J Sign Process Syst
DOI 10.1007/s11265-015-1020-6

A Multistage Architecture for Statistical Inference with Stochastic Signal Acquisition

Ryan M. Corey[1] · **Andrew C. Singer**[1]

Received: 31 January 2015 / Accepted: 23 June 2015

Abstract We describe a statistical inference approach for designing signal acquisition interfaces and inference systems with stochastic devices. A signal is observed by an array of binary comparison sensors, such as highly scaled comparators in an analog-to-digital converter, that exhibit random offsets in their reference levels due to process variations or other uncertainties. These offsets can limit the performance of conventional measurement devices. In our approach, we build redundancy into the sensor array and use statistical estimation techniques to account for uncertainty in the observations and produce a more reliable estimate of the acquired signal. We develop an observational model and find a Cramér-Rao lower bound on the achievable square error performance of such a system. We then propose a two-stage inference architecture that uses a coarse estimate to select a subset of the sensor outputs for further processing, reducing the overall complexity of the system while achieving near-optimal performance. The performance of the architecture is demonstrated using a simulated prototype for parameter estimation and symbol detection applications. The results suggest the feasibility of using unreliable components to build reliable signal acquisition and inference systems.

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.

✉ Ryan M. Corey
corey1@illinois.edu

[1] University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Keywords Statistical inference · Parameter estimation · Stochastic circuits · Quantization

1 Introduction

This paper addresses the problem of statistical decision making using sensors that exhibit uncertainty in their behavior. In particular, we are concerned with a sensor array that measures an unknown variable by comparing it to a number of reference levels, such as a flash analog-to-digital converter (ADC) that measures a voltage using a series of electronic comparator circuits. The sensors do not have fixed known reference levels; instead, the levels vary randomly around a nominal reference level. Such sensors may have advantages in size, power, speed, or cost compared to more reliable sensors, but cannot be used in conventional deterministic architectures. We propose an alternative architecture that leverages redundancy and statistical inference techniques to build reliable inference systems from such unreliable components.

An important motivation of this work is electronic mixed-signal interfaces built with emerging circuit technologies that exhibit larger behavioral uncertainty than conventional technologies. High-speed analog-to-digital converter circuits, for example, are often limited in speed and power by large comparator circuits. Smaller circuits generally operate faster and consume less power, but are more sensitive to process variations. Small comparators may exhibit uncertain offsets to their reference levels, in part, because of mismatch between transistor threshold voltages [1]. These voltage variations, which can be caused by dopant fluctuations during production, are generally modeled as having a normal distribution with variance inversely proportional to the gate area [2]. If the random offsets are large compared

Published online: 09 July 2015

to the spacing between nominal reference levels, they can cause errors in the output of a conventional converter.

There have been a number of proposed designs for ADCs using comparator circuits with large offsets. Since the offsets are generally static with time, they can be corrected using analog offset cancellation [3] or digital techniques, such as disabling comparators with large offsets [4] or using extra logic for error correction [5]. These designs all correct or compensate for the offsets so that the converter behaves more like a conventional flash ADC. Other recent work has directly leveraged the offset statistics rather than compensating for them. In [6], many comparators were produced with intentionally large variations and only a few were chosen to create a reference ladder. The comparators in [7] were also designed to be highly variable, but the true levels were not measured; instead, the outputs were averaged to produce an estimate based on the offset statistics.

Although this problem has been addressed primarily in the circuits literature, it can be framed as a more general statistical inference problem: there is an unknown parameter (the input voltage) that is observed indirectly using a set of uncertain observations (sensor outputs) whose statistics depend on that parameter. We can therefore apply tools from parameter estimation and machine learning to design a robust inference architecture that incorporates and even exploits device-level uncertainty. The results presented here do not apply only to quantization circuits, but to any problem in which a parameter is observed via uncertain binary thresholds. Indeed, the mathematical model developed in Section 2 is equivalent to a distributed estimation system subject to bandwidth constraints [8, 9]. The results in [9] for parameter estimation over a bandlimited sensor network parallel the results presented here, including the form of the Cramér-Rao lower bound. A key advantage of a statistical inference approach is that it can be applied at the system level: if the sensor array is to be used as a front-end to a decision-making system, such as a classifier, the array outputs can be used directly for inference; there is no need to design a separate quantizer and classifier.

In this paper, which expands on results first presented in [10], we consider a general architecture with an abitrary number of nominal levels and of sensors per level, generalizing both the conventional flash ADC (many levels with one sensor per level) and the design of [7] (few levels with many sensors each). We develop an observational model that relates the array outputs to the input parameter and the offset statistics, then use this model to predict the achievable performance of an optimal parameter estimator. Because one motivation of using unreliable sensors is to improve speed and efficiency, it may not be practical to use such an optimal estimator; fortunately, the statistical model suggests an effective strategy for approximate inference using a multistage design. We will show how the sensor outputs can be reduced to a vector of sufficient statistics and then further reduced to a scalar statistic with little loss of information. This design takes advantage of the device statistics with relatively small computational overhead.

2 System Model

The sensor array architecture is shown in Fig. 1. The array observes an unknown real-valued variable x using a number of binary sensors. There are r sensors at each of n nominal reference levels $v_1, \ldots, v_n$. Each sensor has a random reference level $V_{i,j}$ that is offset from its nominal reference level v_i for $i = 1, \ldots, n$ and $j = 1, \ldots, r$. If $r = 1$, then the architecture is a conventional reference ladder with one sensor assigned to each level. If $n = 1$, then all of the sensors are identical, as in [7]. The choice of n and r enables a tradeoff between performance and complexity.

The offsets are assumed to be independent and identically distributed with a known cumulative distribution function (CDF) F_V. This assumption is reasonable as long as the spatial correlation in offsets and the independent noise on each sensor are small compared to the offsets. For comparator circuits, the offsets are often assumed to be normally distributed and the numerical results, figures, and simulations in this paper will use a normal distribution; however, the analytical results are derived for arbitrary offset distributions.

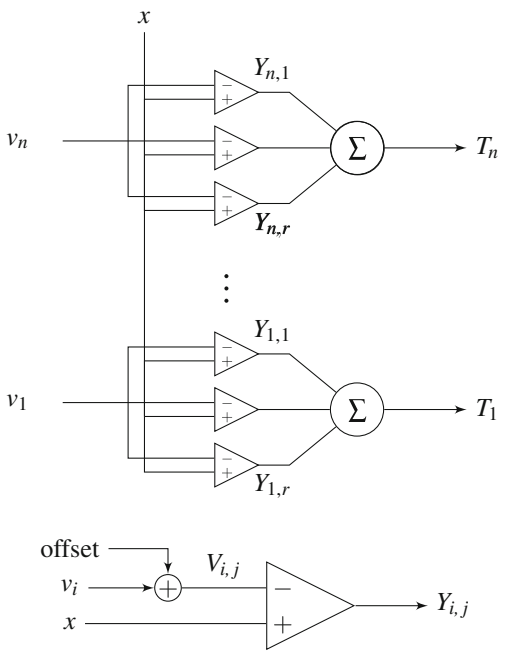

Figure 1 The sensor array (*top*) uses r binary comparison sensors at each of n nominal reference levels. Each sensor (*bottom*) has a random offset to its reference level. The sum of outputs at each nominal level is a sufficient statistic for x.

Each sensor has a binary output $Y_{i,j} = 1$ if $x \geq V_{i,j}$ and 0 otherwise. The set of $Y_{i,j}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, r$ is the observed data to be used for inference about x. Each $Y_{i,j}$ has a Bernoulli probability mass function (PMF):

$$\begin{aligned} p_{Y_{i,j}|X}(1 \mid x) &= \Pr\left\{x \geq V_{i,j}\right\} && (1)\\ &= \Pr\{x \geq v_i + V\} && (2)\\ &= \Pr\{V \leq x - v_i\} && (3)\\ &= F_V(x - v_i)\ . && (4) \end{aligned}$$

For brevity, define $F_i(x) = F_V(x - v_i)$. Similarly, $\bar{F}_i(x) = p_{Y_{i,j}|X}(0 \mid x) = 1 - F_V(x - v_i)$. When F_i is differentiable at x, denote the probability density function (PDF) by $f_i(x) = \frac{\partial}{\partial x} F_i(x)$. These nr binary observations and the corresponding PMFs can be used to estimate x. However, for high-resolution systems with many sensors, it may be impractical to use the full set of nr-dimensional data. Fortunately, the dimensionality of the observation can be significantly reduced by making some reasonable assumptions about the system.

Because the sensors all have independent and identically distributed random offsets, the observations can be reduced to an n-dimensional sufficient statistic $\mathbf{T}$ that fully preserves the information about x. Let $T_i = \sum_{j=1}^{r} Y_{i,j}$ for $i = 1, \ldots, n$. As the sum of independent Bernoulli variables, each T_i has a binomial conditional PMF:

$$p_{T_i|X}(t_i \mid x) = \binom{r}{t_i} F_i(x)^{t_i}\, \bar{F}_i(x)^{r-t_i}\ . \tag{5}$$

The conditional mean of T_i is given by

$$\mathbb{E}_x[T_i] = r F_i(x)\ , \tag{6}$$

where $\mathbb{E}_x[\cdot]$ denotes the conditional expectation given x, and its conditional variance is

$$\mathrm{Var}_x(T_i) = r F_i(x)\, \bar{F}_i(x)\ . \tag{7}$$

Because each T_i is conditionally independent given x, the PMF of the total observation is the product of these PMFs for $i = 1, \ldots, n$. If the offset density has finite support, as in Fig. 2, then it is possible that some or many of the T_i are deterministic; in particular, if $F_i(x) = 1$ then $T_i = r$ and if $F_i(x) = 0$ then $T_i = 0$. These components provide information by restricting the range of x for which the observation has nonzero probability, but do not otherwise contribute to the conditional PMF. For the remaining components whose level densities have support at x, and

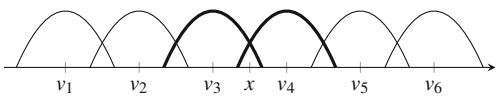

Figure 2 If the offset distributions have finite support, then only levels with support at x (*bold curves*) contribute to the PMF; the other observations are deterministic given x.

therefore satisfy $\bar{F}_i(x) \neq 0$, the PMF can be expressed in exponential form as

$$\begin{aligned} p_{\mathbf{T}|X}(\mathbf{t} \mid x) &= \prod_{i:\bar{F}_i(x)\neq 0} \binom{r}{t_i} \bar{F}_i(x)^r \left(\frac{F_i(x)}{\bar{F}_i(x)}\right)^{t_i} && (8)\\ &= h(\mathbf{t}) \exp\left\{\sum_{i:\bar{F}_i(x)\neq 0} t_i \eta_i(x) - A_i(x)\right\}\ , && (9) \end{aligned}$$

where $h(\mathbf{t}) = \prod_{i=1}^{n} \binom{r}{t_i}$, $A_i(x) = -r \ln \bar{F}_i(x)$, and

$$\eta_i(x) = \ln F_i(x) / \bar{F}_i(x)\ . \tag{10}$$

If F_i is differentiable at x and $f_i(x) > 0$ for all i included in the sum, then (9) forms a curved exponential family [11] with natural parameter $\boldsymbol{\eta}(x)$.

Although $\mathbf{T}$ is a sufficient statistic, it may not be a complete sufficient statistic; that is, there may exist a lower-dimensional statistic that also preserves the observed information about x. This is the case, in particular, when each $\eta_i(x)$ is a linear function of x. If $\eta_i(x) = a_i x + b_i$, then the PMF can be expressed

$$\begin{aligned} p_{\mathbf{T}|X}(\mathbf{t} \mid x) &= h(\mathbf{t}) \exp\left\{\sum_{i=1}^{n} a_i x t_i + b_i t_i - A_i(x)\right\} && (11)\\ &= h^*(\mathbf{t}) \exp\left\{\left(\sum_{i=1}^{n} a_i t_i\right) x - A_i(x)\right\}\ , && (12) \end{aligned}$$

where the terms containing b_i have been absorbed into $h^*(\mathbf{t})$. By the completeness theorem for exponential families [12], the sum $S = \sum_{i=1}^{n} a_i T_i$ is a complete sufficient statistic. The natural parameter has a linear form if the random offsets have a logistic distribution; in that case, the sum of the outputs of all the sensors is a complete sufficient statistic and the observations can be represented by this scalar with no loss of information about x. In Section 5, we will use this special case to find approximate estimators for more general distributions.

3 Statistical Inference

In many inference applications, the system is designed to infer the value of x using the observations from the sensor array. There are several metrics that can be used to characterize the performance of the inference system. If x is known to take values in a discrete set, such as a finite alphabet of symbols, then we are often concerned with the probability of detection error. If x is drawn from a continuous set, then we often evaluate the performance of an estimator $\hat{x}(\mathbf{T})$ at a particular value of x by the bias, $\mathbb{E}_x\left[\hat{x}(\mathbf{T}) - x\right]$, and variance, $\mathrm{Var}_x\left(\hat{x}(\mathbf{T})\right)$, of the estimator.

In both cases, a reasonable decision-making strategy is the maximum likelihood rule, which selects the value of x

that maximizes the conditional probability of the observation given x:

$$\hat{x}_{\mathrm{ML}}(\mathbf{T}) = \arg\max_x \ln p_{\mathbf{T}|X}(\mathbf{T} \mid x) \tag{13}$$

$$= \arg\max_x \sum_{i:\bar{F}_i(x)\neq 0} \eta_i(x)\, T_i - A_i(x) \tag{14}$$

If $\eta_i(x)$ were linear for all i, then the maximum could be expressed in terms of the scalar sufficient statistic:

$$\hat{x}_{\mathrm{ML}}(\mathbf{T}) = \arg\max_x \sum_{i=1}^{n} a_i x T_i + b_i T_i - A_i(x) \tag{15}$$

$$= \arg\max_x Sx - \sum_{i=1}^{n} A_i(x) \tag{16}$$

3.1 Maximum Likelihood Parameter Estimation

First consider the problem of estimating x over a continuous set using the maximum likelihood estimator (MLE). Assume that x is in the support of at least one level density, i.e., $f_i(x) > 0$ for at least one $i \in \{1, \ldots, n\}$, so that the parameter is identifiable and the log-likelihood is differentiable. If the log-likelihood function (14) is strictly concave in x, its maximum is the solution to the likelihood equation

$$0 = \frac{\partial}{\partial x} \ln p_{\mathbf{T}|X}(\mathbf{T} \mid x) \tag{17}$$

$$= \sum_{i:f_i(x)>0} \frac{\partial}{\partial x}\left(\eta_i(x)\, T_i - A_i(x)\right) \tag{18}$$

$$= \sum_{i:f_i(x)>0} \eta_i'(x)\left(T_i - r F_i(x)\right) . \tag{19}$$

Substituting the mean of the binomial distribution (6), the likelihood equation can also be expressed

$$\sum_{i:f_i(x)>0} \eta_i'(x)\, T_i = \sum_{i:f_i(x)>0} \eta_i'(x)\, \mathbb{E}_x[T_i] . \tag{20}$$

In the special case where each $\eta_i(x)$ is linear, (20) reduces to $S(\mathbf{T}) = \mathbb{E}_x[S(\mathbf{T})]$.

3.2 Achievable Performance

We can use tools from information theory to bound the achievable performance of the estimator. An estimator is called unbiased if $\mathbb{E}_x\left[\hat{x}(\mathbf{T})\right] = x$. For any unbiased estimator $\hat{x}_{\mathrm{UE}}$, the Cramér-Rao Lower Bound (CRLB) on the conditional variance is

$$\mathrm{Var}_x\left(\hat{x}_{\mathrm{UE}}(\mathbf{T})\right) \geq I(x)^{-1} , \tag{21}$$

where $I(x)$ is the Fisher information provided by $\mathbf{T}$ about x [12]. The Fisher information is well defined for the reference levels for which $F_i(x)$ is differentiable at x and $f_i(x) > 0$. It is given by

$$I(x) = -\mathbb{E}_x\left[\left(\frac{\partial^2}{\partial x^2} \ln p_{\mathbf{T}|X}(\mathbf{T} \mid x)\right)\right] \tag{22}$$

$$= -\sum_{i:f_i(x)>0} \mathbb{E}_x\left[\frac{\partial^2}{\partial x^2} \ln p_{T_i|X}(T_i \mid x)\right] \tag{23}$$

$$= \sum_{i:f_i(x)>0} \mathbb{E}_x\left[\left(\frac{\partial}{\partial x}\left(\eta_i(x)\, T_i - A_i(x)\right)\right)^2\right] \tag{24}$$

$$= \sum_{i:f_i(x)>0} r\frac{f_i(x)^2}{F_i(x)\,\bar{F}_i(x)} \tag{25}$$

$$= r \sum_{i:f_i(x)>0} \eta_i'(x)\, f_i(x) . \tag{26}$$

Figure 3 shows the Fisher information contribution of a single sensor as a function of the distance between the nominal level v_i and the parameter x for normally distributed offsets with mean zero and variance σ^2. The sensors provide the most Fisher information about signals near their nominal reference levels and little information about signals far from their levels.

To find a simple, approximate expression for the CRLB of a high-resolution sensor array, notice that Eq. 26 has the form of a probability-weighted mean of $\eta_i'(x)$ which resembles an expectation. Suppose that the nominal reference levels are densely spaced a uniform distance $\Delta v \ll \sigma$ apart and that x is far from the smallest and largest reference levels. Let V be a random variable distributed according to the offset PDF f_V. The total Fisher information (26) can be approximated as

$$I(x) = \frac{r}{\Delta v} \sum_i \eta'(x - v_i)\, f(x - v_i)\, \Delta v \tag{27}$$

$$\approx \frac{r}{\Delta v} \sum_i \eta'(x - v_i) \Pr\{V \in (x - v_i, x - v_{i+1})\} \tag{28}$$

$$\approx \frac{r}{\Delta v} \mathbb{E}\left[\eta'(V)\right] . \tag{29}$$

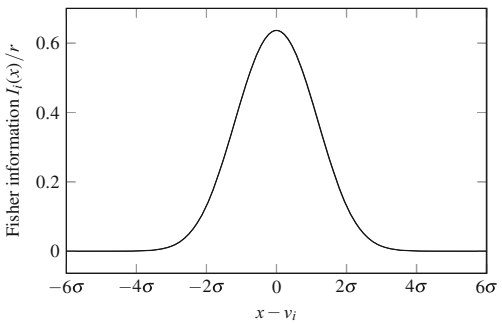

Figure 3 Fisher information contributed by observation T_i with normally distributed offsets. The information is largest when $x = v_i$.

For the normal distribution with zero mean and variance σ^2, $\mathbb{E}\left[\eta'(V)\right] \approx 1.806/\sigma$ by numerical integration. For a logistic distribution with variance σ^2, it is $\pi/(\sqrt{3}\sigma)$ in closed form. There are some densities, such as the uniform density, for which the Fisher information curve does not have finite area and therefore this approximation is not valid. Figure 4 shows the exact Fisher information as a function of x and Δv for a sample sensor array with normally distributed offsets. The approximation (29) is close when $\Delta v \leq \sigma$. Thus, the high-resolution CRLB appears to be proportional to $\sigma \Delta v / r$, the ratio of the standard deviation of the offsets to the average density of sensor levels.

4 A Multistage Architecture

Solving the likelihood equation exactly requires the full observation vector $\mathbf{T}$; yet Fig. 3 shows that only a subset of the sensor measurements contribute significant information about x. The calculation could be simplified significantly if the low-information observations were removed. Therefore, we propose the two-step estimation architecture shown in Fig. 5. A simple coarse estimator makes a rough estimate of x and uses it to select a subset of the observations. That lower-dimensional observation vector is then passed to a more complex inference function that makes the final decision or estimate.

4.1 Subset Selection

Denote the selected subset by $\mathbf{T}^*$. Its length $n^* \leq n$ is a design choice that will be discussed later. For simplicity, we will assume that n^* is fixed, though in implementation it may be smaller for x values near the boundaries of the level

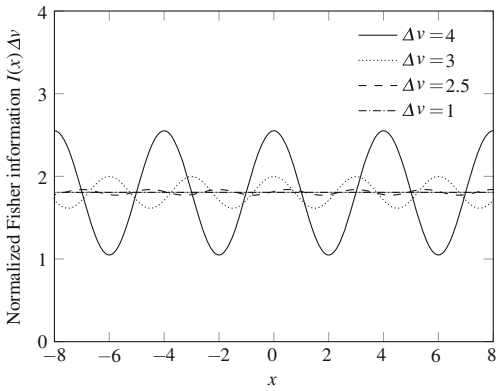

Figure 4 Fisher information $I(x)$ as a function of x for an array of sensors with nominal levels uniformly spaced Δv apart and normally distributed offsets with unit variance. The peaks of each curve correspond to the nominal levels. As the spacing grows smaller, the Fisher information approaches the high-resolution approximation (29) shown by the horizontal gray line $I(x)\,\Delta v = 1.806$.

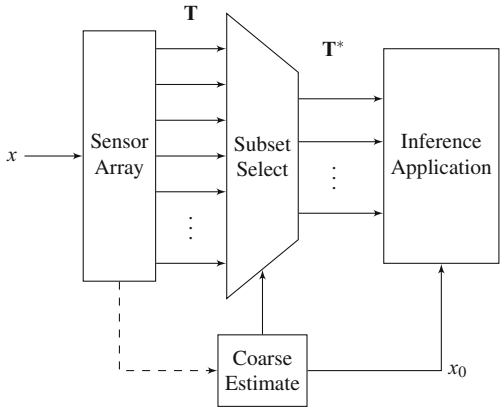

Figure 5 A multistage architecture for statistical inference problems. A coarse estimate is used to select a subset of the output statistics that contains most of the information about the parameter.

range. For the remainder of this paper, we relabel F_k, f_k, A_k and η_k to correspond to the elements T_k of $\mathbf{T}^*$ for $k = 1, \ldots, n^*$. We assume that n^* is small enough that $f_k(x) > 0$ for all k.

The likelihood for $\mathbf{T}^*$ has the same form as that for $\mathbf{T}$ (9):

$$\ln p_{\mathbf{T}^*|X}\left(\mathbf{t}^* \mid x\right) = h\left(\mathbf{t}^*\right) \exp\left\{\sum_{k=1}^{n^*} \eta_k(x)\, t_k - A_k(x)\right\}. \tag{30}$$

Similarly, the Fisher information is

$$I^*(x) = r \sum_{k=1}^{n^*} \eta_k'(x)\, f_k(x)\,. \tag{31}$$

The first stage need not have high resolution, so there are many possible solutions for selecting the subset. One convenient approach is to assign each subset a corresponding coarse estimate and choose one such estimate based on a few of the observations. For example, suppose that the coarse estimator has access to a single sensor output (say, $Y_{i,1}$) from each nominal reference level and that n^* is even. Let $Q = \sum_{i=1}^{n} Y_{i,1}$ be the sum of these outputs. A simple subset selection rule chooses $\mathbf{T}^* = \left\{T_{Q-n^*/2+1}, \ldots, T_{Q+n^*/2}\right\}$ and corresponding coarse estimate $x_0 = v_1 + \Delta v\left(Q - \frac{1}{2}\right)$ for $\frac{n^*}{2} \leq Q \leq n - \frac{n^*}{2}$. This rule is used in the simulations in Section 6. The subset selection stage can be made more or less accurate based on the requirements of the inference stage.

4.2 Linear Model

When the subset is restricted to a sufficiently small range of x values, the nonlinear function $\eta_k(x)$ can be well approximated by a linear function with slope $\eta_k'(x_0)$, where x_0 is the

coarse estimate from the first stage of the estimator. Then the log-likelihood function is approximately

$$\ln p_{\mathbf{T}^*|X}\left(\mathbf{t}^* \mid x\right) \approx h^*\left(\mathbf{t}^*\right) \exp\left\{\sum_{k=1}^{n^*} \eta_k'\left(x_0\right) x t_k - A_k\left(x\right)\right\} \tag{32}$$

where $h^*\left(\mathbf{t}^*\right)$ includes all terms that depend only on $\mathbf{t}^*$ and not x. In that case, the n^*-dimensional vector $\mathbf{T}^*$ can be reduced to the scalar statistic

$$S^*\left(\mathbf{T}^*, x_0\right) = \sum_{k=1}^{n^*} \eta_k'\left(x_0\right) T_k^*\,, \tag{33}$$

which has conditional mean

$$\mathbb{E}_x\left[S^*\left(\mathbf{T}^*, x_0\right)\right] = \sum_{k=1}^{n^*} \eta_k'\left(x_0\right) r F_k\left(x\right) \tag{34}$$

and conditional variance

$$\mathrm{Var}_x\left(S^*\left(\mathbf{T}^*, x_0\right)\right) = \sum_{k=1}^{n^*} \eta_k'\left(x_0\right)^2 r F_k\left(x\right) \bar{F}_k\left(x\right)\,. \tag{35}$$

By the same reasoning from Section 2, if $\eta_k\left(x\right)$ were truly linear, then S^* would be a complete sufficient statistic for $\mathbf{T}^*$. Because $\eta_k\left(x\right)$ is generally a nonlinear function, there is some information loss.

5 Approximate Inference

5.1 Local Linear Estimator

To estimate x using minimal computation, we would like to find a linear estimator of the form $\hat{x}\left(\mathbf{T}^*\right) = w_0 + \sum_{k=1}^{n^*} w_k T_k$, as shown in Fig. 6. Using the linear approximation for $\eta_k\left(x\right)$, the likelihood (19) becomes

$$0 \approx \sum_{k=1}^{n^*} \eta_k'\left(x_0\right)\left(T_k - r F_k\left(x\right)\right) \tag{36}$$

$$= S^* - \sum_{k=1}^{n^*} \eta_k'\left(x_0\right) r F_k\left(x\right)\,. \tag{37}$$

This equation is still nonlinear. Let us make the additional linear approximation $F_k\left(x\right) \approx F_k\left(x_0\right) + f_k\left(x_0\right)\left(x - x_0\right)$. Substituting this approximation into Eq. 37 gives

$$0 \approx S^* - \sum_{k=1}^{n} \eta_k'\left(x_0\right) r\left(F_k\left(x_0\right) + f_k\left(x_0\right)\left(x - x_0\right)\right)\,. \tag{38}$$

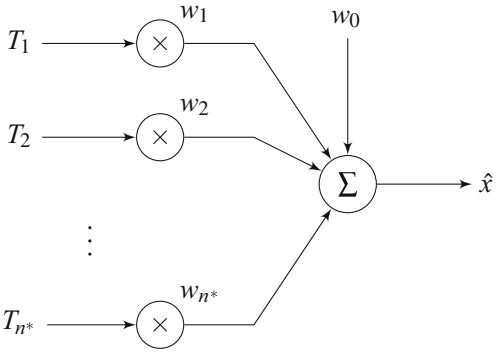

Figure 6 Under the linear approximation of the log-likelihood of a subset of observations $\mathbf{T}^*$, we can derive a linear estimator for x.

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

$$\hat{x}_{\mathrm{LL}}\left(\mathbf{T}^*, x_0\right) = x_0 + \frac{S^* - \sum_{k=1}^{n^*} \eta_k'\left(x_0\right) r F_k\left(x_0\right)}{\sum_{k=1}^{n^*} \eta_k'\left(x_0\right) r f_k\left(x_0\right)} \tag{39}$$

$$= x_0 + \frac{S^* - \mathbb{E}_{x_0}\left[S^*\right]}{I^*\left(x_0\right)} \tag{40}$$

$$= x_0 - \frac{\mathbb{E}_{x_0}\left[S^*\right]}{I^*\left(x_0\right)} + \sum_{k=1}^{n^*} \frac{\eta_k'\left(x_0\right)}{I^*\left(x_0\right)} T_k\,. \tag{41}$$

If the reference levels are uniformly spaced and x_0 is fixed for each subset, such as to the centers of the intervals of nominal reference levels, then the linear estimator weights are the same for each subset.

The performance of the local linear estimator depends on the accuracy of the initial estimate x_0. For a given x_0 and x, the mean of the estimate is

$$\mathbb{E}_x\left[\hat{x}_{\mathrm{LL}}\left(\mathbf{T}^*, x_0\right)\right] = x_0 + \sum_{k=1}^{n^*} \frac{r \eta_k'\left(x_0\right)}{I^*\left(x_0\right)}\left(F_k\left(x\right) - F_k\left(x_0\right)\right) \tag{42}$$

and the variance is

$$\mathrm{Var}_x\left(\hat{x}_{\mathrm{LL}}\left(\mathbf{T}^*, x_0\right)\right) = \sum_{k=1}^{n^*} \frac{\left(\eta_k'\left(x_0\right)\right)^2}{I^*\left(x_0\right)^2} r F_k\left(x\right) \bar{F}_k\left(x\right) \tag{43}$$

$$= \sum_{k=1}^{n^*} \frac{I_k\left(x_0\right)}{I^*\left(x_0\right)^2} \frac{F_k\left(x\right) \bar{F}_k\left(x\right)}{F_k\left(x_0\right) \bar{F}_k\left(x_0\right)}\,. \tag{44}$$

If $x = x_0$, then $\mathbb{E}_x\left(\hat{x}_{\mathrm{LL}}\right) = x$ and $\mathrm{Var}_x\left(\hat{x}_{\mathrm{LL}}\right) = I^*\left(x\right)^{-1}$, so the estimate is unbiased and achieves the CRLB with equality at those points. Otherwise, performance degrades as x_0 grows farther from x. Figure 7 shows the conditional mean and variance of $\hat{x}_{\mathrm{LL}}$ as functions of $x - x_0$ for normally distributed offsets. To achieve good performance, the coarse estimate should be well within one standard deviation of the true parameter.

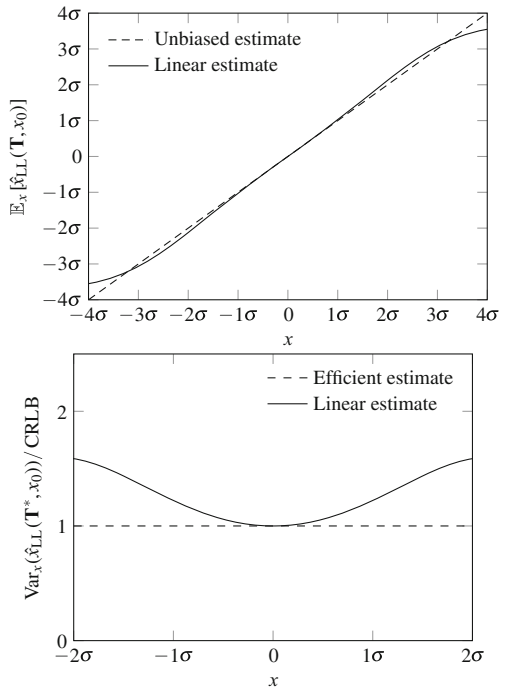

Figure 7 The conditional mean (*top*) and variance (*bottom*) of the linear estimate as a function of x for $x_0 = 0$. The estimator is unbiased and efficient when $x = x_0$.

5.2 Inference with the Scalar Statistic

The previous section describes an architecture for estimating the value of the input signal, much like a conventional analog-to-digital converter. For many inference problems, such as symbol detection in a communication receiver, the value of the input is important only to the extent that it contains information about another variable, in that case the transmitted symbol. For such applications, the observation statistic can be considered a soft output of the sensor array to be used directly for the inference problem; the estimated input is not necessary.

Consider a classification problem in which the input is a random variable X drawn from one of m distributions, $f_{X,1}(x), \ldots, f_{X,m}(x)$, where each distribution corresponds to a class $\theta \in \{1, \ldots, m\}$. The goal is to find the class based on the observations from the sensor array. The maximum likelihood classifier is

$$\hat{\theta}_{\mathrm{ML}}(\mathbf{T}) = \arg \max_{1 \le \theta \le m} p_{\mathbf{T}|\Theta}(\mathbf{T} \mid \theta) \tag{45}$$

$$= \arg \max_{1 \le \theta \le m} \int p_{\mathbf{T}|X}(\mathbf{T} \mid x) f_{X,\theta}(x)\, \mathrm{d}x \tag{46}$$

where the integral is over the support of $f_{X,\theta}$. Using the subset of output statistics and assuming that $f_k(x) > 0$ for all k and all x in the support of $f_{X,\theta}$, the classifier is

$$\hat{\theta}_{\mathrm{ML}}(\mathbf{T}^*) = \arg \max_{1 \le \theta \le m} \int \exp\left\{ \sum_{k=1}^{n^*} (\eta_k(x)\, T_k - A_k(x)) \right\} f_{X,\theta}(x)\, \mathrm{d}x\,. \tag{47}$$

Finally, using the linear approximation from Eq. 32 gives the approximate classifier

$$\hat{\theta}(S^*) = \arg \max_{1 \le \theta \le m} \int \exp\left\{ S^* x - \sum_{k=1}^{n^*} A_k(x) \right\} f_{X,\theta}(x)\, \mathrm{d}x\,. \tag{48}$$

Thus, the scalar output statistic S^* can be used for a classifier that approximates the maximum likelihood classifier. The performance of such a classifier will be evaluated in Section 6.2 for a symbol detection problem.

6 Performance

The performance of the proposed inference architecture is demonstrated using two simulated systems: a parameter estimator and a symbol detector. In both cases, the sensor array levels are uniformly spaced over an interval slightly wider than the range of the input signal with normally distributed offsets. The coarse estimator is that described in Section 4.1. The subset size is given in terms of the interval length instead of the number of elements; a length-2σ subset includes the outputs of all sensors with nominal levels within $\pm\sigma$ of x_0, for example.

6.1 Parameter Estimation

First we consider the analog-to-digital conversion problem of estimating a voltage signal at the input to an array of comparators. The nominal levels are uniformly spaced between 0 V and 1 V. The parameter is estimated using a full maximum likelihood estimator and a two-stage linear estimator with varying subset size. The Monte Carlo simulations used 256 input signals between 100 and 900 mV and 1024 sets of random levels for each data point.

Figure 8 shows the mean square error (MSE) performance of the estimators as a function of the number of nominal levels n for $r = 16$ and $\sigma = 20$ mV. The dashed line shows the high-resolution CRLB $\sigma \Delta v / 1.806r$, where $\Delta v = (1\,\mathrm{V})/(n-1)$. The MLE and linear estimators all have MSE performance that decreases asymptotically linearly with n, as expected. The MLE is close to the high-resolution CRLB for $n > 10$, which corresponds to $\Delta v < 5\sigma$ in this case. The two-stage estimator performance

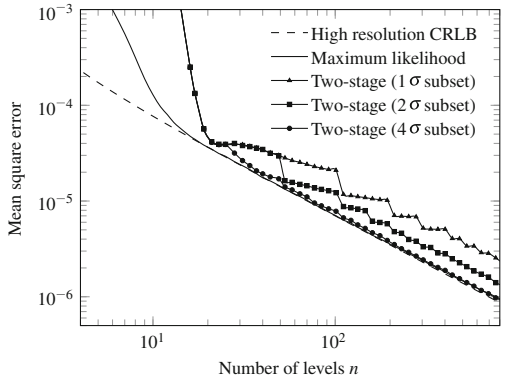

Figure 8 Simulated mean square error performance of the parameter estimator as a function of n for subset interval widths 1σ, 2σ, and 4σ. Here $r = 16$ and $\sigma = 20$ mV.

depends on the size of the subset, with larger subsets performing better. For example, with $n = 401$ ($\Delta v = \sigma/8$), the 1σ subset estimator had MSE roughly three times that of the MLE while using 2 % of the sensor outputs, while the 4σ subset estimator had MSE about 4 % higher using 8 % of the sensors. There is little benefit to using much larger subsets because those observations with nominal levels far from the parameter contribute negligible information.

Figure 9 shows the MSE as a function of σ^2 for level spacing $\Delta v = 5$ mV and varying values of r. Both the MLE and the two-stage estimator achieve performance close to the high-resolution CRLB when σ is comparable to or larger than the spacing between levels but small compared to the input range. In this range, the MSE scales roughly linearly with σ and inversely with r, as expected. Notice that MSE

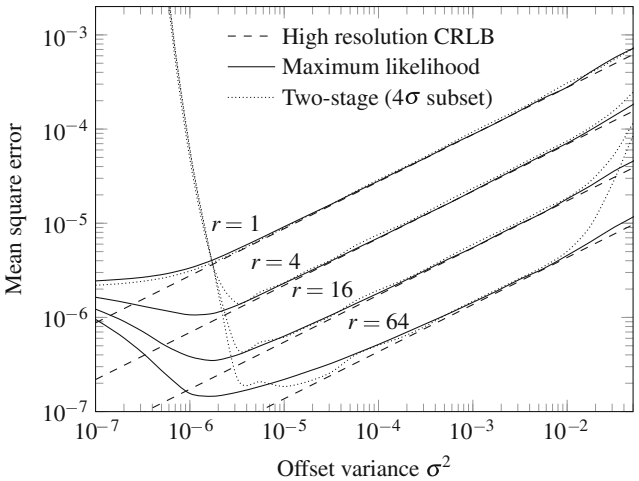

Figure 9 Simulated mean square error performance of the parameter estimator as a function of σ^2 for $\Delta v = 5$ mV and $r = 1, 4, 16$, and 64.

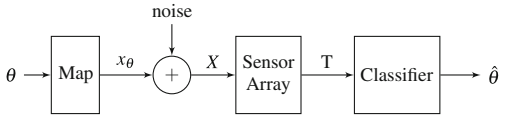

Figure 10 The sensor array is used to detect symbols transmitted through a noisy channel.

performance is worse for $\sigma \ll \Delta v$ when $r > 1$, as the sensors within each nominal level are tightly clustered and do not cover the input range.

6.2 Symbol Detection

Next, we demonstrate the performance of the array for a statistical inference problem, detecting symbols in noise. Figure 10 shows the simulated system: the input is chosen from an alphabet of $m = 16$ real-valued symbols spaced $d = 0.5$ apart and corrupted by normally distributed noise with mean zero and variance ρ^2. The noisy symbol is the input to the sensor array with $\Delta v = 0.05$ and offset variance σ^2. The classifier recovers the transmitted symbol using either a full maximum likelihood calculation (46) or the approximate decision rule (48) with subset size 4σ. If the classifier were able to measure x perfectly and the symbols were transmitted with equal frequency, then the probability of error would be

$$\mathrm{P}_{\mathrm{err,ideal}} = \frac{2m-2}{m} Q\left(\frac{d}{2\rho}\right), \tag{49}$$

where $Q(\cdot)$ is the complementary normal CDF.

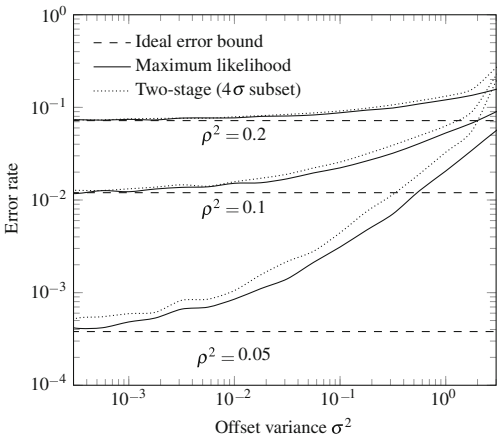

Figure 11 Simulated error probability for detecting noisy symbols with the stochastic sensor array. The dashed lines show the ideal error bounds (49).

The Monte Carlo simulations used 1024 sets of random levels and a total of 2^{17} random noise samples for each data point. The error probability is averaged over each of the 16 symbols. Figure 11 shows the simulated error rate versus offset variance. The dashed line shows the error rate with perfect measurement (49) for each value of the noise power. Both classifiers achieve error rates close to this ideal error rate for small σ^2. The performance of the approximate classifier is close to that of the full ML classifier as long as σ is not much larger than d. The scalar output statistic is therefore a reasonable alternative to the full set of observations for classification.

7 Conclusions

The simulation results suggest that the stochastic signal acquisition architecture performs close to the CRLB for well-chosen design parameters. The performance of the sensor array depends on the Fisher information, which is proportional to $r/\sigma\,\Delta v$ when Δv is comparable to or smaller than σ. If the statistics of the sensor reference levels are known, then the system designer can use the analysis presented here to predict the performance of the sensor array. To design a high-resolution signal acquisition system using sensors with a fixed offset variation, a designer could place the nominal levels with a uniform spacing $\Delta v < \sigma$. To reach the desired level of performance, increase r by replicating the existing sensor array and summing the outputs. With this approach, performance is improved without increasing the dimensionality of the sufficient statistic $\mathbf{T}$, resulting in a scalable design.

The two-stage design further reduces the size of the statistic by finding a coarse estimate and taking a subset of the observations. The subset size is a design choice that trades performance with complexity and depends on the sensor variation. If the range is too narrow, then the estimator will be more sensitive to errors in the coarse estimate. If it is too wide, then the estimator may be needlessly complex. The simulation results shown in Fig. 8 suggest that the subset should cover at least a few standard deviations to achieve performance close to that of the full observation vector for parameter estimation. The output can be further reduced to a scalar using a highly parallel linear combination of the observations. The lower-dimensionality statistics produced by the two-stage design can be used directly for inference applications such as detection, estimation, and classification. These inference algorithms incorporate the randomness of the measurement into the overall observational model as another source of uncertainty.

The stochastic sensor array is useful in applications for which reliable sensors are unavailable or costly and it is preferable to use sensors with higher variability. Whereas many previously proposed designs using unreliable comparators require measurement or calibration, the architecture described here does not measure the true thresholds. As long as the variations in the sensors are independent, the system can average over their measurements to produce a more reliable decision. In contrast to conventional quantization architectures, the mean square error of the estimate from the stochastic mixed-signal interface can be much lower than the variance of the random offsets. By leveraging the natural variation in device characteristics, we can build reliable systems even with unreliable components.

Acknowledgments We wish to acknowledge our collaborators Naveen Verma and Sen Tao of Princeton University for their input on this work.

References

1. Kinget, P. (2005). Device mismatch and tradeoffs in the design of analog circuits. *IEEE Journal Solid-State Circuits*, *40*(6), 1212–1224.
2. Keyes, R.W. (1975). The effect of randomness in the distribution of impurity atoms on FET thresholds. *Applied Physics*, *8*(3), 251–259.
3. Razavi, B., & Wooley, B. (1992). Design techniques for high-speed, high-resolution comparators. *IEEE Journal Solid-State Circuits*, *27*(12), 1916–1926.
4. Donovan, C., & Flynn, M. (2002). A 'Digital' 6-bit ADC in 0.25 μm CMOS. *IEEE Journal Solid-State Circuits*, *37*(3), 432–437.
5. Paulus, C., Bluthgen, H.-m., Low, M., Sicheneder, E., Briils, N., Courtois, A., Tiebout, M., & Thewes, R. (2004). A 4GS/s 6b flash ADC in 0.13 μm CMOS. *VLSI Circuits*, 420–423.
6. Sundstrom, T., & Alvandpour, A. (2009). Utilizing process variations for reference generation in a flash ADC. *IEEE Transactions Circuits System II, Experimentalis Briefs*, *56*(5), 364–368.
7. Weaver, S., Hershberg, B., Kurahashi, P., Knierim, D., & Moon, U.-k. (2010). Stochastic Flash Analog-to-Digital Conversion. *IEEE Transactions Circuits System II, Experimentalis Briefs*, *57*(11), 2825–2833.
8. Papadopoulos, H., Wornell, G.W., & Oppenheim, A. (2001). Sequential signal encoding from noisy measurements using quantizers with dynamic bias control. *IEEE Transactions Information Theory*, *47*(3), 978–1002.
9. Ribeiro, A., & Giannakis, G. (2006). Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case. *IEEE Transactions Signal Processing*, *54*(3), 1131–1143.
10. Corey, R.M., Singer, A.C., Tao, S., & Verma, N. (2014). A low complexity estimation architecture based on noisy comparators. In *Proceedings IEEE Workshop Signal Processing Systems*.
11. Lehmann, E., & Casella, G. (1998). *Theory of Point Estimation*. New York: Springer.
12. Poor, H. (1994). *An Introduction to Signal Detection and Estimation*. New York: Springer.

Ryan M. Corey received the BSE degree in Electrical Engineering from Princeton University in 2012 and the MS degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2014. He currently holds a National Science Foundation Graduate Research Fellowship at Illinois, where he is pursuing a PhD. His research interests include signal processing, information theory, and machine learning.

Andrew C. Singer received the S.B., S.M., and Ph.D. degrees, all in electrical engineering and computer science, from the Massachusetts Institute of Technology. From 1996 to 1998, he was a Research Scientist at Sanders, A Lockheed Martin Company in Manchester, New Hampshire. Since 1998, he has been on the faculty of the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, where he currently holds a Fox Family Endowed Professorship in the Electrical and Computer Engineering department. His research interests include signal processing and communication systems. In 2000, he co-founded Intersymbol Communications, Inc., a fabless semiconductor IC company, based in Champaign, Illinois, which built the worlds fastest signalprocessing-enhanced receivers for 10Gb/s optical communications. In 2007, Intersymbol Communications, Inc. was acquired by Finisar Corporation (FNSR). He continues to work in areas related to signal processing algorithms and their potential to enhance mixed-signal analog and digital circuits both in his research and as Assistant Director of the SRC-funded Systems On Nanoscale Information fabriCs (SONIC) center. He received the National Science Foundation CAREER Award in 2000; in 2001 he received the Xerox Faculty Research Award, and in 2002 he was named a Willett Faculty Scholar. In 2005, he was appointed the Director of the Technology Entrepreneur Center at the University of Illinois, where he directs a wide range of entrepreneurship activities in the College of Engineering. In 2006 he received the IEEE Journal of Solid State Circuits Best Paper Award for the paper entitled “An MLSE Receiver for Electronic Dispersion Compensation of OC-192 Fiber Links”. In 2008, he received the IEEE Signal Processing Magazine Award for the paper entitled “Turbo Equalization”. In 2009, he was elected Fellow of the IEEE “for contributions to signal processing techniques for digital communication,” and in 2014, he was named as a Distinguished Lecturer of the IEEE Signal Processing Society.