

Chapter 2

Sigma-Delta Modulation

2.1 Introduction

The Sigma-Delta Modulator ($\Sigma\Delta\text{M}$) is a type of data converter that, when used as an ADC, uses oversampling, sampling the input signal at a rate several times higher than the Nyquist rate, to achieve higher resolution, at the cost of speed, when compared with ADCs that use the Nyquist rate, such as flash, pipeline, or successive approximation ADCs.

The Nyquist rate is defined as $f_N = 2B$ where B is the signal bandwidth and f_N is the sample rate. This means that two samples are taken over a period of the input signal. In the case of the $\Sigma\Delta\text{M}$, hundreds of samples may be taken over the same period due to oversampling. Higher resolution is achieved by using digital processing techniques, instead of complex and precise analog circuits, making the modulator independent of component matching or precise sample-and-hold circuitry (S/H), needing only a small amount of analog circuitry.

The typical process of an ADC consists in filtering the input signal, to minimize aliasing effects, sampling, quantization, and digital encoding of the input signal. In the case of the oversampling ADC, no dedicated sampling circuit is needed, a modulator does the quantization, and, usually, a digital filter performs the encoding of the signal. Figure 2.1 shows the typical block diagrams for both of these types of ADCs.

One of the main differences between these two types of ADCs is that the oversampling converter does not require a complex high order anti-aliasing filter. That is because it samples the signal at many times its bandwidth. Aliasing happens because when sampled, a signal is reproduced, in the frequency domain, at multiples of the sampling frequency, as band-limited signals. When using Nyquist rate, these signal reproductions are very close together, as seen in Fig. 2.2a. This may lead to

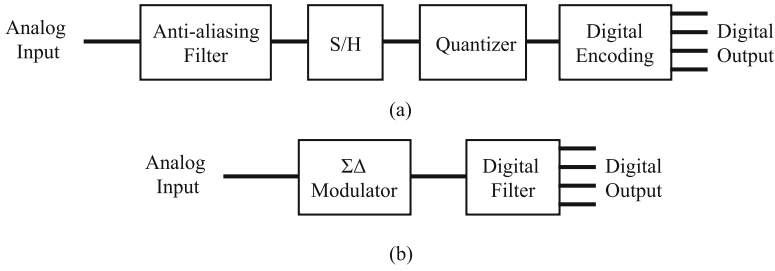


Fig. 2.1 Block diagram for (a) typical Nyquist rate ADCs and (b) for oversampling ADCs. Adapted from [6, Sect. 29.2.6, p. 1008]

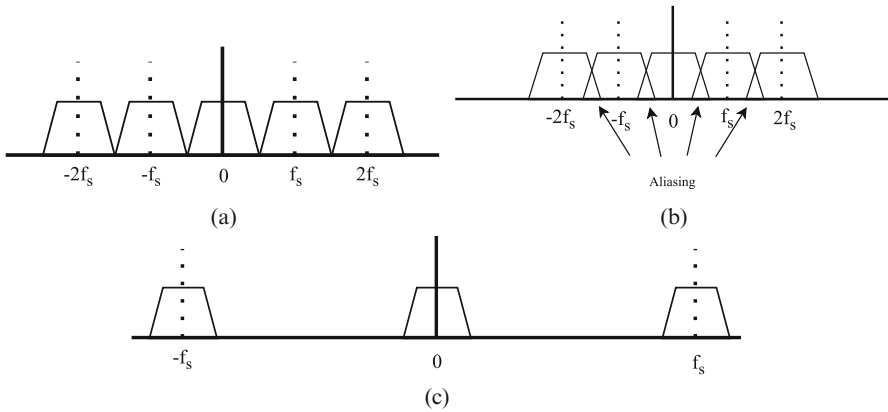


Fig. 2.2 Frequency domain for (a) Nyquist rate ADC, (b) the aliasing effect, and (c) oversampling ADC. Adapted from [6, Sect. 29.2.6, p. 1008]

aliasing problems, which is the overlapping of the signal reproductions, as seen in Fig. 2.2b. This does not occur when oversampling is used, since the signal reproductions are very far apart, as seen in Fig. 2.2c.

Oversampling ADCs are normally constructed using switched-capacitor (SC) circuits and the modulator's output is obtained in real time, so there is no need of a dedicated sample-and-hold circuit.

The modulator quantizes the input as a pulse-density modulated signal. This density represents the average value of the input over a specific period. When the input signal value increases, the amount of high pulses also increases, and vice versa. Figure 2.3 represents the output of the modulator for the positive half of a sine wave, where this behavior is shown.

This output is then processed digitally to filter quantization noise and spurious signals out of the band. Lastly, the signal is downsampled to the Nyquist rate, transforming it in digital data that represents the average value of the analog voltage at the input. The values of the Signal to Noise ratio (SNR) and dynamic range (DR) obtained determine the resolution of the converter [6, Sect. 29.2.6, p. 1007].

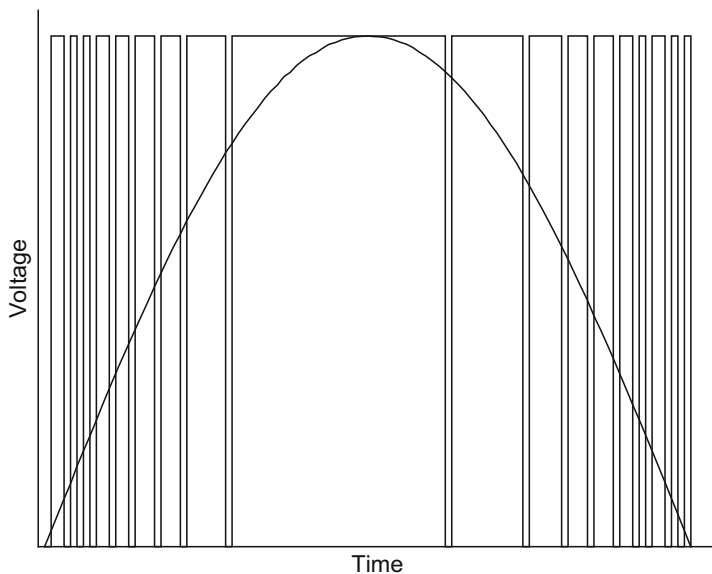


Fig. 2.3 Digital output signal of the $\Sigma\Delta M$ for a positive half sine wave

The SNR value is often used to determine the ENOB of the modulator using Eq. (2.1), which represents the typical relationship between SNR and ENOB for ideal Nyquist rate converters, excited with a sine-wave signal [1, Sect. 1.1, p. 4].

$$\text{SNR} = 6.02\text{ENOB} + 1.76 \quad (\text{dB}) \quad (2.1)$$

An important parameter of the oversampling converter is the oversampling ratio (OSR), shown in Eq. (2.2), where B is the signal bandwidth and f_s is the sampling frequency. This value defines how much faster is the signal sampled by the oversampling converter, when compared with the Nyquist rate one [1, Sect. 1.2, p. 8].

$$\text{OSR} = \frac{f_s}{2B} \quad (2.2)$$

2.2 First Order Sigma-Delta Modulator

The basic structure of a first order $\Sigma\Delta M$ is presented in Fig. 2.4a. It is built using an integrator, a 1-bit ADC and a 1-bit DAC in the feedback path. The 1-bit ADC is a comparator that produces either a high or a low output. The 1-bit DAC uses the comparator output to determine what is summed with the input, either $+V_{\text{REF}}$ or $-V_{\text{REF}}$. The labeled signals are expressed in terms of the sampling time, T , and an integer, k , meaning these are discrete time signals.

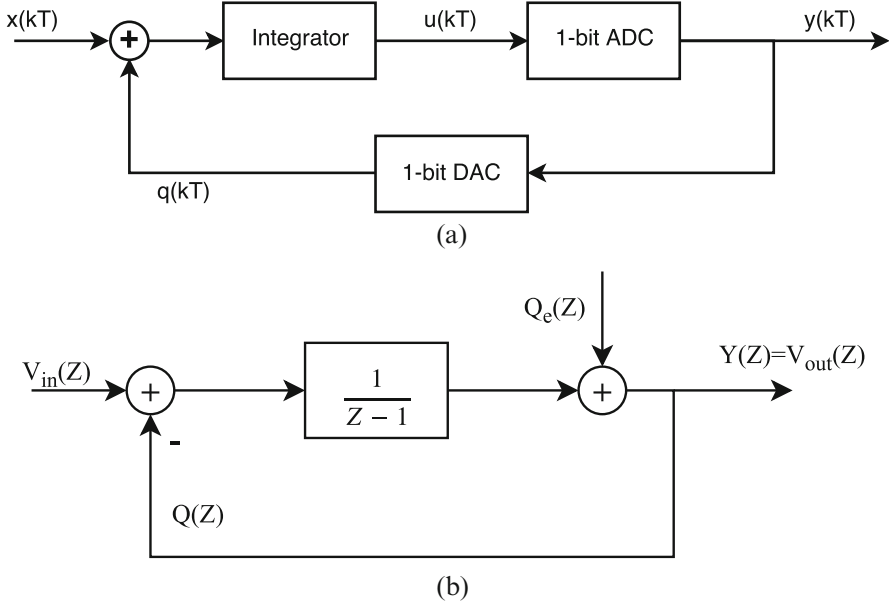


Fig. 2.4 First order $\Sigma\Delta M$ structure (a) and its frequency domain model (b)

To obtain the expression of the output signal, $y(kT)$, ideal components are considered. The integrator simply sums its previous input and output signals:

$$u(kT) = x((k-1)T) - q((k-1)T) + u((k-1)T). \quad (2.3)$$

The ideal 1-bit ADC has a quantization error defined by the difference between its output and its input:

$$Q_e(kT) = y(kT) - u(kT). \quad (2.4)$$

The ideal 1-bit DAC produces a positive or negative reference voltage, according to the logic level of its input. Since these logic levels are defined using the same reference voltages, it's easy to assume that the ideal 1-bit DAC is a unity gain block, its output is equal to its input:

$$q(kT) = y(kT). \quad (2.5)$$

Using Eqs. (2.3) and (2.4) in Eq. (2.5), the output of the first order $\Sigma\Delta M$, $y(kT)$, is defined by:

$$y(kT) = x((k-1)T) + Q_e(kT) - Q_e((k-1)T). \quad (2.6)$$

Equation (2.6) shows that the output of the modulator is equal to the quantized value of the input, delayed by one sampling period, plus the difference between the current and late quantization error. This means that the quantization noise cancels itself to the first order [6, Sect. 29.2.6, p. 1011].

Considering the ideal frequency domain model of the modulator, shown in Fig. 2.4b, these results may once again be reached. The modulator is modeled in the z -domain, in which the integrator is modeled by its transfer function, $\frac{1}{z-1}$, the 1-bit ADC is assumed to be a quantization error source, $Q_e(z)$, and the 1-bit DAC is modeled in a way that $Q(z)$ is equal to the output $Y(z)$, as seen previously. Using feedback theory on the model, the output becomes:

$$V_{\text{out}}(z) = z^{-1}V_{\text{in}}(z) + (1 - z^{-1})Q_e(z). \quad (2.7)$$

The output can be written in the general form:

$$V_{\text{out}}(z) = \text{STF}(z)V_{\text{in}}(z) + \text{NTF}(z)Q_e(z). \quad (2.8)$$

The $\text{STF}(z)$ and $\text{NTF}(z)$ correspond to signal and noise transfer functions, respectively. In Fig. 2.5 is plotted their squared magnitude, by setting $z = e^{j2\pi f}$, and considering the sampling frequency, f_s , equal to 1. This yields:

$$|\text{NTF}(e^{j2\pi f})|^2 = [2\sin(\pi f)]^2. \quad (2.9)$$

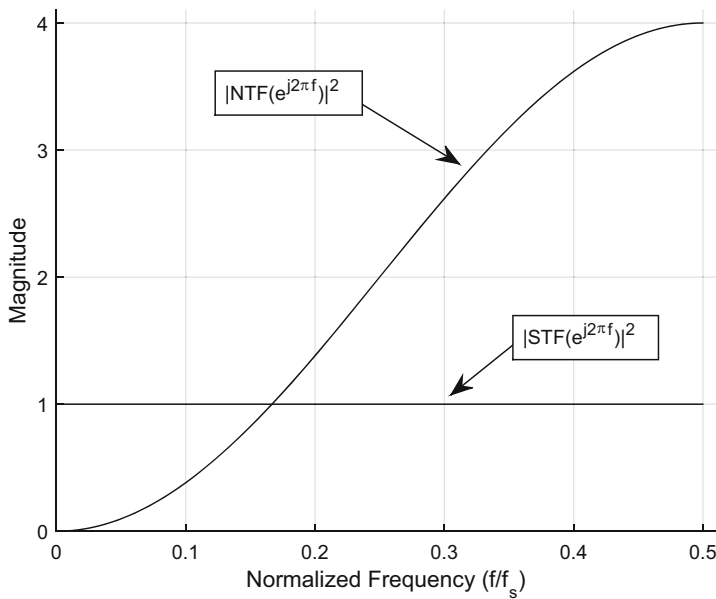


Fig. 2.5 Frequency response of the first order $\Sigma\Delta\text{M}$

Equation (2.9) can be approximated by $|NTF|^2 = (2\pi f)^2$, for $f \ll 1$. Figure 2.5 and Eq. (2.7) show that, in the signal bandwidth, which is as close to 0 as higher the OSR is, the signal maintains its strength, suffering only from a one sample delay, and the quantization noise is heavily attenuated and pushed to higher, out-of-band frequencies, where it is amplified, showing a high-pass characteristic. This is called *noise shaping* and is characteristic of $\Sigma\Delta$ modulation and it is the reason of its effectiveness. A low-pass filter must then be used at the output to reduce this quantization noise and obtain the final high resolution output by downsampling. These operations are most often realized by *sinc filters*, which are implemented in stages [1, Sect. 2.9, p. 54].

Considering a large, randomly varying input at the DAC, the error can be treated as white noise with mean-square value $\sigma_e^2 = \frac{1}{3}$ and a 1-sided power spectral density of $S_e(f) = 2\sigma_e^2 = \frac{2}{3}$. In this scenario, the in-band noise power of the output is approximately:

$$\sigma_q^2 = \int_0^{1/(2 \cdot \text{OSR})} |NTF|^2 \cdot S_e(f) df = \frac{\pi^2}{9(\text{OSR})^3}. \quad (2.10)$$

Assuming a sine wave as input, with peak amplitude A , the output signal power is $\sigma_u^2 = \frac{A^2}{2}$, since $|STF|^2 = 1$. The SNR is then given by:

$$\text{SNR} = \frac{\sigma_u^2}{\sigma_q^2} = \frac{9A^2(\text{OSR})^3}{2\pi^2} \quad (\text{dB}). \quad (2.11)$$

Equation (2.11) shows that when the OSR doubles, the SNR increases by a factor of 8, equivalent to 9 dB, and, according to Eq. (2.1), the resolution, ENOB, is increased by 1.5 bits. This means that, even for a relatively high OSR value, for instance 256, will result in a relatively low SNR, less than 70 dB [1, Sect. 2.4, pp. 36–38].

Its important to note that the nonlinear distortion of the DAC is not affected by the *noise shaping*, thus limiting the overall performance of the modulator. The simplest solution to this problem is using single-bit quantization, using, for instance, a comparator as the ADC. This leads to the DAC's operation becoming inherently linear, since it only consists of two points. In this case, and with an input signal with magnitude below or equal to 1, it is proved that the first order $\Sigma\Delta\text{M}$ is stable [1, Sect. 1.2, p. 7; Sect. 2.7, p. 49].

2.3 Second Order Sigma-Delta Modulator

Figure 2.6a shows the structure of a second order $\Sigma\Delta\text{M}$. It is obtained by replacing the quantizer, the 1-bit ADC, in the first order $\Sigma\Delta\text{M}$ with a copy of a first order $\Sigma\Delta\text{M}$. Considering again ideal components, the output, $y(kT)$, takes the form:

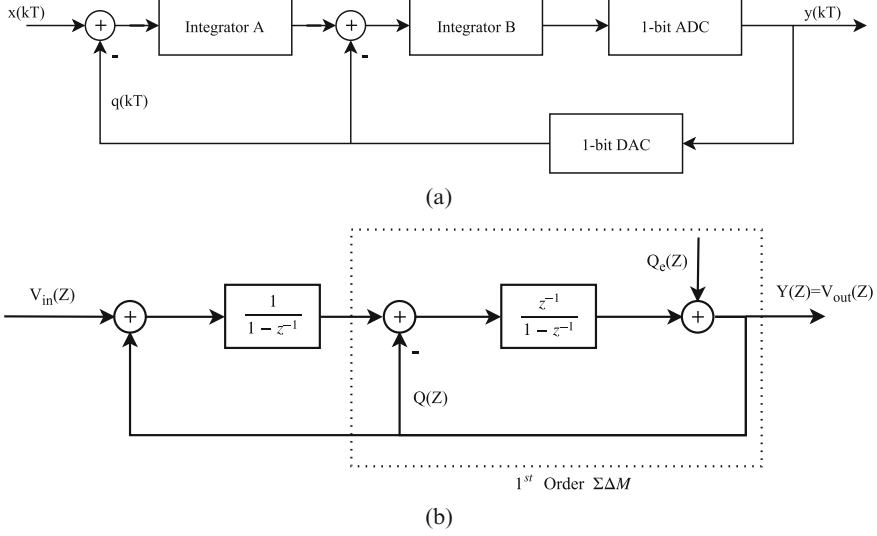


Fig. 2.6 Second order $\Sigma\Delta M$ structure (a) and its frequency domain model (b)

$$y(kT) = x((k-1)T) + Q_e(kT) - 2Q_e((k-1)T) + Q_e((k-2)T). \quad (2.12)$$

Equation (2.12) shows that the output is still a delayed quantized input signal, but this time with a second order differencing of the quantization noise. Figure 2.6b shows its linear z -domain model, from which the NTF and the STF can be obtained:

$$V_{out}(z) = \text{STF}(z)V_{in}(z) + \text{NTF}(z)Q_e(z) = z^{-1}V_{in}(z) + (1-z^{-1})^2Q_e(z). \quad (2.13)$$

The squared magnitude of the NTF is then given by:

$$|\text{NTF}(e^{j2\pi f})|^2 = [(2\sin(\pi f))]^4 \approx (2\pi f)^4, \text{ for } f \ll 1. \quad (2.14)$$

In Fig. 2.7, the squared magnitudes of the STF and NTF are plotted, showing that, like in the first order $\Sigma\Delta M$, the signal is unaffected and that the quantization noise is highly attenuated in the band of the signal, near 0, but is highly boosted in high, out-of-band frequencies, more than in case of the first order $\Sigma\Delta M$. This means that the total quantization noise power at the output of the second order $\Sigma\Delta M$ is greater, but since it is located out of the signal bandwidth, it can be easily filtered, by a digital decimation filter at the output of the system.

The in-band noise power of the modulator can be calculated similarly as in the first order $\Sigma\Delta M$:

$$\sigma_q^2 = \int_0^{1/(2\text{OSR})} |\text{NTF}|^2 \cdot S_e(f) df = \frac{\pi^4}{15(\text{OSR})^5}. \quad (2.15)$$

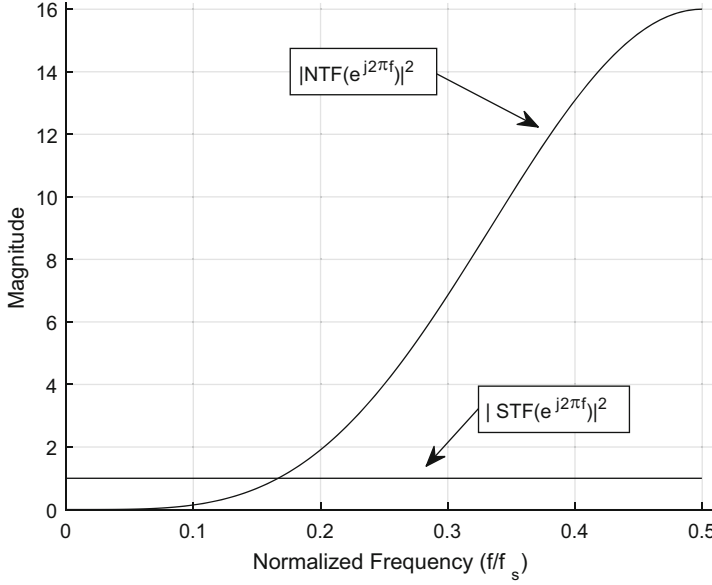


Fig. 2.7 Frequency response of the second order $\Sigma\Delta M$

Considering again a sine wave as an input signal with power $\sigma_u^2 = \frac{A^2}{2}$, the SNR of the second order $\Sigma\Delta M$ is thus:

$$\text{SNR} = \frac{\sigma_u^2}{\sigma_q^2} = \frac{15A^2(\text{OSR})^5}{2\pi^4} \quad (\text{dB}). \quad (2.16)$$

From Eq. (2.16), the doubling of the OSR leads to an increase by a factor of 32, equivalent to 15 dB, of the SNR, which translates, according to Eq. (2.1), in an increase of 2.5 bits in resolution, ENOB. This means that to achieve the same value of SNR, the second order $\Sigma\Delta M$ needs a lower OSR, leading to a lower sampling frequency, when compared with the first order $\Sigma\Delta M$. The downside is the increased total noise power, as stated previously.

If a single bit quantizer is used, the second order $\Sigma\Delta M$ is inherently linear, under the same conditions as the first order $\Sigma\Delta M$, that is the magnitude of the input signal is lower than 1. The nonidealities of the components are treated as noise and are affected by the NTF, and shifts in component values lead to a shift in the pole and zero locations of the STF and NTF, not into nonlinearities [1, Sect. 3.1, pp. 63–66].

The output filtering and downsampling are once again achieved using decimation filters, normally based on cascaded sinc filters. These filters can be used for any L th order modulator, and to achieve adequate results, its order, K , must be higher than the modulator's order, that is $K > L$ [1, Sect. 3.5, p. 86].

2.4 Higher Order Sigma-Delta Modulator

Higher order modulators offer increased resolution and more noise shaping, for the same OSR. The amount of quantization noise added is increased, but more is pushed to higher frequencies, decreasing the noise in the signal band. The output is still a delayed version of the input with a differencing of the quantization noise proportional to the order of the modulator.

The easiest way to construct an n -order modulator is to simply adding $n - 1$ integrators to the structure of the first order SDM. However, because this structure employs feedback loops, stability starts to become a critical issue [6, Sect. 29.2.6, p. 1014].

The range of the input signal magnitude in which the modulator works as intended is called stable input range. This range is determined by the full-scale range of the feedback DAC, usually being a few dB lower for single-bit quantization. This happens because when the input signal of the modulator approaches the edge of the overload region of the quantizer, extra quantization noise added may push the quantizer into the overload region. Because of feedback, the modulator enters a vicious cycle causing the quantizer to overload and leading to the saturation of the active blocks, turning the modulator unstable. Stability may not be restored by returning the input to the stable input range, thus it is often required an external intervention, such as a reset of the integrators in the modulator.

The stability properties of a single bit modulator are ruled by its NTF. However, there are no linear models that can predict stability with total reliability due to the signal dependency of the quantizer gain, which cannot be captured by any linear model. Extensive simulations using worst case input signals must be performed to study stability.

There are some criteria which can help predict instabilities, but are either too conservative or only work for specific modulators. The most widely used is the *Lee criterion* [7], which is neither necessary nor sufficient to determine stability, has no solid theoretical foundations, but, due to its simplicity, it is very popular [1, Sect. 4.2, pp. 97–103].

One way to deal with the stability issues is to use a multi-bit quantizer. However, its complexity rises exponentially with the number of bits, limiting it to 4 or 5 bits maximum.

A different approach is to cancel the quantization noise, rather than filter it, using multi-stage, or cascade, structures. Several such structures exist, like the *Leslie-Singh Structure* [8], and the one used in this work is called MASH (for Multi-stage-noise-SHaping), which will be discussed in the next section [1, Sect. 4.5, p. 122].

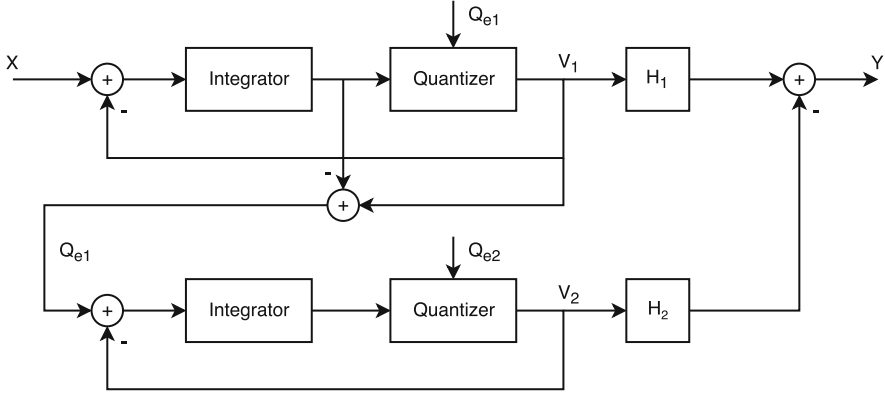


Fig. 2.8 Structure of a 2-stage 1-1 MASH $\Sigma\Delta M$

2.4.1 The MASH Structure

A Multi-stAge-noise-SHaping (MASH) structure is constructed using $\Sigma\Delta M$ in stages, where the input of the next stage is the quantization error introduced by the quantizer of the current stage, that is the difference between its output and input. The outputs of the individual stages are then digitally filtered and combined in a way that the quantization noise of each stage, excluding the last, is canceled at the overall output of the structure.

In Fig. 2.8 is shown a 2-stage MASH structure using a first order $\Sigma\Delta M$ in each stage. In terms of signal (STF) and noise (NTF) transfer functions, the output of the first stage is:

$$V_1(z) = \text{STF}_1(z)X(z) + \text{NTF}_1(z)Q_{e1}(z). \quad (2.17)$$

Considering that the input of the second stage is the quantization error of the first stage, its output is given by:

$$V_2(z) = \text{STF}_2(z)Q_{e1}(z) + \text{NTF}_2(z)Q_{e2}(z). \quad (2.18)$$

The digital filters H_1 and H_2 must be designed in order to cancel the quantization error, Q_{e1} , of the first stage at the output of the structure, Y . To achieve it, the following condition must be true:

$$H_1 \cdot \text{NTF}_1 - H_2 \cdot \text{STF}_2 = 0. \quad (2.19)$$

The simplest way to satisfy Eq. (2.19) is to make $H_1 = \text{STF}_2$ and $H_2 = \text{NTF}_1$. The overall output is then given by:

$$Y = H_1 V_1 - H_2 V_2 = \text{STF}_1 \cdot \text{STF}_2 \cdot X - \text{NTF}_1 \cdot \text{NTF}_2 \cdot Q_{e2}. \quad (2.20)$$

Considering the transfer functions of the first order $\Sigma\Delta\text{M}$, discussed in a previous section, namely $\text{STF} = z^{-1}$ and $\text{NTF} = 1 - z^{-1}$, Eq. (2.20) becomes:

$$Y(z) = z^{-2} \cdot X(z) - (1 - z^{-1})^2 \cdot Q_{e2}(z). \quad (2.21)$$

Equation (2.21) shows that the structure has a performance of a second order $\Sigma\Delta\text{M}$. If second order $\Sigma\Delta\text{Ms}$ were used instead, the overall structure would have the noise shaping performance of a fourth order modulator. This principle can be extended to a 3-stage structure, where if second order $\Sigma\Delta\text{Ms}$ were used, it would perform as a sixth order modulator. The third stage would cancel the quantization noise from the second and the cancellation conditions are found in exactly the same way.

The overall stability of the structure is determined by the modulators used in each stage, so by using only first or second order modulators, stability becomes less of an issue. In practice, the input of each additional stage must be scaled to fit the stable input range, and its inverse must be included in H_2 for the noise cancellation.

Since quantization error, which is similar to noise, is used as input, the quantization error in that stage is very close to true white noise, even if the first stage noise contains tones. This reduces the need for dithering. Also no harmonic distortion of the signal is generated in these stages and the small noise added by the DAC's nonlinearity is tolerable. This structure also allows the use of multi-bit quantizer in the second or higher stages without correction of the DAC nonlinearity, because it is high-pass filtered by the NTF of the previous stage, suppressing it at the baseband.

To achieve good performance, that is a large SNR value, the MASH structure relies on the accurate cancellation of the quantization noise. This requires a good matching between the transfer functions of the modulators used in the stages and the transfer functions of the digital filters, so that the conditions like in Eq. (2.19) holds. Any mismatch results in a degradation of the noise cancellation and leads to lower SNR values. This is called noise leakage and requires accurate behavioral simulations to understand its effect on the performance of the structure and to reduce it to acceptable levels. The increase in number of stages leads to more noise leakage and more complex equations describing it [1, Sect. 4.5.2, pp. 127–132].

<http://www.springer.com/978-3-319-57032-7>

Design of Low Power and Low Area Passive Sigma Delta
Modulators for Audio Applications

Fouto, D.; Paulino, N.

2017, IX, 78 p. 70 illus., 38 illus. in color., Softcover

ISBN: 978-3-319-57032-7