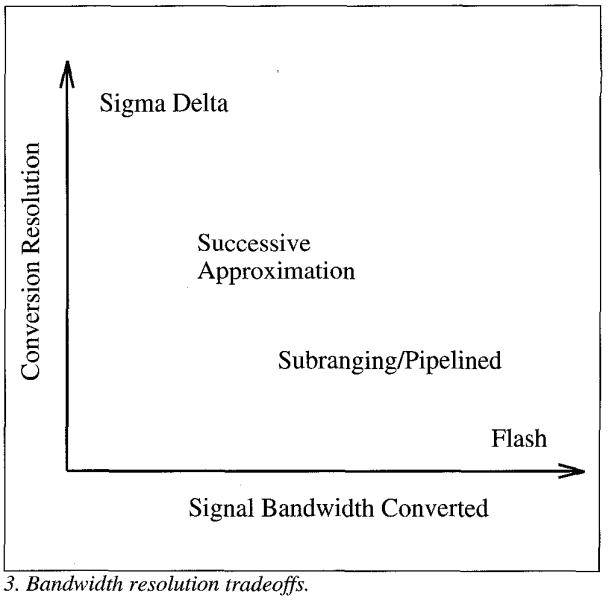
We now point out some general properties of the quantization transfer curves shown in Figs. 2a-2b for a four level (2 bit) "midriser" and a five level (roughly 2 bit) "midtread" ADC. Unlike the midtread ADC, the midriser ADC does not contain a zero output level for a zero input value, effectively creating a DC offset that may be undesirable in some applications. Note that all the transfer characteristics shown in Fig. 2 are symmetric. The midriser needs to have an even number of output levels to produce a completely symmetric transfer curve, whereas the midtread needs an odd number of output levels.

The midriser ADC's symmetric characteristic, with an even number of levels, is an advantage because the number of output levels, Q, can be made a power of two and encoded with exactly $N = \log_2(Q)$ bits. However, the number of output levels, Q, for a symmetric midtread ADC must be odd, and so cannot be made a power of two and encoded as efficiently. The number of bits needed will be $N = \log_2(Q-1) + 1$, where Q-1 is chosen a power of two. If the number of levels for the midtread ADC is forced to be a power of two by using only Q-1 levels, it will no longer have a symmetric transfer characteristic and will distort large amplitude symmetric input signals (e.g., a sinusoid). This distortion, of course, may be negligible when the number of output levels is very large.

These issues may play a role in choosing whether a midriser or midtread quantizer transfer characteristic should be used. For the special case of a 2 level quantizer, a midtread characteristic will not be able to represent both positive and negative output levels, and so will severely distort a signal containing samples of both polarities. For this 2 level case, a midriser characteristic, shown in Fig. 2c, will almost always be used.

Trading Resolution for Bandwidth

Most conventional A/D converters, such as the successive approximation, subranging, and flash converter types quantize signals sampled at, or slightly above, the Nyquist rate. Consequently, these converters are often referred to as Nyquist rate PCM converters. These, and other converters, provide tradeoffs among signal bandwidth, output resolution, and the complexity of the analog and digital hardware.



3. Bandwidth resolution tradeoffs.

Qualitative bandwidth and resolution tradeoffs of some of these A/D techniques, as well as sigma-delta conversion, is shown in Fig 3. As is evident from Fig 3, sigma-delta A/D converters attain the highest resolution for relatively low signal bandwidths. Consequently, sigma-delta techniques are often used in speech applications where the signal bandwidth is only 4 kHz and where up to 14 bits of resolution may be needed. Similarly, sigma-delta ADCs are popular for digital audio applications, where the signal bandwidth is 20-24 kHz and where high fidelity audio requires 16-18 bits of resolution. Flash converters, on the other hand, may be used for broadcast video applications where the signal band is about 5 MHz, but the resolution required is only about 8 bits.

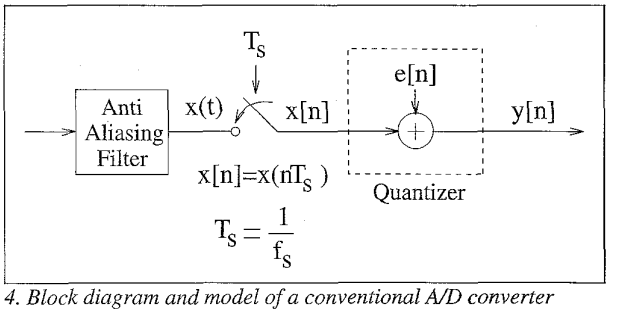
Performance Modeling

Having looked at the sampling and quantization processes, we now examine the A/D converter and characterize its performance. The diagrams in Fig. 2 show the transfer characteristic for typical quantizers with input x and output y.

Let the maximum and minimum quantized output values always be V and $-V$. The least significant bit (LSB) of an ADC with Q quantization levels is equivalent to $2V/(Q-1)$. For both the midriser and midtread type of ADCs of Fig 2, the magnitude of the quantization error ($e = y - x$) between the output and input does not exceed half a LSB, i.e., $|e| \leq \Delta/2$, provided that $|x| \leq V + \Delta/2$. Under these circumstances, the quantizer or ADC is said to be not overloaded. For $|x| > V + \Delta/2$ (hence, $|e| > \Delta/2$), the ADC is said to be overloaded.

The quantizer embedded in any ADC is a non-linear system, which makes its analysis difficult. To make the analysis tractable, the quantizer is often linearized and modeled by a noise source, $e[n]$, added to the signal $x[n]$, to produce the quantized output signal $y[n]$:

$$y[n] = x[n] + e[n] \tag{2}$$



4. Block diagram and model of a conventional A/D converter (ADC) system.

A block diagram of an A/D system showing the sampling process and the quantizer model is shown in Fig 4. To further simplify the analysis of the noise from the quantizer, the following assumptions about the noise process and its statistics are traditionally made [3, p. 120]:

- The error sequence, $e[n]$, is a sample sequence of a stationary random process.
- $e[n]$ is uncorrelated with the sequence $x[n]$.
- The probability density function of the error process is uniform over the range of quantization error, i.e., over $\pm \Delta/2$.
- The random variables of the error process are uncorrelated, i.e., the error is a white noise process.

Under certain conditions, such as when the quantizer is not overloaded, N is large, and the successive signal values are not excessively correlated, these assumptions are reasonable [4]. Consider an N bit ADC with $Q = 2^N$ quantization levels, i.e., with $\Delta = 2V/Q\text{-}1 = 2V/(2^N\text{-}1)$. For a zero mean $e[n]$, its variance σ_e^2 or power is

$$\sigma_e^2 = \frac{\Delta^2}{12} = \left(\frac{2V}{2^N - 1}\right)^2 / 12 \cong \left(\frac{2V}{2^N}\right)^2 / 12 \tag{3}$$

If the signal is treated as a zero mean random process and its power is σ_x^2, then the signal to quantization noise ratio is:

$$SNR = 10\log\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\log\left(\frac{\sigma_x^2}{V^2}\right) + 4.77 + 6.02\ (dB) \tag{4}$$

Note that for each extra bit of resolution in the ADC, i.e., for every increment in N, there is about a 6 dB improvement in the SNR. Thus, there is a direct relationship between the resolution of an ADC and its SNR, and it is common to equate differences in SNR in dB to bits by dividing the dB value by 6. For example, if an ADC has a SNR that is 3 dB better than that of another ADC, the better ADC will be said to have 1/2 bit higher resolution. Also, note that for a given N, the SNR in dB is linearly related to the signal power, σ_x^2, in dB.

Let us now examine the dynamic range of the ADC, which is a measure of the range of input amplitudes for which the ADC produces a positive SNR. For sinusoidal inputs, the dynamic range of the A/D converter is defined as the ratio of the signal power of a full scale sinusoid to the signal power of a small sinusoidal input that results in a SNR of 1 (or 0 dB) [5]. The signal power of a full scale sinusoid is $V^2/2$. A sinusoid with signal power $\sigma_x^2 = \sigma_e^2 = \Delta^2/12$ will result in an SNR of 1, or 0 dB.

The dynamic range, by definition, is

$$\left(\frac{V^2}{2} / \frac{\Delta^2}{12}\right) \cong \left(\frac{V^2}{2} / \frac{(2V/2^N)^2}{12}\right)$$

This expression reduces to a dynamic range value given by:

$$R = 6.02N + 1.76\ (dB) \tag{5}$$

Note that the ratio of $V^2/2$ to $\Delta^2/2$ is just the peak SNR of the ADC for a sinusoidal input. Consequently, the dynamic range of the Nyquist rate ADC is the same as its peak SNR. Sigma-delta converters do not necessarily have their peak SNR equal to their dynamic range. However, by using the dynamic range of a sigma-delta converter in Eq. 5 and calculating the corresponding N, we will be able to determine the resolution of a Nyquist rate PCM converter that would be required to produce the same dynamic range.

Limitations of Nyquist Rate ADCs

For Nyquist rate converters, each signal sample is quantized at the full precision or resolution of the converter. The resolution of such converters implemented on VLSI chips is limited by the technology in which these chips are fabricated. For example, some successive approximation A/D techniques rely on matching of two capacitors to perform a repeated division of a reference voltage by 2.

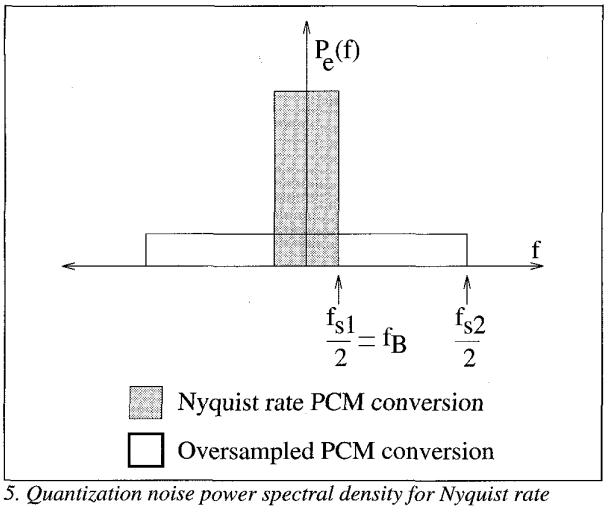
If such a converter is to convert signal values to N bits of resolution, the required matching on the capacitor components needs to be at least one part in 2^N. Matching of components to greater than 10 bits (one part in 2^{10}) or equivalently to more than 0.1% is difficult in VLSI.

High resolution Nyquist rate converters are extremely difficult to attain in current integrated circuit technology without the use of techniques such as laser trimming of components or calibration. Furthermore, if the signal is sampled too close to the Nyquist rate, the anti-aliasing filter must have a very sharp cutoff, which is a non-trivial design requirement for analog filters.

Oversampled PCM Conversion

System Description

Oversampled PCM conversion is a technique that improves the resolution obtained from straightforward Nyquist rate PCM conversion. This improvment is achieved by oversampling the signal, i.e., samples are acquired from the analog waveform at a rate significantly faster than the Nyquist rate. Each of these samples is quantized by an N bit ADC. Since quantization is described by Eq. 2, the total amount of noise power injected into the sampled signal, $x[n]$, is σ_e^2 and is given by Eq. 3.



5. Quantization noise power spectral density for Nyquist rate PCM and oversampled PCM conversion.

Obviously, this is exactly the same noise power produced by a Nyquist rate converter, but its frequency distribution is different because of the higher sampling rate. The performance modeling criteria designated the noise process as white, which means that the noise power is uniformly distributed between $-f_s/2$ to $f_s/2$, where f_s is the sampling frequency.

Fig 5 shows the power spectral density, $P_e(f)$, of the quantization noise for Nyquist rate sampling with rate f_{s1} and oversampling rate f_{s2}. For Nyquist rate sampling where the signal band, $f_B = f_{s1}/2$, all the quantization noise power, represented by the area of the tall shaded rectangle, occurs across the signal bandwidth.

In the oversampled case, the same noise power, represented by the area of the unshaded rectangle has been spread over a bandwidth equal to the sampling frequency, f_{s2}, which is much greater than the signal bandwidth, f_B. Only a relatively small fraction of the total noise power falls in the band $[-f_B, f_B]$, and the noise power outside the signal band can be greatly attenuated with a digital low-pass filter following the ADC.

After the low-pass filtering is performed, the signal can be downsampled to the Nyquist rate without affecting the signal to noise ratio. The collective operation of low-pass filtering and downsampling is known as decimation. The low-pass filter and downsampler are collectively called a decimator. A block diagram of an oversampled PCM system showing the sampling, the ADC model, and the decimator is presented in Fig 6.

Performance Modeling for Oversampled PCM Converters

By taking the Z transform of Eq. 2, we obtain the Z domain relationship between the input and output of an oversampled PCM converter:

$$Y(z) = X(z) + E(z) \tag{6}$$

where Y, X and E are the Z transforms of the output, input signal and the quantization error process, respectively. Based on our two-input linear system model for the quantizer, Eq. 6 states that in the Z domain, the output is the input plus the quantization error or noise. We can observe that X and E both experience a unity transfer function.

A more general way of writing Eq. 6, where X and E do not necessarily experience a unity transfer function, is

$$Y(z) = X(z)H_x(z) + E(z)H_e(z) \tag{7}$$

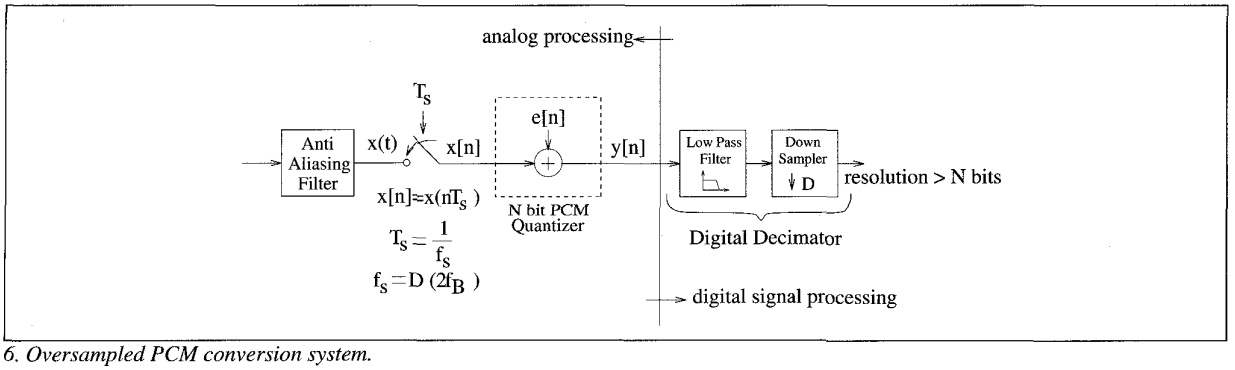
The output is now the input signal modulated by a signal transfer function, denoted by $H_x(z)$, plus the quantization noise modulated by a noise transfer function, denoted by $H_e(z)$. To evaluate the performance of such a converter, we need to find the total signal and noise power at the output of the converter. To do this, we need to evaluate the power spectral densities, $P_{xy}(f)$ and $P_{ey}(f)$, of the signal and noise at the output of the converter, based on the power spectral densities, $P_x(f)$ and $P_e(f)$ of the signal and noise at the input of the converter. We can make use of the fact that if a stationary random process with power spectral density $P(f)$ is the input to a linear filter with transfer function $H(f)$, the power spectral density of the output random process is $P(f)|H(f)|^2$. Consequently,

$$P_{xy}(f) = P_x(f)|H_x(f)|^2$$

$$P_{ey}(f) = P_e(f)|H_e(f)|^2$$

For the oversampled PCM converter, $|H_x(f)| = |H_e(f)| = 1$, and our white noise assumption for $e[n]$ states that $P_e(f) = \sigma_e^2/f_s$, which implies $P_{ey}(f)$ is also σ_e^2/f_s.

Assuming an *ideal* low-pass filter with cutoff frequency



6. Oversampled PCM conversion system.

f_B following the oversampled quantizer, the *in-band* noise power, σ_{ey}^2 at the output of the A/D is

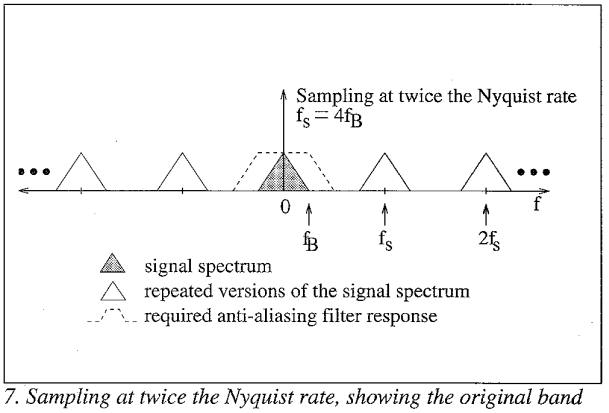
$$\sigma_{ey}^2 = \int_{-f_B}^{f_B} P_{ey}(f)df = 2\int_0^{f_B} P_{ey}(f)df$$
$$= \int_0^{f_B} \frac{2\sigma_e^2}{f_s} df = \sigma_e^2\left(\frac{2f_B}{f_s}\right)$$

Note that some of the noise power is now located outside of the signal band as a result of the oversampling, and so the in-band power σ_{ey}^2 is less than what it would have been without any oversampling (σ_e^2). Since the signal power is assumed to occur over the signal band only, it is not modified in any way and the signal power at the output (σ_{xy}^2) is the same as the input signal power σ_x^2. The maximum achievable SNR in dB is then:

$$SNR = 10\log\left(\frac{\sigma_x^2}{\sigma_{ey}^2}\right)$$
$$= 10\log(\sigma_x^2) - 10\log(\sigma_e^2) + 10\log\left(\frac{f_s}{2f_B}\right) \ (dB) \tag{8}$$

For the case of Nyquist rate sampling where $f_s = 2f_B$, this formula reduces to Eq. 3, which is the SNR for the Nyquist rate PCM quantizer. Letting the oversampling ratio $f_s/2f_B = 2^r$, we obtain,

$$SNR = 10\log(\sigma_x^2) - 10\log(\sigma_e^2) + 3.01r \ (dB) \tag{9}$$



7. Sampling at twice the Nyquist rate, showing the original band limited signal spectrum, repeated versions of the signal spectrum due to sampling, and an anti-aliasing filter that is sufficient to band limit the signal.

For every doubling of the oversampling ratio, i.e., for every increment in r, the SNR improves by about 3 dB, or the resolution improves by one-half bit.

Note that in this scheme, we are trading speed for resolution. The higher resolutions are obtained at the expense of requiring the internal PCM quantizer to quantize samples at the oversampled rate. Analog circuit complexity has also been traded for digital circuit complexity. The analog circuit complexity is simplified, since we have said the resolution of the internal N bit quantizer, an analog circuit, is lower than that of the overall conversion resolution.

Another benefit, which is a direct consequence of the oversampling, is that the analog anti-aliasing filter does not need as sharp a cutoff. This can be seen from Fig. 7, where a signal is sampled at four times the nominal signal bandwidth (twice the Nyquist rate). In this case, the anti-aliasing filter can have a transition band between f_B and $f_s/2$ as long as it provides very good attenuation beyond $f_s/2$. However, a price is paid in the digital domain, since the digital filter must attenuate the remaining quantization noise power (beyond f_B) as much as possible. In the process of filtering out-of-band quantization noise, any other noise that existed in the transition band of the anti-aliasing filter prior to sampling will be attenuated further. The closer the low-pass filter approximates an ideal low-pass filter, the more resources it will need.

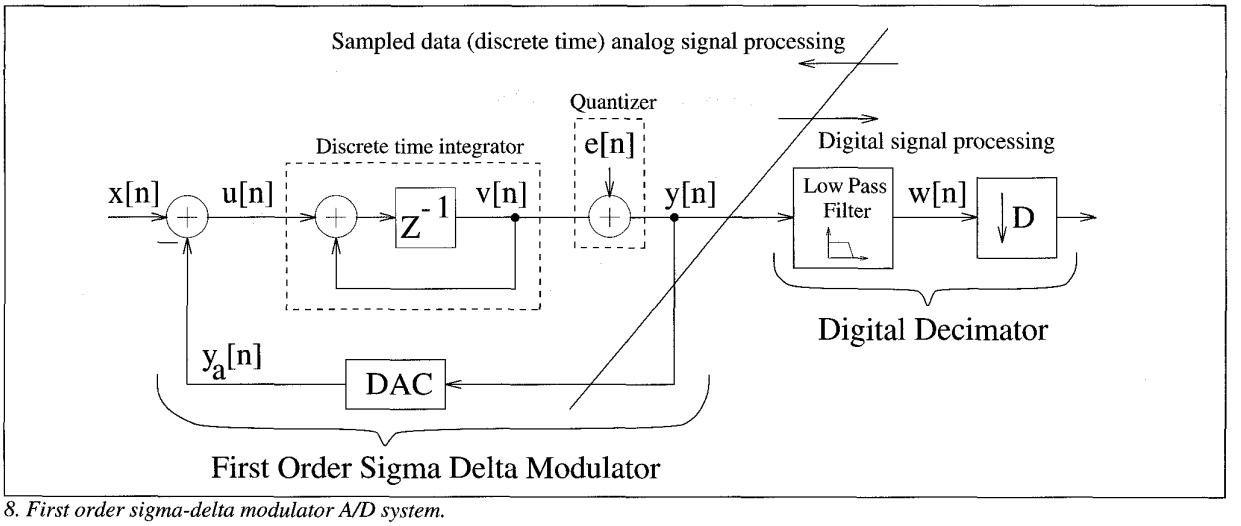
Finally, note that for high resolution conversion, one needs, $f_s >> f_B$, and the signal bandwidth must be small so that f_s does not exceed the maximum circuit speed attainable in the given technology. Let us use Eq. 8 for a simple calculation. Suppose we apply a full scale sinusoid with amplitude $V = 1$ corresponding with signal power $V^2/2 = 0.5$, as the input to an oversampled PCM ADC with a 20 kHz audio range signal band. Let the final desired resolution be 16 bits (for CD quality audio), which corresponds to a 98 dB SNR according to Eq. 4. Now, if we use an 8 bit A/D in an oversampled PCM scheme, i.e., if we use $N = 8$ in the expression for σ_e^2 in Eq. 8, we can calculate the f_s required for SNR = 98 dB and f_B = 20 kHz. The needed f_s is 2.64 GHz! Eight bit ADCs implemented in current CMOS technology certainly can not operate at such a high speed.

Suppose a 12 bit internal ADC is used instead. In this case, the required f_s is about 10 MHz, an operating speed that is not extremely high, per se, but which is still not trivial for a 12 bit ADC to attain. We will later see how sigma-delta modulation A/D conversion allows the use of internal ADCs with as low as 1 bit (i.e., $N = 1$) of resolution to achieve an overall resolution of 16 bits for a 20 kHz audio bandwidth.

Sigma-Delta Modulation A/D Conversion

A general way of writing the Z domain output of an A/D converter was given in Eq. 7, as $Y(z) = X(z)H_x(z) + E(z)H_e(z)$, where H_x is the signal transfer function (STF) and H_e is the noise transfer function (NTF). For oversampled PCM conversion, we saw that $H_x(z) = H_e(z) = 1$. This need not be the case and, in fact, oversampled A/D converters can be designed to incorporate noise shaping, where H_e is designed to be different from H_x such that H_x usually leaves the signal undisturbed but H_e shapes the noise to allow a high resolution output [2, 6].

Although the term delta-sigma ($\Delta\Sigma$) was used by some of the earliest researchers in the field [7], the term sigma-delta ($\Sigma\Delta$) has also become almost synonymous with noise shaping ADCs. We will use the term sigma-delta to describe noise shaping ADCs. As noted, oversampling reduces the quanti-



8. First order sigma-delta modulator A/D system.

zation noise power in the signal band by spreading a fixed quantization noise power over a bandwidth much larger than the signal band. Noise shaping or modulation further attenuates this noise in the signal band and amplifies it outside the signal band. Consequently, this process of noise shaping by the sigma-delta modulator can be viewed as pushing quantization noise power from the signal band to other frequencies. The modulator output can then be low-pass filtered to attenuate the out-of-band quantization noise and finally can be downsampled to the Nyquist rate.

The price of attaining high resolution is again a penalty in speed, as the hardware has to operate at the oversampled rate, and an increased complexity of the digital hardware. For high resolution conversion, the sampling frequency f_s must still be much greater than the signal bandwidth f_B, but, of course, not as great as needed by oversampled PCM conversion.

First Order Sigma-Delta Modulation

Operation and Performance Modeling

A block diagram of a first order sigma-delta modulator A/D system is shown in Fig 8. The system consists of an analog sigma-delta modulator, followed by a digital decimator. The modulator consists of an integrator, an internal A/D converter or quantizer, and a D/A converter (DAC) used in the feedback path.

The signal that is quantized is not the input $x[n]$ but a filtered version of the difference between the input and an analog representation, $y_a[n]$, of the quantized output, $y[n]$. The filter, often called the feedforward loop filter, is a discrete time integrator whose transfer function is $z^{-1}/(1-z^{-1})$.

The integrator and the rest of the sigma-delta analog circuit are typically implemented in sampled data switched capacitor technology. Consequently, the sampling operation is not shown explicitly in Fig. 8 or any other modulator architectures to be described in the rest of this article. Continuous-time versions of the modulator have also been considered [8], but this aspect of the modulators will not be discussed here. The linearized model replaces the quantizer with a noise source, $e[n]$, as shown in Fig 8.

If the DAC is ideal, it is replaced by a unity gain transfer function. The modulator output $Y(z)$ is then given by:

$$Y(z) = X(z)z^{-1} + E(z)(1 - z^{-1}) \tag{10}$$

so that $H_x(z) = z^{-1}$ and $H_e(z) = (1-z^{-1})$. The output is just a delayed version of the signal plus quantization noise that has been shaped by a first order Z domain differentiator or high-pass filter. The corresponding time domain version of the modulator output is

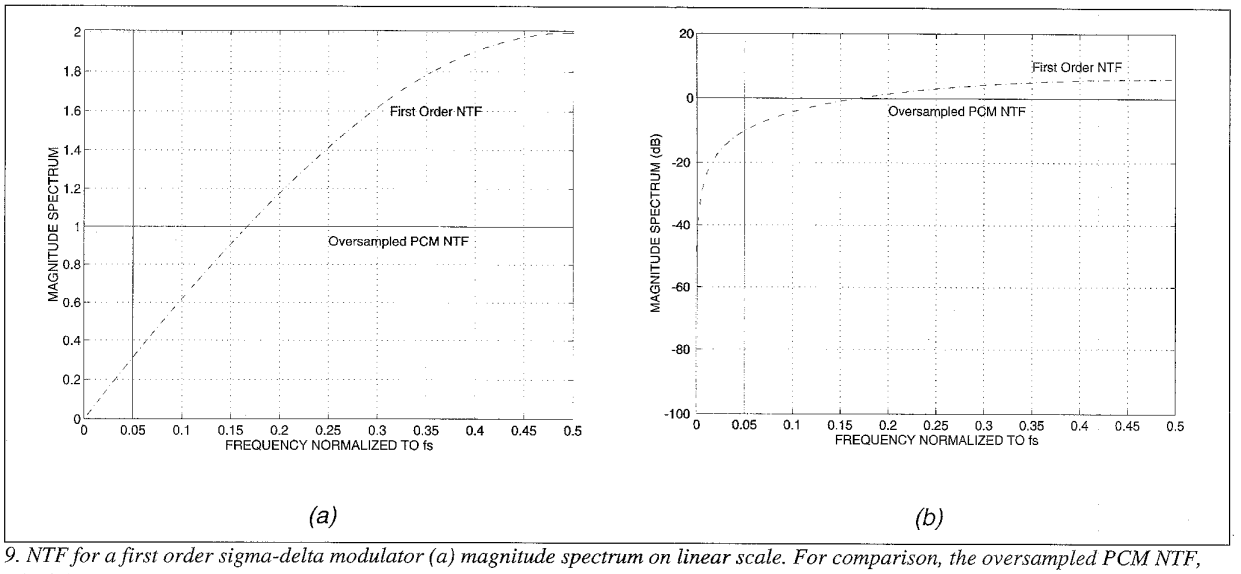
$$y[n] = x[n-1] + e[n] - e[n-1] \tag{11}$$

where the $e[n]$-$e[n-1]$ term is the first order difference of $e[n]$.

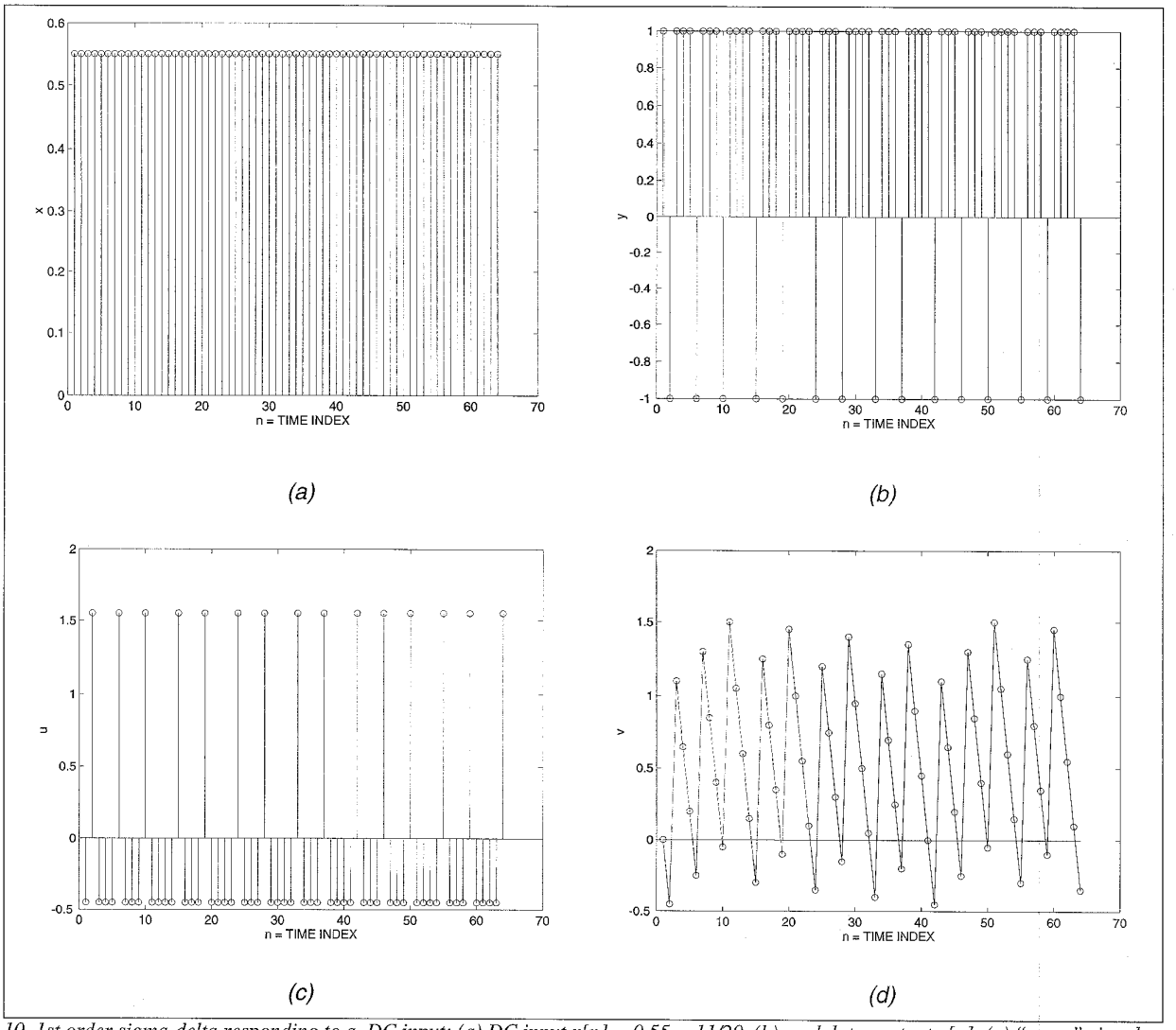
The magnitude spectrum of a first order sigma-delta noise transfer function (NTF) is plotted in Fig 9a, while Fig 9b shows the same plot in dB. The frequency axis has been normalized with respect to the sampling frequency, f_s.

Since $H_e(z)$ contains a zero at $z = 1$, i.e., at DC frequency on the unit circle of the Z plane, note the zero gain or infinite attenuation provided by the NTF at DC frequency. Note the large attenuation at lower frequencies and relative amplification at higher frequencies. For comparison, the oversampled PCM NTF, which has unity gain, is shown in Fig 9(a). The vertical bar demarcates the extent of the signal band, f_B, where $f_B = 0.05\ f_s$. Quantization noise to the left of the bar that contributes to the finite resolution of the modulator is greatly attenuated while noise to the right of the bar is not attenuated as much or is actually amplified. However, noise to the right of the bar can mostly be removed with a digital low-pass filter.

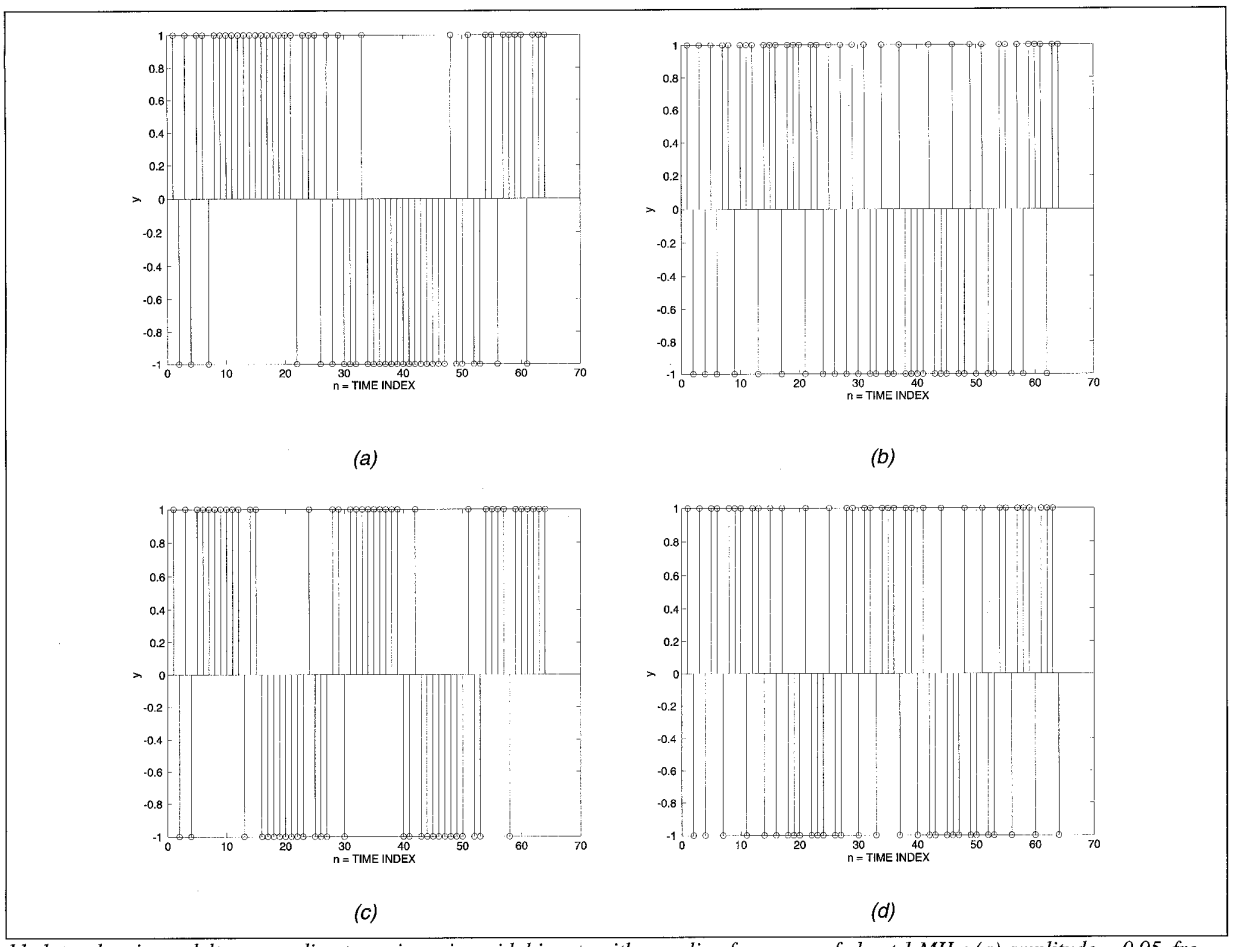
Let us now examine the DAC before going on to examine the performance of the converter. The DAC is required to be nearly as linear as the overall conversion resolution. Any D/A non-linearity can be modeled as an error source that adds directly to the input. This error source benefits from the



9. NTF for a first order sigma-delta modulator (a) magnitude spectrum on linear scale. For comparison, the oversampled PCM NTF, which has unity gain, is shown; (b) magnitude spectrum in dB.



10. 1st order sigma-delta responding to a DC input: (a) DC input x[n] = 0.55 = 11/20. (b) modulator output y[n]; (c) "error" signal u[n]; (d) integrator output v[n]



11. 1st order sigma-delta responding to various sinusoidal inputs with sampling frequency of about 1 MHz: (a) amplitude = 0.95, frequency= 20 kHz; (b) amplitude = 0.5, frequency = 20 kHz; (c) amplitude = 0.95, frequency = 40 kHz; (d) amplitude = 0.5, frequency = 40 kHz.

oversampling but unlike $e[n]$, which models the A/D quantization error, is not subject to the noise shaping.

Since a 1 bit DAC is perfectly linear, it is common to use a 1 bit DAC and a corresponding 1 bit quantizer, which is simply a comparator. Consequently, provided the sampling frequency is high enough, the sigma-delta A/D allows the use of a 1 bit quantizer to achieve high overall resolution. Using $H_x = z^{-1}$, $H_e = (1\text{-}z^{-1})$, and the procedure used for the oversampling PCM A/D, the in-band noise power (i.e., the noise in the frequency range $[-f_B, f_B]$) at the output of a first order sigma-delta modulator is

$$\sigma_{ey}^2 = \sigma_e^2 \frac{\pi^2}{3}\left(\frac{2f_B}{f_s}\right)^3 \tag{12}$$

The SNR in dB is then:

$$SNR = 10\log(\sigma_x^2) - 10\log(\sigma_e^2) - 10\log\left(\frac{\pi^2}{3}\right) + 30\log\left(\frac{f_s}{2f_B}\right)\ (dB) \tag{13}$$

Letting the oversampling ratio, $f_s/2f_B = 2^r$, we obtain:

$$SNR = 10\log\left(\sigma_x^2\right) - 10\log(\sigma_e^2) - 10\log\left(\frac{\pi^2}{3}\right) + 9.03r\ (dB) \tag{14}$$

For every doubling of the oversampling ratio i.e., for every increment in r, the SNR improves by 9 dB, or equivalently, the resolution improves by 1.5 bits.

Let us revisit the example considered at the end of the discussion on oversampled PCM. It was desired to convert a 20 kHz audio band to CD quality resolution of 16 bits. Using Eq. 13, we can compute that the required f_s with a 1 bit internal ADC is 96.78 MHz. A 1 bit ADC or comparator can operate at this speed in current CMOS technology. However, it is not possible for the sampled data analog switched capacitor integrators to operate at such high speeds. Second order sigma-delta modulation can allow the use of a 1 bit quantizer, allowing us to meet the 16 bit 20 kHz target with a much more reasonable f_s.

Qualitative Time Domain Behavior

The sigma-delta modulator can be thought of as a PCM converter with feedback, which attempts to force the output signal $y[n]$ to be equal to the input signal, $x[n]$. Consider the case where a 1 bit A/D converter or comparator is used. The transfer characteristics of this 1 bit ADC with output levels V and $-V$ is shown in Fig 2c.

Assume that $V = 1$, such that the comparator's digital output is 1 or -1, so that $y[n]$ and $y_a[n]$ can be used interchangeably. In the time domain, referring to Fig. 8, we have,

$$v[n] = u[n-1] + v[n-1] \tag{15}$$

$$y[n] = \begin{cases} 1 & v[n] \geq 0 \\ -1 & v[n] < 0 \end{cases} \tag{16}$$

$$u[n] = x[n] - y[n] \tag{17}$$

The "error" between the modulator output and input is $u[n]$. Note that this is not the quantization error, which is given by $e[n] = y[n] - v[n]$.

Since $y[n]$ can take on values of 1 or -1 only, it can never equal the input unless the input happens to be one of these two values exactly. Consequently, except for the mentioned cases, there will always be an error, $u[n] \neq 0$. Consider a DC input for $x[n]$. When $y[n] = 1$, $y[n]$ is greater than the input $x[n]$ and the error $u[n]$ is negative, and so negative values are accumulated by the integrator to produce $v[n]$. After a number of clock cycles, enough negative values will have accumulated to cause the quantizer to produce $y[n] = -1$, thereby changing the sign of the error $u[n]$ to be positive. The error between the output and input has been reduced, in some sense, because the positive errors will now cancel the prior negative errors when averaged over a period of time.

Now with $y[n]=-1$, the errors will be positive, and positive values of the error will be accumulated again until the quantizer output changes, this time back to $y[n] = 1$. Over a period of time, the proportion (or density) of 1's and -1's will be related to the DC input value—the larger the input, the more 1's will be present in the output, and vice versa, for smaller inputs. For this reason, the output of a sigma-delta modulator using a 1 bit quantizer is often said to be in pulse density modulated (PDM) format.

Let us now illustrate the time domain behavior using a few examples. Fig 10a shows a DC input $x[n] = 0.55 = 11/20$, while Fig 10b shows the corresponding modulator output, $y[n]$. Roughly, three fourths of the output values are 1's, and the others are -1's. Fig 10c shows the error signal, $u[n]$, and Fig 10d shows the accumulated error signal or integrator output $v[n]$, whose sign change forces the quantizer output to change. For a DC input of $x[n] = 1$, all the modulator output values will be 1's. For a zero DC input, half the modulator output values will be 1's, half -1's. For a DC input of -1, all the values will be -1's.

By averaging the modulator output over a period of time, we can approximate the input. This averaging operation represents the low-pass filter block in Fig. 8, since averaging is a crude low-pass filtering operation. Using a better low-pass filter will result in the modulator output being a better approximation to the input for a given oversampling ratio.

Finally, let us look at some time domain examples of the modulator output for sinusoidal inputs. Figures 11a-d show the modulator outputs for various sinusoidal inputs. As for the DC input case, the sinusoid amplitude information is encoded in the relative number of 1's vs -1's. The modulator output pulse pattern has periodic components, and the fundamental period encodes the sinusoid frequency. This is particularly clear in Figs. 11a-c.

Implementation Imperfections

The results presented thus far have not considered imperfections in the analog hardware. Let us now discuss the consequences of imperfections in some of the main circuit parameters.

The integrator in the modulator may have a gain of g instead of unity, and may be leaky. For an input $u[n]$, an integrator with gain g and leakage factor α has an output $v[n] = g\,u[n-1] + \alpha v[n-1]$ instead of $v[n] = u[n-1] + v[n-1]$, and the integrator transfer function is $g\,z^{-1}/1-\alpha z^{-1}$ instead of $z^{-1}/1-z^{-1}$.

The D/A gain may also not be perfectly unity, and assuming a gain of d, we find the STF and NTF are $H_x = g\,z^{-1}/(1+(g\,d-\alpha)z^{-1})$ and $H_e = 1-\alpha z^{-1}/(1+(g\,d-\alpha)z^{-1})$.

The original NTF $(1-z^{-1})$, which had a Z domain zero at $z = 1$ (on the unit circle and at DC), now has a zero which is still at DC but is moved inside the unit circle. This degrades the NTF noise attenuation in the signal band and can thus affect the noise performance significantly. The term "leaky" comes from the fact that there is charge leakage in the switched capacitor implementation of the integrator circuit and only a portion of the charge from the input capacitor is transferred to the integrating capacitor.

The leakage factor α is related to the open loop gain, A, of the operational amplifier (opamp) used to implement the switched capacitor integrator such that $1 - \alpha \simeq 1/A$. Ignoring the denominator (which will add a slight ripple to the numerator) of the degraded H_e, i.e., by considering the degraded NTF to be $(1-\alpha z^{-1})$, we find the in-band quantization noise power with integrator leakage is:

$$\sigma_{ey}^2 = \sigma_e^2\left(\frac{2f_B}{f_s}\right)\frac{1}{A^2} + \alpha\sigma_e^2\frac{\pi^2}{3}\left(\frac{2f_B}{f_s}\right)^3 \tag{18}$$

instead of Eq. 12. Note that the noise power now contains a term that is inversely proportional to the oversampling ratio, $f_s/2f_B$, as well as a term inversely proportional to the oversampling ratio cubed. However, the first term is divided by A^2, and it has been found that if the opamp open loop gain, A, exceeds the oversampling ratio, $f_s/2f_B$, leakage causes no significant degradation of the SNR [2]. Consequently, the

circuit constraint required to implement good integrators is not that difficult to meet unless the oversampling ratio is extremely high.

From the linearized analysis, the STF and NTF pole is stable only for $0< gd <2$. Consequently, there is a relatively wide margin over which the two gains may vary from this point of view. It has been reported [2] that variations of g by up to 10% from unity does not degrade the SNR significantly. The gain is implemented in practice as a ratio of two capacitors and so the corresponding precision on the capacitor matching is minimal—one part in 10, or three and a quarter bits.

Now consider an imperfect DAC gain, d, that is slightly different from unity. This can be modeled as a gain of $1/d$ at the input of the modulator. To see this, consider that a gain of $1/d$ is inserted at the input of the integrator. This gain can be moved past the summing node at the modulator input in Fig 8. The result is that the DAC gain d is cancelled by the gain $1/d$ but the input now experiences a gain of $1/d$ before the modulator. Consequently, the STF experiences a slight gain change but there is no great impact on the modulator SNR.

Finally, consider imperfections in the quantizer. Any non-linearity in the quantizer can be modeled as another noise source which adds to $e[n]$, the quantization error. However, the noise from this extra source is subject to noise shaping by the modulator and so its affect on SNR degradation is not significant.

If the 1 bit quantizer or comparator has a non-zero threshold, v_{th}, its output is given by:

$$y[n]=\begin{cases}1 & v[n]-v_{th}\geq 0\\ -1 & v[n]-v_{th}<0\end{cases}$$

This is equivalent to an offset at the input of the comparator, i.e., at the output of the integrator. However, an offset v_{th} at the output $v[n]$ of the integrator corresponds to an impulse at its input $u[n] = x[n]- y[n]$ which amounts to one incorrect output $y[n]$ for a given $x[n]$. One such incorrect value will have a negligible impact on the overall performance of the modulator. The offset can also be modeled as an error source, with mean v_{th} and zero variance, that adds to $e[n]$.

If for any reason, the offset is input dependent or changes with time, this new error source will have non-zero variance. However, it will be subject to the noise shaping property of the modulator just as $e[n]$ and its presence will not degrade the performance of the modulator significantly.

Non-linear Behavior

A sigma-delta modulator is a non-linear system incorporating feedback. Not surprisingly, the modulator may display limit cycle oscillations that result in the presence of periodic (tone) components in the output. This phenomenon is analogous to limit cycles that occur in digital IIR filters operating with finite precision arithmetic, because, like a sigma-delta modulator, such a filter is a non-linear system that employs feedback. The quantizer error spectrum is not white, which is not surprising as the conditions for the white noise assumption are not perfectly satisfied—the quantizer has only two output levels, and due to the oversampling, successive quantizer input samples may be correlated.

Now consider the existence of limit cycles in the modulator, as has been done in [9] for the simple case of a DC input, $x[n] = x$. For a limit cycle of period T, $v[n]$ should be periodic with period T, i.e., $v[n] = v[n+T]$. This clearly implies that $y[n] = y[n+T]$.

For the DC input, the input to the integrator, $u[n] = x - y[n]$, will likewise be periodic with period T. Thus, the modulator behavior can be represented by T equations.

Combining Eqs. 15 and 17, we obtain: $v[n] - v[n-1] = x[n-1] - y[n-1]$, which for a DC input becomes $v[n] - v[n-1] = x - y[n-1]$. Writing this equation for T different time instances starting (arbitrarily) at $n = 1$, and adding up all these equations we obtain,

$$v[T]-v[0]=\sum_{l=0}^{T-1}x-\sum_{l=0}^{T-1}y[l] \tag{19}$$

However, $v[T]=v[0]$ by assumption, and consequently:

$$x=\frac{1}{T}\sum_{l=0}^{T-1}y[l]=\frac{(P-M)}{T}V \tag{20}$$

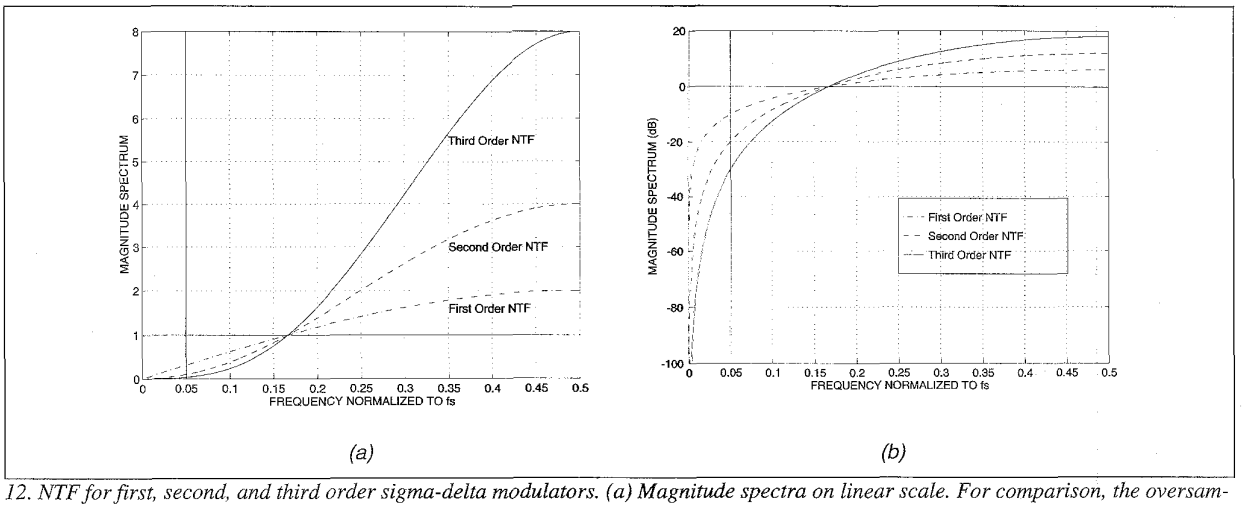
where P is the number of positive quantizer outputs over T output samples, and M is the number of negative quantizer outputs over T output samples. Since $(P - M)$ is an integer as is T by assumption, we have $x = bV/a$, with a and b integers. Thus, the output y consists of a limit cycle with period T, provided that x is a rational multiple of V.

The limit cycle with period T will manifest itself in the output spectrum as tones present at frequency f_s/T and its harmonics. The period is $T = 2a$ if a or b is even. $T = a$ if both a and b are odd [11]. For the prior example of $x = 11/20$ in Fig 10, $a = 20$ and $b = 11$, and as can be seen in the figure, $u[n]$, $v[n]$, and $y[n]$ do have the expected period of 40.

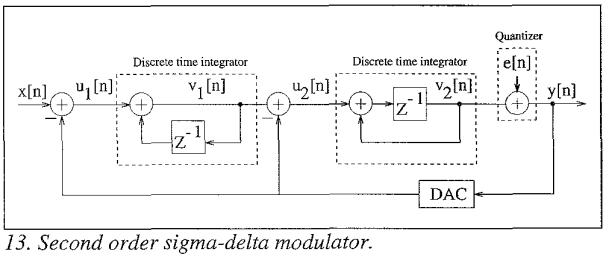
For the special case of $x = 0$, the output oscillates between V and $-V$ and the output spectrum consists of a pure tone at $f_s/2$. More complete results, which are independent of the integrator initial condition, $v[0]$, are provided in [9, 10, 12]. In fact, it is shown in [10] and [12] that even if the DC input is an irrational multiple of V, the quantization noise will not be white and the spectrum at the output of the modulator will be discrete, consisting of tones.

Even for sinusoidal inputs, the quantization error is not white and strong tone components are observed in the output and the strength of the tone distortion components depends on the input amplitudes in a complicated way [14]. The tone structure present at the output of the modulator for very low DC or sinusoidal input amplitudes is often called idle-channel noise.

One other point is worth noting. It can be shown that for comparator output levels of $\pm V$, the output of the integrator can have magnitude of at most $2V$ if the input to the modulator is bounded by $\pm V$ [14]. This is easily seen from Eqs. 15 and



12. NTF for first, second, and third order sigma-delta modulators. (a) Magnitude spectra on linear scale. For comparison, the oversampled PCM NTF, which has unity gain, is shown; (b) magnitude spectra in dB.



13. Second order sigma-delta modulator.

17, from which we have $v[n] = x[n\text{-}1] - (y[n\text{-}1] - v[n\text{-}1])$, which is $v[n] = x[n\text{-}1] - e[n\text{-}1]$. If we assume $|v[n\text{-}1]| \leq 2V$, then from the transfer characteristic of the 1 bit quantizer (Fig. 2c), $|e[n\text{-}1]| \leq V$, i.e., the quantizer is not overloaded. We then have, $|v[n]| = |x[n\text{-}1] - e[n\text{-}1]| \leq |x[n\text{-}1]| + |e[n\text{-}1]| \leq V + V = 2V$. Thus if $|x| \leq V$ and $|v[n\text{-}1]| \leq 2V$, then $|v[n]| \leq 2V$. This can be guaranteed by ensuring that $|v[0]| \leq 2V$, so that $|v[1]| \leq 2V$, and so on.

In practice, because of the significant tone structure present at the output of a first order sigma-delta ADC, it is rarely used in applications such as speech or audio, where the presence of such tones is objectionable even if the oversampling ratio, $f_s/2f_B$, is high enough to provide a good overall SNR based on the linearized white noise model.

Higher Order Sigma-Delta Modulation

The fundamental ideas presented can be extended to create sigma-delta architectures in a variety of ways that provide different tradeoffs among resolution, bandwidth, circuit complexity, and modulator stability. Our discussion will include higher order, multi-bit, and multi-stage (cascaded) architectures. In general, to obtain a performance improvement, most of these converters require analog circuits that need to be more complex and precise than those used in the 1st order sigma-delta modulator. Of course, the precision required must still be significantly less than the overall high conversion resolution.

Second Order Modulation

Operation and Performance Modeling

The standard 2nd order sigma-delta modulator A/D is widely used. This modulator realizes $H_x(z) = z^{-1}$ and $H_e(z) = (1\text{-}z^{-1})^2$, so that

$$Y(z) = X(z)z^{-1} + E(z)(1-z^{-1})^2 \tag{21}$$

The second order modulator noise transfer function (NTF) is shown in Fig 12 along with the NTFs for the first order modulator and the third order modulator (a logical extension of the second order modulator which will be discussed later).

Note that, compared with the first order sigma-delta NTF, the second order NTF provides more quantization noise suppression over the low frequency signal band, and more amplification of the noise outside the signal band. Compared with a first order sigma-delta, more noise power is pushed outside the signal band.

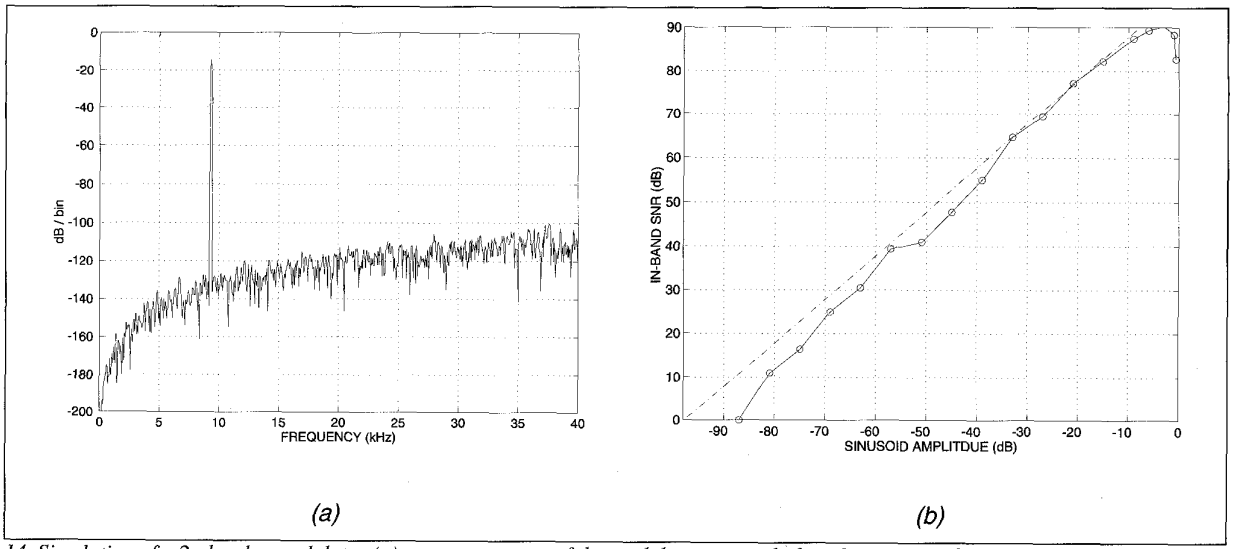
A block diagram of the modulator is shown in Fig. 13. The structure now contains two integrators. The transfer function of the first one is $1/1\text{-}z^{-1}$ and that of the second one is $z^{-1}/1\text{-}z^{-1}$. Assuming the modulator output is filtered by an ideal lowpass filter, the linearized white noise model yields the following for the in-band SNR:

$$SNR = 10\log(\sigma_x^2) - 10\log(\sigma_e^2) - 10\log\left(\frac{\pi^4}{5}\right) + 50\log\left(\frac{f_s}{2f_B}\right) \quad (dB) \tag{22}$$

Again, letting $f_s/2f_B = 2^r$, we obtain:

$$SNR = 10\log(\sigma_x^2) - 10\log(\sigma_e^2) - 10\log\left(\frac{\pi^4}{5}\right) + 15.05r \quad (dB) \tag{23}$$

Thus, for every increment in r or for every doubling of the oversampling ratio, $f_s/2f_B$, the SNR improves by 15 dB, or the



14. Simulation of a 2nd order modulator (a) power spectrum of the modulator output before decimation; (b) in-band SNR vs amplitude of input sinusoid.

equivalent resolution by 2.5 bits, which is 1 bit better than the improvement achieved by a first order sigma-delta.

Now consider the example used previously where a 20 kHz audio band needs to be converted to a resolution of 16 bits. The f_s needed by a 2nd order sigma-delta modulator using only a 1 bit quantizer is, from Eq. 22, only 6.12 MHz in contrast with the 96 MHz needed by the 1st order sigma-delta. This circuit speed is very reasonable in current CMOS technology.

Fig 14a shows the low frequency portion (0 to 40 kHz) of an FFT based power spectral density estimate of the output of a 2nd order modulator with a sinusoidal input frequency of 9.3 kHz and f_s of 6.20 MHz. The tall peak is, of course, the sinusoidal signal. Notice the noise shaping whereby the noise at lower frequencies is greatly attenuated. Such power spectra are often used to numerically calculate the in-band SNR. Here, the in-band SNRs are computed over the 20 kHz audio band. The SNR computation can be repeated for various sinusoid amplitudes to obtain a plot of in-band SNR vs amplitudes. This plot is shown in Fig 14b.

The dashed curve corresponds to the values which are obtained from the linearized white noise model SNR formula (Eq. 22). The discrepancy between the two plots will be discussed later. The dynamic range, R, can be determined from the in-band SNR vs amplitude plot by looking at the amplitude value for which the SNR is 0 dB. The peak SNR is about 90 dB and R is about 88 dB which, from Eq. 5, yields an equivalent resolution, N, of 14.32 bits.

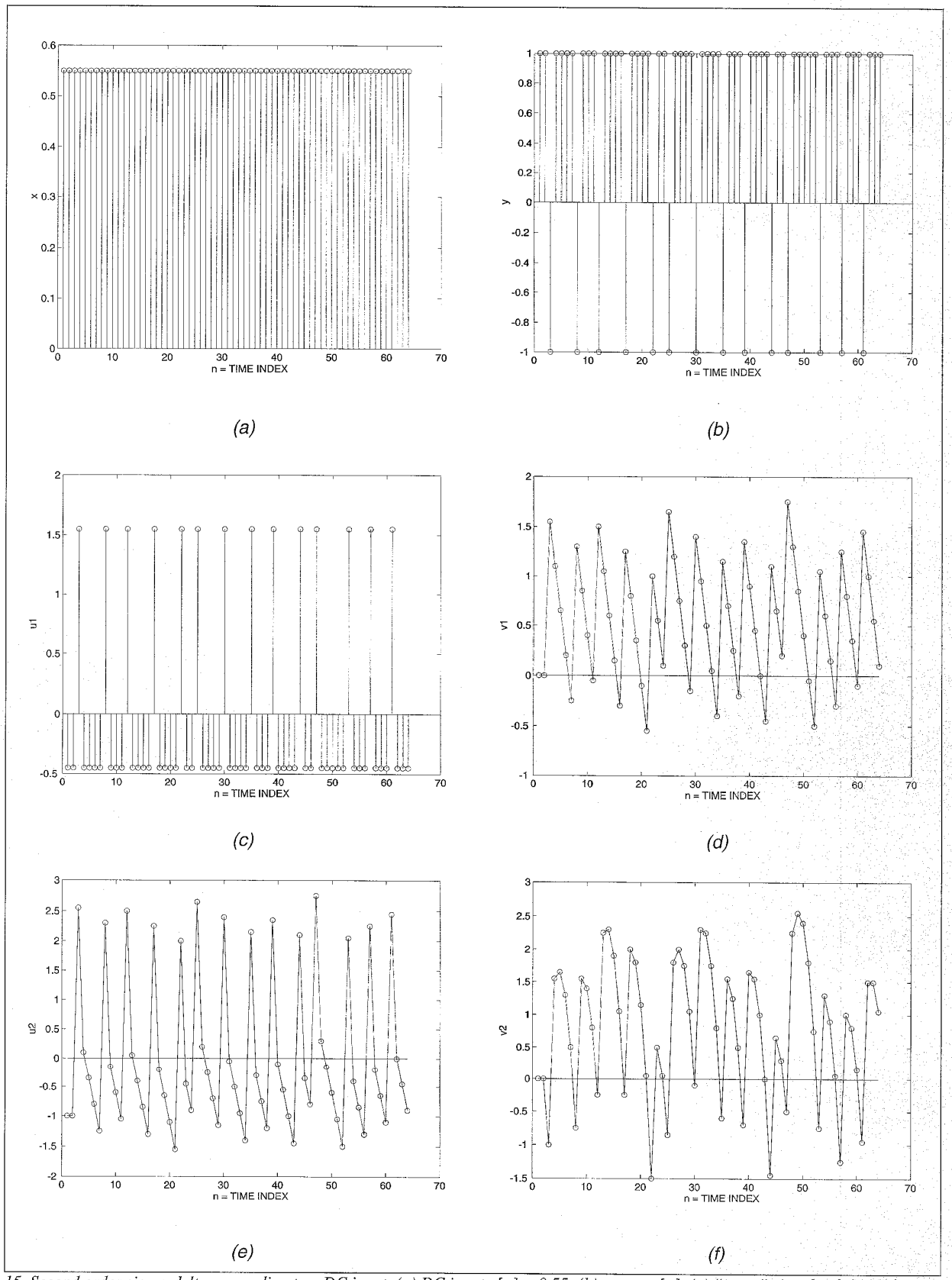
Qualitative Time Domain Behavior

In the time domain, referring to Fig. 13, we have,

$$u_2[n] = v_1[n] - y[n]$$
$$v_2[n] = u_2[n-1] + v_2[n-1]$$
$$y[n] = \begin{cases} 1 & v_2[n] \geq 0 \\ -1 & v_2[n] < 0 \end{cases}$$
$$u_1[n] = x[n] - y[n]$$
$$v_1[n] = u_1[n] + v_1[n-1]$$

Figures 15a and b show plots of a DC input, $x[n] = 0.55$, and the resulting output, $y[n]$, of a second order sigma-delta modulator. The input and output of the first integrator are $u_1[n]$, and $v_1[n]$, respectively, while the input and output of the second integrator are $u_2[n]$, and $v_2[n]$, respectively. The "error" between the modulator input and output is $u_1[n]$, which again is not the quantization error (given by $e[n] = y[n] - v_2[n]$). Looking at $u_2[n]$, we see that it is the difference between an integrated (or low-pass filtered) version, $v_1[n]$, of the modulator "error," $u_1[n]$, and the output, $y[n]$. Thus, $u_2[n]$ can be considered to be a more fine or accurate version of the modulator error. The signal which is quantized, $v_2[n]$, is an integrated version of the "fine error," $u_2[n]$. Consequently, $u_1[n]$, and $v1[n]$ are analogous to the $u[n]$ and $v[n]$ of the 1st order modulator. Outputs $u_2[n]$ and $v_2[n]$ are more accurate representations of $u_1[n]$ and $v_1[n]$, and thus produce an output $y[n]$, which is more accurate than the output of a 1st order modulator. This should be clear from comparing Fig 15c with Fig 15e and Fig 15d with Fig 15f.

In comparing $y[n]$ of the second order sigma-delta of Fig. 15b to the $y[n]$ of the first order sigma-delta of Fig. 10b, the key point is that the distribution of 1's and -1's in Fig. 15b is such their average provides a more accurate representation of the input than the corresponding average of the first order modulator output. In other words, for a given block of output samples, the second order modulator uses its allocation of samples more efficiently to represent the input.



15. Second order sigma-delta responding to a DC input: (a) DC input x[n] = 0.55; (b) output y[n]; (c) "error" signal $u_1[n]$; (d) first integrator output $v_1[n]$; (e) more accurate "error" signal u2[n]; (f) second integrator output $v_2[n]$.

Implementation Imperfections

Compared with a first order sigma-delta modulator, a second order modulator contains an extra integrator. Assuming the same leakage factor, α , for both integrators, the degraded NTF is approximately $(1- \alpha z^{-1})^2$. The leakage factor of the second order integrators can satisfy somewhat less stringent requirements than that of the first order modulator integrator.

One might expect this since despite the NTF zeros being moved inside the unit circle, there are still two of them providing attenuation of quantization noise. The reduced requirement on the second order modulator integrator leakage factors can also be seen if one calculates the in-band noise power, σ_{ey}^2, from the $(1-\alpha z^{-1})^2$ NTF.

The noise term inversely proportional to the oversampling ratio, fs/2fB, is now divided by A4 rather than A2, as was the case for the first order modulator in Eq. 18. Consequently, a lower op-amp gain can suffice for the second order modulator.

The other parameter to consider is the gain of the second integrator. From the linearized analysis, the STF and NTF poles are stable for integrator gains up to 4/3, and so still allow a relatively wide variation from this point of view.

Simulations indeed show that the integrator gains are relatively insensitive to deviations from their nominal values over a wide range of oversampling ratios [8].

Non-linear Behavior

Like the 1st order sigma-delta modulator, a second order modulator may also display limit cycle oscillations [2, 9], and this is easily illustrated in a manner similar to that used for the 1st order modulator. The nature of these limit cycles has been investigated and unlike the first order modulator depend on the initial conditions of the integrator outputs [9, 11].

Most of the exact analyses, e.g., [13-15], which provide an exact description of the spectrum of the quantization error and the modulator output for DC or sinusoidal inputs, have been performed for second order modulators using quantizers with 2 or more bits. In fact, a 1 bit quantizer used in a 2nd order modulator can become overloaded, thereby making the analysis much more difficult. The quantizer is overloaded because the output of the second integrator can significantly exceed values of $\pm 2V$ [2] even with the modulator input bounded by the quantization levels $\pm V$. This is particularly true for large modulator inputs near the quantization levels. However, it has been determined from simulations that a modified 2nd order architecture [5] using a 1 bit quantizer can operate without the integrator outputs having to significantly exceed values of $\pm 2V$.

The spectral properties of overload noise have not been theoretically characterized but simulations have been reported in the literature [16, 17]. The simulations demonstrate that the noise significantly manifests itself as harmonic distortion tones for sinusoidal inputs as well as significant tone components near $f_s/2$.

As with the first order sigma-delta, idle channel tones may be observed for small DC or low amplitude tone inputs. According to the linearized model (Eq. 22), σ_e^2 is fixed and the SNR should increase linearly with signal power σ_x^2. However, due to the presence of overload or idle channel tones, the SNR of a second order modulator, using a 1 bit quantizer, increases linearly with signal power only over a certain range of signal power even though the modulator input may be between the quantization levels $\pm V$. More and more overload noise power is produced with increasing input values. Consequently, above a certain value of high signal power, the SNR will actually start to decrease when the increase in overload noise power dominates the increase in signal power.

This can be seen in Fig 14b for input amplitudes greater than about -5 dB. On the other hand, as the input power becomes low, the SNR decrease is caused both by the decrease in signal power and by the presence of idle channel tone noise in the signal band. This can be clearly be seen in Fig. 14 for low amplitude values.

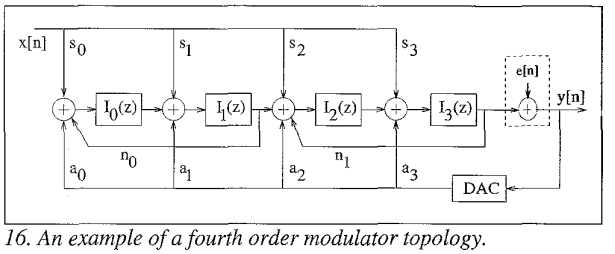
According to Eq. 22 based on the linearized white noise model, the peak SNR and dynamic range should have been about 98 dB. Thus, idle channel tone problems, resulted in a 10 dB degradation of the dynamic range predicted by the linearized model while overload noise prevented the modulator from reaching the peak SNR predicted by the model. Dithering techniques [17] will often break up tone structures including overload and idle channel tones, thereby producing a smoother power spectrum output, a more linear dependence of SNR on signal power, and a dynamic range, which is closer to the value predicted by Eq. 22.

Another factor which affects the in-band SNR is the sinusoid input frequency. This is particularly true if the sinusoid amplitude is such that strong harmonic distortion components are produced in the power spectrum output. In this case, the choice of a higher input frequency will result in these harmonics falling outside the 20 kHz bandwidth and not contributing to the in-band SNR. On the other hand, the choice of a lower input frequency will yield poorer values of in-band SNR because the harmonics will now fall inside the 20 kHz signal band.

Other Types of Higher Order Sigma-Delta Modulation

Sigma-delta converters realizing higher order NTFs achieve even higher resolution by pushing even more noise power outside the signal band. Alternatively, a lower sampling rate can be used to achieve the same resolution for a given signal bandwidth. In this case, the speed requirements on the analog hardware is relaxed.

An order L modulator based on a straightforward extension of the first order sigma-delta realizes a STF given by $H_x = z^{-1}$ and a NTF given by $H_e = (1-z^{-1})^L$, which contains L zeros at $z = 1$ or at DC frequency on the unit circle. A third order modulator structure can be created from the second order structure of Fig. 13 by inserting an integrator with transfer function $1/(1-z^{-1})$ between the the summing node of the modulator input and the first integrator of the second order modulator. The input to this new integrator is now $x[n]$-$y[n]$ and the output of this integrator minus $y[n]$ is fed as the input to



16. An example of a fourth order modulator topology.

what used to be the first integrator of the second order modulator.

The magnitude spectra for a third ($L=3$) order NTF plotted on a linear scale and in dB are shown in Figs. 12a and b. Note that over the signal band, which is $0.05 f_s$ in the figure, the 3rd order NTF provides more attenuation of the quantization noise than the second or first order NTFs and so is capable of pushing more noise power outside the signal band than the second or first order modulators. The ideal in-band SNR achieved by an Lth order modulator is given by

$$SNR = 10\log(\sigma_x^2) - 10\log(\sigma_e^2) - 10\log\left(\frac{\pi^{2L}}{2L+1}\right)$$
$$+(20L+10)\log\left(\frac{f_s}{2f_B}\right) \quad (dB) \tag{24}$$

Thus, for every doubling of the oversampling ratio, this modulator provides an extra $(6L + 3)$ dB of SNR, or an extra $(L+ 1/2)$ bits of resolution. For the CD example used throughout this article, a 3rd order modulator will need a f_s of 1.92 MHz to convert a 20 kHz band to 16 bits of resolution.

Modulator topologies are not constrained to always realize FIR NTFs as we have seen until now or to realize all the NTF zeros which provide attenuation at DC frequency. In fact, distributing the zeros over the signal band rather than placing them all at DC frequency can be more efficient in pushing quantization noise outside the signal band.

Examples of higher order topologies are described in [18-21]. One such fourth order topology described in [19] is shown in Fig 16, where $I_k(z)$ denotes the kth integrator. This structure realizes Eq. 7 in the form

$$Y(z) = X(z)\frac{B_x(z)}{A(z)} + E(z)\frac{B_e(z)}{A(z)} \tag{25}$$

Note that the STF and NTF are IIR transfer functions in this case. The feedforward coefficients s_k realize B_x, which contains the Z domain zeros of the STF. The feedback coefficients n_k between every second integrator realize B_e, which contains the zeros of the NTF. Finally, the feedback coefficients a_k realize A, which contains the poles of both the STF and the NTF.

In order to implement Eq. 25, based on a sigma-delta modulator that incorporates feedback, the z^0 coefficient in B_e and A must be equal to ensure a causal feedback loop (in other words to avoid non-computable delays, [3, pp. 308-309]). If the NTF is FIR, i.e., if $A(z) = 1$, then the causality constraint will require the z^0 coefficient of $B(z)$ to be 1.

Note that the performance will be limited by the degree to which the analog coefficients a_k, s_k, and n_k, match their desired values. For the modulator to be useful, the degree of required matching should be significantly less than the overall resolution of the converter.

Higher order architectures also alleviate some of the tone problems mentioned earlier [2, 18]. The main difficulty with such higher order structures is that such modulators are only conditionally stable when an one bit quantizer is used. Stability may for example depend on the input signal being kept below a certain value or on precise circuit matching needed to satisfy a stability criterion. Stability is often described in the sense of the quantizer not being overloaded. This is useful because any higher order modulator structure can be transformed into an equivalent modulator in the form of the 1st order modulator of Fig 8 with the integrator being replaced by a general "loop filter" $H(z)$. If the quantizer is not overloaded, its input is bounded by $\pm 2V$, and this implies that the loop filter is also operating in a stable manner and all internal signals will be bounded. Stability is, in general, difficult to determine for a modulator using an 1 bit quantizer. One reason is the difficulty in characterizing the gain of the 1 bit quantizer. The gain of the 1 bit quantizer of Fig. 2c is variable—it depends on the quantizer input. The smaller the quantizer input, the larger is its gain (e.g., if the input is zero, the output is V and the gain is infinite). The larger the quantizer input, the smaller is its gain (e.g., if the input is infinite, the output is V and the gain is zero). Another way of looking at this is to observe that if one attempts to linearize the quantizer transfer curve of Fig. 2c by trying to fit a straight line to the curve, the correct slope of the straight line is arbitrary.

As we have seen earlier, the input to the quantizer clearly changes with time, even if the input to the modulator is a DC signal. The manner in which the quantizer gain changes over time will also depend on the type of input applied to the modulator. Consequently, performing a linear system analysis of the modulator signal and noise transfer functions in terms of modulator parameters is inadequate because the poles of the transfer function, which determine stability, depend on a time varying and input dependent quantizer gain. An attempt to characterize the quantizer gain more accurately for DC and sinusoidal inputs has been made for several modulators [22].

The phenomenon of limit cycle oscillations is also connected to stability. This is because the structure of limit cycles may be such that the amplitude of internal modulator variables is large, causing the quantizer to overload. The frequency of such a deleterious limit cycle oscillation can correspond to the point on the unit circle where the modulator transfer function pole crosses the unit circle into the unstable region. Properties of limit cycles in the context of stability have been investigated in [11]. A limit cycle that corresponds to the modulator transfer function pole moving from the inside to the outside of the unit circle may not necessarily result in unstable behavior in the long term, provided integrator outputs do not saturate before stability is restored [23].

Suppose an unstable limit cycle, corresponding to poles moving outside the unit circle, results when the quantizer gain is too high, i.e., when the input to the quantizer is small. If this is the case, growing signal values in the modulator which result from this instability will eventually increase the input to the quantizer thereby reducing its gain and so moving the poles back inside the unit circle [23].

One way to guarantee stability would be to reset the integrators if it was detected, by additional circuitry, that their values were becoming too large. However, this approach may cause a significant decrease in the SNR [24]. Similarly, allowing integrator outputs to clip or saturate may also cause degradations in the SNR performance. In particular, low frequency limit cycles which may introduce distortion components in the signal band may occur (albeit with reduced values of integrator output) and may persist for a long time thereby reducing the SNR [23]. An alternative approach uses local feedback loops in an attempt to gracefully return integrator outputs to their normal operating region [24]. Since the linearized transfer function of the system is modified due to the local feedback loops, the effect of these loops is cancelled in the digital domain.

Several quantitative criteria have also been proposed to characterize stability. The l_1 norm criterion [14, 25, 26] relates the sum of the magnitudes of the modulator NTF impulse response coefficients, the number of bits in the quantizer, and the modulator input level to the no overload stability requirement. This is a sufficient but not necessary condition for stability, and it has unfortunately been found to be too conservative for practical use. An ad-hoc stability criterion which has been proposed [18] and found to be useful [21] is to design the NTF to possess less than 2 to 6 dB of out-of-band gain.

Multi-bit Sigma-Delta Modulation

Until now, we have assumed that the quantizer and DAC inside our sigma-delta modulator were 1 bit devices. However, converters using a multi-bit internal quantizer offer more potential resolution from the internal quantizer. A 2nd order multi-bit sigma-delta converter would look exactly the same as the modulator shown in Fig. 13, except that $e[n]$ in the figure would be the model for an N bit quantizer instead of a 1 bit quantizer, and the DAC would be an N bit DAC instead of a 1 bit DAC. The use of a multi-bit quantizer affects the σ_e^2 term in the expressions for the SNR, where each additional bit used in the quantizer will yield a 6 dB improvement in the SNR. Using Eq. 3 without the approximation, it is easy to see that if a 5 bit internal quantizer is used instead of a 1 bit quantizer, a 30 dB improvement in SNR is possible. Alternatively, the sampling frequency can be reduced by a factor of 4, while keeping the resolution the same. For our CD example, a 2nd order modulator using a 5 bit internal quantizer can use a f_s of 1.53 MHz, rather than the f_s of 6.12 MHz needed by a modulator using a 1 bit quantizer.

The behavior of multi-bit sigma-delta systems more closely follow that predicted by the linearized model (in the extreme case, if the quantizer has an infinite number of bits, there is no non-linearity). Consequently, the stability of higher order modulators using multi-bit quantizers is generally more accurately predicted. Another way of viewing enhanced stability is to consider the gain of the multi-bit quantizer. If a midtread multi-bit quantizer is used, its gain is relatively close to one for most output values (even for zero input, a midtread quantizer will have a zero output and hence unity gain unlike the midriser 1 bit quantizer, which has infinite gain for a zero input). This is because a straight line drawn through the multi-bit quantizer transfer characteristic (e.g., Fig. 2b) can no longer be arbitrarily drawn. Of course, if for any reason the quantizer does start to overload, its gain will start to deviate more and more from unity. Even though the l_1 stability criterion mentioned earlier may be too conservative, it may allow one to obtain an initial idea about the number of bits needed for stable operation for a given NTF and input signal level.

Modulators using multi-bit quantizers also display less of the tone problems associated with the 1st and 2nd order sigma-delta converters using a 1 bit internal quantizer. The main disadvantage is that the multi-bit DAC cannot be easily fabricated in VLSI with sufficient linearity needed for high resolution conversion. Various techniques, examples of which can be found in [27-33], have been proposed to reduce the linearity required for the DAC. The multi-bit output also complicates the digital low-pass filter hardware following the modulator, because for multi-bit processing, the filter requires multi-bit hardware multipliers.

Multi-stage (Cascaded) Sigma-Delta Modulation

Higher order NTFs can also be created by cascading independent modulator stages. This cascading does not adversely affect the stability of the overall modulator, provided the individual stages are stable. They may also suffer fewer of the tone problems than a first or second order sigma-delta alone [2, 14].

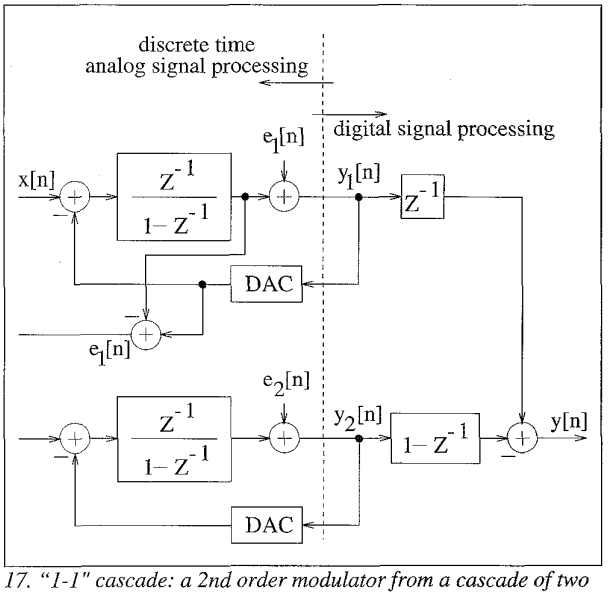
An example of a second order modulator obtained by cascading two 1st order modulators is shown in Fig 17. The signal $x[n]$ is the input to the first modulator in the cascade but the quantization error, $e_1[n]$, of the first modulator is used as the input to the second modulator. Finally, the outputs of the first stage and second stage modulators are added in the digital domain after passing through a digital delay (z^{-1}) and a digital differentiator ($1\text{-}z^{-1}$), respectively.

From the Z domain analysis of the linear system model with the DACs replaced by unity gains, we have,

$$Y_1(z) = X(z)z^{-1} + E_1(z)(1 - z^{-1}) \tag{26}$$

$$Y_2(z) = E_1(z)z^{-1} + E_2(z)(1 - z^{-1}) \tag{27}$$

The output is computed as $Y_1\ z^{-1}$ - $Y_2(1\text{-}z^{-1})$. This sum results in a cancellation of the first order noise term $E_1(z)$ to produce the overall output,



17. "1-1" cascade: a 2nd order modulator from a cascade of two 1st order modulators.

$$Y(z) = X(z)z^{-2} - E_2(z)(1-z^{-1})^2 \tag{28}$$

Except for the sign on the noise that is irrelevant and an extra delay experienced by the input, the modulator realizes the same output as the standard second order sigma-delta modulator. One advantage of using this structure over the second order modulator is the fact that the quantizer in either of the first order modulator sections will never overload for $x[n]$ bound by $\pm V$. However, the cascaded structure requires matching between the analog and digital transfer functions as well matching among the D/A output levels among various stages [2, 34]. In fact, mismatch effects and integrator leakage can lead to the propagation of unshaped or poorly shaped noise from an earlier section to the final output [36, 40].

Assume there are circuit imperfections in the "1-1" cascade of Fig. 17, such that the transfer function of the integrator in the first section is $g\,z^{-1}/(1-\alpha z^{-1})$ instead of $z^{-1}/(1-z^{-1})$. Even if we assume the integrator in the second stage is ideal, the output is then

$$Y(z) = \frac{X(z)gz^{-2}}{1-(g-\alpha)z^{-1}} + \frac{(1-\alpha)z^{-2}E_1(z)}{1-(g-\alpha)z^{-1}} + \frac{(g-\alpha)z^{-2}E_1(z)(1-z^{-1})}{1-(g-\alpha)z^{-1}} - E_2(z)(1-z^{-1})^2 \tag{29}$$

The first term of the equation contains the signal, which is no longer a pure delay but will have a ripple to it determined by the factor $g/(1-(g-\alpha)z^{-1})$. However, since the signal is oversampled, this additional signal transfer function will mostly be flat at lower baseband frequencies. Ignoring the ripple due to the factor $1/(1-(g-\alpha)z^{-1})$, the second term in Eq. 29 is the unshaped noise from the first stage, the third term in Eq. 29 is the first order shaped noise from the first stage, and the fourth term in Eq. 29 is the desired second order shaped noise term. Clearly, for large imperfections, the unshaped noise term might dominate the noise term subject to second order noise shaping. If more than two stages are cascaded, the cumulative effects of such quantization error leakage effects will yield diminishing returns in performance improvement.

Architectures using only first order modulators have been realized [35], as have architectures using second order modulators [37-39]. A comparison of some architectures can also be found in [40]. Finally, note that due to the addition of various single bit intermediate outputs, the architecture has a multi-bit final output, which complicates the decimation filter hardware.

Band-pass Sigma-Delta Systems

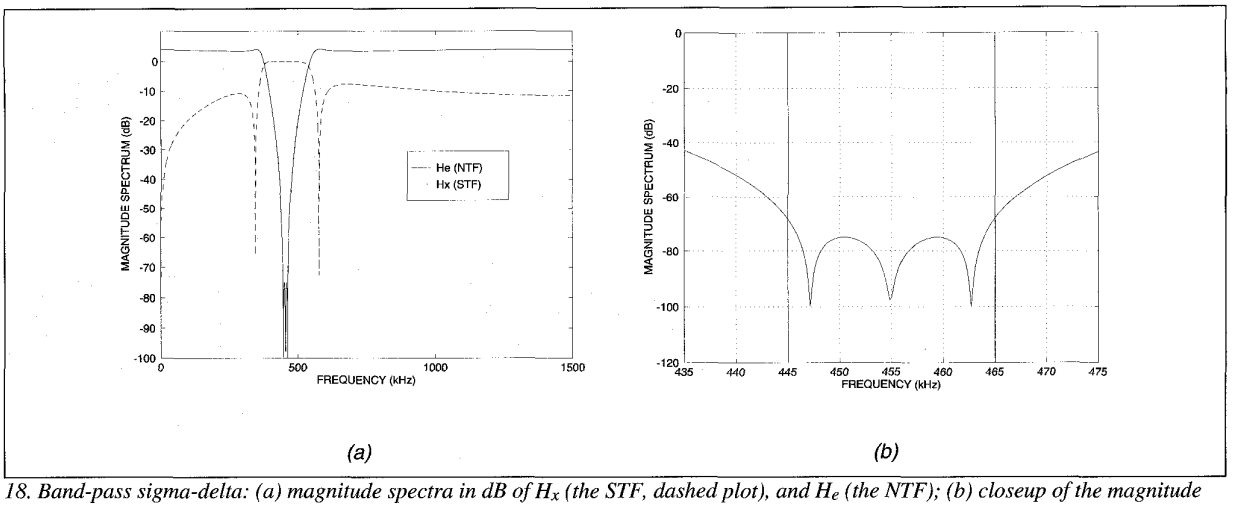
Thus far, we have assumed that the sampling frequency f_s is much greater than the Nyquist rate, which is twice the highest frequency component in the input signal. For low-pass signals, the highest frequency component is also the signal bandwidth f_B. If a signal with bandwidth f_B is band-pass but is located at a center frequency, f_c, its highest frequency is $f_c + f_B/2$. If f_c is large, choosing f_s to be much greater than the highest frequency will lead to an unreasonably large f_s, and does not take advantage of the band-pass nature of the signal.

Band-pass sigma-delta modulation [41] allows high resolution conversion of band-pass signals if f_s is much greater than the signal bandwidth f_B, rather than the highest signal frequency. Band-pass sigma-delta modulators can be used in AM digital radios [41] or receivers for digital cellular mobile radios [45].

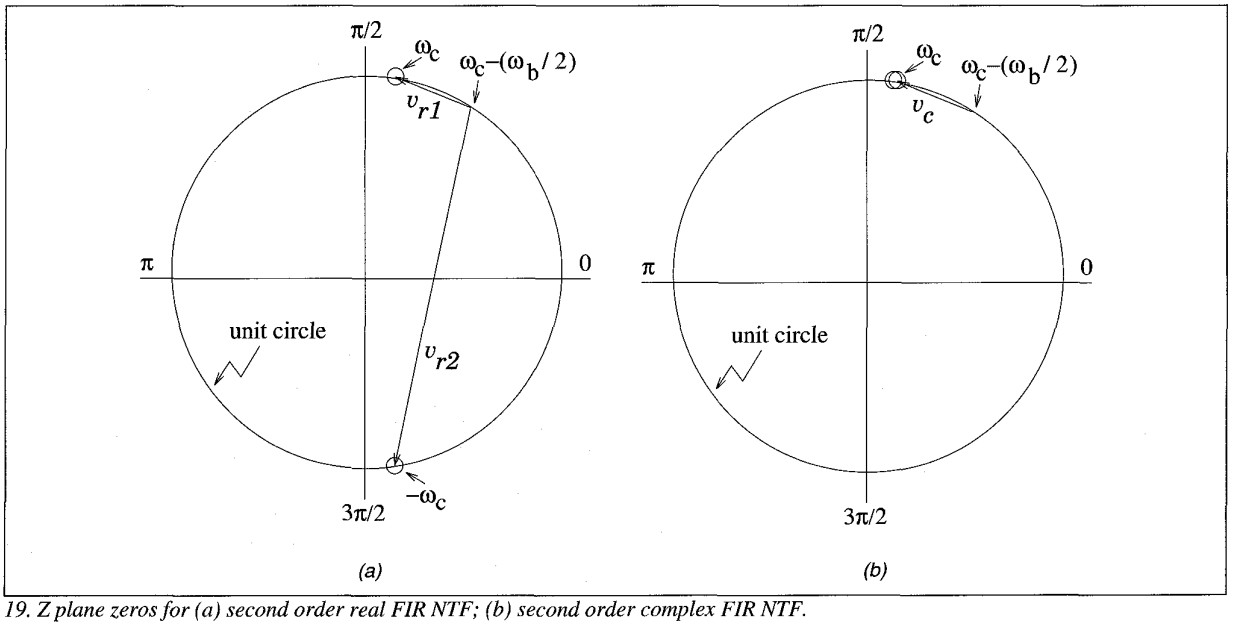
Real Band-pass Modulation

Unlike low-pass sigma-delta modulators, which realize NTF zeros at DC or low frequencies on the unit circle of the Z plane, band-pass modulators have NTFs that realize zeros or notches in the signal band of interest, $[f_c - f_B/2,\ f_c + f_B/2]$. Consequently, quantization noise that occurs over the signal band is attenuated, and noise power is pushed outside this band. Regardless of where the signal band is centered, the smaller the signal band, f_B, relative to the sampling frequency, f_s, there is less in-band noise power for a given NTF. Noise outside the signal band can be attenuated with a digital decimation filter and so high resolution conversion is possible for large oversampling ratios $f_s/2f_B$. The modulator STF and the decimation filter will typically have a band-pass characteristic, providing unity gain over the signal band.

As an example, the NTF and STF magnitude spectra for the design in [42] are shown in Fig 18a. In this example, the signal band has a center frequency of $f_c = 455$ kHz, $f_s = 3$ MHz, $f_B = 20$ kHz, so the oversampling ratio is 75. Figure 18a shows the magnitude spectra in dB of H_x ,and H_e. Figure 18b shows a closeup view of the magnitude spectrum in dB of H_e over the 435 kHz to 475 kHz region. The vertical bars delineate the 20 kHz signal band centered at 455 kHz. The NTF is sixth order, and as should be clear from Fig. 18b, contains three notches or zeros, which provide attenuation of the quantization noise over the 20 kHz signal bandwidth (the other three



18. Band-pass sigma-delta: (a) magnitude spectra in dB of H_x (the STF, dashed plot), and H_e (the NTF); (b) closeup of the magnitude spectrum in dB of H_e.



19. Z plane zeros for (a) second order real FIR NTF; (b) second order complex FIR NTF.

zeros are complex conjugates of these). The STF is band-pass and has minimal ripple and is approximately linear phase in the signal band. Band-pass converters employing Lth order modulators display a SNR performance that improves at the rate of $(3L + 3)$ dB per octave increment with the oversampling ratio $f_s/2f_B$ [42].

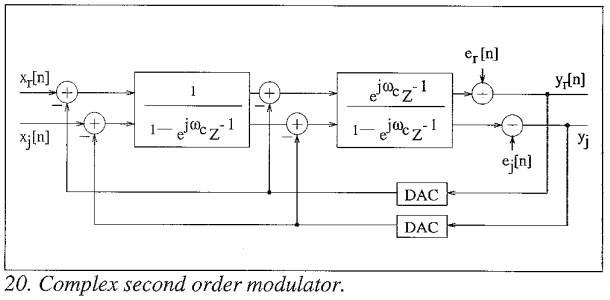
Complex Band-pass Modulation

Most band-pass sigma-delta converters [43-45] use real NTFs. Thus, all the coefficients in the Z domain transfer function are real. Let us now discuss the idea of complex band-pass NTFs, which have been proposed independently in [46] and [47]. The use of complex NTFs can improve the resolution that can be obtained for real band-pass signals [47]. The reason for this improvement is best illustrated through an example using second order FIR NTFs.

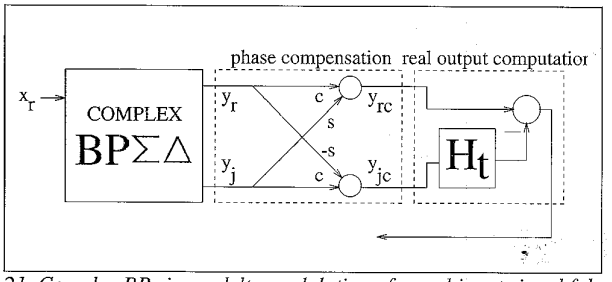
Figure 19a shows the unit circle on the Z plane containing zeros for a real second order FIR NTF. We find it more convenient to use the discrete time frequency $\omega = 2\pi f/f_s$ instead of f in our discussion. A real NTF realizing a zero at center frequency $\omega = \omega_c$ is constrained to also realize one at $-\omega_c$.

For a signal bandwidth ω_b consider the attenuation of the NTF at half the bandwidth away from the zero, that is, at $\omega_c - \omega_b/2$. This attenuation is given by the product of the magnitude of the vectors v_{r1} and v_{r2}, i.e., by $|v_{r1}| \times |v_{r2}|$. The smaller this product, the better the attenuation provided by the NTF, and the better the quantization noise suppression over the signal band, hence a better SNR will be obtained.

Vector v_{r1} represents the contribution from the zero at ω_c, and v_{r2} represents the contribution from the zero at $-\omega_c$. In the figure, the distance between ω_c and $\omega_c - \omega_b/2$ has been exaggerated—in practice, high resolution can be obtained only if



20. Complex second order modulator.



21. Complex BP sigma-delta modulation of a real input signal followed by phase compensation and computation of a real output.

the signal is narrowband and this distance is small. As the center frequency ω_c increases, v_{r2} increases as well, and the attenuation provided by the NTF becomes worse and so there will be less quantization noise suppression.

For the complex case, the zeros of a second order FIR NTF with both zeros located at a center frequency ω_c, are shown in Fig. 19b. The attenuation at ω_c-$\omega_b/2$ is given by the quantity $|v_c| \times |v_c|$, which depends only on ω_b and not on ω_c, as there is no influence from a zero at -ω_c. Consequently, the attenuation provided by the complex NTFs does not suffer any degradation with increasing center frequency. Therefore, at higher center frequencies, complex NTFs can provide more attenuation in the signal band, i.e., better quantization noise suppression and so a better SNR than a real transfer function. Complex band-pass converters employing Lth order modulators display a SNR performance which increases at the rate of $(6L + 3)$ dB per octave increment in the the oversampling ratio $f_s/2f_B$. This is in contrast to the $(3L + 3)$ dB rate of improvement for Lth order real band-pass modulators.

Having discussed the complex NTF, we now discuss the implementation of a complex 2nd order modulator with a band centered at ω_c. Such a modulator can be generated by modulating the Z domain NTFs and STFs of a standard 2nd order modulator with the transformation $z^{-1} \rightarrow z^{-1}\exp(j\omega_c) = z^{-1}\exp(j\,2\pi f_c/f_s)$. A block diagram of such a system obtained from Fig. 13 with the above transformation is shown in Fig. 20. Notice, for example, that the integrator $z^{-1}/(1 - z^{-1})$ of Fig. 13 is replaced with the integrator $(z^{-1}\,e^{j\omega_c})/(1 - e^{j\omega_c}\,z^{-1})$, which will have complex inputs and outputs consisting of real and imaginary parts. Such a complex integrator can be physically realized in switched capacitor technology using two cross coupled integrators [48]. Note there also need to be two physical quantizers, E_r and E_j, one for the "real" channel and the other for the "imaginary" channel. The Z domain output of the complex modulator is given by

$$Y(z) = [X_r(z) + jX_j(z)]z^{-1}\exp(j\omega_c) + [E_r(z) + jE_j(z)][1 - z^{-1}\exp(j\omega_c)]^2 \quad (30)$$

The STF is no longer a pure delay but contains the phase factor $e^{j\omega_c}$ However, this phase factor can be compensated digitally by multiplying the output of the complex modulator with the complex constant $e^{-j\omega_c}$ to obtain the phase compensated output. The phase compensated output is $Y_c(z) = Y(z)e^{-j\omega_c}$. For a real input signal, we have $x_j[n] = 0$ and $X_j(z) = 0$ so that $x[n] = x_r[n]$ and $X(z) = X_r(z)$. In this case,

$$Y_c(z) = X^r(z)z^{-1} + [E_r(z) + jE_j(z)] \cdot \exp(-j\omega_c)[1 - z^{-1}\exp(j\omega_c)]^2 \quad (31)$$

where $Y_c(z) = Y_{rc}(z) + j\,Y_{jc}(z)$.

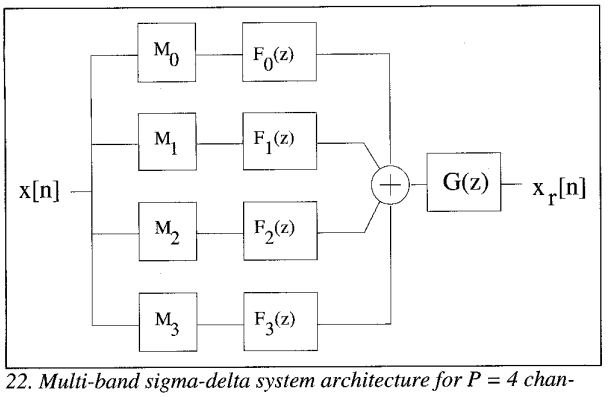
$Y_c(z)$ consists of a signal term $X_r(z)\,z^{-1}$, which we have assumed results from a real signal, plus the shaped quantization noise term. From Eq. 31, we can also see that in the time domain, the signal component in $y_c[n]$ is $x_r[n-1]$. However, $y_c[n]$ will still be complex because of the complex noise and, accordingly, the spectrum of $y_c[n] = y_{rc}[n] + j\,y_{jc}[n]$ will not be symmetric. A real output can be obtained without disturbing the signal or altering the SNR by considering the final output to be $y_{rc}[n] - y_{jc}[n]*h_t[n]$ where h_t is the impulse response of an ideal Hilbert transformer, $H_t(z)$. In the Z domain, the final output is then $[Y_{rc}(z) - Y_{jc}(z)H_t(z)]$. This operation amounts to keeping only positive frequencies (discarding negative frequencies) with the Hilbert transformer and then taking the real part to make the spectrum symmetric by folding the positive frequencies on to the negative frequency axis.

The phase compensation and computation of the real output are shown in Fig. 21 where $c = \cos(\omega_c) = \cos(2\pi f_c/f_s)$ and $s = \sin(\omega_c) = \sin(2\pi f_c/f_s)$. Decimation of the sigma-delta output and demodulation of the band-pass signal to baseband are not shown here.

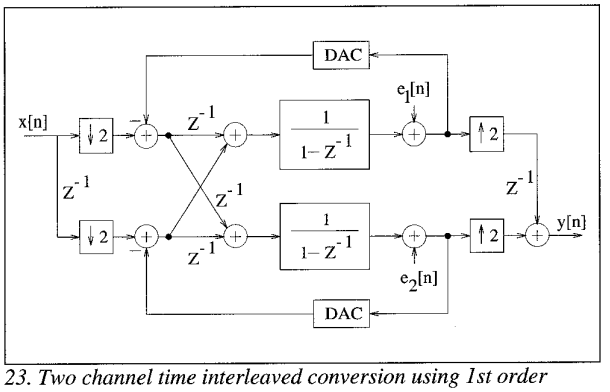
Note that in the case of a real input signal, the imaginary channel of the modulator is not directly connected to the input signal. One could feed $x_r[n]$ into the imaginary channel input as well. This has the benefit of resulting in a $\sqrt{2}$ gain of the STF which in principle will result in a 3 dB improvement in signal power. However, simulations show that with the input being fed to the imaginary channel, the quantizers overload much more often and the increase in quantization noise does not merit the gain in the signal power.

Parallel Sigma-Delta Systems

The use of parallelism for PCM A/D conversion has been considered in [49] and [50]. This section very briefly discusses several schemes that use architectural parallelism to improve the performance of sigma-delta modulators A/D



22. Multi-band sigma-delta system architecture for P = 4 channels.



23. Two channel time interleaved conversion using 1st order sigma-delta modulators.

converters. For a given signal bandwidth, modulator order, and sampling frequency, these architectures can attain higher resolution. The cost of realizing the improved performance lies clearly in the extra hardware needed for each parallel channel.

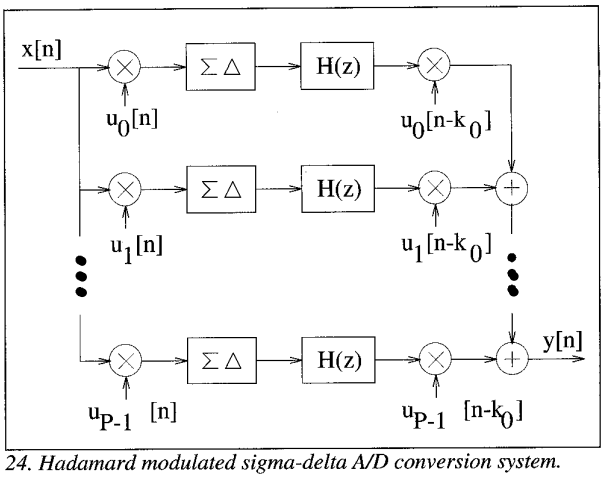
Multi-band Sigma-Delta Modulation

One architecture [51] uses modulators that realize different band reject NTFs for different portions of the signal band. Each band (channel) is converted in parallel. A bank of FIR filters attenuates the out-of-band noise for each band and can achieve perfect reconstruction of the signal component assuming that the modulator STF is a simple delay [52].

A block diagram of the system architecture for four channels is shown in Fig. 22, where M_k denote the modulators, and $F_k(z)$ and $G(z)$ comprise the digital filter bank. Using Lth order complex band-pass modulators, assuming equal sized bands, and assuming the quantization errors from the different channels to be mutually uncorrelated, the total in-band noise power, σ_{er}^2, at the output of the reconstructed signal is

$$\sigma_{er}^2 = \frac{4\sigma_e^2}{2^{2L+1}} \frac{\pi^{2L}}{2L+1} \left(\frac{2f_C}{f_s}\right)^{2L+1} \left(\frac{1}{P}\right)^{2L} \tag{32}$$

where f_B is the total signal bandwidth converted, $f_C = f_B/P$ is the bandwidth per channel, P is the number of channels, and



24. Hadamard modulated sigma-delta A/D conversion system.

σ_e^2 is the quantizer noise power in any one modulator. The SNR improves at a rate of (6L + 3) dB, or the resolution by (L + 1/2) bits per octave increment in the oversampling ratio per channel, $f_s/2f_C$. The SNR and resolution also improve at rates of (6L) dB, and L bits, respectively, per octave increment in the number of channels, P.

Time Interleaved Sigma-Delta Modulation

Another method for incorporating parallelism into sigma-delta converters is through time interleaving [53]. This architecture employs ideas of block filter theory to use P identical, mutually cross coupled, modulators running at a sampling rate f_s to generate the same modulator transfer function, which runs at an equivalent sampling rate of Pf_s. The block diagram for a two channel system using 1st order modulators running in each individual channel is shown in Fig 23. Note that downsampling the signal can result in aliasing, but the choice of the cross coupling terms ensure that the aliasing is cancelled in the final output for the ideal system.

If Lth order low-pass sigma-delta modulators are used in each channel, the ideal in-band SNR improves by (6L + 3) dB, or the resolution by L + 1/2 bits per octave increment in the number of channels, P, since each octave increment in P amounts to an octave increment in the oversampling ratio of the modulators.

Hadamard System

Another parallel sigma-delta system has been described in the literature [54] recently. Here, each channel contains a Hadamard modulator, which multiplies the input signal by a ± 1 sequence, $u_k[n]$. This operation is called Hadamard modulation.

The Hadamard sequences are obtained from repeating the rows of the Hadamard transform matrix. The Hadamard modulated sequence of each channel is then quantized using a standard sigma-delta modulator. The output of each sigma-delta modulator is filtered to attenuate out-of-band noise, and

again multiplied by a Hadamard sequence before all the channel outputs are added to provide the final output.

The block diagram of the system is shown in Fig 24, where k_0 is assumed to be the delay in the STF of the sigma-delta modulators. The overall effect of the system on the signal is to filter it with a subset of the filter coefficients of $H(z)$. The relevant $H(z)$ coefficients affecting the signal can be chosen such that the STF experiences a delay, while the other coefficients can be chosen to maximize the quantization noise attenuation.

The quantization error does not see the first Hadamard sequence and the effect of the second Hadamard sequence modulation is to frequency shift the filtered quantization error power spectral density. Using Lth order low-pass sigma-delta modulators, the SNR using this approach improves by $(6L)$ dB or the resolution improves by L bits per octave increment in the number of channels P.

Applications Using Actual Sigma-Delta Converters

We now present some applications using actual sigma-delta converters that have been fabricated in VLSI. The purpose here is to provide a sense of the final performance achieved by the converters as to resolution and bandwidth, rather than to compare them with respect to theoretical performance or to the multitude of other performance criterion (power, area, topology, technology, etc.). To aid the resolution/bandwidth evaluation, we use the abbreviation *osr* for the oversampling ratio, $f_s/2f_B$, where f_B is the signal bandwidth and f_s is the sampling frequency. To provide an overview of the converters discussed here, we have included all the converters in a common format in the Table.

Data conversion for instrumentation applications may require resolutions up to 19-20 bits, albeit at low bandwidths. One such converter used in instrumentation transducers [55] uses a 5th order modulator topology with a $f_s = 128$ kHz (*osr* $\simeq$128) to a achieve a 118 dB dynamic range, or more than 19 bits of resolution over a 492 Hz bandwidth. Similarly, a converter that is used for seismic activity measurements uses a 128 kHz sampling rate with an *osr* of 128, achieving more than 120 dB peak SNR, or almost 20 bits of resolution over about a 500 Hz bandwidth [56]. A 4th order topology is used. Note that for extremely high resolution such as reported here, the quantization noise floor approaches the level of circuit noise for state of the art technologies. Thus, very careful circuit design and optimization is required to fully take advantage of the potential performance that can be realized by the sigma-delta ADC.

Sigma-delta converters are good candidates for voiceband (speech) applications where the signal bandwidth is 4 kHz and 13-14 bits of resolution is desirable. One such converter [57] actually used a single bit 1st order modulator with $f_s = 4$ MHz, or an *osr* of about 500 to achieve a 79 dB dynamic range, i.e., about 13 bits of resolution. Dithering was required to alleviate the tone problem associated with the first order modulator. Another converter [58] also achieved 13 bits resolution using a $f_s = 1.024$ MHz with an *osr* of 128. The modulator used is a standard second order modulator employing a 1 bit quantizer.

Digital audio applications such Hi-Fi CD and DAT systems often use sigma-delta A/D converters. Consumer quality Hi-Fi audio needs to be digitized at 16 bits of resolution, and audiophiles prefer up to 18-20 bits of resolution. Many converters have been reported in the literature for audio bandwidths of 20-24 kHz. A high resolution audio range converter was reported as early as 1986 and achieved a dynamic range of 106 dB, or almost 18 bits of resolution over a 24 kHz bandwidth using a fourth order modulator and 4 bit internal A/D and D/A converters [59]. The sampling frequency used was 6.144 MHz with an *osr* of 128. Another converter that also uses a fourth order modulator to convert a 24 kHz bandwidth achieves near 16 bit dynamic range using single bit quantizers with an *osr* of 64 or $f_s = 3.072$ MHz [60]. Audio band conversion has also been performed by a standard 2nd order modulator with a 4 bit internal ADC and DAC. The converter achieved nearly 16 bit peak SNR for a 20.5 kHz

Table: Sigma Delta Converters in Use Today

Signal Band (f_B)	Sampling Frequency (f_s)	OSR ($f_s/2f_B$)	Overall Resolution (# bits)	Modulator Structure	Internal Quantizer	Application	Reference
492 Hz	128 kHz	128	20	4th order	1 bit	instrumentation	[55]
500 Hz	128 kHz	128	20	4th order	1 bit	seismic	[56]
4 kHz	4 MHz	500	13	1st order	1 bit	speech	[57]
4 kHz	1.024 MHz	128	13	2nd order	1 bit	speech	[58]
20.5 kHz	5.25 MHz	128	16	2nd order	4 bits	audio	[30]
24 kHz	6.144 MHz	128	18	4th order	4 bits	audio	[59]
24 kHz	3.072 MHz	64	16	2nd order	1 bit	audio	[60]
25 kHz	6.4 MHz	128	17	"2-1" cascade	1 bit	audio	[39]
40 kHz	10.24 MHz	128	14	2nd order	1 bit	ISDN	[16]
40 kHz	2.56 MHz	32	13	"2-1" cascade	1 bit	ISDN	[37]
100 kHz	3.25 MHz	16	15	"2-2-2" cascade	3 level	digital cellular radio	[61]
160 kHz	20.48 MHz	24	16	"2-1" cascade	1 bit	-	[38]
250 kHz	32 MHz	32	14	4th order	1 bit	-	[20]
1 MHz	50 MHz	20	12	"2-1" cascade	1 and 3 bit	ultrasound	[62]

bandwidth using an *osr* of 128, or $f_s = 5.25$ MHz [30]. Finally, an architecture consisting of a cascade of second order and a first order modulator and employing 1 bit quantizers achieves nearly 17 bit performance for a 25 kHz bandwidth, also using an *osr* of 128 with a $f_s = 6.4$ MHz [39].

A sigma-delta converter has also been used as a receiver input of an ISDN U-interface 2B1Q access rate receiver [16]. The converter attains a dynamic range of 89 dB, or a resolution of 14 bits, for a 40 kHz bandwidth. A standard 2nd order modulator was used with a 1 bit quantizer running at a 10.24 MHz sampling frequency, and an *osr* of 128. Another converter [37] used for a similar ISDN U-interface consisted of a cascade of a 2nd order modulator, followed by a 1st order modulator. The resolution was 13 bits, using a f_s of 2.56 MHz, i.e., an *osr* of 32.

A sigma-delta ADC has been used as the baseband converter for a digital cellular radio that required a moderately higher bandwidth of 100 kHz [61]. The converter produced 15 bit peak resolution by using a three stage cascade employing three 2nd order modulators running at $f_s = 3.25$ MHz. The *osr* was about 16.

At somewhat higher bandwidths, high resolution conversion has been attained for 160 kHz and 250 kHz bandwidths. The 160 kHz bandwidth converter achieves a dynamic range of 96 dB, or nearly 16 bit performance using a cascade of second and first order modulators that employ 1 bit quantizers [38]. The sampling rate is $f_s = 20.48$ MHz and the resulting *osr* is 24. The 250 kHz bandwidth converter achieves 14 bit resolution using a 4th order modulator using a 1 bit quantizer [20]. The *osr* is 32 and thus the f_s is 32 MHz.

At high conversion bandwidths of about 1 MHz, a 12 bit converter has been realized with an *osr* of about 24, and a f_s of about 50 MHz [62]. The converter uses a cascade of second order and first order modulators. The second order modulator used in the first stage employs a 1 bit quantizer, while the 1st order modulator used in the second stage utilizes a 3 bit internal ADC and DAC. Such a converter can find use in data acquisition for ultrasound imaging systems.

Finally, sigma-delta ADCs have been integrated with digital signal processor (DSPs) to provide a single chip data conversion/ computation engine solution. Sigma-delta ADCs require relatively imprecise analog circuits and digital decimation filtering, thus making them good candidates for fabrication using digital technology such as CMOS. The task of decimation can be handled entirely by a DSP or shared by a DSP and some extra digital hardware dedicated to performing a portion of the decimation.

Conclusion

We have reviewed the basic principles of A/D conversion with sigma-delta modulators. The techniques of oversampling and noise shaping allows the use of relatively imprecise analog circuits to perform high resolution conversion using only a 1 bit A/D converter. Oversampling reduces the amount of quantization noise power present in the signal band, and noise shaping further attenuates quantization noise in the signal band, thereby pushing noise power to out-of-band frequencies. The use of analog filtering combined with feedback around the 1 bit A/D can be used to implement the noise shaping sigma-delta modulator. The noise power that is pushed outside the signal band can be attenuated by a digital filter such that it has no further effect on the signal.

Various sigma-delta architectures exist and many of these have been used in applications such as instrumentation, speech and Hi-Fi audio digitization, ISDN and digital cellular radio. Sigma-delta techniques are also applicable to the high resolution A/D conversion of narrowband band-pass signals using band-pass sigma-delta modulators. Parallel sigma-delta systems offer the potential for extending high resolution operation to larger signal bandwidths than currently possible with single channel systems.

Pervez M. Aziz is a Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ. Henrik V. Sorensen is Senior Researcher at Ariel Corporation, Highland Park, NJ. Jan Van der Spiegel is Professor, Department of Electrical Engineering, University of Pennsylvania, Philadelphia. This work was completed while the first two authors were at the University of Pennsylvania.

References

1. S. Renukunta and D. Wells, "Optical memory and blue lasers," *IEEE Potentials*, pp. 14-18, Oct/Nov 1994.

2. J. Candy and G. Temes, "Oversampling methods for A/D and D/A conversion" in *Oversampling Delta-Sigma Data Converters*, pp. 1-25, IEEE Press, 1992.

3. A. Oppennheim and R. Schafer:, *Discrete Time Signal Processing*, (Prentice-Hall, 1989).

4. W. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, pp. 446-472, July 1948.

5. Boser and B. Wooley,"The design of sigma-delta modulation analog-to-digital converters," *IEEE Journal of Solid State Circuits*, pp. 1298-1308, December, 1988.

6. B. Leung, "Theory of sigma-delta analog to digital converter," *IEEE International Symposium on Circuits and Systems Tutorials*, pp. 196-223, 1994.

7. H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback," *Proceedings of the IEEE*, pp. 1524-1535, November, 1963.

8. J. Candy, "A use of double integration in sigma delta modulation," *IEEE Transactions on Communications*, pp. 249-258, March, 1985.

9. V. Friedman, "The structure of limit cycles in sigma delta modulation," *IEEE Transactions on Communications*, pp. 972-979, August, 1988.

10. J. Candy, O. Benjamin, "The structure of quantization noise from sigma-delta modulation," *IEEE Transactions on Communications*, pp. 1316-1323, September, 1981.

11. S. Hein and A. Zakhor, "On the stability of sigma delta modulators," *IEEE Transactions on Signal Processing*, pp. 2322-2348, July, 1993.

12. R. Gray, "Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input," *IEEE Transactions on Communications*, pp. 588-599, June, 1989.

13. N. He, F. Kuhlmann, A. Buzo, "Double-loop sigma-delta modulation with dc input," *IEEE Transactions on Communications*, pp. 487-495, April, 1990.

14. R. Gray, "Quantization noise spectra," *IEEE Transactions on Information Theory*, pp. 1220-1244, November, 1990.

15. S. Rangan and B. Leung, "Quantization noise spectrum of double-loop sigma-delta converter with sinusoidal input," *IEEE Transactions on Circuits and Systems II*, pp. 168-173, February, 1994.

16. S. Norsworthy, I. Post, H. Fetterman,"A 14-bit 80kHz sigma-delta A/D converter: modeling, design, and performance evaluation," *IEEE Journal of Solid State Circuits*, pp. 256-266, April, 1989.

17. S. Norsworthy and D. Rich, "Idle channel tones and dithering in delta sigma modulators," *95th Convention of the Audioengineering Society*, preprint 3711, October, 1993.

18. K. Chao, S. Nadeem, W. Lee, and C. Sodini, "A higher order topology for interpolative modulators for oversampling A/D converters," *IEEE Transactions on Circuits and Systems*, pp. 309-318, March, 1990.

19. P. Ferguson, A. Ganesan, R. Adams, "One bit higher order sigma-delta A/D converters," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 890-893, 1990.

20. F. Op't Eynde, G. Yin, W. Sansen, "A CMOS fourth-order 14b 500k-sample/s sigma-delta ADC converter," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 62-63, 1991.

21. R. Adams, "Design aspects of high-order delta-sigma A/D converters," *IEEE International Symposium on Circuits and Systems Tutorials*, pp. 235-259, 1994.

22. S. Ardalan and J. Paulos, "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Transactions on Circuits and Systems*, pp. 593-603, June, 1987.

23. R. Baird and T. Fiez, "Stability analysis of high-order delta-sigma modulation for ADCs," *IEEE Transactions on Circuits and Systems II*, pp. 59-62, January, 1994.

24. S. Moussavi and B. Leung, "High-order single-stage single-bit oversampling A/D converter stabilized with local feedback loops," *IEEE Transactions on Circuits and Systems II*, pp. 19-25, January, 1994.

25. D. Anastassiou, "Error diffusion coding for A/D conversion," *IEEE Transactions on Circuits and Systems*, pp. 1175-1186, September, 1989.

26. R. Schreier and Y. Yang, "Stability tests for single-bit sigma-delta modulators with second-order FIR noise transfer functions," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 1316-1319, 1992.

27. L. Larson, T. Cataltepe, and G. Temes, "Multibit oversampled sigma-delta A/D convertor with digital error correction," *Electronics Letters*, pp. 1051-1052, August 4, 1988.

28. T. Leslie and B. Singh, "An improved sigma-delta modulator architecture," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 372-375, 1990.

29. A. Hairapetian, G. Temes, and Z. Zhang, "Multibit sigma-delta modulator with reduced sensitivity to DAC nonlinearity," *Electronics Letters*, pp. 990-991, May 23, 1991.

30. M. Sarhang-Nejad and G. Temes, "A high-resolution multibit sigma-delta ADC with digital correction and relaxed amplifier requirements," *IEEE Journal of Solid State Circuits*, pp. 648-660, June, 1993.

31. L. Carley, "A noise-shaping coder topology for 15+ bit converters," *IEEE Journal of Solid State Circuits*, pp. 267-273, April, 1989.

32. B. Leung and S. Sutarja, "Multibit sigma-delta A/D converter incorporating a novel class of dynamic element matching technique," *IEEE Transactions on Circuits and Systems II*, pp. 35-51, January, 1992.

33. F. Chen and B. Leung, "A high resolution multi-bit sigma-delta modulator with individual level averaging," Digest of Technical Papers, *IEEE Symposium on VLSI Circuits*, pp. 101-102, June, 1994.

34. Y. Matsuya, K. Uchimura et al, "A 16-bit oversampling A-to-D conversion technology using triple-integration noise shaping," *IEEE Journal of Solid State Circuits*, pp. 921-929, December, 1987.

35. K. Uchimura, T. Hayashi, T. Kimura and A. Iwata, "Oversampling A-to-D and D-to-A converters with multistage noise shaping modulators," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 1899-1905, December, 1988.

36. M. Rebeschini, N. van Bavel, P. Rakers, et al, "A 16-b 160-kHz CMOS A/D converter using sigma-delta modulation," *IEEE Journal of Solid State Circuits*, pp. 431-440, April, 1990.

37. L. Longo and M. Copeland, "A 13 bit ISDN-band oversampled ADC using two-stage third order noise shaping," *Proceedings, IEEE Custom Integrated Circuits Conference*, pp. 21.2.1-21.2.4, 1988.

38. G. Yin, F. Stubbe, W. Sansen, "A 16-b 320-kHz CMOS A/D converter using two-stage third-order sigma-delta noise shaping," *IEEE Journal of Solid State Circuits*, pp. 640-647, June, 1993.

39. L. Williams and B. Wooley, "Third-order sigma-delta modulator with extended dynamic range," *IEEE Journal of Solid State Circuits*, pp. 193-202, March, 1994.

40. D. Ribner, "A comparison of modulator networks for high-order oversampled sigma-delta analog-to-digital converters," *IEEE Transactions on Circuits and Systems*, pp. 145-159, February, 1991.

41. R. Schreier and M. Snelgrove, "Bandpass sigma-delta modulation," *Electronics Letters*, pp. 1560-1561, November 9, 1989.

42. S. Jantzi, R. Schreier and M. Snelgrove, "Bandpass sigma-delta analog-to-digital conversion," *IEEE Transactions On Circuits and Systems*, pp. 1406-1409, November, 1991.

43. L. Longo and B. Horng, "A 15b 30kHz bandpass sigma delta modulator," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 226-227, 1993.

44. S. Jantzi, M. Snelgrove, P. Ferguson, "A 4th-order bandpass sigma-delta modulator," *IEEE Journal of Solid State Circuits*, pp. 282-291, March, 1993.

45. G. Tröster, H Dreβler, et al, "An interpolative bandpass converter on a 1.2-um BiCMOS analog/digital array," *IEEE Journal of Solid State Circuits*, pp. 471-477, April, 1993.

46. S. Jantzi, K. Martin, M. Snelgrove, A. Sedra, "Complex bandpass sigma-delta converter for digital radio," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 453-456, 1994.

47. P. Aziz, H. Sorensen, J. Van der Spiegel, "Performance of complex noise transfer functions in bandpass and multi band sigma delta systems," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 641-644, 1995.

48. Q. Liu, M. Snelgrove, A. Sedra, "Switched-capacitor implementation of complex filters," *Proceedings, IEEE International Symposium on Circuits and Systems*}, pp. 1121-1124, 1986.

49. W. Black and D. Hodges, "Time interleaved converter arrays," *IEEE Journal of Solid State Circuits*, pp. 1022-1029, December, 1980.

50. A. Petraglia and S. Mitra, "High speed A/D conversion using QMF banks," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 2797-2800, 1990.

51. P. Aziz, H. Sorensen, J. Van der Spiegel, "Multiband sigma-delta modulation," *Electronics Letters*, pp. 760-762, April 29, 1993.

52. P. Aziz, H. Sorensen, J. Van der Spiegel, "Multiband sigma delta analog to digital conversion," *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 249-252, 1994.

53. R. Khoini-Poorfard, L. Lim, D. Johns, "Time-interleaved oversampling convertors," *Electronics Letters*, pp. 1673-1674, September 16, 1993.

54. I. Galton and H. Jensen, "Delta-sigma modulator based A/D conversion without oversampling," to appear, *IEEE Transactions on Circuits and Systems*.

55. C. Thompson, S. Bernadas, "A digitally corrected 20b delta-sigma modulator," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 194-195, 1994.

56. D. Kerth, D. Kasha, et al, "A 120 dB linear switched-capacitor delta-sigma modulator," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 196-197, 1994.

57. B. Leung, R. Neff, P. Gray, R. Broderson, "Area-Efficient Multichannel Oversampled PCM Voice-Band Coder," *IEEE Journal of Solid State Circuits*, pp. 1351-1357, December, 1988.

58. V. Friedman, D. Brinthaupt, et al, "A dual-channel voice-band PCM codec using sigma-delta modulation technique," *IEEE Journal of Solid State Circuits*, pp. 274-280, April, 1989.

59. R. Adams, "Design and implementation of an audio 18-bit analog-to-digital converter using oversampling techniques," *Journal of the Audioengineering Society*, pp. 153-166, March, 1986.

60. D. Welland, B. Del Signore, et al, "A stereo 16-bit delta-sigma A/D converter for digital audio," *Journal of the Audioengineering Society*, pp. 476-486, June, 1989.

61. I. Dedic, "A sixth-order triple-loop sigma-delta CMOS ADC with 90 dB SNR and 100 kHz bandwidth," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 188-189, 1994.

62. B. Brandt and B. Wooley, "A CMOS oversampling A/D converter with 12b resolution at conversion rates above 1 MHz," Digest of Technical Papers, *International Solid State Circuits Conference*, pp. 64-65, 1991.