ΣΔ Modulation is a Mapping

Richard Schreier and Prof. W. Martin Snelgrove
Department of Electrical Engineering
10 King's College Road, University of Toronto
Toronto, Canada. M5S lA4
Tel. (416) 978-3381, (416) 978-4185
FAX (4 16) 978-7423

Abstract

In this paper we take the viewpoint that ΣΔ modulation is a mapping in the mathematical sense. We describe the mapping in mathematical terms and find that it is a projection. We give an explicit formula for the set of all inputs which give rise to any given output pattern. The fact that the mapping is a projection suggests the existence of limit-cycles and we investigate these in both lowpass and bandpass modulators. We take this one step further with an examination of the related phenomenon of amplitude quantization in non-ideal modulators.

1 A General ΣΔ Modulator

The essence of a ΣΔ modulator is captured by the equation

$$Y=GU+HE \qquad (1)$$

where Y, U, and E are the Z-transforms of the output signal, *y*, the input signal, u, and the error signal, e, respectively, and G and *H* are the signal and noise transfer functions. This equation applies to any modulator which does linear operations on *u*, y, or *e*, regardless of whether *e* is shown explicitly or not. Thus this equation covers many of the proposed ΣΔ modulator structures [1,2]. It must be emphasized that the equation is exact, provided *e* is chosen appropriately.

Without too great a loss of generality, we take G=l. If $G\neq 1$, one can take it to be *1* and consider the input to be $g\otimes u$.

For a ΣΔ modulator with the describing equation $Y=U+HE$, one possible realization is shown in Figure 1. We advocate this structure for analysis because it makes ΣΔ easy to understand. A necessary condition for the realizability of a ΣΔ modulator with an error transfer function *H* is that $H-1$ be strictly causal. This condition ensures that the diagram of Figure 1 has no delay-free loops and so is well-posed.

The modulator must also be stable, but at present there is no adequate theory for the stability of this system[3,4,5].

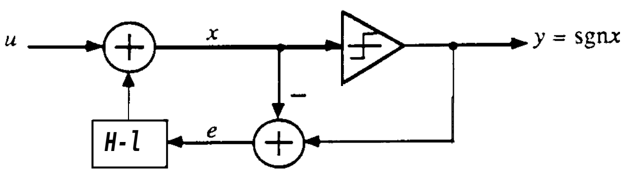

Figure 1: A block diagram of a general ΣΔ modulator with our canonical structure.

2 Properties of the Mapping

A ΣΔ modulator transforms a discrete-time, continuous-amplitude signal into a discrete-time, discrete-amplitude signal. It can thus be viewed as accomplishing a mapping from one signal space to another.

2.1. The mapping is many-to-one and into

One of the fiit and most trivial observations one can make about this mapping is that it is many-to-one. That is, for a given initial state, there are an infinite number of input signals which yield the same output

We say that the mapping is 'into' because the space of discrete-time, discrete-amplitude signals can be considered a subset of the space of discrete-time, continuous-amplitude signals.

2.2. The mapping is idempotent

The second observation is that, for an initial state of zero, the mapping is *indempotent* (*An* operator is said to be idempotent if composition with itself results in the original operator.) This is equivalent to saying that ΣΔ modulation is a projection. Even multi-level ΣΔ modulators have this property.

Proof: We can see from Figure 1 that if the initial state of the $H-1$ block is zero, then $x(0)=u(0)$. If the input is quantized in exactly the same manner as the output, then $y(0)=u(0)$ and $e(0)=0$. Thus the state of the $H-1$ block stays zero and by induction we conclude that $y=x=u$ and $e=0$ for all time. This result falls out quickly because of the way we draw a ΣΔ modulator.

One of the consequences of this property is that it is possible to excite a ΣΔ modulator in such a way that any output pattern is produced. Simply apply the desired pattern to the input.

The assumption that the initial state is zero is a standard one for stable linear systems, wherein the state decays to zero for zero input. In a ΣΔ modulator with zero input, the state is oscillatory and does not decay to zero. Without the zero-initial-state assumption, the mapping would not be idempotent..

2.3. Equivalent Inputs

It was mentioned earlier that many inputs yield the same output. What these signals have in common depends on the modulator under consideration. For an ideal lowpass modulator, H has a zero at DC, so the input and the output must have the same DC value. For a bandpass modulator [6], H has a zero at ω_0, so the input and the output have the same ω_0-component. In general, the mapping preserves some functional, F, of the input, where the form of F is determined by the error transfer function. We can make this explicit by finding the set of all inputs that yield a given output pattern.

We know that applying an output pattern, y, to the input yields y and that in this case x=y. Other inputs which yield y differ from it by amounts small enough such that the sign of the decision sequence, x, does not change.

Consider for a moment the sequence $u(n) = y(n) + \varepsilon h(n)$, where ε is a small number. At time zero $x(0)$ has ε added to it, and this will not change the sign of x iff

$$sgn(x(0)) = sgn(y(0)+\varepsilon) = sgn(y(0)), \tag{2}$$

or

$$\varepsilon \geq -1 \text{ when } y(0) = 1 \text{ and } \varepsilon < 1 \text{ when } y(0) = -1. \tag{3}$$

If this condition holds, then $e(0) = -\varepsilon$. Thereafter, the H-1 block produces $-\varepsilon h$ which precisely cancels the succeeding terms of our perturbation to u. The result is that only one sample of x is altered, and this change is small enough that it does not affect y.

From here it is easy to write a formula for inputs u which yield the output y:

$$u(n) = y(n) - \sum_{i=0}^{\infty} e(i)h(n-i), \tag{4}$$

or

$$u = y - h \otimes e, \tag{5}$$

where

$$e(i) \leq 1 \text{ if } y(i) = 1 \text{ and } e(i) > -1 \text{ if } y(i) = -1. \tag{6}$$

It is possible to show that all inputs which yield y are of this form.

This remarkably simple formula is more than just a trivial rearrangement of the describing equation. It provides profound insight into the behaviour of a $\Sigma\Delta$ modulator.

A traditional interpretation of (5) is that the output is the input plus a filtered error. If the error is assumed to be white and uniformly distributed in $[-1,+1]$, so that its power spectral density is $\frac{1}{3}$, then one can make use of (5) to compute the noise power of y in the band $[\omega_1, \omega_2]$:

$$N^2 = \frac{1}{\pi}\int_{\omega_1}^{\omega_2} \frac{1}{3}\left|H(e^{j\omega})\right|^2 d\omega. \tag{7}$$

This is the method that is used almost universally in analytical calculations of the signal-to-noise ratio.

If we make no assumptions about e, other than its power is below an amount P_e, then we can bound the in-band noise power with

$$N^2 \leq \max_{\omega\in[\omega_1,\omega_2]} \left|H(e^{j\omega})\right|^2 P_e, \tag{8}$$

which shows that as long as P_e is not too large, the in-band noise will be small.

An alternative interpretation of (5) begins with a restriction on the size of e. If we insist that $e(i) \in (-1, 1]$, then the condition (6) is satisfied for any y-pattern. In addition, $|x(i)| \leq 2$ and so the modulator stays stable. Thus (5) can be used to find all the inputs which keep the modulator stable, where stable is defmed to mean $|x(i)| \leq 2$. This set is given in [5] as:

$$U_1 = \left\{ \begin{array}{l} u \mid u = y - (h \otimes e), \\ \quad \text{where } y(n) = \pm 1 \text{ and } e(n) \in (-1, 1] \end{array} \right\}. \tag{9}$$

'For these inputs the modulator is guaranteed to work, in the sense that the error signal is always less than 1 and the bound (8) holds with $P_e = 1$. Any other input will overload the modulator, at least momentarily. If one could find the smallest signal which is not in the set U_1, then one would know the stable input range of the modulator. Clearly, a good understanding of the set U_1 is of practical significance.

3 Limit-cycles and Amplitude Quantization

We know that if one were to excite a $\Sigma\Delta$ modulator with a periodic binary-valued signal, then the output would follow the input. Since many inputs can yield that output sequence, it is conceivable that a steady signal, perhaps the F-component of the periodic signal, would give rise to the original periodic signal. We see by this argument that the projection property of $\Sigma\Delta$ suggests the existence of limit-cycles.

3.1 Ideal Modulators

A great deal is known about the limit-cycles in the ideal first-order and second-order lowpass $\Sigma\Delta$ modulators. Gray [7] explored the behaviour of the first-order modulator with a rational input and showed that the output is always periodic. Friedman [8] did likewise, but also showed that for a rational input, the output of a second-order modulator will only be periodic for particular initial states.

What happens in the general case is more uncertain, but some universal statements can be made. For a steady input to produce a limit-cycle, the input and output must have the same F-component. Thus for a modulator with a zero of the error transfer function at DC, the average value of the input must equal the average value of the limit-cycle, which is a rational number. If the modulator does not have a zero at DC, then the input need not be rational to give rise to a limit-cycle.

In the bandpass case, it is not clear that limit-cycles will exist. Since a bandpass modulator typically has a zero at ω_0, the limit-cycles of interest are those with a non-zero ω_0-**component.** Since the input must have the same ω_0-**component,** the purest such input is a sine wave of frequency ω_0.

Consider a second-order bandpass modulator described by the error transfer function $H = 1 - \sqrt{2}z^{-1} + z^{-2}$. It has a zero at $\omega_0 = \frac{\pi}{4}$ so the period of a limit-cycle with a non-zero ω_0-component must be a multiple of 8. Some possible limit-cycles and their ω_0-components are shown in Table 1. To determine whether or not the supposed pattern can be produced by the modulator, it is necessary to get the modulator in the correct state. We found, by simulation, that it is possible to get the modulator in approximately the right state by initially applying the desired limit-cycle to the input and then slowly decreasing its non-ω_0-components to zero. Table 1 shows the test patterns for which this process succeeded in exposing a limit-cycle. It is possible to verify these results analytically using the discrete Fourier transform if x is assumed to be periodic.

Pattern	ω_0-component	Seen in simulation?
+ + + + + + + −	0.5000	No
+ + + + + + − −	0.9239	No
+ + + + + − − −	1.2071	Yes
+ + + + − − − −	1.3066	Yes
+ + + − − + − −	0.7071	Yes

Table 1: Some possible limit-cycles in a second-order bandpass modulator ($\omega_0 = \frac{\pi}{4}$).

We see that bandpass modulators, like their lowpass counterparts, can indeed exhibit limit-cycles. As a result, one can expect to find all the limit-cycle related phenomena that are known to exist in lowpass modulators, especially the existence of in-band tones which degrade the signal-to-noise ratio for particular inputs.

It is worthy of note that in contrast to the lowpass case, a bandpass modulator can produce an output that is an exact representation (in a narrow band) of an input whose peak value exceeds 1, the peak value of the output. In the case of the modulator above, the largest such signal has a peak value of approximately 1.3. In general, the limit is the magnitude of the ω_0-component of the signal $\mathrm{sgn}(\sin(\omega_0 t))$. At $\omega_0 = \frac{\pi}{4}$, the limit is $\sqrt{2}$ and it decreases to $\frac{4}{\pi}$ as $\omega_0 \to 0$. Of course, the rms value of such a signal cannot exceed 1.

3.2 Non-ideal modulators

In a modulator with a zero at ω_0 a limit-cycle with a non-zero ω_0-component can only exist for specific input phases and amplitudes. If $H(e^{j\omega_0}) \neq 0$, i.e. the modulator is not ideal, then a range of input amplitudes and phases can sustain a limit-cycle. Figure 2 displays the ω_0-component of the output sequence as a function of the amplitude of a sine input for ideal and perturbed versions of the second-order bandpass modulator. The perturbed modulators have their zeros moved radially by a factor $1 - \delta$, and Figure 2 uses the relatively large values of $\delta = \pm 0.1$ to make the non-ideal behaviour readily apparent.

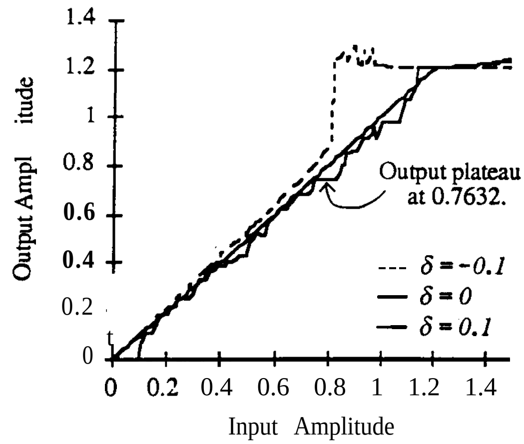

Figure 2: Output ω_0-**component** as a function of input amplitude for ideal and non-ideal bandpass modulators with a sine input

We see that the curves for $\delta > 0$ and $\delta < 0$ are markedly different. For positive δ, where the zeros move inward, the output amplitude tends to increase in steps, whereas for negative δ the curve is much smoother. This occurs because limit-cycles are stable if and only if H^{-1} is stable.

Proof: Suppose y is a limit-cycle corresponding to an input u. Then we can remove the quantizer from Figure 1 and supply y as an input, as shown in Figure 3. The $\Sigma\Delta$ modulator thus becomes a linear system with two inputs u and y.

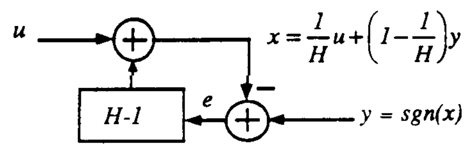

Figure 3: A $\Sigma\Delta$ modulator with the comparator removed and the output considered to be an input.

The transfer functions from u to x and y to x are H^{-1} and $1 - H^{-1}$, respectively. If H^{-1} is unstable, then small perturbations will cause x to change by increasingly large amounts, until y is no longer sgn(x) or x becomes infinite. If H^{-1} is stable, small perturbations will not be magnified to the point where the sign of x is changed, and so the limit-cycle will survive small perturbations.

We can illustrate the use of discrete Fourier transforms, alluded to earlier, in the calculation of the width of the output amplitude plateau at 0.7632. From the time-domain simulations used to produce Figure 2, we find that the limit-cycle has a length of 24 samples:

$$y = + + + + - + - - + - + + - - - - + - + + - + - - \quad (10)$$

The input is a sine wave of amplitud A and frequency $\omega_0 = \frac{\pi}{4}$:

$$u(n) = A\sin(\omega_0 n), \tag{11}$$

and we fmd x via:

$$x = Ax_1 + x_2 \tag{12}$$

where

$$x_1 = H^{-1} \bullet \sin(\omega_0 n), \tag{13}$$

and

$$x_2 = (1 - H^{-1}) \bullet y. \tag{14}$$

The x_1-sequence is a sine wave with an amplitude of $\left|H(e^{j\omega_0})\right|^{-1}$ and a phase of $-\arg\left(H(e^{j\omega_0})\right)$. *The* x_2-sequence is easily computed using the discrete Fourier transform.

For consistency, we must have $y(n) = \operatorname{sgn}(x(n))$ for all n, and this condition determines the rang $[A_1, A_2]$ for which this limit-cycle will exist. For each n where $y(n) = \operatorname{sgn}(x_1(n))$, A must be larger than a certain minimum value, $-\frac{x_2(n)}{x_1(n)}$. The minimum value o A is the largest of these lower limits:

$$A_1 = \max_{n \mid y(n) = \operatorname{sgn}(x_1(n))} \left(-\frac{x_2(n)}{x_1(n)}\right) \tag{15}$$

Likewise the maximum value o A is:

$$A_2 = \min_{n \mid y(n) \neq \operatorname{sgn}(x_1(n))} \left(-\frac{x_2(n)}{x_1(n)}\right) \tag{16}$$

For the limit-cycle under consideration, this procedure yields the A-range [0.*7438,0.8343*], which matches the simulation results from Figure 2.

We know that when $\left|H(e^{j\omega_0})\right| = 0$, the width of the stable range of the limit-cycle is zero. When $\left|H(e^{j\omega_0})\right| \neq 0$, the width of the stable range is not zero, but is roughly proportional to $\left|H(e^{j\omega_0})\right|$. This suggests the possibility of using limit-cycles to measure the quality o H, and so facilitate tuning of the modulator to reduc $\left|H(e^{j\omega_0})\right|$.

4 Conclusions

We have examined ΣΔ modulation from a mathematical viewpoint and have seen that it is a projection. We have been able to derive the set of all inputs which correspond to any given output pattern, for a general class of ΣΔ modulators. This formula can be used to justify the traditional frequency domain arguments used in the analysis of ΣΔ modulators and to determine the inputs which keep the modulator stable. We have also investigated limit-cycles, and found that they are stable only when H^{-1} is stable. This suggests that choosing H *so* that H^{-1} is unstable may rid a modulator of troublesome limit-cycles. In addition, we demonstrated that limit-cycles can exist in bandpass modulators and that they lead to amplitude quantization when H^{-1} is stable. We have shown *how* to quantify this behaviour and suggested its use in tuning a ΣΔ modulator.

Acknowledgment

Thanks to Steve Jantzi for taking the time to discuss these ideas and to preview this document. Your help is greatly appreciated.

References

[1] J.C. Candy, "A Use of Double Integration in Sigma-Delta Modulation, " *IEEE Transactions on Communications,* vol. COM-33, pp. 249-258, March 1985.

[2] W.L. Lee and . "A Topology for Higher Order Interpolative Coders, *Proceedings of the 1990 IEEE International Symposium on Circuits and Systems,* pp. 459-462, May 1987.

[3] W.L. Lee, "A novel higher order interpolative modulator topology for high resolution oversampling A/D converters," Master's thesis, Massachusetts Institute of Technology, pp. 34-36, June 1987.

[4] B.P. Agrawal and K. Shenoi, "Design Methodology for ΣΔM", *IEEE Transactions on Communications,* vol. COM-31, no. 3, pp. 360-369, March 1983.

[5] R. Schreier and W.M. Snelgrove, "Stability in a General ΣΔ Modulator, " *presented at the 1991 International Conference on Acoustics, Speech and SignalProcessing,* Toronto, May 1991.

[6] R. Schreier and W.M. Snelgrove, "Bandpass Sigma-Delta Modulation, "*Electronics Letters*, vol. 25, no. *23,* pp. 1560-1561, Nov. 9 1989.

[7] R.M. Gray, "Oversampled Sigma-Delta Modulation," *IEEE Transactions on Communications, vol. COM-35,* no. 5, pp. 481-489, May 1987.

[8] V. Friedman, "The Structure of the Limit Cycles in Sigma Delta Modulation," *IEEE Transactions on Communications,* vol. *COM-36,* no. 8, pp. 972-979, August 1988.