The Structure of the Limit Cycles in Sigma Delta Modulation

VLADIMIR FRIEDMAN, MEMBER, IEEE

Abstract—**It is shown that when the input to a sigma delta modulator is a dc level which can be expressed as a rational number b/a, when normalized with respect to the quantizer step, the output bit string is periodic with a period which is a multiple of the denominator a. Based on number theory, the structure of these cycles for single loop modulators is determined and the noise contribution is computed. Around such levels the noise has two peaks, for which the maximum value and the width are proportional to the relative signal bandwith and to the inverse of the period of the cycle, respectively. The effect of the limit cycles on the performance of the A/D and D/A converters using sigma delta modulation is discussed. A comparison between single loop and double loop modulators from the point of view of this phenomena is made.**

I. Introduction

THE principle of delta modulation is the spreading of the quantization noise in a band which is much larger than that of the signal by oversampling and shaping the noise. Though delta modulation has been known for a long time, a renewed interest in the subject has shown during recent years because of its utilization in the design of PCM codecs. For VLSI technologies under 2μ, such devices become competitive with charge redistribution codecs [5], [6]. Compared to the charge redistribution method, higher PCM sampling rates [7] or higher resolution can be obtained.

Experimental measurements of the noise versus the dc input level for sigma delta modulators [1] and for digital interpolators based on the same principle [2], show the existence of peaks of noise near dc input values which can be expressed as rational fractions, when normalized with respect to the quantizer step. A characteristic shape of such a noise peak is shown in Fig. 1.

Based on an equivalent model of the sigma delta modulator, Candy and Benjamin [1] have determined the baseband noise and the value of such peaks, as well as their influence on the overall performance for slow changing dynamic inputs. Section II of this paper contains a proof of the fact that, for input dc values which can be expressed as rational fractions, the output bit string is periodic. This phenomenon is reminiscent of the limit cycles of digital filters due to finite signal word length. In Section III, an algorithmic procedure based on the Euclid algorithm is given, which derives the structure and the z transform of these cycles for first-order sigma delta modulators. This makes it possible, for this particular case, to compute the noise spectrum based on the output bit stream sequence. In the general case, the noise contribution is determined on the basis of an equivalent circuit such as the infinite staircase model of van de Weg [3] or the discrete pulse phase model [4]. In Section IV, the noise spectrum and the noise contribution are determined for all these dc levels. Finally, in Section V, the effect of the limit cycles on the performance of the A/D and D/A converters based on sigma delta modulation is discussed.

Paper approved by the Editor for Speech Processing of the IEEE Communications Society. Manuscript received March 16, 1987; revised January 4, 1988.

The author is with AT&T Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 8822472.

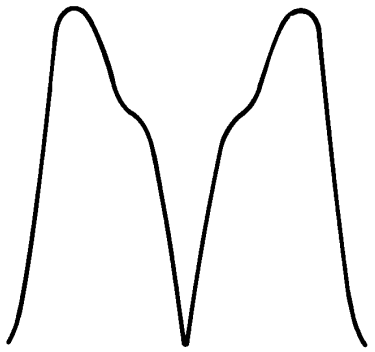


Fig. 1. A characteristic shape of a pair of noise peaks.

II. The Generation of Limit Cycles in Sigma Delta Modulation

We will consider the double loop sigma delta modulator shown in Fig. 2. The z transform of the state variable Y is

$$Y(z)=\frac{G1G2z^{-1}}{(1-z^{-1})^2}X(z)-\frac{G1G2z^{-1}}{(1-z^{-1})^2}Y1(z)-\frac{G2z^{-1}}{1-z^{-1}}Y1(z). \tag{2.1}$$

Taking into account that

$$Y1(n)=\text{sgn } Y(n)$$

where sgn is the sign function defined as

$$\text{sgn } (x)=1 \text{ for } x\geq 0$$

$$\text{sgn } (x)=0 \text{ for } x<0$$

then the following difference equation will describe the behavior of the circuit:

$$Y(n)-2Y(n-1)+Y(n-2)=G1G2(X(n-1)-\text{sgn } \cdot Y(n-1))-G2\ (\text{sgn } Y(n-1)-\text{sgn } Y(n-2)). \tag{2.2}$$

We will consider the case when, for a dc input bias $X(i)=X$, $X\in[0,1]$, there is a limit cycle of period L. Equation (2.2), written for each of the state variables $Y(0)$, $Y(1)$, $\cdots$ $Y(L-1)$, will lead to the following equation system:

$$\begin{aligned}
Y(0)-2Y(L-1)+Y(L-2)&=G1G2(X-\text{sgn } Y(L-1))\\
&\quad -G2\ (\text{sgn } Y(L-1)-\text{sgn } Y(L-2))\\
Y(1)-2Y(0)+Y(L-1)&=G1G2(X-\text{sgn } Y(0))\\
&\quad -G2\ (\text{sgn } Y(0)-\text{sgn } Y(L-1))\\
&\cdots\cdots\\
Y(L-1)-2Y(L-2)+Y(L-3)&=G1G2\\
\cdot\ (X-\text{sgn } Y(L-2))&-G2\ (\text{sgn } Y(L-2)-\text{sgn } Y(L-3)).
\end{aligned} \tag{2.3}$$

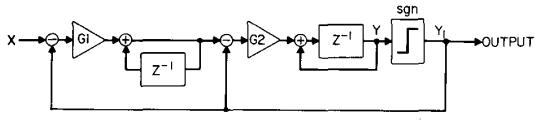


Fig. 2. Second-order sigma delta modulator.

Adding together the equations of the system (2.3), the left side and the last term of the right side will cancel and we will obtain

$$X=\frac{\sum_{i=0}^{i=L-1}\operatorname{sgn} Y(i)}{L}=\frac{P}{L}=\frac{b}{a} \tag{2.4}$$

where P is the number of positive bits of the output limit cycle, and a and b are relative prime integers, i.e., their greatest common divisor is 1.

The necessary condition for the existence of a limit cycle is that the input must be a rational number. The length of the cycle is a multiple of the denominator. These results are independent of the value of the gains $G1$, $G2$. In the relation (2.4), X was normalized with respect to the quantizer step. If in the circuit from Fig. 2 the quantizer level is A instead of unity, then the value of X corresponding to that from the relation (2.4) is

$$X=\frac{b}{a}A. \tag{2.4.a}$$

This result is identical with that previously reported [8] for the single loop modulator (shown in Fig. 3), described by the difference equation

$$Y(n)-Y(n-1)=X(n-1)-\operatorname{sgn} Y(n-1). \tag{2.5}$$

It can be also obtained by using a similar proof as that for the double loop modulators. We will try to determine the sufficient conditions for the existence of the limit cycle. We will consider the set of rational numbers $\{b/a\}$ where a is fixed and b can be any integer. This set is isomorphic with the set of integer numbers. For the single loop delta modulator described by (2.5), if the dc input $X(n)$ is an element of the set $\{b/a\}$, then the right term of the equation will be an element of the same set, since the sign function is an integer. For the initial condition $Y(0) = y_0$, $Y(1)$ will belong to the set $y_0 + \{b/a\}$, and if $Y(n-1)$ is an element of the set $y_0 + \{b/a\}$, then $Y(n)$ will belong to the same set. Therefore, the values $Y(n)$ can take are situated at intervals $1/a$ apart. If the state variable $Y(n)$ is limited to a finite interval of length M, then the number of acceptable values is finite and equal to Ma, after a finite number of steps Y will repeat itself. For a single loop sigma delta modulator, the range the state variable can take is less than two (Fig. 4). But we have seen that the period of the limit cycle must be a multiple of a. Therefore, for a single loop sigma delta modulator, the period of the limit cycle is equal to the denominator a and is independent of the initial condition.

This is not the case for double loop sigma delta modulation. For simplicity, we will consider the modulator shown in Fig. 2 with $G1 = G2 = 1$, described by the equation

$$Y(n)=2Y(n-1)-Y(n-2)$$
$$-2\operatorname{sgn} Y(n-1)+\operatorname{sgn} Y(n-2)+X. \tag{2.6}$$

If X belongs to the set $\{b/a\}$, then the quantity $2 \operatorname{sgn} Y(n-1) - \operatorname{sgn} Y(n-2) - X$ will be an element of the same set, since the sign function takes only integer values. By using the same approach, as for a single loop modulator, it can be easily shown that, with the initial conditions $Y(0) = y_0$, $Y(1) = y_1$,

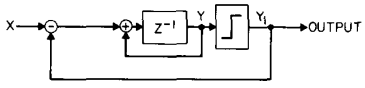


Fig. 3. First-order sigma delta modulator.

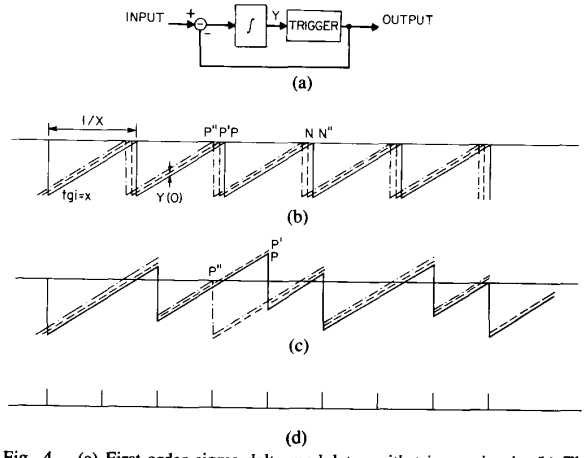


Fig. 4. (a) First-order sigma delta modulator with trigger circuit. (b) The integrated error waveforms for the asynchronous modulator (c) The integrated error waveform for the modulator synchronized by the clock (d).

$Y(n)$ will be equal to

$$Y(n)=ny_1-(n-1)y_0-\frac{b_n}{a}. \tag{2.7}$$

If there is a limit cycle of period L then

$$Y(n+L)=(n+L)y_1-(n+L-1)y_0-\frac{b_{n+L}}{a}=Y(n) \tag{2.8}$$

which leads to

$$L=\frac{b_{n+L}-b_n}{(y_1-y_0)a} \tag{2.9}$$

for any integer $n > 0$. The period of the cycle depends on the initial conditions. For instance, for $X = 0.5$ with the initial condition $Y(1) = 0$ and $Y(0) = 0.5$ and $Y(0) = 0.75$, the period of the cycle will be equal to 4 and 8, respectively. In fact, if the difference

$$y_1-y_0 \tag{2.10}$$

in (2.9) is such that L is not an integer, there will be no limit cycle. This will occur when the difference (2.10) is an irrational number.

III. The Structure of the Limit Cycle

We will consider first the same model of single loop sigma delta modulator analyzed in [1], which is shown in Fig. 4(a). A positive pulse is generated by the trigger circuit when its input is equal to or greater than zero. For the asynchronous modulator which uses no timing clocks when the input is a constant dc level X, then the characteristic waveform of the error Y will be a sawtooth of amplitude unity and period $1/X$. In Fig. 4(b) such a waveform (drawn with unbroken line) for the initial condition $Y(0) = 0$ is represented. If the modulator is synchronized by a clock [Fig. 4(d)], then the decisions concerning the output are taken at fixed intervals determined

by the clock edge [Fig. 4(c)]. The two waveforms for synchronous and asynchronous modulators are related. If a positive pulse is generated for the asynchronous modulator in the interval between two clock edges, then the corresponding pulse for the synchronous modulator will be generated at the next clock edge. Because of this, the error waveform for the synchronous modulator overshoots in the positive region before returning to the same value as in the asynchronous case.

If the initial condition is positive, both error waveforms will slide to the left, the distance from their previous position measured along the vertical axis is equal to $Y(0)$. For very small values of $Y(0)$, the point P which is the nearest to the preceeding clock edge when a pulse is generated by the asynchronous modulator will move closer to the clock edge and become P', but it will remain in the same interval. The same thing will happen with the rest of the peaks. As a result the limit cycle of the synchronous modulator remains unchanged, 10101101 in our case. When $Y(0)$ is sufficiently large, then the point P will move on the clock edge to the position P'', and the corresponding synchronous pulse will be generated one clock period earlier. But the asynchronous and synchronous waveforms generated in this case are identical to those for zero initial condition with the origin translated in P''. The new limit cycle 10110101 can be obtained from the previous one by shifting it three bit positions to the right. Similar considerations can be made for negative initial conditions, for the point N closest to the following clock edge.

In conclusion, for the single loop sigma delta modulation, not only is the period of the cycle invariant of the initial condition, but the structure of the cycle remains unchanged, except for a translation in time.

From Fig. 4(c), it can be noticed that in the case of the synchronous modulator, while the sequence of positive pulses cannot be uniformly distributed because of the constraints imposed by the utilization of the clock, it is as close as possible to such a distribution. There is also another explanation for this assumption. For such a sequence the error between the constant input and the average value of any segment of the output bit stream of length N, i.e.,

$$\left| X - \frac{1}{N} \sum_{n=k}^{n=N+k} \operatorname{sgn}\, Y(n) \right|$$

is minimal. For a complete limit cycle this error is 0 [8]. The structure of the limit cycle will be determined by finding the sequence of length a containing b ones distributed as uniformly as possible among b–a zeros. Dividing the sequence length a by the number of ones b,

$$a = bq_0 + r_0 \tag{3.1}$$

where q_0 and r_0 are integers and $r_0 < b$, it results that the input level $X = 1/(q_0 + (r_0/b))$ belongs to the interval $[1/q_0, 1/q_0 + 1]$, the extremities of which are characterized by the cycles containing a one followed by $q_0 - 1$ zeros, and containing a one followed by q_0 zeros, respectively. The first step of the Euclid algorithm [10], written in a modified form:

$$a = bq_0 + r_0 = (b - r_0)q_0 + r_0(q_0 + 1) \tag{3.2}$$

gives the number of such strings. There are:

— $b - r_0$ strings $S_{0,0}$ of length $l_{0,0} = q_0$ with the structure $1[(q_0 - 1)0]$ and z transform $S_{0,0}(z) = 1$.

— r_0 strings $S_{0,1}$ of length $l_{0,1} = q_0 + 1$ with the structure $1[q_0 0]$ and z transform $S_{0,1}(z) = 1$ where $n0$ means the zero value repeated n times. The number of ones is equal to the total number of strings b, each string containing only a one value, while the relation (3.2) ensures that the total number of bits contained in these strings is equal with the period a. The way the two strings were chosen ensures the uniform distribution of ones by subsequent concatenation.

The total number of strings is b, r_0 of them are of the type $S_{0,1}$. In order that the ones in the output sequence of the modulator be distributed as uniformly as possible, the same thing should happen with the strings $S_{0,0}$, $S_{0,1}$. The second step of the Euclid algorithm,

$$b = r_0 q_1 + r_1 = (r_0 - r_1)q_1 + r_1(q_1 + 1),$$

will lead to the generation of more complex structures. $r_0 - r_1$ strings $S_{1,0}$ with the structure $[(q_1 - 1)S_{0,0}]S_{0,1}$ and r_1 strings $S_{1,1}$ with the structure $[q_1 S_{0,0}]S_{0,1}$ will be generated where $[nS_i]S_j$ means that the string S_i is repeated n times and concatenated with the string S_j. The length of the two strings is

$$l_{1,0} = (q_1 - 1)l_{0,0} + l_{0,1} = q_1 q_0 + 1$$

$$l_{1,1} = q_1 l_{0,0} + l_{0,1} = q_1 q_0 + q_0 + 1.$$

Continuing the same procedure, i.e., dividing r_0 by r_1, more complex strings will be generated. Let $n + 1$ be the number of steps of the Euclid algorithm. Since a and b are relative prime then $r_n = 1$ and the last two steps will be

$$r_{n-2} = r_{n-1}q_n + 1 = (r_{n-1} - 1)q_n + (q_n + 1)$$

$$r_{n-1} = q_{n+1}. \tag{3.3}$$

The strings $S_{n,0}$, $S_{n,1}$ will have the structure

$$S_{n,0} = [(q_n - 1)S_{n-1,0}]S_{n-1,1}$$

$$S_{n,1} = [q_n S_{n-1,0}]S_{n-1,1} \tag{3.4}$$

and the lengths are given by

$$l_{n,0} = (q_n - 1)l_{n-1,0} + l_{n-1,1}$$

$$l_{n,1} = q_n l_{n-1,0} + l_{n-1,1}. \tag{3.5}$$

The z transform of the string $S_{n,0}$ is

$$\begin{aligned} S_{n,0}(z) &= S_{n-1,0}(z) \\ &\quad \cdot [1 + z^{-l_{n-1,0}} + z^{-2l_{n-1,0}} + \cdots + z^{-(q_n-2)l_{n-1,0}}] \\ &\quad + S_{n-1,1}(z) z^{-(q_n-1)l_{n-1,0}} \\ &= \frac{1 - z^{-(q_n-1)l_{n-1,0}}}{1 - z^{-l_{n-1,0}}} S_{n-1,0}(z) \\ &\quad + z^{-(q_n-1)l_{n-1,0}} S_{n-1,1}(z) \end{aligned} \tag{3.6}$$

and similarly

$$S_{n,1}(z) = \frac{1 - z^{-q_n l_{n-1,0}}}{1 - z^{-l_{n-1,0}}} S_{n-1,0}(z) + z^{-q_n l_{n-1,0}} S_{n-1,1}(z).$$

Finally, from (3.3), the limit cycle C will consist of $r_{n-1} - 1$ strings $S_{n,0}$ and one string $S_{n,1}$.

$$S_{n+1,0} = C = [(r_{n-1} - 1)S_{n,0}]S_{n,1}. \tag{3.7}$$

Its length a and z transform $C(z)$ will be equal to

$$l_{n+1,0} = a = (r_{n-1} - 1)l_{n,0} + l_{n,1} \tag{3.8}$$

$$C(z) = \frac{1 - z^{-(r_{n-1}-1)l_{n,0}}}{1 - z^{-l_{n,0}}} S_{n,0}(z) + z^{-(r_{n-1}-1)l_{n,0}} S_{n,1}(z). \tag{3.9}$$

As an example, for the input X equal to 35/81, the algorithmic procedure described will lead to the following results:

—1st step: 24 strings $S_{0,0} = 10$ and 11 strings $S_{0,1} = 100$.

—2nd step: 9 strings $S_{1,0} = (2S_{0,0})S_{0,1}$ and 2 strings $S_{1,1} = (3S_{0,0})S_{0,1}$.

—3rd step: one string $S_{2,0} = (4S_{1,0})S_{1,1}$ and one string $S_{2,1} = (5S_{1,0})S_{1,1}$. $C = S_{2,0}S_{2,1} = (4S_{1,0})S_{1,1}(5S_{1,0})S_{1,1} = 4[(2S_{0,0})S_{0,1}](3S_{0,0})S_{0,1}5[(2S_{0,0})S_{0,1}](3S_{0,0})S_{0,1}$. This result is identical with that obtained by computer simulation. We could have chosen as well, $S_{0,0} = 01$, $S_{0,1} = 001$. This corresponds to a one bit position shift to the left, or $S_{0,0} = 01$, $S_{0,1} = 010$, which corresponds to a one bit position shift to the right. Obviously, $S_{0,0} = 01$, $S_{0,1} = 100$, would not work because by concatenation the distribution of ones among zeros will not be uniform. Similar choices can be made at every step, and they will correspond to translations of the limit cycle in time.

IV. The Output Spectrum and the Noise Contribution

The limit cycle is characterized by a periodic output sequence. The Fourier coefficients can be derived from the DFT of this sequence

$$A(k)=\frac{1}{a}\,|C(e^{jk\omega_0 T})|=\frac{1}{a}\,|C(\alpha_k)| \quad \alpha_k=\frac{k}{a} \tag{4.1}$$

where $\omega_0 = 2\pi f_s/a$, f_s is the sampling frequency.

The dc component $A(0)$ represents the input level X, and the harmonics $A(k)$ constitute the noise. If k_0 is the highest harmonic in the baseband, then the noise power will be equal to

$$N=\sum_{k=1}^{k_0} A(k)^2. \tag{4.2}$$

We will define first-order spectra as those output spectra for which the Euclid algorithm applied to the dc input X consists of one equation.

$$a=bq_0+1 \tag{4.3}$$

$$X=\frac{b}{a}=\frac{1}{q_0+\dfrac{1}{b}}. \tag{4.4}$$

The limit cycle of such spectra has the structure

$$C=[(b-1)S_{0,0}]S_{0,1}. \tag{4.5}$$

The noise contribution for $X < X_c = 1/q_0$, can be examined, letting b take values on the set of natural numbers. Substituting $S_{n,0}(z) = S_{n,1}(z) = 1$ in (3.9) we have

$$C(z)=\frac{1-z^{-bq_0}}{1-z^{-q_0}} \tag{4.6}$$

$$A(k)=\frac{1}{a}\left|\frac{\sin\,(bq_0\alpha_k\pi)}{\sin\,(q_0\alpha_k\pi)}\right|. \tag{4.7}$$

In Fig. 5(a), (b), the spectra for $q_0 = 2$ and $b = 3$ and $b = 7$ are shown, the dc value, ($\alpha_0 = 0$) b/a is equal to X, the distance between the kth zero of the envelope $k/a - 1$ and the preceeding spectral line $\alpha_k = k/a$ is $d_k = k/a(a - 1)$. If b is very large, then the period of the cycle a is large, and the noise spectral lines are canceled by the zeros of the envelope function. Decreasing the value of b causes the power of the noise spectral lines to increase (this explains the noise peaks from Fig. 1), and at the same time the number of these components in the baseband will decrease. The discontinuities in the peak region are due to such components leaving the band, and finally for values of b sufficiently low there is no noise component in the band of interest. In the majority of the applications the cutoff frequency f_c of the low-pass filter is very small compared to the sampling frequency, ($\alpha_c = f_c/f_s$

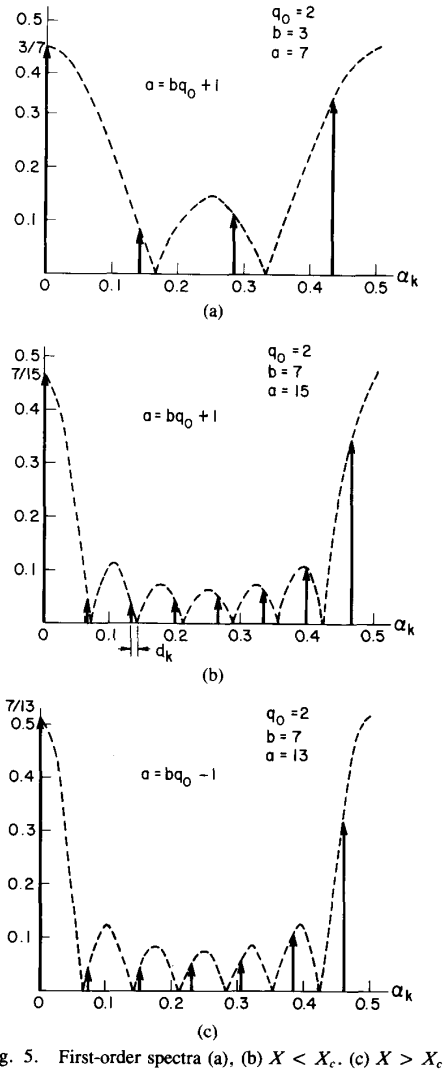


Fig. 5. First-order spectra (a), (b) $X < X_c$. (c) $X > X_c$.

$\ll 1$). If $q_0\alpha_c \ll 1$, the relation (4.7) becomes

$$A(k)=\frac{1}{a}\left|\frac{\sin\left(1-\dfrac{1}{a}\right)k\pi}{\sin\dfrac{q_0 k}{a}\pi}\right|\approx\frac{1}{aq_0}=\frac{1}{aa_c}. \tag{4.8}$$

$a_c = q_0$ is the period of the cycle corresponding to the center value X_c of the peak region. The number of spectral components in the baseband is $[\alpha_c a]$, the brackets indicating the integer part. The noise power (4.2) will be equal to

$$N=\sum_{k=1}^{[\alpha_c a]}\frac{1}{a^2q_0^2}=\frac{[\alpha_c a]}{a^2q_0^2}\approx\frac{\alpha_c}{aq_0^2}=\frac{\alpha_c}{aa_c^2}. \tag{4.9}$$

Taking into account that $1/a \leq \alpha_c$, the noise power will have a maximum for $a = 1/\alpha_c$.

$$N_{\text{peak}}=\frac{\alpha_c^2}{q_0^2}=\frac{\alpha_c^2}{a_c^2}. \tag{4.10}$$

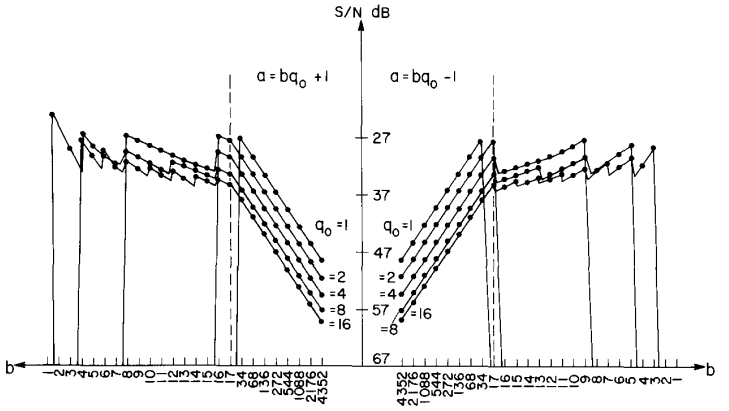


Fig. 6. Signal-to-noise ratio for first-order spectra noise peaks, f_s = 256 KHz, f_c = 4 KHz.

The value of the input dc level at which the peak noise power occurs is

$$X_e=\frac{b}{a}=\frac{1}{q_0}-\frac{1}{aq_0}=\frac{1-\alpha_c}{q_0}. \tag{4.11}$$

Therefore, the width of the noise peak δx will be equal to

$$\delta x=X_c-X_e=\frac{\alpha_c}{q_0} \tag{4.12}$$

The relations (4.10) and (4.12) are identical to the results obtained by Candy and Benjamin [1], except for a factor of two in (4.10) due to their use of single-sided spectra. On the left side of Fig. 6, the signal-to-noise ratio $S/N = X^2/N$, with N computed using the relation (4.2), is represented as a function of b for diverse values of q_0. It can be noticed that for $q_0 = 16$, the approximation made in determining the relation (4.8) does not hold anymore. The right side peak can be analyzed as a degeneracy of the second-order spectra, for which the Euclid algorithm has two steps:

$$a=bq_0+r_0$$
$$b=r_0q_1+1 \tag{4.13}$$

$$X=\frac{b}{a}=\frac{1}{q_0+\dfrac{1}{q_1+\dfrac{1}{r_0}}}=\frac{q_1r_0+1}{r_0(q_0q_1+1)+q_0} \tag{4.14}$$

with the limit cycle $\{[(r_0-1)[[(q_1-1)S_{0,0}]S_{0,1}]\}\cdot[q_1S_{0,0}]S_{0,1}$. For $q_1 = 1$

$$X=\frac{r_0+1}{(q_0+1)(r_0+1)-1}=\frac{1}{q-\dfrac{1}{b}}=\frac{b}{bq-1}=\frac{b}{a}$$
$$q=q_0+1,\ b=r_0+1. \tag{4.14a}$$

By a translation in time the structure of the limit cycle

$$C=[(r_0-1)S_{0,1}]S_{0,0}S_{0,1}=[r_0S_{0,1}]S_{0,0}$$
$$=[(b-1)S_{0,1}]S_{0,0} \tag{4.14b}$$

becomes identical with that for the first-order spectra (4.5) in which the strings $S_{0,0}$, $S_{0,1}$ are interchanged. An identical relation with (4.7) will be obtained for the right side peak, the proof is given in the Appendix. The difference consists of the period of the limit cycle which is $a = bq - 1$. As a result the relative position of the zeros of the envelope and the spectral lines is reversed [Fig. 5(c)], and there is a slight asymetry between the two peaks (Fig. 6). Fig. 7 shows the noise power plotted against the dc input level. Most of the discontinuities from the plots from Fig. 6, in Fig. 7 are smoothed by the characteristic of the real filters, which do not have the abrupt cutoff assumed by the theory [1].

In the Appendix, the Fourier coefficients for the nondegenerate second-order spectra are computed, the dc component is found to be equal to $b/a = X$, and the noise components are given by

$$A(k)=\frac{1}{a}\left[\frac{\sin[(q_1q_0+q_0+1)\alpha_k\pi]}{\sin[(q_1q_0+1)\alpha_k\pi]}\frac{\sin(q_1q_0\alpha_k\pi)}{\sin(q_0\alpha_k\pi)}-\frac{\sin[(q_0q_1+1)\alpha_k\pi]}{\sin(q_0\alpha_k\pi)}\right] \quad \text{for } k=1, 2\cdots. \tag{4.15}$$

When the center value X_c is in the vicinity of a first-order spectra noise peak, its noise contribution will correspond to the characteristic of that peak. An infinite number of second-order spectra points will be generated between two values belonging to the first-order spectra by making r_0 in (4.14) take values on the set of natural numbers. Such intermediate values are shown in Fig. 7. Otherwise, new peaks similar to those from Fig. 7 will be generated. In Fig. 8, the structure of the noise peaks based on the order of the spectra is represented.

We will analyze the noise region in the general case based on the following assumptions:

—the maximum value of the noise peaks is attained before the last noise component leaves the baseband.

$$\alpha_c=\alpha_1=\frac{1}{a}=\frac{1}{(r_{n-1}-1)l_{n,0}+l_{n,1}} \tag{4.16}$$

—the cutoff frequency of the low-pass filter is very small relative to the sampling frequency.

In the general case, the input X is equal with

$$X=\cfrac{1}{q_0+\cfrac{1}{q_1+\cdots\cdots+\cfrac{1}{q_n+\cfrac{1}{r_{n-1}}}}}. \tag{4.16.a}$$

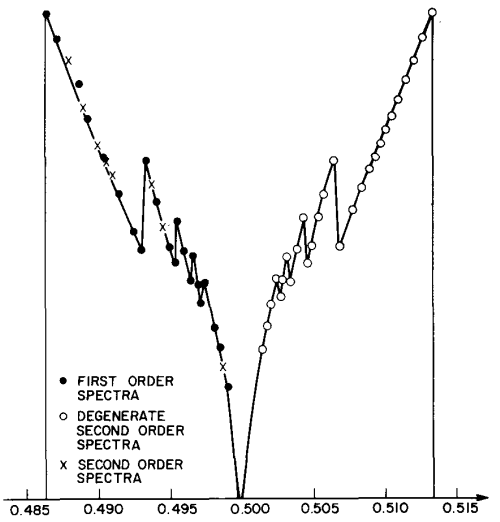


Fig. 7. The pair of noise peaks around $x_c = 0.5$, $f_s = 256$ KHz, $f_c = 3.5$ KHz.

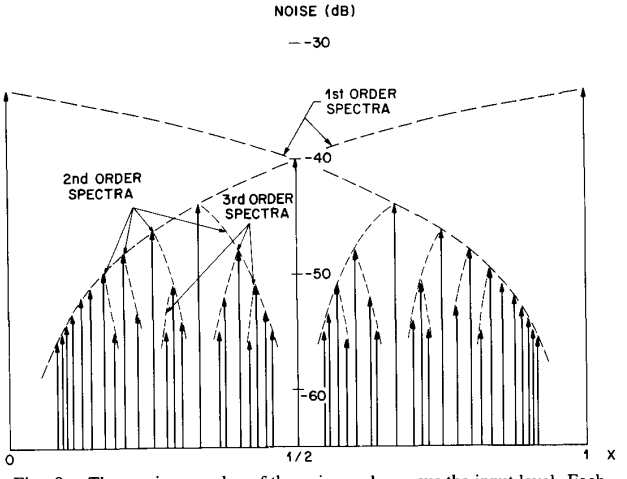


Fig. 8. The maximum value of the noise peaks versus the input level. Each pair of peaks is represented by a spectral line. $f_s = 256$ KHz, $f_c = 3.5$ KHz.

Substituting $z = e^{j\alpha_k\pi}$ in the z transform (3.6) of the strings $S_{n,0}$, $S_{n,1}$, the following expressions for the spectrum of these strings can be found:

$$S_{n,0}(\alpha_k) = e^{-j(q_n-2)l_{n-1,0}\alpha_k\pi} \cdot \frac{\sin[(q_n-1)l_{n-1,0}\alpha_k\pi]}{\sin(l_{n-1,0}\alpha_k\pi)} S_{n-1,0}(\alpha_k) + e^{-j2(q_n-1)l_{n-1,0}\alpha_k\pi} S_{n-1,1}(\alpha_k).$$

$$S_{n,1}(\alpha_k) = e^{-j(q_n-1)l_{n-1,0}\alpha_k\pi} \cdot \frac{\sin(q_n l_{n-1,0}\alpha_k\pi)}{\sin(l_{n-1,0}\alpha_k\pi)} S_{n-1,0}(\alpha_k) + e^{-j2q_n l_{n-1,0}\alpha_k\pi} S_{n-1,1}(\alpha_k). \quad (4.17)$$

For small values of α_k, these relations become

$$S_{n,0}(\alpha_k) = (q_n-1)S_{n-1,0}(\alpha_k) + S_{n-1,1}(\alpha_k)$$
$$S_{n,0}(\alpha_k) = q_n S_{n-1,0}(\alpha_k) + S_{n-1,1}(\alpha_k). \quad (4.18)$$

The relations (4.18) are identical with (3.5) used for determining the string length. Both the length of the string and its spectral value could be determined using the same recursive procedure derived from (4.18):

$$S_{n,0}(\alpha_k) = q_n S_{n-1,0}(\alpha_k) + S_{n-2,0}(\alpha_k)$$
$$S_{n,1}(\alpha_k) = S_{n,0}(\alpha_k) + S_{n-1,0}(\alpha_k). \quad (4.19)$$

The difference between the two series is given by the initial conditions. For the length of the cycle, they are $l_{0,0} = q_0$, $l_{0,1} = q_1q_0 + 1$, this series being known in the theory of numbers [10] as the numerators P_n of the continued fractions expansion of the number representation of the input X. The first two elements of the series $S_{n,0}(\alpha_k)$ are 1 and q_0, respectively, and they will generate the denominators Q_n of the continued fractions. The following relations from the theory of the Euclid algorithm [10], transcribed into our notation, will be used for our proof:

$$l_n S_{n-1,0}(\alpha_k) - l_{n-1} S_{n,0}(\alpha_k) = (-1)^n \quad (4.20)$$

$$a = l_{n+1,0} = r_{n-1}l_{n,0} + l_{n-1,0}$$

$$b = S_{n+1,0}(\alpha_k) = r_{n-1}S_{n,0}(\alpha_k) + S_{n-1,0}(\alpha_k). \quad (4.21)$$

The width of the peak region is given by the difference between the dc input levels X_c, X_e corresponding to the center of the noise peaks $r_{n-1} \to \infty$ and to its extremities [r_{n-1} from (4.16)].

$$\delta x = X_c - X_e = \frac{S_{n,0}(\alpha_k)}{l_{n,0}} - \frac{r_{n-1}S_{n,0}(\alpha_k) + S_{n-1,0}(\alpha_k)}{r_{n-1}l_{n,0} + l_{n-1,0}} = \frac{(-1)^{n+1}}{l_{n,0}(r_{n-1}l_{n,0} + l_{n-1,0})} = (-1)^{n+1}\frac{\alpha_c}{a_c}. \quad (4.22)$$

In (4.22) $a_c = l_{n,0}$. The sign indicates that for odd order spectra $X_c > X_e$ the noise peaks are situated in the left side. The right side peaks are a degenerate case of the next higher order spectra for which the input X is equal with (4.16.a):

$$X = \cfrac{1}{q_0 + \cfrac{1}{q_1 + \cdots\cdots + \cfrac{1}{q_n + \cfrac{1}{q_{n+1} + \cfrac{1}{r_n}}}}}.$$

In the degenerate case $q_{n+1} = 1$, this expression for the input X

$$X = \cfrac{1}{q_0 + \cfrac{1}{q_1 + \cdots\cdots + \cfrac{1}{(q_n+1) - \cfrac{1}{(r_n+1)}}}}$$

becomes identical with (4.16.a) with $q_n' = q_n + 1$ and $r_{n-1}' = -(r_n + 1)$. Similar conclusions can be drawn for even order spectra.

Computing the Fourier coefficients from (4.1) using (4.16)

$$A(k) = \frac{1}{a}\left|C(e^{j2\pi\alpha_k})\right| = \frac{1}{a}\left|\frac{\sin(l_{n,1}\alpha_k\pi)}{\sin(l_{n,0}\alpha_k\pi)} S_{n,0}(\alpha_k) - e^{-jl_{n,1}\alpha_k\pi} S_{n,1}(\alpha_k)\right|$$

and with the assumptions made will become

$$A(k)=\frac{1}{a}\left|\frac{l_{n,1}}{l_{n,0}}S_{n,0}-S_{n,1}\right|=\frac{1}{al_{n,0}}=\frac{1}{aa_c}. \qquad (4.23)$$

This result is identical to (4.8), and as a consequence the formulas (4.10), (4.12) are valid in the general case.

V. The Effects of the Limit Cycles on the Performance of Sigma Delta Modulators

For analog inputs the modulator structures from Figs. 2 and 3 are usually implemented by switched capacitor circuits. The same structures implemented using adders and registers ([6] for single loop, [11] for double loop) can be used in digital-to-analog converters. The PCM input is converted by these modulators in a one-bit stream, which is filtered by analog filters. The performance of these digital modulators is very close to that of the ideal model because their operation is not perturbed by parasitics or thermal noise as in the case of the analog sigma delta modulators.

Because of the finite bit length of the arithmetic units, the inputs to such digital sigma delta modulators are represented by rational numbers. When dc values are applied to the input, such modulators will invariably end up in a limit cycle.

For the single loop sigma delta modulator doubling the sampling rate will result in a 9 dB gain in the signal-to-noise ratio for a variable input signal. However, for a dc input the worst value, given by the amplitude of the noise peaks (4.10), will decrease only with 6 dB. That means, the difference in the performance of the system with a random signal as an input, versus a dc value, will increase with 3 dB for each doubling of the sampling rate. Moreover, for certain applications as voice channels, such limit cycles are perceived as tones and are noticeable at signal levels much lower than the white noise or the shaped noise characteristic to the sigma delta modulators.

For double loop modulators the mechanism of generating the limit cycles is much more complex due to the fact that the period of the limit cycles depends on the initial conditions. If the fundamental frequency is outside of band of interest it will be eliminated by the subsequent filtering of the output of the modulators. There will be no noise contribution from the modulator in this case. This fundamental frequency can be much lower than the sampling frequency as the spectrum from Fig. 9 indicates.

No theoretical results equivalent with those obtained in Section IV, which would permit the characterization of the limit cycles of the double loop modulators have been obtained, so far. The simulation and experimental data seem to indicate that with each doubling of the sampling rate the amplitude of the limit cycles decreases with 10–12 dB. Therefore, the observations made for single loop modulators remain valid in this case.

For analog double loop sigma delta modulators the idle channel condition ($X = 0.5$) can generate such limit cycles. The dithering effect of the noise inherent to the analog circuits can help to eliminate this problem. For designs with very low noise levels such spectrums exists.

VI. Conclusions

Two methods for the study of the limit cycles in sigma delta modulation were presented in this paper.

The first method, described in Section II, permits to determine the necessary conditions for the existence of the limit cycles. The one bit quantizer is characterized by the sign function. The difference equations which describe the circuit behavior, written for every sampling interval of the limit cycle are added together, so that the terms corresponding to the state variables are cancelled. This method was used successfully to more complex architectures like Candy's double interpolative D/A converter [2] and extensions of this architecture [12]. These architectures contain two, respectively, three one-bit quantizers, the range of the output values is not limited only to binary values. The results were identical to that from Section II. The input must be a rational number, which is equal with the sum of the values of the output during the cycle divided by the cycle length.

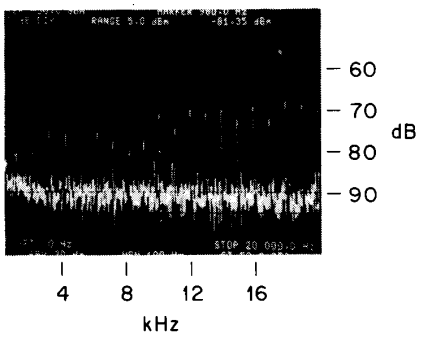


Fig. 9. The spectrum of the limit cycle, in the 0-20 kHz band, at the output of a double loop digital sigma delta modulator. f_s = 1 MHz. The input is idle code (X = 0.5).

In Sections III and IV the structure of the limit cycle was determined for single loop sigma delta modulators and based on it, the spectrum and the noise contribution were computed. This method could be used in the case of idle channel noise in double loop modulators, which presents interest in some telecommunication applications. The applicability of this method is limited by the fact that for double loop modulators the structure of the limit cycle depends on the initial conditions. The simulation results show that this is also true for other sigma delta modulator architectures ([2], [12]).

Appendix

The Fourier Coefficients of the Second-Order Spectra

The structure of the limit cycle for second-order spectra (3.7) is

$$C=[(r_0-1)S_{1,0}]S_{1,1}. \qquad (A.1)$$

The z transforms of the strings $S_{1,0}$, $S_{1,1}$ (3.6) and their corresponding spectra are similar to those of the limit cycle for the first-order spectra.

$$S_{1,l}(z)=\frac{1-z^{-q_0(q_1+l)}}{1-z^{-q_0}}$$

$$S_{1,l}(\alpha_k)=e^{-jq_0(q_1+l-1)\alpha_k\pi}\frac{\sin[q_0(q_1+l)\alpha_k\pi]}{\sin[q_0\alpha_k\pi]} \quad l=0,\ 1 \qquad (A.2)$$

$$\alpha_k=\frac{k}{r_0(q_0q_1+1)+q_0}. \qquad (A.3)$$

Substituting (A.2) in (3.9), taking into account that $l_{1,0} = q_1q_0 + 1$ and that α_k is given by (A.3) and neglecting a phase component the following expression is obtained for the Fourier coefficients:

$$A(k)=\frac{1}{a}\left[\frac{\sin[(r_0-1)(q_1q_0+1)\alpha_k\pi]}{\sin[(q_0q_1+1)\alpha_k\pi]}\frac{\sin(q_1q_0\alpha_k\pi)}{\sin(q_0\alpha_k\pi)}+(-1)^k\frac{\sin[q_0(q_1+1)\alpha_k\pi]}{\sin(q_0\alpha_k\pi)}\right]. \qquad (A.4)$$

For the degeneracy case $q_1 = 1$

$$A(k)=\frac{1}{a}\left[\frac{\sin\left[(r_0-1)(q_0+1)\alpha_k\pi\right]}{\sin\left[(q_0+1)\alpha_k\pi\right]}+(-1)^k\frac{\sin\left(2q_0\alpha_k\pi\right)}{\sin\left(q_0\alpha_k\pi\right)}\right] \tag{A.5}$$

where $\alpha_k = k/r_0(q_0 + 1) + q_0$.

With the substitution

$$\begin{aligned}&\sin\left[(r_0+1)(q_0+1)\alpha_k\pi\right]-\sin\left[(r_0-1)(q_0+1)\alpha_k\pi\right]\\&\quad=2\sin\left[(q_0+1)\alpha_k\pi\right]\cos\left[r_0(q_0+1)\alpha_k\pi\right]\\&\quad=2\sin\left[(q_0+1)\alpha_k\pi\right]\cos\left[(r_0(q_0+1)+q_0)\alpha_k\pi-q_0\alpha_k\pi\right]\\&\quad=2\sin\left[(q_0+1)\alpha_k\pi\right]\cos\left(k\pi-q_0\alpha_k\pi\right)\\&\quad=2(-1)^k\sin\left[(q_0+1)\alpha_k\pi\right]\cos\left(q_0\alpha_k\pi\right).\end{aligned}$$

(A.5) becomes

$$A(k)=\frac{1}{a}\frac{\sin\left[(r_0+1)(q_0+1)\alpha_k\pi\right)}{\sin\left[(q_0+1)\alpha_k\pi\right]}. \tag{A.6}$$

The relation (A.6), with the notations (4.14.a) is identical with (4.7).

For the nondegenerate second-order spectra the dc component is $A(0) = b/a = X$ in (A.4). In computing the noise components, the relation (A.4) can be simplified, using (A.3)

$$\begin{aligned}&(r_0-1)(q_1q_0+1)\alpha_k\pi\\&\quad=\left[r_0(q_1q_0+1)+q_0-q_1q_0-1\right]\alpha_k\pi\\&\quad=k\pi-(q_1q_0+q_0+1)\alpha_k\pi.\end{aligned}$$

With this substitution and neglecting the sign, (A.6) becomes (4.15).

Acknowledgment

The author would like to thank J. C. Candy and E. M. Fields for their helpful advice.

References

[1] J. C. Candy and O. J. Benjamin, "The structure of quantization noise from sigma delta modulation," *IEEE Trans. Commun.*, vol. COM-29, pp. 1316–1323, Sept. 1981.

[2] J. C. Candy and A. Huynh, "Double interpolation for digital-to-analog conversion," *IEEE Trans. Commun.*, vol. COM-34, pp. 77–81, Jan. 1986.

[3] N. van de Weg, "Quantizing noise of a single delta modulation system with an N-digit code," Philips Res. Rep., Oct. 1953, pp. 367–385.

[4] J. E. Flood and M. J. Hawksford, "Exact model for delta modulation process," *Proc. IEEE*, vol. 118, no. 9, pp. 1155–1161, Sept. 1971.

[5] J. C. Candy, B. A. Wooley, and O. J. Benjamin, "A voiceband codec with digital filtering," *IEEE Trans. Commun.*, vol. COM-29, pp. 815–830, June 1981.

[6] T. Misawa, J. E. Iwersen, L. J. Loporcaro, and J. G. Ruch, "Single-chip per channel codec with filters utilizing delta-sigma modulation," *IEEE J. Solid-State Circuits*, vol. SC-16, pp. 333–341, Aug. 1981.

[7] R. Koch and B. Heise, "A 120 KHz sigma delta A/D converter," in *Proc. Conf. Dig. Tech. Papers, 1986 Int. Solid-State Circuits Conf.*, Feb. 1986, pp. 138–139.

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

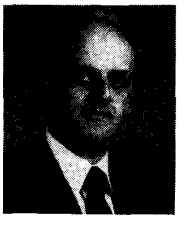
[9] ——, "Spectral analysis of sigma-delta quantization noise," submitted for publication.

[10] I. M. Vinogradov *An Introduction to the Theory of Numbers.* New York: Pergamon, 1961, pp. 7–11.

[11] B. P. Agrawal and K. Shenoi, "Design methodology for ΣΔM," *IEEE Trans. Commun.*, vol. COM-31, pp. 360–370, Mar. 1983.

[12] Y. Matsuya, K. Uchimura, A. Iwata, T. Kobayashi, and M. Ishikawa, "A 16b oversampling A/D conversion technology using triple integration noise shaping," in *Proc. Int. Solid-State Circuits Conf.*, Feb. 1987, pp. 48–49.

★



Vladimir Friedman (M'79) was born in Romania in 1944. He received the Dipl. Eng. degree from Polytechnic Institute of Bucharest in 1966 and M.S.E.E. degree from Polytechnic Institute of New York in 1980.

In 1970 he joined the Telecommunication Research Institute of Bucharest, where he worked in the area of CW wide-band frequency modulated systems. In 1978 he joined ITT Advanced Technology Center, Shelton, CT, being involved in the design of VLSI circuits for switching systems and digital signal processing. Since 1985 he is with AT&T Bell Laboratories, where his present interests are in the area of A/D and D/A conversion using sigma delta modulation.