Granular Quantization Noise in the First-Order Delta–Sigma Modulator

Ian Galton, *Member, IEEE*

Abstract—Delta–sigma ($\Delta\Sigma$) modulators are attractive candidates for oversampling analog-to-digital (A/D) converters because they are amenable to VLSI implementation and have low component sensitivity. However, because they are nonlinear systems, they have proven difficult to analyze. Rigorous analyses have been performed only for a small number of artificial input sequences such as constant, sinusoidal, and Gaussian white noise input sequences. By allowing for the inevitable presence of small amounts of noise in the $\Delta\Sigma$ modulator circuitry, a general framework is developed which extends the repertoire of tractable input sequences to a large class of stochastic sequences in addition to handling many input sequences for which results have been previously presented. Under the assumptions that some circuit noise is present and that the input sequence does not cause overload, a simple autocorrelation expression is developed that is only locally dependent upon the input sequence. Ergodic properties are derived and various examples are presented.

Index Terms—Sigma-delta, delta-sigma, oversampling, analog-to-digital conversion, quantization.

I. Introduction

THE first-order $\Delta\Sigma$ modulator [1] is the simplest of a class of systems generally referred to as $\Delta\Sigma$ modulators that employ sampled-data filters and coarse quantizers within feedback loops. They are widely used in high-precision oversampling A/D converters because they are well suited to VLSI implementation and tend to be robust with respect to nonideal components. Accordingly, they have received much attention from both academic and industrial researchers. Nevertheless, most of the previously published rigorous theoretical analyses of $\Delta\Sigma$ modulators apply only to a small set of input sequences. In the current work, we concentrate on the first-order $\Delta\Sigma$ modulator and provide rigorous results for a large class of input sequences.

The first-order $\Delta\Sigma$ modulator consists of a sampled-data integrator, a uniform midrise quantizer [2], and a negative feedback loop surrounding the integrator and quantizer, as shown in Fig. 1(a). The system operates on a sampled-data input $x(n)$ and produces a quantized output $y(n)$. The quantizer can be interpreted as an additive quantization noise source, as depicted in Fig. 1(b). A straightforward linear systems analysis shows that the input sequence sees the one-sample delay $S(z) = z^{-1}$ while the quantization noise sequence sees the high-pass filter $N(z) = 1 - z^{-1}$. Thus, as shown in Fig. 1(c), the output consists of two components: a component corresponding to the input sequence, and a component corresponding to the quantization noise sequence.

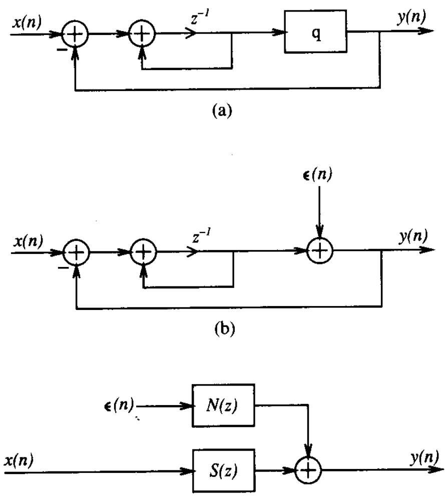

Fig. 1. (a) The first-order $\Delta\Sigma$ modulator. (b) An equivalent form of the system with the quantizer represented as an additive quantization noise source. (c) An equivalent form of the system showing the different filters that act on the input and quantization noise sequences, respectively.

Note that $N(z)$ is a high-pass filter with a zero at zero frequency. This causes the spectral energy of the quantization error at the output of the $\Delta\Sigma$ modulator to be weighted toward the high-frequency end of the spectrum for most input sequences [3]. It is this property of the $\Delta\Sigma$ modulator that makes it useful in oversampling A/D converters.

An oversampling A/D converter consists of a $\Delta\Sigma$ modulator followed by a low-pass decimation filter, as shown in Fig. 2. The input to the $\Delta\Sigma$ modulator $x(n)$ is obtained by sampling a bandlimited analog signal at a rate Nf where N is a positive integer and f is the Nyquist rate. Therefore, the spectrum of $x(n)$ is nonzero only on $(-\pi/N, \pi/N)$ where 2π corresponds to the sampling rate. Provided N is sufficiently large, the spectral energy of the quantization error will fall mostly outside $(-\pi/N, \pi/N)$. The low-pass filter removes the out-of-

Manuscript received December 10, 1991; revised March 12, 1993. This work was supported by a grant from Pacific Bell.

The author is with the Department of Electrical and Computer Engineering, University of California, Irvine, CA 92717.

IEEE Log Number 9212918.

0018-9448/93$03.00

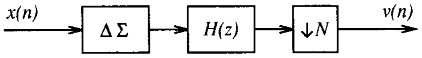

Fig. 2. A ΔΣ modulator based on oversampling A/D converter.

band quantization error, and the decimator reduces the output sequence to the Nyquist rate.

Although conceptually simple, the system has proven difficult to analyze because of the nonlinearity introduced by the quantizer. As will be shown, the quantization noise has a complicated structure that is globally dependent upon the input sequence. If two input sequences differ at just one sample time, say $n = n_0$, then the corresponding quantization noise sequences will appear very different for all $n > n_0$.

The quantizer imposes the following nonlinearity on its input:

$$q(x) = \begin{cases} \Delta\left\lfloor \dfrac{x}{\Delta} \right\rfloor + \dfrac{\Delta}{2} & \text{if } -\gamma \le x < \gamma \\ \gamma - \dfrac{\Delta}{2} & \text{if } x \ge \gamma \\ -\gamma + \dfrac{\Delta}{2} & \text{if } x < -\gamma \end{cases} \tag{1}$$

where γ is usually an integer multiple of Δ. When the input to the quantizer has absolute value greater than γ, the quantizer is said to *overload*. It is desirable to avoid the overload condition because the resulting distortion tends to be severe and difficult to characterize [4]–[10]. Most of the existing ΔΣ modulator analyses, including ours, assume that the overload condition is avoided. Since the quantizer will not overload provided the ΔΣ modulator input sequence is bounded in absolute value by $\gamma - \Delta/2$ [8], this is not an unreasonable assumption. Furthermore, simulations show that if the overload condition occurs, but does so only rarely, then the performance of the ΔΣ modulator is not significantly degraded [11]. We can therefore expect any exact results obtained under the no-overload assumption to approximately hold if the overload condition has a low frequency of occurrence.

Even under the no-overload restriction, the system does not yield to a straightforward analysis. Most analyses rely on approximations [1], [3], [12], or apply only to specific input sequences such as constant [4], [5], sinusoidal [6], or Gaussian white noise sequences [7].

In the current work, we develop rigorous results by assuming that the input sequence contains an additive independent identically distributed (iid) random component. The assumption is not very restrictive because the random component can have an arbitrarily small variance. Moreover, since thermal noise in the analog front end of the ΔΣ modulator can be modeled as an iid random sequence, the assumption is reasonable in practice. The approach has the benefit that it can be applied to a large class of input sequences. We develop a simple expression for the autocorrelation of the quantization error $R_{ee}(n, p)$, and show that it is equal in probability to the corresponding time-averaging autocorrelation. The autocorrelation expression is convenient because it is only locally dependent on the input sequence. This property makes tractable many desired input sequences that cannot be handled using previously presented theory, such as the class of arbitrary stochastic sequences respecting the no-overload constraint.

In Section II, we derive the theory outlined above, and in Section III we apply it to specific input sequences. By considering constant and sinusoidal inputs, the theory is shown to contain many of the existing results concerning the first-order ΔΣ modulator as special cases, although new observations are also presented. In particular, for a sinusoidal input, we develop a closed-form expression for the quasi-stationary autocorrelation of the quantization error. Additional classes of sequences, which heretofore have not been rigorously analyzed in conjunction with the ΔΣ modulator, are then considered. We also present simulation results to support our theoretical analysis.

II. Theoretical Analysis

Instead of considering the ΔΣ modulator in isolation, our system of study will be the ΔΣ modulator followed by a causal, stable, linear time-invariant digital filter with transfer function $H(z)$ and impulse response $h(n)$, as shown in Fig. 3. The reason for not considering the ΔΣ modulator in isolation is that, in practice, it is almost always followed by a filter and, as we will show, the statistics of the output are dependent upon the filter. Since we could choose $H(z) = 1$, the isolated ΔΣ modulator is a special case of our system.

We will distinguish between the *quantization noise sequence*, $\epsilon(n)$, and the *quantization error sequence*, $e(n)$. As shown in Fig. 1(b), the quantization noise sequence is the difference between the output and the input of the quantizer. It is the noise injected into the system by the quantizer. The quantization error sequence is the component of the output of the system in Fig. 3 corresponding to the quantization noise. As mentioned above, the ΔΣ modulator subjects the quantization noise sequence to the filter $N(z) = 1 - z^{-1}$. Thus, the quantization error sequence is equivalent to the output of the filter $(1 - z^{-1})H(z)$ when driven by the quantization noise sequence. From the argument leading to Fig. 1(c), it follows that we can write the output of the system in Fig. 3 as

$$r(n) = w(n) + e(n), \tag{2}$$

where $w(n)$ can be interpreted as the response of the filter $z^{-1}H(z)$ to the input sequence $x(n)$.

As alluded to above, we will assume that the input sequence seen by the ΔΣ modulator consists of a *desired input sequence*, $x_d(n)$, plus an *input noise sequence*, $\{\eta_n\}$:

$$x(n) = x_d(n) + \eta_n. \tag{3}$$

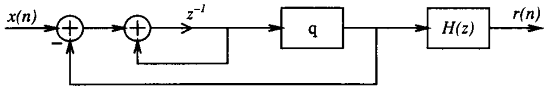

Fig. 3. A first-order $\Delta\Sigma$ modulator followed by the filter $H(z)$.

We require that the η_n are independent and identically distributed with a distribution that has a density. The desired input sequence is the sampled-data signal that is to be converted into a digital sequence by the $\Delta\Sigma$ modulator (e.g., the music signal, the video signal, etc.), and the input noise sequence is an unrelated sequence that is assumed to be present in the analog front-end of the $\Delta\Sigma$ modulator. The assumption is realistic in practice because thermal noise which is ubiquitous in analog circuity can be accurately modeled as an iid random sequence in sampled data systems.

A. An Expression for the Quantization Error Sequence

In the calculations to follow, we will consider the $\Delta\Sigma$ modulator to have been "turned on" at a specific time in the past. For all $n \le a$, we will take the input sequence and all storage elements in the $\Delta\Sigma$ modulator and filter to be zero. In some cases, we will consider the system in the limit as $a \to -\infty$. This corresponds to a system that has always been running.

Gray [8] has shown that the quantization sequence can be written as

$$\epsilon(n) = \frac{\Delta}{2} - \Delta\left\langle \frac{1}{\Delta}\sum_{i=1}^{n-a}\left[x(n-i) + \frac{\Delta}{2}\right]\right\rangle \tag{4}$$

provided $n > a$.[1] For convenience, we will take $\epsilon(n) = 0$ whenever $n < a$.

Since $H(z)$ is causal, its impulse response, $h(n)$, is zero for all $n < 0$. Therefore, for $n > a$, we can write the quantization error sequence as

$$e(n) = \sum_{k=0}^{\infty}[h(k) - h(k-1)]\epsilon(n-k). \tag{5}$$

Again, for convenience, we will take $e(n) = 0$ whenever $n < a$. Combining these two equations gives

$$e(n) = \Delta\sum_{k=0}^{n-a-1}[h(k) - h(k-1)] \cdot\left\langle \frac{1}{\Delta}\sum_{i=1}^{n-k-a}\left[x(n-k-i) + \frac{\Delta}{2}\right]\right\rangle. \tag{6}$$

Although (6) is an exact formula for the quantization error sequence, it does not give great insight into the long-term behavior of the quantization error sequence. In particular, the specific quantization error sequence obtained for a given input sequence is globally dependent upon each value of the input sequence. For example, consider two input sequences $x_1(n)$ and $x_2(n)$, which are identical except for their first value at time $n = a$. That is, suppose

$$x_2(n) = \begin{cases} x_1(n), & \text{if } n \ne a \\ x_1(n) + \beta, & \text{if } n = a \end{cases}$$

for some nonzero $\beta \in \boldsymbol{R}$ (such that the no-overload condition is maintained). Then the quantization error sequence associated with $x_1(n)$ is

$$e_1(n) = \Delta\sum_{k=0}^{n-a-1}[h(k) - h(k-1)] \cdot\left\langle \frac{1}{\Delta}\sum_{i=1}^{n-k-a}\left[x_1(n-k-i) + \frac{\Delta}{2}\right]\right\rangle,$$

while the quantization error sequence associated with $x_2(n)$ is

$$e_2(n) = \Delta\sum_{k=0}^{n-a-1}[h(k) - h(k-1)] \cdot\left\langle \beta\frac{1}{\Delta} + \frac{1}{\Delta}\sum_{i=1}^{n-k-a}\left[x_1(n-k-i) + \frac{\Delta}{2}\right]\right\rangle.$$

Because of the presence of β, each term of the first sum in the equation for $e_2(n)$ differs in a complicated fashion from the corresponding term in $e_1(n)$; the two quantization error sequences typically look very different.

B. Quantization Noise Statistics

Chou and Gray [13] have investigated the statistics of the quantization noise sequence of the first-order $\Delta\Sigma$ modulator under the assumptions that overload is avoided and that the input sequence consists of a deterministic sequence plus a so-called *dither sequence* that is iid with a density. Mathematically, the dither sequence assumption is equivalent to (3); our input noise sequence plays the role of the dither sequence. The reason that we do not refer to the input noise sequence as a dither sequence is that the term dither is usually applied to sequences that are intentionally introduced. From a practical point of view, we are making the opposite assumption. We consider the presence of the input noise sequence to be an inevitable result of the $\Delta\Sigma$ modulator having an analog front-end. The practical consequence of our distinction is that the results presented in this paper hold regardless of whether a dither sequence is intentionally added. Nevertheless, the results presented by Chou and Gray can be applied directly to our system.

In particular, they proved that the quantization noise sequence converges in distribution to a random variable that is uniformly distributed on $(-\Delta/2, \Delta/2)$ and is independent of the desired input sequence. In order to extend their work, it is convenient to begin by stating this result in a slightly different form. We do this in the following lemma.

[1] The angle brackets denote the fractional part operator. This operator is defined as $\langle x\rangle = x - \lfloor x\rfloor$ for all $x \in \boldsymbol{R}$.

Lemma 1: For each $r = 1, 2, \cdots$, let

$$U_r = \langle \mu_r + \sum_{i=1}^{r} c\eta_i \rangle \tag{7}$$

where $\{\mu_r\}$ is any deterministic sequence, c is any nonzero real number, and $\{\eta_i\}$ is a sequence of independent, identically distributed random variables whose distribution has a density. Then, as $r \to \infty$, U_r converges in distribution to a random variable U that is uniformly distributed on [0, 1).

Proof: The proof is essentially the same as that presented in [13]. ■

The following lemma generalizes this result to stochastic sequences $\{\mu_r\}$.

Lemma 2: Let $\{U_r\}$, c, and $\{\eta_i\}$ be as defined in the hypothesis of Lemma 1. Let $\{\mu_r\}$ be any stochastic sequence that is independent of $\{\eta_i\}$. Then, as $r \to \infty$, U_r converges in distribution to a random variable U that is uniformly distributed on [0, 1).

Proof: The moments of U_r are defined as $E(U_r^n)$, for $n = 1, 2, \cdots$. Because $U_r \le 1$ with probability 1, each moment exists and has absolute value less than or equal to one. Thus, the distribution of U_r is uniquely determined by its moments, and it is sufficient to show that the moments of U_r converge to the corresponding moments of U as $r \to \infty$.[2]

Since the sequences $\{\eta_k\}$ and $\{\mu_k\}$ are independent, for any integer n we can write

$$E(U_r^n) = E[E(U_r^n \mid \mu_1, \mu_2, \cdots)].$$

By Lemma 1, for any deterministic real sequence $\{a_1, a_2, \cdots\}$,

$$E(U_r^n \mid \mu_1 = a_1, \mu_2 = a_2, \cdots) \to E(U^n)$$

as $r \to \infty$. It follows that

$$E(U_r^n \mid \mu_1, \mu_2, \cdots) \to E(U^n)$$

with probability 1 as $r \to \infty$.

By definition, $E(U_r^n \mid \mu_1, \mu_2, \cdots) \le 1$ with probability 1. Therefore, it follows from the Lebesgue dominated convergence theorem that

$$E[E(U_r^n \mid \mu_1, \mu_2, \cdots)] \to E(U^n)$$

as $r \to \infty$. ■

In accordance with the usual definitions, we will take the mean and autocorrelation of the quantization noise sequence to be

$$M_\epsilon(n) = \lim_{a \to -\infty} E[\epsilon(n)],$$

and

$$R_{\epsilon\epsilon}(n, p) = \lim_{a \to -\infty} E[\epsilon(n)\epsilon(n + p)],$$

respectively. We will take the cross correlation of the quantization noise sequence and the desired input sequence to be

$$R_{x_d\epsilon}(n, p) = \lim_{a \to -\infty} E[x_d(n)\epsilon(n + p)].$$

We will take the mean, autocorrelation, and cross correlation of the quantization error sequence, namely, $M_e(n)$, $R_{ee}(n, p)$, and $R_{x_de}(n, p)$ to be analogously defined.

The following theorem is an extension of a result proven by Chou and Gray. They proved the result under the restriction that the desired input sequence is deterministic. The current result holds for deterministic and stochastic desired input sequences.

Theorem 3: For deterministic or stochastic desired input sequences, $M_\epsilon(n)$ and $R_{x_d\epsilon}(n, p)$ are zero. Consequently, $M_e(n)$ and $R_{x_de}(n, p)$ are also zero.

Proof: If $M_\epsilon(n)$ and $R_{x_d\epsilon}(n, p)$ are zero, then by the linearity of $H(z)$, it follows that $M_e(n)$ and $R_{x_de}(n, p)$ are zero. Therefore, it is sufficient to show that $M_\epsilon(n)$ and $R_{x_d\epsilon}(n, p)$ are zero.

From (4), for each $n > a$, we can write $\epsilon(n) = (\Delta/2) - \Delta U_{n-a}$ where U_{n-a} corresponds to U_r in Lemma 1 with $c = 1/\Delta$ and

$$\mu_r = \frac{1}{\Delta} \sum_{i=1}^{r} \left[x_d(n - i) + \frac{\Delta}{2} \right].$$

From Lemma 2,

$$\lim_{r \to \infty} E(U_r) = \tfrac{1}{2}, \tag{8}$$

which implies that $M_\epsilon(n) = 0$. Moreover, (8) holds regardless of the value taken on by $x_d(n)$. Hence,

$$\lim_{r \to \infty} E[x_d(n)U_r] = \tfrac{1}{2}E[x_d(n)],$$

which implies that $R_{x_d\epsilon}(n, p) = 0$. ■

Assuming for now that autocorrelation functions for $e(n)$ and $w(n)$ exist, Theorem 3 in conjunction with (2) implies that the autocorrelation of the output of the system of Fig. 3 can be written as

$$R_{rr}(n, p) = R_{ww}(n, p) + R_{ee}(n, p).$$

Therefore, the significance of Theorem 3 is that the autocorrelation of the quantization error sequence, if it exists, characterizes the second-order statistics of the quantization error.

The following theorem shows that $R_{ee}(n, p)$ indeed exists and provides a convenient expression for its evaluation.

Theorem 4: The autocorrelation of the quantization noise sequence can be written as

$$R_{\epsilon\epsilon}(n, p) = E[r(n, n + p)], \tag{9}$$

[2] See, for example, 14, Theorems 30.1 and 30.2].

where

$$r(n,m) = \begin{cases} \dfrac{\Delta^2}{12}, \\ \text{if } n = m \\ \dfrac{1}{2}\left[\dfrac{\Delta}{2} - \Delta\left\langle \dfrac{1}{\Delta}\sum_{i=m}^{n-1}\left[x(i) + \dfrac{\Delta}{2}\right]\right\rangle\right]^2 - \dfrac{\Delta^2}{24}, \\ \text{if } n > m \\ \dfrac{1}{2}\left[\dfrac{\Delta}{2} - \Delta\left\langle \dfrac{1}{\Delta}\sum_{i=n}^{m-1}\left[x(i) + \dfrac{\Delta}{2}\right]\right\rangle\right]^2 - \dfrac{\Delta^2}{24}, \\ \text{if } n < m. \end{cases} \tag{10}$$

Consequently, the autocorrelation of the quantization error sequence can be written as

$$R_{ee}(n,p) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}[h(j) - h(j-1)][h(k) - h(k-1)] \cdot R_{\epsilon\epsilon}(n-k, n+p-j). \tag{11}$$

Proof: As in the proof of Theorem 3, write $\epsilon(n) = (\Delta/2) - \Delta U_{n-a}$. Then,

$$E[\epsilon(n)\epsilon(n+p)] = \frac{\Delta^2}{4} - \frac{\Delta^2}{2}E(U_{n-a}) - \frac{\Delta^2}{2}E(U_{n+p-a}) + \Delta^2 E(U_{n-a}U_{n+p-a}).$$

From Lemma 2, it follows that

$$\lim_{a\to-\infty} E[\epsilon(n)\epsilon(n+p)] = -\frac{\Delta^2}{4} + \Delta^2 \lim_{a\to-\infty} E(U_{n-a}U_{n+p-a}).$$

Therefore, to prove (9), it is sufficient to show that

$$\lim_{a\to-\infty} E(U_{n-a}U_{n+p-a}) = \frac{1}{\Delta^2}E[r(n,n+p)] + \frac{1}{4}. \tag{12}$$

If $p = 0$, (12) holds as a direct consequence of Lemma 2. Therefore it is sufficient to prove that (12) holds for $p \geq 1$.

The fractional part operator has the property that for any $x, y \in \boldsymbol{R}$, $\langle x + y\rangle = \langle\langle x\rangle + y\rangle$. It follows that for $p \geq 1$.

$$U_{n+p-a} = \left\langle U_{n-a} + \frac{1}{\Delta}\sum_{i=0}^{p-1}\left[x(n+i) + \frac{\Delta}{2}\right]\right\rangle.$$

Therefore, by Lemma 2,

$$\lim_{a\to-\infty} E(U_{n-a}U_{n+p-a}) = E\left[U\left\langle U + \frac{1}{\Delta}\sum_{i=0}^{p-1}\left[x(n+i) + \frac{\Delta}{2}\right]\right\rangle\right], \tag{13}$$

where U is uniformly distributed on [0, 1). The expectation on the right side of (13) can be evaluated in closed form. However, the algebra is messy so it has been relegated to the Appendix as Lemma A1. Applying Lemma A1 to (13) for $p \geq 1$ gives

$$\lim_{a\to-\infty} E(U_{n-a}U_{n+p-a}) = \frac{1}{3} + \frac{1}{2}E\left\langle \frac{1}{\Delta}\sum_{i=0}^{p-1}\left[x(n+i) + \frac{\Delta}{2}\right]\right\rangle^2 - \frac{1}{2}E\left\langle \frac{1}{\Delta}\sum_{i=0}^{p-1}\left[x(n+i) + \frac{\Delta}{2}\right]\right\rangle.$$

This can be rearranged as (12) so the proof of (9) is complete.

Combining (5) and the autocorrelation definition gives

$$R_{ee}(n, n+p) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}[h(j) - h(j-1)][h(k) - h(k-1)] \cdot \lim_{a\to-\infty} E[\epsilon(n-j)\epsilon(n+p-k)],$$

which is equivalent to (11). The term-by-term multiplication of the series for $e(n)$ and $e(n+p)$ and the interchange of the limit and the sums are justified because the impulse response, $h(n)$, is absolutely summable (because $H(z)$ is stable). Hence, (11) is a direct consequence of (9) and the stability of the filter. ■

Several observations can be made regarding (9). Note that $r(n,m)$ is formally a constant offset plus the squared quantization error of a uniform midrise quantizer operating upon a finite partial sum of the input sequence. Thus, the quantization error autocorrelation is the weighted sum of the mean-squared errors of multiple uniform quantizers operating on various partial sums of the input sequence.

Another observation which we anticipated in the Introduction is that the quantization error autocorrelation is only locally dependent upon the input sequence. That is, for a given p, the dependence of $R_{ee}(n,p)$ upon the set of input values $\{x(k): k < n - N\}$ can be made arbitrarily small by increasing N. This is a consequence of the impulse response of the filter, $h(n)$, being absolutely summable. If $H(z)$ is an FIR filter, then the stronger assertion can be made that $R_{ee}(n,p)$ is only dependent upon a finite number of values from the input sequence for a given p.

Suppose we were given M systems, all operating on the same desired input sequence at the same time. Each system would produce a quantization error sequence $e_i(n)$ that would differ from the other quantization error sequences because of the random variables η_n. In this case, by the law of large numbers,

$$\frac{1}{M}\sum_{i=0}^{M-1} e_i(n)e_i(n+p) \approx R_{ee}(n, n+p) \tag{14}$$

provided M is large (and $a \ll n$).

If Theorem 4 were useful only to the extent that we could predict the average behavior of many identical systems per (14), it would be of limited use. It is more often the case in practice that we are interested in the long-term time-average behavior of $e(n)e(n+p)$. The question therefore arises as to what bearing the statistical autocorrelation has on the time-average behavior of $e(n)e(n+p)$.

Relationships between time and ensemble averages are usually referred to as ergodic properties [15], [16]. The following theorem and corollary present ergodic results which greatly extend the utility of Theorem 4.

Theorem 5: The following equations:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\epsilon(n)=0 \tag{15}$$

and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}x_d(n)\epsilon(n+p)=0, \tag{16}$$

hold in probability. Moreover, whenever one of the limits exists,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\epsilon(n)\epsilon(n+p)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}R_{\epsilon\epsilon}(n,p) \tag{17}$$

holds in probability. In particular, the limits exist if the desired input sequence is quasi-stationary.

Proof: As in the proof of Theorem 3, we begin by writing $\epsilon(n)=(\Delta/2)-\Delta U_{n-a}$. Without loss of generality, we will assume that $a=0$. It follows from the proof of Theorem 3 that

$$E(U_k \mid \eta_0,\cdots,\eta_j,\mu_0,\mu_1,\cdots)\to\frac{1}{2}$$

with probability 1 (and, consequently, in probability) as $k-j\to\infty$ with $k>j\geq 0$.

Applying Lemma A2 (presented in the Appendix) with $U_n-\frac{1}{2}$ playing the role of X_n gives

$$\frac{1}{N}\sum_{n=0}^{N-1}U_n\to\frac{1}{2}$$

in probability. Therefore, (15) holds in probability. The argument that (16) holds in probability is almost identical.

To show that (17) holds in probability provided either limit exists, it is sufficient to show that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}[\epsilon(n)\epsilon(n+p)-R_{\epsilon\epsilon}(n,p)]=0 \tag{18}$$

holds in probability. Since $\epsilon(n)=(\Delta/2)-\Delta U_n$, a sufficient condition for (18) to hold in probability is that both

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left[U_n-\frac{1}{2}\right]\to 0 \tag{19}$$

and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left[U_nU_{n+p}-\frac{1}{\Delta^2}R_{\epsilon\epsilon}(n,p)-\frac{1}{4}\right]\to 0 \tag{20}$$

hold in probability. We have already shown that (19) holds in probability, so we can conclude that (18) holds in probability if (20) holds in probability.

Define

$$X_n=U_nU_{n+p}-\frac{1}{\Delta^2}R_{\epsilon\epsilon}(n,p)-\frac{1}{4}.$$

From the proof of Theorem 4, it follows that

$$E(X_k \mid \eta_0,\cdots,\eta_j,\mu_0,\mu_1,\cdots)\to 0$$

with probability 1 as $k-j\to\infty$ with $k>j\geq 0$. It follows from Lemma A2 that (20) holds in probability, which completes the proof that (17) holds in probability provided either limit exists.

If the desired input sequence is quasi-stationary, then the quantization noise sequence is also quasi-stationary [13], and so the limit on the right side of (17) exists. In this case, since (18) holds in probability, (17) must hold in probability. ■

Corollary 6: The following equations

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}e(n)=0 \tag{21}$$

and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}x_d(n)e(n+p)=0 \tag{22}$$

hold in probability. Moreover, whenever one of the limits exists,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}e(n)e(n+p)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}R_{ee}(n,p) \tag{23}$$

holds in probability. In particular, the limits exist if the desired input sequence is quasi-stationary.

Proof: Each equation follows formally by expanding $e(n)$ with (5) and applying Theorem 5. In each case, the various limits and sums can be interchanged because the impulse response of $H(z)$ is absolutely summable. ■

There are various ergodic theorems that have been or can be applied to the first-order ΔΣ modulator for specific classes of input sequences [16]–[19]. However, the published ergodic theorems do not apply to the class of input sequences considered in the current work.

We now develop procedures for applying the theory presented thus far to arbitrary input sequences. Recall that our theory requires the input sequence to contain the random variables η_n. Therefore, even if the desired input sequence is deterministic, the actual input sequence is stochastic. As argued in the Introduction, this assumption is realistic in practice. However, many of the existing results for the ΔΣ modulator are in terms of purely deterministic signals. So that we can later compare our theory to existing work, we first develop a systematic approach to annexing the deterministic case into our theory. We then consider the more important class of arbitrary stochastic input sequences with known statistics. Finally, we present a procedure for obtaining approximate results when the statistics of the desired input sequence are not fully known. To avoid cluttering the development, specific examples are deferred to the next section.

C. Deterministic Input Sequences

Most of the treatments concerning deterministic input sequences involve the evaluation of the quasi-stationary autocorrelation $R_\epsilon(p)$. In this case,

$$R_\epsilon(p) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} [\epsilon(n)\epsilon(n+p)]$$

since the input sequence does not contain a random component (see [8] and [20] for a discussion of quasi-stationary processes).

To circumvent the restriction that the input sequence contain the random variables η_n, we take the limiting case as the distribution function of the η_n approaches a unit step function at the origin (i.e., as the η_n converge in distribution to a random variable that is zero with probability 1). From Theorem 5 and the definition of the quasi-stationary autocorrelation,

$$R_\epsilon(p) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} R_{\epsilon\epsilon}(n,p) \tag{24}$$

in probability. Consider a sequence of absolutely continuous probability distribution functions (i.e., probability distributions with densities) $\{P_k(x)\}$ such that

$$\lim_{k\to\infty} P_k(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Let $R_{\epsilon\epsilon}(n,p)\mid_{P_k(x)}$ be the value of $R_{\epsilon\epsilon}(n,p)$ corresponding to η_n with distribution function $P_k(x)$. From (9) and (10), we have

$$\lim_{k\to\infty} R_{\epsilon\epsilon}(n,p)\mid_{P_k(x)} = \begin{cases} \dfrac{\Delta^2}{12}, \\ \quad \text{if } p = 0 \\ \dfrac{1}{2}\left[\dfrac{\Delta}{2} - \Delta\left\langle \dfrac{p}{2} + \dfrac{1}{\Delta}\displaystyle\sum_{i=n}^{n+p-1} x_d(i)\right\rangle\right]^2 - \dfrac{\Delta^2}{24}, \\ \quad \text{if } p > 0 \\ \dfrac{1}{2}\left[\dfrac{\Delta}{2} - \Delta\left\langle \dfrac{p}{2} + \dfrac{1}{\Delta}\displaystyle\sum_{i=n+p}^{n-1} x_d(i)\right\rangle\right]^2 - \dfrac{\Delta^2}{24}, \\ \quad \text{if } p < 0. \end{cases} \tag{25}$$

Define

$$\hat{R}_\epsilon(p) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \lim_{k\to\infty} R_{\epsilon\epsilon}(n,p)\mid_{P_k(x)}.$$

Applying (25) gives

$$\hat{R}_\epsilon(p) = \lim_{N\to\infty} \frac{\Delta^2}{2N} \sum_{n=0}^{N-1} \left[\frac{1}{2} - \left\langle \frac{p}{2} + \frac{1}{\Delta} \sum_{i=n}^{n+|p|-1} x_d(i)\right\rangle\right]^2 - \frac{\Delta^2}{24}. \tag{26}$$

Care must be taken to properly interpret $\hat{R}_\epsilon(p)$. It is tempting to consider it to *be* the quantization noise autocorrelation corresponding to the deterministic input signal $x_d(n)$, without any contribution from the η_n. As we shall see, in some cases, this interpretation is valid, and in other cases, it is not. In general, what can be said is that given a deterministic input signal, for any $\epsilon > 0$ there is an uncountable infinity of deterministic signals, each of which has a mean-squared difference from the original signal of less than ϵ and a quasi-stationary autocorrelation equal to $\hat{R}_\epsilon(p)$. In cases where $R_\epsilon(p) = \hat{R}_\epsilon(p)$, the existence of the limit in (25) implies that the ideal result for the purely deterministic input sequence is approximately valid if some noise is present. It becomes increasingly accurate as the noise level is reduced. Of course, this can be determined from simulations and observations of actual systems, but the argument above provides a theoretical basis for the behavior. In cases where $R_\epsilon(p) \neq \hat{R}_\epsilon(p)$, we should be wary of applying the deterministic analysis to a physical $\Delta\Sigma$ modulator implementation. With the slightest amount of noise per (3), the autocorrelation will differ from that of the noiseless case and be very close to $\hat{R}_\epsilon(p)$ in probability. In this sense, the purely deterministic result is not a physically stable solution. Examples of such physically unstable deterministic solutions are presented in Section III.

D. Stochastic Input Sequences

In conjunction with existing results concerning uniform quantizers, our theory can be used to handle stochastic desired input sequences respecting the no-overload constraint. The first task is to evaluate the autocorrelation function $R_{ee}(n,p)$ of the quantization error sequence.

As is evident from Theorem 4 and the observations following it, to evaluate this expression, we must evaluate the mean-squared quantization error corresponding to various quantized partial sums of the input sequence. It is in solving this part of the problem that we benefit from existing results concerning uniform quantizers. If the resulting expression for $R_{ee}(n,p)$ is not dependent upon n, then the quantization error sequence is wide-sense stationary, and we are done. Otherwise, we may perform a time average of $R_{ee}(n,p)$ to obtain the quasi-stationary autocorrelation function. In either case, Corollary 6 ensures that the resulting function, if it exists, converges to the time average of $e(n)e(n+p)$ in probability.

For a given input sequence, the success of our approach depends upon evaluating the mean-squared quantization error of partial sums of the input sequence. Fortunately, considerable attention has been devoted to analyzing the effect of uniform quantization upon stochastic sequences [8], [21]–[24]. In particular, Sripad and Snyder [24] have derived an exact expression for the probability density function of a quantized sequence. If the statistics of $x(n)$ are known, then, using Sripad and Snyder's expression, $R_{ee}(n, p)$ can be evaluated easily. In particular, if the filter has length M and we know all $2M$ and lower joint probability distribution functions of the input sequence, we can calculate the mean-squared quantization error $\sigma_e^2 = R_{ee}(n, n)$. If we know the $2M + N$ and lower joint probability distribution functions of the input sequence, we can calculate $R_{ee}(n, p)$ for all $|p| \leq N$. As will be demonstrated in the next section, the first step is to apply Sripad and Snyder's expression to calculate the probability distributions of the quantized partial sums. It is then straightforward to evaluate $R_{\epsilon\epsilon}(n, p)$.

E. Approximate Analysis

Sometimes, the statistics of the input sequence are not known or are only partially known. For many such input sequences, our theory gives rise to approximate analyses. As with deterministic and stochastic input sequences, we benefit from having reduced the problem to one of evaluating the mean-squared error of a uniform quantizer operating on partial sums of the input sequence.

If the input to a uniform midrise quantizer is sufficiently "busy" or "active" on a scale that is larger than the quantization step size, Δ, it is common to approximate the quantization noise as uniformly distributed on $(-\Delta/2, \Delta/2)$ [8], [21]–[24]. In many cases, the partial sums of such a sequence also satisfy this property. Indeed, even if the individual members of the sequence do not satisfy the property, it is possible that partial sums of several members do satisfy the property.

Because (10) is an offset plus the squared quantization error of a partial sum of the input sequence, it follows that if all the partial sums are busy in the sense described above, then

$$R_{\epsilon\epsilon}(n, p) \approx \begin{cases} \dfrac{\Delta^2}{12}, & \text{if } p = 0 \\ 0, & \text{otherwise.} \end{cases}$$

In this case, from (11), we have

$$R_{ee}(n, p) \approx \frac{\Delta^2}{6}[(h(n) * h(-n))(p) - (h(n) * h(-n))(p + 1)].$$

In a ΔΣ modulator based oversampling A/D converter, it is not likely that the individual members of the desired input sequence are busy on a scale that is larger than Δ. However, the desired input sequence might be busy on a smaller scale. In this case, sequences of partial sums containing many terms might be busy on a scale that is larger than Δ. Thus, if we know only enough of the low-order statistics of the desired input sequence to evaluate (10) for all the cases where the uniform quantization noise approximation is not valid, we can obtain a good approximation to $R_{\epsilon\epsilon}(n, p)$.

Of course, the accuracy of the uniform quantization noise approximation is highly dependent upon the nature of the input sequence, and must be assessed on an individual basis. Fortunately, there exists a large body of work addressing issues of applicability and accuracy associated with the approximation.

III. Application to Specific Input Sequences

A. Constant-Amplitude Input Sequences

Although constant-amplitude input sequences have been considered by several people, Gray [5] was the first to perform an exact analysis. With $\Delta = 1$ and for an input sequence $x(n) = x$, where x is an irrational number bounded in absolute value by $\frac{1}{2}$, he showed that the quantization error sequence has a quasi-stationary autocorrelation given by

$$R_{\epsilon}(p) = \frac{1}{12} - \frac{1}{2}\left\langle p\left(\frac{1}{2} + x\right)\right\rangle\left(1 - \left\langle p\left(\frac{1}{2} + x\right)\right\rangle\right). \tag{27}$$

An equivalent result can be obtained from our theory. From (26), for any $x \in (-\frac{1}{2}, \frac{1}{2})$,

$$\begin{aligned} \hat{R}_{\epsilon}(p) &= \lim_{N \to \infty} \frac{1}{2N} \sum_{n=0}^{N-1} \left[\frac{1}{2} - \left\langle p\left(\frac{1}{2} + x\right)\right\rangle\right]^2 - \frac{1}{24} \\ &= \frac{1}{12} - \frac{1}{2}\left\langle p\left(\frac{1}{2} + x\right)\right\rangle\left(1 - \left\langle p\left(\frac{1}{2} + x\right)\right\rangle\right), \end{aligned}$$

which agrees with (27). Therefore, provided x is irrational, $R_{\epsilon}(p) = \hat{R}_{\epsilon}(p)$.

The form of $\hat{R}_{\epsilon}(p)$ is not dependent on whether x is rational or irrational. However, (27) does not hold for rational x [5]. It follows that $R_{\epsilon}(p) \neq \hat{R}_{\epsilon}(p)$ if x is rational. Since it is not possible to generate a perfectly constant rational voltage, there is little practical significance to this discrepancy. Adding a vanishingly small amount of noise to a rational constant input in accordance with (3) causes the quasi-stationary autocorrelation to approach $\hat{R}_{\epsilon}(p)$ in probability. Hence, the purely deterministic result is not a physically stable solution in the case of a rational constant input.

B. Sinusoidal Input Sequences

Because the ΔΣ modulator is not a linear system, its overall performance cannot be completely characterized by its response to sinusoidal inputs. Nevertheless, sinusoidal inputs are often used to test and partially characterize A/D converters. Therefore, sinusoidal input sequences have received considerable attention in the ΔΣ modulator literature. Gray *et al.* [6] were the first to perform an exact analysis. As in the constant-input case, they showed that the quantization noise sequence is quasi-stationary, and derived an exact expression for the quasi-stationary autocorrelation function. Unlike the constant input case, their expression is not in closed form. In contrast, our theory does yield a closed-form result.

Suppose $x_d(n) = A \cos n\omega_0$ where $|A| < \gamma - (\Delta/2)$. From (26), we have

$$\hat{R}_\epsilon(p) = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=0}^{N-1} \cdot \left[\frac{\Delta}{2} - \Delta \left\langle \frac{p}{2} + \frac{1}{\Delta} \sum_{i=n}^{n+|p|-1} A \cos i\omega_0 \right\rangle \right]^2 - \frac{\Delta^2}{24}.$$

After some trigonometric manipulation, this becomes

$$\hat{R}_\epsilon(p) = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=0}^{N-1} \left[\frac{\Delta}{2} - \Delta \left\langle \frac{1}{\Delta} \left\{ B(\omega_0, p) \cdot \sin[\omega_0 n + \theta(\omega_0, p)] + \frac{p\Delta}{2} \right\} \right\rangle \right]^2 - \frac{\Delta^2}{24},$$

where

$$B(\omega_0, p) = A \frac{\sin(\omega_0 |p|/2)}{\sin(\omega_0/2)}, \tag{28}$$

and

$$\theta(\omega_0, p) = \frac{1}{2}(|p| - 1)\omega_0 - \frac{\pi}{2}.$$

Note that for each even p, $\hat{R}_\epsilon(p)$ is equal to a constant plus the average squared error of a quantizer operating on a sinusoid. Similarly, for each odd p, it is equal to a constant plus the averaged squared error of a quantizer operating on a sinusoid offset by $\Delta/2$. Closed-form expressions for these quantities have been derived by Clavier *et al.* [25] and reformulated by Gray [8]. Our theory has thus reduced the problem to one that can be solved by applying results from another problem that has a known solution.

When $\omega_0/2\pi$ is an irrational number, we can apply the results to obtain

$$\hat{R}_\epsilon(p) = \begin{cases} \dfrac{\Delta^2}{12}, & p = 0 \\ \Delta^2 \left\{ \left(R - \dfrac{1}{2}\right)^2 + \dfrac{\zeta^2}{2} - \dfrac{2\zeta}{\pi} - \dfrac{2}{\pi} \displaystyle\sum_{k=1}^{R-1} \left[2k \sin^{-1}\left(\dfrac{k}{\zeta}\right) + 2\sqrt{\zeta^2 - k^2} \right] \right\}, & p \text{ even} \\ \Delta^2 \left\{ S^2 + \dfrac{\zeta^2}{2} - \dfrac{4\zeta}{\pi} - \dfrac{2}{\pi} \displaystyle\sum_{k=1}^{R-1} \left[(2k+1) \sin^{-1}\left(\dfrac{k + \frac{1}{2}}{\zeta}\right) + 2\sqrt{\zeta^2 - \left(k + \dfrac{1}{2}\right)^2} \right] \right\}, & p \text{ odd} \end{cases}$$

where $\zeta = (1/\Delta)B(\omega_0, p)$, $R = \lceil \zeta \rceil$, and $S = \lfloor \zeta + \frac{1}{2} \rfloor$ (when comparing these results to those in [8], note that various algebraic errors have been corrected). Similar results apply to the less important case in which $\omega_0/2\pi$ is a rational number.

Although not an intuitive result, the expression for $\hat{R}_\epsilon(p)$ is a closed-form expression and is simple to evaluate. From (28), it follows that $B(\omega_0, p) \le A|p|$ for ω_0. Thus, each sum has at most $(A|p|/\Delta) + 1$ terms. Since the autocorrelation is most interesting for values of p near the origin, and since A/Δ is usually less than one in practice, the sums rarely involve many terms.

Once again, it is interesting to answer the question of whether $\hat{R}_\epsilon(p)$ and $R_\epsilon(p)$ are equal. Because the expression for $R_\epsilon(p)$ presented in [6] is a double infinite summation of Bessel functions, a quantitative comparison of the two functions for all values of p is difficult. A simpler approach is to compare the functions when $p = 0$. In this case,

$$R_\epsilon(0) = \frac{\Delta^2}{12} - \Delta^2 \sum_{l=0}^{\infty} \frac{1}{(\pi 2l)^2} (-1)^l J_0(2\pi l \zeta / \sin(\omega_0/2)).$$

whereas $\hat{R}_\epsilon(0) = \Delta^2/12$. Clearly, the two expressions are not equal. A similar, but more involved analysis shows that they are not equal for most finite values of p. As in the case of rational constant inputs, that $R_\epsilon(p)$ and $\hat{R}_\epsilon(p)$ differ indicates that the purely deterministic result is not a physically stable solution. For example, the slightest amount of noise added according to (3) causes the second term in the expression for $R_\epsilon(0)$ to vanish in probability.

C. A Simple Class of Stochastic Input Sequences

In the following, we will assume that the variance of the input noise sequence is so small that we can ignore its effect when evaluating (9). This is not a necessary assumption, but it makes the calculations simpler. In an actual $\Delta\Sigma$ modulator, the assumption is equivalent to assuming that the circuit noise floor is significantly below the quantization noise floor.

As a first test of our theory for stochastic input sequences, suppose that $x_d(n)$ is a sequence of independent random variables with characteristic functions satisfying $\Phi_{x_d(n)}(2\pi n/\Delta) = 0$ for all $n \neq 0$. For example, a sequence satisfies this property if each member has a mean of β_n and is uniformly distributed on $[\beta_n - \Delta/2, \beta_n + \Delta/2)$ (respecting the no-overload constraint). Such a sequence might be created by adding a stochastic dither sequence d_n, satisfying the characteristic function equation above, to a deterministic sequence, β_n.

Since the members of the sequence are independent, adding them together corresponds to multiplying their characteristic functions. Hence, any partial sum of the form

$$S_{n,m} = \sum_{i=m}^{n-1}\left[x(i) + \frac{\Delta}{2}\right]$$

has a characteristic function satisfying $\Phi_{S_{n,m}}(2\pi n/\Delta) = 0$ for all $n \neq 0$. This is a necessary and sufficient condition for the error produced by quantizing $S_{n,m}$ to be uniformly distributed [24], [26]. Applying this result to evaluate (10) gives

$$E[r(n,m)] = \begin{cases} \dfrac{\Delta^2}{12}, & \text{if } n = m \\ 0, & \text{if } n \neq m. \end{cases}$$

For the case of a ΔΣ modulator with a 1-bit quantizer (i.e., the case of $\gamma = 1$ and $\Delta = 1$), Chou and Gray [13] have presented an equivalent result. They pointed out that the result is of limited use because in order to satisfy the no-overload constraint, the input sequence must have zero mean (i.e., $\beta_n = 0$ must hold for all n). Although the ΔΣ modulator is most commonly implemented with a 1-bit quantizer, multibit quantizers are sometimes used (see, for example, [27]). In such cases, the restriction on the input sequence (when each member of the input sequence is uniformly distributed on $[\beta_n - \Delta/2, \beta_n + \Delta/2])$ is that $-\gamma + \Delta < \beta_n < \gamma - \Delta$ for all n. Since $\Delta = 2\gamma/(2^R - 1)$, where R is the number of quantizer bits, this is not necessarily a severe restriction. For example, if an 8-bit quantizer with $\gamma = \frac{1}{2}$ is used in the ΔΣ modulator, then the allowed range of the input sequence means is $[-(\Delta/2) + (1/255), (\Delta/2) - (1/255)]$, which is *99.6%* of the full dynamic range of the input.

D. Gaussian Input Sequences

As outlined in the previous section, the general procedure for determining $R_{ee}(n,p)$ involves calculating the mean-squared quantization error of various partial sums of the input sequence. In the example just considered, this was particularly simple because all the partial sums had uniformly distributed quantization error. For arbitrary nonstationary stochastic input sequences, the method is still straightforward, but the calculations can be tedious because each partial sum may have a different distribution. In such cases, the calculations are most easily performed using a computer.

To illustrate the general method, but nevertheless obtain results that can be verified by hand, we consider the case of a stationary Gaussian desired input sequence. Because all partial sums of such sequences have Gaussian distributions, we need not explicitly determine the distribution of each partial sum so the tedium mentioned above is avoided.

The specifics of the example are as follows. Let $x_d(n)$ be a stationary Gaussian sequence with autocorrelation $R_{x_d x_d}(p) = (\sigma^2\alpha^{|p|}$ where σ^2 is the variance of the sequence and $|\alpha| < 1$. Let the ΔΣ modulator have $\gamma = \infty$ and $\Delta = 1$, and let $H(z)$ have the form $H(z) = F^2(z)$, where

$$F(z) = \frac{1}{M}\sum_{n=0}^{M-1} z^{-1}.$$

It is necessary to have $\gamma = \infty$ so that the overload condition is avoided. If γ were finite, the ΔΣ modulator would be sure to overload sooner or later because Gaussian distributions do not have finite support. However, the example can be applied as a good approximation when γ is finite provided $\gamma \gg \sigma^2$ because, in such cases, the overload condition is rare.

Using the easily derived fact that each partial sum S_N of N consecutive samples of the desired input sequence has a Gaussian distribution with variance

$$\sigma_N^2 = \sigma^2\frac{N(\alpha - \alpha^{-1}) - 2\alpha^N + 2}{2 - \alpha - \alpha^{-1}},$$

it is straightforward to evaluate (10) using standard techniques (see, e.g., [24]). Fig. 4 shows the autocorrelation so calculated for the case of $M = 16$, $\sigma^2 = 0.05$, and $\alpha = -0.8$, along with the autocorrelation as found by computer simulation. As is evident from the figure, the theoretical and simulated autocorrelations are in close agreement.

As shown in the previous section, if the statistics of the desired input sequence are only partially known, it is often possible to obtain approximate results. We illustrate this by applying the approximation to the previous example.

For arbitrary stochastic desired input sequences, in order to calculate $R_{\epsilon\epsilon}(n,p)$ using Theorem 4, it is necessary to know all $|p|$ and lower order statistics of the desired input sequence. For example, if we only know the fifth and lower order statistics, we can only calculate $R_{\epsilon\epsilon}(n,p)$ when $|p| \leq 5$. In this case, to apply the approximation of the previous section, we would set $R_{\epsilon\epsilon}(n,p) = 0$ whenever $|p| > 5$. Although we know all the statistics of the Gaussian desired input sequence considered in the previous example, we can nevertheless apply the approximation and compare the approximate results to the exact result.

Fig. 5 shows the results of the approximation for the cases where $R_{\epsilon\epsilon}(n,p)$ is only calculated for $|p| \leq 5$ and for $|p| \leq 10$. The curves are labeled in terms of the order of statistics that would generally be required to obtain the corresponding approximations if the input sequence were not Gaussian. The approximation is quite good in both cases.

IV. Conclusion

We have presented a unified approach to analyzing the granular quantization error of the first-order ΔΣ modulator. The approach handles many of the previously analyzed input sequences in addition to a large class of new input sequences. By averaging over the arbitrarily small

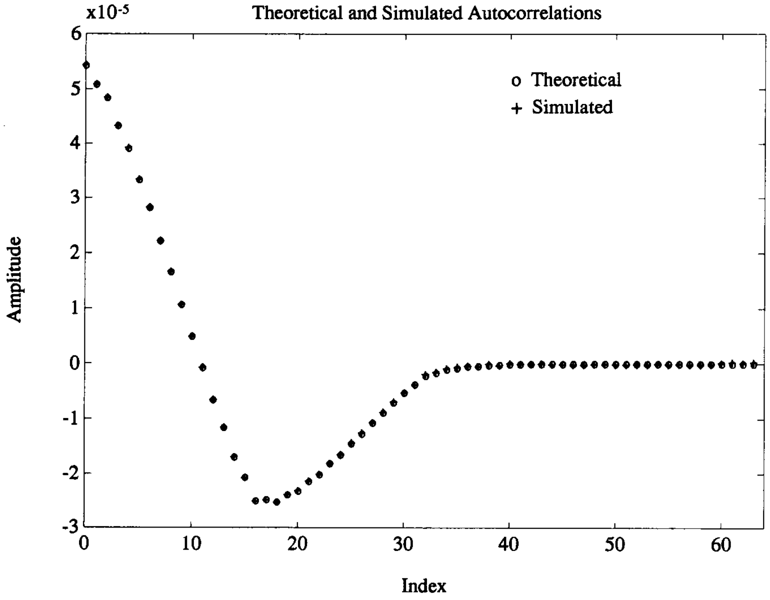

Fig. 4. The autocorrelation function as predicted by theory and as obtained by simulation. In this example, $M = 16$, $\sigma^2 = 0.05$, and $\alpha = -0.8$. Each point of the simulation was generated by averaging two-million consecutive points of the form $e(n)e(n + p)$.

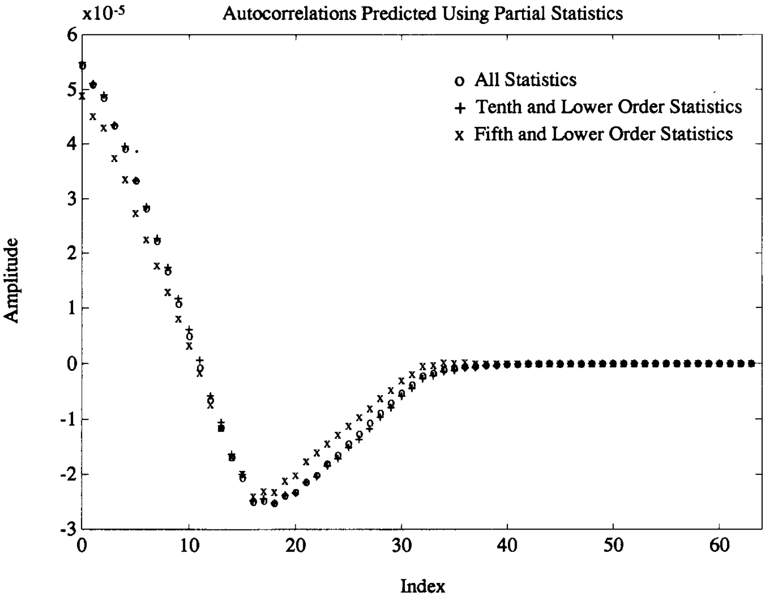

Fig. 5. The theoretically predicted autocorrelation function of Fig. 4 and approximations to it as predicted by the theory when only tenth and lower order and fifth and lower order statistics are known.

amount of circuit noise assumed to be present at the analog input to the $\Delta\Sigma$ modulator, we have derived a simple expression for the autocorrelation of the quantization error. Each term in the expression is formally equal to the quantization error of a nonoverloaded uniform quantizer operating upon a finite partial sum of consecutive input sequence samples. Hence, existing results concerning uniform quantizers are directly applicable in evaluating the autocorrelation expression for specific input sequences. In particular, if the statistics of the desired input sequence are known, then the autocorrelation can be calculated using standard techniques. If only partial statistics are known, an approximate result can be obtained. The theory is also applicable to deterministic input sequences, and has been applied to obtain a new closed-form result for sinusoidal input sequences. We have presented ergodic results which assert that under mild conditions, the autocorrelation equals the time-average autocorrelation in probability. We have applied the theory to various input sequences, some of which have been previously considered and some of which are new. Simulation results have been presented that closely support the theory.

Appendix
Supporting Lemmas

Lemma A1: Let α be a random variable that is uniformly distributed on [0, 1). Then, for any $x, y \in \boldsymbol{R}$.

$$E(\langle \alpha + x \rangle \langle \alpha + y \rangle) = \frac{1}{3} + \frac{1}{2}\left(\langle x - y \rangle^2 - \langle x - y \rangle\right).$$

Proof: We will prove the lemma in two steps. In the first step, we derive the relation

$$E(\langle \alpha + x \rangle \langle \alpha + y \rangle) = \frac{1}{3} + \frac{1}{2}(\langle x \rangle - \langle y \rangle)^2 - \frac{1}{2}|\langle x \rangle - \langle y \rangle|. \tag{29}$$

In the second step, we show the surprising result that given $u, v \in \boldsymbol{R}$,

$$(\langle u + v \rangle - \langle u \rangle)^2 - |\langle u + v \rangle - \langle u \rangle| = \langle v \rangle^2 - \langle v \rangle. \tag{30}$$

The lemma follows by combining (29) and (30) with $u = y$ and $v = x - y$.

To prove (29), we proceed as follows. By the properties of the fractional part operator,

$$\begin{aligned} E(\langle \alpha + x \rangle \langle \alpha + y \rangle) &= \int_0^1 \langle \alpha + x \rangle \langle \alpha + y \rangle \, d\alpha \\ &= \int_0^1 \langle \alpha + \langle x \rangle \rangle \langle \alpha - \langle y \rangle \rangle \, d\alpha. \end{aligned}$$

Defining $u_1 = \min\{1 - \langle x \rangle, 1 - \langle y \rangle\}$ and $u_2 = \max\{1 - \langle x \rangle, 1 - \langle y \rangle\}$, we can write

$$\begin{aligned} &E(\langle \alpha + x \rangle \langle \alpha + y \rangle) \\ &\quad = \int_0^{u_1} (\alpha + \langle x \rangle)(\alpha + \langle y \rangle) \, d\alpha \\ &\qquad + \int_{u_1}^{u_2} (\alpha + \max\{\langle x \rangle, \langle y \rangle\} - 1) \\ &\qquad \cdot (\alpha + \min\{\langle x \rangle, \langle y \rangle\}) \, d\alpha \\ &\qquad + \int_{u_2}^{1} (\alpha + \langle x \rangle - 1)(\alpha + \langle y \rangle - 1) \, d\alpha. \end{aligned}$$

Expressing the integrands in terms of u_1 and u_2 gives

$$\begin{aligned} E(\langle \alpha + x \rangle \langle \alpha + y \rangle) &= \int_0^{u_1} (\alpha - u_1 + 1)(\alpha - u_2 + 1) \, d\alpha \\ &\quad + \int_{u_1}^{u_2} (\alpha - u_1)(\alpha - u_2 + 1) \, d\alpha \\ &\quad + \int_{u_2}^{1} (\alpha - u_1)(\alpha - u_2) \, d\alpha. \end{aligned}$$

Expanding the integrands and collecting terms gives

$$\begin{aligned} E(\langle \alpha + x \rangle \langle \alpha + y \rangle) &= \int_0^1 (\alpha^2 - \alpha u_1 - \alpha u_2 + u_1 u_2) \, d\alpha \\ &\quad + \int_0^{u_2} (\alpha - u_1) \, d\alpha \\ &\quad + \int_0^{u_1} (\alpha + 1 - u_2) \, d\alpha. \end{aligned}$$

Evaluating the integrals, collecting terms, and expanding u_1 and u_2 gives

$$\begin{aligned} E(\langle \alpha + x \rangle \langle \alpha + y \rangle) &= \frac{1}{3} + \frac{1}{2}u_1 - \frac{1}{2}u_2 + \frac{1}{2}(u_1 - u_2)^2 \\ &= \frac{1}{3} + \frac{1}{2}(\langle x \rangle - \langle y \rangle)^2 \\ &\quad - \frac{1}{2}\max\{\langle x \rangle, \langle y \rangle\} + \frac{1}{2}\min\{\langle x \rangle, \langle y \rangle\} \end{aligned}$$

from which (29) follows.

It remains to prove (30). Because $\langle u + v \rangle = \langle \langle u \rangle + v \rangle$, without loss of generality, we can assume $u \in [0, 1)$. For convenience, define

$$f(v, u) = (\langle u + v \rangle - \langle u \rangle)^2 - |\langle u + v \rangle - \langle u \rangle|.$$

Choose any $v \in \boldsymbol{R}$. Then, there exists some integer P such that $v \in [P, P + 1)$. Hence, we must have either $v + u \in [P, P + 1)$ or $v + u \in [P + 1, P + 2)$. In the first case,

$$\begin{aligned} f(v, u) &= (u + v - P - u)^2 - |u + v - P - u| \\ &= (v - P)^2 - (v - P) \\ &= \langle v \rangle^2 - \langle v \rangle. \end{aligned}$$

In the second case,

$$\begin{aligned} f(v, u) &= (u + v - (P + 1) - u)^2 - |u + v - (P + 1) - u| \\ &= v^2 - 2v(P + 1) + (P + 1)^2 - (P + 1) + v \\ &= (v - P)^2 - (v - P) \\ &= \langle v \rangle^2 - \langle v \rangle. \end{aligned}$$

Hence, $f(v, u) = \langle v \rangle^2 - \langle v \rangle$ for all $u, v \in \boldsymbol{R}$. ■

Lemma A2: For each $k = 1, 2, \cdots$, let $f_k\colon \boldsymbol{R}^{2k+2} \to \boldsymbol{R}$ be a measurable function that has absolute value less than β. Let $\{\eta_0, \cdots, \eta_k\}$ and $\{\mu_0, \cdots, \mu_k\}$ be two sequences of random variables where the η_n are independent of each other and independent of the μ_n, and let $X_k = f_k(\eta_0, \cdots, \eta_k, \mu_0, \cdots, \mu_k)$. Suppose that

$$E(X_k \mid \eta_0, \cdots, \eta_j, \mu_0, \mu_1, \cdots) \to 0 \tag{31}$$

in probability as $k - j \to \infty$ with $k > j \geq 0$. Then,

$$\frac{1}{N}\sum_{n=0}^{N-1} X_n \to 0$$

in probability as $N \to \infty$.

Proof: Define the random variable S_N as

$$S_N = \frac{1}{N}\sum_{n=0}^{N-1} X_n.$$

For each $\epsilon > 0$, Chebyshev's inequality [14] gives

$$\operatorname{Prob}\{|S_N| \geq \epsilon\} \leq \frac{E(S_N^2)}{\epsilon^2}. \tag{32}$$

By the linearity of the expectation operator,

$$E(S_N^2) = \frac{1}{N^2} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} E(X_j X_k). \tag{33}$$

By hypothesis,

$$|E(X_j X_k)| \leq \beta^2, \tag{34}$$

for all $j, k \geq 1$. However, to find a tighter upper bound on $E(S_N^2)$, it is necessary to consider $|E(X_j X_k)|$ when $|k - j|$ is large. Since the η_n are independent and X_j is independent of the variables $\{\eta_n : n > j\}$, for each $k > j$ we can write

$$E(X_j X_k) = E\left[X_j E(X_k \mid \eta_0, \cdots, \eta_j, \mu_0, \mu_1, \cdots)\right].$$

Because of (31),

$$X_j E(X_k \mid \eta_0, \cdots, \eta_j, \mu_0, \mu_1, \cdots) \to 0$$

in probability as $k - j \to \infty$ with $k > j \geq 0$.

By definition, $X_j E(X_k \mid \eta_0, \cdots, \eta_j, \mu_0, \mu_1, \cdots) < \beta^2$ with probability 1. Therefore, it follows from the Lebesgue dominated convergence theorem that $E(X_j X_k) \to 0$ as $k - j \to \infty$ with $k > j \geq 0$. In particular, this means that there exists a positive integer M such that

$$|E(X_j X_k)| < \frac{\epsilon^3}{2} \quad \text{whenever } |j - k| \geq M \quad \text{and} \quad j, k \geq 0. \tag{35}$$

Choose

$$N' = \max\left\{\left\lceil \frac{(2M + 1)\beta^2}{\epsilon^3/2} \right\rceil, M\right\}.$$

Dividing the terms on the right side of (33) into two groups corresponding to $|j - k| < M$ and $|j - k| \geq M$ and applying the upper bounds (34) and (35), respectively, with $N \geq N'$ gives $E(S_N^2) < \epsilon^3$. From (32), it follows that for each $N \geq N'$,

$$\operatorname{Prob}\{|S_N| \geq \epsilon\} \leq \epsilon$$

This implies that $S_N \to 0$ in probability as $N \to \infty$. ■

Acknowledgment

The author would like to thank Professor Howard G. Tucker for his suggestions leading to improvements of Lemma A2. The author is also grateful to the late Professor Edward C. Posner whose support, encouragement, and friendship made this work possible.

References

[1] H. Inose, Y. Yasuda, and J. Murakami, "A telemetering system code modulation—Δ-Σ modulation," *IRE Trans. Space Electron. Telemetry*, vol. SET-8, pp. 204–209, Sept. 1962.

[2] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[3] J. C. Candy, "A use of limit cycle oscillations to obtain robust analog to digital converters," *IEEE Trans. Commun.*, vol. COM-22, pp. 298–305, Mar. 1974.

[4] R. M. Gray, "Oversampled sigma–delta modulation," *IEEE Trans. Commun.*, vol. COM-35, pp. 481–489, May 1987.

[5] ____, "Spectral analysis of quantization noise in a single-loop sigma–delta modulator with dc input," *IEEE Trans. Commun.*, vol. COM-37, pp. 588–599, June 1989.

[6] R. M. Gray, W. Chou, and P. W. Wong, "Quantization noise in single-loop sigma–delta modulation with sinusoidal inputs," *IEEE Trans. Inform. Theory*, vol. 35, pp. 956–968, Sept. 1989.

[7] P. W. Wong and R. M. Gray, "Sigma-delta modulation with I.I.D. Gaussian inputs," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 784–798, July 1990.

[8] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1220–1244, Nov. 1990.

[9] E. N. Protonotarios, "Slope overload noise in differential pulse code modulation systems," *Bell Syst. Tech. J.*, vol. 46, no. 9, pp. 2119–2177, Nov. 1967.

[10] D. J. Goodman and L. J. Greenstein, "Quantizing noise of ΔM/PCM encoders," *Bell Syst. Tech. J.*, vol. 52, pp. 183–204, Feb. 1973.

[11] J. C. Candy, "A use of double integration in sigma–delta modulation," *IEEE Trans. Commun.*, vol. COM-33, pp. 249–258, Mar. 1985.

[12] S. H. Ardalan and J. J. Paulos, "An analysis of nonlinear behavior in delta–sigma modulators," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 593–603, June 1987.

[13] W. Chou and R. M. Gray, "Dithering and its effects on sigma–delta and multistage sigma–delta modulation," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 500–513, May 1991.

[14] P. Billingsley, *Probability and Measure*. New York: Wiley, 1986.

[15] A. M. Laglom, *An Introduction to The Theory of Stationary Random Functions*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[16] K. Petersen, *Ergodic Theory*. Cambridge: Cambridge Univ. Press, 1983.

[17] J. C. Kieffer, "Analysis of dc input response for a class of one-bit feedback encoders," *IEEE Trans. Commun.*, vol. COM-38, pp. 337–340, Mar. 1990.

[18] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. New York: Wiley, 1974.

[19] D. F. Delchamps, "Exact asymptotic statistics for sigma-delta quantization noise," in *Proc. 28th Annu. Allerton Conf. Commun., Contr., Computing*, Oct. 1990.

[20] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[21] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[22] B. Widrow, "A study of rough amplitude quantization by means of Nyquist sampling theory," *IRE Trans. Circuit Theory*, vol. CT-3, pp. 266–276, 1956.

[23] ____, "Statistical analysis of amplitude quantized sampled data systems," *Trans. Amer. Inst. Elect. Eng., Pt II: Appl. Ind.*, vol. 79, pp. 555–568, 1960.

[24] A. B. Sripad and D. L. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 442–448, Oct. 1977.

[25] A. G. Clavier, P. F. Panter, and D. D. Grieg, "Distortion in a pulse count modulation system," *AIEE Trans.*, vol. 66, pp. 989–1005, 1947.

[26] L. Schuchman, "Dither signals and their effects on quantization noise," *IEEE Trans. Commun. Technol.*, vol. COM-12, pp. 162–165, Dec. 1964.

[27] B. P. Brandt and B. A. Wooley, "A 50-MHz multibit sigma-delta modulator for 12-b 2-MHz A/D conversion," *IEEE J. Solid-State Circuits*, vol. 26, no. 12, Dec. 1991.