FORMAL SPECTRAL THEORY FOR IDEAL SIGMA-DELTA QUANTIZATION WITH STATIONARY TIME-VARYING INPUTS

Nguyen T. Thao

Dept. of Electrical Engineering, CUNY, New York, NY 10031, email: thao@ee.ccny.cuny.edu

ABSTRACT

Because a $\Sigma\Delta$ modulator is a nonlinear feedback system, its rigorous signal analysis escapes from the standard theories used in signal processing and communications. We introduce mathematical tools that are new to these two areas and that lead to systematic and high-level methods for the spectral error analysis of $\Sigma\Delta$ modulators. They include the use of dynamical system techniques, of the roots of ergodic theory and the spectral properties of unitary operators in a Hilbert space. In this paper, we show their application to the case of ideal $\Sigma\Delta$ modulation, thus providing a formal and concise theory for the extensive derivations performed by Gray et al. and He et al. in this special case some 15 years ago. We finally point to the potential extension of these new methods to a substantially larger class of $\Sigma\Delta$ modulators.

Index Terms — Sigma-Delta modulation, feedback systems, nonlinear systems, error analysis, spectral analysis.

1. INTRODUCTION

In the signal processing area, the transfer function of discrete-time feedback systems is easily analyzed when they are linear and time-invariant. Such systems are classically reduced to equations of the type

$$\begin{cases} \mathbf{u}[k] &= \mathbf{L}\,\mathbf{u}[k-1] + x[k]\,\mathbf{a} \\ y[k] &= \mathbf{b}\cdot\mathbf{u}[k] \end{cases} \tag{1}$$

where $x[k]$ and $y[k]$ are the input and output of the considered system, $\mathbf{L}$ is an $m\times m$ matrix, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{u}[k]$ are m-dimensional vectors, and where $\cdot$ designates the inner-product between vectors. However, this technique collapses as soon as a nonlinear function is involved in the feedback. This is the case of $\Sigma\Delta$ modulators whose structure is shown in Figure 1 and which include a scalar quantizer as nonlinear function. These systems have been successfully used in modern analog-to-digital conversion, but do not enjoy the support of an existing system theory. Most of the current signal analysis of $\Sigma\Delta$ modulation is based on empirical models. The main instance of rigorous error analysis was performed by Gray et al. [1, 2, 3] and He at al. [4] some 15 years ago in the ideal case where

$$H(z) = (1-z^{-1})^m$$

and the quantizer is uniform and not overloaded (it is shown in [4] that non-overloading is guaranteed by making the resolution of the quantizer at least m-bit). As one easily derives from the block diagram of Figure 1 that the system error[1] is

$$x[k] - q[k] = h[k] * e[k] \tag{2}$$

This work has been supported by the National Science Foundation Grants DMS-0219053 and CCF-0515252.

[1] In this paper, we take the opposite to the usual system definition of "error", but this convention will appear to be convenient from a dynamical system point of view.

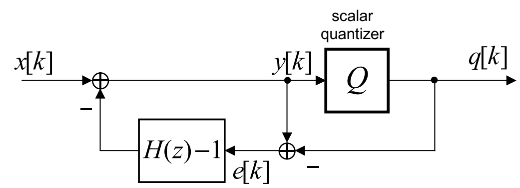

Fig. 1. $\Sigma\Delta$ modulator in its error diffusion form.

where $h[k]$ is the inverse z-transform of $H(z)$ and $e[k] := y[k] - q[k]$ is the quantizer error sequence, this prior work concentrated on the spectral analysis of $e[k]$. The derivations were however heavy in algebra (resulting in multiple journal publications) and very specific to the considered architectures.

We propose here to establish a new concise system approach for this type of nonlinear feedback circuit. As an extension to the equation structure (1), we show that the input-output relation of an ideal $\Sigma\Delta$ modulator can be put in the form of the following system of equations:

$$\begin{cases} \mathbf{e}[k] &= \big\langle \mathbf{L}\,\mathbf{e}[k-1] + x[k]\,\mathbf{i} \big\rangle \\ e[k] &= p(\mathbf{e}[k]) \end{cases} \tag{3}$$

where p is some function from $\mathbb{R}^m$ to $\mathbb{R}$ and $\langle\cdot\rangle$ is an m-dimensional modulo function, by necessity nonlinear. Contrary to linear difference equations, the first equation of (3) is however difficult to iterate, especially with a time-varying input $x[k]$. Assuming an input of the type

$$x[k] = x_0 + \tilde{x}(k\tau) \tag{4}$$

where x_0 is a constant component and $\tilde{x}(t)$ is a bandlimited 1-periodic zero-mean continuous-time signal, we show that the system can be equivalently described by equations of the type

$$\begin{cases} \mathbf{v}[k] &= B(\mathbf{v}[k-1]) \\ e[k] &= p''(\mathbf{v}[k]) \end{cases} \tag{5}$$

where $\mathbf{v}[k]$ is a state vector of higher dimension $m+1$, B is a nonlinear but fixed mapping of $\mathbb{R}^{m+1}$ that only depends on x_0, and p'' is a nonlinear function from $\mathbb{R}^{m+1}$ to $\mathbb{R}$ that depends on $\tilde{x}(t)$. The transformation from (3) to (5) relies on techniques typical to dynamical systems.

Given this new setting of equations, we revisit the derivations performed in the prior work of the time-averaged autocorrelation sequence

$$r_e[n] := \operatorname*{mean}_{k\geq 0}\ e[k]\,e[k+n] \tag{6}$$

where $\operatorname*{mean}_{k\geq 0}\ v[k] := \lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} v[k]$. Its Fourier transform $R_e(\omega)$ is the spectrum of the error sequence $e[k]$ in the time-averaging sense. After finding the mathematical conditions of ergodicity of the mapping B, we show that $r_e[n]$ takes the concise form of

$$r_e[n] = \big(p'', \mathcal{U}^n p''\big)_H \tag{7}$$

where $\big(\cdot,\cdot\big)_H$ is a Hilbert space inner product and $\mathcal{U}$ is a unitary operator. Then, with standard techniques from the spectral theory of unitary operators, the properties of the spectrum $R_e(\omega)$ that were previously derived by extensive derivations, can be retrieved here in a concise manner. Finally, we explain how the derivations shown in this paper can be extended to more complicated inputs, and more importantly to a substantially larger class of $\Sigma\Delta$ modulators.

2. BASIC DYNAMICAL SYSTEM EQUATIONS

As part of the conditions of ideal $\Sigma\Delta$ modulation, it is assumed that the quantizer is uniform and not overloaded. For convenience, we normalize the signal amplitude so that the quantization step size is 1. This implies that at any instant k, $q[k]$ is an integer[2] and the quantizer error satisfies

$$e[k] \in I := [-\tfrac{1}{2}, \tfrac{1}{2}). \tag{8}$$

Meanwhile, equation (2) implies in the z-domain

$$X(z) - Q(z) = (1 - z^{-1})^m E(z).$$

Let us recursively construct the sequences $e_i[k]$ such that in the z-domain

$$E_{i-1}(z) = (1 - z^{-1}) E_i(z) \tag{9}$$

with $e_0[k] := x[k] - q[k]$. Obviously $E_m(z) = E(z)$. Now, (9) implies in the time domain that $e_i[k] = e_i[k-1] + e_{i-1}[k]$, which recursively leads to

$$e_i[k] = e_i[k-1] + e_{i-1}[k-1] + \cdots + e_1[k-1] + (x[k] - q[k]).$$

Then, the m-dimensional column vector

$$\mathbf{e}[k] := \big[e_1[k] \; e_2[k] \; \cdots \; e_m[k] \big]^\top,$$

satisfies the recursive relation

$$\mathbf{e}[k] = \mathbf{L}\,\mathbf{e}[k-1] + (x[k] - q[k])\mathbf{i} \tag{10}$$

where $\mathbf{L}$ is the $m \times m$ lower triangular matrix

$$\mathbf{L} := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{i} := \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{11}$$

Solving the recursive relation (10) is the key to evaluating $r_e[n]$ in (6) since at every instant

$$e[k] = e_m[k] = p(\mathbf{e}[k]) \tag{12}$$

where $p(\mathbf{e})$ is by definition the projection of $\mathbf{e}$ onto its last mth component. The difficulty of equation (10) however is that $q[k]$ is itself a nonlinear function of $\mathbf{e}[k-1]$, which can be easily proved. Now, because of (8), we have $\mathbf{e}[k] = \langle \mathbf{e}[k] \rangle$ if we define $\langle\cdot\rangle$ to be the unique 1-periodic function of $\mathbb{R}^m$ that is invariant on the unit hypercube I^m. This formally, means that $\langle \mathbf{u} + \mathbf{k} \rangle = \langle \mathbf{u} \rangle$ for all $u \in \mathbb{R}^m$ and all $\mathbf{k} \in \mathbb{Z}^m$, and $\langle \mathbf{u} \rangle = \mathbf{u}$ for all $\mathbf{u} \in I^m$. Since $\mathbf{i}$ and $q[k]$ are all composed of integer numbers, then (10) and (12) lead to the system of equations of (3).

[2] Quantizers in $\Sigma\Delta$ modulation are commonly of mid-riser type, implying that $q[k]$ is an integer plus $\frac{1}{2}$. For simplicity here, we will omit here the $\frac{1}{2}$ offset, which is not fundamental in the equations.

3. PSEUDO-PERIODIC INPUT

When $x[k]$ is a time-varying input, it is however difficult to derive $r_e[n]$ from (6), because the explicit determination of $\mathbf{e}[k]$ is still difficult. Now, assuming that $x[k]$ is of the form (4) where $\tilde{x}(t)$ is a 1-periodic zero-mean continuous-time signal, we can proceed by performing an unusual operation to the signal processing area. Instead of taking $\mathbf{e}[k]$ as the state vector of the system, consider the new $m+1$-dimensional state vector

$$\mathbf{u} := (t, \mathbf{e}) \;\in I^{m+1}.$$

Consider then the sequence $\mathbf{u}[k]$ such that

$$\mathbf{u}[k] = A(\mathbf{u}[k-1])$$

where A is the mapping from I^{m+1} to I^{m+1} defined by

$$A(t, \mathbf{e}) := \big(\, \langle t + \tau \rangle \, , \, \langle \mathbf{L}\,\mathbf{e} + (x_0 + \tilde{x}(t+\tau))\mathbf{i} \rangle \, \big) \tag{13}$$

and where $\langle t \rangle$ designates the 1-dimensional modulo function invariant in I. By taking the initial vector $\mathbf{u}[0] := (0, \mathbf{e}[0])$, one can easily check that for all $k \in \mathbb{Z}$, $\mathbf{u}[k] = (\langle k\tau \rangle, \mathbf{e}[k])$ where $\mathbf{e}[k]$ satisfies (3). We thus end up with the new system of equations

$$\begin{cases} \mathbf{u}[k] &= A(\mathbf{u}[k-1]) \\ e[k] &= p'(\mathbf{u}[k]) \end{cases} \tag{14}$$

where $p'(t, \mathbf{e}) := p(\mathbf{e})$. The contribution of this new dynamical system is that its state vector $\mathbf{u}[k]$ is recursively obtained through the mapping A that is independent of the time index k. The study of such mapping was first introduced in [5] in the first order case $m = 1$, and in the language of dynamical systems, is obtained by *skew-product* from the mapping of (3). Since $\mathbf{u}[k+n] = A^n(\mathbf{u}[k])$, then,

$$e[k]e[k+n] = f_n(\mathbf{u}[k])$$

where

$$f_n(\mathbf{u}) := p'(\mathbf{u})\, p'(A^n(\mathbf{u})).$$

Finally, since $\mathbf{u}[k] = A^k(\mathbf{u}_0)$ with $\mathbf{u}_0 := \mathbf{u}[0]$, (6) implies

$$r_e[n] := \operatorname*{mean}_{k \geq 0} \; f_n\big(A^k(\mathbf{u}_0)\big). \tag{15}$$

4. PSEUDO-PERIODIC BANDLIMITED INPUT

In (15), the mapping A is constant with respect to the iteration index k, but is itself a "heavy" operator as it contains the whole input waveform $\tilde{x}(t)$ in its definition (see (13)). This difficulty can be removed thanks to the additional assumption that $\tilde{x}(t)$ is bandlimited. We show that the mapping A can be reduced to the new mapping

$$B(t, \mathbf{e}) := \big(\, \langle t + \tau \rangle \, , \, \langle \mathbf{L}\,\mathbf{e} + x_0 \mathbf{i} \rangle \, \big) \tag{16}$$

thanks to yet another change of state vector. Consider a mapping of the type

$$T(t, \mathbf{e}) := \big(\, t \, , \, \langle \mathbf{e} + \mathbf{y}(t) \rangle \, \big) \tag{17}$$

where $\mathbf{y}(t)$ is some m-dimensional and 1-periodic function. The goal is to check whether there exists a function $\mathbf{y}(t)$ such that

$$B = T^{-1} \circ A \circ T. \tag{18}$$

As we have

$$(A \circ T)(t, \mathbf{e}) = \big(\langle t + \tau \rangle, \langle \mathbf{L}\big(\mathbf{e} + \mathbf{y}(t)\big) + \big(x_0 + \tilde{x}(t+\tau)\big)\mathbf{i} \rangle \big)$$
$$(T \circ B)(t, \mathbf{e}) = \big(\langle t + \tau \rangle, \langle \mathbf{L}\mathbf{e} + x_0 \mathbf{i} + \mathbf{y}(t+\tau) \rangle \big).$$

one can see that (18) is satisfied when

$$\mathbf{y}(t+\tau) = \mathbf{L}\,\mathbf{y}(t) + \tilde{x}(t+\tau)\mathbf{i}. \tag{19}$$

Writing the real Fourier expansion of the bandlimited signal $\tilde{x}(t)$ as $\tilde{x}(t) = \sum_{k=1}^{K} a_k \cos(2\pi kt + \theta_k)$, we show in [6] that a solution to (19) is the function $\mathbf{y}(t) = [y_1(t)\, y_2(t)\, \cdots\, y_m(t)]^\top$ where $y_n(t) = \sum_{k=1}^{K} a_{n,k} \cos(2\pi kt + \theta_{n,k})$ with $a_{n,k} := \frac{a_k}{2^n \sin^n(\pi k\tau)}$ and $\theta_{n,k} := \theta_k + \pi n(k\tau - \frac{1}{2})$. With this choice of function $\mathbf{y}(t)$, (18) is then satisfied. Then, by performing the change of variable

$$\mathbf{v} := T^{-1}(\mathbf{u}) \,\in I^{m+1}$$

(14) becomes the system of equation (5) where $p''(\mathbf{v}) := p'\big(T(\mathbf{v})\big)$. With (17), we have explicitly,

$$p''(t,\mathbf{e}) = p'\big(t, \langle \mathbf{e} + \mathbf{y}(t)\rangle\big) = p\big(\langle \mathbf{e} + \mathbf{y}(t)\rangle\big) = \langle e_m + y_m(t)\rangle. \tag{20}$$

Then, similarly to (6), we find

$$r_e[n] := \operatorname*{mean}_{k\geq 0}\; g_n\big(B^k(\mathbf{v}_0)\big) \tag{21}$$

where

$$g_n(\mathbf{v}) := p''(\mathbf{v})\, p''(B^n(\mathbf{v})) \tag{22}$$

and B is given in (16). Of course, the autocorrelation $r_e[n]$ still depends on the waveform $\tilde{x}(t)$, but this dependence is all contained in the function p'' and no longer in the mapping B that needs to be iterated infinitely in (21).

5. AUTOCORRELATION UNDER ERGODICITY

The next important step is to transform the discrete average of (21) into a continuous integral. This is performed thanks to ergodic theory, from which we recall two properties [7].

Lemma 5.1 *A measure-preserving mapping B from I^N to I^N is* ergodic *if and only if, for all function $f \in L^1(I^N)$ that satisfies $f \circ B = f$, f is a constant function.*

Theorem 5.2 [Birkhoff] *Consider an ergodic measure-preserving mapping B from I^N to I^N. Then, for any $g \in L^1(I^N)$ and almost every $\mathbf{v}_0 \in I^N$,*

$$\operatorname*{mean}_{k\geq 0}\; g\big(B^k(\mathbf{v}_0)\big) = \int_{I^N} g(\mathbf{v})\mathrm{d}\mathbf{v}.$$

First, the mapping B of (16) is indeed measure preserving as $\det(\mathbf{L}) = 1$ and $\mathbf{L}$ is only composed of integer coefficients. The second test is to check whether B is ergodic. The answer is in the next two theorems.

Theorem 5.3 *Consider the mapping B of* (16)*. A function $f \in L^1(I^{m+1})$ satisfies $f \circ B = \lambda f$ for some $\lambda \in \mathbb{C}$ if and only if there exist $(k,\ell) \in \mathbb{Z}^2$ such that $f(t,\mathbf{e}) = e^{j2\pi(kt+\ell e_1)}$ and $\lambda = e^{j2\pi(k\tau+\ell x_0)}$, where e_1 designates the first component of the m-dimensional vector* $\mathbf{e}$*.*

This is proved in [6] by Fourier expansions. As an easy consequence of Definition 5.1 and Theorem 5.3, we have the following property.

Theorem 5.4 *The mapping B of* (16) *is* ergodic *if and only if*

$$\forall (k,\ell) \in \mathbb{Z}^2 \backslash \{(0,0)\}, \qquad k\tau + \ell x_0 \notin \mathbb{Z}. \tag{23}$$

Let us consider for now that we are in the ergodicity condition of (23). By combining Theorem 5.4, Theorem 5.2, (21) and (22), we obtain

$$r_e[n] = \int_{I^{m+1}} g_n(\mathbf{v})\mathrm{d}\mathbf{v} = \int_{I^{m+1}} p''(\mathbf{v})\, p''\big(B^n(\mathbf{v})\big)\,\mathrm{d}\mathbf{v} \tag{24}$$

where p'' is given in (20) and B in (16). As it is standard in the spectral theory of dynamical systems, let $\mathcal{U}$ be the operator on $L^2(I^{m+1})$ defined by

$$\mathcal{U}g = g \circ B. \tag{25}$$

Then, (24) yields the autocorrelation expression of (7), where $\big(\cdot,\cdot\big)_H$ denotes the inner product of the Hilbert space $H := L^2(I^{m+1})$.

6. SPECTRAL PROPERTIES OF A UNITARY OPERATOR

The expression (7) shows that properties of the autocorrelation $r_e[n]$ will result from the general analysis of the sequence

$$s_f[n] := \big(f, \mathcal{U}^n f\big)_H \tag{26}$$

where $f \in L^2(I^{m+1})$. The operator $\mathcal{U}$ of (25) is easily shown to be unitary (i.e. $\big(\mathcal{U}f,\mathcal{U}g\big)_H = \big(f,g\big)_H$) from the fact that B is a mapping of I^{m+1} that preserves measure. Thus, a complete set of techniques suddenly becomes available thanks to the standard spectral theory of unitary operators [8]. Any function $f \in L^2(I^{m+1})$ has a unique decomposition

$$f = \dot{f} + \bar{f}, \tag{27}$$

where $\dot{f}$ is the orthogonal projection of f onto the subspace $\mathcal{E}$ spanned by the eigenfunctions of $\mathcal{U}$ and $\bar{f} \in \mathcal{E}^\perp$. Because $\mathcal{U}$ is unitary, $\mathcal{E}$ and $\mathcal{E}^\perp$ are both invariant by $\mathcal{U}$. As a result, one easily derives that

$$s_f[n] = s_{\dot{f}}[n] + s_{\bar{f}}[n]. \tag{28}$$

In fact, the eigenfunctions of $\mathcal{U}$ are given by Theorem 5.3. From the form of these eigenfunctions, on concludes that $\mathcal{E}$ is exactly the space of functions $f \in L^2(I^{m+1})$ that depend only on t and e_1. As a result, the orthogonal projection of f onto $\mathcal{E}$ is the function

$$\dot{f}(t,\mathbf{e}) = \dot{f}(t,e_1) = \int_{I^{m-1}} f(t,e_1,w_2,\cdots,w_m)\,\mathrm{d}w_2\cdots\mathrm{d}w_m. \tag{29}$$

Remark 6.1 *In the particular case where $m = 1$, $\mathcal{E} = L^2(I^2)$. Hence $f = \dot{f}$ and $\bar{f}$ is the zero function.*

Given that $\dot{f}(t,\mathbf{e}) = \dot{f}(t,e_1)$, then one easily finds from (16) that $(\dot{f} \circ B^n)(t,\mathbf{e}) = \dot{f}(t+n\tau, e_1+nx_0)$. So,

$$s_{\dot{f}}[n] = \int_{I^2} \dot{f}(t,e_1)\dot{f}(t+n\tau,e_1+nx_0)\,\mathrm{d}t\mathrm{d}e_1 = a_{\dot{f}}(n\tau, nx_0),$$

where $a_{\dot{f}}(t,e_1)$ is the two-variable autocorrelation function of the 1-periodic function $\dot{f}(t,e_1)$. This relation gives the interpretation that $s_{\dot{f}}[n]$ is the sampled version of the autocorrelation of $\dot{f}(t,e_1)$ at the vector sampling period (τ,x_0). If we denote the Fourier coefficients of $\dot{f}(t,e_1)$ by $\dot{F}_{k,\ell}$, then it is known that the Fourier coefficients of the autocorrelation $a_{\dot{f}}$ are $\big|\dot{F}_{k,\ell}\big|^2$. Then,

$$a_{\dot{f}}(t,e_1) = \textstyle\sum_{k,\ell} \big|\dot{F}_{k,\ell}\big|^2 e^{2\pi j(kt+\ell e_1)},$$

which implies that $s_{\dot{f}}[n] = \sum_{k,\ell} \big|\dot{F}_{k,\ell}\big|^2 e^{2\pi j(k\tau+\ell x_0)n}$. We conclude that the Fourier transform $S_{\dot{f}}(\omega)$ of $s_{\dot{f}}[n]$ is purely discrete and equal to

$$S_{\dot{f}}(\omega) = \textstyle\sum_{k,\ell\in\mathbb{Z}} \big|\dot{F}_{k,\ell}\big|^2 \delta_{\omega_{k,\ell}}(\omega) \tag{30}$$

where $\omega_{k,\ell} := 2\pi(k\tau + \ell x_0)$ and $\delta_{\omega_{k,\ell}}(\omega)$ denotes the Dirac peak located at frequency $\omega_{k,\ell}$. Meanwhile, the computation of $s_{\bar{f}}[n]$ is not easy in general, but it is known that its Fourier transform is by necessity a continuous measure [8].

7. APPLICATION TO IDEAL $\Sigma\Delta$ MODULATION

We assume again condition (23).

Case $m = 1$: From (7) and (26), $r_e[n] = s_{p''}[n]$, but due to Remark 6.1, p'' is equal to its projection $\dot{p}''$ onto $\mathcal{E}$ ($\bar{p}'' = 0$). Then, from (30), we find

$$R_e(\omega) = S_{p''}(\omega) = \sum_{k,\ell\in\mathbb{Z}} \left|P''_{k,\ell}\right|^2 \delta_{\omega_{k,\ell}}(\omega) \qquad (31)$$

where $P''_{k,\ell}$ are the Fourier coefficients of $p''(t, e_1) = \langle e_1 + y_1(t)\rangle$.

For illustration, let us derive the spectrum coefficients $|P''_{k,\ell}|^2$ in the simple case where $\tilde{x}(t)$ is a single sinusoid $\tilde{x}(t) = a\cos(2\pi t)$. Then, from Section 4, we find $y_1(t) = b\sin(2\pi(t + \frac{\tau}{2}))$ with $b := \frac{a}{\sin(\pi\tau)}$. We have the Fourier expansions $\langle v\rangle = \sum_{\ell\neq 0} \frac{j}{2\pi\ell}\, e^{j2\pi\ell v}$ and $e^{j\alpha\sin(2\pi t)} = \sum_k J_k(\alpha)\, e^{j2\pi kt}$, where $J_k(\alpha)$ is the Bessel function of order k. By using these expansions in the function $\langle e_1 + y_1(t)\rangle$, one extracts $P''_{k,\ell}$ and finds $|P''_{k,\ell}|^2 = \frac{J_k^2(2\pi\ell b)}{4\pi^2\ell^2}$ when $\ell \neq 0$ and $|P''_{k,0}|^2 = 0$.

Case $m \geq 2$: It is easy to see from (20) and (29) that $\dot{p}'' = 0$, since $\langle\cdot\rangle$ is a 1-periodic and zero-mean function. Then, $p'' = \bar{p}''$ and $r_e[n] = s_{\bar{p}''}$. In fact, one can in this case directly obtain $r_e[n]$ from (24). Let us show that this yields the following result:

$$r_e[n] = \frac{1}{12}\delta[n]. \qquad (32)$$

Consider $n \neq 0$. From (16), $B^n(t, \mathbf{e}) = (\langle t + n\tau\rangle, \langle\mathbf{L}^n\mathbf{e} + x_0\mathbf{i}_n\rangle)$, where $\mathbf{i}_n := \sum_{k=0}^{n-1}\mathbf{L}^k\mathbf{i}$. Then from (20), $p''(B^n(t,\mathbf{e})) = p\big(\langle\mathbf{L}^n\mathbf{e} + x_0\mathbf{i}_n + \mathbf{y}(t + n\tau)\rangle\big)$. By looking at the definition of $\mathbf{L}$, one easily finds that $p''(B^n(t,\mathbf{e})) = \big\langle ne_{m-1} + h_n(t, e_1, \cdots, e_{m-2}, e_m)\big\rangle$ where h_n is some function. Since $\langle e\rangle$ is a 1-periodic and zero-mean function of e, then $\int_I p''(B^n(t,\mathbf{e}))\mathrm{d}e_{m-1} = 0$. From (20), $p''(t,\mathbf{e})$ does not depend on e_{m-1}. So $\int_I p''(t,\mathbf{e})p''(B^n(t,\mathbf{e}))\mathrm{d}e_{m-1} = 0$. This proves that $r_e[n] = 0$ for any $n \neq 0$. Meanwhile $r_e[0] = \int_{I^{m+1}} |p''(t,\mathbf{e})|^2\mathrm{d}t\mathrm{d}\mathbf{e}$. Now, $p''(t,\mathbf{e}) = \langle e_m + y_m(t)\rangle$ from (20). Using again the 1-periodicity of $\langle\cdot\rangle$, we have

$$\int_I |\langle e_m + y_m(t)\rangle|^2\mathrm{d}e_m = \int_I |\langle e_m\rangle|^2\mathrm{d}e_m = \int_I e^2\mathrm{d}e = \tfrac{1}{12}.$$

Then, $r_e[0] = \int_{I^m} \frac{1}{12}\,\mathrm{d}t\mathrm{d}e_1\cdots\mathrm{d}e_{m-1} = \frac{1}{12}$. This proves (32).

The autocorrelation of (32) then implies the typical white quantization noise spectrum $R_e(\omega) = \frac{1}{12}$. This result is independent from the choice of bandlimited input waveform $\tilde{x}(t)$.

8. GENERALIZATIONS

We briefly indicate the possible extensions of the derivations presented in this paper. More complete or detailed explanations are contained in [6].

Mixed input: The techniques shown in this paper can be also applied to inputs of the form $x[k] = x_0 + \tilde{x}[k] + \tilde{x}(k\boldsymbol{\tau})$, where $\tilde{x}[k]$ is a *q-periodic* zero-mean sequence, $\tilde{x}(\mathbf{t})$ is a real zero-mean 1-periodic and bandlimited function of the time vector $\mathbf{t}\in\mathbb{R}^d$, and $\boldsymbol{\tau}$ is a fixed sampling vector of $\mathbb{R}^d$. Under some ergodic condition similar to (23), it is found again that $R_e(\omega)$ is purely discrete in the case $m = 1$, with Diracs located at frequencies $\omega_{p,\mathbf{k},\ell} := 2\pi(\frac{p}{q} + \mathbf{k}\cdot\boldsymbol{\tau} + \ell x_0)$ with $p \in \{0, 1, \cdots, q-1\}$, $\mathbf{k}\in\mathbb{Z}^d$ and $\ell\in\mathbb{Z}$. At high orders $m \geq 2$, it is also found that $R_e(\omega) = \frac{1}{12}$, which is the spectrum of a white noise.

Zero-dc input: This case also requires extra derivations as the ergodicity condition of (23) rules out all inputs with a rational dc component. As shown in [6], this case is solved by writing the dynamical system equations satisfied by the state vector of smaller dimension $\mathbf{e}'[k] := [e_2[k]\ \cdots e_m[k]]^\top$. These equations are found to have the same form as (3) at the dimension $m-1$ with an input $x'[k]$ of the form $x'_0 + \tilde{x}'[k] + \tilde{x}'(k\boldsymbol{\tau})$ where x'_0 depends on the initial state value $e_1[0]$. Then under some irrational condition on the initial condition, ergodicity is shown to be re-established, and the same derivations apply to this dynamical system of dimension $m-1$. As a result, it is found that $R_e(\omega)$ is purely discrete at $m = 2$ and white for $m \geq 3$, which coincides with the results of [2, 3, 4].

More general $\Sigma\Delta$ configurations: Among other extensions, it is shown in [6] how similar derivations apply to the case of a mid-riser quantizer (where $q[k]\in\mathbb{Z}+\frac{1}{2}$), or to the use of cascaded ideal $\Sigma\Delta$ modulators [2, 3]. But the ultimate generalization is the extension of this analysis to the much larger class of $\Sigma\Delta$ modulators where the quantizer can be overloaded (thus including the single-bit case when $m \geq 2$) and where $H(z)$ has the more general form of $H(z) = (1 - z^{-1})^m / A(z)$, $A(z)$ being a polynomial of z^{-1} of maximum degree m. This is based on the *tiling* property of the invariant set of the dynamical system mapping, previously shown in [9, 8] with dc inputs, and potentially extendable to the case time-varying inputs [6].

9. ACKNOWLEDGMENT

We would like to thank Sinan Güntürk for his great help in this research.

10. REFERENCES

[1] R.M.Gray, W.Chou, and P.-W.Wong, "Quantization noise in single-loop sigma-delta modulation with sinusoidal input," *IEEE Trans. Communications*, vol. COM-37, pp. 956–968, Sept. 1989.

[2] P.-W.Wong and R.M.Gray, "Two-stage sigma-delta modulation," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 38, pp. 1937–1952, Nov. 1990.

[3] W.Chou, P.-W.Wong, and R.M.Gray, "Multistage sigma-delta modulation," *IEEE Trans. Information Theory*, vol. 35, pp. 784–796, July 1989.

[4] N.He, F.Kuhlmann, and A.Buzo, "Multi-loop sigma-delta quantization," *IEEE Trans. Information Theory*, vol. IT-38, pp. 1015–1028, May 1992.

[5] A.Teplinsky, E.Condon, and O.Feely, "Driven interval shift dynamics in sigma-delta modulators and phase-locked loops," *IEEE Trans. Circuits and Systems I*, vol. 52, pp. 1224–1235, June 2005.

[6] N.T.Thao, "Formal spectral theory for ideal $\Sigma\Delta$ quantization with stationary inputs," *IEEE Trans. Acoust., Speech, Signal Proc.* in preparation.

[7] K.Petersen, *Ergodic Theory*. Cambridge England: Cambridge University Press, 1983.

[8] S.Güntürk and N.T.Thao, "Ergodic dynamics in $\Sigma\Delta$ quantization: tiling invariant sets and spectral analysis of error," *Advances in Applied Math.*, no. 34, 2005.

[9] N.T.Thao, "Breaking the feedback loop of a class of $\Sigma\Delta$ A/D converters," *IEEE Trans. on Signal Proc.*, vol. 52, pp. 3378–3393, Dec. 2004.