Fundamental limitations of switched-capacitor sigma-delta modulators

V F. Dias
G. Palmisano
P. O'Leary
F. Maloberti

Indexing terms: Switched capacitor, Modulators

Abstract: In recent years oversampling techniques have become very popular for implementing high resolution analogue-to-digital (A/D) converters. To obtain high resolution the order of the modulator L and the oversampling ratio M are commonly seen as degrees of freedom since no theoretical limit exists for the maximum achievable signal-to-quantisation-noise ratio (SNR). However, in practical cases two fundamental constraints need to be considered in the design of the modulator, namely the thermal noise generated by the switches and the amplifiers and the settling error due to finite values of the gain-bandwidth product and the slew rate. The paper presents a closed form relation for the SNR performance of generic L-order N-bit modulators designed with switched-capacitor (SC) techniques. The relation takes into account all the above constraints. One of the major conclusions is that state-of-the-art 2nd-order SC modulators are sufficient to achieve competitive SNR performance.

1 Introduction

In both research and industrial environments it has begun to be generally accepted that sigma-delta (ΣΔ) is probably the most suitable technique to implement high-resolution A/D converters (>16 bits) [1–9]. The preference for this kind of convertor is justified by the fact that oversampled noise-shaping modulators can efficiently trade speed for accuracy. Indeed, the required analogue processing is reduced to simple integration and low resolution A/D and D/A operations, a fact that is considered very useful by circuit designers since the usual precautions for nonideal effects of Nyquist-rate A/D convertors can be significantly relaxed.

As illustrated in Fig. 1, a ΣΔ-based A/D conversion system is composed of three main blocks:
(a) a low-selectivity continuous-time filter that removes the spurious components which would result in aliasing in the signal band
(b) an oversampling ΣΔ modulator that shapes the signal and the quantisation noise differently
(c) a digital decimator that suppresses out-of-band noise and reduces the sampling rate.

![Fig. 1 Block diagram of ΣΔ A/D convertor](image)

The joint effect of oversampling modulation and decimation operations gives rise to an increase of the in-band SNR. Compared with the simple oversampling technique, the ΣΔ approach provides a faster increase of the SNR value as a function of the oversampling ratio. However, as will be shown later, in high resolution ΣΔ A/D convertors this faster increase is not always true if real characteristics for the circuit elements are considered. It will be shown that for a fixed signal bandwidth an oversampling factor exists over which modulators of any order begin to behave as common oversampled circuits.

Up to now, research efforts on oversampling techniques have been concentrated on three main fields:
(i) development of exact and/or approximate criteria to help modulator and decimator circuit design [10–13]
(ii) development of modulator and decimator architectures to obtain high-order noise filtering and/or reduce nonideal effects associated with circuit elements [1–3]
(iii) development of behavioural simulators to aid ΣΔ designers in evaluating the effects of the real characteristics of circuit elements [14–17].

The work presented here falls into the first and third areas, since the closed form relation obtained for the overall SNR is closely related to the circuit simulation models developed by the authors [17, 18].

Because of the nature of ΣΔ modulators, the first integrator in the loop is the most critical block of the circuit. Indeed, the noise sources introduced at this point behave as an input signal and thus cannot be processed differently by the modulator. Design efforts need to be concentrated on understanding and reducing such spurious effects.

In this paper, a closed form relation is developed which defines the fundamental limits of ΣΔ modulators designed with SC techniques. A model has been derived to quantitatively evaluate the effect of the settling error due to the operational amplifier finite gain-bandwidth product and the slew rate, and careful attention has been given to the quantisation and thermal noise sources.
2 Quantisation noise

The block diagrams of both 1-bit 2nd-order and N-bit L-order ΣΔ modulators are shown in Fig. 2. Apart from the multistage approach and the N-bit quantisation and one bit feedback structure proposed by Leslie et al. [26, 9], an N-bit L-order modulator is composed of a loop

that contains an interconnection of L integrators terminated by an N-bit A/D converter in the forward path and an N-bit D/A converter in the feedback path. For 1-bit modulators, the A/D and D/A converters simply use an analogue comparator that selects one of two reference voltages. The 1-bit 2nd-order modulator presents some advantages over those of higher order N-bit modulators, principally because of the simplicity of the circuit implementation, the stability of the loop, the approximately white quantisation noise [12] and the independence of the SNR from the accuracy of the D/A quantisation levels (two levels only).

Assuming that the quantisation noise is random and uncorrelated with the input signal, a linearised model can be adopted for the quantiser; see Fig. 3a. The output signal ψ(n) is assumed to be the superposition of the quantiser input ω(n) and the quantisation error q(n). As illustrated in Fig. 3b, T(n) and S(n) are intended to model the relevant noise sources of the first integrator in the modulator.

The principal objective of a ΣΔ modulator is to leave unchanged the input signal and to introduce a highpass filtering function on the quantisation noise. A suitable design of an L-order modulator gives rise to a Z-domain relation between the input and the output signals:

\[ Y(z) = z^{-1} \left( X(z) + T(z) \right) + S(z) \left( 1 - z^{-1} \right) Q(z) \]  

(1)

where \( X(z) \) and \( Y(z) \) are, respectively, the input and the output of the modulator and \( Q(z) \) is the quantisation noise associated with the quantiser. If it is assumed that the signal to be quantised is uniformly distributed within each quantisation step, then a noise power \( \sigma^2 = (4/3)A^2 \cdot 2^L \) can be obtained for \( q(n) \) [19], having an associated power spectral density

\[ N(f) = \frac{2}{3F_s} \cdot 2^{-2\cdot\text{ord}^2} \cdot F_s/2 < f < F_s/2 \]  

(2)

where \( F_s \) is the oversampling frequency. If the approximation \( \sin(nB_wF_s) \approx nB_wF_s (B_w \ll F_s) \) is used and the oversampling ratio \( M \) is defined as \( F_s/2B_w \), the shaped contribution at the modulator output results in an in-band power given by

\[ P_0 = \frac{4}{3} \cdot \frac{A^2 \cdot 2^{2L}}{2L + 1} \cdot \frac{1}{M^{L+1}} \]  

(3)

This equation indicates that the quantisation noise power decreases \( 3(2L + 1) \) dB per octave of oversampling and 6 dB per additional bit of resolution of the quantiser — see Fig. 4. For example, a quantisation noise of \(-100 \) dB

\[ \text{quantising noise power} \leq \text{over sampling ratio} \]

(4)

\( (A = 1, N = 1) \) requires an oversampling ratio of about 25 for a 4th-order modulator, whereas an oversampling ratio as much as 145 would be required for a 2nd-order modulator.

3 Thermal noise

An expression for the thermal noise power of oversampled ΣΔ modulators is now derived. This kind of noise represents one of the most limiting factors for the maximum achievable dynamic range (DR) [20, 21]. The noise sources that contribute principally to the overall thermal noise power are the switch resistance \( R_s \), both in the sampling and integrating clock phases, and the input-referred operational amplifier broadband noise. During the sampling phase the noise power frozen in the input capacitor is limited by a 1st-order lowpass filtering of the power spectrum. However, in the integrating phase, the integrator input-referred noise power which comes from the amplifier and the switches is limited by a combined filtering due to the amplifier bandwidth and

IEEE PROCEEDINGS-G, Vol. 139, No. 1, FEBRUARY 1992
the RC paths. Flicker noise is not taken into account since its effects can be strongly reduced by applying special compensation techniques [22]. Only transconductance amplifiers with a single high-impedance node placed at the output are considered here, the frequency compensation being performed by the series of the sampling and integrating capacitors.

3.1 Switch thermal noise
Let us consider the simple lossless integrator in Fig. 5a with the output sampled during phase \( \Phi_2 \). In the sampling phase \( \Phi_1 \), Fig. 5b, the lowpass-type frequency response \( H_s(j\omega) \) of the circuit composed of the two \( R_{n1} \) switch resistances and the sampling capacitor \( C_i \) defines the shaping for noise integration. Since the two noise sources corresponding to the switches \( S_1 \) and \( S_2 \) can be assumed as uncorrelated, the bilateral power spectral density is equal to \( 2 + 2kTR_{n1} \) and the noise power

\[
P_{S_{1,2}} = \frac{kT}{2 \pi} \int_{-\infty}^{\infty} 2(2kTR_{n1}) |H_s(j\omega)|^2 d\omega
\]

is, as known, \( kT/C_i \). Assuming that the cut-off frequency of \( H_s(j\omega) \) is much larger than the sampling frequency \( F_s \), the \( kT/C_i \) noise power is uniformly distributed in the Nyquist band \(-F_s/2 < f < F_s/2\), because of the aliasing phenomena.

As shown in Fig. 5c, in the integrating phase \( \Phi_2 \), the filtering function \( H_s(j\omega) \) for the power spectrum associated with the switches \( S_3 \) and \( S_4 \) depends on the operational amplifier bandwidth, on the input time constant \( 2R_{n1}C_i \), and on the output time constant \( R_{n1} \). The latter contribution is defined by the capacitor \( C_n \) and by both the switch on-resistance and the amplifier closed-loop output resistance \( R_{o_1} \) (see Appendix 9). The amplifier frequency response is \( A_0(j\omega) = A_0(1 + j\omega/C_0) \), where \( A_0 \) is the pole frequency given by \( \sqrt{C_0 C_i} / C_0 \), \( C_0 \) is the transconductance gain of the amplifier and \( C_i \) is the sum of \( C_n \) and \( C_0 \). In the applications considered here the effect of the parasitic capacitances charging the input and the output nodes of the amplifier is negligible since their values are smaller than \( C_i \) and \( C_0 \).

Furthermore, the value of \( C_n \) can be made much smaller than \( C_i \), since in high resolution \( \Sigma \Delta \) modulators the \( kT/C \) noise contributions in integrators, apart from the first one, are highpass shaped.

In the following we analyze the three distinct cases that can be considered for \( H_s(j\omega) \).

Case 1: \( R_{n1} = 0 \) and \( \omega_{g1}, \omega_{g2} \gg \) BW (equivalent bandwidth of the integrator)
Let us neglect, as a first approximation, the output resistance \( R_{o_1} \) and assume equal values for the poles \( \omega_{g1} = (2R_{n1}C_i)^{-1} \) and \( \omega_{g2} = (2R_{n1}C_n)^{-1} \), verifying the condition \( \omega_{g1}, \omega_{g2} \gg \) BW. Under these assumptions, \( H_s(j\omega) \) can be simplified to

\[
H_s(j\omega) = \frac{C_i/C_0}{1 + j\omega/BW}
\]

where

\[
BW = \frac{C_i}{C_0 + C_i} \frac{\omega_{g1}}{\omega_{g2}} \frac{\omega_{g2}}{\omega_{g1}} = \frac{C_i}{C_i}
\]

is the equivalent integrator bandwidth. After integrating as in eqn 3 the power spectral density of switches \( S_3 \) and \( S_4 \), the resulting input-referred noise power is

\[
P_{S_{3,4}} = \frac{kT}{C_i} 2R_{n1}\omega_{g1}
\]

However, the hypothesis that \( \omega_{g1}, \omega_{g2} \gg \) BW is equivalent to having \( 2R_{n1}\omega_{g1} \ll 1 \) and therefore \( P_{S_{3,4}} \ll P_{S_{1,2}} \).

Case 2: \( R_{n1} \neq 0 \) and \( \omega_{g1}, \omega_{g2} \gg \) BW
In reality, a modified output pole frequency given by \( \omega_{g0} = [(2R_{n1} + R_{n2})(2R_{n1} + R_{n2}C_n)^{-1}] \) needs to be considered, where the output resistance \( R_{o_1} = (C_0 + C_i)/C_i \) is now taken into account (see Appendix 9 for frequencies over the amplifier open-loop cut-off frequency). However, in most \( \Sigma \Delta \) modulators \( C_0 + C_i/C_i \approx 1 \), i.e. \( R_{o_1} \approx 1/\omega_{g0} \) and, since \( C_n \ll C_i \), the pole frequency \( \omega_{g0} \) is again much larger than the integrator bandwidth, BW.

Case 3: \( \omega_{g1}, \omega_{g2} \gg \) BW
The optimal sizing for the switches should probably be such that assumes \( \omega_{g1}, \omega_{g2} \gg \) BW. In this way a good trade-off can be obtained between speed requirements and charge injection effects. Under these conditions, a 3rd-order noise filtering function with three poles of about equal frequency can be assumed and which can be approximate by an ideal lowpass filter with cut-off frequency equal to BW. The resulting input-referred noise power is now

\[
P_{S_{3,4}} = \frac{kT}{\pi C_i} 2R_{n1}\omega_{g1} \approx \frac{kT}{C_i}
\]

because \( 2R_{n1}\omega_{g1} \approx 1 \).

Given these conditions, the noise power spectral density of the sampling and integrating switches is

\[
N_{n_s}(f) = \frac{1}{2} \frac{kT}{F_s} \frac{1}{C_i} - F_s/2 < f < F_s/2
\]

3.2 Amplifier thermal noise
A third contribution to the overall thermal noise power in SC integrators is the broad-band noise of the operational amplifier. This noise is usually represented as a
source with power spectral density $2kTR_\text{eq}$ at the positive input terminal [23, 24], where

$$R_\text{eq} = \frac{4}{3} \frac{1}{g_m} \left( 1 + \frac{\alpha}{g_m} \right)$$  

(10)

is the equivalent thermal noise resistance. The parameter $\alpha$ depends on the particular amplifier scheme and takes into account thermal noise sources associated with the amplifier transistors other than those of the differential input stage. As shown in Fig. 6, the effect of this noise source can be calculated by splitting the noise into two distinct sources, placed at the input and output of the integrator [22], and which power results from the integration of the power spectral density adequately filtered. The former represents the noise contribution processed by the integrator while the latter holds for the buffer effect from the amplifier input to output. By assuming the same filtering conditions as for the integrating switches (Case 3), the thermal noise power spectral density associated with both the input and output noise sources is

$$N_{\text{IN,0}}(f) = \frac{1}{F_s} \frac{K T}{C_1} R_\text{eq} g_m = \frac{4}{3} \frac{K T}{F_s} \frac{C_1}{C_1} \left( 1 + \frac{\alpha}{g_m} \right)$$  

(11)

Finally, on using eqns. 1, 9 and 11, the overall in-band thermal noise power of the modulator can be expressed as

$$P_{\text{th}} = \frac{1}{M} \left[ 2 + \frac{4}{3} \left( 1 + \frac{\alpha}{g_m} \right) \right]$$  

(12)

In this equation, the thermal noise source at the output of the first integrator was not considered, since its contribution to the modulator noise is highpass shaped.

As can be seen from eqn. 12, the thermal noise power simply decreases by 3 dB per octave of increase in the oversampling ratio $M$ or in the input sampling capacitor value $C_1$. For the example in Section 2, the 4th order modulator has 8 dB more thermal noise power than the 2nd-order modulator, if equal sized capacitors are used.

4 Setting error

Speed constraints are imposed by the operational amplifier on $\Sigma A$ modulators. The two parameters that characterise the amplifier dynamic performance in the linear and current-limited operation modes are the gain bandwidth product and the slew rate, respectively. As referred to by the authors of References 1 and 2, the simultaneous effect of GBW and SR leads to the loss of proportionality between the amplitude of the input samples and the setting error, causing an increase of the in-band noise power in the output signal. The approach developed here assumes that the setting error is uncorrelated from sample to sample and uniformly distributed in a range $[-\varepsilon_{\text{max}}, \varepsilon_{\text{max}}]$, where $\varepsilon_{\text{max}}$ is the maximum settling error.

In this way, quantitative results can be obtained by modelling the settling error with a white noise source at the output of the integrator. Let us consider the integrator in Fig. 6 during the integrating phase. By applying the charge conservation principle to the virtual ground node during the hold-to-integrating transition, a closed form relation can be obtained for the output voltage $v_d(t)$ [18]

$$v_d(t) = a V_{\text{om}} + b G_i V_{\text{step}}(1 - e^{-t/\tau})$$  

(13)

during the $n$th time-slot with width $T = 1/2F_s$, and where a simple 1-pole model is assumed for the amplifier frequency response. Constant factors $a$ and $b$ depend on the amplifier dc-gain and on the capacitive values, and they approximate unity for $A_p \to \infty$. Factor $G_i$ is the gain of the integrator defined as $C_i/C_0$, $\tau$ is the time constant given by $\tau = 1/BW$, $V_{\text{om}}$ is the initial output voltage and $V_{\text{step}}$ is the input step voltage resulting from both the input and feedback signals.

Slowing limitations arise when the derivative of eqn. 13 is higher than the slew rate $SR$. In this case the output rises linearly up to a point in which equality exists between the SR and the recalculated derivative of the exponential (it is same as imposing a continuity in the derivative). This time instant defines the transition point between the current-limited and the exponential model. By simultaneously modelling these two parameters, the settling error can be calculated as

$$\alpha(t) = SR t e^{-t/(1 - G_i V_{\text{step}}/(SR \tau))}$$  

(14)

defined as the difference between the limit value of $v_d(t)$ for $t \to \infty$ and its value for $t \to T$. Only for input steps that satisfy the condition $G_i V_{\text{step}} < SR$, the time response is purely exponential and the settling error $\alpha(t) = G_i V_{\text{step}} \exp(-t/\tau)$ is found to be proportional to the input steps.

To calculate the effect of the settling error it is necessary to know its power spectral density and the way it is correlated or not with the input signal. A first simplifying assumption could consist in assuming the error samples $\alpha(n)$ uncorrelated with the modulator input signal. Indeed, the integrator input steps result from the difference between the modulator input and output signals and thus they are precisely the samples of the shaped quantisation noise; these samples have been assumed to be uncorrelated with the signal. The settling error power spectral density could subsequently be obtained by means of the known spectra of the input steps and the transfer function that relates them to the error samples $\alpha(n)$ [25]; as we have seen this relation is nonlinear, see eqn. 14.

A set of simplifying assumptions can however be made which consists in assuming that the settling error is uniformly distributed in the range $[-\varepsilon_{\text{max}}, \varepsilon_{\text{max}}]$ and the power spectral density is constant in the interval $[-F_s/2, F_s/2]$. These assumptions form a worst case condition for the settling error since simulation results indicate that, on one hand, the probability density function decreases for increasing amplitudes of $\alpha(n)$ an, on the other, the corresponding spectra is slightly highpass shaped. However, as we shall see later, this model represents quite a good tradeoff between simplicity and accuracy. In these conditions, the power spectral density is as follows

$$N_\alpha(f) = \frac{1}{3} \frac{\varepsilon_{\text{max}}^2}{F_s}$$  

(15)

From eqns. 1, 14 and 15, the in-band settling error noise

IEE PROCEEDINGS-G, Vol. 139, No. 1, FEBRUARY 1992
power at the output of the modulator can be expressed as

\[ P_{st} = K_{sw} \frac{1}{M} e^{-\frac{2\Delta t}{M}} \]  

\[ \text{(16)} \]

where

\[ K_{sw} = \frac{1}{3} \left( \frac{1}{g_{m}} \right) (1 + G)^3 e^{\frac{-2\Delta t}{(1 + G)t_{sw}}} \]  

\[ \text{(17)} \]

In this case, a maximum input step voltage equal to \( 2\Delta \) is assumed, where \( T = \frac{1}{4B_0} \), \( M \) is the integrating time slot, and \( t_{sw} = \text{SR} \), is the slew rate output current.

In Fig. 7 the noise power defined by eqn. 16 is plotted as a function of the oversampling ratio \( M \). Each curve refers to a distinct slew rate value with a time constant of 3 ns (BW \( \approx 330 \text{ Mrad/s} \)), \( B_0 = 22 \text{ kHz} \) and \( \Delta = 1 \text{ V} \). Eqn. 16 shows a double dependence on the oversampling ratio \( M \). The noise reduction factor \( M^{-1} \) is provided by the increase in the sampling frequency and the exponential factor is due to the reduction of the settling time-slot width or, equivalently, due to the increased time percentage in the slew-rate operation mode. For small values of \( M \) the increment in the noise power is very sharp, and the exponential factor prevails; but for large values of \( M \) this increase tends towards saturation. However, for each curve there exists an oversampling ratio value over which the amplifier never leaves the current-limited operation mode. On the other hand, for increasing values of the slew rate value SR the improvement in the noise power saturates, since the time response to the input step approaches the exponential form (in this case the model is no longer valid). What is interesting in this result is that the settling error constitutes a hard barrier in terms of noise power, defining by itself the maximum speed of operation of the circuit.

5 Discussion

Three of the most important limiting factors of switched-capacitor \( \Sigma \Delta \) modulators have been analysed, namely, the quantisation noise \( P_Q \) introduced by the quantiser, the amplifier and switches thermal noise \( P_{th} \), and the settling error noise \( P_{st} \). In Fig. 8, each one of these contributions is illustrated as a function of the oversampling ratio \( M \). Here a 1-bit 2nd-order modulator is considered with signal bandwidth \( B_0 = 22 \text{ kHz} \), quantisation levels \( \Delta = \pm 1 \text{ V} \) and integrator parameters given by \( g_{m} = 2 \text{ mA/V} \), \( C_1 = 6 \text{ pF} \), \( G_1 = 1/2 \) and \( \alpha = 0 \). As can be seen, there are three distinct working regions for the modulator: a noise-shaping limited region, \( P_Q > P_{th} \), a thermal noise limited region, \( P_{th} > P_Q \), \( P_{st} \), and a settling noise limited region, \( P_{st} > P_{th} \). 

![Fig. 8 Quantisation \( P_Q \), thermal \( P_{th} \) and settling error \( P_{st} \) noise power in 1-bit second-order \( \Sigma \Delta \) modulator](image)

By considering a sinusoidal input signal with a full dynamic amplitude \( \Delta \), the maximum achievable SNR can be written as

\[ \text{SNR} = \frac{\Delta^2}{P_Q + P_{th} + P_{st}} \]  

\[ \text{(18)} \]

By replacing \( P_Q \), \( P_{th} \) and \( P_{st} \) with eqns. 3, 12 and 16, the family of curves in Fig. 9 is obtained. They represent the resolution in bit as a function of the oversampling ratio for three different modulators using the integrator specified above. The three distinct working regions now correspond to:

(a) a region where the resolution increases by \( (L + 1/2) \) bit per octave of \( M \) (noise-shaping limited region)

(b) a region where the resolution simply increases by \( 1/2 \) bit per octave of \( M \) (thermal noise limited region)

(c) a region in which the resolution begins to decrease (settling noise limited region).

After the first transition oversampling ratio, defined by \( P_{st} = P_{th} \), all the modulators behave as common oversampled circuits and tend to a common asymptote. Subsequently, the settling error defines a barrier over which a decrease rather than an increase results for the overall modulator resolution. In the example considered the maximum speed of operation occurs for a sampling frequency which is in close agreement with the simulation results referred to in Reference 2. In the present design example, the maximum achievable resolution of \( \approx 17.5 \) bit can be obtained using either 2nd- or 3rd- or 4th-order modulators, all of them working in the oversampling to settling-limited transition.
In Fig. 10, the same curves as those given in Fig. 8 are shown with an increased value for the sampling capacitor C (12 pF, one octave above). The new value of C leads to a 3 dB reduction in the thermal noise power but, because of the increased load capacitance, a lower M for the oversampling to settling-limited transition results.

6 Conclusions

Intrinsic limitations to the maximum achievable SNR performance of switched capacitor ΣΔ modulators have been considered. An equation taking into account both the switches and the operational amplifier broad-band noise was derived. Setting error has been modelled by a white noise source placed inside the first integrator. Based on the assumptions made, three working regions were identified for the modulator:
(a) noise-shaping limited
(b) thermal noise limited and
(c) setting noise limited regions.

The limitations imposed by the thermal noise source and the settling error lead to the conclusion that at the state-of-the-art of SC modulators the 2nd-order noise-shaper constitutes a competitive structure to implement high resolution A/D converters.

7 Acknowledgments

The authors are grateful to the EEC-Joint Research Center and to Piano Nazionale Microelettronica for their financial support.

8 References