Sequential Signal Encoding from Noisy Measurements Using Quantizers with Dynamic Bias Control

Haralabos C. Papadopoulos, Gregory W. Wornell, *Member, IEEE*, and Alan V. Oppenheim, *Fellow, IEEE*

***Abstract*—Signal estimation from a sequential encoding in the form of quantized noisy measurements is considered. As an example context, this problem arises in a number of remote sensing applications, where a central site estimates an information-bearing signal from low-bandwidth digitized information received from remote sensors, and may or may not broadcast feedback information to the sensors. We demonstrate that the use of an appropriately designed and often easily implemented additive control input before signal quantization at the sensor can significantly enhance overall system performance. In particular, we develop efficient estimators in conjunction with optimized random, deterministic, and feedback-based control inputs, resulting in a hierarchy of systems that trade performance for complexity.**

***Index Terms*—Data fusion, distributed estimation, dithering, maximum likelihood (ML), quantization, stochastic resonance, wireless networks.**

I. Introduction

In this paper, we consider a particular problem of estimating an information-bearing signal from noisy measurements where system constraints force us to rely on a quantized description of those measurements. Problems of this kind arise in the context of data fusion in a very broad and diverse range of applications, including distributed sensing for military applications [1], data-based management systems [2], target tracking and surveillance for robot navigation [3], [4], radar applications [5], and medical imaging [6].

Recently, data fusion has attracted considerable attention in the context of distributed sensing problems, due to the continuing reduction in the cost of sensors and computation, and the performance improvements that inherently emanate from the use of multiple sensors [7]. Unlike classical multisensor fusion where the data collected by the sensors are communicated in full to a central processor, it is often desirable to perform some form of decentralized processing at the sensor before communicating the acquired information to the central processor in a condensed and often lossy form.

Manuscript received November 1, 1997; revised August 1, 2000. The material in this paper was presented in part at the the International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, May 1998.

H. C. Papadopoulos is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: babis@eng.umd.edu).

G. W. Wornell and A. V. Oppenheim are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: gww@allegro.mit.edu; avo@allegro.mit.edu).

Communicated by R. Laroia, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(01)01345-1.

Various challenging signal detection and estimation problems that have surfaced in such distributed sensing applications have been addressed. For instance, it is important to determine the extent to which decentralized preprocessing limits performance and to develop low-complexity methods for performing decentralized data fusion. As Hall *et al.* [8] show in the context of decentralized estimation, depending on the particular scenario, distributed data processing may range from being optimal, in the sense that no loss in performance is incurred by simply communicating the local estimates computed at each sensor, to being catastrophic, in the sense that not sufficiently careful preprocessing at each sensor can completely eliminate the underlying structure in the joint set of sensor measurements. Similar performance characteristics are exhibited in decentralized signal detection problems [9], [10]. Although for many important cases of practical interest decentralized signal detection and estimation methods have been formed for locally optimized processing at each sensor and subsequent efficient data fusion at the host (see [8]–[18] and the references therein), a number of real-time decentralized fusion problems are still largely unexplored.

The distributed sensing problem that is considered in this work arises, for example, in the context of wireless sensor networks and involves a central site estimating a remotely acquired analog signal from an efficiently encoded digital description constructed at the remote sensors. In such a network, the local measurements made at each sensor must be communicated effectively without delay to a host over a wireless channel, where they must be effectively combined to decipher the information-bearing signal. Since bandwidth must often be shared across such a sensor network, the effective data rate at which each sensor can reliably communicate to the host over the wireless channel may be severely limited, often to a few bits of information per each acquired sensor measurement. The need for power-efficient design may also place constraints in the available processing complexity at each sensor, but usually not at the host, which typically possesses more processing power than each individual sensor. Depending upon bandwidth availability in these wireless networks, the host may or may not broadcast information back to the remote sensors, so as to improve the quality of the future sensor data it receives. Finally, as delays in transferring information across such a network may be a critical constraint, it may also be desirable to employ techniques that provide sequences of sequentially refined signal estimates at the host.

Similar problems arise in applications involving sensing devices or measurement apparatuses intrinsically limited by de-

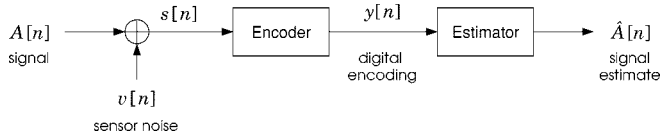

Fig. 1. Block diagram of encoding the noisy measurements at the sensor and signal estimation from these encodings at the host.

sign. In many cases, the sensing devices exhibit only a finite set of possible outputs and there is limited flexibility in terms of affecting or biasing these outputs. Networks of resolution-limited sensors are also employed by a number of biological systems for performing vital sensory tasks, suggesting that the type of processing performed by these systems somehow corresponds to an efficient use of resources [19]–[21]. For instance, it has been conjectured that certain types of crayfish enhance the ability of their crude sensory neurons to reliably detect weak signals sent by their predators by exploiting remarkably simple and, at first sight, counterintuitive preprocessing [20].

In developing methods for overcoming the power/bandwidth constraints that may arise across a sensor network, or the dynamic range and resolution constraints at each sensor, it is instructive to first examine the single-sensor problem. In fact, this special case captures many of the key design and performance issues that arise in the context of networks of sensors. The block diagram corresponding to a single sensor is shown in Fig. 1, where $A[n]$ denotes the information-bearing signal at time n, $v[n]$ represents sensor noise, $s[n]$ denotes the sensor measurement sequence, and $y[n]$ denotes the sequence of M-ary symbols encoded at the sensor and time n and used at the host to obtain a signal estimate $\hat{A}[n]$. In general, the task is to design the encoder at the sensor (subject to existing available processing power, transmit power, bandwidth, and delay constraints) and the associated estimator from the encodings at the host so as to optimize the host estimate quality.

This problem can be viewed as one of lossy encoding of a noisy source where in addition to the bandwidth/rate constraints, there exist delay and processing power constraints. Since the end objective is to obtain an accurate signal reconstruction of the information-bearing signal in the measurements, the metric we employ for evaluating the encoding performance is based on the fidelity of the underlying signal estimate resulting from the encoding, rather than on the quality of the approximate representation of the source (sensor) data obtained at the host from the encoding [22, p. 78]. A number of lossy encoding methods for noisy sources have been developed in the literature for a variety of signal-in-noise models. These methods range from the information-theoretic solutions for the achievable rate-distortion regions in the absence of complexity constraints [22, p. 124], [23], to more practical approaches [24], [25]. For instance, in [25], practical solutions are developed whereby the problem of lossy encoding of a noisy source is mapped to an equivalent standard lossy encoding problem and modified distortion measures are exploited to develop systems based on vector quantization. Various related decentralized detection and estimation problems in the context of bandwidth/rate constraints have been examined in the literature; see [26]–[31], and the references therein.

To illustrate some of the key issues that may arise in the encoder design, it is insightful to consider the static case, i.e., the case where the signal $A[n]$ is varying slowly enough that we may view it as static over the observation interval. Given a fixed time instant N, we can easily devise a method for efficiently encoding the N sensor measurements $s[1], s[2], \ldots, s[N]$, into a sequence of N M-ary symbols $y[1], y[2], \ldots, y[N]$ provided N is large. Specifically, consider the following algorithm:

At the sensor:
i) compute an estimate of the static information-bearing signal using the N sensor measurements;
ii) quantize the estimate using a uniform quantizer with M^N quantization levels;
iii) communicate to the host the quantized level by means of the N M-ary symbols $y[1], y[2], \ldots, y[N]$.

At the host:
reconstruct the "quantized" estimate using $y[1], y[2], \ldots, y[N]$.

Clearly, since the number of available quantization levels in step ii) of the encoder grows exponentially with the number of available observations N, the error between the "quantized" estimate used at the host and the original sensor estimate produced in step i) of the encoder (i.e., the estimate prior to quantization) decays exponentially fast with N.

A major disadvantage of such an encoding scheme, however, is that it is not sequentially refinable, namely, it provides an one-shot description; no encodings are available to the host for forming estimates before time N, and no encodings are available after time N to further refine the quality of the host estimate. Furthermore, this encoding scheme assumes that there is absolute freedom in designing the M^N-level quantizer. However, this is often not the case such as in problems where the sensors are intrinsically limited by design. For these reasons, unlike, for instance, the work in [25] and [26], in this paper we focus on designing sequentially refinable encoding strategies. This design problem can be viewed, in some sense, as the analog of the detection problem considered in [28] in the context of estimation. In addition, some of the refinable strategies we develop are shown to have interesting connections to the technique proposed in [30].

One of simplest refinable encoding strategies that can be constructed consists of quantizing each noisy measurement at the sensor by means of an M-level quantizer. As we show, however, this simple encoding scheme can have very poor performance characteristics, in terms of overcoming the power/bandwidth constraints across the network, or the dynamic range and resolution constraints at the sensor. As a means for improving the effective digital encoding we may consider the use of a control input added to the information-bearing signal prior to quantization at the sensor. The block diagram corresponding to a single sensor in the context of such a remote-sensing estimation environment is shown in Fig. 2, where $A[n]$ denotes the slowly varying information-bearing signal, $v[n]$ represents sensor noise, $w[n]$ is a control input, and $y[n]$ denotes the quantized signal that is sent to the central site. The operation of

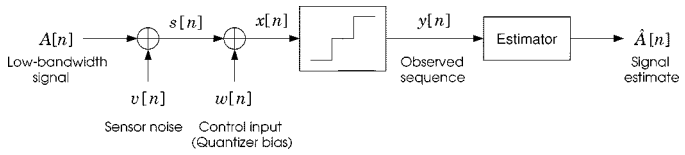

Fig. 2. Signal estimation via quantized observations, in the context of low-complexity encoding methods that are based on an additive control input.

adding a control input prior to quantization is often used in applications involving lossy compression such as video and audio coding where it is commonly referred to as dithering [32], [33]. In some (but not all) of the cases we consider, the control input plays a role very similar to dithering.

In this paper, we focus on the static case of the estimation problem depicted in Fig. 2 in which $A[n] = A$, i.e., we examine the problem of estimating a noise-corrupted unknown parameter A via quantized observations. This case reveals several key features of signal estimation from quantized observations obtained via a system comprising a control input and a quantizer; extensions of our analysis corresponding to the case where $A[n]$ corresponds to a sample path of an autoregressive moving average process are developed in [34].

Several basic variations of the estimation problem in Fig. 2 can arise in practice, which differ in the amount of information about the control input that is available for estimation and the associated freedom in the control input selection. In this paper, we develop effective control input selection strategies and associated estimators for several such important scenarios. In particular, for random control inputs that are well modeled as independent and identically distributed (i.i.d.) processes and whose statistical characterization alone is exploited at the receiver, we show that there is an optimal power level for minimizing the mean-square estimation error (MSE). The existence of a nonzero optimal control input power level reveals strong connections to the phenomenon of stochastic resonance, which is encountered in a number of physical nonlinear systems where thresholding occurs and is often exploited for signal enhancement [20], [35], [36]. In addition, it possesses strong connections to pseudorandom dithering techniques often exploited for image and audio compression [32]. Performance can be further enhanced if detailed knowledge of the applied control waveform is exploited at the receiver. In this scenario, we develop methods for judiciously selecting the control input from a suitable class of periodic waveforms for any given system. Finally, for scenarios where feedback from the quantized output to the control input is available, we show that, when combined with suitably designed receivers, these signal quantizers come within a small loss of the quantizer-free performance.[1] In the process we develop a framework for constructing the control input from past observations and design computationally efficient estimators that effectively optimize performance in terms of MSE.

The outline of the paper is as follows. In Section II, we describe the parameter estimation problem associated with the system depicted in Fig. 2. In Section III, we develop the estimation performance limits for a number of important scenarios. In Section IV, we design control inputs and associated estimators for each of these distinct scenarios, which achieve the performance limits developed in Section III. Finally, in Section V, we examine a particular network generalization of the scenario depicted in Fig. 2, in which signal estimation is based on quantized observations collected from multiple sensors.

[1]Although the feedback loop can be entirely implemented at the sensor, sensor complexity is reduced by having the feedback information come from the central site. This may be especially attractive in wireless sensor networks where there are less stringent power resources available at the central site, provided bandwidth is available for broadcasting high-resolution control information to the sensors.

II. System Model

As outlined in Section I, we consider the problem of estimating an unknown parameter A from observation of

$$y[n] = F(A + v[n] + w[n]), \qquad n = 1, 2, \ldots, N \tag{1}$$

where the sensor noise $v[n]$ is an i.i.d. process, $w[n]$ is a control input, and the function $F(\cdot)$ is an M-level quantizer, with the quantized output $y[n]$ taking M distinct values $Y_1, Y_2, \ldots, Y_M$, i.e.,

$$F(x) = \begin{cases} Y_i, & \text{if } X_{i-1} \le x < X_i \text{ for } 2 \le i \le M \\ Y_1, & \text{otherwise} \end{cases} \tag{2a}$$

where $X_0 = -\infty$ and $X_M = \infty$. We note that two sets of encodings produced by quantizing the same sequence

$$x[n] = A + v[n] + w[n]$$

via two M-level quantizers with the *same* set of X_i's and different (but distinct) quantization levels are equivalent from the point of view of signal estimation. Hence, without loss of generality, we assume that the quantizer levels are uniformly spaced, i.e.,

$$Y_i = -(M+1) + 2i, \qquad i = 1, 2, \ldots, M. \tag{2b}$$

Any other set of distinct quantization levels leads to a set of measurements $y[n]$ that is equivalent to the one generated via the quantization levels (2b), in the sense that the two quantized measurement sets are related by means of an invertible transformation. For convenience, we shall often consider the intermediate *measurement* sequence

$$s[n] \triangleq A + v[n] = A + \sigma_v \tilde{v}[n]. \tag{3}$$

We shall frequently be interested in a measure of predicted performance for a family of sensor noises parameterized by σ_v in (3), arising from scaling an i.i.d. noise sequence $\tilde{v}[n]$. We use the notation $p_z(\cdot)$ to denote the probability density function (pdf) of any sample of an i.i.d. sequence $z[n]$, and $C_z(\cdot)$ to denote one minus the corresponding cumulative distribution, i.e.,

$$C_z(x) = \int_x^\infty p_z(t)\, dt.$$

For convenience, we shall refer to an i.i.d. noise process as *admissible* if the associated pdf is nonzero and smooth (i.e., C^1) almost everywhere. Throughout the paper, we assume that all noise processes are admissible, including $v[n]$ as well as $w[n]$, when $w[n]$ is viewed as another random process. Furthermore,

when referring to a Gaussian process we assume it is i.i.d. and zero-mean, unless we specify otherwise.

III. Performance Limits

In this section, we quantify the performance degradation that results from estimating A based on observation of $y[n]$ instead of $s[n]$. We first introduce the concept of *information loss*, which we will use as a figure of merit to design quantizer systems and evaluate the associated estimators. We then present a brief preview of performance limits based on this notion for a number of important scenarios and finally develop these performance limits in Sections III-A–C.

The quality of the signal encoding is evaluated by comparing the limits of the estimate fidelity based on the encoding to that of the estimate based on the original measurements. Specifically, we define the information loss of an encoder comprising a control input followed by a quantizer as the ratio of the Cramér–Rao bounds for unbiased estimates of the parameter A obtained via $y[n]$ and $s[n]$, respectively, i.e.,

$$\mathcal{L}(A) \triangleq \frac{\mathcal{B}(A;\, \boldsymbol{y}^N)}{\mathcal{B}(A;\, \boldsymbol{s}^N)} \tag{4}$$

where $\mathcal{B}(A;\, \boldsymbol{y}^N)$ is the Cramér–Rao bound [37, p. 66] for unbiased estimation of A from[2]

$$\boldsymbol{y}^N \triangleq [\, y[1] \quad y[2] \quad \cdots \quad y[N] \,]^T \tag{5}$$

where $y[n]$ is given by (1), and where $\mathcal{B}(A;\, \boldsymbol{s}^N)$ and $\boldsymbol{s}^N$ are defined similarly. We often consider the information loss (4) in decibels [i.e., $10 \log_{10} \mathcal{L}(A)$]; it represents the additional MSE in decibels that arises from observing $y[n]$ instead of $s[n]$ in the context of efficient estimation of A. From this perspective, better systems achieve smaller information loss over the range of parameter values of interest.

Taking into account the inherent dynamic range limitations of these signal quantizers, we assume that the unknown parameter takes values in the range $(-\Delta,\, \Delta)$, with Δ assumed to be known. Often, the degradation of the estimation quality is conveniently characterized in terms of the ratio $\chi = \Delta/\sigma_v$, which we may view as a measure of peak signal-to-noise ratio (peak SNR).

Given that the signal parameter is assumed to be unknown, worst case performance is used to characterize the overall system. Accordingly, we define the worst case Cramér–Rao bound and worst case information loss via

$$\mathcal{B}_{\max}(\Delta) \triangleq \sup_{|A|<\Delta} \mathcal{B}(A;\, \boldsymbol{y}^N) \tag{6}$$

and

$$\mathcal{L}_{\max}(\Delta) \triangleq \sup_{|A|<\Delta} \mathcal{L}(A) \tag{7}$$

respectively. Both the worst case Cramér–Rao bound and the worst case information loss are functions of other system parameters, such as σ_v and $F(\cdot)$, the dependence on which is suppressed for convenience in the above definitions.

[2]Referring to $\mathcal{L}(A)$ in (4) as "information loss" is reasonable as (4) is also equal to the inverse of the ratio of the associated Fisher information quantities.

TABLE I
Order of Growth of Worst Case Information Loss as a Function of Peak SNR $\chi = \Delta/\sigma_v$ for Large χ and for Any M-Level Quantizer. The Quantity Δ Denotes the Dynamic Range of the Unknown Parameter, and σ_v Is the Sensor Noise Power Level. The Gaussian Case Refers to Gaussian Sensor Noise of Variance σ_v^2. The General Case Refers to Any Admissible Sensor Noise

Control Input	Information loss growth order	
	Gaussian case	General case
Control-free case	$e^{\chi^2/2}$	$> \chi^2$
Random input	χ^2	χ^2
Known input	χ	χ
Feedback control	1	1

As a consequence of the linear model (3), the Cramér–Rao bound $\mathcal{B}(A;\, \boldsymbol{s}^N)$ is independent of the parameter value A, i.e., $\mathcal{B}(A;\, \boldsymbol{s}^N) = \mathcal{B}(0;\, \boldsymbol{s}^N)$ for any A. Furthermore, the bound $\mathcal{B}(A;\, \boldsymbol{s}^N)$ is proportional to σ_v^2; by letting

$$\tilde{s}[n] = A + \tilde{v}[n]$$

and using (3), we obtain

$$\mathcal{B}(A;\, \boldsymbol{s}^N) = \sigma_v^2 \mathcal{B}(0;\, \tilde{s})/N \tag{8}$$

where $\mathcal{B}(0;\, \tilde{s})$ denotes the Cramér–Rao bound for estimating A based on any one sample of the i.i.d. sequence $\tilde{s}[n]$.[3] Hence, since $\mathcal{B}(A;\, \boldsymbol{s}^N)$ from (8) is independent of A, both $\mathcal{B}_{\max}(\Delta)$ and $\mathcal{L}_{\max}(\Delta)$ can be used interchangeably as figures of merit for assessing the performance of quantizer systems.

Table I summarizes the performance limits as described by the worst case information loss for a number of important scenarios. As we will show, in any of these scenarios the worst case information loss can be conveniently characterized as a function of peak SNR χ. According to Table I, random control inputs with properly chosen power levels provide performance improvements over control-free systems in any admissible noise. Specifically, for random control inputs, the control input power level can be selected so that the worst case information loss grows only quadratically with χ, while it can be shown to grow faster than quadratically in the control-free case for any admissible sensor noise. When the control input is known for estimation, the associated worst case loss can be made to grow as slow as χ with proper control input selection. Finally, if feedback information from the encoder output is available and properly used in the selection of the control input, a fixed small information loss can be achieved, which does not grow with increasing χ. In the remainder of Section III, we develop the performance limits shown in Table I, while in Section IV we develop control selection methods and associated estimators that achieve these limits.

[3]There exist i.i.d. sensor noises for which estimators based on the sensor measurements can be constructed with MSE that decays faster than $1/N$. For instance, in cases that the sensor noise is i.i.d. with uniformly distributed marginals, by using a maximum-likelihood (ML) estimator the MSE can be made to decay as $1/N^2$. In all these cases, the Cramér–Rao bound (8) does not exist. These noises are nonadmissible according to the definition in Section II, and their treatment is beyond the scope of this paper.

A. Random Control Inputs

In this section, we consider signal quantizers with control inputs $w[n]$ that correspond to sample paths of an i.i.d. process, independent of the sensor noise process $v[n]$, and determine the performance limits in estimating the unknown parameter A based on observation of $\boldsymbol{y}^N$ from (5), by simply exploiting the statistical characterization of $w[n]$ at the receiver. A similar type of control signal is often exploited in the context of lossy compression and is commonly referred to as nonsubtractive dithering; it has been shown to provide compression/distortion improvements in the context of encoding of images [38] and audio [32], and more generally in the context of lossy compression (see [39], [40], [32] and the references therein).

In general, we may consider families of random control inputs parameterized by means of a scale parameter σ_w, where $w[n] = \sigma_w \tilde{w}[n]$, and where $\tilde{w}[n]$ is an admissible i.i.d. noise sequence with pdf $p_{\tilde{w}}(\cdot)$. Our goal is to select the random control scaling parameter σ_w so as to optimize performance in terms of the associated worst case information loss.[4]

The Cramér–Rao bound for all unbiased estimates of the parameter A based on observation of the vector $\boldsymbol{y}^N$ is defined as [37, p. 66]

$$\mathcal{B}(A;\, \boldsymbol{y}^N) = -\left(E\left[\frac{\partial^2 \ln P(\boldsymbol{y}^N;\, A)}{\partial A^2} \right] \right)^{-1}$$

where $P(\boldsymbol{y}^N;\, A)$ is the associated likelihood function, denoting the probability that the particular vector $\boldsymbol{y}^N$ is observed from (1) given that the unknown parameter takes the value A. In particular, the log-likelihood function satisfies

$$\ln P(\boldsymbol{y}^N;\, A) = \sum_{i=1}^{M} \mathcal{K}_{Y_i}(\boldsymbol{y}^N) \ln \Pr(y[n] = Y_i;\, A) \tag{9}$$

where $\mathcal{K}_{Y_i}(\boldsymbol{y}^N)$ denotes the number of entries in $\boldsymbol{y}^N$ that are equal to Y_i. Since the aggregate noise

$$\alpha[n] = v[n] + w[n] \tag{10}$$

is an i.i.d. sequence, $\mathcal{B}(A;\, \boldsymbol{y}^N)$ satisfies the condition

$$\mathcal{B}(A;\, \boldsymbol{y}^N) = \frac{1}{N}\, \mathcal{B}(A;\, y) \tag{11}$$

where $\mathcal{B}(A;\, y)$ corresponds to the Cramér–Rao bound for estimating A based on any one sample of the i.i.d. sequence $y[n]$. Finally, by taking the second partial derivative of (9) with respect to A followed by an expectation, we obtain

$$\mathcal{B}(A;\, y) = \left(\sum_{i=1}^{M} \frac{[p_\alpha(X_{i-1} - A) - p_\alpha(X_i - A)]^2}{C_\alpha(X_{i-1} - A) - C_\alpha(X_i - A)} \right)^{-1}. \tag{12}$$

For the system corresponding to the symmetric two-level quantizer $(M = 2, X_1 = 0)$, i.e.,

$$F(x) = \operatorname{sgn}(x) \tag{13}$$

the Cramér–Rao bound (12) reduces to

$$\mathcal{B}(A;\, y) = C_\alpha(-A)[1 - C_\alpha(-A)][p_\alpha(-A)]^{-2}. \tag{14}$$

When, in addition, the pdf $p_\alpha(\cdot)$ is an even function of its argument, (14) further specializes to

$$\mathcal{B}(A;\, y) = \mathcal{B}(-A;\, y) = C_\alpha(-A) C_\alpha(A) [p_\alpha(A)]^{-2}. \tag{15}$$

We next consider the special case where $v[n]$ and $w[n]$ are i.i.d. Gaussian processes and $F(\cdot)$ is the symmetric two-level quantizer, and determine the random control power level that minimizes the worst case information loss. We then consider the general case, i.e., the case $M \geq 2$ where $v[n]$ and $w[n]$ are any i.i.d. processes.

1) Special Case—Gaussian Noises and $M = 2$: For the system $M = 2$ where $v[n]$ and $w[n]$ are independent i.i.d. Gaussian noise sequences with variances σ_v^2 and σ_w^2, respectively, the Cramér–Rao bound (15) reduces to

$$\mathcal{B}(A;\, y) = 2\pi\sigma_\alpha^2 Q\left(\frac{A}{\sigma_\alpha}\right) Q\left(-\frac{A}{\sigma_\alpha}\right) \exp\left(\frac{A^2}{\sigma_\alpha^2}\right) \tag{16}$$

where

$$\sigma_\alpha = \sqrt{\sigma_w^2 + \sigma_v^2}$$

and

$$Q(x) = \int_x^\infty (1/\sqrt{2\pi}) e^{-t^2/2}\, dt.$$

Fig. 3 depicts the associated information loss (4) as a function of A for $\Delta = 1$, $\sigma_v = 0.1$, and various σ_w levels. Observation of Fig. 3 reveals several key characteristics of this type of quantizer-based processing. Specifically, in this Gaussian sensor noise scenario, the minimum achievable information loss occurs for $A = 0$ and $\sigma_w = 0$ and equals $10 \log 10(\pi/2) \approx 2$ dB. In addition, for any random control power level σ_w^2 the information loss is an increasing function of $|A|$. This property is shared by many other admissible i.i.d. processes with common marginals, such as the i.i.d. Laplacian and Cauchy random processes. More important, as the figure reveals, proper use of a random control input $(\sigma_w \neq 0)$ can have a major impact on performance in terms of reducing the associated worst case information loss.

The sensitivity of performance with respect to the optimal control noise power level for the Gaussian noise scenario is examined in Fig. 4, where we depict the additional worst case information loss (in decibels) due to suboptimal selection of the control noise level versus σ_α/Δ. Note that 0-dB additional worst case information loss corresponds to the optimal random control power level selection. From the figure we see that the optimal aggregate noise level is well approximated by

$$\sigma_\alpha^{\text{opt}} \approx \frac{2}{\pi}\, \Delta \tag{17}$$

so that the optimal random control power level satisfies

$$\sigma_w^{\text{opt}} = \begin{cases} \sqrt{(\sigma_\alpha^{\text{opt}})^2 - \sigma_v^2}, & \text{if } \sigma_v < \sigma_\alpha^{\text{opt}} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

[4]The scaling factor σ_w can be viewed as a measure of the strength of the control input process $w[n]$. For cases where the control input variance exists, we may pick the prototype $\tilde{w}[n]$ as the unit-variance member in the family, in which case σ_w^2 corresponds to the power level of the control input signal.

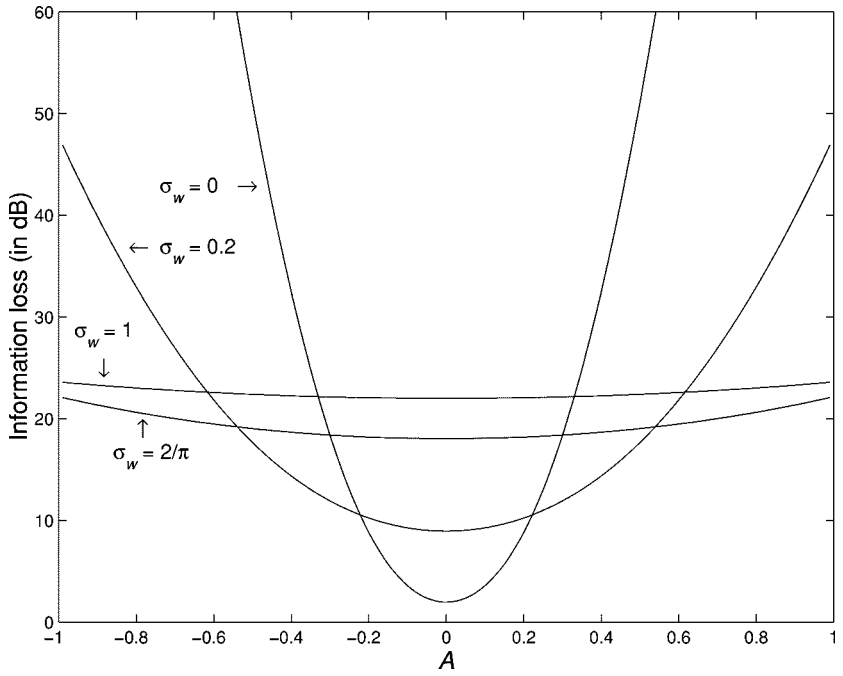

Fig. 3. Information loss for a system comprising a two-level quantizer and an i.i.d. Gaussian control input, for various control signal power levels σ_w^2. The sensor noise is i.i.d. Gaussian with variance $\sigma_v^2 = 0.01$.

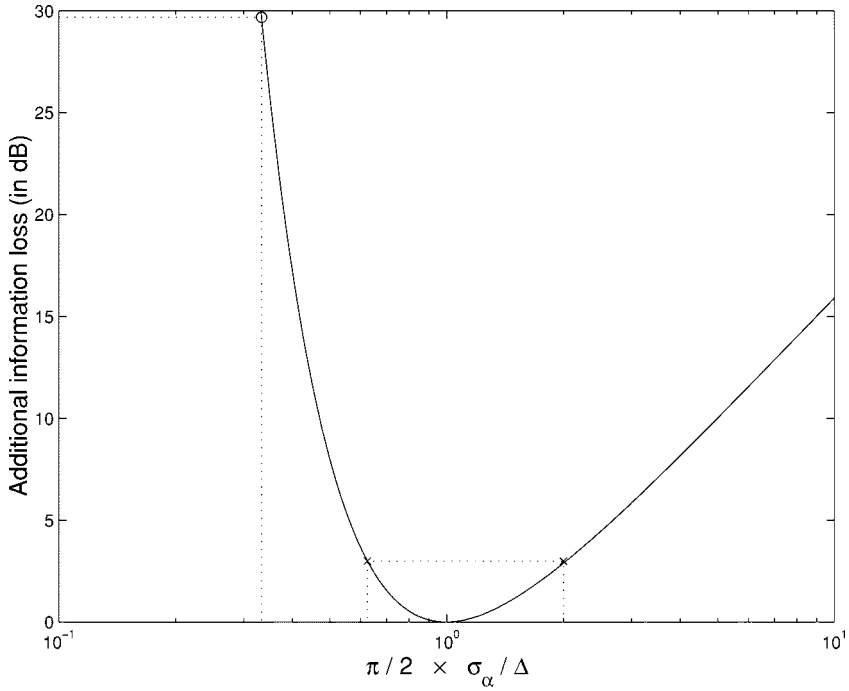

Fig. 4. Additional worst case information loss (solid) due to suboptimal random control power level selection for a two-level quantizer. The net noise sequence $\alpha[n] = v[n] + w[n]$ is Gaussian with variance σ_α^2. The "×" marks depict the additional information loss for net noise levels $5/8\sigma_\alpha^{\rm opt}$ and $2\sigma_\alpha^{\rm opt}$. The "∘" mark depicts the additional information loss at $\sigma_\alpha^{\rm opt}/3$.

If $\sigma_v \ll \Delta$, Fig. 4 reveals that for the fairly wide range of control input power levels

$$\frac{5}{8}\sigma_\alpha^{\rm opt} \le \sigma_w \le 2\sigma_\alpha^{\rm opt}$$

the associated performance is inferior to that corresponding to the optimal random control power level by less than 3 dB. However, the performance degrades rapidly as the control input power level is reduced beyond $5\sigma_\alpha^{\rm opt}/8$. For instance, for $\sigma_w = \sigma_\alpha^{\rm opt}/3$, there is nearly 30 dB of additional loss incurred by the suboptimal selection of the control input power level.

The information loss associated with the optimal random control power level corresponds to the best achievable performance by a particular family of random control inputs—in this particular example, the family of zero-mean normal distributions. For the optimal choice of σ_w in (18), the worst case information

loss can be completely characterized by means of peak SNR χ. In particular, by using (17), (18) with (16) in (4) we obtain the optimal worst case information loss for the Gaussian scenario with random control, namely

$$\mathcal{L}_{\max}^{\text{rand}}(\chi) = \begin{cases} 2\pi Q(\chi)Q(-\chi)e^{\chi^2}, & \text{if } 0 \le \chi \le \frac{\pi}{2} \\ \frac{8}{\pi} Q\left(\frac{\pi}{2}\right) Q\left(-\frac{\pi}{2}\right) e^{\frac{\pi^2}{4}}\chi^2, & \text{if } \frac{\pi}{2} \le \chi \end{cases} \tag{19}$$

where we indicate explicitly that in this case the worst case information loss is a function of χ.

As (19) reveals, for parameter estimation in Gaussian noise via a two-level quantizer system, the worst case information loss can be made to grow quadratically with peak SNR by judicious selection of a Gaussian control input. For comparison, the worst case information loss in the absence of control input grows exponentially with peak SNR. In particular, by substituting $\mathcal{B}(A;\, y)$ from (16) in (7), we obtain

$$\mathcal{L}_{\max}^{\text{free}}(\chi) = 2\pi Q(\chi)Q(-\chi)e^{\chi^2} \tag{20}$$

which grows as $\exp(\chi^2/2)$ for large χ. The results in (19) and (20) extend to quantizers with $M > 2$, i.e., the worst case information loss grows as $\exp(\chi^2/2)$ for control-free systems, while it can be made to grow as χ^2 for appropriately chosen Gaussian control inputs.

2) General Case: Arbitrary Admissible Noises and $M \ge 2$:

As we next show, proper use of an admissible random control input $w[n]$ can improve performance at high SNR over the control-free system in any (admissible) sensor noise $v[n]$ and for any M-level quantizer. Substituting (8) and (11) in (4) reveals that the associated information loss is independent of N. Thus, we may focus on the case $N = 1$ without any loss of generality. We next use $\mathcal{B}_{\max}(\Delta;\, \sigma_v,\, \sigma_w,\, \{X_i\}_{i=1}^{M-1})$ to denote the worst case Cramér–Rao bound (6), in order to make its dependence on σ_v, σ_w, and the quantizer thresholds explicit. Also, we suppress the dependence of $\mathcal{B}_{\max}(\cdot)$ on the quantizer thresholds when there is no ambiguity.

For admissible $\tilde{v}[n]$, the Cramér–Rao bound (12) is continuous in the σ_v variable, and hence so is $\mathcal{B}_{\max}(\Delta;\, \sigma_v,\, \sigma_w)$. Thus, given any fixed $\sigma_w > 0$ and Δ, for small enough σ_v we have

$$\mathcal{B}_{\max}(\Delta;\, \sigma_v,\, \sigma_w) \approx \mathcal{B}_{\max}(\Delta;\, 0,\, \sigma_w). \tag{21}$$

Substituting (21) and (8) in (7) while keeping Δ fixed and letting $\sigma_v^2 \to 0$ reveals that $\mathcal{L}_{\max}^{\text{rand}}(\chi) \sim \chi^2$ is achievable for large χ. Furthermore, since $\mathcal{B}_{\max}(\Delta;\, \sigma_v,\, \sigma_w)$ is also continuous in σ_w, for any $F(\cdot)$ with fixed $M < \infty$

$$\inf_{\sigma_w \in [0,\, \infty)} \mathcal{B}_{\max}(\Delta;\, 0,\, \sigma_w) > 0 \tag{22}$$

for any $\Delta > 0$. In addition, given that the sequence $y[n]$ does not change if we scale both the input to the quantizer and the quantizer thresholds by $1/\Delta$, the Cramér–Rao for estimating A based on $y[n]$ is Δ^2 times the Cramér–Rao for estimating $\overline{A} = A/\Delta$ based on the sequence generated by quantizing $x[n]/\Delta$ via an M-level quantizer with thresholds $\overline{X}_i = X_i/\Delta$, i.e.,

$$\begin{aligned} &\mathcal{B}_{\max}(\Delta;\, \sigma_v,\, \sigma_w,\, \{X_i\}_{i=1}^{M-1}) \\ &\qquad = \Delta^2 \mathcal{B}_{\max}\left(1;\, \frac{1}{\chi},\, \frac{\sigma_w}{\Delta},\, \left\{\frac{X_i}{\Delta}\right\}_{i=1}^{M-1}\right) \end{aligned} \tag{23}$$

which in conjunction with (8), (21), and (22) implies that the worst case information loss cannot be made to grow slower than χ^2 for random control inputs. Therefore, at high peak SNR the worst case information loss for random control inputs $\mathcal{L}_{\max}^{\text{rand}}(\chi)$ can be made to grow at best as slow as quadratically with peak SNR for random control inputs. In general, the sensor noise level may be fixed, in which case we are interested in selecting the random control level σ_w as a function of the dynamic range Δ so as to minimize the worst case information loss. From (21)–(23) the optimal worst case information loss rate can be achieved by selecting $\sigma_w = \lambda\Delta$ for some $\lambda > 0$. This is in agreement with our conclusions for the Gaussian scenario in the special case $M = 2$, as (17)–(19) clearly demonstrate. For comparison, in Appendix A, we show that for control-free systems corresponding to $F(\cdot)$ in (2a), (2b), and for any sensor noise the worst case information loss $\mathcal{L}_{\max}^{\text{free}}(\chi)$ grows faster than χ^2 for large χ. Remarkably, random control inputs with appropriately selected power levels provide performance improvements over the control-free systems for any admissible sensor noise at high peak SNR.

B. Known Control Inputs

We next develop performance limits for scenarios where the estimator can exploit detailed knowledge of a suitably designed control waveform. In particular, we determine the minimum possible growth rate of the worst case information loss as a function of χ, and develop control input selection strategies that achieve the minimum possible rate.

The Cramér–Rao bound for unbiased estimates of A based on $\boldsymbol{y}^N$ and given knowledge of the associated N samples of $w[n]$ is denoted by $\mathcal{B}(A;\, \boldsymbol{y}^N;\, \boldsymbol{w}^N)$ and satisfies

$$\begin{aligned} \mathcal{B}(A;\, \boldsymbol{y}^N,\, \boldsymbol{w}^N) &= -\left(E\left[\frac{\partial^2 \ln P(\boldsymbol{y}^N;\, A,\, \boldsymbol{w}^N)}{\partial A^2}\right]\right)^{-1} \\ &= \left[\sum_{n=1}^{N} [\mathcal{B}(A + w[n];\, y)]^{-1}\right]^{-1} \end{aligned} \tag{24}$$

where $\mathcal{B}(A;\, y)$ is given by (12), with α replaced by v, and where $P(\boldsymbol{y}^N;\, A,\, \boldsymbol{w}^N)$ denotes the associated likelihood function. As expected, the associated worst case Cramér–Rao bound and worst case information loss are functions of the control waveform $\boldsymbol{w}^N$. In Appendix B, we show that, for any known control waveform selection strategy, the worst case information loss associated with any M-level signal quantizer grows at least as fast as χ for any sensor noise distribution. This includes the optimal scheme, which selects the waveform $w[n]$ that results in minimizing the worst case information loss for any given set $\{\Delta,\, \sigma_v,\, p_{\tilde{v}}(\cdot),\, F(\cdot)\}$.

Classes of periodic waveforms parameterized by the period K are attractive candidates for known control inputs, since they are easy to construct and can be chosen so that the worst case information loss grows at the minimum possible rate. In constructing these classes of periodic sawtooth waveforms, we will use as a figure of merit the worst case information loss as $N \to \infty$; extensions to the finite N case are developed in Appendix

B. From (24), the Cramér–Rao bound for estimating A based on $\boldsymbol{y}^N$, where N is a multiple of the period K, is given by

$$\mathcal{B}(A;\, \boldsymbol{y}^N,\, \boldsymbol{w}^N) = \frac{1}{N}\,\frac{K}{\sum_{n=1}^{K}[\mathcal{B}(A+w[n];\, y)]^{-1}}. \tag{25}$$

As we will show, in order to achieve the minimum possible growth rate it suffices to select $w[n]$ from properly constructed K-periodic classes for which there is a one-to-one correspondence between each element in the class and the period K. Optimal selection of the control input in this case is equivalent to selecting the period K that minimizes the associated worst case information loss, or equivalently, the worst case Cramér–Rao bound from (25)

$$K_{\rm opt}(\Delta,\, \sigma_v) \triangleq \arg\min_{K}\ \sup_{A\in(-\Delta,\,\Delta)}\ \frac{K}{\sum_{n=1}^{K}[\mathcal{B}(A+w[n];\, y)]^{-1}} \tag{26}$$

where $\mathcal{B}(A;\, y)$ is given by (12) with α replaced by v. We next develop a method for selecting the control waveform from properly constructed classes of K-periodic waveforms for the case $M = 2$, which results in achieving the optimal growth rate of worst case information loss. Then, we extend our method to quantizers with $M > 2$.

1) Optimized Periodic Waveforms for Signal Quantizers with $M = 2$: The construction of the elements of the K-periodic class in the case $M = 2$ is based on the observation that in the control-free scenario the worst case information loss grows with Δ for fixed σ_v. This observation suggests that the information loss is typically largest for parameter values that are furthest from the quantizer threshold. This is strictly true, for instance, for Gaussian sensor noise, since $\mathcal{B}(A;\, y)$ in (16) is an increasing function of $|A|$. Since our objective is to optimize over the worst case performance, a potentially effective strategy is to construct the K-periodic waveform $w[n]$ so as to minimize the largest distance between any A in $(-\Delta,\, \Delta)$ and the closest *effective* quantizer threshold. For this reason, we consider K-periodic control inputs, which have the form of the sawtooth waveform

$$w[n] = \delta_w\left(-\frac{K-1}{2} + n \bmod K\right) \tag{27}$$

where the effective spacing between thresholds is given by $\delta_w = 2\Delta/(K-1)$. The net effect of the periodic control input (27) and the symmetric two-level quantizer (13) is equivalent to a two-level quantizer with a periodically time-varying threshold; it is important to observe that the time-varying quantizer threshold comes within at least $\delta_w/2$ of any possible parameter value once every K samples.

For the system with $F(\cdot)$ given by (13) and $w[n]$ given by (27), the optimal period $K_{\rm opt}$ is completely characterized by means of peak SNR χ; using (14) in (26) reveals that $K_{\rm opt}$ satisfies

$$K_{\rm opt}(\Delta,\, \sigma_v) = K_{\rm opt}(\mu\Delta,\, \mu\sigma_v)$$

for any $\mu > 0$. For this reason, we will use the one-variable function $K_{\rm opt}(\chi)$ to refer to the optimal period from (26) for a particular χ.

In the context of the sawtooth K-periodic inputs (27), strategies that select K so as to keep a fixed sawtooth spacing δ_w achieve the minimum possible growth rate. In particular, in Appendix B we show that, for any given χ, if we select the period K in (27) according to

$$K = \lceil \lambda\chi + 1 \rceil \tag{28}$$

where λ can be any positive constant, the associated worst case information loss grows linearly with χ. In general, there is an optimal λ for any particular noise pdf $p_{\tilde{v}}(\cdot)$, resulting in an optimal *normalized* sawtooth spacing. Specifically, consider the normalized spacing between successive samples of $w[n]$ in (27), namely

$$d(\chi;\, K) \triangleq \frac{\delta_w}{\sigma_v} = \frac{2\chi}{K-1}. \tag{29}$$

In addition, let $d_{\rm opt}(\chi)$ denote the normalized spacing associated with the optimal period $K_{\rm opt}(\chi)$ from (26), i.e.,

$$d_{\rm opt}(\chi) \triangleq d(\chi;\, K_{\rm opt}(\chi)). \tag{30}$$

In Appendix B, we outline a method for finding the asymptotic optimal normalized spacing

$$d_\infty \triangleq \lim_{\chi\to\infty} d_{\rm opt}(\chi) \tag{31}$$

associated with a particular sensor noise pdf. For purposes of illustration, we also show in Appendix B that in the special case that the sensor noise is Gaussian with variance σ_v^2

$$d_\infty \approx 2.5851 \tag{32}$$

while the associated worst case information loss is well approximated by

$$\mathcal{L}_{\max}^{\rm per}(\chi) \approx 1.4754\left(\frac{2\chi}{d_\infty} + 1\right) \tag{33}$$

for large χ. In this Gaussian scenario, if we select $w[n]$ as in (27) with $K = \lceil 2\chi/d_\infty + 1 \rceil$, the worst case information loss is given by (33) and achieves the optimal growth rate for known control waveforms. We next extend the above analysis to quantizers with $M > 2$.

2) Optimized Periodic Waveforms for Signal Quantizers with $M > 2$: As we have seen in the preceding section, selection of $w[n]$ according to (27) for $M = 2$ results in a two-level quantizer with periodically time-varying thresholds uniformly spaced in $[-\Delta,\, \Delta]$. This selection method minimizes the maximum distance between the parameter value and the closest of the time-varying thresholds, over the dynamic range $(-\Delta,\, \Delta)$. The same strategy can be used for $M > 2$, although the availability of multiple thresholds allows for reduction of the dynamic range that $w[n]$ needs to span. We assume that all quantizer thresholds are within the dynamic range, i.e., $-\Delta < X_i < \Delta$, for $i = 1,\, 2,\, \ldots,\, M-1$. In this case, the effective dynamic range $\Delta_{\rm eff}$ that $w[n]$ needs to span is given by

$$\Delta_{\rm eff} = \max_i \delta x_i$$

where

$$\delta x_i = \begin{cases} X_1 + \Delta, & \text{if } i = 1 \\ X_i - X_{i-1}, & \text{if } 2 \le i \le M-2 \\ \Delta - X_{M-1}, & \text{if } i = M-1. \end{cases}$$

In particular, we consider using the control input (27) where the effective spacing between thresholds δ_w is given in terms of Δ and the quantizer thresholds $X_1, X_2, \ldots, X_{M-1}$ as follows:

$$\delta_w = \max_i \delta w_i \tag{34a}$$

where

$$\delta w_i = \begin{cases} \dfrac{2\delta x_i}{K-1}, & \text{if } i = 1, M-1 \\ \dfrac{2\delta x_i}{K}, & \text{if } 2 \le i \le M-2. \end{cases} \tag{34b}$$

For any A in $(-\Delta, \Delta)$, this selection guarantees that at least one of the $M-1$ time-varying quantizer thresholds is within $\delta_w/2$ of the parameter, where δ_w is given by (34a). One can, in principle, perform the optimization (26) to obtain $K_{\text{opt}}(\Delta, \sigma_v)$ for any $F(\cdot)$ with $M > 2$. We should emphasize, however, that at high SNR we may often obtain an approximate estimate of performance via our results for the case $M = 2$. For instance, for $\Delta_{\text{eff}}/\sigma_v$ large and small enough λ in (28), the optimal normalized spacing and the corresponding worst case information loss for a quantizer with $M > 2$ are approximately given by the respective quantities for the symmetric two-level quantizer, with χ replaced by $\chi_{\text{eff}} = \Delta_{\text{eff}}/\sigma_v$.

If, in addition, there is freedom in selecting the $M-1$ quantizer thresholds, these can be selected so that $\delta w_i = \delta w_j$ for all i and j in (34b) which implies that $\delta_w = \Delta/[(M-1)K-1]$. This selection guarantees that for every K successive observations, the collection of all MK associated quantizer thresholds form a uniformly spaced collection in $[-\Delta, \Delta]$. For instance, in the special case that the sensor noise is Gaussian, the optimal normalized spacing and the worst case loss for large χ are given by (32) and (33), respectively, with $\chi/(M-1)$ replacing χ on the left-hand side of (33). In summary, simply constructed classes of periodic control waveforms achieve the optimal information loss growth rate with peak SNR.

C. Control Inputs in the Presence of Feedback

In this section, we consider the scenario where, in addition to knowing the control waveform, the estimator has the option of using feedback from past output observations in the selection of the present control input. Specifically, we develop performance bounds for the problem of estimation of A based on $\boldsymbol{y}^N$, where the control input sequence $w[n]$ is a function of all past quantized observations $\boldsymbol{y}^{n-1}$. This scenario is depicted in Fig. 5 where $w[n] = g(\boldsymbol{y}^{n-1})$.

We next show that the worst case information loss for any feedback-based control input strategy is lower-bounded by the *minimum* possible information loss for the same quantizer system with $w[n] = 0$; in Section IV, we develop feedback-based control selection algorithms that effectively achieve this lower bound. Examination of the Cramér–Rao bound (24) reveals that for any A in $(-\Delta, \Delta)$ we can obtain information loss equal to $\mathcal{L}(A_\circ)$ by selecting $w[n] = A_\circ - A$.

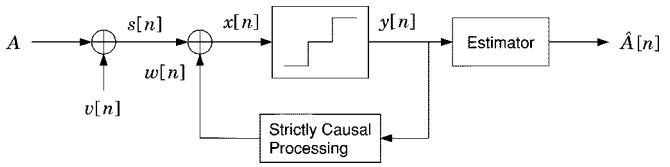

Fig. 5. Estimation based on observations from a signal quantizer, where feedback from the quantized output is used in the selection of the control input.

In particular, if there exists a parameter value A_* for which $\mathcal{B}(A; y) \ge \mathcal{B}(A_*; y)$ for all A in $(-\infty, \infty)$ and where $\mathcal{B}(A; y)$ is given by (12) with α replaced by v, then using (24) we obtain

$$\mathcal{B}(A; \boldsymbol{y}^N, \boldsymbol{w}^N) \ge \mathcal{B}(A_*; y)/N \tag{35}$$

with equality achieved for $w[n] = A_* - A$ for $n = 1, 2, \ldots, N$. This control input results in

$$\mathcal{L}(A; \boldsymbol{w}^N) \ge \mathcal{L}(A; A_* - A) = \mathcal{L}(A_*) \tag{36}$$

where $\mathcal{L}(A)$ is given by (4), and where $\mathcal{B}(A; y)$ is given by (12) with α replaced by v.

The minimum information loss from (36) decreases as the number of quantization levels increases. In Appendix C, we show that as we would expect, the minimum information loss $\mathcal{L}(A_*)$ tends to zero as the number of quantization levels approaches infinity for any sensor noise.

For a number of common sensor noises the control-free information loss for the system corresponding to $M = 2$ is minimized at the negative of the median of the pdf $p_v(\cdot)$, i.e., $C_v(-A_*) = 1/2$. The corresponding minimum information loss (36) can be obtained by evaluating (4) at $A = A_*$, while employing (8) and (14) for $\sigma_w = 0$, namely

$$\mathcal{L}(A_*) = [4p_{\tilde{v}}^2(-A_*/\sigma_v)\mathcal{B}(0; \tilde{s})]^{-1} \tag{37}$$

which is independent of σ_v and Δ, since $-A_*/\sigma_v$ equals the median of the pdf of $\tilde{v}[n]$.

1) Special Case: Gaussian Sensor Noise: In the case that the sensor noise is Gaussian, the minimum information loss (36) decays rapidly to zero as more quantization levels are introduced. In Fig. 6 we plot the minimum possible information loss through any uniform M-level quantizer for various values of M, in the presence of i.i.d. Gaussian noise. From the figure, it is apparent that a few quantization levels suffice to effectively eliminate the minimum information loss due to quantizer-based processing.

For the two-level quantizer (13) in this Gaussian scenario, use of (16) for $\sigma_\alpha = \sigma_v$ in (7) reveals that $A_* = 0$. In this case, (35) reduces to

$$\mathcal{B}(A; \boldsymbol{w}^N, \boldsymbol{y}^N) \ge \mathcal{B}(0; y)/N = \frac{\pi\sigma_v^2}{2N} \tag{38}$$

while from (37) the information loss for any parameter value A is lower-bounded as follows:

$$\mathcal{L}(A; \boldsymbol{w}^N) \ge \mathcal{L}(0) = \frac{\pi}{2} \tag{39}$$

which corresponds to a 2-dB information loss.

Fig. 7 depicts the worst case information loss for the system corresponding to $M = 2$ in the context of Gaussian sensor noise and the various control input scenarios that we have examined. As reflected in the figure, the performance of the control-free

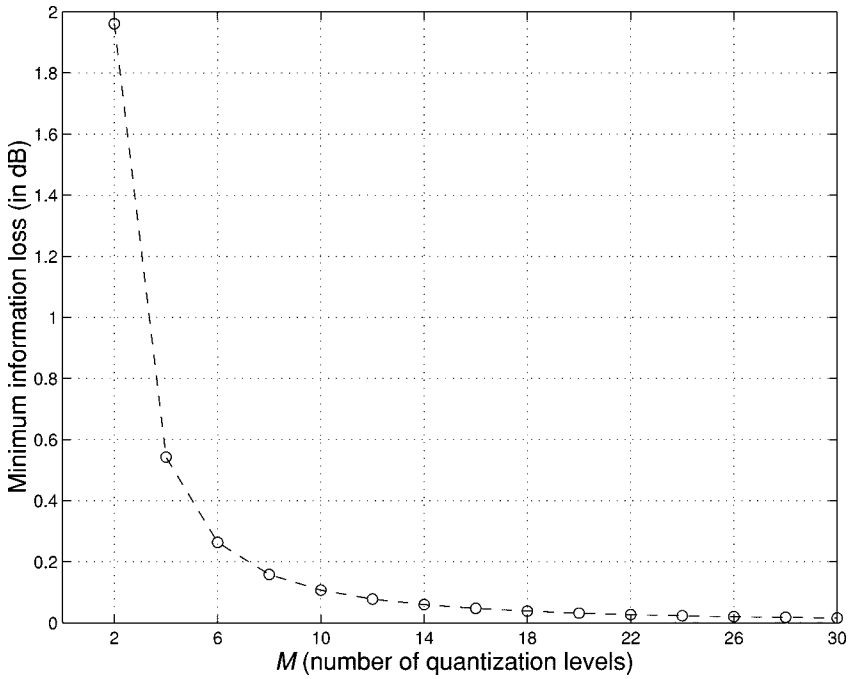

Fig. 6. Minimum possible information loss as a function of quantization levels for a uniform quantizer in i.i.d. Gaussian noise. For any given M, the threshold spacing is selected so as to minimize this loss.

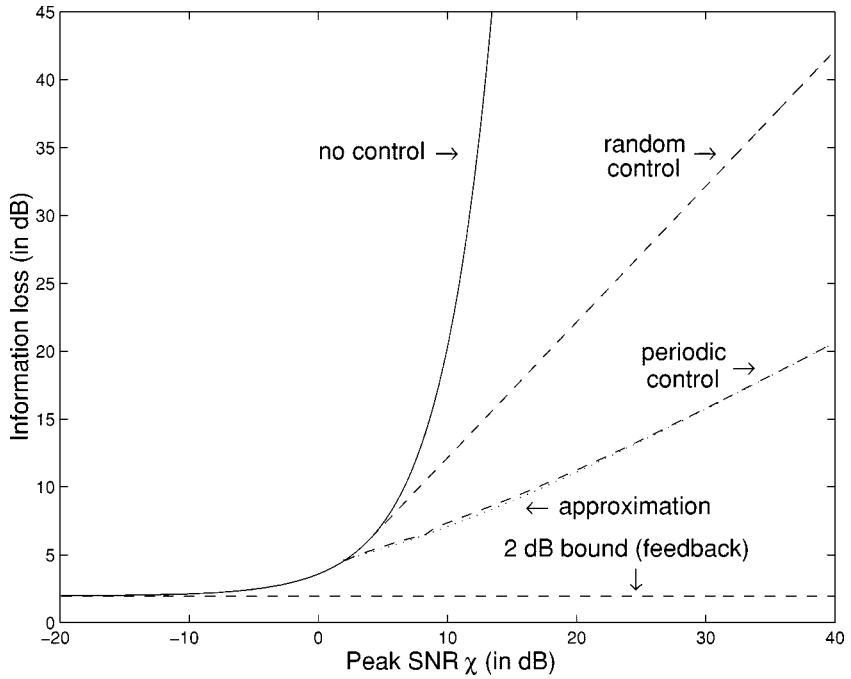

Fig. 7. Worst case information loss over $|A| < \Delta$ for a two-level quantizer in zero-mean i.i.d. Gaussian noise of variance σ_v^2, with no control input (solid), random control inputs (upper dashed), and known periodic control waveforms (middle dashed). The dotted curve depicts approximation (33). The lower dashed line depicts the minimum possible information loss ($\approx$ 2 dB) for any control input scheme.

system (solid curve) degrades rapidly as the peak SNR is increased. The benefits of random control inputs (upper dashed curve) at high peak SNR are clearly evident, and known periodic control inputs provide additional performance benefits (middle dashed curve) over random control inputs. In particular, the associated worst case information loss increases linearly with peak SNR as the accurate approximation (33) reveals. Finally, in the presence of feedback from the quantized output to the control input, the performance is lower-bounded by the minimum possible information loss of 2 dB, which is independent of χ. In Section IV, we develop control selection strategies and associated estimators that meet all these bounds.

IV. Efficient Estimation

In this section, we develop control input selection strategies and associated estimators which achieve the performance limits obtained in Section III. A natural measure of performance of a specific system, comprising a control input a quantizer and a particular estimator, is the *MSE loss*; it is defined as the ratio of the actual MSE of a particular estimator of A based on observation of $\boldsymbol{y}^N$, divided by the Cramér–Rao bound for estimating A from observation of $\boldsymbol{s}^N$. Whenever an efficient estimator of A based on $\boldsymbol{s}^N$ exists, the notion of the MSE loss of any given estimator of A given $\boldsymbol{y}^N$ has an alternative, instructive interpretation: it represents the additional MSE in decibels that arises from estimating A using this particular estimator on $\boldsymbol{y}^N$, instead of efficiently estimating A via $\boldsymbol{s}^N$. Analogously to $\mathcal{L}_{\max}$ in (7), the worst case MSE loss of an estimator is defined as the supremum of the MSE loss function over the range $|A| < \Delta$.

In this section, we construct estimators for which the corresponding MSE loss asymptotically $(N \to \infty)$ achieves the associated information loss, for each of the control input scenarios of Section III. We examine the control-free and random control scenarios first, and then develop estimators applicable to known K-periodic control inputs. Finally, in the context of feedback, we develop control input selection strategies and associated estimators which achieve the minimum possible information loss for any given scalar quantizer system.

A. Random Control Inputs

For random control inputs, the ML estimator of A based on $\boldsymbol{y}^N$ and restricted over the dynamic range $|A| \leq \Delta$ satisfies

$$\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \Delta) = \arg \max_{|\theta| \leq \Delta} \ln P(\boldsymbol{y}^N; \theta) \tag{40}$$

where $\ln P(\boldsymbol{y}^N; \theta)$ is the log-likelihood function given by (9). We first examine ML estimation for the system with $M = 2$, and then construct estimators for signal quantizers with $M > 2$. Estimators of A for control-free systems can be readily obtained as a special case of the estimators of A for the associated systems with random control inputs by setting $\sigma_w = 0$.

1) ML Estimation for Signal Quantizers with $M = 2$ in i.i.d. Noise: If $F(\cdot)$ is given by (13) and $\alpha[n]$ is admissible, the ML estimator (40) can be found in closed form, by setting to zero the partial derivative of the log-likelihood function (9) with respect to A, viz.,

$$\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \Delta) = \mathcal{I}_\Delta(\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \infty)) \tag{41}$$

where $\mathcal{I}_\Delta(\cdot)$ is the following piecewise-linear limiter function:

$$\mathcal{I}_\Delta(x) = \begin{cases} x, & \text{if } |x| \leq \Delta \\ \Delta \mathrm{sgn}(x), & \text{otherwise.} \end{cases} \tag{42}$$

The function $\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \infty)$ denotes the ML estimate of A from $\boldsymbol{y}^N$ when there are no restrictions imposed in the dynamic range of the unknown parameter A.[5] In particular

$$\begin{aligned} \hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \infty) &= \arg \max_{\theta} \ln P(\boldsymbol{y}^N; \theta) \\ &= -C_\alpha^{-1}\left(\frac{\mathcal{K}_1(\boldsymbol{y}^N)}{N}\right) \end{aligned} \tag{43}$$

where $C_\alpha^{-1}(\cdot)$ in (43) is the inverse of $C_\alpha(\cdot)$, and $\mathcal{K}_{Y_i}(\boldsymbol{y}^N)$ denotes the number of elements in $\boldsymbol{y}^N$ that are equal to Y_i. In the special case that $w[n]$ and $v[n]$ are zero-mean i.i.d. Gaussian noise sequences with variances σ_w^2 and σ_v^2, respectively, (43) reduces to

$$\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \infty) = -\sigma_\alpha Q^{-1}\left(\frac{\mathcal{K}_1(\boldsymbol{y}^N)}{N}\right). \tag{44}$$

For any parameter value A in the range $(-\Delta, \Delta)$, the Cramér–Rao bound (14) is a reasonable predictor of the MSE performance of the ML estimator (41)–(43) provided that the number of observations N is large enough. Indeed, as shown in Appendix D, for any $A \in (-\Delta, \Delta)$, the ML estimator (41)–(43) is asymptotically efficient in the sense that it achieves the Cramér–Rao bound for unbiased estimates (14) for large enough N, i.e.,

$$\lim_{N \to \infty} NE[(\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \Delta) - A)^2] = \mathcal{B}(A; y).$$

Although the ML estimate (41)–(43) is asymptotically unbiased and efficient for any A in $(-\Delta, \Delta)$, the associated MSE does not converge uniformly to the Cramér–Rao bound in the parameter A with N. Specifically, for any fixed N, no matter how large, there exist parameter values close enough to the boundaries $\pm\Delta$ for which the ML estimator has significant bias,[6] in which case (14) should not be expected to accurately predict the associated MSE of the ML estimator. This is clearly reflected in Fig. 8, where the actual MSE loss for $\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \Delta)$ is also depicted alongside the associated information loss for the Gaussian noise scenario. In particular, the dashed and solid lines depict the MSE loss from Monte Carlo simulations for the ML estimator (41)–(43), in the absence $(\sigma_w = 0)$ and presence $(\sigma_w = 2/\pi)$ of a random control input, respectively, for $\sigma_v = 0.1$, $\Delta = 1$, and $N = 100, 10^4$. As we can see in Fig. 8, when the random control level is $\sigma_w = 2/\pi$, the worst case MSE loss is about 21 dB. However, in the absence of a control input, the worst case MSE loss is about 36 dB for $N = 100$, and 55 dB for $N = 10^4$. For both values of N, the Cramér–Rao bound (14) is applicable for only a subset of the dynamic range, whose size increases with N. In fact, since the ML estimator is asymptotically efficient for any $|A| < \Delta$ with respect to the Cramér–Rao bound (14) for unbiased estimates, the worst case MSE loss for the control-free system increases with N toward the associated worst case information loss (20), which is approximately 211 dB.[7]

2) ML Estimation for Signal Quantizers with $M > 2$ in i.i.d. Gaussian Noise: For the estimation problem (1), (2a), (2b), where $F(\cdot)$ is an M-level quantizer and $\alpha[n]$ is an i.i.d. sequence, the set of sufficient statistics reduces to $\mathcal{K}_{Y_1}(\boldsymbol{y}^N), \ldots, \mathcal{K}_{Y_{M-1}}(\boldsymbol{y}^N)$ [see (9)]. For the special case that $\alpha[n]$ is Gaussian with variance σ_α^2, we develop in Appendix E an expectation–maximization (EM) algorithm [41] for obtaining the ML estimate (40). This algorithm takes the form

[5] Note that (41) does not necessarily hold for $M > 2$.

[6] By incorporating the bias of the ML estimator (41)–(43) it is possible to obtain a Cramér–Rao bound that directly applies to the associated MSE. An even tighter bound is obtained by properly combining three separate Cramér–Rao bounds, each describing the effects of a piecewise-linear region of the soft limiter $\mathcal{I}_\Delta(\cdot)$ on $\hat{A}_{\mathrm{ML}}(A; \infty)$ in (41) [34]. These bounds, however, are beyond the scope of this paper.

[7] Inevitably, the pseudonoise samples $w[n]$ generated in the computer simulations are uncorrelated but not i.i.d.; the ML estimator achieves the bound (16) as expected (see Appendix A). However, in the case the $w[n]$'s are uncorrelated but not i.i.d., the actual Cramér–Rao bound may be less than the one corresponding to the i.i.d. case (16).

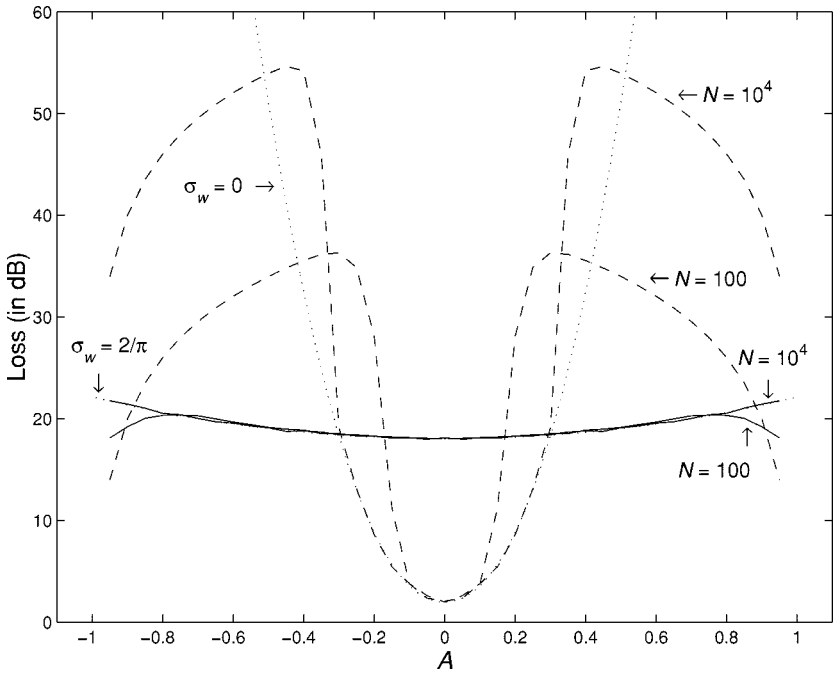

Fig. 8. MSE loss from Monte Carlo simulations for a system comprising a Gaussian control input (pseudonoise in the simulations), a two-level quantizer, and the ML estimator (41)–(43) for $\Delta = 1$, $\sigma_v = 0.1$, and various control input power levels. The dashed curves depict the MSE loss of $\hat{A}_{\rm ML}(\boldsymbol{y}^N;\, \Delta)$ in the absence of control input (i.e., $\sigma_w = 0$); upper curve: $N = 10^4$, lower curve: $N = 100$. The solid curves depict the MSE loss of $\hat{A}_{\rm ML}(\boldsymbol{y}^N;\, \Delta)$ for $\sigma_w = 2/\pi$, and for $N = 100,\ 10^4$. For comparison, the associated information loss functions are depicted by the dotted curves (also shown in Fig. 3).

found in (45) at the bottom of this page, where it is initialized with $\hat{A}_{\rm EM}^{(0)} = 0$. Provided that the log-likelihood function does not possess multiple local minima, (45) provides the ML estimate (40), i.e.,

$$\hat{A}_{\rm ML}(\boldsymbol{y}^N;\, \Delta) = \lim_{k\to\infty} \hat{A}_{\rm EM}^{(k)}.$$

Empirical evidence suggests that $\lim_{k\to\infty} \hat{A}_{\rm EM}^{(k)}$ obtained via the algorithm (45) is asymptotically efficient, i.e., it achieves (12) for large N. Consequently, use of information loss as an accurate predictor of the MSE loss is also justified in this scenario.

3) Efficient Estimation for Signal Quantizers with $M > 2$ in i.i.d. Noise: In general, there is no computationally efficient method for obtaining the ML estimate (40) of A in non-Gaussian noise via a signal quantizer with $M > 2$. In this section, we present an alternative class of elementary estimators which can be shown to be asymptotically efficient for any admissible noise pdf $p_\alpha(\cdot)$, in the sense that for any $|A| < \Delta$ the MSE of the estimator approaches the bound (12) for large N.

Without loss of generality we may view the output of the quantizer $F(\cdot)$ in (2a) and (2b) as the collection of the outputs of $M-1$ two-level quantizers generating the following observed sequences:

$$y_i[n] = \operatorname{sgn}(x[n] - X_i), \qquad i = 1, 2, \ldots, M-1$$

where $x[n] = s[n] + \alpha[n]$ (cf. Fig. 2) and the X_i's are the thresholds of the quantizer. Consider the ML estimates of A formed from each of these binary sequences, namely

$$\hat{A}_i = \mathcal{I}_\Delta(\hat{A}_{\rm ML}(\boldsymbol{y}_i^N;\, \infty) + X_i), \qquad i = 1, 2, \ldots, M-1 \tag{46}$$

where

$$\boldsymbol{y}_i^N \triangleq [\, y_i[1] \quad y_i[2] \quad \cdots \quad y_i[N] \,]^T$$

and where $\mathcal{I}_\Delta(\cdot)$ is given by (42), and $\hat{A}_{\rm ML}(\cdot;\, \infty)$ is given by (43) with α replaced by v. In Appendix F, we show that the joint cumulative distribution of

$$\hat{\boldsymbol{A}} = [\, \hat{A}_1 \quad \hat{A}_2 \quad \cdots \quad \hat{A}_{M-1} \,]^T \tag{47}$$

approaches the cumulative distribution of a Gaussian random vector with mean $A\mathbf{1}$ (where $\mathbf{1}$ denotes a vector of 1's) and covariance matrix C/N, whose inverse is given by (91). We also show in the appendix that if we use

$$\hat{A} = (\mathbf{1}^T \hat{C}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \hat{C}^{-1} \hat{\boldsymbol{A}} \tag{48}$$

$$\hat{A}_{\rm EM}^{(k+1)} = \mathcal{I}_\Delta \left(\hat{A}_{\rm EM}^{(k)} + \frac{\sigma_\alpha}{\sqrt{2\pi} N} \sum_{m=1}^{M} \mathcal{K}_{Y_m}(\boldsymbol{y}^N) \frac{\exp\left(-\frac{\left(X_{m-1} - \hat{A}_{\rm EM}^{(k)}\right)^2}{2\sigma_\alpha^2} \right) - \exp\left(-\frac{\left(X_m - \hat{A}_{\rm EM}^{(k)}\right)^2}{2\sigma_\alpha^2} \right)}{Q\left(\frac{X_{m-1} - \hat{A}_{\rm EM}^{(k)}}{\sigma_\alpha} \right) - Q\left(\frac{X_m - \hat{A}_{\rm EM}^{(k)}}{\sigma_\alpha} \right)} \right). \tag{45}$$

where $\hat{C} = C(\hat{A}_i)$ for some $1 \le i \le M-1$, the estimator $\hat{A}$ is asymptotically efficient, i.e.,

$$\lim_{N\to\infty} NE[(\hat{A}-A)^2;\, A] = \mathcal{B}(A;\, y) \tag{49}$$

where $\mathcal{B}(A;\, y)$ is given by (12). In practice, in computing $\hat{C}$ we may select the value of i for which $\mathcal{B}(\hat{A}_i;\, \boldsymbol{y}_i^N)$ is minimum, so as to expedite the MSE convergence to the asymptotic performance predicted by (49). In summary, the estimator first obtains the set (47) by means of (46) as well as (42) and (43), it then selects the value of i for which $\mathcal{B}(\hat{A}_i;\, \boldsymbol{y}_i^N)$ is minimized and forms $\hat{C} = C(\hat{A}_i)$, and, finally, substitutes $\hat{A}_i$ and $\hat{C}$ in (48) to obtain the asymptotically efficient estimate $\hat{A}$.

B. Known Control Inputs

In this section, we construct estimators that exploit detailed knowledge of the applied control waveform. In particular, in the context of K-periodic control inputs that are known for estimation, we develop estimators that are asymptotically efficient in the sense that they asymptotically achieve (24).

For i.i.d. Gaussian sensor noise, the ML estimate of A from $\boldsymbol{y}^N$ given a control vector $\boldsymbol{w}^N$, where $w[n]$ is a K-periodic sequence and N is a multiple of K, can be obtained as a special case of the EM algorithm presented in Appendix E. In particular, the EM algorithm takes the form of (50a) and (50b) at the bottom of this page, where $\overline{N} = N/K$, and $\boldsymbol{y}^{\overline{N}}[\ell]$ is the $\overline{N}\times 1$ vector comprised of the elements of the ℓth K-decimated subsequence, i.e.,

$$\boldsymbol{y}^{\overline{N}}[\ell] \triangleq [\,y[\ell] \quad y[K+\ell] \quad \cdots \quad y[N-K+\ell]\,]^T, \qquad \ell = 1,\, 2,\, \ldots,\, K. \tag{51}$$

Empirical evidence suggests that the estimate resulting from the EM algorithm (50) is asymptotically efficient, i.e., it achieves the Cramér–Rao bound (25) for large enough N.

Asymptotically efficient estimators in the context of non-Gaussian sensor noises can be obtained in a fashion similar to those developed in Appendix F. Specifically, in the case $M = 2$, we may consider the vector $\hat{\boldsymbol{A}}$ in (47) where we use for $\hat{A}_i$ the ML estimate of A given the ith K-decimated subsequence from (51), i.e.,

$$\hat{A}_i = \mathcal{I}_\Delta(\hat{A}_{\mathrm{ML}}(\boldsymbol{y}^{\overline{N}}[i];\, \infty) - w[i]), \qquad i = 1,\, 2,\, \ldots,\, K \tag{52}$$

and where $\mathcal{I}_\Delta(\cdot)$ and $\hat{A}_{\mathrm{ML}}(\cdot;\, \infty)$ are given by (42) and (43), respectively. The $\hat{A}_i$'s from (52) are independent random variables, since for any $i \ne j$, $\boldsymbol{y}^{\overline{N}}[i]$ and $\boldsymbol{y}^{\overline{N}}[j]$ are independent random vectors. Therefore, the corresponding vector $\hat{\boldsymbol{A}}$ from (47) is asymptotically Gaussian (in terms of its cumulative distribution), with diagonal covariance matrix C/N; the $(i,\, i)$th entry of the matrix C equals $\mathcal{B}(A + w[i];\, y[i])$, where $\mathcal{B}(A;\, y)$ is given by (12) with α replaced by v. Consequently, an asymptotically efficient estimate is provided by $\hat{A}$ from (48); the estimate covariance matrix that is used for faster MSE convergence to the asymptotic performance is given by $\hat{C} = C(\hat{A}_i)$ where i is the index that minimizes $\mathcal{B}(\hat{A}_i + w[i];\, \boldsymbol{y}^{\overline{N}}[i])$.

Asymptotically efficient estimators can also be constructed for signal quantizers with $M > 2$ and known K-periodic inputs in non-Gaussian sensor noise. Specifically, for each M-ary subsequence $\boldsymbol{y}^{\overline{N}}[\ell]$ from (51) we may first apply the algorithm (46)–(48) to obtain K statistically independent estimates of A. By combining these K estimates in a fashion similar to the method used in the case $M = 2$ for combining the estimates (52), we obtain an asymptotically efficient estimator of A based on $\boldsymbol{y}^N$ given $\boldsymbol{w}^N$.

C. Control Inputs in the Presence of Feedback

In Section III-C, we have shown that the worst case information loss of a system composed of a signal quantizer and an additive control input is lower-bounded by the minimum possible information loss of the same system in the control-free case. In this section, we develop control input selection strategies based on past quantized output samples and construct associated estimators which effectively achieve this bound.

1) Feedback Control and Estimation for Signal Quantizers with $M = 2$: We first examine the Gaussian sensor noise scenario with $M = 2$ in detail. As (39) reveals, the associated control-free information loss is minimized for $w[n] = -A$. Although this control input selection is not permissible, it suggests a viable control input selection method based on past quantized observations. Specifically, if $\hat{A}[n]$ is any consistent estimator of A based on $\boldsymbol{y}^n$, a reasonable choice for the control input sequence is as follows:

$$w[n] = -\hat{A}[n-1]. \tag{53}$$

Assuming the control sequence is selected according to (53), the ML estimator at time n satisfies

$$\hat{A}_{\mathrm{ML}}[n] = \arg\max_{|\theta|\le\Delta} \sum_{m=1}^{n} \ln Q(y[m](\hat{A}_{\mathrm{ML}}[m-1] - \theta)).$$

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \mathcal{I}_\Delta\left(\hat{A}_{\mathrm{EM}}^{(k)} + \sum_{\substack{1\le\ell\le K\\ 1\le m\le M}} \frac{\sigma_v \mathcal{K}_{Y_m}(\boldsymbol{y}^{\overline{N}}[\ell])}{\sqrt{2\pi}N} \frac{\exp\left(-\frac{\left(X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}-w[\ell]\right)^2}{2\sigma_v^2}\right) - \exp\left(-\frac{\left(X_m-\hat{A}_{\mathrm{EM}}^{(k)}-w[\ell]\right)^2}{2\sigma_v^2}\right)}{Q\left(\frac{X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}-w[\ell]}{\sigma_v}\right) - Q\left(\frac{X_m-\hat{A}_{\mathrm{EM}}^{(k)}-w[\ell]}{\sigma_v}\right)}\right) \tag{50a}$$

and

$$\hat{A}_{\mathrm{ML}} = \lim_{k\to\infty} \hat{A}_{\mathrm{EM}}^{(k)}. \tag{50b}$$

In Appendix E, we show that in the Gaussian scenario, the ML estimate of A based on $\boldsymbol{y}^n$ for $n = 1, 2, \ldots$ can be obtained using the following EM algorithm:

$$\hat{A}_{\mathrm{EM}}^{(k+1)}[n] = \mathcal{I}_\Delta \left(\hat{A}_{\mathrm{EM}}^{(k)}[n] + \frac{\sigma_v}{\sqrt{2\pi} n} \sum_{m=1}^{n} y[m] \frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{ML}}[m-1] - \hat{A}_{\mathrm{EM}}^{(k)}[n]\right)^2}{2\sigma_v^2} \right)}{Q\left(y[m] \frac{\hat{A}_{\mathrm{ML}}[m-1] - \hat{A}_{\mathrm{EM}}^{(k)}[n]}{\sigma_v} \right)} \right) \tag{54a}$$

initialized with $\hat{A}_{\mathrm{EM}}^{(0)}[n] = \hat{A}_{\mathrm{ML}}[n-1]$ and $\hat{A}_{\mathrm{ML}}[0] = 0$, where for any n

$$\hat{A}_{\mathrm{ML}}[n] = \lim_{k \to \infty} \hat{A}_{\mathrm{EM}}^{(k)}[n]. \tag{54b}$$

Although empirical evidence suggests that the ML estimator obtained by means of the EM algorithm in (54a) and (54b) achieves the 2-dB information loss bound (39) for any A in $(-\Delta, \Delta)$ for a moderate number of observations,[8] it is rather computationally intensive; for any additional observed sample an EM algorithm has to be employed. In addition, even though the number of iterations necessary for adequate convergence of the EM algorithm appears to be small for large n, the algorithm may still be impractical.

We next develop algorithms that achieve the bound (39) and have the additional advantage that they can be implemented very efficiently. These are based on the observation that once the estimate $\hat{A}[n-1]$ is not changing significantly with n (i.e., the changes are small with respect to σ_v) we may assume that $s[n] - \hat{A}[n-1]$ is in the regime where the information loss is small, and a low-complexity estimator can be constructed that approaches the 2-dB bound (39). Specifically, let $z = Q(A/\sigma_v)$ and assume that $|A/\sigma_v| < 0.1$. In this regime, the truncated power series expansion provides a reasonable approximation for $Q^{-1}(z)$, i.e.,

$$Q^{-1}(z) \approx \sqrt{\frac{\pi}{2}} (1 - 2z). \tag{55}$$

We can use (55) to form a linear estimator as follows. Assuming that the estimation error is inversely proportional to the measurements (which, for admissible sensor noises, implies that the asymptotic MSE loss is not infinite), the estimate at time n is given as a weighted sum of the estimate at time $n-1$ and an estimate arising from using the nth measurement $y[n]$ alone, i.e.,

$$\hat{A}_{\mathrm{L}}[n] = \frac{n-1}{n} \hat{A}_{\mathrm{L}}[n] + \frac{1}{n} \hat{A}[n|y[n]] \tag{56}$$

[8]There are a number of other control input selection methods and associated estimators which can approach arbitrarily close to the 2-dB bound; the systems developed in this paper for the case $M > 2$ and non-Gaussian noise are such an example. However, the associated MSE of these algorithms converges to the bound (39) considerably slower than the algorithms of this section. In fact, the number of samples required so that the MSE of (54a) and (54b) with $w[n]$ as in (53) effectively achieves the 2-dB bound (39) increases linearly with $\ln(\chi)$.

where the estimate based on the nth measurement alone is given by using (55) in (54), (54b) (by setting σ_w to 0), and the fact that $w[n] = -\hat{A}_{\mathrm{L}}[n-1]$, i.e.,

$$\hat{A}[n|y[n]] = \hat{A}_{\mathrm{L}}[n-1] + \sigma_v \sqrt{\frac{\pi}{2}}\, y[n]. \tag{57}$$

By incorporating (57) in (56) this linear estimator takes the following iterative form:

$$\hat{A}_{\mathrm{L}}[n] = \hat{A}_{\mathrm{L}}[n-1] + \sigma_v \sqrt{\frac{\pi}{2}}\, \frac{y[n]}{n}. \tag{58}$$

In order to obtain an algorithm that converges much faster than (58) to the 2-dB bound (39), we employ the EM algorithm (54a), (54b) for $n \le n_o$ and the recursive algorithm (58) for $n > n_o$, i.e.,

$$\check{A}[n] = \begin{cases} \hat{A}_{\mathrm{ML}}[n] \text{ from (54)}, & \text{if } n \le n_o \\ \mathcal{I}_\Delta \left(\check{A}[n-1] + \sigma_v \sqrt{\frac{\pi}{2}}\, \frac{y[n]}{n} \right), & \text{if } n > n_o \end{cases} \tag{59}$$

where the control input $w[n]$ is given by (53) provided that we substitute $\check{A}[n-1]$ for $\hat{A}[n-1]$, and where we also incorporated the dynamic range information by means of $\mathcal{I}_\Delta(\cdot)$.

Selection of an appropriate value for n_o is related to the peak SNR χ. Since, in principle, the larger the peak SNR, the longer (in terms of the number of observations) it takes $A - \hat{A}_{\mathrm{ML}}[n]$ to reach the linear regime (55), we consider the case $\Delta \gg \sigma_v$. For instance, assume we are interested in selecting n_o so that the $\sqrt{\mathrm{MSE}}$ in $\check{A}[n_o]$ is less than a given fraction of σ_v (so that the truncated series approximation is valid), for example, $\sigma_v/8$. For small enough n_o, the maximum MSE from n_o observations is roughly given as the square of $\Delta 2^{-n_o}$. In summary, this crude-MSE-based rule of thumb for selecting n_o reduces to $n_o \ge \log_2(\Delta/\sigma_v) + 3$.

The solid and dashed curves in Fig. 9 depict the MSE of the ML estimator obtained by means of the EM algorithm in (54a), (54b) and of the computationally efficient estimator (59) with $n_o = 10$, respectively, based on Monte Carlo simulations. The system parameters for this simulation are $\Delta = 1$, $\sigma_v = 0.1$, resulting in $\log_2(\Delta/\sigma_v) \approx 6.6$, while $A = 0.4$. In both cases, the control sequence is selected according to (53). The lower and upper dotted lines depict $\mathcal{B}(A; \boldsymbol{s}^N)$ and the right-hand side of (38), respectively. As we can see in this figure, both estimates effectively achieve the 2-dB loss bound (39) for a moderate number of observations.

In terms of the actual implementation of the estimator (59), for a given n_o there are 2^{n_o} possible values of $\hat{A}_{\mathrm{ML}}[n_o]$. These 2^{n_o} estimate values can be precomputed and stored in a lookup table. This results in a computationally efficient implementation, whereby given n_o or fewer observations the estimate is obtained from a lookup table, while once the number of observations exceeds n_o, a recursive linear estimator is employed. Since n_o grows logarithmically with χ, the number of lookup table entries for storing all possible values of $\hat{A}_{\mathrm{ML}}[n_o]$ grows only linearly with peak SNR χ.

A similar strategy can be used in the context of quantizer systems using feedback in any sensor noise. In the general case, A_* in (36) may not equal zero. A reasonable extension of the control input selection method (53) for nonzero A_* is as follows:

$$w[n] = A_* - \hat{A}[n-1]. \tag{60}$$

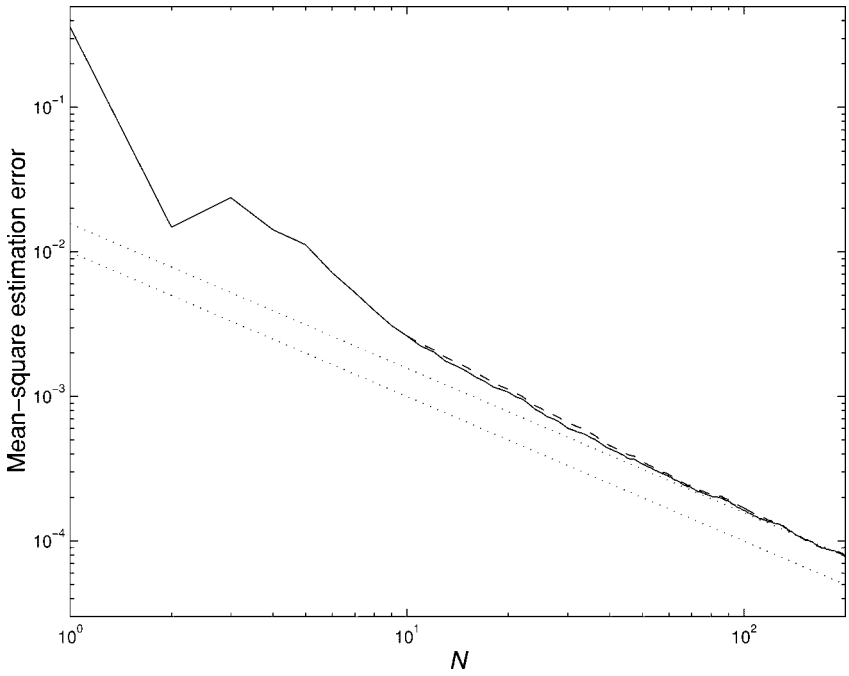

Fig. 9. MSE from Monte Carlo simulations for $\hat{A}_{\mathrm{ML}}[n]$ (solid) and $\check{A}[n]$ with $n_o = 10$ (dashed), based on observations from a signal quantizer with $M = 2$ exploiting feedback according to (53). The lower dotted line represents the Cramér–Rao bound for estimating A based on $s[n]$, while the upper dotted line is the 2-dB bound (39); Parameters: $\sigma_v = 0.1$, $\Delta = 1$, and $A = 0.4$.

An estimator similar to (59) can be used to estimate A in this case. Specifically, for $n \leq n_o$ the estimator may consist of a precomputed lookup table, while for $n > n_o$ a recursive estimator resulting from a truncated series expansion of $C_v^{-1}(z)$ around $z = A_*$ can be employed, namely

$$\check{A}[n] = \mathcal{I}_\Delta\left(\check{A}[n-1] + \frac{1}{n}\,\frac{y[n] + 1 - 2C_v(-A_*)}{2p_v(-A_*)}\right).$$

In particular, if A_* is the median of $p_v(\cdot)$, in which case $\mathcal{L}(A_*)$ is given by (37), we have

$$\check{A}[n] = \mathcal{I}_\Delta\left(\check{A}[n-1] + \frac{y[n]}{2np_v(-A_*)}\right), \qquad \text{for } n > n_o.$$

In general, empirical evidence suggests that the MSE loss of these algorithms practically achieves the associated $\mathcal{L}(A_*)$ for a moderate number of observations.

2) Feedback Control and Estimation for Signal Quantizers with $M > 2$: For the Gaussian sensor noise scenario, the EM algorithm (54a), (54b) can be extended to $F(\cdot)$ with $M > 2$; the resulting algorithm is a special case of the one presented in Appendix E. Empirical evidence suggests that it is also asymptotically efficient. Assuming flexibility in selecting the thresholds of the M-level quantizer, the corresponding information loss (36) can be obtained from Fig. 6. For instance, for the optimal selection of the quantizer thresholds for $M = 6$ we have $A_* = 0$; if the control input is selected according to (60), the EM algorithm in Appendix E yields a worst case MSE loss of about 0.25 dB. Similarly to $\mathcal{L}_{\max}$, the asymptotic MSE loss is independent of σ_v and Δ.

For signal quantizers with $M > 2$, where $v[n]$ is any non-Gaussian noise, we may use the following two-stage approach that effectively achieves $\mathcal{L}(A_*)$. For the first N_1 observations, we may employ any consistent estimator $\hat{A}_1[n]$ of A. For instance, we may use one of the feedback-based algorithms corresponding to the system $M = 2$ by ignoring all but two of the M levels of the quantized output. In the second stage, we fix $w[n] = A_* - \hat{A}_1[N_1]$ for all $n > N_1$. The number N_1 determines the accuracy of the approximation

$$\mathcal{L}(A_* + A - \hat{A}_1[N_1]) \approx \mathcal{L}(A_*).$$

For any given $n > N_1$, we can then obtain an estimate $\hat{A}_2[n]$ of A from

$$[\,y[N_1+1] \quad y[N_1+2] \quad \cdots \quad y[n]\,]^T$$

by means of (46)–(48), which is asymptotically efficient with respect to $\mathcal{L}(A_* + A - \hat{A}_1[N_1])$. For faster convergence, the overall estimate can be a weighed sum of the estimates $\hat{A}_1[N_1]$ and $\hat{A}_2[n]$. Although the associated asymptotic MSE loss can be made to approach $\mathcal{L}(A_*)$ arbitrarily closely, these algorithms typically require significantly larger data sets to effectively achieve the desired information loss, as compared to the algorithms for $M = 2$ of the previous section.

The estimators developed in this section possess close connections to a well-known class of oversampled analog-to-digital (A/D) converters. Specifically, for large Δ/σ_v and for small n, these estimators can be effectively viewed as generalized versions of successive approximation A/D converters, which take into account the sensor noise characteristics. Unlike the conventional successive-approximation A/D converters, once the successive approximation step is small enough as compared to the noise level (i.e., for large enough n), these estimators also incorporate the noise characteristics in obtaining their running estimate.

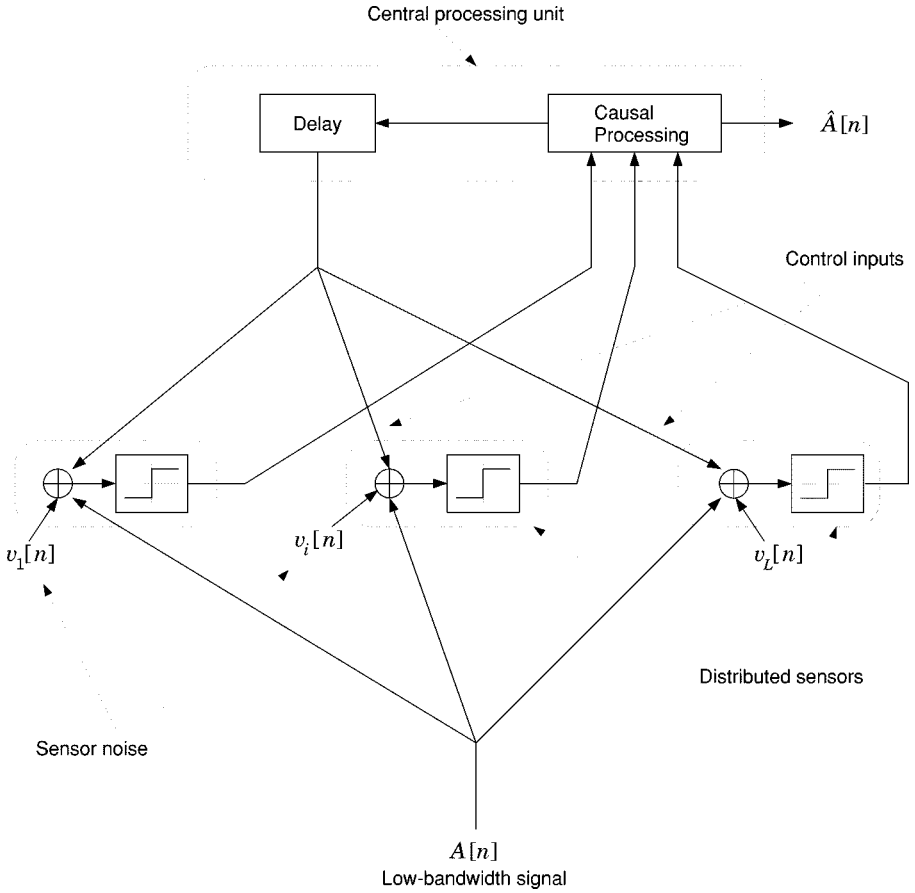

Fig. 10. Block diagram of a network of distributed signal quantizers using feedback in the context of signal estimation.

V. Multiple Sensors

In this section, we examine a network generalization of the estimation problem (1), (2a), and (2b), namely, estimating an unknown parameter A from observation of

$$y_i[n] = F(A + v_i[n] + w_i[n])$$

for $n = 1, 2, \ldots, N$, and $i = 1, 2, \ldots, L$, and where $F(\cdot)$ is given by (2a), (2b), the $v_i[n]$'s are i.i.d. processes, and the $w_i[n]$'s denote the applied control input sequences. For simplicity, we assume that the noise processes are independent of one another.[9] Such networks may provide reasonably accurate models for a number of distributed estimation problems that involve possibly remote sensors that are not colocated. In Fig. 10, for instance, we show the block diagram of a special case of such distributed estimation network, which uses feedback in the selection of the control inputs; however, distributed estimation networks without feedback are also considered.

Straightforward extensions of the single-sensor systems developed in Sections III and IV yield network generalizations that can be analyzed by means of the tools developed for the single-sensor case. For the remainder of this section, we restrict our attention to two-level quantizers in i.i.d. Gaussian sensor noise, which we use as a representative example to illustrate the extensions of the single-sensor results to the associated multisensor settings. Similar extensions can be derived for all the other scenarios we developed in Sections III and IV.

A. Random Control Inputs

We may consider, for instance, a network of signal quantizers for which the control inputs are i.i.d. sequences with known statistical description, such that $w_i[n] \sim \mathcal{N}(0, \sigma_w^2)$, and which can be adequately modeled as statistically independent of one another and of the sensor noises. In the case that all sensor noises have equal strength σ_v^2, the collection of L $N \times 1$ observation vectors $\{\boldsymbol{y}_i^N\}$ can be viewed as a single $(NL) \times 1$ observation vector collected from a single sensor, in which case all the analysis of Sections III-A and IV-A applies.

If the overall noise levels (summarizing the effects of the sensor noise and the random control component) have variable strengths, Cramér–Rao bounds and corresponding ML estimators can be formed with minor modifications of the single-sensor problem. In general, for random control inputs at high peak SNR (i.e., $\Delta \gg \sigma_{v_i}$ for $i = 1, 2, \ldots, L$) the worst case information loss still grows quadratically with dynamic range for any fixed-length network with a fixed number of quantization levels and sensor noise components with fixed power levels.

B. Control Inputs in the Presence of Feedback

Networks exploiting feedback from the quantized outputs to the control inputs can also be analyzed using the associated

[9]The most interesting case in which the sensor noises are correlated conditioned in the signal component is not considered in this paper; the case $L = 2$ with correlated sensor noises is considered in [26] in the context of encodings in the form of vector quantization.

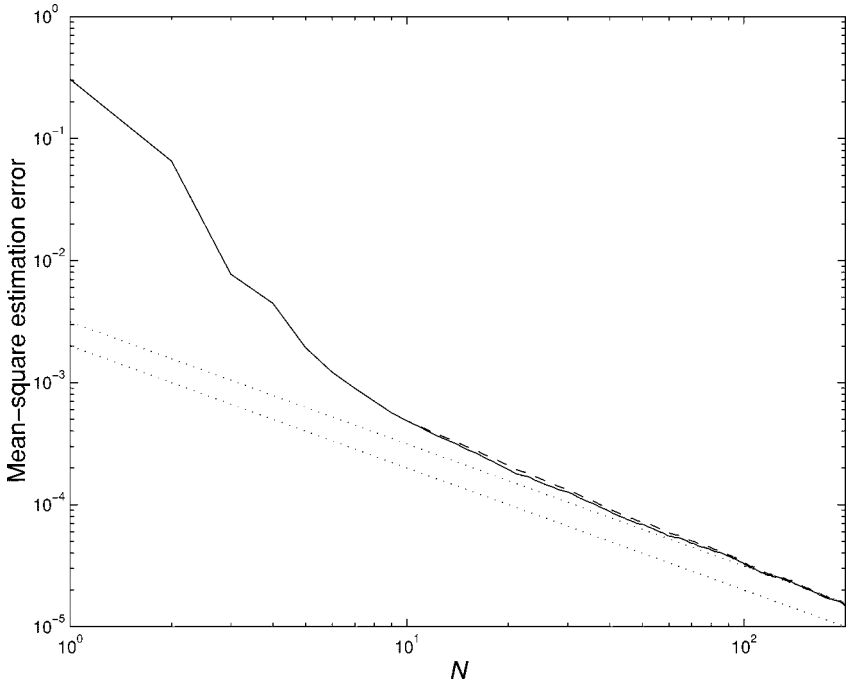

Fig. 11. MSE for $\hat{A}_{\mathrm{ML}}[N]$ and $\check{A}[N]$ for a network of $L = 5$ two-level quantizers, using feedback in the selection of the control input, and associated Cramér–Rao bounds (see also caption of Fig. 9). The sensor noise levels are 0.08, 0.08 0.08, 0.2, and 0.4, while $A = 0.4$ and $\Delta = 1$.

single-sensor principles. Specifically, the control input can be selected using (53), where $\hat{A}[n-1]$ denotes the estimate of A based on observations collected from all L sensors up to and including time $n-1$. It is worth noting that this type of feedback is analogous to case explored in [28] in the context of decentralized detection.

In Fig. 11, we show the MSE performance of the network extensions of the ML estimator (54a), (54b) and $\check{A}[N]$ given by (59) for a network of $L = 5$ sensors and where the sensor noises are independent i.i.d. Gaussian random processes with spatially nonuniform sensor noise power levels. As in the single-sensor case, the MSE of each estimator practically achieves the bound corresponding to a 2-dB information loss for moderate N.

In the Gaussian scenario, for networks of two-level signal quantizers with feedback, the associated information loss can be directly obtained using appropriate interpretation of Fig. 7 describing the single-sensor case. Similar extensions of the associated single-sensor problem can be obtained for any set of sensor noises for $M \geq 2$. For instance, if feedback is available and properly used in the multisensor setting shown in Fig. 10, a small worst case information (and MSE) loss can be achieved, independent of the dynamic range and the noise power levels.

There exist certain interesting connections between our strategies and the one considered in [30]. In [30] Gray *et al.* consider the estimator arising for $n = 1$ in the limit $L \to \infty$ and in the case where both A and the sensor noises are both nonnegative. For a given fixed $L = N^2$, the authors construct a sequence of quantizers $F_i(\cdot)$ for $i = 1, 2, \ldots, L$ with $X_i = i/\sqrt{L}$ and employ it in conjunction with a simple decoding rule whose MSE is shown to tend to zero as $L \to \infty$. The resulting decoder exploits only knowledge of $E(v_i[n])$, unlike our strategies which exploit knowledge of $F_v(\cdot)$. The strategy used in [30] is analogous to one of our strategies, namely, one using symmetric binary quantizers (i.e., with $X_i = 0$) with known control inputs, where $w_i[1]$ denotes one period of a sawtooth waveform (in i) spanning $[1/\sqrt{L}, \sqrt{L}]$ with spacing $1/\sqrt{L}$. It can easily be verified that the resulting encoded sequences are the same in both cases. Consistent with our analysis, our estimator (a variant of the one described in Appendix E) would exploit knowledge of the $w_i[n]$ but also of $p_v(\cdot)$ to also result in MSE that tends to 0 as $L \to \infty$. In fact, $\lim_{L\to\infty} \mathrm{MSE} = 0$ is also achieved with properly designed random control inputs $w_i[1]$ in i (with a simple variation of our algorithm in Section V-A); however, the information loss in that case is greater, implying that a larger L would be required to achieve the same MSE performance as we would with a sequence of known sawtooth (in i) waveform.

VI. Summary and Concluding Remarks

In this paper, we have examined the problem of parameter estimation based on observations from an M-level quantizer in the context of additive controlled perturbation of the quantizer thresholds. We have developed a methodology for evaluating these sequential quantization-based systems by means of a figure of merit which we refer to as the information loss; it is defined as the increase in decibels that is incurred in the Cramér–Rao bound for unbiased estimates by a particular combination of control input and M-level quantizer. In general, for control-free systems the performance rapidly degrades with peak SNR χ, where χ is defined as the ratio of the parameter dynamic range Δ to the sensor noise power level σ_v. In particular, for a wide class of i.i.d. sensor noises, the worst case information loss grows faster than χ^2 if no control input is used.

A number of important scenarios may arise in practice which differ in terms of the available knowledge about the control waveform for estimation and the associated freedom in the control input selection. For scenarios where only the statistical characterization of the control input can be exploited for estimation, we have shown that random control inputs can provide significant performance benefits, in the sense that the worst case information loss grows quadratically with peak SNR. If knowledge of the particular control input is exploited for estimation, even higher performance can be achieved. In particular, we developed methods for selecting the control input from a suitably designed class of periodic waveforms, for which the worst case information loss grows linearly with peak SNR. Finally, for cases where feedback is available we developed control waveform selection strategies and corresponding computationally efficient estimators that asymptotically achieve the best possible performance for quantizer-based systems with additive control inputs. Specifically, these estimators achieve the minimum possible information loss for the associated quantizer-based system, which is independent of peak SNR. We should emphasize that the preceding discussion applies to any M-level quantizer and a wide class of i.i.d. sensor noises. Furthermore, our methodology can be generalized to scenarios involving networks of these quantizer-based systems.

Complementary to this work, performance analysis and system design in [34] reveals the performance rates in Table I remain unaffected if system design is based on *average* rather worst case performance criteria, when A is a normally distributed random variable. In addition, in this work we have assumed that the sequence used at the host for estimation is the same as the one encoded at the sensor. The case where the sensor encodings are observed at the host through a discrete memoryless channel (DMC) is examined in [42], where it is shown that, given any fixed (nonpathological) DMC, the performance rates in Table I also remain unaffected. As is also shown in [42], however, the performance-optimizing encoders and estimators in each case do depend on the DMC quality.

This preliminary work suggests a number of important and practical extensions for further investigation. As one example, it is important to study the performance that is achievable based on (short) finite-length observation windows. Such analysis can potentially be beneficial for a number of applications involving signal quantizers, such as various distributed estimation problems that arise in practical settings, reliable sensing of subthreshold signals such as those encountered in cellular systems, and design of oversampled A/D converters and coupled A/D converter arrays. In the context of A/D conversion in particular, it is noteworthy that, besides the interesting connections between the feedback schemes to successive approximation A/D converters, all the control input encoding techniques we have considered have similarities to nonsubtractive and subtractive dithering techniques [32], [33], [38], [43], [44]. For instance, in both the known control input case and the feedback case, knowledge of the control input is exploited by the estimator and compensated after quantization, although, unlike subtractive dithering, the control input is *not* simply subtracted off the quantized signal, as the associated estimators (49), (53), and (58) reveal.

Appendix A
Worst Case Information Loss for Control-Free Signal Quantizers

In this appendix, we show that the worst case information loss of any signal quantizer grows faster than χ^2 for large χ in the absence of a control input. We first consider the case $M = 2$ and show by contradiction that $\mathcal{L}_{\max}^{\text{free}}(\chi) \not\sim o(\chi^2)$ as $\chi \to \infty$, i.e., we show that

$$\lim_{\chi\to\infty} \frac{\mathcal{L}_{\max}^{\text{free}}(\chi)}{\chi^2} = 0 \tag{61}$$

cannot be true. Letting $\chi \to \infty$ is equivalent to fixing Δ and letting $\sigma_v \to 0^+$, since the control-free information loss for $M = 2$ is completely characterized by χ. Let $\mathcal{B}_{\max}(\Delta, \sigma_v; y)$ denote the worst case Cramér–Rao bound for estimating A from one sample of the i.i.d. sequence $y[n]$, for $|A| < \Delta$, and noise level σ_v. Then, (61) implies that

$$\lim_{\sigma_v\to 0^+} \mathcal{B}_{\max}(\Delta, \sigma_v; y) = 0 \tag{62}$$

where we used (4), (7), and (8). However, (62) suggests that, as $\sigma_v \to 0^+$, we can estimate *any* A in $(-\Delta, \Delta)$ with infinite accuracy from a 1-bit observation $y[n]$, which is not possible. Thus, (61) is false, i.e., $\mathcal{L}_{\max}^{\text{free}}(\chi)$ has to grow at least as fast as χ^2.

Similarly, we can also show that $\mathcal{L}_{\max}^{\text{free}}(\chi)$ grows faster that χ^2, in the sense that $\mathcal{L}_{\max}^{\text{free}}(\chi) \not\sim O(\chi^2)$. We show this by first assuming that $\mathcal{L}_{\max}^{\text{free}}(\chi) \sim O(\chi^2)$, i.e., that we can find $D < \infty$ and χ_o, such that for $\chi > \chi_o$ we have $\mathcal{L}_{\max}^{\text{free}}(\chi) \le D\chi^2$, and arriving to a contradiction. The condition $\mathcal{L}_{\max}^{\text{free}}(\chi) \sim O(\chi^2)$ is equivalent to the statement that there exists $D < \infty$ such that

$$\limsup_{\chi\to\infty} \frac{\mathcal{L}_{\max}^{\text{free}}(\chi)}{\chi^2} = D. \tag{63}$$

Again using (4) as well as (7) and (8) in (63) we obtain the following equivalent statement:

$$\limsup_{\sigma_v\to 0^+} \mathcal{B}_{\max}(\Delta, \sigma_v; y) = D' \tag{64}$$

where $D' < \infty$. Since the sequence $y[n]$ is i.i.d., (64) implies that as $\sigma_v \to 0^+$, the Cramér–Rao bound $\mathcal{B}(A; \boldsymbol{y}^N)$ is upper-bounded by D'/N, which goes to 0 as $N \to \infty$. However, for any $A \neq 0$, in the limit $\sigma_v \to 0^+$ we have $y[n] = \text{sgn}(A)$ with probability 1 for all n, which, in turn, implies that $\mathcal{B}(A; \boldsymbol{y}^N)$ cannot go to 0 as $N \to \infty$. Thus, we must have $D' = \infty$ in (64), which proves that the control-free worst case information loss is not $O(\chi^2)$.

We can show that $\mathcal{L}_{\max}^{\text{free}}(\chi) \not\sim O(\chi^2)$ for signal quantizers with $M > 2$, by using our results for $M = 2$. Specifically, if Δ is fixed, in which case $\chi \to \infty$ is equivalent to $\sigma_v \to 0^+$, the arguments used for the $M = 2$ case still apply with minor modifications. Next consider fixing σ_v, in which case $\chi \to \infty$ is equivalent to $\Delta \to \infty$. As usual, let X_1, X_2, X_{M-1} denote the quantizer thresholds. By rescaling by $1/\Delta$, this problem can be mapped to an equivalent one where $\Delta' = 1$, $\sigma_v' = \sigma_v/\Delta \to 0^+$, and where the new quantizer thresholds are $X_1/\Delta, X_2/\Delta, X_{M-1}/\Delta$. The arguments used to show that $\mathcal{L}_{\max}^{\text{free}}(\chi) \not\sim O(\chi^2)$ in the $M = 2$ case are still valid in this case with minor modifications.

APPENDIX B
WORST CASE INFORMATION LOSS FOR KNOWN CONTROL INPUTS

We first show that for any known control input scenario, the worst case information loss grows at least as fast as χ. This is true for any admissible sensor noise distribution and for any $M \geq 2$. For convenience, we denote by $p_w(\cdot)$ the type (empirical probability distribution) of the known sequence $w[n]$ [45, p. 279]. The associated Cramér–Rao bound for estimating A based on $y[n]$ for a particular $p_w(\cdot)$ is given by

$$\mathcal{B}(A;\, \boldsymbol{y}^N,\, p_w(\cdot)) = \frac{1}{N}\left(E[\{\mathcal{B}(A+w;\, y)\}^{-1}]\right)^{-1} \tag{65}$$

where the expectation is with respect to $p_w(\cdot)$. For instance, if the periodic sequence (27) is represented by an empirical pdf consisting of K Kronecker delta functions located at $w[n]$ for $n = 0, 1, \ldots, K-1$ and each with area $1/K$, then (65) and (25) are equivalent.

For convenience, we consider the inverse of the Cramér–Rao bound in (65), namely, the Fisher information of A given $y[n]$. We denote the Fisher information in the control-free case by $\mathcal{F}(A;\, y)$. The worst case Fisher information $\mathcal{F}_{\min}(\Delta;\, p_w(\cdot))$ for a control input $\boldsymbol{w}^N$ with an empirical pdf $p_w(\cdot)$ is defined as

$$\mathcal{F}_{\min}(\Delta;\, p_w(\cdot)) \triangleq \inf_{|A|<\Delta} E[\mathcal{F}(A+w;\, y)]$$

where the expectation is with respect to $p_w(\cdot)$. Consider the optimal selection of $p_w(\cdot)$, which results in maximizing $\mathcal{F}_{\min}(\Delta;\, p_w(\cdot))$, i.e.,

$$\mathcal{F}_{\text{opt}}(\Delta) \triangleq \max_{p_w(\cdot)} \mathcal{F}_{\min}(\Delta;\, p_w(\cdot)).$$

The growth of the optimal worst case (i.e., the minimized maximum) information loss equals the decrease of the inverse of the optimal worst case (i.e., the maximized minimum) Fisher information defined above.

We will make use of the fact that the control-free worst case information loss grows strictly faster than χ^ρ for $\rho < 2$ [cf. the generalization of (63) to quantizers with $M \geq 2$]. Without loss of generality, we may set $\sigma_v = 1$, in which case $\Delta = \chi$. Since $\mathcal{B}(A;\, s)$ is independent of A (and thus Δ), the control-free worst case Fisher information of A based on $y[n]$ decays faster than $1/\Delta^\rho$ with increasing Δ for any $\rho < 2$. Thus, there exist $D > 0$ and $\delta > 0$, such that for any $|A| > \delta$

$$\mathcal{F}(A;\, y) = [\mathcal{B}(A;\, y)]^{-1} < \min\{D|A|^{-\rho},\, [\mathcal{B}(A;\, s)]^{-1}\} \tag{66}$$

for any given $\rho < 2$. For convenience, we pick ρ so that $1 < \rho < 2$. Also, let

$$\mathcal{P}_k(A;\, p_w(\cdot)) \triangleq \int_{k\delta \leq |w+A| < (k+1)\delta} p_w(w)\, dw.$$

For any empirical pdf $p_w(\cdot)$ and any Δ satisfying $\Delta > \delta$, we must have

$$\inf_{|A|<\Delta} \mathcal{P}_k(A;\, p_w(\cdot)) < \frac{2\delta}{\Delta}. \tag{67}$$

We can establish (67) via proof by contradiction; if the inequality in (67) is reversed, for any A in $(-\Delta,\, \Delta)$ we have

$$\mathcal{P}_k(A;\, p_w(\cdot)) \geq \frac{2\delta}{\Delta}. \tag{68}$$

Let $A_j = j\delta$ for $j = 0, \pm 1, \ldots, \pm j_o$, where j_o is the largest index j satisfying $A_j < \Delta$. Note that $j_o > \Delta/(2\,\delta)$. Applying (68) for $A = A_j$, and summing over all j yields

$$\sum_{j=-j_o}^{j_o} \mathcal{P}_k(A;\, p_w(\cdot)) \geq (2j_o+1)\frac{2\delta}{\Delta} \tag{69}$$

which is a contradiction since the left-hand size of (69) is upper-bounded by $2\int_w p_w(w)\, dw$, while $(2j_o+1)(2\delta)/\Delta > 2$. We can similarly derive the following generalization of (67):

$$\inf_{|A|<\Delta} \sum_k \beta_k \mathcal{P}_k(A;\, p_w(\cdot)) < \frac{2\delta}{\Delta} \sum_k \beta_k \tag{70}$$

where $\beta_k \geq 0$ and at least one of the β_k's is nonzero. We have

$$\begin{aligned}
\mathcal{F}_{\text{opt}}(\Delta) &= \max_{p_w(\cdot)} \inf_{|A|<\Delta} \sum_{k=0}^{\infty} \int_{k\delta \leq |w+A| < (k+1)\delta} p_w(w) \\
&\qquad \cdot \mathcal{F}(A+w;\, y)\, dw \\
&< \max_{p_w(\cdot)} \inf_{|A|<\Delta} \left([\mathcal{B}(0;\, s)]^{-1} \mathcal{P}_0(A;\, p_w(\cdot)) \right. \\
&\qquad \left. + \sum_{k=1}^{\infty} \frac{D}{(k\delta)^\rho}\, \mathcal{P}_k(A;\, p_w(\cdot)) \right) && \text{(71a)} \\
&< \frac{2\delta}{\Delta} \left([\mathcal{B}(0;\, s)]^{-1} + \frac{D}{\delta^\rho} \sum_{k=1}^{\infty} \frac{1}{k^\rho} \right) && \text{(71b)} \\
&\leq \frac{C\delta}{\Delta} && \text{(71c)}
\end{aligned}$$

where $C < \infty$, since $\sum_{k=1}^{\infty} k^{-\rho}$ is a convergent series for $\rho > 1$. To obtain (71a) and (71b) we used (66) and (70), respectively. As (71c) reveals, for large Δ the optimal worst case information loss grows at least as fast as χ (since $\chi = \Delta$ for $\sigma_v = 1$).

We next show that simple periodic control input schemes can be constructed for which the worst case information loss (for $N \to \infty$) grows linearly with χ. It suffices to consider signal quantizers with $M = 2$, since signal quantizer with $M > 2$ provide additional information and would thus perform at least as well. In particular, we next show that K-periodic waveforms given by (27), where K is given by (28) for a fixed $\lambda > 0$, achieve the optimal growth rate for any admissible sensor noise and a symmetric two-level quantizer. Let $\overline{\mathcal{B}}(A;\, \sigma_v)$ denote the Cramér–Rao bound (14) with α replaced by v. Note that since

$$\overline{\mathcal{B}}(A;\, \sigma_v) = \sigma_v^2 \overline{\mathcal{B}}(A/\sigma_v;\, 1) \tag{72}$$

we also have

$$\mathcal{B}_{\max}(\Delta;\, \sigma_v) = \sigma_v^2 \mathcal{B}_{\max}(\Delta/\sigma_v;\, 1)$$

which in conjunction with (8) reveals that the associated information loss is completely characterized by the ratio $\chi = \Delta/\sigma_v$. Since K also solely depends on χ, we may fix $\Delta = 1$ without loss of generality. Note that the class (27) remains invariant to changes in σ_v. Hence, we may use $w[n;\, K]$ to denote the

unique K-periodic sequence from the class (27) corresponding to $\Delta = 1$. For $\sigma_v < \lambda$, we have $3\lambda/\sigma_v > K$, and

$$\begin{aligned}
&\mathcal{B}_{\max}(1,\, \sigma_v) \\
&\quad = \sup_{A\in(-1,\,1)} \frac{K}{\sum_{n=1}^{K} [\overline{\mathcal{B}}(A + w[n;\, K];\, \sigma_v)]^{-1}} && \text{(73a)} \\
&\quad < \frac{3\lambda}{\sigma_v} \sup_{A\in(-1,\,1)} \min_{n\in\{1,\,2,\,\ldots,\,K\}} \overline{\mathcal{B}}(A + w[n;\, K];\, \sigma_v) && \text{(73b)} \\
&\quad \le 3\lambda\sigma_v \sup_{A'\in(-1/\sigma_v,\,1/\sigma_v)} \min_{n\in\{1,\,2,\,\ldots,\,K\}} \overline{\mathcal{B}}(A' + w'[n;\, K];\, 1) && \text{(73c)} \\
&\quad \le 3\lambda\sigma_v \sup_{A\in(-1/\lambda,\,1/\lambda)} \overline{\mathcal{B}}(A;\, 1) && \text{(73d)}
\end{aligned}$$

where

$$w'[n;\, K] = w[n;\, K]/\sigma_v$$

and where we used (72) to obtain (73c) from (73b). To verify (73d) from (73c), note that for any fixed A' in $(-1/\sigma_v,\, 1/\sigma_v)$, the minimum of $\overline{\mathcal{B}}(A' + w'[n;\, K];\, 1)$ over n is upper-bounded by $\overline{\mathcal{B}}(A' + w'[n';\, K];\, 1)$, where n' is the value of n for which $|A' + w'[n;\, K]|$ is the smallest. Since the spacing $\delta_{w'}$ of the sawtooth waveform $w'[n;\, K]$ satisfies $\delta_{w'} = \delta_w/\sigma_v \le 2/\lambda$, $|A' + w'[n';\, K]|$ is upper-bounded by $\delta_{w'}/2 \le 1/\lambda$ for any $|A'| < 1/\sigma_v$, verifying (73d). Since $\mathcal{B}(A;\, s) \sim \sigma_v^2$ from (8) and by using (73d), the worst case information loss for known $w[n]$ given by (27) with K given by (28) is inversely proportional to σ_v for small σ_v. Hence, this control selection method achieves the optimal worst case information loss growth rate.

We next determine the optimal λ in (28) for the case where $v[n]$ is Gaussian with variance σ_v^2. We use $\hat{\mathcal{B}}_{\mathcal{N}}(x;\, \chi,\, K)$ to denote the Cramér–Rao bound (25) for $A = x\Delta$ in order to make its dependence on χ and on the period K in (27) explicit. The optimality of (28) suggests that K_{opt} from (26) is a nondecreasing function of χ for large χ. Indeed, there is a sequence χ_k where $k \ge 3$, such that, $K_{\text{opt}}(\chi) = k$, if $\chi_k < \chi < \chi_{k+1}$. If $\chi = \chi_k$, both $K = k$ and $K = k + 1$ minimize (26), i.e.,

$$\sup_{x\in(-1,\,1)} \hat{\mathcal{B}}_{\mathcal{N}}(x;\, \chi_k,\, k) = \sup_{x\in(-1,\,1)} \hat{\mathcal{B}}_{\mathcal{N}}(x;\, \chi_k,\, k+1). \quad (74)$$

For large k the left-hand side of (74) is maximized at $x = 1$ [i.e., $A = \Delta$ in (25)], while the right-hand side is maximized at $x = 1 - d(\chi_k;\, k)/2$ with $d(\cdot;\, \cdot)$ given by (29). Assuming that $d_{\text{opt}}(\chi)$ in (30) converges for large χ to a limit, i.e., that $d_\infty = \lim_{\chi\to\infty} d_{\text{opt}}(\chi)$ exists, (74) reduces to

$$\sum_{n=-1}^{\infty} [\mathcal{B}_{\mathcal{N}}((n + 1/2)d_\infty;\, 1)]^{-1} = \sum_{n=0}^{\infty} [\mathcal{B}_{\mathcal{N}}(nd_\infty;\, 1)]^{-1} \quad (75)$$

where $\mathcal{B}_{\mathcal{N}}(A;\, \sigma)$ denotes $\overline{\mathcal{B}}(A;\, \sigma)$ for $v[n]$ Gaussian, and is given by (16) for $\sigma_\alpha = \sigma$. Both infinite series in (75) are convergent; in fact, only a few terms from each series are required to obtain an accurate estimate of d_∞ such as the one given in (32). Using d_∞ from (32) in conjunction with (74) and (25) yields (33). Similar results hold for non-Gaussian sensor noise pdfs. Specifically, a relation of the form (75) holds for d_∞ defined in (31), where $\mathcal{B}_{\mathcal{N}}(\cdot;\, \cdot)$ is replaced by the associated $\overline{\mathcal{B}}(\cdot;\, \cdot)$. The resulting infinite series in (75) are both convergent since their terms decay faster than $1/n^\rho$ for $1 < \rho < 2$ [recall that $\mathcal{B}(A;\, y)$ grows faster than A^ρ for the control-free scenario]. Clearly, the value of d_∞ depends on the particular noise pdf.

Extensions of the preceding control selection strategies can be developed, which achieve the optimal growth rate of the worst case information loss for finite N. Let $\overline{\boldsymbol{w}}^N$ denote the control vector associated with the finite-N strategy, which is assumed known for estimation. Given a set of $w[n]$ and K selected according to the infinite-N scheme, a finite-N method that achieves the same information loss for any A selects $\overline{\boldsymbol{w}}^N$ randomly from a set of K equally likely vectors $\mathcal{W}(N,\, K) = \{\boldsymbol{w}_i^N : 1 \le i \le K\}$, where the nth element of the $N \times 1$ vector $\boldsymbol{w}_i^N$ is given by $w_i[n] = w[i\,N + n]$.

APPENDIX C
INFORMATION LOSS FOR SIGNAL QUANTIZERS WITH $M \to \infty$

We consider a uniform quantizer with $M = 2(K+1)$ levels. Given K, we select the quantizer thresholds as $X_k = k/\sqrt{K}x$, where $k = -K,\, \ldots,\, K$, and $x > 0$. For convenience, we let $X_{-K-1} = -\infty$ and $X_{K+1} = \infty$. We next examine the Cramér–Rao bound (12) for $w[n] = 0$, where $v[n]$ is admissible. We may rewrite (12) as

$$\mathcal{B}(A;\, \boldsymbol{y}^N) = \frac{1}{N} \left[\sum_{k=-K}^{K+1} \xi_k \right]^{-1} \quad \text{(76a)}$$

where

$$\xi_k = \frac{[p_v(X_k - A) - p_v(X_{k-1} - A)]^2}{C_v(X_{k-1} - A) - C_v(X_k - A)}. \quad \text{(76b)}$$

Note that as $K \to \infty$, both $\xi_{-K} \to 0$ and $\xi_{K+1} \to 0$. By letting

$$m_k = (X_k + X_{k-1})/2 - A$$

for large K and for $k = -K + 1,\, \ldots,\, K$ we have

$$\begin{aligned}
p_v(X_k - A) - p_v(X_{k-1} - A) &\approx p_v'(m_k)x/\sqrt{K} \\
C_v(X_{k-1} - A) - C_v(X_k - A) &\approx p_v(m_k)x/\sqrt{K}
\end{aligned}$$

which imply that

$$\xi_k \approx \frac{[p_v'(m_k)]^2}{p_v'(m_k)} \frac{x}{\sqrt{K}}. \quad \text{(77)}$$

Approximation (77) becomes an equality as $K \to \infty$. Letting $K \to \infty$ in (76a), (76b), and using (77) yields

$$\begin{aligned}
&\lim_{K\to\infty} \mathcal{B}(A;\, \boldsymbol{y}^N) \\
&\quad = \frac{1}{N} \left[\lim_{K\to\infty} (\xi_{-K} + \xi_{K+1}) + \lim_{K\to\infty} \sum_{k=-K+1}^{K} \xi_k \right]^{-1} \\
&\quad = \frac{1}{N} \left[\int_{t=-\infty}^{\infty} \frac{[p_v'(t - A)]^2}{p_v(t - A)}\, dt \right]^{-1} \\
&\quad = \frac{1}{N} \left[\int_{t=-\infty}^{\infty} \left(\frac{\partial \ln p_v(t - A)}{\partial A} \right)^2 p_v(t - A)\, dt \right]^{-1} \\
&\quad = \mathcal{B}(A;\, \boldsymbol{s}^N).
\end{aligned}$$

APPENDIX D
ASYMPTOTIC EFFICIENCY OF ML ESTIMATOR FOR CASE $M = 2$

In this appendix, we show that $\hat{A}_{\mathrm{ML}} = \hat{A}_{\mathrm{ML}}(\boldsymbol{y}^N; \Delta)$ given by (41)–(43) achieves (14) for N large, if $\alpha[n]$ is admissible. Let $\tilde{k}$ denote the binomial random variable $\tilde{k}(\boldsymbol{y}^N) = \mathcal{K}_1(\boldsymbol{y}^N)/N$. Then

$$\hat{A}_{\mathrm{ML}} = \begin{cases} -\Delta, & \text{if } \tilde{k} \le C_\alpha(\Delta) \\ g(\tilde{k}) = -C_\alpha^{-1}(\tilde{k}), & \text{if } C_\alpha(\Delta) < \tilde{k} < C_\alpha(-\Delta) \\ \Delta, & \text{if } \tilde{k} \ge C_\alpha(-\Delta). \end{cases}$$

For large N, the following approximation is valid in the cumulative sense [46, p. 214]:

$$\tilde{k} \sim \mathcal{N}(p, \tilde{\sigma}_N) \tag{78}$$

where $p = C_\alpha(-A)$ and $\tilde{\sigma}_N = \sqrt{p(1-p)/N}$. Since $g(\cdot)$ is invertible [$C_\alpha(\cdot)$ is strictly monotone almost everywhere], the pdfs of $\hat{A}_{\mathrm{ML}}$ and $\tilde{k}$ are related as follows [46, p. 93]:

$$p_{\hat{A}_{\mathrm{ML}}}(\hat{A}) \approx \begin{cases} \delta(\hat{A}-\Delta)Q\left(\sqrt{N}\beta_+\right), & \text{if } \hat{A} = \Delta \\ p_{\tilde{k}}(C_\alpha(-\hat{A}))p_\alpha(-\hat{A}), & \text{if } -\Delta < \hat{A} < \Delta \\ \delta(\hat{A}+\Delta)Q\left(\sqrt{N}\beta_-\right), & \text{if } \hat{A} = -\Delta \end{cases} \tag{79}$$

where

$$\beta_+ = \frac{C_\alpha(-\Delta) - p}{\sqrt{p(1-p)}} \qquad \beta_- = \frac{p - C_\alpha(\Delta)}{\sqrt{p(1-p)}}. \tag{80}$$

Note that the pdf of $\hat{A}_{\mathrm{ML}}$ in (79) consists of a sum of Kronecker delta functions.

We first consider $p_{\hat{A}_{\mathrm{ML}}}(\hat{A})$ for $|\hat{A}| < \Delta$. If N is large enough, so that (78) is valid and also $\tilde{\sigma}_N \ll \sigma_\alpha$, the following approximations for $p_{\hat{A}_{\mathrm{ML}}}(\hat{A})$ are valid in the regime $(-\Delta, \Delta)$, in the sense that for any A in $(-\Delta, \Delta)$ the values of the corresponding cumulative distribution functions are approximately equal (and where the approximation generally improves as N increases):

$$p_{\hat{A}_{\mathrm{ML}}}(\hat{A}) \approx \frac{1}{\sqrt{2\pi}\tilde{\sigma}_N} \exp\left(-\frac{(C_\alpha(-\hat{A}) - C_\alpha(-A))^2}{2\tilde{\sigma}_N^2}\right) p_\alpha(-\hat{A}) \tag{81a}$$

$$\approx \frac{1}{\sqrt{2\pi}\sqrt{\tilde{\sigma}_N^2 (p_\alpha(-A))^{-2}}} \cdot \exp\left(-\frac{(C_\alpha(-\hat{A}) - C_\alpha(-A))^2}{2\tilde{\sigma}_N^2}\right) \tag{81b}$$

$$\approx \frac{1}{\sqrt{2\pi}\left(\gamma/\sqrt{N}\right)} \exp\left(-(p_\alpha(-A))^2 \frac{(\hat{A}-A)^2}{2\tilde{\sigma}_N^2}\right) \tag{81c}$$

$$\approx \frac{1}{\sqrt{2\pi}\left(\gamma/\sqrt{N}\right)} \exp\left(-\frac{(\hat{A}-A)^2}{2\left(\gamma/\sqrt{N}\right)^2}\right) \tag{81d}$$

where

$$\gamma^2 = [1 - C_\alpha(-A)]C_\alpha(-A)(p_\alpha(-A))^{-2}.$$

Approximation (81a) results from using (78) in (79). To verify (81b) note that in the region that

$$\exp(-[C_\alpha(-\hat{A}) - C_\alpha(-A)]^2/[2\tilde{\sigma}_N^2])$$

is essentially nonzero, we have

$$p_\alpha(-\hat{A}) \approx p_\alpha(-A).$$

For $\tilde{\sigma}_N \ll \sigma_\alpha$, the following approximation is valid for the exponent in (81b):

$$[C_\alpha(-\hat{A}) - C_\alpha(-A)]^2 \approx (p_\alpha(-A))^2(\hat{A}-A)^2$$

which when substituted in (81b) results in (81c), and (81d). From (81d), for large N we have $\hat{A}_{\mathrm{ML}} \sim \mathcal{N}(A, \gamma^2/N)$ in the regime $(-\Delta, \Delta)$. Provided N is large enough, $\gamma^2/N \ll \Delta - |A|$, in which case the MSE term contributed from $\hat{A}_{\mathrm{ML}} \in (-\Delta, \Delta)$ approaches the Cramér–Rao bound γ^2/N. Next, consider the two other regimes, where $\hat{A}_{\mathrm{ML}} = \pm\Delta$. Let

$$\rho_- = \exp(-\beta_-^2/2)$$

and

$$\rho_+ = \exp(-\beta_+^2/2)$$

where β_- and β_+ are given by (80). For large enough N

$$Q(\sqrt{N}\beta_+) \approx c_1\rho_+^N/\sqrt{N}$$

and

$$Q(\sqrt{N}\beta_-) \approx c_2\rho_+^N/\sqrt{N}.$$

Since $0 < \rho_+, \rho_- < 1$, the corresponding MSE terms go to zero much faster than $1/N$ for large N, so their contribution to the MSE becomes negligible for large N as compared to γ^2/N. Hence, $\hat{A}_{\mathrm{ML}}$ achieves the Cramér–Rao bound (12) for large N.

APPENDIX E
EM ALGORITHM FOR PARAMETER ESTIMATION IN GAUSSIAN NOISE VIA SIGNAL QUANTIZERS

In this appendix, we present the derivation of an EM algorithm that can be used to obtain the ML estimator of A from a network of signal quantizers. The ith observation y_i is given by

$$y_i = F_i(x_i), \qquad i = 1, 2, \ldots, I$$

where

$$x_i = A + v_i + w_i \tag{82}$$

A is the unknown parameter of interest, $v_i \sim \mathcal{N}(0, \sigma_i^2)$, w_i is the selected (known) control input, and $F_i(\cdot)$ is the ith quantizer and is given by (2a) and (2b). We use $\underline{X}_i(\cdot)$ and $\overline{X}_i(\cdot)$ to denote the functions mapping each quantizer level Y_m of the ith quantizer $F_i(\cdot)$ to the associated lower and upper thresholds X_{m-1} and X_m, respectively.

We select as the complete set of data the set x_i in (82). For convenience, let

$$\boldsymbol{x} = [x_1 \quad x_2 \quad \cdots \quad x_I]^T \quad \text{and} \quad \boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_I]^T.$$

The EM algorithm selects $\hat{A}_{\mathrm{EM}}^{(k+1)}$, the $k+1$st estimate of A based on $\hat{A}_{\mathrm{EM}}^{(k)}$ and $\boldsymbol{y}$, according to

$$\hat{A}_{\mathrm{EM}}^{(k+1)} = \underset{|\theta|<\Delta}{\arg\max}\, U\left(\theta; \hat{A}_{\mathrm{EM}}^{(k)}\right) \tag{83}$$

where

$$U\left(\theta;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)=E\left[\ln p(\boldsymbol{x};\,\theta)|\boldsymbol{y};\,\hat{A}_{\mathrm{EM}}^{(k)}\right]. \tag{84}$$

The log-likelihood function $\ln p(\boldsymbol{x};\,\theta)$ satisfies

$$\begin{aligned}\ln p(\boldsymbol{x};\,\theta)&=C-\sum_{i=1}^{I}\frac{1}{2\sigma_i^2}\left(x_i-w_i-\theta\right)^2\\&=h(\boldsymbol{x})+\theta\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\left(x_i-w_i\right)-\theta^2\sum_{i=1}^{I}\frac{1}{2\sigma_i^2}.\end{aligned} \tag{85}$$

If we substitute (85) for $\ln p(\boldsymbol{x};\,\theta)$ in (84) we obtain

$$U\left(\theta;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)=E\left[h(\boldsymbol{x})|\boldsymbol{y};\,\hat{A}_{\mathrm{EM}}^{(k)}\right]-\theta E[k]+\frac{\theta^2}{2}\,\mu \tag{86}$$

where $\mu=\sum_{i=1}^{I}\sigma_i^{-2}$, and

$$E[k]=\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\,E_i[k]=\sum_{i=1}^{I}\frac{1}{\sigma_i^2}\left(E\left[x_i|\boldsymbol{y};\,\hat{A}_{\mathrm{EM}}^{(k)}\right]-w_i\right).$$

Substituting in (83) the expression for $U(\theta;\,\hat{A}_{\mathrm{EM}}^{(k)})$ in (86) we obtain

$$\hat{A}_{\mathrm{EM}}^{(k+1)}=\mathcal{I}_{\Delta}(E[k]/\mu). \tag{87}$$

Let $\underline{x}_i=\underline{X}_i(y_i)$ and $\overline{x}_i=\overline{X}_i(y_i)$. Using

$$\begin{aligned}&p\left(x_i|\boldsymbol{y};\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\\&\quad=p\left(x_i|y_i;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\\&\quad=p\left(y_i|x_i;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)p\left(x_i;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\left[p\left(y_i;\,\hat{A}_{\mathrm{EM}}^{(k)}\right)\right]^{-1}\end{aligned}$$

we obtain for $E_i[k]$ the first expression shown at the bottom of this page, which when substituted in (87) results in (88), also shown at the bottom of this page. Several special cases of (88) are of interest. In particular, if $F_i(x)=F(x)=\mathrm{sgn}(x)$, (88) reduces to

$$\hat{A}_{\mathrm{EM}}^{(k+1)}=\mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)}+\frac{1}{\sqrt{2\pi}\sum_{i=1}^{I}\sigma_i^{-2}}\cdot\left[\sum_{i=1}^{I}\frac{y_i}{\sigma_i}\,\frac{\exp\left(-\frac{\left(\hat{A}_{\mathrm{EM}}^{(k)}+w_i\right)^2}{2\sigma_i^2}\right)}{Q\left(-\frac{\hat{A}_{\mathrm{EM}}^{(k)}+w_i}{\sigma_i}\,y_i\right)}\right]\right)$$

of which (54a), (54b) is a special case. Next, consider the special case where N observations are collected from a single M-level quantizer [i.e., $F_i(x)=F(x)$ and $I=N$]. If, in addition, $w_i=w$ and $\sigma_i=\sigma$ for all i, (88) reduces to (89) found at the bottom of the page; note that only the sufficient statistics $\mathcal{K}_{Y_1}(\boldsymbol{y})$, $\mathcal{K}_{Y_2}(\boldsymbol{y})$, $\ldots$, $\mathcal{K}_{Y_{M-1}}(\boldsymbol{y})$ are used in (89) to obtain $\hat{A}_{\mathrm{ML}}$.

Appendix F
Asymptotically Efficient Estimation for Random Control Inputs

We develop a class of asymptotically efficient estimators of the parameter A, based on observation of (1) where $F(\cdot)$ is

$$\begin{aligned}E_i[k]&=\frac{\frac{1}{\sqrt{2\pi}\sigma_i}\int_{x=\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}^{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}\left(x+\hat{A}_{\mathrm{EM}}^{(k)}+w_i\right)\exp\left(-\frac{x^2}{2\sigma_i^2}\right)dx}{Q\left(\frac{\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)-Q\left(\frac{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)}-w_i\\&=\hat{A}_{\mathrm{EM}}^{(k)}+\frac{\sigma_i}{\sqrt{2\pi}}\,\frac{\exp\left(-\frac{\left(\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\sigma_i^2}\right)-\exp\left(-\frac{\left(\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\sigma_i^2}\right)}{Q\left(\frac{\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)-Q\left(\frac{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)}.\end{aligned}$$

$$\hat{A}_{\mathrm{EM}}^{(k+1)}=\mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)}+\frac{1}{\sqrt{2\pi}\sum_{i=1}^{I}\sigma_i^{-2}}\sum_{i=1}^{I}\frac{\exp\left(-\frac{\left(\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\sigma_i^2}\right)-\exp\left(-\frac{\left(\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i\right)^2}{2\sigma_i^2}\right)}{\sigma_i\left[Q\left(\frac{\underline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)-Q\left(\frac{\overline{x}_i-\hat{A}_{\mathrm{EM}}^{(k)}-w_i}{\sigma_i}\right)\right]}\right). \tag{88}$$

$$\hat{A}_{\mathrm{EM}}^{(k+1)}=\mathcal{I}_{\Delta}\left(\hat{A}_{\mathrm{EM}}^{(k)}+\sum_{m=1}^{M}\frac{\sigma\mathcal{K}_{Y_m}(\boldsymbol{y})}{\sqrt{2\pi}N}\,\frac{\exp\left(-\frac{\left(X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}-w\right)^2}{2\sigma^2}\right)-\exp\left(-\frac{\left(X_m-\hat{A}_{\mathrm{EM}}^{(k)}-w\right)^2}{2\sigma^2}\right)}{Q\left(\frac{X_{m-1}-\hat{A}_{\mathrm{EM}}^{(k)}-w}{\sigma}\right)-Q\left(\frac{X_m-\hat{A}_{\mathrm{EM}}^{(k)}-w}{\sigma}\right)}\right). \tag{89}$$

given by (2a) and (2b), and where $\alpha[n]$ in (10) is an i.i.d. admissible noise process. In the absence of a control input, $\alpha[n]$ equals $v[n]$. Consider the following collection of binary sequences:

$$y_i[n] = F_i(A + \alpha[n]) = \operatorname{sgn}(A + \alpha[n] - X_i)$$

for $i = 1, 2, \ldots, M-1$. The observed output $y[n]$ is equivalent to the collection $y_1[n], \ldots, y_{M-1}[n]$, since $y[n] = \sum_i y_i[n]$ and $y_i[n] = \operatorname{sgn}(y[n] - Y_i)$. The ML estimate of A based on

$$\boldsymbol{y}_i^N = [\, y_i[1] \quad y_i[2] \quad \cdots \quad y_i[N] \,]^T$$

is given by $\hat{A}_i$ in (46). We have

$$\hat{A}_i = \mathcal{I}_\Delta(-C_\alpha^{-1}(T_i)) \tag{90}$$

where

$$T_i \triangleq \mathcal{K}_1(\boldsymbol{y}_i^N) = \frac{1}{2N} \sum_n y_i[n] + \frac{1}{2}.$$

The estimators we develop next are based on the vector $\hat{\boldsymbol{A}}$ defined in (47). Note that, although the collection of T_i's is a set of sufficient statistics for the problem, $\hat{\boldsymbol{A}}$ is not, in general, a sufficient statistic due to the limiting operation in (47). As a first step in obtaining the distribution of $\hat{\boldsymbol{A}}$ for large N, we obtain the distribution of the vector

$$\boldsymbol{T} = [\, T_1 \quad T_2 \quad \cdots \quad T_{M-1} \,]^T.$$

For convenience, let

$$p_i \triangleq C_\alpha(X_i - A)$$

and

$$f_i \triangleq p_\alpha(X_i - A).$$

First note that the distribution of the vector

$$[\, \mathcal{K}_{Y_1}(\boldsymbol{y}^N) \quad \mathcal{K}_{Y_2}(\boldsymbol{y}^N) \quad \cdots \quad \mathcal{K}_{Y_M}(\boldsymbol{y}^N) \,]^T$$

is multinomial, and approaches a Gaussian vector in the cumulative sense [46, p. 214]. The T_i's are linear combinations of the $\mathcal{K}_{Y_i}(\boldsymbol{y}^N)$'s, since

$$T_i = \sum_{j=i+1}^{M} \mathcal{K}_{Y_j}(\boldsymbol{y}^N).$$

Consequently, $\boldsymbol{T}$ also approaches a Gaussian vector in the cumulative sense, i.e., $\boldsymbol{T} \sim \mathcal{N}(\overline{\boldsymbol{T}}, R_{\boldsymbol{T}})$, where

$$\overline{\boldsymbol{T}} = [\, p_1 \quad p_2 \quad \cdots \quad p_{M-1} \,]^T$$

$R_{\boldsymbol{T}} = R/N$, and

$$R = \begin{bmatrix} p_1(1-p_1) & p_2(1-p_1) & \cdots & p_{M-1}(1-p_1) \\ p_2(1-p_1) & p_2(1-p_2) & \cdots & p_{M-1}(1-p_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_{M-1}(1-p_1) & p_{M-1}(1-p_2) & \cdots & p_{M-1}(1-p_{M-1}) \end{bmatrix}.$$

In a manner analogous to the case $M = 2$ described in Appendix E, by using the theorem for the pdf of transformation of variables [46, p. 93], and in the limit as $N \to \infty$ (where we invoke the central limit theorem and ignore the boundary effects due to $|\hat{A}_i| \le \Delta$), we have $\hat{\boldsymbol{A}} \sim \mathcal{N}(A\mathbf{1}, C/N)$ in the cumulative sense, where $C = F^{-1}RF^{-1}$, and

$$F = \operatorname{diag}(f_1, f_2, \ldots, f_{M-1}).$$

It can be readily verified that

$$C^{-1} = \begin{bmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{M-2} \\ 0 & \cdots & 0 & b_{M-2} & a_{M-1} \end{bmatrix} \tag{91}$$

where

$$a_i = \frac{f_i^2}{p_{i-1} - p_i} + \frac{f_i^2}{p_i - p_{i+1}} \quad \text{and} \quad b_i = \frac{f_i f_{i+1}}{p_{i+1} - p_i}.$$

If C were available, then the optimal estimate (in terms of minimizing the MSE) would be

$$\overline{A} = \boldsymbol{\lambda}^T \hat{\boldsymbol{A}} = (\mathbf{1}^T C^{-1} \mathbf{1})^{-1} \mathbf{1}^T C^{-1} \hat{\boldsymbol{A}}$$

while the associated MSE would satisfy

$$\lim_{N\to\infty} NE[(\overline{A} - A)^2] = (\mathbf{1}^T C^{-1} \mathbf{1})^{-1} = \mathcal{B}(A; y)$$

i.e., this estimator would be asymptotically efficient. However, C is a function of the unknown parameter A. Instead, note that $C(\hat{A}_i)$ approaches $C(A)$ for large N for any i. Specifically, $\hat{A}_i$ is a consistent estimate since it is an asymptotically efficient estimate with respect to the Cramér–Rao bound for a two-level quantizer with threshold X_i. Hence, set $i = 1$ and consider

$$\hat{A} = \boldsymbol{\lambda}(\hat{A}_1)^T \hat{\boldsymbol{A}}$$

where $\boldsymbol{\lambda}(\theta) = (\mathbf{1}^T C^{-1}(\theta)\mathbf{1})^{-1}\mathbf{1}^T C^{-1}(\theta)$. Let $\overline{z} = \overline{A} - A$, $z = \hat{z} - A$, and $\hat{\boldsymbol{z}} = \hat{\boldsymbol{A}} - A\mathbf{1}$. Also, let $\Delta\boldsymbol{\lambda} = \boldsymbol{\lambda}(\hat{A}_1) - \boldsymbol{\lambda}(A)$, and denote by $\Delta\lambda_i$ the ith element of $\Delta\boldsymbol{\lambda}$. Then

$$\begin{aligned} \lim_{N\to\infty} NE[(\hat{A} - A)^2;\, A] &= \lim_{N\to\infty} NE[z^2;\, A] \\ &= \lim_{N\to\infty} NE[(\overline{z} + \Delta\boldsymbol{\lambda}^T \hat{\boldsymbol{z}})^2;\, A] \\ &= \mathcal{B}(A; y) + \lim_{N\to\infty} N \sum_{i,j} (\beta_{i,j} + \zeta_{i,j}) \end{aligned} \tag{92}$$

where

$$\beta_{i,j} = E[\Delta\lambda_i \Delta\lambda_j \hat{z}_i \hat{z}_j]$$

and

$$\zeta_{i,j} = E[\Delta\lambda_i \lambda_j \hat{z}_i \hat{z}_j].$$

Note that in Appendix E we have shown that

$$\lim_{N\to\infty} NE[(\hat{A}_1 - A)^2] = \mathcal{B}(A; y_1).$$

Since $p_\alpha(\cdot)$ is admissible, for large N we have

$$\Delta\lambda_i \approx (\hat{A}_1 - A)\lambda_i'(A).$$

Also, since $\hat{A}_1 - A$ is Gaussian for large N (see Appendix E), so is $\Delta\lambda_i$. In addition, there exists $G > |\lambda_i'(A)|$ for all i, which implies that $E[\Delta\lambda_i^2] \leq G/N$, and $E[\Delta\lambda_i^4] \leq 3G^2/N^2$. There also exists U such that $E[\hat{z}_i^2] \leq U/N$ for all i, for N large enough. Finally, let $\lambda_{\max} = \max_i \lambda_i(A)$. Repeated applications of the Schwarz inequality yield

$$|\beta_{i,j}| \leq (E[\Delta\lambda_i^4]\, E[\Delta\lambda_j^4]\, E[\hat{z}_i^4]\, E[\hat{z}_j^4])^{1/4} \leq \frac{3GU}{N^2}$$

and

$$|\zeta_{i,j}| \leq \lambda_{\max}(E[\Delta\lambda_i^2])^{1/2}(E[\hat{z}_i^4]\, E[\hat{z}_j^4])^{1/4} \leq \frac{\lambda_{\max}\sqrt{3GU}}{N\sqrt{N}}$$

which, when substituted in (92), result in

$$\begin{aligned}
&\lim_{N\to\infty} NE[(\hat{A}-A)^2;\, A] \\
&\quad \leq \mathcal{B}(A;\, y) + \sum_{i,j} \lim_{N\to\infty} N(|\beta_{i,j}| + |\zeta_{i,j}|) \\
&\quad \leq \mathcal{B}(A;\, y) + \sum_{i,j} \lim_{N\to\infty} \left(\frac{3GU}{N} + \frac{\lambda_{\max}\sqrt{3GU}}{\sqrt{N}} \right) \\
&\quad \leq \mathcal{B}(A;\, y). \qquad (93)
\end{aligned}$$

Since $\hat{A}$ is asymptotically unbiased (as a sum of asymptotically unbiased estimates), for N large we have

$$E[(\hat{A}-A)^2;\, A] \geq \mathcal{B}(A;\, y)/N$$

which in conjunction with (93) yields the desired (49).

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewer for several helpful comments and suggestions.

REFERENCES

[1] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, pp. 98–101, Jan. 1986.

[2] R. T. Antony, "Database support to data fusion automation," *Proc. IEEE*, vol. 85, pp. 39–53, Jan. 1997.

[3] M. Kam, X. Zhu, and P. Kalata, "Sensor fusion for mobile robot navigation," *Proc. IEEE*, vol. 85, pp. 108–119, Jan. 1997.

[4] B. S. Rao, H. F. Durrant-Whyte, and J. A. Sheen, "A fully decentralized multi-sensor system for tracking and surveillance," *Int. J. Robot. Res.*, vol. 12, no. 1, pp. 20–44, Feb. 1993.

[5] M. Yeddanapudi, Y. Bar-Shalom, and K. R. Pattipati, "IMM estimation for multitarget-multisensor air traffic surveillance," *Proc. IEEE*, vol. 85, pp. 80–94, Jan. 1997.

[6] R. Chellappa, Q. Zheng, P. Burlina, C. Shekhar, and K. B. Eom, "On the positioning of multisensor imagery for exploitation and target recognition," *Proc. IEEE*, vol. 85, pp. 120–138, Jan. 1997.

[7] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors—Part I: Fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, Jan. 1997.

[8] E. B. Hall, A. E. Wessel, and G. L. Wise, "Some aspects of fusion in estimation theory," *IEEE Trans. Inform. Theory*, vol. 37, pp. 420–422, Mar. 1991.

[9] R. S. Blum and S. A. Kassam, "Optimum distributed detection of weak signals in dependent sensors," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1066–1079, May 1992.

[10] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Contr. Sig. Syst.*, vol. 1, no. 2, pp. 167–182, 1988.

[11] P.-N. Chen and A. Papamarkou, "Error bounds for parallel distributed detection under the Neyman-Pearson criterion," *IEEE Trans. Inform. Theory*, vol. 41, pp. 528–533, Mar. 1995.

[12] D. Warren and P. Willett, "Optimal decentralized detection for conditionally independent sensors," in *Amer. Contr. Conf.*, 1989, pp. 1326–1329.

[13] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors—Part II: Advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[14] K. C. Chou, A. S. Willsky, and A. Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 464–478, Mar. 1994.

[15] D. A. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 418–425, May 1985.

[16] M. M. Daniel and A. S. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proc. IEEE*, vol. 85, pp. 164–180, Jan. 1997.

[17] H. V. Poor and J. B. Thomas, "Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, Sept. 1977.

[18] Z.-Q. Luo and J. N. Tsitsiklis, "Data fusion with minimal communication," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1551–1563, Sept. 1994.

[19] M. DeWeese and W. Bialek, "Information flow in sensory neurons," *Nuovo Cimento Soc. Ital. Fys.*, vol. 17D, no. 7–8, pp. 733–741, July–Aug. 1995.

[20] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," *Nature*, vol. 365, pp. 337–340, Sept. 1993.

[21] J. Levin and J. Miller, "Broadband neural encoding in the cricket sensory system enhanced by stochastic resonance," *Nature*, vol. 380, no. 6570, pp. 165–168, Mar. 1996.

[22] T. Berger, *Rate Distortion Theory*, ser. Prentice-Hall Series in Information and System Sciences. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[23] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 406–411, July 1970.

[24] J. M. Morris, "The performance of quantizers for a class of noise-corrupted signal sources," *IEEE Trans. Commun.*, vol. COM-24, pp. 184–189, Feb. 1976.

[25] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory*, vol. 34, pp. 826–834, July 1988.

[26] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, Nov. 1987.

[27] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 241–255, Mar. 1990.

[28] V. V. Veeravalli, T. Basar, and H. V. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433–442, Mar. 1993.

[29] H. R. Hashemi and I. B. Rhodes, "Decentralized sequential detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 509–520, May 1989.

[30] R. M. Gray, S. Boyd, and T. Lookabaugh, "Low rate distributed quantization of noisy observations," in *Proc. Allerton Conf. Communication, Control and Computing*, Urbana-Champaign, IL, 1985, pp. 354–358.

[31] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 887–902, May 1996.

[32] R. A. Wannamaker, S. P. Lipshitz, and J. Vanderkooy, "A theory of nonsubtractive dither," *IEEE Trans. Signal Processing*, vol. 48, pp. 499–516, Feb. 2000.

[33] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, vol. 45, pp. 106–117, 1999.

[34] H. C. Papadopoulos, "Efficient digital encoding and estimation of noisy signals," Ph.D. dissertation, MIT, Cambridge, MA, May 1998.

[35] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys.*, vol. A14, pp. L453–L457, 1981.

[36] Z. Gingl, L. B. Kiss, and F. Moss, "Non-dynamical stochastic resonance: Theory and experiments with white and arbitrarily colored noise," *Europhys. Lett.*, vol. 29, no. 3, pp. 191–196, Jan. 1995.

[37] H. L. Van Trees, *Detection, Estimation and Modulation Theory—Part I.* New York: Wiley, 1968.

[38] L. R. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 145–154, Feb. 1962.

[39] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: A theoretical survey," *J. Audio Eng. Soc.*, vol. 40, no. 5, pp. 355–375, May 1992.

[40] R. M. Gray and T. G. Stockham Jr., "Dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 39, pp. 805–812, May 1993.

[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Ann. Roy. Statist. Soc.*, vol. 39, pp. 1–38, Dec. 1977.

[42] H. C. Papadopoulos, "Sequential signal encoding and estimation for wireless sensor networks," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000.

[43] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. Commun. Technol.*, vol. COM-12, pp. 162–165, Dec. 1964.

[44] A. Gersho, "Principles of quantization," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 427–436, July 1978.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[46] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.