Over Sampling Converters
Franco Maloberti, University of Pavia, Italy
Paul O'Leary Jonnaeum Research Graz, Austria

7.1 Introduction
Conventional high accuracy data converters require the use of accurate analogue components. In general, precise components are obtained by the use of specific technological tricks that led to what is commonly referred to as "analogue technology". However, for high-performance digital applications it is necessary to optimize the technology in directions that are often opposite to the ones indicated by analogue requirements, and, since the digital section of a complex circuit is dominant, a compromise that is very favourable to digital needs must be adopted. As a specific example it is worthwhile remembering that in order to get good capacitors it is necessary to use double poly layer. However, since the number of interconnection layers for digital circuits is extremely important, it is preferable to make the additional technological effort (and masks) to get one additional metall layer.

The increasing requirement for accuracy in recent years has stimulated intensive research to find suitable analogue design techniques that are capable of achieving good analogue performances while using "digital technologies". For data conversion, the oversampling technique, described in this chapter, seems to be a very profitable solution for low (instrumentation and audio-band) and medium frequency applications. The speed of operation used is much higher than the band of interest and this feature is exploited to improve accuracy [1]. In conventional A/D converters (the so-called Nyquist rate category) the input signal is sampled at a rate that is only twice that of the band of the input signal itself. The digital output, generated at the same rate, following to the sampling theorem, retains all the informative contents of the input signal which represents it. However, in order to avoid aliasing, the input signal must be band-limited by an anti-aliasing filter before sampling. In oversampling converters, by contrast, the input analogue signal is sampled and processed at a rate which is significantly higher than twice the input band. The ratio between the sampling rate and twice the band of interest is usually referred to as the oversampling factor, Rs. Oversampling as large as many hundreds is often used.

As we will study in this chapter, the important effect of oversampling is that the power of the error associated with the quantization (quantization noise)
Over Sampling Converters
Franco Maloberti, University of Pavia, Italy
Paul O'Leary Jonnaeum Research Graz, Austria

7.1 Introduction
Conventional high accuracy data converters require the use of accurate analogue components. In general, precise components are obtained by the use of specific technological tricks that led to what is commonly referred to as "analogue technology". However, for high-performances digital applications it is necessary to optimize the technology in directions that are often opposite from the ones indicated by analogue requirements, and, since the digital section of a complex circuit is dominant, a compromise that is very favourable to digital needs must be adopted. As a specific example it is worthwhile remembering that in order to get good capacitors it is necessary to use a double poly layer. However, since the number of interconnection layers for digital circuits is extremely important, it is preferable to make the additional technological effort (and masks) to get one additional metal layer.

The increasing requirement for accuracy in recent years has stimulated intensive research to find suitable analogue design techniques that are capable of achieving good analogue performances while using "digital technologies". For data conversion, the oversampling technique, described in this chapter, seems to be a very profitable solution for low (instrumentation and audio-band) and medium frequency applications. The speed of operation used is much higher than the band of interest and this feature is exploited to improve accuracy [1]. In conventional A/D converters (the so-called Nyquist rate category) the input signal is sampled at a rate that is only twice that of the band of the input signal itself. The digital output, generated at the same rate, following to the sampling theorem, retains all the informative contents of the input signal which represents it. However, in order to avoid aliasing, the input signal must be band-limited by an anti-aliasing filter before sampling. In oversampling converters, by contrast, the input analogue signal is sampled and processed at a rate which is significantly higher than twice the input band. The ratio between the sampling rate and twice the band of interest is usually referred to as the oversampling factor,Ros, oversampling as large as many hundreds is often used.

As we will study in this chapter, the important effect of oversampling is that the power of the error associated with the quantization (quantization noise)
is spread over a band that is much wider than the band of interest. Thus the digital output of an oversampled A/D converter contains a given fraction of the quantization noise that is pushed out of the band of interest. This fraction can be filtered out without affecting the informative content associated with the input signal. After filtering, the power of the quantization noise is reduced and, in turn, the resolution of the converter is increased.

Different categories of oversampling converters perform the spread (or the shaping) of the quantization noise in the band and outside of the band of interest with different efficiencies; consequently, they determine the need for post filtering at different levels of complexity. This chapter, after considering basic oversampling converters, concentrates on the sigma delta schemes.

7.2 Intuitive Introduction to Oversampling Data Converters

Before looking at oversampling converters analytically, it is important to gain an intuitive understanding of how they function. Most publications on oversampling converters start with very tedious analytical examinations of their functionality. This has given oversampling converters the reputation of being complicated, which is not true.

To help the reader gain an intuitive understanding of oversampling converters, this introduction starts by looking at digital to analogue conversion. Digital to analogue conversion has been chosen because it is understood and perceived more simply than analogue to digital conversion. However, the principles and problems of oversampling are similar for both D/A and A/D converters.

Let us consider to begin with the conversion system known as pulse width modulation, of which a possible generated waveform is shown in Figure 7.1 the output of the converter can have only one of two voltage levels which are either $V_{\text{ref}}$ or ground. The conversion time $T_c$ is long with respect to the clock period $T_c$. Consequently, there are a large number of clock cycles, say $m$, available for each conversion. The desired output voltage is approximated by the repetition of a pulse, whose duty-cycle is digitally controlled. The average level of the pulse corresponds to the output analogue signal. In the example shown in Figure 7.1, the output voltage is set to $V_{\text{ref}}$ for four of the $m$ possible time slots. Consequently, its average $V_{\text{out, av}}$ is equal to $(4/8)V_{\text{ref}}$. The average output voltage can be generated by either integrating the DAC output over the conversion time interval, $T_c$, or by a low-pass filtering of the pulse stream. The latter is more to be preferred.

![Figure 7.1 Possible pulse placement for an oversampling converter](image)

As can be seen from the principle of operation the output voltage is always smaller than the reference; moreover, given in clock period $T_c$ for conversion, the precision of the output voltage turns out to be equal to one part per $m$ that corresponds to $\log(m)$ equivalent number of bits.

Since the clock frequency is much higher than the conversion frequency the described technique falls into the oversampling data conversion category. The oversampling ratio $R_{\text{os}}$, defined as the ratio of the sampling frequency $f_s$ to the required Nyquist frequency, $f_{\text{Nyq}} = f_c$, is equal to $m$. Consequently, the equivalent number of bits $b_{\text{eq}}$ in such an oversampled converter is given by $\log(m)$.

This is an important result, since it represents a fundamental limit for the accuracy of pure oversampling modulators. Doubling the clock frequency while keeping the signal band constant results in a doubling of the number of time slots available per conversion and, in turn, allows a doubling of the achievable accuracy: the converter gains half a bit of accuracy for each doubling of the clock frequency.

It is now important to examine the spectral properties of the digital to analogue converter output. For this, consider a pulse width modulator with 32 time slots (5 bits) during 7 of which the output is high i.e. connected to $V_{\text{ref}}$. Since the output of the PWM is periodic, the spectrum will consist of discrete spectral lines spaced evenly at multiples of the conversion frequency $f_c$. The amplitude envelope of these spikes is given by the Fourier transform of the output during one period of the conversion. In the present case this results in $\sin(x)/x$ envelope. A plot of the Fourier transform for this example is shown in Figure 7.2. The $\sin(x)/x$ envelope and the DC level of $7/32$ can be seen.

The value at DC is the desired signal, all other frequency components must be considered as disturbances. The disturbances with the largest magnitude are at lower frequencies. These noise components do not represent a problem when driving systems with long time constants or which have large momentum. This explains why pulse width modulators are popular when
driving motors and other mechanical devices. Furthermore, only one rising edge and one falling edge is required per conversion, thus making it possible to reduce the switching transients in power systems. However, for other applications the undesired discrete spectral components must be removed via a low pass filter. This requires a filter with a very steep transition from pass band to rejection band. Additionally, a large stop-band rejection is required because the first unwanted frequency components have almost the same amplitude as the desired signal.

The generation of a pulse width modulated signal can be implemented with simple circuitry. A possible solution is shown in Figure 7.3. The input data is stored in a register, the output of which is compared with the output of a binary counter once per clock cycle. This is equivalent to comparing the data word to a sawtooth signal, as shown in Figure 7.4. The output bit is high so long as the data value is greater than that of the counter.

In order to avoid large disturbances at low frequency a different placement of the pulses required to construct the output signal within the available clock periods can be used. One possibility is shown in Figure 7.5, where the pulses are uniformly distributed in time. This pulse placement is often called "pulse density modulation" (PDM) and corresponds to the output from a "first order sigma-delta modulator" or a "charge balancing converter" as they are also known.

The accuracy which can be achieved with this pulse placement is the same as for pulse width modulation, however, the spectral properties of the output signal are different. By evaluating the output spectrum of this pulse stream under the same conditions as for the PWM i.e. 7 from 32 pulses high, we come to the spectrum shown in Figure 7.6. The desired output DC level has
seen maintained, while the disturbances have been moved to higher frequencies. However, it is important to note that, since m time slots are used to determine the average output value, the achieved accuracy of the DC level remains unchanged. In other terms, the different modulation algorithms modify the signal to noise ratio but not accuracy.

The pulse density modulated bit stream can be generated using simple circuitry, see Figure 7.7. The input digital data is stored in a register (B) and

![Figure 7.7 Pulse density modulator implementation (first order sigma-delta DAC)](image)

at every clock cycle it is summed up with the contents of register A, which, in turn, is the delayed version of the output of the accumulator. The system generates an output pulse whenever the accumulator reaches its full scale carry out. It is easy to verify that for a full scale equal to 32 and 7 applied at the input, the full scale is obtained, assuming the accumulator at the beginning to be empty, after 5, 9, 13, 18, 22, 27, 32 clock periods.

The required pulses could also be placed in a random manner. Such modulators are known as “stochastic” converters. If the pulses are truly random then the quantization error is spread uniformly in the frequency domain. The output spectrum of a stochastic modulator (shown in Figure 7.8) is “whiter” than for the previous examples of PWM and PDM. These converters are advantageous in applications where the system is sensitive to disturbances with discrete frequency components. The implementation of the stochastic modulator is very similar to that shown in previous examples, Figure 7.9. A pseudo random (PN) generator is the key to stochastic converters, it must generate numbers which are truly random and having a uniform probability density function. Any deviation from these desired properties will result in an output spectrum where the noise is not white.

![Figure 7.8 Output spectrum of a stochastic modulator for 7 from 32 high time pulses](image)

![Figure 7.9 Implementation of a stochastic DAC](image)

It is obvious that there is an almost infinite number of possibilities of placing the pulses. The three examples considered had the aim of giving the reader an insight into oversampling techniques. More generally the high order sigma–modulators which will be analyzed later are an extension of the above ideas.

7.3 Quantization

Quantization is the process of converting a signal with continuous amplitude to a signal with discrete amplitude levels [2][3]. During this process the signal is changed resulting in a modified spectrum. The change in signal power is referred to as quantization noise. However, the error due to the quantization process can be assimilated to noise only under a number of assumptions. These assumptions have certain weaknesses and it is important to examine them beforehand.
Assumption 1: All quantization levels are exercised with equal probability.

This assumption is almost never fulfilled. For a sine wave, the quantization levels have a non-uniform probability distribution function. For DC signals, this condition is not even approximated. Only signals with uniform amplitude distribution are compatible with this assumption, e.g., triangular waves with maximum amplitude.

Assumption 2: The quantization steps are uniform.

This requires an ideal analogue-to-digital converter. This assumption may be fulfilled for low-resolution converters. If the steps are not uniform, then the error is a function of the input signal and cannot be regarded as a simple additive noise component.

Assumption 3: The quantization error is not correlated with the input signal.

This is one of the weakest assumptions in the calculation of the signal-to-noise ratio. On the basis of this premise, it is assumed that the quantization error can be handled as noise with a white spectrum. The correlated nature of the quantization can be demonstrated with a simple example of a sine wave and its 4-bit quantized equivalent. See Figure 7.10. It can be seen that the resulting quantization error, Figure 7.11, is correlated with the sine wave at the input.

Assumption 4: A large number of quantization levels are used.

This assumption is usually fulfilled. However, in the case of one-bit oversampling converters, this assumption is violated. It is important to be aware of the above limitations when calculating the performance of a data conversion system. If all the above assumptions can be regarded as fulfilled, then quantization can be dealt with as an additive process.

The transfer curve for an ADC and the resulting quantization error is shown in Figure 7.12. It corresponds to an ideal analogue-to-digital converter with the transfer characteristics shifted by half LSB; the error is limited to the interval $-1/2Q + 1/2Q$, where $Q$ is the quantization step equivalent to one least significant bit.

The probability distribution function of the quantization error, $p(x)$, has the following limitation:

$$\int_{-\infty}^{\infty} p(x) \, dx = 1$$
Now if we apply assumption 1 that all the allowed levels are exercised with equal probability we get:

\[ p(x) = \frac{1}{Q} \quad \text{for} \quad x \in \left[ -\frac{Q}{2}, \frac{Q}{2} \right] \]

\[ p(x) = 0 \quad \text{otherwise} \]

The time average power for the quantization error, \( N_Q \), can now be calculated:

\[ N_Q = \int_{-\frac{Q}{2}}^{\frac{Q}{2}} x^2 \, p(x) \, dx = \frac{Q^2}{12} \]

The signal to noise ratio in dB is defined as:

\[ SNR = 10 \log_{10} \left( \frac{N_S}{N_Q} \right) \, dB \]

Where \( N_S \) is the signal power.

In order to give the reader a feeling for the signal dependence of this parameter let us evaluate the SNR for three different signal types.

Case 1: Consider an input signal with a uniform amplitude probability distribution function i.e. triangular wave with amplitude \( A_0 \). The signal power is given by:

\[ N_S = \frac{A_0^2}{12} \]

If the triangular wave has maximum amplitude \( A_0 \) corresponding to the full scale of the \( n \) bits ADC, we get:

\[ N_S = \frac{(2^n - 1)Q^2}{12} \]

It follows that the signal to noise ratio is:

\[ SNR = 20 \log_{10} \left( \frac{2^n - 1}{12} \right) = n \log_{10} (2) - 6.02 \, dB \]

Case 2: Consider a maximum amplitude sine wave. The signal power is:

\[ N_S = \frac{(2^n - 1)Q^2}{8} \]

Proceeding as with the previous example:

\[ SNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{8} \right) - 6.02 \, dB \]

Case 3: The input signal is a random variable with a Gaussian distribution where the four sigma points correspond to the ADC full scale range. This represents a very common requirement for an ADC, for example, a weighing scale. The signal power is:

\[ N_S = \frac{(2^n - 1)Q^2}{64} \]

Once again expressing the SNR in dB:

\[ SNR = 6.02 \, dB \]

Following the analysis of these three cases it is important to note that the signal to noise ratio at the output of an ADC is dependent on the input signal. Specifications which depend on signal to noise ratios should be handled very carefully as all boundary conditions must be specified.

The transfer curve for an ADC and the resulting quantisation error is shown in figure 7.13.

7.3.1 Spectral properties of quantisation noise

In analysing random signals it is now assumed that quantisation noise is a random, sampled data variable, there are two very important mathematical functions: the auto-correlation function and the power spectral density.
The auto-correlation function of a given variable \( v(t) \) is defined by:

\[
R(r) = \lim_{T \to \infty} \frac{1}{T} \int_{-T}^{T} v(t) v(t+r) \, dt
\]

and gives a measurement of the randomness of the variable \( v(t) \). From the auto-correlation function it is possible to derive the power spectral density of a signal:

\[
G(f) = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi \tau f} \, d\tau
\]

The power spectral density function is a statistical evaluation for the power of a random signal and its distribution in the frequency domain. For the assumed assumptions on the quantization noise \( \epsilon_q \) its auto correlation function is given by (Fig. 7.13):

\[
R_q(\tau) = \text{RMS}_{\epsilon_q}^2 \left( \frac{|\tau|}{T_s} \right) \quad \text{for} \quad |\tau| \leq T_s
\]

\[
R_q(\tau) = 0 \quad \text{for} \quad |\tau| > T_s
\]

Where \( \text{RMS}_{\epsilon_q} \) is the root mean square value of the noise and \( T_s \) is the sampling period. The assumptions which lead to this result are that the quantization error is constant during one sampling period and that the quantization noise is uncorrelated from one sample to the next. For example, the auto correlation function is zero for \( \tau \) greater than \( T_s \).

The spectral density of the quantization noise \( G_q(f) \) can be calculated by making the Fourier transform of the auto-correlation function shown above:

\[
G_q(f) = \text{RMS}_{\epsilon_q}^2 T_s \cdot \sin^2(\pi f T_s)
\]

Where \( \sin(x) = \sin(x) / x \).

Since this function will only be evaluated in a range where \( |f| \leq 1 / 2 \) a further simplification can be made by neglecting the action of the sine function. Thus the noise level is independent of frequency i.e. the quantization noise has a white spectrum. Moreover, the noise floor is proportional to the sampling period, conversely the higher the sampling frequency the lower the noise floor.

The total noise power in a given band width \( N_{\text{OH}} \) can now be calculated:

\[
N_{\text{OH}} = \int_{-B}^{B} \text{RMS}_{\epsilon_q}^2 T_s \, df = 2B \text{RMS}_{\epsilon_q}^2 T_s
\]
signal with only discrete amplitudes. In order words, it superimposes an additive component onto the signal: quantization noise. Under the assumption that the addition of quantization noise $e_k$ is an additive process, the schematic in Figure 7.14(a) can be transformed into its linearised version, shown in Figure 7.14(b). The system has two inputs and correspondingly two transfer functions: the signal transfer function, $T(z)$, and the noise transfer function, $N(z)$, they are given respectively by:

$$T(z) = \frac{V_{in}e^{zH(z)}}{V_{in}e^{zH(z)}} = \frac{H(z)}{1 + H(z) F(z)} \tag{7.19}$$

$$N(z) = \frac{V_{in}e^{zH(z)}}{V_{in}e^{zH(z)}} = \frac{1}{1 + H(z) F(z)} \tag{7.20}$$

In order to get the desired effect, the signal transfer function should not insert any attenuation, at least in the band of interest, and by contrast, the noise transfer function should reject signals in the input band. A solution to this is obtained by using $F(z) = 1$ and $H(z)$ of integrator-type. If $H(z)$ is a simple integrator the modulator is referred to as a first order noise-shaper; for more complex functions, typically cascades or a more complex interconnection of integrators, second-order or high-order noise shapers result.

When 1-bit A/D and D/A converters are used around the noise shaping loop the circuit is usually referred to as a "sigma delta modulator".

**7.5 First order sigma delta modulators**

A continuous time implementation of a first order sigma–delta modulator is shown in Figure 7.15. The 1-bit A/D converter is realized very simply by means of a comparator and a latched flip-flop. The two complementary outputs of the flip-flop drive two switches which connect the non inverting input of the integrator to ground or to the reference voltage; in this way the 1-bit D/A converter is implemented. Depending on the value of the sign of the output of the analog integrator (measured with respect to $V_{ref}$) the current that is injected into the virtual ground is positive or negative; hence, determining a falling or a rising of the output of the op-amp. Now, since the op-amp is in the integrating configuration and the loop is assumed to be stable, the mean value of the injected current must be zero. Thus, assuming $V_{in}$ constant (or varying very slowly in time), it must be, over a large number of clock cycles, $n$:

$$V_{ref} \cdot k_1 = V_{ref} \cdot k_0 = V_{ref} \cdot n - k_1 \tag{7.21}$$

where $k_1$ and $k_0$ is the number of clock cycles for which the output of the flip-flop is 1 or 0 respectively. From 7.21 we get:

$$V_{in} = \frac{V_{ref} \cdot k_1}{n} \tag{7.22}$$

Thus, the digital pulses at the output of the modulator contain the information on the amplitude of the input signal. Moreover, after one or more 1 pulses the system will react with more or one 0 pulses to compensate the negative injected current with an equivalent positive one. Thus the pulses are uniformly distributed in time.

The switched capacitor implementation of a first order sigma delta modulator is shown in Figure 7.16. The integrator is realized by a conventional switched capacitor structure where the same value of
capacitance is used in the input structure and the feedback element. The linearised diagram of the circuit is shown in Figure 7.17. The quantization noise, as usual, is represented by the symbol \( \epsilon_Q \). In the time domain the system is described by [6]:

\[
\begin{align*}
\eta(t) & = \eta(t) + \eta(t-1) - x(t) \\
\gamma(t) & = \eta(t) + \epsilon_Q(t)
\end{align*}
\]

that, in the z-domain results in:

\[
\gamma(z) = \eta(z)z^{-1} + (1 - z^{-1})\epsilon_Q(z)
\]

The transfer function of the signal is only a simple delay \((z^{-1})\); by contrast, the noise is modified by \( N(z) = (1 - z^{-1}) \) that corresponds to an high pass action:

\[
N(z) = 1 - e^{-\pi T/2} e^{-\pi /2} - e^{-\pi /2} + 2e^{-\pi /2} \sin \frac{\pi t}{2}
\]

where \( T \) is the sampling period.

The white spectrum of the quantization noise \( S_Q(f) \), spread over the band \( \pm f_S/2 \), is

\[
S_Q(f) = \frac{1}{12} \left| \frac{1}{f_S} \right|^2
\]

It is filtered by the square module of the noise transfer function, and results in the shaped noise spectrum:

\[
S_N(f) = S_Q(f) |2 \sin \pi f T|^2
\]

If we assume that \( B_n < < f_s \), equation 7.26 can be approximated by using \( \sin(x) = x \); in this case the total power in the bandwidth \( B_n \) is given by:

\[
N_{RMS} = \int_{B_n} S_N(f) (2\pi f)^2 df = \frac{1}{12} \pi^2 (2\pi T)^2 = \frac{1}{12} \pi^2 (R_{eq})^2
\]

This equation states that the quantization noise power is reduced by a factor of eight for a doubling of the sampling frequency. That is, the resolution of the data converter could be increased by one and a half bit by doubling the clock frequency. The advantage with respect to the simple oversampling operation is evident; however, the noise shaping transfer function, shown in Figure 7.18, amplifies the noise spectrum in the high frequency range. If

\[
\text{Noise gain} = 3
\]

\[
\text{Normalized frequency} (f / f_s)
\]

only for this reason, the specifications of the digital filter used to reject the undesired noise will be more severe than for the one used in a simple oversampled converter; besides rejecting an amplified noise it must also avoid the folding in baseband of noise power (due to aliasing) at a level which is negligible with respect to the one left by the more efficient shaping.

The SNR of a first order sigma delta converter, employing an ideal low pass digital filter with cut off frequency set at \( B_{in} \), can be calculated by the use of equation 7.27. Assuming an input signal with uniform amplitude probability, we get:
\[ \text{SNR} = 9.1 + 4.3 \log_{10} R_c + 5.3 \text{dB} \]

The above results have been obtained under the hypothesis that justifies the linearized model. In reality, since there are nonlinear elements around the loop, and since the assumptions taken for assimilating the quantization error to white noise have not been verified well, the derived equations give only an approximate description of the system's behaviour. The divergence between theory and reality can be shown by appropriate computer simulations \cite{7,8}. Figure 7.19 shows the spectrum of the quantization noise for a first order sigma delta with a 1 kHz sinewave applied at the input. It is obvious that its spectrum is not strictly white; however, in the low frequency range the spectrum is reasonably constant. Figure 7.20 shows the spectrum of the noise at the output of the first order modulator; it can be noted that the shaping follows the behaviour that was foreseen, however, discrete frequency components with relevant amplitude affect the spectrum.

### 7.6 Second order sigma delta modulator

The noise shaping of the modulator considered in the previous section can be further improved by using the cascade of two integrators around the noise shaper loop \cite{9}. This configuration is referred to as "second order modulator"; its typical block diagram is shown in Figure 7.21. It consists of the cascade of two sampled data integrators, the first without delay, the second with delay. The second integrator, by the feedback connection from the 1 bit D/A converter, is damped. This configuration avoids stability problems otherwise present when the cascade of two integrators is used in a feedback loop. The analysis of the schematic, in the z-domain, gives:

\[
\begin{align*}
\sigma(z) &= [\sigma(z) - \mathcal{Y}(z)] \frac{1}{1 - z^{-1}} \\
\mathcal{Y}(z) &= [\sigma(z) - \mathcal{Y}(z)] z^{-1} \\
\mathcal{Y}(z) &= \sigma(z) + \epsilon_d z^{-2} 
\end{align*}
\]

by solving the system we get:

\[
\mathcal{Y}(z) = \sigma(z) z^{-1} + \epsilon_d z^{-2} (1 - z^{-1})^2 
\]

It should be noted that the signal is passed through a simple delay, while the quantization noise is passed through the square of the transfer function of the already analyzed first order modulator.
The same transfer functions expressed by (7.30) can also be obtained with different schematics; different architectures are often derived and used in order to optimize the dynamic range at the output of the two integrators. For the schematic shown in Figure 7.21 the voltage at the output of the first op-amp can extend up to \( \pm 5 \text{V ref} \) and the voltage of the second op-amp up to \( \pm 4 \text{V ref} \). Such large output swings can result in a limitation in practical circuits which can be removed by adopting a different modulator scheme.

As for the already considered first order modulator it is worthwhile calculating the in-band total power of the shaped quantization noise. Again using the approximation \( \sin(x) \approx x \) (justified by the condition \( B_{in} \ll f_s \)) we get:

\[
N_{in,x} = \int_{-f_s}^{f_s} \frac{1}{2} (2nR)^2 \left\{ \frac{1}{12} \pi f_s^3 + \frac{1}{12} \pi f_s^4 (R_{in})^2 \right\}
\]

The signal to noise ratio, SNR, for input signals with uniform amplitude probability

\[
C_e \times n_{in} = \text{SNR} / 12
\]

is calculated by:

\[
\text{SNR} = 15.05 \log_{10}(C) - 12.0 \text{ [dB]}
\]

Hence with a second order modulator the converter can gain two and a half bits of accuracy by doubling the clock frequency. The strong reduction in the quantization noise in the input band is achieved at the expense of increasing the requirements on the specifications of the digital filter which rejects the noise out of band. Since the output data must be decimated, the contribution that will be folded back in band base (aliasing) must be a negligible fraction of the term left in the band base by the shaping.

7.7 Multistage sigma–delta modulator

The modulator analyzed in the previous section was derived from a first order scheme by adding a second integrator to the feedback loop. The scheme of the modulator was designed in such a way that the signal transfer function is a simple delay while the noise transfer function becomes the squared of a sine shaping. The procedure can be generalized to improve the reduction of the noise in the low frequency band \([10][11]\). However, it is not advisable to put three integrators in a feedback loop because of the stability problems involved. Higher–order modulators are usually realized with a cascade of low–order modulators (1st or 2nd order). The outputs of the modulators are suitably combined to obtain a high–order noise transfer function. This approach is advantageous because it allows modularity in the architecture, improves the dynamic range and allows unconditional stability of the circuit. However, the high–order noise shaping is achieved by a perfect matching of elements. Errors due to mismatch rapidly degrade the performance of the modulator.

In order to explain the technique, let us consider the scheme shown in Figure 7.22. It is a third order modulator, popularly known by its acronym MASH \([12]\). It is composed of three first order modulators which provide both the digital data and the quantization noise at their output (Fig. 7.23). The quantization noise of the first modulator is used as an input of the second modulator and the quantization noise of the second modulator is an input for the third one. The three outputs are then combined by a digital filter.

The architectures used in the first order modulators give rise to the following equations:

\[
\begin{align*}
C_1 &= X + (1 - z^{-1}) \omega_0 & C_3 &= X + (1 - z^{-1}) \omega_0 \\
C_2 &= \omega_0 + (1 - z^{-1}) \omega_0 & C_4 &= \omega_0 + (1 - z^{-1}) \omega_0
\end{align*}
\]

the output of the digital filter is:

\[
\begin{align*}
C_4 &= C_1 + (1 - z^{-1}) C_2 + (1 - z^{-1})^2 C_3 = X + (1 - z^{-1})^2 \omega_0 \\
C_5 &= X + (1 - z^{-1}) (1 - z^{-1}) \omega_0
\end{align*}
\]

The result is that the contributions of the quantization noise of the first and the second modulator are cancelled and the noise of the third integrator is shaped with the transfer function \((1 - z^{-1})^2\).
The method described is also used with basic blocks made of second order modulators. As for the use of first order cells, it is necessary to generate the quantization noise that is then processed by a successive stage. From knowledge of the signal and noise transfer functions it is possible to find a given digital processing to cancel the effect of the quantization noise of the inner stages. It is evident that the weak point of the method lies in the generation of the quantization noise; as shown in the schematic of Figure 7.23 it is obtained by subtracting the output of the op-amp from the

![Schematic of the basic cell of the sigma-delta architecture](image)

**Figure 7.23 Schematic of the basic cell of the sigma-delta architecture**

A converter data. Any error in this operation is reflected in an imperfect cancellation of the noise. For example, if in the first integrator in Figure 7.22 the generated quantization error is 

\[ e_t(F) \]

instead of the correct value 

\[ e_t(1) \]

the residual part 

\[ e_t(2) \]

will be filtered out only by the noise transfer function 

\[ (1-z^{-1}) \]

Thus, in general, the residual quantization error of the first modulator is shaped only by a first-order transfer function, the residual error of the second integrator by a second order transfer function, and so on.

On the basis of the above observation it seems advisable to use second order modulators as elementary stage of the system.

**7.8 Non-ideal effects in sigma delta modulators**

The basic blocks used in sigma delta modulators have been, up to now, considered as ideal elements. However, all the real elements are affected by limitations which reflect non-idealities in the system where they are utilized. We will concentrate on switched capacitor modulators, where operational amplifiers, comparators, switches and capacitors are employed.

A real operational amplifier is limited because its DC gain, bandwidth and slew-rate are finite: its output impedance is far from zero (very often transconductance operational amplifiers, OTAs, are utilized) and because the op-amp itself is noisy (the noise performances are described by an input referred noise generator). A comparator is affected by limitations similar to the ones of the op-amp, with, in addition, the presence of hysteresis in the input-output characteristic. Switches are limited by a finite on-resistance, at which a white noise source is associated. Moreover, for switches, a coupling between the driving logic and the analog section (clock-feedthrough effect) always affects the operation. The matching in integrated capacitors is quite good, however small inaccuracies, as well as non-linearities and parasitic elements, can become significant for high resolution data converters [13].

The non-idealities recalled above determine some effect on the modulator performances. In this section the more important of them are considered.

**7.8.1 Finite gain and leakage in integrators**

Let us consider the conventional switched capacitor integrator shown in Figure 7.24. As known, in the ideal case, the injecting capacitor \( C_1 \) is pre-charged at the input signal (phase 1) and during the phase 2 it is completely discharged through the virtual ground. However, in a real case, because of the finite value of the DC gain the bandwidth and the slew rate of the op-amp, the charge is not completely transferred. Moreover, a mismatch between the capacitors \( C_1 \) and \( C_2 \) determines an additional error. If the above limits are taken into account, the integrator, in the time domain, is described by:

\[ V_{in}(n+1) = (1-i)\cdot V_{in}(n) - \frac{C_2}{C_1} \cdot i \cdot V_{in}(n) \]

**Figure 7.24 Switched capacitor integrator**

where \( i \) and \( n \) are suitable parameters; in the \( z \) domain, assuming \( C_1 \) and \( C_2 \) nominally equal, it results:

\[ H(z) = \frac{V_{in}(z)}{V_{in}(z)} = \frac{(1-n)z^{-1}}{1-(1-\delta)z^{-1}} \]

where \( n \) is the virtual ground and \( \delta \) is the mismatch.

\[
\text{where } \frac{C_1}{C_2} = (1-n).
\]
With respect to the ideal transfer function, equation 7.35 reveals a gain error $\epsilon_a$ and a leakage error $\epsilon_b$. A similar equation can be derived for the non-inverting integrator.

Using the above result in the first order modulator shown in Figure 7.17, the output signal, as a function of the input $v(z)$ and the quantization noise $v_Q$, becomes:

$$y(z) = \frac{1}{1 + (b_1 + b_2)z^{-1}} \left[ (v(z)z^{-1} + [1 - z^{-1}] + \delta z^{-2}) \right]'$$

at low frequency ($f < f_s$) the effect of the denominator is negligible. Thus, beside the expected shapings, the quantization noise is also multiplied by the term $b_1 z^{-1}$. It corresponds to an additional unshaped noise component proportional to the leakage error.

If gain and leakage errors are taken into account in the second order modulator we get:

$$y(z) = \frac{1}{D(z)} \left[ (v(z)z^{-1} + [(1 - z^{-1}) + \delta z^{-2}])'(1 - z^{-1}) + \delta z^{-1} \right]'$$

where $b_1$ and $b_2$ are the leakage factors of the two integrators and $D(z)$ is a second order function. Again, at low frequency, $D(z)$ has negligible effects. The resulting noise transfer function is made of three terms:

$$N(z) = (1 - z^{-1})^2 = (b_1 + \delta z^{-1})'(1 - z^{-1}) + \delta_2 z^{-2}$$

they correspond to three additive components in the output noise. The first one is the same as in an ideal modulator. The second one corresponds to the quantization noise attenuated by $(b_1 + \delta_2)$ and shaped by a first order modulator; the third term is the quantization noise only attenuated by $b_1 b_2$.

### 7.8.2 Electronic noise

The relevant noise sources in sigma delta modulators are the input referred noise generator of operational amplifiers and the noise source associated to the on-resistance of switches. Since the modulator is a discrete time network, all the noise sources are sampled every clock cycle. Then, their effect is transferred to the output by a proper sampled data (noise) transfer function. It is worthwhile to observe that, before the sampling, the noise generators are filtered by a continuous-time action. It takes place, very often, because of the finite on-resistance of switches and of finite bandwidth of the op-amps. However, the noise band is limited at frequencies that are well beyond the sampling frequency, consequently, the spectrum will be aliased into the band base. The simple case where the band
e limitation comes from the on-resistance of the switch and the sampling capacitor, is studied in this book in the Chapter dealing on "Data Converters". For it, the thermal noise of the on-resistance results into a white sampled-data noise which power over the Nyquist band is $kT/C$ ($k$ is the Boltzmann constant, $T$ is the absolute temperature, $C$ is the sampling capacitor) [14][15].

The sampled-data transfer functions acting on the noise critically depend on the point where the noise itself is injected. If the noise is applied immediately after the quantizer it will be shaped in the same way as the quantization noise. By contrast, if the noise is injected at the input of the modulator it will be treated like signal. It immediately turns out that the noise sources the the first stage of the modulator are much more critical than the ones in successive stages, since they are not favourably processed.

To be more specific, let us consider the first order modulator shown in Figure 7.25. The effect of the noise sources associated to the two switches exercised during the phase 1 is considered. As already pointed out, the continuous time filtering and the successive sampling results into a white sampled-data noise which power is $kT/C_1$. The related noise transfer function, being the switches in series with the input signal, is just the signal transfer function, $z^{-1}$. Therefore, the only rejection of the $kT/C_1$ power comes from the digital filter, used after the analog modulator. Assuming an ideal low pass response, as shown in Figure 7.25, the output noise that is contributed by the two considered switches is given by:

$$\sigma_{in_{mod}}^2 = \frac{kT \cdot 2B_{in}}{C_1 \cdot R_{in}} = \frac{kT}{C_1 \cdot R_{in}}$$

for $C_1 = 1 pF$ and $R_{in} = 128$, $V_{in_{mod}}$ is equal 5.7 mV, a value that becomes comparable to 1/21 LSB for a 16 bit (Vref = 1V) data converter. Thus, for high resolution it is necessary to use, at least in the input stage of the modulator, large capacitances.
Like the noise injected into inner points of the noise shaping loop, outcomes of non idealities, such as, for example hysteresis and threshold variations in the comparator are attenuated in the low frequency range.

### 7.8.3 Sampling jitter

The input stage of a modulator samples the input analogue signal at the oversampled frequency. The exact sampling instant occurs when the sampling switch goes off. Since this switch is realized by an MOS transistor, it goes off when the gate voltage is larger (or smaller depending on the transistor type) than the input signal by just the threshold. If the falling edge (or the rising edge) of the driving signal is not negligible the sampling instant will depend on the input amplitude. Moreover, the driving signal can displace a jitter. These effects results in a non uniform sampling that is usually described with an additional noise source: the jitter noise, $\xi_j$. In order to estimate its relevance, let us consider the sampling of a sinusoidal input. If the sampling instant, $T$, is affected by an error, $\delta$, its equivalent jitter noise, $\xi_j$, is given by:

$$
\xi_j = V_{in}(T + \delta) - V_{in}(T) = A \left( \sin(\omega_m (T + \delta)) - \sin(\omega_m T) \right) \equiv
$$

$$
= \omega_m \delta A \cos \omega_m T
$$

which indicates that the random variable $\delta$ is modulated at the input frequency. However, if $\delta$ is a Gaussian process its spectrum is white and the modulation is irrelevant.

The jitter noise, introduced in the circuit, is transferred to the output in the same way as the electrical noise sources: it is passed through a suitable sampled-data transfer function. Remembering the results discussed for the electrical noise we can immediately state that the dominant spur arises from jitter in the input stage: the sampled data transfer function does not introduce any low frequency attenuation and the noise power is only reduced by the decimation process.

### 7.9 Digital decimation

The purpose of a digital filter cascaded to a sigma delta modulator is twofold [16]:

- to remove the shaped quantization noise out of the base band
- to prevent the noise aliasing in the successive decimation. After the decimation the rate of the digital data is twice the band of the input analogue signal and is represented by a given number of bits; they are completely meaningful only if the total residual noise is smaller than 1/2 LSB.

A very common architecture for digital decimators in sigma delta converters is shown in Figure 7.26. It is made of the cascade of many stages.

![Figure 7.26 Architecture of a typical digital decimator](image)

Each of them introduces a given decimation factor. Usually, $M_1 > M_2 > \ldots > M_n$. The obtained decimation is, of course:

$$
M = \prod_{i=1}^{n} M_i
$$

Every decimator is preceded by a low pass filter for getting in each stage the two before mentioned actions.

In order to intuitively understand the low pass filtering needs, Figure 7.27 shows examples of the outcome of different filters followed by a by-4 decimator. The considered in put is the quantization noise shaped by a first-order sigma-delta modulator (Fig. 7.27a). Without any filtering, the decimation results in the band-base spectrum shown in Figure 7.27b. The noise is strongly dominated by the terms that are folded from the upper bands, $b_2, b_3, b_4$; they are summed up in almost a white spectrum. It can be concluded that the advantage of using a sigma delta modulator is completely vanishes. The filtering action shown in Figure 7.27c determines an improvement of the situation: thanks to a zero placed at $\sin/4$ the noise that is folded at DC frequency is completely cancelled; however, in the low frequency spectrum the folded upper bands. The stronger filtering action shown in Figure 7.27c permits to push folded components well below the unfolded term; thus allowing full benefit by the modulator shaping.

As shown above it is necessary to have in the transfer function of the digital filter single or multiple zeros. This is get by the use of FIR architectures. Very often in the first stage of the decimator chain it is used a sinc filter (comb filter), since its realization does not need multipliers. The transfer function of a sinc filter is expressed by:

$$
D(z) = \frac{1}{N} \left( \frac{1 - z^{-N}}{1 - z^{-1}} \right)^{k}
$$

where $k$ is the order of the sinc-filter. In order to have a negligible amount of aliased noise the order of the sinc-filter must be at least 1 larger than the modulator order. The simple sinc filter gives rise to some attenuation in the base band. The effect is usually corrected in the following stages.
7.10 References


