

Chapter 2

$\Sigma\Delta$ ADCs: Basic Concepts, Topologies and State of the Art

SIGMA-DELTA MODULATION HAS DEMONSTRATED TO BE VERY SUITED FOR the implementation of Analog-to-Digital (A/D) interfaces in many different electronic systems, covering a large number of applications from instrumentation to telecom [Nors97a, Mede99, Geer02, Schr05b, Ortm06]. This type of A/D converters (ADCs), composed of a low-resolution quantizer embedded in a negative feedback loop, uses oversampling (a sampling frequency much larger than the Nyquist frequency) to reduce the quantization error and $\Sigma\Delta$ modulation to push this noise out of the signal band [Inos62]. The combined use of redundant temporal data (oversampling) and filtering (noise shaping) results in high-resolution, robust ADCs which have lower sensitivity to circuit parasitics and tolerances, and are more suited for their implementation in modern standard CMOS technologies [Rodr03, Schr05b].

The aim of this chapter is to introduce the foundations of $\Sigma\Delta$ ADCs. The fundamentals of sampling and quantization processes—inherent to the A/D conversion—are revised in Sect. 2.1 for the purpose of juxtaposing the operation principles of $\Sigma\Delta$ modulation (oversampling and quantization noise shaping). Sect. 2.2 presents the basic scheme of a $\Sigma\Delta$ ADC, together with its ideal behavior and a definition of its most important performance metrics. A classification of practical implementations of $\Sigma\Delta$ modulators is then presented in Sect. 2.3. Sect. 2.4 describes single-loop $\Sigma\Delta$ architectures, discussing stability issues and practical topologies to implement high-order loops. Cascade $\Sigma\Delta$ architectures are presented in Sect. 2.5 as an architectural alternative to implement unconditionally-stable high-order $\Sigma\Delta$ modulators. Sect. 2.6 revises $\Sigma\Delta$ modulators with multi-bit embedded quantizers, analyzing their pros and providing techniques to circumvent their impact on the modulator linearity, such as dual-quantization or dynamic element matching techniques. Finally, the state of the art in $\Sigma\Delta$ ADCs is reviewed from integrated circuit implementations reported in open literature.

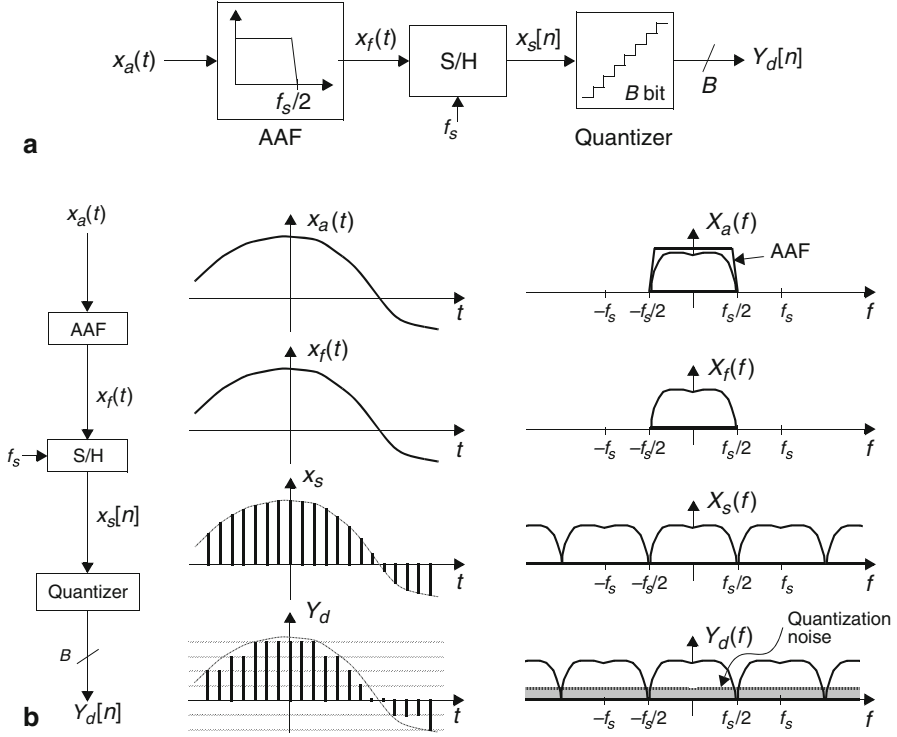


Fig. 2.1 Analog-to-digital converter: **a** generic scheme, **b** signal processing involved

2.1 Fundamentals of the A/D Conversion

An ADC is a system that transforms signals which are continuous in time and amplitude (analog signals) into signals which are discrete in time and amplitude (digital signals). Fig. 2.1a shows the generic scheme of an ADC intended for the conversion of low-pass signals that includes the following blocks: an anti-aliasing filter (AAF), a sampling-and-hold circuit (S/H) and a quantizer. The signal processing involved in the operation of the ADC blocks is illustrated in Fig. 2.1b both in time and frequency domains. First, the analog input signal $x_a(t)$ of the ADC passes through the AAF in order to remove spectral components above one half of the sampling frequency f_s of the S/H. Otherwise, according to the Nyquist theorem, out-of-band components would be folded back into the signal band during the subsequent sampling process, therefore corrupting the signal information. The resulting band-limited signal $x_f[t]$ is then sampled at a rate f_s by the S/H, yielding a discrete-time signal $x_s[n] = x_f[nT_s]$, where $T_s = 1/f_s$. Finally, the quantizer maps the range of amplitudes of $x_s[n]$ into a discrete set of levels using B bits; i.e., each sample of the continuous-valued input is coded onto the closer discrete-valued level out of the 2^B levels that cover the variation interval of the input signal. This process yields the converter digital output $Y_d[n]$.

The fundamental operations involved in the A/D conversion are sampling and quantization, as illustrated in Fig. 2.1. On the one hand, the sampling process performs the continuous-to-discrete conversion of the input signal in the time domain. On the other, the quantization process performs the continuous-to-discrete conversion of the input signal in amplitude. These two transformations inherently impose limitations to the performance of an ADC, even if its implemented using ideal components.

2.1.1 Sampling

Sampling imposes a limit on the bandwidth of the analog input signal. According to the Nyquist theorem, the minimum frequency f_s required for sampling a signal with no loss of information is twice the signal bandwidth, BW; i.e., $f_N = 2BW$, also called the Nyquist frequency. Based on this criterion, the ADCs in which the analog input signal is sampled at minimum rate ($f_s = f_N$) are called Nyquist ADCs. Fig. 2.1b illustrated the signal processing involved assuming a Nyquist ADC. Since the input signal bandwidth equals $f_s/2$, aliasing will occur if $x_a(t)$ contains frequency components above $f_s/2$. High-order AAFs are therefore required in Nyquist ADCs in order to implement a sharp transition band and remove the out-of-band components with no significant attenuation of the signal band.

2.1.2 Quantization

Quantization also limits the performance of an ideal ADC, since the process itself of mapping continuous-valued levels into a set of discrete levels degrades the quality of the input signal. In the process an error is generated, called quantization error.

This process is illustrated in Fig. 2.2 for the case of a 3-bit ideal quantizer ($B = 3$). As the input signal changes from x_{min} to x_{max} , it is ‘rounded’ to one out of the eight (2^B) discrete levels. For a B -bit quantizer, the separation between adjacent levels is defined by the quantization step Δ as

$$\Delta = \frac{X_{FS}}{(2^B - 1)} \quad (2.1)$$

with X_{FS} being the quantizer full scale.

The quantizer operation can be therefore described mathematically by a linear model

$$y = g_q x + e(x) \quad (2.2)$$

where g_q stands for the quantizer gain—the slope of the line intersecting the code transitions—and $e(x)$ stands for the quantization error. This error is a non-linear

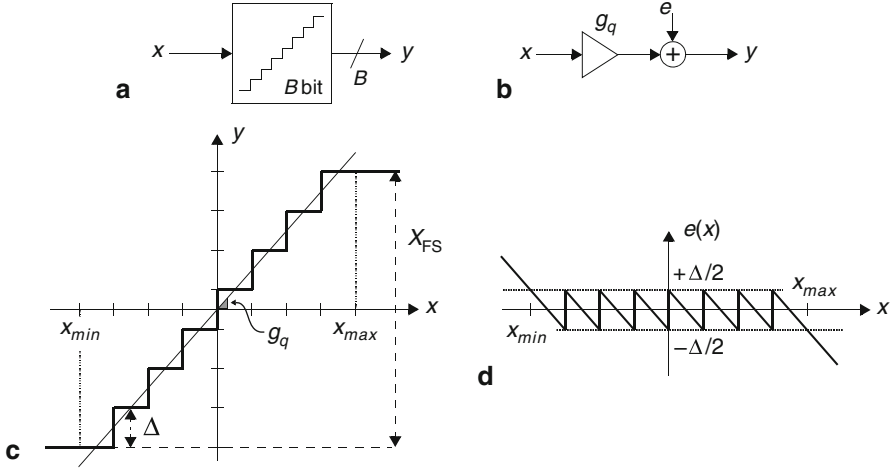


Fig. 2.2 Ideal quantization process: **a** symbolic representation, **b** linear model of an ideal quantizer, **c** ideal transfer characteristic, **d** quantization error of a 3-bit quantizer

function of the input x , as shown in Fig. 2.2d. Note that, if x is confined in the interval $[x_{min}, x_{max}]$, the quantization error is bounded by $\pm \Delta/2$. For inputs outside that interval, the absolute value of the quantizer error grows monotonically. This situation is known as quantizer overload.

2.1.3 White Noise Approximation of Quantization Error

In order to evaluate the performance of an ideal quantizer, some assumptions are usually made on the statistical properties of the quantization error which are collectively called the ‘additive white noise approximation’. As shown in Fig. 2.2d, the quantization error is strongly dependent on the input signal value. Nevertheless, if x is assumed to change randomly from sample to sample in the interval $[x_{min}, x_{max}]$ and the number of levels in the quantizer is large, the quantization error can be assumed uncorrelated from sample to sample. Quantization can therefore be viewed as a random process, with the quantization error exhibiting a uniform probability density function (PDF) in the range $[-\Delta/2, +\Delta/2]$ [Benn48, Srip77]. The power associated to the quantization error is then

$$\sigma^2(e) = \int_{-\infty}^{+\infty} e^2 \text{PDF}(e) de = \int_{-\Delta/2}^{+\Delta/2} e^2 \frac{1}{\Delta} de = \frac{\Delta^2}{12} \quad (2.3)$$

This power will be uniformly distributed in the band $[-f_s/2, +f_s/2]$ as the quantized signal is sampled at rate f_s and the quantization error will therefore exhibit a constant

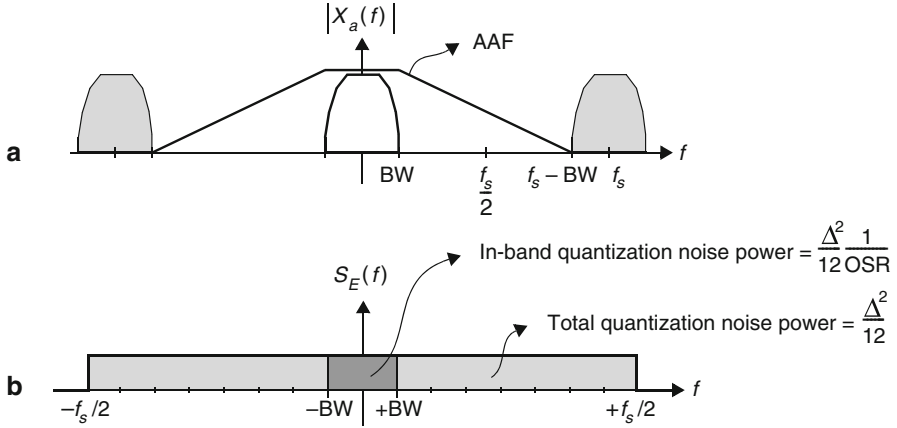


Fig. 2.3 Benefits of oversampling on the: **a** AAF, **b** in-band quantization noise

power spectral density (PSD) in that frequency interval

$$S_E(f) = \frac{\sigma^2(e)}{f_s} = \frac{\Delta^2}{12f_s} \quad (2.4)$$

The degradation introduced by the quantizer in the performance of an ADC can be expressed through the in-band quantization ‘noise’ power P_Q , calculated as

$$P_Q = \int_{-BW}^{+BW} S_E(f) df = \frac{\Delta^2}{12} \quad (2.5)$$

Note that f_s equals $2BW$ in a Nyquist ADC and all the quantization noise power therefore falls inside the signal band and passes to the ADC output as a part of the signal itself.

2.1.4 Oversampling

Oversampling consists in sampling a signal faster than the minimum rate imposed by the Nyquist theorem to avoid aliasing. How much faster than required the signal is sampled is expressed through the oversampling ratio, defined as $OSR = f_s/(2BW)$.

Oversampling has two noticeable effects in an ADC. First, as illustrated in Fig. 2.3a, since f_s is larger than the Nyquist rate, the images of the input created by the sampling process are more separated than in a Nyquist ADC. Spectral components of the input signal in the range $[BW, f_s - BW]$ do not alias within the signal band and, consequently, the transition band of the AAF can be smoother in an oversampling ADC, what greatly simplifies its design. Second, as illustrated in Fig. 2.3b, when an

oversampled signal is quantized, the quantization noise is uniformly distributed in the range $[-f_s/2, +f_s/2]$ and only a fraction of the total power lays within the signal band. The in-band quantization noise power can therefore be calculated as

$$P_Q = \int_{-BW}^{+BW} S_E(f) df = \int_{-BW}^{+BW} \frac{\Delta^2}{12f_s} df = \frac{\Delta^2}{12OSR} \quad (2.6)$$

and decreases with OSR at a rate of 3 dB/octave.

2.1.5 Quantization Noise Shaping

The quantization noise power within the signal band can be further decreased through the processing of the quantization error. Let us consider the quantization of an over-sampled signal. If OSR is large, the input signal value will only slightly change from one sample to another and most of the changes in the quantization error will occur at high frequencies—i.e., low-frequency components of consecutive samples of the quantization error $e[n]$ will be similar. Hence, low-frequency in-band components of the quantization error can be attenuated by subtracting the previous sample from the current one

$$e_{HPF}[n] = e[n] - e[n-1] \quad (2.7)$$

or further reduced by involving a larger number of previous samples in the error processing

$$\begin{aligned} e_{HPF,1}[n] &= e[n] - e[n-1], & \text{1st - order error processing} \\ e_{HPF,2}[n] &= e[n] - 2e[n-1] + e[n-2], & \text{2nd - order error processing} \\ e_{HPF,3}[n] &= e[n] - 3e[n-1] + 3e[n-2] - e[n-3], & \text{3rd - order error processing} \\ \dots & & \end{aligned} \quad (2.8)$$

This procedure can be formulated in an unified manner in Z-domain as

$$E_{HPF,L}(z) = (1 - z^{-1})^L \cdot E(z) \quad (2.9)$$

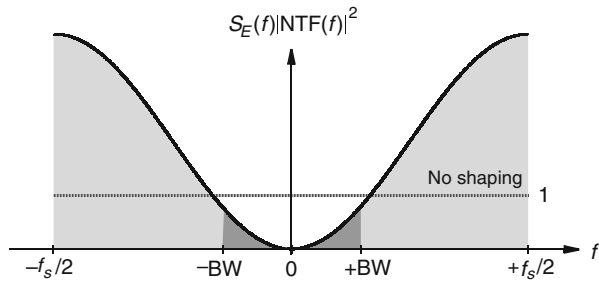
indicating that the processed error is a (high-pass) filtered version of the original. The filtering transfer function on the quantization error due to this processing, called noise transfer function (NTF), is therefore obtained as

$$NTF(z) = (1 - z^{-1})^L \quad (2.10)$$

where L denotes the order of the filtering. If OSR is large enough, NTF takes small values within the signal band and can be approximated to

$$\left| NTF\left(e^{j2\pi f/f_s}\right) \right|^2 = \left| 1 - e^{-j2\pi f/f_s} \right|^{2L} = 2^{2L} \sin^{2L}\left(\frac{\pi f}{f_s}\right) \approx 2^{2L} \left(\frac{\pi f}{f_s}\right)^{2L} \quad (2.11)$$

Fig. 2.4 Illustration of the quantization noise shaping



Hence, the in-band power of the filtered quantization error results in

$$P_Q = \int_{-BW}^{+BW} S_E(f) |NTF(f)|^2 df \approx \frac{\Delta^2}{12} \cdot \frac{\pi^{2L}}{(2L+1) \text{OSR}^{(2L+1)}} \quad (2.12)$$

that is much smaller than only applying oversampling. The resulting error reduction is illustrated in Fig. 2.4.

2.2 Basics of $\Sigma\Delta$ A/D Converters

In contrast to Nyquist ADCs, oversampling sigma-delta ADCs—usually referred to as $\Sigma\Delta$ ADCs—make use of oversampling and noise shaping to decrease the quantization noise power within the signal band and increase the accuracy of the A-to-D conversion. Fig. 2.5 illustrates the basic scheme of a $\Sigma\Delta$ ADC, as well as the signal processing involved. As shown, a $\Sigma\Delta$ converter comprises three main blocks:

- **Anti-Aliasing Filter (AAF).** Its function is the same as in Nyquist ADCs; i.e., to band limit the input signal in order to avoid aliasing during sampling. As stated above, oversampling considerably relaxes the attenuation requirements for this analog filter and smooth transition bands are sufficient (see Fig. 2.3a).
- **Sigma-Delta Modulator ($\Sigma\Delta\text{M}$).** It simultaneously performs the oversampling and quantization of the band-limited input signal. Quantization error is also high-pass filtered by means of a given noise-shaping technique. This is accomplished by placing an appropriate loop filter $H(z)$ before a low-resolution quantizer and closing a negative feedback loop around them. The in-band quantization noise is therefore greatly decreased in comparison to that of the embedded quantizer. The output of the $\Sigma\Delta\text{M}$ is a B -bit digital stream at f_s sampling rate.
- **Decimator.** It reduces the rate of the $\Sigma\Delta\text{M}$ output stream down to the Nyquist rate. Jointly, the word length increases from B to N in order to preserve resolution as the word rate decreases. Although the block scheme of a decimator may differ in practice from that illustrated in Fig. 2.5, it conceptually consists of a high-selectivity digital filter and a downsampler. Frequency components of the stream

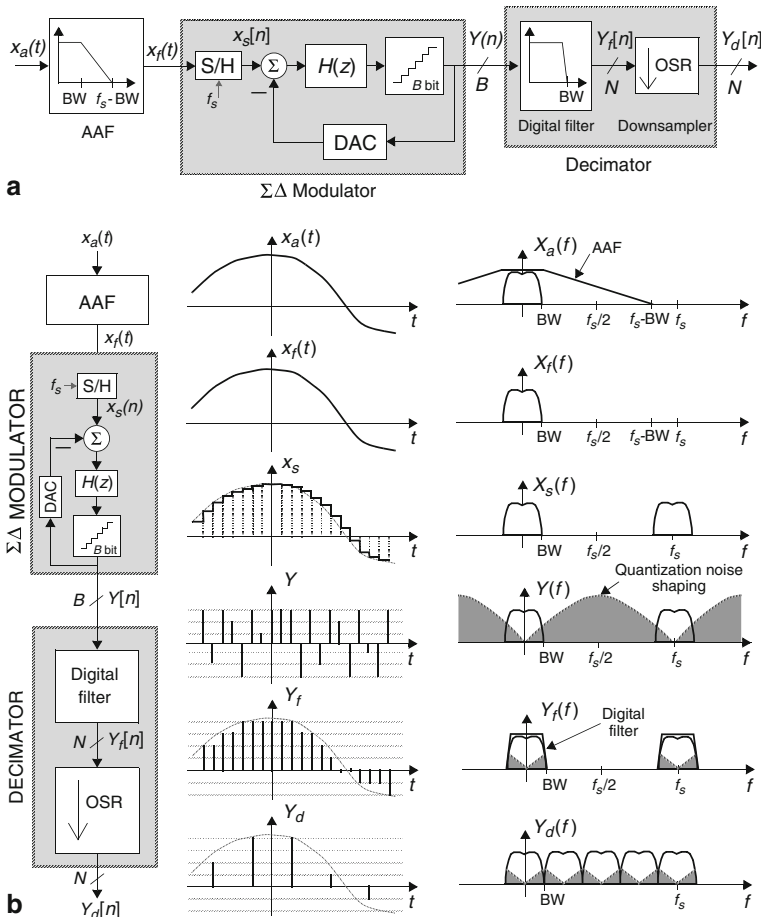


Fig. 2.5 $\Sigma\Delta$ ADC: **a** generic scheme, **b** signal processing involved

above BW are removed¹—and, therefore, most part of the shaped quantization error—to avoid aliasing during the subsequent downsampling, in which the stream rate is divided by OSR keeping only one out of every OSR samples.

The $\Sigma\Delta$ modulator is the block that has most influence upon the ADC performance, basically because it is the responsible of the sampling and quantization processes and, therefore, ultimately limits the accuracy of the A-to-D conversion.

¹ A large steepness in the transition band is demanded usually for the filter in order to avoid degrading the signal band. However, this specification is imposed on a digital filter and is a priori easier to fulfill than for the analog AAF of a Nyquist ADC.

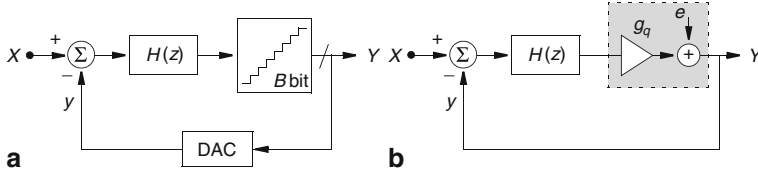


Fig. 2.6 $\Sigma\Delta$ M architecture: **a** basic scheme, **b** linear model

2.2.1 Signal Processing in a $\Sigma\Delta$ Modulator

Figure 2.6a shows the basic scheme of a $\Sigma\Delta$ modulator. It consists of a feed-forward path formed by a loop filter $H(z)$ and a B -bit quantizer and a negative feedback path around them using a B -bit DAC [Inos62]. The operation of the $\Sigma\Delta$ M can be explained as follows. Assume that $H(z)$ exhibits large gain inside the signal band and small gain outside of it. Due to the negative feedback, the error signal $x - y$ will become practically null in the signal band; i.e., the input signal x and the analog version of the output y will practically coincide within this band. Most of the differences between x and y will therefore be placed at higher frequencies, shaping quantization error and pushing it outside the signal band.

Figure 2.6b shows the linear model of a $\Sigma\Delta$ M, in which the DAC is assumed to be ideal, the quantizer is replaced by the model in Fig. 2.2b and the additive white noise approximation is considered for the quantization error. This way, the modulator can be viewed as a two-input system whose output is represented in Z -domain as

$$Y(z) = \text{STF}(z)X(z) + \text{NTF}(z)E(z) \quad (2.13)$$

where $X(z)$ and $E(z)$ are the Z -transform of the input signal and the quantization noise, respectively, and $\text{STF}(z)$ and $\text{NTF}(z)$ are the respective transfer functions, given by

$$\text{STF}(z) = \frac{g_q H(z)}{1 + g_q H(z)} \quad \text{NTF}(z) = \frac{1}{1 + g_q H(z)} \quad (2.14)$$

Since the signal and the noise pass through different transfer functions, $H(z)$ can be chosen so that the noise shaping does not affect the signal. Using a loop filter with large gain within the signal band, the signal and noise transfer functions can be approximated in that range to

$$\text{STF}(z) \approx 1 \quad \text{NTF}(z) \approx \frac{1}{g_q H(z)} \ll 1 \quad (2.15)$$

The noise-shaping functions in (2.10) can be built with proper selection of $H(z)$. The easiest loop filter that exhibits the desired frequency performance is an integrator, whose Z -domain transfer function is

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \quad (2.16)$$

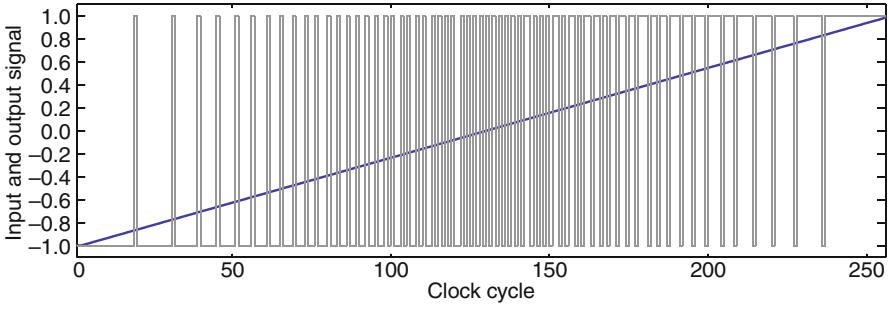


Fig. 2.7 PDM output stream of a 1st-order $\Sigma\Delta$ M for an input ramp

Assuming that the quantizer gain g_q equals unity, the $\Sigma\Delta$ M output yields

$$Y(z) = z^{-1}X(z) + (1 - z^{-1})E(z) \quad (2.17)$$

and the modulator is called a 1st-order $\Sigma\Delta$ M, referring to the order of the noise shaping.

Figure 2.7 shows the output of a 1st-order $\Sigma\Delta$ M with a 1-bit embedded quantizer for a ramp input signal. Due to the combined action of oversampling and negative feedback, the $\Sigma\Delta$ M output is a pulse-density modulated (PDM) signal that locally tracks the input on average: when the input level is low, the $\Sigma\Delta$ M output contains more -1 's than $+1$'s; when it is high, the $+1$'s are dominant; and when the input signal is close to zero, the density of $+1$'s and -1 's practically coincides. If the quantizer resolution is larger, the output tracks the input much closer, since the separation between the discrete levels decreases.

2.2.2 Performance Metrics of $\Sigma\Delta$ Modulators

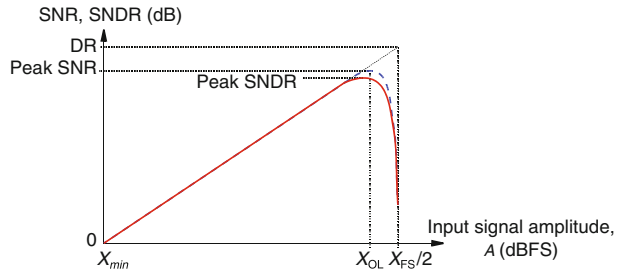
For the sake of clarity, it is convenient at this time to define the most important parameters commonly used to quantify the performance of $\Sigma\Delta$ modulators; namely:

- **Signal-to-noise ratio, SNR.** It is the ratio of the output power at the frequency of an input sinusoid to the uncorrelated in-band error power. Due to non-idealities of the circuitry that implements the modulator, other (linear and non-linear) errors apart from quantization noise contribute to the in-band error. SNR accounts for the linear performance of the modulator the in-band power associated to harmonics is therefore not included. For an ideal $\Sigma\Delta$ modulator and taking only quantization error into account, the SNR can be approximated to

$$\text{SNR}\Big|_{\text{dB}} = 10\log_{10}\left(\frac{A^2}{2P_Q}\right) \quad (2.18)$$

where A is the amplitude of the output sinusoid.

Fig. 2.8 Illustration of the performance parameters of a $\Sigma\Delta$ M on a typical SNR/SNDR curve



- **Signal-to-(noise + distortion) ratio, SNDR.** It is defined as the ratio of the output power at the frequency of an input sinusoid to the total in-band error power, taking therefore into account harmonics at the modulator output.
- **Dynamic range, DR.** It is defined as the ratio of the output power at the frequency of an input sinusoid with maximum amplitude to the output power for a small input for which $\text{SNR} = 0$ dB (i.e., so that it cannot be distinguished from the error). Ideally, a sinusoid with maximum amplitude at the modulator input $X_{\text{FS}}/2$ will provide an output sinusoid sweeping the full scale of the $\Sigma\Delta$ quantizer and hence

$$\text{DR}_{\text{dB}} = 10 \log_{10} \left[\frac{(X_{\text{FS}}/2)^2}{2P_Q} \right] \quad (2.19)$$

- **Effective number of bits, ENOB.** Since the DR of an ideal N -bit Nyquist ADC equals $6.02N + 1.76$, a similar relationship is established for $\Sigma\Delta$ ADCs for comparison purposes [Bose88]:

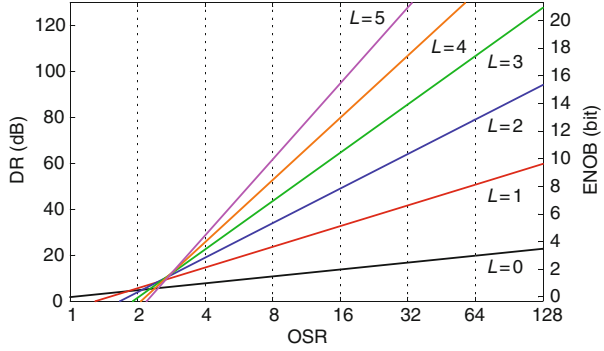
$$\text{ENOB} = \frac{\text{DR}_{\text{dB}} - 1.76}{6.02} \quad (2.20)$$

where the ENOB represents the number of bits needed for an ideal Nyquist ADC to achieve the same DR as the $\Sigma\Delta$ converter.

- **Overload level, X_{OL} .** The SNR of a $\Sigma\Delta$ does not increase monotonously for input amplitudes in the range $[0, X_{\text{FS}}/2]$, where X_{FS} stands for the quantizer full scale. In practice, the embedded quantizer overloads for large amplitudes close to $X_{\text{FS}}/2$, causing an increase in the in-band error and a sharp drop in the SNR. The maximum value of the SNR before that drop defines the peak SNR and the corresponding input level is defined as the overload level X_{OL} of the $\Sigma\Delta$.

The performance parameters defined above are illustrated in Fig. 2.8 on typical SNR and SNDR curves of a $\Sigma\Delta$ as a function of the amplitude of an input sinusoid signal. Both curves usually coincide up to medium input levels, since distortion is submerged into the modulator noise floor. For large input levels, harmonic distortion becomes more evident and degrades performance, causing the deviation of the SNDR curve.

Fig. 2.9 DR and ENOB versus OSR for different orders of an ideal $\Sigma\Delta$ modulator. (A single-bit embedded quantizer is assumed)



2.2.3 Ideal Performance of $\Sigma\Delta$ Modulators

The dynamic range of an ideal L th-order $\Sigma\Delta$ M with a B -bit embedded quantizer that operates at a given oversampling ratio can be obtained from (2.1), (2.12) and (2.19) as

$$\text{DR} \approx 10\log_{10} \left(\frac{3}{2} (2^B - 1)^2 \cdot \frac{(2L + 1)\text{OSR}^{(2L+1)}}{\pi^{2L}} \right) \quad (2.21)$$

and can therefore be increased if L , OSR and/or B are augmented. The pros and cons of each possibility are discussed below:

- **Increasing the modulator order L** considerably improves the performance of a $\Sigma\Delta$ M, since quantization error will be more attenuated at low frequencies and pushed to high frequencies. For a given OSR, the rise in DR when increasing L in one leads from (2.21) to

$$\Delta\text{DR}_{\text{dB}} \approx 10\log_{10} \left[\frac{2L + 3}{2L + 1} \cdot \left(\frac{\text{OSR}}{\pi} \right)^2 \right] \quad (2.22)$$

This means, e.g., that the DR of a 4th-order $\Sigma\Delta$ M with OSR = 32 is improved in 21.3 dB (3.5 bit) in comparison to a 3rd-order one. However, stability problems arise from using high-order shapings ($L > 3$). These problems can be circumvented using different techniques [Nors97a], but at the price of reducing DR in comparison to the ideal value given by (2.21).

- **Increasing the oversampling ratio OSR** leads, according to (2.21), to an increase in the dynamic range of $3(2L + 1)$ dB/octave for an ideal L th-order $\Sigma\Delta$ M. This is shown in Fig. 2.9, where DR and ENOB are plotted versus OSR for different modulator orders. Note that for OSR > 4, the combined action of oversampling and noise shaping considerably improves performance. However, for a given signal band, larger OSRs lead to higher sampling frequencies and a faster operation of the modulator internal circuitry. The latter, if achievable, penalizes power dissipation.

- **Increasing the resolution B of the modulator embedded quantizer** leads to an increase in the DR of approximately 6 dB (1 bit) per extra bit in the quantizer [Geer02]. However, $\Sigma\Delta$ Ms with an internal multi-bit quantizer require a multi-bit DAC in the feedback loop that—contrary to a single-bit one, with only two levels—is not inherently linear. Note from Fig. 2.6 that non-linearities in the multi-bit DAC will be directly added to the modulator input. The linearity required in the DAC equals therefore in practice that wanted for the $\Sigma\Delta$ modulator.

2.3 Classification of $\Sigma\Delta$ Modulators

The above-mentioned strategies can be combined in many different ways giving rise to a pleiad of $\Sigma\Delta$ M topologies reported in literature and that can be grouped attending to different classification criteria [Rodr03]:

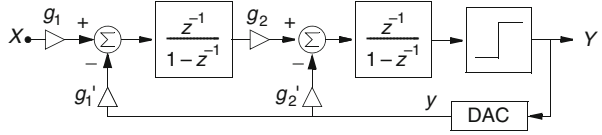
- The nature of the signals being converted: low-pass versus band-pass $\Sigma\Delta$ Ms.
- The type of dynamics of the loop filter. Historically, most of $\Sigma\Delta$ modulators employed only discrete-time (DT) loop filters, as has been assumed in previous sections. Per contra, continuous-time (CT) $\Sigma\Delta$ modulators are also actually implemented. They use CT loop filters but DT quantizers. Also hybrid CT-DT modulators have been more recently reported.
- The number of quantizers employed. $\Sigma\Delta$ Ms employing only one quantizer are called single-loop structures. Those employing several quantizers have different names: cascades, dual-quantizer $\Sigma\Delta$ Ms, etc.
- The number of bits in the embedded quantizer. Historically, $\Sigma\Delta$ modulators employed mostly single-bit quantizers because they are inherently linear. Today, multi-bit $\Sigma\Delta$ Ms—i.e., those using multi-bit quantizers—are widely spread.
- Type of circuitry employed, devices available in the fabrication process, voltage supply, etc. Most of the reported DT implementations employ switched-capacitor (SC) circuits with high-quality passive capacitors—mixed-signal technology options—, but others employ capacitors available on standard CMOS technologies, active capacitors built with MOS transistors, switched-current (SI) circuits, etc.

Describing all possible $\Sigma\Delta$ M architectures derived from previous classification criteria goes beyond the scope of this book. A detailed study of them can be found in many papers and books [Nors97a, Mede99, Geer02, Schr05b, Ortm06]. Instead, we will hereafter focus on low-pass DT $\Sigma\Delta$ architectures implemented using SC techniques, field to which the main contributions of this book refer to.

2.4 Single-Loop $\Sigma\Delta$ Architectures

The fundamentals and ideal performance of generic DT $\Sigma\Delta$ Ms have already been introduced. The 1st-order $\Sigma\Delta$ M has been also presented in Sect. 2.2.1 for illustrating the operation of the simplest $\Sigma\Delta$ topology. However, its scope of application is very

Fig. 2.10 Second-order $\Sigma\Delta$ modulator



limited in practice due to the high correlation between the quantization error and the input signal, which severely deviates from the white noise approximation and leads to non-linear dynamic phenomena. This section is dedicated to modulator topologies with a larger number of integrators and only one quantizer, called single-loop $\Sigma\Delta$ Ms. Their linear performance—thanks to a better decorrelation of quantization error—will be discussed together with some aspects that are not covered by the white noise approximation, such as instabilities. A single-bit embedded quantizer will be assumed.

2.4.1 Second-Order $\Sigma\Delta$ Modulator

We will start the review of basic modulator architectures with the single-loop second-order topology [Cand85] shown in Fig. 2.10. Assuming a linear model for the single-bit quantizer (comparator)² as in Fig. 2.6b, the modulator output in Z-domain yields

$$Y(z) = \frac{g_1 g_2 g_q z^{-2} X(z) + (1 - z^{-1})^2 E(z)}{1 + (g_2' g_q - 2)z^{-1} + (1 + g_1' g_2 g_q - g_2' g_q)z^{-2}} \quad (2.23)$$

so that the following conditions must be fulfilled for a pure 2nd-order NTF:

$$\left. \begin{aligned} g_1' g_2 g_q &= 1 \\ g_2' &= 2g_1' g_2 \end{aligned} \right\} \Rightarrow Y(z) = \frac{g_1}{g_1'} \cdot z^{-2} X(z) + (1 - z^{-1})^2 E(z) \quad (2.24)$$

Note that the modulator can provide signal gain $G = g_1/g_1'$ if different coefficients are used in the first integrator for the input and the feedback paths. However, usually $g_1 \equiv g_1'$ and it will be so assumed hereinafter.

The conditions shown in (2.24) for pure 2nd-order shaping define relationships between the integrator coefficients, but not the values themselves. In general, selecting the coefficients of a $\Sigma\Delta$ M involves solving several trade-offs between architectural, circuit and technological aspects of the practical implementation; namely:

- Keeping the integrator outputs bounded to ensure the modulator stability³.
- Maximizing the overload level X_{OL} of the $\Sigma\Delta$ M to ensure a high peak SNR.

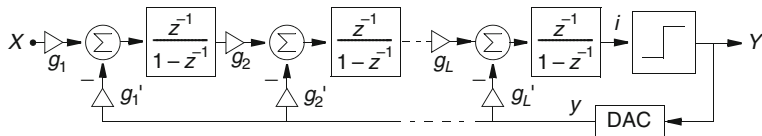
² Note that a two-level quantizer has no inherent gain and, therefore, there is no obvious value for g_q . The linear model used henceforth for comparators assumes that g_q is such that the product of the loop gain factors is forced to unity by the feedback loop [Will91]. This model for comparators in a $\Sigma\Delta$ M is empirical, but its results usually compare well to computer simulations using the true non-linear quantizer function.

³ A $\Sigma\Delta$ M is considered stable if, for bounded inputs and whatever integrators initial conditions, the internal state variables (integrator outputs) remain also bounded over time.

Table 2.1 Some reported coefficients for 2nd-order single-bit $\Sigma\Delta$ Ms

Coefficients	[Bose88]	[Yin94]	[Mede99]	[Marq98]
g_1, g'_1	0.5, 0.5	0.25, 0.25	0.25, 0.25	1/3, 1/3
g_2, g'_2	0.5, 0.5	0.5, 0.25	1, 0.5	0.6, 0.4
OS/Δ (at -4 dBFS)	1.50	0.75	1.25	1.00
Unitary capacitors ^a	6 (3 + 3)	10 (4 + 6)	9 (5 + 4)	20 (4 + 16)

^a Capacitor sharing between coefficients in an integrator is assumed

**Fig. 2.11** L th-order single-loop $\Sigma\Delta$ M with distributed feedback

- Minimizing the required signal swing at the integrator outputs; i.e., the integrator output swing (OS) specifications should be, first, achievable in practice, and second, as low as possible in order to reduce power consumption and to ease circuit design.
- Simplifying the implementation of the coefficients. In SC $\Sigma\Delta$ Ms they are implemented as capacitor ratios using unitary elements for improved matching. A set of coefficients involving a small number of unitary capacitors leads to a saving of silicon area.

For comparison purposes, Table 2.1 shows some sets of coefficients for 2nd-order $\Sigma\Delta$ reported in literature. All of them exhibit an overload level $X_{OL} \approx -4$ dBFS (-4 dB below the full-scale amplitude, $\Delta/2$). The requirements on integrator OS and the total number of unitary capacitors are also included.

2.4.2 High-Order $\Sigma\Delta$ Modulators

The simplest way to extend a $\Sigma\Delta$ M to an arbitrary L th-order shaping consists of including L integrators before the quantizer [Ritc77]. Extending the 2nd-order $\Sigma\Delta$ M in Fig. 2.10, the architecture in Fig. 2.11 can be obtained, often called L th-order single-loop $\Sigma\Delta$ M with distributed feedback. Using a linear model, its output would be given by

$$Y(z) = z^{-L}X(z) + (1 - z^{-1})^L E(z) \quad (2.25)$$

if a set of relationships between the coefficients were fulfilled, similar to those in (2.24) for 2nd-order $\Sigma\Delta$ Ms. The in-band quantization error for this NTF would ideally be given by (2.12), therefore achieving very low P_Q for large L , even for low OSR.

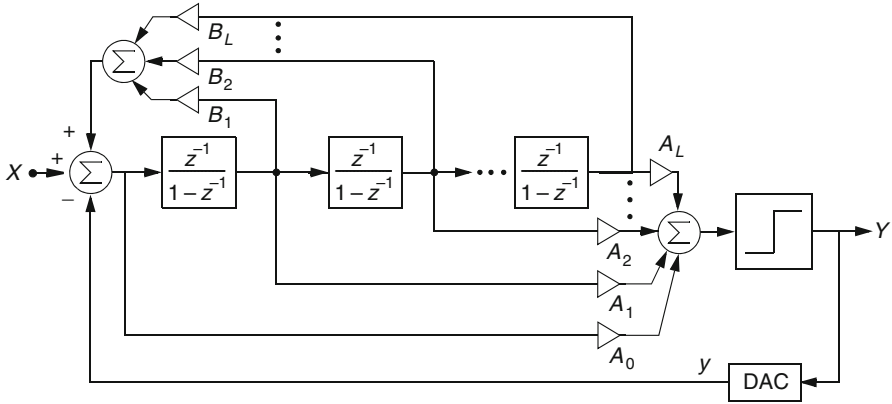


Fig. 2.12 Lee-Sodini L th-order $\Sigma\Delta$ modulator

Nevertheless, this performance is not achievable in practice because $\Sigma\Delta$ Ms with pure differentiator NTFs—i.e., FIR filters like $(1 - z^{-1})^L$ —are prone to instability⁴ if $L > 2$, exhibiting unbounded states and poor SNR in comparison to that predicted by the linear model. This tendency to instability can be qualitatively explained as follows [Adam97]. For a $\Sigma\Delta$ M to be stable, the input to the quantizer must not be allowed to become too large. Since the quantizer input in Z -domain is given by

$$I(z) = \text{STF}(z)X(z) + [\text{NTF}(z) - 1]E(z) \quad (2.26)$$

the gain of $\text{NTF}(z) - 1$, or simply $\text{NTF}(z)$, must not be too large. However, the gain of NTFs of the form $(1 - z^{-1})^L$ rapidly increases for high frequencies if $L > 2$, having a maximum $\|\text{NTF}\|_{\infty} = \max[\text{NTF}(z)] = 2^L$ at $z = -1$ ($f = f_s/2$).

The determination of exact conditions to guarantee the stability of higher-order ($L \geq 3$) $\Sigma\Delta$ Ms is still an open question, but, despite the absence of general stability conditions, high-order $\Sigma\Delta$ Ms have been successfully designed since the late 1980s.

2.4.2.1 Optimization of NTFs

Figure 2.12 illustrates the so-called interpolative architecture introduced by Lee and Sodini [Lee87], which allows to obtain high-order stable designs. Thanks to the large set of coefficients in multiple feed-forward and feedback paths, more complex high-pass NTFs can be built with sufficiently low gain at the high-frequency region. Let us first consider that the feedback coefficients B_i are set to zero. In this case, the following IIR NTF is obtained

$$\text{NTF}(z) = \frac{(z - 1)^L}{D(z)} \quad (2.27)$$

⁴ Instability appears at the modulator output as a large-amplitude low-frequency oscillation, leading to long strings of alternating +1's and -1's [Adam97].

where all zeros are located at $z = 1$ (DC) and $D(z)$ is a polynomial determined by the feed-forward coefficients A_i . These coefficients can be adjusted to build a high-pass Butterworth or Chebyshev filter for NTF, with cutoff frequency beyond the signal band and almost flat gain in the filter pass-band. For the $\Sigma\Delta$ M to remain stable, this gain must be adjusted to satisfy⁵

$$\|NTF\|_{\infty} = \max[NTF(z)] \sim 1.5 \quad (2.28)$$

However, with all zeros at DC, NTF rises monotonically in the signal band like an L th-order function and its gain at the end of the signal band will therefore practically determine the total in-band error power. The feedback coefficients B_i in the Lee-Sodini $\Sigma\Delta$ M are in practice used to modify the NTF in (2.27) and place notches in the signal band for further shaping of the quantization error. The values of B_i are set small in comparison to A_i in order to obtain NTF poles mostly controlled by the latter coefficients, preserving therefore the flat out-of-band gain of NTF and the modulator stability.

In [Lee87] the zeros are fixed to obtain an NTF with equal-ripple response over the signal band, but other alternatives are also feasible. The approach followed in [Schr93] leads to minimizing P_Q by optimally placing complex-conjugate zeros of NTF at frequency positions obtained by solving the minimization problem

$$\min \left[\int_0^{\text{BW}} |NTF(f)|^2 df \right] \Rightarrow \min \left[\begin{cases} \int_0^1 \prod_{i=1}^{L/2} (f^2 - f_{z_i}^2)^2 df, & L \text{ even} \\ \int_0^1 f^2 \prod_{i=1}^{(L-1)/2} (f^2 - f_{z_i}^2)^2 df, & L \text{ odd} \end{cases} \right] \quad (2.29)$$

where f_{z_i} represents the location of the complex zeros normalized to BW. The solutions for (2.29) up to $L=8$ can be found in [Schr93]. Fig. 2.13 illustrates these alternatives for implementing stable high-order NTFs for a 5th-order $\Sigma\Delta$ M with $\text{OSR} = 64$. Note from Fig. 2.13a that the position of the poles has been fixed following a Butterworth configuration, so that the cutoff frequency of NTF is beyond the signal band—vertical line in Fig. 2.13b—and its out-of-band gain is 1.5 (3.5 dB). Two notches can be introduced in the signal band by slightly moving four of the zeros at $z = 1$ along the unit circle. As shown in Fig. 2.13b, this considerably reduces $|NTF(f)|$ at the upper edge of the signal band and therefore the final P_Q (18-dB reduction for the case considered).

Figure 2.14 shows the maximum SNR achievable by single-loop $\Sigma\Delta$ Ms of various orders following these two approaches. The ideal performance for pure differentiator

⁵ This empirical condition is commonly accepted as a rule of thumb for designing stable high-order NTFs [Adam97]. Other stability criteria also exist, but none of them can ensure the stability of an arbitrary high-order single-bit $\Sigma\Delta$ M [Schr93]. Computer simulation is still the most reliable method to verify stability.

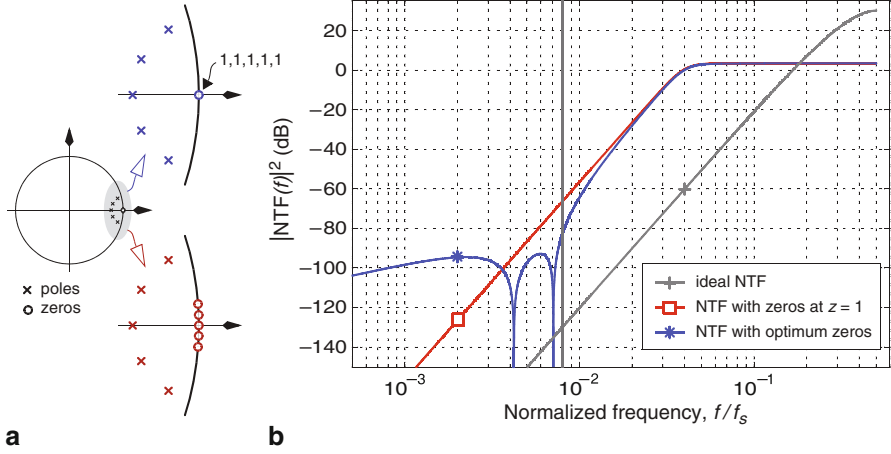


Fig. 2.13 Comparison of different implementations of a 5th-order NTF: **a** illustration of the pole-zero placement in the unit circle, **b** magnitude response of the corresponding NTFs. (OSR = 64 is assumed for the design of the Butterworth poles)

NTFs is included for comparison purposes in this figure as in Fig. 2.13, indicating that the SNRs obtained in practice are much lower. Nevertheless, as shown in Fig. 2.14b, the optimal placement of notches within the signal band can lead to stable high-order $\Sigma\Delta$ Ms with high SNR at moderate OSR.

2.4.2.2 Common High-Order Topologies

Many $\Sigma\Delta$ modulator topologies have been developed for implementing high-order NTFs with the characteristics of most common IIR filter families. Most of them make use of multiple weighted feedback or feed-forward paths (or both) [Adam97]. The two architectures of 5th-order $\Sigma\Delta$ Ms illustrated in Fig. 2.15 can be considered typical examples.

Figure 2.15a illustrates a topology with feed-forward summation of the integrator outputs before the quantizer. If $\gamma_1 = 0$, a high-pass NTF is obtained with all zeros at DC. By adding small negative feedback locally, around pairs of integrators in the loop filter, pairs of zeros can be moved along the unit circle to create notches in $|NTF(f)|$ at frequencies $2\pi f_i/f_s \approx \sqrt{\gamma_i}$. Once NTF is set for the desired noise shaping, the modulator topology fixes the signal transfer function to $STF(z) = 1 - NTF(z)$.

The architecture in Fig. 2.15b that includes distributed feedback and feed-forward paths can be used if a certain degree of freedom is desired in designing both NTF and STF. In this topology, the zeros of STF can be fixed with coefficients b_i without affecting the pole placement. Local resonator feedbacks can be also included to set notches in $|NTF(f)|$.

A common drawback of most high-order topologies implementing IIR filters for NTF is the increased circuit complexity due to the many feedback and feed-forward

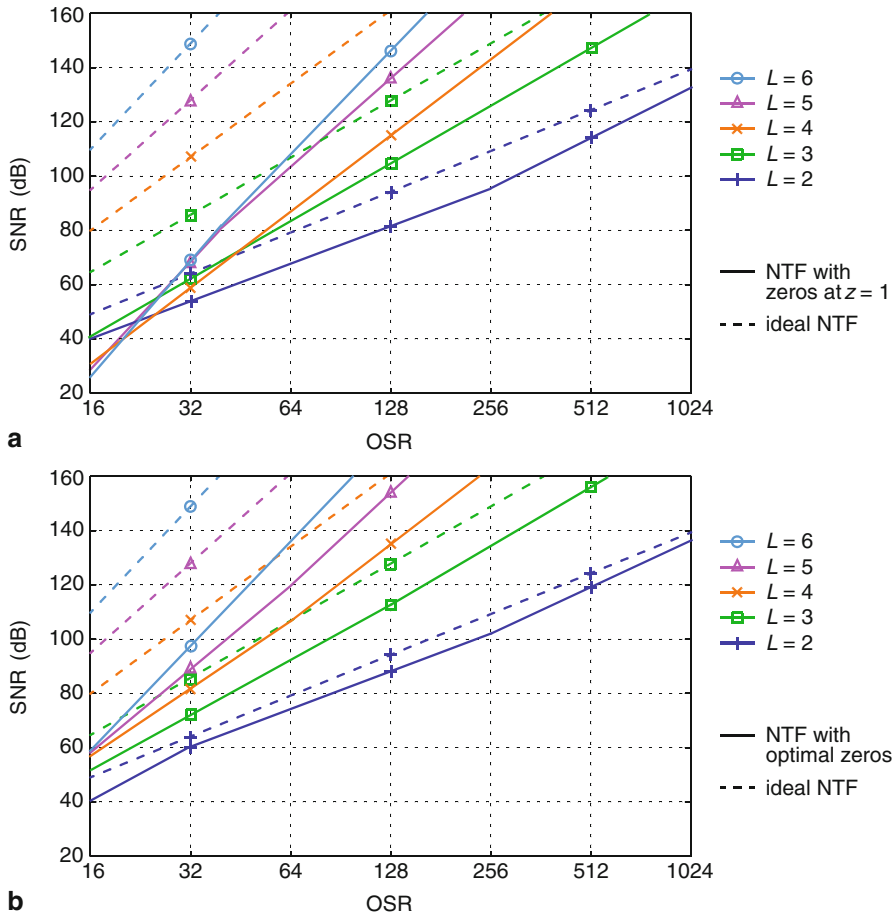


Fig. 2.14 Maximum SNR achievable by L th-order single-loop $\Sigma\Delta$ Ms versus OSR: **a** NTF with all zeros at $z = 1$, **b** NTF with zeros optimally spread over the signal band. (Data taken from [Schr93])

coefficients required. Furthermore, some of the coefficients may be considerably small [Chao90, Kuo99], leading to large capacitor ratios in the final SC implementation and, therefore, to great power consumption and area occupation.

Another strategy for designing stable high-order single-loop $\Sigma\Delta$ Ms basically consists of extensively exploring the design space by means of behavioral simulations searching for sets of coefficients that are easy to implement and maximize the modulator SNR, while achieving a reasonable overload level. This procedure is followed in [OptE90] and [Marq98] for the distributed feedback topology in Fig. 2.11. High-pass IIR NTFs are obtained with zeros at DC, but the obtained placement of poles does not follow any particular filter family configuration.

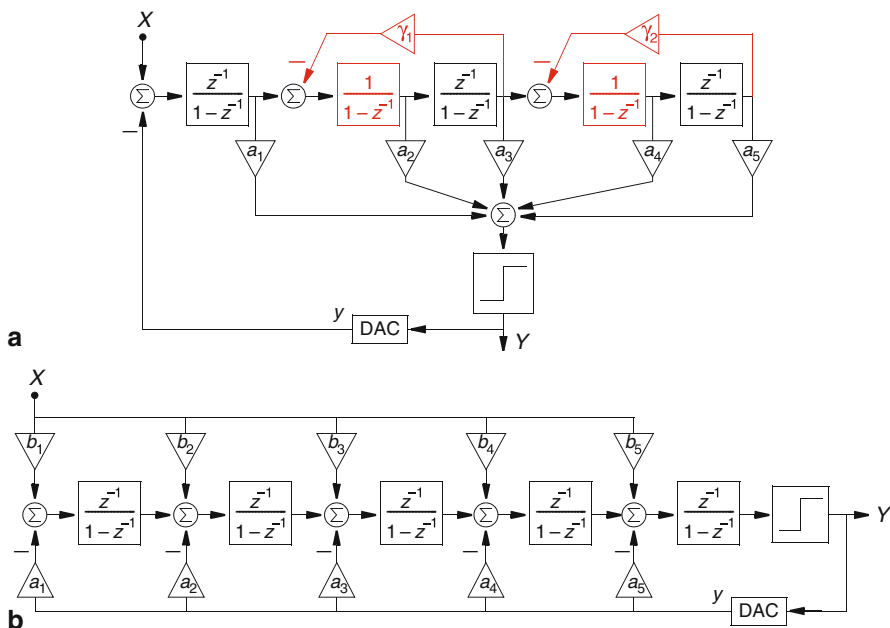


Fig. 2.15 5th-order single-loop $\Sigma\Delta$ Ms: **a** with feed-forward summation and local resonator feedbacks, **b** with distributed feedback and distributed feed-forward input paths

2.4.2.3 Non-linear Stabilization Techniques

In spite of the procedures described above, the resulting high-order single-loops are only conditionally stable and instabilities may therefore appear for inputs above certain bounds or for certain initial conditions. Non-linear techniques can be used to ensure global stable operation. Limiters can be included at the integrators in order to preclude their outputs from exceeding ‘safe’ values of the state-variable, previously identified during stable operation of the modulator [OptE90]. Instead of clipping the integrators outputs, they can be reset to zero or some other initial condition at power-up of the system or when unstable operation is detected [Mous94]. The detection of instability can be done at the integrator level, by placing comparators to determine if a state variable has surpassed a certain limit, or by monitoring the length of consecutive +1’s and −1’s at the modulator output.

2.5 Cascade $\Sigma\Delta$ Architectures

As stated in the previous section and illustrated in Fig. 2.14, precluding instabilities in high-order single-loop $\Sigma\Delta$ modulators leads to attainable values of SNR that are considerably far from ideal. An alternative to circumvent instabilities while obtaining

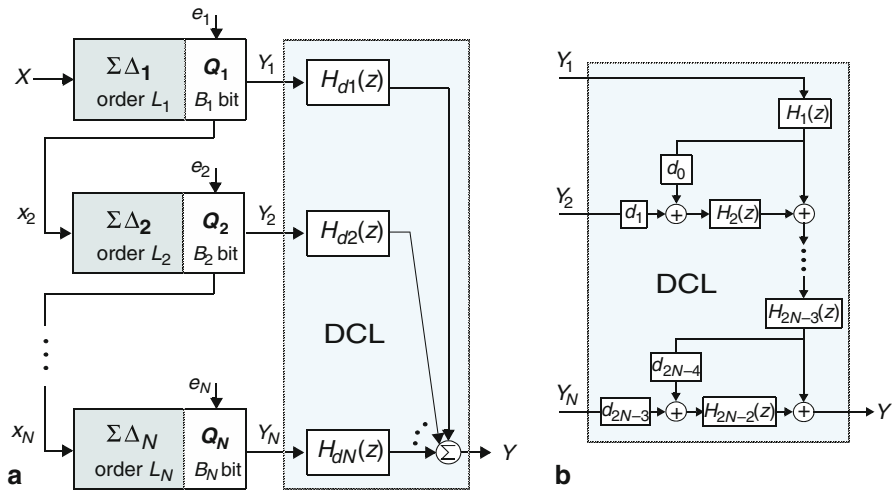


Fig. 2.16 **a** Generic N -stage cascade $\Sigma\Delta$ modulator; **b** structure of the digital cancellation logic (DCL)

high SNR can be found in the so-called MASH (multi-stage noise-shaping) $\Sigma\Delta$ Ms, often referred to as cascade $\Sigma\Delta$ Ms [Mats87, Long88]. The generic scheme of a cascade $\Sigma\Delta$ M is shown in Fig. 2.16. It consists of several low-order $\Sigma\Delta$ modulators or stages, in which each stage remodulates a signal containing the quantization error generated in the previous one. The outputs Y_i of the low-order $\Sigma\Delta$ Ms are properly processed and combined in digital domain in order to cancel out in the overall cascade output Y the traces of the quantization errors of all the stages, but the last one. This latter error appears at the overall modulator output, after the digital cancellation logic (DCL block in Fig. 2.16), with a noise shaping of order equal to the summation of the orders of the cascaded stages. Furthermore, since all feedback loops are local and there is no inter-stage feedback, unconditionally-stable high-order shaping can be obtained provided that only $\Sigma\Delta$ Ms with order $L \leq 3$ are cascaded. The performance of a MASH $\Sigma\Delta$ M is, therefore, similar to that of an ideal high-order loop with no stability issues.

For an N -stage cascade $\Sigma\Delta$ M like that in Fig. 2.16, by means of adequately processing in digital domain the stages outputs, only the modulator input signal $X(z)$ and the last-stage quantization error $E_N(z)$ remain in the overall modulator output, yielding

$$Y(z) = \text{STF}(z)X(z) + \text{NTF}_N(z)E_N(z) = z^{-L}X(z) + d_{2N-3}(1 - z^{-1})^L E_N(z) \quad (2.30)$$

where $L = L_1 + L_2 + \dots + L_N$ and d_{2N-3} is a scaling factor related to the transmission of the input signal from one stage to another (in order to avoid overloading of the stages) that amplifies the last-stage quantization error. The in-band quantization error

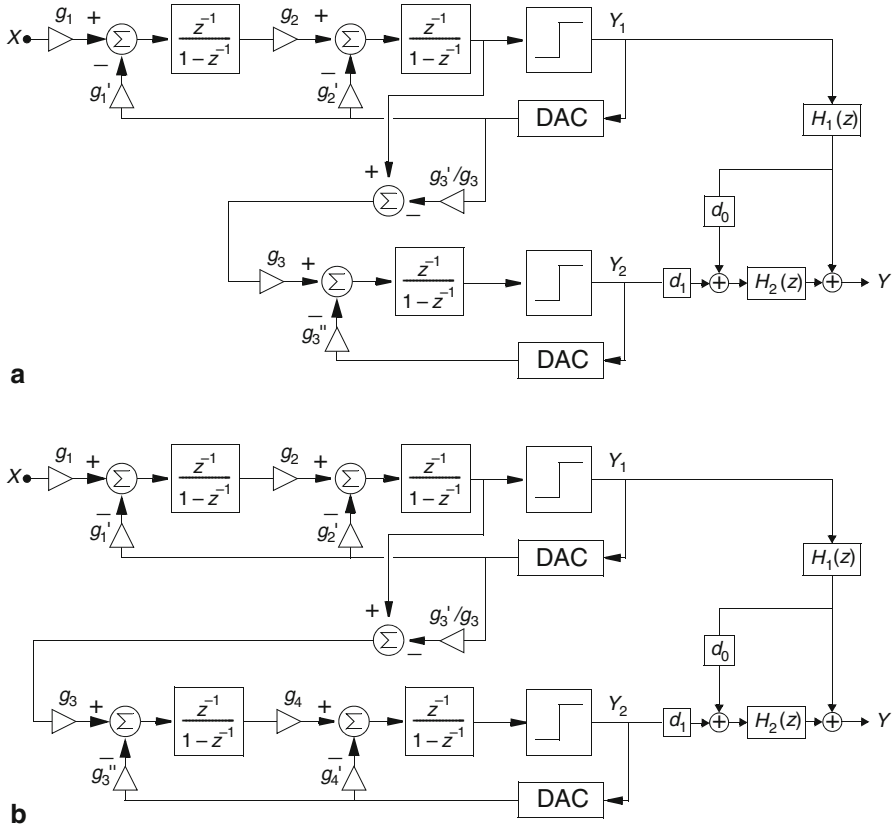


Fig. 2.17 Illustration of 2-stage cascade topologies: **a** 3rd-order 2-1 $\Sigma\Delta$ M, **b** 4th-order 2-2 $\Sigma\Delta$ M

power of a N -stage cascade is, therefore, given by

$$P_Q \cong d_{2N-3}^2 \cdot \frac{\Delta_N^2}{12} \cdot \frac{\pi^{2L}}{(2L+1)\text{OSR}^{(2L+1)}} \quad (2.31)$$

with Δ_N being the level spacing in the B_N -bit quantizer of the N th modulator stage. Hence, the performance corresponds to that of an ideal L th-order B_N -bit $\Sigma\Delta$ M, except for the scalar d_{2N-3} that causes a systematic loss of performance. Common values for this amplifying factor are 2 and 4, which lead to a reduction in the attainable SNR of 6 dB (1 bit) and 12 dB (2 bit), respectively. These performance losses that are inherent to cascade $\Sigma\Delta$ Ms, are however considerably lower than those resulting for optimized high-order single loops (see Fig. 2.14). Moreover, in the case of MASH $\Sigma\Delta$ Ms, they are independent of OSR.

The aforementioned benefits have favoured the development of a great number of cascade $\Sigma\Delta$ topologies. Fig. 2.17 illustrates two different 2-stage cascades. A 3rd-order $\Sigma\Delta$ M that is formed by a 2nd-order stage followed by a 1st-order stage

Table 2.2 Relationships to be fulfilled for digital cancellation in the 2-1 and the 2-2 $\Sigma\Delta$ Ms

2-1 $\Sigma\Delta$ M			2-2 $\Sigma\Delta$ M		
Analog	Digital		Analog	Digital	
$g'_2 = 2g'_1g_2$	$d_0 = \frac{g'_3}{g'_1g_2g_3} - 1$	$H_1(z) = z^{-1}$	$g'_2 = 2g'_1g_2$	$d_0 = \frac{g'_3}{g'_1g_2g_3} - 1$	$H_1(z) = z^{-2}$
	$d_1 = \frac{g''_3}{g'_1g_2g_3}$	$H_2(z) = (1-z^{-1})^2$	$g'_4 = 2g''_3g_4$	$d_1 = \frac{g''_3}{g'_1g_2g_3}$	$H_2(z) = (1-z^{-1})^2$

[Long88] (so-called 2-1 $\Sigma\Delta$ M) is shown in Fig. 2.17a, whereas a 4th-order cascade built up by two 2nd-order stages [Kare90, Bahe92] (2-2 $\Sigma\Delta$ M) is shown in Fig. 2.17b. Table 2.2 also summarizes the relationships between integrator coefficients (analog coefficients) and the blocks in the digital cancellation logic (digital coefficients and filters) that must be fulfilled for a proper performance of the cascades in accordance with (2.30).

Many other MASH topologies have been proposed: a 4th-order 3-stage cascade (2-1-1 $\Sigma\Delta$ M) [Yin94], 5th-order cascades implemented as a 2-2-1 $\Sigma\Delta$ M [Vleu01] or as a 2-1-1-1 $\Sigma\Delta$ M [Rio00], a 6th-order 2-2-2 architecture [Dedi94], etc. Although cascade $\Sigma\Delta$ Ms can a priori be extended to whatever number of stages to increase the order of the noise shaping, non-idealities in the analog circuitry of the final implementation (e.g., mismatching) limit in practice the effectiveness of this approach. They cause deviations in the performance of the analog section of the cascade (integrator coefficients and transfer functions) which impede that the DCL completely cancels the NTFs of the first $N - 1$ stages at the overall output. This effect, called noise leakage, causes traces of the lower-order quantization errors to appear at the cascade output, where they can dominate the overall in-band error power⁶ [Rebe97].

Note that the relationships in Table 2.2 for the digital section of the cascade assume that the digital cancellation logic is structured according to Fig. 2.16b. For other kind of structures, the equations may be different, but anyhow the digital part will have to fulfil given relationships with the analog one. Once these relationships are fulfilled, the value of the free analog coefficients is fixed for minimum loss of performance and simpler circuit implementation of the cascade $\Sigma\Delta$ M. Proper concerns for their selection are:

- Minimizing the loss of performance in comparison to an ideal $\Sigma\Delta$ M.
- Maximizing the modulator overload level to ensure high peak SNR.
- Minimizing the integrators output swing, especially in low-voltage scenarios.
- Easing the implementation of the analog coefficients as capacitor ratios with unitary elements by means of enabling capacitor sharing within the integrators, reducing the total number of unitary capacitors to save silicon area, etc.

⁶ Obviously, non-idealities in the analog circuitry also affect the practical performance of single-loop $\Sigma\Delta$ Ms. Nevertheless, their operation does not rely in the cancellation of quantization errors and their sensitivity to noise leakages is therefore much lower than for cascade $\Sigma\Delta$ Ms.

Table 2.3 Some sets of analog coefficients reported for the 4th-order 2-2 $\Sigma\Delta$ M

Coefficients	[Miao98]	[Marq98]-A	[Marq98]-B
g_1, g'_1	0.25, 0.25	0.5, 0.5	0.5, 0.5
g_2, g'_2	0.5, 0.25	0.5, 0.5	0.5, 0.5
g_3, g'_3, g''_3	0.5, 0.125, 0.25	1, 0.5, 0.5	0.5, 0.25, 0.5
g_4, g'_4	0.5, 0.25	0.5, 0.5	0.5, 0.5
d_0, d_1	1, 4	1, 2	1, 4
Δ SNR in comparison to ideal	−12 dB (−2 bit)	−6 dB (−1 bit)	−12 dB (−2 bit)
$X_{OL}/(\Delta/2)$	−2 dBFS	−5 dBFS	−2 dBFS
Unitary capacitors ^a	28 (4 + 6 + 12 + 6)	13 (3 + 3 + 4 + 3)	16 (3 + 3 + 7 + 3)

^a Capacitor sharing between coefficients in an integrator is assumed

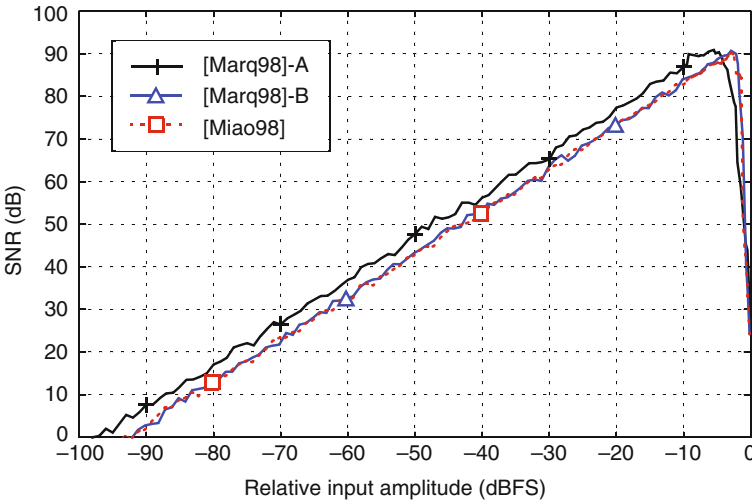


Fig. 2.18 SNR curves of the 2-2 cascade $\Sigma\Delta$ Ms in Table 2.3 for OSR = 32

- Facilitating the implementation of the digital cancellation logic; e.g., power-of-two coefficients are often preferred in order to use simple shift registers.

For comparison purposes, Table 2.3 shows some sets of analog coefficients reported for the 4th-order 2-2 cascade $\Sigma\Delta$ M. Related aspects, like nominal performance loss, overload level and minimum number of unitary capacitors, are also included. Obviously, many trade-offs exist between the aforementioned aspects, but once they are worked out the SNR obtained with a cascade $\Sigma\Delta$ M can be considerably larger than with its single-loop counterpart. Fig. 2.18 shows the SNR curves of the three 2-2 $\Sigma\Delta$ Ms in Table 2.3 for an oversampling ratio of 32, in which peak SNRs of approximately 91 dB are obtained. This value surpasses in 30 dB the peak SNR achieved by a stable 4th-order single-loop $\Sigma\Delta$ M with all NTF-zeros at DC (see Fig. 2.14a). The improvement is still around 10 dB in comparison to optimized 4th-order NTFs with two notches within the signal band (see Fig. 2.14b).

2.6 Multi-bit $\Sigma\Delta$ Architectures

As previously stated, the dynamic range of a $\Sigma\Delta$ modulator can be enhanced by means of increasing the order of the quantization noise shaping. Nevertheless, the expected improvement in performance for large modulator orders can vanish due to instabilities in single-loop $\Sigma\Delta$ architectures (see Sect. 2.4.2) or due to noise leakages in cascade $\Sigma\Delta$ topologies (see Sect. 2.5). In accordance with (2.19), an alternative way for further increasing the modulator DR consists of using embedded quantizers with larger resolution. The main advantages of multi-bit $\Sigma\Delta$ modulators are:

- The in-band quantization error power is roughly reduced 6 dB per additional bit in the embedded quantizer thanks to the smaller internal quantization step Δ .
- Internal non-linearities are weaker in multi-bit $\Sigma\Delta$ Ms than for single-bit counterparts. The quantizer operation better fits the white noise approximation and phenomena caused by non-linear dynamics are less evident.
- For a given order in the loop filter, the stability properties for multi-bit $\Sigma\Delta$ Ms are better than for single-bit $\Sigma\Delta$ architectures.

The aforementioned benefits suggest that, for a targeted modulator performance, multi-bit quantization can be traded for noise shaping and/or oversampling. Indeed, multi-bit $\Sigma\Delta$ Ms are often employed in wideband applications, since the oversampling ratio can be lower than in their single-bit counterparts. This helps to reduce the operation frequency and, therefore, the power consumption not only in the $\Sigma\Delta$ M itself, but also in the decimation filter. Nevertheless, besides the increase in circuit complexity when moving from single- to multi-bit quantization, other aspects related to linearity requirements have a strong impact on the modulator operation. Contrary to 1-bit quantizers that are intrinsically linear because only two levels are used in the quantization process, multi-bit quantizers exhibit in practice some non-linearities in their transfer characteristic mostly due to device mismatching. As we indicate below, these errors have a significant influence on modulator performance and may represent an important drawback that counters the aforementioned advantages.

2.6.1 Impact of DAC Non-linearities

Figure 2.19 illustrates the linear model of a multi-bit $\Sigma\Delta$ M, in which errors related to the multi-bit conversion are added to the quantization error e that has been taken into account so far; namely, an error e_{ADC} associated to the A-to-D conversion process and an error e_{DAC} in the subsequent D-to-A conversion required to reconstruct the analog feedback signal. Note that e_{ADC} is injected in the same path as the quantization error e and is also therefore attenuated by the gain of the loop filter within the signal band. However, DAC errors are injected in the feedback path and the corresponding non-linearities are therefore directly added to the modulator input where they are not

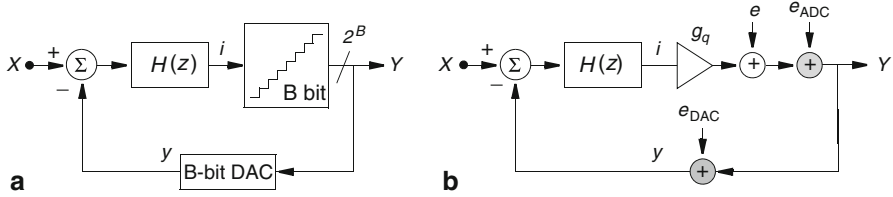


Fig. 2.19 Impact of errors in $\Sigma\Delta$ M with multi-bit embedded quantizers: **a** basic scheme, **b** corresponding linear model

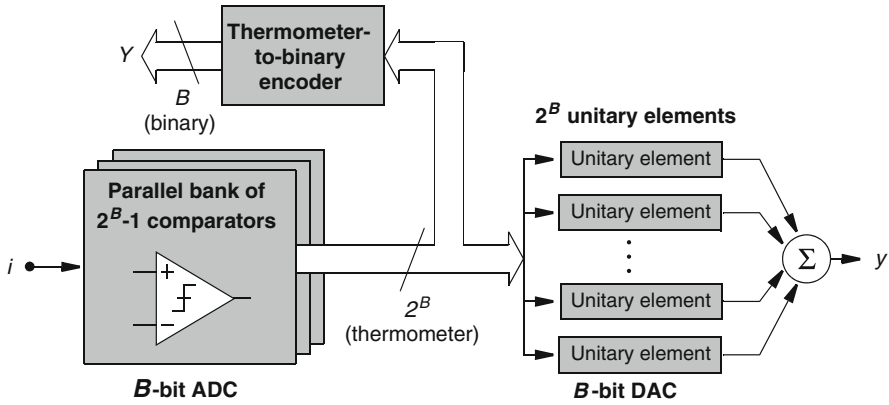


Fig. 2.20 Typical scheme of the multi-bit ADC and DAC embedded in a $\Sigma\Delta$ M

be mitigated by noise shaping; i.e., they pass to the modulator output as the input signal does. Consequently, the linearity of a multi-bit $\Sigma\Delta$ M will be no better than that of the multi-bit embedded DAC [Carl97] and the latter must be designed to reach the linearity targeted for the whole $\Sigma\Delta$ ADC, what is quite challenging due to the influence of component mismatching.

Figure 2.20 shows the architecture typically employed for the embedded quantizer and the feedback DAC in multi-bit $\Sigma\Delta$ modulators, which is fully parallel since the number of internal bits is normally low ($B \leq 5$). The B -bit ADC consists of a bank of $2^B - 1$ comparators that digitizes the output of the loop filter into thermometer code, which is subsequently coded into binary at the back end. On the other hand, the DAC employs $2^B - 1$ unitary elements (capacitors, resistors, current sources, etc.) to reconstruct the analog feedback signal using 2^B levels. The i th analog output level is generated by activating i unitary elements and adding their outputs (charges or currents). Errors in the DAC are originated by the mismatching between its unitary elements, which causes the deviation of the DAC output levels from their nominal values. Assuming that the actual value of each unitary element follows a Gaussian distribution, the worst-case relative error in the DAC output y can be

estimated as [Carl97]

$$\sigma\left(\frac{\Delta y}{y}\right) \approx \frac{1}{2\sqrt{2^B}} \sigma\left(\frac{\Delta U_e}{U_e}\right) \quad (2.32)$$

where $\sigma(\Delta U_e/U_e)$ is the relative error in the value of the unitary element. Obviously, the accuracy in the DAC increases with the number of unitary elements thanks to the parallel topology. However, for a $\Sigma\Delta$ M with 4-bit internal quantization to achieve 16-bit linearity, the required matching of the unitary elements in the DAC should be better than 0.01% (13 bits). Unfortunately, the matching of devices that can be achieved in present-day CMOS processes is in the range of 0.1% (10 bits) and the required accuracy in the unitary components could only be obtained through the parallel connection of many more elements. This clearly means that achieving linearities larger than 12 or 13 bits in multi-bit $\Sigma\Delta$ Ms, while relying only on standard device matching, leads to prohibitive area occupation.

A straight-forward method to improve the standard device matching consists of trimming their values [Carl97], what can be sometimes done at the foundry, but at the expense of additional fabrication and/or measurement steps and increased cost. Analog calibration running periodically or in background can also be used to this purpose and to compensate for drifts or aging, but at the cost of a significant increase in circuit complexity. Another strategy consists of converting DAC errors into digital form and correcting them in digital domain using look-up tables [Carl97].

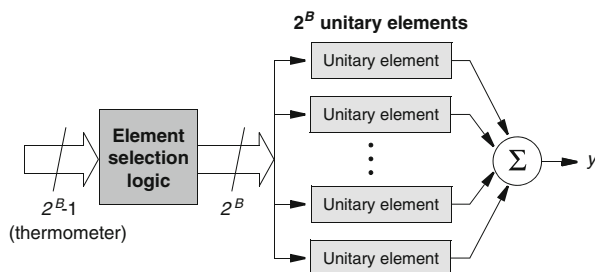
Among the different alternatives that have been developed through the years for achieving high-linear multi-bit $\Sigma\Delta$ Ms, two of them clearly prevail because of the modest component matching and circuit complexity involved. These are discussed below.

2.6.2 *Dynamic Element Matching*

As previously stated, mismatching between the unitary elements of the feedback DAC in a multi-bit $\Sigma\Delta$ M causes non-linear errors to be directly added at its input, therefore limiting the overall modulator linearity. For a given DAC with the structure in Fig. 2.20, there is a univocal correspondence between the thermometric input code and the respective DAC error, because the same unitary elements are always used to generate a given DAC output level. The fundamental idea underlying dynamic element matching (DEM) consists of breaking this direct correspondence by varying over time the set of elements that are used to generate a given DAC output level, therefore transforming its fixed error into a time-varying one. As shown in Fig. 2.21, a digital block is added to control the selection of elements at each clock cycle according to an algorithm that tries to drive the average error in each DAC level to zero over time. Part of the DAC error power that laid in the low-frequency range will therefore be moved to higher frequencies, where it will be removed by the decimation filter.

Many of the different DEM techniques that have been developed are detailed in [Geer02]. They can roughly be grouped in the following clusters:

Fig. 2.21 Architectural concept of DEM



- Randomization algorithms, in which the DAC unitary elements are selected according to pseudo-randomly configured networks (e.g., butterfly structures). Harmonic distortion induced by the DAC is transformed into white noise, whose out-of-band power will be removed by decimation. The DAC error power laying within the signal band will however increase the modulator noise floor.
- Rotation algorithms, in which the DAC unitary elements are selected in a periodic manner for shifting harmonic distortion out of the signal band. The modulator noise floor is not increased, but the processing can originate mixed frequency components that fold back into the baseband. Clocked averaging (CLA) is an example of this kind of DEM techniques.
- Mismatch-shaping algorithms, in which the DAC unitary elements are selected according to algorithms that conform the mismatching error in order to push most of its power to higher frequencies. The order of the mismatch shaping is normally limited to one or two. Individual level averaging (ILA), data weighted averaging (DWA and its many modifications) and data directed scrambling (DDS) pertain to these kind of algorithms, whose use is very extended.

2.6.3 Dual-Quantization Techniques

An alternative approach to reduce the impact of DAC non-linearities in multi-bit $\Sigma\Delta$ Ms can be found in dual-quantization techniques, in which both single- and multi-bit quantizers are used at a time in the same $\Sigma\Delta$ M. The underlying idea is to combine their benefits: two-level quantization for its the intrinsic linearity and multi-bit quantization for its reduced error power. Architectural examples of these techniques will be discussed below.

2.6.3.1 Leslie-Singh Architecture

Figure 2.22 shows the general scheme of the dual-quantizer $\Sigma\Delta$ M proposed by Leslie and Singh [Lesl90]. Single-bit quantization is used in the $\Sigma\Delta$ feedback loop because of its intrinsic linearity and a path containing only a multi-bit quantizer is connected

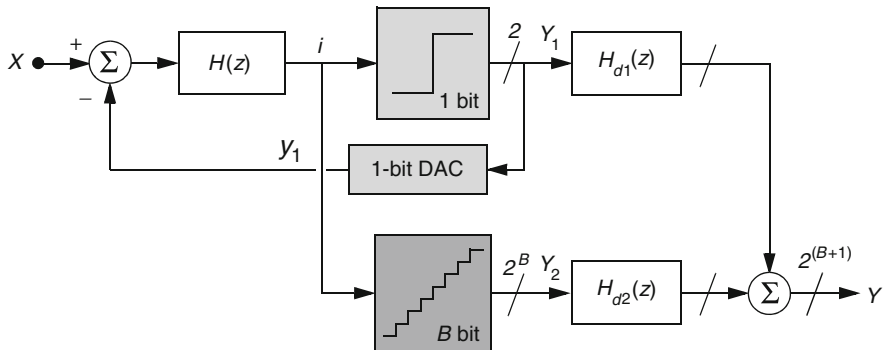


Fig. 2.22 Leslie-Singh $\Sigma\Delta$ modulator

to the loop filter output. The two quantizer outputs are then properly combined in digital domain to reduce the quantization error at the output to that of the multi-bit quantizer. Note that the scheme in Fig. 2.22 can be simplified in practice by removing the comparator and feeding the most-significant bit of the B -bit quantizer back to the $\Sigma\Delta$ loop.

The Leslie-Singh topology can be viewed as a MASH $\Sigma\Delta$ M, in which the first stage is an L th-order single-bit $\Sigma\Delta$ M and the second stage is a 0-order B -bit $\Sigma\Delta$ M. It therefore requires perfect cancellation of the 1st-stage 1-bit quantization error at the modulator output and also suffers from noise leakage problems. Note also that, although the modulator output ideally contains only the quantization error coming from the multi-bit quantizer, the stability of the architecture does not improve as in standard multi-bit $\Sigma\Delta$ Ms, since the loop is closed through single-bit feedback. The NTF cannot therefore be optimized to more aggressive high-order shapings without jeopardizing stability.

2.6.3.2 Dual-Quantizer Single Loops

An alternative way of employing dual quantization in high-order $\Sigma\Delta$ Ms is illustrated in Fig. 2.23. In the 3rd-order $\Sigma\Delta$ M shown, the first two integrators are fed by a 1-bit DAC, while the third one is fed by a multi-bit DAC. The modulator linearity is not menaced, since DAC non-linearities are suppressed by the gain of the first two integrators. At the same time, the topology benefits from improved stability thanks to the multi-bit feedback in the last integrator. In practice, the most-significant bit of the B -bit quantizer can be used to close the 1-bit feedback. Under linear analysis, the modulator output is ideally obtained as

$$Y(z) = z^{-1}X(z) + 2(1 - z^{-1})^3 E_2(z) - 2z^{-1}(1 - z^{-1})^2 E_{\text{DAC}}(z) \quad (2.33)$$

yielding 3rd-order shaping for the B -bit quantization error, $E_2(z)$, and 2nd-order shaping for the DAC errors, $E_{\text{DAC}}(z)$. Note that the topology suffers however from noise leakages.

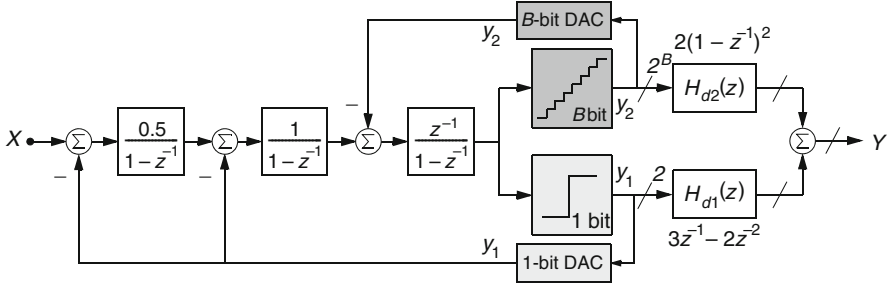


Fig. 2.23 A 3rd-order single-loop $\Sigma\Delta$ M with dual quantization. [Hair91]

The concept in Fig. 2.23 can be generalized to higher-order $\Sigma\Delta$ Ms. As the order of the loop filter increases, the number of back-end integrators with multi-bit feedback can be traded off with aggressive noise shaping (for improved stability) and linearity requirements of the multi-bit DAC.

2.6.3.3 Dual-Quantizer Cascades

Dual-quantization techniques can be easily incorporated to generic MASH $\Sigma\Delta$ Ms. As shown in (2.30), the output of a cascade $\Sigma\Delta$ M ideally contains only the input signal and the last-stage quantization error, whereas the quantization errors from the remaining stages are removed by the DCL. The dynamic range of the $\Sigma\Delta$ M can therefore easily be increased by using multi-bit quantization only in the last stage of the cascade. The remaining quantizers are usually single-bit to retain linear feedback in the front-end stages. This way non-linearities in the multi-bit DAC are injected in the back-end stage of the cascade and they will appear at the modulator output with the attenuation provided by the integrators in the preceding stages. This is usually sufficient for achieving a good linearity performance in the MASH $\Sigma\Delta$ M, with no need for correction techniques in the multi-bit DAC.

The resulting topology is that illustrated in Fig. 2.16, considering $B_i = 1$ for $i = 1, \dots, N-1$ and $B_N = B$. Under ideal linear analysis, the output of this generic N -stage cascade $\Sigma\Delta$ M with multi-bit quantization at the back end yields

$$\begin{aligned}
 Y(z) &= \text{STF}(z)X(z) + \text{NTF}_N(z)E_N(z) + \text{NTF}_{\text{DAC}}(z)E_{\text{DAC}}(z) \\
 &= z^{-L}X(z) + d_{2N-3}(1-z^{-1})^L E_N(z) + d_{2N-3}(1-z^{-1})^{(L-L_N)} E_{\text{DAC}}(z)
 \end{aligned}
 \tag{2.34}$$

where L is the summation of the stage orders, d_{2N-3} is the scaling factor due to inter-stage coupling (usually 2 or 4), $E_N(z)$ is the last-stage quantization error and $E_{\text{DAC}}(z)$ is the error in the multi-bit DAC. Note that the DAC error will be $(L - L_N)$ th-order shaped—i.e., the overall cascade order minus that of the back-end stage.

Figure 2.24 illustrates the block diagram of the first reported cascade employing dual quantization [Bran91]. It is a 3rd-order cascade $\Sigma\Delta$ M employing a 2nd-order

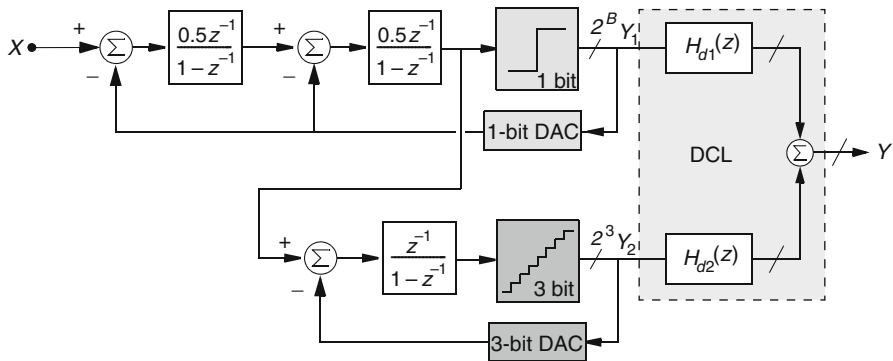


Fig. 2.24 Dual-quantizer 2-1 cascade $\Sigma\Delta$ M. [Bran91]

front-end stage with 1-bit quantization for linear feedback and a 1st-order back-end stage with a 3-bit embedded quantizer for enhanced dynamic range. Many integrated cascade $\Sigma\Delta$ Ms using this dual-quantization scheme can be found in the literature [Nors97a, Rodr03].

Cascade $\Sigma\Delta$ Ms employing multi-bit quantization in all stages have been also reported [Fuji00, Vleu01]. Note that, under ideal conditions, the quantization errors of the first stages are cancelled out at the modulator output by the DCL. Multi-bit quantization can be used in these stages for the main purpose of reducing the corresponding quantization errors that will in practice leak to the output. Being that the case, DEM techniques can be incorporated to the multi-bit DAC of the front-end stage in order to achieve the required modulator linearity (if required), whereas succeeding multi-bit stages often rely just in the in-band attenuation provided by the preceding integrators [Vleu01]. Using multi-bit quantization in all stages of a cascade $\Sigma\Delta$ M presents an additional appealing feature: the coupling factors between stages can be increased in comparison to a 1-bit approach without overloading the quantizers [Fuji00]. This way, the scaling factor d_{2N-3} that amplifies the last-stage quantization error [see (2.34)] can be smaller than unity, leading to an improvement of the global performance⁷.

Cascades using tri-level (1.5-bit) quantizers can also be found in literature [Dedi94] in order to reduce quantization errors in comparison with 1-bit quantization—over 3-dB SNR improvement. Although tri-level coding is not inherently linear, it is often used in fully-differential SC $\Sigma\Delta$ Ms, since highly-linear tri-level DACs can be easily implemented using just one extra switch [Reut02].

⁷ In [Fuji00], 4-bit quantization is used in all stages of a 2-1-1 cascade $\Sigma\Delta$ M. The resulting scaling factor is $d_3 = 1/32$, which leads to a 30-dB increase in DR in comparison to an ideal 4th-order 4-bit $\Sigma\Delta$ M.

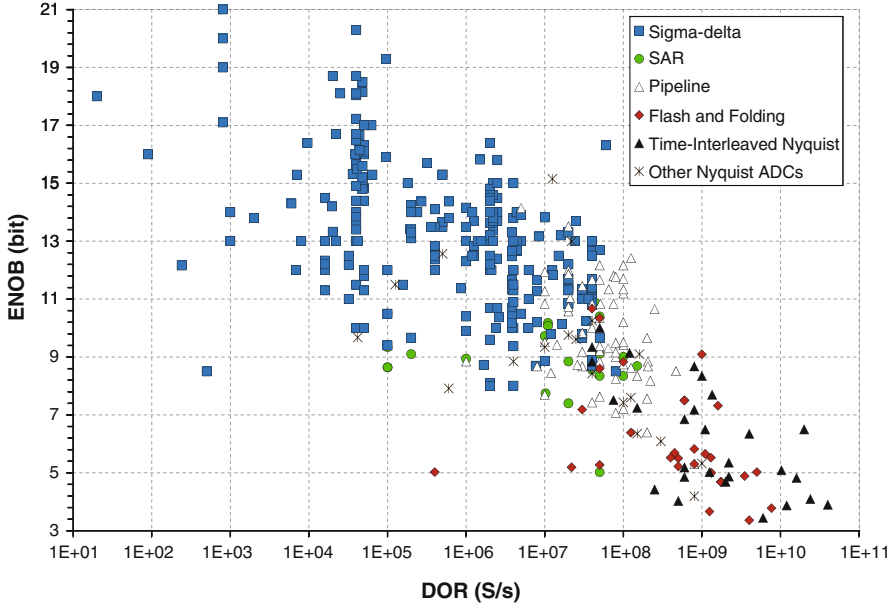


Fig. 2.25 State-of-the-art CMOS ADCs in the ENOB-DOR plane

2.7 State of the Art in $\Sigma\Delta$ ADCs

Although $\Sigma\Delta$ converters were originally conceived for low-frequency, high-resolution applications (like audio and precision measurement) in which they clearly outperform other existing A/D conversion techniques, their use has progressively extended to medium- and high-frequency applications with the development of VLSI technologies. Fig. 2.25 illustrates the state of the art in ADCs implemented in CMOS processes reported up to year 2010 and places them in the ENOB-DOR plane, where $\text{DOR} = 2 \text{ BW}$ stands for the digital output rate of the ADC (i.e., the Nyquist rate). Data in Fig. 2.25 corresponding to CMOS Nyquist-rate ADCs has been taken from [Murm10]. It can be noted that $\Sigma\Delta$ ADCs cover a wide frequency range, ranging from 10 Hz to 50 MHz. Larger conversion bandwidths are still dominated by Nyquist-rate ADCs, especially flash and folding. Oversampling techniques are less efficient in these applications because of the excessive operation speed that is required in the analog blocks. They however coexist with algorithmic, subranging and especially pipeline converters in communication applications.

As discussed in previous sections, many different alternatives exist for the realization of $\Sigma\Delta$ modulators. Fig. 2.26 looks closely at the reported $\Sigma\Delta$ Ms taking their topologies into account. For comparison purposes, they have been classified into three categories—single- and multi-bit single loops and cascades—distinguishing between discrete-time (DT) and continuous-time (CT) implementations. Note from Fig. 2.26 that, although the majority of reported ICs correspond to DT $\Sigma\Delta$ Ms,

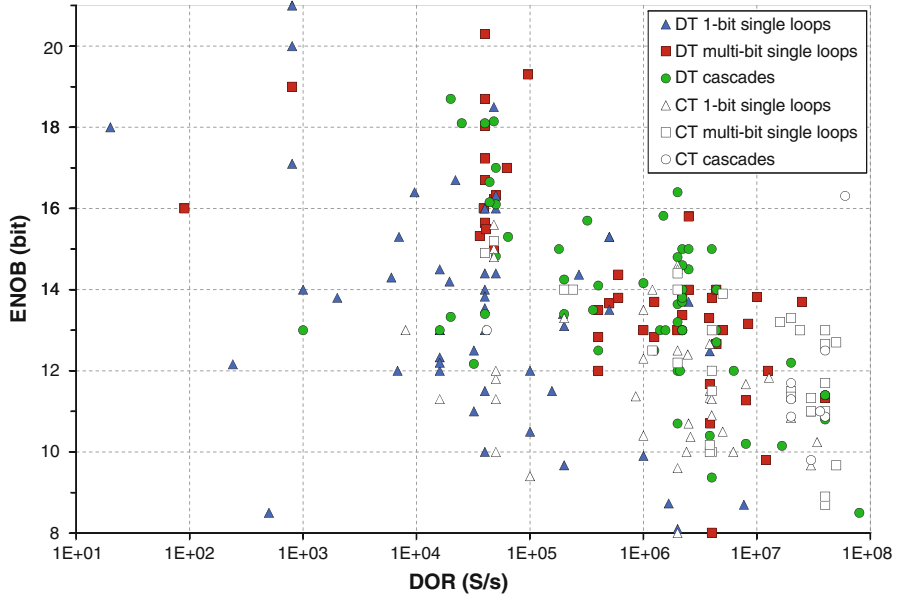


Fig. 2.26 State-of-the-art CMOS $\Sigma\Delta$ Ms in the ENOB-DOR plane

CT $\Sigma\Delta$ Ms almost reach up to 35%. DT implementations are clearly dominant for high resolutions (>16 bit), whereas CT $\Sigma\Delta$ Ms dominate for large bandwidths ($\text{DOR} > 10 \text{ MS/s}$). Nevertheless, both types of implementations coexist for large-bandwidth, medium-resolution (10–14 bit) applications.

The performance of $\Sigma\Delta$ Ms is globally quantified in terms of their main specifications (effective resolution, signal bandwidth and power consumption of the circuit) through the following figures of merit (FoMs), respectively proposed by [Good96, Rabi97],

$$\text{FoM}_1|_{\text{pJ/conv}} = \frac{\text{Power(W)}}{2^{\text{ENOB(bit)}} \text{DOR(S/s)}} 10^{12} \quad (2.35)$$

$$\text{FoM}_2 = 2kT \frac{3 \cdot 2^{\text{ENOB(bit)}} \text{DOR(S/s)}}{\text{Power(W)}} \quad (2.36)$$

where k is the Boltzmann constant and T is the temperature (measured in Kelvin, K). Note that FoM_1 emphasizes power consumption, whereas FoM_2 stresses effective resolution, represented by the ENOB. Therefore, the smaller the FoM_1 value and the larger the FoM_2 value, the “better” the $\Sigma\Delta$ M is. For comparison purposes, the aforementioned information is graphically illustrated in Figs. 2.27 and 2.28.

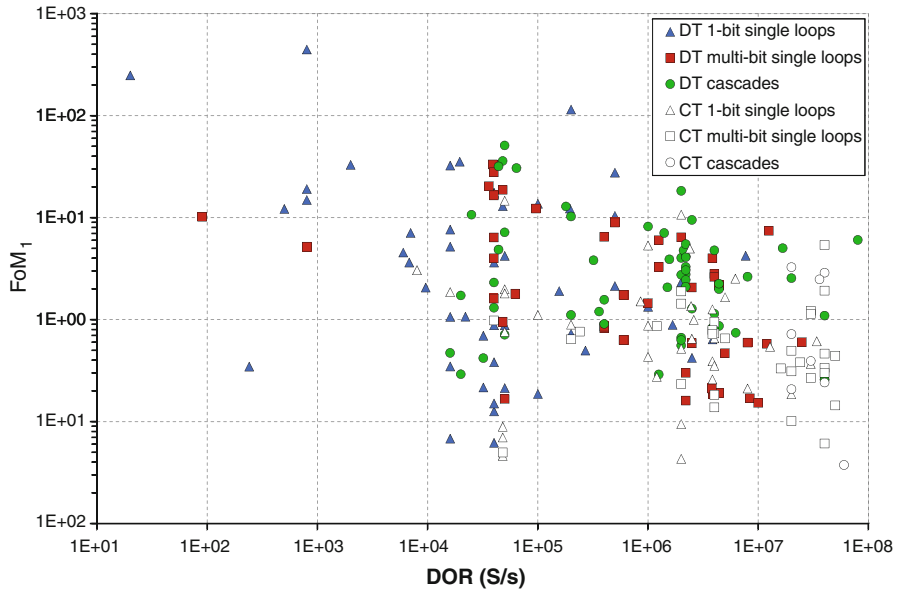


Fig. 2.27 FoM_1 versus DOR of the reported low-pass $\Sigma\Delta$ ICs

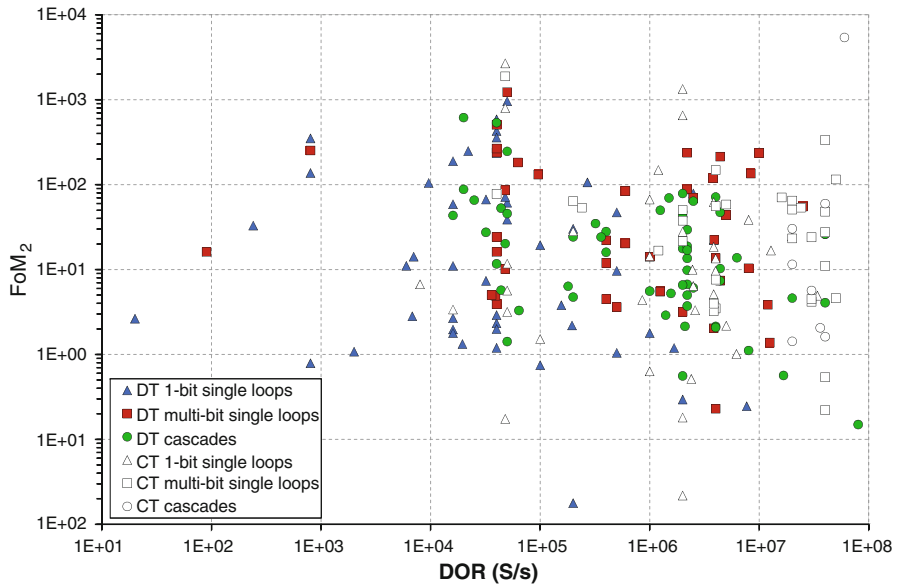


Fig. 2.28 FoM_2 versus DOR of the reported low-pass $\Sigma\Delta$ ICs

2.8 Summary

The basic principles of $\Sigma\Delta$ modulation have been presented in this chapter. The benefits of oversampling and noise shaping on the ADC performance have been discussed and the generic scheme, ideal performance and metrics of $\Sigma\Delta$ converters have been defined and compared with Nyquist ADCs.

Topological alternatives for the practical implementation of $\Sigma\Delta$ modulators have been also presented, addressing the realization of stable high-order $\Sigma\Delta$ Ms by means of optimized single loops or by cascading of low-order stages. The use of multi-bit internal quantization has also been presented as an alternative to enhance the modulator effective resolution. Besides reducing quantization error, it provides better stability properties to single-loop architectures, but jeopardizes linearity. Non-linearity error in the multi-bit DAC due to component mismatching has been discussed, together with techniques to palliate its impact on the modulator performance, such as DEM and dual-quantization schemes.

The many design alternatives for the practical implementation of $\Sigma\Delta$ ADCs are finally summarized in the state of the art of reported low-pass ICs.

Nanometer CMOS Sigma-Delta Modulators for Software
Defined Radio

Morgado, A.; del Río, R.; de la Rosa, J.M.

2012, XVIII, 288 p., Hardcover

ISBN: 978-1-4614-0036-3