Noise-Shaped Coding

by

Richard Schreier

A Thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy in the University of Toronto.

Richard Schreier

University of Toronto

Department of Electrical Engineering

Noise-shaped Coding

ABSTRACT

The motivation for this work comes from the exciting development of circuits which use ΣΔ modulation to achieve hitherto impossible levels of performance: analog-to-digital converters with 20 bits of linearity, ***untrimmed,*** have been reported. Despite this impressive success, there remains room for improvement in both the circuits which implement them and the methods used to design them.

This document makes several contributions to ΣΔ modulation. The first is a simplified way of looking at how and why ΣΔ modulation works. Using this simplified description, the results from this work and those of others are presented in a very easy-to-understand format.

The second contribution is the development and proof-by-simulation of bandpass ΣΔ modulation. This technique extends the application range of ΣΔ modulation by allowing the frequency band of interest to be centered on any frequency, not just zero. These modulators achieve high oversampling ratios without requiring that the sampling rate be much greater than the upper frequency of interest.

The third major contribution is an examination of noise-shaped coding from a mathematical perspective. It is shown that, under certain conditions, ΣΔ modulation is an idempotent, surjective mapping, and a formula is given for the equivalence classes induced by this

mapping. In addition, a necessary condition for limit-cycle stability is proven, suggesting that modulators can be designed which are free from the undesirable effects of limit-cycles.

Finally, the problem of ΣΔ modulator stability is studied. A formula for the set of all inputs which keep any modulator stable is given, but it defies attempts to use it as a test for stability. Previous researchers have proposed simple tests for stability, but it is shown that these are neither necessary, nor in some cases sufficient. Modification of one of these criteria yields the most general analytical test for stability known at this time. In addition, a numerical test for stability is proposed, and its predictions are compared with simulations for a very special set of ΣΔ modulators.

Acknowledgments

There are many who made this work possible. Firstly, thanks go to my supervisor, Martin Snelgrove, for introducing me to the wondrous world of sigma-delta and for giving me a free hand in the conduct of this research. I would not have done it if you had not been here.

Secondly, David Johns and Steve Jantzi deserve thanks for their willingness to listen to me ramble on about my latest breakthrough, even though it turned out disappointingly often that I was wrong. My enthusiasm would have run out long ago had it not been for our many conversations.

Thirdly, Bruce Francis is greatly appreciated for his careful reading of Appendix A and his comments on terminology.

Finally, I would like to thank those who supported me during my studies. The generous financial support of both the Natural Sciences and Engineering Research Council and Bell-Northern Research removed all financial worries. The loving emotional support of Paddy Duncan and our two families has been a source of strength in times of need.

Thanks to you all.

Table of Contents

0 Notation

This chapter collects the definitions of the common symbols used in subsequent chapters in one convenient place. It is not expected that the reader fully understand the meanings of all these quantities until they are presented in the main part of the text.

f_s sampling frequency, in Hertz.

ω normalized angular frequency, $\omega = \pi$ corresponds to $f_s/2$.

z z-transform variable: z^{-1} is a unit delay and $z = e^{j\omega}$ for physical frequencies.

t time (discrete), an integer.

ω_B normalized bandwidth of the band-of-interest.

R the oversampling ratio, $R = \pi/\omega_B$.

sgn() the signum function, $\mathrm{sgn}(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases}$

u the input sequence, $u = (u(0), u(1), u(2)\ldots)$, $u(t) \in \Re$.

x the decision sequence, i.e. the input to the quantizer.

y the output sequence. For the bulk of this work, $y(t) = sgn(x(t)) = \pm 1$.

e the error sequence, $e = y - x$.

g the impulse response of the signal transfer function.

h the impulse response of the error transfer function, $h(0) = 1$ and $y = g * u + h * e$.

U, X, Y, E, G, H the z-transforms of the corresponding sequences, $U(z) = \sum_{t=0}^{\infty} z^{-t} u(t)$.

$\|H\|_2$ the 2-norm of H, $\|H\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} \left|H(e^{j\omega})\right|^2 d\omega$.

$\|h\|_2$ the 2-norm of h, $\|h\|_2^2 = \sum_{t=0}^{\infty} h(t)^2$. Parseval's Theorem states that $\|h\|_2^2 = \|H\|_2^2$.

$\|h\|_1$ the 1-norm of h, $\|h\|_1 = \sum_{t=0}^{\infty} |h(t)|$.

$\|H\|_\infty$ the ∞-norm of H, $\|H\|_\infty = \max_{\omega \in [0, 2\pi]} \left|H(e^{j\omega})\right|$.

$\|e\|_\infty$ the ∞-norm of e, $\|e\|_\infty = \sup_{t \geq 0} |e(t)|$.

$h * e$ the convolution of h with e, $(h * e)(t) = \sum_{i=0}^{\infty} h(i) e(t - i)$.

He a concise notation for the above. He is the output of a linear system with transfer function $H(z)$ and input e. This form is avoided unless its brevity is needed.

1 Introduction

Noise-shaped coding is another name for sigma-delta, or delta-sigma, modulation. The latter terms are historically more popular, but have become misnomers now that modulators have evolved beyond the original single-loop and double-loop structures. In addition, the technique of error feedback in digital filters [Diniz and Antoniou 1985] and the half-toning of digital images [Anastassiou 1989] are essentially equivalent to sigma-delta modulation in that they all seek to reduce quantization noise in one frequency band at the expense of the remaining frequencies. The title "Noise-Shaped Coding" was chosen to reflect this fact.

Sigma-delta ($\Sigma\Delta$) modulation forms the foundation of, and provides the motivation for, this work. It is therefore necessary for the reader to understand sigma-delta modulation: what it is, how it works and why it is interesting. The full depth and breadth of the $\Sigma\Delta$ modulation literature is beyond the scope of this document and so this chapter presents only a selection of this literature designed to give the reader an adequate context for the results of this thesis.

1.1 First-Order Sigma-Delta Modulator

The simplest $\Sigma\Delta$ modulator is the first-order lowpass modulator, shown in Figure 1.1 [Candy and Benjamin 1981]. The input, u, is a discrete-time, continuous-amplitude (analog) signal; the output, y, is a discrete-time, binary-valued signal. For convenience, we scale the circuit so that the output of the comparator is ± 1. Example waveforms are shown in Figure 1.2.

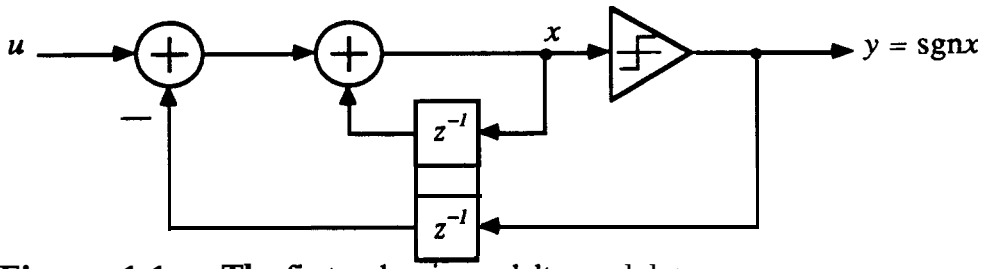

Figure 1.1: **The first-order sigma-delta** modulator.

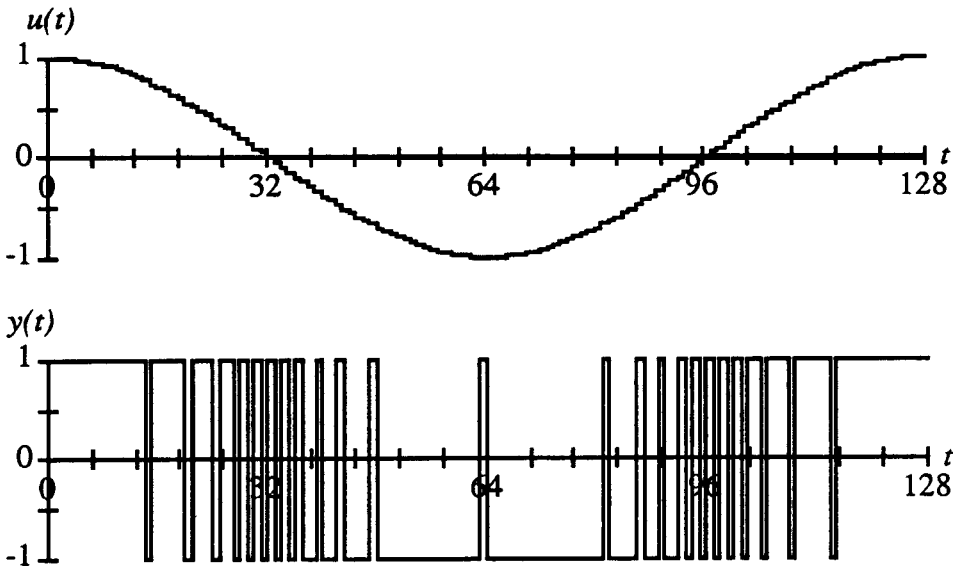

Figure 1.2: Input and output waveforms for a first-order lowpass sigma-delta modulator with a sine wave input.

Referring to Figure 1.1, we see that the circuit computes the difference, or A, between the input and the delayed output, then feeds the result to a discrete-time summer, Σ, the output of which feeds a comparator to produce y. From this description it is readily apparent how the name $\Sigma\Delta$ reflects the structure of the first-order **modulator.**[1]

The transformation from input to output is necessarily non-linear, but it nonetheless preserves the low-frequency content of *u.* To see how it accomplishes this feat, consider a constant input. It can be readily shown (see Section 4.5.1) that for inputs in the range

[1] The etymology is contentious, see [Hauser 1981].

[- 1,+ 1], the output of the summer, x, is bounded. In order for this to happen, the input to the summer must have an average value of zero. This in turn requires that the average value of the output equals the input. Thus, in the case of a constant input in [-1,+ 1], we can recover the input exactly by finding the average of the output.

The averaging operation can be done digitally, and this is essentially all it takes to make an analog-to-digital converter out of a $\Sigma\Delta$ modulator. Some advantages of this approach are:

1) The analog circuitry is trivial. All that is required are a discrete-time summer and differencer, a unit delay, and a comparator. These are readily available in switched-capacitor form, so integrating the modulator is easy.

2) The conversion is inherently linear. Since the transformation from two digital codes to two analog levels exactly fits a straight line no matter what the analog levels are, the binary quanitzer can only possess gain and offset errors. In audio applications, these errors are not as objectionable as nonlinearities which induce harmonic distortion [Hauser 1991]. In practice, nonlinearities in the analog components limit the linearity of the converter. Nonetheless, the achievable linearity is high enough that a20-bit A-to-D has recently been reported [DelSignore, Kerth, Sooth, and Swanson 1990].

3) The anti-aliasing requirements are lessened by oversampling. We shall see that the sampling rate is much greater than the upper frequency of interest, and that the high frequencies are removed by post-filtering. The attenuation requirements of the anti-aliasing filter apply only to frequencies which alias down to the band of interest, and since these frequencies are remote from that band, the anti-aliasing filter's specifications can be relaxed.

4) Accuracy increases quickly with sampling rate. We shall see that for an n^{th} order converter, each doubling of the sampling rate adds $n + \frac{1}{2}$ bits to the accuracy of the converter.

A more complete understanding of the first-order modulator can be had by a little analysis. If the quantizer is modeled as an additive noise source, so that y = x + e, the circuit becomes a two-input linear system. The output may then be written as the sum of two terms, one from the input, u, and one from the error, e:

$$y = u + h * e \tag{1.1}$$

h is the impulse response of the noise transfer function; its z-transform is H(z) = $1 - z^{-1}$. In contrast, the signal transfer function is unity. Thus we see that the output is equal to the input plus an error term whose spectrum is ***shaped*** by H. ***H*** has a zero at DC, so at low frequencies the error term is small and consequently the first order lowpass modulator preserves the low frequency content of u. Applying a digital lowpass filter to y removes the bulk of the noise, yielding a high-quality digital representation of the input; see Figures 1.3 and 1.4.

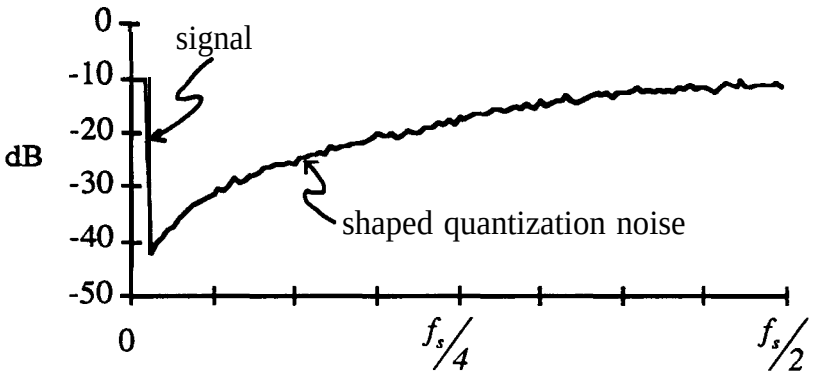

Figure 1.3: An FFT of the output of a first-order modulator ***with*** a -10dB input spread uniformly across the band $[0, \frac{f_s}{64}]$. The shaping of the noise spectrum is clearly evident.

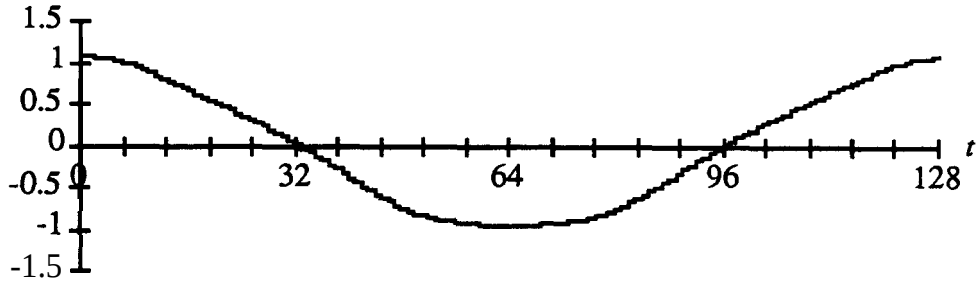

Figure 1.4: The output waveform of Figure 1.2 filtered by an ideal digital lowpass filter with a cutoff just above one-thirty-second of the sampling frequency. This cotresponds to an oversampling ratio of 16.

To quantify the conversion quality, we need to know something about the spectrum of e. It turns out that e can have a wide variety of spectra, but it is nevertheless customary to assume that e is white with a power σ_e^2. With this assumption, one can immediately write the noise power in the frequency band $[0, \frac{\pi}{R}]$ as:

$$N_0^2 = \frac{\sigma_e^2}{\pi} \int_0^{\frac{\pi}{R}} \left| H\left(e^{j\omega}\right) \right|^2 d\omega. \tag{1.2}$$

R is the ***oversampling ratio;*** $1/R$ is the fraction of the spectrum we are interested in, the remainder acting as a dumping ground for the noise energy. R is typically large, on the order of 100, so we can approximate

$$\begin{aligned} \left| H\left(e^{j\omega}\right) \right|^2 &= \left| 1 - e^{-j\omega} \right|^2 \\ &= \left| 1 - \cos\omega + j\sin\omega \right|^2 \\ &= (1 - \cos\omega)^2 + \sin^2\omega \\ &\approx \omega^2 \qquad \text{for } \omega \ll 1. \end{aligned}$$

Substituting this in equation (1.2) yields

$$N_0^2 = \frac{\sigma_e^2}{\pi} \frac{1}{3} \left(\frac{\pi}{R} \right)^3 = \frac{\sigma_e^2 \pi^2}{3R^3},$$

and we arrive at the startling result that doubling R reduces the noise power by a factor of eight. With an oversampling ratio of 100 and ideal filtering on y, we can expect rms noise levels on the order of 10^{-3}, or -60dB.

This result is succinctly stated in the 9dB/octave rule: in a first-order ΣΔ modulator, the SNR increases by 9dB (lOlog8), or 1.5 bits ($\frac{1}{2}\log_2 8$), for every octave increase in *R.*

Figure 1.5 plots the simulated SNR of a first-order ΣΔ modulator as a function of the oversampling ratio, *R. The* input signal was a sine wave with a peak amplitude of *A=O.* 1, 20dB below full-scale. The signal and in-band noise powers were determined from an FFT of the Hann-windowed output. As can be seen in the figure, the SNR thus determined does roughly follow the 9dB/octave rule. The discrepancies are mostly due to the inaccuracy of our e-is-white-noise assumption.

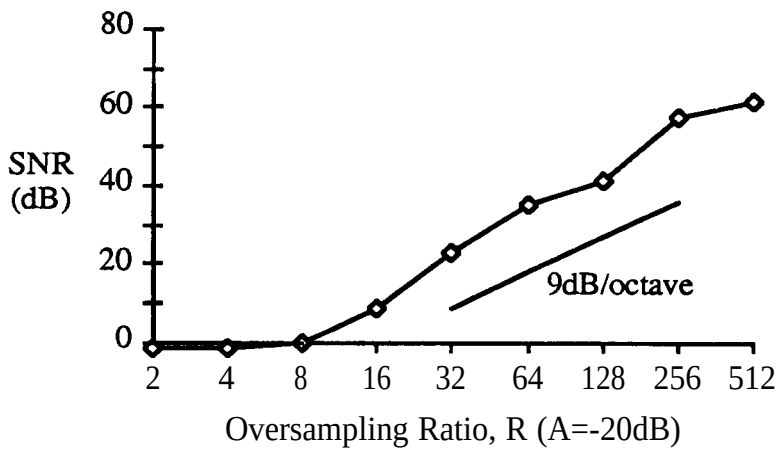

Figure 1.5: The signal-to-noise ratio of the first-order modulator increases by approximately 9dB when the oversampling ratio is doubled.

It is precisely the lack of the validity of this whiteness assumption that is responsible for one of the major drawbacks of first-order modulators. It has been found that, in practice, the error signal has a highly-colored, discrete spectrum which results in disturbing whistles and squeaks [Candy 1974][Norsworthy 1990]. In particular, it has been shown that a

constant input of value $\frac{a}{b}$, where a and b are relatively-prime integers, can result in a repetitive output where the period is a multiple of b [Friedman 1988]. If b is large, an appreciable amount of tonal energy can be located in the band-of-interest, and this may very well be audible.

One way to break up such limit-cycles is to randomize the quantizer slightly, for example by adding a small amount of noise (dither) to its input [Chou and Gray 1990]. This noise is added at the same point in the loop as e, so it too gets shaped by H, and consequently degrades the SNR only marginally. Another method for producing a whiter e-signal involves the use of a finer quantizer. By giving the quantizer more levels, the error signal becomes smaller and whiter. The main drawback of this method is that a multi-bit quantizer is likely to have appreciable non-linearity and so the modulator would no longer be inherently linear. A more satisfying solution uses a higher-order modulator to produce error signals with much whiter spectra.

Before we move on to fancier modulators, it is worthwhile to note that ΣΔ modulators can also be used to make digital-to-analog converters. In this application, the modulator is digitaI, with a PCM digital signal replacing the analog input and a two-level analog output. Analog filters are needed to do the requisite lowpass filtering of the output. Figure 1.6 shows a digital ΣΔ modulator in which some simplifications have been made [Candy and Huynh 1986]. The comparator operation is accomplished simply by stripping off the most significant bit, and the remaining bits form the negative of the error signal. Thus the output may be written by inspection as $Y = U + E - z^{-1}E = U + (1 - z^{-1})E$, so that the error transfer function is $1 - z^{-1}$. This is exactly the same as in the modulator of Figure 1.1, and it should come as no surprise that a simulation of one is identicaI to a simulation of the other.

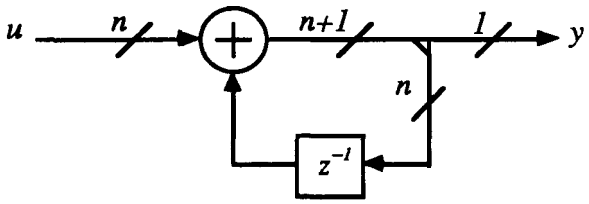

Figure 1.6: A digital first-order lowpass sigma-delta modulator. This structure can be used to make a digital-to-analog converter.

This is true in general- a simulation of a ΣΔ modulator intended for an A-to-D produces the same output as a simulation of a similar ΣΔ modulator intended for a D-to-A, if the inputs are identical. This last condition needs to be stated because, typically, the former modulator samples a constantly-changing analog signal, whereas the latter repeatedly samples a low-speed digital signal which does not change with each tick of the modulator's clock. Nevertheless, when simulating a modulator it is not necessary to know whether it is for an A-to-D or a D-to-A: the simulation models are identical.[2]

1.2 Second-Order Sigma-Delta Modulator

An important step in the evolution of sigma-delta modulation was the invention of the double-loop modulator shown in Figure 1.7 [Candy 1985]. By observing that a comparator is a kind of analog-to-digital converter, one might consider replacing it with a better kind of converter. Figure 1.7 results from replacing the comparator in Figure 1.1 with a first-order modulator.

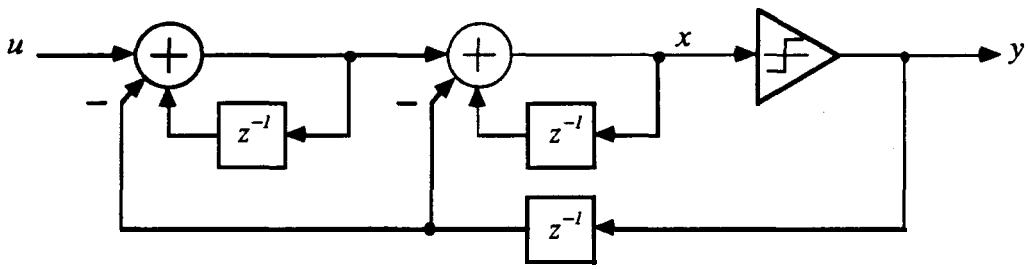

Figure 1.7: A second-order lowpass sigma-delta modulator.

---

[2] If we ignore non-idealities such as round-off noise and sensitivity,

Proceeding as before, we model the quantizer with an additive noise source and find that

$$y = u + h * e$$

where h now has a z-transform of $H(z) = \left(1 - z^{-1}\right)^2$. Assuming the noise is white, with a power σ_e^2, the noise power in the band of interest is

$$\begin{aligned} N_0^2 &= \frac{\sigma_e^2}{\pi} \int_0^{\frac{\pi}{R}} \left| H\left(e^{j\omega}\right) \right|^2 d\omega \\ &\approx \frac{\sigma_e^2}{\pi} \int_0^{\frac{\pi}{R}} \omega^4 d\omega, \qquad R \gg \pi \\ &= \frac{\sigma_e^2 \pi^4}{5R^5} \end{aligned}$$

and we see that an octave increase in R now reduces the noise power by a factor of 32. Stated another way, the second-order modulator achieves a 15dB (2.5 bit) increase in the SNR per octave increase in R – 6dB (1 bit) per octave better than the first-order modulator.

Experience with real second-order modulators has shown that they are much less susceptible to limit cycles and tonal behavior than first-order modulators [Candy 1985][Candy and Huynh 1986]. This is thought to be a result of the x-signal appearing more random because of the extra complexity in the feedback loop.

Experience has also shown that a second-order modulator is not as robust as a first-order one. The double-loop modulator is slow to recover from inputs which go outside the [-l,+l] range. In addition, the noise power begins to increase at a lower level of input signal than in the single-loop modulator. A plot of the SNR versus the amplitude of the input, Figure 1.8, shows this phenomenon as a flattening out and eventual peaking of the SNR curve just below 0dB.

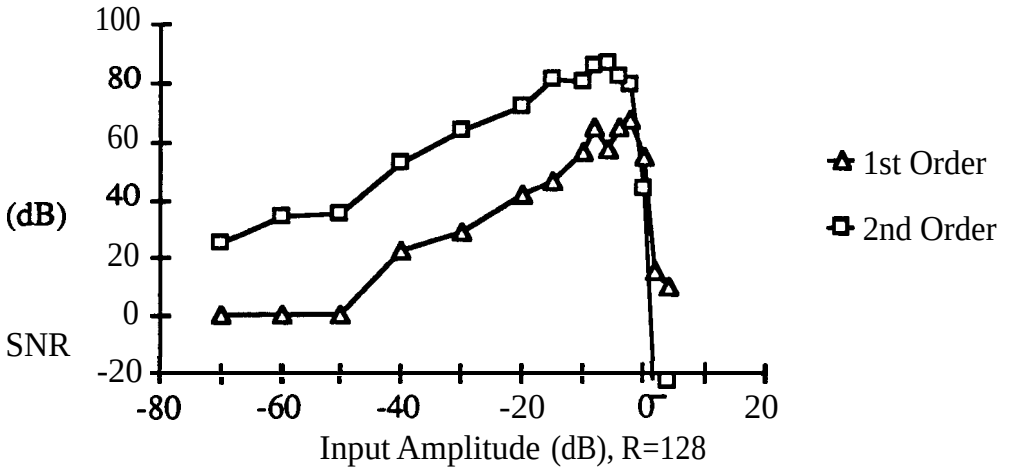

Figure 1.8: The signal-to-noise ratio of the first- and second-order modulators as a function of the amplitude of the input signal. The input limit on the second order modulator is a few dB less than that of the first-order modulator.

1.3 Higher-Order Sigma-Delta Modulators

We have seen that the error transfer function of the first-order modulator is $1-z^{-1}$ and that of the second order modulator is $(1-z^{-1})^2$. Given the superiority of the second-order modulator, one would naturally like to continue this progression and try to build a third order modulator with an error transfer function $(1-z^{-1})^3$.

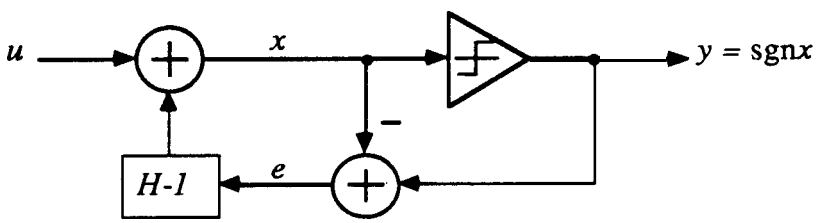

Figure 1.9: A sigma-delta modulator with an arbitrary error transfer function, H.

Figure 1.9 shows one way to build a ΣΔ modulator with an arbitrary error transfer function, H. By inspection,

$$Y=X+E$$

$$X=U+(H-1)E$$

$$\therefore Y=U+HE$$

This diagram has no delay-free loops if the $H-1$ block is strictly causal[3], and so the diagram is well-posed (in the mathematical sense) if this condition is met. In addition, this condition is necessary for the realizability of a ΣΔ modulator with the given error transfer function. Consequently the diagram is capable of realizing any ΣΔ modulator which possesses a realization.

The structure of Figure 1.9 is, because of its generality, readily suited to analysis and so will be used repeatedly. Its simplicity makes ΣΔ easier to understand than the complicated topologies often shown in the literature. There is no reason not to use this structure in a simulation since its input-output behavior is identical to any other structure with the same noise transfer function and a signal transfer function of unity.

With this structure one can build a third order modulator, but simulation reveals it to be unstable: internal signals become arbitrarily large and the spectrum of the output does not resemble the input. This holds true for any binary modulator with an error transfer function of the form $(1-z^{-1})^n$ for n 2 3. At first glance, it would appear that high-order modulators do not work.

Fortunately, this is not so. One way to make a high-order modulator stable is to use a multi-level quantizer [Hawksford 1985], although it may then suffer from linearity problems. A more satisfactory solution is to allow the use of a more general transfer function [Chao et al. 1990].

Until now, we have been dealing with FIR transfer functions: those with only a z^n in the denominator. We gain design freedom if the poles are allowed to be arbitrary. Placing the

---

[3] Requiring the $H-1$ block to be strictly causal is equivalent to requiring that $H-1$ be strictly proper, $h(0)=1$ or $H(\infty) = 1$.

poles on top of the zeros is guaranteed to produce a stable zeroth-order modulator, and it is reasonable to expect that perturbing this system slightly, for example by moving the poles towards the origin, would not make it unstable. This is indeed the case: putting all the poles at $z=0.5$ results in an apparently stable third-order modulator which has 21dB/octave performance, as shown in Figure 1.10. **As** with the second-order modulator, the maximum input amplitude is slightly below 0dB.

It has been found [Chao et al. 1990][Hein and Zakhor 1991] that high-order modulators are susceptible to self-sustaining, large-amplitude, low-frequency oscillations. These oscillations are detrimental to the operation of the converter since they consume dynamic range and cause the error signal to be large. In practice, most high-order single-bit modulators employ some form of limiting or even resetting circuitry to extinguish such oscillations.

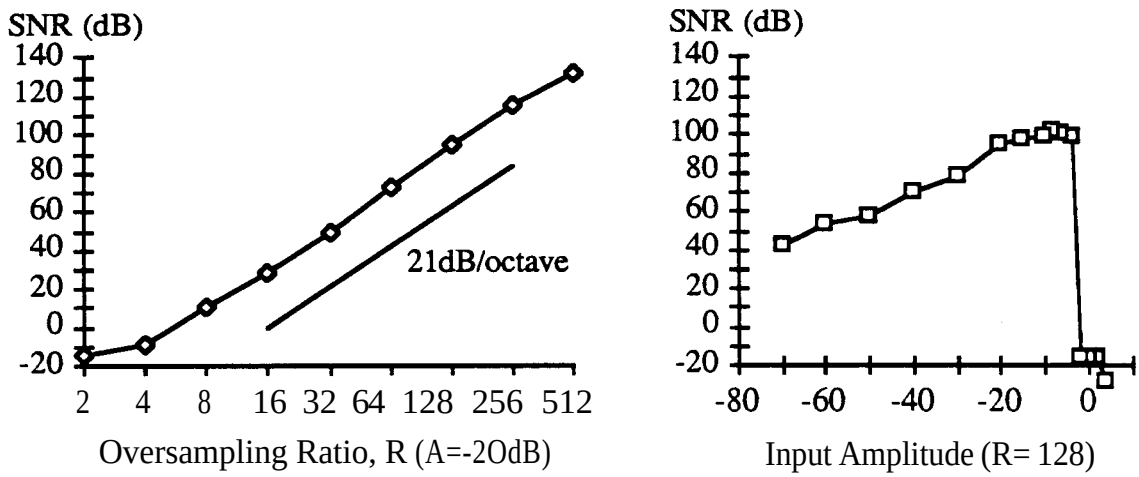

Figure 1.10: **SNR** plots for a stable third-order modulator, $H(z) = \frac{(z-1)^3}{(z-0.5)^3}$.

An &-order lowpass modulator with an error transfer function of the form

$$H(z) = \frac{(z-1)^n}{D(z)}$$

may be approximated near $\omega=0$ as

$$|H(e^{j\omega})| \approx \frac{\omega^n}{|D(1)|}$$

so the noise power is approximately

$$N_0^2 \approx \frac{\sigma_e^2 \pi^{2n}}{|D(1)|^2 (2n+1) R^{2n+1}}$$

and we see from this that the SNR ought to increase by $(6n+3)$dB, or $n+\frac{1}{2}$ bits for each doubling of R.

For a particular R and n, we can actually do better than this formula predicts by allowing the zeros to spread across the entire band. To make the best modulator of a given order, H must be chosen to minimize N_0^2, subject to the realizability condition and a stability[4] constraint. Lee [Lee 1987] claims that if $\|H\|_\infty$, the maximum gain of the error transfer function, is less than 2, then the modulator will be stable. Agrawal and Shenoi use the rule $\|H\|_2^2 < 3$, where $\|H\|_2^2$ is the power gain of H [Agrawal and Shenoi 1983]. If either rule were reliable, the design of nearly optimal $\Sigma\Delta$ modulators of any order would be routine: simply use an optimization program to minimize the white-noise-based estimate of N_0^2 while satisfying the realizability and stability constraints. Unfortunately, these rules are not completely reliable[5] and designers must resort to lengthy simulations to confirm the stability of their designs.

The resultant transfer function can be realized in a number of ways, including that of Figure 1.9. The choice would be guided by sensitivity and noise considerations. The

---

4 A definition of stability which ensures that the modulator essentially works as desired is that a modulator is stable if and only if x, or equivalently e, is bounded. Stability is discussed at length in Chapter 4.

5 A further disadvantage of these rules is that they are too conservative. Both rule out the most popular modulator, the second-order modulator.

cascade-of-resonators structure [Ferguson, Ganesan and Adams 1990] has been shown to be a strong contender [Jantzi, Schreier and Snelgrove 1991].

1.4 Other Sigma-Delta Modulators

The discussion of ΣΔ modulation has thus far focussed on its purest incarnation: a single-bit quantizer embedded in an otherwise linear system. It is the purpose of this section to make the reader aware of some variations of the basic modulator, their advantages and their drawbacks.

1.4.1 Continuous-Time Filtering

The discrete-time filter in the forward path of the modulator can be replaced by a continuous-time filter, or a mixed continuous/discrete-time filter [Del Signore, Kerth], Sooch and Swanson 1990]. The input to the modulator is then a continuous-time analog signal, and the act of sampling occurs inside the loop. In general, such a mixed modulator can be modelled by an equivalent discrete-time modulator fed by the sampled output of a continuous-time filter. For example, if the discrete-time integrator of Figure 1.1 is replaced by a continuous-time integrator whose output is sampled by the comparator, an equivalent system is an ordinary first-order ΣΔ modulator fed with the sampled output of a moving average filter, illustrated in Figure 1.11.

Thus the continuous-time approach allows one to incorporate a crude anti-aliasing filter in the modulator itself. The disadvantage of this approach is that the modulator is now sensitive to the behavior of the implicit digital-to-analog converter in the feedback path over the entire sampling period, and not just to its value at the end of the sampling period. In particular, care must be taken to ensure that the effect of the second positive pulse in a pair of successive positive pulses is the same as that of the first, lest this be a source of distortion.

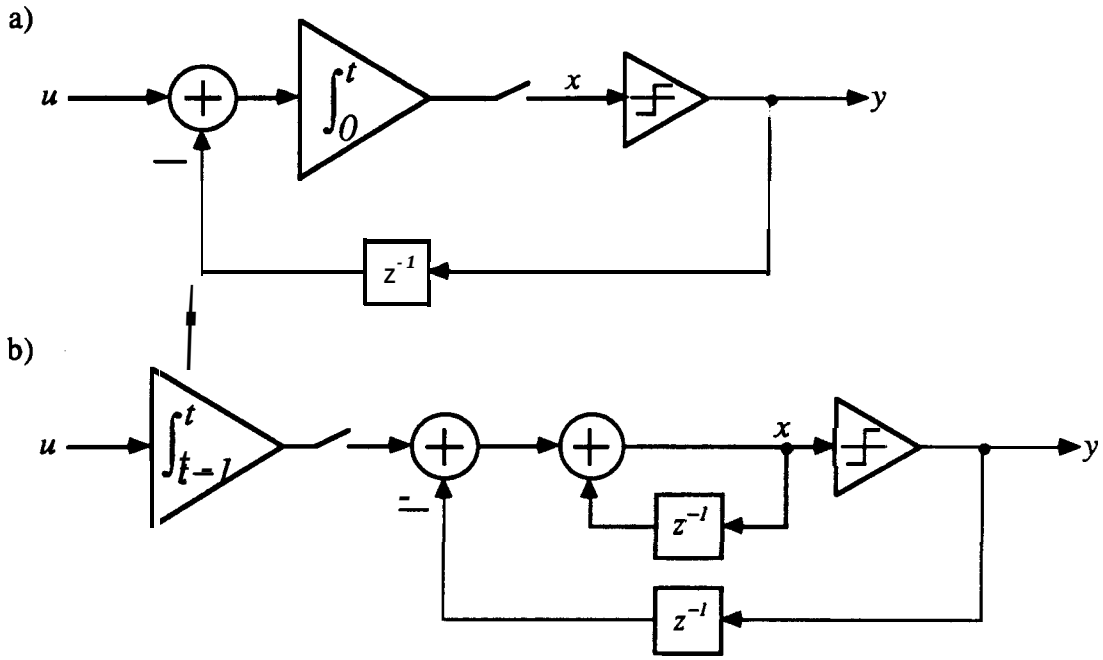

Figure 1.11: a) A first-order modulator with a continuous-time integrator in place of the discrete-time integrator, and b) an equivalent system employing a discrete-time modulator. The sampling rate is assumed to be 1 Hz.

1.4.2 The MASH Modulator

Historically, the first improvement to the basic modulator was the development of the Multistage Noise Shaping (MASH) Modulator, shown in Figure 1.12 [Uchimura, Hayashi, Kimura and Iwata 1988]. The basic idea is to use a second $\Sigma\Delta$ modulator to digitize the error signal of the first. The output of this secondary modulator can be digitally filtered and subtracted from the primary modulator's output to leave quantization noise which has been shaped by the product of two noise-shaping functions. This is the two-stage MASH modulator.

One can increase the number of stages to any desired level by repeatedly digitizing the error signal in the last modulator. The resulting cascade produces a stable modulator with an order equal to the sum of the orders of the individual modulators.

The primary disadvantages of this approach are that it relies on the precise cancellation of terms derived from two separate circuits, one analog and one digital, and that there is added complexity on the digital side.

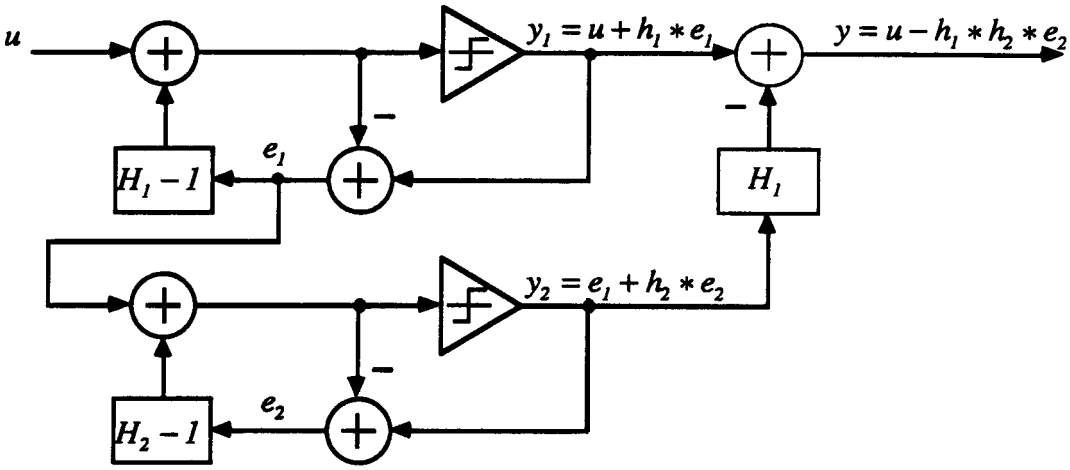

Figure 1.12: A blockdiagramof a two-StageMASHmodulator.

1.4.3 Innovations with Multi-bit. Quantizers

The use of a multi-bit quantizer in a **ΣΔ** modulator is desirable since it reduces the size of the error signal and consequently makes the SNR higher. An added benefit, as mentioned in Section 1.3, is that high-order modulators are easier to design with multi-bit quantizers since the circuit better approximates a linear system. The problem is that these quantizers may possess differential nonlinearity which degrades the performance of the modulator.

The differential nonlinearity of a multi-bit quantizer can be compensated for by digital means [Larsen, Cataltepe and Temes 1988]. If one regards the discrete levels of the quantizer as being exact, then it is a simple matter of using the correct digital representation of that level on the digital side. These corrected codes can be determined during a self-calibration cycle.

A second innovation uses a multi-bit A-to-D, but only single-bit feedback [Leslie and Singh 1990]. This hybrid approach does not suffer from nonlinear distortion, and its performance has been shown to approach that of an ideal system with a full n-bit quantizer. Figure 1.13 illustrates the structure of this interesting modulator.

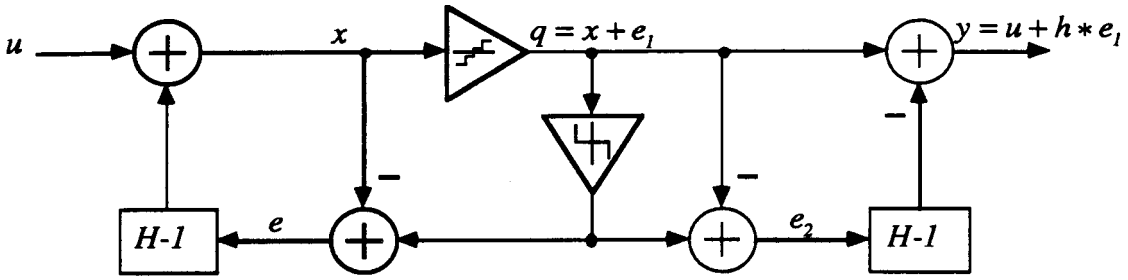

Figure 1.13: A modulator employing a multi-bit A-to-D and single-bit feedback.

From the diagram,

$$X = U + (H - 1)E$$
$$Q = X + E_1$$
$$E = E_1 + E_2$$
$$Y = Q - (H - 1)E_2$$

whence

$$Y = U + HE_1.$$

Thus the noise-shaped residual of the multi-bit A-to-D is the only error term in the output. Consequently this structure achieves the same performance as a full multi-bit modulator without endangering the inherent linearity property of the basic modulator. Unfortunately, it suffers from the same drawbacks as the MASH modulator: increased complexity on the digital side and the need to match a digital transfer function to an analog one.

1.4.4 Nonlinear supporting circuitry

A particularly intriguing possibility in the development of ΣΔ modulators is the use of additional nonlinear elements in the circuit. The modulator in Figure 1.14 has been shown to be more tolerant of overload conditions than the regular double-loop modulator [Stonick, Rulla, Ardalan and Townsend 1990]. Moreover, loose bounds on the α, λ and μ parameters have been found which guarantee the stability of this structure.

This circuit was derived from the standard second-order modulator by toying with its describing equations. It is up to the designer to select values for the α, λ and μ parameters which result in satisfactory performance. These two facts make it unlikely that this method can be made into a general design procedure, and so the design of nonlinear feedback circuits for ΣΔ modulators is lacking a solid foundation.

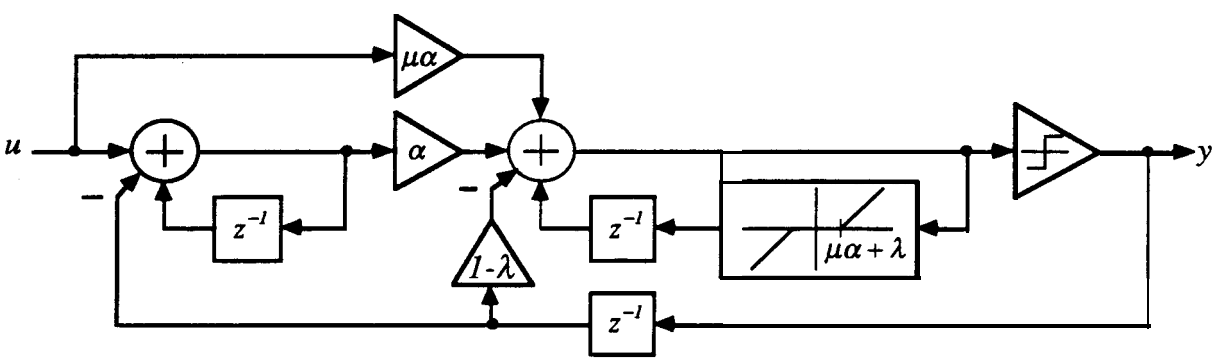

Figure 1.14: A modulator employing a second nonlinear element.

1.5 The Art of Analysis

A critical requirement of the circuit design process is the ability to analyze the circuits under consideration. Since a ΣΔ modulator is a nonlinear system, analysis can be done with varying degrees of correctness. The simplest analysis technique pretends that the quantizer is just a source of additive white noise; the most complete analysis treats the quantizer exactly. At the present time, there does not exist an analysis technique which combines the

insightful simplicity of the former with the accuracy of the latter. The designer must find a way to cope in the absence of an adequate theory.

1.51 The Linear Model

Replacing the quantizer operation, $y = \text{sgn}\, x$, with the linear equation $y = x + e$ is an exact transformation, provided $e = \text{sgn}\, x - x$. The approximation occurs when it is assumed that e has certain convenient properties, such as a white spectrum or a uniform distribution over $[-1, +1]$. With these assumptions, the modulator becomes a simple linear circuit upon which one can apply the powerful techniques of linear circuit analysis. This method has already been used to predict the in-band noise for several modulators, and the results have agreed surprisingly well with simulations.

The problem is that such analyses are predicated upon a lie and so lead to inconsistencies. A linear analysis of the first-order modulator indicates that the noise power is independent of the amplitude of the input signal. However, it is clear from the plot of SNR against input amplitude, Figure 1.8, that the noise power increases drastically as the input exceeds 0 dB. The linear model cannot predict this important phenomenon.

Another pitfall of the linear model is that it leads to a paradox. Figure 1.15 illustrates what happens when we insert a linear gain, $k > 0$, in front of the quantizer. The behavior of the circuit doesn't change, but the linear model does! The signal transfer function is no longer unity and the noise transfer function has different poles. Despite this apparent disaster, it is encouraging that the zeros of the noise transfer function stay the same, giving us some hope that the noise-shaping property is robust.

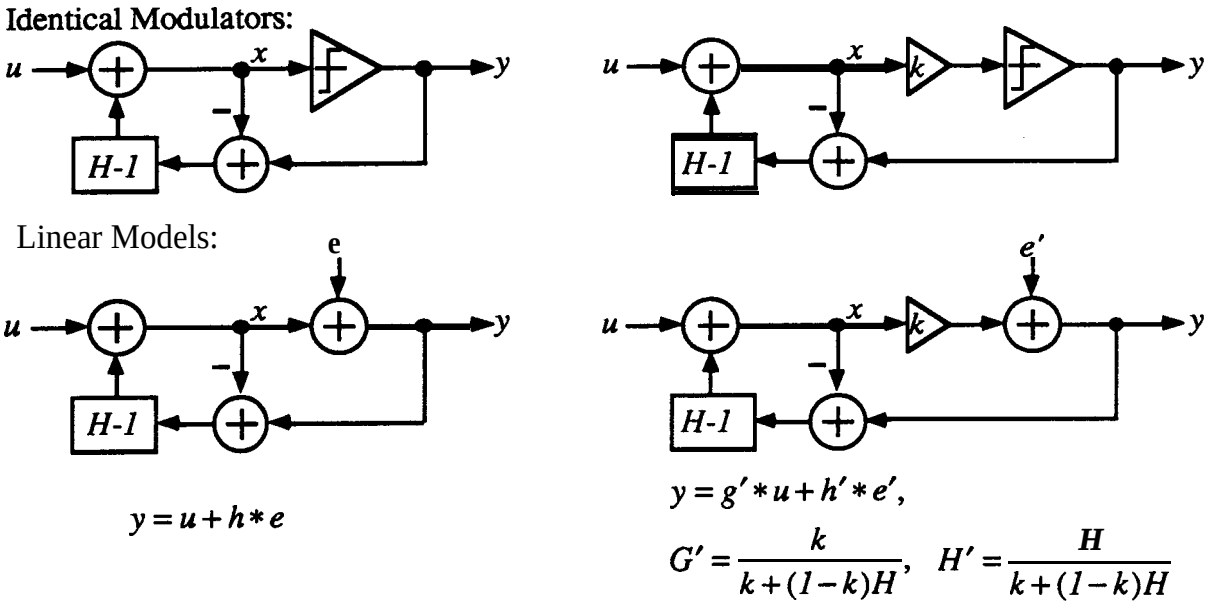

Figure 1.15: **The quantize-gain paradox: placing** a gain element in front of the quantizer leads to a different linear model.

The best linear model of a sigma-delta modulator is one that tries to minimize the error term. The minimum of the mean-square value of e occurs when e is uncorrelated with x. Mathematically,

$$
\begin{aligned}
& \mathrm{E}(ex) = 0 \\
\Leftrightarrow \quad & \mathrm{E}\big((y - kx)x\big) = 0 \\
\Leftrightarrow \quad & \mathrm{E}\big(x\,\mathrm{sgn}x - kx^2\big) = 0 \\
\Leftrightarrow \quad & k = \frac{\mathrm{E}(|x|)}{\mathrm{E}\big(x^2\big)}
\end{aligned}
$$

We see that the optimal ***k*** for analysis depends on the statistics of x. Consequently the computation of ***k*** will, in general, require simulation. Even worse, the statistics of x depend on the input, so we are back to having to do simulations for each input of interest. It is not surprising that most linear analyses simply assume that $k=1$.

1.5.2 The Describing-Function Method

A more advanced model of the quantizer uses a quasi-linear method based on the describing-function method [Kochenburger 1950]. Various forms of this method have been developed, see for example [Smith 1966] or [Atherton 1981]. Ardalan and Paulos have achieved good results with a moderately sophisticated variant [Ardalan and Paulos1987J. Figure 1.16 illustrates the application of their method to our general ΣΔ modulator.

The input to the comparator, x, is decomposed into two parts: the signal component, x_0, and the noise component, x_1. The signal component is proportional to the input, u, whereas the noise component is uncorrelated with u. The quantizer is modelled as two separate linear gains plus a noise source, n. *The* gains are chosen to minimize the mean-square value of n and consequently depend on the input. The result is that the modulator can be separated into two linear systems: one for the signal component and one for the noise component.

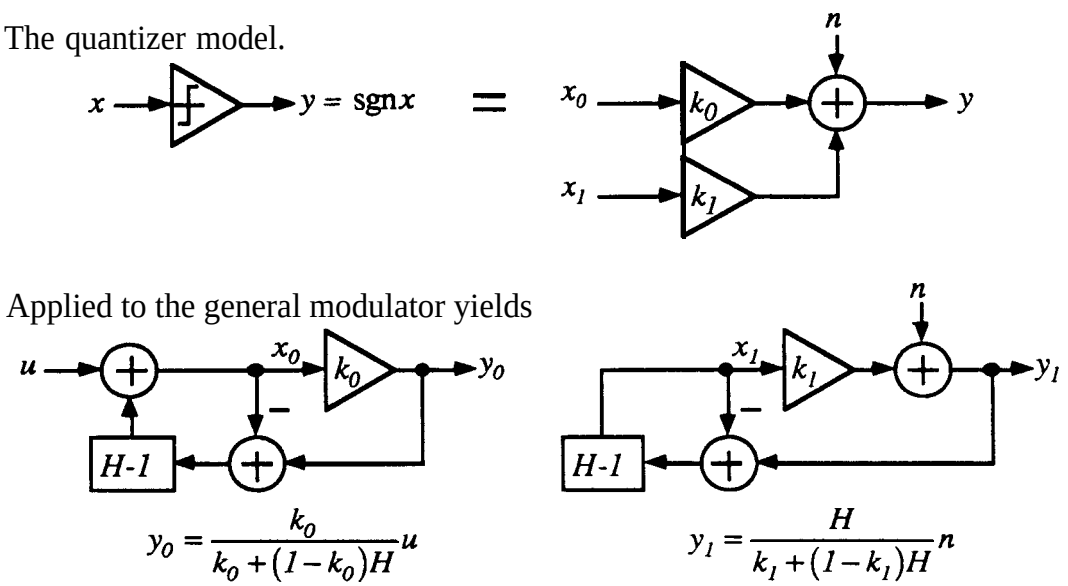

Figure 1.16: The describing-function method applied to the general modulator yields two linear systems: one for the signal and one for the noise.

With the aid of a few assumptions, such as x_1 having a Gaussian probability density function, Ardalan and Paulos were able to derive how the noise power depends on the input signal for constant and sine-wave inputs. This allowed them to predict the shape of the SNR versus input amplitude curve, and to find conditions on the input required for the stability of a second and a third-order lowpass modulator. The main problems with this approach are that the analysis is specific to a given modulator, that it is difficult even with simple inputs, and that it results in only approximate answers. In particular, the conditions for stability are not completely rigorous.

1.5.3 **Exact Analyses**

Gray et al. have been able to perform exact analyses for specific inputs, DC and sine wave, on two specific modulators: first-order lowpass [Gray 1987][Gray 1989][Gray, Chou and Wong 1989] and second-order MASH [Wong and Gray 1990]. Although completely correct, these feats of algebra are too narrow and too opaque to be practical. Knowing the behavior for one input is of little help in understanding the behavior for the sum of two inputs because the circuits are not linear. As well, the analyses are specific to particular modulators and difficult to generalize to others. He, Kuhlmann and Buzo [1990] have succeeded in analyzing the second-order modulator with two-bit quantization, but an exact analysis of the single-bit double loop modulator is still unavailable.

1.6 Summary of Sigma-Delta Modulation

Sigma-delta modulation offers a clever technique for making inherently linear digital-to-analog converters. It employs fast clock rates and digital signal processing to achieve reduced sensitivity to analog components. Modulators with excellent performance are continually being reported.

Despite this clear success, basic theoretical understanding of $\Sigma\Delta$ modulation is lacking. Exotic high-order modulators can be designed, but one must rely on lengthy simulations to verif'y their stability and to determine their dynamic range.

2 Bandpass Sigma-Delta Modulation

We have seen that a ΣΔ modulator can convert an analog signal to a digital one while preserving the low-frequency content. In this chapter, noise-shaped coding is extended by demonstrating that noise-shaping modulators can be designed to preserve the content in any narrow band of frequencies. This enables us to achieve large values of ***R*** while keeping the sampling rate well below 2R times the upper frequency of interest. The resulting system is dubbed a ***bandpass sigmu-delta modulator.***

2.1 The Bandpass Leap

A basic premise of ΣΔ modulation is that the sampling rate is much greater than the highest frequency of interest present in the input. This is necessary because ordinary (lowpass) ΣΔ modulators have zero quantization noise only near DC. If one were instead to null quantization noise at a non-zero frequency, say ω_0, then one would expect to obtain good accuracy near ω_0. Figure 2.1 contrasts the pole/zero placement of the error transfer functions for lowpass and bandpass ΣΔ modulators, and highlights their respective passbands.

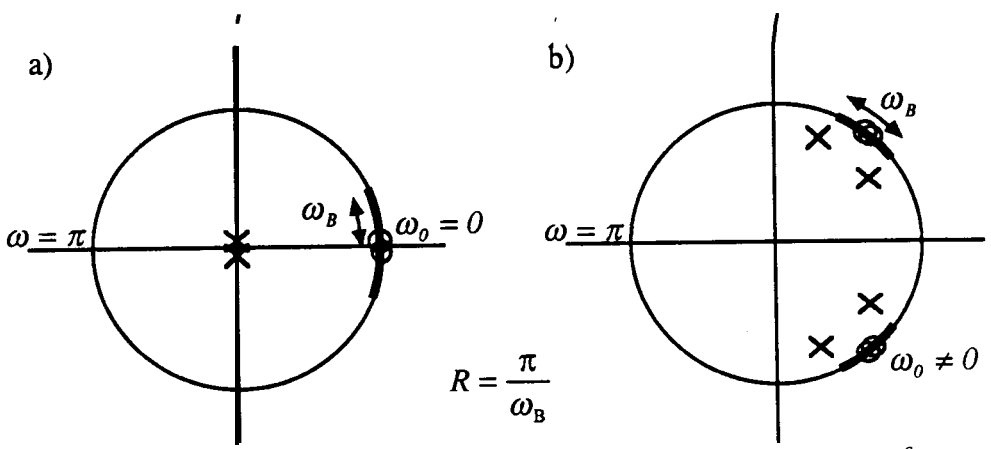

Figure 2.1: Comparison of the pole and zero placements of error transfer functions for (a) an ordinary second-order *lowpass modulator and (b) a fourth-order* bandpass ΣΔ modulator.

The oversampling ratio, R, is defined such that R=l corresponds to a passband which spans the entire unit circle, i.e. $R = \pi/\omega_B$. This is consistent with the definition in the lowpass case, provided the bandwidth, ω_B, is defined to include only positive frequencies.

2.2 Simulation Examples

With the above realization and the knowledge of the first chapter, it is a straightforward task to propose possible bandpass ΣΔ modulators. All we need are candidate noise transfer functions. This section presents the simulation results for two simple noise transfer functions:

$$H_1(z) = 1 - \sqrt{2}z^{-1} + z^{-2} \tag{2.1}$$

$$H_2(z) = \frac{\left(z^2 - \sqrt{2}z + 1\right)^2}{z^4 - 2.1757z^3 + 2.3077z^2 - 1.3054z^1 + 0.3846} \tag{2.2}$$

The first is second-order with single zeros at $z = 1/\sqrt{2} \pm j/\sqrt{2}$, i.e. $\omega_0 = \pi/4$, and two poles at z=0. This modulator is analogous to a first-order lowpass modulator in the sense that the zeros have a multiplicity of one. The second modulator is fourth-order, with zeros of multiplicity two, and so corresponds to a second-order lowpass ΣΔ modulator. For the fourth-order modulator, stability requires that the poles be moved away from the origin. The poles were chosen such that $\|H\|_\infty = 1.625 < 2$, in accordance with the maximum gain stability criterion [Lee 1987] [Chao, Nadeem, Lee and Sodini 1990].

Figure 2.2 plots the signal-to-noise ratio as a function of the input signal amplitude for the above modulators. The band-of-interest is centered on $\omega_0 = \pi/4$, and has a width of $\pi/512$. The SNR was determined by injecting a single tone at the lower edge of the passband for 2^{18} samples and then taking a Harm-windowed FFT of the output to find the ratio of the tone power to the in-band noise power.

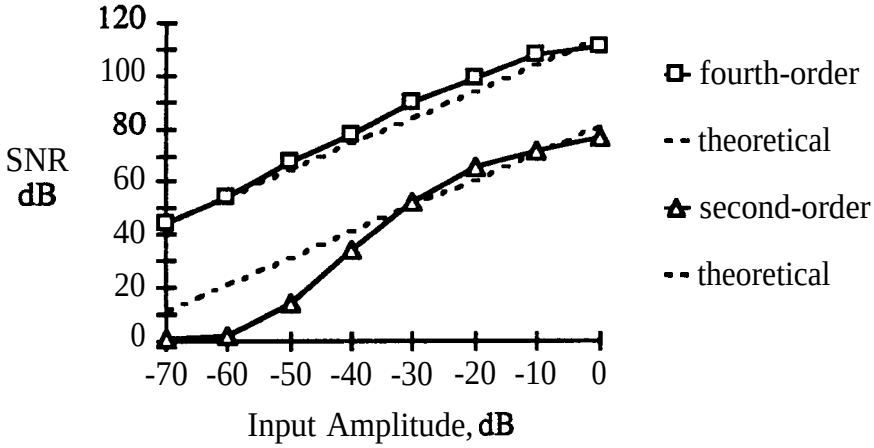

Figure 2.2: SNR as a function of input amplitude for the two candidate bandpass modulators; $R=512$. A sine wave of unit amplitude corresponds to 0dB.

These simulations may be compared to theoretical predictions by proceeding as in Section 1.3. The magnitude of the noise transfer function near a zero of multiplicity m is approximately

$$\left|H\left(e^{j\omega}\right)\right| \approx k\left(\omega-\omega_0\right)^m \tag{2.3}$$

where

$$k=\frac{1}{m!}\frac{d^m\left|H\left(e^{j\omega}\right)\right|}{d\omega^m}. \tag{2.4}$$

Thus the noise power in the band $\left[\omega_0-\pi/2R,\omega_0+\pi/2R\right]$ is

$$\begin{aligned} N_0^2 &\approx \frac{1}{\pi}\int_{\omega_0-\frac{\pi}{2R}}^{\omega_0+\frac{\pi}{2R}} \frac{1}{3}\left|H\left(e^{j\omega}\right)\right|^2 d\omega \\ &\approx \frac{k^2\pi^{2m}}{3(2m+1)(2R)^{2m+1}} \end{aligned} \tag{2.5}$$

where it has been assumed that the error signal is uniformly distributed over $[-1,+1]$ and white, so that its power spectral density is $1/3$. For a bandpass modulator of order n the zeros have multiplicity $m=n/2$ and so the in-band noise power is

$$N_0^2 \approx \frac{k^2 \pi^n}{3(n+1)(2R)^{n+1}}. \qquad (2.6)$$

For the second- and fourth-order modulators under consideration, $k = \sqrt{2}$ and 13, respectively. The dotted lines in Figure 2.2 plot the analytical prediction (2.6) along with the results from simulation. We see that the linear analysis predicts the performance of the fourth-order modulator well except when the input is large. This indicates that the in-band power of the error signal increases when the input is large- a phenomenon which, as we have seen, the linear analysis does not model. For the second-order modulator, the predictions go especially awry when the input is small. This indicates that the error spectrum is distinctly colored, with a large amount of in-band energy. This lack-of-whiteness is reminiscent of the fiit-order lowpass modulator. The fact that we don't see this behavior in the fourth-order bandpass modulator is consistent with the fact that the second order lowpass modulator tends to possess whiter error spectra than the first-order one.

These simulations show that it is indeed possible to design bandpass $\Sigma\Delta$ modulators. As a numerical example, take the clock rate to be 8MHz. Then the fourth-order modulator achieves 17 bits of linearity for a band-of-interest with a center frequency of 1MHz and a bandwidth of 8kHz. These numbers suggest that it may be possible to build a high-quality analog-to-digital converter for narrow-band RF signals, such as AM radio, without the use of a mixer and a conventional A-to-D[Jantzi, Schreier and Snelgrove 1991].

2.3 Decimation and AM Demodulation

In an analog-to-digital converter, the modulator is only half the system. The other half is a digital filter which converts the high-speed bit-stream into a multi-bit output at the Nyquist rate, as shown in Figure 2.3. In lowpass $\Sigma\Delta$ modulation, the first stage of filtering and decimation can be done with a sinc^k filter fashioned out of very simple logic: a counter and

k-1 adders [Candy 1986]. For bandpass ΣΔ modulation, we need to do narrowband filtering on a high-speed bit-stream. At first glance this appears to be a major hurdle but it is the purpose of this section to show how this feat is accomplished.

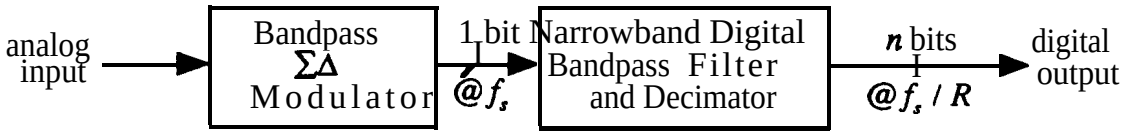

Figure 2.3: An analog-to-digital converter based on a bandpass ΣΔ modulator.

2.3.1 Complex Modulator and Complex Lowpass Filter

The band of interest can be modulated down to DC using a complex modulator, as shown in Figure 2.4. If we implement this scheme directly, two high-speed n-bit multipliers would be needed for the modulator, and as the lowpass filter no longer operates on a bit stream, it would have to be a full-fledged digital filter

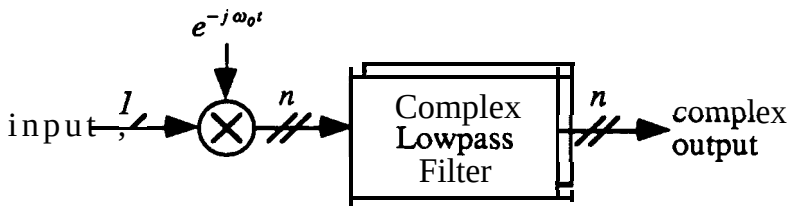

Figure 2.4: A complex modulator followed by a complex lowpass filter The double-slashes denote two n-bit streams.

2.3.2 Simplifications

The complex modulator can be simplified if ω_0 is a simple fraction of the sampling rate. For then the modulation sequence is periodic and we can replace the multipliers with a small look-up table.

As an interesting special case, we see from Figure 2.5 that if we choose $\omega_0 = \pi/4$, the sine and cosine sequences have a very simple structure: each term is either 0, ± 1 or $\pm 1/\sqrt{2}$.

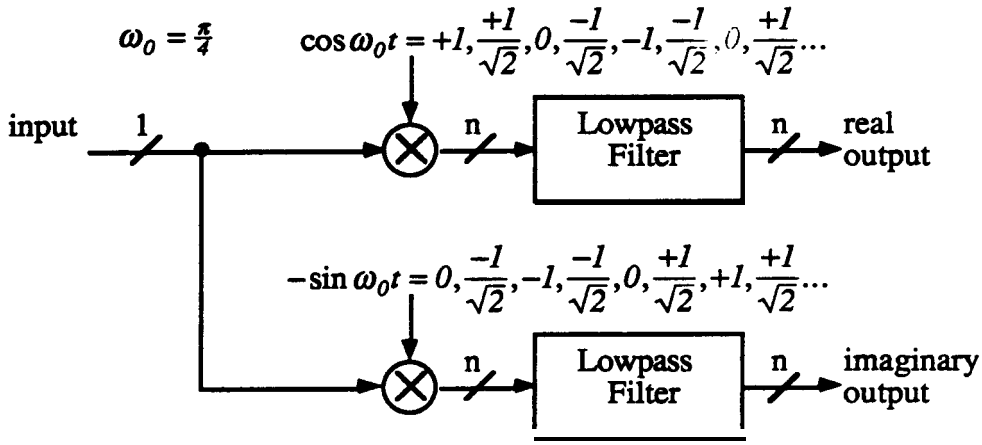

Figure 2.5: Modulator and filter split into real and imaginary parts.

The streams can thus be split in two, one containing integers and the other containing integers times $1/\sqrt{2}$:

$$\sin(\tfrac{\pi}{4}t) = \left(0, \tfrac{1}{\sqrt{2}}, 1, \tfrac{1}{\sqrt{2}}, 0, \tfrac{-1}{\sqrt{2}}, -1, \tfrac{-1}{\sqrt{2}} \ldots\right)$$
$$= (0,0,1,0,0,0,-1,0\ldots) + \tfrac{1}{\sqrt{2}} \times (0,1,0,1,0,-1,0,-1\ldots)$$

Using the linearity property of the multiplication and filtering operations, we can separate the streams into rational and irrational parts, as illustrated in Figure 2.6. Note that the first stage of decimation is now done by lowpass filters operating on a bit stream, just as in the

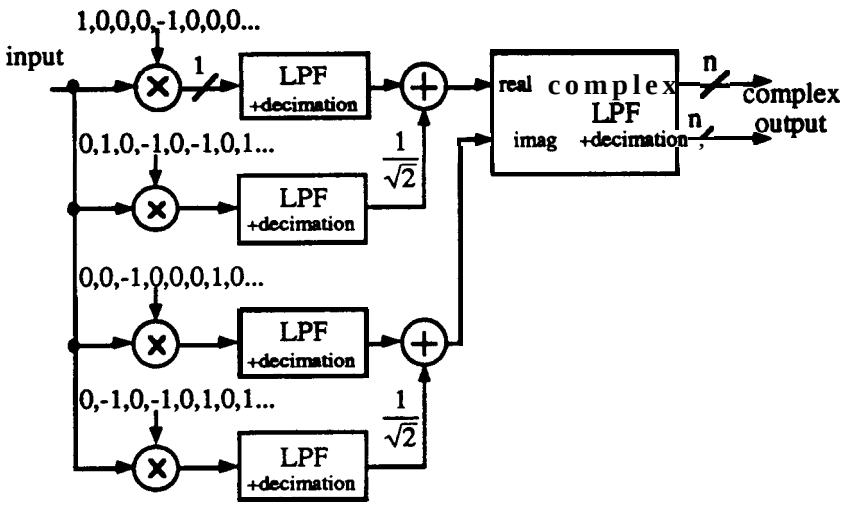

Figure 2.6: The simplified complex modulator and lowpass filter (LPF).

case of lowpass modulation. Also note that the multiplicands are now just 0 and ± 1, and can be effected by simple boolean operations. The last stage of decimation requires a more sophisticated filter, but it is now feasible because it operates at a low speed.

From Figure 2.6, it appears that four preliminary lowpass filter/decimators are now needed, but more simplifications are possible. The presence of so many zeros suggests multiplexing the hardware of a reduced number of lowpass filters. If it were not for the overlap of non-zero terms in the irrational streams of sine and cosine, a single lowpass filter could be multiplexed to implement this structure.

By using the sum and difference of the irrational streams their overlap is eliminated and thus multiplexing a single filter becomes possible, as shown in (2.8) below. The digital signal processor implementing the fmal stage of filtering and decimation is then responsible for calculating the values of thereal and imaginary inputs from the values in the four sets of registers in the primary decimator.

$$\begin{bmatrix} \cos\left(\frac{\pi}{4}t\right) \\ -\sin\left(\frac{\pi}{4}t\right) \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & -1 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \tag{2.8}$$

Figure 2.7 shows the simplified block diagram of a triangular decimator built with this scheme. **Note** that only one adder is required in the implementation of all fourfilters

2.3.3 **Analysis of the First Stage of Decimation**

The goal of the first stage is to reduce the sampling rate as much as possible while maintaining an adequate SNR. In a communications application, the decimation filter must also provide sufficient rejection of out-of-band signals that getaliased to baseband. For this analysis it will be assumed that the only source of noise is shaped quantization noise.

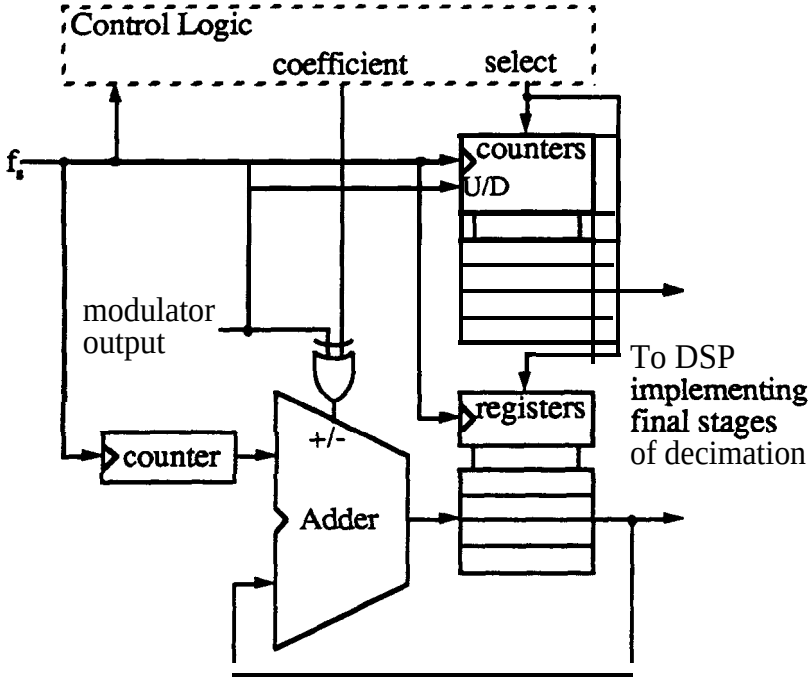

Figure 2.7: Ultimately simplified complex modulator and first-stage triangular lowpass filter/decimator.

To calculate the expected in-band noise power after decimation, one must add the noise power from the various images, N_i^2 , to the original in-band noise, N_0^2. With a sinc^k filter and decimation by a factor N, the noise power from image i is given by

$$N_i^2 = \frac{1}{2\pi} \int_{\frac{2\pi i}{N} - \frac{\omega_B}{2}}^{\frac{2\pi i}{N} + \frac{\omega_B}{2}} \frac{1}{3} \left| \left(\frac{1}{N} \frac{\sin \frac{N\omega}{2}}{\sin \frac{\omega}{2}} \right)^k H\left(e^{j(\omega + \frac{\pi}{4})} \right) \right|^2 d\omega, \tag{2.9}$$

which for small ω_B can be approximated by

$$N_i^2 \approx \frac{2(\frac{\pi}{4R})^{2k+1}}{3\pi(2k+1)} \left(\frac{\left| H\left(e^{j(\frac{2\pi i}{N} + \frac{\pi}{4})} \right) \right|}{\sin^k \frac{\pi i}{N}} \right)^2 \tag{2.10}$$

The total noise power is the sum of these terms: $N_T^2 = \sum_{i=0}^{N-1} N_i^2$.

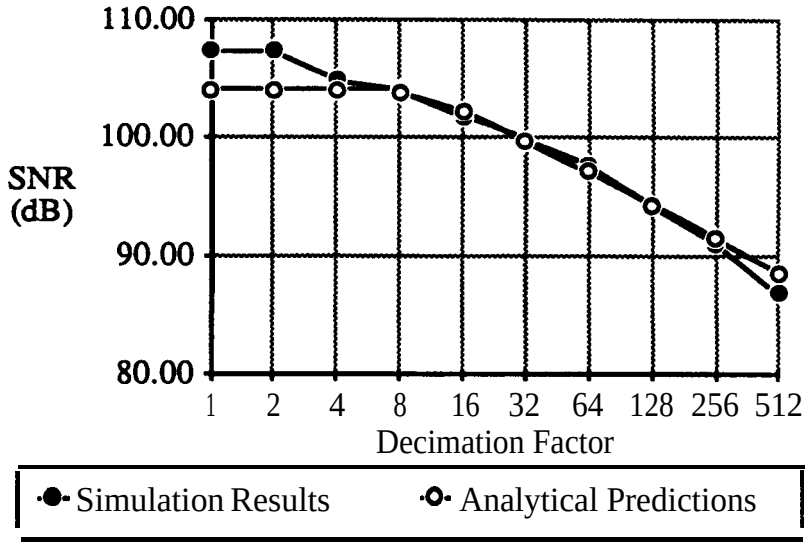

Figure 2.8: Signal-to-noise ratio of the first stage output as a function of the amount of triangular decimation. The input was a -10dB sine wave.

Figure 2.8 shows the trade-off between the SNR and the degree of triangular (k=2) decimation for the fourth-order modulator, assuming the only source of noise is shaped quantization noise. The simulation results are within a few dB of the analytical predictions. With decimation by a factor N=32, the SNR for a -10dB sine wave input is predicted to be 99.7dB; simulations yield 99.8dB.

2.3.4 The Last Stage of Decimation

The final decimation is done with a full-fledged digital filter. By examining the noise spectrum and the requirements for alias rejection, the system designer can derive a set of specifications for this filter and it can then be designed with standard techniques. The design of this filter is no different than the design of the final decimation filter in an A-to-D based on a lowpass $\Sigma\Delta$ modulator.

Figure 2.9 shows the spectrum of the output of the first stage for triangular decimation by a factor of 32 for our running example. The input tone is visible in the middle of the lower sideband. We can use this Figure to estimate the requirements of the final decimation filter

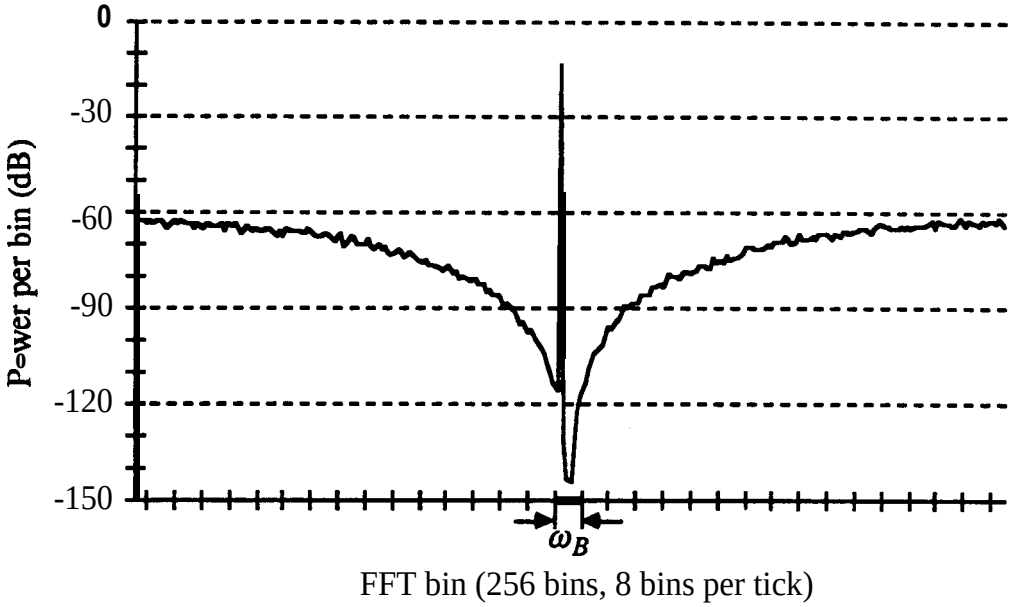

Figure 2.9 : The output spectrum of the fourth-order modulator, after multiplication by a complex exponential and triangular decimation by a factor of 32. The input is a -13dB tone in the lower sideband.

From Figure 2.9 we see that the noise power in each bin is about -63dB. There are 31 images of eight bins each, for a total of nearly 2^8 bins. The combined noise power in the unattenuated images is thus -63dB+(8 x 3)dB=-39dB. The signal power is -13dB, so for a SNR of 95dB we expect to have to attenuate the images by no more than (95+13-39)=69dB.

Simulations show that a fifth order elliptic filter with 0.5dB of passband ripple and 60dB of stopband attenuation has a sufficiently sharp cutoff and sufficiently high stopband attenuation to achieve a SNR of 97.8dB when its output is sub-sampled by a factor of 32. Note that in bandpass $\Sigma\Delta$ we decimate by a combined factor of 2R because of the presence of both real and imaginary channels.

Once the real and imaginary channels are brought down to the minimum rate, they need to be combined to yield the demodulated baseband signal. For AM applications, it is sufficient to take the magnitude of the complex output For single-sideband communications, the final stage must either be a positive-pass or a negative-pass filter. For phase or frequency modulation one must find the argument (angle) of the output. Only in this last case is it automatically mandatory for the decimation filters to be linear phase.

2.4 Summary

Simulations have been used to show that bandpass $\Sigma\Delta$ modulation works. The design of a bandpass $\Sigma\Delta$ modulator makes use of the same techniques as are used in the design of high-order lowpass modulators, and the decimation hardware is only slightly more complex. Beneficial side-effects of the decimation process are rejection of out-of-band signals and translation to baseband.

Although the examples presented here use $\omega_0 = \pi/4$, it is clear that similar results can be achieved for other simple fractions of the sampling rate; $\omega_0 = \pi/2$ is particularly simple. For arbitrary passbands, the decimation circuits would require fancier logic but the modulator design process would not be affected.

Of course, it is also possible to have multiple passbands, or to use bandpass $\Sigma\Delta$ to do narrowband digital-to-analog conversion. Other natural modifications to the modulators considered here include spreading the zeros across the band-of-interest to minimize thein-band noise power, and taking advantage of the available analog circuitry to filter the incoming signal [Jantzi, Schreier and Snelgrove 1991].

3 Sigma-Delta Modulation as a Mapping

This chapter takes an abstract view of **ΣΔ** modulation in order to identify some of its fundamental properties. The viewpoint taken is that the transformation of the input sequence, ***u***, to the output sequence, y, accomplishes a mapping in the mathematical sense.

For convenience, we will consider only binary modulators with a signal transfer function of unity and an initial state of zero. In this case, y may be computed from ***u*** according to the following recursion:

$$x(t) = u(t) + \sum_{i=1}^{\infty} h(i)e(t-i) \qquad t \geq 0 \qquad (3.1)$$

$$y(t) = \mathrm{sgn}(x(t)) \qquad t \geq 0 \qquad (3.2)$$

$$e(t) = \begin{cases} y(t) - x(t) & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad (3.3)$$

These equations could be modified to handle more general modulators as follows. An arbitrary signal transfer function could be incorporated in these equations by replacing ***u*** in Equation 3.1 with its filtered version. Also, the signum function in Equation 3.2 could be replaced by a staircase to provide an n-level encoding. Finally, a non-zero initial state could be captured by altering the definition of e for ***t<0 in*** Equation 3.3.

3.1 The Mapping is Many-to-One

The first and most trivial observation one can make about this mapping is that it appears to be many-to-one. The intuition which justifies this observation is that the input is a sequence of real numbers whereas the output is a just a sequence of bits. Thus there is a massive reduction in the information content of the signal, suggesting that there may be an infinite number of input signals which yield the same output. This argument is not rigorous, but Section 3.3 will present a result which validates our intuition.

3.2 The Mapping is Idempotent

The second observation is that the mapping is *idempotent*[1].

Proof: Suppose u is the output of a ΣΔ modulator, i.e. $u(t)=\pm 1$. We show by induction that e=O.

Basis: $e(t)=0$ for $t<0$ is given.

Recursion: Assume $e(t)=0$ for $t<T-1$. We desire to show that $e(T)=0$.

By Equation 3.3, $x(T)=u(T)$ since $e(t)=0$ for $t<T-1$.
Since $u(T)=\pm 1$, $y(T)=\mathrm{sgn}(x(T))=x(T)$, and consequently $e(T)=0$.

Thus the induction step is satisfied and we conclude e=O. Consequently $y=x=u$, and we have shown that the image of a binary signal, such as the output of a ΣΔ modulator, is the original signal.

Alternate Proof (for the visually-oriented): It can be seen from Figure 3.1 that if the initial state of the H-l 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=O for all time. This result falls out quickly because the way we draw a ΣΔ modulator directly reflects Equations 3.1 to 3.3.

The idempotence property is shared by multi-level ΣΔ modulators, but the zero initial state and the unity signal transfer function requirements cannot be relaxed.

---

[1] An operator is said to be idempotent if composition with itself yields the same operator. If the transformation were linear, this would be equivalent to saying that ΣΔ modulation is a projection.

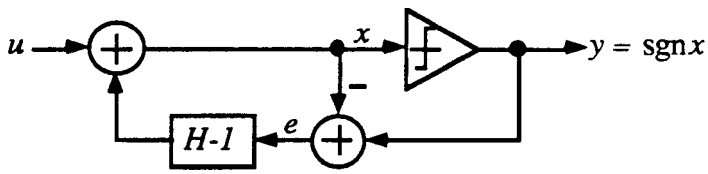

Figure 3.1: A sigma-delta modulator with an error transfer function *H* and a signal transfer function of one. Replicated from Figure 1.9.

One of the consequences of this property is that it is possible to excite a $\Sigma\Delta$ modulator in such a way that any binary output sequence is produced. Simply apply the desired sequence to the input. A mapping wherein for each element of the range there exists a member of the domain which maps to that element is said to be surjective.

3.3 Equivalent Inputs

It was claimed 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 and all signals corresponding to the same output must have the same DC value. For a bandpass modulator, *H* has a zero at ω_0 and so the input and the output must have the same ω_0-component. Can we be more specific? What can we say about the general case?

We know from the preceding section that applying a binary signal, y, to the input yields an output of 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 x does not change.

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

$$sgn(x(0)) = sgn(y(0) + \varepsilon) = sgn(y(0)), \tag{3.4}$$

or

$$\varepsilon \geq -1 \text{ when } y(0) = 1 \text{ and } \varepsilon < 1 \text{ when } y(0) = -1. \tag{3.5}$$

If this condition holds, then $e(0) = -\varepsilon$. Thereafter, the $H\text{-}1$ 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.

We can add many similar perturbations to u, each affecting precisely one value of x, by shifting εh along the time axis. Collecting these perturbations allows us to write a formula for inputs u which yield the output y:

$$u(t) = y(f) - \sum_{i=0}^{\infty} e(i)h(t-i), \tag{3.6}$$

or

$$u = y - h * e, \tag{3.7}$$

where

$$e(z) \leq 1 \text{ if } y(i) = 1 \text{ and } e(i) > -1 \text{ if } y(i) = -1. \tag{3.8}$$

It is possible to show that all inputs which yield y are of this form. Thus we have a formula for the set of all inputs which give rise to any output and this formula provides a rigorous justification of the intuition of Section 3.1.

One interpretation of (3.7), which yields a familiar result, 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 $1/3$, then one can make use of (3.7) to compute the noise power of y in the band $[\omega_1, \omega_2]$:

$$N_0^2 \approx \frac{1}{\pi} \int_{\omega_1}^{\omega_2} \frac{1}{3} \left| H\left(e^{j\omega}\right) \right|^2 d\omega. \tag{3.9}$$

This is precisely the method we used in the preceding chapters and the one that is used almost universally in analytical calculations of the signal-to-noise ratio.

If no assumptions are made about e, other than its power being bounded by P_e, then an upper bound may be put on the in-band noise power,

$$N_0^2 \leq \max_{\omega \in [\omega_1, \omega_2]} \left|H\left(e^{j\omega}\right)\right|^2 P_e, \tag{3.10}$$

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

An alternative interpretation of (3.7) begins with a restriction on the size of e. If we insist that $e(i) \in (-1, 1]$, then the condition (3.8) is satisfied, no matter whether y $= +1$ or y = -1. In addition, $|x(i)| \leq 2$ and so the modulator is stable. Thus (3.7) can be used to find all the inputs which keep the modulator 1 -*stable*, where *1*-stable is defmed to mean $|e(i)| \leq 1$. This set is:

$$U_1 = \left\{ u \mid u = y - h * e, \quad \text{where } y(r) = \pm 1 \text{ and } e(t) \in (-1, 1] \right\}. \tag{3.11}$$

For these inputs the modulator is guaranteed to work, in the sense that the error signal is always less than 1 and so the bound (3.10) holds with $P_e = 1$. Any other input will overload the modulator, at least momentarily. If one could fmd the smallest signal which is not in the set U_1, then one would know the input limit of the modulator- a very important parameter. Clearly, a good understanding of the set U, is of practical significance. More will be said about this set and others like it in the next chapter.

3.4 Limit-Cycles and Amplitude Quantization

We know that if a ΣΔ modulator is excited with a binary-valued signal, then the output follows the input. If, in addition, the input is chosen to be periodic, then the output will be periodic as well. Since many inputs can yield that output sequence, it is conceivable that other periodic signals, perhaps just the ω_0-component of the periodic signal, would give rise to the same periodic output, or limit-cycle. We see by this argument that the idempotence property of ΣΔ suggests the existence of limit-cycles.

3.4.1 Ideal Modulators

A great deal is known about limit-cycles in the ideal first-order and second-order lowpass $\Sigma\Delta$ modulators. Gray [Gray 1987] explored the behavior of the first-order modulator with a rational input and showed that the output is always periodic. Friedman [Friedman 1988] 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 an input to produce a limit-cycle, the input and output must have the same ω_0-component if the modulator has a zero at ω_0. Thus for a modulator with a zero of the error transfer function at $\omega_0 = 0$, the average value of the input must equal the average value of the limit-cycle, which can only be a rational number. If the modulator does not have a zero at DC, then a constant input need not be rational to give rise to a limit-cycle.

In the bandpass case, it is not clear that a limit-cycle can be supported by simple input, such as a sine wave at ω_0. Consider a second-order bandpass modulator with an error transfer function $H = 1 - \sqrt{2}z^{-1} + z^{-2}$. It has a zero at $\omega_0 = \pi/4$ so the period of a limit-cycle with a non-zero ω_0-component must be a multiple of 8. Some possible length-8 limit-cycles and their ω_0-components are shown in Table 3.1. To determine whether or not the supposed sequence can be produced by the modulator, it is necessary to get the modulator in the correct state. It was 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 3.1 shows for which test patterns this process succeeded in exposing a limit-cycle. We shall see that it is possible to verify these results analytically using the discrete Fourier transform.

Period of limit-cycle	ω_0-component	Seen in simulation?
+++++++−	0.5000	No
++++++−−	0.9239	No
+++++−−−	1.2071	Yes
++++−−−−	1.3066	Yes
+++−−+−−	0.7071	Yes

Table 3.1: Some possible limit-cycles in a second-order bandpass modulator ($\omega_0 = \pi/4$).

Bandpass modulators, like their lowpass counterparts, do indeed exhibit limit-cycles. As a result, one can expect to fmd all the limit-cycle related phenomena that areknown 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 (the fourth entry in Table 3.1). In general, the limit is the magnitude of the ω_0-component of the limit-cycle $\mathrm{sgn}(\sin(\omega_0 t))$. At $\omega_0 = \pi/2$, the limit is $\sqrt{2}$ and it decreases to $4/\pi$ as $\omega_0 \to 0$. Of course, the root-mean-square value of such a signal cannot exceed 1.

3.4.2 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 $H\left(e^{j\omega_0}\right) \neq 0$, i.e. the modulator is not ideal, then a range of input amplitudes and phases can sustain a limit-cycle. Figure 3.2 displays the ω_0-component of the output sequence as a function of the amplitude of a sine input for

ideal and non-ideal versions of the second-order bandpass modulator. The non-ideal modulators have their zeros moved radially by a factor $1-\delta$, and Figure 3.2 uses the relatively large values of $\delta = \pm 0.1$ to make the non-ideal behavior readily apparent.

We see that the curves for $\delta > 0$ and $\delta < 0$ are qualitatively 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, but saturates at a much lower input level. The flat portions exist for $\delta > 0$ but not for δ c 0 because limit-cycles are persistent[2] if and only if H^{-1} is stable.

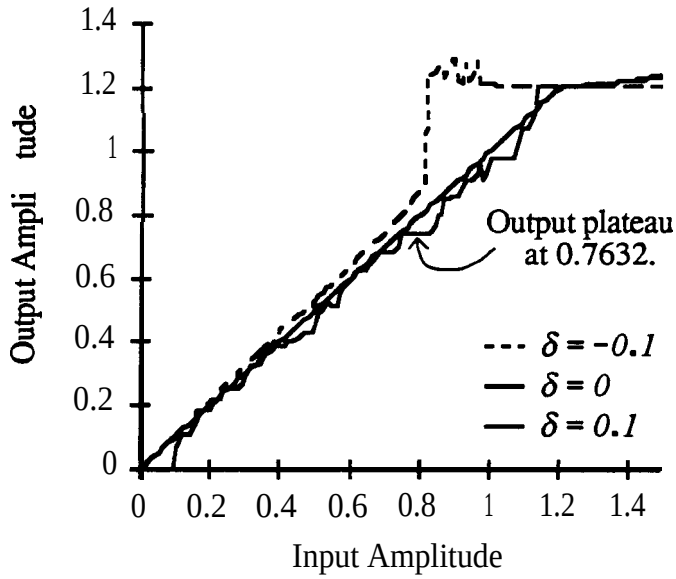

Figure 3.2: **Output** amplitude as a function of input amplitude for ideal and non-ideal bandpass modulators with a sine input.[3]

---

2 Persistent means that the limit-cycle is able to survive small perturbations. The terms stable or attracting are often used in the nonlinear oscillator literature to denote similar ideas. The only limit-cycles which are observable in practice are those which are persistent

3 This Figure was produced by applying sine wave inputs of various amplitudes to a modulator with an initial state of zero. If, instead, a sine wave whose amplitude was swept across the range indicated were applied to the input, the plots would likely show hysteresis.

This last statement is actually a special case of a more general result: the output of a $\Sigma\Delta$ modulator is predictable[4] if and only if H^{-1} is stable.

Essence of the proof: Suppose y is an output corresponding to an input u. Since y is assumed to be known exactly, we can remove the quantizer from Figure 3.1 and supply y as an input, as shown in Figure 3.3[5]. The $\Sigma\Delta$ modulator thus becomes a linear system with two inputs, u and y.

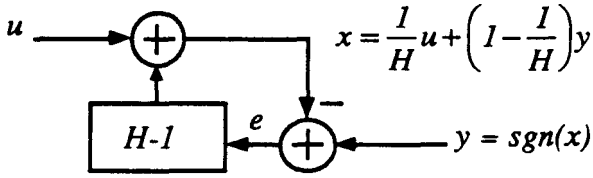

Figure 3.3: A $\Sigma\Delta$ modulator with the comparator removed and the output considered to be an input.

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 in u, caused for example by noise, will cause x to change by increasingly large amounts, until y is no longer $\mathrm{sgn}(x)$ or x becomes infinite. If H^{-1} is stable, perturbations that are small enough will not be magnified to the point where the sign of x is changed, and so the output will not change. See Appendix A for a full proof.

A straightforward technique, which uses the discrete Fourier transform, exists for calculating the maximum width of a given output amplitude plateau or for verifying that a limit cycle exists. An example will be used to illustrate the technique.

---

4 A system is said to have a predictable output if inputs which are close to one-another produce outputs which are likewise close. In the present context, since the output takes only discrete values, outputs which areclose are taken to be identical.

5 This technique of "opening the loop" was popularized by Tsypkin, and the author acknowledges Søren Hein [Hein and Zakhor 1991] for pointing out the originator of this technique. The feature of the signum nonlinearity which this technique exploits is that the signum function is piece-wise constant.

From the time-domain simulations used to produce Figure 3.2, it was found that the limit-cycle corresponding to the output amplitude plateau at 0.7632 has a length of 24 samples:

$$y = \dot{+}+++-+--+-++----+-++-+-\dot{-} \tag{3.12}$$

The input is a sine wave of amplitude *A* and frequency $\omega_0 = \pi/4$:

$$u(t) = A\sin(\omega_0 t), \tag{3.13}$$

and x is found via

$$x = Ax_1 + x_2 \tag{3.14}$$

where

$$x_1 = H^{-1}\frac{u}{A} = H^{-1}(\sin(\omega_0 t)),^6 \tag{3.15}$$

and

$$x_2 = (1 - H^{-1})y. \tag{3.16}$$

The x_1-sequence is a sine wave with an amplitude of $\left|H\left(e^{j\omega_0}\right)\right|^{-1}$ and a phase of $-\arg\left(H\left(e^{j\omega_0}\right)\right)$, and the x_2-sequence is easily computed using the discrete Fourier transform. With these sequences known, it is a simple matter to determine limits on *A*, $[A_1, A_2]$, which satisfy the consistency condition y(r) = sgn(x(t)).

For each t where $y(t) = \text{sgn}(x_1(t))$, *A* must be larger than a certain minimum value, $-\frac{x_2(t)}{x_1(t)}$. The minimum value of *A* is the largest of these lower limits:

$$A_1 = \max_{t|y(t)=\text{sgn}(x_1(t))}\left(-\frac{x_2(t)}{x_1(t)}\right) \tag{3.17}$$

---

6 A term of the form Lx, where L is a transfer function and x is a sequence, denotes the sequence (signal) which is the output of a system with a transfer function L and an input x.

Likewise the maximum value of A is:

$$A_2 = \min_{t|y(t)\neq sgn(x_1(t))} \left(-\frac{x_2(t)}{x_1(t)} \right) \tag{3.18}$$

For the limit-cycle under consideration, this procedure yields the *A-range* $(0.7438, 0.8343)$, which has a width of 0.0905. The simulations used to produce Figure 3.2 give the width of the plateau as 0.09 ± 0.01, which agrees with the calculated value. This is somewhat fortuitous, since the analytical calculation assumes that the modulator is in the correct state for supporting the desiredlimit-cycle- there is no guarantee that an input of the correct amplitude is sufficient to produce the limit-cycle when the initial state is zero.

The numerical details of the foregoing calculations have been relegated to Appendix **A.**

This procedure may be repeated for non-zero phases of the input sine wave, to find the amplitudes and phases which can support this limit-cycle. The resulting set is shown in Figure 3.4. The boundary has comers since the critical values of t in (3.17) and (3.18)

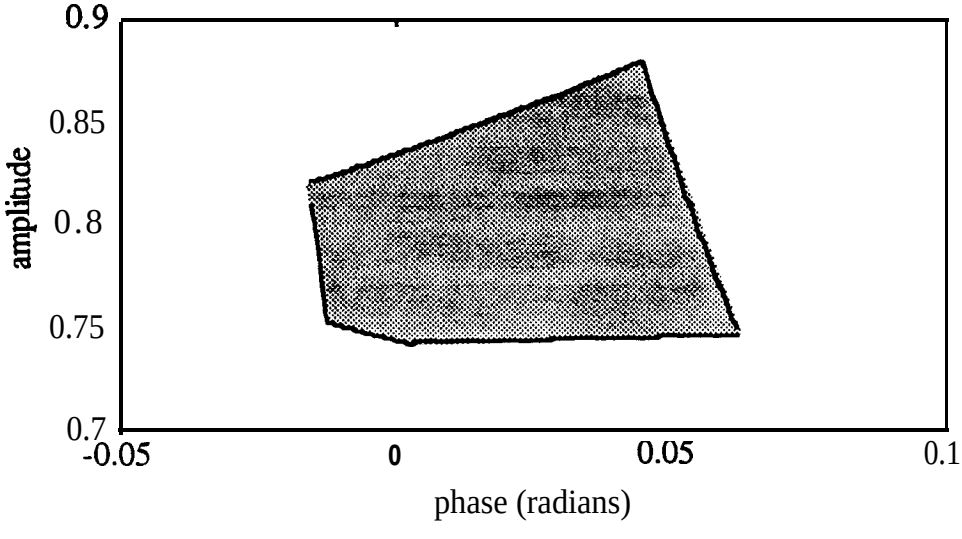

Figure 3.4: **The** set of amplitudes and phases of a sine wave capable of supporting the limit-cycle in Equation (3.12).

change as the phase changes, but each segment is approximately linear over the small phase-ranges seen here.

An interesting feature of this Figure is that the phase range is rather narrow: about *4* degrees. An innovative designer might find that a simple, low-quality, bandpass modulator such as the one used here could serve as a crude sort of phase-detector, one with a digital output.

3.5 Summary

This chapter examined ΣΔ modulation from a mathematical viewpoint and showed that it accomplishes an idempotent, surjective mapping. The inputs which yield a particular output form an equivalence class, and a formula for the elements of these classes was given. 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.

This chapter also investigated limit-cycles, showing that they exist in bandpass modulators and that they lead to amplitude quantization when H^{-1} is stable. An efficient procedure for calculating the maximum width of an output plateau and for verifying the existence of a limit-cycle was given.

When H^{-1} is not stable, limit-cycles are not persistent and so cannot be observed in practice. In fact, when H^{-1} is not stable, the output is not even predictable: arbitrarily small changes in the input or in the state of the modulator can cause the output to eventually change from $+1$ to -1 or vice versa. Aperiodic behavior and a lack of predictability are two of the hallmarks of chaos [Gleick 198 1][Parker and Chua 1987]. It is remarkable that the rule for the transition from periodic behavior to aperiodic and possibly chaotic behavior in a ΣΔ modulator has such a simple form: H^{-1} stable or not.

4 Stability in a Noise-Shaped Modulator

A modulator is said to be ***stable if the*** input to the quantizer is bounded, or equivalently, if the error signal, e, is bounded In a modulator which is operating stably, e has finite power and the modulator essentially works as desired since the in-band component o f $h * e$ is small. In addition, circuits which implement the describing equations of a $\Sigma\Delta$ modulator can only be built if e is bounded. Clearly, stability is a very desirable property.

Stability is a pressing issue in the design of noise-shaped modulators. Simulations are widely used to determine if a modulator is stable, and if so, to find its dynamic range. This approach is serviceable, but does not possess the comforting certainty one would have were there an adequate analytical test for stability. Several investigators have proposed simple criteria based on the magnitude of the frequency response of the error transfer function, but this chapter will show that these are wrong. To counter this disappointment, a rigorous test based on the impulse response and the size of the input is presented and the coverage of this rule is illustrated.

4.1 Stable Inputs

The investigation into stability begins with the observation that stability is conditional upon the input. From the preceding chapter, we have seen that an input u will give rise to an output y if and only if there exists e such that

$$u(t) = y(t) - (h * e)(t) \tag{4.1}$$

and

$$y(t) = \operatorname{sgn}\big(y(t) - e(t)\big). \tag{4.3}$$

The set of all inputs which ensure $e(t) \in (-M, M]$, called the set of ***M-stable inputs,*** is given by

$$U_M = \{ u \mid \text{u satisfies (4.1), y satisfies (4.2) and } e(t) \in (-M, M] \}. \tag{4.3}$$

The case M=1 is a natural choice since one often assumes in the linear analysis of a $\Sigma\Delta$ modulator that e is uniformly distributed over [-1,1]. This yields the set presented in Section 3.3, and is repeated here for comparison with (4.3).

$$U_1 = \left\{ u \mid u(t) = y(t) - (h * e)(t), \quad \text{where y(r)} = \pm 1 \text{ and } e(t) \in (-1, 1] \right\}. \tag{4.4}$$

These formulae show how stability depends on the input. Every modulator has inputs for which it is stable in the sense that e remains bounded, and the set of all such inputs has a very special structure. The set consists of non-overlapping islands, one for each possible output sequence, which may touch along their edges. In the case M=1, ***the h* e*** term is independent of y, so the islands all have the same shape with the output sequence at the center of each one.

To make this more concrete, consider these sets in two dimensions: the first two samples of u. If $e(t) = 0$ for $t \leq 0$, (4.1) becomes

$$\begin{aligned} u(0) &= y(0) - h(0)e(0) \\ u(1) &= y(1) - h(0)e(1) - h(1)e(0) \end{aligned} \tag{4.5}$$

and for $M = 1$, (4.2) becomes

$$e(0), e(1) \in (-1, 1]. \tag{4.6}$$

Figure 4.1 plots the set of l-stable inputs for the first-order lowpass modulator, $h = (1,-1,0\ldots)$, and the second-order lowpass modulator, $h = (1, -2,1,0\ldots)$. The set of I-stable inputs is a union of identical, disjoint hyperparallelepipeds[1] centered on every possible output sequence. The key question is whether this set covers a useful set of signals or not.

The input u=O is particularly crucial as it is likely the center of any useful set of signals. It appears to be well within the *1*-stable set for the first-order modulator, but only on the boundary for the second-order modulator.

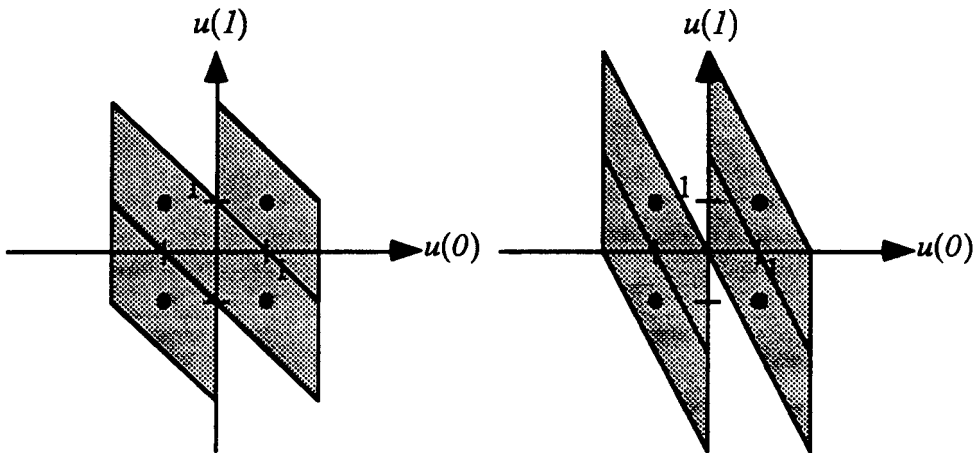

Figure 4.1: The *1*-stable set, U_1, for the standard first and second-order lowpass modulators. The dots represent the four possible output sequence: $(+1,+1),(+1,-1),(-1,+1)$ and $(-1,-1)$.

Figure 4.2 shows the set of M-stable inputs for the two modulators when M is increased to **2. The** parallelograms grow outward, along some of their exposed faces and so no longer remain centered on the output sequences. Note that the origin is covered in both diagrams,

---

[1] A hyperparallelepiped is the n-dimensional analog of a parallelogram (2 dimensions) or a parallelepipcd (3 dimensions).

indicating that zero may belong to the set of 2-stable inputs for both modulators (it in fact does).

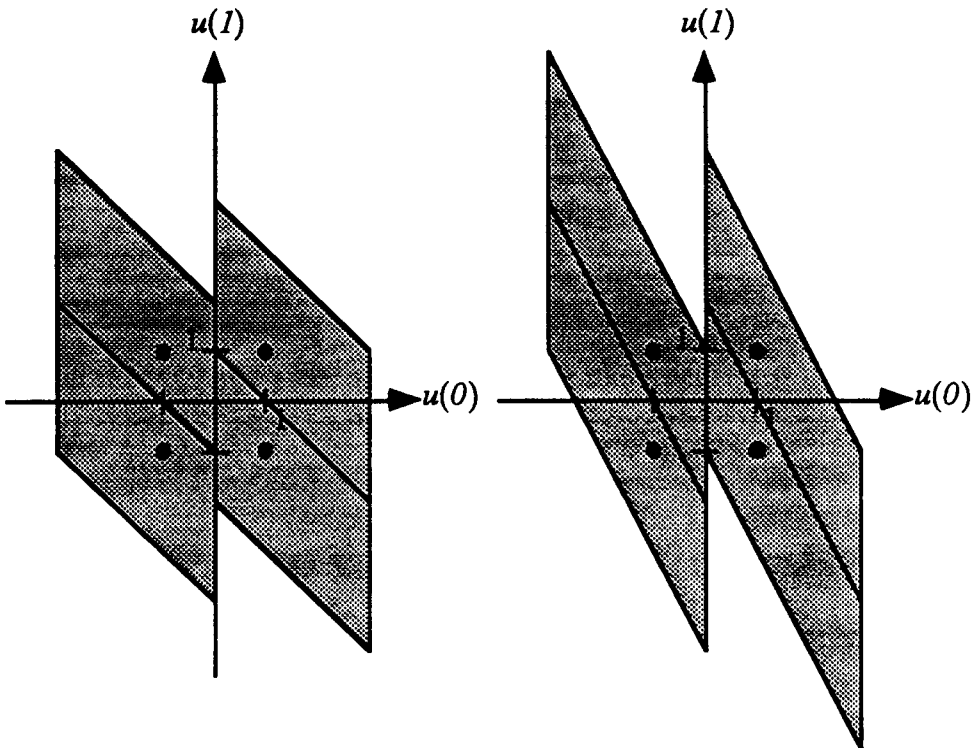

Figure 4.2: The 2-stable set, U_2, for the standard first and second-order lowpass modulators.

The pictures are somewhat misleading because they only deal with two instants in time. When the number of samples becomes infinite, some of the impressions induced by Figures 4.1 and 4.2 are wrong. For example, it appears that any input which is sufficiently close to an output sequence ought to belong to the set of stable inputs which give rise to that output. However, we know from section 3.4.2 that when H^{-1} is unstable and y is assumed fixed, arbitrarily small perturbations can be magnified to the point where either x becomes infinite or y becomes inconsistent. In either case, this implies that points arbitrarily close to the output pattern may not belong to the set of stable inputs which correspond to that output: the islands have zero width in almost all directions and consequently the idempotence property is not robust.

The pictures of the stable input set are quite simple in two dimensions, but as the number of dimensions grows, they quickly become impossible to visualize. Despite the fact that (4.3) provides a compact expression for the set of stable inputs, it is very difficult to estimate the coverage of this set and so it has been abandoned as a method for testing stability.

4.2 Zero-Input Stable Modulators

The preceding section showed that stability is conditional upon the input, with zero being especially important. This signal is "in the middle" of all the islands of stability and in that sense is equally remote from all of them. In addition, it is certainly the case that any useful modulator should be stable with zero input. This section examines the set of zero-input stable modulators.

A ΣΔ modulator is characterized by its noise transfer function. The set of all noise transfer functions is infinite-dimensional but one can only draw pictures of finite-dimensional sets. For the purposes of the bulk of this chapter, it suffices to examine the three-dimensional set: $h=(1, h(1), h(2), h(3), 0...)$. For each h chosen from this set, a simulation of the modulator corresponding to that h was run and the maximum value of e was noted. Figures 4.3, 4.4 and 4.5 plot the maximum value of e, $\|e\|_\infty$, for various $h(1)$ and $h(2)$, when $h(3)$ is -0.5, 0 and +0.5, respectively.

These pictures show that we are dealing with a very complicated object. There exist a few simple criteria in the ΣΔ literature which have been used to predict the stability of a ΣΔ modulator, but it will be shortly seen that the simple regions they delineate bear only a rough resemblance to the complex object seen here.

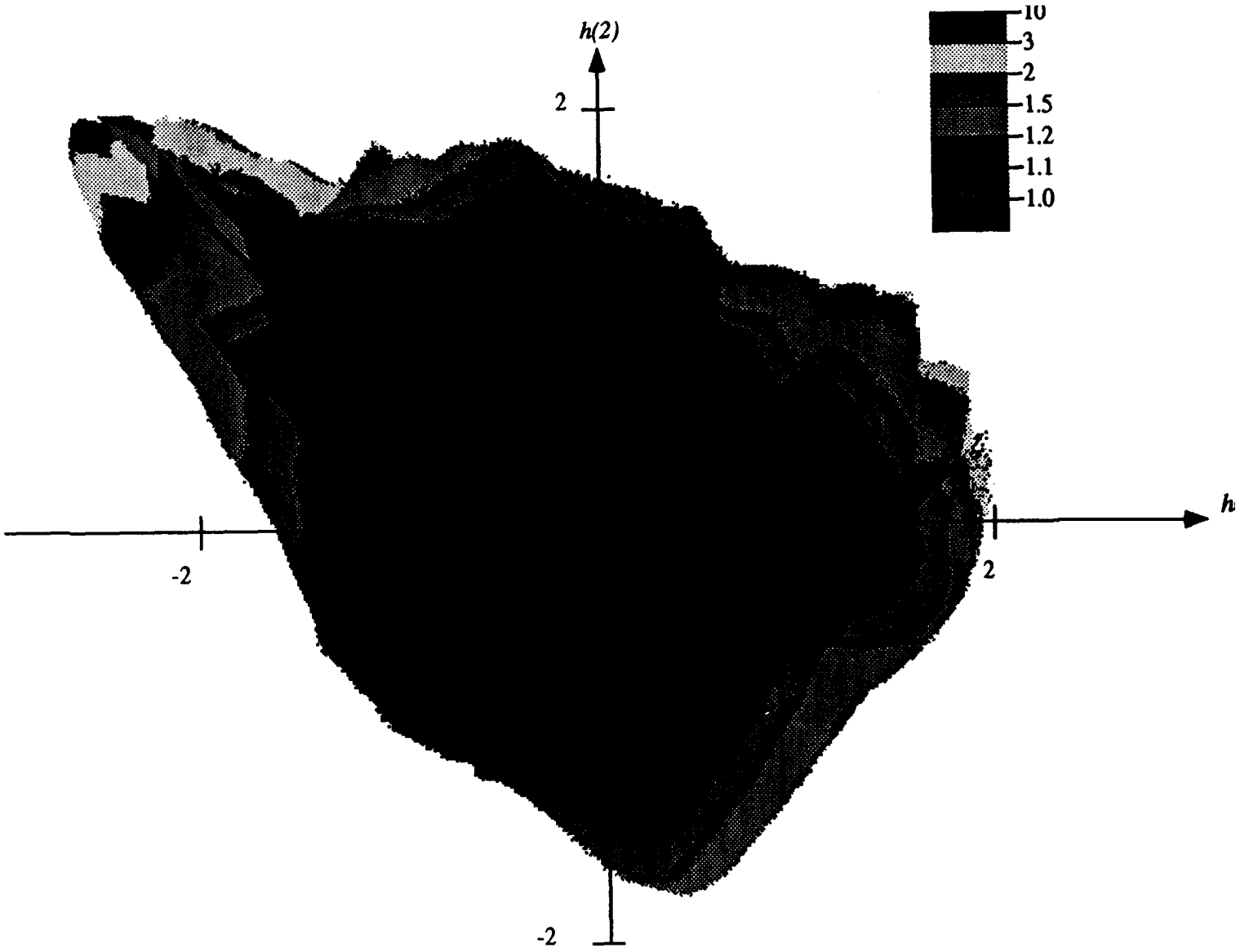

Figure 4.3: The magnitude of the error signal in modulators of the form *h=(1,a,b,-0.5,0,0...)*, with an input of zero.

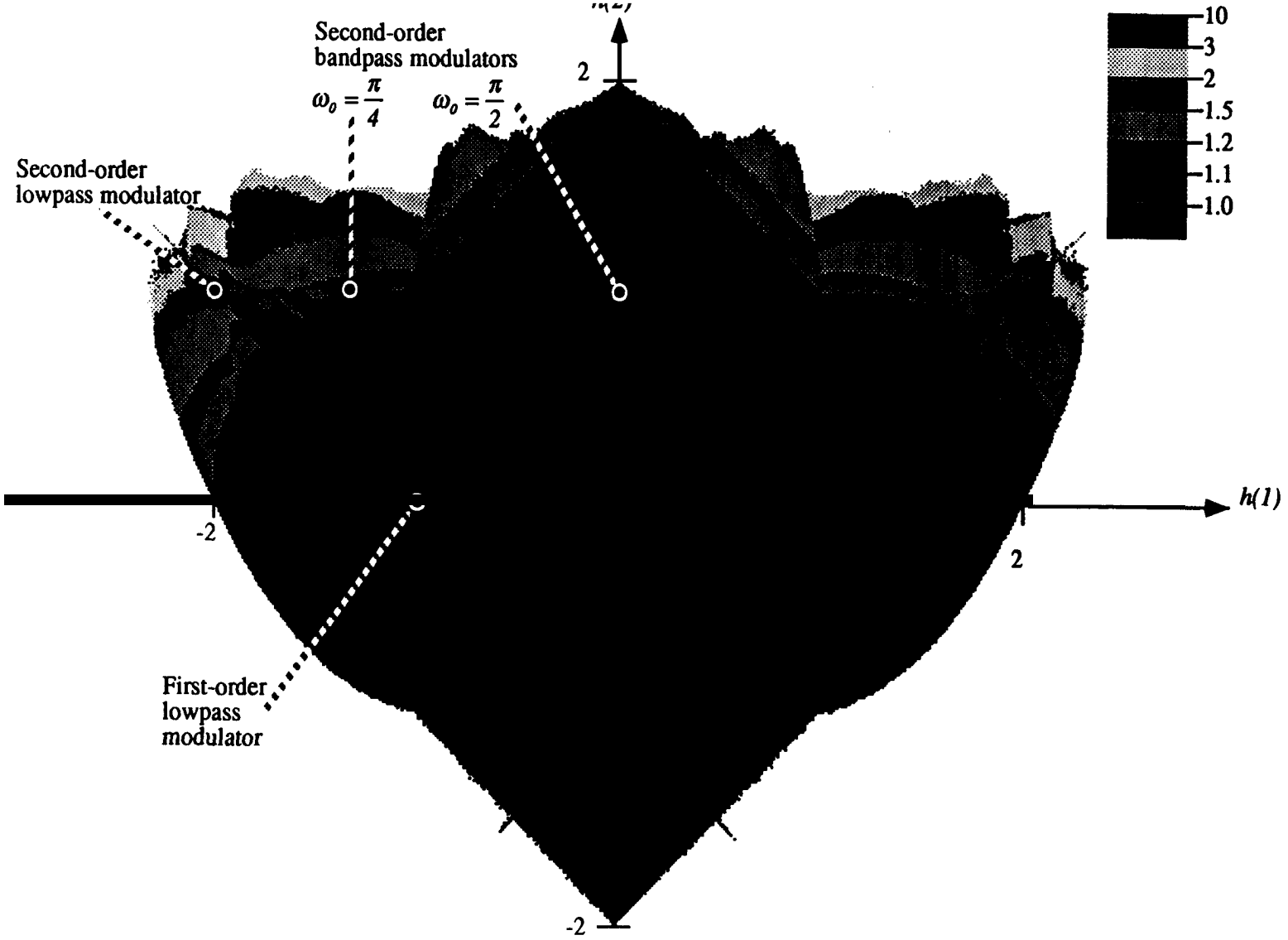

Figure 4.4: The magnitude of the error signal in modulators of the form $h=(1,a,b,0,0,0...)$, with an input of zero.

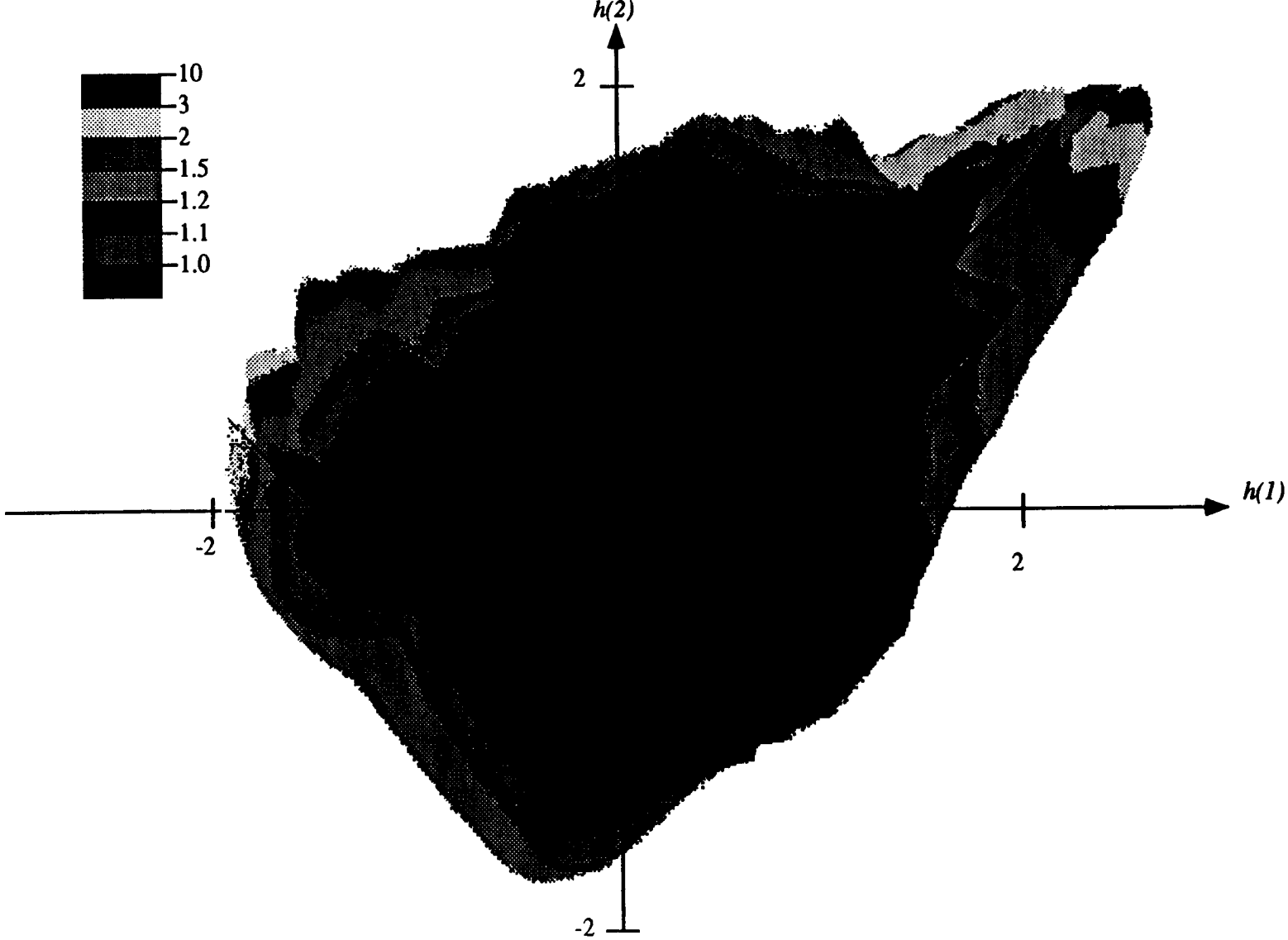

Figure 4.5: The magnitude of the error signal in modulators of the form $h=(1,a,b,+0.5,0,0...)$, with an input of zero.

4.3 Rules of Thumb

The brute force stability test of the preceding section required a modulator to be simulated for infinite time. A shortcut for gauging the stability of a $\Sigma\Delta$ modulator is sorely needed, and several authors have provided such criteria based on the magnitude of the frequency response of *H. This* section presents these rules of thumb and shows them to be both insufficient and unnecessary.

4.3.1 The Power Gain, or L_2-Norm, Criterion

Agrawal and Shenoi [Agrawal and Shenoi 1983] give an argument akin to the following for their test. If it is desired that the error signal be white and uniformly distributed over $[-1,+1]$, it must have a power of $1/3$. The power in the output signal is 1, so it appears that the power gain of the the error transfer function, $\|H\|_2^2$, ought to be less than 3. For the parameterization of the error transfer function by the samples of its impulse response, this rule of thumb corresponds to a sphere, and its circular boundaries are shown on Figure *4.6*.

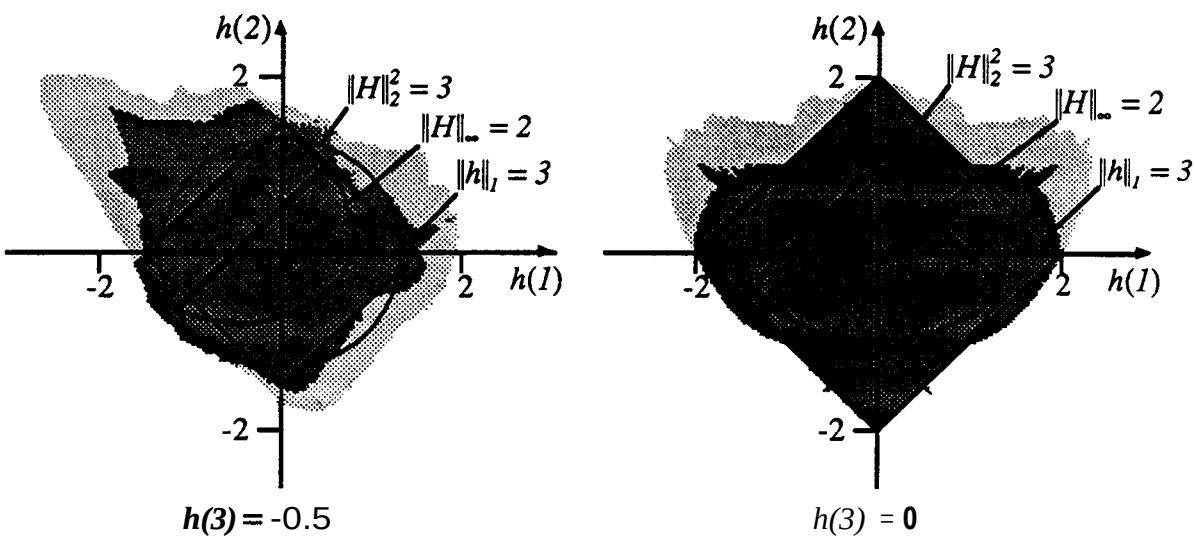

Figure 4.6: The set of modulators stable (light gray) and I-stable (dark gray) with zero input, plus several stability criteria gleaned from the literature. No rule completely covers either set.

4.3.2 The Maximum Gain, or L_∞-Norm, Criterion

Lee [Lee 1987] argues that if the gain of the error transfer function at every frequency is less than 2, the resultant modulator will be stable. The boundary of this region is an irregular curve, possibly with linear segments, and is also plotted in Figure 4.6.

This Figure indicates that the power gain and the maximum gain criteria are not necessary conditions for zero-input stability. It will be rigorously established in Section 4.5 that $\|h\|_1 \leq 3$ is a sufficient test for zero-input stability, and Figure 4.6 shows that this test passes modulators which fail both frequency-domain criteria.

In Figure 4.6, the maximum gain criterion, $\|H\|_\infty \leq 2$, appears to be more conservative than the power gain criterion, $\|H\|_2^2 \leq 3$, but in practice it can be the reverse. This happens because noise transfer functions designed to satisfy $\|H\|_\infty \leq 2$ generally have gains near 2 at most frequencies and consequently have $\|H\|_2^2 \approx 4$. This is not apparent in Figure 4.6 because it deals with FIR error transfer functions of extremely low order.

4.3.3 Counter-Examples

Even worse than not being necessary, the power gain and maximum gain criteria are not sufficient to ensure stability. Figure 4.7 shows a close-up of a part of modulator space wherein completely unstable modulators pass both criteria.

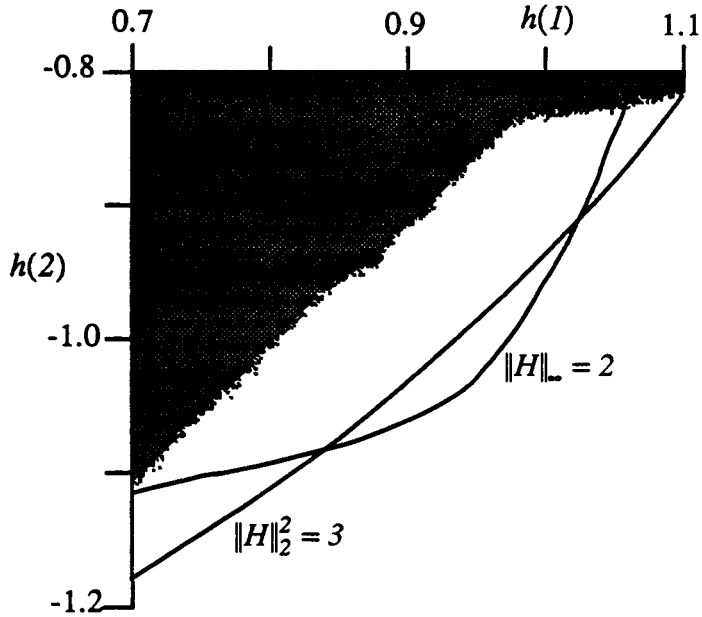

Figure 4.7: Modulators of the form $h=(1, a, b, 0.15, -0.3, 0...)$ which are zero-input stable. Both the maximum gain and the power gain criteria include modulators which are unstable.

4.4 Relations Among Error Transfer Functions

Before embarking on the development of rigorous stability criteria, it is worthwhile to make some general observations on the relationships which exist among zero-input stable modulators.

The equations which describe a ΣΔ modulator with an initial state of zero and an input of zero may be considered to form a set of nonlinear recursion equations:

$$x(t) = \sum_{i=1}^{\infty} h(i)e(t-i), \qquad t \geq 0 \tag{4.7}$$

$$e(t) = \begin{cases} \operatorname{sgn}(x(t)) - x(t), & t \geq 0 \\ 0, & t < 0. \end{cases} \tag{4.8}$$

For a particular ***h, the x*** and ***e*** sequences can be calculated using (4.7) and (4.8). Are these sequences related to those for other values of ***h?***

4.4.1 Alternate Negation

$h' = (1, -h(1), h(2), -h(3), \ldots)$ has $e' = (e(0), -e(1), e(2), -e(3), \ldots)$ and so ***h'*** yields a zero-input stable modulator if and only if ***h*** does and $\|e'\|_\infty = \|e\|_\infty$.

Proof: If e and x satisfy the Equations 4.7 and 4.8 for the parameter vector ***h,*** then e' $= (e(0), -e(1), e(2), -e(3), \ldots)$, $x' = (x(0), -x(1), x(2), -x(3), \ldots)$ satisfy these equations for the parameter vector h'. ***This will be*** shown by induction.

Basis: $x'(0) = 0 = x(0)$ and e'(O) = 1= e(O).

Recursion: Assume $x'(t) = (-1)^t x(t)$ and e'(r) $= (-1)^t e(t)$ for $t < T$. Then

$$\begin{aligned} x'(T) &= \sum_{t=1}^{\infty} h'(t)e'(T-t) \\ &= \sum_{t=1}^{\infty} (-1)^t h(t) \cdot (-1)^{T-t} e(T-t) \\ &= (-1)^T \sum_{t=1}^{\infty} h(t)e(T-t) \\ &= (-1)^T x(T) \end{aligned}$$

If T is even, $x'(T) = x(T)$ and so $e'(T) = e(T)$.

If T is odd, $x'(T) = -x(T)$ and so $e'(T) = -e(T)$ iff $x(T) \neq 0$. ***The*** possibility that $x(T) = 0$ for T odd is a very singular one, and so will be discounted. (If $x(T) = 0$, then arbitrarily close to the given ***h*** there is another value of ***h*** which does not suffer from this perversity.)

Thus we can multiply the odd terms in the x, ***h*** and e sequences by -1 and the nonlinear recursion equations will (generically) be satisfied. This explains why Figure 4.5 is a mirror image of Figure 4.3 and why Figure 4.4 is symmetrical about the vertical axis.

When the input is non-zero, these relations continue to hold if alternate terms in the input are also multiplied by -1.

4.4.2 Zero-padding

$h' = (1,0,h(1),0,h(2),0,h(3),\ldots)$ has $\|e'\|_\infty = \|e\|_\infty$.

Proof: If e and x satisfy the nonlinear recursion equations for the parameter vector h, then it is readily apparent that the sequences e' $= (e(0),e(0),e(1),e(1),e(2),e(2),\ldots)$ and $x' = (x(0),x(0),x(1),x(1),x(2),x(2),\ldots)$ likewise satisfy (4.7) and (4.8) for the parameter vector h'.

Physically, this transformation corresponds to replacing all z^{-1} delays with z^{-2} delays. The result is that the modulator behaves in the same manner as two copies of the original modulator, with all sequences interleaved.

Corollary: It is possible to put k zeros between all the samples in the impulse response and not alter the stability of the equations. The error sequence then consists of k interleaved copies of the original sequence.

A moment's consideration is sufficient to verify that a similar relation holds when the input is non-zero.

4.4.3 Alternative definitions of e

From the linear model paradox of Section 1.5.1, it can be observed that a $\Sigma\Delta$ modulator with an error transfer function of the form $H' = \dfrac{H}{k + (1-k)H}$, where $k>0$, *can* be made from one with an error transfer function H merely by altering the definition of e. The signal transfer function changes, but for an input of zero this is irrelevant, and we conclude that

H' is zero-input stable iff $\boldsymbol{H}$ is. This relationship cannot be verified in Figures 4.3-4.5 since it converts FIR error transfer functions into IIR ones and $\|e'\|_\infty \neq \|e\|_\infty$.

4.5 Rigorous Criteria

4.51 The ℓ_1-Norm Criterion

Anastassiou [Anastassiou 1989] shows that if $\|u\|_\infty \leq 1$, then the modulator is guaranteed to be I-stable if

$$\sum_{i=1}^{\infty} |h(i)| \text{ Il.} \tag{4.9}$$

A more general version can be proved as follows. From Equation 3.1,

$$\begin{aligned} |x(t)| &= \left| \sum_{i=1}^{\infty} h(i)e(t-i) + u(t) \right| \\ &\leq \sum_{i=1}^{\infty} |h(i)e(t-i)| + |u(t)| \\ &\leq \sum_{i=1}^{\infty} |h(i)| + \|u\|_\infty \qquad \text{if } |e(t-i)| \leq 1. \end{aligned}$$

To ensure that $|e(t)| \leq 1$, it is sufficient to ensure that $\sum_{i=1}^{\infty} |h(i)| + \|u\|_\infty \leq 2$. Therefore, by induction, the modulator is I-stable if

$$\sum_{i=1}^{\infty} |h(i)| \leq 2 - \|u\|_\infty \tag{4.10}$$

or, since $h(0) = 1$,

$$\|h\|_1 = |h(0)| + \sum_{i=1}^{\infty} |h(i)| \leq 3 - \|u\|_\infty. \tag{4.11}$$

The diamonds corresponding to this test are plotted along with the two rules of thumb in Figure 4.6. The $\|h\|_1$-test is the only criterion mentioned in the literature which is correct. Unfortunately, it can be the most conservative of all. As before, this is not apparent in Figure 4.6 because it uses FIR error transfer functions of such low order.

Figure 4.6 clearly shows that none of the criteria examined completely covers what appears to be the set of zero-input stable modulators. In particular, none of the stability criteria says that a second-order lowpass modulator, $h=(1,-2,1,0...)$, ought to be stable, yet empirically it is stable and is in fact a very popular modulator. This omission motivates a more detailed examination of the conditions necessary for stability in a $\Sigma\Delta$ modulator.

4.5.2 **First-Order FIR Criteria**

The starting point is the simplest modulator imaginable: $h = (1, h(1),0,...)$. Here there is only one free parameter and this allows (4.7) and (4.8) to be combined into a single non-linear equation:

$$e(t) = sgn(h(1)e(t-1)) - h(1)e(t-1). \tag{4.12}$$

Due to the symmetry present (Section 4.4.1), one may assume $h(1)>0$. In this case, (4.12) can be simplified to

$$e(t) = sgn(e(t-1)) - h(1)e(t-1). \tag{4.13}$$

Although this is a nonlinear equation, its simplicity allows one to determine the range of $h(1)$ for which $e(t)$ is bounded. From (4.10), $h(1) \le 2$ ensures $|e| \le 1$. If $h(1) > 2$ then $|e|$ grows without bound. For

$$\begin{aligned} |e(t)| &= |sgn(e(t-1)) - h(1)e(t-1)| \\ &\ge |h(1)e(t-1)| - 1 \\ &> 2|e(t-1)| - 1 \end{aligned}$$

together with

$$e(0) = 1$$

implies that $|e|$ increases monotonically and without bound.

From Section 3.4.2, where it was shown that limit-cycles can be persistent if H^{-1} is stable, we suspect that limit-cycles exist for $|h(1)|$ c I. It turns out that the limit-cycles have period 1, i.e. e converges to a limit, e. :

$$e_* = sgn(e_*) - h(1)e_*$$

which, if it is assumed that $e_* > 0$, leads to

$$e_* = 1 - h(1)e_*,$$

or

$$e_* = \frac{1}{1+h(1)}.$$

It can be proved that if $e(-1) \geq 0$, then the e-sequence converges to e_* when $h(1) < 1$. For if $e(t-1) \geq 0$,

$$\begin{aligned} e(f) - e_* &= sgn(e(t-1)) - h(1)e(t-1) - e_* \\ &= 1 - h(1)e(t-1) - \frac{1}{1+h(1)} \\ &= -h(1)(e(t-1) - e_*) \end{aligned}$$

and consequently e converges to e_* if and only if $h(1) < 1$. For $e(-1) < 0$, e converges to $-e_*$.

Figure 4.8 plots the value of the state $e(t)$ for various values of $h(l)$, for $t > 1000$. Note that e does indeed converge to e_* when $0 < h(1)$ c 1, and that it takes on a multitude of values when $h(1) > 1$. For $h(1) > 2$, the iterations blow up. The above results are able to explain these primitive aspects of Figure 4.8, but are unable to explain the finer detail, in particular the unusual distribution of e when $|h(1)| > 1$.

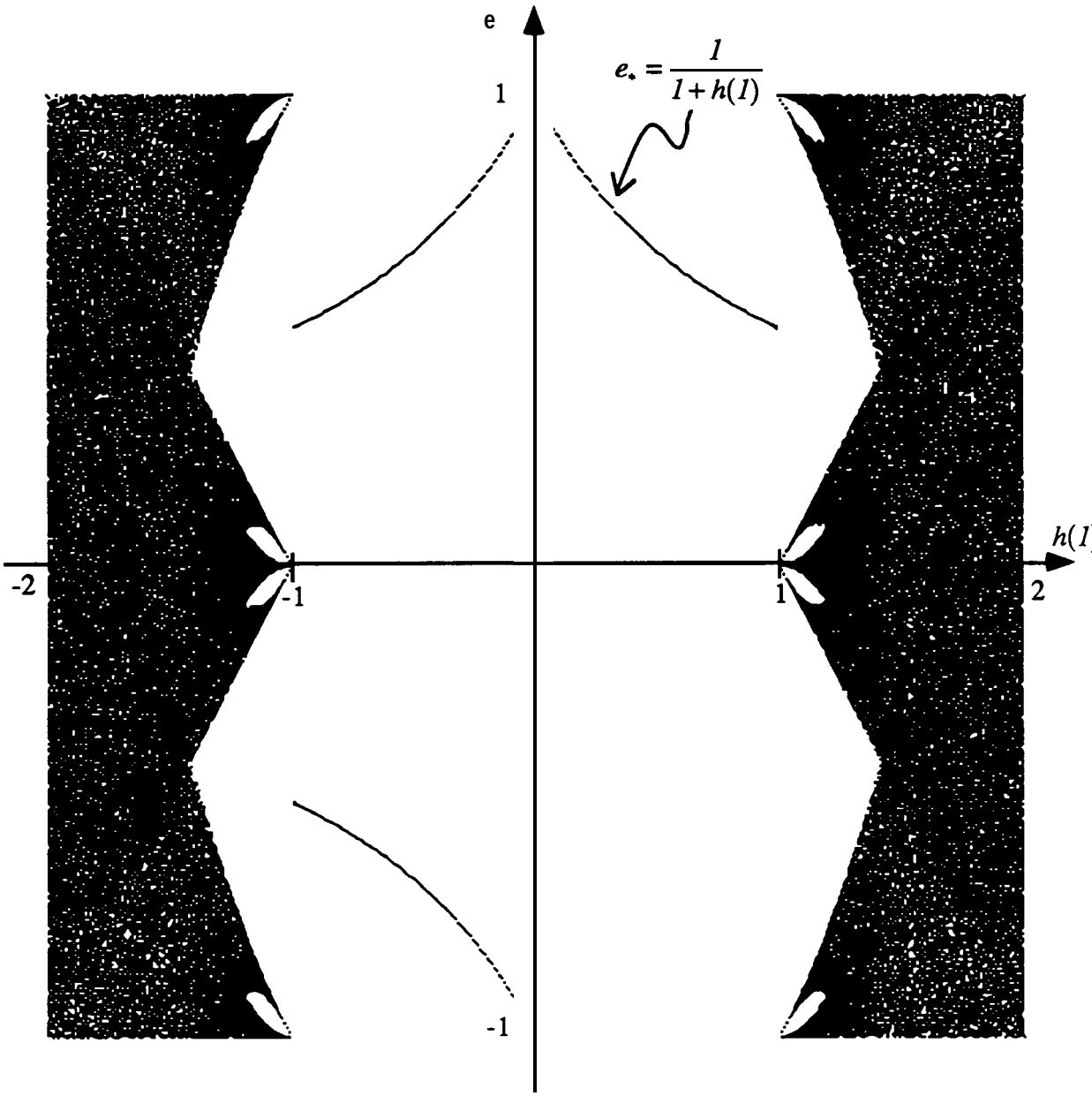

Figure 4.8: The values of e for different values of $h(1)$, for an initial state of zero and after the initial transient.

This plot is very different from those usually associated with a transition to chaos. Typically, such a bifurcation diagram shows a "period-doubling route to chaos" [Gleick 198 1], but our system appears to go directly from period-1 behavior to chaotic behavior at $h(1) = 1$, or from period-2 to chaos at $h(1) = -1$.

4.5.3 Second-Order FIR Criteria for 1-Stability

This section studies the nonlinear recursion **(4.7)** and **(4.8) with two** parameters in or&r to gain a better understanding of Figure 4.4.

From the derivation of the condition $\|h\|_1 \leq 3$, which for the present may also be written as $|h(1)| + |h(2)| \leq 2$, it is clear that this condition would be both necessary and sufficient for I-stability if it were possible for $(e(t-1), e(t-2))$ to take on any value in the unit square. However, Figure 4.4 shows that there exist I-stable modulators with $|h(1)| + |h(2)| > 2$, so it follows that for these modulators the state of the modulator does not range over the whole unit square.

Figure 4.9 plots the values of the state of the modulator parameterized by $h(1)=1.7$ and $h(2)=0.6$. We see that indeed, $\{(e(t-1), e(t-2))\}$ is not the whole square; much of the square is inaccessible. When $|h(1)| + |h(2)| > 2$, as in this case, there are two critical triangles on opposite sides of the unit square, S : one where $x(t) = h(1)e(t-1) + h(2)e(t-2) > 2$ and one where $x(t)$ c -2. Should the state wander into one of these triangles, $|e(t)|$ will exceed 1 and so the state will subsequently map outside S.

Figure 4.9 suggests a technique for finding I-stable modulators: just find a subset of S that maps into itself with each iteration of (4.7) and (4.8). We know that this subset cannot include the critical triangles, so the largest it could be is S minus the critical triangles. Figure 4.10 shows this set and its image when $h(1) > h(2) > 0$. We see that the shape of the image is such that it may be a subset of the original set We shall determine conditions on $h(1)$ and $h(2)$ for which this is true.

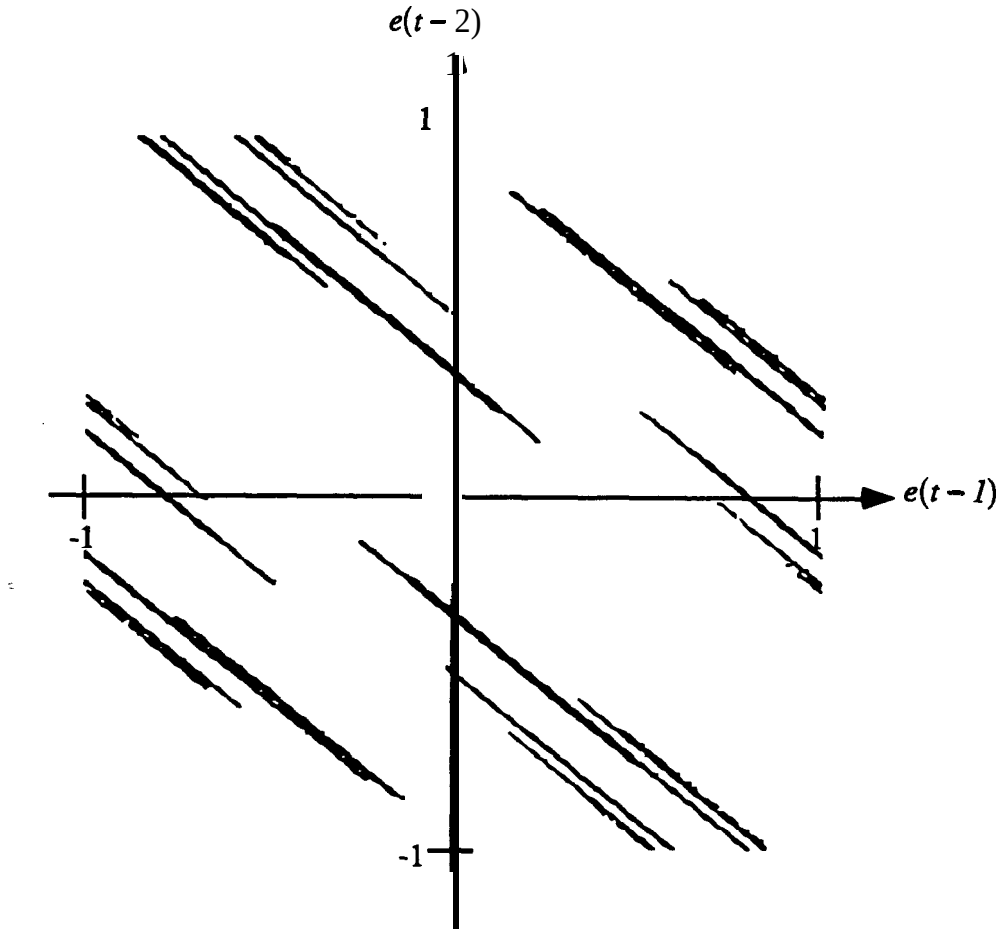

Figure 4.9: The inhabited region of state-space for a second-order FIR ΣΔ modulator with $h(1)=1.7, h(2)=0.6$ and zero input. The state does not occupy the entire unit square

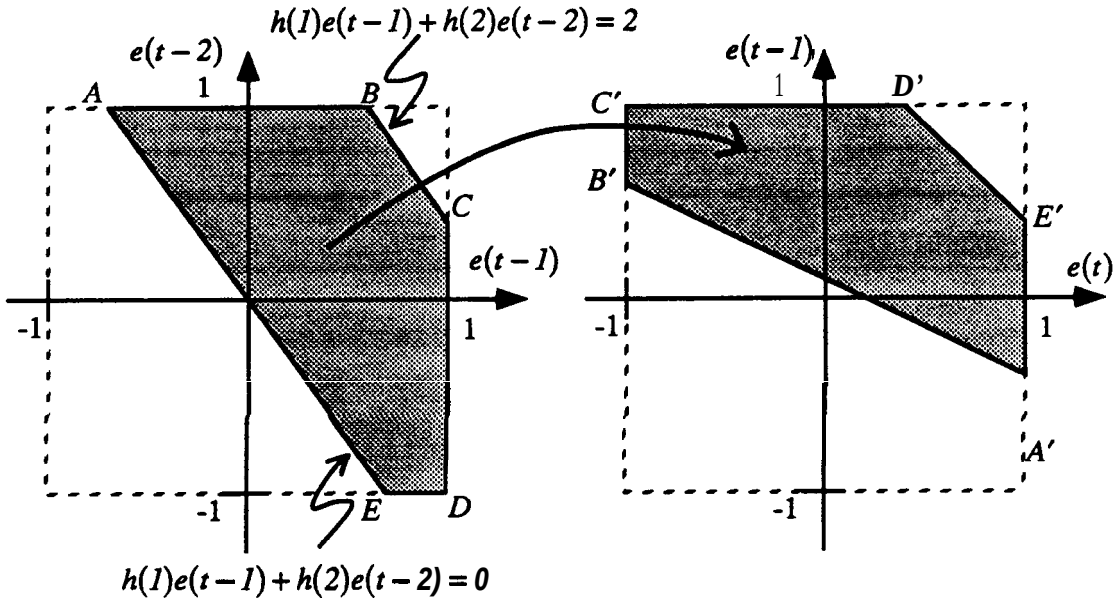

Figure 4.10: Proving I-stability requires finding a subset of the unit square which maps into itself. The unit square minus the critical triangles may map to a subset of itself when $h(1) > h(2) > 0$.

As a result of symmetry, we can restrict our attention to the case $h(1) > 0$. S is divided into two halves by the line $x(t) = h(1)e(t-1) + h(2)e(t-2) = 0$; on one side of the line $\mathrm{sgn}(x(t)) = 1$, on the other $\mathrm{sgn}(x(t)) = -1$. Denote these halves by S^+ and S^-. The symmetry of this bisection implies that the image of only one of the pieces need be examined.

In each piece, $\mathrm{sgn}(h(1)e(t-1) + h(2)e(t-2))$ is constant, so the mapping from one state to the next is affine. Consequently straight lines map to straight lines, and so it is sufficient to look at the vertices of half the candi&te polygon, S_1^+, and their images. Table 4.1 lists the vertices of S_1^+, ***their*** images, and the conditions necessary for the images to be within S_1.

The most restrictive of these conditions is that for the ***E vertex.*** Rearranging this condition yields that S, maps to a subset of itself if $h(1) > h(2) > 0$ and

$$(h(1) - 1)^2 + h(2)^2 \le 1. \tag{4.14}$$

This equation describes a circle of radius *1* and center *(1,0)*.

Vertex		Image	Image a member of S_1 iff
A	$\left(-\frac{h(2)}{h(1)}, 1\right)$	$\left(1, -\frac{h(2)}{h(1)}\right)$	$\frac{h(2)}{h(1)} \le 1$
B	$\left(\frac{2-h(2)}{h(1)}, 1\right)$	$\left(-1, \frac{2-h(2)}{h(1)}\right)$	no restriction
C	$\left(1, \frac{2-h(1)}{h(2)}\right)$	$(-1, 1)$	no restriction
D	$(1, -1)$	$(1-h(1)+h(2), 1)$	$-1 \le 1-h(1)+h(2) \le \frac{2-h(2)}{h(1)}$
E	$\left(\frac{h(2)}{h(1)}, -1\right)$	$\left(1, \frac{h(2)}{h(1)}\right)$	$\frac{h(2)}{h(1)} \le \frac{2-h(1)}{h(2)}$

Table 4.1: The vertices of S_1^+, their images, and the conditions necessary for the images to be inside \s\do5(S_1) when $h(1) > h(2) > 0$.

For the case $h(2) < 0$, it is convenient to assume $h(1) < 0$, so that the critical comers remain at $(1,1)$ and $(-1,$ -1). It turns out that the image of the unit square minus the critical triangles protrudes into the critical triangles, so it is necessary to shave off more than just these triangles.

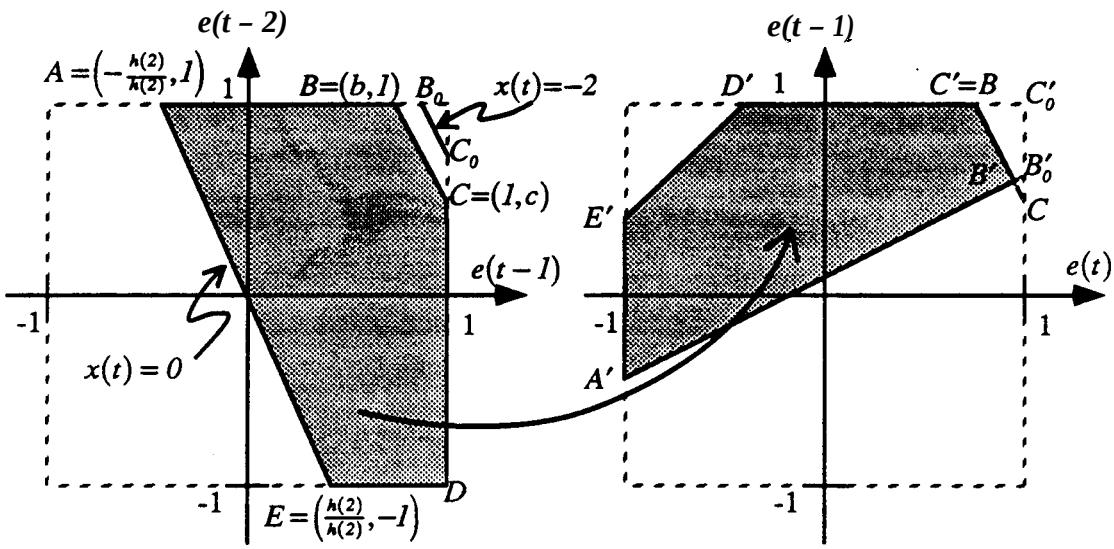

Figure 4.11: When $h(1) < h(2) < 0$, it is necessary to shave off more than just the critical comer in order for the subset property to hold..

Figure 4.11 shows the negative half, S_2^-, of the candidate region, S_2, and its image when point C is shifted down from its original position C_0. C must be shifted far enough so that C' is no longer to the right of B, so the least restrictive requirement is that $C' = B$. Thus

$$C' = (1,c)' = (-1 - h(1) - ch(2), 1) = B = (b,1)$$

or

$$b = -1 - h(1) - ch(2) \tag{4.15}$$

In addition, it is necessary to shift B left to bring B' within S_2. Putting B' on CB makes S_2 as large as possible while still allowing the subset property to hold. This leads to the requirement that

$$B' = (b,1)' = (-1 - h(1)b - h(2), b)$$

beontheline

$$\frac{e(t) - 1}{b - 1} = \frac{e(t-1) - c}{1 - c}. \tag{4.16}$$

Solving (4.15) and (4.16) yields values for b and c. Consistency in the diagrams and the subset property require that

$$\frac{h(2)}{h(1)} \le c \le c_0 \qquad \text{and} \qquad -\frac{h(2)}{h(1)} \le b \le b_0.$$

The most severe of these requirements turns out to be

$$\frac{h(2)}{h(1)} \le c. \tag{4.17}$$

Figure 4.12 plots the regions covered by (4.17), by (4.14), by their symmetric counterparts and by the $|h(1)| + |h(2)| \le 2$ condition. These rules combine to give quite good coverage of the set of second-order FIR, zero-input, I-stable modulators.

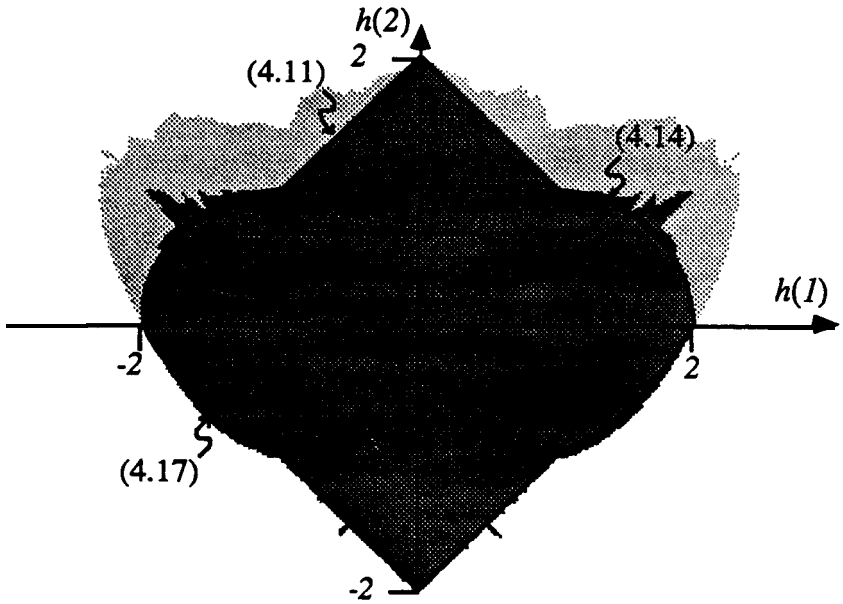

Figure 4.12: **The** extended rules for I-stability cover many of the second-order FIR modulators I-stable with an input of zero.

Figure 4.12 shows that (4.17) is conservative. This is to be expected since (4.17) comes from the requirement that the image of *E* be inside S_2, which is overly restrictive because the image of S_2 may not include *E. This* suggests that cutting the comers at *A* and *E* would result in less restrictive requirements on $h(1)$ and $h(2)$, a possibility which is too tedious to explore here.

4.5.4 **Second-Order FIR Criteria for M-Stability**

Figure 4.12 shows that many zero-input stable modulators, those M-stable with $M>1$, still lack an analytical justification. This can be remedied by examining how the critical triangle is mapped by repeated application of (4.7) and (4.8). Figure 4.13 shows that when *h(2)* **>** *0 the* triangle may map back inside S, whereas for *h(2)* **c** 0 the triangle maps further and further outside S. This explains why when *h(2)* **c** 0 modulators which are not I-stable are not stable, whereas when $h(2) > 0$ modulators can be stable without being I-stable.

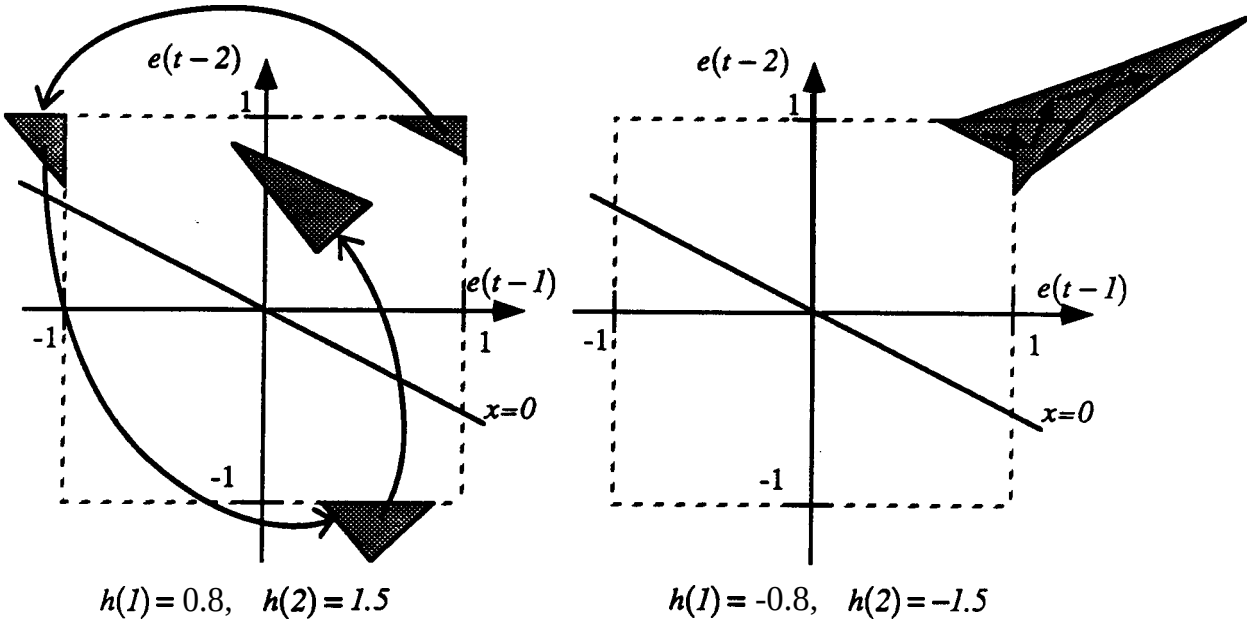

Figure 4.13: When $h(2) > 0$, the critical triangle may map back inside S, but when $h(2) \leq 0$, it maps further and further outside S.

We can keep track of the image of the critical triangle to see if it eventually returns to the unit square. If it does, the zero-input stability of the given modulator has been proven. Noting the maximum value that e attains during this process gives an upper bound for M.

This is difficult to do analytically, since it is likely that an image will be cut by the lin$x=0$, necessitating a cleaving of the image into two polygons, whose images may, in turn, need splitting. A computer may be programmed to carry out the required calculations, and the result of doing so is shown in Figure 4.14. This Figure closely resembles Figure 4.4, which is remarkable considering that a maximum of only 20 iterations were made for each modulator in Figure 4.14, whereas in Figure 4.4 10^5 iterations were needed for each stable modulator to achieve the degree of smoothness in the boundaries of that Figure. Of the modulators which appeared zero-input stable after a brute force simulation of 10^5 time

steps, 94%[2] were *proven* zero-input stable by following the critical triangle for 20 iterations.

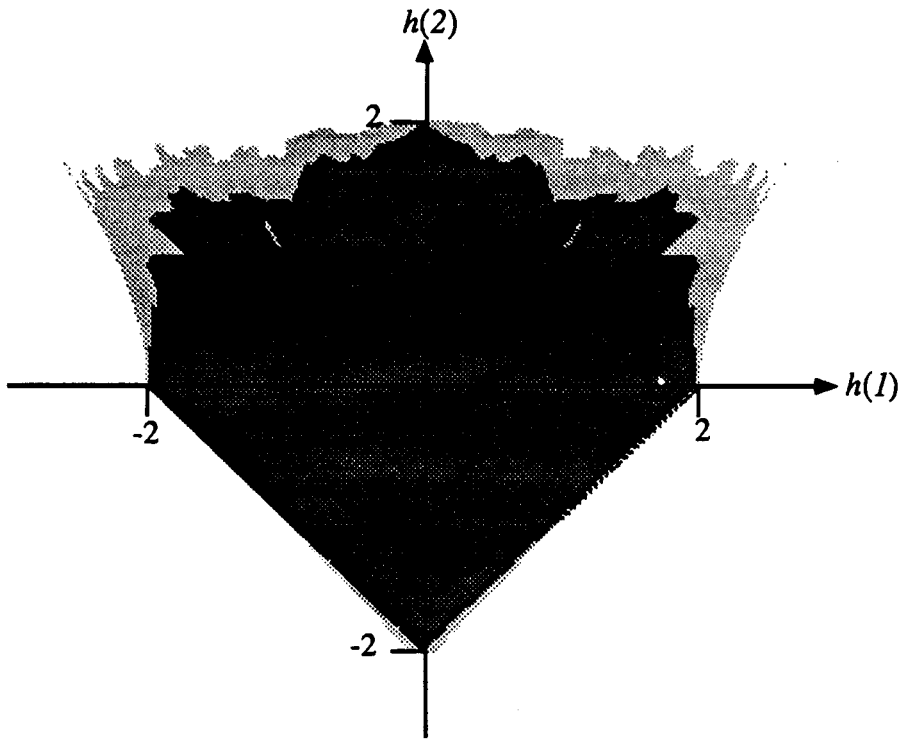

Figure 4.14: **The** set of modulators which can be proven zero-input stable (dark gray), by following the critical comer for 20 iterations. Modulators for which an image of the critical triangle went beyond $e=100$ are white, i.e. the algorithm judged these modulators unstable. The remaining indeterminate modulators are colored light gray.

Conceptually, this process may be generalized to higher dimensions in a straightforward manner. The critical triangle becomes a **simplex**[3], the repeated cleaving of which can yield

---

[2] The standard double-loop modulator is in the indeterminate area.

[3] A simplex is an n-dimensional object of non-zero volume with $n+1$ *vertices.*

convex polytopes[4] with any number of vertices. However, writing a program to carry out the calculations in the *n*-dimensional case would not be as straightforward as it was in the two-dimensional case considered here.

A more intriguing possibility is to generalize the process to IIR error transfer functions. The process would essentiaIly be unchanged, except that the "home" region of state-space would no longer be the unit square. The home region, defined as the set of states that the $H-1$ block could reach if the input were restricted to [-1,1], could be determined from a state-space description of this block. It is hoped that the home region is a convex polytope, so that testing another polytope for inclusion in the home region could be done by simply checking that the vertices of the test polytope are inside the home region. Once the home region is known, the critical polytope could be readily found and its images tracked.

To include the effect of a non-zero input, one would have to augment the iteration by allowing the input to cause a point in state-space to map to many possible destination states. This would have the effect of enlarging images beyond their zero-input values, again creating extra vertices.

Writing a program to correctly and efficiently carry out these operations would be quite a chalIenge. Efficiency is important since the cleaving which can occur at each iteration can double the number of polytopes which need to be examined and simultaneously add many vertices to them. This challenge will not be taken up at this time.

---

[4] A polytope is the n-dimensional analog of a polygon (two dimensions) or a polyhedron (three dimensions).

4.6 Summary

This chapter developed the notion of stability in a $\Sigma\Delta$ modulator, and explored some of the properties of stability in a $\Sigma\Delta$ modulator characterized by its error transfer function, It was found that every modulator is stable for some inputs, but not for others, and a formula for the set of all inputs which keep a particular modulator stable was given. Although the formula is explicit, it is difficult to use it to decide whether the set of stable inputs forms a useful set of signals or not.

In search of a practical test for stability, the maximum value of the error signal was plotted for modulators belonging to a three-dimensional subspace of modulator space, with an input of zero. The resulting object showed surprising complexity and consequently dashed any hope for finding a simple analytical test for stability in a general modulator which is both necessary and sufficient.

Several rules of thumb have been mentioned in the literature, and their simple predictions were compared to the complex object seen here. In general, the shapes given by these criteria bear little resemblance to the shape observed. Many modulators, including the very popular second-order lowpass modulator, do not pass these tests. In the case of the power gain and maximum gain criteria, counter-examples were given to show that these criteria are insufficient as well as unnecessary.

The $\|h\|_1$ -rule was extended by including a bound on the maximum value of the input. This has yielded the most general analytical test for stability, with a sound theoretical basis, of which the author is aware.

In the case of modulators with second-order FIR error transfer functions and an input of zero, this chapter has been able to go beyond the $\|h\|_1$-rule. Mechanisms for l-stability when the $\|h\|_1$-rule is violated have been explored and these were able to cover the bulk of

I-stable modulators of this type. An algorithm was described which was able to rapidly prove the zero-input stability of 94% of the second-order FIR modulators stable with an input of zero.

Suggestions for generalizing this test to higher-order FIR and IIR error transfer functions and for allowing non-zero inputs were made, but the algorithm was not implemented.

5 Conclusions

5.1 Contributions

In the first chapter, a simplified model of ΣΔ was presented. Existing modulators were put into this framework and found to be readily comprehensible.

In the second chapter, simulations were used to show that bandpass ΣΔ modulation is possible and that decimation is not difficult. Bandpass ΣΔ modulation may find application in the simultaneous analog-to-digital conversion and demodulation of narrowband RF signals.

The third chapter explored some fundamental mathematical properties of ΣΔ modulation. It was shown that ΣΔ modulation is an idempotent operation if the signal transfer function is unity and the initial state is zero, that limit-cycles are persistent if H^{-1} is stable, and that limit-cycles exist in bandpass modulators. A formula for the set of all inputs which yield the same output was presented. The existence of amplitude quantization in non-ideal modulators was noted and a technique for calculating the widths of the amplitude plateaus was illustrated with an example. It was suggested that designing H^{-1} to be unstable might rid a ΣΔ modulator of the troublesome effects of limit-cycles.

The fourth chapter concentrated on the stability problem The equivalent inputs formula of the previous chapter was applied to find the set of all inputs which keep a modulator stable, but it was not seen how to estimate the size of this set. The problem was simplified by asking “Which modulators are stable for zero input?‘, and this set was shown to be very complicated.

Two tests for stability, taken from the literature, were shown to be incorrect by means of a family of counter-examples. The $\|h\|_1$-test, also taken form the literature, was extended to

include the input range and this resulted in the most general, provably sufficient, analytical condition for stability known at this time.

A detailed examination of a two-dimensional slice of the set of zero-input stable modulators provided the insight necessary for the development of a program capable of proving the zero-input stability of modulators with second-order FIR error transfer functions.

5.2 Future Work

Bandpass ΣΔ modulation needs further development. The next step involves the synthesis of a real circuit and its installation in a real system. Structures, testing and tuning are all fertile areas for research. Work is under way in the author's lab, and in the labs of several prominent researchers and corporations to develop this important area.

Despite all this activity, the basic problem of stability still lacks an adequate solution. The method for proving the stability of high-order FIR and IIR modulators with non-zero inputs, suggested at the end of the last chapter, ought to be tried. The resulting program could be incorporated into an existing filter design system to aid the design of noise-shaped modulators.

The author is indebted to Prof. W. M. Wonham for pointing out a recent paper [Rachid 1991] which uses techniques reminiscent of those used in Chapter 4. Rachid considers the problem of finding positive invariant domains [1] for linear systems with both modelling uncertainties and constrained inputs. Although the domains he examines are hyperparallelepipeds, which would have to be generalized to polytopcs, the linear system

---

[1] A positive invariant domain is a region in state-space, D, where for any state in D the next state is also in D.

requirement is not a major problem since, in a ΣΔ modulator, the state-space is divided into two halves, wherein the next state of the modulator is an affme function of the current state. Investigation of the applicability of the positive invariant domain concept to ΣΔ is yet another avenue for research.

Appendix A: Predictability and Limit Cycles

The purpose of this appendix is to elaborate on the verbal descriptions of Section 3.4.2 in an unambiguous mathematical manner. Some notation and a few definitions need to be established before the main theorems can be stated and proved. Patience on the part of the reader is required in order to reach these objectives.

In order for this Appendix to be self-contained, the relevant defiitions of Chapter 0 will be repeated, with a few minor modifications.

t time, an integer $t \geq 0$.

sgn() the signum function, $\mathrm{sgn}(a) = \begin{cases} +1, & a \geq 0 \\ -1, & a < 0 \end{cases}$

u the input sequence, $u = (u(0), u(1), u(2)\ldots), u(t) \in \Re$. All sequences used in this appendix are one-sided.

x the decision sequence, i.e. the input to the quantizer.

y the output sequence.

e the error sequence.

h the impulse response of the error transfer function, the realizability constraint requires $h(0) = 1$.

δ the discrete-time delta function, $\delta(t) = \begin{cases} 1, & t = 0. \\ 0, & t > 0 \end{cases}$

$\|h\|_1$ the 1-norm of h, $\|h\|_1 = \sum_{t=0}^{\infty} |h(t)|$.

$\|e\|_\infty$ the ∞-norm of e, $\|e\|_\infty = \sup_{t \geq 0} |e(t)|$.

$h * e$ the convolution of h with e, $(h * e)(t) = \sum_{i=0}^{t} h(i) e(t-i)$.

He a concise notation for the above. H is a linear system with impulse response h; He is the output of this system when the input is e.

There are a number of equivalent ways to describe a ΣΔ modulator. One description is pictorial: the canonical diagram presented in Chapter 1 and replicated below.

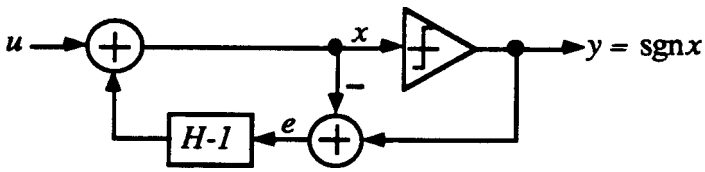

Figure A.1: A ΣΔ modulator with an error transfer function H and a signal transfer function of one.

An alternative, more precise, description defines the x, y and e sequences in terms of u and h according to the following nonlinear recursion:

$$x(t) = \begin{cases} u(0) & t = 0 \\ u(t) + \sum_{i=1}^{t} h(i)e(t-i) & t > 0 \end{cases} \tag{A.1}$$

$$y(t) = \mathrm{sgn}(x(t)) \tag{A.2}$$

$$e(t) = y(t) - x(t)$$

These equations mimic the behavior of Figure A.1 when the initial state of the H-l block is zero.

For compactness, we also write the above as

$$x = u + (h - \delta) * e \tag{A.4}$$

$$Y = \mathrm{sgn}(x) \tag{A.5}$$

$$e = y - x. \tag{A.6}$$

and further economy of notation can be had by writing A.4 as

$$x = u + (H - I)e. \tag{A.7}$$

where 1 represents the identity system.

By A.6, y = x + ***e***, and we can use A.7 to arrive at a familiar description of a ΣΔ modulator:

$$y = u + He. \qquad \text{(A.8)}$$

Note that, for a fiied ***h,*** equations A.l-A.3 provide definitions for functions from ***u*** to x, y and ***e,*** and that these function are well-defined. That is, there is a unique sequence x such that

$$x = u + (H - I)[\mathrm{sgn}(x) - x].$$

Throughout this appendix, we will consider ***h*** to be fixed, and define x, y and e in terms of ***u*** according to A.l-A.3. In addition, the sequences x', ***y'*** and ***e'*** are defined in terms of an input ***u' (h*** is fixed) in a manner analagous to A. 1-A.3.

Lemma 1: Let ***a be*** a sequence. Then there exists a unique sequence ***b*** such that ***a=Hb.***

Paraphrased, this lemma states that H^{-1} exists.

Proof:

Recall that $h(0) = 1$. Then

$$a = H\,b$$

$$\Leftrightarrow \quad a(t) = \sum_{i=0}^{t} h(i)b(t-i)$$

$$\Leftrightarrow \quad b(t) = \begin{cases} a(0) & t = 0 \\ a(t) - \sum_{i=1}^{t} h(i)b(t-i) & t > 0 \end{cases}$$

QED

To arrive at the final important equation relating the signals in a ΣΔ modulator, we substitute A.6 into A.7,

$$x = u + (H - 1)y - (H - 1)x$$

rearrange,

$$Hx = u + (H - 1)y$$

and apply Lemma 1

$$x = H^{-1}u + \left(1 - H^{-1}\right)y \qquad \text{CA.9)}$$

Note that ifx and y are sequences satisfying A.5 and A.9, then they are indeed the x and y sequences corresponding to the input u in a ΣΔ modulator parameterized by h.

Definition: The ΣΔ modulator with input u is stable if $\|x\|_\infty$ c ∞.

Definition: The ΣΔ modulator with input u is *predictable ii* $\exists \varepsilon > 0$ s. t. $\|u' - u\|_\infty < \varepsilon \Rightarrow$ $y' = y$ and $\|x'\|_\infty < \infty$.

The paraphrased version of this definition is that changes in the input which are small enough don't affect the output and don't cause the modulator to become unstable.

Remark Predictability requires that the modulator be stable for the given input. To see this, *take* \s\do2($u' = u$), observe that \s\do2($x' =$ x) and conclude by the definition of predictability that $\|x\|_\infty$ c ∞.

Theorem 1: If the ΣΔ modulator is predictable with some input u, then H^{-1} is bounded-input bounded-output (BlBO) stable.

P *roof* (by contradiction):

Assume that the modulator is predictable with input u, but H^{-1} is not BIB0 stable.

H^{-1} unstable $\Rightarrow \exists v \in \ell_\infty$ s.t. $\left\|H^{-1}v\right\|_\infty = \infty$.

Predictibility $\Rightarrow \exists \varepsilon > 0$ s. t. $\|u' - u\|_\infty < \varepsilon \Rightarrow$ y' = y and $\|x'\|_\infty < \infty$.

Choose such an ε and set $u' = u + \dfrac{\varepsilon}{\|v\|_\infty + 1} v$.

Then $\|u' - u\|_\infty < \varepsilon$ and by predictability we conclude $y' = y$ and $\|x'\|_\infty < \infty$.

But by A.9,

$$\begin{aligned} x' &= H^{-1}u' + \left(1 - H^{-1}\right)y' \\ &= H^{-1}u + \frac{\varepsilon}{\|v\|_\infty + 1} H^{-1}v + \left(1 - H^{-1}\right)y \\ &= x + \frac{\varepsilon}{\|v\|_\infty + 1} H^{-1}v. \end{aligned}$$

And by the triangle inequality,

$$\begin{aligned} \|x'\|_\infty &\geq \frac{\varepsilon}{\|v\|_\infty + 1} \|H^{-1}v\|_\infty - \|x\|_\infty \\ &= \infty \end{aligned}$$

Which contradicts $\|x'\|_\infty < \infty$.

QED

When interpreting the foregoing proof, the reader must not be misled into believing that H^{-1} unstable will imply that $\|x\|_\infty = \infty$ for almost all inputs. The conclusion $\|x\|_\infty = \infty$ came from assuming that y' = y, and this assumption is wrong when H^{-1} is unstable.

What happens if H^{-1} is stable? Is the modulator predictable? Yes, if x is bounded away from zero.

Theorem 2: If H^{-1} is BIB0 stable, $\|u\|_\infty$ c ∞ and x is such that $\inf_{t \geq 0} |x(t)| > 0$, then the modulator with input u is predictable.

Proof:

Let $a = \inf_{t \geq 0} |x(t)|$ and let h_1 be the impulse response of H^{-1}.

Since H^{-1} is BIBO stable, $\|h_1\|_1 < \infty$.

Let $\varepsilon = \frac{a}{\|h_1\|_1}$. Then a > 0 and $\|h_1\|_1$ c ∞ imply $\varepsilon > 0$.

Let u' be any sequence s.t. $\|u' - u\|_\infty < \varepsilon$.

Defme x_1 by

$$\begin{aligned} \text{x,} &= H^{-1}u' + (1 - H\text{-}')y \\ &= x + H\text{-}'(d\text{-}u). \end{aligned}$$

But, $\forall t$

$$|H^{-1}(u' - u)(t)| \leq \|h_1\|_1 \|u' - u\|_\infty < a \leq |x(t)|$$

THUS

$$\begin{aligned} \operatorname{sgn}(x_1(t)) &= \operatorname{sgn}\left(x(t) + H^{-1}(u' - u)(t)\right) \\ &= \operatorname{sgn}(x(t)) \\ &= y(t) \end{aligned}$$

Consequently x' and y satisfy the describing equations A.9 and A.5 and we conclude that

$$\begin{aligned} x_1 &= x' \\ y &= y' \ . \end{aligned}$$

It remains to show that $\|x'\|_\infty$ c ∞.

H^{-1} BIBO stable, $\|u\|_\infty$ c ∞ and $\|y\|_\infty$ = I imply

$$\begin{aligned} |x\|_\infty &= \left\|H^{-1}u + (1 - H^{-1})y\right\|_\infty \\ &< \infty \end{aligned}$$

Thus

$$\|x'\|_\infty \le \|x\|_\infty + \left\|H^{-1}(u'-u)\right\|_\infty$$
$$\le \|x\|_\infty + a < \infty$$

QED

Are there inputs u which satisfy the conditions of Theorem 2?

Yes. From Section 3.2, any input with $u(t) = \pm 1 \; \forall t \ge 0$ has $x = u$ and consequently $\forall t \ge 0, \; |x(t)| = 1 > 0$.

Other examples are described below, where we consider the concepts of limit-cycles and their persistence.

Definition: ***A*** sequence y is ***eventually*** periodic ***with period $T>0$ if*** $\exists t_1$ s. t. $t \ge t_1 \implies y(t+T) = y$

Definition: A ***limit-cycle*** is the periodic part of an eventually periodic output.

Remark: In much of the literature, the definition of a limit-cycle is restricted to periodic outputs for zero or constant inputs. We make the extension to arbitrary inputs so that we can talk about limit-cycles in bandpass modulators, wherein the analog of a constant input is a sine wave at the center frequency.

Definition: The modulator with input u is eventually ***predictable if***
$\exists \varepsilon > 0$ and $\exists t_1$ s. t. $\left(\|u'-u\|_\infty \text{ c } \varepsilon \text{ and } \forall t < t_1 \; u'(t) = u(t)\right) \implies (y' = y \text{ and } \|x'\|_\infty < \infty)$.

Definition: Assume the modulator with input u is eventually predictable and the output y is eventually periodic. Then the periodic part of y is called a ***persistent*** limit cycle.

Discussion: We know from Theorem 1 that when H^{-1} is not BIBO stable, the modulator is not predictable for any input, and hence limit-cycles cannot be persistent. In this case, arbitrarily small errors in the input get magnified by H^{-1} and will eventually cause the output to break the limit-cycle. Subsequently the output may appear to settle into another

repetitive pattern, perhaps a shifted version of the supposed limit-cycle, but this cannot persist for very long. Eventually, noise at the input will again become the dominant component of x and cause the new output pattern to be broken. As this happens repeatedly in a random, unpredictable manner, the output is, in the ordinary sense of the word, unpredictable.

If these random events occur sufficiently often, then the output will appear chaotic. If these random events happen sufficiently rarely, then the output may appear periodic over short intervals

Lemma 2: If u and y are eventually periodic with period T and H^{-1} is BIB0 stable, then x approaches a ***period-T*** sequence. Furthermore, $\|x\|_\infty < \infty$.

Plausibility Argument:

This follows directly from the linear equation $x = H^{-1}u + \left(1 - H^{-1}\right)y$, by appealing to the well-established notion of the steady-state solution in a stable linear system. Note that since u is eventually periodic, $\|u\|_\infty < \infty$, and consequently $\|x\|_\infty < \infty$.

Definition: Denote by $\tilde{x}$ the **period-T** sequence to which x converges in Lemma 2.

Theorem 3: If H^{-1} is BIB0 stable, u and y are eventually periodic and $\tilde{x}$ satisfies $\forall t \;\; \tilde{x}(t) \neq 0$, then the periodic part of y is a persistent limit-cycle.

Proof:

The proof parallels that of Theorem 2. We must show that the modulator with input u is eventually predictable.

Since $\tilde{x}$ is periodic and $\forall t \; \tilde{x}(t) \neq 0, \min \tilde{x}(t) > 0$. Let $a = \min \tilde{x}(t)$.

Let $\varepsilon = \frac{a}{2\|h_1\|_1}$, where, as before, h_1 is the impulse response of H^{-1}.

For the same reasons as in Theorem 2, $\varepsilon > 0$.

Choose t_1 large enough that $\forall t \geq t_1, |\tilde{x}(t) - x(t)|$ c $\frac{a}{2}$. Then $\forall t \geq t_1$,

$$|x(t)| = |\tilde{x}(t) - (\tilde{x}(t) - x(t)]$$
$$\geq |\tilde{x}(t)| - |\tilde{x}(t) - x(t)| > a - \frac{a}{2} = \frac{a}{2}.$$

Let u' be any sequence s. t. $\|u' - u\|_\infty < \varepsilon$ and $\forall t$ c $t_1 u'(t) = u(t)$.

Defme x_1 by

$$\text{x,} = H^{-1}u' + (1 - H\text{-}')Y$$
$$= x + H\text{-}'(d\text{-}u).$$

But, $\forall t \geq t_1$

$$|H^{-1}(u' - u)(t)| \leq \|h_1\|_1 \|u' - u\|_\infty < \frac{a}{2} < |x(t)|.$$

Thus

$$\operatorname{sgn}(x'(t)) = \operatorname{sgn}(x(t) + H^{-1}(u' - u)(r))$$
$$= \operatorname{sgn}(x(t))$$
$$= y(t)$$

Consequently x' and y satisfy the describing equations A.9 and A.5 and we conclude that

$$x_1 = x'$$
$$y = y' \ .$$

Once again, H^{-1} BIBO stable, $\|u\|_\infty$ c ∞ and $\|y\|_\infty = 1$ imply

$$\|x\|_\infty = \|H^{-1}u + (1 - H^{-1})y\|_\infty$$
$$< \infty$$

So that

$$\|x'\|_\infty \le \|x\|_\infty + \left\|H^{-1}(u'-u)\right\|_\infty$$
$$\le \|x\|_\infty + \frac{a}{2} < \infty.$$

QED

Theorem 3 shows that in order to establish a particular finite sequence of plus and minus ones as a persistent limit cycle for a given input, three things suffice:

1) *H-l* stable.
2) Consistency: the eventually periodic input produces an output whose periodic part is the desired limit cycle.
3) The steady-state value of x is non-zero at all times.

With this in mind, let us now re-examine the numerical example of Section 3.3 in more detail The perturbed modulator has an error transfer function given by

$$H(z) = \frac{z^2 - 0.9\sqrt{2}z + 0.9^2}{z^2}$$

which has zeros at $z = 0.9e^{\pm j\frac{\pi}{4}}$, inside the unit circle, so H^{-1} is stable and we suspect that persistent limit cycles exist for this modulator. In order to verify our suspicion, we need to fmd an eventually periodic input which gives rise to an output that is eventually periodic.

First, let us fmd the steady-state value of x for the period-24 limit-cycle of Section 3.3

$$y = (+++-+--+-++---+-++-+--),$$

where + denotes *+1* and − denotes *−1*, when the input is a sampled sine wave of amplitude *A* and frequency $\pi/4$. Strictly speaking, a tilde (~) should be atop all the signals since we will only be dealing with their periodic parts. To reduce clutter, this diacritical mark is

omitted.

From A.9,

$$\begin{aligned} x &= H^{-1}u + (1 - H^{-1})y \\ &= AH^{-1}\sin\left(\frac{\pi}{4}t\right) + (1 - H^{-1})y \\ &= Ax_1 + x_2 \end{aligned}$$

The +-component is given by

$$x_1 = \left|H^{-1}\left(e^{j\frac{\pi}{4}}\right)\right| \sin\left(\frac{\pi}{4}t + \arg\left(H^{-1}\left(e^{j\frac{\pi}{4}}\right)\right)\right),$$

where $\left|H^{-1}\left(e^{j\frac{\pi}{4}}\right)\right| = 7.4329$ and $\arg\left(H^{-1}\left(e^{j\frac{\pi}{4}}\right)\right) = -0.7328$.

We can compute the x_2 -component via the discrete Fourier transform.

Let **Y** be the length-24 DFT of y.

$$Y(n) = \sum_{t=0}^{23} e^{-j\frac{2\pi n}{24}t} y(t)$$

Defme X_2 likewise.

Then X_2 is related to Y by

$$X_2(n) = \left(1 - H\left(e^{j\frac{2\pi n}{24}}\right)^{-1}\right) Y(n)$$

Table A. 1 gives the numerical results of the DFT-based calculations.

n	$H\left(e^{j\frac{2\pi n}{24}}\right)$	$Y(n)$	$X_2(n)$
0	0.5372	0	0
1	0.4721-0.0756i	3.8637-1.0353i	-4.4589-0.1746i
2	0.3027-0.0651i	0	0
3	0.1000+0.0900i	0.3431-8.8284i	42.3456+41.6537i
4	-0.0414+0.4008i	0	0
5	-0.0309+0.8244i	1.0353-3.8637i	5.7623-2.7851i
6	0.1900+1.2728i	0	0
7	0.6279+1.6344i	-1.0353-3.8637i	1.2367-3.6242i
8	1.2314+1.8038i	0	0
9	1.9000+1.7100i	11.6569+3.1716i	7.4372+5.3000i
10	2.5073+1.3379i	0	0
11	2.9309+0.7344i	-3.8637-1.0353i	-2.5400-1.0137i
12	3.0828+0.0000i	0	0
13	2.9309-0.7344i	-3.8637+1.0353i	-2.5400+1.0137i
14	2.5073-1.3379i	0	0
15	1.9000-1.7100i	11.6569-3.1716i	7.4372-5.3000i
16	1.2314-1.8038i	0	0
17	0.6279-1.6344i	-1.0353+3.8637i	1.2367+3.6242i
18	0.1900-1.2728i	0	0
19	-0.0309-0.8244i	1.0353+3.8637i	5.7623+2.7851i
20	-0.0414-0.4008i	0	0
21	0.1000-0.0900i	0.3431+8.8284i	42.3456-41.6537i
22	0.3027+0.0651i	0	0
23	0.4721+0.0756i	3.8637+1.0353i	-4.4589+0.1746i

Table A.1: **The** calculation of X_2 via the discrete Fourier transform.

The sequence x_2 may be calculated from X_2 via the inverse discrete Fourier transform:

$$x_2(t) = \frac{1}{24}\sum_{n=0}^{23} e^{j\frac{2\pi}{24}n} X_2(n)$$

Table A.2 lists the values of they, x, and x_2 sequences.

t	$u=\sin(\frac{\pi}{4}t)$	y	x_1	x_2	Condition on A
0	0	1	-4.9723	4.1486	A<0.8343
1	0.7071	1	0.3908	-0.2251	A>0.5760
2	1	1	5.5249	-4.1096	A>0.7438
3	0.7071	1	7.4226	-5.5112	A>0.7425
4	0	-1	4.9723	-4.1486	A<0.8343
5	-0.7071	1	-0.3908	1.2666	A<3.2412
6	-1	-1	-5.5249	2.8896	A>0.5230
7	-0.7071	-1	-7.4226	4.7348	A>0.6379
8	0	1	-4.9723	4.1486	A<0.8343
9	0.7071	-1	0.3908	-0.6377	A<1.6319
10	1	1	5.5249	-2.0892	A>0.3781
11	0.7071	1	7.4226	-4.2254	A>0.5693
12	0	-1	4.9723	-4.1486	A<0.8343
13	-0.7071	-1	-0.3908	0.2251	A>0.5760
14	-1	-1	-5.5249	4.1096	A>0.7438
15	-0.7071	-1	-7.4226	5.5112	A>0.7425
16	0	1	-4.9723	4.1486	A<0.8343
17	0.7071	-1	0.3908	-1.2666	A<3.2412
18	1	1	5.5249	-2.8896	A>0.5230
19	0.7071	1	7.4226	-4.7348	A>0.6379
20	0	-1	4.9723	-4.1486	A<0.8343
21	-0.7071	1	-0.3908	0.6377	A<1.6319
22	-1	-1	-5.5249	2.6892	A>0.3781
23	-6.7671	-1	-7.4226	4.2254	A>0.5693

Table A.2: **The** calculation of the two components of $\boldsymbol{x}$ and hence the allowable limits on A.

In order for the supposed limit-cycle to exist, we need to ensure that the consistency condition, A.5, holds. We want

$$y = \mathrm{sgn}(x)$$

where

$$x = Ax_1 + x_2.$$

If $\mathrm{sgn}(y) = \mathrm{sgn}(x_1)$, then A must be large enough to ensure that the term Ax_1 dominates. The critical value of A is that which sets x to zero: $-x_2/x_1$.

Likewise, if $\mathbf{sgn}(y) \neq \mathbf{sgn}(x_1)$, then *A* must be small enough to ensure that the x_2-term dominates. Once again, the critical value of *A* is that which sets x to zero: $-x_2/x_1$.

Table A.2 lists the conditions on *A* required for consistency at each time step. The overall result is that the limit-cycle can be supported by any sine wave whose amplitude falls in the range $(0.7438, 0.8343)$.

This is almost all that we need in order to prove that y is indeed a persistent limit-cycle for sine wave inputs of the appropriate amplitude. The remaining detail requires that we fmd an input which takes the modulator from an initial state of zero to a state which will support the limit cycle. For this detail, we take the pragmatic viewpoint that simulations have shown that for this modulator and this limit-cycle, a sine wave input of amplitude $A=0.75$, 0.76, 0.77, 0.78, 0.79, 0.80, 0.81, 0.82, or 0.83 will eventually cause the output to settle into the limit-cycle under consideration.

Bibliography

[Agrawal and Shenoi 1983]
B. P. Agrawal and K. Shenoi, "Design Methodology for ΣΔM," ***IEEE Transactions on Communications,*** vol. COM-3 1, pp. 360-370, March 1983.

[Anastassiou 1989]
D. Anastassiou, "Error Diffusion Coding for A/D Conversion," ***IEEE Transactions on Circuits and Systems,*** vol. CAS-36, pp. 11751186, Sept. 1989.

[Ardalan and Paulos 1987]
S. H. Ardalan and J. J. Paulos, "An Analysis of Nonlinear Behavior in Delta-Sigma Modulators," ***IEEE Transactions on Circuits and Systems,*** vol. CAS-34, pp. 593-603, June 1987.

[Atherton 1981]
D. P. Atherton, ***Stability of Nonlinear Systems.*** Chichester: Research Studies Press, 1981.

[Boser and Wooley 1988]
B. E. Boser and B. A. Wooley, "The Design of Sigma-Delta Modulation Analog-to-Digital Converters," ***IEEE Journal of Solid-State Circuits,*** vol. ***23,*** no. ***6,*** pp. 1298-1308, Dec 1988.

[Candy 1974]
J. C. Candy, "A Use of Limit Cycle Oscillations to Obtain Robust Analog-to-Digital Converters," ***IEEE Transactions on Communications, vol. COM-22,*** no. ***3,*** pp. 298-305, March 1974.

[Candy, Wooley and Benjamin 1981]
J. C. Candy, B. A. Wooley and 0. J. Benjamin, "A Voiceband Codec with Digital Filtering," ***IEEE Transactions on Communications,*** vol. COM-29, no. ***6,*** pp. 815-830, June 1981.

[Candy and Benjamin 1981]
J. C. Candy and 0. J. Benjamin, "The Structure of Quantization Noise from Sigma-Delta Modulation", ***IEEE Transactions on Communications,*** vol. COM-29, no. 9, pp. 1316-1323, Sept. 1981.

[Candy 1985]
J. C. Candy, "A Use of Double Integration in Sigma-Delta Modulation," ***IEEE Transactions on Communications,*** vol. COM-33, no. 3, pp. 249-258, March 1985.

[Candy 1986]
J. C. Candy, "Decimation for Sigma Delta Modulation," ***IEEE Transactions on Communications,*** vol. COM-34, no. 1, pp. 72-76, Jan. 1986.

[Candy and Huynh 1986]
J. C. Candy and A. Huynh, "Double Interpolation for Digital- to-Analog Conversion," ***IEEE Transactions on Communications,*** vol. COM-34, no. 1, pp. ***77-8 1,*** Jan. 1986.

[Chao, Nadeem, Lee and Sodini 1990]
K. C.-H. Chao, S. Nadeem, W. L. Lee and C. G. Sodini, "A Higher Order Topology for Interpolative Modulators for Oversampling A/D Converters" IEEE ***Transactions on Circuits and Systems,*** vol. CAS-37, no. 3, pp. 309-318, March 1990.

[Chou and Gray 1990]
W. Chou and R. M. Gray, "Dithering and Its Effects on Sigma Delta and Multistage Sigma Delta Modulation," ***Proceedings of the 1990 IEEE International Symposium on Circuits and Systems,*** vol. ***3,*** pp. 368-371, May 1990.

[Del Signore, Kerth, Sooch and Swanson 1990]
B. P. Del Signore, D. A. Keith, N. S. Sooch and E. J. Swanson, "A Monolithic 20-b Delta-Sigma A/D Converter," ***IEEE Journal of Solid-State Circuits,*** vol. ***25,*** no. ***6,*** pp. 1311-1317, Dec. 1990.

[Diniz and Antoniou 1985]
P. S. R. Diniz and A. Antoniou, "Low-sensitivity digital-filter structures which are amenable to error-spectrum shaping" ***IEEE Transactions on Circuits and Systems,*** vol. CAS-32, no. 10, pp. 1000-1007, October 1985.

[Ferguson Ganesan and Adams 1990]
P. F. Ferguson, A. Ganesan and R. W. Adams, "One Bit Higher Order Sigma-Delta AID Converters," ***Proceedings of the 1990 IEEE International Symposium on Circuits and Systems, vol. 2,*** pp. 890-893, May 1990.

[Friedman 1988]
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.

[Gailus, Tumey and Yester 1989]
P. H. Gailus, W. J. Tumey and F. R. Yester, Jr., "Method and Arrangement for a Sigma Delta Converter for Bandpass Signals," US Patent number 4,857,928 August 15, 1989, Motorola, Inc., Schaumburg, Ill.

[Gelb and Vander Velde 1963]
A. Gelb and W. E. Vander Velde, "On Limit Cycling Control Systems," ***IEEE Transactions on Automatic Control, vol. 8,*** pp. 142- 157, April 1963.

[Gleick 198 1]
J. Gleick, Chaos New York, NY: Penguin, 1981.

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

[Gray 1989]
R. M. Gray, "Spectral Analysis of Quantization Noise in a Single-Loop Sigma-Delta Modulator with dc Input," ***IEEE Transactions on Communications,*** vol. COM-37, no. 6, pp. 588-599, June 1989.

[Gray, Chou and Wong 1989]
R. M. Gray, W. Chou and P. W. Wong, "Quantization Noise in Single-Loop Sigma-Delta Modulation with Sinusoidal Inputs," ***IEEE Transactions on Communications, vol. COM-37,*** no. 9, pp. 956-968, September 1989.

[Hauser 199 1]
M. W. Hauser, "Principles of Oversampling A/D Conversion," ***Journal of the Audio Engineering Society,*** vol. 39, no. 1/2, pp. 3-26, January/February 1991.

[Hawksford 1985]
M. J. Hawksford, "N-th order recursive sigma-ADC machinery at the analogue-digital gateway," presented at Audio Engineering Society Convention, May 1985.

[He, Kuhlmann and Buzo 1990]
N. He, F. Kuhlmann and A. Buzo, "Double-Loop Sigma-Delta Modulation with dc Input," ***IEEE Transactions on Communications,*** vol. COM-38, no. 4, pp. 487-495, April 1990.

[Hein and Zakhor 1991]
S. Hein and A. Zakhor, "On the Stability of Interpolative Sigma Delta Modulators," ***Proceedings of the 1991 IEEE International Symposium on Circuits and Systems,*** vol. 3, pp. 1621-1624, June 1991.

[Horrocks 1991]
D. H. Horrocks, "A Second-order Oversampled Sigma-Delta Modulator for Bandpass Signals," ***Proceedings of the 1991 IEEE International Symposium on Circuits and Systems,*** vol. 3, pp. 1653-1656, June 1991.

[Höfelt 1979]
M. H. H. Höfelt, "On the stability of a l-bit quantized feedback system," ***Proceedings of the 1979 IEEE International Conference on Acoustics, Speech and Signal Processing,*** Washington, DC, pp. 844-848, May 1979.

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

[Inose and Yasuda 1963]
H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback," ***Proceedings of the IEEE,*** vol. 51, pp. 1524-1535, Nov. 1963.

[Jantzi, Schreier and Snelgrove 1991]
S. A. Jantzi, R. Schreier and M. Snelgrove, "A Bandpass $\Sigma\Delta$ A/D Convertor for a Digital AM Receiver," ***accepted at IEE 1991 ADDAC, Swansea,*** UK.

[Jury and Lee 1964]
E. I. Jury and B. W. Lee, "On the Stability of a Certain Class of Nonlinear Sampled-Data Systems," ***IEEE Transactions on Automatic Control,*** vol. AC-9, no. 1, pp. 51-61, Jan. 1964.

[Kochenburger 1950]
R J. Kochenburger, "A frequency response method for analyzing and synthesizing contactor servomechanisms," *Trans. AIEE,* vol. 69 (*Appl and Ind., pt. I),* pp. 270-284, 1950.

[Larsen, Cataltepe and Temes 1988]
L. E. Larsen, T. Cataltepe and G. C. Temes, "Multi-bit Oversampled ΣΔ A/D Converter with Digital Error Correction," *Electronics Letters,* vol. *24,* pp. 1051-1052, Aug. 1988.

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

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

[Leslie and Singh 1990]
T. C. Leslie and B. Singh, "An Improved Sigma-Delta Modulator Architecture", *Proceedings of the I990 IEEE International Symposium on Circuits and Systems,* vol. 1, pp. 372-375, May 1990.

[Lindorff 1965]
D. P. Lindorff, *Theory of Sampled-Data Control Systems,* Chapter 9. New York, NY: John Wiley & Sons, 1965.

[Norsworthy, Post and Fetterman 1989]
S. R. Norsworthy, I. G. Post and H. S. Fetterman, "A 14-bit 80-kHz Sigma-Delta A/D Converter: Modeling, Design and Performance Evaluation," *IEEE Journal of Solid-State Circuits,* vol. *SC-24,* pp. 256-266, April 1989.

[Norsworthy 1990]
S. R. Norsworthy, "Oversampled Sigma-Delta Data Converters," Pm-conference tutorial at the 1990 IEEE International Symposium on Circuits and Systems, New Orleans, LA, April 30 1990.

parker and Chua 1987]
T. S. Parker and L. 0. Chua, "Chaos: A Tutorial for Engineers," *Proceedings of the IEEE, vol. 75,* no. 8, pp. 982-1008, August 1987.

[Rachid 199 1]
A. Rachid, "Positively Invariant Polyhedral Sets for Uncertain Discrete Time Systems," *Control Theory and Advanced Technology,* vol. 7, no. 1, pp 191-200, March 1991.

[Ritoniemi, Karema and Tenhunen 1990]
T. Ritoniemi, T. Karema and H. Tenhunen, "The Design of Stable High Order 1-Bit Sigma-Delta Modulators," *Proceedings of the 1990 IEEE International Symposium on Circuits and Systems, vol. 4,* pp 3267-3270, May 1990.

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

[Schreier and Snelgrove 1990]
R Schreier and W. M. Snelgrove, "Decimation for Bandpass Sigma-Delta Analog-to-Digital Conversion," *Proceedings of the 1990 IEEE International Symposium on Circuits and Systems,* vol. *3,* pp. 1801-1804, May 1990.

[Schreier and Snelgrove 199 1 a]
R. Schreier and M. Snelgrove, "Stability in a General ΣΔ Modulator," *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech and Signal Processing,* vol. *3,* pp. 1769- 1772, May 199 1.

[Schreier and Snelgrove 1991b]
R. Schreier and W. M. Snelgrove, "ΣΔ Modulation is a Mapping," *Proceedings of the 1991 IEEE International Symposium on Circuits and Systems,* vol. *5,* pp. 2415-2418, June 1991.

[Stonick, Rulla, Ardalan and Townsend 1990]
J. T. Stonick, J. L. Rulla, S. H. Ardalan and J. K. Townsend, "A New Architecture for Second Order Σ – A Modulation", *Proceedings of the I990 IEEE International Symposium on Circuits and Systems,* vol. 1, pp. 360-363, May 1990.

[Smith 1966]
H. W. Smith, ***Approximate analysis of randomly excited nonlinear controls.*** Cambridge, Massachusetts: The M.I.T. Press, 1966.

[Stacey, Frost and Ware 1991]
N. D. Stacey, R. L. Frost and G. A. Ware, "Error Spectrum Shaping Quantizers with Non-Ideal Reconstruction Filters and Saturating Quantizers", *Proceedings of the 1991 IEEE International Conference* on *Acoustics, Speech and Signal Processing,* vol. *3,* pp. 1905-1908, May 1991.

[Tewksbury and Hallock 1978]
S. K. Tewksbury and R. W. Hallock, "Oversampled, linear predictive and noise-shaping coders of order *N*>1" *IEEE Transactions on Circuits and Systems, vol. CAS-25,* no. *7,* pp. 436-447, July 1978.

[Uchimura, Hayashi, Kimura and Iwata 1988]
K. Uchimura, T. Hayashi, T. Kimura and A. Iwata, "Oversampling A-to-D and D-to-A Converters with Multistage Noise Shaping Modulators", *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 36, no. 12, pp. 1899-19005 , Dec. 1988.

[Wolff and Carley 1989]
C. Wolff and L. R. Carley, "Calculating the stability range, SNR and distortion of delta-sigma modulators", *Proceedings of the 1989 IEEE International Symposium on Circuits and Systems,* vol. *2,* pp. 1423-1426, May 1989.

[Wong and Gray 1990]
P. W. Wong and R. M. Gray, "Two-Stage Sigma-Delta Modulation", *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 38, no. 11, pp. 1937-1952 , *Nov.* 1990.