

1992

# Characterization of quantization noise in oversampled analog to digital converters

Eric W. Multanen  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Signal Processing Commons](#)

Let us know how access to this document benefits you.

---

## Recommended Citation


Multanen, Eric W., "Characterization of quantization noise in oversampled analog to digital converters" (1992). *Dissertations and Theses*. Paper 4424.  
<https://doi.org/10.15760/etd.6302>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

AN ABSTRACT OF THE THESIS OF Eric W. Multanen for the Master of Science in Electrical Engineering presented October 29, 1992.

Title: Characterization of Quantization Noise in Oversampled Analog to Digital Converters.


APPROVED BY THE MEMBERS OF THE THESIS COMMITTEE:



Y. C. Jenq, Chair



Robert Daasch



Faryan Etesami

The analog to digital converter (ADC) samples a continuous analog signal and produces a stream of digital words which approximate the analog signal. The conversion process introduces noise into the digital signal. In the case of an ideal ADC, where all noise sources are ignored, the noise due to the quantization process remains. The resolution of the ADC is defined by how many bits are in the digital output word. The amount of quantization noise is clearly related to the resolution of the ADC. Reducing the quantization noise results in higher effective resolution.

The traditional method of increasing resolution by increasing the number of

levels in the quantizer becomes impractical when the resolution becomes high due to limitations in analog circuit technology. A popular method of circumventing these limitations is to oversample the input using a low resolution quantizer and increase the effective resolution of the digital signal with digital signal processing.

An important factor in the analysis of oversampled ADC's is the characteristics of the quantization noise. It is often assumed that the quantization noise has characteristics similar to white noise. This assumption makes analysis of oversampled ADC's simpler, but it can provide misleading results.

This study focuses on the digital signal produced by two types of oversampled ADC's. A low resolution traditional ADC and the sigma delta modulator, which uses a one bit quantizer and a feedback configuration. Simulated signal and noise performance is compared against predicted performance based on the white noise assumptions. Further work is done to develop an understanding of the quantization noise dynamics for the different types of ADC configurations. Comparisons are made with theoretical models which provide exact descriptions of the the quantization noise.

It is discovered that the low resolution traditional ADC performance is very different than the performance predicted using the white noise assumption. The performance of the sigma delta modulator compares much better to the predictions made with the white noise assumption. However, simulations and exact theories point out areas of signal amplitudes and frequencies where the performance is not as good.

CHARACTERIZATION OF QUANTIZATION NOISE IN OVERSAMPLED  
ANALOG TO DIGITAL CONVERTERS

by  
ERIC W. MULTANEN

A thesis submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE  
in  
ELECTRICAL ENGINEERING

Portland State University  
1992

TO THE OFFICE OF GRADUATE STUDIES

The members of the Committee approve the thesis of Eric W. Multanen  
presented October 29, 1992.

[Redacted Signature]

Y. C. Jenq, Chair

[Redacted Signature]

Robert Daasch

[Redacted Signature]

Faryar Etesami

APPROVED:

[Redacted Signature]

Rolf Schaumann, Chair, Department of Electrical Engineering

[Redacted Signature]

Roy W. Koch, Vice Provost for Graduate Studies and Research

## TABLE OF CONTENTS

	PAGE
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	vii
CHAPTER	
I INTRODUCTION TO ANALOG TO DIGITAL CONVERSION	1
I.1 Sampling Fundamentals . . . . .	3
I.2 ADC Noise Sources . . . . .	5
I.3 Reducing Quantization Noise . . . . .	7
I.4 Oversampled ADC's . . . . .	8
I.5 Scope of Study . . . . .	11
II UNIFORM ADC THEORY AND SIMULATIONS . . . . .	14
II.1 The White Noise Assumption . . . . .	14
II.2 Oversampled Uniform ADC's . . . . .	17
III SIMULATION TECHNIQUES . . . . .	20
III.1 Simulation Methods and Techniques . . . . .	20
III.2 The Simulation Process . . . . .	21
III.3 Simulation of Input Signal . . . . .	22
III.4 Computing SQNR and Total Noise Power . . . . .	23

III.5	Computing the Power Spectrum . . . . .	24
III.6	Filtering the Output Signal . . . . .	26
IV	SIMULATIONS OF THE UNIFORM ADC . . . . .	30
IV.1	Simulation Results . . . . .	30
IV.2	An Improved Uniform ADC Theory . . . . .	32
V	THE SIGMA DELTA MODULATOR . . . . .	41
V.1	Basic $\Sigma\Delta M$ Operation . . . . .	42
V.2	Formal Analysis of the $\Sigma\Delta M$ . . . . .	46
V.3	$\Sigma\Delta M$ Analysis Using White Noise Assumption . . . . .	52
VI	EXACT DC ANALYSIS OF THE $\Sigma\Delta M$ . . . . .	54
VI.1	$\Sigma\Delta M$ Output Signal Structure for dc Inputs . . . . .	55
VI.2	Recursive Structure of Output Signal . . . . .	57
VI.3	$\Sigma\Delta M$ Noise Characteristics for dc Inputs . . . . .	60
VI.4	Fundamental Definitions and Results . . . . .	63
VI.5	Moments of Irrational Inputs . . . . .	64
VI.6	Moments of Rational Inputs . . . . .	66
VI.7	Computing the Spectrum and the Bohr-Fourier Series . . . . .	67
VI.8	Spectrum Results for Irrational and Rational Inputs . . . . .	69
VI.9	Equality of Irrational and Rational Results . . . . .	69
VI.10	Total Quantization Error Spectrum . . . . .	71
VII	SIMULATION OF $\Sigma\Delta M$ WITH DC INPUTS . . . . .	72
VII.1	Setup of Simulations . . . . .	72

VIII	EXACT ANALYSIS OF $\Sigma\Delta M$ WITH SINUSOID INPUTS . . .	82
	VIII.1 Exact Analysis for Sinusoidal Inputs . . . . .	82
IX	SIMULATION OF THE $\Sigma\Delta M$ WITH SINUSOID INPUTS . . .	90
	IX.1 Setup of Sinusoidal Simulations . . . . .	91
	IX.2 Results of Sinusoidal Simulations . . . . .	92
	IX.3 Simulations and the Exact Theory . . . . .	96
X	CONCLUSION . . . . .	110
	X.1 Summary of Results . . . . .	111
	X.2 Applications and Extensions of Results . . . . .	113
	REFERENCES CITED . . . . .	115
	APPENDIX . . . . .	116



## LIST OF TABLES

TABLE		PAGE
I	Rational and Irrational Theory Results . . . . .	71
II	Total Quantization Noise Power . . . . .	75

## LIST OF FIGURES

FIGURE	PAGE
1. Transfer characteristic of a 3 bit ADC. . . . .	3
2. The discrete time frequency spectrum. . . . .	4
3. Typical ADC circuit configuration. . . . .	5
4. 3 Bit ADC Transfer Function Including Errors. . . . .	7
5. Model of quantizer error in the uniform ADC. . . . .	15
6. Uniform noise spectrum for different $f_s$ . . . . .	19
7. SQNR curves for 4 bit Uniform ADC. . . . .	31
8. SQNR curves for 10 bit Uniform ADC. . . . .	33
9. Spectrum of 4 bit ADC at various $M$ for $f_x = 979Hz$ . . . . .	34
10. Decomposition of highly oversampled sinusoid. . . . .	36
11. SQNR at $M = 128$ for a 4 bit ADC. . . . .	40
12. $\Sigma\Delta M$ circuit diagram. . . . .	42
13. Discrete time analog integrator. . . . .	44
14. $\Sigma\Delta M$ output with sinusoidal input. . . . .	45
15. $\Sigma\Delta M$ discrete time model. . . . .	47
16. $\Sigma\Delta M$ discrete time signal model. . . . .	49
17. $\Sigma\Delta M$ discrete time noise model. . . . .	49
18. Spectrum shaping effect of $T(z)$ . . . . .	50

19.	Example with dc input of $\frac{3}{40}$ . . . . .	61
20.	Total noise power vs dc input. . . . .	76
21.	$\Sigma\Delta M$ noise harmonic positions. . . . .	78
22.	$\Sigma\Delta M$ structural sub-frequencies. . . . .	79
23.	SQNR curves for full scale sine input. . . . .	93
24.	SQNR curves for 80 % full scale sine input. . . . .	94
25.	SQNR surface for $M = 128$ . . . . .	96
26.	Theoretical spectrum $f_x = 8687Hz$ . . . . .	99
27.	Simulated spectrum $f_x = 8687Hz$ . . . . .	100
28.	Theoretical spectrum $f_x = 23537Hz$ . . . . .	100
29.	Simulated spectrum $f_x = 23537Hz$ . . . . .	101
30.	Harmonics located in the signal bandwidth. . . . .	102
31.	Amplitudes of first harmonics in the signal bandwidth. . . . .	105
32.	$N$ vs Amplitude for several frequencies for $M = 256$ . . . . .	109

## CHAPTER I

### INTRODUCTION TO ANALOG TO DIGITAL CONVERSION

This is a study about some of the issues involved in the process of converting an analog signal to a digital signal. An analog signal, for the purposes of this study, will be defined as a continuous voltage signal. A digital signal is defined as a sequence of binary values with a specified number of bits, or word length. The goal of analog to digital conversion is to obtain a sequence of binary numbers which accurately represents the analog input signal. The primary interest of this study is to gain an understanding of the issues that affect the accuracy of the analog to digital conversion process. The important advantage of analog to digital conversion is that the virtually limitless number of phenomena which can be represented by analog voltage signals can be transformed to a signal which is accessible to the power and flexibility of digital computers. Application areas which use analog to digital conversion include audio, video, communications, data acquisition, and many more.

The analog to digital converter, referred to from here on as an ADC, is an electronic device which is used to perform analog to digital conversion. The key component of the ADC is the quantizer, which takes the analog input value and produces a digital output value. There are a couple basic input and output

characteristics of a quantizer which define its operation. One of these is the valid input voltage range. Voltages within the range will be accurately converted to the corresponding digital value. If the input goes outside of the valid range, the output will be clipped at the minimum or maximum digital value.

Another primary characteristic of a quantizer is the resolution of the output. If the digital output word has  $n$  bits, then the quantizer is said to have a nominal  $n$  bit resolution. The resolution of a quantizer indicates that it is capable of resolving the input signal range into  $2^n$  evenly spaced levels. Figure 1 shows the transfer relationship between the analog input and digital output for a quantizer with 3 bit resolution. Since error is introduced into the output when the input is rounded to one of the output levels, the resolution of a quantizer is related to the amount of quantization error that will be present in the output. Higher resolution quantizers will introduce less noise into the output than low resolution quantizers. In general, a quantizer with a specific resolution has an associated nominal amount of quantization noise. In practice, an ADC with an  $n$  bit quantizer may, for any number of reasons, produce an output which has more noise than would be expected for  $n$  bit resolution. In this case, the effective resolution of the ADC is less than the nominal resolution. Since most real world devices are not ideal, the effective resolution of a quantizer will usually be lower than the nominal resolution. On the other hand, this study will be devoted to examining methods for increasing the effective resolution of an ADC above the nominal resolution of the quantizer.

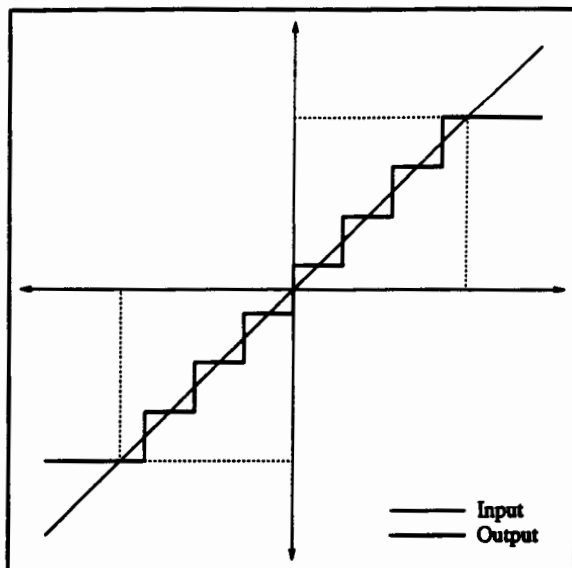


Figure 1. Transfer characteristic of a 3 bit ADC.

## I.1 SAMPLING FUNDAMENTALS

The ADC is typically part of a sampling system in which the digital values must be produced at a specified rate or frequency. The conversion process samples the analog signal to produce a discrete time analog signal which is then converted to a digital signal. The well known nyquist sampling theorem states that a continuous signal must be sampled at a frequency which is at least twice the input signal bandwidth in order for the original signal to be reconstructed, with no distortion, from the discrete signal. The input signal bandwidth is defined as the frequency range from dc to the highest frequency that is considered a valid input. If the highest frequency of the signal bandwidth is  $f_{max}$ , then the nyquist frequency is defined as  $f_n = 2f_{max}$ . Thus, the minimum reasonable value for the sampling frequency,  $f_s$ , is  $f_n$ . The sampling bandwidth will be defined as the frequency

range from dc to  $\frac{f_s}{2}$ .

Since the output signal is only capable of representing frequencies less than  $\frac{f_s}{2}$ , if the input contains frequencies greater than  $\frac{f_s}{2}$ , then these frequencies will appear in the output as lower frequencies. This is known as aliasing. Figure 2 shows the discrete frequency spectrum of a discrete signal. The spectrum is centered around zero, and is periodic with period  $f_s$ . The negative side of the spectrum is identical to the positive side. If the spectrum of the input signal extends beyond  $\frac{f_s}{2}$ , then that part of the input spectrum will be aliased back from  $\frac{f_s}{2}$ . For example, an input frequency of  $\frac{4}{6}f_s$  will be aliased to  $\frac{f_s}{2} - \frac{1}{6}f_s$  in the output. Figure 3 shows a typical ADC circuit configuration. The analog low pass filter at the front end is used to remove frequencies above the signal bandwidth in order to prevent aliasing.

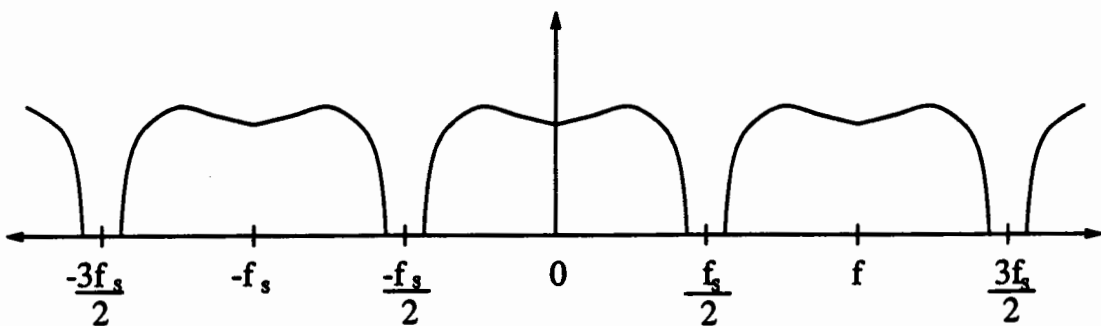


Figure 2. The discrete time frequency spectrum.

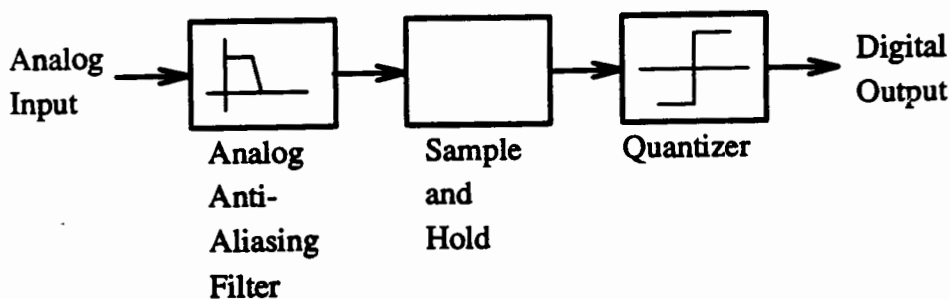


Figure 3. Typical ADC circuit configuration.

## 1.2 ADC NOISE SOURCES

According to the nyquist theorem, a discrete time signal is a perfect representation of the analog signal if the conditions of the theorem are satisfied. However, the digital signal produced by the ADC is not a perfect representation of the analog signal even when the sampling requirements are met. This is a result of the quantization of the analog samples. When the analog sample is converted to a digital word, a small amount of error is introduced. The amount of error will be somewhere in the range of  $\pm \frac{\Delta}{2}$ , where  $\Delta = \frac{1}{2^n - 1}$ . The process of quantization adds noise to the signal. The error caused by the quantization process is referred to as the quantization error or quantization noise.

There are several other error sources which will introduce noise in the output signal. These other sources of noise will be mentioned here and then dismissed for



the rest of the study. If the input signal is outside of the input range, the ADC output will be clipped at its minimum or maximum value. This type of error is known as saturation error. It will be assumed for this study that the input signal has its amplitude limited to fall within the valid input range. Offset error occurs when the digital signal does not match the analog signal by a fixed offset value. That is, if the input line in the transfer characteristic does not pass through the origin, then there is an offset error. Gain error occurs when the output reaches saturation either sooner or later than it should. Gain error can also be explained by noting that the input line in the characteristic will have a slope not equal to one. If the input line in the transfer characteristic is not straight, then the quantizer has a linearity error. Error can also occur if the quantizer levels, or  $\Delta$ 's, are not uniformly spaced. An ADC where all the  $\Delta$ 's are equal is called a uniform ADC. Another source of noise that could occur in an actual ADC could be related to the input low pass analog filter. If the passband of the filter is not flat, then the input will be distorted before it is even quantized. If the passband does not attenuate the signal strongly enough by half of the sampling bandwidth, then high frequency input signals will be aliased to low frequencies, causing yet another source of error. Figure 4 shows an ADC transfer function which includes some of the errors listed here. Many of the errors here are caused by imperfections in the implementation of the ADC. If the quantizer is ideal and the input signal is bounded by the valid input range of the quantizer and the valid frequencies of the signal bandwidth, then all the noise sources can be eliminated except for the quantization noise.

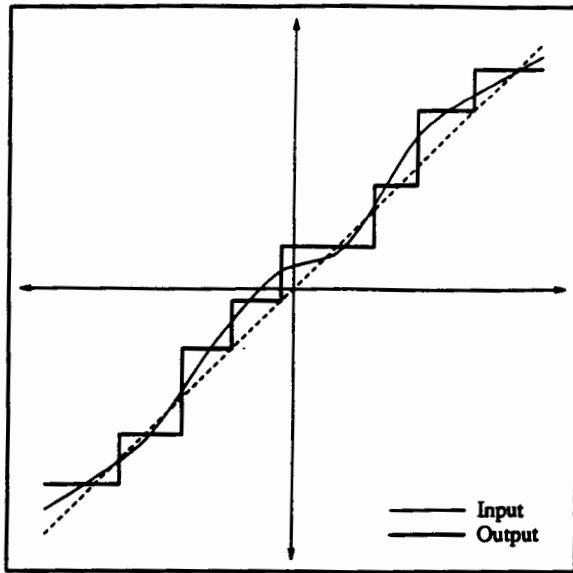


Figure 4. 3 Bit ADC Transfer Function Including Errors.

### I.3 REDUCING QUANTIZATION NOISE

There is no way to avoid the quantization error, even in an ideal ADC. An obvious method for reducing the quantization error in an ADC is to increase the resolution of the quantizer. Unfortunately, quantizer circuits become increasingly difficult to implement as the resolution becomes high. One of the difficulties is the design of the anti-aliasing analog filter. For high resolution quantizers, the passband of the anti-aliasing needs to have a very flat response and the stop band must have very strong attenuation after  $\frac{f_s}{2}$ . If  $f_n = f_s$ , then the transition between passband and stopband should be as narrow as possible in order to avoid wasting bandwidth. It is a very challenging task to design a good anti-aliasing analog filter [3]. The sample and hold circuit must be able to accurately sample and hold the analog voltage level. The major errors caused by the sample and hold circuit are

timing errors and amplitude errors. The sample and hold circuit must be able to quickly settle to the correct amplitude during the sample mode, and then be able to maintain the correct amplitude during the hold mode. A common method for performing the analog to digital conversion, or quantization, is by a method called successive approximation. This quantization method works by testing each bit of the  $n$  bit word and converting the word back to an analog value and comparing it with the input value to determine whether the bit belongs in the output. As the resolution increases, more tests are required, which reduces the time allowed for the test to occur.

Another method of quantization which is simpler than the successive approximation approach is called flash conversion. The flash converter directly compares the input voltage level against the voltage levels of a resistive ladder. Digital logic then converts the results of the comparisons into a digital value. The drawback to this method is that requires  $2^n - 1$  comparators. Also, the resistors in the resistive ladder must all be very accurate so that the quantizer levels are uniformly spaced. So, for high resolution quantizers, the flash converter is not suitable.

#### I.4 OVERSAMPLED ADC'S

Since there are practical limitations to the resolution of a quantizer using traditional methods, other methods of increasing the resolution have been devised. One successful method is to oversample the input signal. Oversampling means that the signal is sampled more than necessary. Since it is necessary to sample at

least at  $f_n$ , oversampling means that  $f_s > f_n$ . The oversampling ratio is defined as

$$M = \frac{f_s}{f_n}.$$

The basic idea of oversampling is that instead of increasing the quantizer resolution to reduce noise, the sampling frequency is increased. The goal is to take many low resolution digital values and construct, through digital filtering and decimation, a high resolution value at the desired rate. For example, taking the average of every  $M$  low resolution values to produce the final output should result in a much better approximation of the input at  $\frac{1}{M}$ <sup>th</sup> the frequency. The digital filtering and decimation circuits will take an input with a small digital word length and high frequency and produce an output with a larger word length and lower frequency. Thus, oversampling can make the effective resolution of a quantizer higher than its nominal resolution.

The advantage of oversampling is that the analog circuitry can be made with cruder parts tolerances, and most of the work to produce high resolution output can be performed by digital signal processing circuits. Since  $f_s$  is much higher than  $f_n$ , the amount of bandwidth available for the transition band of the anti-aliasing analog input filter is larger. This relaxes the need for a high-order low pass filter which can be difficult to implement [3] and allows the use of a relatively crude analog filter which is easier to implement. The high performance filtering tasks are handed over to the digital circuitry after quantization. The technological advantage is that the digital circuitry can be implemented more reliably in comparison to implementing very precise analog filters. Another implementation advantage is

that a low resolution quantizer can be used, such as a fast low resolution flash converter.

The increase in effective resolution of a low resolution quantizer by oversampling is achieved by filtering out all the frequencies above the signal bandwidth. If it is assumed that the amount of noise energy is constant for a particular quantizer resolution, independent of the sampling rate, and if the noise energy is fairly evenly distributed across that entire sampling bandwidth, then oversampling and filtering out the noise energy above the signal bandwidth will remove more noise from the output than if no oversampling is done. Thus, the effective resolution of the quantizer is increased above its nominal, or nyquist rate, resolution.

Further improvements in the performance of the oversampled quantizer can be achieved by adding a feedback loop around the quantizer. A popular configuration known as the sigma delta modulator is one such circuit. It is common for the sigma delta modulator to use a one bit quantizer. The purpose of the feedback loop is to attempt to cause more of the noise energy to fall in the frequencies above the signal bandwidth where it can be filtered out. The feedback loop acts in a way such that low frequency noise is reduced and high frequency noise is increased. The effect of the feedback loop on the noise spectrum is known as noise shaping. Noise shaping quantizer systems will realize even greater increases in effective resolution due to oversampling than just oversampling a traditional quantizer. Both of these cases will be presented in this study.

## I.5 SCOPE OF STUDY

The primary purpose of this study is to examine a couple types of oversampled ADC's and gain an understanding of how the quantization noise affects their performance. A crucial factor in determining the effectiveness of an oversampling system is understanding the characteristics of the quantization noise. Because quantization is a highly nonlinear process, various assumptions and simplifications are often used when doing analysis. The most common assumption is that the quantization noise is white across the sampling bandwidth. Given a deterministic input, such as a sinusoid, this is clearly not true since the error will also be deterministic and closely related to the input. However, this assumption has been shown to be true, or at least a good approximation, given that certain conditions hold. One of these conditions is that the resolution of the quantizer is high.

When analyzing oversampling systems, the distribution of noise energy in the spectrum is important. If the noise is white and uniform, than noise shaping configurations will be successful in shaping the spectrum so that most of the noise falls in the high frequencies. An important point to note is that oversampled ADC systems often use low resolution quantizers, going as low as one bit resolution. In these cases, the assumption that the quantization noise is white will clearly be untrue. The question that remains is how good an approximation the white noise assumption provides. It is important to study the characteristics of the quantization noise, since actual noise characteristics could cause performance problems

which would not be expected if simplistic assumptions about the noise distribution are made. For example, even if actual noise power is equivalent to the noise power predicted by the white noise assumption, the actual performance can be unsatisfactory if all of the actual noise power is located in a few harmonics instead of being spread evenly over the signal bandwidth. One of the most important characteristics of the noise is the power spectrum of the noise. The power spectrum shows how the noise energy is distributed in the sampling bandwidth, and knowing what the spectrum is will provide indications of how well oversampling and noise shaping work.

This study will take a detailed look at the characteristics of the quantization noise of two oversampling ADC systems which utilize low resolution quantizers. First a low resolution traditional quantizer, and secondly, a first order, or one feedback loop, sigma delta modulator with a one bit quantizer. The format of the study will be to present theoretical predictions of the ADC performance and then to verify the predictions by examining simulated results. Theories based on the assumption of white quantization noise will be presented first. Improved theories which deal more directly with the quantization process and predict exact results for some of the noise characteristics will also be presented and verified. Simulations will then be performed and compared to the theoretical results. Since the more complicated exact theories do not always readily provide intuitive understanding of the expected results, the results of the simulations are also used to provide insights into the theoretical results.

In many papers that discuss this topic, simulated results are given for a couple examples and compared to whatever theory is being presented. Since the behavior of the quantization noise is complex and dependent on the input frequency and the oversampling ratio, the approach taken in this study is to characterize the behavior of the quantization noise over the entire signal bandwidth. This requires a lot of simulations, but the simulations coupled with the various theories help provide more insight into the behavior and performance of oversampled ADC systems.



## CHAPTER II

### UNIFORM ADC THEORY AND SIMULATIONS

The uniform ADC will be studied first. Since the uniform ADC is about as simple as an ADC can get, understanding its operation and noise characteristics will provide a basis for the study of sigma delta modulators. The low resolution uniform ADC is the focus here because sigma delta modulators typically use low resolution quantizers. Although the uniform ADC is a conceptually simple device, it can be difficult to analyze since the quantizing process is nonlinear. Since nonlinear systems are generally more difficult to solve, various methods are often employed to linearize the uniform ADC to simplify analysis. The most common assumptions and approximations will be developed here. It will also be demonstrated that the linearized uniform ADC analysis is limited in its usefulness. A more complex analysis will then be developed to provide a better explanation of the observed performance.

#### II.1 THE WHITE NOISE ASSUMPTION

The most common method used in analysis of the uniform ADC is to model the quantizer as an additive noise source, as shown in figure 5. The following equation can be written to describe this system.

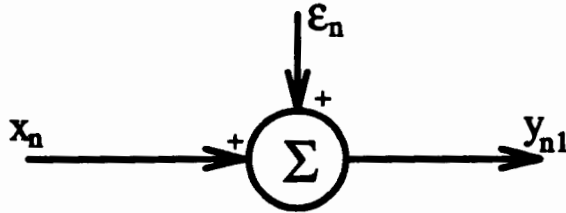


Figure 5. Model of quantizer error in the uniform ADC.

$$y_n = x_n + \epsilon_n \quad (\text{II.1})$$

The error signal  $\epsilon_n$  is the result of a several different processes. As it was mentioned in chapter I, it will be assumed that the quantizer is ideal and that the input signal stays within the valid range, so the only error process that will be considered is the quantization error. Thus,  $\epsilon_n$  represents the quantization error. The critical factor affecting the ease of analysis is the nature of  $\epsilon_n$ . Although it is not apparent in figure 5,  $\epsilon_n$  is dependent on the value of  $x_n$ . This makes it difficult to determine the affect of the quantization error on system performance without knowing the input signal in advance. If it can be assumed that  $\epsilon_n$  is independent of  $x_n$ , then the performance of a particular quantizer can be analyzed without worrying that a different input signal will result in a completely different performance.

It has been shown that  $\epsilon_n$  can be assumed to be independent of  $x_n$  when  $x_n$  meets certain requirements [6]. These requirements are:

- The quantizer does not overload.

- The quantizer has a high resolution.
- $\Delta$  is small.
- The probability distribution of pairs of input samples is given by a smooth probability density function.

The independence of  $\epsilon_n$  from  $x_n$  allows for a linear analysis of the quantizer. To ease the analysis further, it is often assumed that  $\epsilon_n$  is a uniformly distributed white noise source. In other words, the density of  $\epsilon_n$  over the sampling bandwidth is uniform. With these assumptions it is possible to calculate the total quantization noise power for an  $n$  bit quantizer.

The following characteristics are known for a white noise source. The sample average mean is

$$M\{\epsilon_n\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \epsilon_n = 0 \quad (\text{II.2})$$

and the sample average power is

$$M\{\epsilon^2\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \epsilon_n^2 = \frac{\Delta^2}{12} \quad (\text{II.3})$$

The sample autocorrelation is defined as

$$r_c(k) = M\{\epsilon_n \epsilon_{n+k}\}$$

and  $r_c(0) = \frac{\Delta^2}{12}$  and  $r_c(k) = 0$  for  $k \neq 0$ . These characteristics of the white noise quantization error are a basic result because they represent the most common assumptions made when analysis of ADC's is performed. The moments and autocorrelation of the white noise spectrum presented here are important because they

will be used throughout the rest of this study as a basis for comparison with more complex theoretical and simulated noise characteristics.

The power of  $\epsilon_n$ , as defined in equation II.3, depends only on the resolution of the quantizer. A value that can be used to measure the performance of a uniform ADC is the signal to quantization noise ratio, or SQNR. The SQNR will be used when there are sinusoidal inputs to the ADC. It will be assumed that the sinusoidal inputs have a full scale amplitude, where full scale amplitude is  $\frac{1}{2}$ . The SQNR will be measured in decibels.

Using equation II.3, an expression for the SQNR can be derived.

$$SQNR = 10 \log \left( \frac{\frac{1}{8}}{M \{\epsilon^2\}} \right) \quad (II.4)$$

which reduces to

$$SQNR = 6.02N + 1.76db \quad (II.5)$$

Equation II.5 shows that the SQNR increases by about  $6db$  with each additional bit of resolution in the quantizer.

## II.2 OVERSAMPLED UNIFORM ADC'S

Now that the uniform ADC has been analyzed using the white noise assumption, it is time to include oversampling in the system and see what effect it will have. Intuitively, oversampling should reduce the amount of quantization noise that will be seen in the output. Equation II.3 implies that the total quantization noise power for the uniform ADC is constant no matter what  $f_s$  is. Thus, since

the noise is assumed to be uniformly distributed, as  $M$  is increased, the amount of quantization noise inside the signal bandwidth decreases. So, by digitally filtering out the noise above the signal bandwidth, the error due to quantization noise is reduced and the effective resolution of the quantizer is increased. Figure 6 illustrates how this looks in the frequency domain.

All that remains to do is to develop the equations with oversampling incorporated into them. It will be assumed that an ideal lowpass filter with a cutoff at  $\frac{f_n}{2}$  is used. Along with the assumptions listed above, it is also assumed that the quantization noise for the oversampled uniform ADC is uniform over the range  $\pm \frac{f_s}{s}$ . The SQNR for the oversampled uniform ADC can be computed based on equation II.4. Since the oversampling ratio is  $M$ , only  $\frac{1}{M}$ <sup>th</sup> of the uniform noise spectrum will remain in the signal bandwidth.

$$SQNR = 10 \log \left( \frac{\frac{1}{8}}{\frac{1}{M} M \{e^2\}} \right) \quad (\text{II.6})$$

which reduces to

$$SQNR = 6.02n + 10 \log(M) - 1.25 \text{db} \quad (\text{II.7})$$

Equation II.7 shows that the  $SQNR$  increases by about  $3 \text{db}$  every time  $f_s$ , or  $M$ , is doubled. This demonstrates how oversampling can be used to reduce the quantization noise. The reduction for the uniform ADC is relatively modest. The sigma delta modulator, which will be examined later, shapes the noise spectrum so that most of noise is located out of the signal bandwidth and can be filtered out.

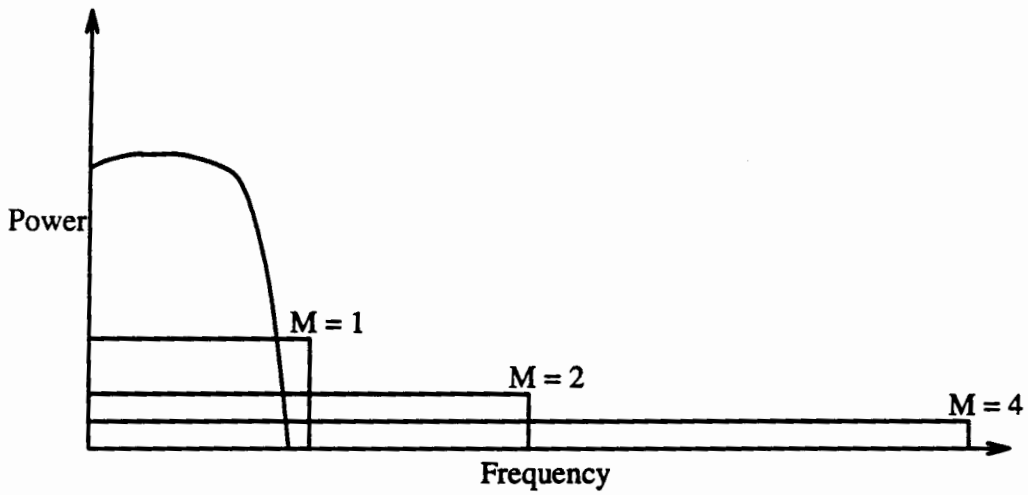


Figure 6. Uniform noise spectrum for different  $f_s$ .

## CHAPTER III

### SIMULATION TECHNIQUES

Now that a theory has been developed for the noise characteristics of the uniform ADC, it is time to see how well the theory stands up against simulations. Before the simulations of the uniform ADC are performed, the methods used to simulate the ADC will be presented in detail. Most of the techniques introduced here will apply to the simulation of the sigma delta modulator as well.

#### III.1 SIMULATION METHODS AND TECHNIQUES

The goal of simulating the operation of an ADC is to obtain an understanding of how the actual ADC will perform. The results of the simulation can be processed and compiled in various ways to provide measurements of the performance of the simulated ADC. As mentioned before, the goal of this study is to understand the quantization noise characteristics of oversampled ADC's with low resolution quantizers. Understanding the quantization noise will provide an idea of how much gain in effective resolution can be expected from oversampling.

One of the most important characteristics for predicting ADC performance is the frequency spectrum of the quantization noise. The frequency spectrum provides the information needed to compute the total noise power for a given

input signal. The two types of input signals that will be considered are dc inputs and sinusoidal inputs. The dc input is important because it represents the system in a quiet state. The sinusoid input is important because it is typically used as a standard signal for system analysis. It will be of interest to discover how the noise spectrum behaves as these inputs vary across the valid ranges. For dc inputs, the range will be from 0 to  $\pm 0.5$ . For sinusoidal inputs, the frequency of the input signal can range from 0 hz to  $\frac{f_n}{2}$  hz, which is the signal bandwidth.

In order to analyze the performance of ADC's, it is necessary to simulate their operation to obtain sample output sequences. Using the output sequences, it is possible to compute the frequency spectrum and frequency power spectrum of the output. This enables the computation of the total noise power and total signal power. These quantities can be used to calculate the signal to quantization noise ratio, or SQNR. The SQNR is defined as the ratio of the total signal power and the total quantization noise power. The SQNR will be used when the input signal is a sinusoid since the frequency spectrum of a sinusoid is a spike and is easily found. When dc inputs are used, the total quantization noise power will be examined.

### III.2 THE SIMULATION PROCESS

The simulation process will first be described in general terms. Then the issues involved in performing each step of the simulation will be described in detail. The general steps in the simulation process are as follows.



- Generate an input signal to represent the analog input. This can either be a dc or sinusoidal input.
- Sample the input signal at  $f_s$  and generate a quantized output signal.
- Lowpass filter the output signal, when oversampling, to remove high frequency noise.
- Analyze the filtered output signal to find quantities of interest, such as the total quantization noise and the SQNR.

### III.3 SIMULATION OF INPUT SIGNAL

The front end of the simulations, which involves generating an input signal and producing an output signal, were performed using the MIDAS software system developed at Stanford University [9]. MIDAS is a functional simulator for mixed digital and analog sampled-data systems. MIDAS provides the ability to specify and configure the functional modules of a system. The modules of particular interest here are modules to generate test signals and modules which perform a quantizing function. Other useful functions which can be used are adders and delays. Other system parameters like the input frequency, sampling frequency, and nyquist frequency are specified. After a particular system is set up, MIDAS is used to simulate the system and produce the output of the quantizer. The full capabilities of MIDAS were not utilized in the simulation process. MIDAS was used only as a convenient method to set up various ADC configurations. Actually,

MIDAS was used primarily for the uniform ADC. By the time simulations for the sigma delta modulator were performed, greater performance was desired and C programs were written to generate the inputs and compute the outputs.

### III.4 COMPUTING SQNR AND TOTAL NOISE POWER

When an ADC which uses oversampling is simulated, it is necessary to filter out all frequencies above  $\frac{f_n}{2}$  in order to obtain the benefits of oversampling. The frequency spectrum of the filtered output sequence will be the source of most of the calculations which will measure ADC performance. By calculating the Discrete Fourier Transform (DFT) of the output sequence, the frequency spectrum can be found. The average power spectrum of the output sequence can be computed by finding the magnitude squared of each frequency component. With the power spectrum, it is possible to compute the total quantization noise power and the total signal power. All that needs to be done is to identify the sections of the spectrum which belong to each part. With sinusoidal inputs, it is expected that there will be a spike of power in the spectrum at the frequency which corresponds to the input frequency. It is assumed that the power spike corresponding to the input frequency will be the largest spike in the spectrum. To calculate the total quantization noise power, the spectrum is integrated, excluding the part of the spectrum which contains the spike of the input signal. The opposite is done to compute the signal power. These values, which have been found directly from the simulated results, can then be used to compute the simulated SQNR. For dc inputs,

the input signal is obviously the dc part of the spectrum, so the quantization noise power can be computed by integrating the spectrum, not including the dc part.

### III.5 COMPUTING THE POWER SPECTRUM

The previous section described how the power spectrum is used to compute various performance measurements. Now it is time to look at how the power spectrum itself is computed. As it was stated in the previous section, the DFT can be used to find the frequency spectrum of the output signal. In practice, it is much more efficient and speedy to compute the DFT using the Fast Fourier Transform (FFT) algorithm. The only practical limitation this imposes is that the number of elements in the sequence should be a power of two. This is no problem, since any number of output samples can be generated. If a sequence of  $2^n$  real values in the time domain are transformed by the FFT, then the result will be a sequence of  $2^{n-1}$  complex values in the frequency domain which represent the sinusoidal frequency components of the input sequence. The  $n$  frequency components are divided evenly in the range of 0 to  $\frac{f_s}{2}$ . Actually, they range from  $-\frac{f_s}{2}$  to  $\frac{f_s}{2}$ , but the negative components are normally combined with the positive components. To find the power of a frequency component, the square of the magnitude of the complex frequency component is computed. From here on, the frequencies represented by the FFT will be referred to as bins.

There are some more complexities involved with using the FFT. As men-

tioned above the FFT produces bins at frequencies which are multiples of  $\frac{f_s}{n}$ . The number of bins in the range  $\frac{f_s}{2}$  can be thought of as the resolution of the FFT. Frequency components of the input signal which fall exactly on a bin frequency will have all of their spectral energy located at that bin. However, if the signal contains frequencies which are not equal to one of the bin frequencies, then the spectral energy of those frequency components will leak out to the surrounding bins. This is known as spectral leakage. Spectral leakage results because a finite number of samples are used and the DFT assumes that the set of samples is periodic. If, for example, a non-integer number of periods of a sinusoid is represented by the input sample, then there is a discontinuity in the sample. This discontinuity results in the leakage of the input power throughout the signal bandwidth. A method used to control spectral leakage is known as windowing. The  $n$  samples which are fed into the FFT can be thought of as an infinite sequence of samples which are multiplied by a rectangular window which has a value of 1 for  $n$  points and a value of zero everywhere else. This rectangular window preserves the discontinuity in the input sample. Various windows have been devised which rise up gradually from zero. This has the tendency to reduce the discontinuities at the ends of the input sample. Clearly, windowing affects the input data and it will have associated errors. The error caused by windowing is called spectral smearing. The error is called smearing because a frequency which falls exactly on a bin value and would normally be a spike in the spectrum will be smeared. Instead of a spike, the signal

power will be spread across the surrounding bins. The proper choice of windows will limit the smearing to a set of localized bins and reduce the leakage to distant bins.

All of the FFT calculations performed for this study were windowed by the Blackman-Harris window function [8, 4]. The particular form of Blackman-Harris window used is designed to reduce spectral leakage by 92 dB within 5 bins from a spike. In other words, a frequency which would have a spike of a single bin with rectangular windowing will now be a nine bin wide peak. The original bin with leakage to  $\pm 4$  bins around it. However, by the fifth bin, the leakage is virtually non-existent. So, although the resolution is degraded for frequencies which fall exactly on bins, the situation is improved and predictable for all other frequencies which fall somewhere between bins.

### III.6 FILTERING THE OUTPUT SIGNAL

Developing a filtering technique for the simulation was one of the most difficult steps in setting up the simulation process. As stated above, the goal is to lowpass filter the ADC output signal, when oversampling is being used, to remove high frequency noise, and hopefully, increase the effective resolution of the ADC. The filtered output sequence is then available for processing, such as computing an FFT. In a real oversampled ADC, the output signal would be digitally filtered and decimated back down to  $f_n$ . For this study, the decimation step was ignored since the primary interest is the quantization noise characteristics.

Initial attempts at filtering were made by using digital filters. One such attempt used an FIR (Finite Impulse Response) filter. This proved to be troublesome because the order of the filter must increase as the sampling rate increases in order to maintain the same width transition band between the pass band of the filter and the stop band of the filter. This is because as the sampling rate increases, the size of the transition band, which remains constant, decreases relative to the sampling bandwidth. It was found that the required filter order increased beyond practicality for higher oversampling ratios of interest (around 64 to 256).

Another attempt was made using FIR filters with fixed order as the oversampling ratio was increased. In this case the size of the transition band relative to the sampling bandwidth remains constant. However, it also results in a doubling of the size of the transition band between the passband and the stopband as the sampling frequency is doubled. No matter how narrow the transition band is made for low sampling frequencies, as the sampling frequency increases, the transition band will soon dominate the signal bandwidth. That is, partially filtered noise power from the transition band will overwhelm the power contained in the signal bandwidth, which is decreasing relative to the sampling bandwidth as the oversampling ratio is increased.

In real oversampled ADC systems, the filtering will most likely be implemented as a multi-stage digital filter which combines the decimation process [3]. A brief attempt was made at using a multi-stage decimation filter. In retrospect, it appeared to have worked correctly, but since the investigation was beginning to get

bogged down in filter design instead of the specified goals, a simple and effective filtering scheme was developed.

The filtering technique finally arrived at does not attempt to simulate a digital filter that might be used in an ADC system. Instead, the goal is to get the same results without the adding to the simulation the complications of a multi-stage digital filter. Instead of filtering the output signal and then computing the FFT to get the spectrum, the filtering stage is ignored and the FFT is computed for the raw output signal itself. Filtering is achieved by simply truncating the spectrum at the desired cutoff frequency, thereby achieving an ideal low pass filter. Actually, this method is not truly ideal since the FFT spectrum is distorted a bit by the spectral leakage and smearing problems described above. However, the truncated spectrum should be just as good as the spectrum produced from the FFT of a digitally filtered signal.

There is one complication involved with using this method when studying oversampled systems. If there are  $n$  points in the output sample, then the FFT will produce  $\frac{n}{2}$  bins. If  $n$  remains constant as  $f_s$  is increased, then the number of bins representing the signal bandwidth will decrease. That is, the bin resolution of the signal bandwidth decreases. This is a problem because of the spectral leakage problem of the FFT. Since all the FFT's computed for this study use the Blackman-Harris window discussed before, it is known that the energy of a sinusoid will spread out over nine bins. Now consider an example where  $n$  is 1024 and  $M$  is 128. For this case, the part of the FFT spectrum which corresponds to the signal

bandwidth will be the first  $\frac{512}{128} = 4$  bins. Clearly, the signal power which should be contained in the signal bandwidth will not even fit inside the signal bandwidth, which will obviously cause computation errors.

In order to prevent this problem, the following procedure was developed. For  $M = 1$ , decide how many bins should represent the signal bandwidth in the spectrum. If this number is 256, for example, then  $n = 512$  samples need to be produced by the simulation of the ADC and fed into the FFT. As  $M$  increases, it is desirable that the number of bins representing the signal bandwidth remains constant. This can be accomplished doubling  $n$  every time  $M$  doubles. Thus,  $n = M \times 512$  samples need to be generated for a particular value of  $M$ . Using this technique keeps the bin resolution of the signal bandwidth constant, although it does require using large values of  $n$  to compute the FFT when  $M$  gets large. This method is obviously not a model for the operation of a real ADC, but the end result should be the same.



## CHAPTER IV

### SIMULATIONS OF THE UNIFORM ADC

#### IV.1 SIMULATION RESULTS

Now that a theory has been developed and the simulation technique has been discussed, it is time to simulate the uniform ADC. Various parameters of the ADC need to be defined. A four bit quantizer will be used and  $f_n = 48KHz$ . The input signal will be sinusoids with frequencies in the range from  $(0, \frac{f_n}{2}]$ . After running the simulations for a range of input frequencies and computing the SQNR, the results are plotted, SQNR versus  $M$ . Figure 7 shows results for several input frequencies spanning the range of the signal bandwidth. The theory predicts that the SQNR curve will be linear with a slope of 3 db. However, it is observed that the curve plateaus at a level lower than the predicted level for low frequencies. The level of the plateau rises with frequency until it suddenly disappears and then the SQNR curve rises much higher than the predicted linear curve for higher values of  $M$ .

Clearly, the simulated results do not match the white noise theory closely. At low frequencies, the SQNR is too low and at high frequencies it is too high. The SQNR dependency on the input frequency shows clearly that the assumptions made about the independence of the quantization noise with the input are not

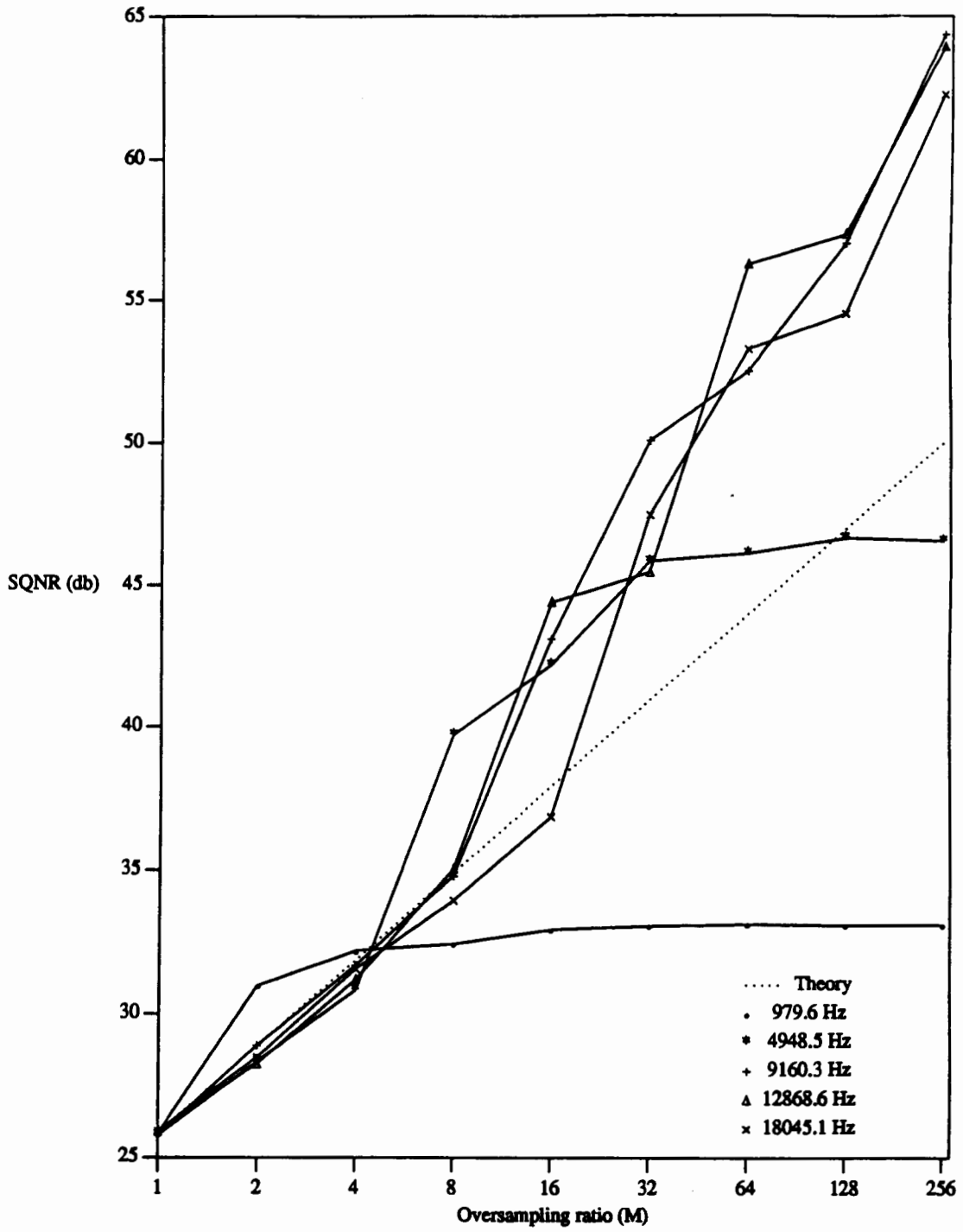


Figure 7. SQNR curves for 4 bit Uniform ADC.

valid for this example. Depending on the type of application, this type of behavior could be very undesirable. In this case the primary factor affecting the SQNR behavior is the resolution of the quantizer. Four bits is a fairly low resolution. If the simulations are run again using a ten bit quantizer it can be seen that the SQNR matches the predicted performance much more closely over the range of possible inputs. Figure 8 shows results for several frequencies spanning the signal bandwidth using a 10 bit quantizer. However, as stated above, this study will focus on low resolution quantizers and the affect that oversampling has on them.

#### IV.2 AN IMPROVED UNIFORM ADC THEORY

The most interesting feature of the simulations is the plateau in the curves at low frequencies. By examining the simulations it is observed that the plateau effect abruptly disappears around the input frequency of  $8KHz$ . To explain this behavior it is observed that as the sampling frequency becomes much higher than the input frequency, the output signal begins to take on a staircase-like shape. As  $f_s$  is increased, the staircase shape becomes more defined and sharp. Thus, once the value of  $f_s$  is high relative to the input frequency, the output signal becomes essentially fixed and periodic in nature. So, it could be predicted that the SQNR will stop increasing since if the signal does not change, then the spectrum will not change. Figure 9 illustrates this behavior for a low input frequency and several values of  $M$ . Note that as  $M$  increases, the noise floor drops, but the harmonic spikes remain virtually constant in height. The power in the noise floor eventually

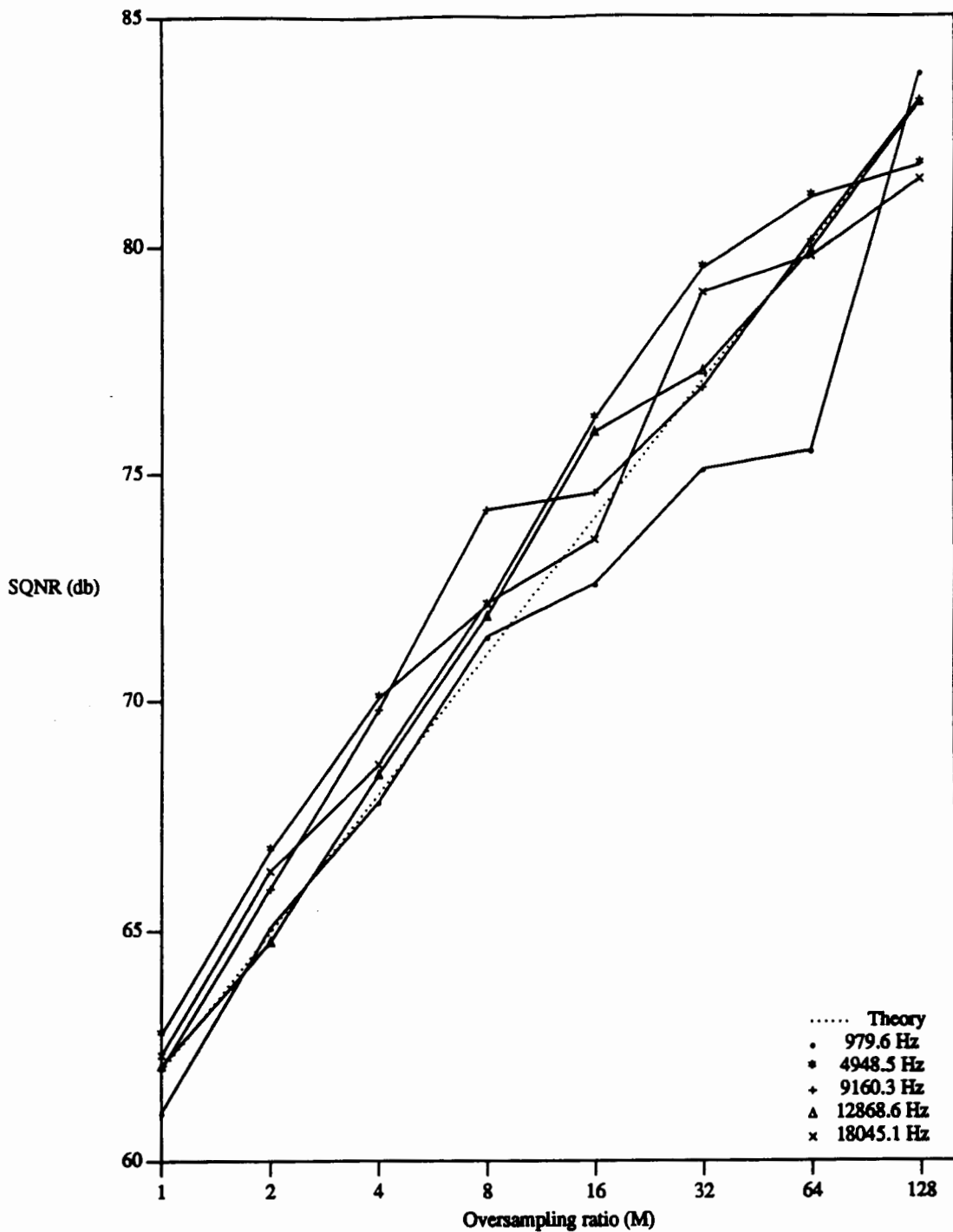


Figure 8. SQNR curves for 10 bit Uniform ADC.

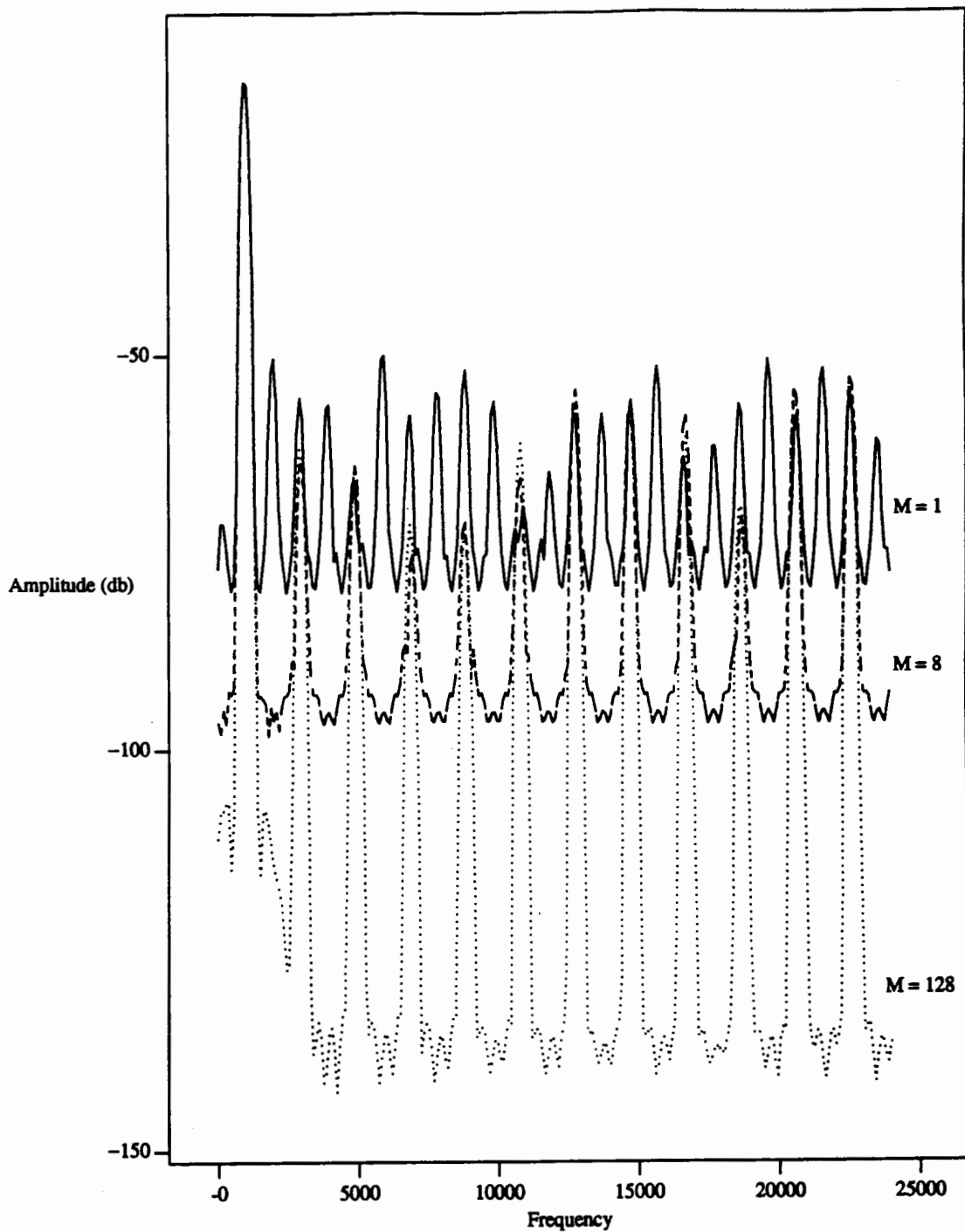


Figure 9. Spectrum of 4 bit ADC at various  $M$  for  $f_x = 979Hz$ .

becomes insignificant compared to the unchanging noise power contained in the harmonics and the SQNR flattens out for high values of  $M$ . [See A.2] By analyzing the structure of the output signal, an improved theory for the behavior of the low resolution quantizer uniform ADC can be developed.

If the input sinusoid is sampled and quantized at a frequency approaching infinity, then the output becomes a periodic staircase wave. This staircase can be decomposed into a set of rectangular pulse waves as shown in figure 10. The reason for doing this is because the frequency spectrum for a rectangular pulse wave is easy to find using Fourier Series analysis. In order to compute the frequency spectrum of the staircase wave, all that needs to be done is to sum the spectrums of the rectangular waves that compose the staircase wave. Computing this spectrum will determine that limiting spectrum that the simulation results are approaching as  $M$  increases. This spectrum should provide a good estimate of ADC performance for high values of  $M$ .

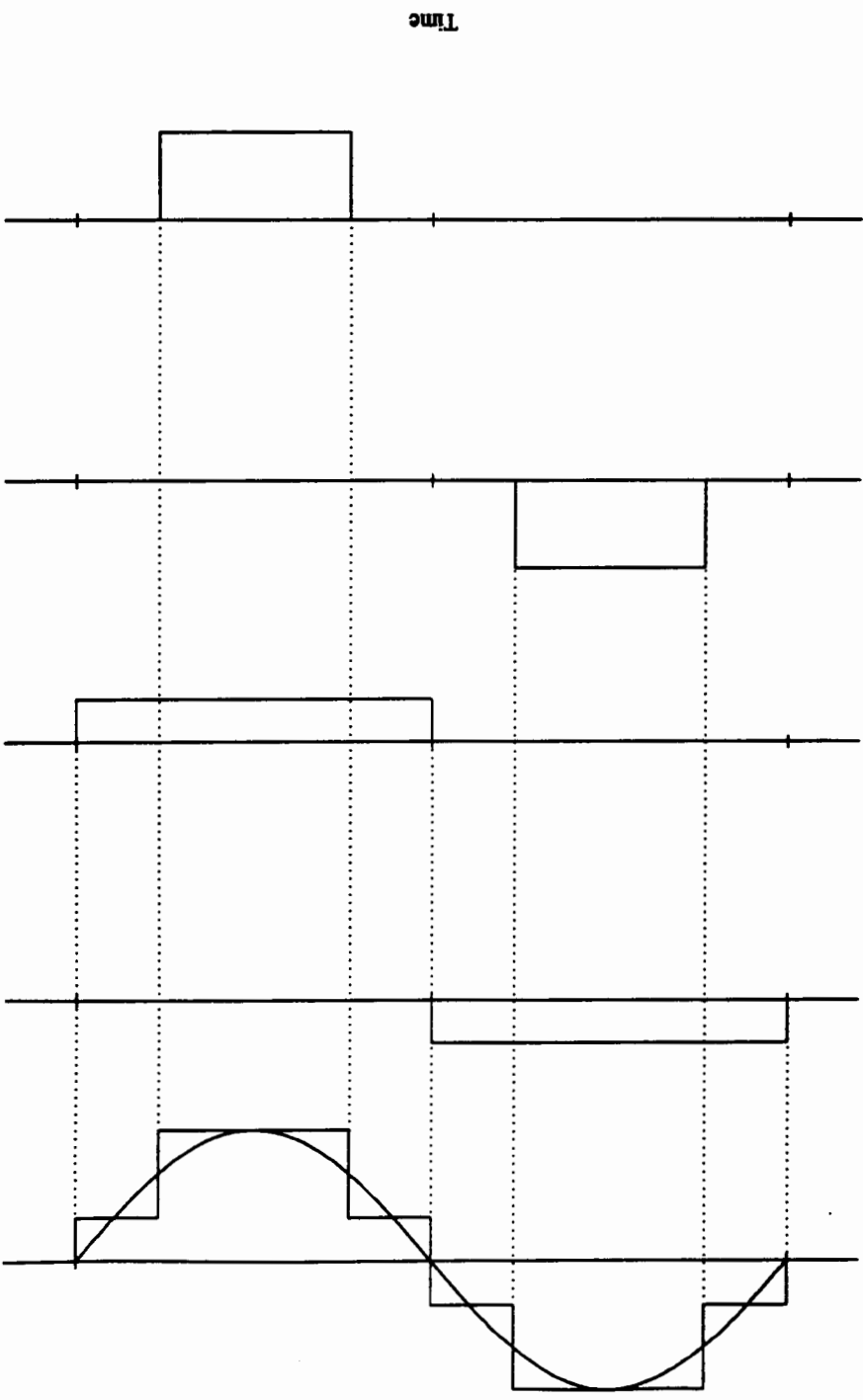
The Fourier Series for a periodic function is defined as [8]

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk\omega_0 t} \quad (\text{IV.1})$$

$$c_k = \frac{1}{P} \int_{-\frac{P}{2}}^{\frac{P}{2}} f(t) e^{-jk\omega_0 t} dt \quad (\text{IV.2})$$

$P$  is the period of the function. Let  $P = 2\pi$  for this analysis. Also, define  $X_1$  to be the start of the pulse,  $X_2$  to be the end of the pulse, and  $h$  to be the height of the pulse.  $X_1$  and  $X_2$  are constrained to be within the range  $(0, 2\pi]$ . A bit of math

Figure 10. Decomposition of highly oversampled sinusoid.



will show that the values of  $c_k$  for this rectangular pulse are

$$c_k = \frac{h}{2\pi k} [(\sin(2\pi k X_2) - \sin(2\pi k X_1)) + j(\cos(2\pi k X_2) - \cos(2\pi k X_1))] \quad (\text{IV.3})$$

For convenience a sine wave with zero phase and period  $2\pi$  and amplitude 1 will be used. The height will be  $\frac{\Delta}{2}$  for the base square pulse, and  $\Delta$  for all the smaller width rectangular pulses. For each positive height pulse, there is a negative pulse with a phase shift of  $\pi$ . For the positive pulses

$$X_{1_p} = \sin^{-1}(\alpha) \quad (\text{IV.4})$$

$$X_{2_p} = \pi - X_{1_p} \quad (\text{IV.5})$$

And for negative pulses

$$X_{1_n} = X_{1_p} + \pi \quad (\text{IV.6})$$

$$X_{2_n} = -X_{1_p} \quad (\text{IV.7})$$

Where  $\sin^{-1}(\alpha)$  are the points at which pulses begin.

$$\alpha = 0, \Delta, 2\Delta, 3\Delta, \dots, (2^{N-1} - 1)\Delta \quad (\text{IV.8})$$

By using the values from equations IV.4 through IV.7 in equation IV.3 several simplifications can be made. It turns out that the real parts of the pulses for odd values of  $k$  and the imaginary parts for even values of  $k$  are always zero. Furthermore, for even values of  $k$ , the real parts of a pair of positive and negative pulses cancel each other out. The imaginary parts for odd values of  $k$  are equal for



a pair of positive and negative pulses. So, equation IV.3 can be reduced to

$$c_k = -j4 \frac{h}{2\pi k} \cos(kX_{1p}) \quad (\text{IV.9})$$

for odd values of  $k$  and positive values of  $h$  only. All of the other terms are zero.

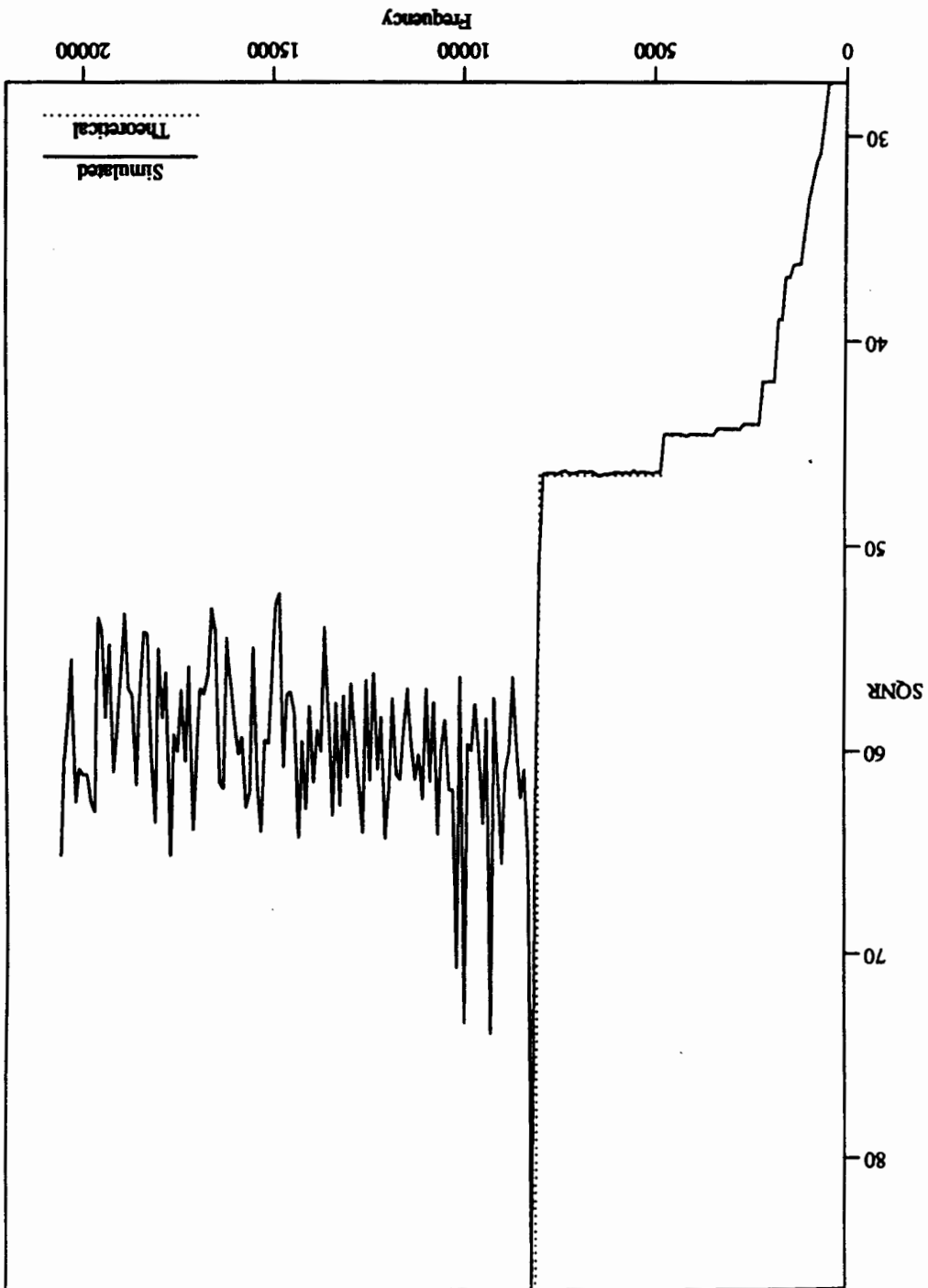
If the results of equation IV.9 are added together for all the pulses that make up a quantized sine wave, then the power spectrum of the quantized sine wave can be computed by taking the magnitude squared of the sequence of  $c_k$ 's. By limiting  $k$  to be less than or equal to  $\frac{f_n}{2f_x}$ , the computed spectrum can be effectively low pass filtered with cutoff at  $\frac{f_n}{2}$ .

The resulting spectrum consists of the fundamental frequency and zero or more of the odd numbered harmonics depending on the value  $f_x$ . This spectrum can be used to predict what the upper limit of the SQNR should be for any input frequency. For low values of  $f_x$  there will be many harmonics included in the spectrum. As  $f_x$  increases, the harmonics will be filtered out one by one. Since the power of each harmonic does not change as  $f_x$  changes, the SQNR will stay constant until another harmonic is filtered out. Then the SQNR will jump to a higher level and stay constant until the next harmonic is filtered out. A close look at the simulation results will confirm that this occurs. All of the harmonics will be filtered out when  $f_x > \frac{f_n}{6}$ . At this point the SQNR would theoretically become infinite. In the simulations, the value of  $f_x$  when this occurs is  $8KHz$ , which is exactly the point where a sudden switch is observed from the SQNR

curve flattening to the SQNR curve rising sharply. The SQNR jumps much higher for  $f_x$  greater than  $8KHz$ , but it does not go to infinity since  $f_s$  is obviously not infinite. Figure 11 shows the SQNR over the frequency range for the simulated results for  $M = 128$  and for the theoretical prediction. Up to the point of  $8KHz$ , the simulated and theoretical results match very closely.

The Fourier Series analysis of the ideal quantized sine wave matches very closely with the simulation results where the sampling frequency is high. The lesson learned here is that the assumption that the error signal is random and independent from the input breaks down as the sampling frequency increases. Instead, the error signal becomes very deterministic and periodic in nature. This effect is very noticeable for the simulation results which used a four bit quantizer. Increasing the number of bits in the quantizer will reduce these effects and make the original white noise assumptions more valid, as was seen for the case of a ten bit quantizer.

Figure 11. SQNR at  $M = 128$  for a 4 bit ADC.



## CHAPTER V

### THE SIGMA DELTA MODULATOR

As shown in chapter IV, an oversampled uniform ADC with a low resolution quantizer has quantization noise characteristics that are highly dependent on the frequency of the input sinusoid. Other ADC circuit configurations have been developed which have better performance. The rest of this study will focus on a particular ADC known as the sigma delta modulator, or  $\Sigma\Delta M$ . The  $\Sigma\Delta M$  is a popular ADC circuit which uses oversampling and a feedback loop to improve the effective resolution of a one bit quantizer. The general configuration of a first order, or one loop,  $\Sigma\Delta M$  is shown in figure 12. Keeping in mind that the goal of the study is to understand the quantization noise characteristics, it is important to clarify exactly what the quantization noise is for the  $\Sigma\Delta M$ . There are two important quantization noise signals to be considered for the  $\Sigma\Delta M$ . The first one, which will be called the quantization noise, is the difference between the signals  $y_n$  and  $u_n$ . This is the error of the quantizer by itself. Also of interest is the quantization noise of the whole  $\Sigma\Delta M$ . This second quantizer noise signal will be referred to as the total quantization noise. The total quantization noise is defined as the difference between  $y_n$  and  $x_n$ . The total quantization noise is the error between the output and the input to the  $\Sigma\Delta M$ .

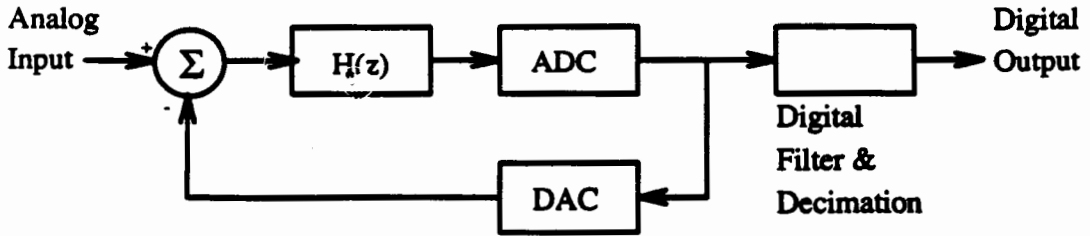


Figure 12.  $\Sigma\Delta M$  circuit diagram.

### V.1 BASIC $\Sigma\Delta M$ OPERATION

While the uniform ADC was simple to understand conceptually, the operation of the  $\Sigma\Delta M$  is more difficult to understand because of the feedback loop. Inspection of the circuit and some intuitive reasoning can provide a general understanding of how the  $\Sigma\Delta M$  works. The key to understanding is to start by looking at the quantizer. The analog signal  $u_n$  is quantized to produce the digital output signal  $y_n$ . The output  $y_n$  provides an estimate of the input. However, with a one bit quantizer, any one output value will very likely be a very poor approximation of the input. But, since the input is oversampled, a moving average of  $M$  output values will tend to track the input much better. The digital filtering and decimation block in figure 12 bring the frequency of the output back down to the desired frequency and, at the same time, filter out all noise energy above the signal band-

width. For the purposes of this study, it will be assumed that the low pass digital filter is ideal and that the decimation process is ideal. So, it should be sufficient to study the  $\Sigma\Delta M$  output signal  $y_n$  in order to understand the characteristics of the quantization noise.

Now, back to the discussion on the operation of the  $\Sigma\Delta M$ . The output  $y_n$  is converted back to an analog signal in the feedback loop and subtracted from the input  $x_n$  to produce  $w_n$ . The transfer function  $H(z)$  in the circuit will be assumed to be a simple discrete time analog integrator for the purposes of this study. Figure 13 shows the diagram of an integrator circuit. Note that the discrete time analog integrator is actually summing up the difference between the input and the output signal. This explains the  $\Sigma\Delta$  in the name of the circuit. Actually, since the integrator has a delay, the output signal lags one time unit behind the input. Nevertheless, if the input signal changes little between samples, then  $w_n$  will be a close approximation to the opposite of the total quantization error. If the oversampling ratio  $M$  is high and the input signal is limited to the signal bandwidth, then  $x_n$  will change slowly between samples.

By subtracting the total quantization error from  $u_n$ , the  $\Sigma\Delta M$  is continually attempting to correct the output value. Since the output of a one bit quantizer is most likely in error at any given instant, whenever a particular output state is reached, the  $\Sigma\Delta M$  immediately attempts to correct the error by heading towards the opposite output state. The number of cycles it takes to reverse the output depends on what the total quantization error is. Note that if level of  $x_n$  is close to

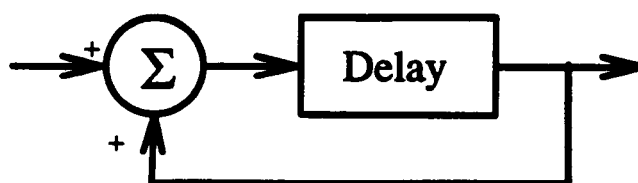


Figure 13. Discrete time analog integrator.

one of the output levels, then the total quantization error will be very small for one output value and very large for the other. Thus, when the output level nearest the input is reached, it may take many cycles for  $u_n$  to change sign. When it does, the output changes state and a large quantization error is subtracted from  $u_n$ , causing it to immediately change sign again. The output signal favors the output state closest to the level of the input value.

If a signal is close to zero, or midway between the quantizer outputs, then the output will tend to oscillate evenly between the two output states. Figure 14 shows an input sinusoid with its corresponding output signal. The behavior just described can be clearly observed in this figure. Whatever the input signal is, the  $\Sigma\Delta M$  has a tendency to oscillate. The frequency of the oscillation is determined by the level of the input signal. Thus, the input signal is being modulated by  $f_s$ .

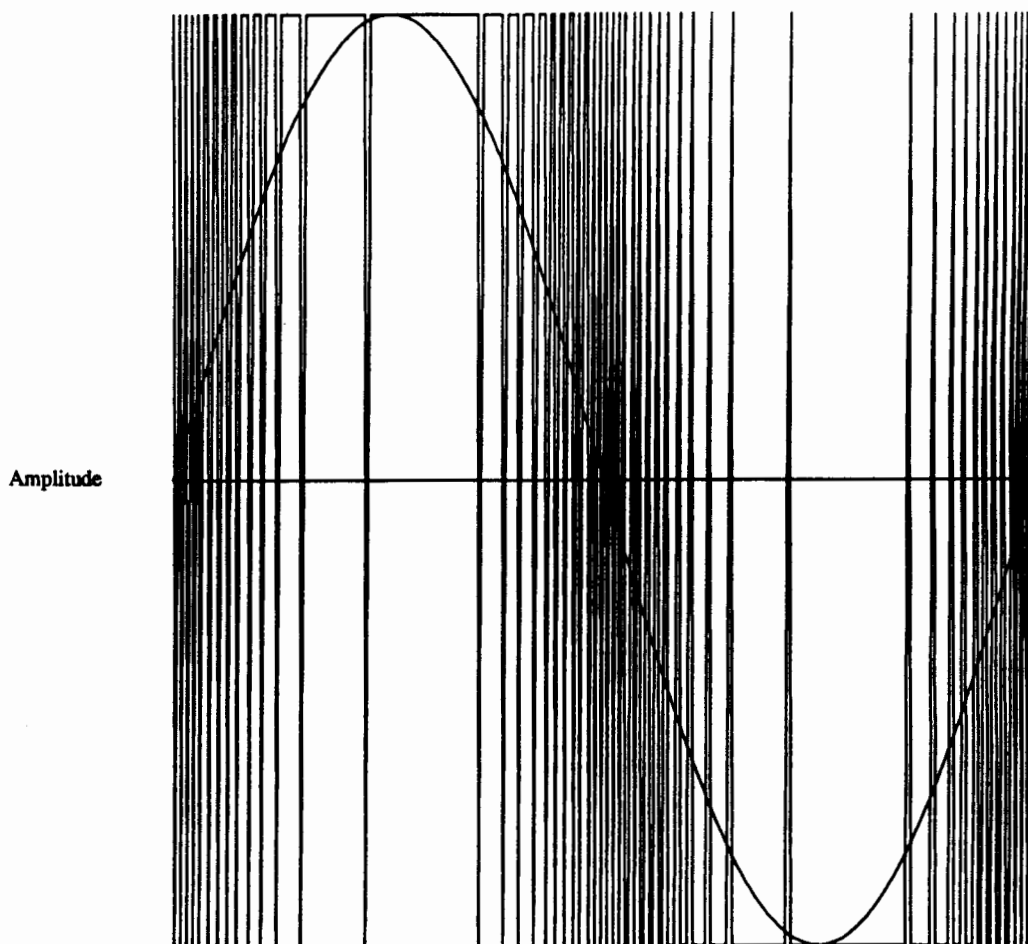


Figure 14.  $\Sigma\Delta M$  output with sinusoidal input.

Hence, the remainder of the circuits name becomes clear.

The purpose of this study is to understand the nature of the quantization noise. Given that the  $\Sigma\Delta M$  has a one bit quantizer, a lot of quantization noise is generated. However, if the input signal tends to change slowly and the output of the quantizer tends to oscillate, then the total quantization error will tend to be an active, rapidly changing signal. Consequently, it is likely that a significant portion of the spectral energy will be concentrated at higher frequencies. Since the output



signal will be ideally lowpass filtered with a cutoff at  $\frac{f_n}{2}$ , most of the quantization noise will be removed from the output signal. Recall how oversampling the uniform ADC resulted in a relative low frequency quantization noise signal which could not be filtered out for low frequency inputs. The  $\Sigma\Delta M$  eliminates this problem by shaping the quantization noise so that noise in the signal bandwidth is reduced and high frequency noise is increased. Noise shaping enables the  $\Sigma\Delta M$  to achieve greater performance through oversampling than the uniform ADC. However, it is important to study the quantization noise characteristics since the noise may not behave in the way it is assumed to behave, if, for example, the quantization noise is assumed to be white.

## V.2 FORMAL ANALYSIS OF THE $\Sigma\Delta M$

Now it is time to develop a more rigorous analysis of the  $\Sigma\Delta M$ , so that the general insights above can be validated. For the purposes of analysis the circuit can be simplified to the discrete time model shown in Figure 15. The quantizer has been modeled as an additive noise source. By writing down some of the difference equations that describe this model, further insight into the  $\Sigma\Delta M$  can be developed. The quantization error signal can be written as follows.

$$\epsilon_n = y_n - u_n. \tag{V.1}$$

With some algebraic manipulation, it can be shown that the difference equation describing the output of the  $\Sigma\Delta M$  in terms of the input and the quantizer

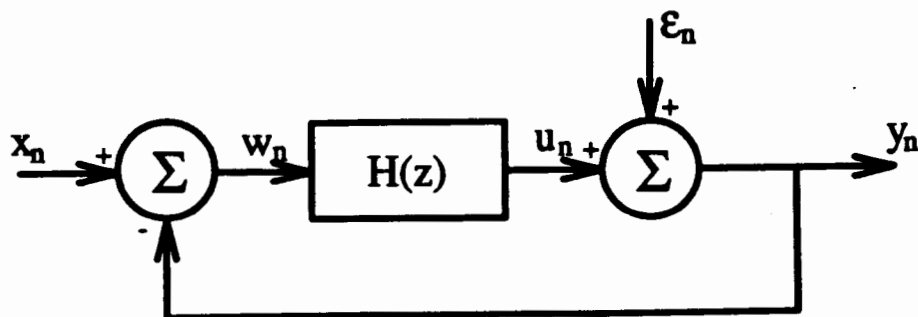


Figure 15.  $\Sigma\Delta M$  discrete time model.

error is

$$y_n = x_{n-1} + \epsilon_n - \epsilon_{n-1}. \quad (\text{V.2})$$

This equation clearly shows the relationship between the quantization error and the total quantization error. The total quantization error is  $\epsilon_n - \epsilon_{n-1}$ , which is just the difference between the current and previous quantization noise values. Based on the general discussion above of the operation of the  $\Sigma\Delta M$ , it is desirable that the spectrum of  $\epsilon_n - \epsilon_{n-1}$  have its energy concentrated mostly in the high frequencies.

In order to calculate the spectrum of the quantization error, it is useful to transform the time based difference equations into the frequency domain using the z-transform. Using the concept of superposition, the circuit model of figure 15 can be split into two models, one for each signal [1]. The signal model is shown

in figure 16 and the noise model is shown in figure 17. The outputs of these two models can be added together to get the output for the total model. Using these models, two transfer functions can be developed. One between the input and the output ( $W(z)$ ) and one between the noise and the output ( $T(z)$ ). These transfer functions are

$$W(z) = \frac{H(z)}{1 + H(z)} \quad (\text{V.3})$$

$$T(z) = \frac{1}{1 + H(z)} \quad (\text{V.4})$$

The output of the  $\Sigma\Delta M$  is the summation of  $x_n$  modified by  $W(z)$  and  $\epsilon_n$  modified by  $T(z)$ .

There are some general desired properties of the transfer functions  $W(z)$  and  $T(z)$  which are necessary to provide optimal noise shaping behavior. For the best response,  $W(z)$  should be flat in the low frequencies since the signal bandwidth should be passed through with as little disturbance as possible. On the other hand, to provide as much noise attenuation in the low frequency range as possible, the  $T(z)$  should be a high pass function which attenuates low frequencies.

For the  $\Sigma\Delta M$  configuration under consideration, the transfer function  $H(z)$  is the only part of the system available for modification. It is necessary to specify  $H(z)$  in order to analyze the affect it will have on  $W(z)$  and  $T(z)$ . A typical  $H(z)$  for the  $\Sigma\Delta M$  is the integrator, as shown in figure 13.

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \quad (\text{V.5})$$

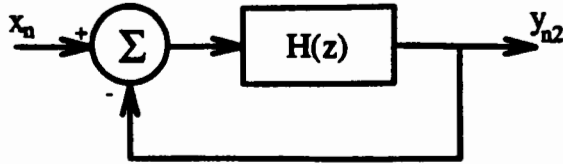


Figure 16.  $\Sigma\Delta M$  discrete time signal model.

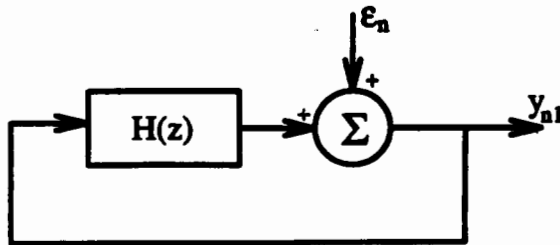


Figure 17.  $\Sigma\Delta M$  discrete time noise model.

Using this  $H(z)$ ,  $W(z)$  and  $T(z)$  can be found as follows.

$$W(z) = \frac{1}{z} \quad (\text{V.6})$$

$$T(z) = \frac{z-1}{z} \quad (\text{V.7})$$

Computing the magnitude of these transfer functions gives the frequency response of the transfer functions.

$$|W(z)|^2 = 1 \quad (\text{V.8})$$

$$|T(z)|^2 = 4 \sin^2\left(\pi \frac{f}{f_s}\right) \quad (\text{V.9})$$

$W(z)$  meets the requirement that it have a low pass frequency response. It allows the input signal to pass straight through. The  $T(z)$ , which is of greater

interest because it describes how the noise will behave, also has the desired characteristics. As figure 18 shows, it cuts off low frequencies and passes the high frequencies.

If  $|q(z)|^2$  and  $|\epsilon(z)|^2$  represent the power spectrum of the total quantizer noise and quantizer noise respectively, then

$$\begin{aligned} |q(z)|^2 &= |T(z)|^2 |\epsilon(z)|^2 \\ &= 4 \sin^2\left(\pi \frac{f}{f_s}\right) |\epsilon(z)|^2 \end{aligned} \quad (\text{V.10})$$

Equation V.10 describes how the quantizer noise spectrum is shaped to obtain the total quantizer noise spectrum.

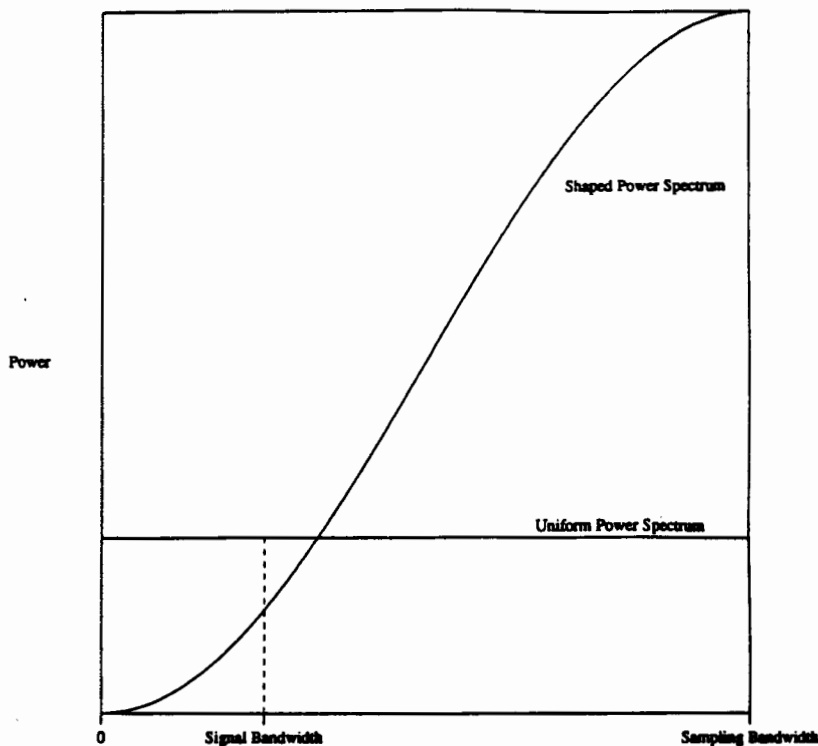


Figure 18. Spectrum shaping effect of  $T(z)$ .

The crucial factor which determines the total quantization noise is the quantization noise. If  $\epsilon(z)$  is known, then  $|q(z)|^2$  can be integrated over the baseband to obtain the total quantization noise power. If  $|\epsilon(z)|^2$  has a simple form, then a closed expression can be derived, otherwise, the integration can be performed numerically.

At this point, a couple different approaches can be taken. The simplest is to make the assumption that the quantizer noise has a uniform density, as was done for the uniform ADC. That is, the quantization noise is assumed to be white. If the input and the noise are uncorrelated, then the performance of the system depends only on the characteristics of the noise and  $H(z)$ . The more difficult approach is to assume that the noise is not independent of the input. In this case the characteristics of the noise will change as the input changes, making the performance of the  $\Sigma\Delta M$  dependent on the input signal as well. This results in a nonlinear system which is difficult to analyze.

Both of these approaches will be covered. First, an analysis based on the assumption of white quantization noise will be developed. The white noise analysis will assume that the input signal is a sinusoid. After this, analysis based on the actual operation of the  $\Sigma\Delta M$  will be developed. This will be divided into several parts based on the type of input signal. The reason for this is that in actual operation, the  $\Sigma\Delta M$  quantization noise has different characteristics for different kinds of inputs. First, dc inputs will be examined. The coverage of dc inputs will include theoretical analysis and simulations to compare against the theory. After

dc inputs are discussed, sinusoidal inputs will be covered in a similar fashion.

### V.3 $\Sigma\Delta M$ ANALYSIS USING WHITE NOISE ASSUMPTION

Simulation results for the uniform ADC showed that the assumption of white quantizer noise was not valid for low resolution quantizers, although it did a fairly good job for high resolution quantizers. The  $\Sigma\Delta M$  only has a one bit quantizer, but due to modulation, it produces an active output signal which should result in an active noise signal. This is different than the situation for the uniform ADC where the noise signal actually became less active as the oversampling ratio increased. Thus, it is reasonable to at least consider the possibility that assuming the quantizer noise is white for the  $\Sigma\Delta M$  might be able to provide a useful approximation to  $|\epsilon(z)|^2$ . Since the generation of noise in the  $\Sigma\Delta M$  is caused by a different process than in the uniform ADC, this assumption may not be valid. It will be necessary to validate the theories based on this assumption by simulation.

From equation II.3, it is known that the total quantization noise power for a quantizer with a uniform distribution is  $\frac{\Delta^2}{12}$ . Therefore,

$$|\epsilon(z)|^2 = \frac{\Delta^2}{12} \frac{1}{f_s} \quad (\text{V.11})$$

An equation for the SQNR, assuming a sinusoidal input signal, can be developed. The total quantization noise can be expressed as follows given that  $\Delta = 1$ .

$$|q_u(z)|^2 = (2 \sin \pi \frac{f}{f_s})^2 \frac{1}{12 f_s} \quad (\text{V.12})$$

To calculate the total quantization noise power, it is necessary to integrate over the baseband. This assumes that all spectral energy above the baseband will be filtered out. To simplify the integration, it is assumed that  $f_s \gg f_n$ . This will allow the assumption that  $\sin(x) \approx x$ .

$$\begin{aligned} P_n &= \int_{-\frac{f_n}{2}}^{\frac{f_n}{2}} |q_u(z)|^2 df \\ &= \frac{\pi^3}{36} \left(\frac{f_n}{f_s}\right)^3 \end{aligned} \quad (\text{V.13})$$

The total signal power is

$$P_s = \frac{a^2}{2} \quad (\text{V.14})$$

where  $a$  is the amplitude of the sinusoid, and so the SQNR is

$$SQNR = 18a^2 \frac{M^3}{\pi^3} \quad (\text{V.15})$$

If  $a$  is equal to the maximum input level of  $\frac{1}{2}$ , then

$$SQNR = 10 \log \left( \frac{9M^3}{2\pi^3} \right) = 30 \log(M) + 10 \log(9) - 10 \log(2\pi^3) \text{ db} \quad (\text{V.16})$$

Equation V.16 says that the SQNR increases by 9 db every time  $M$  is doubled. This is an improvement over the Uniform ADC which predicted 3 db improvement. Again, this theory must be verified by experiment since the assumption that the quantizer noise is white has not been determined to be true. In fact, it will be shown that the quantizer noise is not white. However, it is useful to find out how valid equation V.16 is, since it provides a intuitively simple explanation of how the noise shaping of the  $\Sigma\Delta M$  works.



## CHAPTER VI

### EXACT DC ANALYSIS OF THE $\Sigma\Delta M$

The assumption of white quantization noise provides a simple method of modeling the quantization noise characteristics of the  $\Sigma\Delta M$ . It is an assumption which has been shown to be reasonable for uniform ADC's under certain conditions. However, simulations of the uniform ADC demonstrated that the quantization noise is definitely not white for a four bit quantizer. Even though the  $\Sigma\Delta M$  has a one bit quantizer, the feedback loop should act in such a way to make the noise signal more active than it was for the uniform ADC. Thus, while the white noise assumption may prove to predict the SQNR behavior of the  $\Sigma\Delta M$  fairly well, it still does not provide any information about the actual noise characteristics.

If the white quantization noise assumption is not used, then another approach to analysis is to model the  $\Sigma\Delta M$  in a more exact way. This is done by solving for the noise characteristics of the actual output sequences for given input sequences instead of assuming the noise characteristics. To understand how the  $\Sigma\Delta M$  works in an exact way, it is useful to begin by examining how it works for a dc input. The dc input is an important signal to consider because it is the simplest signal to analyze and can be used later to develop an understanding of slowly varying signals. The dc performance is also important in some applications,

such as audio. After an exact analysis has been done for dc inputs, the analysis will be expanded to include sinusoidal inputs. The analysis of dc inputs will be presented in two parts. First will be an analysis of the structure of the output signal for dc inputs. This is a more formal way of describing some of the intuitive ideas discussed in the previous chapter. Also, it is useful to understand how the output signal is generated because it can provide insights into how the  $\Sigma\Delta M$  will operate for more complicated signals. Secondly, an analysis that predicts the exact structure of the quantization noise spectrum will be presented. Knowledge of the spectrum provides theoretical predictions for how much noise will be generated for a specific dc input.

## VI.1 $\Sigma\Delta M$ OUTPUT SIGNAL STRUCTURE FOR DC INPUTS

A  $\Sigma\Delta M$  with a 1 bit quantizer modulates the dc input at the rate  $f_s$  to produce an output signal which can be thought of as a sequence of square pulses of height  $\frac{\Delta}{2}$ . For convenience, it will be assumed that  $\Delta = 1$ . A moving average of  $M$  output values will approximate the dc input. It was noted earlier that the  $\Sigma\Delta M$  operates by adding the opposite of the total quantization error to the integrator at each time interval. For the case of dc inputs, there are only two possible values for the total quantization error,  $x + \frac{1}{2}$  and  $x - \frac{1}{2}$ , where  $x$  is the dc input value and  $\pm\frac{1}{2}$  are the output levels. Let

$$\alpha = x - \frac{1}{2} \tag{VI.1}$$

$$\beta = x + \frac{1}{2} \quad (\text{VI.2})$$

The  $\Sigma\Delta M$ 's operation is then described as follows.

$$u_n \geq 0 : u_{n+1} = u_n + \alpha \quad (\text{VI.3})$$

$$u_n < 0 : u_{n+1} = u_n + \beta \quad (\text{VI.4})$$

Note that  $\alpha \leq 0$  and  $\beta \geq 0$  for all  $x$  in the valid range.

A little thought will reveal that for any value of  $x$  in the valid range, the output will always stay at one of the output states for exactly one cycle. The output state this occurs on is the one farthest from  $x$ . For example, if  $x < 0$ , then  $|\alpha| > |\beta|$ . So, if  $u_n < 0$  and  $u_{n+1} > 0$ , then  $u_{n+1} < \beta$ . Thus,  $u_{n+2}$  will always be negative. If  $u_{n+1} = 0^+$ , then  $u_{n+2} = \alpha$ . It will take exactly  $|\frac{\alpha}{\beta}|$  cycles for  $u_n$  to become positive again. Of course,  $|\frac{\alpha}{\beta}|$  is most likely not an integer value and this is a discrete time system, so  $|\frac{\alpha}{\beta}|$  represents the average number of cycles it will take for  $u_n$  to become positive.

For  $x < 0$ , the output signal will consist of a sequence of, on average,  $|\frac{\alpha}{\beta}|$  negative pulses followed by one positive pulse. So, the output is a can be thought of as a square wave which has a frequency of

$$\begin{aligned} f_x &= \frac{1}{|\frac{\alpha}{\beta}| + 1} f_s \\ &= \beta f_s \end{aligned} \quad (\text{VI.5})$$

This equation holds for all values of  $x$  in the valid range. If the equation had been

developed based on an example where  $x > 0$ , then the equation would be

$$\begin{aligned} f_x &= \frac{1}{|\frac{\beta}{\alpha}| + 1} f_s \\ &= -\alpha f_s \end{aligned} \tag{VI.6}$$

Taking into account the frequency folding about  $\frac{f_s}{2}$  due to aliasing

$$f_x = \beta f_s = -\alpha f_s \tag{VI.7}$$

For  $x < 0$ , the number of negative pulses will actually be either  $\langle |\frac{\alpha}{\beta}| \rangle$  or  $\langle |\frac{\alpha}{\beta}| \rangle + 1$  where  $\langle a \rangle$  is the integer part of  $a$ . The signal obviously contains noise since the input is a constant and the output is an oscillating signal. It would appear that  $f_x$  could be the fundamental frequency of this noise. It will be shown that this is true.

## VI.2 RECURSIVE STRUCTURE OF OUTPUT SIGNAL

The previous analysis describes the average behavior of the output signal for dc inputs to the  $\Sigma\Delta M$ . That is, if the ratio of  $\frac{\alpha}{\beta}$  is not a whole number or the reciprocal of a whole number, then the frequency  $f_x$  can only be the average frequency of the square wave and not a true description of the output signal. The analysis can be extended to give a description of the output behavior to any desired precision.

The description of the output signal turns out to be recursive in nature. The original dc input  $x_0$  results in an output signal composed of the two output states

of the quantizer. The output wave will be a periodic sequence with  $\mathcal{R}$  outputs of the output state closest to the dc input and one output of the other state, where

$$\mathcal{R} = \begin{cases} \left\lfloor \frac{|\alpha|}{|\beta|} \right\rfloor & \text{if } |\alpha| > |\beta| \\ \left\lfloor \frac{|\beta|}{|\alpha|} \right\rfloor & \text{if } |\beta| > |\alpha| \end{cases}$$

If  $\mathcal{R}$  is not an integer, then this is not physically possible, so the output will consist of two patterns. One pattern is  $\langle \mathcal{R} \rangle$  of the output state closest to the dc input and one output of the other state. The other pattern is  $\langle \mathcal{R} \rangle + 1$  of the output state closest to the dc input and one output of the other state. The pattern that  $\mathcal{R}$  is closest to will tend to occur more often. This is analogous to the original dc input. So, if  $[\mathcal{R}] - \frac{1}{2}$  is thought of as an input to a  $\Sigma\Delta M$  and the two patterns are thought of as the two output states, then the first level of recursion has been described. This process can be continued until  $\mathcal{R}_n$  is an integer value. If the original dc input is rational, then the recursion will eventually end and the output signal will be periodic. If the dc input is irrational, then the recursion will continue infinitely and the output signal will never repeat itself.

The recursive nature of the  $\Sigma\Delta M$  output signal for a dc input can be described by the following notation. The initial level is described first. The  $\mathcal{L}$  terms are used to denote the output states of the current recursion level.

$$x_0 = \text{dc Input Level} \tag{VI.8}$$

$$\alpha_0 = x_0 - \frac{1}{2} \tag{VI.9}$$

$$\beta_0 = x_0 + \frac{1}{2} \tag{VI.10}$$

$$\mathcal{R}_0 = \begin{cases} \left| \frac{\alpha_0}{\beta_0} \right| & \text{if } |\alpha_0| > |\beta_0| \\ \left| \frac{\beta_0}{\alpha_0} \right| & \text{if } |\beta_0| > |\alpha_0| \end{cases} \quad (\text{VI.11})$$

$$\mathcal{L}_{min_0} = \frac{-1}{2} \quad (\text{VI.12})$$

$$\mathcal{L}_{max_0} = \frac{1}{2} \quad (\text{VI.13})$$

$$\mathcal{L}_{one_0} = \begin{cases} \mathcal{L}_{min_0} & \text{if } x_0 > 0 \\ \mathcal{L}_{max_0} & \text{if } x_0 < 0 \end{cases} \quad (\text{VI.14})$$

$$\mathcal{L}_{many_0} = \begin{cases} \mathcal{L}_{max_0} & \text{if } x_0 > 0 \\ \mathcal{L}_{min_0} & \text{if } x_0 < 0 \end{cases} \quad (\text{VI.15})$$

$$f_{x_0} = \frac{1}{|\mathcal{R}_0| + 1} f_s \quad (\text{VI.16})$$

Now the  $n^{\text{th}}$  level of recursion is defined.

$$x_n = [\mathcal{R}_{n-1}] - \frac{1}{2} \quad (\text{VI.17})$$

$$\alpha_n = x_n - \frac{1}{2} \quad (\text{VI.18})$$

$$\beta_n = x_n + \frac{1}{2} \quad (\text{VI.19})$$

$$\mathcal{R}_n = \begin{cases} \left| \frac{\alpha_n}{\beta_n} \right| & \text{if } |\alpha_n| > |\beta_n| \\ \left| \frac{\beta_n}{\alpha_n} \right| & \text{if } |\beta_n| > |\alpha_n| \end{cases} \quad (\text{VI.20})$$

$$\mathcal{L}_{min_n} = \langle \mathcal{R}_{n-1} \rangle \mathcal{L}_{many_{n-1}}, \mathcal{L}_{one_{n-1}} \quad (\text{VI.21})$$

$$\mathcal{L}_{max_n} = (\langle \mathcal{R}_{n-1} \rangle + 1) \mathcal{L}_{many_{n-1}}, \mathcal{L}_{one_{n-1}} \quad (\text{VI.22})$$

$$\mathcal{L}_{one_n} = \begin{cases} \mathcal{L}_{min_n} & \text{if } x_n > 0 \\ \mathcal{L}_{max_n} & \text{if } x_n < 0 \end{cases} \quad (\text{VI.23})$$

$$\mathcal{L}_{many_n} = \begin{cases} \mathcal{L}_{max_n} & \text{if } x_n > 0 \\ \mathcal{L}_{min_n} & \text{if } x_n < 0 \end{cases} \quad (\text{VI.24})$$

$$f_{x_n} = \frac{1}{|\mathcal{R}_n| \text{size}(\mathcal{L}_{many_n}) + \text{size}(\mathcal{L}_{one_n})} f_s \quad (\text{VI.25})$$

If  $[\mathcal{R}_n] = 0$ , then  $\mathcal{L}_{max_{n+1}} = \mathcal{L}_{min_{n+1}}$ , and the series of recursions ends.

Observe that the output sequence is actually defined by a sequence of dc inputs which are derived from the fractional remainder of the ratio  $R$ . This series of dc

inputs can be defined as

$$x_0, x_1, x_2, \dots, x_n = [\mathcal{R}_{n-1}] - \frac{1}{2}; n = 1, 2, 3, \dots \quad (\text{VI.26})$$

If the input  $x_0$  is a rational number, then the series of equation VI.26 is finite and the output sequence of the  $\Sigma\Delta M$  is periodic. If the input  $x_0$  is an irrational number, then the series of equation VI.26 is infinite and the output sequence of the  $\Sigma\Delta M$  is not periodic. Thus, the basic structure of the output sequence has been defined. It can be roughly described as a square wave which has a frequency of  $f_{x_0}$ , as discussed above. However, since the  $\Sigma\Delta M$  is a discrete time system and  $f_s$  is unlikely to be an exact multiple of  $f_{x_0}$ , the recursive algorithm detailed above can be used to generate the patterns of highs and lows in the output signal. The average frequency  $f_{x_n}$  for recursion levels greater than 0 have not been observed to indicate anything of significance. These sub-frequencies will be referred to later on. Figure 19 shows an example of how the algorithm works for a specific rational dc input value.

### VI.3 $\Sigma\Delta M$ NOISE CHARACTERISTICS FOR DC INPUTS

The previous analysis provides a detailed description of how the output signal of the  $\Sigma\Delta M$  is structured. However, it does not easily lend itself to providing a description of the power spectrum of the signal. If the dc input is rational, then a Fourier Series analysis similar to the analysis done for the Uniform ADC above could be done since the output sequence is periodic. But, if the dc input

**Level 0**

$$\begin{aligned}
 x_0 &= \frac{3}{40} \\
 \alpha_0 &= \frac{-17}{40}, & \beta_0 &= \frac{23}{40} \\
 \mathcal{R}_0 &= \frac{23}{17} = 1 \frac{6}{17} = 1 \frac{12}{34} \\
 \mathcal{L}_{\min_0} &= \frac{-1}{2}, & \mathcal{L}_{\text{many}_0} &= \frac{1}{2} \\
 \mathcal{L}_{\max_0} &= \frac{1}{2}, & \mathcal{L}_{\text{one}_0} &= \frac{-1}{2} \\
 f_{x_0} &= \frac{1}{\frac{23}{17}+1} f_s = \frac{17}{40} f_s
 \end{aligned}$$

**Level 1**

$$\begin{aligned}
 x_1 &= \frac{12}{34} - \frac{1}{2} = \frac{-5}{34} \\
 \alpha_1 &= \frac{-22}{34}, & \beta_1 &= \frac{12}{34} \\
 \mathcal{R}_1 &= \frac{22}{12} = 1 \frac{10}{12} \\
 \mathcal{L}_{\min_1} &= 1 \cdot \mathcal{L}_{\text{many}_0}, 1 \cdot \mathcal{L}_{\text{one}_0}, & \mathcal{L}_{\text{many}_1} &= \mathcal{L}_{\min_1} \\
 \mathcal{L}_{\max_1} &= 2 \cdot \mathcal{L}_{\text{many}_0}, 1 \cdot \mathcal{L}_{\text{one}_0}, & \mathcal{L}_{\text{one}_1} &= \mathcal{L}_{\max_1} \\
 f_{x_1} &= \frac{1}{\frac{22}{12}(2)+3} f_s = \frac{6}{40} f_s
 \end{aligned}$$

**Level 2**

$$\begin{aligned}
 x_2 &= \frac{10}{12} - \frac{1}{2} = \frac{4}{12} \\
 \alpha_2 &= \frac{-2}{12} = \frac{-1}{6}, & \beta_2 &= \frac{10}{12} = \frac{5}{6} \\
 \mathcal{R}_2 &= 5 \\
 \mathcal{L}_{\min_2} &= 1 \cdot \mathcal{L}_{\text{many}_1}, 1 \cdot \mathcal{L}_{\text{one}_1}, & \mathcal{L}_{\text{many}_2} &= \mathcal{L}_{\max_2} \\
 \mathcal{L}_{\max_2} &= 2 \cdot \mathcal{L}_{\text{many}_1}, 1 \cdot \mathcal{L}_{\text{one}_1}, & \mathcal{L}_{\text{one}_2} &= \mathcal{L}_{\min_2} \\
 f_{x_2} &= \frac{1}{5(2)+5} f_s = \frac{1}{40} f_s
 \end{aligned}$$

**Output Wave** =  $5 \cdot \mathcal{L}_{\text{many}_2}, 1 \cdot \mathcal{L}_{\text{one}_2}$

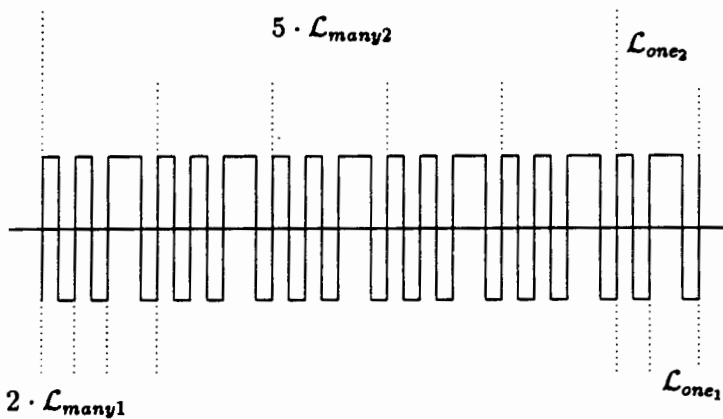


Figure 19. Example with dc input of  $\frac{3}{40}$ .



is irrational, then the output signal is not periodic and a Fourier Series analysis would not give an exact description of the spectrum. Irrational response could be approximated by truncating the output sequence at a suitable point and then continuing the analysis as if the input were rational.

Gray [7] has developed a method of exact analysis which is able to predict the power spectrum of the  $\Sigma\Delta M$  for all dc inputs. Gray splits the analysis into the rational and irrational cases. This seems reasonable since the recursive algorithm above indicates that the output signal has different characteristics depending on the rationality of the input. Gray develops the analysis by first developing an expression to describe the error sequence of the quantizer. The first and second moments are then derived for the error sequence. Gray then develops an expression for the autocorrelation. The next step is to find the fourier transform of the autocorrelation which is well known to represent the power spectral density of the sequence. The resulting power spectrum describes the characteristics of the quantizer error. Gray makes the point that the result is definitely not white noise. The analysis is continued to find the spectrum of the total quantization noise. The spectrum of the total quantization noise can also be found by multiplying the spectrum of the quantization noise by the noise shaping expression from above (equation 4.6b).

## VI.4 FUNDAMENTAL DEFINITIONS AND RESULTS

Here are the results of Gray's development. Gray begins by developing expressions for the normalized quantizer and quantizer error sequences,  $y_n$  and  $\zeta_n$ .

$$y_n = \frac{q(u_n)}{2b} = \langle (n-1)\beta \rangle - \langle n\beta \rangle + \frac{x}{2b} \quad (\text{VI.27})$$

and

$$\zeta_n = \frac{\epsilon_n}{2b} = \frac{1}{2} - \langle n\beta \rangle \quad (\text{VI.28})$$

where  $b = \frac{\Delta}{2}$ ,  $\beta = \frac{x}{2b} + \frac{1}{2}$ , and  $n = 0, 1, 2, \dots$ . Note that  $y_n$  is the output of the  $\Sigma\Delta M$  and  $u_n$  is the output of the integrator. Gray observes that  $y_n$  and  $\zeta_n$  can be described in terms of the following simpler sequence.

$$w_n = \langle n\beta \rangle \quad (\text{VI.29})$$

Thus,

$$\zeta_n = \frac{1}{2} - w_n \quad (\text{VI.30})$$

and

$$y_n = w_{n-1} - w_n + \frac{x}{2b} \quad (\text{VI.31})$$

Gray develops his analysis by studying the sequence  $w_n$ , which is a well known sequence in the field of ergodic theory. The form of the analysis differs depending on the nature of  $\beta$ . As the recursive analysis above demonstrated, if  $\beta$  is rational, then  $w_n$  will be periodic with period  $N$ , if  $\beta = \frac{K}{N}$  where  $\frac{K}{N}$  is in lowest terms. A periodic signal can be analyzed using a fourier series method similar to

the method used to analyze the uniform ADC. If  $\beta$  is irrational, then the sequence  $w_n$  is not periodic and a different method is necessary to analyze it.

The mean or time average of a sequence is an important element of Gray's development. Gray defines the sample average for a time sequence  $\{g_n; n = 0, 1, 2, \dots\}$  as

$$M\{g_n\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^N g_i \quad (\text{VI.32})$$

If the time sequence is produced by a stationary and ergodic random process, then the mean would be defined as an expectation. In this case, the time average of the sequence is being dealt with directly. The operator  $M$  is similar to the expectation and is linear in the sense that  $M\{ag_n + bf_n\} = aM\{g_n\} + bM\{f_n\}$ .

## VI.5 MOMENTS OF IRRATIONAL INPUTS

Another important tool for Gray's analysis when dealing with irrational values of  $\beta$  is a result from ergodic theory known as Weyl's formula. For any integrable function  $f$  and for any irrational  $\beta$ , it is known for any  $y$  that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\langle y - k\beta \rangle) = \int_0^1 f(r) dr \quad (\text{VI.33})$$

Without going into a proof of this, a little thought can provide insight into the correctness of this result. On the left side, as  $n \rightarrow \infty$ , the quantity  $\langle y - k\beta \rangle$  will take on every value in the range  $[0, 1)$  in the limit since  $\beta$  is irrational. So, the left side of equation VI.33 covers the same interval as the right side, just not in the same order.

Using these mathematical tools for analyzing the case for irrational values of  $\beta$  the moments of the sequence  $w_n$  can be derived.

$$M\{w_n\} = \frac{1}{2} \quad (\text{VI.34})$$

$$M\{w_n^2\} = \frac{1}{3} \quad (\text{VI.35})$$

Using these results and the linearity of the  $M$  operator, the moments of the quantizer error sequence  $\zeta_n$  can be derived.

$$M\{\zeta_n\} = 0 \quad (\text{VI.36})$$

and

$$M\{\zeta_n^2\} = \frac{1}{12} \quad (\text{VI.37})$$

The autocorrelation can also be found. For non-negative  $l$ , the autocorrelation function is shown to be

$$r_w(l) = M\{w_k w_{k+l}\} = \frac{1}{3} - \frac{1}{2} \langle l\beta \rangle (1 - \langle l\beta \rangle) \quad (\text{VI.38})$$

From  $r_w(l)$ , the autocorrelation for the quantizer error sequence is shown to be

$$r_\zeta(l) = \frac{1}{12} - \frac{\langle l\beta \rangle}{2} (1 - \langle l\beta \rangle) \quad (\text{VI.39})$$

Gray observes that the quantizer error sequence has mean values which are identical to the values of a uniformly distributed random variable over the same range. This is consistent with the frequent assumption that the quantizer error is a uniformly distributed error. However, the second order properties seen in the

autocorrelation function diverge from this assumption. If the error sequence was truly random, or uncorrelated, then  $r_\zeta(l)$  would be 0 for all nonzero  $l$ . This is obviously not true.

## VI.6 MOMENTS OF RATIONAL INPUTS

In order to study the sequence  $w_n$  for rational values of  $\beta$ , some more mathematical tools need to be developed. Instead of using the result from ergodic theory (eqn VI.33), the following result from the study of finite fields is used. If  $\frac{K}{N}$  is in lowest terms, then the collection of numbers

$$\left\{ \left\langle \frac{kK}{N} \right\rangle; k = 0, 1, 2, \dots, N - 1 \right\}$$

is the same as the collection of numbers

$$\left\{ \frac{k}{N}; k = 0, 1, 2, \dots, N - 1 \right\}.$$

Note that each collection contains the same set of numbers, although the numbers are not generated in the same order. This is similar to the process involved in Weyl's formula, except that for the case of irrational numbers, there are an infinite number of elements in the collection.

Using the result for rational numbers, the moments of the sequence  $w_n$  can be derived.

$$M\{w_n\} = \frac{1}{2} \frac{N - 1}{N} \tag{VI.40}$$

and the second moment is

$$M\{w_n^2\} = \frac{1}{3} - \frac{1}{2N} + \frac{1}{6N^2} \quad (\text{VI.41})$$

For rational input values the first moment of the quantizer error sequence is

$$M\{\zeta_n\} = \frac{1}{2N} \quad (\text{VI.42})$$

and the second moment is

$$M\{\zeta_n^2\} = \frac{1}{12} + \frac{1}{6N^2} \quad (\text{VI.43})$$

The autocorrelation can then be derived for rational values of  $\beta$ .

$$r_w(l) = M\{w_k w_{k+l}\} = \frac{1}{3} + \frac{1}{6N^2} - \frac{1}{2} \langle \beta \rangle (1 - \langle \beta \rangle) \quad (\text{VI.44})$$

From  $r_w(l)$ , the autocorrelation for the quantizer error sequence is shown to be

$$r_\zeta(l) = \frac{1}{12} + \frac{1}{6N^2} - \frac{\langle l\beta \rangle}{2} (1 - \langle l\beta \rangle) \quad (\text{VI.45})$$

The moments and autocorrelation for rational  $\beta$  are similar to the results for irrational  $\beta$  except that the results for rational  $\beta$  also include a term which depends on the period of the error sequence.  $N$  represents the period of the sequence and is closely related to the denominator of the lowest fractional form of the rational input. Note that as  $N$  becomes large, or as the input becomes closer to being irrational, the results for rational  $\beta$  approach the results for irrational  $\beta$ .

## VI.7 COMPUTING THE SPECTRUM AND THE BOHR-FOURIER SERIES

It is a well known result in the analysis of deterministic waveforms or sequences that the fourier transform of the auto-correlation  $r_\zeta(l)$  represents the power

spectral density of the sequence. This result applies to the case where  $\beta$  is rational since  $r_\zeta(l)$  for rational  $\beta$  will be periodic in  $l$  and have a period of  $N$  samples. However, for irrational inputs, the auto-correlation will not be periodic, so the  $r_\zeta(l)$  may not have a Fourier sequence. It can be shown that  $r_\zeta(l)$  is a nearly periodic sequence and that a Bohr- Fourier series can be found for it. Gray concludes that the power spectral density for irrational  $\beta$  can be represented exactly with the Bohr-Fourier sequence.

A sequence  $g_n$  possesses a Bohr-Fourier series if there is a countable sequence of distinct number  $\lambda_l$  in the range  $[0, 1]$  such that

$$\lim_{N \rightarrow \infty} M \left\{ \left| g_k - \sum_{l=-N}^{\infty} a_l e^{2\pi i \lambda_l k} \right|^2 \right\} = 0 \quad (\text{VI.46})$$

If this is true, then

$$g_n = \sum_{l=-\infty}^{\infty} a_l e^{2\pi i n \lambda_l} \quad (\text{VI.47})$$

This is actually a generalized form of a Fourier series, except that it is not necessary that  $g_n$  be periodic. If  $g_n$  is periodic with period  $N$ , then the Bohr-Fourier series reduces to a Fourier series where  $\lambda_l = \frac{l}{N}, l = 0, 1, 2, \dots, N - 1$ .

For the case of irrational  $\beta$  and the sequence  $\zeta_n$ , Gray shows that

$$\zeta_n = \sum_{k=-\infty}^{\infty} a_k e^{2\pi i n \langle k \beta \rangle} \quad (\text{VI.48})$$

where

$$a_k = M \left\{ \zeta_l e^{-i 2\pi k l \beta} \right\} = \begin{cases} 0 & \text{if } k = 0 \\ \frac{i}{2\pi k} & \text{if } k \neq 0 \end{cases}$$

Note that  $\lambda_l$  has been replaced by  $\langle k \beta \rangle$ , which is a countable infinite collection of distinct number in  $[0, 1]$ .

## VI.8 SPECTRUM RESULTS FOR IRRATIONAL AND RATIONAL INPUTS

After proving these results, Gray shows that the power spectral density for irrational values of  $\beta$  is

$$s_{\zeta}(k) = \begin{cases} 0 & \text{if } k = 0 \\ \frac{1}{(2\pi k)^2} & \text{if } k \neq 0 \end{cases} \quad (\text{VI.49})$$

$s_{\zeta}(k)$  is the area of the impulse located at the frequencies  $(k(\frac{1}{2} + x))$ , where the sampling frequency has been normalized. Due to aliasing, the frequencies are the fractional portion of the number and are folded into the range  $[0, 0.5)$ .

For rational inputs the power spectral density is found to be as follows

$$s_{\zeta}(k) = \begin{cases} \frac{1}{2N} & \text{if } k = 0 \\ \frac{1}{N} \frac{1}{\sin^2 \pi \frac{k}{N}} & \text{if } k \neq 0 \end{cases} \quad (\text{VI.50})$$

This result differs from the irrational case, but note that it converges to the irrational result as  $N$  becomes large.

## VI.9 EQUALITY OF IRRATIONAL AND RATIONAL RESULTS

An interesting point which Gray did not bring out is the relationship between the results for rational and irrational cases. He makes the point that the rational result will converge to the irrational result as  $N$  becomes large. However, there is a further observation to make. The first thing to note is that the equation to determine the frequencies of the noise values is the same for both cases. All the noise pulses are harmonics of  $f_{x_0}$ , which was defined earlier as the frequency of the average square wave which could define the output signal. The only differ-



ence between rational and irrational cases is that the irrational spectral frequencies never repeat and the rational frequencies begin to repeat after  $N$  values. The rational result has a sum which goes from 0 to  $N - 1$  while the irrational result goes from  $-\infty$  to  $+\infty$ . There is no reason why the irrational result cannot be used with rational numbers. The only thing to keep in mind is that the total spectral value for a particular frequency is the sum of an infinite number of terms since the frequencies overlap. The following equation should clarify the intended meaning here.

$$\frac{1}{4N^2} \frac{1}{\sin(\pi \frac{K}{N})} = \frac{1}{4\pi^2} \sum_{n=0}^{\infty} \left( \frac{1}{(k + nN)^2} + \frac{1}{(N - k + nN)^2} \right) \quad (\text{VI.51})$$

for  $k = 1, 2, \dots, N - 1$

Equation VI.51 can be shown to be true through the use of a computer program. A mathematical proof of this has not been determined yet. Table I shows computed results for equation VI.51. The results shown are specifically for low values of  $N$ , since Gray's results are most different for low  $N$ .

The results in Table I show clearly that equation VI.51 is valid. The irrational result can be used to find the spectrum for irrational and rational cases. The rational result is just a more convenient form to use when the rational case is being considered. Actually, equation VI.51 does not include the dc component, or the result for  $k = 0$ . At this time, it has not been shown that the rational and irrational results match exactly for the dc component.

TABLE I  
RATIONAL AND IRRATIONAL THEORY RESULTS

N	K	Irrational Theory (k = 1 to 10000)	Rational Theory	Difference
2	1	0.062499	0.062500	0.000001
4	1	0.031250	0.031250	0.000000
	2	0.015625	0.015625	0.000000
	3	0.031250	0.031250	0.000000
8	1	0.026673	0.026674	0.000000
	2	0.007812	0.007812	0.000000
	3	0.004576	0.004576	0.000000
	4	0.003906	0.003906	0.000000
	5	0.004576	0.004576	0.000000
	6	0.007812	0.007813	0.000000
	7	0.026673	0.026674	0.000000

#### VI.10 TOTAL QUANTIZATION ERROR SPECTRUM

To find the power spectrum for the total quantization error, just modify the results by the noise shaping factor developed above. For irrational inputs,

$$P_n(k) = 4 \sin^2\left(\pi \frac{f}{f_s}\right) \frac{1}{(2\pi k)^2}; \text{ if } k \neq 0 \quad (\text{VI.52})$$

For rational inputs,

$$P_n(k) = 4 \sin^2\left(\pi \frac{f}{f_s}\right) \frac{1}{N \sin^2 \pi \frac{k}{N}}; \text{ if } k \neq 0 \quad (\text{VI.53})$$

However, since it has been shown that the irrational theory, which is simpler, can produce the same results as the rational theory, the irrational theory will be used for the simulations of the  $\Sigma\Delta M$ .

## CHAPTER VII

### SIMULATION OF $\Sigma\Delta M$ WITH DC INPUTS

Although dc inputs were not even considered with the uniform ADC, it is important to consider the dc input with the  $\Sigma\Delta M$ . A dc input will produce a fixed error signal with a uniform ADC. However, due to the modulating operation of the  $\Sigma\Delta M$ , a dc input will have a noise signal which, as the theory predicts, will be a series of harmonics with a fundamental period related to the dc value. Simulations were performed for the  $\Sigma\Delta M$  with a series of dc inputs.

#### VII.1 SETUP OF SIMULATIONS

Some comments are in order at this point. First of all, the equations for irrational dc inputs will be used to make the theoretical calculations. Although numbers used in a computer simulation are rational, most will have a long period, so the irrational equation should give a good approximation. Also, it has been shown in the previous chapter that the irrational theory will produce the same answer as the rational theory. The theory predicts that the spectrum of the output can occur anywhere in the continuous frequency range  $[0, \frac{f_s}{2})$ . Simulation results, on the other hand, will be confined to  $\frac{N}{2}$  discrete frequencies in that same range where  $N$  is the number of bins used to compute the DFT. In order to compare theory with

simulation, the frequencies of the theoretical values were rounded to the nearest DFT bin frequency and added to the value for that bin. In order to reduce error caused by this, the DFT's were calculated with 65536 points, which results in a 32768 point spectrum for the range  $[0, \frac{f_s}{2}]$ . The final result is two power spectrums. The simulation spectrum is obtained by computing the DFT of a simulated  $\Sigma\Delta M$  output sequence. The theoretical spectrum is obtain by calculations based on the theory of the previous chapter. Although the theoretical frequencies are quantized to the DFT bin frequencies, this error should not be significant since 32768 bins are used.

With these two spectrums, it is now possible to obtain noise characteristics of the  $\Sigma\Delta M$  for dc inputs at different values of  $M$  (oversampling ratio). The measure of interest here is the total quantization noise power. This is the integration of the power spectrum over the range  $(0, \frac{f_s}{2M})$ . The simulations and theoretical calculations were performed for values of  $M$  ranging from 1 to 256 for 205 evenly spaced dc input values in the range  $[-0.499502, 0.499502]$ . These particular numbers were chosen in order to avoid using rational inputs with small denominators. For each input value, the spectrum is obtained experimentally and theoretically and then each spectrum is integrated over the baseband to find the total noise power in the baseband.

For each input value, a theoretical and simulated spectrum of 32768 points is computed. The 32768 point spectrum can be used to find the total noise power for all values of  $M$  from 1 to 256. This means that the frequency resolution of

the signal bandwidth for  $M = 256$  is 128 points. As  $M$  is reduced, the frequency resolution of the signal bandwidth increases. This point is mentioned only because for the simulations of the uniform ADC with sinusoid inputs, the number of points in the spectrum was based on  $M$  so that there were always 256 points in the signal bandwidth after filtering was performed. The only impact is that low values of  $M$  will have a more finely resolved spectrum, but the total noise power will not change.

A practical consideration for integrating the simulated spectrum is that the first 5 values must be dropped since the spike at dc will spread over the next 4 bins due to spectral smearing caused by the Blackman-Harris data window. For integration of the theoretical spectrum, only the first bin needs to be ignored. The final result is that the total noise power of the experimental spectrum matches the total noise power of the theoretical spectrum exactly to at least 6 decimal places. Since a very high resolution DFT (65536 points) was used, this could be expected. Table II, shows the difference between the theoretical and simulated results for a few example input values.

The fact that the simulated and theoretical total noise power matches indicates that the theory is correct. However, the total noise power is just one characteristic of the quantization noise. Since the total noise power is an average, it does not provide detailed information about about the structure of the noise spectrum. Figure 20 shows the total noise power for the range of dc input values for a number of oversampling ratios. These figures show that the total noise power

TABLE II  
TOTAL QUANTIZATION NOISE POWER

M	DC Input levels			
	0.004897	0.176295	0.249751	0.499502
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000
64	0.000000	0.000000	0.000000	0.000000
128	0.000000	0.000000	0.000000	0.000000
256	0.000000	0.000000	0.000000	0.000000

varies widely depending upon the dc input.

In order to understand why the noise power varies for different inputs, it is necessary to go back to the simulations and theory and look at the spectrum of the noise. The exact theory for describing the quantization noise spectrum is relatively simple. It says that the power spectrum is made up of a fundamental frequency, which is dependent on the input frequency, and all of the harmonics of the fundamental frequency. The power of the harmonics decrease by the square of the harmonic. So, most of the noise power will be located in the fundamental and first few harmonics. The problem at hand is to determine how this noise will contribute to the noise power of the  $\Sigma\Delta M$  system.

An intuitive understanding can be developed by first considering a dc input of zero. With a dc input of zero, the output of the  $\Sigma\Delta M$  will change at every

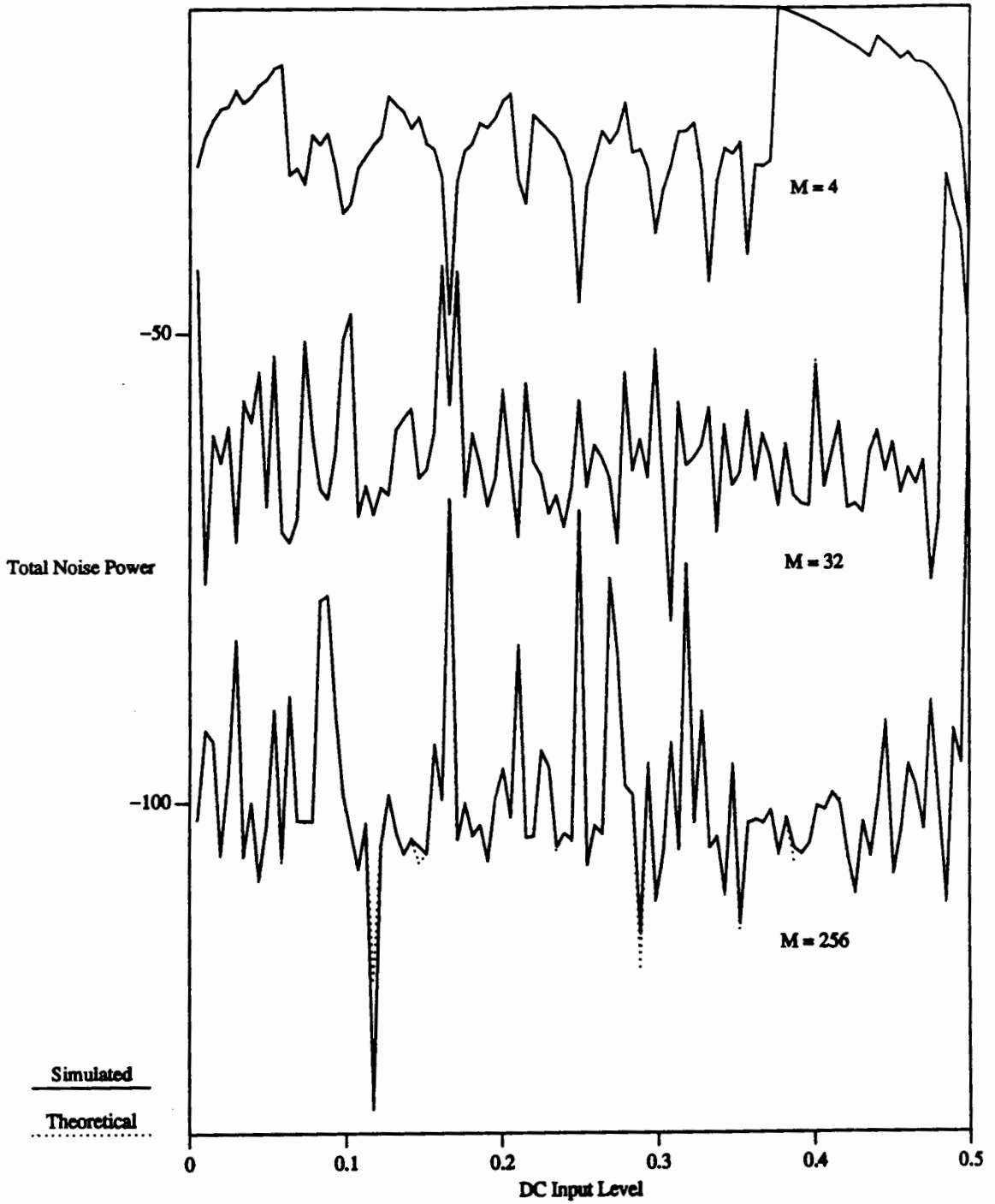
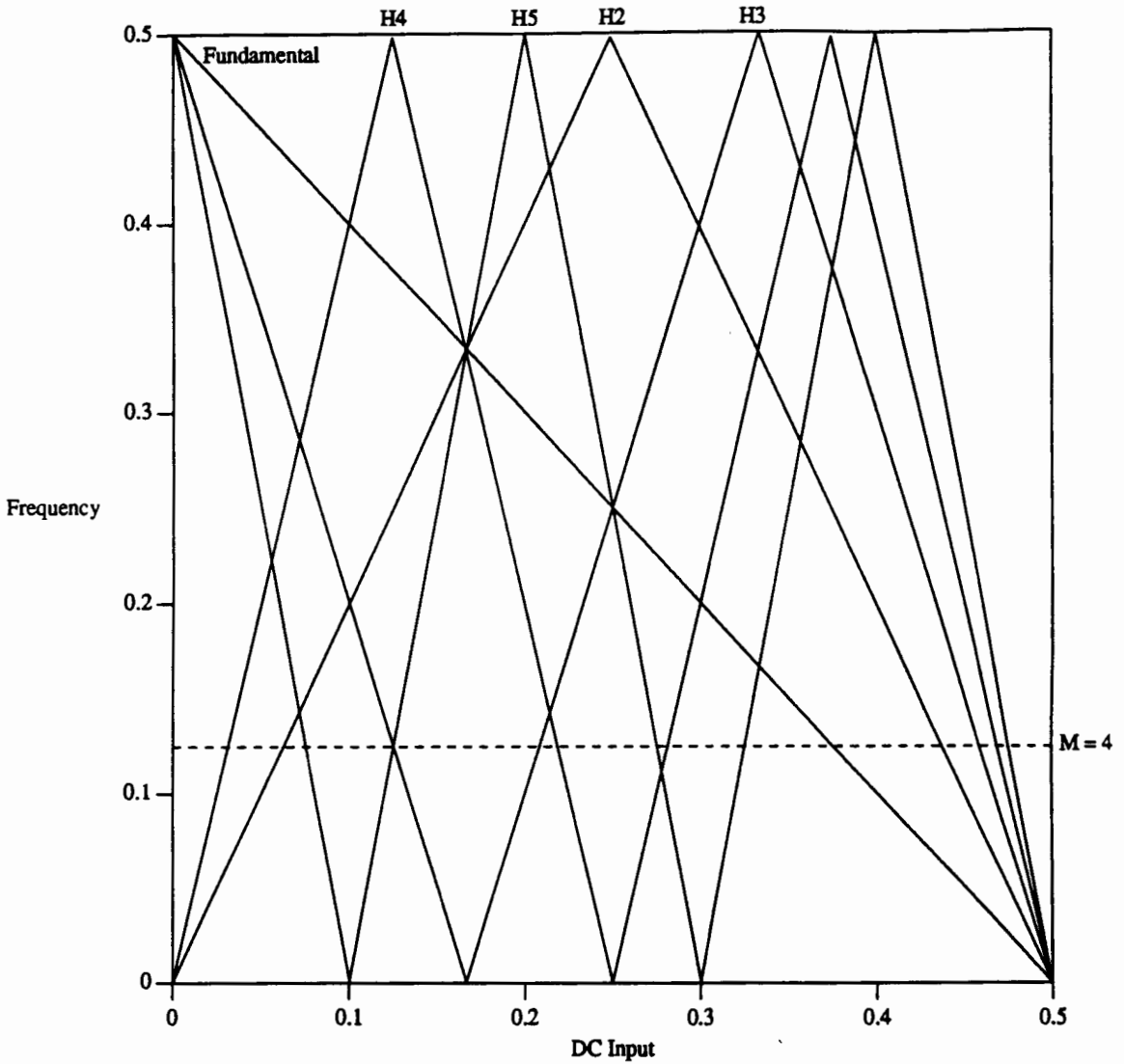


Figure 20. Total noise power vs dc input.

cycle. Since there is no input, all of the output is noise, and the output is a periodic signal with frequency  $\frac{f_s}{2}$ . So, the power spectrum consists of just a spike at  $\frac{f_s}{2}$ . Now, if the dc input value is increased by a small amount, the fundamental noise frequency is predicted to be slightly greater than  $\frac{f_s}{2}$ , which, when aliasing is accounted for, will be slightly less than  $\frac{f_s}{2}$ . Continuing with this line of reasoning will show that the fundamental noise frequency starts at  $\frac{f_s}{2}$  for a zero input, and moves to zero for a full scale dc input. Further thought reveals that all of the harmonics have similar behavior, except that they move faster. The fundamental crosses the sampling bandwidth one time. The second harmonic will cross two times, and so on. So, harmonics of the fundamental appear to bounce back and forth across the sampling bandwidth. Finally, the theory predicts that the even harmonics begin from zero frequency and the odd harmonics begin from  $\frac{f_s}{2}$ . Figure 21 shows the frequency position of the first few harmonics for the range of dc input values.

Figure 22 shows the harmonics from 21, and in addition, it shows the positions of the set of frequencies predicted by the recursive algorithm which described the structure of output sequence of the  $\Sigma\Delta M$  for dc inputs. The fundamental frequency for both exact theories is the same. The sub-frequencies of the structural theory are the average frequencies of the output pattern at each level of recursion. Except for the fundamental, the significance of the sub-frequencies of the structural theory was not determined. Figure 22 does show, however, that they have a definite relationship to the harmonics of the quantization noise.





**Figure 21.**  $\Sigma\Delta M$  noise harmonic positions.

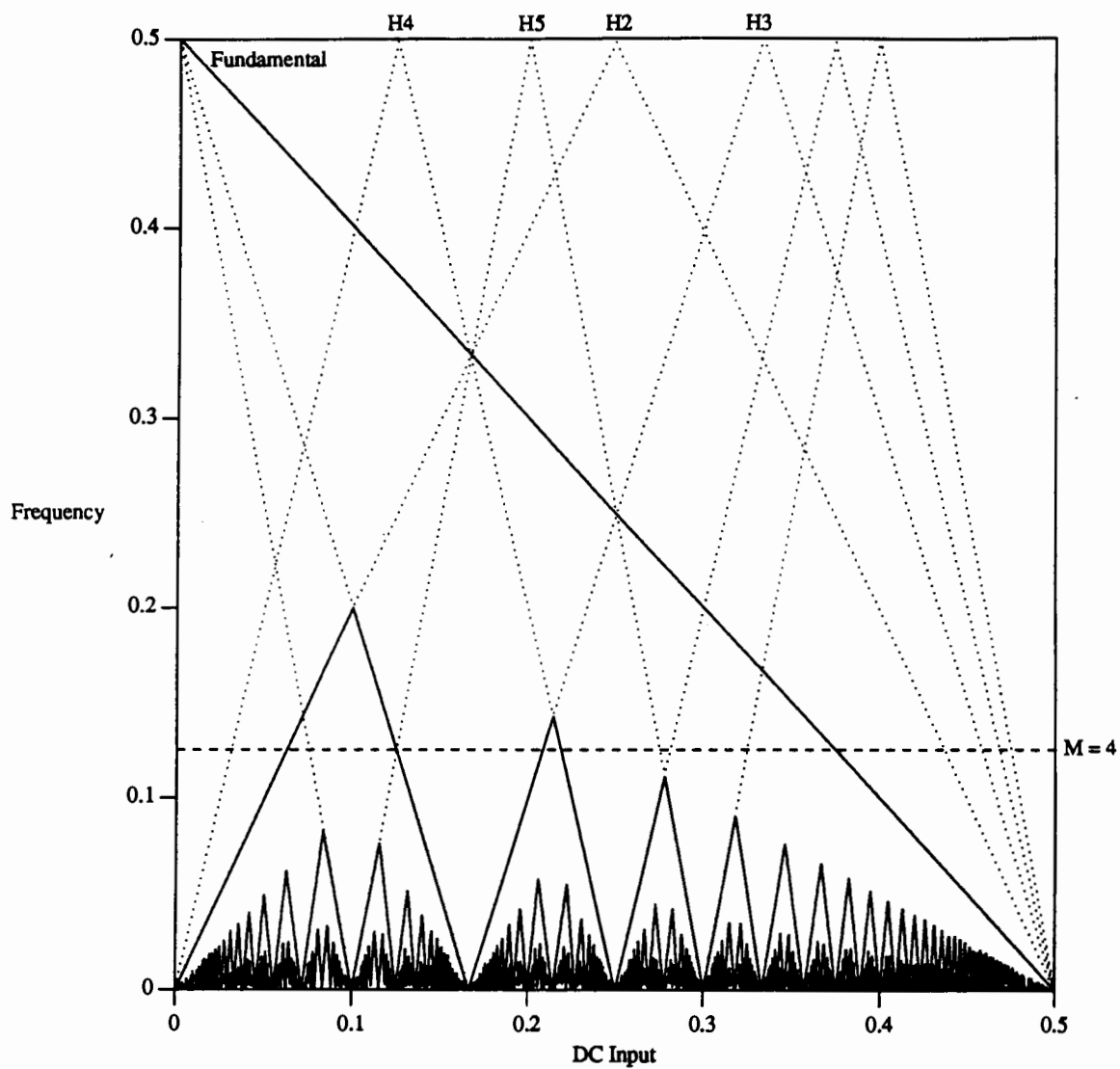


Figure 22.  $\Sigma\Delta M$  structural sub-frequencies.

The power spectrums produced by simulations and theory do in fact demonstrate the behavior just described. Video clips A.3 and A.4 show the spectrums for theory and simulations for a range of dc inputs from zero to full scale. The harmonics bouncing back and forth inside the sampling bandwidth is easy to observe. The similarity in the two clips provides additional verification that the behavior predicted by the theory is correct in simulation.

Understanding the nature of the quantization noise power spectrum of the  $\Sigma\Delta M$  for dc inputs will aid in understanding the total noise power in figure 20. The first thing to note is that any harmonic which does not fall within the signal bandwidth will be filtered out, and will not contribute to the noise power. The next thing to note is that the noise power will likely be higher when the fundamental or low numbered harmonics are located within the signal bandwidth. A qualification to this last point is that the noise frequencies will make their biggest contribution to noise power when they are located at the high end of the signal bandwidth, or  $\frac{f_s}{2}$ . When the harmonic approaches zero frequency, its contribution drops to zero due to the effect of noise shaping.

In figures 20 it is easy to see this in effect, especially for the fundamental noise frequency. The fundamental frequency starts at the high end of the sampling bandwidth for zero input and moves towards zero as the input increases. So, the fundamental frequency will not enter the sampling bandwidth until the input is close to full scale. Since the fundamental contains the greatest noise power, it should be obvious when it enters the signal bandwidth. The point at which the

fundamental enters the signal bandwidth depends on the value of  $M$ . Low values of  $M$  have a wider range of input values which are affected by the fundamental. The range decreases as  $M$  increases.

Thus, the exact theory provides a good way to understand the behavior of the quantization noise for the  $\Sigma\Delta M$  with dc inputs. Knowledge of which harmonics are located in the signal bandwidth for a dc input can provide a good indication of when to expect a higher amount of quantization noise. A good rule of thumb is to expect more noise when the fundamental or low numbered harmonics are inside the signal bandwidth. However, it does not seem to be the case that the highest noise peaks correspond to the order of the noise harmonics. The situation becomes complicated because several higher numbered harmonics in the signal bandwidth could possibly combine to produce greater noise than a lower numbered harmonic. This is possible since the fundamental and second harmonic only contribute to noise power at the ends of the input range. Since the noise power goes by the inverse of the harmonic squared, higher numbered harmonics are closer in power than the low numbered harmonics.

## CHAPTER VIII

### EXACT ANALYSIS OF $\Sigma\Delta M$ WITH SINUSOID INPUTS

#### VIII.1 EXACT ANALYSIS FOR SINUSOIDAL INPUTS

In the case of the exact dc analysis, several characteristics of the quantization noise were found. The characteristics found were the average, power, and autocorrelation. For the general case of irrational dc inputs, it was found that the autocorrelation was a nearly periodic sequence which had a Bohr-Fourier series. The Bohr-Fourier series was interpreted to represent the power spectrum, and simulation results matched this theory very well. The goal in analyzing the noise behavior with sinusoidal inputs is the same. Instead of assuming the characteristics of the quantization noise, the goal is to find an exact solution for the important characteristics.

The exact analysis of the  $\Sigma\Delta M$  for sinusoidal inputs was developed by Gray [5]. The technique for solving the noise characteristics for sinusoidal inputs is similar to the dc input case, although a bit more complicated. To begin, a difference equation that is suitable for solving will be developed for the quantization noise. The input sequence to the quantizer can be written as follows,

$$u_n = x_{n-1} + u_{n-1} - q(u_{n-1}) \quad (\text{VIII.1})$$

The quantization error sequence can be written as

$$e_n = q(u_n) - u_n = x_n - u_{n+1} \quad (\text{VIII.2})$$

As stated above, the desired result is an exact solution for the characteristics of the quantization noise in VIII.2 instead of assuming that the characteristics are like white noise. In a previous chapter, the SQNR performance of the  $\Sigma\Delta M$  has already been predicted for the case where the quantization noise is assumed to be white. The results of the exact analysis here will provide an indication of the accuracy of the white noise assumption.

In order to facilitate the solution, equation VIII.2 is shifted and normalized.

$$z_n = \frac{1}{2} - \frac{e_{n-1}}{2b} = \frac{u_n}{2b} - \frac{1}{2} - \frac{x_{n-1}}{2b}; n = 1, 2, \dots \quad (\text{VIII.3})$$

The following equations show how the various normalized  $\Sigma\Delta M$  sequences can be represented in terms of  $z_n$ . The quantization error sequence.

$$\zeta_n = \frac{e_{n-1}}{2b} = \frac{1}{2} - z_{n+1} \quad (\text{VIII.4})$$

The quantization input sequence.

$$\frac{u_n}{2b} = z_n - \frac{1}{2} + \frac{x_{n-1}}{2b} \quad (\text{VIII.5})$$

The output sequence.

$$q_n = \frac{q(u_n)}{2b} = z_n - z_{n+1} + \frac{x_{n-1}}{2b} \quad (\text{VIII.6})$$

Equation VIII.6 is actually the same as V.2. The total total quantization noise of the  $\Sigma\Delta M$  is the difference between two consecutive quantization noise

values. Equation VIII.5 provides more insight into the operation of the  $\Sigma\Delta M$  by interpreting the quantizer input as the input sequence dithered by the previous quantization error. Dithering is a technique in which a dither signal is added to an input signal with the hope that output performance will be improved. Recalling the simulation results of the low resolution, uniform ADC, the SQNR performance was poor for low input frequencies because the quantization noise signal became very periodic as the sampling frequency was increased. Adding a small, random dither signal to the input would have the effect of breaking up the periodicity of the quantization error signal and result in improved performance. Thus, the  $\Sigma\Delta M$  can now be seen as an ADC which is generating its own dither signal, and the question to be solved is how well does the dither signal work to improve performance.

In preparation for solution, equation VIII.6 can be rewritten in terms of the input sequence.

$$z_1 = 0 \quad (\text{VIII.7})$$

$$z_n = z_{n-1} + \frac{x_{n-2}}{2b} - \frac{q(2by_{n-1} - b + x_{n-2})}{2b}; n = 2, 3, \dots \quad (\text{VIII.8})$$

Equation VIII.8 can be rewritten in a recursive form as follows,

$$z_n = \left\langle z_{n-1} + \frac{1}{2} + \frac{x_{n-2}}{2b} \right\rangle; n = 2, 3, \dots \quad (\text{VIII.9})$$

which can be rewritten as follows,

$$z_n = \left\langle \sum_{k=0}^{n-2} \left( \frac{1}{2} + \frac{x_k}{2b} \right) \right\rangle = \langle s_n \rangle; n = 1, 2, \dots \quad (\text{VIII.10})$$

The important characteristic of equation VIII.10 is that it can be used to find the sequences in equations VIII.4 to VIII.6 for any input sequence  $x_n \in [-b, b]$ . Note that in the case of a dc input,  $x_n = x$ ,

$$z_n = \left\langle (n-1) \left( \frac{1}{2} + \frac{x}{2b} \right) \right\rangle \quad (\text{VIII.11})$$

Equation VIII.11 is identical to VI.29, which was the basis for the exact dc input analysis. So, the development here is the same as before, although it has been written in more general terms.

Now that the difference equations describing the quantization noise of the  $\Sigma\Delta M$  have been defined in a general way, it is time to solve for the characteristics of the quantization noise. Equation VI.32 defined a time average operator, which is an expectation like operator, which can be used to find the desired characteristics. The characteristics desired are the mean, power and sample autocorrelation for the normalized quantization error signal. These can be written as follows.

$$M\{\zeta_n\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \zeta_n = \frac{1}{2} - M\{z_{n+1}\} \quad (\text{VIII.12})$$

$$M\{\zeta_n^2\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \zeta_n^2 = \frac{1}{4} - M\{z_{n+1}\} + M\{z_{n+1}^2\} \quad (\text{VIII.13})$$

$$r_\zeta(k) = M\{\zeta_n \zeta_{n+k}\} = r_z(k) - \frac{1}{4} + M\{\zeta_n\} \quad (\text{VIII.14})$$

Thus, the characteristics of  $\zeta_n$  can be computed if the characteristics of  $z_n$  can be solved.

$z_n$  can be written as  $z_n = g(s_n)$  where  $g(x) = \langle x \rangle$ . The function  $g$  is a



periodic function with period 1, so it has a Fourier series representation,

$$g(x) = \sum_{l=-\infty}^{\infty} \hat{g}(l)e^{2\pi jl x} \quad (\text{VIII.15})$$

The error signal can now be written as

$$z_n = \sum_{l=-\infty}^{\infty} \hat{g}(l)e^{2\pi j l s_n} \quad (\text{VIII.16})$$

Now, take this result and combine it with the sample average operator.

$$M\{z_n\} = M\left\{\sum_{l=-\infty}^{\infty} \hat{g}(l)e^{2\pi j l s_n}\right\} = \sum_{l=-\infty}^{\infty} \hat{g}(l)M\{e^{j2\pi l s_n}\} \quad (\text{VIII.17})$$

The big trick here is the interchange of sum and the sample average operator. This interchange is not valid in every situation, so it needs to be shown to be valid for any particular case.

A simple average characteristic function can be defined as follows

$$\Phi(l) = M\{e^{j2\pi l s_n}\}. \quad (\text{VIII.18})$$

So, the sample average of the noise sequence is

$$M\{z_n\} = \sum_{l=-\infty}^{\infty} \hat{g}(l)\Phi(l) \quad (\text{VIII.19})$$

This technique can be used to develop the rest of the desired characteristics.

The sample power can be written as

$$M\{z_n^2\} = \sum_{l=-\infty}^{\infty} \hat{h}(l)\Phi(l) \quad (\text{VIII.20})$$

where  $\hat{h}(l)$  are the Fourier series coefficients for  $h(x) = \langle x^2 \rangle$ . The sample auto-correlation can be written as

$$r_z(k) = \sum_{i=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \hat{g}(i)\hat{g}(l)\Phi_k(i, l) \quad (\text{VIII.21})$$

where the sample joint characteristic function is defined as

$$\Phi_k(i, l) = M\{e^{j2\pi(is_n + lsn + k)a}\} \quad (\text{VIII.22})$$

For all of these characteristics, the infinite sums of the Fourier series expansions of the functions  $g$  and  $h$  have been interchanged with the time average operator  $M$ . This interchange can be shown to be valid under certain conditions. For the case of a sinusoidal input sequence, the interchange is valid.

The previous equations show the desired noise characteristics in general terms for a set of situations where the required interchanges can be shown to be valid. In order to obtain exact solutions for the characteristics given a sinusoidal input, the interchange must be shown valid and the equations must be solved.

For an input sequence of the form

$$x_n = a \cos n\omega \quad (\text{VIII.23})$$

it can be shown that the interchange is valid. All that remains to do is to solve the equations. For the input sequence of equation VIII.23, it can be shown that

$$s_n = \frac{n-1}{2} + \frac{a}{4b} + \alpha \sin\left(n\omega - 3\frac{\omega}{2}\right) \quad (\text{VIII.24})$$

where

$$\alpha = \frac{a}{4b \sin \frac{\omega}{2}}$$

. This can be plugged into the characteristic functions in the equations above to solve for the quantization noise characteristics. Skipping over some of the intermediate steps, the final results follow.

The sample average for a sinusoidal input is

$$M\{\zeta_n\} = \frac{1}{2\pi} \sum_{l=1}^{\infty} \frac{1}{l} J_0(4\pi l\alpha) \sin \pi l \frac{a}{b} \quad (\text{VIII.25})$$

The sample power, or second moment, is

$$M\{\zeta_n^2\} = \frac{1}{12} - \frac{1}{4\pi^2} \sum_{l=1}^{\infty} \frac{1}{l^2} J_0(4\pi l\alpha) \cos \pi l \frac{a}{b} \quad (\text{VIII.26})$$

The autocorrelation is shown to be

$$r_z(k) = \sum_{m=-\infty}^{m=\infty} e^{jk m \omega} (-1)^m c_e(m)^2 + \sum_{m=-\infty}^{m=\infty} e^{jk(m\omega - \pi)} (-1)^m c_o(m)^2 \quad (\text{VIII.27})$$

where

$$c_e(m) = \begin{cases} \frac{1}{2} - \frac{1}{\pi} \sum_{l=1}^{\infty} \frac{J_0(4\pi\alpha l)}{2l} \sin\left(\pi l \frac{a}{b}\right); & m = 0 \\ -\frac{1}{\pi} \sum_{l=1}^{\infty} \frac{J_m(4\pi\alpha l)}{2l} \sin\left(\pi l \frac{a}{b}\right); & m \text{ even} \\ \frac{j}{\pi} \sum_{l=1}^{\infty} \frac{J_m(4\pi\alpha l)}{2l} \cos\left(\pi l \frac{a}{b}\right); & m \text{ odd} \end{cases} \quad (\text{VIII.28})$$

$$c_o(m) = \begin{cases} -\frac{1}{\pi} \sum_{l=1}^{\infty} \frac{J_m(4\pi\alpha(2l-1))}{2l-1} \sin\left(\pi(2l-1) \frac{a}{2b}\right); & m \text{ even} \\ \frac{j}{\pi} \sum_{l=1}^{\infty} \frac{J_m(2\pi\alpha(2l-1))}{2l-1} \cos\left(\pi(2l-1) \frac{a}{2b}\right); & m \text{ odd} \end{cases} \quad (\text{VIII.29})$$

Equations VIII.27 - VIII.29 represent the autocorrelation of the output sequence  $z_n$ . Gray notes that the important fact about these equations is that they can be written in the form

$$r_z(k) = \sum_{l=-\infty}^{\infty} s_l e^{j2\pi k \lambda_l} \quad (\text{VIII.30})$$

Equation VIII.30 defines the Bohr-Fourier series of the sequence of  $r_z(k)$ . The frequencies of the spectrum are represented by  $\lambda_l$ , which is normalized in  $[0, 1)$ , and the amplitudes are represented by  $s_l$ . It is useful to consider the indexes  $l$

in VIII.30 to have the form  $l = (m, i); m = \dots, -1, 0, 1, \dots, i = 1, 2$ . Then the following can be written.

$$s_{(m,1)} = (-1)^m c_e(m)^2 \quad (\text{VIII.31})$$

$$s_{(m,2)} = (-1)^m c_o(m)^2 \quad (\text{VIII.32})$$

$$\lambda_{(m,1)} = \left\langle m \frac{\omega}{2\pi} \right\rangle \quad (\text{VIII.33})$$

$$\lambda_{(m,2)} = \left\langle m \frac{\omega}{2\pi} - \frac{1}{2} \right\rangle \quad (\text{VIII.34})$$

These are the results which comprise the exact analysis for sinusoidal inputs to the  $\Sigma\Delta M$ .

If it is assumed that  $a = b$ , then the results can be simplified. This represents the case of a full scale sinusoid input. The autocorrelation can be represented as

$$r_z(k) = \sum_{m=-\infty}^{\infty} s_m e^{j2\pi k \lambda_m} \quad (\text{VIII.35})$$

where

$$s_m = \begin{cases} \frac{1}{2}; & m = 0 \\ \left( \frac{1}{\pi} \sum_{l=1}^{\infty} \frac{J_m(2\pi\alpha(2l-1))}{2l-1} (-1)^l \right)^2; & m \text{ even} \\ \left( \frac{1}{\pi} \sum_{l=1}^{\infty} \frac{J_m(4\pi\alpha l)}{2l} (-1)^l \right)^2; & m \text{ odd} \end{cases} \quad (\text{VIII.36})$$

and

$$\lambda_m = \begin{cases} \left\langle m \frac{\omega}{2\pi} - \frac{1}{2} \right\rangle; & m \text{ even} \\ \left\langle m \frac{\omega}{2\pi} \right\rangle; & m \text{ odd} \end{cases} \quad (\text{VIII.37})$$

These formulas for the full scale sinusoid will be the starting point for checking simulation results against the theory.

## CHAPTER IX

### SIMULATION OF THE $\Sigma\Delta M$ WITH SINUSOID INPUTS

Two theories describing the behavior of the  $\Sigma\Delta M$  with sinusoidal inputs have been presented. The first theory makes the assumption that the quantization noise of the quantizer is white. The result is a simple prediction, equation V.16, that the SQNR will increase by 9 db as  $M$  is doubled. It was observed that the assumptions used to develop this theory were carried over from analysis of uniform ADC's under conditions where it is reasonable to assume that the quantization noise was white. As the second theory, which performs an exact analysis of the  $\Sigma\Delta M$  quantization error sequence predicts, the spectrum of the quantization noise is far from white. Both theories will be compared against the simulation results. The white noise theory predicts the SQNR and can be compared directly with simulation SQNR results. The exact theory predicts the spectrum of the quantization noise. Since it is computationally expensive to compute this theoretical spectrum, due to the infinite sums of Bessel functions, it is not practical to use the exact theory to compute SQNR values. Thus, the use of the exact theory will center around how the predicted error spectrum matches the actual spectrum for a couple examples. The primary usefulness of the exact theory will be to use its formulas to explain some of the features that will be found in the simulated results.

## IX.1 SETUP OF SINUSOIDAL SIMULATIONS

Simulations were performed using a sinusoidal input to the  $\Sigma\Delta M$ . In order to get a complete picture of the behavior of the  $\Sigma\Delta M$  with a sinusoidal input, simulations were performed for oversampling ratios of 1 to 128, for the complete range of amplitudes and frequencies in the signal bandwidth. The 102 amplitudes used are evenly spaced in the range (0.0, 0.499502]. Zero amplitude was not used. These amplitudes are the same values used for the dc input simulations. The 103rd amplitude is the full scale amplitude of 0.5. To cover the signal bandwidth, 241 frequencies were used based on the following expression.

$$f_x = 74 + 99i; i : 1, 2, 3, \dots, 241$$

To speed up the simulation process, a C program was used to simulate the  $\Sigma\Delta M$  instead of using MIDAS. As in the case of the uniform ADC, increasing numbers of samples were generated as  $M$  increased so that the same number of FFT bins will represent the signal bandwidth. For the most part, the following discussion of the results will focus on the results for  $M = 128$ . The two results of interest for the simulations will be the spectrum of the output signal and the SQNR. The SQNR was computed for all the cases mentioned above. Obviously, due to the method used to calculate the experimental SQNR, the spectrum is computed for every case, but the SQNR is a much more compact result than the spectrum. So, the actual output spectrum will be examined for a couple examples to illustrate the drawback of using only the SQNR value as a measure of the  $\Sigma\Delta M$

performance.

## IX.2 RESULTS OF SINUSOIDAL SIMULATIONS

A large number of simulations were performed on the  $\Sigma\Delta M$  with sinusoidal inputs. To begin examining the results, a couple cases which show SQNR results over the range of oversampling ratios will be discussed. Figure 23 shows results for a few input frequencies and for full scale amplitude. Figure 24 shows results for a few input frequencies and for an amplitude about 80% of full scale. Both of these figures include the theoretical SQNR curve as predicted by the white noise theory. The first thing to notice is that these results are not too bad. The results of the uniform ADC had serious problems matching the theoretical curve over the signal bandwidth, and they varied widely over the signal bandwidth. Here, the SQNR curves fall much closer together over the signal bandwidth. The results also increase fairly linearly, as the theory predicts, although the slope of increase is a little bit smaller than the theory. One unusual point is that the results for the 80% of full scale amplitude actually have a better SQNR performance than the full scale curves.

At first these results might seem to indicate that the white noise theory works well for the  $\Sigma\Delta M$ . However, there are a few points to consider. The white noise theory is based on the assumption that the quantization noise is white. The exact theory predicts that the  $\Sigma\Delta M$  has a very non-white, discrete quantization noise spectrum. It is important to remember that the SQNR value does not provide

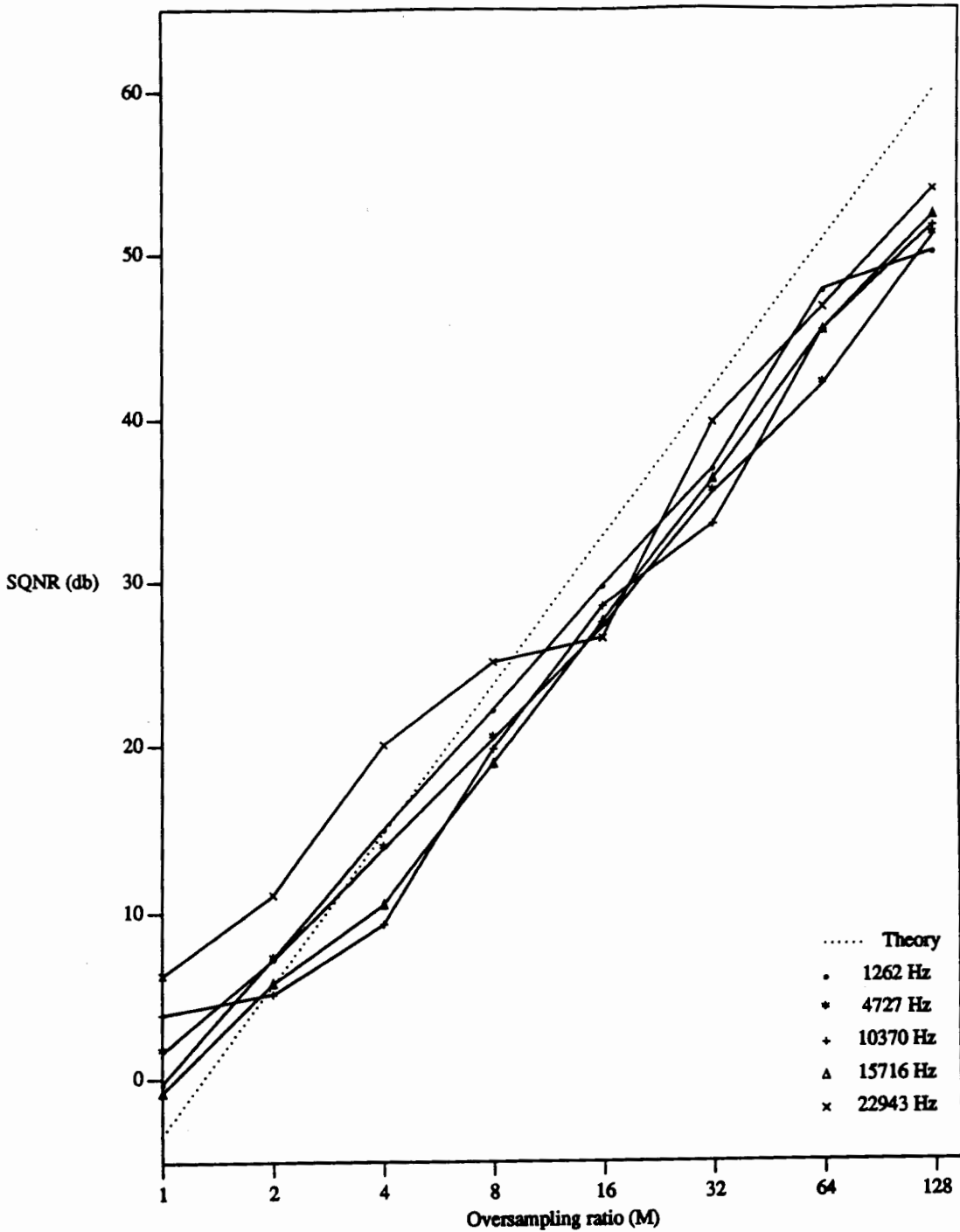


Figure 23. SQNR curves for full scale sine input.



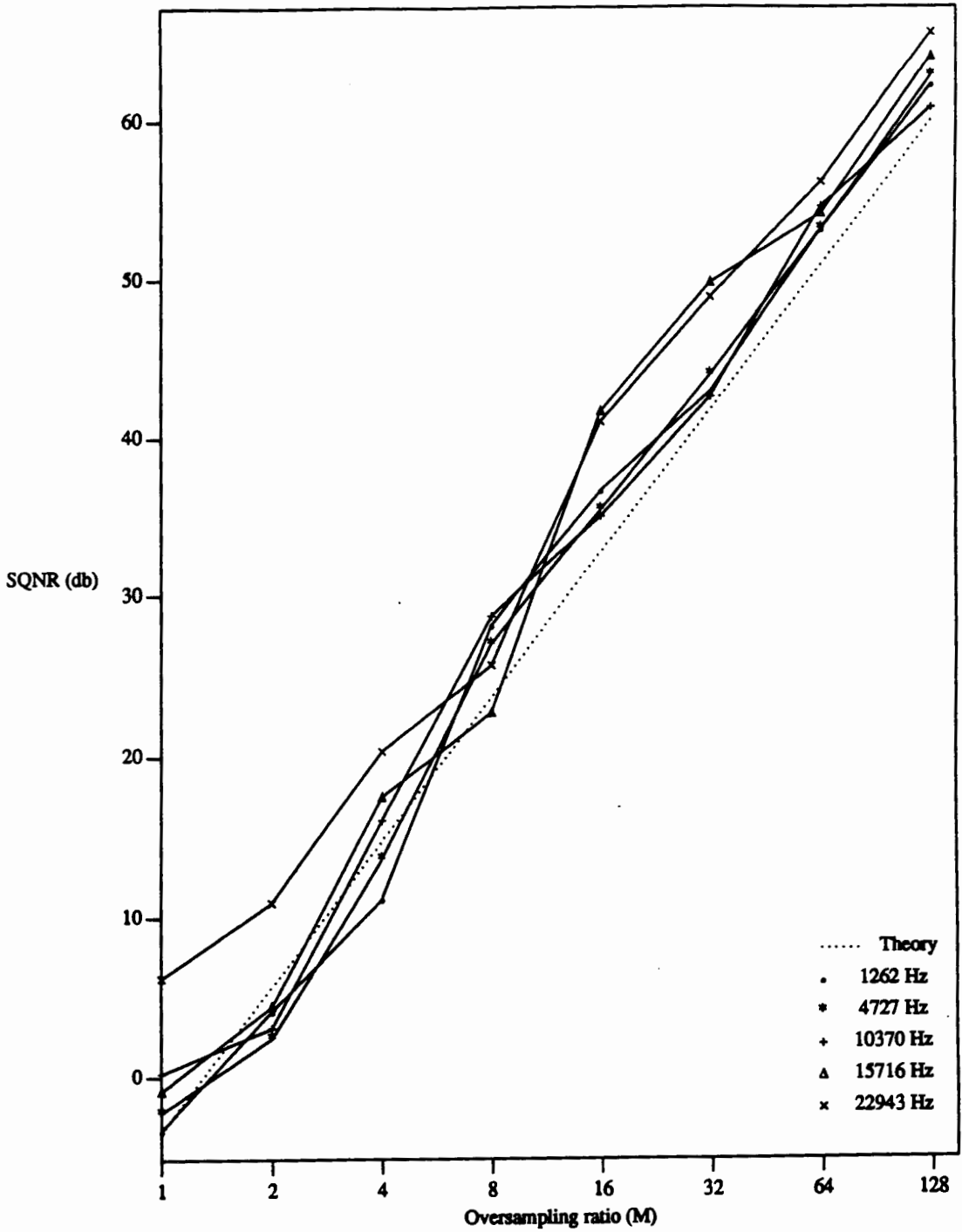


Figure 24. SQNR curves for 80 % full scale sine input.

any information about the spectrum. The same SQNR value could be obtained from a white, evenly distributed noise spectrum, or from a spectrum where all the noise energy is located in a single spike. This kind of behavior was observed in the uniform ADC with high oversampling ratios for frequencies under  $\frac{f_n}{6}$ . Under those conditions, the noise spectrum was dominated by noise spikes on the odd harmonics of the input frequency.

Since the exact theory predicts a discrete noise spectrum, it is a good idea to examine the simulation results more closely. Video clip A.5 fill in the details for figure 23 by showing the SQNR curves for the whole signal bandwidth. There is some interesting behavior in these curves. For low frequencies, they remain relatively stable, but when the frequency gets higher, the curves start to jump around a bit, mostly to higher SQNR values. Close observation shows that the curves begin jumping around after about 8000 Hz, which is  $\frac{f_n}{6}$ . This is reminiscent of the uniform ADC's harmonics.

Figure 25 shows the complete set of SQNR results for  $M = 128$ . The results are plotted in the form of a surface with signal amplitude and frequency being the two horizontal axes, and SQNR representing the height. [See A.6] This surface clearly shows that interesting behavior is occurring in the error spectrum. The surface shows fairly regular behavior for low frequencies. At  $\frac{f_n}{6}$ , there is a sudden shift, or ridge, that occurs for all amplitudes. From  $\frac{f_n}{6}$  to  $\frac{f_n}{4}$  the surface becomes higher and rougher. After  $\frac{f_n}{4}$ , there is another clear shift and the surface becomes even higher and rougher. Throughout the high frequency side of the surface, there

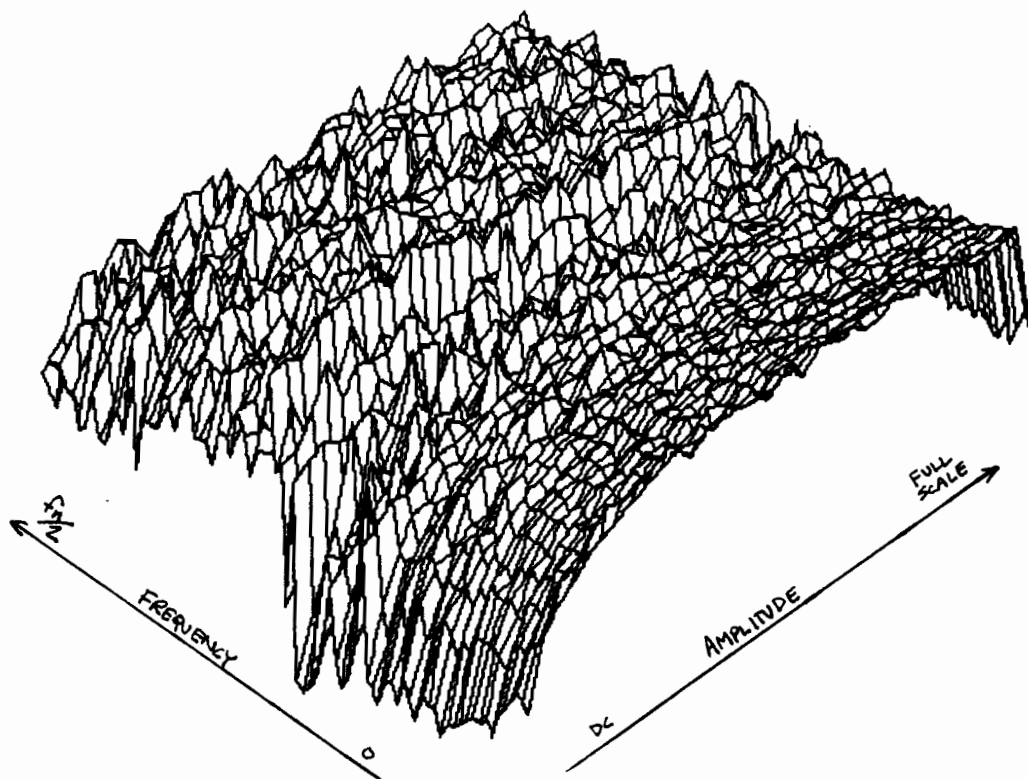


Figure 25. SQNR surface for  $M = 128$ .

are clear ridges that occur at certain frequencies. Another obvious feature of the surface is the significant drop in SQNR as the input signal reaches full amplitude. This drop occurs across the whole signal bandwidth.

The SQNR surface illustrates that the SQNR and the white noise theory alone cannot explain the characteristics of the quantization noise in  $\Sigma\Delta M$  with sinusoidal inputs.

### IX.3 SIMULATIONS AND THE EXACT THEORY

Now it is time to bring the exact theory into the discussion and see if it can provide insight into understanding the results. The SQNR alone does not give

any information about the structure of the noise spectrum, since it is basically an average. The exact theory predicts the nature of the spectrum.

First, it is necessary to discuss how the theoretical results were computed and how they will be used. The exact theory for sinusoid inputs is related to the theory of dc inputs, but it is significantly more complicated. The noise power for a harmonic in the dc theory was a simple result based on the number of the harmonic. For the sinusoid case, the power for each term is an infinite sum of bessel functions multiplied by a sinusoid. It is difficult to obtain any intuitive feeling for how this theory works. Also, it is much more computationally expensive to compute the noise spectrum since bessel functions take a long time to evaluate when the bessel number becomes high.

Thus, the approach taken with the exact theory for sinusoids is to compute a few sample results to verify that the theory works since it would take too long to compute a complete set of theoretical results. After showing that the theory works, the primary function of the theory will be to explain some of the behavior and characteristics which were observed in the simulation results. Taken together, the simulations and theory will be able to shed some light on each other. That is, the theory can be used to explain some of the simulation results, and the simulation results can be used to identify the important parts of the theory.

It was a relatively challenging job to get the theory to match the simulation results. One factor is the need to compute the sums of bessel functions. The decision was made to truncate the summation. Typically several hundred terms

were computed. Another factor involved in computing the theory was determining which frequencies would contribute to the quantization noise. Since it was so expensive to compute the spectrum, only the part of the spectrum that will appear in the signal bandwidth was computed so as not to waste time working on parts of the spectrum that get filtered out. The frequencies work much the same as the dc case. One difference is that the fundamental frequency of the noise is the same as the input sinusoid. This means that the fundamental of the noise ranges only within the signal bandwidth instead of the sampling bandwidth in the dc case. In order to find which harmonics fall in the signal bandwidth, the values of  $m$  for which the frequencies defined in equations VIII.33 and VIII.34 are less than  $\frac{1}{2M}$  must be found. Some thought will reveal that for mid to high range input frequencies, there are long gaps in which the harmonics are outside the signal bandwidth.

Once the correct harmonics are determined and computed, a spectrum for the quantizer noise has been created. The goal is to match this theoretical spectrum with the spectrum from simulation. In theory, it should be possible to match the theoretical noise spectrum to the simulated spectrum with the exception of the fundamental frequency. In order to do this, the theoretical spectrum must be shaped by the noise shaping function. When this is done correctly, it is found that the theoretical spectrum does in fact match the simulated results quite well. Figures 26 – 29 show the spectrums for some sample cases. In general, the theory was able to pick out the dominant harmonics. In some cases, it looks as if some of the harmonics are missing. Since the dominant harmonics are predicted the lack

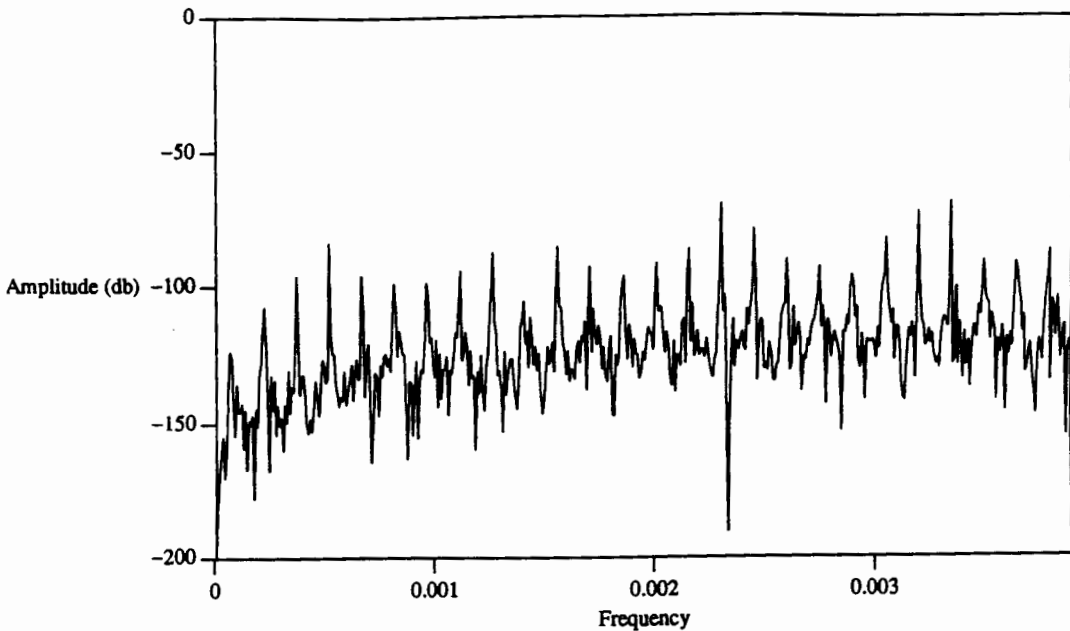


Figure 26. Theoretical spectrum  $f_x = 8687 Hz$ .

of some of the harmonics is not significant. A theory for why some harmonics do not show up is that computational error dominated the computation for that term. Anyway, the theory checks out against the simulated results, so now the theory can be examined to see if it can predict some of the characteristics observed in the simulation results.

One of the most obvious features is how the SQNR surface appears to have bands of activity in it depending on frequency. Low frequencies are fairly stable and high frequencies have much higher SQNR values. Some clues to this can be obtained by looking more closely at the harmonic structure of the signal bandwidth. Equation VIII.33 predicts that all harmonics of the input frequency will appear in the noise spectrum, and equation VIII.34 predicts that all harmonics shifted by  $\pi$  will appear in the noise spectrum. The frequencies for VIII.34 can be disregarded

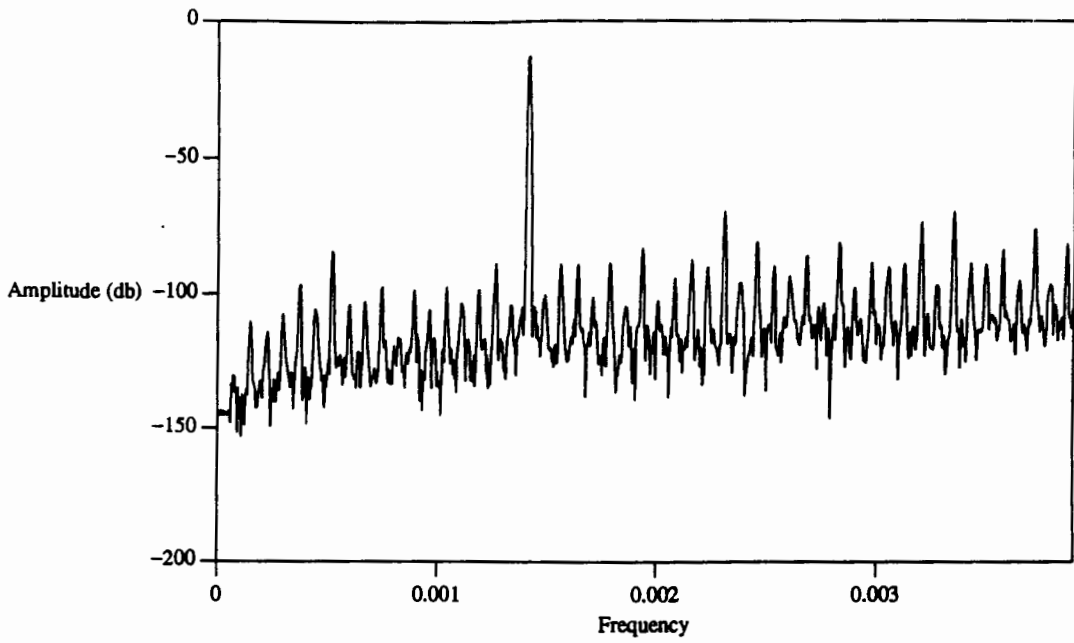


Figure 27. Simulated spectrum  $f_x = 8687 Hz$ .

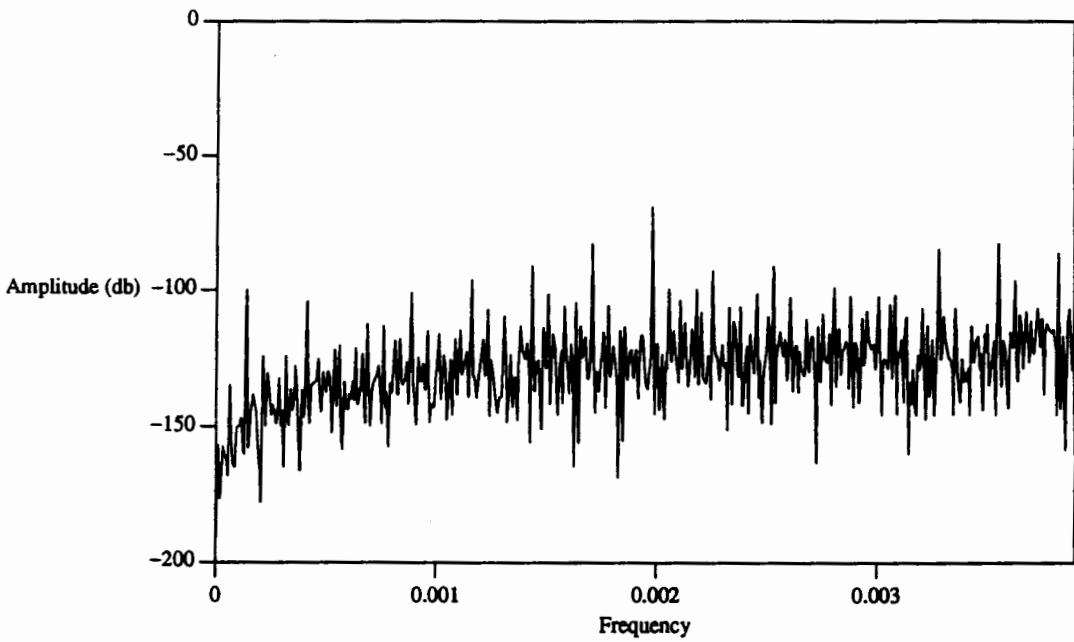


Figure 28. Theoretical spectrum  $f_x = 23537 Hz$ .

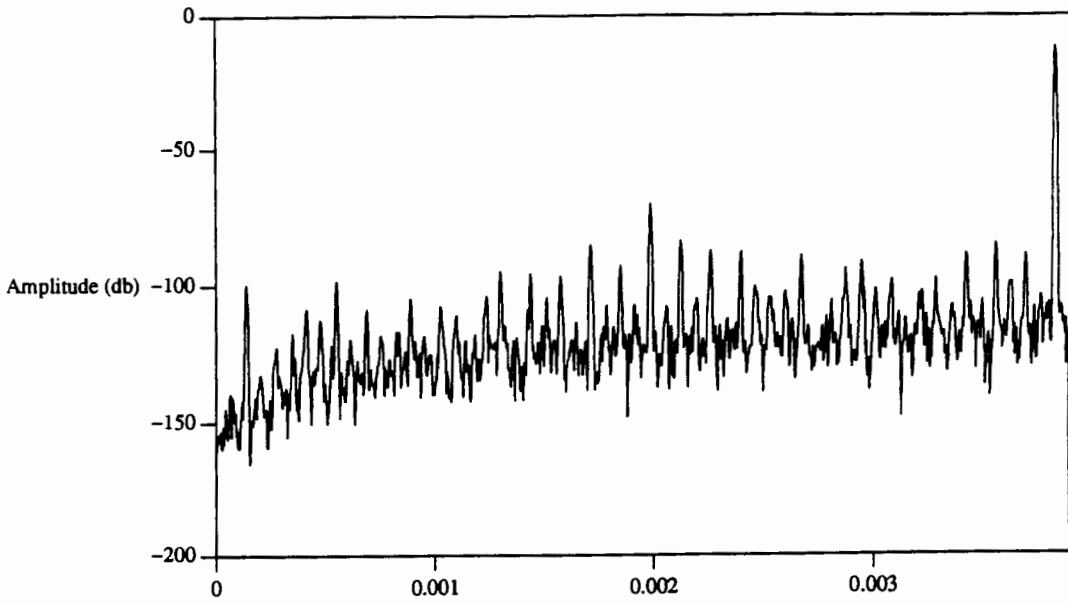


Figure 29. Simulated spectrum  $f_x = 23537 Hz$ .

right away since the shift by  $\pi$  moves them far away from the signal bandwidth. Figure 30 shows a partial picture of which harmonics contribute to the signal bandwidth for  $M = 128$ . The end result is that for low frequencies, harmonics of the frequency will appear in the signal bandwidth. Although the sums of Bessel functions are not highly intuitive, it seems reasonable to make a general assumption that the power in low numbered harmonics will be higher than the power in high numbered harmonics. The second and third harmonics will be present in the signal bandwidth for lower frequencies, so it could be assumed that these frequencies will show higher noise power.

The feature present in the SQNR surface does show that it is likely that the second and third harmonics are dominant noise frequencies. This is concluded from the fact that significant changes are observed in the SQNR surface as  $\frac{f_n}{6}$  and



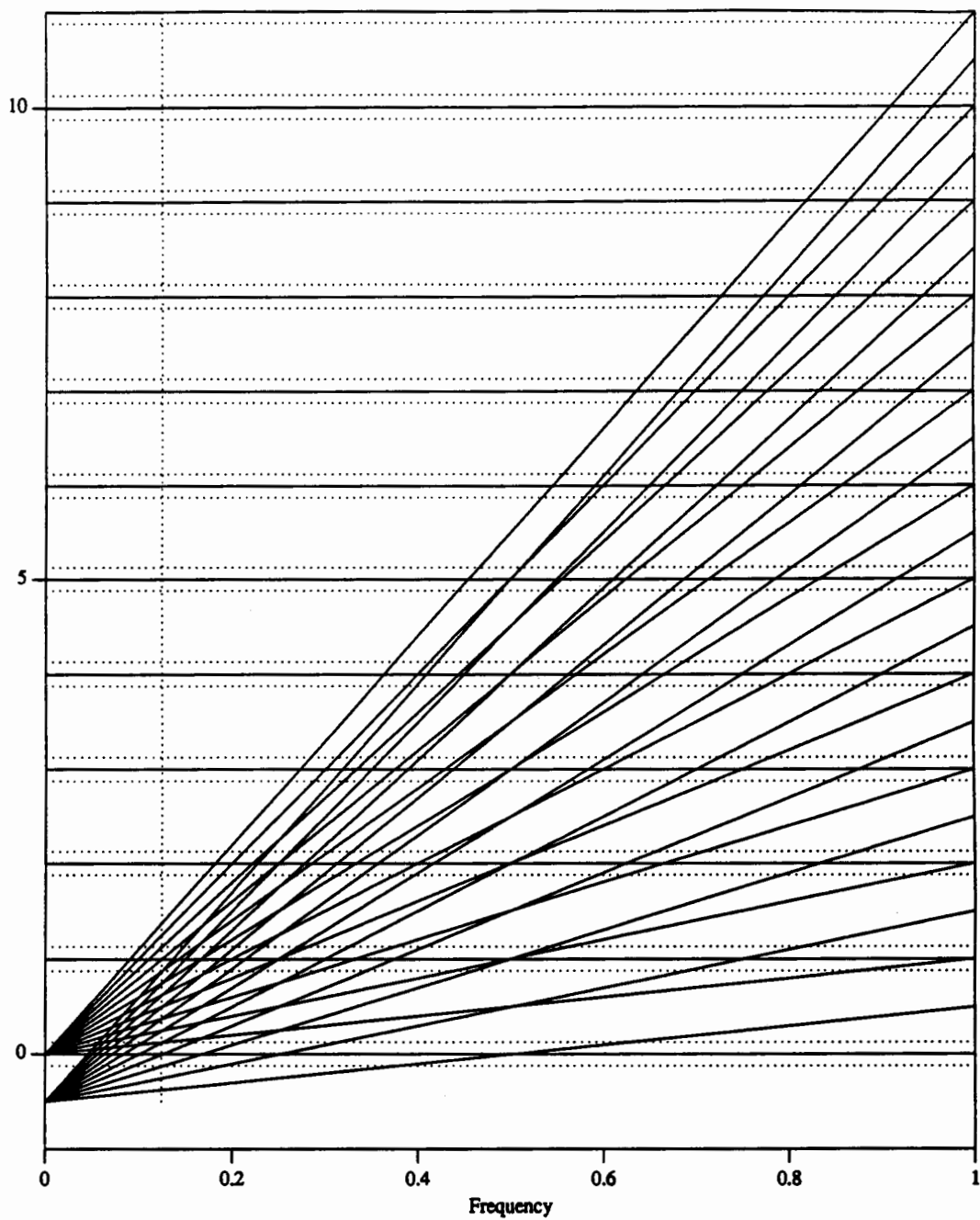


Figure 30. Harmonics located in the signal bandwidth.

$\frac{f_n}{4}$  are reached. As these two harmonics pass out of the signal bandwidth, the SQNR jumps up to higher values. Since the SQNR is fairly constant below  $\frac{f_n}{6}$ , it is hypothesized that these first two harmonics dominate over the others. This is similar to the uniform ADC, where the SQNR jumped significantly when the the third harmonic passed out of the signal bandwidth. An important difference in this case is that the harmonics are not the same for different values of  $M$ . The SQNR still rises with  $M$  for low frequencies, unlike the uniform ADC case. Video clips A.7 A.8 shows the spectrum in the signal bandwidth for the range of input frequencies. A close observation will show that there are some higher valued harmonics for the lower end of the signal bandwidth. After the input frequency passes the halfway point, the noise appears to drop somewhat, which matches with the results shown in the SQNR surface.

A feature in the SQNR surface which is more difficult to explain are the ridges of high SQNR values which appear along some of the higher frequencies for all values of amplitude. There are two parts of the theory that depend on frequency. The input frequency determines which harmonics will be present in the signal bandwidth, and the bessels functions have a term which includes the frequency. A general assumption was made above that the harmonic power values decrease. In the dc case this was true. However, it is not obvious for the sinusoid case. The ridges occur in the high frequency side of the signal bandwidth, so it is known from above that only high numbered harmonics which will appear. Figure 31 shows the

first harmonics that appear in the signal bandwidth for a few frequencies. The important observation here is that in some cases, higher harmonics have greater power than lower harmonics. Thus, the frequency interacts in a complex way with the sums of bessels to produce the noise power. No method for explaining a ridge has been developed except to say that the frequencies must have been just right.

Finally, the other significant feature of the SQNR output is the drop in SQNR as the amplitude of the sinusoid reaches full scale. This implies that the total noise energy increases as the amplitude reaches full scale. It is difficult to look at the equations for the exact theory and explain why this occurs. The amplitude affects both the bessel function and the sinusoid factor. Determining the behavior of the sum of the product of these two factors in terms of the amplitude would be a complex and difficult task. The drop in SQNR does not seem to be related to frequency since it occurs across the frequency range. The sinusoid factor actually simplifies when the amplitude reaches full scale and half of the terms drop out for both sets of frequency harmonics. Still, there is a significant drop in SQNR. The explanation for this drop will be based on the equations for the frequency positions of the noise signal components, some intuitive reasoning based on the operation of the  $\Sigma\Delta M$ , and the results from the uniform ADC.

The increase in noise power for sinusoidal inputs as the amplitude reaches full scale is reminiscent of the increase in noise power that occurred for dc inputs when the input approached full scale. However, there are some major differences. For dc inputs, the increase in noise was caused by the fundamental frequency of

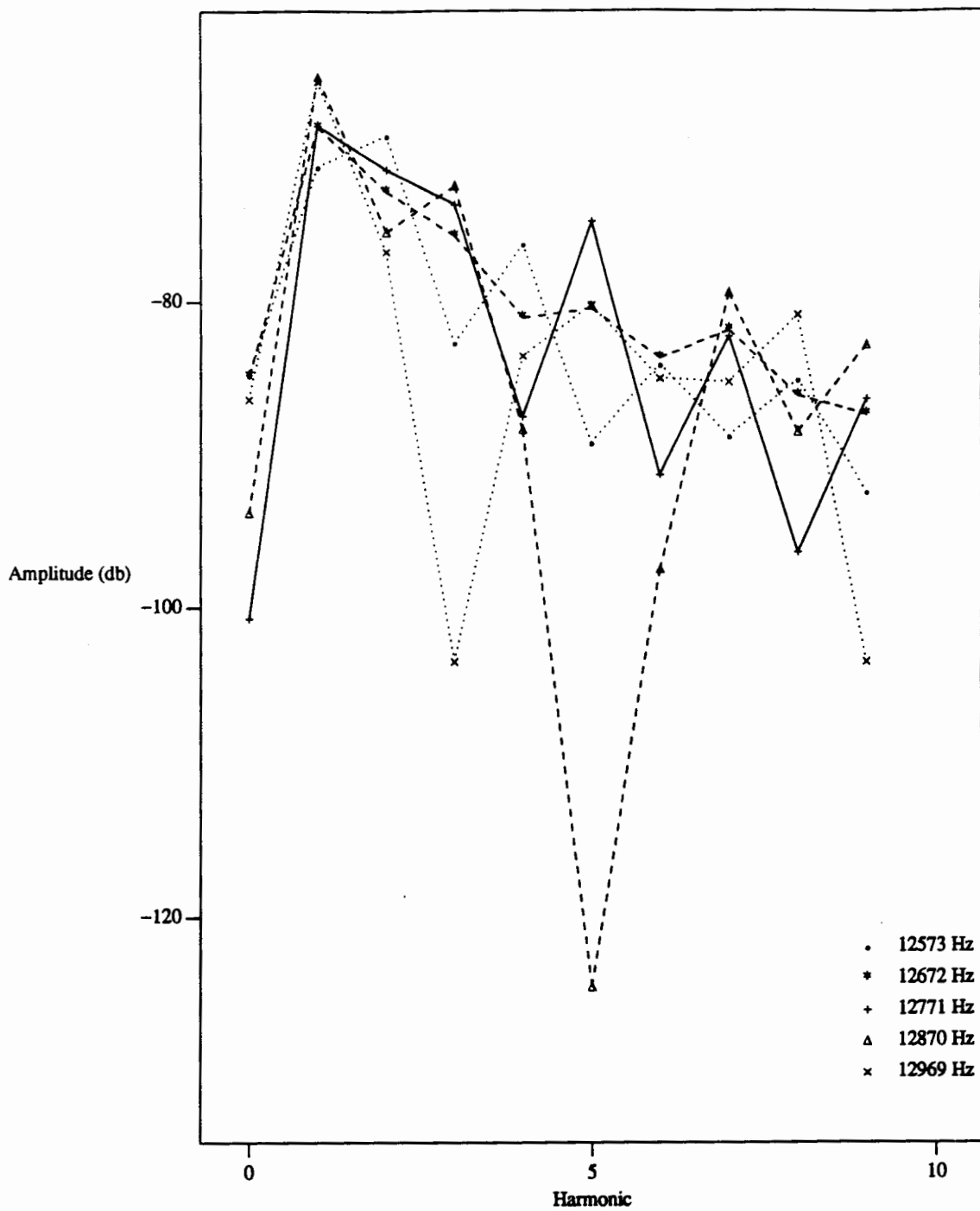


Figure 31. Amplitudes of first harmonics in the signal bandwidth.

the noise signal entering the signal bandwidth shortly before the dc input reaches full scale. As the dc input reaches full scale, the fundamental frequency of the noise reaches zero frequency. In fact, all of the harmonics of the dc signal approach zero frequency as the dc input nears full scale. As they move down through the signal bandwidth, they are reduced further by noise shaping. So, at exactly full scale, the noise actually disappears. On the other hand, the noise signal appears to be greatest at full scale amplitude for a sinusoidal input. For sinusoidal inputs, all of the harmonics are not in the signal bandwidth as the amplitude reaches full scale. The fundamental frequency of the noise is the only major component of the noise signal that remains in the signal bandwidth over the length of the signal bandwidth.

From the discussion of the operation of a  $\Sigma\Delta M$ , it is known that as the input value approaches one of the full scale limits, the output will tend to oscillate less and spend more time at the value close to the input. In the extreme case of a dc input equal to full scale, the output does not oscillate at all. So, for a sinusoid that has an amplitude which approaches full scale, the output value will tend to stay constant when the sinusoid is at a peak. Since the peaks of the sinusoid change slowly, it seems reasonable to assume that the output stops modulating for a significant period of time. When this occurs, the error signal will tend to develop a periodic segment. The period is the fundamental. This is similar to the noise structure of the uniform ADC, except that in this case, only a periodic segment of the noise has this structure. Also, from the uniform ADC discussion, it is known

that the harmonics of the noise signal move out of the signal bandwidth by  $\frac{f_n}{6}$ , so the fundamental must be the factor in the SQNR reduction.

This makes sense because the periodic segment of the noise has a fundamental frequency equal to the input frequency. From before, it was shown that the fundamental of the noise signal will remain in the signal bandwidth over the entire range. Since the error signal will have a sign opposite to the sign of the sinusoid peak, it seems clear that the noise power will act to reduce the power of the signal frequency. This is consistent with the results. Thus, a sinusoid with a full scale amplitude will experience a drop in gain.

An analytical study can be made here to validate the arguments being made. The claim is that the output of the  $\Sigma\Delta M$  will stop oscillating for a while when the sinusoid is near a peak. The goal of the analysis will be to obtain some measure of the length of time that the oscillation stops. The expected result is that the length of no oscillation will rise significantly as the amplitude approaches full scale. A couple assumptions are made for this analysis. The first is that the period of no oscillation is centered about the peak of the sinusoid. The other assumption is that at the beginning of the no oscillation period,

$$u_n = x_{n-1} - q(u_{n-1}) + u_{n-1} = x_{n-1} - q(u_{n-1}).$$

The output will be changing to the no oscillation value for  $u_n$ , so  $u_{n-1}$  will have an opposite sign from  $u_n$ . The assumption being made is that  $u_{n-1}$  is very close to zero. The result of this assumption is that maximum possible length of the no

oscillation period will be found by the analysis. This is acceptable since the goal is to obtain a general measure or feel for what is going on. The measure is obtained by calculating how long it takes for  $u_n$  to change sign again. This will occur when the sum of all the  $x_{n-1} - q(u_{n-1})$  terms, which will be opposite in sign from  $u_n$ , to become greater than the initial value of  $u_n$ .

Assuming that the input is of the form  $x_n = a \cos 2\pi n f$  and that the length of no oscillation is  $2N$ , then the relation that represents the analysis can be written as

$$a \cos \omega N + b \geq 2 \sum_{n=0}^N (b - a \cos \omega n) - b + a \quad (\text{IX.1})$$

By finding the largest  $N$  for which this equation is true for different values of  $a$ , a measure of the length of the no oscillation period can be obtained. Equation IX.1 was solved by computer and results were obtained for a number of frequencies as  $a$  approaches full scale.

Figure 32 shows the results. The value of  $N$  clearly rises as  $a$  approaches full scale. An interesting feature is that the rise is much greater for lower frequencies. The overall period would seem to be longer, however, noise shaping will attenuate the power in the fundamental at lower frequencies, so the periodicity of the low frequencies should be stronger and longer so that the power in the fundamental will be greater.

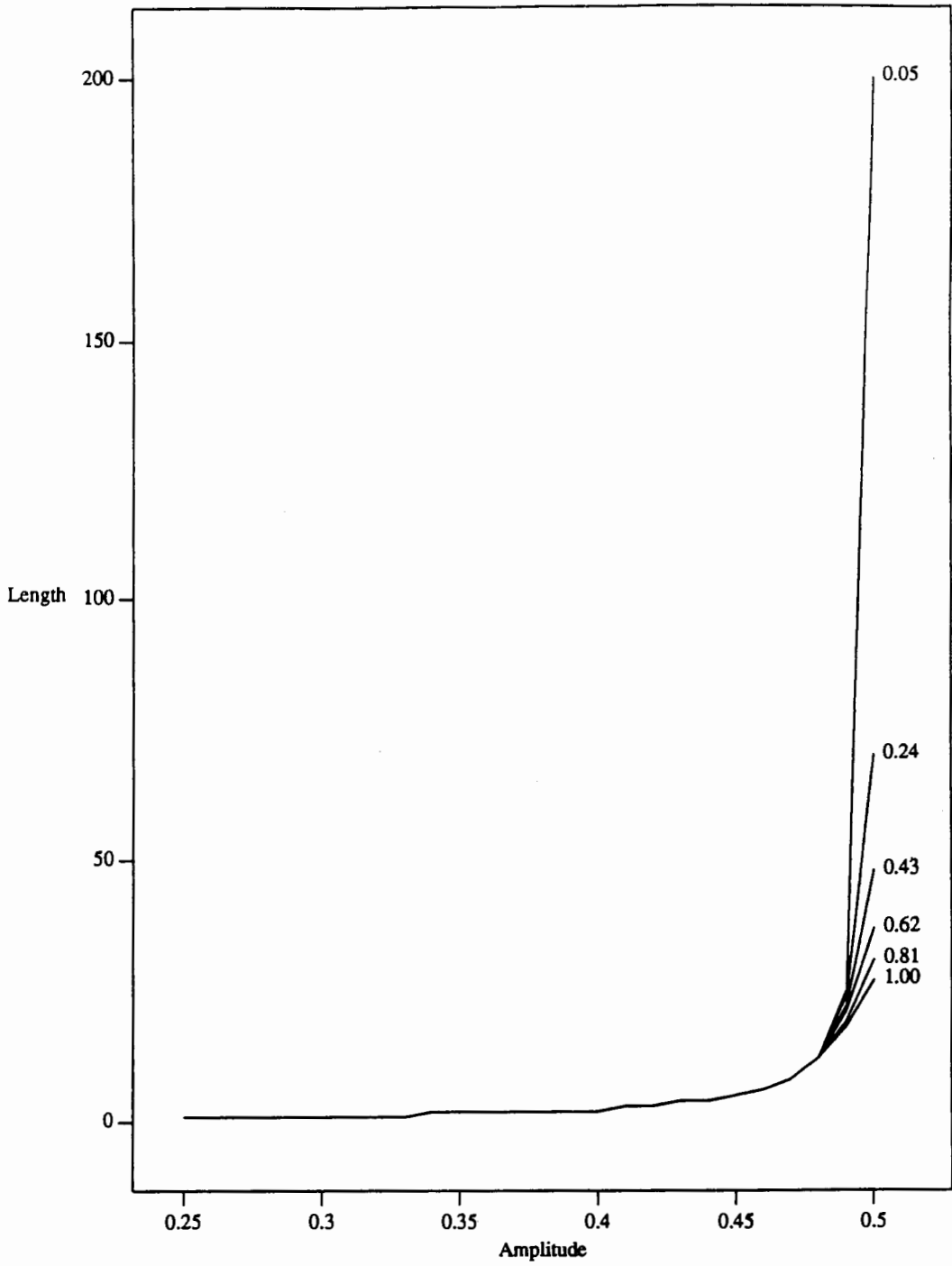


Figure 32.  $N$  vs Amplitude for several frequencies for  $M = 256$ .



## CHAPTER X

### CONCLUSION

The goal of this study has been to understand the quantization noise characteristics of a couple types of oversampled analog to digital converters. Although there are other sources of noise due to implementation issues and circuit imperfections, the quantization noise cannot be eliminated since the values of a digital signal are discrete by definition. The only way to reduce the quantization noise is to increase the resolution of the digital signal. One method to achieve this is to increase the physical resolution of the quantizer by making it able to distinguish between more levels. As discussed in chapter I, it becomes increasingly difficult to implement quantizers as the resolution increases due to the increasingly severe specifications placed on the circuit components, such as the low pass analog anti-aliasing filter.

Instead of increasing the physical resolution of the quantizer, oversampling and spectral shaping can be used to increase the effective resolution of low resolution quantizers. Oversampling increases the signal bandwidth which allows the use of a simpler analog anti-aliasing filter. The increase in effective resolution is made possible because the quantization noise is spread across the signal bandwidth and the part of the sampling bandwidth that is greater than the signal bandwidth can

be filtered out with digital filters after the quantization process. The advantage of digital processing is that it is precise and not touchy like highly precise analog components. Circuits like the  $\Sigma\Delta M$  use spectral shaping techniques to move more of the noise into high frequencies so that more of it will be filtered out and even more increase in effective resolution will be achieved.

## X.1 SUMMARY OF RESULTS

This study did not address the issues involved in the digital filtering and decimation process. Assuming that the digital processing is ideal, or at least very good, then the important characteristics of the quantization noise will be apparent in the oversampled output of the quantizer. Two oversampled ADC's were studied. The uniform ADC and the  $\Sigma\Delta M$ . It was found that a uniform ADC with a low resolution quantizer did not have good quantization noise characteristics over the signal bandwidth. With a high oversampling ratio, the quantization error becomes highly periodic and produces strong odd numbered harmonics. So, the output power spectrum for input frequencies in the first third of the signal bandwidth will have significant harmonic content. After that, the harmonics move into the region of the sampling bandwidth which is filtered out. As a result, the SQNR across the signal bandwidth is highly irregular. It was shown that higher resolution quantizers performed better. This would indicate that oversampling a high resolution uniform ADC could be a practical method to achieve modest increases in resolution. However, this study intentionally focuses on low resolution

quantizers since the feature ADC considered is the  $\Sigma\Delta M$ , which typically uses a one bit quantizer. So, understanding the quantization noise characteristics of low resolution oversampled quantizers provides a foundation that provided some insights into the operation of the more complicated  $\Sigma\Delta M$ . A Fourier series analysis was developed to explain the behavior of the uniform ADC for high oversampling ratios. It was useful in explaining the observed behavior.

One way to look at the  $\Sigma\Delta M$  is to consider it as a one bit quantizer which provides its own dither signal. This dither signal can also be viewed as a correction signal since the input to the quantizer is modified by the opposite of the previous total quantization error. The result is that the quantization noise signal becomes much more active and has most of its energy located high in the sampling bandwidth. This shaping of the power spectrum of the quantization noise results in greater increases of the SQNR for the  $\Sigma\Delta M$  in comparison to the uniform ADC.

Some theoretical analyses of the  $\Sigma\Delta M$  for dc and sinusoid inputs were discussed. The theories were compared against simulated results which spanned the range of the signal bandwidth. The simulated results matched the theoretical predictions. The simulation results also brought to light behavior in the quantization noise signal which was not readily apparent from the theoretical predictions. An important feature of this study was the use of simulations over the range of valid input signals to bring out behavior that would not have been obvious from the theoretical predictions or from a few isolated simulated results. The use of extensive simulations over the valid signal ranges also emphasized the limitations

of the commonly used assumption that the quantization noise is white. The white noise assumption was clearly inadequate for the oversampled, low resolution uniform ADC. Although the SQNR results for the  $\Sigma\Delta M$  seemed to indicate a better match with the white noise theory, simulations and exact theoretical predictions showed some definite frequency and amplitude dependent spectral structure.

## X.2 APPLICATIONS AND EXTENSIONS OF RESULTS

The base of knowledge of this study is relatively fundamental. First, the quantization noise characteristics of the oversampled, low resolution uniform ADC were studied and understood. Then the characteristics for the  $\Sigma\Delta M$  were developed for dc and sinusoid inputs. It was found that knowledge of the simpler ADC's and simple inputs provided useful insight into the operation of the most complicated case considered, which was the sinusoid input to the  $\Sigma\Delta M$ .

The one feedback loop, or first order,  $\Sigma\Delta M$  considered here is just the beginning. ADC's using higher order  $\Sigma\Delta M$ 's can be made. Configurations which cascade more than one  $\Sigma\Delta M$  are also possible. Other possible variations include using a function other than the discrete time integrator, or increasing the number of bits of the quantizers. Whatever the configuration, the results produced in this study should be able to provide some insights into the behavior of more complex circuits that would not be obtained from assumptions of white quantization noise.

The important result from this study is that the noise characteristics of the oversampled ADC are highly dependent on the input signal. It is important to

understand the noise behavior because actual results could be unexpectedly poor if simplified assumptions, such as white quantization noise, are made. In particular, an application using the  $\Sigma\Delta M$  should probably be designed so that the input signal does not approach too close to the limits of the input range. The results of the simulations showed that the SQNR dropped for sinusoidal inputs as the amplitude reached full amplitude. The noise power for dc inputs also reached a maximum when the dc input was close to the limit. The simulation results also indicate that sinusoidal inputs with frequencies in the lower half of the signal bandwidth tend to have more noise power localized in a few harmonics. While the SQNR may be acceptable, it may be necessary to examine whether the power in these harmonics is acceptable or not.

Although the noise characteristics of the two ADC's examined in this study may not be good enough for some applications where very accurate conversion is required, knowledge of the noise characteristics could provide the information needed to determine if one of the simple ADC's is good enough for a particular application.

## REFERENCES CITED

- [1] Bhagwati P. Agrawal and Kishan Shenoi. Design methodology for  $\Sigma\Delta M$ . *IEEE Transactions on Communication*, COM-31(3):360–370, March 1983.
- [2] Bernhard E. Boser and Bruce A. Wooley. Quantization error spectrum of sigma-delta modulators. Technical report, Stanford University
- [3] Timothy F. Darling and Malcolm O. J. Hawksford. Oversampled analog-to-digital conversion for digital audio systems. *J. Audio Eng. Soc.*, 38(12):924–943, December 1990.
- [4] David J. DeFatta, Joseph G. Lucas, and William S. Hodgkiss. *Digital Signal Processing: A System Design Approach*. John Wiley and Sons, Inc., 1988.
- [5] Robert M. Gray. Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input. *IEEE Transactions on Communications*, 37(6):588–599, June 1989.
- [6] Robert M. Gray. Quantization noise spectra. *IEEE Transactions on Information Theory*, 36(6):1220–1244, November 1990.
- [7] Robert M. Gray, Wu Chou, and Ping W. Wong. Quantization noise in single-loop sigma-delta modulation with sinusoidal inputs. *IEEE Transactions on Communications*, 37(9):956–968, September 1989.
- [8] B. Hutchins and W. H. Ku. *Introduction to Signal Processing*. Electronotes, 1985.
- [9] Louis A. Williams III, Bernhard E. Boser, Edward W. Y. Liu, and Bruce A. Wooley. *MIDAS User Manual*. Center for Integrated Systems, Stanford University, June 1990.

## APPENDIX

### VIDEO CLIPS

A number of animated video clips were generated which demonstrate the behavior of some of the processes studied. One of the items of interest in this study is to understand how the various processes behave over the range of valid inputs. These clips provide a visual illustration to aid in understanding the processes.

The level of polish on the clips varies. The first couple are finished the most. Most of the clips were generated quickly for the purpose of identifying behavior that was not obvious from the analysis of the processes. As a result, descriptions of what the clips show and what the features of interest are follows.

#### A.1 SQNR FOR UNIFORM ADC WITH 4 BIT QUANTIZER

This segment shows the evolution of the SQNR for oversampling values from 1 to 128 for a frequency range covering the signal bandwidth. The quantizer used has 4 bit nominal resolution. The primary feature of interest is the flattening of the curve for low frequencies. As frequency increases, the flat portion of the curve rises. After the frequency passes  $\frac{f_n}{6}$ , the flattening effect disappears and the SQNR jumps around at values higher than what is predicted by the white noise theory.

## A.2 SIGNAL BANDWIDTH SPECTRUM FOR 4 BIT QUANTIZER

This clip shows the evolution of the frequency spectrum of a 4 bit quantizer output, with an oversampling ratio of 128, over the range of valid input frequencies. The primary feature of interest is the harmonics present in the spectrum for frequencies lower than  $\frac{f_n}{2}$ . These harmonics explain the flattening effect that appeared in A.1. As the harmonics move out of the signal bandwidth, the SQNR value will rise.

## A.3 THEORETICAL SPECTRUM OF DC INPUT TO $\Sigma\Delta M$

This clip shows the theoretical spectrum of the  $\Sigma\Delta M$  output for a dc input over the range of valid dc inputs. The entire sampling bandwidth is shown. The clip starts with a dc input of zero and increases to full scale. The fundamental noise frequency begins at  $\frac{f_s}{2}$  and moves down towards zero. The higher harmonics can be seen bouncing back and forth at ever increasing speeds. Also of interest is the effect of noise shaping. This can be seen as the fundamental approaches zero, its amplitude begins to decrease.

## A.4 SIMULATED SPECTRUM OF DC INPUT TO $\Sigma\Delta M$

This clip shows the simulated spectrum of the  $\Sigma\Delta M$  output for a dc input over the range of valid dc inputs. The entire sampling bandwidth is shown. The primary importance of this clip is that it shows exactly the same behavior that is



observed in A.3.

#### A.5 SQNR CURVES FOR $\Sigma\Delta M$ WITH SINUSOID INPUT

This segment shows the evolution of the SQNR of the  $\Sigma\Delta M$  output for oversampling values from 1 to 128 for a frequency range covering the signal bandwidth. The sinusoidal input has a full scale amplitude. This clip is similar in nature to A.1. Note that the flattening effect does not occur here. However, a close look will show that for low frequencies, the curve appears to be fairly stable. As the frequency increases, the curve jumps around more at higher values. This behavior is analogous to the flattening behavior in A.1.

#### A.6 SQNR SURFACE FOR $\Sigma\Delta M$ WITH $M = 128$

This clip shows the SQNR surface of 25. It is rotated in this clip to provide a clear picture of the nature of the surface.

#### A.7 SPECTRUM FOR $\Sigma\Delta M$ WITH FULL SCALE SINUSOID INPUT

This segment shows the evolution of the frequency spectrum of the  $\Sigma\Delta M$  output for an oversampling ratio of 128 for a frequency range covering the signal bandwidth. The sinusoidal input has a full scale amplitude. The feature of interest are the harmonics that can be picked out for the lower frequencies. Also, as the frequency becomes higher, the noise floor appears to drop somewhat. This verifies the behavior seen in A.6.

## A.8 SPECTRUM FOR $\Sigma\Delta M$ WITH 80 % SCALE SINUSOID INPUT

This segment shows the evolution of the frequency spectrum of the  $\Sigma\Delta M$  output for an oversampling ration of 128 for a frequency range covering the signal bandwidth. The sinusoidal input has a amplitude of about 80 % full scale. The motivation behind this clip is to see if there are any obvious differences from A.7, since the SQNR surface drops significantly.