Rigorous analysis of delta-sigma modulators for fractional-N PLL frequency synthesis.

Mücahit Kozak\textsuperscript{1}

Izzet Kale\textsuperscript{2}

\textsuperscript{1} Department of Electrical and Computer Engineering, University of Rochester, USA

\textsuperscript{2} School of Informatics, University of Westminster
Rigorous Analysis of Delta–Sigma Modulators for Fractional-\(N\) PLL Frequency Synthesis

Mücahit Kozak, Member, IEEE, and İzzet Kale, Member, IEEE

Abstract—In this paper, rigorous analyses are presented for higher order multistage noise shaping (MASH) Delta–Sigma (\(\Delta \Sigma\)) modulators, which are built out of cascaded first-order stages, with rational dc inputs and nonzero initial conditions. Asymptotic statistics such as the mean, average power, and autocorrelation of the binary quantizer error are formulated using a nonlinear difference equation approach. An important topic of interest considered here is the fractional-\(N\) phase-locked-loop frequency synthesis applications, where the input to the modulator has to be a rational constant. It has been mathematically shown that, regardless of the initial conditions, first-order and second-order MASH \(\Delta \Sigma\) modulators with rational dc inputs cannot sufficiently randomize the quantization error samples, and, therefore, are not suitable for fractional-\(N\) synthesis applications. An irrational initial condition imposed on the first accumulator of a third or higher order MASH modulator, on the other hand, annihilates the tones throughout the whole output spectrum, and provides a very smooth noise shaping. Simulation results are provided to support the theoretically derived results. Implementation issues of the irrational initial condition in the digital domain are also discussed and investigated together with the effect of finite accumulator size on the noise-shaping quality factor.

Index Terms—DC input, delta–sigma (\(\Delta \Sigma\)) modulators, fractional-\(N\) phase-locked loop (PLL), frequency synthesis, rigorous analysis.

I. INTRODUCTION

DELTA–SIGMA (\(\Delta \Sigma\)) modulation is an attractive means for the implementation of high-resolution A/D and D/A converters [1]–[3]. Although they constitute simpler circuitry, which is less sensitive to circuit imperfections such as component variations and mismatches, the exact analysis of \(\Delta \Sigma\) modulators is not a straightforward task, because of the severe nonlinearity introduced by the single-bit quantizer [4]. A commonly employed approach that attempts to overcome this difficulty is the linear approximation method, which involves modeling the quantizer error as an additive white noise source. Despite the fact that the linearized model can provide reasonable results with respect to signal-to-noise ratio (SNR) performance, it completely fails to determine the exact spectral shaping of the modulator, which is highly likely to be discrete and colored.

A more advanced linearized model for analyzing \(\Delta \Sigma\) modulators is the so-called describing function method, in which the quantizer is modeled as a linear gain (chosen in a minimum mean-square-error sense) followed by an input dependent additive white noise source [5]. This improved model gives crude explanations on the input dependent stability and large amplitude limit cycle behavior of the modulators. However, it is still an approximation in the sense that, except the fundamental component (assuming a sinusoidal input), all other harmonics generated by the quantizer are perfectly removed by linear filters within the modulator [4], [5]. The exact quantization noise shaping performance of \(\Delta \Sigma\) modulators is an important characteristic, as in certain applications the spectral spikes contained in the output spectrum may be strictly objectionable, for instance in the fractional-\(N\) phase-locked-loop (PLL) frequency synthesis applications [6], [7].

Fractional-\(N\) PLL frequency synthesizers utilize an all-digital \(\Delta \Sigma\) modulator to generate accurately defined frequencies and desired channel selection in wireless transceivers [6]–[9]. Fig. 1 illustrates such a PLL synthesizer where the output of the \(\Delta \Sigma\) modulator is used to modulate the instantaneous division ratio of a multi-modulus frequency divider. The division ratio alterations take place very rapidly, and in a random/pseudo-random fashion so that the phase noise and spurious contents are pushed to higher offset frequencies, where the existing PLL loop filter can easily cut them off [10]. Consequently, the phase error (analogous to the quantizer error in data conversion applications) spectrum is shaped by the high-pass noise transfer function of the modulator, so that the phase noise in the vicinity of the carrier (analogous to the baseband frequency in data conversion applications) is small. This technique permits narrow step sizes compared to the reference frequency and fast settling times, while improving the phase noise performance of the PLL frequency synthesizers [11]–[13].

Fig. 1. Fractional-\(N\) PLL synthesizer using \(\Delta \Sigma\) modulation.
The difficulty involved in the realization of fractional-$N$ frequency synthesizers is that the ΔΣ modulator is driven by a worst case input, e.g. a constant input that is the desired fractionality. The synthesized output frequency at the PLL, $f_{\text{out}}$, is given by $f_{\text{out}} = f_{\text{ref}} \left( N_{\text{int}} + X \right)$ [10], where $N_{\text{int}}$ is the integer part of the division ratio of the multimodulus divider, and $X \in [0, 1)$, the input to the modulator, is a constant number indicating the desired fractionality. It is a well-known fact that constant inputs are the worst case inputs to the oversampling ΔΣ modulators, as the successful operation of these modulators relies on the assumption that the successive input samples are uncorrelated (i.e., busy inputs), and under the static input condition, this assumption is definitely violated [14]. These complications are worsened in the context of fractional-$N$ synthesis, because the modulator is implemented in the digital domain, and due to the finite-word-length of the arithmetic, the input as well as all the internal states are represented by rational numbers. It is commonly observed that simple rational dc inputs to the ΔΣ modulator produce purely limit cycle oscillations and tones, and hence, no noise shaping is present at all [15], [16].

Another crucial point, is that, unlike data conversion applications, in fractional-$N$ synthesis, it is very dangerous to generate out-of-band1 limit cycles and tones as these components are very likely to fold back to the baseband frequencies, because of the nonlinearity in the implementation of the phase detector [17]. Consequently, suitable ΔΣ modulators for fractional-$N$ synthesis should produce tone-free output spectra throughout the whole of the spectrum, not only in the baseband. This fact draws attention to the importance of obtaining a very smooth noise shaping (i.e., a noise shaping without discrete spectral spikes throughout the whole of the spectrum) from a ΔΣ modulator when deployed in a fractional-$N$ synthesis application. This requirement is the main motivation for the exact analysis of ΔΣ modulators with rational dc inputs performed in this paper.

Exact analysis of the first-order ΔΣ modulator with dc inputs was performed by Gray in [18], where it was shown that the output spectrum consists of discrete spectral spikes whose locations and amplitudes are input value dependent. The technique was then extended for the higher order multistage noise shaping (MASH) [19] ΔΣ modulators in [20] and [21], using concepts from ergodic theory. All of these exact analysis studies were based on obtaining a nonlinear difference equation from which the asymptotic autocorrelation function of the binary quantizer error was determined. The results in [20] and [21] have mathematically proved that in a higher order MASH ΔΣ modulator, the white noise assumption for the binary quantizer error from the last stage is indeed correct, when irrational dc inputs are employed. The same approach used for the exact analysis of MASH modulators is also used in [22] and [23] to analyze higher order single-stage modulators with irrational dc inputs.

All of the above-cited studies on the exact analysis of modulators assume that the initial conditions in the accumulators (for a digital modulator) or integrators (for an analog modulator) are zero. In fact, for the case of irrational dc inputs the asymptotic behavior of the quantizer error is not affected by initial conditions [4], [18], [20], [21]. However, when the input is a rational constant (as it is the case in fractional-$N$ synthesis), such initial conditions may be very effective in randomizing the binary quantizer error. It is, therefore, the aim of this paper to provide exact analysis results for the higher order MASH ΔΣ modulators with rational dc inputs and nonzero initial conditions. In our analysis, we follow similar steps to those performed in [18], [20], and [21], and make use of theorems from ergodic theory as well as the characteristic function method combined with the solution of the nonlinear recursive difference equation. Our results show that an irrational initial condition imposed on the first accumulator guarantees a tone free output spectrum for a third-order or higher order MASH ΔΣ modulator when driven by rational dc inputs.

The rest of this paper is organized as follows. Section II introduces the recursive nonlinear difference equations for the first-order ΔΣ and higher order MASH ΔΣ modulators. Section III presents theorems for the mean, average power, and autocorrelation of the binary quantizer error sequence in the first-order, second-order MASH, and higher order MASH ΔΣ modulators. Simulation results, which justify our theoretical developments, are presented in Section IV. Implementation issues concerning the realization of irrational initial condition in the digital domain are discussed in Section V. Conclusions are given in Section VI, followed by appendices that presents a synopsis of the proofs of the theorems.

II. NONLINEAR DIFFERENCE EQUATIONS

A. First-Order ΔΣ Modulator

Consider a first-order modulator as shown in Fig. 2, which is described by the following:

$$v[n] = \begin{cases} v_0, & n = 0 \\ u[n - 1] + v[n - 1], & n = 1, 2, \ldots \end{cases} \quad (2.1)$$

$$y[n] = Q(v[n]) = \begin{cases} +b, & v[n] \geq 0 \\ -b, & v[n] < 0 \end{cases} \quad (2.2)$$

$$u[n] = x[n] - y[n], \quad n = 0, 1, 2, \ldots \quad (2.3)$$

where $v_0$ is the initial condition on the accumulator, $+b$ denotes the output levels of the single-bit (or binary) quantizer, and $Q(\cdot)$ is the quantization function. Let us define the binary quantization error sequence such that $e[n] = y[n] - v[n]$. A nonlinear difference equation, which relates the binary quantizer error sequence to the input sequence, is stated in the following theorem.
Theorem 1: For the first-order $\Delta\Sigma$ modulator shown in Fig. 2 with an arbitrary input $x[n] \in [-b, b]$ for all $n$, and an arbitrary initial condition in the accumulator $v_0 \in [-2b, 2b]$, the normalized binary quantizer error sequence is given by

$$\varepsilon[n] = \frac{e[n]}{2b} = \frac{1}{2} \left\langle \frac{v_0}{2b} + \sum_{k=0}^{n-1} \left( \frac{1}{2} + \frac{x[k]}{2b} \right) \right\rangle, \quad n = 0, 1, 2, \ldots$$

(2.4)

where $\pm b$ is the output levels of the single-bit quantizer, and $e[n] = y[n] - v[n]$ for all $n$.

The angle brackets in (2.4) denote the fractional part operator, which is defined as $\langle x \rangle = x \mod 1$. Similar derivations are given in [24], and therefore, details of the proof for this theorem are omitted here. For a dc input $x[n] = X$ for all $n$ (i.e., the input is held at a constant for all time), the normalized binary quantizer error sequence can be written as

$$\varepsilon[n] = \frac{1}{2} - \left\langle \frac{v_0}{2b} + n \left( \frac{1}{2} + \frac{X}{2b} \right) \right\rangle, \quad n = 0, 1, 2, \ldots$$

(2.5)

As a matter convenience, we follow the same notation as used in [18], and define

$$\beta \equiv \left( \frac{1}{2} + \frac{X}{2b} \right),$$

(2.6)

$$\zeta[n] \equiv \left\langle \frac{v_0}{2b} + n\beta \right\rangle.$$

(2.7)

Hence, $\varepsilon[n] = 1/2 - \zeta[n]$.

B. Higher Order MASH $\Delta\Sigma$ Modulator

The discrete-time model for an $L$th-order MASH $\Delta\Sigma$ modulator, which is formed out of first-order stages, is shown in Fig. 3. The input to the modulator goes into the first stage, and the negative of the binary quantizer error sequence from the first stage is fed forward into the second stage as its input, and versa. The outputs from all stages are processed by an error cancellation network, which removes the binary quantizer error components from all the stages except the last one [1], [2], and [19]. The difference equations describing the $L$th-order MASH $\Delta\Sigma$ modulator are given as follows:

$$v_i[n] = \begin{cases} v_i, & n = 0 \\ u_i[n-1] + v_i[n-1], & n = 1, 2, \ldots \end{cases} \quad i = 1, 2, \ldots, L$$

(2.8)

$$y_i[n] = \left\langle v_i[n] \right\rangle = \begin{cases} +b, & v_i[n] \geq 0 \\ -b, & v_i[n] < 0 \end{cases} \quad \forall n, \ i = 1, 2, \ldots, L$$

(2.9)

$$e_i[n] = y_i[n] - v_i[n] \quad \forall n, \ i = 1, 2, \ldots, L$$

(2.10)

$$w_i[n] = \begin{cases} x[n], & i = 1 \\ -e_{i-1}[n], & i = 2, 3, \ldots, L \end{cases} \quad \forall n$$

(2.11)

$$u_i[n] = w_i[n] - y_i[n]; \quad n = 1, 2, \ldots \quad i = 1, 2, \ldots, L$$

(2.12)

where $v_{i0}$ and $y_i$ ($i = 1, 2, \ldots, L$) are the initial condition on the $i$th accumulator, and the output from the $i$th stage, respectively. As is well known in the art, for an $L$th-order MASH modulator the only noise term that appears in the output is the binary quantizer error from the last stage [1], [19]. The reason for this is that, as mentioned before, the quantizer error from all stages except the last one is cancelled in the error cancellation network. Hence, all that is needed for an exact analysis of a higher order MASH $\Delta\Sigma$ modulator is the behavior of the binary quantizer error from the last stage. Since every stage of the MASH $\Delta\Sigma$ modulator is built out of the first-order ones, we can use Theorem 1 as long as the initial condition on each of the accumulators is bounded within the range $[-2b, 2b]$. Mathematically,
if \( v_{i,0} \in [-2b_i + 2b_i]; \ i = 1, 2, \ldots, L \), then, the normalized binary quantizer error from each stage is given by

\[
\varepsilon_i[n] = \frac{\varepsilon_i[n]}{2b_i} = \begin{cases} 
\frac{1}{2} - \left( \frac{\varepsilon_{i,1} + \sum_{k=0}^{n-1} \left( \frac{1}{2} + \frac{x[k]}{2b_i} \right) }{2b} \right) & \text{if } i = 1 \\
\frac{1}{2} - \left( \frac{\varepsilon_{i,0} + \sum_{k=0}^{n-1} \left( \frac{1}{2} - \frac{\varepsilon[k - 1]}{2b} \right) }{2b} \right) & \text{if } i = 2, 3, \ldots, L,
\end{cases} \quad \forall n.
\]  

(2.13)

Since \( x[n] \in [-b_i + b_i] \) for all \( n \) and \( v_{i,0} \in [-2b_i + 2b_i] \), (2.13) is valid for \( i = 1 \) due to Theorem 1, and furthermore it implies that \( \varepsilon_i[n] \in [-b_i + b_i] \) for all \( n \). With this conclusion, if \( v_{2,0} \in [-2b_2 + 2b_2] \), then, (2.13) is also valid for \( i = 2 \). Similar reasoning can be applied by induction to show that (2.13) is valid for all stages [21]. The following theorem gives a general description for the normalized binary quantizer error sequence from the last stage.

**Theorem 2:** For an \( L \)-th order MASH modulator formed out of first-order stages with any arbitrary input \( x[n] \in [-b_i + b_i] \) for all \( n \), and arbitrary initial conditions on the accumulators such that \( v_{i,0} \in [-2b_i + 2b_i] \), \( i = 1, 2, \ldots, L \), the normalized binary quantizer error from the last stage is given by

\[
\varepsilon_L[n] = \frac{1}{2} - \left( \frac{v_{L,0}}{2b} + \sum_{k=0}^{n-1} \left( \frac{v_{L,1} - 1,0}{2b} + \cdots + \frac{v_{L,k} - 1,0}{2b} + \sum_{k_{1}=0}^{k_{-1}} \beta_k \right) \right),
\]  

(2.14)

The proof of this theorem is accomplished in Appendix I. Note that (2.14) applies to any arbitrary input and initial conditions. In this paper, we consider only rational dc inputs in the analysis, as it is the case in fractional- \( N \) PLL frequency synthesis applications. Once again, let \( x[n] = X \in [-b_i + b_i] \) for all \( n \), and using the definition for \( \beta \) stated in (2.6) we write

\[
\varepsilon_L[n] = \frac{1}{2} - \left( \frac{v_{L,0}}{2b} + \sum_{k=0}^{n-1} \left( \frac{v_{L,1} - 1,0}{2b} + \cdots + \frac{v_{L,k} - 1,0}{2b} + \sum_{k_{1}=0}^{k_{-1}} \beta_k \right) \right),
\]  

(2.15)

For simplicity purposes, we define \( \zeta_L[n] \) such that \( \varepsilon_L[n] = 1/2 - \zeta_L[n] \), where

\[
\zeta_L[n] = \left( \frac{v_{L,0}}{2b} + \sum_{k=0}^{n-1} \left( \frac{v_{L,1} - 1,0}{2b} + \cdots + \frac{v_{L,k} - 1,0}{2b} + \sum_{k_{1}=0}^{k_{-1}} \beta_k \right) \right).
\]  

(2.16)

The nonlinear difference equations in (2.5) and (2.15) indicate that the normalized binary quantizer error sequence has a bias of 1/2 from which the fractional part of a summation, which is composed of the \( L \)-th order and \( (L - \delta) \)-th order discrete-time integrated versions of the dc input and the initial condition on the \( i \)-th accumulator, respectively, is subtracted. These equations will later be used in Appendix II to derive the statistics of the binary quantizer error sequence.

### III. STATISTICS OF QUANTIZER ERROR SEQUENCE

#### A. First-Order \( \Delta \Sigma \) Modulator

This section presents a theorem for the mean value, average power, and autocorrelation function of the binary quantizer error sequence in the first-order \( \Delta \Sigma \) modulator with a rational constant input and arbitrary initial condition on the accumulator. Before presenting the theorem, it is useful to explicitly state the assumptions made throughout this section.

- The input to the modulator is a constant, \( x[n] = X \in [-b_i + b_i] \) for all \( n \). This assumption is required, because in fractional- \( N \) PLL frequency synthesis applications the input to the modulator is a constant.
- The normalized dc input to the modulator is a rational number (i.e., \( X/(2b) \) is a rational number). Note that this assumption implies that \( \beta \) is also a rational number (see (2.6)), and therefore, can be given by \( \beta = K/N \) where \( K \) and \( N \) are relatively prime integer numbers. This assumption is required, because in fractional- \( N \) PLL frequency synthesis applications, the \( \Delta \Sigma \) modulator is implemented in the digital domain where the dc input value can only be represented as a rational fraction of the quantization step size.
- The initial condition on the accumulator is arbitrary (it can be either rational or irrational). Later, we shall see in the case of higher order MASH \( \Delta \Sigma \) modulators that it is required that the initial condition on the first accumulator be an irrational fraction of the quantization step size. For the first-order \( \Delta \Sigma \) modulator case, such a restrictive assumption is not required, and the results of the following theorem are valid for both rational and irrational initial conditions.

**Theorem 3:** Consider the first-order \( \Delta \Sigma \) modulator as shown in Fig. 2 employing single-bit quantization with output levels \( \pm b \), and having an arbitrary initial condition on the accumulator such that \( v_0 \in [-2b_i + 2b_i] \). Assume that the input to the modulator is a constant such that \( x[n] = X \in [-b_i + b_i] \) for all \( n \), where \( X/(2b) \) is a rational number. The mean value, average power, and autocorrelation of the normalized binary quantizer error sequence are, respectively, given by

\[
M(\varepsilon) = \frac{1}{2N} + \left( \frac{1 - v_0}{2b} \right) - \frac{1}{N} \eta_0
\]  

(3.1)

\[
M(\varepsilon^2) = \frac{1}{12} + \frac{1}{6N^2} + \frac{v_0}{2b} \left( \frac{2\eta_0}{N} + \frac{v_0}{2b} - 2 - \frac{1}{N} \right)
\]  

(3.2)

and

\[
r_\varepsilon(\ell) = \frac{1}{12} + \frac{1}{6N^2} + \frac{v_0}{2b} \left( \frac{2\eta_0}{N} + \frac{v_0}{2b} - 2 - \frac{1}{N} \right)
\]  

(3.3)
where
\[
\beta = (1/2 + X/(2b)), \quad \text{and can be written as } \beta = K/N,
\]
where \(K\) and \(N\) are relatively prime integer numbers;
\[
\eta_0 = \lceil (1 - v_0/2b)N \rceil, \quad \text{where the operator } \lceil \cdot \rceil \text{ denotes the smallest integer greater than or equal to } x.
\]

The proof of the theorem is given in Appendix II. Note that Theorem 3 states that the mean value of the normalized binary quantizer error sequence is not zero. It also indicates that the average power is not 1/12. Furthermore, the autocorrelation is nonzero for nonzero lag indexes, implying strong periodic components in the output spectrum. By virtue of Theorem 3, in the first-order \(\Delta \Sigma\) modulator, it is not possible to obtain a quantizer error sequence that is uniformly distributed.

As mentioned before, the first-order \(\Delta \Sigma\) modulator was analyzed before for the rational dc input case with a zero initial condition on the accumulator [18]. For zero initial condition (i.e., \(v_0 = 0\) and \(\eta_0 = N\)), the formulas given in (3.1)–(3.3) reduce to
\[
M(\varepsilon) = 1/2N, \quad M(\varepsilon^2) = 1/12 + 1/6N^2, \quad \text{and } r_\varepsilon(l) = 1/12 + 1/6N^2(1 - l/\beta)/2(1 + l/\beta),
\]
which are identical to the equations derived in [18, eqs. 3.11a, 3.11b, and 3.14], respectively. Theorem 3 is an extension of the previous results presented in [18] with an arbitrary initial condition on the accumulator.

**B. Higher Order MASH \(\Delta \Sigma\) Modulators**

As in the previous section, before presenting the theorems for higher order MASH \(\Delta \Sigma\) modulators, we explicitly state the assumptions made throughout this subsection.

- The input to the modulator is a constant, \(x[n] = X \in [-b,+b]\) for all \(n\). This assumption is required, because in Fractional-\(N\) PLL frequency synthesis applications the input to the modulator is a constant.
- The normalized dc input to the modulator is a rational number (i.e., \(X/(2b)\) is a rational number), implying that \(\beta\) is a rational number (see (2.6)), and thus can be given by
  \[
  \beta = K/N, \quad \text{where } K \text{ and } N \text{ are relatively prime integer numbers.}
  \]
  This assumption is essential due to the use of an all-digital \(\Delta \Sigma\) modulator in Fractional-\(N\) PLL frequency synthesis applications.
- The initial condition on the first accumulator is an irrational fraction of the quantization step size arbitrary (i.e., \(v_{1,0}/(2b)\) is an irrational number). This assumption is required, because in the derivation of the statistics of the binary quantizer error sequence we shall make use of a result from ergodic theory of mathematics that requires some form of irrationality. Unfortunately, a general analysis as for the first-order \(\Delta \Sigma\) modulator case does not easily extend to the higher order MASH \(\Delta \Sigma\) modulators. Therefore, the results of the following theorems are only valid for the irrational initial condition case.

**Theorem 4:** Consider the second-order MASH \(\Delta \Sigma\) modulator employing single-bit quantization with output levels \(\pm b\) and arbitrary initial conditions on the accumulators such that \(v_{i,0} \in [-2b,+2b]\) (i.e., \(v_{1,0}/(2b)\) is an irrational number). The mean value, average power, and autocorrelation of the normalized binary quantizer error sequence from the second stage are, respectively, given by
\[
M(\varepsilon_2) = 0
\]
\[
M(\varepsilon^2_2) = 1/12
\]
and
\[
r_{\varepsilon_2}(l) = \begin{cases} 
\frac{1}{12}, & l = 0 \\
0, & l \neq 0.
\end{cases}
\]

The proof of this theorem is presented in Appendix II. Expressions (3.7)–(3.9) show that in a higher order MASH \(\Delta \Sigma\) modulator (where \(L \geq 3\)), the binary quantizer error sequence from the last stage has a mean value of zero, an average power of 1/12, and an autocorrelation function which is zero for all nonzero lag indexes. Thus, the quantizer error sequence is uniformly distributed and statistically independent of the input, which is completely consistent with the white noise assumption.
IV. SIMULATION RESULTS

In order to verify our theoretical results, we have undertaken extensive simulations on \( \Delta \Sigma \) modulators employing single-bit quantization with output levels \( \pm 1 \). A time-domain difference-equation simulation is run for \( 2^{18} \) number of output samples from which the sample mean, average power, and autocorrelation function of the quantizer error sequence are obtained. Figs. 4–7 show the simulated autocorrelation function of the binary quantizer error sequence along with its analytically derived expression, for various \( \Delta \Sigma \) modulators, dc input values, and initial conditions. The simulated mean value and the average power are also listed on top of these plots.

As seen from these plots, the theory and simulation are in exact agreement justifying the derived formulas. Fig. 4(a) and (b) illustrate the autocorrelation function of the binary quantizer error sequence in the first-order \( \Delta \Sigma \) modulator with a positive and negative initial condition, respectively, where \( X = 0.1 \).

Figs. 5 and 6 show the autocorrelation function in the second-order 1–1 MASH \( \Delta \Sigma \) modulator for various dc input value and initial condition combinations. These figures justify the validity of the formulas in (3.4)–(3.6). Fig. 7(a) and (b) depict that the binary quantizer error sequence from the last stage of the 1–1–1 and 1–1–1–1 MASH \( \Delta \Sigma \) modulator, respectively, has identical first- and second-order statistics as a uniformly distributed white noise, as predicted by (3.7)–(3.9).

Note that only \(-100\) and \(+100\) lag indexes are shown in Fig. 4–7 for clarity. Fig. 8(a) demonstrates the difference between the theoretically derived and simulated autocorrelation function of the binary quantizer error in the first-order modulator for the case shown in Fig. 4(a). This very small (of the order of \( 1 e^{-7} \)) discrepancy between the simulated and the analytically calculated autocorrelation may be attributable to possible numerical inaccuracies and/or drifts in the simulation software environment. The difference between the analytically calculated and simulated autocorrelation function in the second-order 1–1 MASH modulator for the case of Fig. 5(a) is shown in Fig. 8(b). This difference (of the order of \( 1 e^{-5} \)) is mainly a result of the finite number of samples that can be obtained from a digital simulation of \( \Delta \Sigma \) modulators.
V. DIGITAL REALIZATION OF IRRATIONAL INITIAL CONDITION

In Section III, it is shown that the third-order or higher order MASH ΔΣ modulators provide tone free output spectrum, when an irrational initial condition is used in the first accumulator. However, as mentioned before, in fractional-$N$ frequency synthesis applications, the modulator is implemented in the digital domain, and one might rightly consider how the irrational conditional could be realized in the digital domain. The answer to this question relies on the fact that a rational number, which is represented by a ratio of two relatively prime integer numbers, will tend to an irrational number when the number in the denominator is made very large [18]. This fact suggests that, provided the dynamic range of the accumulators is sufficiently large, an odd number initial condition can still act as an irrational number in fixed-point implementations.

Note that here the reason for choosing an odd number is to represent the initial condition by a ratio of two relatively prime integer numbers, with the largest possible denominator.

Another point, which should also be considered here, is that the formulas presented in Section III for the mean value of the binary quantizer error sequence is for infinite time (see Appendix II). This means that if the data were infinitely long, the mean value of the binary quantizer error would approach zero (i.e., the mean value of the output would approach the value of the constant input). However, the fractional-$N$ PLL frequency synthesizer should settle to the desired frequency within a specified finite time (which is the settling time of the synthesizer). This implies that the number of samples that can be taken from the output of the MASH ΔΣ modulator is large, but limited. Arbitrarily choosing a large odd number for the initial condition in the first accumulator, therefore, may result in relatively large error between the mean value of the output and the constant input. This yields a relatively large frequency error in the synthesized PLL output. In order to minimize these frequency deviations, it is desirable to utilize the minimum odd
number, which is “1,” in the first accumulator as the initial condition. This is accomplished by setting “1” LSB in the first accumulator each time the circuit is reset [9]. A digital implementation of the 1–1–1 MASH ΔΣ modulator with the “1” LSB initial condition in the first accumulator is depicted in Fig. 9 [25], [26]. The effect of “1” LSB irrational initial condition was previously observed and successfully utilized to achieve a low spurious performance in [9]. The findings and results reported in this paper have rigorously explained this complex behavior, which was empirically found after the painful experiences of the designers, that reported in the past [9] with some intuitive reasoning and without mathematical justification.

Fig. 10 presents output noise spectra (dc input removed) of the three, four and five-stage MASH ΔΣ modulators, where the dynamic range of accumulators was set to 20-bit, and “1” LSB irrational initial condition is used in the first accumulator. These spectra are obtained from the fixed-point simulations, where 16 units of 50% overlapping modified periodograms, each with a 32K length, are used to estimate the power spectral density. The results clearly indicate the effect of “1” LSB initial condition, as no tones are observed throughout the whole spectrum. Also observe that the noise power as well as its spectral distribution is uniform across dc input values, which indicates the linear operation of ΔΣ modulators, despite the severe nonlinearity introduced by single-bit quantization.
Fig. 9. Digital implementation of the 1–1–1 MASH ΔΣ modulator.

Fig. 10. Output noise spectra for MASH modulators from 20-bit fixed-point simulation. (a) Three-stage. (b) Four-stage. (c) Five-stage.
In order to investigate the effect of “1” LSB initial condition for various accumulation sizes, here a new measure called noise-shaping quality (NSQ) factor is introduced. The NSQ factor of an \(L\)th-order modulator is defined as follows:

\[
\text{NSQ} = 1 - \frac{\text{SAD}}{\text{TP}}
\]

(5.1)

where SAD stands for “the sum of absolute differences in power between the measured output quantization noise and the theoretical noise shaping performance of the modulator,” and TP is “the total power in the noise transfer function.” Mathematically, this is written as

\[
\text{NSQ} = 1 - \frac{\int_0^{f_s} |H_{E}(f)|^2 - P_{E}(f) |H_{E}(f)|^2 |df|}{\int_0^{f_s} |H_{E}(f)|^2 df}
\]

(5.2)

where \(f_s\) is the sampling frequency of the modulator, \(H_{E}(f) = (1 - e^{-j\pi f/f_s})^L\) is the noise-shaping transfer function, and \(P_{E}(f)\) is the power-spectral density of the quantizer error sequence. The NSQ factor gives an indication of how closely distributed the output quantization noise spectrum is with respect to the theoretical noise shaping performance of the modulator.

It is clear, however, from the definition stated in (5.2) that for a white noise quantizer error, the value of NSQ does not equal to unity (in fact it is much smaller than unity). This means that the maximum achievable NSQ factor for a real modulator is limited to that of the truly white noise. Therefore, before extensive fixed-point simulations are undertaken, it is required that the NSQ factor of a white noise is determined and all the raw NSQ measurements are normalized to this value. It can be shown that for an ideal white noise quantizer error the NSQ factor is \(\text{NSQ}_{\text{white}} = \sigma^2\), which is the variance of the white noise [27].

The effect of the finite accumulation size on the NSQ factor is illustrated in Fig. 11, where the solid, dotted, and dashed lines correspond to the third-, fourth-, and fifth-order MASH \(\Delta \Sigma\) modulator. Also note that in each of these fixed-point simulations, “1” LSB of the first accumulator was set at the beginning of the start-up. Panel (a) of this figure was obtained by averaging the NSQ factors over the worst ten inputs (i.e., \(X\) was traced from 0 to 1 with 0.1 increments), while the NSQ simulations in panel (b) were averaged over all the specified inputs in a Fractional–\(N\) synthesis application.\(^3\)

The relatively low value of the NSQ at a low number of bits is a result of the huge truncation error in the internal states. Four- and five-stage modulators, however, demonstrate almost perfect operation (e.g., they produce quantizer error which is really close to white noise), when the number of bits in the implementation is higher than 16. As shown in Fig. 11(a), in the three-stage modulator, as the accumulator dynamic range is increased the NSQ factor begins to deteriorate. The reason for this behavior is that when the number of bits is large the third accumulator does not experience sufficient activity near the LSB positions, resulting in the tonal/structured output spectra. We conclude that fourth-order and higher order modulators, on the other hand, are more robust with respect to the digital implementation of the “1” LSB irrational initial condition.

VI. CONCLUSION

In this paper, rigorous analyses of \(\Delta \Sigma\) modulators with rational dc inputs and nonzero initial conditions are performed. Our theoretical results were verified through extensive simulations. It is concluded that for a first-order \(\Delta \Sigma\) modulator the binary quantizer error exhibits a purely discrete power spectrum, and consequently the white noise assumption is strictly violated. Neither the first nor the second moments comply with the white noise assumption. The second-order 1–1 MASH \(\Delta \Sigma\) modulator produces the same first order statistics as those of the white noise, when an irrational initial condition is imposed on the first accumulator, but it still fails to generate an uncorrelated quantizer error sequence. It has been proved that the quantizer error sequence from the second stage is colored, as its autocorrelation function is not an impulse. The most important conclusion that can be drawn from this paper is the following statement: “When

\(^3\)We assumed that \(f_{\text{ref}} = 20\) MHz and the desired frequency step was \(f_{\text{step}} = 200\) kHz, which is typical [10]. The dc input in this case, therefore, was traced from 0 to 1 with 0.01 increments.
the number of stages in a MASH ΔΣ modulator is greater than or equal to 3 \((L \geq 3)\), an irrational initial condition on the first accumulator produces a binary quantizer error sequence from the last stage that is a uniformly distributed identical distribution random process.” This is because in this case the sample mean, average power and autocorrelation are the same as those produced by a white noise. Additionally, the quantizer error sequence is uncorrelated with the dc input.

Digital implementation of the irrationality is also discussed in this paper. Specifically, a new measure metric, NSQ, is defined for assessing the modulator performance under nonideal conditions. It has been shown that, compared to the third-order MASH ΔΣ modulator, fourth-order or higher order MASH ΔΣ modulators are more robust in terms of the digital approximation of the “1” LSB irrational initial condition when fixed-point arithmetic is used.

It should also be noted that, with the mathematical tools in our hand, the analysis performed for the higher order MASH modulators were restricted only to irrational initial condition on the first accumulator. The reason for this is that results from ergodic theory (i.e., Weyl’s Theorem) require assumptions with respect to the irrationality of some of the input or system parameters. Unfortunately, the analysis technique performed for the first-order ΔΣ modulator does not easily apply to that of the higher order ones.

It is well known that ΔΣ modulators are nonlinear chaotic systems, and two infinitesimally small different initial conditions may result in completely divergent outputs. This paper, however, presents important results that rigorously prove the effectiveness of the irrational initial condition in terms of the tonality performance.

**Appendix I**

**Proof of Theorem 2**

The proof is by induction on \(L\). Assume that \(L = 1\); then, (2.14) coincides with (2.13). Now, assume that (2.14) holds for \(L\) and write \(\varepsilon_{L+1}[n]\) using (2.13)

\[
\varepsilon_{L+1}[n] = \frac{1}{2} - \left( \frac{v_{L+1,0}}{2b} + \sum_{k_{L+1}=0}^{n-1} \left( \frac{1}{2} - \frac{\varepsilon_{L}[k_{L+1}]}{2b} \right) \right).
\]

Inserting (2.14) into (7.1) we obtain the first equation shown at the bottom of the page. Simple algebraic manipulation on the the equation yields (7.2), shown at the bottom of the page, which complies with (2.14). The theorem is proved. \(\square\)

\[
\varepsilon_{L+1}[n] = \frac{1}{2} - \left( \frac{v_{L+1,0}}{2b} + \sum_{k_{L+1}=0}^{n-1} \left( \frac{1}{2} - \frac{\varepsilon_{L}[k_{L+1}]}{2b} \right) \right).
\]

**Appendix II**

**Proof of Main Results**

**A. Preliminaries**

Our primary goal is to evaluate the sample mean, average power, and autocorrelation of the binary quantizer error sequence. For a stationary ergodic random process \(\hat{x}[n] (n = 0, 1, \ldots)\), the sample mean, average power and auto-correlation are, respectively, defined by

\[
M(\hat{x}) = \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{n=0}^{N_t-1} \hat{x}[n] \tag{8.1}
\]

\[
M(\hat{x}^2) = \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{n=0}^{N_t-1} \hat{x}^2[n] \tag{8.2}
\]

and

\[
r_x(l) = \lim_{N_t \to \infty} \frac{1}{N_t} \sum_{n=0}^{N_t-1} \hat{x}[n] \hat{x}[n+l] \tag{8.3}
\]

if the limits exist. Let us define another ergodic random process such that \(\hat{y}[n] = 1/2 - \hat{x}[n]\). Observe that the statistics of \(\hat{y}[n]\) can be recovered from those of \(\hat{x}[n]\) as follows:

\[
M(\hat{y}) = \frac{1}{2} - M(\hat{x}) \tag{8.4}
\]

\[
M(\hat{y}^2) = \frac{1}{4} - M(\hat{x}) + M(\hat{x}^2) \tag{8.5}
\]

and

\[
r_y(l) = \frac{1}{4} - M(\hat{x}) + r_x(l). \tag{8.6}
\]

The following theorem, which states important facts on the distribution of the fractional numbers, is well known in the ergodic theory, and will play a fundamental role throughout our higher order MASH analysis, as it did in [20]–[23].

**Weyl’s Theorem [28]:** Let \(\hat{e}[n] = a_0 + a_1 n + a_2 n^2 + \cdots + a_k n^k\) be a polynomial with real coefficients. If among \(a_1, a_2, \ldots, a_k\) at least one of them, except the constant term \(a_0\), is an irrational number, then for any Riemann integrable function \(f(\cdot)\) the following equality holds:

\[
\lim_{N_t \to \infty} \frac{1}{N_t} \sum_{n=0}^{N_t-1} f(\langle \hat{e}[n] \rangle) = \int f(r) \, dr. \tag{8.7}
\]

For the purpose of later use, several useful properties of the modulo 1 arithmetic are also given here. Let \(a\) and \(b\) be real
numbers. It follows from the definition of the fractional operator that
\[ \langle a + b \rangle = \langle a \rangle + \langle b \rangle - \chi_{[a, b)}(\langle b \rangle) \]  
(8.8)
where the indicator function \( \chi_{[a, c)}(y) \) is defined as
\[ \chi_{[a, c)}(y) = \begin{cases} 
1, & \text{if } y \in [a, c) \\
0, & \text{otherwise}
\end{cases} \]  
(8.9)
with \( 0 \leq a < c \) where \( c < 1 \).

**B. Proof of the Theorem 3**

In this section, a synopsis of the proof of Theorem 3 is presented. A complete proof is given in [27], therefore much of the details are omitted in this paper, for the sake of brevity. Before proceeding to proving the theorem, it is appropriate to restate the assumptions made on the input to the modulator.

- The input to the modulator is a constant \( x[n] = X \in [-b, +b] \) for all \( n \).
- The normalized dc input to the modulator is a rational number (i.e., \( X/(2b) \) is a rational number). Note that this assumption implies that \( \beta \) is also a rational number [see (2.6)], and therefore can be given by \( \beta = K/N \), where \( K \) and \( N \) are relatively prime integer numbers.

Our goal is to derive expressions for the mean, variance, and autocorrelation of the intermediate sequence \( \zeta[n] \) given in (2.7) using (8.1)–(8.3), from which the statistics of the normalized binary quantizer error sequence can be recovered using (8.4)–(8.6). First note that with the assumption \( \beta = K/N \), where \( K \) and \( N \) are relatively prime integer numbers, \( \zeta[n] \) is a periodic sequence with a period of \( N \) [see (2.7)], and thus, the infinite summation in (8.1) reduces to a finite summation. Using this fact and employing the property in (8.8), the sample mean of \( \zeta[n] \) can be written as
\[ M(\zeta) = \frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{v_0}{2b} + n \frac{K}{N} \right) \]  
(8.10)
Before proceeding further, we shall note the following lemma adopted from [18].

**Lemma 1** [18]: If \( \beta = K/N \) where \( K \) and \( N \) are relatively prime integer numbers, then the collection of numbers \( \{nK/N; n = 0, 1, \ldots, N-1\} \) corresponds to the same collection of numbers \( \{n/N; n = 0, 1, \ldots, N-1\} \).

See [18] for a proof of the above lemma. Let us momentarily assume that \( v_0 \geq 0 \), which leads to \( \langle v_0/(2b) \rangle = v_0/(2b) \). Applying **Lemma 1** to (8.10), we get the following expression:
\[ M(\zeta) = \frac{v_0}{2b} + \frac{1}{N} \sum_{n=0}^{N-1} n \frac{K}{N} - \frac{1}{N} \sum_{n=0}^{N-1} \chi_{[1-(v_0/(2b), 1)}(n/N) \]  
(8.11)
The above expression can be evaluated when the appropriate limits are used for the conditional summation (rightmost term) such that
\[ M(\zeta) = \frac{v_0}{2b} + \frac{1}{N} \sum_{n=0}^{N-1} n \frac{K}{N} - \frac{1}{N} \sum_{n=[1-(v_0/(2b)N]}^{N-1} 1 \]  
(8.12)
where \( [x] \) is the smallest integer greater than or equal to \( x \). As a matter of convenience, let us define \( \eta_0 = \lceil 1 - v_0/(2b) \rceil N \).

With this definition
\[ M(\zeta) = \frac{v_0}{2b} + \frac{1}{N} \left( \frac{(N-1)N}{2} - \frac{1}{N} (N-N_0) \right) \]  
(8.13)
By substituting (8.13) into (8.4), we obtain an expression for the mean value of the normalized binary quantizer error sequence, which is given as
\[ M(\zeta) = \frac{1}{2N} + \left( 1 - \frac{v_0}{2b} \right) - \frac{1}{N} \eta_0 \]  
(8.14)
The second moment of \( \zeta[n] \) is given by
\[ M(\zeta^2) = \frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{v_0}{2b} + n \frac{K}{N} \right)^2 \]  
(8.15)
Combining (8.15) with (8.8), and using **Lemma 1** results in
\[ M(\zeta^2) = \frac{-4N^2 + 3N + 1}{6N^2} \]  
(8.16)
By expanding the square term in (8.16), and after some algebraic manipulations, one can achieve
\[ M(\zeta^2) = \frac{v_0}{2b} \left( \frac{2\eta_0}{N} + \frac{v_0}{2b} - 1 - \frac{1}{N} \right) - \frac{1}{N} \left( \frac{1 - N_0}{N} \right) + 1 \]  
(8.17)
Hence, the average power of the binary quantizer error sequence is found using (8.5), (8.13) and (8.17) followed by simple algebraic manipulations,
\[ M(\zeta^2) = \frac{1}{12} + \frac{1}{6N^2} + \frac{v_0}{2b} \left( \frac{2\eta_0}{N} + \frac{v_0}{2b} - 2 \right) - \frac{1}{N} \left( \frac{1 - N_0}{N} \right) + 2 \left( \frac{N + 1}{N} \right) \eta_0 \]  
(8.18)
The autocorrelation function of \( \zeta[n] \) sequence is given by
\[ r_c(l) = \frac{1}{N} \sum_{n=0}^{N-1} \left( n \frac{K}{N} + \frac{v_0}{2b} \right) \left( (n+l) \frac{K}{N} + \frac{v_0}{2b} \right) \]  
(8.19)
Using (8.8), the above expression can be written as
\[ r_c(l) = M(\zeta^2) + \langle \eta, \beta \rangle M(\zeta) \]  
(8.20)
The rightmost term in (8.20) is evaluated as [27]
\[ \frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{nK}{N} + \frac{v_0}{2b} \right) \chi_{[1-(lK/N), 1)}(n/N) \]  
(8.21)
Substituting (8.21), (8.17), and (8.13) into (8.20), we obtain
\[ r_c(l) = \frac{-4N^2 + 3N + 1}{6N^2} + \frac{v_0}{2b} \left( \frac{2\eta_0}{N} + \frac{v_0}{2b} - 1 - \frac{1}{N} \right) - \frac{1}{N} \left( \frac{1 - N_0}{N} \right) + 1 + \langle \eta, \beta \rangle \left( \frac{l\beta}{2} - \frac{1}{2} \right) \]  
(8.22)
which, in conjuction with (8.6) and (8.13), gives an expression for the autocorrelation function of the binary quantizer error sequence that is given by

$$r_c(l) = \frac{1}{12} + \frac{\nu_0}{6b^2} + \frac{\nu_0}{2b} \left( \frac{2\nu_0}{N} + \frac{\nu_0}{2b} - 2 - \frac{1}{N} \right) - \frac{\nu_0}{N} \left( \frac{1-\nu_0}{N} \right) + 2 - \frac{(N+1)}{\nu_0} - \left( \frac{1}{N} \right) \beta \left( 1 - \frac{1}{N} \right).$$

This proves the theorem. \(\square\)

Note that in the analysis performed here, it was assumed that \(\nu_0 \geq 0\). Similar procedures, however, can be undertaken to extend these results for the negative initial condition by using the fact that \(\langle \nu_0/(2b) \rangle = 1 - \langle \nu_0/(2b) \rangle\) when \(\nu_0 \in [-2b, 0]\). The details are omitted here for the sake of brevity. Nevertheless, the results are the same for negative initial condition, and consequently the formulas in (8.14), (8.18), and (8.23) are still valid. This fact is justified by computer simulations given in Section IV (see Fig. 4). It should also be emphasized that the above development does not put any restriction on the nature of the initial condition, hence the results are valid for both rational and irrational initial condition.

### C. Proof of Theorem 4

In order to evaluate the moments of the normalized binary quantizer error sequence in a MASH modulator, one needs to determine a closed form for the nested sums in (2.16). In achieving this, we follow similar notations used in [21] and [23]. Let \(P_k(n)\) represents “generically” a polynomial of degree \(k\) in \(n\). The qualifier “generically” indicates that \(P_k(n)\) represents any polynomial with degree \(k\) [21]. Effectively, one can write for any positive integer numbers \(k\) and \(m\) with \(m \leq k\)

$$P_k(n) = aP_k(n) + bP_m(n)$$

(8.24)

where \(a\) and \(b\) are two real numbers with \(a \neq 0\). Similarly, we define \(R_k(n)\) and \(3L(n)\) to represent generic polynomials of degree \(k\) with all rational and irrational coefficients, respectively. Note that (8.24) is also valid for rational and irrational polynomials. We continue by noting a lemma from [21].

**Lemma 2:** The \(L\)-th order sum of the constant of “1” is a polynomial of degree \(L\) with all rational coefficients. Mathematically

$$\sum_{k_0=0}^{n-1} \sum_{k_1=0}^{k_0-1} \sum_{k_2=0}^{k_1-1} \cdots \sum_{k_{L-1}=0}^{k_{L-2}-1} 1 = R_L(n).$$

(8.25)

See [21] for proof of this lemma. With the aid of Lemma 2, a closed form for the nested summation in (2.16) is given by

$$\sum_{k_0=0}^{n-1} \sum_{k_1=0}^{k_0-1} \sum_{k_2=0}^{k_1-1} \cdots \sum_{k_{L-1}=0}^{k_{L-2}-1} \left( \frac{\nu_0}{2b} + \frac{\nu_1}{2b} + \cdots + \frac{\nu_{L-1}}{2b} \right) = \frac{\nu_0}{2b} + \frac{\nu_1}{2b} R_1(n) + \cdots + \frac{\nu_{L-1}}{2b} R_{L-1}(n) + \beta R_L(n).$$

(8.26)

Using (8.26) and (8.1), the mean value of the sequence \(\zeta_L[n]\) is given by

$$M(\zeta_L) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{\nu_0}{2b} + \frac{\nu_1}{2b} R_1(n) + \cdots + \frac{\nu_{L-1}}{2b} R_{L-1}(n) + \beta R_L(n) \right).$$

(8.27)

Throughout the developments in this subsection, we consider the case where the constant input is a rational fraction of the quantizer step size (i.e., \(\beta\) is rational), and the initial condition on the first accumulator is an irrational fraction of the quantizer step size (i.e., \(\nu_1/(2b)\) is irrational). With this irrational initial condition assumption, the infinite limit in (8.27) is evaluated with the aid of Weyl’s Theorem stated in (8.7). Applying (8.24) to (8.27), and using (8.7), one can write

$$M(\zeta_L) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left( 3L - 1(n) + R_L(n) \right) \int_0^r dr = \frac{1}{2} \int_0^r dr = \frac{1}{3}$$

(8.28)

provided \(L \geq 2\). Note that the Weyl’s Theorem can only be utilized when \(L \geq 2\), because we need an irrational polynomial of degree at least one. Similarly, the second moment of \(\zeta_L[n]\) is written as:

$$M(\zeta_L^2) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=0}^{N-1} \left( 3L - 1(n) + R_L(n) \right)^2 \int_0^r dr = \frac{1}{3}$$

(8.29)

if \(L \geq 2\). The sample mean and average power of the binary quantizer error sequence from the last stage directly follows from (8.4) and (8.5), respectively

$$M(\varepsilon_L) = 0, \quad M(\varepsilon_L^2) = \frac{1}{12}.$$

(8.30)

Note that (8.30) is valid for any \(L \geq 2\). However, the autocorrelation of the binary quantizer error sequence in a second-order MASH \(\Sigma\Delta\) modulator needs some special treatments; therefore, the results for the autocorrelation function significantly differs from the third-order or higher order \((L \geq 3)\) case.

Using (2.16) and (8.3), the autocorrelation of \(\zeta_2[n]\) is given by (8.31) at the bottom of the page. In order to be able to evaluate the above limit, we make use of the characteristic function method as described in [21], [27]. Therefore, the autocorrelation of \(\zeta_2[n]\) can be written as

$$r_{\zeta_2}(l) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \hat{g}(\hat{i}) \hat{g}(k) \Phi_\theta(i, k)$$

(8.32)

where \(\hat{g}(\gamma)\) are the Fourier series coefficients of the periodic function \(g(\gamma) \equiv (\gamma)\), and are given by [4]

$$\hat{g}(i) = \left\{ \begin{array}{ll}
\frac{2}{2i\pi^2}, & i = 0 \\
\frac{i}{2i\pi^2}, & i \neq 0
\end{array} \right.$$ (8.33)

\(\Phi_\theta(i, k)\) is the joint characteristic function of \(\zeta_2[n]\) sequence, and defined as \((i\text{ and }k\text{ are integer numbers})\) [21], as shown

$$r_{\zeta_2}(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{n=0}^{N-1} \left( \frac{\nu_0}{2b} + \frac{\nu_1}{2b} (n + \beta) + \frac{\nu_2}{2b} (n + \beta)^2 - \frac{\nu_1}{2b} \right) \left( \frac{\nu_0}{2b} + \frac{\nu_1}{2b} (n + \beta) + \frac{\nu_2}{2b} (n + \beta)^2 - \frac{\nu_1}{2b} \right).$$

(8.31)
in (8.34) at the bottom of the page. An application of Weyl’s Theorem to (8.34) results in
\[
\Phi_t(i, k) = \begin{cases} 
    e^{j2\pi k(l(v_{1,0}/2b + \beta/2(l-1))}, & i + k = 0 \text{ and } \beta k l = \text{integer otherwise.} 
\end{cases}
\] 
\tag{8.35}

Substituting (8.35) and (8.33) into (8.32), results in (8.36) shown at the bottom of the page. Recall that [29]
\[
\sum_{k=-\infty}^{\infty} \frac{1}{4\pi^2 k^2} = \frac{1}{12}. 
\tag{8.37}
\]

Using (8.37) in (8.36), and after simplifications, we get (8.38), shown at the bottom of the page. In order to evaluate the above expression, one needs to determine the set of integer numbers \(k\), which makes \(\beta k l\) an integer. Note that by virtue of our rational dc input assumption, \(\beta\) is given by a ratio of two integers numbers such that \(\beta = K/N\), where \(K\) and \(N\) are relatively prime numbers with \(K < N\). The set of integer numbers \(k\) which makes \(\beta k l = 0\) (where \(l \neq 0\)) is then given by \(k = m\mu; m = 1, 2, \ldots\), where \(\mu\) is a positive integer number which is a function of \(N\) and \(l\), and is given by the following relationship:
\[
\mu = \begin{cases} 
    \frac{N}{l}\mu, & \text{when } \left\lfloor \frac{N}{l} \right\rfloor = 0 \text{ and } |l| \leq N \\
    \frac{N}{\text{GCD}(N,l)}, & \text{when } \left\lfloor \frac{N}{l} \right\rfloor \neq 0 \text{ and } |l| \leq N \\
    1, & \text{when } \left\lfloor \frac{N}{l} \right\rfloor = 0 \text{ and } |l| > N \\
    \frac{N}{\text{GCD}(N,l)}, & \text{when } \left\lfloor \frac{N}{l} \right\rfloor \neq 0 \text{ and } |l| > N 
\end{cases}
\tag{8.39}
\]

where GCD\((a,b)\) represents the greatest common divisor of positive integer numbers \(a\) and \(b\). The proof of (8.39) can be accomplished by using results from the elementary number theory, and therefore the details are omitted [27]. Using (8.39), we can write the following equality:
\[
\frac{1}{2\pi^2} \sum_{k=-\infty}^{\infty} \frac{1}{k^2} \cos \left(2\pi k l \left(\frac{v_{1,0}}{2b} + \beta/2(l-1)\right)\right)
= \frac{1}{2\pi^2} \sum_{m=1}^{\infty} \frac{1}{m^2 \mu^2} \cos \left(2\pi m \mu l \left(\frac{v_{1,0}}{2b} + \beta/2(l-1)\right)\right)
\]
for \(l \neq 0\) and \(\beta k l = 0\). We note here a useful identity found in [29, eq. A360]
\[
\sum_{m=1}^{\infty} \frac{1}{m^2} \cos(m\theta) = \frac{1}{4} \left(\theta - \pi\right)^2 - \frac{1}{12} \pi^2 
\tag{8.40}
\]
where \(0 \leq \theta < 2\pi\). Let us call
\[
\gamma = 2\pi \left(\frac{v_{1,0}}{2b} + \beta/2(l-1)\right), 
\]
then using (8.40), (8.38), (8.28), and (8.6) the autocorrelation of the binary quantizer error sequence from the second stage can be found as
\[
r_{\zeta_2}(l) = \begin{cases} 
    \frac{1}{12}, & l = 0 \\
    \frac{1}{2\pi^2 \mu^2} \left(\frac{1}{4} \left(\gamma - \pi\right)^2 - \frac{1}{12} \pi^2\right), & l \neq 0
\end{cases} 
\tag{8.41}
\]
Hence, the theorem is proved. \(\square\)

D. Proof of Theorem 5

Note that (8.30) is valid for any \(L \geq 2\). Therefore, to prove Theorem 5, all we need to derive is an expression for the autocorrelation of the binary quantizer error sequence from the last stage. To this effect, we again utilize the characteristic function method [21] as discussed in Appendix II. Using the Fourier series representation for the modulo 1 operator, the autocorrelation of \(\zeta_L[n]\) is given by
\[
r_{\zeta_L}(l) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(i) \delta(k) \Phi_t(i,k)
\tag{8.42}
\]
where the joint characteristic function of \(\zeta_L[n]\) sequence is defined as follows:
\[
\Phi_t(i,k) = \lim_{N_L \to \infty} \frac{1}{N_L} \sum_{n=-N_L}^{N_L} e^{-2\pi i \delta(n)} \delta(n+i) \delta(n+k)
\]
\tag{8.43}
The above expression can be evaluated using the following lemma [27].

Lemma 3 [27]: Consider an irrational polynomial \(\zeta_{L-1}(n)\) of degree \((L - 1)\) in \(n\). For all integer numbers \(i, k, l\), the following equality holds:
\[
i \zeta_{L-1}(n) + k \zeta_{L-1}(n+l) 
= \begin{cases} 
    \zeta_{L-1}(n), & \text{if } i + k \neq 0, \forall l \\
    \zeta_{L-2}(n), & \text{if } i + k = 0, kl \neq 0 \\
    0, & \text{if } i + k = 0, kl = 0
\end{cases}
\tag{8.44}
\]

\[
\Phi_t(i,k) = \lim_{N_L \to \infty} \frac{1}{N_L} \sum_{n=0}^{N_L-1} e^{2\pi i (v_{2,0}/2b+v_{1,0}/2b+v_{1,0}/2b+\beta/2(n+4))/2(n+4))} \cos(2\pi k (v_{2,0}/2b+v_{1,0}/2b+\beta/2(n+4))/2(n+4))
\tag{8.34}
\]

\[
r_{\zeta_2}(l) = \begin{cases} 
    \frac{1}{4} + \sum_{k=-\infty}^{\infty} \frac{1}{k^2}, & l = 0 \\
    \frac{1}{4} + \sum_{k=-\infty}^{\infty} \frac{1}{k^2} e^{2\pi i k l (v_{1,0}/2b+\beta/2l(l-1))}, & l \neq 0 \text{ and } \beta k l = \text{integer}
\end{cases}
\tag{8.36}
\]

\[
r_{\zeta_2}(l) = \begin{cases} 
    \frac{1}{4} + \frac{1}{8b}; & l = 0 \\
    \frac{1}{4} + \sum_{k=-\infty}^{\infty} \frac{1}{k^2} \cos \left(2\pi k l \left(\frac{v_{1,0}}{2b} + \beta/2(l-1)\right)\right), & l \neq 0 \text{ and } \beta k l = \text{integer}
\end{cases}
\tag{8.38}
\]
By making use of Weyl’s Theorem and Lemma 3, one can evaluate (8.43) as
\[ \Phi_T(i, k) = \begin{cases} 1, & i + k \equiv 0 \text{ and } k = 0 \\ 0, & \text{otherwise} \end{cases} \quad (8.45) \]

Note that in the above development, in order to be able to utilize Weyl’s Theorem the order of the MASH $\Delta \Sigma$ modulator must be at least three (i.e., $L \geq 3$). When $L = 2$, (8.45) is no longer valid thus this case needs special treatment as was performed in Appendix II-C. The autocorrelation of the binary quantizer error sequence from the last stage can be found by substituting (8.45) and (8.33) into (8.42), and making use of (8.6), (8.37), and (8.28), to arrive at
\[ r_{\xi\xi}(l) = \begin{cases} \frac{1}{2r}, & l = 0 \\ 0, & l \neq 0 \end{cases} \quad (8.46) \]

This proves the theorem!

Acknowledgment

The authors wish to thank the anonymous reviewers for helping to improve the presentation of the paper.

References


Mücahit Kozak (S’80–M’01) was born in Izmir, Turkey, in 1973. He received the B.Sc., M.Sc., and Ph.D. degrees from Bilkent University, Ankara, Turkey, the Başkent University, Ankara, Turkey, and the University of Westminster, London, U.K., in 1996, 1998, and 2001, respectively, all in electrical and electronics engineering. His Ph.D. thesis was on the analysis and design of delta–sigma modulator based data conversion systems. During his Ph.D. study, he undertook an industrial consultancy work in working in the field of Fractional-N phase-locked loop design and implementation, for Nokia Networks, Farnborough, U.K., which he joined in 2001 as a Digital Signal Processing (DSP) Research Engineer, working on developing core DSP functions as well as their efficient implementation on fixed-point processors. In 2002, he joined the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, as a Post-Doctoral Fellow. His current research interests are CMOS analog and mixed-signal system design (including image sensors, analog–digital and digital–analog converters), delta–sigma modulators, and low-voltage analog design techniques. Dr. Kozak received the Fellowship Awards of Bilkent University, Başkent University, and University of Westminster, during his B.Sc., M.Sc., and Ph.D. studies, respectively.

İzett Kale (M’88) was born in Akıncilar, Cyprus. He received the B.Sc. (honors) degree in electrical and electronic engineering from the Polytechnic of Central London, London, U.K., the M.Sc. degree in design and manufacture of microelectronic systems from Edinburgh University, Edinburgh, U.K., and the Ph.D. degree in techniques for reducing digital filter complexity from the University of Westminster, (formerly the Polytechnic of Central London) London, U.K. He joined the staff of the University of Westminster in 1984 and he has been with them since. He is currently Professor of Applied Digital Signal Processing (DSP) and very large-scale integration (VLSI) Systems, Founder and Leader of the Applied DSP and VLSI Research Group at the University of Westminster, where he has undertaken numerous applied research and development contracts for European, U.S. and Japanese corporations. His research and teaching activities include DSP, analog signal processing, silicon circuit and system design, digital filter design and implementation, and analog–digital and digital–analog sigma–delta converters. He is currently working on efficiently implementable low-power DSP algorithms/architectures and sigma–delta modulator structures for use in the communications and biomedical industries.