Analysis of $\Delta\Sigma$ Modulators with Zero Mean Stochastic Inputs

Ramin Khoini-Poorfard, *Student Member, IEEE,* and David A. Johns, *Member, IEEE*

***Abstract*—In this paper, a new framework for the analysis of $\Delta\Sigma$ modulators with stochastic inputs is proposed. The framework is based on assuming that the input to the one-bit quantizer is a Gaussian random process with zero mean and is thus able to interrelate the autocorrelation and cross-correlation of different signals of the modulator. Two main equations describing the behavior of two different $\Delta\Sigma$ topologies are derived. These two equations are generally nonlinear and can be of arbitrary order, hence approximations are used to study some interesting cases. First, the nonlinear equations are linearized and solved analytically for first-order modulator with white inputs and numerically for colored inputs both with and without dithering. Also, a numerical iterative approach is used for second and fourth order modulators with white and colored inputs. In these cases, the variance of the one-bit quantizer input is found as a function of modulator input power. Next, the variance of the one-bit quantizer input is calculated when large amplitude oscillations are present assuming a large amplitude limit cycle to have a sinusoidal autocorrelation. Finally, an attempt is made to estimate the modulator's critical input power level beyond which these large amplitude limit cycles start.**

I. Introduction

THOUGH at first look, $\Delta\Sigma$ modulators seem to be very straight forward structures, they are difficult to analyze rigorously. The main reason behind this difficulty is the existence of a one-bit quantizer in a feedback loop introducing strong nonlinearity. Since their introduction in the early 1960's [1], there have been many approaches to study the general behavior of these modulators with the most popular approach being a linear model method [2]. In this method, the quantizer is replaced by a summation node injecting white noise inside the circuit. With these assumptions, the whole modulator becomes a simple linear circuit and straight forward linear systems analysis can be applied. Although this method predicts the in-band noise surprisingly well, it doesn't describe other important aspects of these modulators, such as their input-dependent stability.

Another popular approach for analyzing $\Delta\Sigma$ modulators is the describing-function method which uses a more advanced model for the quantizer [4]. Ardalan and Paulos [6] achieved good results by decomposing the input to the quantizer into two parts, the signal component and the noise component, and then defining two separate gains for each part and finally adding a noise source to complete the quantizer model. Hein and Zakhor [3] also used the describing-function method to evaluate the amplitude and dc bias of large amplitude limit cycles in interpolative $\Delta\Sigma$ modulators with dc inputs. In [7], the same method is used to study the effect of quantizer hysteresis on the overall performance of a $\Delta\Sigma$ modulator. The main problem with the describing-function approach is that it is suitable only for dc or sinusoidal inputs.

The third approach is an exact analysis. In this approach a difference equation is found for the input to the quantizer. Then by solving this equation and assuming that the quantizer is never overloaded, one can derive the exact error spectrum as well as SNR with respect to the input. Though completely correct, this method is restricted to lower order modulators (up to second-order) and simple inputs (dc and sinusoids) [8], [9], [10].

Each of the above methods has its own advantages and disadvantages. The major disadvantage is their lack of ability to deal with higher order modulators as well as more general inputs. In this paper, we introduce a new approach based on an approximation which gets better as the order of $\Delta\Sigma$ structure is increased. It is also better suited to stochastic inputs. The motivation behind using this stochastic approach is the observation that the output bitstream of a $\Delta\Sigma$ modulator resembles a pseudorandom sequence, especially in higher-order modulators. This pseudorandom sequence is fed back and combined with the input. Hence, all the internal states of the modulator as well as the input to the one-bit quantizer also appear to be pseudorandom, particularly when the input is a stochastic signal. In addition to assuming stochastic inputs, we further require that the input to be zero-mean. Although this restricts the cases where this method could be applied to, there are several practical cases where the input is a zero-mean stochastic input such as a multicarrier modulated signal [12], speech signal, or the quantization noise of the first stage of a cascaded multistage architecture being fed to the next stage [11] when the first stage has a zero-mean input.

It has been noted in the literature that overloading the quantizer causes the quantization noise to increase which in turn may cause odd order harmonics to appear [2]. As well, in practical implementations, large internal states would result in harmonic distortion due to clipping effects. Thus, it is important to have estimates of the signal variance at various nodes throughout a modulator given input signal and dithering statistics. In this paper, a formulation is proposed which allows one to estimate the signal variance of the 1-bit quantizer input. Although the main results are all given for statistics of the 1-bit quantizer input, the results can be easily generalized to all internal state statistics. Also, it is worth mentioning that in

Manuscript received Jully 7, 1993; revised April 27, 1994. This paper was recommended by Associate Editor L. R. Carley.

The authors are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 1A4 Canada.

IEEE Log Number 9407649.

1057-7130/95$04.00 © 1995 IEEE

many architectures, the output of the last integrator before the quantizer has the largest variation.

To make the stochastic framework more tractable, we will make some simplifying assumptions. First, we assume that all ΔΣ internal states as well as the input signal are wide-sense stationary (WSS) random processes. Second, we assume that the quantizer input is a Gaussian random process with zero mean. We will study the validity of this assumption in more detail in Section II.

This paper is organized as follows. Section II contains a discussion on the validity of the Gaussian probability density function (pdf) assumption for the quantizer input. We show through simulation that dithering improves the Gaussianity assumption. Next, in Section III, two main equations describing the behavior of two different ΔΣ topologies are derived. These equations relate the quantizer input variance and the autocorrelation and cross-correlation for various signals both with and without dithering. These two equations are nonlinear and difficult to solve in the general case, hence approximations are made in the remainder of the paper to study some interesting cases. In Section IV, the nonlinear equations are linearized and solved analytically for a first-order modulator with a white input. The same linearized equations are then solved numerically for a first-order modulator with a colored input, a second-order modulator with white and colored inputs, and a fourth-order modulator again with white and colored inputs. In Section V, a large amplitude limit cycle is assumed to have a sinusoidal autocorrelation. With such an assumption, the two equations are solved to estimate quantizer input variance as a function of the modulator's input power. Finally, an attempt is made to estimate the modulator's critical power level beyond which large amplitude limit cycles are started. All of the theoretical results derived in Sections IV and V are compared to simulation results to verify their credibility. Finally, concluding remarks are given in Section VI.

II. Gaussianity Assumption

The main assumption throughout this paper, and used elsewhere in the literature, is that the input to the one-bit quantizer is a Gaussian random process with zero mean [5], [6]. As we are considering zero-mean inputs to the modulator, it is easily seen that the input to the one-bit quantizer will also be a zero mean random process[1]. In this section we will have a close and thorough look at the Gaussianity assumption.

A. Fundamentals

Consider a typical interpolative ΔΣ modulator as shown in Fig. 1. A nonrigorous argument on $x(n)$ appearing Gaussian would be as follows. Suppose that $\nu(n)$ has an arbitrary pdf. Then if the order of $H_i(z)$ is high enough and the samples of $\nu(n)$'s are relatively independent, the pdf of $x(n)$ will tend to be Gaussian similar to the application of the central limit theorem. However, one can argue that the samples of $\nu(n)$'s are not necessarily independent and identically distributed (i.i.d.) and hence the Gaussianity assumption may be stronger or weaker for different inputs. This conjecture was observed true as it was seen through simulations that increasing the power of $u(n)$ resulted in the pdf of $x(n)$ to be more Gaussian. Fig. 2(a) and 2(b) show two typical pdf's for a fourth-order interpolative modulator with a white Gaussian input having two different power levels for $u(n)$ and no dithering. As can be seen, the situation improves when a higher power for $u(n)$ is used. In fact, the best situation was observed to be when the input power is high enough to put the modulator at the verge of instability. The authors believe that by increasing the input power, the samples of $\nu(n)$ become more independent and hence the central limit theorem's condition is better satisfied.

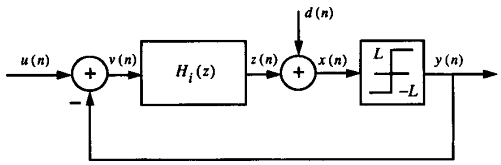


Fig. 1. Interpolative ΔΣ modulator. Note that $d(n)$ is a dither signal which may or may not be present.

Another interesting case to study is dithering. Consider the interpolative architecture with dithering $(d(n))$ added to the one-bit-quantizer input as shown in Fig. 1 [15]. Assuming that $d(n)$ is independent of $z(n)$, we conclude:

$$f_x(x) = f_z(x) * f_d(x) \tag{1}$$

where $*$ is the convolution operator and $f_x(x), f_z(x)$ and $f_d(x)$ are the pdf's of $x(n), z(n)$ and $d(n)$, respectively. One can predict that the above convolution will smooth any raggedness of $f_z(x)$ if $f_d(x)$ is wide enough, a fact verified by simulation results as shown in Fig. 2(c) and 2(d). Here, the applied dithering is uniformly distributed on $[-\frac{L}{2}, \frac{L}{2}]$ where the output levels of the modulator are $\pm L$. Note that even for the low power case, the pdf is nearly Gaussian while for the high power case, the pdf appears extremely close to Gaussian.

However, there is another issue to be resolved. A true Gaussian random process, $x(n)$, also has a characteristic that if one chooses any arbitrary set of random variables, $\{x(n), x(n+1), \ldots, x(n+k)\}$, they are jointly Gaussian for all n's and k's. Throughout our formulations, which appear in the following sections, the joint Gaussianity of $\{x(n), x(n+k)\}$ for all n's and k's is important. In other words, we need $x(n)$ and $x(n+k)$ to have a 2-D Gaussian pdf so that the input and output autocorrelations of a one-bit quantizer satisfy the *ArcSine Law* as follows [17].

$$R_{yy}(k) = \frac{2L^2}{\pi} \times \text{Arcsin}\left(\frac{R_{xx}(k)}{R_{xx}(0)}\right). \tag{2}$$

As well, the input-output cross-correlation will then satisfy the following,

$$R_{xy}(k) = \sqrt{\frac{2}{\pi R_{xx}(0)}} \times L \times R_{xx}(k) \tag{3}$$

[1] Note that by assuming all internal states to be zero-mean random processes, one implicitly assumes that all initial states are zero-mean random variables as well.

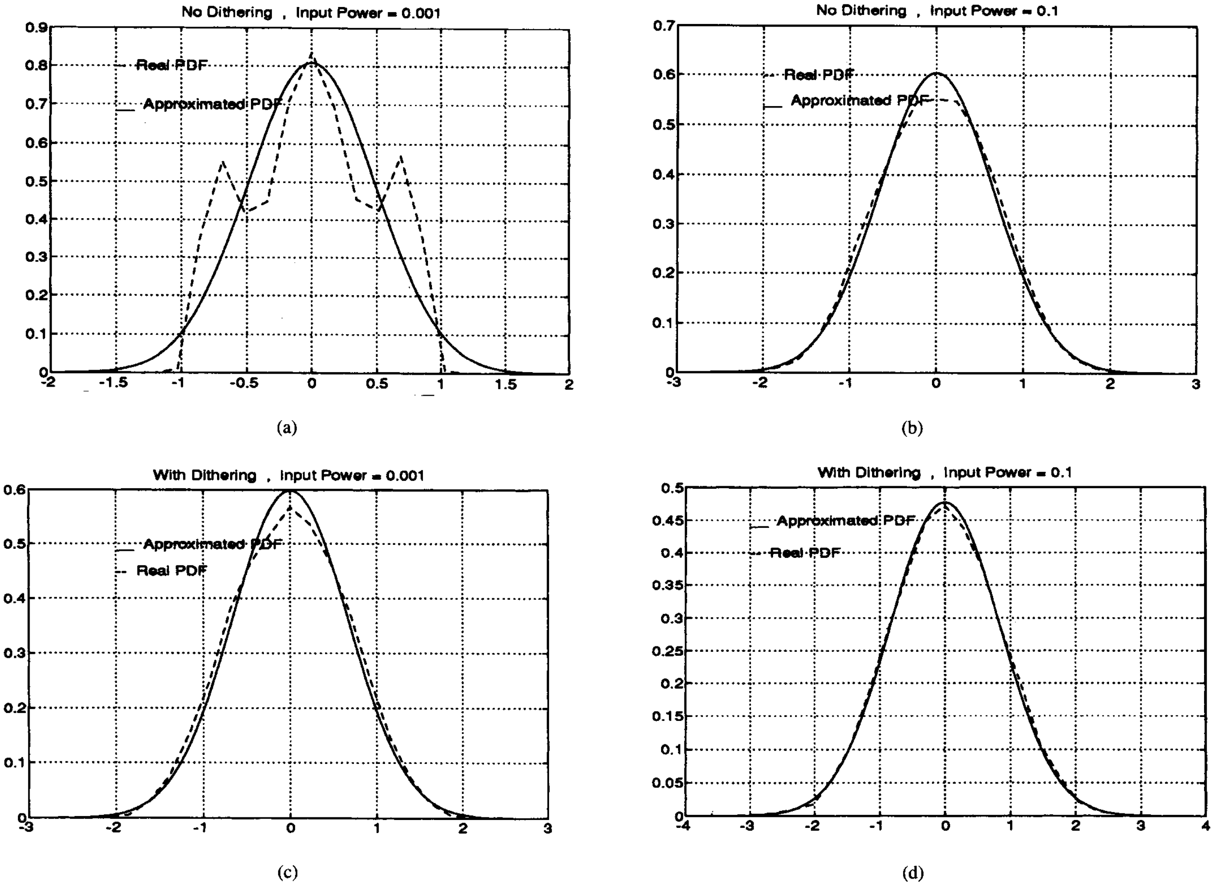


Fig. 2. Pdf of $x(n)$ for $L = 1$. (a) No dithering, input power = 0.001. (b) No dithering, input power = 0.1. (c) With dithering, input power = 0.001. (d) With dithering, input power = 0.1.

where $R_{xx}(k), R_{yy}(k)$ and $R_{xy}(k)$ are quantizer input autocorrelation, quantizer output autocorrelation and quantizer input-output cross-correlation, respectively.

To verify this joint Gaussianity assumption, we examined the 2-D pdf of $\{x(n), x(n+k)\}$ for the same fourth-order interpolative modulator having a white Gaussian input, both with and without dithering. The results for the modulator without dithering are shown in Fig. 3(a) and 3(b) and those with dithering are shown in Fig. 3(c) and 3(d). Although not shown here, it was found that the Gaussianity assumption improved for larger values of k.

By examining the results carefully one can see that the Gaussianity assumption is improved when dithering is applied and the input signal has high power. In summary, one can best rely on the Gaussianity assumption when the above conditions are met and one should be aware that this assumption is unreliable when those conditions are violated. To illustrate the errors, Fig. 4 shows the comparison between simulated and calculated results for the autocorrelation $R_{yy}(k)$ for the same fourth-order modulator. Here, we see good agreement for $k \neq \pm1, \pm2$, whereas for $k = \pm1, \pm2$ there is a discrepancy between the two results. These results show that for $k = \pm1, \pm2$ the 2-D Gaussianity assumption for $x(n)$ and $x(n+k)$ is poor whereas the assumption is reasonable for other values of k.

B. Gaussianity Assumption Method versus Linear Model Method

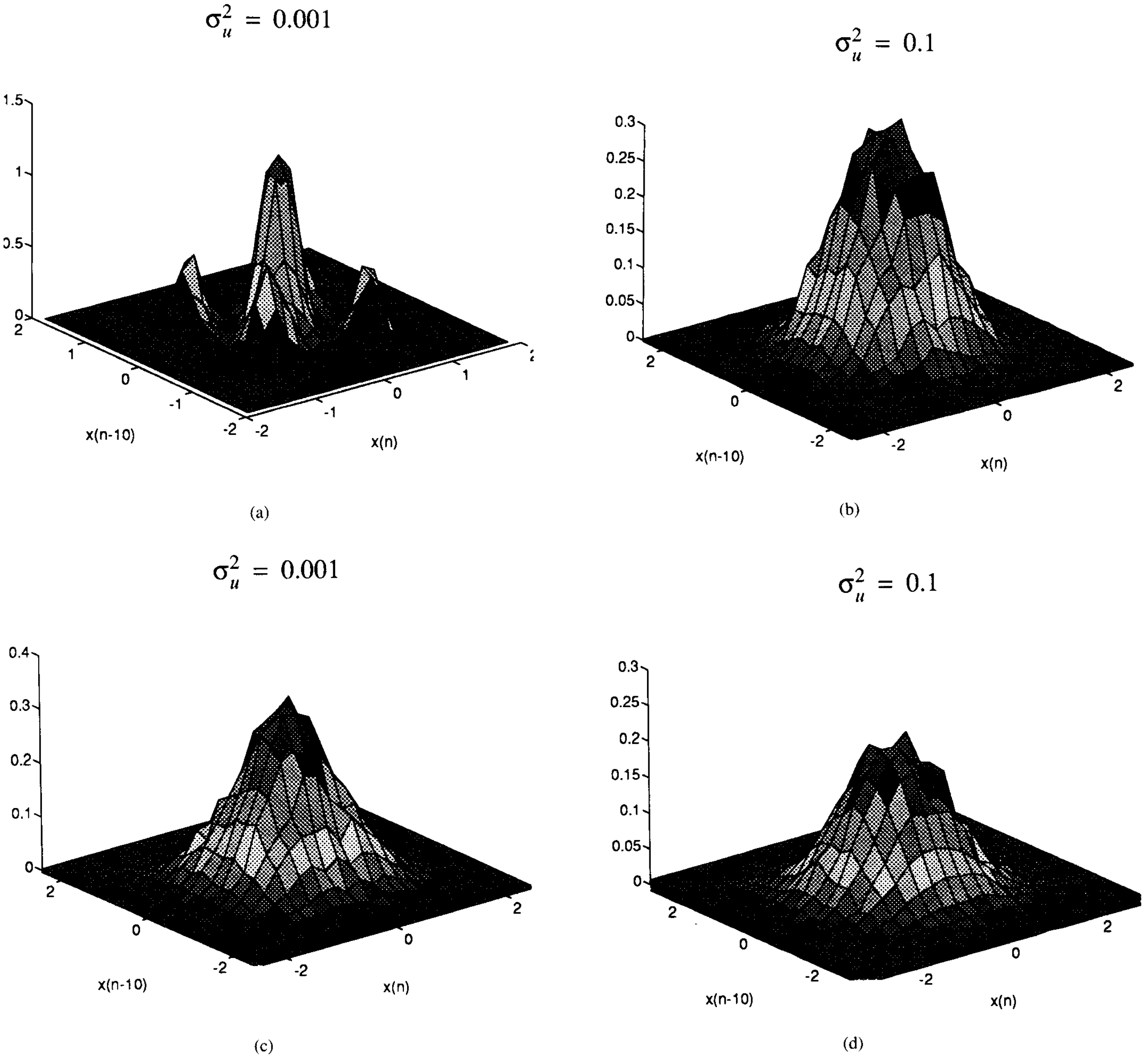
In this section a comparison is made between the Gaussianity assumption method and the linear model method [2]. Specifically, it is shown that under certain conditions, the Gaussianity assumption is approximately equivalent to the linear model.

Defining $e(n) = y(n) - x(n)$, one has the following autocorrelation and cross-correlation relationship

$$R_{ee}(k) = R_{yy}(k) + R_{xx}(k) - R_{xy}(k) - R_{yx}(k). \tag{4}$$

Inserting (2) and (3) into (4) results in

$$R_{ee}(k) = \frac{2L^2}{\pi} \times \text{Arcsin}\left(\frac{R_{xx}(k)}{R_{xx}(0)}\right) + R_{xx}(k) - 2\sqrt{\frac{2}{\pi R_{xx}(0)}} \times L \times R_{xx}(k) \tag{5}$$



(a) (b)

(c) (d)

Fig. 3. Joint probability density function for a fourth-order modulator with white Gaussian input. (a) No dithering, input power = 0.001. (b) No dithering, input power = 0.1. (c) With dithering, input power = 0.001. (d) With dithering, input power = 0.01.

for $k = 0$ we find:

$$\sigma_e^2 = \sigma_x^2\left(1 - 2L\sqrt{\frac{2}{\pi\sigma_x^2}}\right) + L^2 \qquad (6)$$

where $R_{ii}(0)$ is defined to be σ_i^2 for notational convenience. Based on (6) the following corollaries can be derived.

Corollary 1: The minimum value for σ_e^2 is $\left(1 - \frac{2}{\pi}\right)L^2$.

Proof: Differentiating (6) with respect to σ_x^2, we derive

$$\begin{aligned}\frac{d}{d\sigma_x^2}\sigma_e^2 &= 1 - L\sqrt{\frac{2}{\pi\sigma_x^2}} = 0 \Rightarrow \sigma_x^{*2}\\ &= \frac{2L^2}{\pi} \Rightarrow \sigma_{e_{\min}}^2 = \left(1 - \frac{2}{\pi}\right)L^2. \qquad (7)\end{aligned}$$

Note that based on the linear model method, $\sigma_e^2 = \frac{L^2}{3}$, whereas according to corollary 1 the minimum value for σ_e^2 is $(1-\frac{2}{\pi})L^2$ which is slightly greater than $\frac{L^2}{3}$.

Corollary 2: If $\sigma_x^2 = \frac{2L^2}{\pi}$, then x and e are uncorrelated.

Proof: The cross-correlation between e and x is as follows.

$$\begin{aligned}R_{ex}(k) &= E[(y_{k+n} - x_{k+n})x_n]\\ &= R_{yx}(k) - R_{xx}(k)\\ &= \left(\sqrt{\frac{2}{\pi R_{xx}(0)}} \times L - 1\right)R_{xx}(k). \qquad (8)\end{aligned}$$

Letting $R_{xx}(0) = \frac{2L^2}{\pi}$, we see that $R_{ex}(k) = 0$.

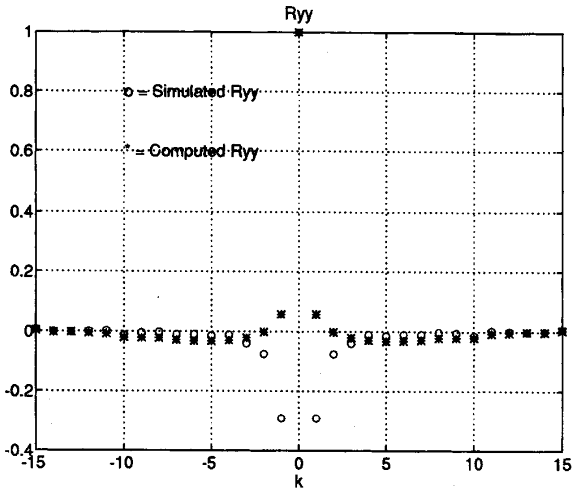


Fig. 4. An example of quantizer output autocorrelation.

Corollary 3: If $\sigma_x^2 = \frac{2L^2}{\pi}$, then the error signal's spectrum would be much wider than that of the quantizer input.

Proof: Replacing $\sigma_x^2 = \frac{2L^2}{\pi}$ into (5) we have:

$$R_{ee}(k) = -R_{xx}(k) + \frac{2L^2}{\pi}\text{Arcsin}\Big(\frac{\pi}{2L^2}R_{xx}(k)\Big). \tag{9}$$

As $|\frac{R_{xx}(k)}{R_{xx}(0)}| \leq 1$, we can use a Taylor expansion for approximating $\frac{2L^2}{\pi}\text{Arcsin}[\frac{R_{xx}(k)}{R_{xx}(0)}]$. Using a third-order approximation, we have

$$\text{Arcsin}(x) = x + \frac{x^3}{6} + O(x^5) \tag{10}$$

and therefore,

$$\begin{aligned}&\frac{2L^2}{\pi}\text{Arcsin}\Big(\frac{\pi}{2L^2}R_{xx}(k)\Big)\\ &\quad = R_{xx}(k) + \frac{\pi^2}{24L^4}R_{xx}^3(k) + O\bigg(\Big(\frac{R_{xx}(k)}{\pi/(2L^2)}\Big)^5\bigg).\end{aligned} \tag{11}$$

Replacing (11) into (9) we derive

$$R_{ee}(k) \approx \frac{\pi^2}{24L^4}R_{xx}^3(k). \tag{12}$$

Or in the frequency domain,

$$S_{ee}(e^{i\omega}) \approx \frac{\pi^2}{24L^4}S_{xx}(e^{i\omega}) * S_{xx}(e^{i\omega}) * S_{xx}(e^{i\omega}). \tag{13}$$

Hence, the error signal's spectrum is at least three times wider than that of the input signal to the quantizer.

Based on corollaries 1–3 we conclude that the linear model assumptions will be most accurate when $\sigma_x^2 = \frac{2L^2}{\pi}$. For this value of σ_x, the error signal is uncorrelated with respect to the input of the quantizer, its variance is $(1 - \frac{2}{\pi})L^2$ which is near $\frac{L^2}{3}$ and finally it is much whiter than the input signal. Also, it should be noted that when $\sigma_x^2 = \frac{2L^2}{\pi}$, the quantizer describing function for stochastic inputs, defined as DF $= \frac{R_{xy}}{R_{xx}}$ [4], is equal to 1, which is another assumption made in the linear model method.

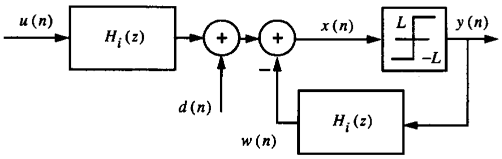


Fig. 5. Equivalent block diagram for an interpolative modulator.

III. General Formulation

In this section, we consider two general topologies for $\Delta\Sigma$ modulators covering all single quantizer architectures—the interpolative [13] and error-feedback architectures [14], [16]. Although these topologies are related,[2] they are considered separately here to simplify the application of results. Also, in both cases, general relations are found for both with and without dithering.

A. Interpolative Modulators

To simplify the derivations, consider an equivalent block diagram for the interpolative topology as depicted in Fig. 5 in which,

$$R_{ww}(k) = R_{yy}(k) * h_i(k) * h_i(-k) \tag{14}$$

$$R_{wx}(k) = R_{yx}(k) * h_i(k) \tag{15}$$

and for the input summation node one has:

$$u(n) * h_i(n) + d(n) = w(n) + x(n) \tag{16}$$

which in terms of autocorrelations and cross-correlations would result in (note that $R_{du}(k) = 0$)

$$\begin{aligned}&R_{uu}(k) * h_i(k) * h_i(-k) + R_{dd}(k)\\ &\quad = R_{ww}(k) + R_{xx}(k) + R_{wx}(k) + R_{xw}(k).\end{aligned} \tag{17}$$

Replacing (2), (3), (14) and (15) into (17) we derive the following formula.

$$\begin{aligned}&R_{uu}(k) * h_i(k) * h_i(-k) + R_{dd}(k)\\ &\quad = \frac{2L^2}{\pi}\text{Arcsin}\left[\frac{R_{xx}(k)}{R_{xx}(0)}\right] * h_i(k) * h_i(-k) + R_{xx}(k)\\ &\quad + \sqrt{\frac{2}{\pi R_{xx}(0)}} \times L \times R_{xx}(k) * [h_i(k) + h_i(-k)].\end{aligned} \tag{18}$$

The above formula in its most general form is a nonlinear difference equation, in which given $R_{uu}(k), R_{dd}(k)$ and $h_i(k), R_{xx}(k)$ should be found.

Solving (18), one can compute $R_{xx}(0)$ or equivalently the variance of $x(n)$ which we also define as σ_x^2 for notational simplicity. The benefit of estimating σ_x^2 is that one can find the probability of the quantizer being overloaded. For example, if the output levels of the quantizer is $\pm L$, then the quantizer-input threshold levels before overloading would be $\pm 2L$. Hence if $\sigma_x = \frac{2L}{2.567}$, 99% of times there is no overloading.

[2] The interpolative topology can be modified to cover the error-feedback topology by adding an extra filter on the input signal and vise versa.

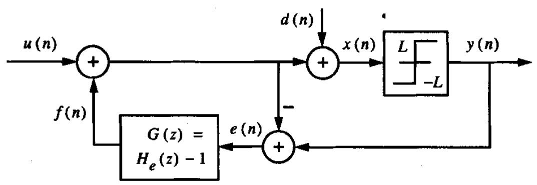


Fig. 6. Error-feedback topology.

As mentioned earlier, this method can also deal with state (i.e., integrator outputs) statistics. Using the definition of intermediate functions [18], $F_i(z) \equiv \frac{S_i(z)}{V(z)}$ in which $s_i(n)$ is the i^{th} state of $H_i(z)$ and $\nu(n)$ is the input to $H_i(z)$ and assuming $d(n) = 0$ (Fig. 1), we conclude that:

$$R_{\nu\nu}(k) * h_i(k) * h_i(-k) = R_{xx}(k) \tag{19}$$

$$R_{s_i s_i}(k) = R_{\nu\nu}(k) * f_i(k) * f_i(-k). \tag{20}$$

B. Error-Feedback Topology

Fig. 6 depicts the error-feedback topology. Based on a similar approach as that of the previous section, the general equation for the error-feedback topology is found to be given by:

$$\begin{aligned} &R_{uu}(k) + R_{dd}(k) * h_e(k) * h_e(-k) \\ &\quad = R_{xx}(k) + \Bigg[R_{xx}(k)\left(1 - 2L\sqrt{\frac{2}{\pi\sigma_x^2}}\right) \\ &\quad + \frac{2L^2}{\pi}\text{Arcsin}\left(\frac{R_{xx}(k)}{R_{xx}(0)}\right)\Bigg] * g(k) * g(-k) \\ &\quad - \left(\sqrt{\frac{2}{\pi\sigma_x^2}} \times L - 1\right) R_{xx}(k) * [g(k) + g(-k)] \end{aligned} \tag{21}$$

where $G(z) = H_e(z) - 1$.

IV. Quantizer Input Variance

One technique used in solving general nonlinear equations is the linear approximation method. In this section, we apply this method to (21) and (18) and compare results with simulations.

The main idea is the use of a first-order Taylor expansion for Arcsin(x) noting that the absolute value of the argument in our case is always less than 1. Hence,

$$\frac{2L^2}{\pi}\text{Arcsin}\left(\frac{R_{xx}(k)}{R_{xx}(0)}\right) \approx \frac{2L^2}{\pi} \times \frac{R_{xx}(k)}{R_{xx}(0)}. \tag{22}$$

Inserting (22) into (21) we have

$$\begin{aligned} &R_{uu}(k) + R_{dd}(k) * h_e(k) * h_e(-k) \\ &\quad = R_{xx}(k) + \Bigg[R_{xx}(k)\left(1 - 2L\sqrt{\frac{2}{\pi\sigma_x^2}}\right) \\ &\quad + \frac{2L^2}{\pi} \times \left(\frac{R_{xx}(k)}{R_{xx}(0)}\right)\Bigg] * g(k) * g(-k) \\ &\quad - \left(\sqrt{\frac{2}{\pi\sigma_x^2}} \times L - 1\right) R_{xx}(k) * [g(k) + g(-k)]. \end{aligned} \tag{23}$$

Linearizing the above equation enables the use of linear transformation techniques, such as the Fourier-transform, to solve the equation. Taking the Fourier-transform of (23), we have

$$\begin{aligned} &S_{uu}(e^{i\omega}) + S_{dd}(e^{i\omega})H_e(e^{i\omega})H_e(e^{-i\omega}) \\ &\quad = S_{xx}(e^{i\omega}) + S_{xx}(e^{i\omega})\left(1 - 2L\sqrt{\frac{2}{\pi\sigma_x^2}} + \frac{2L^2}{\pi\sigma_x^2}\right) \\ &\quad \times G(e^{i\omega})G(e^{-i\omega}) \\ &\quad - \left(\sqrt{\frac{2}{\pi\sigma_x^2}}L - 1\right) S_{xx}(e^{i\omega})[G(e^{i\omega}) + G(e^{-i\omega})]. \end{aligned} \tag{24}$$

Rearranging (24) and factoring it with respect to $S_{xx}(e^{i\omega})$, we derive (25) shown at the bottom of the page, which can be simplified to (26) also shown at the bottom of the page. Using the inverse Fourier-transform, one can compute $R_{xx}(k)$. Specifically,

$$\sigma_x^2 = R_{xx}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{xx}(e^{i\omega})d\omega. \tag{27}$$

The same kind of calculations can be carried out for the interpolative topology. For the sake of brevity, only the final result is presented here as follows.

$$S_{xx}(e^{i\omega}) = \frac{S_{uu}(e^{i\omega})H_i(e^{i\omega})H_i(e^{-i\omega}) + S_{dd}(e^{i\omega})}{\left(1 + L\sqrt{\frac{2}{\pi\sigma_x^2}}H_i(e^{i\omega})\right)\left(1 + L\sqrt{\frac{2}{\pi\sigma_x^2}}H_i(e^{-i\omega})\right)}. \tag{28}$$

Example 1: Consider (26) for a first-order ΔΣ modulator with $G(z) = -z^{-1}$, a white Gaussian input ($S_{uu}(z) = \sigma_u^2$), a white dithering signal ($S_{dd}(z) = \sigma_d^2$) and output levels of ±1.

$$S_{xx}(e^{i\omega}) = \frac{S_{uu}(e^{i\omega}) + S_{dd}(e^{i\omega})H_e(e^{i\omega})H_e(e^{-i\omega})}{1 + \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}} \times L\right)^2 G(e^{i\omega})G(e^{-i\omega}) - \left(\sqrt{\frac{2}{\pi\sigma_x^2}} \times L - 1\right)(G(e^{i\omega}) + G(e^{-i\omega}))} \tag{25}$$

$$S_{xx}(e^{i\omega}) = \frac{S_{uu}(e^{i\omega}) + S_{dd}(e^{i\omega})H_e(e^{i\omega})H_e(e^{-i\omega})}{\left(1 + \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}} \times L\right)G(e^{i\omega})\right)\left(1 + \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}} \times L\right)G(e^{-i\omega})\right)} \tag{26}$$

Hence $H_e(z) = 1 - z^{-1}$ and (26) simplifies to

$$S_{xx}(e^{i\omega}) = \frac{\sigma_u^2 + \sigma_d^2(1 - e^{-j\omega})(1 - e^{j\omega})}{\left(1 - \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}}\right)e^{-i\omega}\right)\left(1 - \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}}\right)e^{i\omega}\right)}$$

$$= \frac{\sigma_u^2 + 2\sigma_d^2(1 - \cos(\omega))}{1 + \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}}\right)^2 + 2\left(\sqrt{\frac{2}{\pi\sigma_x^2}} - 1\right)\cos(\omega)}. \quad (29)$$

Recalling $\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{xx}(e^{i\omega})d\omega = \sigma_x^2$, and using the following relationships

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{d\omega}{A + B\cos(\omega)} = \frac{1}{\sqrt{A^2 - B^2}} \quad (30)$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\cos(\omega)d\omega}{A + B\cos(\omega)} = \frac{1}{B} - \frac{A}{B\sqrt{A^2 - B^2}} \quad (31)$$

one can derive

$$\sigma_x^2 = \frac{\sigma_u^2 + 2\sigma_d^2}{2\sqrt{\frac{2}{\pi\sigma_x^2}} - \frac{2}{\pi\sigma_x^2}} - \frac{2\sigma_d^2}{2\left(\sqrt{\frac{2}{\pi\sigma_x^2}} - 1\right)} + \frac{2\sigma_d^2\left(1 + \left(1 - \sqrt{\frac{2}{\pi\sigma_x^2}}\right)^2\right)}{\left(\sqrt{\frac{2}{\pi\sigma_x^2}} - 1\right)\left(2\sqrt{\frac{2}{\pi\sigma_x^2}} - \frac{2}{\pi\sigma_x^2}\right)}. \quad (32)$$

Defining $y \equiv \sqrt{\frac{2}{\pi\sigma_x^2}}$ into (32) and after some algebraic manipulations, (32) can be simplified to

$$2\sigma_d^2 y^2 + \left(\sigma_u^2 + \frac{2}{\pi}\right)y - \frac{4}{\pi} = 0. \quad (33)$$

Solving (33) for its positive solution results in

$$\sigma_x = \left(\frac{\pi}{32}\left(\sigma_u^2 + \frac{2}{\pi}\right)^2 + \sigma_d^2\right)^{1/2} + \sqrt{\frac{\pi}{32}}\left(\sigma_u^2 + \frac{2}{\pi}\right). \quad (34)$$

For the case where there is no dithering ($\sigma_d = 0$) (34) reduces to

$$\sigma_x = \sqrt{\frac{\pi}{8}}\left(\sigma_u^2 + \frac{2}{\pi}\right). \quad (35)$$

Making use of these two formulas, a comparison of theory and simulation for the root-variance of the quantizer input is shown in Fig. 7. Note that the results are in reasonable agreement with each other.

Example 2: To check the accuracy of the theoretical predictions for nonwhite input signals, example 1 was repeated with dithering and the input signal was low-pass filtered before being applied to the modulator. Specifically, the low-pass filter had the transfer-function $\frac{1-p}{1-pz^{-1}}$ which has a dc gain of unity and a pole at p. Such a low-pass filter results in the input signal spectrum being given by

$$S_{uu}(e^{i\omega}) = \sigma_u^2 \frac{(1-p)^2}{1 + p^2 - 2p\cos(\omega)}. \quad (36)$$

Unfortunately, deriving an analytical result for this case is extremely difficult and thus a numerical approach was used for computing the root-variance of the quantizer input, σ_x. The results are shown in Fig. 8 where it should be mentioned that here the input power is $\sigma_u^2(\frac{1-p}{1+p})$. Once again note that the simulation and theoretical results are in reasonable agreement.

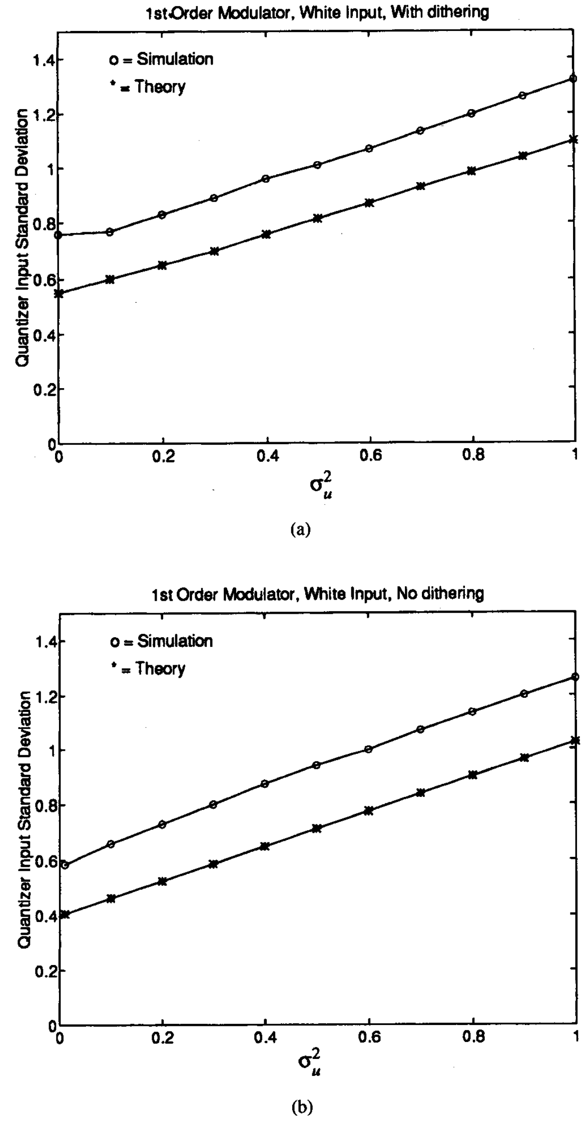


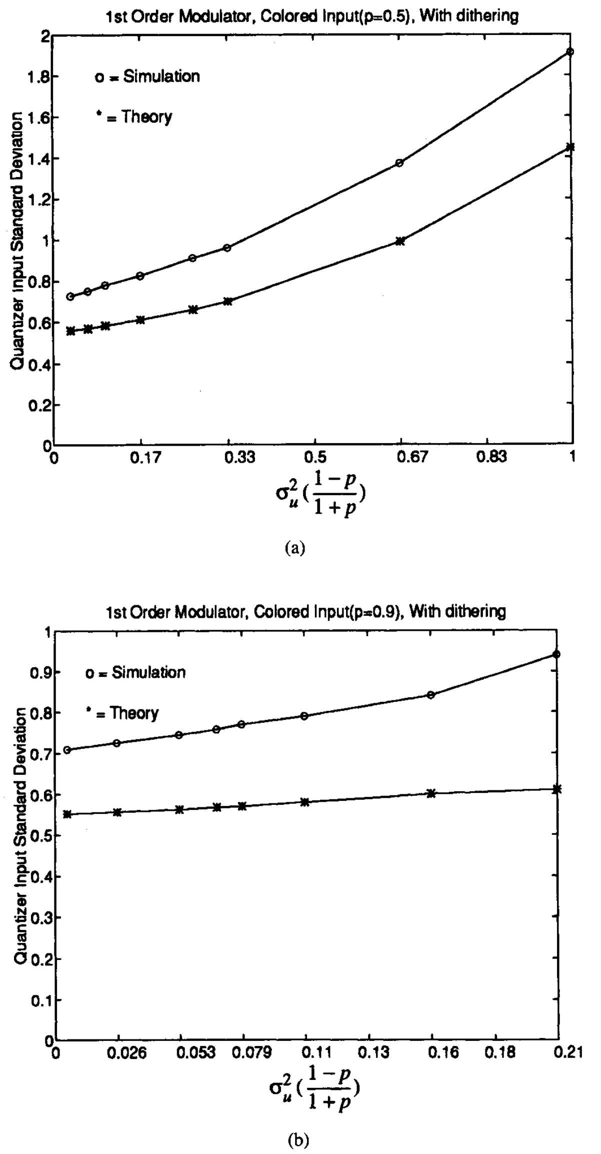
Fig. 7. Quantizer input root-variance for a first-order modulator with a white input having a variance of σ_u^2. (a) With uniformly distributed dithering between $[-0.5, 0.5]$. (b) Without dithering.

Example 3: A second-order modulator with $G(z) = -2z^{-1} + z^{-2}$ and quantizer output levels of ±1 is considered in this example. Once again, a numerical iterative approach was used to compute σ_x due to the difficulty in deriving an analytical formula. The results for white as well as colored inputs are shown in Fig. 9.

Example 4: In this example, a fourth-order interpolative modulator also used in [3], [11] is simulated where $H_i(z)$ is given by

$$H_i(z) = \frac{\sum_{n=0}^{N} A_n(z-1)^{N-n}}{(z-1)^N - \sum_{n=1}^{N} B_n(z-1)^{N-n}} \quad (37)$$

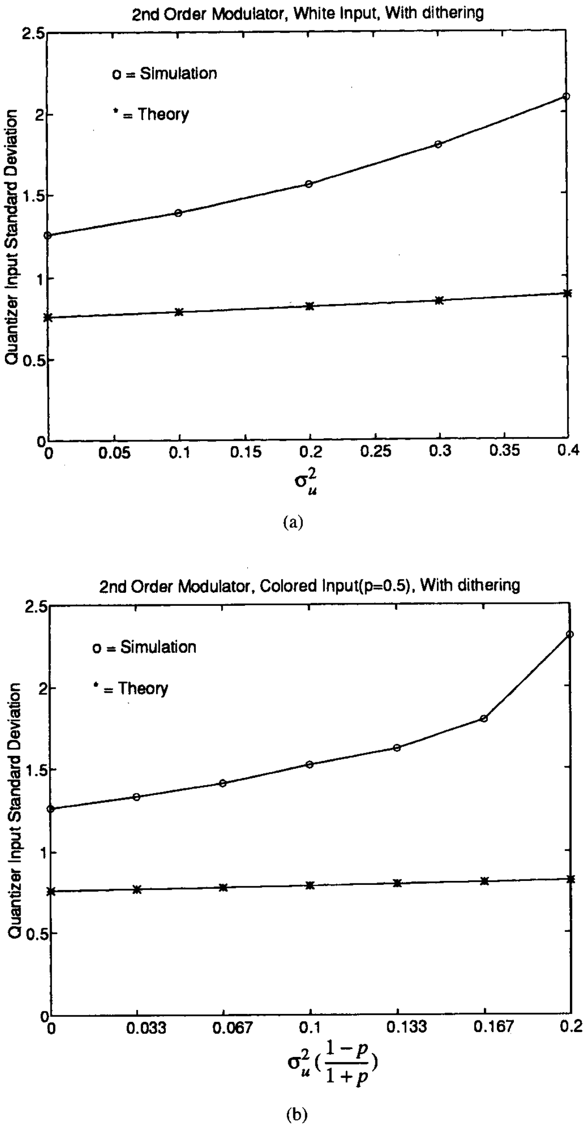
in which $N = 4$, $(A_0, A_1, A_2, A_3, A_4) = (0.8653, 1.1920, 0.3906, 0.06926, 0.005395)$ and $(B_1, B_2, B_3, B_4) =$



(a)

(b)

Fig. 8. Quantizer input root-variance for a first-order modulator with colored input versus input power (a) $p = 0.5$, (b) $p = 0.9$.



(a)

(b)

Fig. 9. Quantizer input root-variance for a second-order modulator versus input power. (a) White input. (b) Colored input.

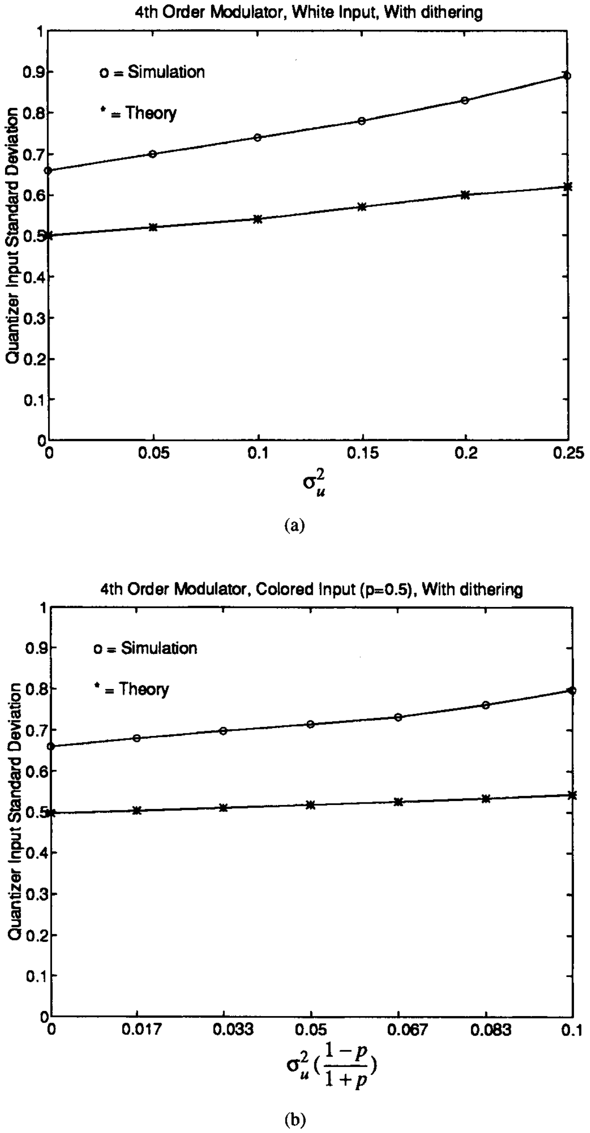
$(-8.347 \times 10^{-2}, -6.010 \times 10^{-3}, -1.752 \times 10^{-4}, -3.053 \times 10^{-6})$. Note that the poles are inside the unit circle and the poles radii are 0.98. The simulated and theoretical results are shown in Fig. 10 for both white and colored inputs.

V. Large Amplitude Limit Cycles During Instability

In the ΔΣ modulator literature, the term *limit cycle* is used to describe two distinct phenomena. One behavior is the audible and annoying tones in the band of interest which persist for a short or a long time-interval [15], [19]. The other behavior is the large amplitude oscillation at the quantizer input which occurs when the modulator is unstable [3]. In this section, the latter case is considered where an effort is taken to estimate the quantizer input variance during instability. The fascinating point about this limit cycle is that up to a certain input power level, no limit cycle exists and the system is stable. However, by increasing the input power above this critical point, suddenly huge oscillations appear and the structure is no longer considered stable.

To theoretically predict the behavior of these limit cycles, suppose that large amplitude limit cycles are present, hence the quantizer input, $x(n)$, has a periodic behavior with frequency ω_0. Also, assume the autocorrelation of $x(n)$ can be approximated by

$$R_{xx}(k) = R_{xx}(0)\cos(\omega_0 k) \qquad (38)$$



(a)

(b)

Fig. 10. Quantizer input root-variance for a fourth-order modulator versus input power. (a) White input. (b) Colored input.

where ω_0 is the resonance frequency of the modulator or mathematically, $\angle H_i(e^{i\omega})\big|_{\omega=\omega_0} = 180°$. Inserting (38) into (18) we derive

$$\begin{aligned} R_{uu}(k) * h_i(k) * h_i(-k) + R_{dd}(k) &= \frac{2L^2}{\pi}\text{Arcsin}(\cos(\omega_0 k)) * h_i(k) * h_i(-k) \\ &+ \sqrt{\frac{2}{\pi R_{xx}(0)}} \times L \times R_{xx}(0)\cos(\omega_0 k) \\ &* [h_i(k) + h_i(-k)] + R_{xx}(0)\cos(w_0 k). \end{aligned} \tag{39}$$

The term $\frac{2L^2}{\pi}\text{Arcsin}(\cos(\omega_0 k))$ is a periodic function, hence its Fourier series representation can be computed as follows,

$$\frac{2L^2}{\pi}\text{Arcsin}(\cos(\omega_0 k)) = \frac{8L^2}{\pi^2}\sum_{n=0}^{\frac{N-1}{2}} \frac{\cos[(2n+1)\omega_0 k]}{(2n+1)^2} \tag{40}$$

where N is the largest integer less than $\frac{2\pi}{\omega_0}$ and is proportional to the oversampling ratio.[3] Therefore,

$$\frac{2L^2}{\pi}\text{Arcsin}(\cos(\omega_0 k)) * h_i(k) * h_i(-k) = \frac{8L^2}{\pi^2}\sum_{n=0}^{\frac{N-1}{2}} \frac{\cos[(2n+1)\omega_0 k]}{(2n+1)^2}\left|H_i(e^{i(2n+1)\omega_0})\right|^2. \tag{41}$$

Also,

$$R_{uu}(k) * h_i(k) * h_i(-k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 e^{-j\omega k}\,d\omega \tag{42}$$

$$\begin{aligned} \cos(\omega_0 k) * [h_i(k) + h_i(-k)] &= \text{Re}\{\left[H_i(e^{i\omega_0}) + H_i(e^{-i\omega_0})\right]e^{i\omega_0 k}\} \\ &= 2\left|H_i(e^{i\omega_0})\right|\cos\left[\angle H_i(e^{i\omega_0})\right]\cos(\omega_0 k) \end{aligned} \tag{43}$$

and noting that $\angle H_i(e^{i\omega_0}) = 180°$ we conclude:

$$\cos(\omega_0 k) * [h_i(k) + h_i(-k)] = -2\left|H_i(e^{i\omega_0})\right|\cos(\omega_0 k). \tag{44}$$

Putting (41)–(44) into (39) we have

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 e^{-i\omega k}\,d\omega + R_{dd}(k) &= R_{xx}(0)\cos(\omega_0 k) \\ &+ \frac{8L^2}{\pi^2}\sum_{n=0}^{\frac{N-1}{2}} \frac{\cos[(2n+1)\omega_0 k]}{(2n+1)^2}\left|H_i(e^{i(2n+1)\omega_0})\right|^2 \\ &- 2L\sqrt{\frac{2}{\pi R_{xx}(0)}}R_{xx}(0)\left|H_i(e^{i\omega_0})\right|\cos(\omega_0 k). \end{aligned} \tag{45}$$

Finally recalling that $\sigma_x^2 = R_{xx}(0)$ and putting $k = 0$ into (45), we derive the following relationship

$$\begin{aligned} \sigma_x^2 - 2L\sqrt{\frac{2}{\pi}}\left|H_i(e^{i\omega_0})\right|\sigma_x &+ \left(\frac{8L^2}{\pi^2}\sum_{n=0}^{\frac{N-1}{2}} \frac{\left|H_i(e^{i(2n+1)\omega_0})\right|^2}{(2n+1)^2}\right. \\ &\left. - \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 d\omega - \sigma_d^2\right) = 0 \end{aligned} \tag{46}$$

where σ_d^2 is the variance of the dithering signal.

It is worth mentioning that the pdf of $x(n)$, when limit cycles are present, is no longer completely Gaussian. In

[3] Note that we are using the Fourier series representation of $\frac{2L^2}{\pi}$Arcsin $(\cos(\omega_0 t))$ as an approximation for DFT. The typical high oversampling ratios in $\Delta\Sigma$ modulators justify this assumption.

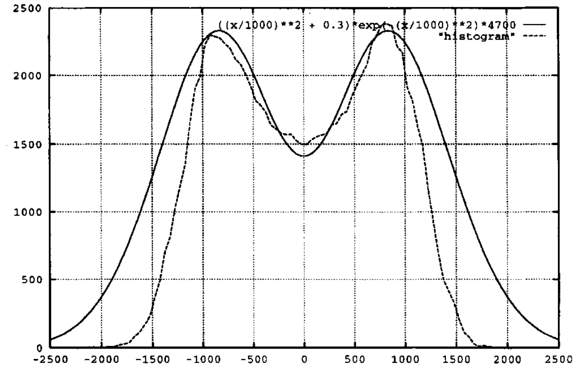


Fig. 11. Typical pdf of quantizer input when limit cycles are present.

fact, through simulations we found that the pdf is better approximated by

$$f_x(x) = K\left[\left(\frac{x}{a}\right)^2 + b\right]e^{-(\frac{x}{a})^2}. \tag{47}$$

A typical pdf for this case is depicted in Fig. 11. Therefore the Gaussianity assumption is not completely valid here but is reasonable enough for determining estimates of the variance of the quantizer input.

Example 5: The same fourth-order ΔΣ modulator as in example 4 is used for this example. Here, the input is assumed to be a zero-mean Gaussian random process with the following spectrum

$$S_{uu}(e^{i\omega}) = \frac{\sigma_u^2|H_i(e^{i\omega})|^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi}|H_i(e^{i\omega})|^2 d\omega}. \tag{48}$$

In other words, the input signal has the same spectral shape as the feed-forward filter. Typical quantizer waveforms for $\sigma_u = 0.237$ (limit cycle present) and $\sigma_u = 0.199$ (no limit cycles) are shown in Fig. 12(a) and (b), respectively. Using (46), σ_x^2 was found and the results are compared versus simulation results in Fig. 13. Once again, the accuracy between the results is reasonable.

Finally, an attempt is made to estimate the critical input power level ($\sigma_{u,\text{critical}}$) beyond which instability occurs. As long as (38) is a valid solution to (18), there exists a limit cycle throughout the system. Whenever such a solution fails to satisfy (18), there can't be any sustained Cosine solution. To have a valid solution to (18), the quadratic equation (46) should have a real valued solution which in turn implies the following inequality.

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 d\omega$$
$$\geq \left(\frac{8L^2}{\pi^2}\sum_{n=0}^{\frac{N-1}{2}}\frac{\left|H_i(e^{i(2n+1)\omega_0})\right|^2}{(2n+1)^2} - \frac{2L^2}{\pi}\left|H_i(e^{i\omega_0})\right|^2 - \sigma_d^2\right). \tag{49}$$

Assuming $H_i(z)$ is heavily attenuating the higher harmonics of the resonant frequency, (49) can be approximated as follows.

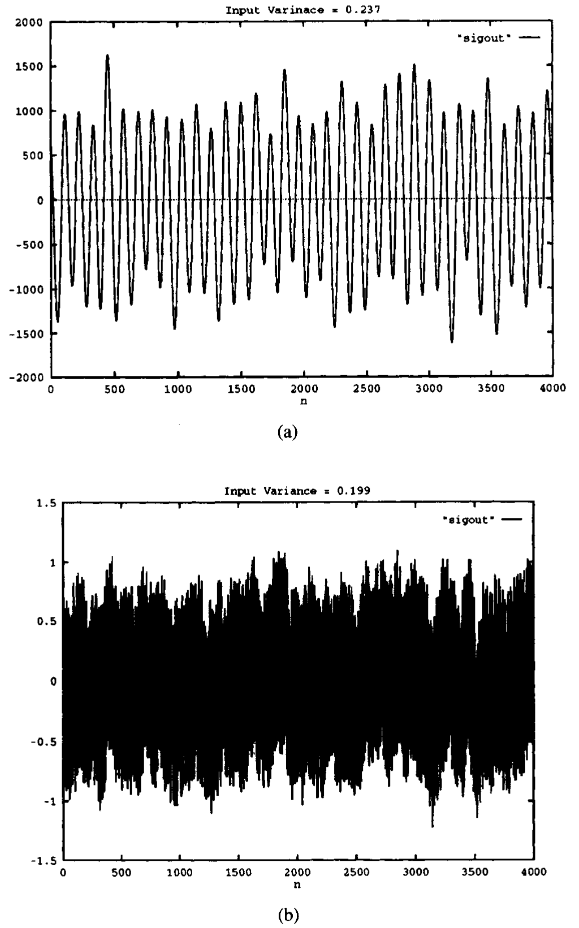


Fig. 12. Example of the quantizer input. (a) Limit cycle present. (b) No limit cycles (Note the difference in the vertical axis scales).

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 d\omega$$
$$\geq \left(\frac{8L^2}{\pi^2} - \frac{2L^2}{\pi}\right)\left|H_i(e^{i\omega_0})\right|^2 - \sigma_d^2. \tag{50}$$

Naming the normalized input spectrum as $S_{uu}^{\text{norm}}(e^{i\omega})$, we have

$$S_{uu}(e^{i\omega}) = \sigma_u^2 S_{uu}^{\text{norm}}(e^{i\omega}) \tag{51}$$

where σ_u^2 is the input power. Hence,

$$\sigma_{u,\text{critical}}^2 = \frac{\left(\frac{8L^2}{\pi^2} - \frac{2L^2}{\pi}\right)\left|H_i(e^{i\omega_0})\right|^2 - \sigma_d^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi} S_{uu}^{\text{norm}}(e^{i\omega})\left|H_i(e^{i\omega})\right|^2 d\omega}. \tag{52}$$

Based on (52) we see that among the equi-power input signals, those with more power inside the passband of the loop filter make the system less stable. In fact, for these signals, a larger portion of the signal power lies inside the filter passband, hence the denominator of (52) is larger, i.e., $\sigma_{u,\text{critical}}^2$ is less. In other words, a system exhibits limit cycles at lower levels of input signal power when its power is concentrated at low frequencies. Conversely, if the input signal has most of its power outside the passband of the loop filter, very large levels of input power can be applied. An interesting example of this case is the cascaded multistage modulator architecture,

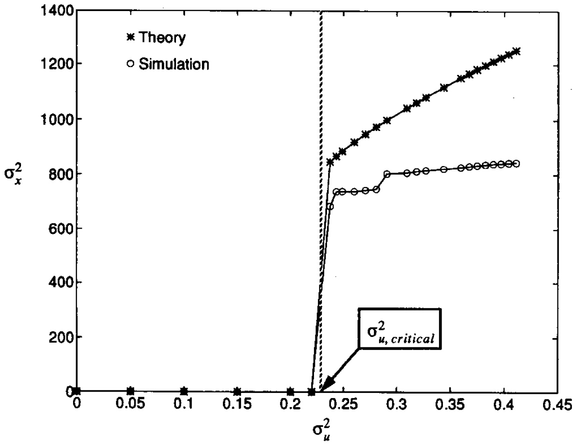


Fig. 13. Quantizer input variance versus modulator input variance. Note that (46) is used to find theoretical values for $\sigma_u^2 > \sigma^2_{u,\text{critical}}$ where a valid solution exists. For $\sigma_u^2 < \sigma^2_{u,\text{critical}}$ no large amplitude limit cycles exist and hence (28) is used in this region.

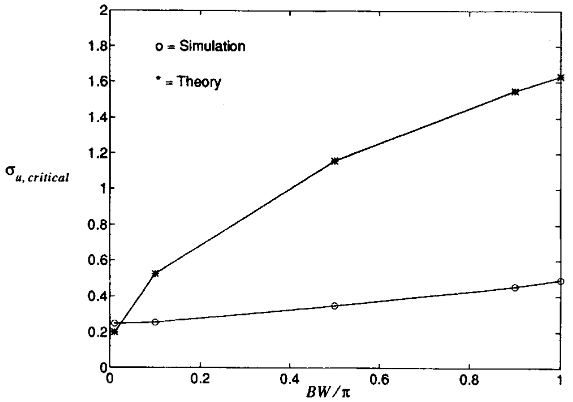


Fig. 14. $\sigma_{u,\text{critical}}$ for input signals having different bandwidths.

where the input to the second stage is the quantization error of the first stage [11]. Though the input to the second stage has much more power than the input to the first stage, the second stage is not necessarily prone to instability since both stages have the same in-band power from their inputs. To verify the credibility of (52), simulations and predicted values are compared in Fig. 14 for inputs with different bandwidths. As one can see, both simulation and theory indicate that by widening the input signal's spectra, $\sigma_{u,\text{critical}}$ is getting larger. The theoretical results predict simulations well for low bandwidth inputs, however, the matching between simulation and theory deteriorates at higher bandwidths.

VI. Conclusion

A qualitative approach based on a stochastic point of view was presented to study the behavior of $\Delta\Sigma$ modulators. Assuming that all the signals throughout the modulator are random processes with zero means, two main equations were derived for two different topologies relating the autocorrelation of the signal at the quantizer input to the autocorrelation of the input signal, a dithering signal, and the loop filter parameters. These equations are general and apply to arbitrary $\Delta\Sigma$ modulators having zero-mean stochastic inputs with an arbitrary spectrum.

An analytical solution was found for the case of the first-order $\Delta\Sigma$ modulator having a white input both with and without dithering and the results are in good agreement with simulation. Also the same modulator was considered with a colored input where the equations were solved numerically and again the results are in agreement with simulation results. This same numerical approach was also applied to a second and a fourth-order modulator to verify applicability of the proposed method to higher-order modulators. In all these cases the standard deviation of the one-bit quantizer input is found as a function of modulator input power. Next, based on the same method, numerical results were derived for the quantizer input variance when large amplitude limit cycles are present assuming a large amplitude limit cycle to have a sinusoidal autocorrelation. Numerical as well as simulation results were presented. Finally, an attempt was made to estimate the critical power level for the input signal's power beyond which the modulator starts having limit cycles.

References

[1] H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback," *Proc. IEEE*, vol. 51, pp. 1524–1533, Nov. 1963.

[2] J. Candy and G. Temes, "Oversampling methods for A/D and D/A conversion," *Oversampling Delta-Sigma Data Converters*, J. C. Candy and G. C. Temes, Eds. New York: IEEE Press, 1991.

[3] S. Hein and A. Zakhor, "On the stability of sigma-delta modulators," *IEEE Trans. Signal Process.*, vol. 41, no. 7, pp. 2322–2348, July 1993.

[4] A. Gelb and W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*. New York: McGraw-Hill, 1968.

[5] H. W. Smith, *Approximate Analysis of Randomly Excited Nonlinear Controls*. Cambridge, MA: M.I.T. Press, 1966.

[6] S. H. Ardalan and J. J. Paulos, "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 287–301, Mar. 1987.

[7] R. Khoini-Poorfard and D. A. Johns, "On the effect of comparator hysteresis in interpolative $\Delta\Sigma$ modulators," in *Proc. 1993 IEEE Int. Symp. Circuits Syst.*, May 1993, pp. 1148–1151.

[8] R. M. Gray, "Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input," *IEEE Trans. Commun.*, vol. 37, no. 6, pp. 588–599, June 1989.

[9] R. M. Gray, W. Chou, and P. W. Wong, "Quantizations noise in single-loop sigma-delta modulation with sinusoidal inputs," *IEEE Trans. Inform. Theory*, vol. 35, no. 9, pp. 956–968, Sept. 1989.

[10] N. He, F. Kuhlmann, and A. Buzo, "Double-loop sigma-delta modulation with dc inputs," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 487–495, Apr. 1990.

[11] Y. Matsuya *et al.*, "A 16-bit oversampling A-to-D conversion technology using triple-integration noise shaping," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 921–929, Dec. 1987.

[12] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, pp. 5–14, May 1990.

[13] K. C.-H. Chao, S. Nadeem, W. L. Lee, and C. G. Sodini, "A higher order topology for interpolative modulators for oversampling A/D converters," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 309–318, Mar. 1990.

[14] D. Anastassiou, "Error diffusion coding for A/D conversion," *IEEE Trans. Circuits Syst.*, vol. 34, pp. 593–603, June 1987.

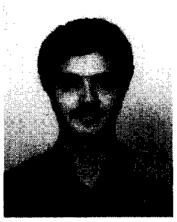
[15] S. R. Norsworthy, "Effective dithering of sigma-delta modulators," *IEEE Proc. ISCAS '92*, vol. 3, pp. 1304–1307, May 1992.

[16] R. Schreier, "Noise-shaped coding," Ph.D. dissertation, University of Toronto, 1991.

[17] A. Papoulis, *Probability Random Variable and Stochastic Process*. New York: McGraw-Hill, 1991, third ed.

[18] W. M. Snelgrove and A. S. Sedra, "Synthesis and analysis of state-space active filters using intermediate transfer functions," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 593–603, June 1987.

[19] M. W. Hauser, "Principles of oversampling A/D conversion," *J. Audio Eng. Soc.*, pp. 2–26, Jan./Feb. 91.



Ramin Khoini-Poorfard (S'90) was born in Tehran, Iran, on July 31, 1965. He received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Tehran, Iran, in 1987 and 1989, respectively.

Since September 1990, he has been a Research and Teaching Assistant at the University of Toronto where he is currently pursuing the Ph.D. degree. He was a recipient of the University of Toronto Open Doctoral Fellowship in 1991 and the Ontario Graduate Scholarship from 1992 to 1994. His research interests include analog and digital signal processing, delta-sigma modulation, and adaptive systems.

David A. Johns (M'88) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto, Canada, in 1980, 1983, and 1989, respectively.

From 1980 to 1981 he worked as an applications engineer in the Semiconductor Division of Mitel Corp., Ottawa, Canada. From 1983 to 1985 he was an analog IC designer at Pacific Microcircuits Ltd., Vancouver, Canada. Upon completion of his doctoral work, he joined the University of Toronto where he is currently an Associate Professor. His research interests are in the areas of analog CMOS and BiCMOS circuit design, oversampling, and adaptive systems.

Dr. Johns is an Associate Editor for IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing.