Timing Offset and Quantization Error Trade-off in Interleaved Multi-Channel Measurements

by

Joseph Gary McMichael

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Author .....
Department of Electrical Engineering and Computer Science
May 13, 2011

Certified by .....
Alan V. Oppenheim
Ford Professor of Engineering
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Chair, Department Committee on Graduate Theses

Timing Offset and Quantization Error Trade-off
in Interleaved Multi-Channel Measurements

by

Joseph Gary McMichael

Submitted to the Department of Electrical Engineering and Computer Science
on May 13, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

Abstract

Time-interleaved analog-to-digital converters (ADCs) are traditionally designed with equal quantization granularity in each channel and uniform sampling offsets. Recent work suggests that it is often possible to achieve a better signal-to-quantization noise ratio (SQNR) with different quantization granularity in each channel, non-uniform sampling, and appropriate reconstruction filtering. This thesis develops a framework for optimal design of non-uniform sampling constellations to maximize SQNR in time-interleaved ADCs.

The first portion of this thesis investigates discrepancies between the additive noise model and uniform quantizers. A simulation is implemented for the multi-channel measurement and reconstruction system. The simulation reveals a key inconsistency in the environment of time-interleaved ADCs: cross-channel quantization error correlation. Statistical analysis is presented to characterize error correlation between quantizers with different granularities. A novel ADC architecture is developed based on weighted least squares (WLS) to exploit this correlation, with particular application for time-interleaved ADCs. A "correlated noise model" is proposed that incorporates error correlation between channels. The proposed model is shown to perform significantly better than the traditional additive noise model for channels in close proximity.

The second portion of this thesis focuses on optimizing channel configurations in time-interleaved ADCs. Analytical and numerical optimization techniques are presented that rely on the additive noise model for determining non-uniform sampling constellations that maximize SQNR. Optimal constellations for critically sampled systems are always uniform, while solution sets for oversampled systems are larger. Systems with diverse bit allocations often exhibit "clusters" of low-precision channels in close proximity. Genetic optimization is shown to be effective for quickly and accurately determining optimal timing constellations in systems with many channels.

Finally, a framework for efficient design of optimal channel configurations is formulated that incorporates statistical analysis of cross-channel quantization error correlation and solutions based on the additive noise model. For homogeneous bit allocations, the framework proposes timing offset corrections to avoid performance degradation from the optimal scenario predicted by the additive noise model. For diverse bit allocations, the framework proposes timing corrections and a "unification" of low-precision quantizers in close proximity. This technique results in significant improvements in performance above the previously known optimal additive noise model solution.

Thesis Supervisor: Alan V. Oppenheim
Title: Ford Professor of Engineering

Acknowledgments

First, I would like to thank my research advisor, Professor Al Oppenheim. Al's guidance and wisdom were vital for the development of this thesis. Our weekly meetings significantly shaped my approach to research. I am very appreciative of the support he has given me throughout my time at the Digital Signal Processing Group (DSPG) at MIT.

I would also like to express my sincere thanks to Shay Maymon for his helpful contribution to this thesis. Shay has been a close research collaborator and friend over the last two years, even taking the time to review my draft thesis. I wish him all the best in his future endeavors.

It has been a privilege working with the DSPG over the last two years. I would like to thank the past and present members of DSPG, including Tom Baran, Ballard Blair, Petros Boufounos, Sefa Demirtas, Dan Dudgeon, Xue Feng, Kathryn Fischer, Zahi Karam, Jon Paul Kitchens, Jeremy Leow, Shay Maymon, Martin McCormick, Milutin Pajovic, Charles Rohrs, Eric Strattman, and Dennis Wei. The weekly group meetings were lively and creative. Special thanks go to Dennis for his guidance related to research. Also, I would like to express thanks to Professor Muriel Médard for being a great teacher and providing advice and support.

My summer at Bose was a great experience. I want to thank my friend and mentor, Davis Pan, for truly caring about the development and well-being of his interns. It was a privilege to work with him.

Finally, closer to home, I thank my amazing wife Sarah for walking with me every step of the way and reminding me about the most important things in life. Thanks also to my parents and my brother, Jon, for their love and support.

Ad maiorem Dei gloriam.

To Sarah

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Accurate, rapid measurement of an analog continuous-time signal is an essential requirement for many applications of digital signal processing. Typically, a signal is sampled and quantized by a single measurement device to obtain an inherently lossy digital representation for processing. It is often possible to obtain more information about the signal by sampling with multiple devices, or "channels," distributed over space or time. For example, each node in a spatially distributed sensor network can independently measure, with analog-to-digital converters (ADCs), a radio-frequency signal transmitted from an unknown location. A base station collects this information from each sensor and reconstructs the approximate position of the transmitter by examining signal arrival time and strength at each sensor.

ADCs can also be distributed in time to obtain information about a signal. A common scenario referred to as time-interleaved A/D conversion involves interleaving sampling times of multiple channels to achieve a faster overall sample rate than the individual rate in each channel. These devices distribute the sampling task over many converters operating at a moderate rate, alleviating some of the limitations of individual ADCs. In practice, the sample rate achievable for a fixed resolution in a time-interleaved ADC is limited by non-linearities including thermal noise and sampling jitter caused by component mismatch [6]. Methods for mismatch estimation and compensation have been extensively studied ([3],[22],[6]).

In spite of these limitations, time-interleaved ADCs provide a powerful distributed architecture often employed to sample high bandwidth signals. Example applications include wireless communication technologies such as CDMA and software-defined radio, synthetic aperture radar, and 3-D

ultrasonic imaging ([23],[13]). For moderate bandwidth signals, the devices are utilized to achieve a high oversampling ratio [3]. Oversampling is a technique commonly used to mitigate the effects of quantization. This process takes advantage of a trade-off between the oversampling ratio ρ and the necessary quantization granularity to achieve a fixed signal-to-quantization noise ratio (SQNR). In the simplest form, a low pass filter with cutoff frequency $\omega_c = \pi/\rho$ and decimator follow the quantizer. As the oversampling ratio ρ increases, less of the quantization noise spectrum overlaps with the signal spectrum, allowing large bands to be removed without altering the signal ([19],[4]).

Traditionally, time-interleaved ADCs have often been designed to reconstruct non-uniform samples with only cursory treatment of the effect of quantization error [6]. The presence of quantization error can seriously degrade performance during reconstruction. This thesis investigates a novel technique for constructing time-interleaved ADCs that would require fewer overall bits than traditional interleaved ADCs to achieve a fixed SQNR. Specifically, we explore the appropriate choice of timing offsets τ_m and bit allocations b_m to achieve such a gain with the optimal reconstruction filters $G_m(e^{j\omega})$.

Even ADCs operating at exactly the same geographic location or time instant can be combined to obtain more information about the underlying analog signal than a single device. For example, consider two ADCs with different granularities simultaneously operating on the same analog sample. The resulting quantized samples can be combined to produce a better estimate of the analog signal.

The research in this thesis has a variety of applications in distributed sampling, beyond the context of time-interleaved A/D conversion. The problem formulation can be generalized so that time distances between channels correspond to geographic distances between ADCs. In this context, the problem involves determining optimal geographic locations of a set of different precision measurement sensors to maximize SQNR for a spatially bandlimited input signal.

1.1 Thesis Outline

This thesis considers the effects of quantization in interleaved, oversampled multi-channel measurements and the design of optimal channel configurations. Specifically, we explore the relationship between channel timing offsets and bit allocations. We begin by formulating the problem setup for the multi-channel interleaved measurement and reconstruction system and defining variables in Chapter 2. In Chapter 3, we describe a software implementation of the time-interleaved ADC that

will allow us to simulate performance with arbitrary numbers of channels, timing offsets, and bit allocations.

Chapters 4 and 5 investigate quantization error in time-interleaved ADCs. Specifically, Chapter 4 presents experimental evidence of a significant discrepancy between the additive noise model and uniform quantizers: cross-channel correlation between quantizer error signals. This results from close proximity of channel sampling instants. Correlation is also measured when the quantizers have different granularity. An experimental analysis of the effects of cross-channel quantizer error correlation on system performance is presented. We also experimentally determine a set of acceptable distances to avoid the effects of correlation. To conclude the chapter, we propose a "correlated noise model" that will model error correlation between channels. The new model is shown to perform significantly better than the traditional additive noise model for channels in close proximity.

In Chapter 5, we undertake a statistical analysis that gives a theoretical explanation for the measured cross-channel error correlation. The result is a simple expression for the error correlation coefficient for quantizers differing by n bits. We employ the Weighted Least Squares (WLS) estimator to formulate an architecture that will exploit the extra information from d quantized versions of a signal with adjacent numbers of bits. This architecture will prove useful for obtaining the best possible performance from a time-interleaved ADC. Finally, we show theoretically that the new architecture can achieve an effective resolution one bit greater than the highest resolution component quantizer. The results are verified with both random and audio signals.

The remaining portion of the thesis focuses on optimization of system parameters to achieve maximum signal-to-quantization noise ratio (SQNR). Chapter 6 considers the appropriate choice of timing offsets τ_m for a fixed bit allocation b_m. The predicted output error variance function is optimized both analytically and numerically for various numbers of channels and bit allocations. A key result of this thesis is a framework for efficient design of optimal channel configurations presented in Chapter 7. The framework incorporates knowledge from the previous chapters to correct predictions from the additive noise model. The corrected results are experimentally shown through simulation to perform well.

Chapter 8 considers a similar design scenario: optimal bit allocation b_m given fixed timing offsets τ_m. Through analytical optimization, we build an intuition for optimal assignment of high-

and low-precision quantizers. Again, the results are experimentally verified through simulation.

Chapter 2

Multi-Channel Sampling and Reconstruction

In this chapter, we describe the general problem setup for the multi-channel interleaved measurement and reconstruction system. The continuous-time input to the system $x(t)$ is assumed to be bandlimited with cutoff frequency Ω_c. We will typically classify the input signal $x(t)$ by its Nyquist rate $1/T_N$, which is twice the bandwidth Ω_c.

The goal of this system is to maximize the signal-to-noise ratio (SNR) of the reconstructed input signal given a set of quantized measurements. The system is composed of a measurement stage and a reconstruction stage. The basic multi-channel interleaved measurement and reconstruction system considered in this thesis is shown in Figure 2-1.

2.1 Measurement Stage

The purpose of the measurement stage is to acquire accurate digital measurements of the analog input signal $x(t)$. The system consists of parallel channels. In each channel, the bandlimited signal $x(t)$ with Nyquist rate $1/T_N$ is undersampled at a rate of $1/T = 1/LT_N$. This operation is performed in M channels ($M > L$). If no two measurements are obtained at the same sampling instant, this corresponds to an oversampling factor of $\rho = M/L > 1$.

The resulting samples can be thought of as an interleaved combination of M sequences of uniformly-spaced samples taken at one L^{th} of the Nyquist rate of $x(t)$. We denote the clock timing

offsets in one recurrence period by τ_m. The complete set of sampling points is given by:

$$nL \cdot T_N + \tau_m \qquad \text{for } \{n \in \mathbb{Z}\} \text{ and } \{m \in \mathbb{Z} \mid 0 \leq m \leq M-1\} \tag{2.1}$$

A uniform sampling grid is obtained if the timing offsets correspond to complex exponentials $e^{j2\pi\frac{\tau_m}{LT_N}}$ equally spaced on the complex unit circle. Otherwise, the interleaved grid can be characterized as recurrent non-uniform sampling ([12],[26]). An example of recurrent non-uniform sampling is depicted in Figure 2-2 for $M = 3$ and $L = 2$. The sample points from each channel have recurrence period $L \cdot T_N$ and are denoted with $\times$, $\diamond$, and $\bullet$, respectively.

After sampling, the discrete-time analog signal is passed through a quantizer to obtain an approximation for each sample from a discrete set of values. This completes the measurement stage of Figure 2-1.

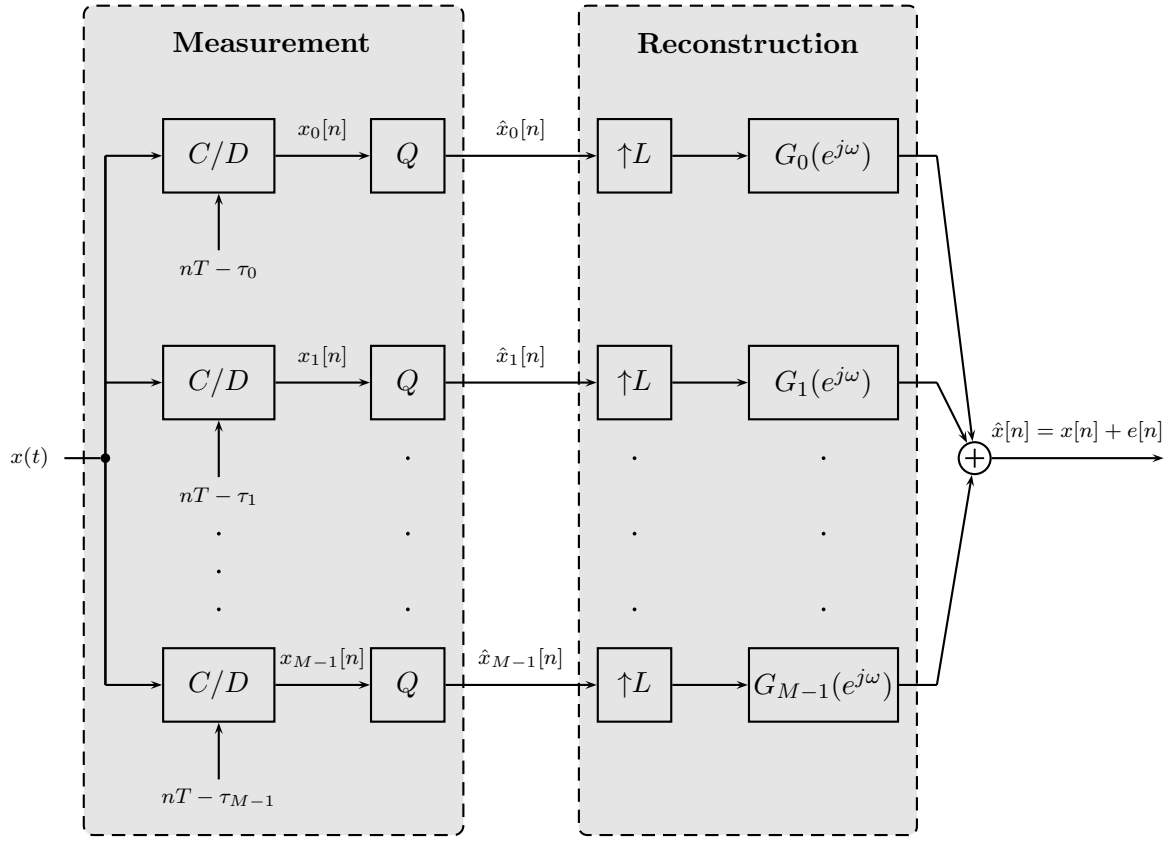


Figure 2-1: Multi-channel interleaved system with reconstruction filters.

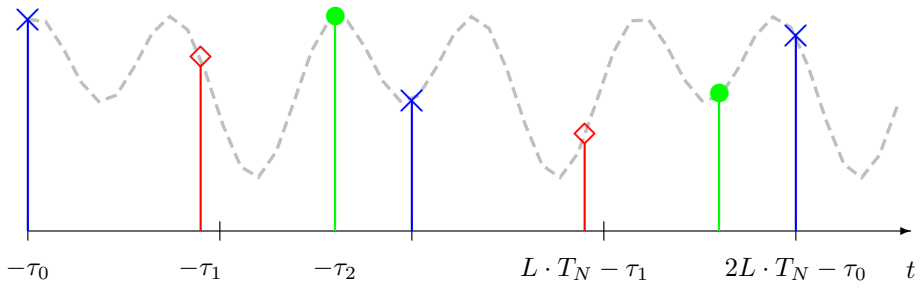


Figure 2-2: Example of recurrent non-uniform samples for $M = 3$ and $L = 2$.

2.2 Signal Reconstruction from Recurrent Non-Uniform Samples

The goal of the reconstruction stage in Figure 2-1 is to obtain a high fidelity representation of the input signal $x(t)$ sampled at its Nyquist rate by compensating for non-uniform timing offsets and quantization error.

In this section, we examine the case of reconstruction in the absence of quantization error, and present two approaches for designing perfect reconstruction systems from recurrent non-uniform samples.

2.2.1 Lagrange Interpolation

In [26], Yen showed that a bandlimited signal $x(t)$ can be perfectly reconstructed from samples in the form of equation (2.1). In other words, a bandlimited signal is completely specified by recurrent non-uniform samples if the average sampling rate exceeds the Nyquist rate. Yen provided an explicit reconstruction formula for the case of recurrent non-uniform sampling when $M = L$. Later, Yao and Thomas ([25]) applied a family of polynomials known as Lagrange interpolation functions to reconstruct bandlimited signals from non-uniform samples. Yen's reconstruction formula for the recurrent non-uniform sampling pattern is given by:

$$x_c(t) = \sum_{n=-\infty}^{\infty} \sum_{m=0}^{M-1} x_c(nL \cdot T_N + \tau_m) \frac{a_m (-1)^{nM} \prod_{q=0}^{M-1} \sin(\pi(t - \tau_q)/LT_N)}{\pi \left(t - nL \cdot T_N - \tau_m\right)/LT_N} \tag{2.2}$$

where

$$a_m = \frac{1}{\prod_{q=0, q \neq m}^{M-1} \sin(\pi(t_m - t_q)/LT_N)}. \tag{2.3}$$

Equation (2.2) performs perfect reconstruction in the absence of quantization error. However, the operation is considerably more complex than reconstruction from uniform samples. The process can be efficiently implemented using a discrete-time filter bank, as shown by Eldar and Oppenheim [7].

Lagrange interpolation functions provide a way to reconstruct from recurrent non-uniform samples. However, there are alternative ways to derive a system that performs perfect reconstruction from recurrent non-uniform samples. We will examine the constraints necessary for perfect reconstruction in the next section. For the case when $M = L$, the system that results from imposing perfect reconstruction is equivalent to the system implied by Lagrange interpolation.

2.2.2 Constraints for Perfect Reconstruction

In the absence of quantization error, a set of spectral constraints necessary to obtain perfect reconstruction is applied to each filter $G_m(e^{j\omega})$. In this section, we will examine these spectral constraints. First we will consider the set of M signals $x_m[n]$ that are the inputs to M reconstruction stages. $x_m[n]$ is the result of time-delaying $x(t)$ and undersampling by L. Thus, we expect $x_m[n]$ to be significantly distorted by the effects of aliasing. The relation between the discrete-time Fourier transform and the continuous-time Fourier transform of the input signal is defined in equation (2.4).

$$X_m(e^{j\omega}) = \frac{1}{L \cdot T_N} \sum_{k=-\infty}^{\infty} X_c \left[j \left(\frac{\omega}{L \cdot T_N} - \frac{2\pi k}{L \cdot T_N} \right) \right] \tag{2.4}$$

From equation (2.4), we conclude that the aliased signal $\hat{x}_m[n]$ in each channel consists of the superposition of $2L - 1$ shifted replicas of the original spectrum of $x(t)$. This is illustrated in Figure 2-3 for $L = 3$. It is important to note that at each frequency ω, only L aliased replicas

contribute to the spectrum. Thus, $L-1$ aliased replicas must be removed to obtain the desired spectrum.

Maymon and Oppenheim ([18],[16],[17]) show that in the absence of quantization error, the summed output $\hat{x}[n]$ corresponds to uniform samples of $x(t)$ at the Nyquist rate when

$$\sum_{m=0}^{M-1} G_m(e^{j\omega}) \cdot \left[\frac{1}{T} \cdot \sum_{k=-\infty}^{\infty} X\left(\frac{\omega - \frac{2\pi}{L}k}{T_N} \right) \cdot e^{-j(\omega - \frac{2\pi}{L}k)\frac{\tau_m}{T_N}} \right] = \frac{1}{T_N} X\left(\frac{\omega}{T_N} \right) \text{ for } |\omega| < \pi. \tag{2.5}$$

The condition in equation (2.5) formalizes the fact that the spectrum of the filtered and summed signal $\hat{x}[n]$ must be identical to the spectrum of the signal $x(t)$ sampled at its Nyquist rate. The authors in [18] rearrange the constraint in equation (2.5) to be of the form shown in equation (2.6). In Figure 2-3, we see that while only L aliased replicas contribute at a given frequency, the specific replicas of interest are dependent on the frequency region. This form explicitly partitions the frequency spectrum into regions in each of which the same L aliased replicas contribute.

$$\begin{aligned} &\sum_{m=0}^{M-1} G_m(e^{j\omega}) \cdot e^{-j(\omega - \frac{2\pi}{L}k)\tau_m / LT_N} = L \cdot \delta[k] \qquad \text{for } \omega \in \Delta\omega_i \\ &\text{where } k = -i, -i+1, \ldots, L-1-i \qquad , i = 0, 1, \ldots, L-1 \\ &\text{and } \Delta\omega_i = \left[\pi - (i+1)\frac{2\pi}{L}, \pi - i\frac{2\pi}{L} \right] \end{aligned} \tag{2.6}$$

These constraints can also be viewed in the time domain. Intuitively, the filters $G_m(e^{j\omega})$ are chosen to compensate for deviations in timing offsets τ_m from the uniform grid in each channel. This operation can also be regarded as a fractional delay to transform the interleaved sample pattern to a uniform grid so that traditional sinc interpolation may be used to recover the continuous-time input signal $x(t)$.

2.2.3 Compensating for Quantization Error

In this section, we discuss design of optimal reconstruction filters $G_m(e^{j\omega})$ that achieve perfect reconstruction in the absence of quantization error and minimize quantization error variance in the presence of quantizers under the set of constraints in equation (2.6). The effect of quantization error

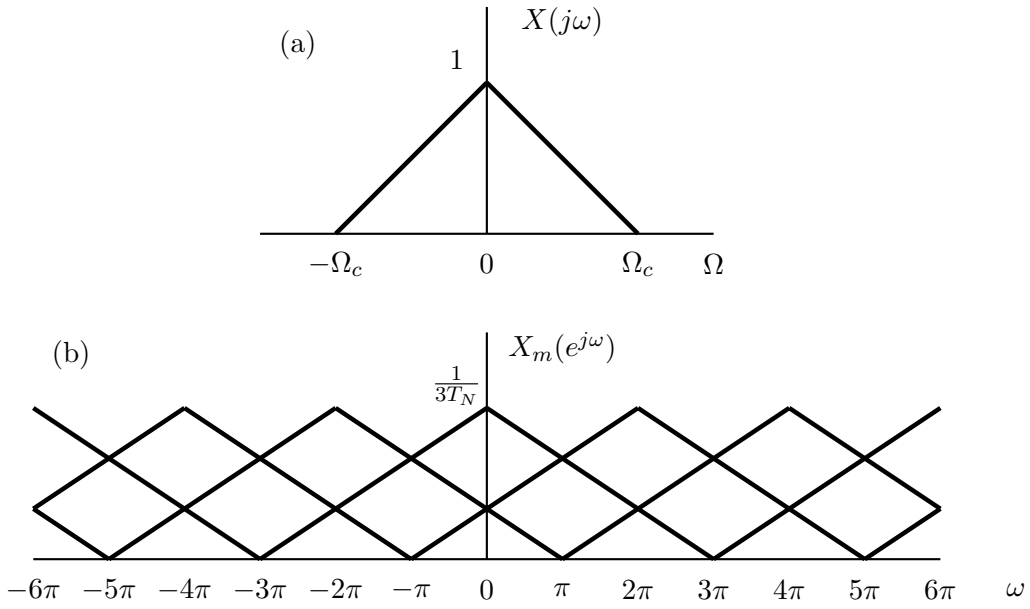


Figure 2-3: (a) Fourier transform of bandlimited input signal. (b) Fourier transform $X_m(e^{j\omega})$ of sampled input with aliasing for $L = 3$ and $\tau_m = 0$.

has often received only cursory treatment in recurrent non-uniform reconstruction [6]. However, quantization error can seriously degrade performance ([18],[16]).

A well-established analysis technique developed by Widrow and extended by others ([24],[21],[1],[9]) models quantization error as additive uniformly distributed white noise. Specifically, the quantizer error in each channel is assumed to be a random process uniformly distributed between $\Delta_m/2$ and uncorrelated with the input, where Δ_m denotes the quantizer step size. The uniform probability density function (PDF) of quantization error modeled with the additive noise model is illustrated in Figure 2-4.

Addition of independent noise is not a perfect representation of quantization error. The quantization error and input are deterministically related and are not uncorrelated, much less statistically independent. However, theoretical analysis has shown that the quantizer may be replaced for the purposes of moment calculation with additive independent uniformly distributed noise if certain characteristics of the input signal are satisfied [24]. In particular, if Quantizing Theorem I or II derived by Widrow in [24] are satisfied, one can assume that quantization error is uncorrelated with the signal being quantized, and all moments correspond exactly. The necessary and sufficient condition for uniformly distributed quantization noise was derived by Sripad and Snyder [21].

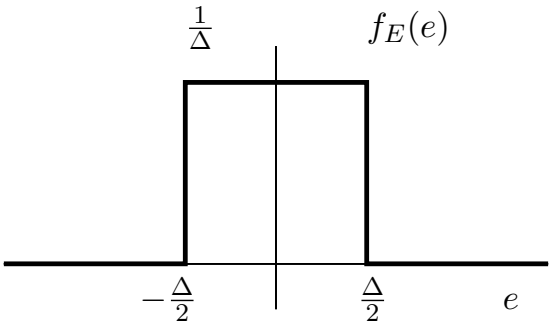


Figure 2-4: PDF of quantization error modeled by additive noise model.

The quantizing theorems formulated by Widrow are listed below. Consider a random variable x with PDF $f_x(x)$. The characteristic function (CF) $\Phi_x(u)$ is defined as the Fourier transform of the PDF. The random variable x is quantized to obtain x', with quantization grain size Δ.

Widrow Quantizing Theorem I ([24])

If the CF of x is bandlimited so that

$$\Phi_x(u) = 0, \qquad |u| > \frac{\pi}{\Delta}, \tag{2.7}$$

then:

- *the replicas contained in $\Phi_{x'}(u)$ will not overlap*
- *the CF of x can be derived from the CF of x'*
- *the PDF of x can be derived from the PDF of x'.*

Widrow Quantizing Theorem II ([24])

If the CF of x is bandlimited so that

$$\Phi_x(u) = 0, \qquad |u| > \frac{2\pi}{\Delta}, \tag{2.8}$$

then the moments of x can be calculated from the moments of x'.

Random variables that occur in practice do not satisfy these conditions. However, in [1] Bennett showed that the additive noise model performs quite well under most circumstances, particularly when the quantizer is not saturated, the granularity is small, and the PDF of the input is smooth.

Thus, the quantizer output $\hat{x}_m[n]$ is modeled as $\hat{x}_m[n] = x_m[n] + q_m[n]$ where $q_m[n]$ is an independent uniform random process specified above. The error signal $q_m[n]$ is defined to be the quantized sample minus the analog sample, $q_m[n] = \hat{x}_m[n] - x_m[n]$, while σ_m^2 denotes the variance of the additive noise in each channel. We denote the overall error signal after reconstruction as $e[n]$.

Analysis in [18] and [16] shows that under the additive noise model for quantization error, the time and ensemble average power of $e[n]$ is equivalent to equation (2.9).

$$\sigma_e^2 = \frac{1}{L}\sum_{n=0}^{L-1} E(e^2[n]) = \frac{1}{L}\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m=0}^{M-1}\sigma_m^2\cdot|G_m(e^{j\omega})|^2\,d\omega \tag{2.9}$$

The optimal reconstruction filters are chosen to minimize this expression under the set of constraints asserted in equation (2.6). The form of these filters will be introduced in the next chapter.

Given a recurrent non-uniform sampling pattern τ_m, the overall SQNR is maximized with different quantizer granularities in each channel. However, for the case where the timing offsets τ_m correspond to uniform sampling, the signal-to-quantization noise ratio (SQNR) is maximized with equal quantizer granularity in each channel [18].

Chapter 3

Simulation Implementation

The derivation of the optimal reconstruction filters $G_m(e^{j\omega})$ in [18] and [16] relies on the additive noise model for quantization error. Although this is a good model in many cases, discrepancies in performance may arise when quantization error in different channels is correlated or when the error PDF is non-uniform. To explore the discrepancies in more detail, we have implemented a simulation of the multi-channel sampling and reconstruction system using MATLAB.

3.1 Simulation Architecture

The architecture of the software simulation consists of several stages, depicted in Figure 3-1. Some technical issues must be addressed in implementing a practical simulation, which will be discussed in the following subsections.

3.1.1 Input and Sampling

The simulation accepts inputs in the form of analytical continuous-time functions $x(t)$ or discrete-time waveform audio (WAV) recordings $x[n]$. Discrete-time inputs require extra signal conditioning to acquire samples of the continuous signal at an arbitrary time through interpolation. Finally, the signal is sampled in each channel with rate $1/T = 1/LT_N$ and timing offset τ_m as described in Chapter 2.

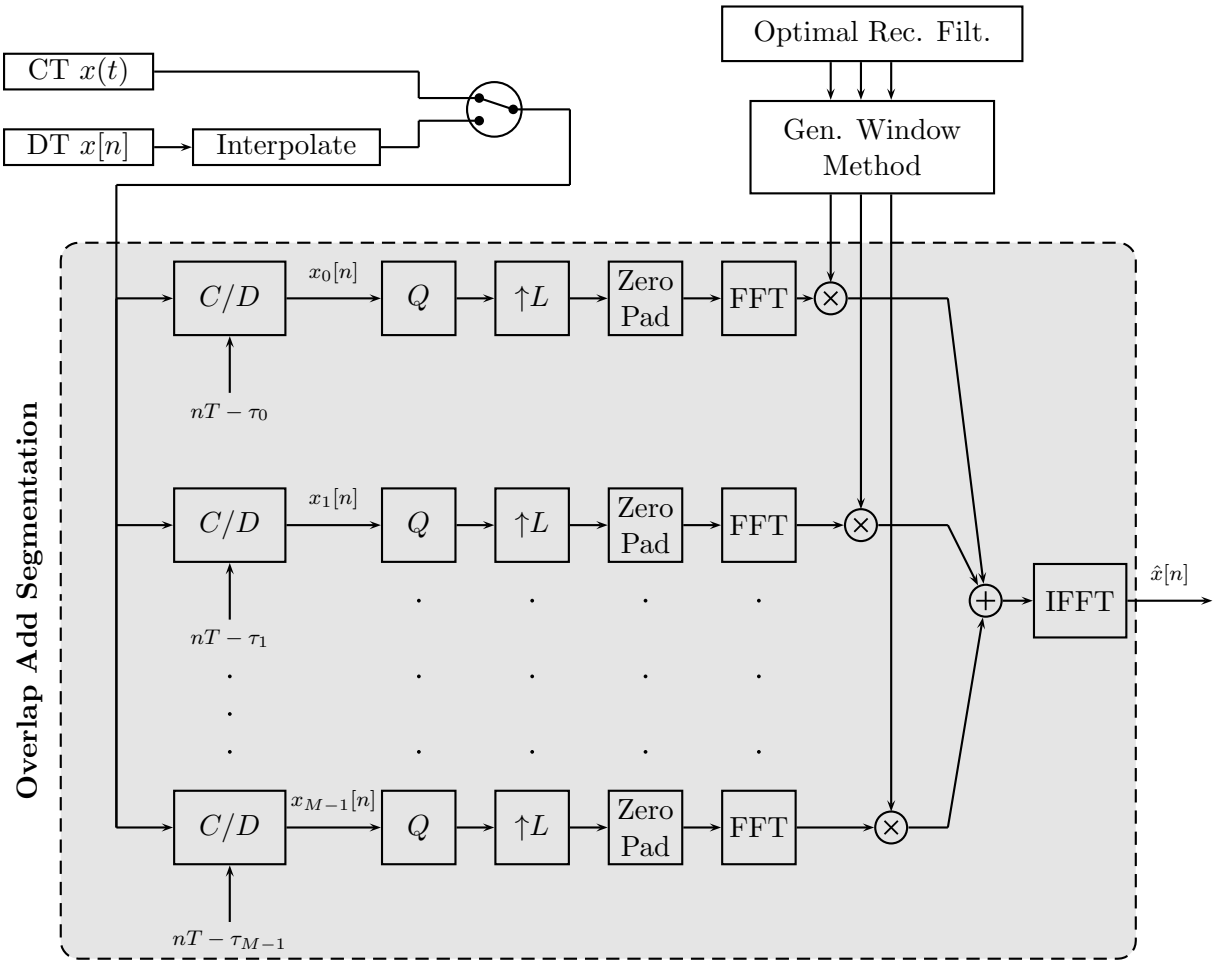


Figure 3-1: Simulation architecture.

3.1.2 Quantization

Quantization error will take three possible forms in this simulation. First, additive uniformly distributed noise sources $q_m[n]$ with arbitrary variance σ_m^2 in each channel have been implemented as described earlier. Second, ideal uniform quantizers are implemented. The term "ideal" is used to distinguish these from real, non-ideal ADCs that are subject to production errors. The word length corresponds to the total number of bits, while the number of bits after the binary point is referred to as the fraction length. For word length w and fraction length f, the dynamic range is from -2^{w-f-1} to $2^{w-f-1} - 2^{-f}$ and the precision is 2^{-f}. The expected noise variance according to the additive noise model is given by:

$$\sigma_m^2 = \frac{\Delta_m^2}{12} = \frac{(2^{-f_m})^2}{12}. \tag{3.1}$$

We will denote the number of bits in each channel as b_m. An alternative notation is the $(1 \times M)$ vector $\mathbf{b}$, where the m^{th} entry corresponds to the number of bits in channel m. We consider only fixed point, uniform quantizers, where the fraction length is one less than the word length. Thus, a k bit quantizer has precision:

$$\Delta = 2^{-(k-1)} \tag{3.2}$$

with 2^k total levels. Approximating the quantization error with a uniformly distributed random variable as discussed in Chapter 2 results in the following expression for error variance:

$$\sigma^2 = \frac{2^{-2(k-1)}}{12}. \tag{3.3}$$

The third form for representing the quantizer is a modified additive noise model with cross-channel correlation. Specifically, we model the quantization error as a Gaussian random process with an arbitrary positive-definite covariance matrix. We obtain these random processes through a Cholesky decomposition [10], discussed further in Chapter 4.

3.1.3 Reconstruction

Prior to reconstruction, the signal is upsampled by L to the Nyquist rate of the original signal. As depicted in Figure 3-1, the function labeled "Optimal Reconstruction Filter" calculates the optimal frequency response given in equation (3.4) at a grid of desired frequency values. This is a computationally intensive task.

$$G_m(e^{j\omega}) = \frac{1}{\sigma_m^2} \cdot e^{\frac{j\omega\tau_m}{T_N}} \cdot \Lambda^{(i)}(e^{j\omega_m}) \qquad \omega \in \left[\pi - (i+1)\frac{2\pi}{L}, \pi - i\frac{2\pi}{L}\right] \tag{3.4}$$

We define $\Lambda^{(i)}(e^{j\omega_m})$ to be the discrete-time Fourier transform of the finite-length sequence $\left\{\lambda_k^{(i)}\right\}_{k=-i}^{L-1-i}$ sampled at $\omega_m = \frac{2\pi\tau_m}{LT_N}$.

For $i = 0, 1, \ldots, L-1$ the sequence $\underline{\lambda}^{(i)} = \left\{\lambda_k^{(i)}\right\}_{k=-i}^{L-1-i}$ is defined as the solution to the following set of equations:

$$A_M \cdot \underline{\lambda}^{(i)} = L \cdot \underline{e}_i \tag{3.5}$$

where $\underline{e}_i$ is an indicator vector whose i^{th} entry is one and all other entries zeros. A_M is a $L \times L$ Hermitian Toeplitz matrix such that:

$$A_M = \sum_{m=0}^{M-1} (\underline{v}_m \cdot \underline{v}_m^H)/\sigma_m^2 \tag{3.6}$$

where

$$\underline{v}_m^H = [1, e^{-j2\pi \frac{\tau_m}{LT_N}}, \ldots, e^{-j2\pi \frac{\tau_m}{LT_N}(L-1)}]. \tag{3.7}$$

As shown in Figure 3-1, filtering is carried out in the frequency domain by multiplying the discrete Fourier transform (DFT) of the input signal by the sampled frequency response of the optimal filters. In general, multiplication in the DFT domain corresponds to circular convolution. However, circular convolution is identical to linear convolution if N, the length of the DFTs, satisfies $N \geq W + P - 1$, where W and P are the lengths of the input sequences [19]. After zero-padding the signal to length $N \geq W + P - 1$, the operation corresponds to linear convolution with an approximately optimal reconstruction filter.

In general, the calculated optimal frequency response corresponds to an infinite length, non-causal impulse response. We must precisely limit the impulse response to satisfy the condition discussed earlier and obtain linear convolution. This is accomplished using smoothing in the frequency domain (or windowing in the time domain) with a procedure known as the "Generalized Window Method" [20]. The procedure involves calculating the inverse fast Fourier transform (IFFT) of the impulse response, multiplying by a window that includes the majority of the impulse energy, and calculating the fast Fourier transform (FFT) to obtain the corrected frequency response.

3.1.4 Overlap-Add Method

While we could store the samples and subsequently compute a single very large DFT, this method introduces significant delay and may be impractical to compute. We have implemented overlap-add processing, which distributes computation into smaller tasks by constructing the output from short filtered blocks of samples [19].

3.2 Verification of the Simulation

To verify that the simulation behaves as expected, both narrowband and wideband signals are processed in the absence of quantization error. The overall error introduced by the system is defined in equation (3.8).

$$e[n] = \hat{x}[n] - x[n] \tag{3.8}$$

The unbiased sample variance of the overall error signal $e[n]$ with N elements is denoted σ_e^2 and calculated as follows, where $\bar{e}$ is defined as the sample mean of $e[n]$.

$$\sigma_e^2 = \frac{1}{N-1} \sum_{i=1}^{N} (e[i] - \bar{e})^2 \tag{3.9}$$

A successful test will yield an acceptably small error variance σ_e^2 in the absence of quantization error. This measurement is an important property of the simulation that depends on several approximation parameters, such as the length of the impulse response of the reconstruction filters and the length of the impulse response of the interpolation used during pre-conditioning. In particular, the measurement provides a lower bound on the variance of the additive noise that can be introduced into each channel. Below this bound, any additive noise will be masked by error from the simulation.

The signal-to-noise ratio (SNR) is defined in equation (3.10), expressed in dB.

$$SNR_Q = 10 \cdot \log_{10} \left(\frac{\sigma_x^2}{\sigma_e^2} \right) \tag{3.10}$$

One of our primary test signals is a high-quality recording of the Norwegian Chamber Orchestra performing "Beethoven's Symphony No. 1 in C," sampled at 96 kHz with 24 bits per sample. For verification, we consider the case when $M = 3$, $L = 2$, and $\boldsymbol{\tau} = [-\frac{2}{3}T_N,\ 0,\ \frac{2}{3}T_N]$ (for which the sampling system is equivalent to uniform sampling).

In the absence of quantization error, with a reconstruction impulse length of 300,000 samples, the measured SNR is 99.5 dB. According to [19], the SNR for a given number of quantizer bits B operating on an audio signal that has been scaled to avoid clipping is approximately equal to:

$$SNR_Q \approx 6B - 1.25\text{dB}. \tag{3.11}$$

Thus, results from the simulation will be conclusive for quantizers with fewer than 16.8 bits. Note that this can be improved by lengthening the reconstruction impulse response.

The relationship between reconstruction impulse length and SNR in the absence of quantization error is illustrated in Figure 3-2. It is important to note that an approximately exponential increase in reconstruction impulse length is necessary to obtain a fixed increase in SNR. One of the reasons for this is that the majority of the energy of the impulse response tends to be concentrated close to zero.

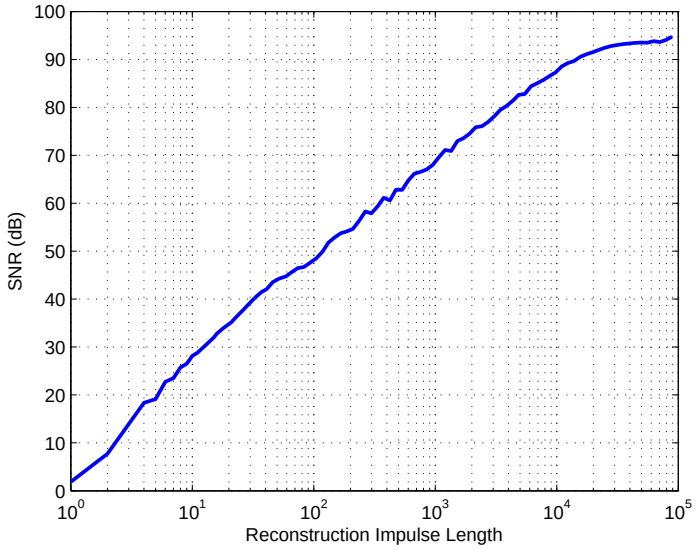


Figure 3-2: SNR as a function of the reconstruction impulse response length, in the absence of quantization error.

Chapter 4

Quantization Error in Time-Interleaved ADCs

Although the additive noise model is a well-established representation for quantization error that holds under many circumstances, it is important to experimentally verify its validity in the environment of distributed measurement systems. In this chapter, we present a key discrepancy that is particularly significant for time-interleaved ADCs: cross-correlation between quantization error in different channels. We quantify the effects of error cross-correlation on noise reduction performance. We also develop a correlated noise model and experimentally verify that it is a better representation of quantization error when sampling instants fall in close proximity to each other. To illustrate this discrepancy, first consider the following example in section 4.1.

4.1 Comparison of Additive Noise Model with Uniform Quantizers

Analysis in [18] based on the additive noise model suggests that with equal quantizer step size in each channel, the SQNR is maximized when the timing offsets τ_m correspond to uniform sampling. To verify this result, we model the quantization error in each channel as additive uniformly distributed noise with variance σ_m^2 and denote $\sigma_{e_{min}}^2$ as the time and ensemble average power of the error after optimal reconstruction filtering. $\sigma_{e_{min}}^2$ is an important measure of performance. For normalization purposes, we also introduce a measure of performance defined in [18] as the "reduction factor" γ.

γ represents the reduction in average power between the quantization noise sources and the output error after optimal reconstruction filtering:

$$\gamma = \frac{\frac{1}{M}\sum_{m=0}^{M-1}\sigma_m^2}{\sigma_{e_{min}}^2}. \tag{4.1}$$

Figure 4-1 displays the simulated reduction factor γ as a function of τ_1 and τ_2 for $M = 3$, $L = 2$, $\boldsymbol{b} = [10,\ 10,\ 10]$, and $\tau_0 = 0$. This simulation was performed with the additive noise model, operating on the high quality audio recording described in Chapter 3. As predicted, maximum noise reduction γ is achieved for $\tau_1 = 2/3 \cdot T_N$ and $\tau_2 = -2/3 \cdot T_N$, which corresponds to uniform sampling. As $\tau_1 \to 0$ and $\tau_2 \to 0$, the noise reduction capability is significantly reduced. This suggests that less information about the analog input signal is obtained when two channels with the same quantizer granularity are placed in close proximity.

Now consider the same experiment with uniform quantizers in place of the additive noise model. Figure 4-2 displays the simulated reduction factor γ as a function of τ_1 and τ_2. It is encouraging that the additive noise model matches the quantizer well over the rest of the surface. The results reveal a large reduction in performance along the "groove" where the sampling instants of two channels fall in close proximity to each other. In this chapter, we explain and quantify this phenomenon.

4.2 Correlation Analysis

Under the additive noise model, quantizer error is independent from one sample to the next. This is the case even when two or more channels sample at exactly the same instant in time. In reality, samples of a bandlimited signal obtained near the same instant in time are correlated. Correlation between samples obtained with a particular set of timing offsets is dependent on the autocorrelation of the original signal. This correlation tends to increase when a signal is highly oversampled.

Often correlation between quantizer inputs results in correlation between their respective error residue signals. As an example, consider the case when the input $x(t) = 1 \ \ \forall t$ and all quantizers have the same granularity. In this case, the quantizer inputs are completely correlated, causing the deterministic error between channels to be completely correlated as well. It is clear that we cannot combine measurements in this case to obtain a better estimate of the analog value, because all observations contain exactly the same information.

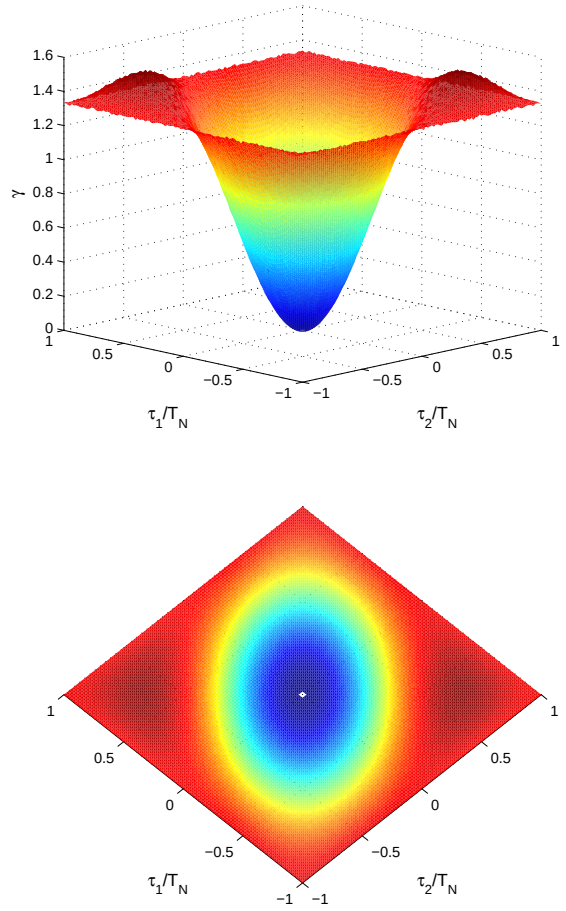


Figure 4-1: Reduction factor γ as a function of delays τ_1 and τ_2 with the additive noise model for $\boldsymbol{b} = [10,\ 10,\ 10]$.

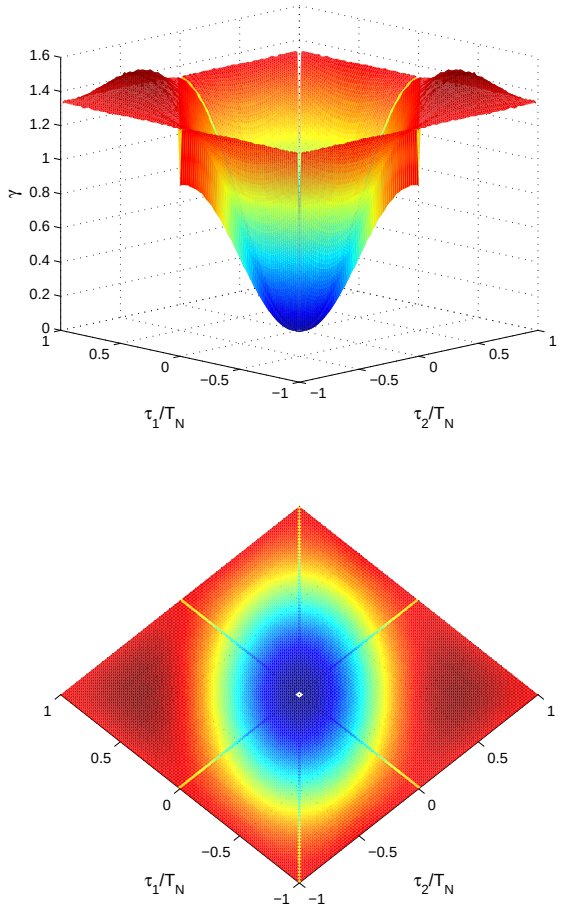


Figure 4-2: Reduction factor γ as a function of delays τ_1 and τ_2 with uniform quantizers for $\boldsymbol{b} = [10,\ 10,\ 10]$.

Correlation between quantization error in different channels is not accounted for in the additive noise model. In this section, we quantify the effects of correlation between quantization error in different channels.

4.2.1 Correlation between Quantizer Inputs in Different Channels

In a time-interleaved ADC, correlation between quantizer inputs in different channels is dependent on the continuous-time autocorrelation of the input signal and the timing offsets τ_m. We will examine different cases that lead to strong correlation between input channels.

For a bandlimited signal $x(t)$, correlation between quantizer inputs tends to increase for sampling

instants in close proximity. This may occur through specification of several timing offsets very close to the same value or through oversampling. As the oversampling ratio increases, samples become more finely spaced. Since the signal is bandlimited, correlation between quantizer input samples will also increase, approaching complete correlation as the oversampling factor approaches infinity.

To illustrate this effect, we will sample an approximately bandlimited white Gaussian random process with three channels ($M = 3$) at half of the Nyquist rate ($L = 2$) in each channel, corresponding to an overall oversampling ratio of $\frac{M}{L} = \frac{3}{2}$. Beginning with uniform timing offsets $\tau_0 = -\frac{2}{3}T_N$, $\tau_1 = 0$, and $\tau_2 = \frac{2}{3}T_N$, we reduce τ_2 so it approaches $\tau_1 = 0$. Intuitively, one expects the correlation coefficient between the quantizer inputs in channels 1 and 2, denoted $\rho_{1,2}$, to increase as the sampling instants draw nearer to each other. The correlation coefficient is defined as:

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}. \tag{4.2}$$

Figure 4-3 displays $\rho_{1,2}$, the correlation coefficient of the unquantized input samples in channels 1 and 2, as a function of the time difference between channels in units of the Nyquist sampling period T_N. The results of the same experiment with an audio signal are also included. Note that the Gaussian process has a slightly lower correlation coefficient for a given time difference. This can be explained by considering the autocorrelation functions of the two signals. The autocorrelation of the approximately bandlimited white Gaussian random process takes the form of a sinc. The audio signal is not white, and therefore contains additional correlation between samples.

From Figure 4-3, it is apparent that channel time differences of less than $0.1 \cdot T_N$ will have an input correlation coefficient approximately adherent to $\rho \geq 0.984$ for the configuration defined above. Differences of less than $0.05 \cdot T_N$ will have an input correlation coefficient approximately adherent to $\rho \geq 0.996$. This insight will allow us to quantify performance mismatches in future sections.

The same phenomenon occurs when the signal is oversampled. Essentially this means that the Nyquist rate $1/T_N$ has been artificially increased beyond the necessary sampling frequency for a given signal. We can parameterize this scenario by defining the artificial Nyquist rate as $\frac{\beta}{T_N}$ and sampling according to equation (2.1), where β is the oversampling parameter. As the oversampling factor increases, the correlation coefficient between quantizer inputs will approach unity, which indicates nearly identically varying samples.

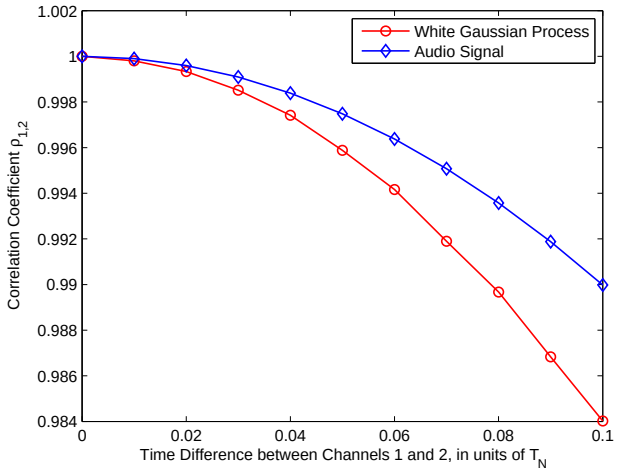


Figure 4-3: Measured correlation coefficient of the unquantized input samples $\rho_{1,2}$ as a function of time difference between channels 1 and 2 for approximately bandlimited Gaussian white noise and audio signal.

4.2.2 Correlation between Quantization Error in Different Channels

Often, correlation between quantizer inputs results in correlation between their respective error residue signals. The amount of correlation is dependent on the quantization granularity in each channel. In [24], Widrow shows that the correlation between the error signals that result from quantizing with different numbers of bits is usually quite small, and can be ignored. However, we will present some notable exceptions.

Correlation between quantization error in different channels tends to be most noticeable when the quantizer inputs are highly correlated and the difference in the number of bits in each quantizer is small. Each additional bit in the quantizer measurement corresponds to an exponential increase in the complexity of the input-output characteristic. We consider the case of highly correlated input samples for channels that differ by zero or one bit. To experimentally verify our earlier statements, consider a two channel system with b_0 and b_1 bit quantizers in each channel, respectively. To simulate a highly correlated input signal, we generate samples $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ from a multivariate Gaussian distribution with covariance matrix $\mathbf{\Sigma}$. This is accomplished through a Cholesky decomposition, described in section 4.3.1.

In particular, we increase the correlation between the quantizer inputs in two channels and measure the resulting correlation between their error residue signals for various quantizer config-

urations. First, consider the case where both channels have identical quantizers with the same number of bits. Intuitively, one expects that as the input correlation coefficient $\rho_{in:0,1}$ approaches unity the quantization error correlation coefficient $\rho_{error:0,1}$ will also approach unity in this case. In other words, quantizing the same sample with identical quantizers should yield identical error samples.

In [24], Widrow derives an approximation for the covariance between error signals when the two quantizers have the same granularity, denoted Δ. The inputs are assumed to be two Gaussian variables with zero mean, variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and correlation coefficient ρ_{x_1,x_2}. For positively correlated input samples, the covariance can be approximated by equation (4.3):

$$\mathrm{cov}(e_1, e_2) \approx \frac{\Delta^2}{2\pi^2} \sum_{l=1}^{\infty} \frac{e^{-4\pi^2\left((1+\rho_{x_1,x_2})l^2 \frac{\sigma^2}{\Delta^2}\right)}}{l^2}. \tag{4.3}$$

When $\rho_{x_1,x_2} = 1$, this expression yields the equation:

$$\mathrm{cov}(e_1, e_2) \approx \frac{\Delta^2}{12}. \tag{4.4}$$

Normalizing by the variance to obtain the correlation coefficient, we obtain the intuitive result that the correlation coefficient for the error signals is equal to one when $\rho_{x_1,x_2} = 1$.

Figure 4-4 displays the measured correlation coefficient of the quantization error signal $\rho_{error:0,1}$ as a function of the correlation coefficient of quantizer input samples $\rho_{in:0,1}$ for $\boldsymbol{b} = [3,\ 3]$ bits, $\boldsymbol{b} = [4,\ 4]$ bits, and $\boldsymbol{b} = [5,\ 5]$ bits. Note that considerably higher quantizer input correlation is necessary to observe a fixed level of error correlation for more precise quantizers. For the $\boldsymbol{b} = [3,\ 3]$ bit configurations, $\rho_{error:0,1} = 0.1$ at $\rho_{in:0,1} = 0.997$. For the $\boldsymbol{b} = [4,\ 4]$ and $\boldsymbol{b} = [5,\ 5]$ bit configurations, the same level is reached for $\rho_{in:0,1} = 0.9991$ and $\rho_{in:0,1} = 0.9998$, respectively. This is a result of the exponentially increasing complexity of the quantizer error function as the number of bits increases.

Some correlation between error signals is present even if the quantizers are not identical. As we will derive in Chapter 5, the correlation coefficient for two quantizers with k bits and $k+n$ bits that operate on the same sample is $\rho = -2^{-n-1}$. Since the correlation decays so quickly, we will consider only the case when the number of bits differs by one.

Figure 4-5 displays the measured correlation coefficient of the quantization error signal $\rho_{error:0,1}$

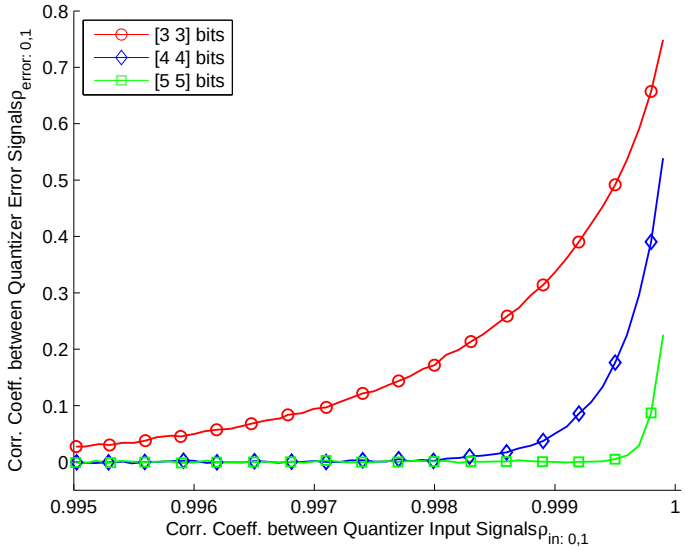


Figure 4-4: Measured correlation coefficient of quantization error signal $\rho_{error:0,1}$ as a function of correlation coefficient of quantizer input samples $\rho_{in:0,1}$ for identical quantizers in each channel. Input signal is multivariate Gaussian.

as a function of the correlation coefficient of quantizer input samples $\rho_{in:0,1}$ for $\boldsymbol{b} = [3,\ 4]$ bits, $[4,\ 5]$ bits, and $[5,\ 6]$ bits. Considerably higher input correlation is necessary to observe a fixed level of error correlation, especially with small quantizer granularities. Note that the measured values of $\rho_{error:0,1}$ approach $-1/4$, the predicted value mentioned earlier.

4.2.3 System Performance with Quantization Error Correlation

As described in the previous section, correlation between quantization error in each channel is not necessarily negligible in time-interleaved ADCs. In this section, we discuss the ramifications of the presence of error correlation on quantization error reduction. This correlation has the potential to enhance quantization error reduction capability or hinder it, depending on the configuration of quantizers.

4.2.3.1 Close Proximity of Quantizers with the Same Granularity

We have shown that when channels with identical quantizers sample in close proximity, the correlation coefficient of their error signals will approach unity. Intuitively, this positive error correlation is expected to hinder error reduction capability because less information is available about the

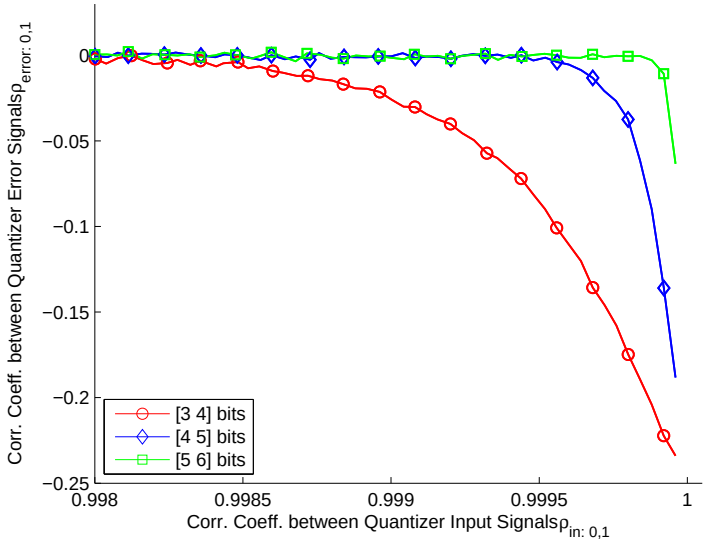


Figure 4-5: Measured correlation coefficient of quantization error signal $\rho_{error:0,1}$ as a function of correlation coefficient of quantizer input samples $\rho_{in:0,1}$ for different quantizers in each channel.

underlying signal. We will experimentally investigate these effects for $M = 3$, $L = 2$, and bit allocation $\boldsymbol{b} = [3,\ 3,\ 3]$. The timing offsets will be defined as $\tau_0 = -\beta T_N$, $\tau_1 = 0$, and $\tau_2 = \frac{2}{3}T_N$, so that β is a positive real number that defines the time difference between channels 0 and 1.

Figure 4-6 displays the predicted reduction factor γ based on the additive noise model and the measured value with uniform quantizers as a function of the time difference between channels 0 and 1 in units of T_N for this system. We observe that the mismatch between predicted and measured γ grows larger as the time difference approaches zero, because of the error correlation described in the previous section.

To quantify the change in performance, we measure the time difference for which the predicted and measured values of γ differ by ± 0.1dB. Equation (4.5) defines the decibel measure. For example, for $\boldsymbol{b} = [3,\ 3,\ 3]$ the ± 0.1dB point is observed when the time difference between channels 0 and 1 is $-0.047 \cdot T_N$.

$$L_{dB} = 10 \cdot \log_{10}\left(\frac{\gamma_{predicted}}{\gamma_{measured}}\right) \tag{4.5}$$

Of course, the change in overall performance resulting from correlation in two channels is also dependent on the properties of the remaining channels. When the two channels in close proximity

produce very little error (because of high-precision quantizers) compared to the third channel, the overall change in performance will be small. This is because the correlated error sources with high precision did not contribute much to the overall error signal to begin with. When the two channels in close proximity have relatively large error variance compared to the third channel, the overall change in performance will be greatest. The two correlated channels now dominate the overall error signal, and any effects of correlation will not be mitigated.

Consider the same system for $\boldsymbol{b} = [3,\ 3,\ 4]$. In this case, the error variance of the first two channels is considerably larger than the error variance of the third channel. As shown in Figure 4-6, the performance reduction is observed slightly earlier, with a ± 0.1dB point of $-0.052 \cdot T_N$. When $\boldsymbol{b} = [4,\ 4,\ 3]$, however, the opposite occurs. The error variance of the first two channels is much smaller than the third channel, so they do not contribute as much to the overall performance reduction factor. The ± 0.1dB point for this case is $-0.016 \cdot T_N$.

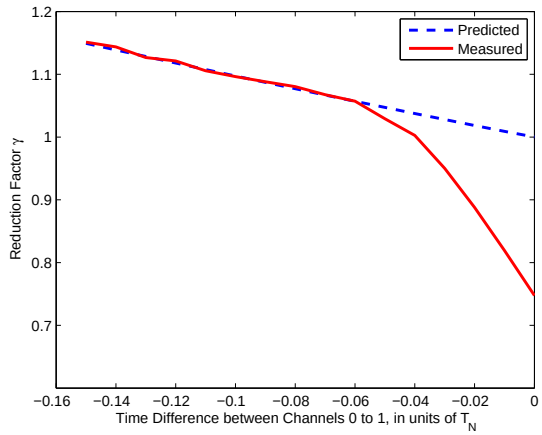
These properties are dependent on the autocorrelation of the input signal and choice of timing offsets. Generally, as the number of bits increases, much smaller time differences are necessary to observe a fixed change in performance. For example, when $\boldsymbol{b} = [4,\ 4,\ 4]$ the ± 0.1dB point drops to $-0.023 \cdot T_N$. Table 4.1 lists the time differences that produce a ± 0.1dB deviation between the predicted and measured value of γ for the bit allocations we have considered.

b_0	b_1	b_2	± 0.1dB Time Diff.
3	3	3	$-0.047 \cdot T_N$
3	3	4	$-0.052 \cdot T_N$
4	4	3	$-0.016 \cdot T_N$
4	4	4	$-0.023 \cdot T_N$

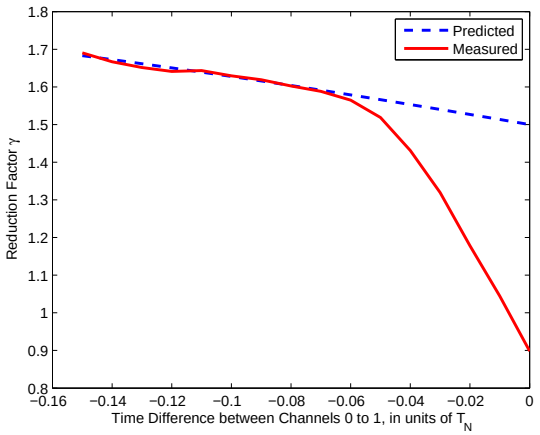
Table 4.1: Time differences between identical quantizers producing a ± 0.1dB deviation between the predicted and measured value of γ.

When the quantizer corresponding to the third channel has much lower resolution than the quantizers in the first two channels that have correlation, it will dominate the overall error signal. Often the effects of correlation will be diluted to the point where they are negligible by the large differences in resolution. Figure 4-7 displays the predicted and measured reduction factor γ for the bit allocation $\boldsymbol{b} = [8,\ 8,\ 3]$. The third channel dominates the overall error signal variance, causing the effects of correlation in the first two channels to be negligible.

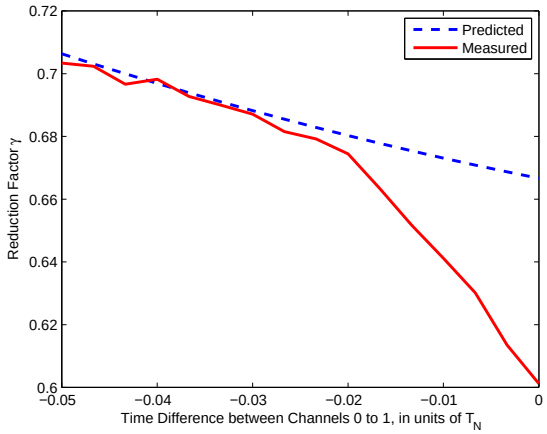
To illustrate that these properties extend to more than three channels, we consider the case



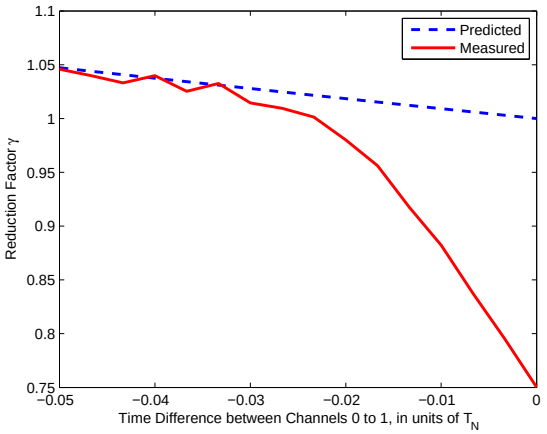
$\boldsymbol{b} = [3,\ 3,\ 3]$



$\boldsymbol{b} = [3,\ 3,\ 4]$



$\boldsymbol{b} = [4,\ 4,\ 3]$



$\boldsymbol{b} = [4,\ 4,\ 4]$

Figure 4-6: Predicted reduction factor γ based on additive noise model and measured value with uniform quantizers as a function of time difference between channels 0 and 1 in units of T_N for various bit allocations.

where $M = 9$ and $L = 6$ with timing offsets $\boldsymbol{\tau} = [-\frac{8}{3} \cdot T_N, -\frac{6}{3} \cdot T_N, -\frac{4}{3} \cdot T_N, -\frac{2}{3} \cdot T_N, \tau_5 \cdot T_N, \frac{2}{3} \cdot T_N, \frac{4}{3} \cdot T_N, \frac{6}{3} \cdot T_N, \frac{8}{3} \cdot T_N]$. The bit allocation is defined as $\boldsymbol{b} = [10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10]$. This corresponds to a uniform sampling grid, except that τ_5 is variable.

Figure 4-8 displays the predicted and measured reduction factor γ as a function of τ_5. The discrepancies in performance confirm our earlier conclusions. When τ_5 is close to another timing offset, positive error correlation causes the reduction factor γ to decrease.

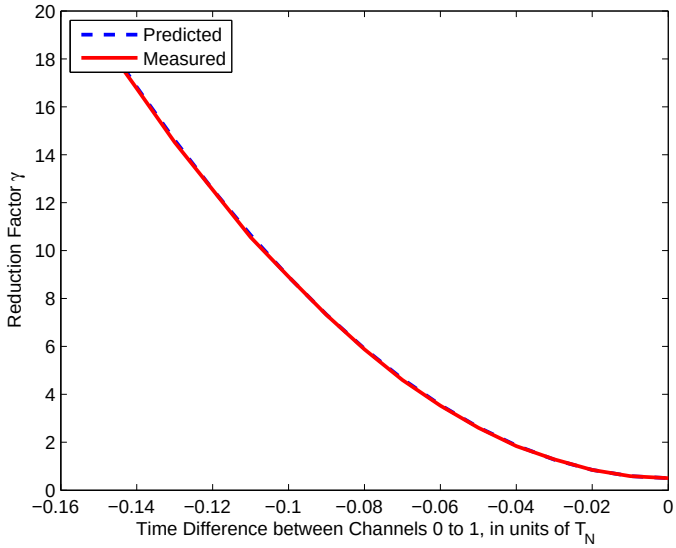


Figure 4-7: Predicted and measured reduction factor γ as a function of time difference between channels 0 and 1 in units of T_N for $\boldsymbol{b} = [8,\ 8,\ 3]$.

4.2.3.2 Close Proximity of Quantizers with Different Granularity

Negative correlation between error signals occurs when quantizers with differing granularities have sampling instants in close proximity. This was shown experimentally in the previous section and will be derived formally in Chapter 5. Using the same experimental system as in the previous section, we measure the $\pm$0.1dB points for $\boldsymbol{b} = [3,\ 4,\ 3]$ and $\boldsymbol{b} = [3,\ 4,\ 4]$. The same general properties hold here. For the former bit allocation, the negative error correlation from the first two channels will be somewhat "diluted" by the high error variance of the third channel. For the latter bit allocation, the third channel has lower variance, so we expect that the sampling instants will be more separated to achieve a fixed deviation in performance. This intuition is verified in Table 4.2.

b_0	b_1	b_2	$\pm$0.1dB Time Diff.
3	4	3	$-0.010 \cdot T_N$
3	4	4	$-0.019 \cdot T_N$

Table 4.2: Time differences between quantizers differing by one bit that produce a $\pm$0.1dB deviation between the predicted and measured value of γ.

Figure 4-9 displays the predicted and measured reduction factor γ as a function of time difference between channels 0 and 1 in units of T_N for the bit allocations $\boldsymbol{b} = [3,\ 4,\ 3]$ and $\boldsymbol{b} = [3,\ 4,\ 4]$.

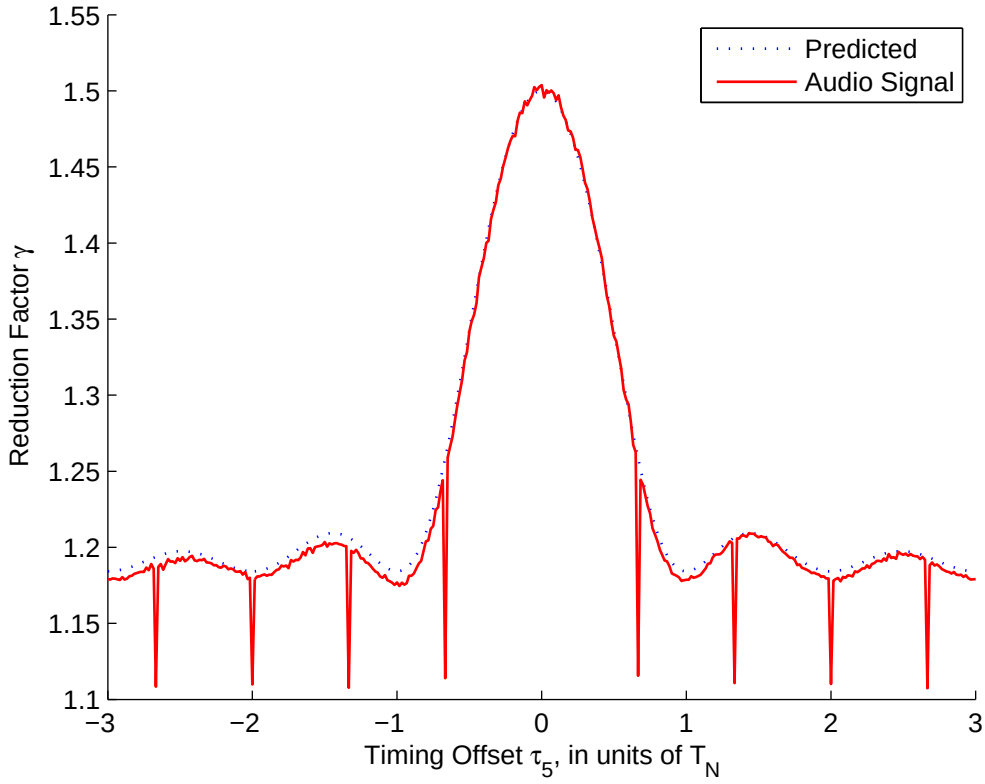


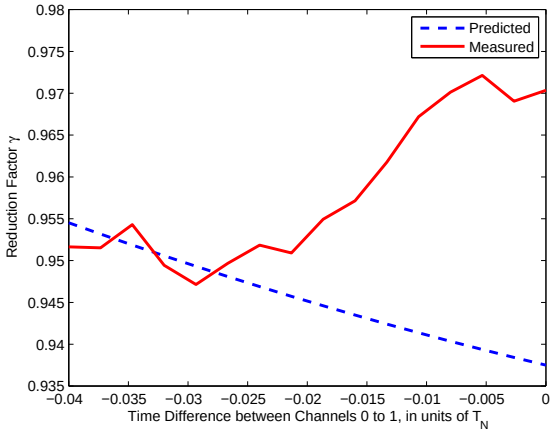
Figure 4-8: Predicted and measured reduction factor γ for $M = 9$ and $L = 6$ as a function of τ_5.

The results reveal that the negative error correlation between quantizers that differ by one bit may actually improve error reduction capability. This is a surprising result that we will consider further in Chapter 5.

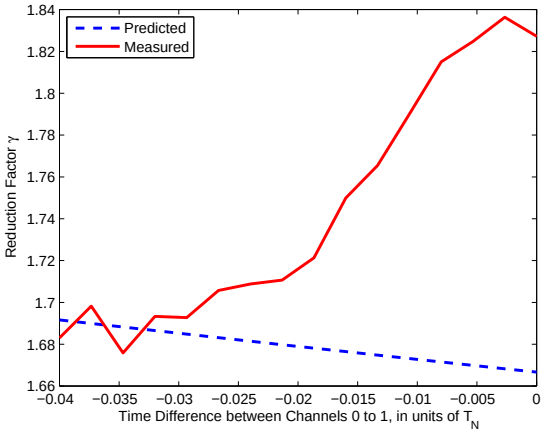
For example, consider the system where $M = 3$, $L = 2$, $\boldsymbol{b} = [10,\ 10,\ 9]$. Figure 4-10 and 4-11 display the reduction factor γ for the additive noise model and ideal quantizers as a function of τ_1 and τ_2 with $\tau_0 = 0$. As the channel with 9 bits draws near each of the 10 bit channels, the negative error correlation results in increased error reduction γ.

To illustrate that these properties extend to more than three channels, we again consider the nine channel scenario mentioned in the previous section. We modify the bit allocation to be $\boldsymbol{b} = [9,\ 9,\ 9,\ 9,\ 10,\ 11,\ 11,\ 11,\ 11]$. The timing offsets τ_m correspond to a uniform sampling grid, except that τ_5 is variable.

Figure 4-12 displays the predicted and measured reduction factor γ as a function of τ_5. The discrepancies in performance again confirm our earlier conclusions. As the fifth channel with 10



$\boldsymbol{b} = [3,\ 4,\ 3]$



$\boldsymbol{b} = [3,\ 4,\ 4]$

Figure 4-9: Predicted and measured reduction factor γ as a function of time difference between channels 0 and 1 in units of T_N for various bit allocations.

bits draws near each of the 9 bit channels, the negative error correlation results in increased error reduction γ. Similarly, γ is improved by sampling near the 11 bit channels. The improvement resulting from sampling near the 9 bit channels is noticeably greater than the improvement from sampling near the 11 bit samples. As we stated earlier, the overall error signal is more heavily dependent on the 9 bit channels because they have higher variance. Any correlation between those channels will have a greater influence on the overall error variance.

4.2.3.3 Acceptable Time Differences to Avoid Performance Degradation

In this section, we formulate some general guidelines for designing time-interleaved ADCs that avoid positive error correlation. We have determined that the quantizer error correlation caused by close proximity between sampling instants is most pronounced when the remaining channel quantizers have relatively high precision. In other words, the largest deviation from the additive noise model occurs when the two channels have coarse quantization granularity and the remaining channels have fine quantization granularity.

Thus, we will examine the bit allocations $\boldsymbol{b} = [3,\ 3,\ 10]$, $\boldsymbol{b} = [4,\ 4,\ 10]$, up to $\boldsymbol{b} = [7,\ 7,\ 10]$ to determine acceptable time differences for these "worst-case" configurations for the case of an approximately bandlimited white Gaussian noise input signal. Table 4.3 lists the time differences that produce a ± 0.1dB deviation between the predicted and measured value of γ. When both

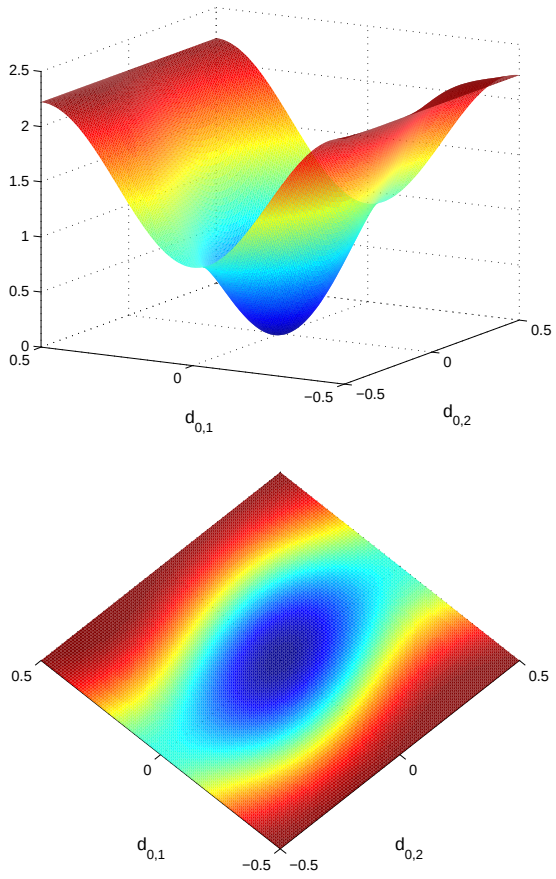


Figure 4-10: Reduction factor γ as a function of delays τ_1 and τ_2 with the additive noise model for $\boldsymbol{b} = [10,\ 10,\ 9]$.

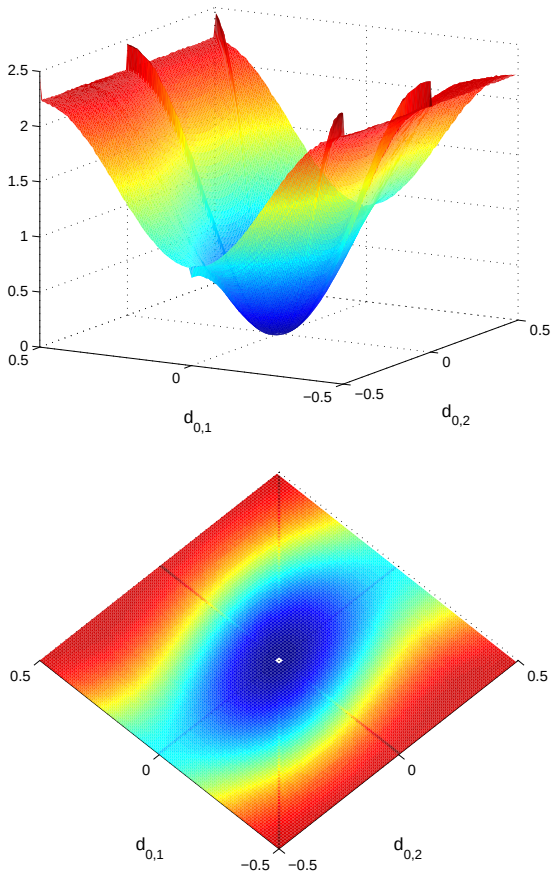


Figure 4-11: Reduction factor γ as a function of delays τ_1 and τ_2 with uniform quantizers for $\boldsymbol{b} = [10,\ 10,\ 9]$.

channels of interest have 3 bits, the acceptable time difference is $-0.0547 \cdot T_N$. As the number of bits increases by one, the acceptable time difference is approximately halved. As discussed earlier, the complexity of the quantizer input-output characteristic increases exponentially as additional bits are added. Note that these cases provide an acceptable bound for all other configurations where the third channel has lower resolution. The lower resolution will simply dilute the effects of correlation.

If the number of bits for the channels in close proximity differs by one (or more), error reduction performance will be improved. As shown in Chapter 5, the correlation coefficient between quantizer error signals in this case is always negative. Thus, we will not consider this case in developing acceptable time differences to avoid performance degradation. In some cases, the slight improvement

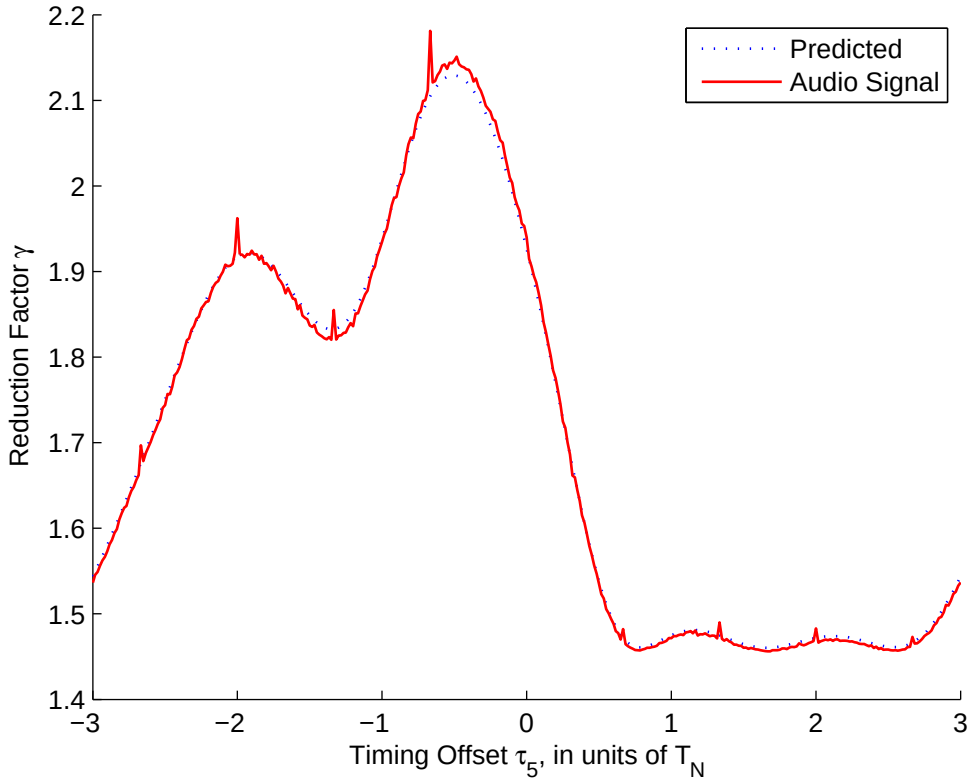


Figure 4-12: Predicted and measured reduction factor γ for $M = 9$ and $L = 6$ as a function of τ_5.

that results from "unifying" quantizers with different granularities is outweighed by the benefit of placing the two channels further apart. In general, systems with diverse bit allocations will gain from unifying quantizers.

b_0	b_1	b_2	$\pm$0.1dB Time Diff.
3	3	10	$-0.0547 \cdot T_N$
4	4	10	$-0.0274 \cdot T_N$
5	5	10	$-0.0136 \cdot T_N$
6	6	10	$-0.0070 \cdot T_N$
7	7	10	$-0.0038 \cdot T_N$

Table 4.3: Acceptable time differences that produce a $\pm$0.1dB deviation between the predicted and measured value of γ.

We conclude that for this scenario, a time difference of $0.06 \cdot T_N$ is acceptable to avoid performance changes due to correlation for quantizers with three or more bits and an approximately bandlimited white Gaussian input signal. As shown in Table 4.3, this is a very conservative estimate

for cases with more than three bits. It is important to note that these calculations were carried out for a specific bit allocation with an input that is approximately bandlimited white Gaussian noise. The acceptable time differences may need to be increased if the autocorrelation of the input signal differs significantly from the autocorrelation of a white random process.

4.3 Extension of Noise Model: Correlated Noise

In the previous section, we explored the effects of correlation in quantization error signals and acceptable distances to avoid performance mismatches with the additive white noise model. In this section, we extend the traditional noise model for quantization to model correlation between channels. In particular, we now model quantization error as a multivariate Gaussian with zero mean and covariance matrix $\mathbf{\Sigma}$. The entries of the Gaussian covariance matrix $\mathbf{\Sigma}$ will be defined as the measured covariance between channels.

Although the Gaussian PDF is not always a good model for quantization error, we show that the simulations performed with this revised noise model match observed phenomenon much more closely than the additive noise model. To generate samples from a multivariate Gaussian with arbitrary covariance matrix, we employ a method known as Cholesky decomposition.

4.3.1 Cholesky Decomposition

To model correlated quantization error, we generate samples $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ from the n-dimensional multivariate Gaussian distribution $\mathbf{X} \sim N(0, \Sigma)$. Let $\mathbf{C}$ be a $(n \times n)$ matrix and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ where Z_i is defined as the standard normal random variable $Z \sim N(0, 1)$. We can express the desired distribution of $\mathbf{X}$ as shown in equation (4.6), because a linear combination of i.i.d. Gaussian random variables is also Gaussian.

$$\mathbf{C}^T\mathbf{Z} \sim N\left(0, \mathbf{C}^T\mathbf{C}\right) \tag{4.6}$$

Thus, we must find the matrix C that satisfies equation (4.7).

$$\mathbf{C}^T\mathbf{C} = \Sigma \tag{4.7}$$

As shown in [10], any symmetric positive-definite matrix $\mathbf{M}$ may be decomposed into an upper triangular matrix $\mathbf{U}$ and a diagonal matrix with positive elements $\mathbf{D}$.

$$\mathbf{M} = \mathbf{U}^T\mathbf{D}\mathbf{U} \tag{4.8}$$

Using this relationship, the symmetric positive-definite covariance matrix $\mathbf{\Sigma}$ can be written as

$$\begin{aligned}\mathbf{\Sigma} &= \mathbf{U}^T\mathbf{D}\mathbf{U} \\ &= (\mathbf{U}^T\sqrt{\mathbf{D}})(\sqrt{\mathbf{D}}\mathbf{U}) \\ &= (\sqrt{\mathbf{D}}\mathbf{U})^T(\sqrt{\mathbf{D}}\mathbf{U}).\end{aligned} \tag{4.9}$$

The matrix $\mathbf{C} = \sqrt{\mathbf{D}}\mathbf{U}$ represents the Cholesky decomposition of $\mathbf{\Sigma}$. The matrix multiplication $\mathbf{C}^T\mathbf{Z}$ is then used to easily generate a sequence of samples with the desired covariance matrix $\mathbf{\Sigma}$.

4.3.2 Experimental Results with Correlated Noise Model

To experimentally test the results of the revised noise model, we consider a system with $M = 3$ and $L = 2$ operating on approximately bandlimited Gaussian white noise. The timing offsets are chosen to be $\boldsymbol{\tau} = [-\frac{\frac{2}{3}\cdot T_N}{\beta}, 0, \frac{\frac{2}{3}\cdot T_N}{\beta}]$ with recurrence period $\frac{L\cdot T_N}{\beta}$ and bit allocation $\boldsymbol{b} = [3,\ 3,\ 3]$. This corresponds to oversampling the bandlimited input signal by a factor of β. We will consider the case where $\beta = 20$, making the timing difference between channels $\frac{1}{30}\cdot T_N$. The timing differences do not satisfy acceptable levels to avoid correlation.

Note that we are essentially creating an artificial Nyquist rate with oversampling parameter β. This is outside of the scope of the original problem formulation in [18], and the reconstruction filters are not necessarily optimal in this case. In fact, if the oversampling factor is high enough, it may be possible to reconstruct the signal from a single channel. Thus, we will not use the results of this particular experiment in our future discussion of optimal noise reduction performance. We use this experiment only to evaluate the accuracy of the correlated noise model compared with the additive noise model and uniform quantizers.

Table 4.4 displays the measured average quantizer error variance, output error variance, and reduction factor γ for uniform quantizers and the two noise models. The system with uniform quantizers performs poorly as a result of error cross-correlation. This is not accounted for in the

additive noise model, which erroneously leads us to conclude the system will perform quite well. The covariance matrix $\mathbf{\Sigma}$ for the correlated noise model was constructed using measured covariances obtained from uniform quantizers. Although the results based on the correlated noise model do not match the uniform quantizer exactly, they are considerably closer than those based on the traditional noise model.

The remaining discrepancy between the correlated noise model and uniform quantizers is likely caused by differences in the error PDF. The marginal PDF of each quantization error stream is usually assumed to be uniform, not Gaussian. Nevertheless, it is clear that the correlated Gaussian error model provides a better fit than the traditional additive error model when dependence is present between error samples. The correlated noise model gives us insight into performance of the time-interleaved ADC when the traditional noise model cannot.

	Measured $\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2$	Measured $\sigma_{e_{min}}^2$	Measured γ
Uniform quantizer	0.00520	0.00403	1.291
Additive white noise model	0.00521	0.00347	1.500
Correlated noise model	0.00521	0.00385	1.353

Table 4.4: Comparison of uniform quantizer error, additive white noise model, and correlated noise model for an approximately bandlimited Gaussian input.

Table 4.5 displays the results of the same experiment with an audio signal input. In this case, $\beta = 5$ and the bit allocation is $\boldsymbol{b} = [4,\ 4,\ 4]$. Again, the correlated noise model provides a better representation of the effects of quantization error than the additive noise model.

	Measured $\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2$	Measured $\sigma_{e_{min}}^2$	Measured γ
Uniform quantizer	0.0011	0.0009	1.152
Additive white noise model	0.0013	0.0009	1.503
Correlated noise model	0.0011	0.0009	1.229

Table 4.5: Comparison of uniform quantizer error, the additive white noise model, and the correlated noise model for an audio input signal.

Chapter 5

Exploiting Negative Correlation in Quantization Error Signals

In the previous chapter, we examined the differences between the additive noise model and the uniform quantizer error signal. When two sampling instants associated with different quantizer granularities have nearly the same value, we observed that there is often an unpredicted performance gain. We concluded that this was due to negative correlation between error signals. This chapter explores why negative correlation between error signals occurs, and how to design a quantizer that exploits this negative correlation to achieve higher effective resolution.

5.1 Uniform Quantizer Error Function

A quantizer can be conveniently expressed as a nonlinear operator. The input-output staircase relationship of a basic uniform, mid-tread quantizer is shown in Figure 5-1, with step size Δ. Mid-tread quantizers reserve a quantization level at zero. For the purposes of this section, we will assume that the input does not saturate the quantizer.

Recall that the quantization error is defined as the quantized signal minus the analog input $e[n] = \hat{x}[n] - x[n]$. Increasing the number of bits by one will double the number of levels and halve the maximum quantization error, as shown in Figure 5-2.

The quantization input-output error relationship is a deterministic non-linear function. For a quantizer with k bits and step size $\Delta = 2^{-(k-1)}$, the error signal is "wrapped" to the interval

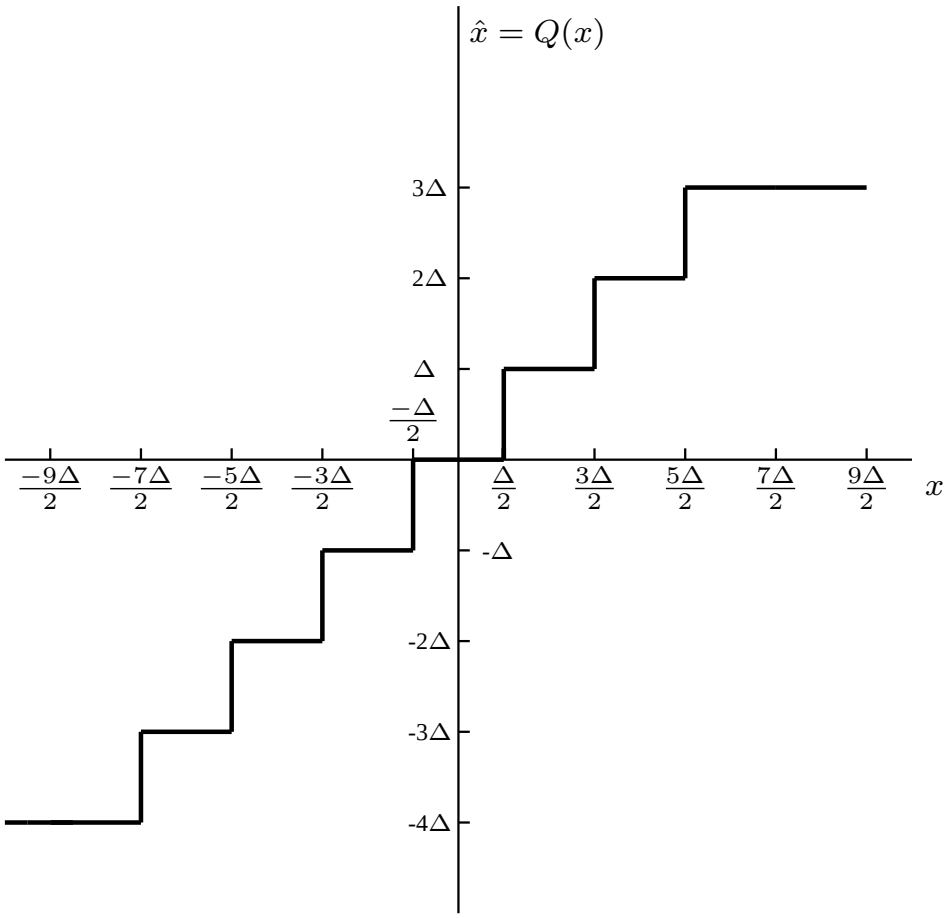


Figure 5-1: A basic uniform, mid-tread, rounding quantizer input-output characteristic.

$\pm\frac{\Delta}{2}$. We will consider the non-linear error function introduced by quantizing the same sample with different granularities.

First, we model the quantization error for a k bit quantizer with a random variable E_k uniformly distributed between $-\frac{\Delta_k}{2}$ and $\frac{\Delta_k}{2}$, as suggested by the additive noise model. The error for a sample quantized with k bits is deterministically related to the error for the same sample quantized with $k+1$ bits, denoted E_{k+1}. The non-linear error function $E_{k+1}(e_k)$ for a sample value of E_k can be determined by "wrapping" to $\pm\frac{\Delta_{k+1}}{2}$.

$$E_{k+1}(e_k) = \begin{cases} e_k + \Delta_{k+1} & \text{if } -\Delta_{k+1} \le e_k < -\frac{\Delta_{k+1}}{2} \\ e_k & \text{if } -\frac{\Delta_{k+1}}{2} \le e_k < \frac{\Delta_{k+1}}{2} \\ e_k - \Delta_{k+1} & \text{if } \frac{\Delta_{k+1}}{2} \le e_k < \Delta_{k+1} \\ \text{undefined} & \text{otherwise} \end{cases} \tag{5.1}$$

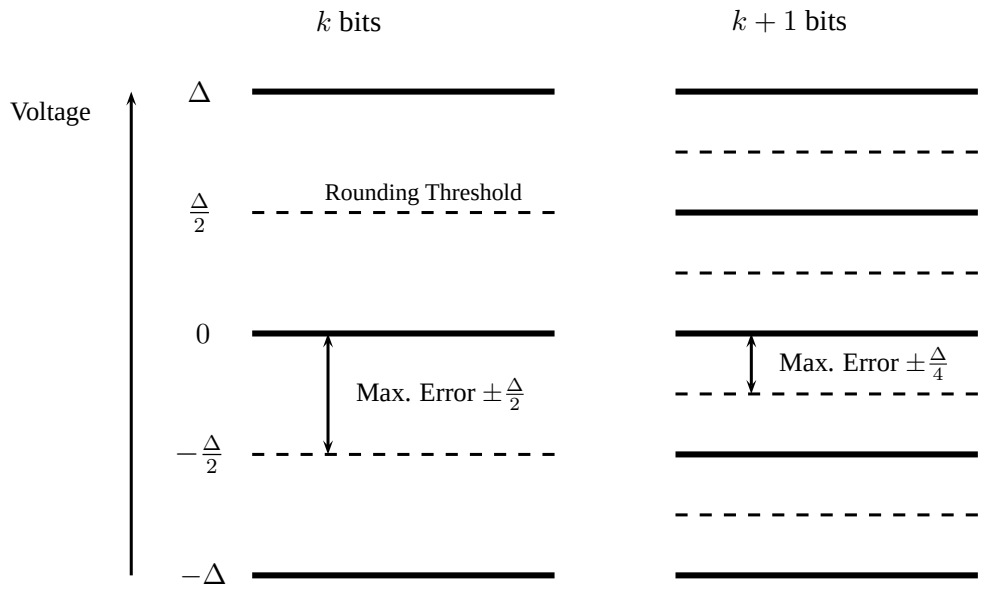


Figure 5-2: Quantization regions and rounding thresholds (dashed lines) for k and $k+1$ bits.

The error signals for k bits and $k+n$ bits are also deterministically related. Again, we assume that E_k is uniformly distributed. To find the function $E_{k+2}(e_k)$, we first map the k bit error to $k+1$ bits using equation (5.1) above. Next, we take the composition of this function with $E_{k+2}(e_{k+1})$, which can easily be derived from equation (5.1). This operation corresponds to wrapping the error signal twice.

$$E_{k+2}(e_k) = \begin{cases} e_k + 2\Delta_{k+2} & \text{if } -2\Delta_{k+2} \le e_k < -\frac{3\Delta_{k+2}}{2} \\ e_k + \Delta_{k+2} & \text{if } -\frac{3\Delta_{k+2}}{2} \le e_k < -\frac{\Delta_{k+2}}{2} \\ e_k & \text{if } -\frac{\Delta_{k+2}}{2} \le e_k < \frac{\Delta_{k+2}}{2} \\ e_k - \Delta_{k+2} & \text{if } \frac{\Delta_{k+2}}{2} \le e_k < \frac{3\Delta_{k+2}}{2} \\ e_k - 2\Delta_{k+2} & \text{if } \frac{3\Delta_{k+2}}{2} \le e_k < 2\Delta_{k+2} \\ \text{undefined} & \text{otherwise} \end{cases} \tag{5.2}$$

The relationship for $E_{k+3}(e_k)$ can be derived in a similar procedure, corresponding to a triple wrapping of the original uniformly distributed error signal. In general, the relationship between E_{k+n} and E_k can be found by "wrapping" the original error function n times. Note that each linear segment has a positive slope equal to Δ_{k+n}.

$$E_{k+3}(e_k) = \begin{cases} e_k + 4\Delta_{k+3} & \text{if } -4\Delta_{k+3} \le e_k < -\frac{7\Delta_{k+3}}{2} \\ e_k + 3\Delta_{k+3} & \text{if } -\frac{7\Delta_{k+3}}{2} \le e_k < -\frac{5\Delta_{k+3}}{2} \\ e_k + 2\Delta_{k+3} & \text{if } -\frac{5\Delta_{k+3}}{2} \le e_k < -\frac{3\Delta_{k+3}}{2} \\ e_k + 1\Delta_{k+3} & \text{if } -\frac{3\Delta_{k+3}}{2} \le e_k < -\frac{\Delta_{k+3}}{2} \\ e_k & \text{if } -\frac{\Delta_{k+3}}{2} \le e_k < \frac{\Delta_{k+3}}{2} \\ e_k - 1\Delta_{k+3} & \text{if } \frac{\Delta_{k+3}}{2} \le e_k < \frac{3\Delta_{k+3}}{2} \\ e_k - 2\Delta_{k+3} & \text{if } \frac{3\Delta_{k+3}}{2} \le e_k < \frac{5\Delta_{k+3}}{2} \\ e_k - 3\Delta_{k+3} & \text{if } \frac{5\Delta_{k+3}}{2} \le e_k < \frac{7\Delta_{k+3}}{2} \\ e_k - 4\Delta_{k+3} & \text{if } \frac{7\Delta_{k+3}}{2} \le e_k < 4\Delta_{k+3} \\ \text{undefined} & \text{otherwise} \end{cases} \tag{5.3}$$

These deterministic relationships are displayed in Figure 5-4. Note that as the difference in bits n increases, more diagonal linear segments appear. We will separate these into three categories. The "base" region includes the linear segment centered around zero, from $-\frac{\Delta_{k+n}}{2}$ to $+\frac{\Delta_{k+n}}{2}$. The "middle" regions on each side contain $2^{n-1}-1$ shifted linear segments. Finally, the two "outermost" regions extend to $\pm 2^{n-1}\Delta_{k+n}$ with length $\frac{\Delta_{k+n}}{2}$. The outermost regions have half the length of the base and middle regions.

5.2 Statistical Analysis of Correlation between Quantizer Error Signals

In this section, we consider the statistical properties of the error resulting from quantizing the *same* input sample with different granularities. Specifically, we consider the covariance and correlation coefficient of E_k and E_{k+n}. We begin with the definition of covariance in equation (5.4). E_k is a zero-mean uniform random variable, causing the second term to be zero (equation (5.5)).

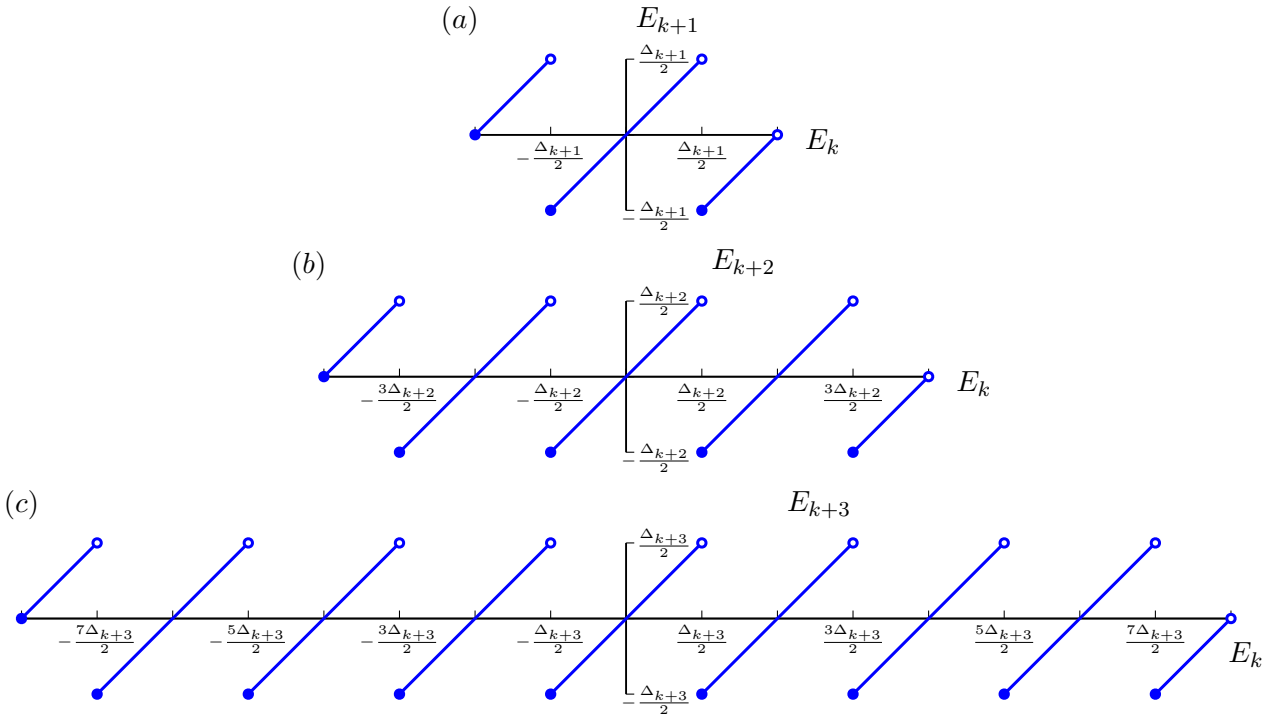


Figure 5-3: Deterministic quantizer error functions for k bits to (a) $k+1$ bits, (b) $k+2$ bits, and (c) $k+3$ bits.

$$\operatorname{cov}(E_k, E_{k+n}) = \mathbb{E}[E_k E_{k+n}] - \mathbb{E}[E_k]\mathbb{E}[E_{k+n}] \tag{5.4}$$

$$= \mathbb{E}[E_k E_{k+n}] \tag{5.5}$$

$$= \int_{\frac{-\Delta_k}{2}}^{\frac{\Delta_k}{2}} e_k E_{k+n}(e_k) f_{E_k}(e_k)\, \mathrm{d}e_k \tag{5.6}$$

Using our previous formulation of the function $E_{k+n}(e_k)$ in equation (5.3), we multiply by E_k to obtain the desired random variable $E_k E_{k+n}$. $E_k E_{k+n}$ contains quadratic segments over the same regions as $E_{k+n}(e_k)$. Figure 5-4 depicts $E_k E_{k+3}$. The area under each of the quadratic segments has been shaded to emphasize that the expectation operation amounts to integration of this function scaled by a density constant for a uniform random variable.

Equation (5.6) can be simplified by noting that the PDF $f_{E_k}(e_k)$ is uniform, and thus evaluates to a constant as shown in equation (5.7).

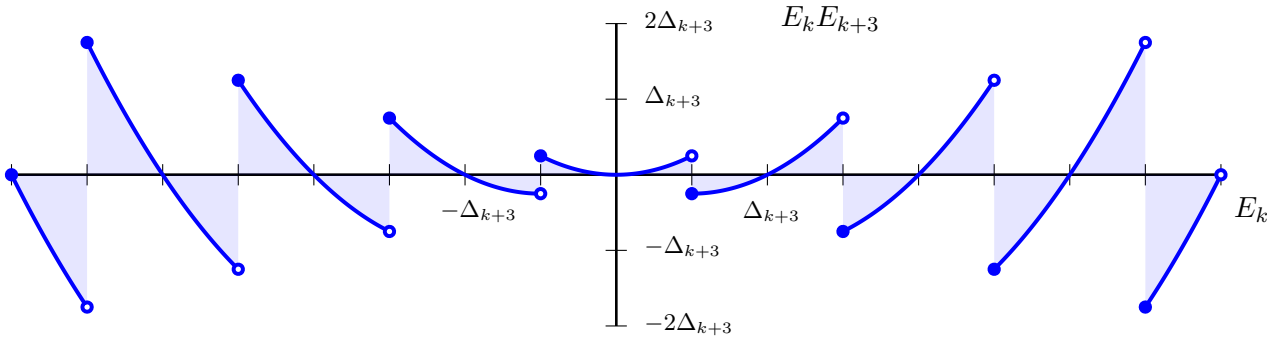


Figure 5-4: E_kE_{k+3} as a function of E_k, where area has been shaded to emphasize that expectation reduces to integration.

$$f_{E_k}(e_k) = \begin{cases} \frac{1}{2^n\Delta_{k+n}} & \text{if } -2^{n-1}\Delta_{k+n} \leq e_k < 2^{n-1}\Delta_{k+n} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

As mentioned earlier, the expectation operation reduces to an integral over $E_{k+n}(e_k)E_k$ scaled by a constant (equation (5.8)). We separate the integral into the three regions, as shown in equation (5.9). Note from Figure 5-4 that the product $E_{k+n}(e_k)E_k$ is an even function. By symmetry, corresponding regions on either side of the y axis contain the same area. It is convenient to perform the integral over the positive regions, then multiply by two (Eq 5.9).

$$\begin{aligned} \text{cov}(E_k, E_{k+n}) &= \int_{\frac{-\Delta_k}{2}}^{\frac{\Delta_k}{2}} e_k E_{k+n}(e_k) f_{E_k}(e_k)\, \mathrm{d}e_k \\ &= \frac{1}{2^n\Delta_{k+n}} \int_{-2^{n-1}\Delta_{k+n}}^{2^{n-1}\Delta_{k+n}} e_k E_{k+n}(e_k)\, \mathrm{d}e_k \end{aligned} \tag{5.8}$$

$$\operatorname{cov}(E_k, E_{k+n}) = \frac{1}{2^n \Delta_{k+n}} \left\{ \overbrace{\int_{-\frac{\Delta_{k+n}}{2}}^{\frac{\Delta_{k+n}}{2}} e_k^2 \, \mathrm{d}e_k}^{\text{base region}} + 2 \overbrace{\sum_{i=1}^{2^{n-1}-1} \int_{\Delta_{k+n}\left(i-\frac{1}{2}\right)}^{\Delta_{k+n}\left(i+\frac{1}{2}\right)} e_k^2 - i\Delta_{k+n} e_k \, \mathrm{d}e_k}^{\text{middle regions}} \right.$$
$$\left. + 2 \overbrace{\int_{\left(2^{n-1}-\frac{1}{2}\right)\Delta_{k+n}}^{2^{n-1}\Delta_{k+n}} e_k^2 - 2^{n-1}\Delta_{k+n} e_k \, \mathrm{d}e_k}^{\text{outermost regions}} \right\} \tag{5.9}$$

Counting each of the middle regions requires a summation of an integral over $2^{n-1} - 1$ distinct intervals. The base and outermost regions only require a single integral. To explore the contributions of the three types of regions, we will evaluate each independently. Performing the integration on the base region and combining terms, we obtain equation (5.10).

$$\overbrace{\int_{-\frac{\Delta_{k+n}}{2}}^{\frac{\Delta_{k+n}}{2}} e_k^2 \, \mathrm{d}e_k}^{\text{base region}} = \frac{1}{3}\left[\left(\frac{\Delta_{k+n}}{2}\right)^3 - \left(-\frac{\Delta_{k+n}}{2}\right)^3\right] = \frac{\Delta_{k+n}^3}{12} \tag{5.10}$$

The integration for the middle regions requires more attention. After some simplification each of the index variables i cancel, and the summation reduces (equation (5.13)).

$$\overbrace{2 \sum_{i=1}^{2^{n-1}-1} \int_{\Delta_{k+n}\left(i-\frac{1}{2}\right)}^{\Delta_{k+n}\left(i+\frac{1}{2}\right)} e_k^2 - i\Delta_{k+n} e_k \, \mathrm{d}e_k}^{\text{middle regions}} = 2 \sum_{i=1}^{2^{n-1}-1} \left[\frac{1}{3}\Delta_{k+n}^3 \left(\left(i+\frac{1}{2}\right)^3 - \left(i-\frac{1}{2}\right)^3\right)\right. \tag{5.11}$$
$$\left. - \frac{1}{2} i \Delta_{k+n}^3 \left(\left(i+\frac{1}{2}\right)^2 - \left(i-\frac{1}{2}\right)^2\right)\right]$$
$$= 2 \sum_{i=1}^{2^{n-1}-1} \left[\Delta_{k+n}^3 \left(i^2 + \frac{1}{12}\right) - i^2 \Delta_{k+n}^3\right] \tag{5.12}$$
$$= (2^n - 2) \frac{\Delta_{k+n}^3}{12} \tag{5.13}$$

For the outermost regions, the integration is similar. Combining terms, we arrive at equation (5.14). Note that the outermost region is the only one that contributes a negative value to the overall expectation.

$$\overbrace{2\int_{\left(2^{n-1}-\frac{1}{2}\right)\Delta_{k+n}}^{2^{n-1}\Delta_{k+n}} e_k^2 - 2^{n-1}\Delta_{k+n}e_k \, \mathrm{d}e_k}^{\text{outermost regions}} = \frac{2}{3}\Delta_{k+n}^3\left(2^{3n-3} - \left(2^{n-1}-\frac{1}{2}\right)^3\right)$$
$$- 2^{n-1}\Delta_{k+n}^3\left(2^{2n-2} - \left(2^{n-1}-\frac{1}{2}\right)^2\right)$$
$$= 2\Delta_{k+n}^3\left(2^{2n-3} - 2^{n-3} + \frac{1}{24}\right) - 2\Delta_{k+n}^3\left(2^{2n-3} - 2^{n-4}\right)$$
$$= \left(1 - 3\cdot 2^{n-1}\right)\frac{\Delta_{k+n}^3}{12} \tag{5.14}$$

Substituting these results into equation (5.9) and multiplying by the probability density constant, we obtain equation (5.15). Equation (5.15) gives insight into the contributions to the covariance of each region. The base region and middle regions simply contribute the original positive variance of E_{k+n} weighted by their respective probabilities. However, the outermost region contributes a negative scaled version of the original variance. Intuitively, one expects this to occur because in the outermost region, the sign of E_k is always the opposite of E_{k+n}.

Combining terms, we arrive at equation (5.16) and equation (5.17), from which we conclude that the overall covariance $\mathrm{cov}(E_k, E_{k+n})$ is simply negative one half of the variance of E_{k+n}.

$$\mathrm{cov}(E_k, E_{k+n}) = \frac{\Delta_{k+n}^2}{12}\left(\overbrace{\frac{1}{2^n}}^{\text{base region}} + \overbrace{\left(1-\frac{1}{2^{n-1}}\right)}^{\text{middle regions}} + \overbrace{\left(1-3\cdot 2^{n-1}\right)\left(\frac{1}{2^n}\right)}^{\text{outermost regions}}\right) \tag{5.15}$$
$$= (-1/2)\,\frac{\Delta_{k+n}^2}{12} \tag{5.16}$$
$$= (-1/2)\,\mathrm{var}(E_{k+n}) \tag{5.17}$$

The correlation coefficient ρ can be viewed as a scaled version of the covariance from -1 to 1. If $\rho > 0$, then the values $(x - \mathbb{E}[X])$ and $(y - \mathbb{E}[Y])$ tend to have the same sign. Again, the correlation coefficient is defined as:

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}. \tag{5.18}$$

To calculate the correlation coefficient for the derived covariance between the error in signals quantized with an adjacent number of bits, we begin by substituting the variances of the respective uniform error signals to obtain equation (5.19). Using the relationship $\Delta_k = 2^n \Delta_{k+n}$, we arrive at equation (5.20).

$$\rho = \frac{\left(-\frac{1}{2}\right)\frac{\Delta_{k+n}^2}{12}}{\sqrt{\left(\frac{\Delta_{k+n}^2}{12}\right)\left(\frac{\Delta_k^2}{12}\right)}} \tag{5.19}$$

$$= \frac{\left(-\frac{1}{2}\right)\frac{\Delta_{k+n}^2}{12}}{\sqrt{\left(\frac{\Delta_{k+n}^2}{12}\right)\left(\frac{(\Delta_{k+n}2^n)^2}{12}\right)}}$$

$$= \frac{\left(-\frac{1}{2}\right)\frac{\Delta_{k+n}^2}{12}}{\frac{\Delta_{k+n}^2 2^n}{12}}$$

$$= -2^{-n-1} \tag{5.20}$$

This reveals a simple formula that defines the correlation between the error in adjacent bits. As the difference in the number of bits n increases, the negative correlation is less pronounced. Equation (5.20) matches the results of the simulation well. Table 5.1 displays the predicted correlation coefficient $\rho_{\text{predicted}}$ between error signals resulting from quantization with k bits and $k \pm n$ bits. It also displays the measured correlation coefficient $\rho_{\text{measured random}}$ for a uniformly distributed random input signal scaled to avoid saturation. To verify our assumptions, we also include the measured correlation coefficient $\rho_{\text{measured audio}}$ for the WAV audio recording used previously in this thesis.

Figure 5-5 illustrates the predicted and measured correlation coefficient between a 0-20 bit quantizer output and a 10 bit quantizer output. The figure reveals that the correlation coefficient is symmetric around the 10 bit quantizer. As expected, there is a discrepancy when $n = 0$ because the quantizer error function is simply a line with no discontinuities. As discussed earlier, in this case the correlation coefficient $\rho = 1$.

Note that the derivation for $\rho_{\text{predicted}}$ depends on the validity of the additive noise model, and thus we assume that the input signal does not saturate the quantizer.

$k+n$ bits	k bits	$\rho_{\text{predicted}}$	$\rho_{\text{measured random}}$	$\rho_{\text{measured audio}}$
10	9	-0.2500	-0.2511	-0.2517
10	8	-0.1250	-0.1238	-0.1233
10	7	-0.0625	-0.0621	-0.0609
10	6	-0.0313	-0.0312	-0.0299
10	5	-0.0156	-0.0150	-0.0150

Table 5.1: Negative Correlation Coefficient between Quantizer Error Signals

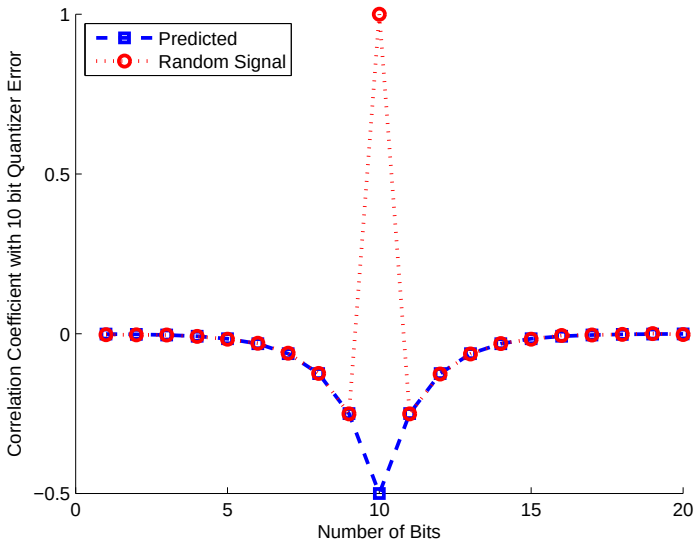


Figure 5-5: Predicted and measured correlation coefficient between 0-20 bit quantizer error and 10 bit quantizer error.

5.3 Combining Inhomogeneous Measurements

We have demonstrated that simultaneously quantizing the same sample with different numbers of bits will result in correlated error signals. Each of the error signals E_k is a uniform random variable that is deterministically related to the other error signals. Each error signal has a different variance $\frac{\Delta_k^2}{12}$.

We will refer to these quantized samples as "inhomogeneous measurements." Given a set of d inhomogeneous measurements, in this section we calculate the Weighted Least Squares (WLS) estimate of the mean. Each measurement consists of the true analog value of the signal and an additive zero-mean error component due to quantization. Specifically, we define d to be the number of quantized samples with different numbers of bits available to generate a composite estimate. For

example, if all quantized versions of a signal from 10 bits through 5 bits are available, then $d = 6$. For a traditional quantizer, $d = 1$. For this section, we will constrain the bits to be adjacent, although in general any different quantizers may be used.

For a vector of observations $\boldsymbol{x}$ and weighting matrix $\boldsymbol{W}$, the Weighted Least Squares criterion minimizes the sum of the weighted squares of the residuals to obtain an estimate for $\boldsymbol{\mu}$ [14].

$$\mu_{WLS} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)^T \boldsymbol{W} \left(\boldsymbol{x} - \boldsymbol{\mu}\right) \tag{5.21}$$

Note that this is equivalent to the Maximum Likelihood Estimate (MLE) when the noise is multivariate Gaussian and the weighting matrix is the inverse of the covariance matrix [2]. Setting $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1}$, we obtain equation (5.22).

$$\mu_{WLS} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \left(\boldsymbol{x} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma^{-1}} \left(\boldsymbol{x} - \boldsymbol{\mu}\right) \tag{5.22}$$

The mean $\boldsymbol{\mu}$ is a $d \times 1$ vector, $\boldsymbol{x}$ is a $d \times 1$ sample vector, and the covariance matrix $\boldsymbol{\Sigma}$ is a $d \times d$ symmetric matrix. In this case, d is the number of simultaneous observations. Each entry in $\boldsymbol{\mu}$ is equal to the same constant, because each quantizer is operating on the same analog sample, which represents the mean. For the covariance matrix $\boldsymbol{\Sigma}$, we will use the covariance relationship defined in equation (5.17).

For the next step, we will define the precision matrix $\boldsymbol{J} = \boldsymbol{\Sigma}^{-1}$. Taking the partial derivative with respect to μ and setting equal to zero yields the expression for μ_{WLS} in equation (5.23).

$$\mu_{WLS} = \frac{\sum_{i=1}^{d} x_i \sum_{j=1}^{d} J_{i,j}}{\sum_{i=1}^{d} \sum_{j=1}^{d} J_{i,j}} \tag{5.23}$$

This can also be expressed as a weighted average of the observations, as shown in equation (5.24).

$$\mu_{WLS} = \sum_{i=1}^{d} w_i x_i \quad \text{where} \quad w_i = \frac{\sum_{j=1}^{d} J_{i,j}}{\sum_{i=1}^{d} \sum_{j=1}^{d} J_{i,j}} \tag{5.24}$$

5.4 Designing Quantizers to Exploit Error Correlation

5.4.1 Table of WLS Weights

In the previous section, we demonstrated that the WLS estimate of an analog sample based on inhomogeneous observations from different quantizers is a weighted average. The weights are dependent on the covariance matrix of deterministic error functions, which were also defined previously. Using equation (5.17) to define each non-diagonal entry, we arrive at the general expanded covariance matrix in Table 5.2.

$$\Sigma = \begin{array}{c} \text{bits} \\ k-d+1 \\ k-d+2 \\ k-d+3 \\ k-d+4 \\ \vdots \\ k-3 \\ k-2 \\ k-1 \\ k \end{array} \begin{array}{c} \begin{array}{cccccccccc} k-d+1 & k-d+2 & k-d+3 & k-d+4 & \ldots & k-3 & k-2 & k-1 & k \end{array} \\ \begin{bmatrix} +2^{2d-2} & -2^{2d-5} & -2^{2d-7} & -2^{2d-9} & \ldots & -2^{5} & -2^{3} & -2^{1} & -2^{-1} \\ -2^{2d-5} & +2^{2d-4} & -2^{2d-7} & -2^{2d-9} & \ldots & -2^{5} & -2^{3} & -2^{1} & -2^{-1} \\ -2^{2d-7} & -2^{2d-7} & +2^{2d-6} & -2^{2d-9} & \ldots & -2^{5} & -2^{3} & -2^{1} & -2^{-1} \\ -2^{2d-9} & -2^{2d-9} & -2^{2d-9} & +2^{2d-8} & \ldots & -2^{5} & -2^{3} & -2^{1} & -2^{-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ -2^{5} & -2^{5} & -2^{5} & -2^{5} & \ldots & +2^{6} & -2^{3} & -2^{1} & -2^{-1} \\ -2^{3} & -2^{3} & -2^{3} & -2^{3} & \ldots & -2^{3} & +2^{4} & -2^{1} & -2^{-1} \\ -2^{1} & -2^{1} & -2^{1} & -2^{1} & \ldots & -2^{1} & -2^{1} & +2^{2} & -2^{-1} \\ -2^{-1} & -2^{-1} & -2^{-1} & -2^{-1} & \ldots & -2^{-1} & -2^{-1} & -2^{-1} & +1 \end{bmatrix} \end{array} \times \frac{\Delta_k^2}{12}$$

Table 5.2: Covariance Matrix for error in $k-d+1$ to k bit quantizers.

Table 5.2 includes a scalar factor for the variance of E_k. Using the matrix inversion property in equation (5.25), we find that the scalar variance term is simply inverted when we calculate the information form parameter $\boldsymbol{J}$. From the form of equation (5.24), it is easy to see that a scalar factor multiplied by $\boldsymbol{J}$ will be normalized out. We conclude that the scalar is unnecessary and may be removed.

$$(kA)^{-1} = k^{-1}A^{-1} \tag{5.25}$$

Also, the weights calculated from the matrix in Table 5.2 are applicable for arbitrary k with

a fixed number of quantized signals d. However, when the number of observations d changes, we must recalculate the weights. We can do this in an efficient way by incorporating Recursive Least Squares to update the weights [14].

In Table 5.3, weights are listed for up to 15 simultaneous signals quantized with adjacent numbers of bits. For $d = 1$ (first row of entries), a single weight of one is defined. This reduces to a traditional quantizer. For $d = 15$ (the fifteenth row of entries) 15 weights are defined. This might correspond to quantizing the same sample with 20 bits through 6 bits simultaneously. Note that the weight w_1 corresponds to the highest number of bits and w_{15} corresponds to the lowest number of bits.

Intuitively, it makes sense that the quantized signals with highest accuracy should be assigned the largest weights.

d ↓	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}	w_{13}	w_{14}	w_{15}
1	1	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2	$\frac{3}{4}$	$\frac{1}{4}$	*	*	*	*	*	*	*	*	*	*	*	*	*
3	$\frac{2}{3}$	$\frac{1}{4}$	$\frac{1}{12}$	*	*	*	*	*	*	*	*	*	*	*	*
4	$\frac{5}{8}$	$\frac{1}{4}$	$\frac{3}{32}$	$\frac{1}{32}$	*	*	*	*	*	*	*	*	*	*	*
5	$\frac{3}{5}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{3}{80}$	$\frac{1}{80}$	*	*	*	*	*	*	*	*	*	*
6	$\frac{7}{12}$	$\frac{1}{4}$	$\frac{5}{48}$	$\frac{1}{24}$	$\frac{1}{64}$	$\frac{1}{192}$	*	*	*	*	*	*	*	*	*
7	$\frac{4}{7}$	$\frac{1}{4}$	$\frac{3}{28}$	$\frac{5}{112}$	$\frac{1}{56}$	$\frac{3}{448}$	$\frac{1}{448}$	*	*	*	*	*	*	*	*
8	$\frac{9}{16}$	$\frac{1}{4}$	$\frac{7}{64}$	$\frac{3}{64}$	$\frac{5}{256}$	$\frac{1}{128}$	$\frac{3}{1024}$	$\frac{1}{1024}$	*	*	*	*	*	*	*
9	$\frac{5}{9}$	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{7}{144}$	$\frac{1}{48}$	$\frac{5}{576}$	$\frac{1}{288}$	$\frac{1}{768}$	$\frac{1}{2304}$	*	*	*	*	*	*
10	$\frac{11}{20}$	$\frac{1}{4}$	$\frac{9}{80}$	$\frac{1}{20}$	$\frac{7}{320}$	$\frac{3}{320}$	$\frac{1}{256}$	$\frac{1}{640}$	$\frac{3}{5120}$	$\frac{1}{5120}$	*	*	*	*	*
11	$\frac{6}{11}$	$\frac{1}{4}$	$\frac{5}{44}$	$\frac{9}{176}$	$\frac{1}{44}$	$\frac{7}{704}$	$\frac{3}{704}$	$\frac{5}{2816}$	$\frac{1}{1408}$	$\frac{3}{11264}$	$\frac{1}{11264}$	*	*	*	*
12	$\frac{13}{24}$	$\frac{1}{4}$	$\frac{11}{96}$	$\frac{5}{96}$	$\frac{3}{128}$	$\frac{1}{96}$	$\frac{7}{1536}$	$\frac{1}{512}$	$\frac{5}{6144}$	$\frac{1}{3072}$	$\frac{1}{8192}$	$\frac{1}{24576}$	*	*	*
13	$\frac{7}{13}$	$\frac{1}{4}$	$\frac{3}{26}$	$\frac{11}{208}$	$\frac{5}{208}$	$\frac{9}{832}$	$\frac{1}{208}$	$\frac{7}{3328}$	$\frac{3}{3328}$	$\frac{5}{13312}$	$\frac{1}{6656}$	$\frac{3}{53248}$	$\frac{1}{53248}$	*	*
14	$\frac{15}{28}$	$\frac{1}{4}$	$\frac{13}{112}$	$\frac{3}{56}$	$\frac{11}{448}$	$\frac{5}{448}$	$\frac{9}{1792}$	$\frac{1}{448}$	$\frac{1}{1024}$	$\frac{3}{7168}$	$\frac{5}{28672}$	$\frac{1}{14336}$	$\frac{3}{114688}$	$\frac{1}{114688}$	*
15	$\frac{8}{15}$	$\frac{1}{4}$	$\frac{7}{60}$	$\frac{13}{240}$	$\frac{1}{40}$	$\frac{11}{960}$	$\frac{1}{192}$	$\frac{3}{1280}$	$\frac{1}{960}$	$\frac{7}{15360}$	$\frac{1}{5120}$	$\frac{1}{12288}$	$\frac{1}{30720}$	$\frac{1}{81920}$	$\frac{1}{245760}$

Table 5.3: WLS weights to exploit negative correlation for d simultaneous quantized signals.

A similar table of weights can be easily calculated for arbitrary numbers of bits d using equa-

tion (5.24) and Table 5.2. In this work we will avoid using quantizers with one or two bits because the saturation levels severely limit the input range.

These concepts can be incorporated into an ADC architecture. The system consists of several quantizers with different numbers of bits in parallel. Figure 5-6 depicts this system architecture, where Q_k denotes quantization with k bits. In general, this is expensive in terms of die cost and power usage. However, it becomes practical in the environment of time-interleaved ADCs. In the system formulated in Chapter 2, M independent quantizers are already present. The WLS weighted average can be calculated easily by simply scaling the amplitude of the digital reconstruction filters. As we will see in Chapter 6, employing WLS in time-interleaved ADCs may increase the overall performance of the system, although it requires certain design concessions.

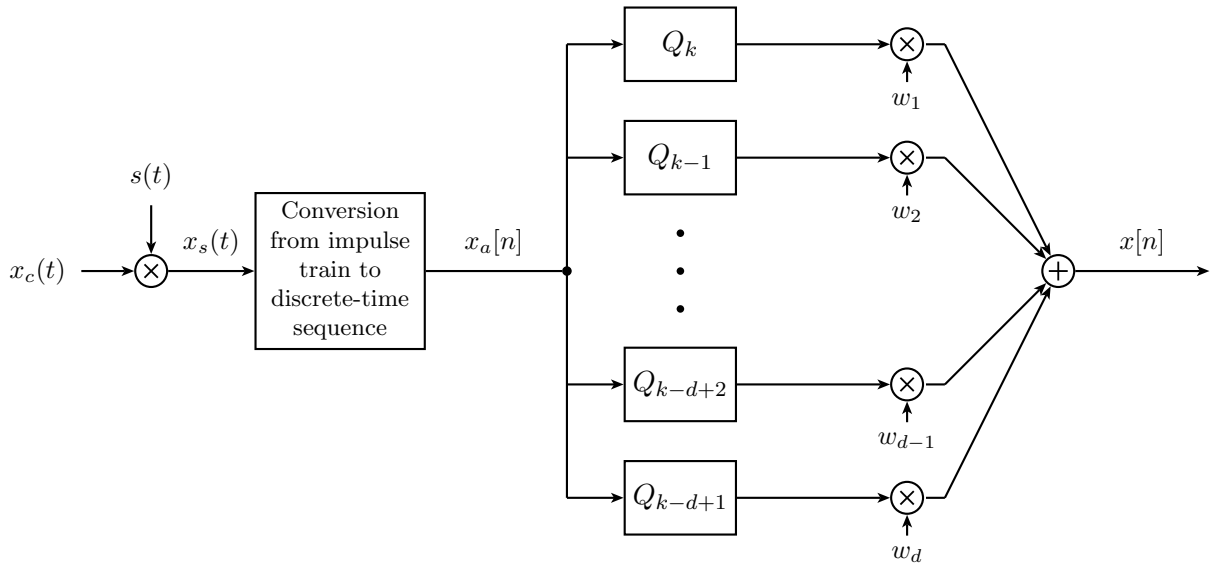


Figure 5-6: ADC architecture exploiting negative correlation between quantizer error signals. Consists of conventional C/D converter followed by a weighted average.

5.4.2 Performance Gains

In this section, we evaluate the predicted performance gain from exploiting negative correlation with a weighted average of observations. The variance of a weighted average of dependent random variables is defined by equation (5.26).

$$\text{var}\left(\sum_{i=1}^{d} w_i X_i\right) = \sum_{i=1}^{d} w_i^2 \text{var}(X_i) + 2\sum_{i=1}^{d}\sum_{j=i+1}^{d} w_i w_j \text{cov}(X_i, X_j) \tag{5.26}$$

From Table 5.3 and Table 5.2 it is simple to solve for the predicted covariance of the composite signal in MATLAB. Again, the WLS composite signal contains weighted samples from d quantizers (k bits to $k-d+1$ bits). From Figure 5-7, it is apparent that the nearest numbers of bits provide the majority of the performance improvement.

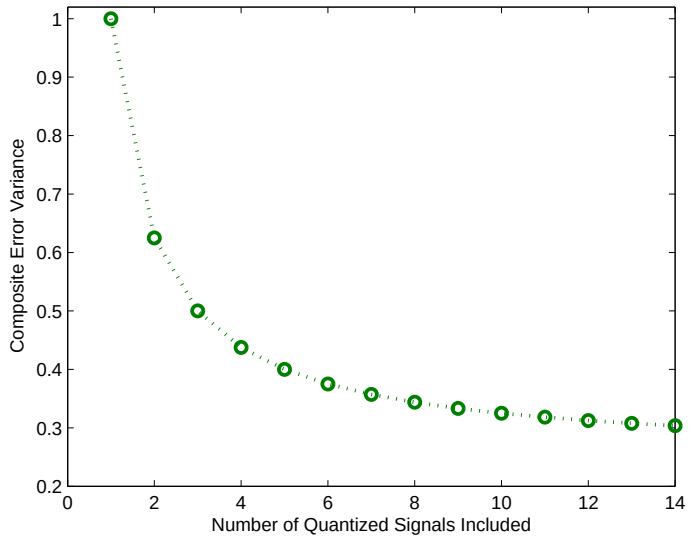


Figure 5-7: Predicted composite error variance as a multiple of the error variance from k bits. Displayed as a function of the number of adjacent quantizer error signals included in the weighted sum.

It is also useful to analyze the effective resolution in bits gained. A k bit quantizer has error variance defined in equation (5.27). Solving for k, we obtain equation (5.28), which defines the equivalent number of bits for a given variance.

$$\sigma_e^2 = \frac{2^{-2(k-1)}}{12} \tag{5.27}$$

$$k_{eff} = 1 - \frac{1}{2}\log_2(12\sigma_e^2) \tag{5.28}$$

To calculate the equivalent number of bits gained, we take the difference between the traditional

and modified architecture error variances in equation (5.28), finally obtaining equation (5.29).

$$\begin{aligned}\text{bits gained} = k_{eff_{mod}} - k_{eff_{trad}} &= \frac{1}{2}\left\{\log_2\left(12\sigma^2_{trad}\right) - \log_2\left(12\sigma^2_{mod}\right)\right\} \\ &= \frac{1}{2}\log_2\left(\frac{\sigma^2_{trad}}{\sigma^2_{mod}}\right) \end{aligned} \tag{5.29}$$

Table 5.4 and Figure 5-8 display the effective bits gained as a function of the number of quantized signals available. As the number of signals quantized with adjacent numbers of bits approaches 30, the effective number of bits gained nears one. This is a reduction by a factor of four from the original error variance with a traditional quantizer.

# of Quantized Signals	Effective Bits Gained
1	0.000
2	0.339
3	0.500
4	0.596
5	0.661
6	0.708
7	0.743
8	0.770
9	0.793
10	0.811
11	0.826
12	0.839
13	0.850
14	0.860

Table 5.4: Predicted effective number of bits gained as a function of the number of signals quantized with adjacent numbers of bits included in the weighted sum.

5.5 Experimental Verification

In this section, we experimentally verify the predicted performance gains. Two signals are tested that are scaled to avoid clipping above two bits. First, a uniformly distributed random variable is simultaneously quantized with 16 through 3 bits in MATLAB. The WLS weighted average of the observations is calculated using the weights in Table 5.3. A WAV audio recording is also tested

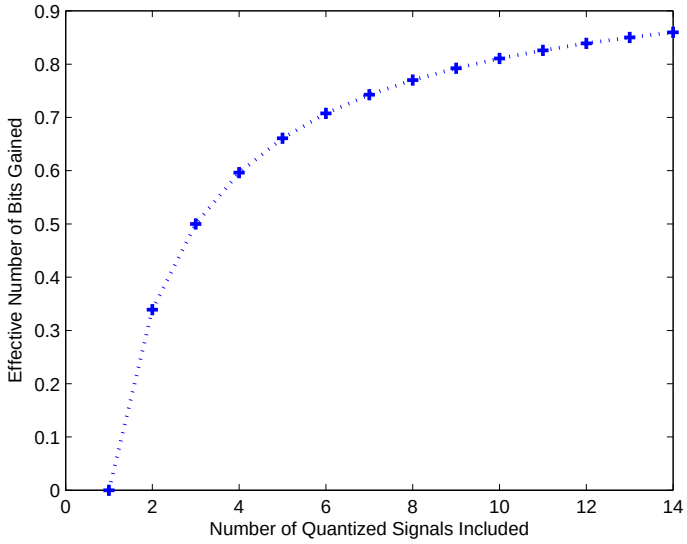


Figure 5-8: Predicted effective number of bits gained as a function of the number of signals quantized with adjacent numbers of bits included in the weighted sum.

using the same process.

In both cases, the results match the predicted error variance and equivalent number of bits closely. These results are plotted in Figure 5-9 and Figure 5-10. The upper and lower bounds in Figure 5-9 denote the variance for traditional 16 and 17 bit quantizers, respectively. The composite signal using 16 through 3 bits approaches the variance of a traditional 17 bit quantizer.

Note that there is a small discrepancy in the audio recording as the weighted average begins to include very coarse quantizers. This is because the assumption of uniformly distributed error becomes less valid.

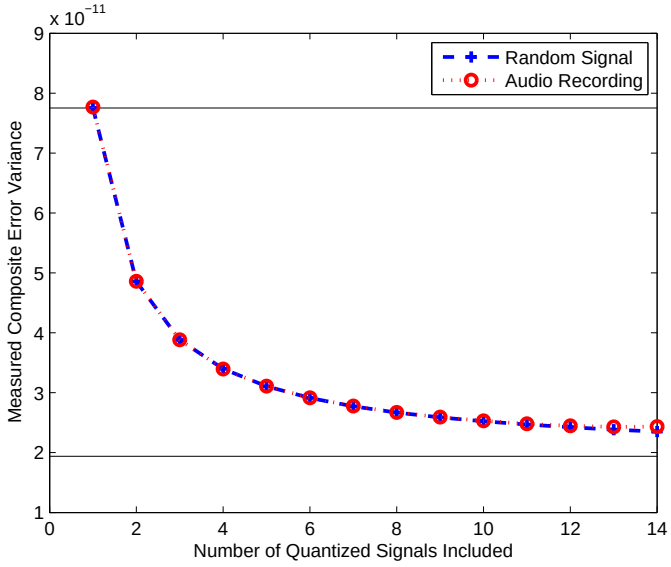


Figure 5-9: Measured composite error variance as a multiple of the error variance from $k = 16$ bits. Displayed as a function of the number of quantized signals with adjacent bits included in the weighted sum.

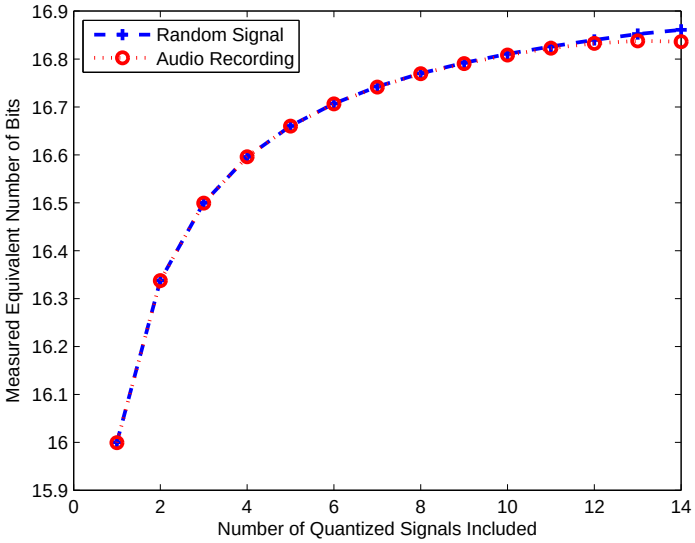


Figure 5-10: Measured effective number of bits for $k = 16$ as a function of the number of quantized signals with adjacent bits included in the weighted sum.

Chapter 6

Optimization of Timing Offsets for Fixed Bit Allocation

In Chapters 4 and 5 of this thesis, we examined discrepancies between the additive noise model and uniform quantizers that arise for time-interleaved ADCs. The remaining portion of the thesis will focus on optimal design of time-interleaved ADCs to mitigate quantization error. Our earlier analysis will prove to be very useful for bridging gaps between the predicted optimal configuration with the additive noise model and the actual optimal configuration.

In this chapter, we investigate the relationship between timing offsets and bit allocations for the time-interleaved ADC structure described in Chapter 2. This is a high-dimensional problem with M timing offsets τ_m, M bit allocations b_m, and a performance measurement associated with each permutation.

One design scenario for time-interleaved ADCs involves choosing timing offsets τ_m for M channels with fixed bit allocation b_m that will maximize the overall SQNR. We discuss how to appropriately choose the optimal timing offsets τ_m and the correction necessary to account for the effects of quantization error correlation.

6.1 Predicted Output Error Variance

The optimal reconstruction filters $G_m(e^{j\omega})$ introduced in Chapter 3 were designed to minimize quantization error power while achieving perfect reconstruction in the absence of quantization.

Substituting $G_m(e^{j\omega})$ into the average output error variance expression equation (2.9) from Chapter 2, we obtain the minimum achievable value for σ_e^2.

$$\sigma_{e_{min}}^2 = \frac{1}{L}\sum_{i=0}^{L-1}\left(\frac{1}{L}|\Lambda^{(i)}(e^{j\omega_m})|^2/\sigma_m^2\right) \tag{6.1}$$

As shown in [18] and [16], this expression can be written compactly in terms of the trace and inverse of the A_M matrix defined in equation (3.6).

$$\sigma_{e_{min}}^2 = tr(A_M^{-1}) \tag{6.2}$$

A lower bound for the minimum achievable error is given in equation (6.3). The conditions for equality are given in [18] and [16].

$$\sigma_{e_{min}}^2 \geq \frac{L}{\sum_{m=0}^{M-1} 1/\sigma_m^2} \tag{6.3}$$

Maymon has shown that with equal quantization granularity in each channel, the uniform sampling grid achieves minimum output quantization noise power, equivalent to $(L/M)\cdot\sigma^2$. This is not a unique solution; often many non-uniform timing configurations achieve the same result. However, with different quantization granularities in each channel, a recurrent non-uniform sampling pattern can often achieve a better SQNR than the uniform timing configuration.

6.2 Analytical Optimization of Predicted Error

The conditions for equality for the lower bound given in equation (6.3) are often unattainable for a given bit allocation b_m. For this reason, the conditions given in [18] and [16] do not necessarily suggest an optimal timing configuration for an arbitrary bit allocation. To optimize the timing offsets τ_m for an arbitrary bit allocation b_m, we refer to the original expression for predicted output error variance.

In equation 6.2, we introduced a compact expression for the predicted output error variance $\sigma_{e_{min}}^2$. Expanding A_M, we arrive at the matrix in equation (6.4).

$$\sigma^2_{e_{min}} = tr\left(\begin{bmatrix} \sum_{m=0}^{M-1}\frac{1}{\sigma_m^2} & \sum_{m=0}^{M-1}\frac{e^{-j2\pi\tau_m/LT_N}}{\sigma_m^2} & \cdots & \sum_{m=0}^{M-1}\frac{e^{-(L-1)j2\pi\tau_m/LT_N}}{\sigma_m^2} \\ \sum_{m=0}^{M-1}\frac{e^{j2\pi\tau_m/LT_N}}{\sigma_m^2} & \sum_{m=0}^{M-1}\frac{1}{\sigma_m^2} & \cdots & \sum_{m=0}^{M-1}\frac{e^{-(L-2)j2\pi\tau_m/LT_N}}{\sigma_m^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{m=0}^{M-1}\frac{e^{(L-1)j2\pi\tau_m/LT_N}}{\sigma_m^2} & \sum_{m=0}^{M-1}\frac{e^{(L-2)j2\pi\tau_m/LT_N}}{\sigma_m^2} & \cdots & \sum_{m=0}^{M-1}\frac{1}{\sigma_m^2} \end{bmatrix}^{-1}\right) \tag{6.4}$$

In general, equation (6.4) is difficult to optimize analytically because of the inverse matrix operation. However, we will provide a closed form solution for the simple case when $M = 3$ and $L = 2$ to gain insight for more complex scenarios.

6.2.1 Error Minimization for the Case of $M = 3$ and $L = 2$

When $M = 3$ and $L = 2$, equation (6.4) reduces to equation (6.5). Using the closed form for the inverse of a 2×2 matrix and gathering complex exponentials into sinusoids yields equation (6.6).

$$\sigma^2_{e_{min}} = tr\left(\begin{bmatrix} \sum_{m=0}^{2}\frac{1}{\sigma_m^2} & \sum_{m=0}^{2}\frac{e^{-j2\pi\tau_m/LT_N}}{\sigma_m^2} \\ \sum_{m=0}^{2}\frac{e^{j2\pi\tau_m/LT_N}}{\sigma_m^2} & \sum_{m=0}^{2}\frac{1}{\sigma_m^2} \end{bmatrix}^{-1}\right) \tag{6.5}$$

$$= \frac{2\left(\sum_{m=0}^{2}\frac{1}{\sigma_m^2}\right)}{\left(\sum_{m=0}^{2}\frac{1}{\sigma_m^2}\right)^2 - \left(\sum_{m=0}^{2}\frac{e^{-j2\pi\tau_m/LT_N}}{\sigma_m^2}\right)\left(\sum_{m=0}^{2}\frac{e^{j2\pi\tau_m/LT_N}}{\sigma_m^2}\right)}$$

$$= \frac{2\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^2 - \frac{1}{\sigma_0^2}^2 - \frac{1}{\sigma_1^2}^2 - \frac{1}{\sigma_2^2}^2 - \frac{2\cos(2\pi(\tau_0-\tau_1)/LT_N)}{\sigma_0^2\sigma_1^2} - \frac{2\cos(2\pi(\tau_0-\tau_2)/LT_N)}{\sigma_0^2\sigma_2^2} - \frac{2\cos(2\pi(\tau_1-\tau_2)/LT_N)}{\sigma_1^2\sigma_2^2}} \tag{6.6}$$

At this stage, minimization of the overall error variance for a given bit allocation depends only on the cosine terms in the denominator. Furthermore, each cosine term is only dependent on the difference in time between two offsets. This realization allows us to simplify the problem by defining time differences between timing offsets.

We define the time difference between channels a and b as $d_{a,b}$, shown in equation (6.7). The

concept is illustrated in Figure 6-1. The time difference $d_{a,b}$ is bounded by $-1/2$ and $1/2$, because of the recurrent sampling pattern.

$$d_{a,b} = \frac{\tau_a - \tau_b}{L \cdot T_N} - i \qquad \text{where } i \text{ is an integer such that } -1/2 < d_{a,b} \leq 1/2 \tag{6.7}$$

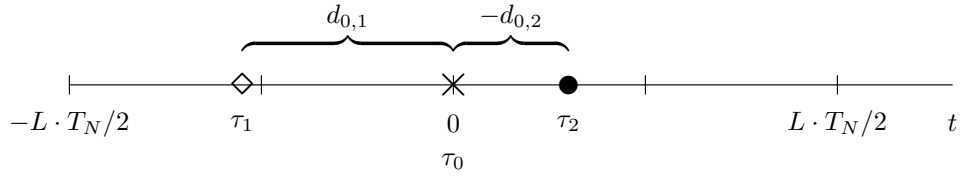


Figure 6-1: Example of distance measurement for $M = 3$ and $L = 2$.

It is possible to uniquely specify a sampling constellation of M channels with $M - 1$ distances to within a time shift. As illustrated in equation (6.6), this time shift is irrelevant in terms of predicted error variance, and only relative distances between channels are important.

Substituting distance variables into equation (6.6) and noting that $d_{1,2} = \tau_1 - \tau_2 = -\tau_0 + \tau_1 + \tau_0 - \tau_2 = d_{0,2} - d_{0,1}$, we arrive at equation (6.8).

$$\sigma^2_{e_{min}} = \frac{2\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^2 - \frac{1}{\sigma_0^2}^2 - \frac{1}{\sigma_1^2}^2 - \frac{1}{\sigma_2^2}^2 - \frac{2\cos(2\pi d_{0,1})}{\sigma_0^2\sigma_1^2} - \frac{2\cos(2\pi d_{0,2})}{\sigma_0^2\sigma_2^2} - \frac{2\cos(2\pi(d_{0,2}-d_{0,1}))}{\sigma_1^2\sigma_2^2}} \tag{6.8}$$

Because the bit allocations are given and all variances are positive, the optimal distances $d_{0,1}$ and $d_{0,2}$ are described by equation (6.9).

$$\underset{d_{0,1},d_{0,2}}{\text{argmin}} \quad \sigma^2_{e_{min}} = \underset{d_{0,1},d_{0,2}}{\text{argmin}} \quad \frac{\cos(2\pi d_{0,1})}{\sigma_0^2\sigma_1^2} + \frac{\cos(2\pi d_{0,2})}{\sigma_0^2\sigma_2^2} + \frac{\cos(2\pi(d_{0,2} - d_{0,1}))}{\sigma_1^2\sigma_2^2} \tag{6.9}$$

If one of the distances is given, say $d_{0,2}$ in this case, then the optimization reduces even further to equation (6.10). Note that we have multiplied the argument by negative one, but the result remains the same because cosine is an even function. Combining the sum of sinusoids and minimizing results in the general solution for $d_{0,1}$ expressed in equation (6.11), where atan2 is the four-quadrant inverse tangent function.

$$\underset{d_{0,1}}{\text{argmin}} \quad \sigma^2_{e_{min}} = \underset{d_{0,1}}{\text{argmin}} \quad \frac{\cos(2\pi d_{0,1})}{\sigma_0^2} + \frac{\cos(2\pi(d_{0,1} - d_{0,2}))}{\sigma_2^2} \tag{6.10}$$

$$d_{0,1} = \frac{1}{2} - \frac{1}{2\pi}\text{atan2}\left(-\sin(2\pi d_{0,2}), \frac{\sigma_2^2}{\sigma_0^2} + \cos(2\pi d_{0,2})\right) \tag{6.11}$$

Figure 6-2 displays the predicted optimal distance $d_{0,1}$ as a function of σ_2^2 for $d_{0,2} = 1/3$ and $\sigma_0^2 = \sigma_1^2 = 1$. The constellation with $d_{0,2} = 1/3$ and $d_{0,1} = -1/3$ is equivalent to uniform sampling. When $\sigma_2^2 = 1$, all three error sources have equal variance, and the optimal sampling constellation is uniform. This matches our earlier result. When $\sigma_2^2 \to 0$, the optimal distance $d_{0,1} \to -1/6$ which maximizes the distance between the low precision and high precision error sources $d_{1,2} = -d_{0,1} + d_{0,2} \to 1/2$. As σ_2^2 increases, $d_{0,1} \to -1/2$, now maximizing the distance between the two high precision error sources.

This example suggests a general characteristic of optimal sampling constellations. Channels with the largest time separation require the highest accuracy.

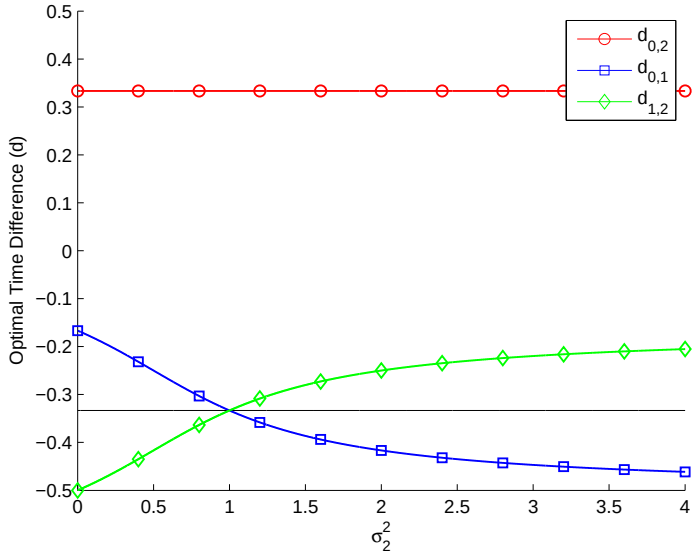


Figure 6-2: Predicted optimal distance $d_{0,1}$ as a function of σ_2^2 for $d_{0,2} = 1/3$ and $\sigma_0^2 = \sigma_1^2 = 1$.

6.2.1.1 Large Discrepancies in Quantization Error Variance

From Figure 6-2, it is apparent that even a small mismatch in channel error variances will produce a disproportionate change in the optimal time differences predicted by the additive noise model. The case of large discrepancies between quantizer error variances is important to consider for practical design scenarios. Time-interleaved ADCs with uniform quantizers are constrained to have an integer number of bits. The error variance of a quantizer is reduced by a factor of four for every added bit. Thus, the traditional design space of available quantizer error variances is quite limited. If there is a difference between two quantizers, their error variance will always differ by a factor of four.

We will consider the same example as in the previous section, except with no constraints on $d_{0,2}$. Recall that $\sigma_0^2 = \sigma_1^2 = 1$ and σ_2^2 varies. Figure 6-3 displays the numerically optimized error variance function, which reveals some interesting properties of the optimal sampling locations.

First, we consider the region where σ_2^2 is much lower than σ_0^2 and σ_1^2, indicating that channel 2 corresponds to a high precision quantizer. In this case, the distances between low precision and high precision channels $d_{0,2}$ and $d_{1,2}$ are maximized. This comes at the expense of sampling in very close proximity with the low precision devices ($d_{0,1} = 0$). This result is easily verified by evaluating equation (6.9) as $\sigma_0^2 \gg \sigma_2^2$ and $\sigma_1^2 \gg \sigma_2^2$. The first term becomes negligible, allowing us to factor out the shared term σ_2^2 to obtain equation (6.12).

$$\underset{d_{0,1},d_{0,2}}{\operatorname{argmin}} \quad \sigma_{e_{min}}^2 \approx \underset{d_{0,1},d_{0,2}}{\operatorname{argmin}} \quad \frac{\cos(2\pi d_{0,2})}{\sigma_0^2} + \frac{\cos(2\pi d_{1,2})}{\sigma_1^2} \tag{6.12}$$

The two terms in this equation can be minimized independently, leaving us with the optimal distances $d_{0,2} = \pm\frac{1}{2}$ and $d_{1,2} = \pm\frac{1}{2}$.

As σ_2^2 increases, channels 0 and 1 become the high precision measurement devices, and $d_{0,1}$ is maximized in Figure 6-3. Now, the distance between the high precision channels is maximized. This can be verified by evaluating equation (6.9) as $\sigma_2^2 \gg \sigma_0^2$ and $\sigma_2^2 \gg \sigma_1^2$ in equation (6.9). Now only the first term remains, as shown in equation (6.13).

$$\underset{d_{0,1},d_{0,2}}{\operatorname{argmin}} \quad \sigma_{e_{min}}^2 \approx \underset{d_{0,1},d_{0,2}}{\operatorname{argmin}} \quad \cos(2\pi d_{0,1}) \tag{6.13}$$

This minimization yields $d_{0,1} = \pm\frac{1}{2}$. Equation (6.13) also suggests that the other distances are essentially meaningless. Of course, this is only the case for extremely large σ_2^2.

When all three error variances are equivalent, the optimal distance between each of the channels is $d_{0,1} = d_{0,2} = d_{1,2} = \frac{1}{3}$, which corresponds to uniform sampling. This can be verified through the same method using equation (6.9). Since all of the sinusoid weights are equal in this case, we obtain equation (6.14).

$$\underset{d_{0,1},d_{0,2}}{\text{argmin}} \quad \sigma^2_{e_{min}} = \underset{d_{0,1},d_{0,2}}{\text{argmin}} \quad \cos(2\pi d_{0,1}) + \cos(2\pi d_{0,2}) + \cos(2\pi(d_{0,2} - d_{0,1})) \tag{6.14}$$

The optimal distances suggested by the minimum of this function are $d_{0,1} = -d_{0,2} = \pm\frac{1}{3}$, which agree with our numerical results.

These three configurations are illustrated in Figure 6-4, where (H) denotes a high precision channel and (L) denotes a low precision channel. We can think of the distance functions as a circle with circumference one where $\frac{1}{2} = -\frac{1}{2}$, because of the wrapping properties mentioned in equation (6.7) that result from the recurrent sampling structure.

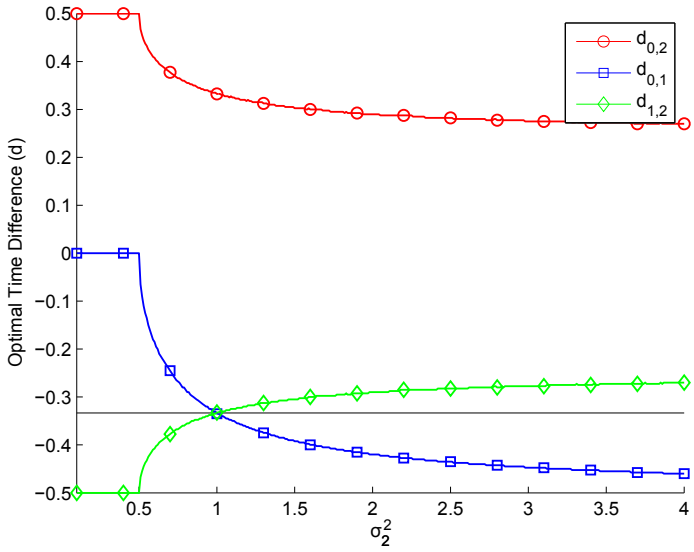
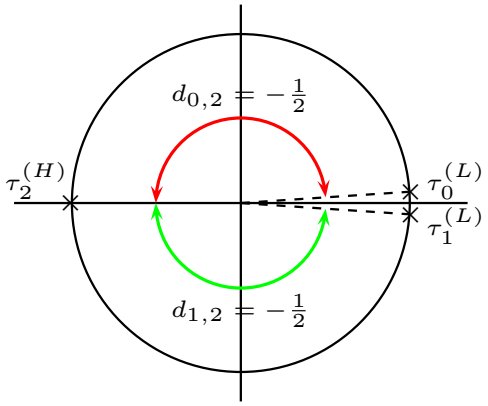


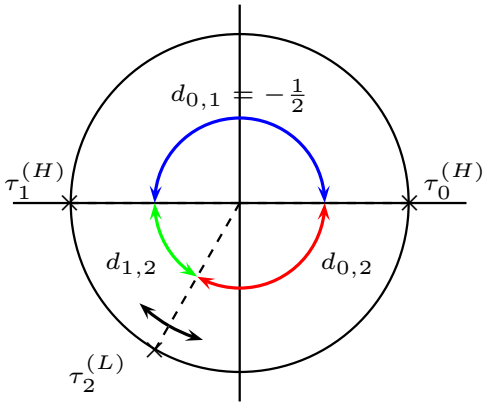
Figure 6-3: Predicted optimal distance $d_{0,1}$ and $d_{0,2}$ as a function of σ_2^2 for $\sigma_0^2 = \sigma_1^2 = 1$.

6.2.1.2 Table of Optimal Offsets

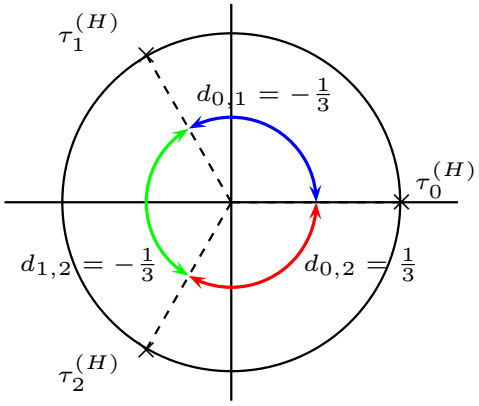
Understanding the three configurations in Figure 6-4 will allow us to easily design an optimal system for virtually any set of bit allocations when $M = 3$ and $L = 2$. Because channel error variances are typically the same or differ by multiples of four, many bit allocations are well approximated by one



(a) One (H) and two (L)



(b) Two (H) and one (L). Black arrow indicates timing offset is irrelevant.



(c) Three (H)

Figure 6-4: Optimal distances between different configurations of high precision channels denoted by (H) and low precision channels denoted by (L).

of the three bit allocations mentioned in the previous section.

Table 6.1 displays the predicted optimal time differences for the three configurations. As the discrepancy in the number of quantizer bits grows larger, the approximations derived earlier will become more valid. In particular, when the second row corresponds to $\mathbf{b} = [10,\ 10,\ 8]$, $d_{0,1} = 0.5$

and the placement of $d_{0,2}$ is virtually irrelevant.

b_0	b_1	b_2	$d_{0,1}$	$d_{0,2}$
10	9	9	0.5	-0.5
10	10	9	0.46	-0.27
10	10	10	0.33	-0.33

Table 6.1: Predicted optimal time differences for different bit allocations for $M = 3$ and $L = 2$.

Even with a discrepancy of only one bit, the approximations appear to hold closely. Consider the case when $M = 3$, $L = 2$, and $b_0 = b_1 = k$ with the additive noise model. We evaluate all possible configurations of timing offsets (on a discrete grid) for b_2 between $k - 2$ and $k + 2$ bits. Figure 6-5 displays the reduction factor γ as a function of these time distances between channels. The optimization surface hardly changes from $b_2 = k - 2$ to $b_2 = k - 1$, while there is a large difference between $b_2 = k - 1$ and $b_2 = k$. This fact allows us to consider only the three cases mentioned in Table 6.1. We will discuss this example in more detail in the next chapter.

6.2.2 Error Minimization for the Case of $L = 1$

It is relatively simple to analytically optimize equation (6.4) for the case when $L = 1$. In this case, the predicted error function reduces to equation (6.15).

$$\sigma^2_{e_{min}} = tr\left(\left[\sum_{m=0}^{M-1} \frac{1}{\sigma_m^2}\right]^{-1}\right) = \frac{1}{\sum_{m=0}^{M-1} \frac{1}{\sigma_m^2}} \tag{6.15}$$

This result reveals that for a given bit allocation with any number of channels and $L = 1$, the location of the timing offsets does not matter in terms of performance. In this case, the signal is critically sampled in each channel, enabling perfect reconstruction in absence of quantization error from a single channel. Intuitively, the high density of sampling locations eases constraints on the timing offsets. For high oversampling factors $M > L = 1$, we will see that this property of "volatility" among optimal sampling configurations is common.

6.2.3 Error Minimization for the Case of $L = M = 2$

We have already obtained a closed form solution for the case when $L = 2$ in equation (6.6). When $M = 2$, the simplified form contains two cosine terms. These terms are easily reduced, yielding equation (6.16).

$$\underset{d_{0,1}}{\operatorname{argmin}} \quad \sigma^2_{e_{min}} = \underset{d_{0,1}}{\operatorname{argmin}} \quad \frac{\cos(2\pi d_{0,1})}{\sigma_0^2\sigma_1^2} + \frac{\cos(2\pi(-d_{0,1}))}{\sigma_0^2\sigma_1^2} = \underset{d_{0,1}}{\operatorname{argmin}} \quad \cos(2\pi d_{0,1}) \tag{6.16}$$

The resulting optimal distance is $d_{0,1} = \pm\frac{1}{2}$, which corresponds to uniform sampling. We will see that when the oversampling factor $\frac{M}{L} = 1$, the optimal configuration is always uniform (with the exception of $L = M = 1$).

When $M = L$, the optimal reconstruction filters correspond to the traditional Lagrange interpolation function discussed in Chapter 2 ([18],[16]). In [26], Yen showed that the maximum value of the Lagrange interpolation impulse response becomes very large when channel timing offsets "bunch" together. This is especially true for the impulse responses of channels near a bunching, or a gap produced by bunching. This phenomenon is caused by the $(t_m - t_q)$ term in the denominator of equation 2.3. Yao and Thomas concluded that such expansions are unstable, because small corruptions of sample values may lead to large changes in the reconstructed function [25].

Figure 6-6 displays the optimal reconstruction impulse responses for $M = L = 3$ and $\boldsymbol{b} = [10, 10, 10]$ with different timing constellations. The top row depicts impulse responses for channels 0-2 with $\boldsymbol{\tau} = \left[-\frac{2}{3}T_N, 0, \frac{2}{3}T_N\right]$, which corresponds to uniform sampling. Note that the responses all have approximately the same maximum amplitude. The bottom row depicts impulse responses for channels 0-2 with $\boldsymbol{\tau} = \left[-\frac{1}{10}T_N, 0, \frac{1}{10}T_N\right]$. This constellation consists of a "bunch" of timing offsets near the same point. In order to achieve perfect reconstruction using the Lagrange kernel, the maximum amplitudes are now sharply increased.

If the reconstruction impulse response has a large value, any error in the sample value will be amplified. Intuitively, for homogeneous bit allocations it is not beneficial to rely on any sample more than any other, because all quantizers have the same resolution. For this reason, the uniform sampling case is optimal, because the impulse responses give equal importance to all measurements. For a diverse bit allocation, it is intuitively less clear that the uniform sampling configuration is optimal. We will show numerically in the next section that this is the case.

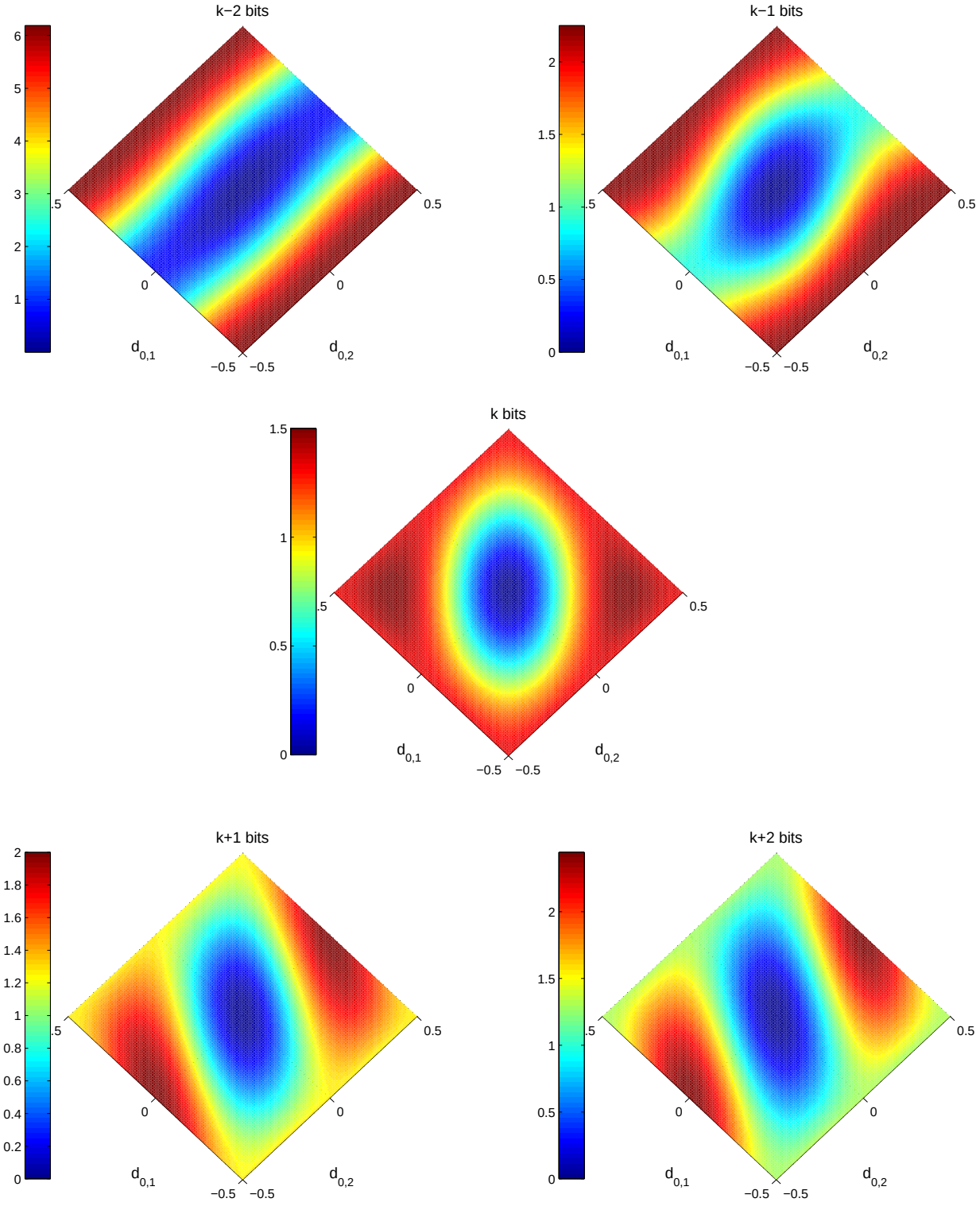


Figure 6-5: Reduction factor γ for additive white noise model as a function of distances for k bits in channels 0 and 1 and $k-2$ to $k+2$ bits in channel 2.

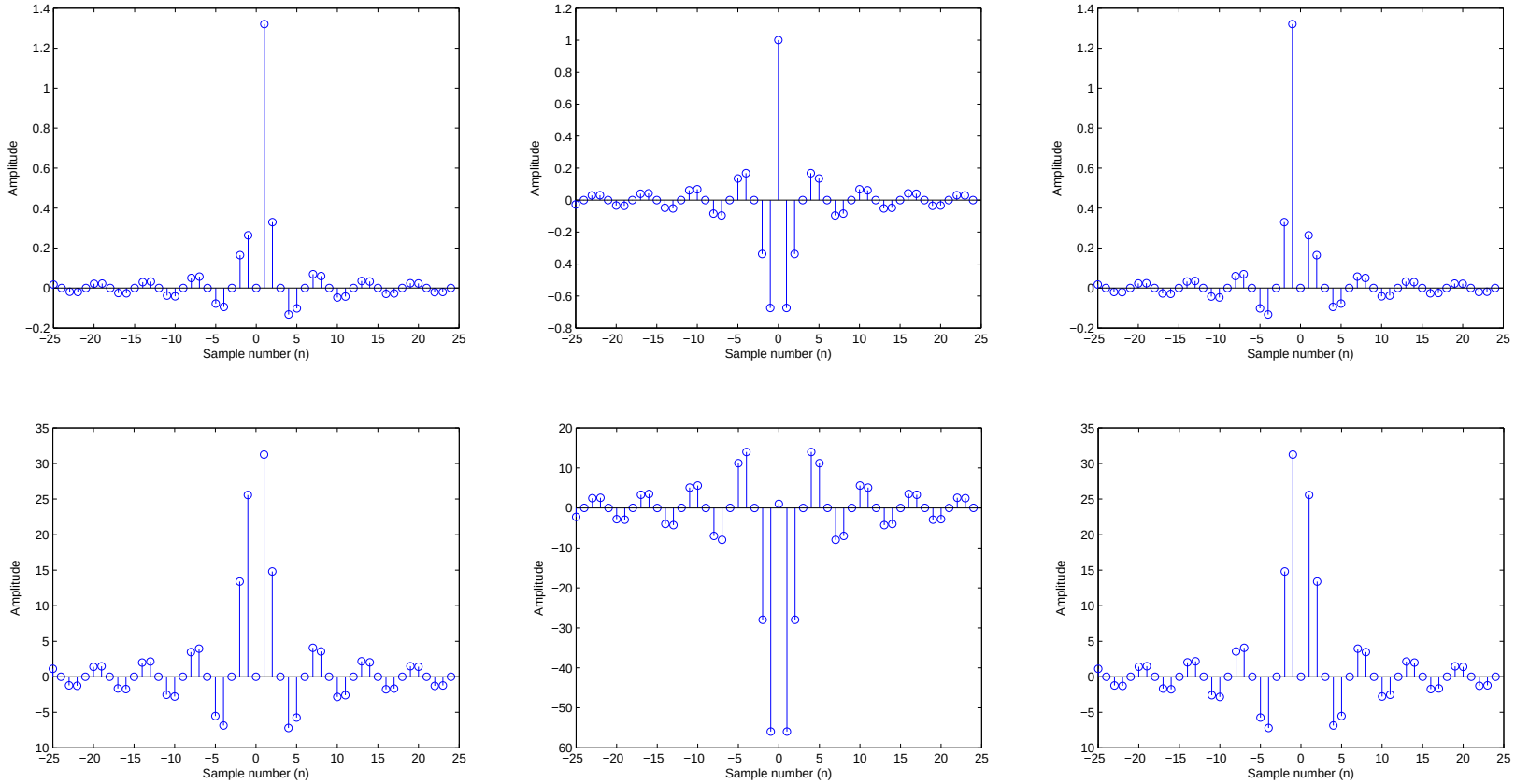


Figure 6-6: Optimal reconstruction impulse responses for $M = L = 3$ and $\boldsymbol{b} = [10, 10, 10]$ with different timing constellations. The top row depicts impulse responses for channels 0-2 with $\boldsymbol{\tau} = \left[-\frac{2}{3}T_N, 0, \frac{2}{3}T_N\right]$. The bottom row depicts impulse responses for channels 0-2 with $\boldsymbol{\tau} = \left[-\frac{1}{10}T_N, 0, \frac{1}{10}T_N\right]$.

6.3 Numerical Optimization of Predicted Error

For $L > 2$, the predicted error variance equation (6.4) is difficult to optimize analytically because of the inverse matrix operation. In this section, we will discuss numerical methods for optimization and their results.

Optimization of equation (6.4) is highly dependent on the parameters M and L that determine the oversampling ratio $\frac{M}{L}$. We show that the oversampling factor $\frac{M}{L}$ is a crucial factor that determines the number of optimal solutions and their overall performance. When the oversampling is one, corresponding to critical sampling, the optimal configurations are exactly equivalent to the uniform sampling grid. As the oversampling factor increases, the set of optimal configurations becomes much larger and much more variable.

6.3.1 Genetic Algorithm

One way to numerically find the global minimum of a cost function is to evaluate the function sufficiently finely that one of the measurements is approximately equal to the analytical optimum. This requires an exhaustive search that is very expensive in terms of time and computation power. As the dimensionality of the optimization surface grows, the volume of the search space increases exponentially.

Many more sophisticated optimization techniques exist that avoid these difficulties. The "genetic algorithm" models biological processes to optimize complex functions. We have chosen to use the genetic algorithm because it does not require derivative information, is suited for extremely complex cost surfaces, and deals with a large number of parameters [11]. The algorithm is initialized with a population of randomly chosen samples. The "fitness level," or associated cost of each sample, is used to select the set of best performing samples while eliminating the rest. The "next generation" of samples is created by "mutating" existing samples with random noise or combining the vector entries of two surviving samples. From here, the process repeats until the change in the average fitness level is less than a pre-designated tolerance or the maximum number of generations is exceeded. For a good review of the literature concerning genetic algorithms, one can refer to [8] or [5].

For the purposes of this thesis, we will use the MATLAB function *ga* for finding the minimum

of a function using the genetic algorithm, with default settings. The random initial population is chosen with a uniform distribution, and two samples are guaranteed to survive each iteration. The fraction of children produced by crossover combinations of parents is 0.8. The mutation is modeled with additive Gaussian noise with zero mean. The algorithm is said to converge if the average fitness level changes by less than 10^{-6} between iterations. The maximum number of generations is 100.

6.3.2 Near Critically Sampled Systems

In the previous section, we analytically showed that the optimal timing offsets τ_m can be almost arbitrarily chosen for a multi-channel system with a high oversampling factor ($L = 1$) and given bit allocation b_m. In other words, the choice of timing offsets has no significant bearing on the performance in this extreme case. Now, we consider the opposite extreme: a system that is critically sampled ($M = L$). Recall that we examined the case where $M = L = 2$ analytically and found that the optimal timing offsets always correspond to uniform sampling.

To explore the behavior of systems with more than two channels ($M = L > 2$), we utilize the genetic algorithm discussed earlier. To obtain more accurate results, we run the algorithm 1000 times and keep only the best-performing set of timing offsets. We will use different numbers of bits in each channel to verify that the solution is independent of bit allocation in this special case. Specifically, we will set the bit allocation $\mathbf{b} = [10,\ 11,\ 12,\ 13]$ for $M = L = 4$, $\mathbf{b} = [10,\ 11,\ 12,\ 13,\ 14]$ for $M = L = 5$, and $\mathbf{b} = [10,\ 11,\ 12,\ 13,\ 14,\ 15]$ for $M = L = 6$.

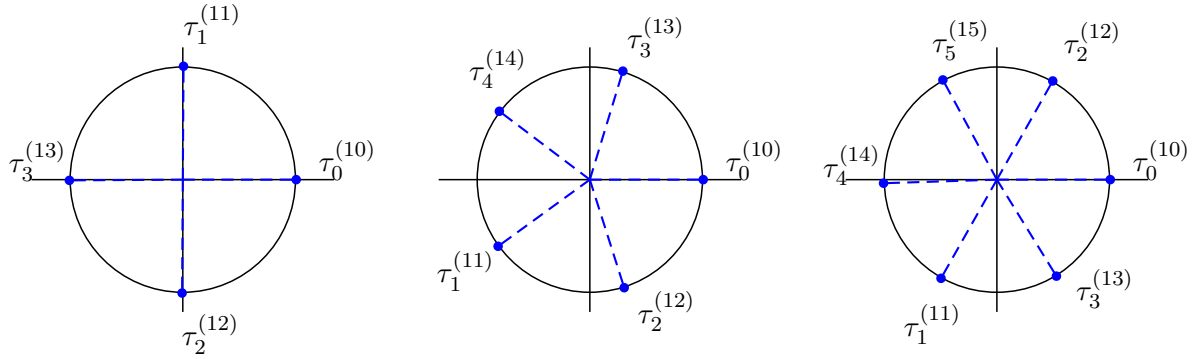
	$d_{0,1}$	$d_{0,2}$	$d_{0,3}$	$d_{0,4}$	$d_{0,5}$
$M = L = 4$	0.249	-0.251	-0.499	*	*
$M = L = 5$	-0.400	-0.201	0.203	0.397	*
$M = L = 6$	-0.332	0.168	-0.162	-0.495	0.329

Table 6.2: Numerically optimized offsets for $M = L = 4$ when $\mathbf{b} = [10,\ 11,\ 12,\ 13]$, $M = L = 5$ when $\mathbf{b} = [10,\ 11,\ 12,\ 13,\ 14]$, and $M = L = 6$ when $\mathbf{b} = [10,\ 11,\ 12,\ 13,\ 14,\ 15]$.

Table 6.2 lists the best sampling patterns after 1000 executions of the genetic algorithm. The sampling configurations are depicted in Figure 6-7. It is clear that each case corresponds to uniform sampling. The uniform sampling pattern occurs even though the channels have drastically different quantizer resolutions. Also, the optimal assignment does not seem to depend on the quantizer resolutions. For example, when $M = L = 6$, the two highest resolution quantizers are also nearest

to each other.

An explanation for this phenomenon is that the Lagrange interpolation functions grow very large for non-uniform timing constellations, amplifying quantization error. An oversampled system eases this restriction, allowing for more sophisticated constellations of timing offsets.



(a) $M = L = 4$ (b) $M = L = 5$ (c) $M = L = 6$

Figure 6-7: Optimal distances for $M = L$ always correspond to a uniform sampling pattern.

6.3.3 Clusters of Low-Precision Channels

One common property of optimal sampling configurations is that low-precision channels tend to form "clusters" of closely spaced samples in time. Often, these "clusters" behave in the same way as a single high-precision channel. For example, a common optimal sampling configuration with k high-precision quantizers and n clusters of low-precision quantizers will feature the two types equally spaced around the unit circle. In this section, we investigate the behavior of optimal timing offset constellations for various combinations of low- and high-precision channels. We will also develop an intuitive explanation for the observed behavior.

To gain insight about typical optimal sampling patterns, we will numerically optimize the predicted error variance for $M = 6$ and $L = 4$. Again, we will retain the best performing result after 1000 executions of the genetic algorithm. Table 6.3 displays the results for various bit allocations. Each case includes a different number of high- and low-precision channels. Table 6.4 lists the error variance and reduction factor associated with each bit allocation. From [18], we know that the analytically optimal case for $\mathbf{b} = [10,\ 10,\ 10,\ 10,\ 10,\ 10]$ is the uniform sampling pattern with reduction factor $\gamma = \frac{3}{2}$. It is encouraging to note that the genetic algorithm produces almost exactly

the same result even for six dimensions, with $\gamma = 1.499$.

For comparison purposes, Table 6.4 lists the best performing configuration for each bit allocation when the timing offsets are constrained to be uniform. Although the offsets are uniform, it is necessary to consider all permutations of bit allocations. For example, one ordering consists of all high-precision quantizers next to each other, while in another ordering the high- and low-precision are interleaved.

Figure 6-8 displays the best reduction factor for uniform and non-uniform timing offsets with each bit allocation. Bit allocations with more diversity among quantizer granularities generally have optimal timing configurations that differ the most from the uniform sampling grid. In these cases, performance can be improved significantly by allowing non-uniform sampling.

Note that the reduction factor γ is a measure that compares relative *noise reduction capability*, rather than overall noise variance at the output.

b	$d_{0,1}$	$d_{0,2}$	$d_{0,3}$	$d_{0,4}$	$d_{0,5}$
[10, 9, 9, 9, 9, 9]	-0.209	-0.348	0.340	-0.496	0.206
[10, 10, 9, 9, 9, 9]	-0.252	-0.489	0.493	0.253	0.246
[10, 10, 10, 9, 9, 9]	0.500	0.249	-0.251	-0.247	-0.242
[10, 10, 10, 10, 9, 9]	0.266	0.494	-0.235	0.136	-0.384
[10, 10, 10, 10, 10, 9]	-0.432	-0.239	0.379	0.190	-0.117
[10, 10, 10, 10, 10, 10]	0.168	-0.158	-0.491	0.337	-0.323

Table 6.3: Numerically optimized offsets for $M = 6$ and $L = 4$ with various bit allocations.

	Uniform		Non-Uniform	
b	$\sigma^2_{e_{min}}$	γ	$\sigma^2_{e_{min}}$	γ
[10, 9, 9, 9, 9, 9]	7.06e-7	1.575	6.83e-7	1.630
[10, 10, 9, 9, 9, 9]	5.65e-7	1.688	4.77e-7	1.998
[10, 10, 10, 9, 9, 9]	4.35e-7	1.829	3.44e-7	2.307
[10, 10, 10, 10, 9, 9]	3.18e-7	2.00	2.83e-7	2.249
[10, 10, 10, 10, 10, 9]	2.65e-7	1.800	2.42e-7	1.967
[10, 10, 10, 10, 10, 10]	2.12e-7	1.500	2.12e-7	1.499

Table 6.4: Predicted reduction factor for $M = 6$ and $L = 4$ with optimal uniform and non-uniform timing offsets, for various bit allocations.

The results of the numerical optimization are easier to interpret in Figure 6-9. With a single high-precision and five low-precision channels, the low-precision quantizers are distributed with a

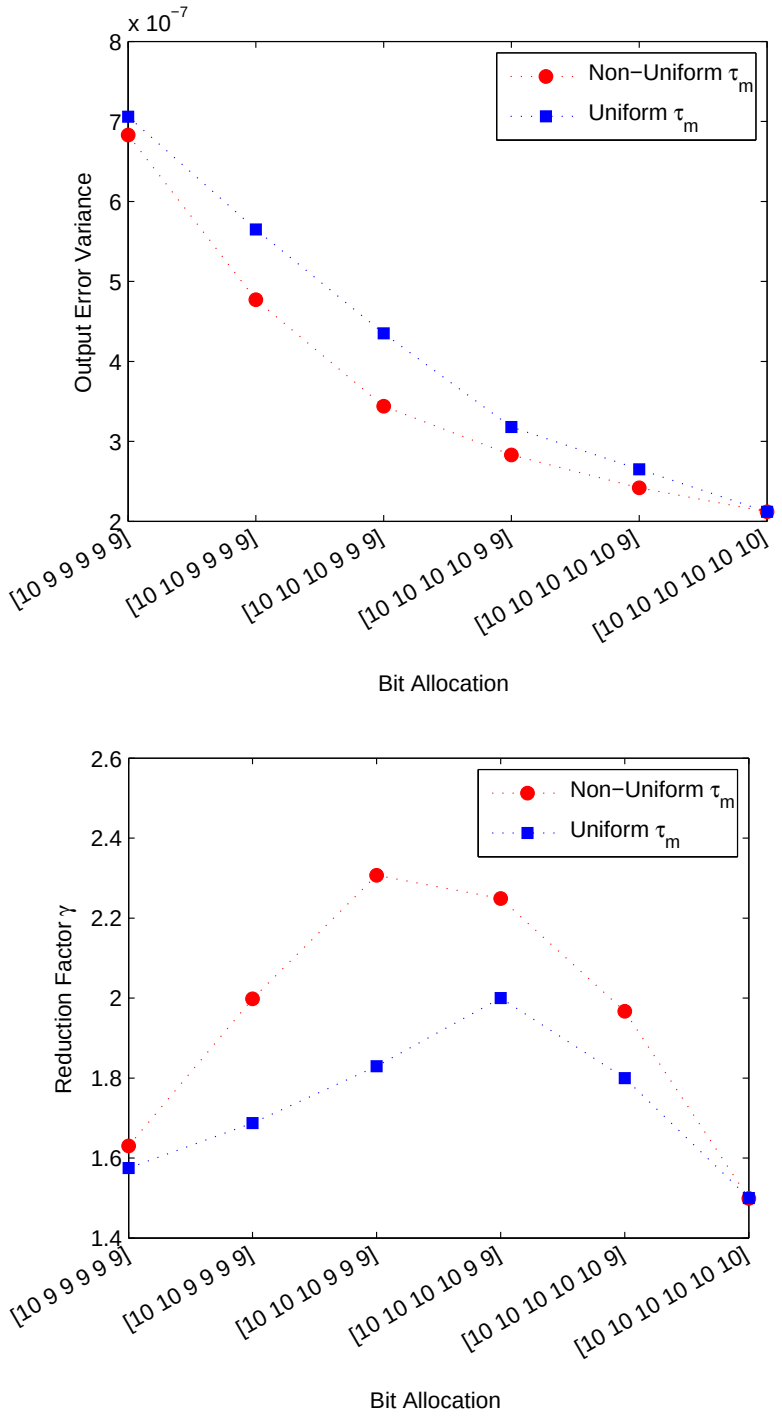


Figure 6-8: Comparison between predicted optimal timing offsets τ_m and uniform offsets τ_m for $M = 6$ and $L = 4$.

uniform distance between them that is smaller than the distance to the high-precision quantizer. This is illustrated in Figure 6-9a. Intuitively, we expect the large number of low-precision quantizers to provide measurement capability that is significant, even when compared with the single high-precision channel.

However, when two or three high-precision channels are present, the measurement capability of the low-precision quantizers is dwarfed. To account for this mismatch, "clusters" of low-precision quantizers are formed that act as a single high-precision channel. For example, in Figure 6-9c three nine bit quantizers act as a cluster equivalent to a single ten bit quantizer. It is interesting to note in Figures 6-9b and 6-9c that the optimal sampling pattern resembles a four-point constellation that corresponds to critical sampling (because $L = 4$).

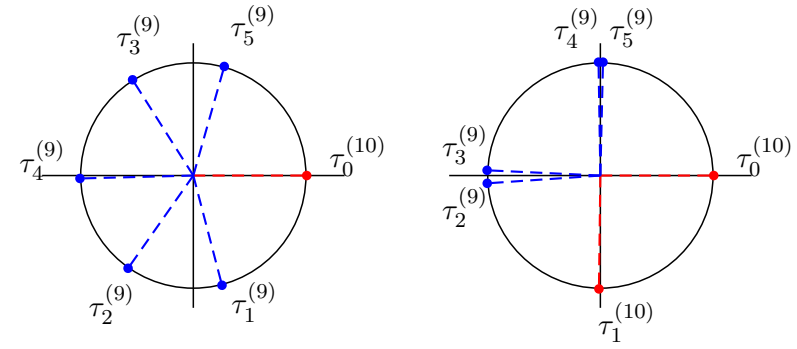
As the number of high-precision increases, the measurement capability of the low-precision channels becomes almost negligible. Figures 6-9d and 6-9e show that the high-precision channels are separated uniformly, while the low-precision channels are simply assigned to the midpoint between two high-precision channels.

Finally, when the channels all correspond to high-precision quantizers, the optimal timing offset constellation is uniform, as expected. This constellation is depicted in Figure 6-9f.

In summary, the measurement capability of low-precision channels becomes significant when they greatly outnumber high-precision channels. In this case, low-precision quantizers will be separated with some uniform distance that is slightly less than the distance to the high-precision quantizers. When the number of low-precision and high-precision channels is approximately equal, "clusters" of low-precision quantizers will form that act as a single high-precision channel. When high-precision channels outnumber low-precision channels, the high-precision channels will have approximately uniformly spaced timing offsets.

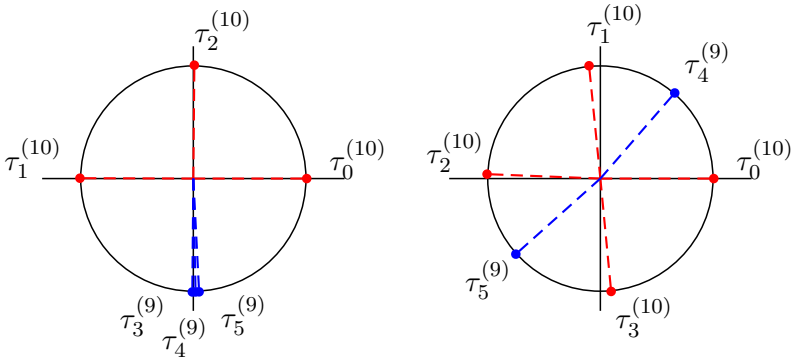
A small oversampling factor $\frac{M}{L}$ will tend to bias solutions to the uniform grid and reduce the "clustering" behavior.

To verify these conclusions, we will consider the same experiment for $M = 9$ and $L = 6$. Tables 6.5 and 6.6 display the best numerical optimization results after 1000 executions of the genetic algorithm. Note that the oversampling factor M/L is the same as the previous experiment. As we discussed earlier, a lower oversampling factor would bias optimal solutions towards the uniform sampling pattern. The optimal timing offset constellations are depicted in Figure 6-11. It



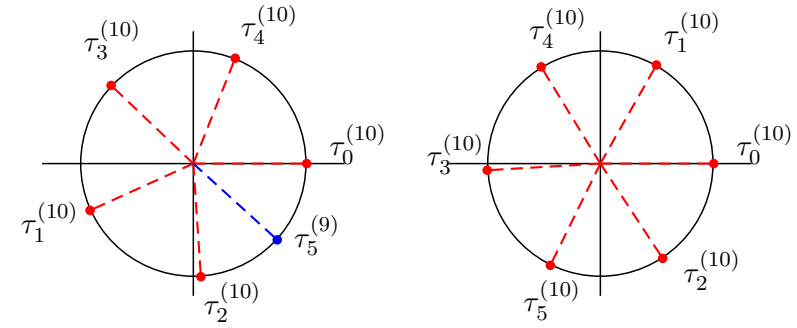
(a) One (10), Five (9)

(b) Two (10), Four (9)



(c) Three (10), Three (9)

(d) Four (10), Two (9)



(e) Five (10), One (9)

(f) Six (10)

Figure 6-9: Predicted optimal distances for $M = 6$ and $L = 4$ found through numerical optimization, for various bit allocations. The number of low- and high-precision channels are given in each subcaption. The number in parentheses represents the number of quantizer bits.

is reassuring to observe that our conclusions hold. We observe the same "clustering" behavior in this case. For example, in Figure 6-11e, four nine bit channels are placed in close proximity and

act as a single ten bit channel.

b	$d_{0,1}$	$d_{0,2}$	$d_{0,3}$	$d_{0,4}$	$d_{0,5}$	$d_{0,6}$	$d_{0,7}$	$d_{0,8}$
[1 (10), 8 (9)]	-0.146	-0.463	-0.343	-0.243	0.455	0.341	0.130	0.228
[2 (10), 7 (9)]	-0.147	0.322	-0.296	0.431	0.141	-0.460	0.220	0.228
[3 (10), 6 (9)]	0.343	-0.331	-0.148	0.171	-0.477	-0.161	0.178	0.496
[4 (10), 5 (9)]	-0.155	0.337	0.181	0.491	-0.425	-0.316	-0.307	0.490
[5 (10), 4 (9)]	-0.334	-0.496	0.328	-0.164	0.168	0.161	0.182	0.135
[6 (10), 3 (9)]	-0.364	0.330	-0.180	0.153	0.481	0.261	-0.095	-0.275
[7 (10), 2 (9)]	-0.277	-0.433	0.273	0.425	0.134	-0.131	0.396	-0.385
[8 (10), 1 (9)]	0.220	-0.133	0.117	0.344	-0.377	0.492	-0.259	0.432
[9 (10), 0 (9)]	-0.107	0.463	-0.432	-0.219	0.118	0.347	-0.321	0.221

Table 6.5: Numerically optimized offsets for $M = 9$ and $L = 6$ with various bit allocations.

	Uniform		Non-Uniform	
b	$\sigma^2_{e_{min}}$	γ	$\sigma^2_{e_{min}}$	γ
[1 (10), 8 (9)]	7.53e-7	1.547	7.41e-7	1.573
[2 (10), 7 (9)]	6.59e-7	1.607	6.21e-7	1.705
[3 (10), 6 (9)]	5.65e-7	1.688	4.81e-7	1.981
[4 (10), 5 (9)]	4.78e-7	1.772	3.90e-7	2.174
[5 (10), 4 (9)]	3.96e-7	1.872	3.19e-7	2.324
[6 (10), 3 (9)]	3.18e-7	2.000	2.84e-7	2.242
[7 (10), 2 (9)]	2.83e-7	1.875	2.55e-7	2.075
[8 (10), 1 (9)]	2.47e-7	1.714	2.32e-7	1.826
[9 (10), 0 (9)]	2.12e-7	1.500	2.13e-7	1.493

Table 6.6: Predicted reduction factor for $M = 9$ and $L = 6$ with optimal uniform and non-uniform timing offsets, for various bit allocations.

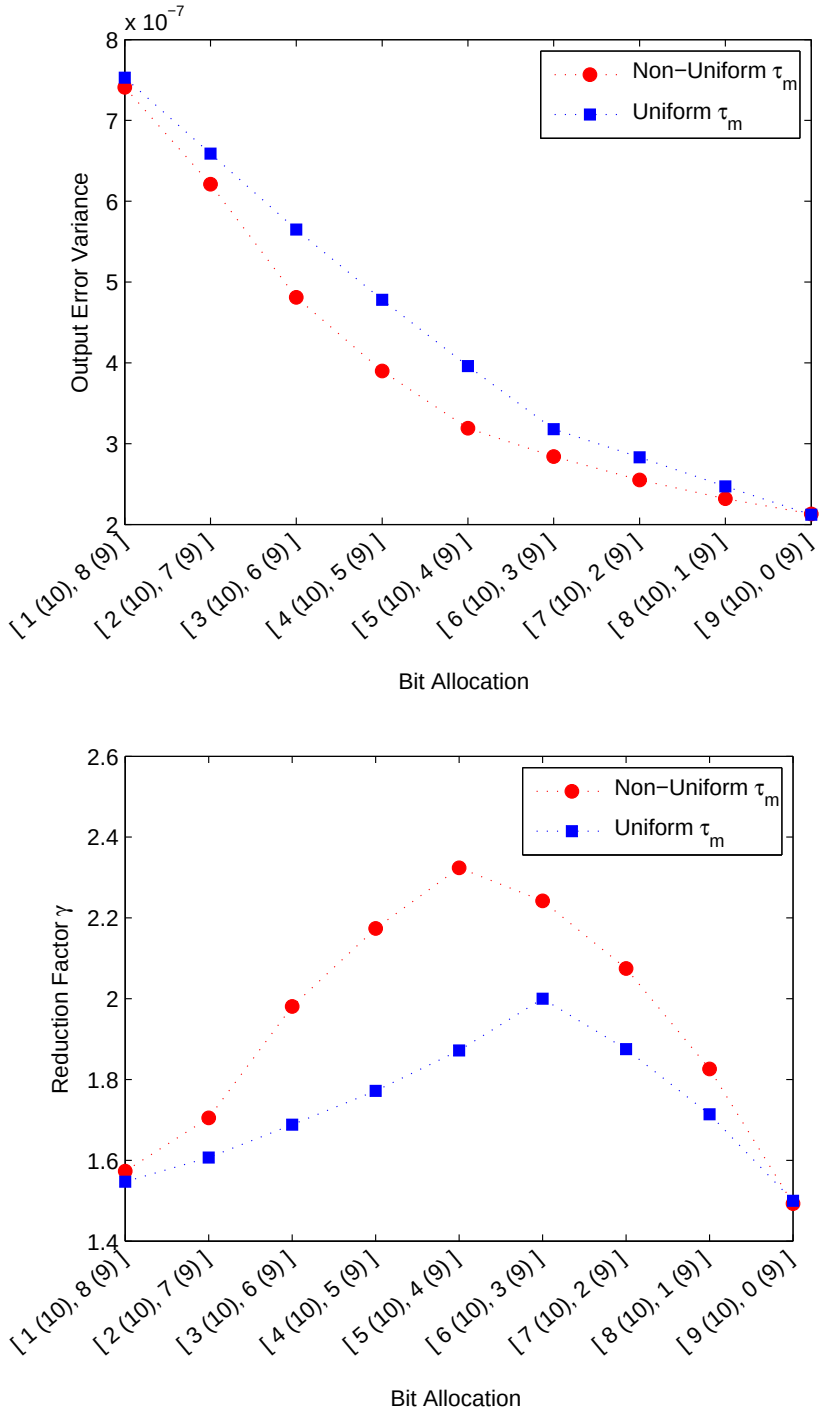
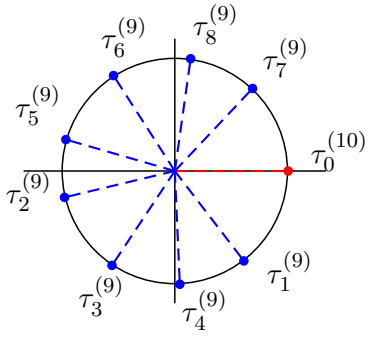
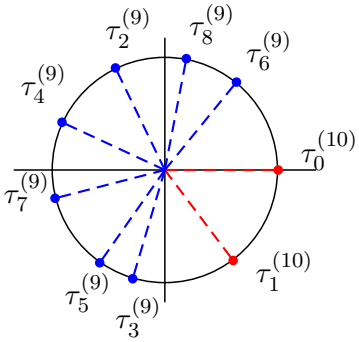


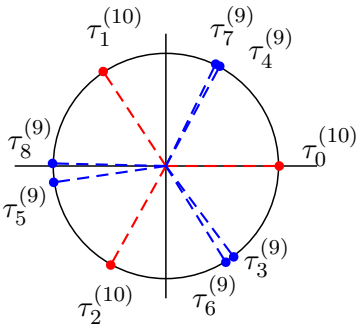
Figure 6-10: Comparison between predicted optimal timing offsets τ_m and uniform offsets τ_m for $M = 9$ and $L = 6$.



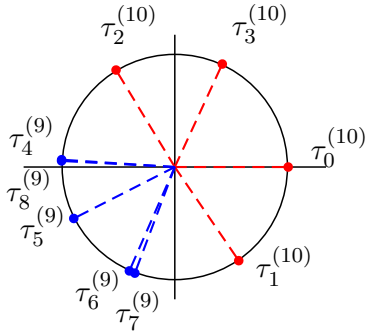
(a) One (10), Eight (9)



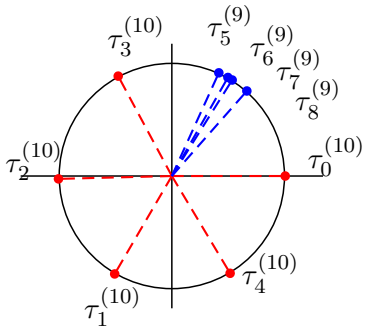
(b) Two (10), Seven (9)



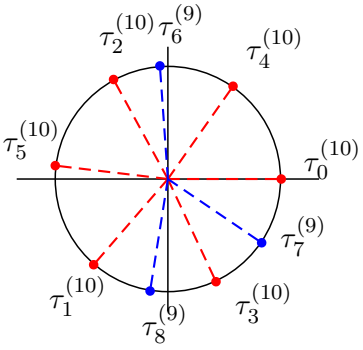
(c) Three (10), Six (9)



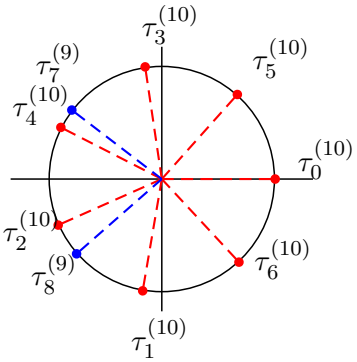
(d) Four (10), Five (9)



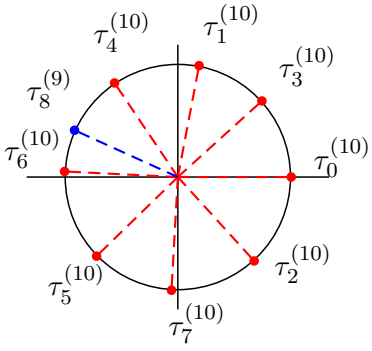
(e) Five (10), Four (9)



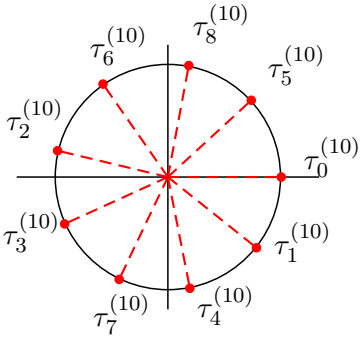
(f) Six (10), Three (9)



(g) Seven (10), Two (9)



(h) Eight (10), One (9)



(i) Nine (10)

Figure 6-11: Predicted optimal distances for $M = 9$ and $L = 6$ found through numerical optimization, for various bit allocations. The number of low- and high-precision channels are given in each subcaption. The number in parentheses represents the number of quantizer bits.

Chapter 7

A Framework for ADC Design Incorporating Correlation Correction

In Chapter 6, we developed an understanding of optimal timing constellations based on analysis using the additive noise model. In this chapter, we extend the design process to account for uniform quantizer error sources. In Chapter 4, we derived a set of acceptable time differences to avoid performance degradation from quantizer error correlation. The acceptable time differences in Table 4.3 are dependent on the quantizer granularity associated with the two channels of interest and the autocorrelation of the input signal. We found that close proximity in time between two identical quantizers will degrade performance, while close proximity between two quantizers with different numbers of bits can improve performance.

We concluded that the additive noise model captures the effects of uniform quantization well in the absence of cross-channel error correlation. Also, the optimization surface for the additive noise model is generally smooth. To avoid performance degradation due to error cross-correlation, we introduce a correction term that separates offsets within the minimal acceptable distance listed in Table 4.3. The design process is outlined below.

Quantizers with the same granularity should not have timing offsets in close proximity. Usually a tiny correction factor will separate the two channels to avoid performance degradation. The acceptable distances were experimentally determined for an approximately bandlimited white Gaussian process with critical sampling. These distances may need to be increased to account for the autocorrelation of the signal or a high oversampling factor.

When the optimal timing configuration for the additive noise model includes two quantizers with different granularities in close proximity, we statistically derived a method to unify the two with a weighted average. Unifying quantizers with different granularity in this way will take advantage of the extra information that their staggered comparator constellations provide. This can also be regarded as negative correlation between their error signals (refer to Chapter 5). This performance increase is not predicted by the additive noise model. Also, while simply assigning two different quantizers to the same sampling instant may yield some performance increase, the reconstruction filters were not designed to account for these cases since the quantization error is assumed to be uncorrelated. In general, the method derived in Chapter 5 will perform better.

After unifying quantizers with different granularities, we are left with a system with less than M channels. The effective resolution of the unified quantizer is now a fractional number of bits, given in Table 5.4. Thus, we must use this fractional number of bits with $M - 1$ channels to calculate the optimal reconstruction filters.

Framework for ADC Design:

1. **Optimize predicted sampling configuration for the additive noise model.**

   If $L = 2$, begin with the table of analytically derived optimal offsets from Table 6.1. If $L > 2$, numerically optimize the error variance in equation (6.2) for a given bit allocation b_m. Often, near-optimal configurations can be constructed by creating clusters of low-precision channels that behave as a single high-precision channel.

2. **Separate identical quantizers that violate acceptable distances.**

   If two channels with identical quantizers violate acceptable distances, we must manually separate the timing offsets. In Table 4.3, acceptable distances are given for a specific bit allocation with an input that is an approximately bandlimited white Gaussian random process. This is usually quite a small correction to ensure that channels satisfy the acceptable time distances.

3. **Unify different granularity quantizers that are in close proximity.**

   If two channels with different quantizers violate acceptable distances, we must unify the quantizers to obtain a better estimate of the underlying analog value. Again, Table 4.3 specifies acceptable distances for a specific bit allocation and input signal. Specifically, we combine the two quantized signals with the WLS weighted average defined in Chapter 5. The unified quantizer will have an effective number of bits defined in Table 5.4. Note that the optimal reconstruction filters must be calculated with fewer than M channels.

7.1 Simulated Results

7.1.1 Simulated Performance of Design Framework

In this section, we verify that the design technique outlined above produces good results by simulating the time-interleaved ADC. As a practical design example, consider the case when $M = 3$, $L = 2$, and $\mathbf{b} = [10,\ 9,\ 9]$ for an audio input signal. Table 6.1 lists the predicted optimal configuration as $d_{0,1} = -d_{0,2} = 0.5$, which means that channels 1 and 2 have identical sampling instants. Since both channels have the same granularity, we must separate them to avoid correlation. Although the acceptable distance for 9 bits is not listed in Table 4.3, it is clear that the acceptable distance for 3 bits of $0.06 \cdot T_N$ is a conservative upper bound that is still valid. To express this in the form of time distances (equation (6.7)), we divide by $L \cdot T_N$. This yields a normalized separation distance of 0.03. Thus, the corrected timing configuration can be written as $d_{0,1} = 0.485$ and $d_{0,2} = -0.485$.

Table 7.1 displays the measured results of the uncorrected and corrected timing offsets with uniform quantizers. As predicted, we observe a large degradation in performance if the channels corresponding to identical quantizers are not separated. However, just a minor correction results in near optimal performance.

	$\boldsymbol{\tau}$	Measured γ
Additive noise model	$[0,\ 0.5,\ -0.5] \cdot LT_N$	2.002
Uniform quantizer (uncorrected)	$[0,\ 0.5,\ -0.5] \cdot LT_N$	1.203
Uniform quantizer (with correction)	$[0,\ 0.485,\ -0.485] \cdot LT_N$	1.997

Table 7.1: Comparison of uncorrected and corrected timing configurations with uniform quantizers, for $M = 3$, $L = 2$, and $\mathbf{b} = [10,\ 9,\ 9]$ with an audio input signal.

Next, consider the case when $M = 9$, $L = 6$, and $\mathbf{b} = [10,\ 10,\ 10,\ 10,\ 9,\ 9,\ 9,\ 9,\ 9]$ for an audio input signal. The predicted optimal configuration is depicted in Figure 6-11d. Table 6.5 lists the predicted optimal configuration as $d_{0,4} = 0.491$ and $d_{0,8} = 0.490$. Since both channels have the same granularity, we must separate them to avoid correlation. In this example, we will separate the two channels by a slightly more aggressive normalized distance of 0.03. We can express the corrected distances as $d_{0,4} = -0.4945$ and $d_{0,8} = 0.4755$. Table 7.2 displays the measured results of the uncorrected and corrected timing offsets with uniform quantizers. Again, just a minor correction rectifies the discrepancy and results in near optimal performance.

	$\boldsymbol{\tau}$	Measured γ
Additive noise model	$d_{0,4} = 0.491,\ d_{0,8} = 0.490$	2.174
Uniform quantizer (uncorrected)	$d_{0,4} = 0.491,\ d_{0,8} = 0.490$	2.000
Uniform quantizer (with correction)	$d_{0,4} = -0.4945,\ d_{0,8} = 0.4755$	2.175

Table 7.2: Comparison of uncorrected and corrected timing configurations with uniform quantizers, for $M = 9$, $L = 6$, and $\mathbf{b} = [10,\ 10,\ 10,\ 10,\ 9,\ 9,\ 9,\ 9,\ 9]$ with an audio input signal.

Finally, we will examine a more complex scenario. Consider the case when $M = 9$, $L = 6$, and $\mathbf{b} = [10,\ 10,\ 10,\ 9,\ 9,\ 9,\ 8,\ 8,\ 8]$. The predicted optimal configuration is listed in Table 7.3 and depicted in Figure 7-1. Note that the low-resolution channels form "clusters" that behave as a single high-resolution channel. Specifically, the 8 and 9 bit channels lie in very close proximity. To obtain the optimal configuration for actual quantizers, we must unify these 8 and 9 bit quantizers. This correction will take advantage of the negative correlation discussed in Chapter 5.

Figure 7-2 displays the reduction factor γ for various levels of design sophistication. When the

timing offsets are constrained to be uniform, the best possible reduction factor is $\gamma = 2.65$. Allowing non-uniform timing offsets boosts performance significantly to $\gamma = 3.31$. Correcting the 8 and 9 bit channel timing offsets so that each pair has the same sampling instant yields another significant performance improvement to $\gamma = 3.91$. Finally, keeping the previous corrections and performing a weighted average defined in Chapter 5 will yield another small improvement, to $\gamma = 3.98$. Note that the last case actually involves a system with $M = 6$ and $L = 6$, where each of the unified quantizers has an effective resolution of 9.339 bits. These values were used to compute the optimal reconstruction filters.

In terms of the reduction factor γ, the unified configuration yields a 20.2% performance improvement versus the predicted optimal timing configuration, and a 50.2% performance improvement versus the uniform sampling pattern. This corresponds to a 16.8% reduction in output error variance versus the predicted optimal timing configuration, and a 33.4% reduction in output error variance versus the uniform sampling pattern.

$d_{0,1}$	$d_{0,2}$	$d_{0,3}$	$d_{0,4}$	$d_{0,5}$	$d_{0,6}$	$d_{0,7}$	$d_{0,8}$
-0.1595	0.3265	0.1667	-0.4949	-0.3227	-0.3380	-0.4990	0.1542

Table 7.3: Numerically optimized offsets for $M = 9$ and $L = 6$ with $\mathbf{b} = [10,\ 10,\ 10,\ 9,\ 9,\ 9,\ 8,\ 8,\ 8]$.

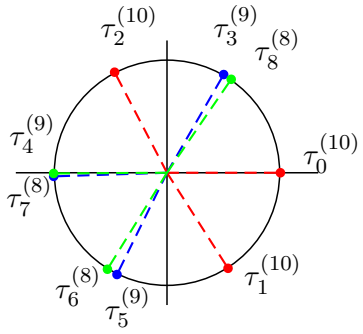


Figure 7-1: Predicted optimal timing offsets for $M = 9$ and $L = 6$ found through numerical optimization, for $\mathbf{b} = [10,\ 10,\ 10,\ 9,\ 9,\ 9,\ 8,\ 8,\ 8]$. The number of low- and high-precision channels are given in each subcaption. The number in parentheses represents the number of quantizer bits.

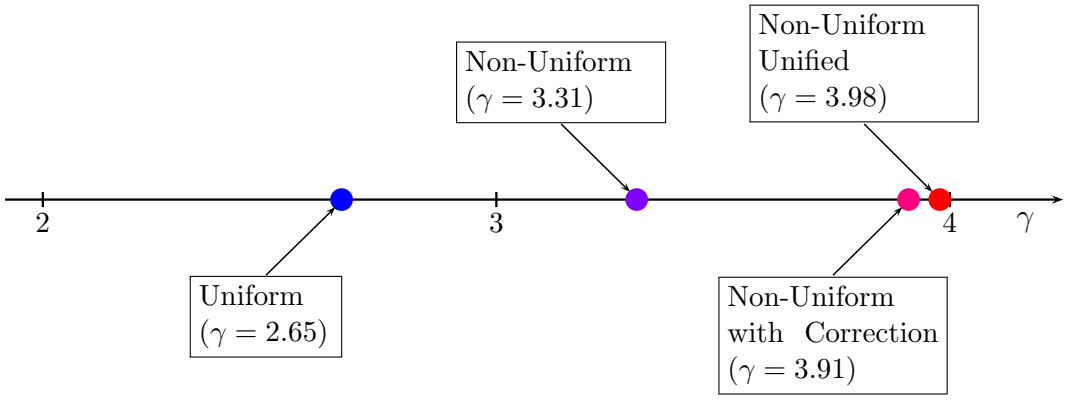


Figure 7-2: Best reduction factor γ for $\mathbf{b} = [10, 10, 10, 9, 9, 9, 8, 8, 8]$, when the timing offsets are: (a) constrained to be uniform, (b) non-uniform, (c) corrected so that quantizers with different granularities in close proximity are placed on the same time instant, and (d) unified so that quantizers with different granularities are combined with the WLS weighted average.

7.1.2 Simulated Optimization Surface

In the previous section, we considered some specific examples to verify our proposed design technique. It is also important to experimentally examine how the optimization surface changes as we introduce uniform quantizers instead of the additive noise model. Earlier we considered the case where $M = 3$, $L = 2$, and $b_0 = b_1 = k$ with the additive noise model. We will simulate the system with uniform quantizers and evaluate all possible configurations of timing offsets (on a discrete grid) for b_2 between $k-2$ and $k+2$ bits. For this simulation we will use $k = 10$. Figure 7-3 displays the simulated reduction factor γ with uniform quantizers as a function of distances for k bits in channels 0 and 1 and $k-2$ to $k+2$ bits in channel 2.

When $b_2 = k$, three narrow "grooves" of performance degradation are present when the sampling instants fall close to each other. Since each channel has identical quantizer granularity, we observe a decrease in performance. However, no correction term is necessary because the optimal timing offsets are already sufficiently separated.

When $b_2 = k + 1$, we observe grooves in the same locations. The most noticeable occurs when $d_{0,1} = 0$, causing a degradation in performance because channels 0 and 1 both have 10 bits. The optimal configuration lies near this groove, so we must add a corrective separation. A small separation easily resolves the performance degradation so that it is consistent with the additive

noise model. The other two grooves actually correspond with a performance increase, because of negative correlation between the 10 and 11 bit channels.

When $b_2 = k-1$, two grooves with positive performance are present near the predicted optimal constellation. Note that selecting a constellation on these lines will actually increase performance above the predicted maximum for the additive noise model. A third groove decreases performance, but this does not affect the optimal sampling configuration. No correction term is necessary.

It is encouraging to note that the optimization surface does not change significantly when uniform quantizers are introduced. In particular, the surface is still smooth, which means that a slight correction to timing offsets to avoid correlation will still remain near the optimal solution.

7.1.3 Extension to Nine Channels

We have also verified our earlier conclusions for more than three channels. Specifically, we have also simulated nine channels, to allow for comparison with current high performance ADCs designed in a traditional manner. With a higher number of channels, the sampling instants will be more densely packed.

Consider the case when $M = 9$, $L = 6$, $\mathbf{b} = [10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10,\ 10]$, and $\boldsymbol{\tau} = [-\frac{8}{3}T_N,\ -\frac{6}{3}T_N,\ -\frac{4}{3}T_N,\ -\frac{2}{3}T_N,\ \tau_1,\ \frac{2}{3}T_N,\ \frac{4}{3}T_N,\ \tau_2,\ \frac{8}{3}T_N]$, with uniform quantizers. We will vary the parameters τ_1 and τ_2 over a discrete grid of timing offsets.

Figure 7-4 displays the reduction factor γ as a function of delays τ_1 and τ_2. From [18], we expect that the optimal timing offsets will correspond to uniform sampling. This is verified by the simulation results.

Again, we observe the grooves of performance degradation when two channels fall in close proximity to each other. This provides further evidence that the same characteristic occurs for high numbers of channels.

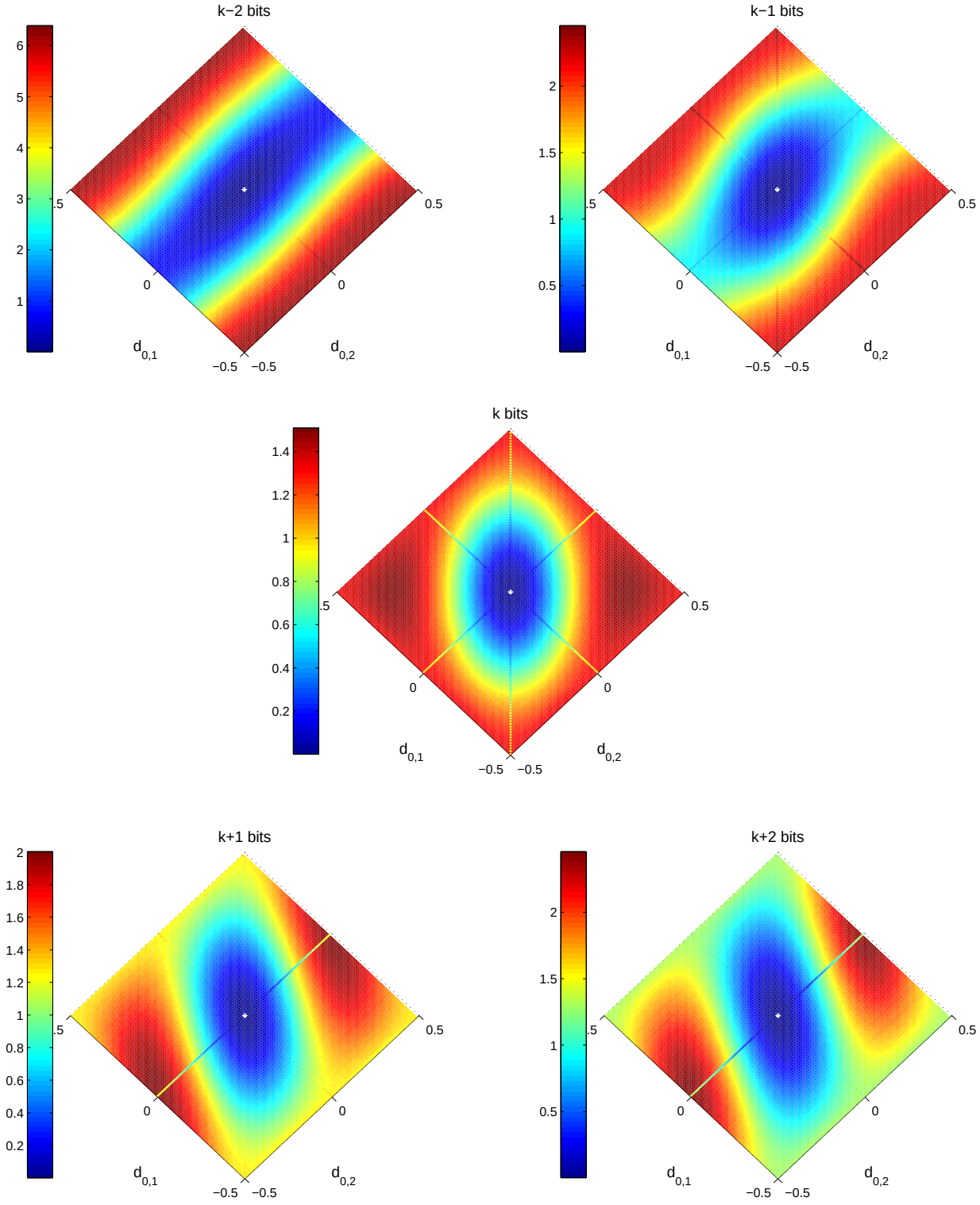


Figure 7-3: Simulated reduction factor γ with uniform quantizers as a function of distances for k bits in channels 0 and 1 and $k-2$ to $k+2$ bits in channel 2.

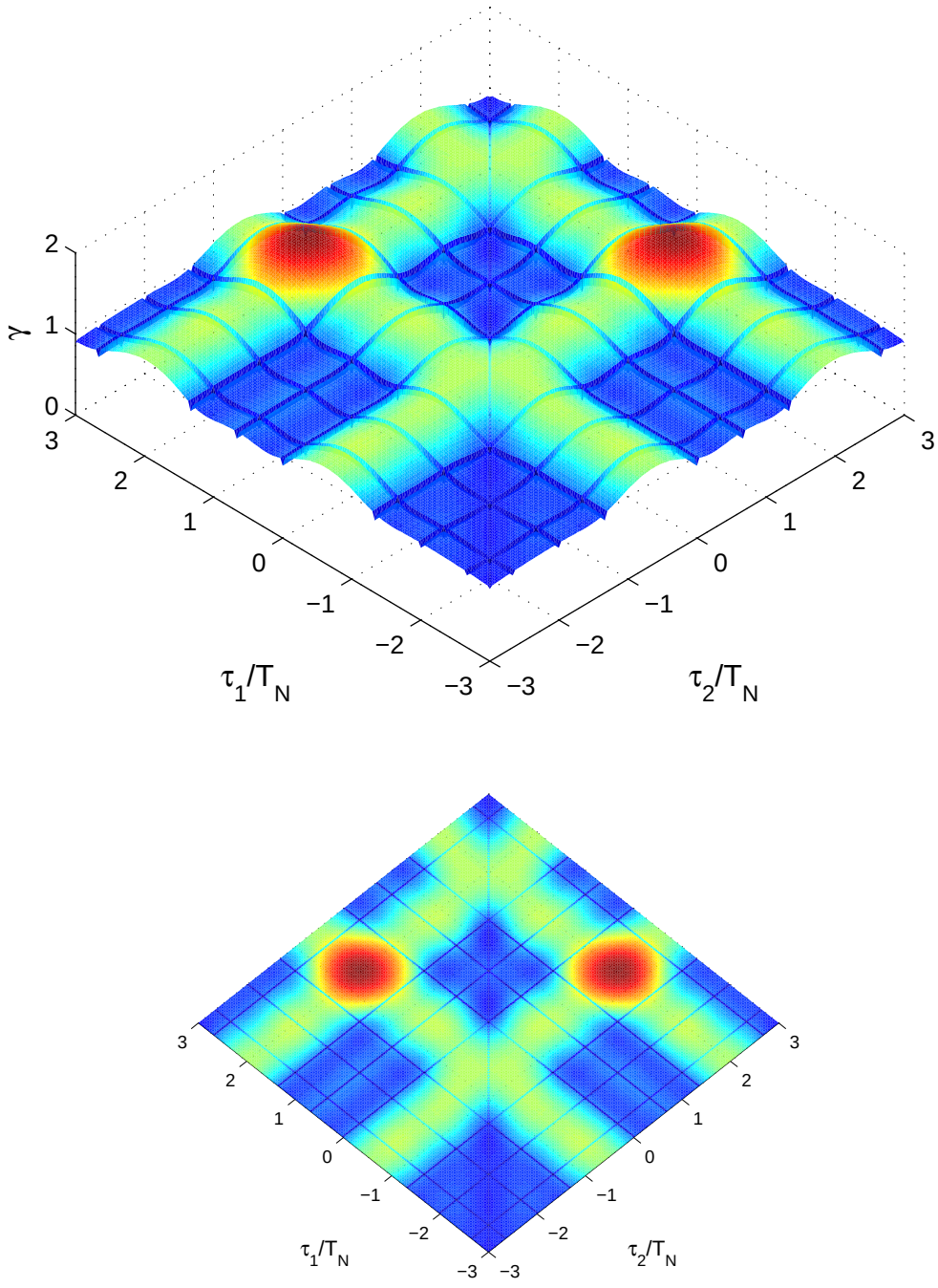


Figure 7-4: Reduction Factor γ as a function of delays τ_1 and τ_2 with uniform quantizers for $M = 9$ and $L = 6$.

Chapter 8

Optimization of Bit Allocation for Fixed Timing Offsets

In the previous chapter a framework has been developed to choose timing offsets τ_m given a fixed bit allocation b_m to maximize the overall SQNR. In this chapter, we discuss the problem of optimal choice of bit allocation b_m to maximize SQNR for given timing offsets τ_m. As shown in [18] and [16], with fixed timing offsets τ_m the SQNR is maximized by selecting a bit allocation to compensate for mismatches in the delays.

8.1 Analytical Optimization of Predicted Error

Intuitively, we might expect that the relative distance between channel timing offsets will define their optimal bit allocations. To verify this intuition analytically, we refer to equation (6.8). We attempt to optimize output error variance $\sigma^2_{e_{min}}$ by choosing the best set of quantizers for a given sampling constellation τ_m. Of course, we must add a constraint on the quantizers so that the result is not unreasonable. For example, setting $\sigma^2_m = 0$ in each channel will allow for perfect reconstruction, but this is not a practical design. One common constraint is to limit the total number of bits used for quantization in all channels (i.e. $\sum_{m=0}^{M-1} b_m = c$, where c is a positive constant). This is a good constraint in the environment of digital communications, where the goal is to achieve the lowest error variance with a fixed number of bits for transmission.

However, the cost of creating quantizers is usually not linear in the number of bits. For example,

an N bit flash ADCs will require $2^N - 1$ comparators. Also, circuit component matching costs increase significantly beyond a certain threshold. We will consider the case in which we are given M quantizers with predetermined resolutions and timing offsets τ_m, and desire to find the optimal assignment of quantizers to timing offsets. Thus, all comparisons will involve the same M quantizers and will have the same manufacturing cost.

8.1.1 Uniform Sampling and Concurrent Sampling Instants

Intuitively, when the timing offsets τ_m fall at the same instant, the assignment of the quantizers will not matter. The same is true for uniform sampling, because all distances between channels are equal. This can be verified by examining equation (6.8) when $M = 3$ and $L = 2$. Several of the terms are constant, because we are only considering different assignments of the same M quantizers. The problem reduces to equation (6.9), where the distances are given. From here it is easy to see that the numerators are equal when $d_{0,1} = d_{0,2} = 0$ or $d_{0,1} = -d_{0,2} = \frac{1}{3}$. With equal weights, we have:

$$\underset{b_0,b_1,b_2}{\operatorname{argmin}} \quad \sigma_{e_{min}} = \underset{b_0,b_1,b_2}{\operatorname{argmin}} \quad \frac{1}{\sigma_0^2\sigma_1^2} + \frac{1}{\sigma_0^2\sigma_2^2} + \frac{1}{\sigma_1^2\sigma_2^2}. \tag{8.1}$$

Equation (8.1) will evaluate to a constant for any assignments of quantizers. This result reveals that quantizer assignment is unimportant for uniform sampling or when sampling instants fall on the same time instant. This intuitively makes sense, because all of the relative distances between quantizers are preserved during a switch of this type.

8.1.2 Large Time Difference between Sampling Instants

From the previous chapter, we expect channels that are separated by small amounts of time to be assigned fewer bits than channels with large time differences. To analytically verify this for the case when $M = 3$ and $L = 2$, we again refer to equation (6.8), which reduces to equation (6.9). We will consider the case when $d_{0,1} = 0$ and $d_{0,2} = 0.5$. In other words, channels 0 and 1 fall on the same time instant, while channel 2 is separated by the maximum possible time difference. Evaluating the cosine functions, we obtain:

$$\underset{b_0,b_1,b_2}{\operatorname{argmin}} \quad \sigma_{e_{min}} = \underset{b_0,b_1,b_2}{\operatorname{argmin}} \quad \frac{1}{\sigma_0^2\sigma_1^2} - \frac{1}{\sigma_0^2\sigma_2^2} - \frac{1}{\sigma_1^2\sigma_2^2}. \tag{8.2}$$

To minimize equation (8.2), we aim to assign quantizers in such a way that the product $\sigma_0^2\sigma_1^2$ is large while $\sigma_0^2\sigma_2^2$ and $\sigma_1^2\sigma_2^2$ are small. The only variable that can be assigned independently is σ_2^2. Since both terms require minimization of σ_2^2, we assign the highest precision quantizer to the second channel. This matches our intuition well, because channel 2 is the furthest apart from the other channels.

8.2 Simulated Results

To verify this intuition, we simulate the system with uniform quantizers for the case when $M = 3$, $L = 2$, $\tau_0 = 0$, $\tau_1 = T_N/8$, and $\tau_2 = -(3/4)T_N$. Keeping the total number of bits equal to 30, we consider all permutations of the bit allocations where only one bit is shifted from one channel to another. Table 8.1 shows the performance gain for different bit allocations as compared to the case where each of the channels is allocated 10 bits. The left three columns represent the bit allocation b_m in each channel, while the right column represents the performance gain compared to the case where each channel is allocated 10 bits.

In this case, the sampling instant of channel 2 is relatively far from the other two channels. Also, the difference between channel 0 and channel 2 is smaller than the offset between channel 1 and channel 2. The results in Table 8.1 fit our intuition well, showing that allocating the most bits to channel 2, less to channel 1, and least to channel 0 is optimal.

b_0	b_1	b_2	$(\sigma^2_{e_{min}})_{(b_0=b_1=b_2=10)}/\sigma^2_{e_{min}}$
10	10	10	1.00
9	10	11	1.46
10	9	11	1.36
9	11	10	1.26
11	9	10	1.14
10	11	9	0.41
11	10	9	0.39

Table 8.1: Performance Gain for Different Bit Allocations

8.3 Variable Timing Offsets and Variable Bit Allocation

One way to visualize the underlying relationship is to set the bit allocation in two channels and vary the third channel's bit allocation. For each permutation of bit allocations, we consider all possible configurations of timing offsets (from a discrete vector of possible sampling points) and record only the optimal sampling configuration. This reveals an underlying structure in the optimal sampling configurations. As the third channel gains more bits, the other two channels' sampling instants draw nearer to each other. The opposite is true if the third channel contains few bits.

In Chapters 4 and 6, we considered the case when $M = 3$, $L = 2$, and $b_0 = b_1 = 10$ with the additive noise model and uniform quantizers. In Figures 6-5 and 7-3, we evaluated all possible configurations of timing offsets (on a discrete grid) for b_2 between 8 and 12 bits. We concluded that the key parameter that determines performance in a given constellation of timing offsets τ_m is the *distance* between sampling instants, not the constellation itself. For example, several different configurations may have equivalent performance because they are shifted versions of the same sampling pattern.

Figure 8-1 depicts the predicted and simulated distances that result in maximum SQNR for each bit allocation. It is important to note that a distance of $d = 0.5$ is equivalent to an offset of $d = -0.5$ in this case. As the number of bits in channel 2 increases, the distance between channels 0 and 1 decreases monotonically. This supports the intuition developed earlier.

Some important insight can be gathered from Figure 8-1. First, note that the optimal distance for the uniform quantizer $d_{0,1}$ differs slightly from the additive noise model when $b_2 = k + 1$ or $b_2 = k+2$. This is because the additive noise model distance $d_{0,1}$ is zero at these points. This would introduce error cross-correlation and the associated performance degradation, so the correction term discussed earlier is necessary. (Also note that the plot has a resolution of $\frac{1}{30}$, so a much smaller correction term would also be adequate.)

One of the biggest discrepancies between the additive noise model and uniform quantizer occurs for $d_{0,2}$ when $b_2 = k-1$ or $b_2 = k-2$. In this case, the performance benefit of placing two quantizers with different granularities in the same location actually outweighs the original optimal solution. This is clearly evident in Figure 7-3.

When $b_2 = k$, all three channels have identical quantizers. As predicted, the simulated result

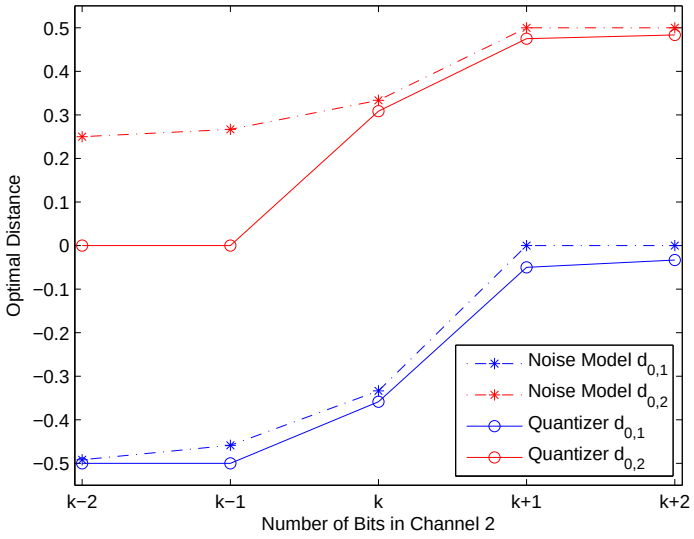


Figure 8-1: Optimal distances for noise model and uniform quantizers as a function of the number of bits in channel 2 with k bits in channels 0 and 1.

is an approximately uniform sampling pattern. It is encouraging to note that as the number of bits associated with b_2 increases, the optimal distance $d_{0,1}$ monotonically decreases. Similarly, $d_{0,2}$ monotonically increases. This fits our intuition well, because the channel 2 dominates the measurement capability for the system, and thus requires the largest time difference.

Chapter 9

Conclusions & Future Work

In this thesis, we have developed a framework for optimal design of non-uniform sampling constellations in time-interleaved ADCs. We presented analytical and numerical optimization techniques that rely on the additive noise model for determining non-uniform sampling constellations that maximize SQNR. Systems that critically sample the bandlimited input signal ($M = L$) always have optimal timing offsets that are uniform. As the oversampling factor increases, the set of optimal sampling configurations becomes much larger. One common characteristic of optimal timing constellations in systems with diverse bit allocations is "clustering" of low-precision channels. Often, the optimal timing offsets corresponding to two or more low-precision channels are close to the same value, and the "cluster" behaves in the same way as a single high-precision channel. Genetic optimization algorithms were shown to be an effective solution for determining optimal timing constellations in systems with large numbers of channels quickly and accurately.

In order to investigate discrepancies between the additive noise model and uniform quantizers, we implemented a software simulation of the time-interleaved ADC. The simulation revealed a key discrepancy in the environment of time-interleaved ADCs: cross-channel quantization error correlation. We presented statistical analysis to characterize error correlation between quantizers with different granularities. We developed a novel ADC architecture based on Weighted Least Squares (WLS) to exploit this correlation, with particular application for time-interleaved ADCs. We also proposed a "correlated noise model" that incorporates error correlation between channels. The proposed model was shown to perform significantly better than the traditional additive noise model for channels in close proximity.

A key result of the thesis is a framework for efficient design of optimal channel configurations. The framework incorporates our earlier statistical analysis to incorporate cross-channel quantization error correlation with the optimal solution predicted by the additive noise model. The corrected results are experimentally shown through simulation to perform well. For homogeneous bit allocations, the framework dictates timing offset corrections to avoid performance degradation from optimal scenario predicted by the additive noise model. For diverse bit allocations, the framework dictates timing corrections and a "unification" of low-precision quantizers in close proximity. This results in significant improvements in performance above the previously-known optimal additive noise model solution.

Several questions will require future investigation. In the thesis, we performed an experimental analysis of quantization error correlation. The results are dependent on the autocorrelation properties of the input signal and the overall bit allocation of the system. A theoretical analysis of acceptable distances between channels to avoid performance degradation based on the autocorrelation of the input and system properties would be useful.

Another area of future interest is non-ideal quantizers. In practice, the comparators in quantizers are not perfectly calibrated, preventing the piecewise linear segments in the quantizer error residue plot from lining up exactly. It would be of practical interest to investigate whether the negative correlation in quantization error derived in Chapter 5 holds for non-ideal quantizers.

Jitter, or uncertainty in the sampling instant, is currently a significant limitation for high-rate time-interleaved ADCs ([23],[15]). A large amount of research has been done in recent years to develop calibration algorithms to mitigate error due to jitter. The simulation platform from this thesis could potentially be adapted to compare timing skew correction algorithms. This would be an interesting experiment that would also allow us to combine the jitter correction algorithms with the discrete-time quantization minimization reconstruction filtering in a single system.

The research in this thesis has a variety of applications in distributed sampling, beyond the context of time-interleaved A/D conversion. The problem formulation can be generalized so that time distances between channels correspond to geographic distances between ADCs. In this context, the problem involves determining optimal geographic locations of a set of different precision measurement sensors to maximize SQNR for a spatially bandlimited input signal.

Overall, the framework presented in this thesis provides a technique for designing optimal

channel configurations that exploit quantization error correlation, which is not usually found in designs based on the traditional additive noise model. We hope that this framework will provide a foundation for future work in this field.

Appendix A

Alternative Architecture: Intersection of Quantizer Regions

We have established that signals quantized with different numbers of bits contain different information about the underlying signal, and statistically derived a method to combine the observations. In this appendix, we outline an alternative method for combining observations. Fundamentally, a quantizer measurement specifies a region that contains the analog value of interest. The size of this region is dependent on the constellation of comparators used for measurement. Recall that each quantizer is operating on the same analog sample, so the analog value must lie within each given region.

A set of d measurements with different granularities will specify d overlapping regions. Intuitively, we expect that we can create a better composite estimate than any single quantizer by taking the intersection between the d regions. This process essentially combines the comparators from all available measurements to create a non-uniform quantizer.

An example of this process is illustrated in Figure A-1. For an analog input sample, the corresponding regions for each quantizer are shaded. The intersection between these regions gives a better estimate of the analog value than any single quantizer. The mean of this composite region is taken as the new estimate.

This alternative architecture is formulated in Figure A-2. Further analysis on this subject will be left for future work.

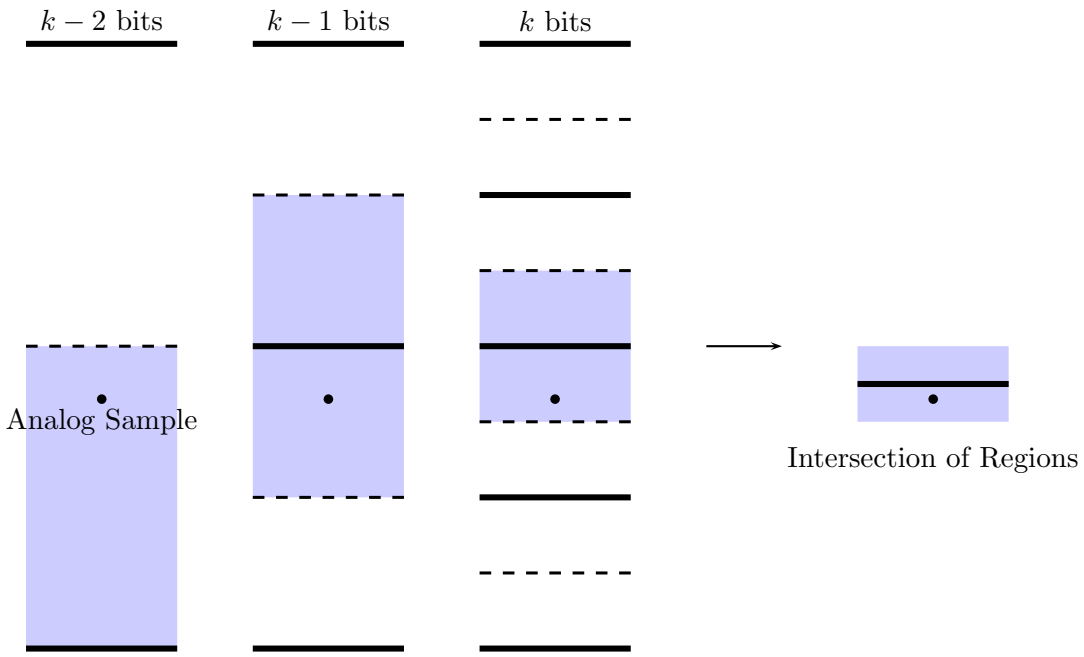


Figure A-1: For an analog input sample, the corresponding regions for each quantizer are shaded. The intersection between these regions gives a better estimate of the analog value than any single quantizer. The mean of this composite region is taken as the new estimate.

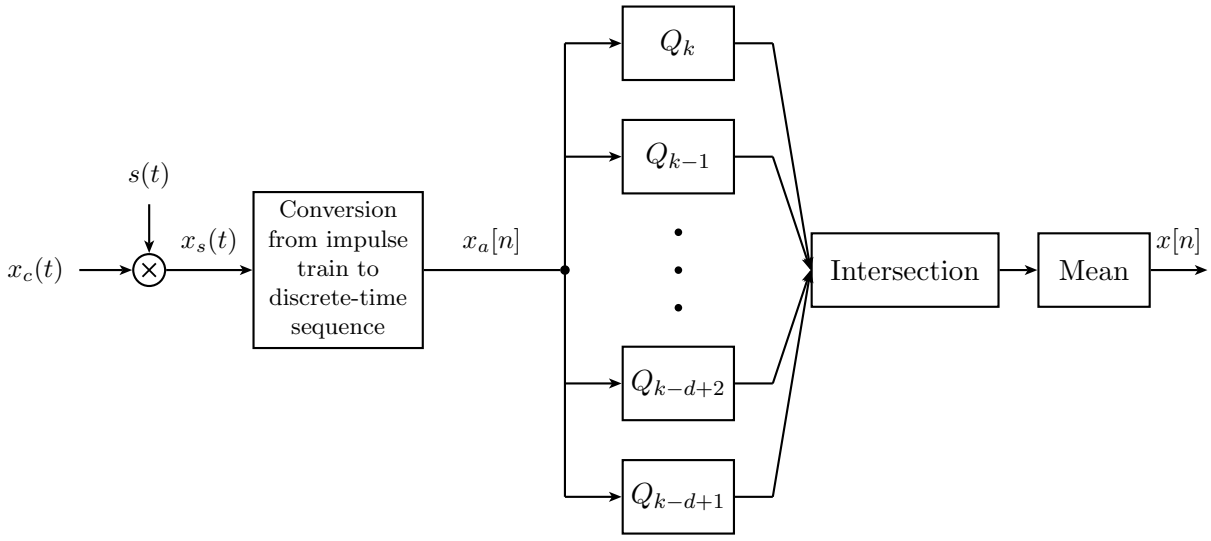


Figure A-2: Alternative ADC architecture for combining quantized signals that takes the intersection between quantizer regions.

Bibliography

[1] W.R. Bennett. Spectra of quantized signals. *Bell Syst. Tech. J*, 27(3):446–472, 1948.

[2] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to probability*. Athena Scientific, 2008.

[3] WC Black and DA Hodges. Time interleaved converter arrays. *Solid-State Circuits, IEEE Journal of*, 15(6):1022–1029, 2002.

[4] J.C. Candy, G.C. Temes, Institute of Electrical, Electronics Engineers, IEEE Circuits, and Systems Society. *Oversampling delta-sigma data converters: theory, design, and simulation*. IEEE press New York, 1992.

[5] L. Davis. *Genetic algorithms and simulated annealing*. Research notes in artificial intelligence. Pitman, 1987.

[6] V. Divi and G.W. Wornell. Blind calibration of timing skew in time-interleaved analog-to-digital converters. *Selected Topics in Signal Processing, IEEE Journal of*, 3(3):509 –522, 2009.

[7] Y.C. Eldar and A.V. Oppenheim. Filterbank reconstruction of bandlimited signals from nonuniform and generalized samples. *Signal Processing, IEEE Transactions on*, 48(10):2864–2875, 2002.

[8] D.E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Artificial Intelligence. Addison-Wesley Pub. Co., 1989.

[9] R.M. Gray. Quantization noise spectra. *Information Theory, IEEE Transactions on*, 36(6):1220–1244, 2002.

[10] M. Haugh. The Monte Carlo framework, examples from finance and generating correlated random variables. *Course Notes*, 2004.

[11] R.L. Haupt and S.E. Haupt. *Practical genetic algorithms*. Wiley-Interscience publication. John Wiley, 2004.

[12] A.J. Jerri. The Shannon sampling theorem - Its various extensions and applications: A tutorial review. *Proceedings of the IEEE*, 65(11):1565–1596, 2005.

[13] K. Kaviani, O. Oralkan, P. Khuri-Yakub, and B.A. Wooley. A multichannel pipeline analog-to-digital converter for an integrated 3-D ultrasound imaging system. *Solid-State Circuits, IEEE Journal of*, 38(7):1266–1270, 2003.

[14] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation theory*. Prentice Hall signal processing series. Prentice-Hall PTR, 1998.

[15] Bin Le, T.W. Rondeau, J.H. Reed, and C.W. Bostian. Analog-to-digital converters. *Signal Processing Magazine, IEEE*, 22(6):69 – 77, 2005.

[16] S. Maymon. *Sampling and Quantization for Optimal Reconstruction*. PhD thesis, Massachusetts Institute of Technology, 2011.

[17] S. Maymon and A.V. Oppenheim. Quantization and compensation in sampled interleaved multi-channel systems. Submitted for publication to IEEE Transactions on Signal Processing.

[18] S. Maymon and A.V. Oppenheim. Quantization and compensation in sampled interleaved multi-channel systems. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 3622–3625. IEEE, 2010.

[19] A.V. Oppenheim and R.W. Schafer. Discrete-time signal processing. *Prentice Hall Signal Processing*, 2009.

[20] Julius O. Smith. *Spectral Audio Signal Processing, October 2008 Draft*.

[21] A. Sripad and D. Snyder. A necessary and sufficient condition for quantization errors to be uniform and white. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 25(5):442–448, 2003.

[22] C. Vogel. Comprehensive, error analysis of combined channel mismatch effects in time-interleaved ADCs. In *Instrumentation and Measurement Technology Conference, 2003. IMTC '03. Proceedings of the 20th IEEE*, volume 1, pages 733 – 738, May 2003.

[23] R.H. Walden. Analog-to-digital converter survey and analysis. *Selected Areas in Communications, IEEE Journal on*, 17(4):539 –550, April 1999.

[24] B. Widrow and I. Kollár. *Quantization noise: roundoff error in digital computation, signal processing, control, and communications*. Cambridge University Press, 2008.

[25] K. Yao and J. Thomas. On some stability and interpolatory properties of nonuniform sampling expansions. *Circuit Theory, IEEE Transactions on*, 14(4):404–408, 2002.

[26] J. Yen. On nonuniform sampling of bandwidth-limited signals. *Circuit Theory, IRE Transactions on*, 3(4):251–257, 1956.