STATISTICAL INFERENCE WITH UNRELIABLE BINARY OBSERVATIONS

BY

RYAN MICHAEL COREY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Adviser:

Professor Andrew Carl Singer

ABSTRACT

We describe a novel statistical inference approach to data conversion for mixed-signal interfaces. We propose a data conversion architecture in which a signal is observed by a set of sensors with uncertain parameters, such as highly scaled comparator circuits in an analog-to-digital converter. These sensor outputs are not used to form a quantized representation of the signal but are used directly to make decisions in statistical inference problems such as parameter estimation, classification, and signal detection. We derive a mathematical model of this system and apply information-theoretic tools to describe the achievable performance of such a converter in information processing systems. In particular, we find asymptotic expressions for Fisher information and Kullback-Leibler divergence in terms of the design parameters and the sensor statistics. Simulations of parameter estimation, classification, and symbol detection systems show that these architectures can achieve strong performance even when the devices have significant process variations. We also discuss practical system design and implementation issues including sensor calibration and propose a lower-complexity suboptimal estimation architecture. The analytical and simulation results suggest that it is both possible and practical to build information processing systems using unreliable mixed-signal components.

ACKNOWLEDGMENTS

This work would not have been possible without the support of a number of people at the University of Illinois and elsewhere. It has been a great privilege to work with the faculty at the University of Illinois; I particularly wish to thank Professors Scott Carney, Naresh Shanbhag, Venugopal Veeravalli, and my adviser, Andrew Singer, for their outstanding teaching and mentorship. I am grateful for the advice and friendship of my fellow students, particularly Andrew Bean, Peter Kairouz, and Jonathan Ligo. The administrative staff in the Coordinated Science Laboratory and SONIC Center, especially Peggy Wells and Brooke Newell, have been invaluable. This project was initially proposed by Professor Naveen Verma of Princeton University. He and his student Sen Tao have made numerous contributions as collaborators in this work. Finally, I wish to thank my family for their tireless support.

This work was supported in part by Systems on Nanoscale Information fabriCs (SONIC), one of the six STARnet centers sponsored by MARCO and DARPA. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant Number DGE-1144245.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This thesis addresses the problem of statistical decision making using binary comparison observations of a signal from sensors that have random variations in their parameters. These sensors may have significant advantages in size, power, speed, or cost, but cannot be used in deterministic systems. We present an architecture that leverages redundancy and statistical inference techniques to build reliable information systems from unreliable components. We model this architecture mathematically and give asymptotic bounds and approximations for performance in information processing applications. The goal of this thesis is to provide circuit and system designers with new models, metrics, and design principles for data conversion with unreliable sensors.

1.1 Motivation

As semiconductor technology advances, systems are increasingly affected by device-level uncertainty. Emerging technologies, such as highly scaled CMOS circuits, offer improved performance and efficiency but exhibit random variation. These technologies can offer system designers a tradeoff between efficiency and reliability. Information processing systems that deal with inherently uncertain signals, such as communication receivers and machine learning classifiers, for example, can be readily adapted to account for device-level uncertainty. To best integrate unreliable devices into reliable systems, the systems cannot be thought of as collections of individual components connected together; instead, each component must be designed with system-level performance in mind.

This work addresses mixed-signal interfaces, such as analog-to-digital converters and sensor arrays, built with highly variable devices. We will focus in particular on quantization circuits, which have been extensively studied and

which could benefit significantly from these emerging technologies. High-speed, high-resolution analog-to-digital converters (ADCs) are often limited in speed and efficiency by large comparator circuits. Smaller circuits generally operate faster and consume less energy, but are more sensitive to process variations. These variations are unacceptable in a deterministic quantizer; in information systems, however, a mixed-signal interface need not behave deterministically. By replacing the deterministic ADC with a stochastic data converter that gathers uncertain observations about a signal, we can take advantage of new technologies while still achieving reliable system-level performance.

1.2 Background

1.2.1 Quantization architecture

Quantization is the problem of mapping a signal from a space of higher cardinality to one of lower cardinality, for example from a continuous space to a discrete space. Given an input signal x, the quantizer selects a codeword from a finite set. Each codeword corresponds to a reconstruction point $\hat{x}$ in the original signal space. Generally, the quantizer is designed to minimize an error metric between the input value and the reconstruction point, such as the mean square error (MSE). A typical scalar quantizer works by dividing the input space into intervals. All signals within a region are mapped to a single codeword and the corresponding reconstruction point is the centroid on that interval. The boundaries between regions are the midpoints between reconstruction points. Such a design minimizes MSE.

For voltage signals, this type of quantizer can be implemented using a flash ADC architecture, shown in Figure 1.1. The flash ADC consists of a number of comparators with different reference levels corresponding to the boundaries between quantization intervals. When the same input signal is applied to each comparator, their outputs form a thermometer code: a string of 1's followed by a string of 0's. A priority encoder converts the thermometer code to a digital codeword. If the reference levels are uniformly spaced distance Δv

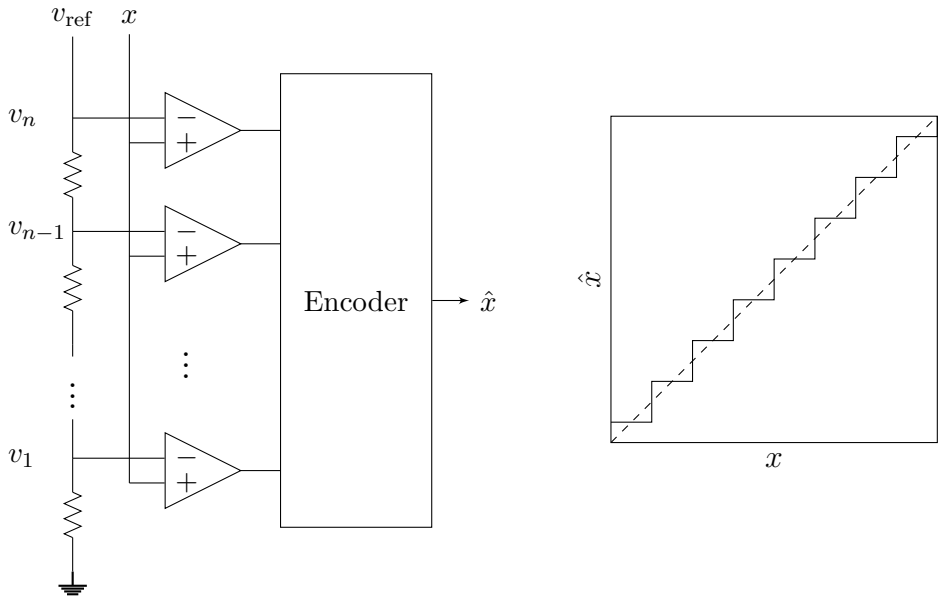

Figure 1.1: A deterministic flash ADC architecture. The signal is compared to many reference levels to generate a codeword corresponding to a quantized value.

apart, then the MSE of the ideal uniform quantizer is

$$\text{MSE}_{\text{ideal}} = \frac{(\Delta v)^2}{12}. \tag{1.1}$$

1.2.2 Process variations

For a flash ADC to function correctly, each comparator must compare the signal to the correct reference level. However, no comparator circuit is perfectly accurate: the reference level is subject to random offsets due to process variations. If these variations are small compared to the spacing between levels, as shown in Figure 1.2(a), then they will have a relatively small impact on quantization performance. If the variations are large so that the offsets can be larger than the spacing between levels, as in Figure 1.2(b), then the true levels may be non-monotonic, causing errors in the encoder.

The random offset in a comparator reference level can be due to a mismatch between the threshold voltages of nominally identical transistors [1]. These threshold voltage variations can be caused by dopant fluctuations in the active area of the transistor for deeply scaled CMOS circuits. These have been thoroughly studied in the circuits literature and are generally characterized

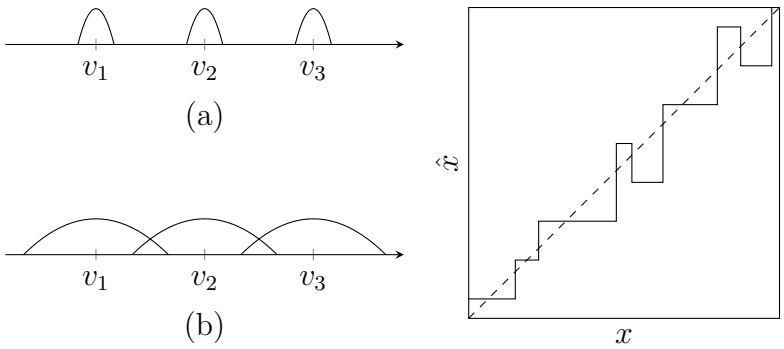

Figure 1.2: When process variations are small, the distributions of comparator reference levels do not overlap (a); when variations are large, the distributions can overlap (b) and cause errors in the quantizer output.

as having a normal distribution with variance inversely proportional to the transistor gate area [2]. Thus, more advanced semiconductor processes with smaller gate areas will likely produce comparators with greater variation. Deterministic ADCs, which cannot tolerate large offsets, must therefore use large transistors for comparator circuits and cannot take advantage of newer, smaller device technologies.

1.3 Prior Work

1.3.1 ADC calibration

Several authors have explored the possibilities of designing flash ADCs using comparator circuits with large offsets. Early designs, such as the offset cancellation stage in [3], used analog calibration techniques to directly correct the comparator offsets. Other designs used digital techniques, such as reassigning or disabling comparators with large offsets [4]. A hybrid approach with analog trimming and digital calibration logic was proposed in [5]. In [6], a conventional thermometer-to-binary converter was modified with fault-tolerant digital logic to handle non-monotonicity in the thresholds. The design in [7] used a fully reconfigurable comparator array with a large memory to assign comparator output codes to digital codewords. All of these proposals attempt to compensate for the offsets so that the architecture functions like a conventional flash ADC with an accurate set of reference

levels. They add significant complexity to the circuit and require an extra calibration step after production of the circuit.

1.3.2 Stochastic ADC

More recent work has directly leveraged the variations in comparator thresholds rather than merely compensating for them. In [8], the reference ladder was eliminated entirely; instead, the levels were allowed to spread randomly and a subset was selected that matched an ideal flash architecture as closely as possible. This approach eliminates the worst variations but requires an extra calibration step. In [9], the comparators were assigned to only two nominal levels with intentionally high variation. The true levels were not measured; instead, the outputs were summed and used to compute an estimated signal based on the offset distribution.

These designs, which directly leverage the comparator offset distributions, anticipate a statistical approach to data conversion. To date, this problem has been addressed primarily in the circuits and devices literature and has been framed in terms of conventional ADC architectures and metrics, with minimal statistical analysis. However, these architectures can be thought of as inference systems that use observations subject to uncertainty to gather information. In this work, we approach the problem from the perspective of statistical detection and estimation, using information-based metrics to predict system-level performance. We aim to provide a more mathematically rigorous framework for data conversion with unreliable circuits.

1.3.3 Distributed sensing

The mathematical model for estimation from comparators with large variations has been studied before, but in a different context. A comparator that compares a signal to a randomly varying threshold level is mathematically equivalent to a sensor that takes a noisy measurement and then quantizes it to one bit using a fixed threshold. Thus, the problem of stochastic data conversion is equivalent to decentralized fusion in bandlimited sensor networks using one-bit quantized observations [10, 11]. The results in [11] parallel several results presented here: the authors find a Cramér-Rao lower bound for

parameter estimation over a bandwidth-constrained sensor network that has the same form as in our problem. They also explore the relationship between estimation performance and spacing between threshold levels, arguing that spacing on the order of the standard deviation of the noise is sufficient; we reach the same conclusion here.

1.4 A Statistical Inference Approach

In this work, we approach the problem of data conversion with unreliable comparators using tools from detection and estimation theory. Our system model consists of an arbitrary number of nominal levels and sensors per level, generalizing both the conventional flash ADC reference ladder (many levels and one comparator per level) and the design of [9] (few levels with many comparators). One critical difference between our analysis and that in the circuits literature is that we do not require the converter to output a digital codeword representing the value of the input signal; instead, we are concerned with the information contained in the vector of comparator outputs or in sufficient statistics of those outputs. In many applications for which mixed-signal interfaces are used, including sensing, classification, and signal detection, the decision-making algorithm can work with these statistics directly. In this thesis, therefore, we will not assess converter performance in terms of conventional metrics such as integral or differential nonlinearity or effective number of bits. Instead, we will use information-based metrics such as Fisher information and Kullback-Leibler divergence and application-level metrics including mean square error and symbol error rate.

In Chapter 2, we present the stochastic data converter architecture and describe the statistical relationship between the input signal and sensor outputs. We use this observational model to derive expressions for the Kullback-Leibler divergence and Fisher information, which are useful for predicting system-level performance. In Chapter 3, we demonstrate the performance of the converter architecture for several decision-making applications using Monte Carlo simulations. We also present a highly parallel two-stage estimation architecture with low computational complexity and consider the performance benefits of calibration. In Chapter 4, we discuss the implications of the performance results for converter design.

CHAPTER 2

SYSTEM MODEL AND ANALYSIS

In information systems, the analog signals collected by converters are often used to make decisions. The role of the converter is to collect information from a signal that is relevant to the particular application for which it is used. Conventional ADCs output quantized measurements of an analog signal and are designed to minimize mean square quantization error. Metrics such as linearity, dynamic range, and effective number of bits do not necessarily predict system-level performance. It is more useful to consider information-based metrics such as Kullback-Leibler divergence and Fisher information. In this chapter, we develop an observational model for our proposed stochastic converter architecture and use that model to characterize performance using these metrics.

2.1 System Architecture

We consider a data converter with a generalized flash ADC architecture, shown in Figure 2.1. Like a conventional flash ADC, it observes an unknown real-valued variable x using a number of binary sensors. There are n nominal reference levels $v_1, \ldots, v_n$. To compensate for uncertainty in the thresholds, the architecture incorporates r duplicate sensors at each nominal level. Each of the nr sensors has a random true reference level $V_{i,j}$ that is offset from its nominal level v_i for $i = 1, \ldots, n$ and $j = 1, \ldots, r$. If $r = 1$, then the architecture is a conventional flash ADC with one comparator at each nominal level. If $n = 1$, then the comparators are all assigned the same nominal level, as in [9]. The architecture can be regarded as making r trials of an n-dimensional observation vector. Our results will show that the choice of n and r involves a tradeoff between performance and complexity: a system with many nominal levels provides information over a broad range

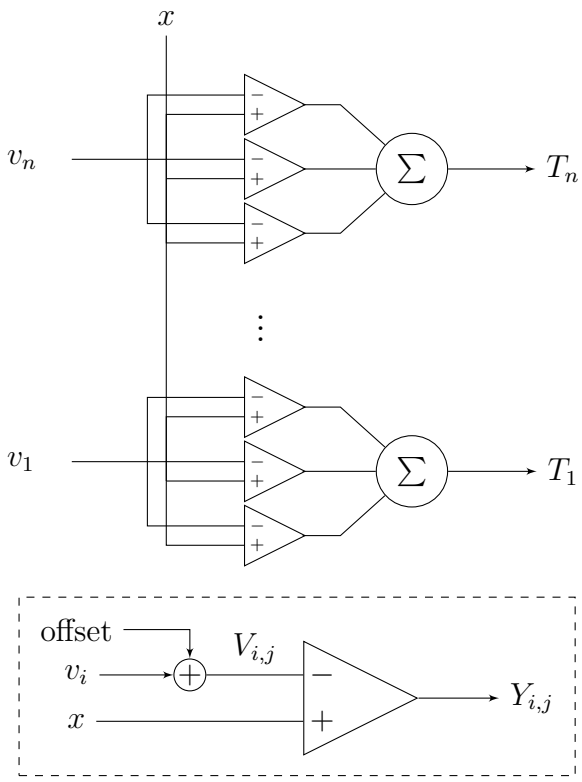

Figure 2.1: The stochastic data converter uses r comparators at each of n nominal reference levels. Each comparator has a random offset in its reference level.

of input values, while a more parallel design improves performance without significantly increasing complexity.

We assume that the offsets for all comparators are independent and identically distributed according to a known distribution. Furthermore, the input x is the same for all comparators. These assumptions are reasonable as long as the spatial correlation between offsets and the time-varying noise on each sensor are negligible. Denote the cumulative distribution function (CDF) of the random offsets by F_V. Then each level $V_{i,j}$ has CDF $\Pr\{V_{i,j} \leq x\} = F_V(x - v_i)$. For brevity, let $F_i(x) = F_V(x - v_i)$ and let $\bar{F}_i(x) = 1 - F_V(x - v_i)$. Where F_i is differentiable at x, denote the corresponding probability density function (PDF) by $f_i(x) = \frac{\partial}{\partial x} F_i(x)$.

Each comparator produces a binary output

$$Y_{i,j} = \begin{cases} 1, & \text{if } x \geq V_{i,j} \\ 0, & \text{if } x < V_{i,j}\,. \end{cases} \tag{2.1}$$

Let $\mathbf{Y}$ denote the set of these outputs for $i = 1, \ldots, n$ and $j = 1, \ldots, r$. These are the observations that will be used for inference about x. In most of this analysis we assume that only one set of observations is available. More complex systems could improve performance by estimating the offsets using past observations; we briefly explore the performance improvements achievable by calibration in Section 3.4.

2.2 Observational Model

Because the comparator outputs are nondeterministic, we must use statistical inference tools to recover information about x from the observations. These tools rely on a statistical model that relates the observed random variable $\mathbf{Y}$ to the unobserved variable x.

2.2.1 Likelihood

Given x, each $Y_{i,j}$ has a Bernoulli probability mass function (PMF),

$$\begin{aligned} p_{Y_{i,j}|X}(1 \mid x) &= \Pr\{x \geq V_{i,j}\} && (2.2)\\ &= F_i(x) && (2.3)\\ p_{Y_{i,j}|X}(0 \mid x) &= \Pr\{x < V_{i,j}\} && (2.4)\\ &= \bar{F}_i(x)\ . && (2.5) \end{aligned}$$

Because the offsets are independent and identically distributed, the rn observations can be reduced to n sufficient statistics that preserve the information about x. For each $i = 1, \ldots, n$, let $T_i = \sum_{j=1}^{r} Y_{i,j}$. As a sum of conditionally independent Bernoulli random variables, each T_i has a binomial conditional distribution:

$$p_{T_i|X}(t_i \mid x) = \binom{r}{t_i} F_i(x)^{t_i} \bar{F}_i(x)^{r-t_i}\ . \tag{2.6}$$

The mean of T_i is given by

$$\mathbb{E}[T_i \mid X = x] = rF_i(x)\,, \tag{2.7}$$

where $\mathbb{E}[\cdot]$ denotes the expectation. If $\bar{F}_i(x) \neq 0$, which is true as long as $f_i(x) > 0$, then the PMF can be written

$$p_{T_i|X}(t_i \mid x) = \binom{r}{t_i} \bar{F}_i(x)^r \left(\frac{F_i(x)}{\bar{F}_i(x)}\right)^{t_i}\ . \tag{2.8}$$

If $\bar{F}_i(x) \neq 0$, then the full vector $\mathbf{T} = \{T_i\}_{i=1}^n$ is a sufficient statistic for $\mathbf{Y}$ given x and has PMF

$$\begin{aligned} p_{\mathbf{T}|X}(\mathbf{t} \mid x) &= \prod_{i=1}^{n} p_{T_i|X}(t_i \mid x) && (2.9)\\ &= \prod_{i=1}^{n} \binom{r}{t_i} \bar{F}_i(x)^r \left(\frac{F_i(x)}{\bar{F}_i(x)}\right)^{t_i}\ . && (2.10) \end{aligned}$$

On the set of inputs for which $F_i(x)$ is differentiable and $f_i(x) > 0$, (2.10) forms an exponential family of distributions [12]. An exponential family is a

set of conditional probabilities parametrized by x that have the form

$$p_{\mathbf{T}|X}\left(\mathbf{t} \mid x\right) = h\left(\mathbf{t}\right) \exp\left(\sum_{i=1}^{n} \eta_i\left(x\right) t_i - A\left(x\right)\right) . \tag{2.11}$$

The exponential form is convenient for statistical inference problems because the log-likelihoods, sufficient statistics, and information metrics of the observational model can be easily expressed in terms of η_i, t_i, and A.

For the converter likelihood function (2.10) with $\bar{F}_i(x) \neq 0$, the terms in (2.11) are $h\left(\mathbf{t}\right) = \prod_{i=1}^{n} \binom{r}{t_i}$, $A\left(x\right) = -r \sum_{i=1}^{n} \ln \bar{F}_i\left(x\right)$, and $\eta_i\left(x\right) = \ln\left(F_i\left(x\right) / \bar{F}_i\left(x\right)\right)$. The vector $\boldsymbol{\eta}\left(x\right) = \left\{\eta_i\left(x\right)\right\}_{i=1}^{n}$ is called the natural parameter of the family. The first derivative of $\eta_i\left(x\right)$ is

$$\eta_i'\left(x\right) = \frac{f_i\left(x\right)}{F_i\left(x\right)\bar{F}_i\left(x\right)} . \tag{2.12}$$

The natural parameter and its derivative determine a number of useful properties of exponential families for detection and estimation problems and will be used throughout this thesis. Note that because the natural parameter $\boldsymbol{\eta}$ has higher dimension than the true parameter x, $\left\{p_{\mathbf{T}|X}\left(\cdot \mid x\right)\right\}_x$ is a curved exponential family [13]. Therefore, while $\mathbf{T}$ is a sufficient statistic, it may not be a complete sufficient statistic: that is, there may exist a lower-dimensional sufficient statistic that preserves its information.

2.2.2 Special case: The logistic distribution

Indeed, there is an offset distribution for which the likelihood belongs to a one-dimensional exponential family. The logistic distribution with mean μ and scale parameter β has the CDF

$$L\left(v\right) = \frac{\exp\left(\frac{v-\mu}{\beta}\right)}{1 + \exp\left(\frac{v-\mu}{\beta}\right)} . \tag{2.13}$$

The variance of a logistic random variable is $\sigma^2 = \beta^2\pi^2/3$. Figure 2.2 compares the logistic distribution with the normal distribution, which is often used to model process variations and which will be used in most of the simulations presented here. If $F_i\left(x\right)$ is logistic with mean v_i, then the natural

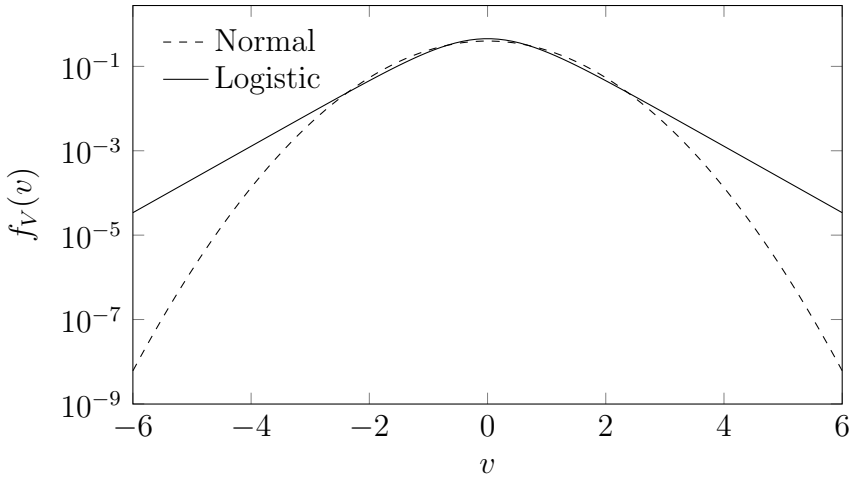

Figure 2.2: Normal and logistic distributions with unit variance. The normal distribution is often used to model process variations. The logistic distribution induces a convenient form on the observation likelihood.

parameter components are $\eta_i(x) = (x - v_i)/\beta$, which are linear functions of the true parameter x. The likelihood (2.11) can be written

$$p_{\mathbf{T}|X}(\mathbf{t} \mid x) = h(\mathbf{t}) \exp\left(-\sum_{i=1}^{n} \frac{v_i t_i}{\beta}\right) \exp\left(x \sum_{i=1}^{n} \frac{t_i}{\beta} - A(x)\right). \tag{2.14}$$

Then the scaled sum of all the comparator outputs, $S(\mathbf{T}) = \sum_{i=1}^{n} \eta_i'(x) T_i = \sum_{i=1}^{n} T_i/\beta$, is a complete sufficient statistic and its conditional distribution forms a one-dimensional exponential family with natural parameter x. We will use this special case to find simple closed-form asymptotic expressions for the information-based metrics in this chapter.

2.3 Kullback-Leibler Divergence

In a statistical inference problem, an algorithm must use observations to choose from a set of hypotheses about an unobserved variable. To ensure an accurate decision, the observations must contain information that distinguishes between different hypotheses. A useful performance metric is the Kullback-Leibler (KL) divergence, which quantifies the discrepancy between two probability distributions. If p_1 and p_2 are two PMFs or PDFs with the

same support, then the KL divergence is defined as

$$D\left(p_1 \| p_0\right) = \mathbb{E}_{p_1}\left[\ln \frac{p_1\left(\mathbf{T}\right)}{p_0\left(\mathbf{T}\right)}\right], \tag{2.15}$$

where $\mathbb{E}_p\left[\cdot\right]$ denotes the expectation of the argument under distribution p. Note that $D\left(p_1 \| p_0\right) \neq D\left(p_0 \| p_1\right)$ in general. For the distributions considered here, however, the divergence is symmetric, so we will restrict our attention to $D\left(p_1 \parallel p_0\right)$.

The performance of the data converter in distinguishing between values of x depends on the KL divergence of the likelihood function. First, consider two fixed points x_0 and x_1 which both lie in the support of all the level distributions for which the observations have positive likelihood. The likelihood functions are

$$p_0\left(\mathbf{t}\right) = p_{\mathbf{T}|X}\left(\mathbf{t} \mid x_0\right) \tag{2.16}$$

$$p_1\left(\mathbf{t}\right) = p_{\mathbf{T}|X}\left(\mathbf{t} \mid x_1\right). \tag{2.17}$$

The KL divergence can be derived using the exponential family form of the likelihood (2.11):

$$D\left(p_1 \| p_0\right) = \mathbb{E}_{p_1}\left[\ln \frac{p_{\mathbf{T}|X}\left(\mathbf{T} \mid x_1\right)}{p_{\mathbf{T}|X}\left(\mathbf{T} \mid x_0\right)}\right] \tag{2.18}$$

$$= \mathbb{E}_{p_1}\left[\sum_{i=1}^{n}\left[\eta_i\left(x_1\right) T_i - A\left(x_1\right)\right] - \sum_{i=1}^{n}\left[\eta_i\left(x_0\right) T_i - A\left(x_0\right)\right]\right] \tag{2.19}$$

$$= \sum_{i=1}^{n}\left[\left(\eta_i\left(x_1\right) - \eta_i\left(x_0\right)\right) \mathbb{E}_{p_1}\left[T_i\right]\right] - \left(A\left(x_1\right) - A\left(x_0\right)\right) \tag{2.20}$$

$$= r \sum_{i=1}^{n}\left(\eta_i\left(x_1\right) - \eta_i\left(x_0\right)\right) F_i\left(x_1\right) + \ln \frac{\bar{F}_i\left(x_1\right)}{\bar{F}_i\left(x_0\right)}. \tag{2.21}$$

2.3.1 High resolution limit

To assess the performance scaling of the data converter, consider a high-resolution architecture with many closely-spaced levels. Suppose that the n nominal levels are uniformly spaced at a distance Δv apart from v_1 to v_n and assume for simplicity that $f_i\left(x\right) > 0$ for all $i = 1, \ldots, n$ and all $x \in \left(v_1, v_n\right)$. In (2.21), the divergence is the sum of a function of v_i. If both sides of (2.21)

are multiplied by Δv then, in the limit as Δv approaches zero, the sum can be replaced by the Riemann integral of a function of v:

$$\lim_{\Delta v \to 0} D\left(p_1 \| p_0; \Delta v\right) \Delta v = \lim_{\Delta v \to 0} r \sum_{i=1}^{n} \Bigg[\left(\eta\left(x_1 - v_i\right) - \eta\left(x_0 - v_i\right)\right) F_V\left(x_1 - v_i\right) + \ln \frac{\bar{F}_V\left(x_1 - v_i\right)}{\bar{F}_V\left(x_0 - v_i\right)} \Bigg] \Delta v \tag{2.22}$$

$$= r \int_{v_1}^{v_n} \Bigg[\left(\eta\left(x_1 - v\right) - \eta\left(x_0 - v\right)\right) F_V\left(x_1 - v\right) + \ln \frac{\bar{F}_V\left(x_1 - v\right)}{\bar{F}_V\left(x_0 - v\right)} \Bigg] \mathrm{d}v \,. \tag{2.23}$$

The integrand is nonnegative for all v. If v_1 and v_n are far from x_0 and x_1, then a useful upper bound for the high-resolution KL divergence is

$$\lim_{\Delta v \to 0} D\left(p_1 \| p_0; \Delta v\right) \Delta v \leq r \int_{-\infty}^{\infty} \Bigg[\left(\eta\left(x_1 - v\right) - \eta\left(x_0 - v\right)\right) F_V\left(x_1 - v\right) + \ln \frac{\bar{F}_V\left(x_1 - v\right)}{\bar{F}_V\left(x_0 - v\right)} \Bigg] \mathrm{d}v \,. \tag{2.24}$$

If the offsets have a logistic distribution, then (2.24) can be evaluated exactly. Let $d = x_1 - x_0$ and let $u = (v - x_1)/\beta$. Then, in the limit as $\Delta v \to 0$, the divergence is bounded by

$$\lim_{\Delta v \to 0} D\left(p_1 \| p_0; \Delta v\right) \Delta v$$

$$\leq r \int_{-\infty}^{\infty} \left[\frac{d}{\beta} \frac{\exp\left(\frac{x_1 - v}{\beta}\right)}{1 + \exp\left(\frac{x_1 - v}{\beta}\right)} + \ln \frac{1 + \exp\left(\frac{x_0 - v}{\beta}\right)}{1 + \exp\left(\frac{x_1 - v}{\beta}\right)} \right] \mathrm{d}v \tag{2.25}$$

$$= r\beta \int_{-\infty}^{\infty} \left[\frac{d}{\beta} \frac{\exp\left(-u\right)}{1 + \exp\left(-u\right)} + \ln \frac{1 + \exp\left(-u - d/\beta\right)}{1 + \exp\left(-u\right)} \right] \mathrm{d}u \tag{2.26}$$

$$= \frac{r\beta}{2} \left(\frac{d}{\beta}\right)^2 \tag{2.27}$$

$$= \frac{rd^2}{2\beta} \tag{2.28}$$

$$= \frac{\pi}{2\sqrt{3}} \frac{rd^2}{\sigma} \,. \tag{2.29}$$

For normally distributed offsets, the equivalent integral does not have a closed form and appears to grow slightly faster with d. For $d = \sigma$, the integral can be computed numerically to find $\lim_{\Delta v \to 0} D\left(p_1 \| p_0\right) \Delta v \leq 0.920 r d^2/\sigma$.

This bound is valid only in the limit as $\Delta v \to 0$, but it gives a useful approximation for $D(p_1 \| p_0)$ even with finite resolution. If Δv is sufficiently small and if v_1 and v_n are sufficiently far from x_0 and x_1, then the KL divergence is approximated by

$$D\left(p_1 \| p_0\right) \approx \frac{\pi}{2\sqrt{3}} \frac{r d^2}{\sigma \Delta v} \tag{2.30}$$

for offsets with the logistic distribution and by

$$D\left(p_1 \| p_0\right) \approx 0.920 \frac{r d^2}{\sigma \Delta v} \tag{2.31}$$

for normally distributed offsets.

This approximation is most useful when x_0 and x_1 are far apart compared to the spacing between nominal levels but close relative to the offset deviation. Figure 2.3 shows the exact divergence as a function of Δv for normally distributed offsets. The plot suggest that the high-resolution approximation (2.31) is accurate for $d \leq \sigma$ and $\Delta v \leq 2\sigma$.

2.4 Fisher Information

The KL divergence is useful for predicting performance for classification problems in which the hypothesis space is discrete. For regression problems, in which a continuous parameter is to be estimated, a useful metric is the Fisher information [12]. The Fisher information quantifies the information provided by the observations about the parameter at a particular point in the parameter space. Subject to certain regularity conditions, the Fisher information contributed by an observation T about the unknown parameter x is

$$I\left(x\right) = -\mathbb{E}\left[\frac{\partial^2}{\partial x^2} \ln p_{T|X}\left(T \mid x\right)\right] . \tag{2.32}$$

The Fisher information determines the achievable performance of unbiased estimators of x, as explained in Section 3.2.

For the data converter outputs, the regularity conditions that ensure the

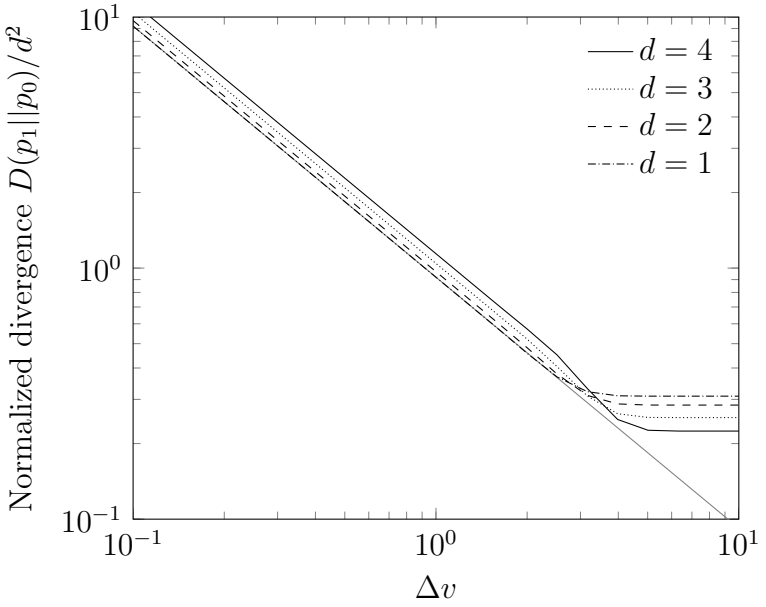

Figure 2.3: KL divergence between two point distributions separated by d for $\sigma = 1$ and $r = 1$. The gray curve is the high-resolution approximation $D\left(p_1 \parallel p_0\right) \approx 0.920 d^2/\Delta v$.

Fisher information is well-defined are satisfied for those T_i for which the corresponding $f_i\left(x\right)$ exists and is strictly positive, that is, where the reference level distribution has support. For those observations, the Fisher information is given by

$$
\begin{aligned}
I_i\left(x\right) &= -\mathbb{E}\left[\frac{\partial^2}{\partial x^2} \ln p_{T_i|X}\left(T_i \mid x\right)\right] && (2.33)\\
&= \mathbb{E}\left[\left(\frac{\partial}{\partial x} \ln p_{T_i|X}\left(T_i \mid x\right)\right)^2\right] && (2.34)\\
&= \mathbb{E}\left[\left(\frac{\partial}{\partial x}\left(\eta_i\left(x\right) T_i - A_i\left(x\right)\right)\right)^2\right] && (2.35)\\
&= \mathbb{E}\left[\left(\eta_i'\left(x\right) T_i - \eta_i'\left(x\right) \mathbb{E}\left[T_i\right]\right)^2\right] && (2.36)\\
&= \left(\eta_i'\left(x\right)\right)^2 \mathrm{Var}_x\left(T_i\right) && (2.37)\\
&= r \frac{f_i\left(x\right)^2}{F_i\left(x\right) \bar{F}_i\left(x\right)}\,. && (2.38)
\end{aligned}
$$

Because the observations are conditionally independent given x, the total Fisher information $I\left(x\right)$ provided by $\mathbf{T}$ about x is the sum of the Fisher

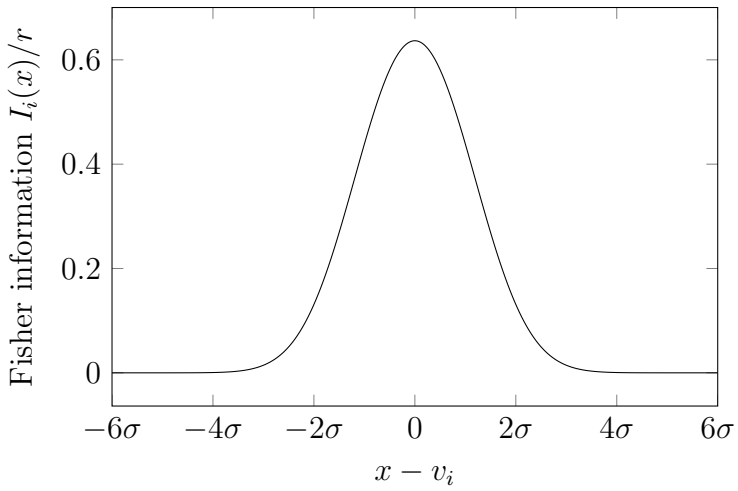

Figure 2.4: The Fisher information provided by each observation about the value of x. The information is maximized when $x = v_i$.

information $I_i(x)$ contributed by each T_i for which the corresponding comparator level distribution has support:

$$I(x) = \sum_{i:f_i(x)>0} I_i(x) \tag{2.39}$$

$$= \sum_{i:f_i(x)>0} r\frac{f_i(x)^2}{F_i(x)\bar{F}_i(x)} \tag{2.40}$$

$$= \sum_{i:f_i(x)>0} r\eta_i'(x) f_i(x) \ . \tag{2.41}$$

Those T_i for which $f_i(x) = 0$ do not contribute Fisher information, but they do restrict the valid space of x values. If no $f_i(x) > 0$ for a particular x, then the Fisher information is not defined but the observations will restrict x to a particular interval, as in a deterministic quantizer. Figure 2.4 shows the Fisher information contributed by a single observation for normally distributed offsets. Nearly all of the information comes from comparators with nominal thresholds within about 3σ of x.

2.4.1 High resolution limit

We can characterize the high-resolution performance of the converter using the approach from Section 2.3.1. Suppose that the n nominal levels are uniformly spaced at a distance Δv apart from v_1 to v_n and assume that

$f_i(x) > 0$ for all $i = 1, \ldots, n$ and all $x \in (v_1, v_n)$. In the limit as Δv approaches zero, the sum (2.41) becomes a Riemann integral:

$$\lim_{\Delta v \to 0} I(x; \Delta v)\, \Delta v = \lim_{\Delta v \to 0} \sum_{i=1}^{n} r\eta'(x - v_i)\, f_V(x - v_i)\, \Delta v \tag{2.42}$$

$$= \int_{v_1}^{v_n} r\eta'(x - v)\, f_V(x - v)\, \mathrm{d}v\,. \tag{2.43}$$

This form can be used to find an upper bound in terms of the natural parameter. Let V be a random variable distributed according to the offset PDF f_V and let $u = x - v$. Because the integrand is nonnegative, (2.43) is bounded by

$$\lim_{\Delta v \to 0} I(x; \Delta v)\, \Delta v \leq r \int_{-\infty}^{\infty} \eta'(x - v)\, f_V(x - v)\, \mathrm{d}v \tag{2.44}$$

$$= r \int_{-\infty}^{\infty} \eta'(u)\, f_V(u)\, \mathrm{d}u \tag{2.45}$$

$$= r\mathbb{E}\left[\eta'(V)\right]. \tag{2.46}$$

This asymptotic bound is a useful approximation when the Fisher information is locally concentrated and the range of the nominal levels is large compared to the standard deviation of the offsets. If Δv is sufficiently small and v_1 and v_n are sufficiently far from x, then the high-resolution Fisher information is approximated by

$$I(x) \approx \frac{r}{\Delta v}\mathbb{E}\left[\eta'(V)\right]. \tag{2.47}$$

For the logistic distribution, $\eta'(x) = \frac{1}{\beta}$ is constant, so the high-resolution Fisher information is approximated by

$$I(x) \approx \frac{r}{\beta \Delta v}\,. \tag{2.48}$$

The high-resolution Fisher information approximation will appear frequently in our performance analysis. For offset distribution F, we denote this limit by γ_F, defined as

$$\gamma_F(\sigma, \Delta v, r) = c_F \frac{r}{\sigma \Delta v}\,, \tag{2.49}$$

where $c_F = \mathbb{E}\left[\sigma\eta'(V)\right]$ is a constant that depends on the distribution. For the logistic distribution, $c_L = \pi/\sqrt{3}$. For the normal distribution, $c_{\mathcal{N}} \approx 1.806$

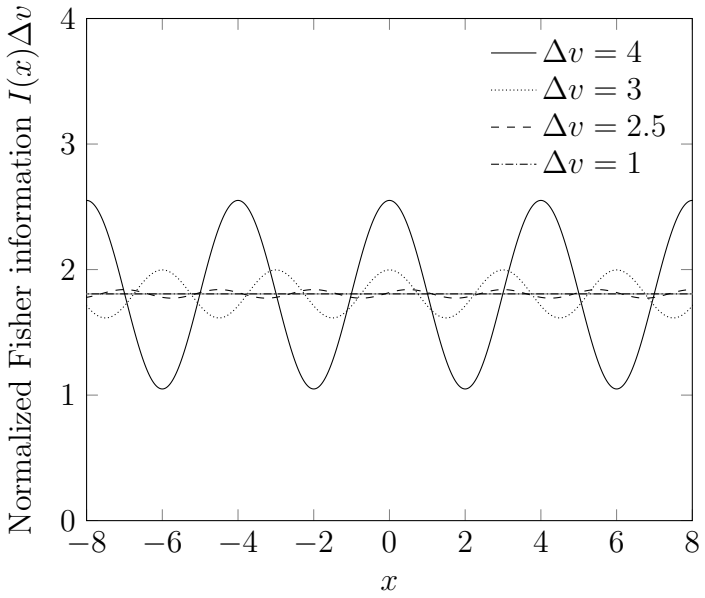

Figure 2.5: Fisher information with levels evenly spaced Δv apart and normally distributed offsets with $\sigma^2 = 1$. The gray horizontal line represents the high-resolution approximation $\gamma_{\mathcal{N}}(\Delta v)\Delta v = 1.806$.

by numerical integration. If c_F is undefined for a particular distribution, such as the uniform distribution, then the Fisher information curve does not have finite area and the high-resolution approximations are not valid.

Figure 2.5 shows the exact Fisher information for normally distributed offsets as a function of x and Δv. The approximation (2.49) is very close when $\Delta v \leq \sigma$. If x is considered to be random with a distribution that is approximately uniform over scales of Δv, then because $I(x)$ fluctuates around γ^{-1}, (2.49) gives the average Fisher information even for $\Delta v > \sigma$.

2.5 Additive Gaussian Noise Approximation

It can be shown [14] that the KL divergence between two normal distributions with means spaced d apart and variance ρ^2 is given by

$$D(q_1 \parallel q_0) = \frac{d^2}{2\rho^2}. \tag{2.50}$$

The high-resolution KL divergence (2.30) has the same form, with the variance replaced by γ^{-1}. This form suggests that we can think of the uncertainty

introduced by the sensor offsets as equivalent to additive Gaussian noise with variance $I(x)^{-1}$.

To see why, consider the scalar statistic $S(\mathbf{T}) = \sum_{i=1}^{n} \eta_i'(x) T_i$. Because S is the sum of many conditionally independent random variables, the conditional density of S given x should converge to a normal distribution by the central limit theorem. It can be seen from (2.37) that the conditional variance of S is $I(x)$. For the logistic distribution, the mean of S is linearly related to x with slope $I(x)$:

$$\mathbb{E}[S(\mathbf{T}) \mid x] = \sum_{i=1}^{n} r\eta_i'(x)\, rF_i(x) \tag{2.51}$$

$$\frac{\partial}{\partial x}\mathbb{E}[S(\mathbf{T}) \mid x] = \sum_{i=1}^{n} r\eta_i'(x)\, f_i(x) \tag{2.52}$$

$$= I(x)\ . \tag{2.53}$$

Therefore, S has a normal distribution with mean $xI(x)$ plus a constant and variance $I(x)$. That is equivalent to a constant plus $I(x)$ times a normally distributed random variable with mean x and variance $I(x)^{-1}$. Therefore, the information provided by the converter output is the same as the information in the signal corrupted by additive Gaussian noise with power $I(x)^{-1}$.

While this approximation only holds exactly for the special case of logistic offsets and in the limit as the number of sensors grows large, it is a useful rule of thumb for predicting performance in more complicated inference problems. For example, consider the problem of distinguishing between inputs drawn from one of two distributions $q_0(x)$ and $q_1(x)$. Suppose that q_0 and q_1 are normal distributions with means x_0 and x_1, respectively, and equal variance ρ^2. Let $d = |x_1 - x_0|$. Figure 2.6 shows the empirical KL divergence for a high-resolution converter with offset variance σ^2 and input noise power ρ^2. The KL divergence is well approximated by $D(p_1 \| p_0) = d^2/2(\rho^2 + \gamma_{\mathcal{N}}(\sigma^2)^{-1})$, indicating that the converter is equivalent to additive noise in its effect on classification performance.

The Gaussian noise approximation can be used as a starting point to predict performance in statistical inference applications. The next chapter explores the performance of the stochastic data converter for detection and estimation applications.

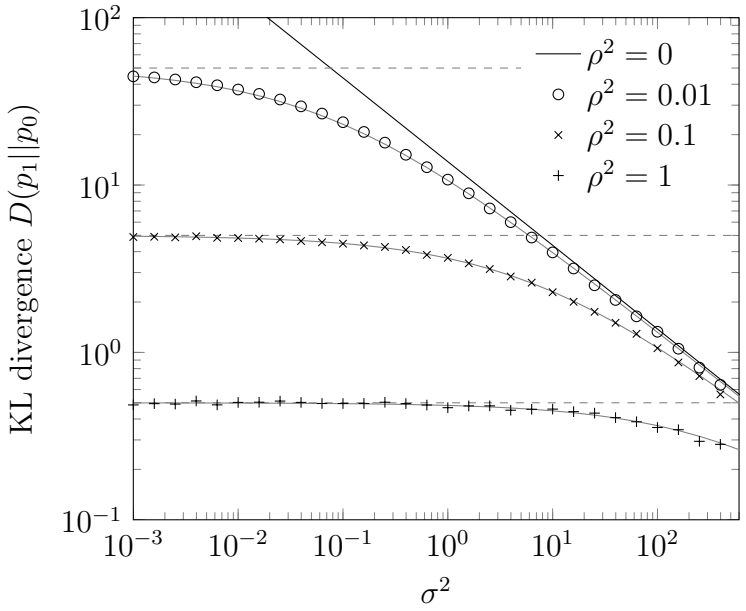

Figure 2.6: Empirical KL divergence between the observations of two normal distributions with means distance $d = 1$ apart, nominal level spacing $\Delta v = 0.2$, and noise power ρ^2 as a function of the offset variance σ^2. The dashed lines show the KL divergence due to noise alone and the gray curves show the additive noise approximation.

CHAPTER 3

STATISTICAL INFERENCE PERFORMANCE

The information-based metrics derived in Chapter 2 suggest that the performance of a stochastic converter in statistical inference applications depends on the ratio of the average density of sensor levels to the standard deviation of the discrepancy between the true reference levels and their nominal values. In this chapter, we describe several inference applications for which a stochastic converter might be used and evaluate performance using Monte Carlo simulations.

3.1 Classification

In classification problems, the system must use the sensor outputs to select one of several hypotheses or classes. Each hypothesis corresponds to a distribution on the input signal X, represented in this section as a random variable. Denote the distribution for class m by $q_m(x)$. The output statistics $\mathbf{T}$ depend on the class indirectly through X via the likelihood function

$$p_m(\mathbf{t}) = \int q_m(x)\, p_{\mathbf{T}|X}(\mathbf{t} \mid x)\, \mathrm{d}x\,. \tag{3.1}$$

A typical performance metric for a classifier is the probability of classification error, $\mathrm{P_e} = \Pr\{\hat{m}(\mathbf{T}) \neq m\}$ where m is the true class. If the classes are equally likely, then $\mathrm{P_e}$ is minimized by the maximum likelihood (ML) classifier, which selects the class for which the observations have the highest log-likelihood:

$$\hat{m}(\mathbf{T}) = \arg\max_m \ln p_m(\mathbf{T})\ . \tag{3.2}$$

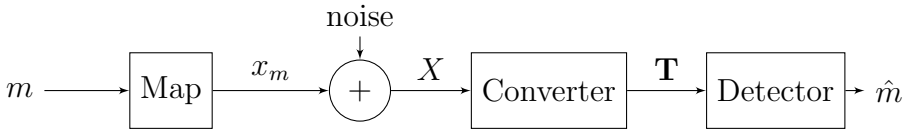

Figure 3.1: The data converter can be used to recover symbols transmitted over a noisy analog channel.

3.1.1 Binary point classification

As a special case, consider the problem of deciding between two points x_0 and x_1. That is, $q_0(x) = \delta(x - x_0)$ and $q_1(x) = \delta(x - x_1)$. The ML decision rule for binary classification takes the form of a likelihood ratio test,

$$\hat{m}(\mathbf{T}) = \begin{cases} 1, & \text{if } \ln \frac{p_1(\mathbf{T})}{p_0(\mathbf{T})} \geq 0 \\ 0, & \text{else.} \end{cases} \tag{3.3}$$

We abbreviate the binary decision rule (3.3) using the notation

$$\ln \frac{p_1(\mathbf{T})}{p_0(\mathbf{T})} \underset{0}{\overset{1}{\gtrless}} 0\,. \tag{3.4}$$

Because $p_m(\mathbf{t}) = p_{\mathbf{T}|X}(\mathbf{t} \mid x_m)$ is an exponential family, the decision rule can be written

$$\ln \frac{h(\mathbf{t}) \exp\left(\sum_{i=1}^{n} \eta_i(x_1) T_i - A(x_1)\right)}{h(\mathbf{t}) \exp\left(\sum_{i=1}^{n} \eta_i(x_0) T_i - A(x_0)\right)} \underset{0}{\overset{1}{\gtrless}} 0 \tag{3.5}$$

$$\sum_{i=1}^{n} (\eta_i(x_1) - \eta_i(x_0)) T_i \underset{0}{\overset{1}{\gtrless}} A(x_1) - A(x_0)\,. \tag{3.6}$$

If the offsets have a logistic distribution, the rule can be simplified further to

$$\sum_{i=1}^{n} \frac{x_1 - x_0}{\beta} T_i \underset{0}{\overset{1}{\gtrless}} A(x_1) - A(x_0) \tag{3.7}$$

$$S(\mathbf{T}) \underset{0}{\overset{1}{\gtrless}} \frac{A(x_1) - A(x_0)}{d}\,. \tag{3.8}$$

3.1.2 Symbol detection in noise

An important classification application for which ADCs are often used is symbol detection in communication receivers. Consider a simple ampli-

tude modulated signal with an alphabet of M equally-probable information symbols $x_1, \ldots, x_M$, spaced d apart. These symbols are passed through a communication channel and corrupted by additive Gaussian noise with power ρ^2, as shown in Figure 3.1. The received signal X therefore has a conditionally normal distribution with mean x_m and variance ρ^2.

If the received signal can be measured exactly, then the ML decision rule is to choose the x_m closest to X. The expected error probability for the ML detector is

$$\mathrm{P}_{\mathrm{e,ideal}} = \frac{2M-2}{M} Q\left(\frac{d}{2\rho}\right), \tag{3.9}$$

where Q is the complementary standard normal CDF.

If, instead, the symbols are measured by a stochastic data converter, there will be additional uncertainty from the observations and the error probability will be higher. If the converter can be thought of as additive Gaussian noise with power γ^{-1}, as proposed in Section 2.5, then the error probability would be

$$\mathrm{P}_{\mathrm{e}} = \frac{2M-2}{M} Q\left(\frac{d}{2\sqrt{\rho^2+\gamma^{-1}}}\right). \tag{3.10}$$

3.1.3 Simulated performance

To demonstrate the symbol detection performance of the converter, a communication link was simulated with an alphabet of $M = 16$ symbols spaced $d = 0.5$ apart. The symbols were transmitted over a Gaussian noise channel with noise power ρ^2 and received by a stochastic data converter with nominal levels spaced $\Delta v = 0.1$ apart, normally distributed offsets with variance σ^2, and $r = 8$ sensors per nominal level. In the receiver, an ML decision rule uses the observations $\mathbf{T}$ directly to select the most likely transmitted symbol; there is no intermediate estimation step. The empirical error probability was computed over all symbols for 4000 sets of random levels with 100 samples each of random channel noise.

The error rate of the ML detector is shown in Figure 3.2. For low signal-to-noise ratios (SNR), the error rate is similar to that of a receiver with an ideal ADC, while the high SNR performance is limited by the converter offset variance. Figure 3.3 shows the same data plotted against an effective signal-to-noise ratio, where the stochastic converter is treated as additive

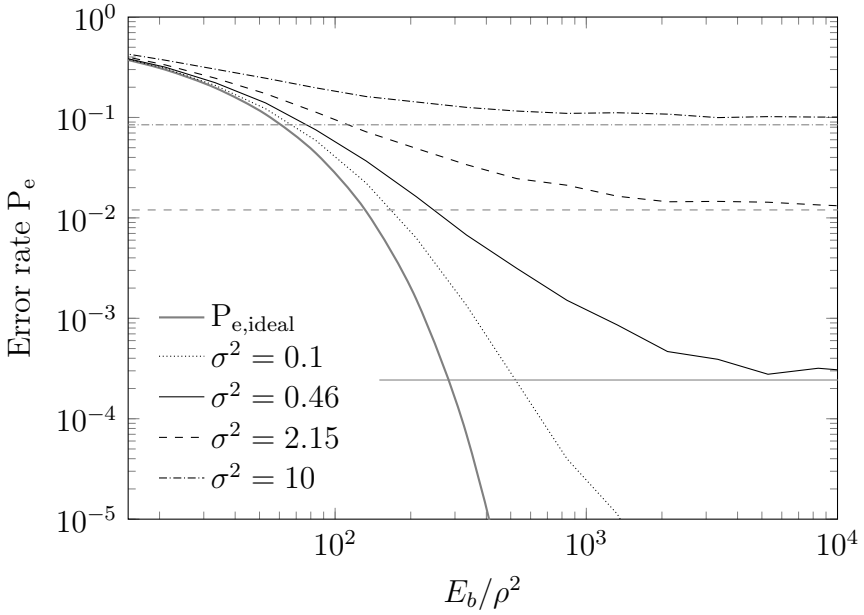

Figure 3.2: Simulated error probability for detecting noisy symbols with a stochastic data converter. The gray lines show the noise-free error rate for each value of the offset variance, σ^2.

white Gaussian noise with power γ^{-1}. The simulation data closely match the error probability predicted by (3.10), suggesting that the Gaussian noise approximation is useful for this problem.

3.2 Regression

In classification problems, the classes are chosen from a finite set. If instead we are interested in a continuous-valued parameter, we must use regression. In this section we consider the problem of measuring the input signal using the observations, so that the parameter of interest is X itself. For this problem we wish to minimize the mean square error, defined as

$$\mathrm{MSE}\left(\hat{x}\right) = \mathbb{E}\left[\left(\hat{x}\left(\mathbf{T}\right) - X\right)^2\right] . \tag{3.11}$$

If the distribution of X is known, then the MSE can be computed exactly. Furthermore, the minimum mean square error (MMSE) estimator is given by

$$\hat{x}_{\mathrm{MMSE}}\left(\mathbf{T}\right) = \mathbb{E}\left[X \mid \mathbf{T}\right] . \tag{3.12}$$

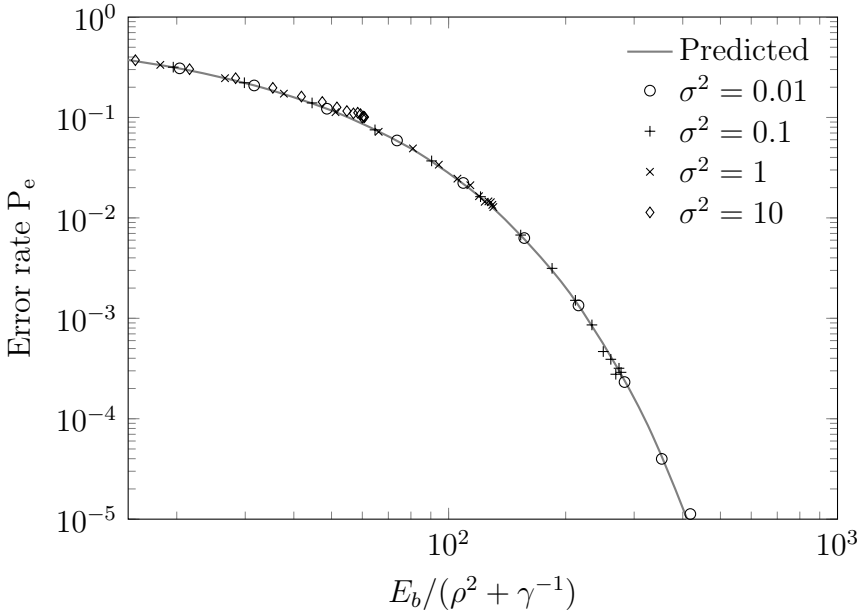

Figure 3.3: Simulated error probability as a function of effective signal-to-noise ratio.

If the prior distribution of X is not known, then it is useful to consider the variance of the estimator at particular points:

$$\mathrm{Var}_x\left(\hat{x}(\mathbf{T})\right) = \mathrm{Var}\left(\hat{x}\left(\mathbf{T}\right) \mid X = x\right) . \tag{3.13}$$

The variance can be made arbitrarily low for biased estimators. An estimator is called biased if $\mathbb{E}\left[\hat{x}\left(\mathbf{T}\right) \mid X\right] \neq X$. An estimator is unbiased if $\mathbb{E}\left[\hat{x}\left(\mathbf{T}\right) \mid X\right] = X$. For any unbiased estimator $\hat{x}_{\mathrm{UE}}$, the Cramér-Rao lower bound (CRLB) [12] on the conditional variance is

$$\mathrm{Var}_x\left(\hat{x}_{\mathrm{UE}}(\mathbf{T})\right) \geq I\left(x\right)^{-1} . \tag{3.14}$$

An unbiased estimator that achieves this bound with equality is called efficient.

One commonly used estimator is the maximum likelihood estimator (MLE), defined by

$$\hat{x}_{\mathrm{ML}}\left(\mathbf{T}\right) = \arg\max_x \ln p_{\mathbf{T}|X}\left(\mathbf{T} \mid x\right) . \tag{3.15}$$

The MLE is neither unbiased nor efficient in general, but often achieves good performance in practice.

3.2.1 Maximum likelihood estimator

We now derive the MLE for x based on the stochastic converter outputs. To ensure that the parameter is identifiable, we take the maximization over the set $\mathcal{X}(\mathbf{T}) = \left\{x : p_{\mathbf{T}|X}(\mathbf{T} \mid x) > 0\right\}$. If x is identifiable, i.e., $\mathcal{X}(\mathbf{T})$ is not empty, the MLE is

$$\begin{aligned}
\hat{x}_{\mathrm{ML}}(\mathbf{T}) &= \arg\max_{x} p_{\mathbf{T}|X}(\mathbf{T} \mid x) && (3.16)\\
&= \arg\max_{x \in \mathcal{X}(\mathbf{T})} \prod_{i:f_i(x)>0} p_{T_i|X}(T_i \mid x) && (3.17)\\
&= \arg\max_{x \in \mathcal{X}(\mathbf{T})} \sum_{i:f(x)>0} \ln p_{T_i|X}(T_i \mid x)\ . && (3.18)
\end{aligned}$$

Since the log-likelihood function for an exponential family is concave, the MLE is the unique solution to the likelihood equation

$$0 = \sum_{i:f_i(x)>0} \frac{\partial}{\partial x} \ln p_{T_i|X}(T_i \mid x)\ . \tag{3.19}$$

Because $\left\{p_{\mathbf{T}|X}(\cdot \mid x)\right\}_x$ is an exponential family of the form (2.11), the derivative of the log-likelihood is given by

$$\begin{aligned}
\sum_{i:f_i(x)>0} \frac{\partial}{\partial x} \ln p_{T_i|X}(T_i \mid x) &= \sum_{i:f_i(x)>0} \eta_i'(x)\, T_i - A'(x) && (3.20)\\
&= \sum_{i:f_i(x)>0} \eta_i'(x) T_i - \sum_{i:f_i(x)>0} r \frac{f_i(x)}{\bar{F}_i(x)} && (3.21)\\
&= \sum_{i:f_i(x)>0} \eta_i'(x) T_i - \sum_{i:f_i(x)>0} r \eta_i'(x) F_i(x)\,. && (3.22)
\end{aligned}$$

Therefore, using (2.7), the likelihood equation can be written

$$\begin{aligned}
\sum_{i:f_i(x)>0} \eta_i'(\hat{x}_{\mathrm{ML}})\, T_i &= \sum_{i:f_i(x)>0} \eta_i'(\hat{x}_{\mathrm{ML}})\, r F_i(\hat{x}_{\mathrm{ML}}) && (3.23)\\
\sum_{i:f_i(x)>0} \eta_i'(\hat{x}_{\mathrm{ML}})\, T_i &= \sum_{i:f_i(x)>0} \eta_i'(\hat{x}_{\mathrm{ML}})\, \mathbb{E}\left[T_i \mid X = \hat{x}_{\mathrm{ML}}\right]\ . && (3.24)
\end{aligned}$$

The likelihood equation (3.24) can often be solved numerically for arbitrary offset distributions. If the offsets have a logistic distribution, the equation

can be simplified in terms of the sufficient statistic:

$$S(\mathbf{T}) = \mathbb{E}\left[S(\mathbf{T}) \mid X = \hat{x}_{\mathrm{ML}}\right] . \tag{3.25}$$

In the high-resolution limit, the mean of the statistic S can be expressed as an integral:

$$\lim_{\Delta v \to 0} \mathbb{E}\left[S(\mathbf{T}) \mid X = x\right] \Delta v = \frac{r}{\beta} \int_{v_1}^{v_n} F_V(x - v)\, \mathrm{d}v . \tag{3.26}$$

Making the substitution $u = x - v$ yields

$$\begin{aligned} \lim_{\Delta v \to 0} \mathbb{E}\left[S(\mathbf{T}) \mid X = x\right] \Delta v &= \frac{r}{\beta} \int_{x-v_n}^{x-v_1} F_V(u)\, \mathrm{d}u & (3.27)\\ &= \frac{r}{\beta} \int_{x-v_n}^{x-v_1} \frac{\exp\left(\frac{u}{\beta}\right)}{1 + \exp\left(\frac{u}{\beta}\right)} \mathrm{d}u & (3.28)\\ &= r \ln \frac{1 + \exp\left(\frac{x-v_1}{\beta}\right)}{1 + \exp\left(\frac{x-v_n}{\beta}\right)} . & (3.29) \end{aligned}$$

If Δv is sufficiently small and x is sufficiently far from v_1 and v_n, the mean is well approximated by the linear relationship

$$\begin{aligned} \mathbb{E}\left[S(\mathbf{T})\right] &\approx \frac{r}{\beta \Delta v} (x - v_1) & (3.30)\\ &= \gamma_L (x - v_1) & (3.31) \end{aligned}$$

and the MLE approaches the linear estimator

$$\begin{aligned} \hat{x}_{\mathrm{ML}}\left(S(\mathbf{T})\right) &= v_1 + \gamma_L^{-1} S(\mathbf{T}) & (3.32)\\ &= v_1 + \frac{\Delta v}{r} \sum_{i=1}^{n} T_i . & (3.33) \end{aligned}$$

In the high-resolution, high-range limit for logistic offsets, the MLE is a linear function of the total number of threshold exceedances. It is straightforward to verify that this estimator is efficient.

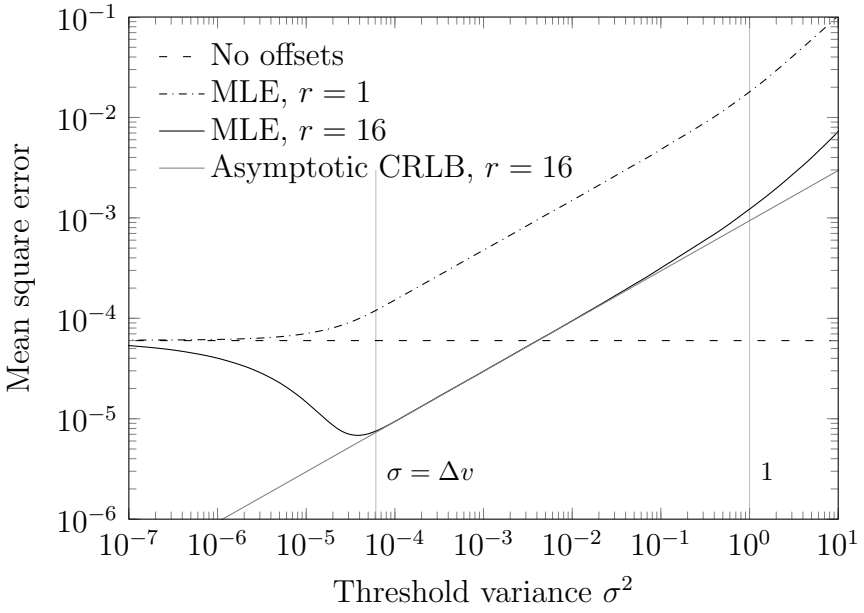

Figure 3.4: Simulated mean square error as a function of offset variance for the MLE with $n = 128$ and normally distributed offsets.

3.2.2 Simulated performance

The theoretical results for estimation performance were verified using Monte Carlo simulations of the converter architecture for a traditional analog-to-digital conversion application. The input signal is drawn from a uniform distribution with unit variance. The converter is designed with n nominal levels uniformly spaced distance $\Delta v = \sqrt{12}/(n-1)$ apart. The comparator outputs are used to compute the MLE $\hat{x}_{\mathrm{ML}}$ of x. The performance metric is the MSE of the estimate averaged over all inputs and all realizations of the converter offsets. For each data point, 600 sets of offsets were randomly generated from a normal distribution with mean 0 and variance σ^2.

Figure 3.4 shows the MSE performance of the MLE as a function of the offset variance for a converter with $n = 128$. The sample MSE is close to the high-resolution CRLB approximation, $\gamma_{\mathcal{N}}^{-1}$, as long as $\sigma^2 > \Delta v^2$. It is close to the quantization error of an offset-free quantizer (1.1) when $\sigma^2 \ll \Delta v^2$. When $r > 1$ and $\sigma^2 < \Delta v^2$, larger offsets improve the MSE performance through a dithering effect. Performance degrades for $\sigma^2 > 1$ because the levels fall outside the range of the input signal.

Figure 3.5 shows the MSE performance as a function of the number of nominal levels n and the number of observations per level r. As long as n

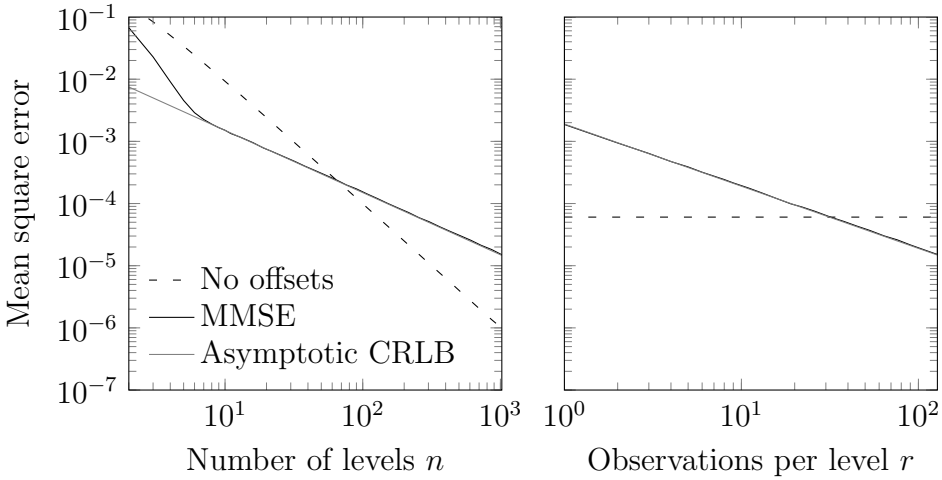

Figure 3.5: Simulated mean square error as a function of n and r for the MLE with normally distributed offsets with $\sigma^2 = 1/64$. Left: $r = 16$. Right: $n = 128$.

is sufficiently large that $\Delta v < \sigma$, the MSE is inversely proportional to both n and r. That is, when the level distributions overlap significantly between nominal levels, performance depends on the total number of comparators in the system.

3.3 A Multistage Measurement Architecture

For a high-resolution converter with a large input range, the number of sensors may be large. However, for any given input, only a subset of the sensors contribute meaningful information about the parameter. To reduce the complexity of the system, the statistical estimation procedure can be split into two stages, as illustrated in Figure 3.6. In the first stage, a low-resolution detector forms a rough estimate of the input signal. This rough estimate is used to select a subset of the outputs from sensors with nominal levels close to the signals.

This architecture simplifies the inference stage by reducing the dimensionality of the sufficient statistic. Furthermore, while the full likelihood function can be difficult to analyze for non-logistic offset distributions, the likelihood tends to be locally well-behaved. In particular, for the set of outputs of sensors with nominal levels close to x and under appropriate regularity conditions, the derivative of the log-likelihood function can be

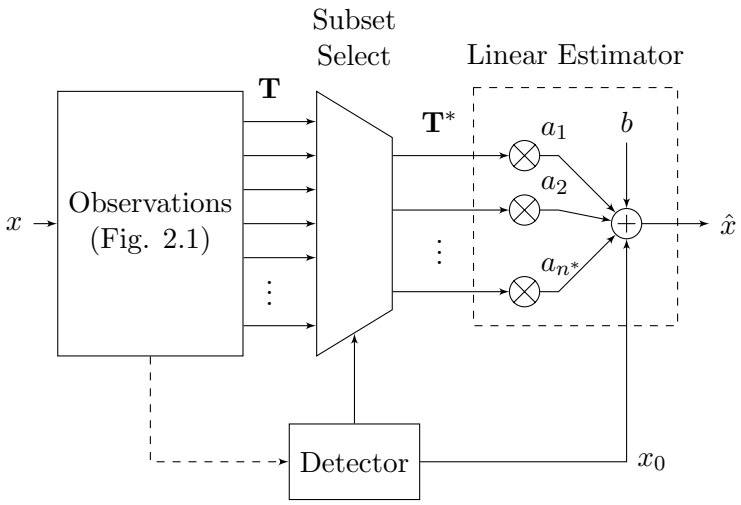

Figure 3.6: The two-stage estimation architecture includes a low-resolution detector and a linear estimator.

closely approximated by a linear function of x. In this section, we apply this approximation to the signal measurement problem and find a simple linear estimator that can perform nearly as well as the full MLE in a two-stage architecture.

3.3.1 Local linear estimator

In Section 3.2.1, we showed that for a logistic offset distribution, the MLE is efficient. That is not true in general for other distributions. However, we can find a linear estimator of a form similar to (3.32) using a local approximation. Suppose that x is known to be close to a point x_0. Then we can locally approximate $\eta_i(x)$ by $\eta_i'(x_0)(x - v_i)$ and $F_i(x)$ by $F_i(x_0) + f_i(x_0)(x - x_0)$. With these approximations, the likelihood equation (3.24) becomes

$$\sum_{i=1}^{n} \eta_i'(x_0) T_i \approx \sum_{i=1}^{n} \eta_i'(x_0) r\left(F_i(x_0) + f_i(x_0)(x - x_0)\right) \quad (3.34)$$

$$\sum_{i=1}^{n} \eta_i'(x_0)\left(T_i - rF_i(x_0)\right) \approx \sum_{i=1}^{n} r\eta_i'(x_0) f_i(x_0)(x - x_0) \ . \quad (3.35)$$

Substituting (2.41), we have

$$\sum_{i=1}^{n} \eta_i'(x_0)\left(T_i - rF_i(x_0)\right) \approx I(x_0)(x - x_0) \ . \quad (3.36)$$

Solving this approximate likelihood equation gives $\hat{x}_{\mathrm{LL}}$, the local linear estimator:

$$\hat{x}_{\mathrm{LL}}\left(\mathbf{T}, x_0\right) = x_0 + I\left(x_0\right)^{-1} \sum_{i=1}^{n} \eta_i'\left(x_0\right)\left(T_i - rF_i\left(x_0\right)\right) . \tag{3.37}$$

If $x_0 = x$, then $\hat{x}_{\mathrm{LL}}$ is unbiased and efficient, as in the logistic case. Otherwise, the estimator has higher variance and may be biased. Its performance depends on the accuracy of the initial estimate x_0.

3.3.2 Estimation with a subset of statistics

The linear estimator (3.32) is a function of the full observation statistic vector $\mathbf{T}$. However, for many unimodal distributions, only a subset of the statistics contribute significantly to the estimate, as shown by Figure 2.4. For normally distributed offsets, more than 99% of the Fisher information is contributed by sensors with nominal levels within 3σ of x. We therefore restrict our attention to the n^* sensors with nominal levels closest to x_0. Let $\mathbf{T}^*$ denote this length-n^* subset of $\mathbf{T}$. For the remainder of this section, we relabel F_k, $\bar{F}_k$, f_k, η_k, and I_k to correspond to the elements T_k of $\mathbf{T}^*$. It is assumed that $f_k\left(x_0\right) > 0$ for $k = 1, \ldots, n^*$. The Fisher information for $\mathbf{T}^*$ **is** $I^*\left(x\right) = \sum_{k=1}^{n^*} I_k\left(x\right)$. On this subset, the local linear estimator (3.37) becomes

$$\hat{x}_{\mathrm{LL}}\left(\mathbf{T}^*, x_0\right) = x_0 + I^*\left(x_0\right)^{-1} \sum_{k=1}^{n^*} \eta_k'\left(x_0\right)\left(T_k - rF_k\left(x_0\right)\right) . \tag{3.38}$$

The bias and variance of $\hat{x}_{\mathrm{LL}}$ depend on the initial estimate x_0. The conditional mean of $\hat{x}_{\mathrm{LL}}$ for uniformly spaced levels and normal offsets is shown in Figure 3.7. The conditional variance is shown in Figure 3.8. The estimator is minimum variance unbiased when $x_0 = x$. Both the bias and variance increase as x_0 grows farther from x. To achieve good performance, therefore, the coarse estimator should be well within one standard deviation of the true parameter.

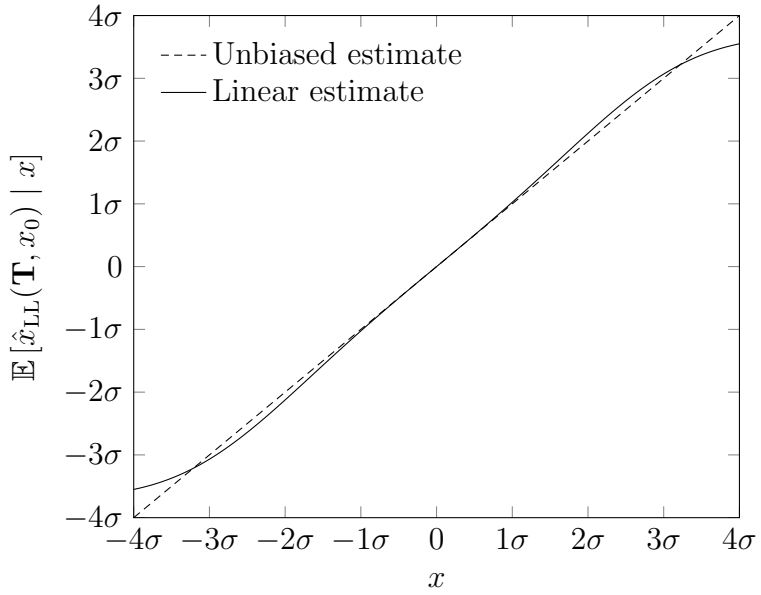

Figure 3.7: Mean of the local linear estimate with $x_0 = 0$ for closely spaced uniform levels and normally distributed $(0, \sigma^2)$ level offsets. The dashed line represents an unbiased estimator.

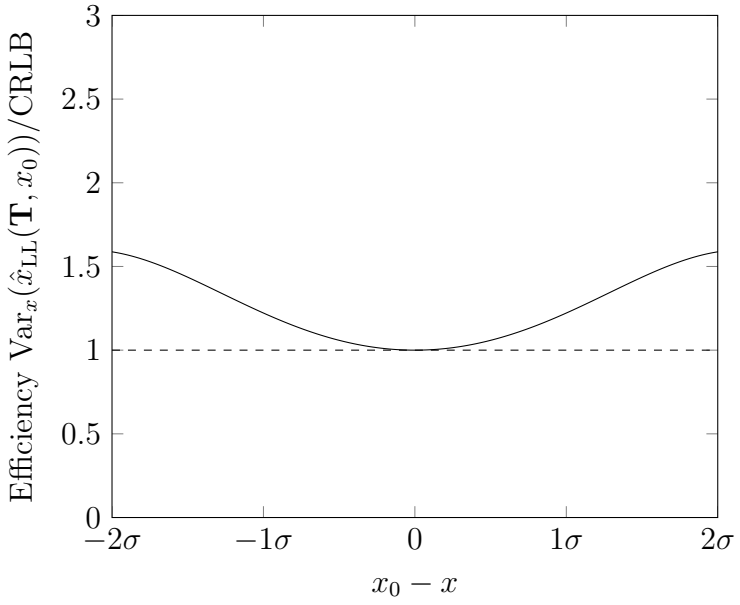

Figure 3.8: Variance of the local linear estimate with $x_0 = 0$ for closely spaced uniform levels and normally distributed $(0, \sigma^2)$ level offsets. The dashed line represents an efficient estimator.

3.3.3 Simulated performance

We will demonstrate the performance of the proposed two-stage architecture using a simulated example. The n nominal references are uniformly spaced $\Delta v = 1/(n-1)$ apart from 0 to 1. The sensors have normally distributed offsets with mean 0 and $\sigma = 0.05$. For the low resolution detector, we use the sum of the first output at each level,

$$x_0(\mathbf{Y}) = v_1 + \Delta v \sum_{i=1}^{n} Y_{i,1} \,. \tag{3.39}$$

The local linear estimator uses a subset of outputs with nominal levels within $\pm\delta$ of x_0, or the nearest two levels if there are no levels within $\pm\delta$. If x is near the boundaries of the level range, fewer levels are included and the weights are normalized accordingly.

In the Monte Carlo simulation, the number of levels was varied from $n = 4$ to 250 and the range parameter was varied from $\delta = 0.025$ to 0.150. For each configuration, 2000 input signals were drawn uniformly at random from $[0, 1]$ and 2000 sets of offsets were drawn from the normal distribution with $\sigma = 0.05$. For comparison, each signal was also estimated using the MLE with the full vector of observations. The results, shown in Figure 3.9, are averaged over all inputs and all sets of offsets.

The simulation results confirm that the two-stage estimator performs nearly as well as the MLE and close to the CRLB as long as the spacing between nominal levels is small compared to σ. The size of the subset is a tradeoff between complexity and performance. If the range is too narrow, more observations will be required to achieve the desired level of performance and the estimator will be more sensitive to error in the initial estimate. On the other hand, there is little benefit to including observations for sensors whose levels are far from the signal. Figure 3.9 suggests that the range should be at least one standard deviation to achieve reasonable performance.

This architecture is best suited to systems where the offset deviation is small compared to the input range but large compared to the desired error. Then the linear estimator performs nearly as well as a more complex MLE. The choice of design parameters depends on the estimate of the offset deviation σ. If the assumed value of σ is too high, the levels may be spaced too far apart. If it is too low, the coefficients will be too large. Thus, it

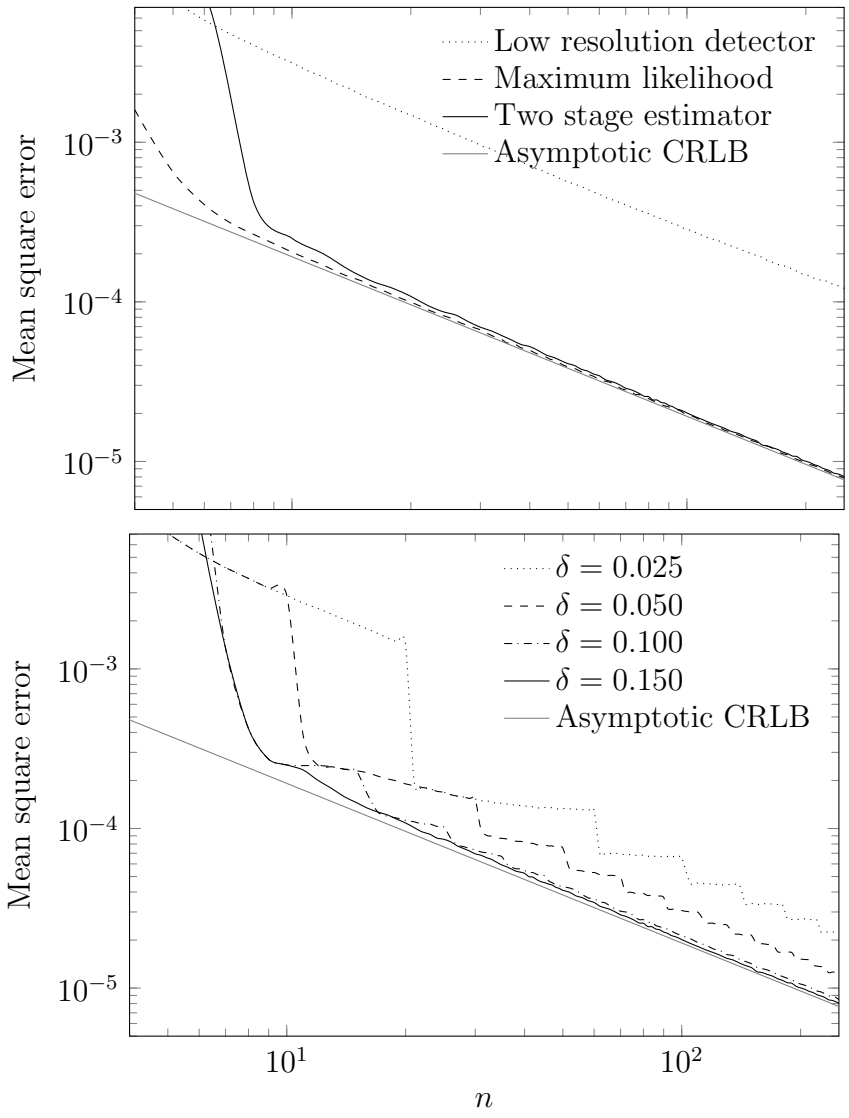

Figure 3.9: Simulated performance of the two-stage architecture. Top: $r = 15$ and $\delta = 0.015$. Bottom: $r = 15$. The gray curves show the asymptotic CRLB.

is best to use a conservative estimate of σ to choose the level spacing and a generous estimate to set the coefficients. If the true value of σ is smaller than the assumed value, the estimator variance will be worse than the best achievable performance but better than the designed performance.

3.4 Calibrated Measurement

The analysis presented in the previous sections is for inference with unknown random levels. This model is useful for time-varying uncertainties such as sensor noise. However, offsets due to process variation can generally be considered static. It should therefore be possible to measure the offsets or to learn them over time; then the converter would become a nonuniform quantizer. In this section, we consider estimation performance for a converter where the levels are randomly generated but are known exactly to the estimator. We can thereby find the best-case performance of an inference system that accounts for the static nature of the offsets. An advantage of this formulation is that we can consider nonuniform distributions of nominal levels and nonuniform input distributions. In this section, the input X is considered a random variable with a known PDF $f_X(x)$.

3.4.1 Mean square error

Because the levels are assumed to be measured, we can express the mean square error of the estimator in terms of the order statistics, $V_{(1)} \leq V_{(2)} \leq \cdots \leq V_{(N)}$, of the levels. Here $N = nr$ is the total number of sensors. For convenience of notation, define $V_{(0)}$ to be the infimum of the support of X and $V_{(N+1)}$ to be the supremum of the support of X. Let $Z = \sum_{i=1}^{n} T_i$ be the total number of exceedances. The MMSE estimator for a quantizer with known levels is

$$\hat{x}_{\mathrm{MMSE}}(\mathbf{Y}) = \mathbb{E}\left[X \mid V_{(Z)} \leq X < V_{(Z+1)}\right] . \tag{3.40}$$

If $f_X(x)$ is uniform on $[V_{(Z)}, V_{(Z+1)})$ then the MMSE estimate is simply the midpoint of the interval.

The mean square error achievable by this quantizer for a given set of levels

is

$$\mathrm{MSE}\left(\mathbf{V}\right) = \sum_{z=0}^{N} \mathrm{Var}\left(X \mid V_{(z)} \leq X < V_{(z+1)}\right) \Pr\left\{X \in [V_{(z)}, V_{(z+1)})\right\} \ . \quad (3.41)$$

If $f_X\left(x\right)$ is slowly varying so that it can be assumed to be approximately constant on each interval $[V_{(z)}, V_{(z+1)})$, the MSE is approximated by

$$\begin{aligned} \mathrm{MSE}\left(\mathbf{V}\right) &\approx \sum_{z=0}^{N} \frac{\left(V_{(z+1)} - V_{(z)}\right)^2}{12} f_X\left(V_{(z)}\right)\left(V_{(z+1)} - V_{(z)}\right) & (3.42) \\ &= \sum_{z=0}^{N} f_X\left(V_{(z)}\right) \frac{\left(V_{(z+1)} - V_{(z)}\right)^3}{12} \, . & (3.43) \end{aligned}$$

The average MSE over all realizations of the stochastic converter is the expectation of (3.43). This expectation can be evaluated in closed form for the special case that both X and all $V_{i,j}$ are drawn from a uniform distribution on an interval of length Δx. In that case,

$$\mathrm{MSE}\left(\Delta x\right) = \frac{1}{12\Delta x} \sum_{z=0}^{N} \mathbb{E}\left[\left(V_{(z+1)} - V_{(z)}\right)^3\right] \ .$$

It can be shown by integration that the third moment of intervals between uniform order statistics is $\left(\Delta x\right)^3 \binom{N+3}{N}^{-1}$. Therefore we have

$$\begin{aligned} \mathrm{MSE}\left(\Delta x\right) &\approx \frac{1}{12\Delta x} \sum_{z=0}^{N} \left(\Delta x\right)^3 \left(\frac{3!N!}{(N+3)!}\right) & (3.44) \\ &= \frac{\left(\Delta x\right)^2}{12}(N+1)\left(\frac{6}{(N+1)(N+2)(N+3)}\right) & (3.45) \\ &= \frac{\left(\Delta x\right)^2}{2(N+2)(N+3)} \, . & (3.46) \end{aligned}$$

For a deterministic quantizer with N levels spaced uniformly at a distance $\Delta v = \Delta x/(N+1)$ apart, the MSE from (1.1) would be

$$\mathrm{MSE}_{\mathrm{ideal}}\left(\Delta x\right) = \frac{(\Delta x)^2}{12(N+1)^2} \, . \quad (3.47)$$

For large N, therefore, the MSE of a uniform stochastic quantizer is roughly six times larger than that of a deterministic quantizer.

3.4.2 High resolution limit

If the converter has a large number of reference levels that are closely spaced, then the MSE can be approximated by an integral, known as the distortion integral [15]. We first derive the distortion integral for the deterministic case. Suppose that the levels are nonuniformly distributed according to a density of $N\lambda(x)$ levels per unit length. That is, the spacing between levels near x is $1/(N\lambda(x))$. We first partition the quantizer range into intervals $[x_m, x_{m+1})$, $m = 1, \ldots, M$, let $\delta x_m = x_{m+1} - x_m$, and assume that the levels are dense enough that both $\lambda(x)$ and $f_X(x)$ can be considered constant on each interval. Taking the average of the MSE over each interval using (3.47) and letting $\delta x_m \to \mathrm{d}x$ so that the sum becomes an integral, the distortion is

$$\begin{aligned} \mathrm{MSE}_{\mathrm{ideal}} &= \sum_{m=1}^{M} \Pr\{X \in [x_m, x_{m+1})\} \frac{\delta x_m^2}{12\left(N\lambda(x_m)\,\delta x_m\right)^2} && (3.48)\\ &\approx \frac{1}{12N^2} \sum_{m=1}^{M} f(x_m)\,\lambda(x)^{-2}\,\delta x_m && (3.49)\\ &\approx \frac{1}{12N^2} \int_{-\infty}^{\infty} f(x)\,\lambda(x)^{-2}\,\mathrm{d}x\,. && (3.50) \end{aligned}$$

For the nondeterministic case, replace the deterministic density $\lambda(x)$ by the average probability density $(\lambda \star f_V)(x)$, where f_V is the offset pdf and $\star$ denotes convolution, and multiply the MSE of each interval by six to account for the randomness of the levels. Then the distortion integral for the stochastic converter is

$$\begin{aligned} \mathrm{MSE} &\approx \sum_{m=1}^{M} \Pr\{X \in [x_m, x_{m+1})\} \frac{\delta x_m^2}{2\left(N(\lambda \star f_V)(x_m)\,\delta x_m\right)^2} && (3.51)\\ &\approx \frac{1}{2N^2} \int_{-\infty}^{\infty} f(x)\left((\lambda \star f_V)(x)\right)^{-2}\,\mathrm{d}x\,. && (3.52) \end{aligned}$$

This approximation is useful for high-resolution quantizers where the offsets are large compared to the spacing between nominal levels so that the levels can be considered to have a smooth average density. Notice that the offset density affects the mean square error only by filtering the nominal density. If the nominal density is fairly uniform, then the particular distribution of the offsets is irrelevant. Furthermore, the variance of the offsets does not appear in (3.52) except as a constant factor of six, whereas it directly affects

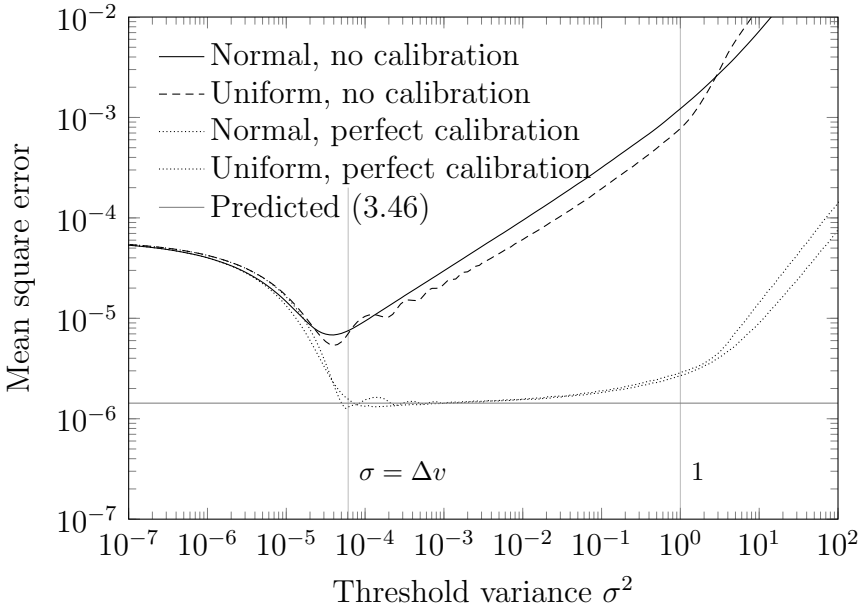

Figure 3.10: Simulated MSE performance of an MMSE estimator with uniformly distributed inputs. There are $n = 128$ nominal levels with $r = 16$ sensors per level and either uniformly or normally distributed offsets.

the performance of the non-calibrated estimator.

3.4.3 Simulated performance

In this section, we compare the simulated performance of two MMSE estimators, one that is given the exact values of the random levels and one that is not. The input was drawn from a uniform distribution. The non-calibrated estimator computes the MMSE estimate using the observation likelihood and the known uniform input density; it is equivalent to the MLE from Section 3.2. The calibrated estimator uses (3.40) and selects the midpoint of the interval containing the input. The simulation results shown here used 600 sets of levels and 512 random inputs.

Figure 3.10 compares the performance of the calibrated and non-calibrated estimators as well as an ideal quantizer. The MSE of the non-calibrated estimator increases with σ and is close to the high-resolution CRLB. The calibrated estimator has nearly constant MSE for $\sigma > \Delta v$. The calibrated MSE is about six times larger than that of an ideal quantizer, as expected. The error of both estimators increases rapidly for $\sigma > x$ because the levels

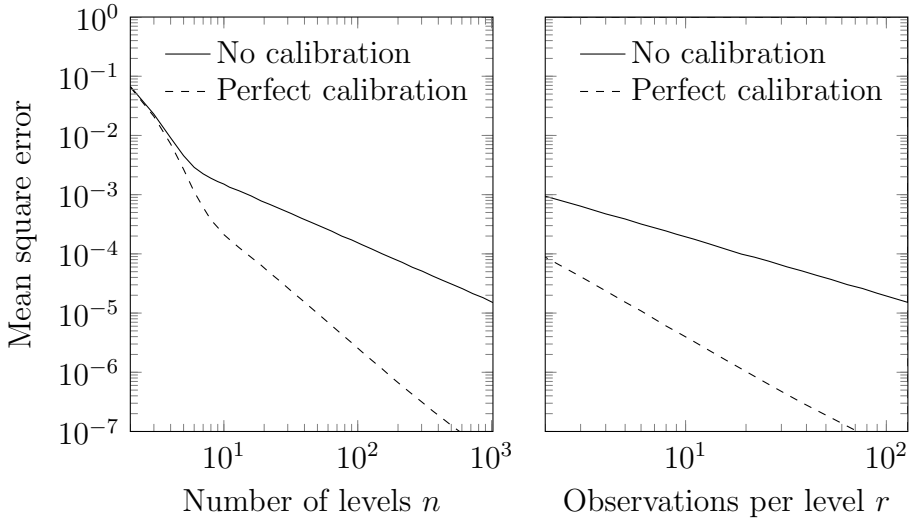

Figure 3.11: Simulated MSE performance of an MMSE estimator with uniformly distributed inputs. The offsets are normally distributed with $\sigma^2 = 1/64$. Left: $r = 16$ sensors per level. Right: $n = 128$ nominal levels.

fall outside the signal range.

Figure 3.11 shows the scaling of both estimators with n and r. For n large enough that $\sigma > \Delta v$, the MSE of the non-calibrated estimator is inversely proportional to N. The MSE of the calibrated estimator scales as $1/N^2$, much like a deterministic quantizer. Thus, if the true levels can be measured or estimated, the MSE performance of the stochastic converter is roughly a constant factor worse than that of an ideal quantizer.

CHAPTER 4

CONCLUSIONS

The analysis and simulation results presented in this thesis suggest that the performance of a high-resolution stochastic data converter depends on the ratio of the average density of sensor levels to the standard deviation of the level offsets. As a rule of thumb for predicting performance, the uncertainty in the stochastic converter can be thought of as additive Gaussian noise with power proportional to $\sigma\Delta v/r$. The analytic results of Chapter 2 show that this approximation holds in the high-resolution limit. The simulation results of Chapter 3 suggest that it is a good approximation as long as the level distributions overlap significantly, i.e. $\sigma > \Delta v$.

4.1 Stochastic Data Converter Design

The design parameters for the data converter architecture presented here are the number of levels and their nominal values, the number of sensors per level, and the variance of the offsets. To design a high-resolution converter using a number of sensors with fixed offset variance, the designer should first select enough nominal levels that $\Delta v > \sigma$. If the offsets are small relative to the spacing, then the design is similar to a conventional quantizer and there is little benefit to adding multiple sensors per level. Once the levels are dense enough that the distributions overlap, performance depends on the total number of sensors. Thus, performance can be improved by replicating the existing sensor array rather than adding new levels. This strategy also simplifies the design of the inference algorithm because the dimension of the sufficient statistic $\mathbf{T}$ does not increase with r.

The majority of the simulation results in this thesis are for the traditional ADC problem of measuring an input signal. However, the proposed architecture is well suited to any statistical decision-making problem. Conventional

ADCs make hard decisions to produce a single discrete representation of an input signal. The stochastic converter outputs a soft decision in the form of a vector of sufficient statistics. These statistics have well defined likelihood functions that can be used directly in soft-input inference algorithms such as maximum likelihood detection. The inference algorithms must then account for the uncertainty in the converter as well as the uncertainty in the input signal itself.

One difficulty in using the output vector for inference is its high dimension. A conventional quantizer outputs a single symbol to encode a quantization cell; the stochastic converter outputs a length-n vector of integers. Section 3.3 showed that a two-stage design can be used to reduce the dimension of the output statistic. If the converter makes an initial rough estimate, it can narrow the output to a few statistics with high information content.

4.2 Advantages and Limitations of Stochastic Data Conversion

The stochastic converter is useful in applications for which reliable sensors are unavailable or costly. The converter can be built with lower-cost or lower-yield sensors because each individual sensor can behave unreliably. This advantage is critical for emerging circuit technologies that could substantially improve efficiency at the cost of reliability. By averaging over many measurements, the converter can achieve reliable average-case performance from unreliable components. On the other hand, because the sensor characteristics are random, the worst-case performance can be arbitrarily poor. If an application calls for worst-case performance guarantees, such an architecture would not be appropriate.

The stochastic converter also reduces design complexity in some ways. Unlike most previously proposed designs for quantization with highly variable comparators, the architecture proposed here requires no calibration after production. There is no dedicated error correction logic; instead, the uncertainty is accounted for by the inference algorithm that processes the converter output. The design is easily scaled: if a higher level of performance is required, the designer can simply replicate the existing circuit several times and combine the outputs. As long as the offsets are independent,

replication is as effective as designing new sensors. The two-stage architecture from Section 3.3 relies on highly parallel linear operations, an important advantage in applications for which latency is critical. On the other hand, the redundancy in the converter significantly increases the number of sensors required compared to a conventional quantizer. The speed and efficiency of the design comes at the expense of area.

4.3 Stochastic Data Conversion for Emerging Inference Applications

This thesis presented a novel architecture for stochastic data conversion. We have generalized previously proposed designs and used a novel statistical inference approach to develop and analyze a mathematical model of data conversion. Our analysis provides new asymptotic performance bounds and design criteria that will be useful to system designers. Some analysis remains to be completed; in particular, we would like to model estimation performance over sequences of observations made with fixed offsets. The analysis of Section 3.4 suggests that if the fixed offsets were perfectly estimated, then performance is a constant factor worse than that of a deterministic quantizer. It is an open research problem to model the performance of the architecture with incomplete knowledge of the true levels.

The results presented here are based on analysis and high-level simulations. To build practical systems, the models should be verified by experimental results. The next stage in research is to build and analyze prototype devices and then integrate them into complete systems. Because most decision-making systems are designed to use hard quantized outputs, existing algorithms must be modified to take advantage of the soft output statistics of the stochastic converter. By designing mixed-signal interfaces from the system level and exploiting the natural variation in device characteristics, we can use emerging device technologies to build reliable systems from unreliable components.

REFERENCES

[1] P. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE J. Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, 2005.

[2] R. W. Keyes, "The effect of randomness in the distribution of impurity atoms on FET thresholds," *Appl. Phys.*, vol. 8, no. 3, pp. 251–259, Nov. 1975.

[3] B. Razavi and B. Wooley, "Design techniques for high-speed, high-resolution comparators," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1916–1926, Dec. 1992.

[4] C. Donovan and M. Flynn, "A 'Digital' 6-bit ADC in 0.25 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 432–437, May 2002.

[5] S. Park, Y. Palaskas, and M. Flynn, "A 4-GS/s 4-bit flash ADC in 0.18-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 1865–1872, 2007.

[6] C. Paulus, H.-m. Bluthgen, M. Low, E. Sicheneder, N. Briils, A. Courtois, M. Tiebout, and R. Thewes, "A 4GS/s 6b flash ADC in 0.13 μm CMOS," *VLSI Circuits*, pp. 420–423, 2004.

[7] D. Daly and A. Chandrakasan, "A 6-bit, 0.2 V to 0.9 V highly digital flash ADC with comparator redundancy," *IEEE J. Solid-State Circuits*, vol. 44, no. 11, pp. 3030–3038, 2009.

[8] T. Sundstrom and A. Alvandpour, "Utilizing process variations for reference generation in a flash ADC," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 56, no. 5, pp. 364–368, 2009.

[9] S. Weaver, B. Hershberg, P. Kurahashi, D. Knierim, and U.-k. Moon, "Stochastic Flash Analog-to-Digital Conversion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 11, pp. 2825–2833, 2010.

[10] H. Papadopoulos, G. W. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.

[11] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[12] H. Poor, *An Introduction to Signal Detection and Estimation*. Springer, 1994.

[13] E. Lehmann and G. Casella, *Theory of Point Estimation*. Springer, 1998.

[14] B. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer, 2008.

[15] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Springer, 1992.