Distributed Adaptive Quantization for Wireless Sensor Networks: From Delta Modulation to Maximum Likelihood

Jun Fang, *Member, IEEE*, and Hongbin Li, *Member, IEEE*

***Abstract*—We consider distributed parameter estimation using quantized observations in wireless sensor networks (WSNs) where, due to bandwidth constraint, each sensor quantizes its local observation into one bit of information. A conventional fixed quantization (FQ) approach, which employs a fixed threshold for all sensors, incurs an estimation error growing exponentially with the difference between the threshold and the unknown parameter to be estimated. To address this difficulty, we propose a distributed adaptive quantization (AQ) approach, which, with sensors *sequentially* broadcasting their quantized data, allows each sensor to adaptively adjust its quantization threshold. Three AQ schemes are presented: 1) AQ-FS that involves distributed delta modulation (DM) with a fixed stepsize, 2) AQ-VS that employs DM with a variable stepsize, and 3) AQ-ML that adjusts the threshold through a maximum likelihood (ML) estimation process. The ML estimators associated with the three AQ schemes are developed and their corresponding Cramér–Rao bounds (CRBs) are analyzed. We show that our 1-bit AQ approach is asymptotically optimum, yielding an asymptotic CRB that is only $\pi/2$ times that of the clairvoyant sample-mean estimator using *unquantized* observations.**

***Index Terms*—Adaptive quantization (AQ), distributed estimation, wireless sensor networks (WSNs).**

I. Introduction

WlockIRELESS sensor networks (WSNs) have attracted much attention over the past few years. Consisting of a large number of small, low-cost sensors with integrated sensing, processing, and communication abilities, WSNs can accomplish a variety of tasks including environment monitoring, battlefield surveillance, target localization and tracking, and many more [1], [2]. Bandwidth and power constraints are two primary issues that need to be addressed in WSNs, as limited communication bandwidth is shared across the entire network and, meanwhile, the sensors are often powered by irreplaceable batteries. As such, a major challenge of the WSN research is to design bandwidth- and power-efficient signal-processing techniques. A multitude of studies along this line have appeared recently in the context of distributed detection (e.g., [3]–[7]), optimal decentralized compression-estimation by exploiting spatial correlation (e.g., [8]–[10]), distributed estimation with quantized observations (e.g., [11]–[21]), and others.

Manuscript received December 13, 2007; revised July 2, 2008. First published July 25, 2008; current version published September 17, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniel P. Palomar. This work was supported in part by the National Science Foundation under Grant CCF-0514938 and the National Natural Science Foundation of China under Grant 60502011.

The authors are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Jun.Fang@stevens.edu; Hongbin.Li@stevens.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.928956

A. The Problem

In this paper, we consider distributed estimation of a deterministic unknown parameter θ from quantized observations in a WSN. Suppose we have N spatially distributed sensors. Each sensor makes a noisy observation of an unknown parameter θ

$$x_n = \theta + w_n, \quad n = 1, 2, \dots, N \tag{1}$$

where w_n denotes the additive observation noise with zero mean and variance σ_w^2, and the noise is assumed independent and identically distributed (i.i.d.) across the sensors. To meet stringent bandwidth/power budgets in WSNs, we consider the case where each sensor uses a 1-bit quantizer

$$b_n = Q_n(x_n) \tag{2}$$

and the binary data $\{b_n\}$ are sent to a fusion center (FC) to form an estimate of θ. The problem of interest is to determine suitable binary quantizers $\{Q_n(\cdot)\}_{n=1}^N$ for each sensor and an estimator for the FC to form an estimate of the unknown parameter θ from $\{b_n\}_{n=1}^N$.

B. Past Related Works

The above distributed estimation problem has been considered in a number of studies. Specifically, by modeling θ as a random parameter, Bayesian techniques were proposed in, e.g., [11]–[13]. Here "Bayesian" reflects the fact that these methods design their quantizers with the aid of the prior distribution of θ explicitly or implicitly. These methods usually require knowledge of the joint distribution of θ and the observed signals for quantizer design (an exception is [12], where a training set of multiple realizations of θ and sensor observations are used instead). Optimum Bayesian quantizers, though, are difficult to obtain because they involve multidimensional search to find the best quantization thresholds; moreover, the optimum Bayesian estimator takes the form of a conditional mean that is usually hard to compute.

Another category of methods treat θ as a deterministic unknown parameter. A notable example is a fixed quantization (FQ) approach, where a common threshold τ is applied at all sensors [14], [15]. Although FQ admits a simple closed-form maximum likelihood estimator (MLE), the fundamental problem of FQ is that the choice of τ is very sensitive: the optimum threshold $\tau = \theta$ is not usable since it is unknown, and the performance of FQ degrades exponentially as $|\tau - \theta|$ increases [14], [15] (also see Section II for an overview). A remedy proposed in [14] is to periodically apply a set of thresholds with equal frequencies (through a periodic control signal or dithering added before quantization), in the hope that one of the thresholds is close to the unknown θ. Multithresholding was employed in several other studies, including, e.g., [15] and [18], the latter focusing on nonparametric estimation. Unlike [14], [15] proposed an unequal-frequency multithresholding strategy that allows some thresholds (in particular those closer to θ) being used more frequently than the others. This, however, requires knowledge of the prior distribution of θ, like the Bayesian methods. Another recent method addressing quantization with deterministically unknown θ was introduced in [20], where the idea is to optimize the worst case performance (by maximizing the minimum asymptotic efficiency between two MLEs using quantized and, respectively, unquantized observations).

C. Contributions of This Paper

The aforementioned quantization schemes can all be considered within the general class of FQ in the sense that the thresholds are precomputed, fixed through the quantization process, and independent of the data $\{x_n\}$. In this paper, we consider a data-dependent distributed adaptive quantization (AQ) approach whereby the threshold is dynamically adjusted from one sensor to another, in a way such that the threshold converges to or near the unknown θ. We assume each sensor sends its quantized data sequentially with the help of a scheduling algorithm, e.g., [22], and while it transmits, the other sensors can listen to the transmission (due to the broadcasting nature of the wireless channel) and use the information to adjust its local quantizer. This is different from earlier distributed estimation techniques with bandwidth constraint (e.g., [11]–[15]), where the sensors do not communicate with each other, although we note that similar sequential transmissions have been adopted in several other recent works for different applications (e.g., [21] and [23]). The AQ approach was initially introduced in [24], where a distributed delta modulation (DM) was used to vary the threshold from sensor to sensor. In this paper, we make several significant extensions leading to improved performance and, in addition, provide analysis offering new insights into the AQ approach.

Specifically, our new contributions are the following. First, generalizing the DM-based scheme of [24], referred to as AQ-FS herein, we introduce two additional AQ schemes, namely, AQ-VS that employs DM with a variable step size and AQ-ML that adjusts the threshold through an ML estimation process. At moderately increased complexity, AQ-VS and AQ-ML converge faster to θ and provide improved performance. Hence, the three AQ schemes offer the flexibility to tradeoff between performance and complexity. Secondly, the MLEs and Cramér–Rao bounds (CRBs) for all three AQ schemes are developed within a unified framework. Thirdly, asymptotic analysis is provided leading to new insights unknown before. In particular, while the CRB for AQ-FS in [24] involves iterative calculations, we derive here an asymptotic CRB in a closed form by using a stationary property of Markov chain. Furthermore, we show that the CRB of AQ-ML asymptotically converges to that of the *best* (although practically infeasible) 1-bit quantizer (i.e., the one that uses the ideal threshold $\tau_{\mathrm{opt}} = \theta$) and is only $\pi/2$ times that of the clairvoyant sample-mean estimator using *unquantized* observations.

The rest of this paper is organized as follows. We first briefly review some basic results of the FQ approach in Section II. The proposed AQ schemes are presented in Section III. Next, we develop the corresponding MLEs and CRBs in Sections IV and V, respectively. Numerical results and comparisons are presented in Section VI, followed by concluding remarks in Section VII.

II. PRELIMINARIES

To motivate and facilitate presentation of our AQ schemes, the FQ approach is briefly reviewed here (see [14] and [15] for details). FQ applies a fixed threshold τ for all sensors

$$b_n = \mathrm{sgn}(x_n - \tau), \quad n = 1, 2, \ldots, N \tag{3}$$

where

$$\begin{cases} \mathrm{sgn}\{x\} = -1, & \text{if } x \le 0 \\ \mathrm{sgn}\{x\} = 1, & \text{if } x > 0 \end{cases}.$$

The CRB of using $\{b_n\}$ to estimate θ is

$$\mathrm{CRB}(\theta; \tau) = \frac{F_w(\tau - \theta)[1 - F_w(\tau - \theta)]}{N p_w^2(\tau - \theta)} \tag{4}$$

where $p_w(x)$ and $F_w(x)$ denote the probability density function (pdf) and the complementary cumulative density function (ccdf) of w_n, respectively. The optimum 1-bit quantizer is obtained by minimizing (4) with respect to τ. For Gaussian and several other noise distributions, including Laplacian and Cauchy, the optimum threshold is $\tau_{\mathrm{opt}} = \theta$, and the minimum CRB is given by

$$\mathrm{CRB}_{\min}(\theta) = \frac{\pi\sigma_w^2}{2N} = \frac{\pi}{2}\mathrm{CRB}_{\mathrm{NQ}} \tag{5}$$

where NQ stands for "no quantization," $\mathrm{CRB}_{\mathrm{NQ}} = \sigma_w^2/N$ denotes the CRB of using the unquantized data $\{x_n\}$ to estimate θ, which is achieved by the *clairvoyant sample mean estimator* that computes the sample mean of the unquantized data

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{6}$$

The optimum 1-bit quantizer is not directly feasible since θ is unknown. On the other hand, as τ deviates from θ, the CRB for the Gaussian case can be tightly bounded by

$$\mathrm{CRB}(\theta; \tau) \le \frac{\pi\sigma_w^2}{2N} e^{(1/2)(\tau - \theta/\sigma_w)^2} \tag{7}$$

which indicates an exponential increase with $[(\tau - \theta)/\sigma_w]^2$. The exponential degradation of performance is also shown in Fig. 1, which plots the CRB (4) as a function of the threshold.

Although the performance of FQ is disappointing, its CRB analysis reveals that to obtain good estimation performance, τ should be placed close to θ. Also, as will be shown theoretically in Section V, the more sensors use a threshold close to θ, the

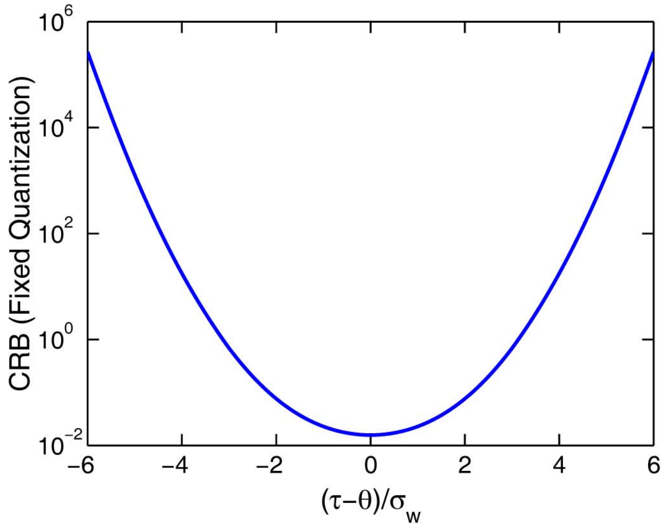


Fig. 1. CRB of the fixed quantization approach.

better the estimation accuracy will be. This motivates us to consider the AQ approach that adaptively adjusts the quantization threshold of each sensor such that it converges in some way to the unknown θ.

III. Proposed AQ Approach

We assume that the sensors use the channel by time sharing and sequentially transmit their quantized data, i.e., one sensor transmits at a time. This can be achieved by using either a centralized scheduler (e.g., each sensor polled by the FC) or a distributed scheduler (e.g., a time-stamp-based scheduling algorithm). We assume that the transmission of each sensor can be heard by the subsequent sensors, due to the broadcasting nature of the wireless channel. To focus on the quantization problem, we assume that the data are received without errors (by using, e.g., a strong error-correction code). Imperfect communication due to noisy channels will affect the performance of all distributed estimation schemes, including ours. Albeit important, we consider this a separate issue. Like most other techniques (e.g., [11]–[20]), we do not assume knowledge of the multilink channels at any of the transmitting nodes, and, as such, the quantizers are independent of the channel (noisy or noiseless). Channel-aware quantization using channel feedback from receiving to transmitting nodes for distributed detection is discussed in [25].

A. AQ-FS

The first AQ scheme, introduced in [24] and referred to here as AQ-FS, is briefly summarized as follows. The 1-bit quantizer at sensor 1 uses an arbitrary, say, $\tau_1 = 0$, to generate b_1

$$b_1 = \text{sgn}\{x_1 - \tau_1\}. \tag{8}$$

Then, b_1 is sent (broadcast) to the FC and all other sensors. After receiving b_1, sensor 2 computes $\tau_2 = \Delta b_1$, where $\Delta > 0$ is a *step-size* parameter whose choice is briefly discussed below, then analyzed in Sections V-B and VI, and generates b_2

$$b_2 = \text{sgn}\{x_2 - \tau_2\}. \tag{9}$$

In general, for sensor n, it first forms a cumulative sum

$$\tau_n = \tau_{n-1} + \Delta b_{n-1} = \Delta \sum_{k=1}^{n-1} b_k \tag{10}$$

and then uses τ_n as a threshold for quantization

$$b_n = \text{sgn}\{x_n - \tau_n\}. \tag{11}$$

One can immediately recognize that the above process is reminiscent of the delta modulation (DM) but is implemented in a distributed fashion to solve a problem notably different from the traditional waveform coding. The dynamic evolution of the thresholds used by AQ-FS is illustrated in Fig. 2(a), where we see τ_n converges near the unknown θ as n increases.

B. AQ-VS

A closer examination of Fig. 2(a) shows that the evolution of the threshold has two distinctive phases: a *transient phase* that brings τ_n near θ and a *convergent phase* where τ_n oscillates around θ. To speed up the convergence, we need a large step size Δ, whereas to reduce the granular noise after convergence, we need a small Δ. This shows the insufficiency of using a fixed Δ and motivates us to consider the following AQ with variable step-size (VS) scheme.

Let b_1 and b_2 be generated the same as in AQ-FS. At sensor $n > 2$, it performs accumulation of the previous bits, weighted by a VS Δ_n

$$\tau_n = \tau_{n-1} + b_{n-1}\Delta_n \tag{12}$$

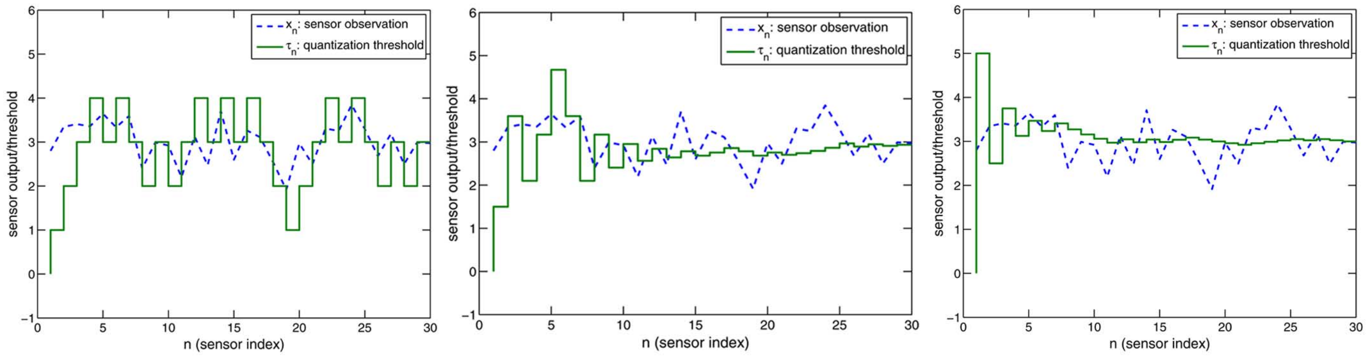
where Δ_n evolves using the following dynamic model:

$$\Delta_n = \Delta_{n-1} K^{b_{n-1}b_{n-2}} \quad n = 3, 4, \ldots \tag{13}$$

where $K > 1$ is a constant. We see that the step size Δ_n is adaptively adjusted based on previous two bits. Specifically, we have

$$\Delta_n = \begin{cases} K\Delta_{n-1}, & \text{if } b_{n-2} \text{ and } b_{n-1} \text{ have same sign} \\ \frac{1}{K}\Delta_{n-1}, & \text{otherwise} \end{cases}. \tag{14}$$

The above adjustment is based on the observation that when successive data bits have identical signs, with a high probability the varying threshold is still in the transient phase and to speed up the convergence, we should increase the step size. On the other hand, alternating signs between successive bits indicate that the thresholds are oscillating around θ and decreasing the step size is desirable. The dynamic evolution of AQ-VS is depicted in Fig. 2(b), where we see that AQ-VS is able to converge faster



(a) (b) (c)

Fig. 2. Dynamic evolution of the quantization threshold versus sensor index. (a) AQ-FS, (b) AQ-VS, and (c) AQ-ML.

and stay closer to θ than AQ-FS. Finally, we note that it is possible to extend the above scheme to use more than two previous bits to provide finer adjustment of the step size.

C. AQ-ML

The above AQ-VS is a *linear* adjustment scheme where the step size is linearly scaled and, moreover, the combining logic XOR that was used to combine b_{n-1} and b_{n-2} is also a linear operator. Yet another AQ scheme, referred to as AQ-ML, can be obtained by using nonlinear ML estimation to adjust the threshold. Specifically, let b_1 be generated as in AQ-FS. Upon receiving b_1, sensor 2 computes $\tau_2 = \Delta b_1$, where Δ is a step size whose choice is discussed shortly and uses it to generate b_2. Based on the received $\{b_1, b_2\}$, sensor 3 computes τ_3 as

$$\tau_3 = \arg\max_{\theta} L_3(\theta; b_1, b_2) \quad (15)$$

where $L_3(\theta; b_1, b_2)$ denotes the likelihood function of θ and τ_3 is recognized as the ML estimate of θ given binary observations b_1 and b_2. The expression of the likelihood function along with the ML solution for a general case is discussed in Section IV and hence not repeated here. The step size Δ used by sensor 2 should be large enough such that b_1 and b_2 have opposite signs. Otherwise, it can be shown that τ_3 obtained above is either infinity or negative infinity (depending on the signs of b_1 and b_2), which should be avoided. Although there is always a nonzero probability for b_1 and b_2 to have identical signs, the probability can be made practically small by choosing Δ sufficiently large. In addition, if, for a chosen Δ, the first two quantized bits are still of an identical sign, the following sensors can keep using AQ-FS or AQ-VS until a binary bit of a different sign is generated, at which point the quantization process is switched to AQ-ML.

In general, for sensor n, it first recovers the previous thresholds $\{\tau_k\}_{k=1}^{n-1}$ from the received quantized data $\{b_k\}_{k=1}^{n-2}$ and then find its own threshold by

$$\tau_n = \arg\max_{\theta} L_n(\theta; b_1, b_2, \ldots, b_{n-1}) \quad (16)$$

where $L_n(\theta; b_1, b_2, \ldots, b_{n-1})$ denotes the likelihood function of θ given $\{b_k\}_{k=1}^{n-1}$. The need to recover prior thresholds is because the likelihood function $L_n(\theta; b_1, b_2, \ldots, b_{n-1})$ is conveniently expressed as a function of $\{\tau_k\}_{k=1}^{n-1}$ (see Section IV). Although sensor n has to perform $n-3$ recursive ML estimations (τ_1 and τ_2 are known), the complexity is moderate for Gaussian noise w_n (see discussions in Section IV). The dynamic evolution of AQ-ML is depicted in Fig. 2(c).

A few comments on the three AQ schemes are now in order. First, all of them are *random thresholding* schemes since their thresholds are random variables that converge toward θ (the convergence of AQ-ML is exact, as shown in Section V-D). After convergence, the thresholds stay near θ, and this, as will be shown in Section V, is critical for getting good estimation performance. Secondly, the thresholds used by each scheme only depend on and can be inferred from the quantized data $\{b_k\}$. No extra bandwidth is needed to communicate the thresholds. Thirdly, the three AQ schemes together offer the flexibility of trading off between estimation accuracy and complexity. AQ-FS and AQ-VS, involving only simple algebraic calculations, are computationally attractive. At moderately increased complexity, AQ-ML offers the best performance among the three and converges to the best 1-bit quantizer, shown in Section V-D. Lastly, we note that AQ-FS and AQ-VS require no knowledge of the distribution of the data and can be considered as *nonparametric quantizers*. On the other hand, AQ-ML needs to know the pdf of the sensor observations.

IV. Maximum Likelihood Estimation

We now develop the ML estimators at the FC to find the final estimate of θ given the binary data $\{b_1, \ldots, b_N\}$ generated by the three AQ schemes. ML estimation is also employed at all but the first two sensors to determine the quantization threshold in AQ-ML, and the implementation is the same. To determine the likelihood function, we note that the binary data $b_1, b_2, \ldots, b_N$ generated by all three AQ schemes are *correlated*. This is in contrast to i.i.d. binary data generated by FQ [see (3)]. Using conditional probabilities, we can write the joint probability mass function (pmf) of $b_1, b_2, \ldots, b_N$ as

$$P(b_1, \ldots, b_N; \theta) = \prod_{n=1}^{N} P(b_n | b_1, \ldots, b_{n-1}; \theta) = \prod_{n=1}^{N} P(b_n | \tau_n; \theta) \quad (17)$$

where the conditional probability of b_n given the threshold τ_n used by sensor n is

$$P(b_n|\tau_n;\theta) = [F_w(\tau_n-\theta)]^{(1+b_n)/2} \times[1-F_w(\tau_n-\theta)]^{(1-b_n)/2}. \quad (18)$$

In the second equality of (17), the condition of the conditional probability is changed since b_n only depends on τ_n, which is a function of $\{b_1,\dots,b_{n-1}\}$ generated by the first $n-1$ sensors. It follows from (17) and (18) that the log-likelihood function (the dependence on $\{b_n\}_{n=1}^N$ is suppressed for notational simplicity)

$$L_{\rm AQ}(\theta)=\sum_{n=1}^N\left\{\left(\frac{1+b_n}{2}\right)\ln[F_w(\tau_n-\theta)] +\left(\frac{1-b_n}{2}\right)\ln[1-F_w(\tau_n-\theta)]\right\}. \quad (19)$$

While the likelihood functions corresponding to the three AQ schemes can all be expressed as in (19), it is important to note that the likelihood functions are different (and so are the MLEs) since the thresholds $\{\tau_n\}$ used by them are different. Specifically, AQ-FS computes $\{\tau_n\}$ from the quantized data via (10), AQ-VS uses(12) and (13), and AQ-ML employs recursive ML estimation as discussed in Section III-C to find its thresholds. We also note that the intermediate likelihood function $L_n(\theta;b_1,b_2,\dots,b_{n-1})$ in (16) can be obtained from the general form (19) by including only the first $n-1$ terms of the sum.

The MLEs corresponding to the three AQ schemes are given by

$$\hat{\theta}=\arg\max_{\theta} L_{\rm AQ}(\theta). \quad (20)$$

In general, (20) admits no closed-form solution, and a searching algorithm has to be utilized. For Gaussian distributed noise w_n, it is easy to show that the likelihood function is concave (e.g., [15]). Therefore, any one-dimensional gradient-based search starting from a random initial estimate is guaranteed to converge to the global maximum, and many efficient routines exist for this type of work (e.g., [26]).

V. Analysis

We evaluate the performance of the proposed AQ schemes through analysis of the corresponding CRB , a lower bound on the mean-squared error (MSE) that is asymptotically achieved by the MLE [27]. We first provide a general expression for the Fisher information that holds for all three AQ schemes, and then consider the CRB for each case. Asymptotic analysis of the CRB is emphasized that leads to insights to the behavior of the AQ approach.

A. Fisher Information

Noting that $F_w'(x)\triangleq(\partial F_w(x)/\partial x)=-p_w(x)$, we can quickly verify that the second-order derivative of $L_{\rm AQ}(\theta)$ is

$$\begin{aligned}\frac{\partial^2 L_{\rm AQ}(\theta)}{\partial\theta^2} &= \sum_{n=1}^N\left\{\left(\frac{1+b_n}{2}\right)\left(-\frac{p_w'(\tau_n-\theta)}{F_w(\tau_n-\theta)}-\frac{p_w^2(\tau_n-\theta)}{F_w^2(\tau_n-\theta)}\right)\right. \\ &\quad -\left(\frac{1-b_n}{2}\right)\left(-\frac{p_w'(\tau_n-\theta)}{[1-F_w(\tau_n-\theta)]}\right. \\ &\quad \left.\left.+\frac{p_w^2(\tau_n-\theta)}{[1-F_w(\tau_n-\theta)]^2}\right)\right\} \\ &\triangleq \sum_{n=1}^N A(b_n,\tau_n,\theta)\end{aligned} \quad (21)$$

where $p_w'(x)\triangleq(\partial p_w(x)/\partial x)$. The Fisher information for the estimation problem is given by (e.g., [27])

$$\begin{aligned}J_{\rm AQ}(\theta) &= -E\left\{\frac{\partial^2 L_{\rm AQ}(\theta)}{\partial\theta^2}\right\} \\ &= -\sum_{n=1}^N E_{b_n,\tau_n}\{A(b_n,\tau_n,\theta)\}\end{aligned} \quad (22)$$

where E_{b_n,τ_n} denotes the expectation with respect to the joint distribution of b_n and τ_n. Since

$$P(b_n,\tau_n;\theta)=P(\tau_n;\theta)P(b_n|\tau_n;\theta) \quad (23)$$

we can write

$$\begin{aligned}J_{\rm AQ}(\theta) &= -\sum_{n=1}^N E_{\tau_n}\left\{E_{b_n|\tau_n}[A(b_n,\tau_n,\theta)]\right\} \\ &\overset{(a)}{=} \sum_{n=1}^N E_{\tau_n}\left[\frac{p_w^2(\tau_n-\theta)}{F_w(\tau_n-\theta)(1-F_w(\tau_n-\theta))}\right] \\ &\overset{(b)}{=} \sum_{n=1}^N \int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n\end{aligned} \quad (24)$$

where E_{τ_n} denotes the expectation with respect to the distribution $P(\tau_n;\theta)$, $E_{b_n|\tau_n}$ denotes the expectation with respect to the conditional distribution $P(b_n|\tau_n;\theta)$, (a) follows from the fact that b_n is a binary random variable with $P(b_n=1|\tau_n,\theta)=F_w(\tau_n-\theta)$ and $P(b_n=-1|\tau_n,\theta)=1-F_w(\tau_n-\theta)$, and we define

$$G(\tau_n;\theta)\triangleq\frac{p_w^2(\tau_n-\theta)}{F_w(\tau_n-\theta)(1-F_w(\tau_n-\theta))}$$

in (b).

Note that for the Gaussian noise (and several other distributions, e.g., Cauchy, Laplacian, etc.), $G(\tau_n;\theta)$ is a unimodal, positive, and symmetric function achieving its maximum at $\tau_n=\theta$. Hence the Fisher information (24) is maximized when $P(\tau_n;\theta)=\delta(\tau-\theta)$, where $\delta(x)$ is the Dirac delta function.

This has the following two implications. First, the best achievable performance of the AQ schemes will not exceed that of the FQ approach with optimum threshold, i.e., $\tau = \theta$. Secondly, more thresholds set close to θ lead to a better performance.

To compute the exact Fisher information (24), we need to determine the distributions of $\{\tau_n\}$, i.e., $\{P(\tau_n;\theta)\}$. Clearly, these three AQ schemes result in different distributions of $\{\tau_n\}$. In the following, we analyze the CRBs of these three AQ schemes, with focus on AQ-FS and AQ-ML.

B. CRB of AQ-FS

For the AQ-FS, the threshold τ_n is a discrete random walk process with increment of Δ and $-\Delta$. For example, $\tau_1 \in \{0\}$, $\tau_2 \in \{\pm\Delta\}$, $\tau_3 \in \{-2\Delta, 0, 2\Delta\}$, $\tau_4 \in \{-3\Delta, -\Delta, \Delta, 3\Delta\}$, and so forth. In general, we have

$$\begin{aligned}\tau_{2n} &\in \{\pm\Delta, \ldots, \pm(2n-1)\Delta\}, \quad n = 1, 2, \ldots \\ \tau_{2n+1} &\in \{0, \pm 2\Delta, \ldots, \pm 2n\Delta\}, \quad n = 1, 2, \ldots. \end{aligned} \tag{25}$$

The pmf of τ_n can be recursively computed as follows. For notational convenience, let

$$P_{n,j} \triangleq P(\tau_n = j\Delta)$$

. It is clear that $P_{1,0} = 1$ for τ_1. Considering τ_2, we have

$$\begin{aligned} P_{2,1} &= P_{1,0} P(\tau_2 = \Delta | \tau_1 = 0) \\ &\quad + P_{1,2} P(\tau_2 = \Delta | \tau_1 = 2\Delta) \\ &= P_{1,0} P(\tau_2 = \Delta | \tau_1 = 0) \\ &= F_w(-\theta) \end{aligned} \tag{26}$$

and

$$\begin{aligned} P_{2,-1} &= P_{1,0} P(\tau_2 = -\Delta | \tau_1 = 0) \\ &\quad + P_{1,-2} P(\tau_2 = -\Delta | \tau_1 = -2\Delta) \\ &= P_{1,0} P(\tau_2 = -\Delta | \tau_1 = 0) \\ &= 1 - F_w(-\theta) \end{aligned} \tag{27}$$

where $P(\tau_{n+1} = j\Delta | \tau_n = i\Delta)$ denotes the transition probability from state $i\Delta$ to $j\Delta$ and is given by

$$P(\tau_{n+1} = j\Delta | \tau_n = i\Delta) = \begin{cases} 1 - F_w(i\Delta - \theta), & \text{if } i = j+1, \\ F_w(i\Delta - \theta), & \text{if } i = j-1, \\ 0, & \text{if } |i-j| > 1. \end{cases} \tag{28}$$

As a generalization, the pmf of τ_n is given by

$$\begin{aligned} P_{n,j} = P_{n-1,j-1} P(\tau_n = j\Delta | \tau_{n-1} = (j-1)\Delta) \\ + P_{n-1,j+1} P(\tau_n = j\Delta | \tau_{n-1} = (j+1)\Delta). \end{aligned} \tag{29}$$

Although the above recursive computation provides the exact solution, which can be used to determine the exact CRB for AQ-FS [24], it is not convenient to use. To gain additional insights, we examine the asymptotic distribution of τ_n as n increases.

Notice that $\{\tau_n\}$ form a Markov chain (see Fig. 3) with the transition probabilities given in (28). The convergence in distributions of $\{\tau_{2n}\}$ and $\{\tau_{2n+1}\}$ follows from a stationarity theorem (see [28, Lemma] for a detailed proof)

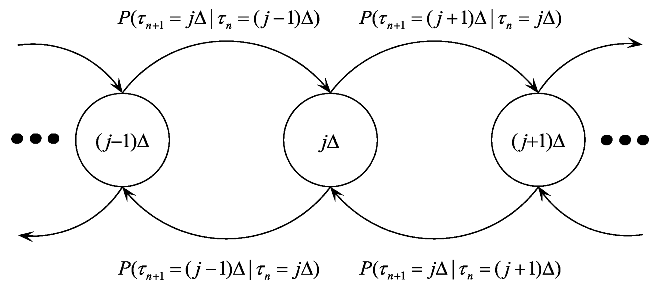


Fig. 3. Markov chain with state variables formed by the possible thresholds used by AQ-FS.

$$\lim_{n\to\infty} P_{2n,k} = P_{e,k}, \quad \lim_{n\to\infty} P_{2n+1,k} = P_{o,k}. \tag{30}$$

It is necessary to distinguish the distributions of $\{\tau_{2n}\}$ and $\{\tau_{2n+1}\}$ because τ_{2n} only contains odd states $\{\pm\Delta, \ldots, \pm(2n-1)\Delta\}$ and τ_{2n+1} only contains even states $\{0, \pm 2\Delta, \ldots, \pm 2n\Delta\}$. In other words, we have $P_{e,2i} = 0$ and $P_{o,2i+1} = 0$ for any integer i. Also, although the number of states grows linearly with n, they are composed of *atypical states* whose steady-state probability diminishes (as n increases) with increasing n and *typical states* whose steady-state probability remains significant with increasing n. Typical steady states are thresholds that are relatively close to θ (see Fig. 2(a)). For asymptotic analysis, we only need consider the typical states and the atypical states can be ignored. In view of this, we form the following vector:

$$\mathbf{p} \triangleq [P_{o,-2k}, P_{e,-2k+1}, P_{o,-2k+2}, P_{e,-2k+3}, \ldots, P_{o,2k}, P_{e,2k+1}]^T \tag{31}$$

where k is chosen large enough to include all typical states of the Markov chain. For Gaussian sensor noise, a choice of k that meets the condition $2k\Delta > |\theta| + 5\sigma_w$, where σ_w denotes the standard deviation of the sensor noise, would be more than enough. The steady-state probability vector $\mathbf{p}$ can be efficiently solved by using the transition equivalence principle and the unit probability constraints, whose details are addressed in Appendix A.

We plot $\mathbf{p}$ under different values of θ in Fig. 4, with $\sigma_w = 1$ for the Gaussian noise and $\Delta = 0.2$. We see that for different values of θ, the asymptotic distributions of $\{\tau_n\}$ preserve the same unimodal structure and achieve the maximum at θ. This justifies the effectiveness of our proposed AQ-FS scheme since it guarantees that with a sufficiently large number of sensors, the thresholds $\{\tau_n\}$ are around the unknown θ with a high probability.

With the above-derived asymptotic distribution of $\{\tau_n\}$, we are able to compute the asymptotic CRB of the AQ-FS scheme. We have the following results.

Proposition 1: The asymptotic CRB of the AQ-FS scheme is given by

$$\mathrm{CRB}_{\text{AQ-FS}}(\theta) \to \frac{2}{N\mathbf{g}^T\mathbf{p}} \tag{32}$$

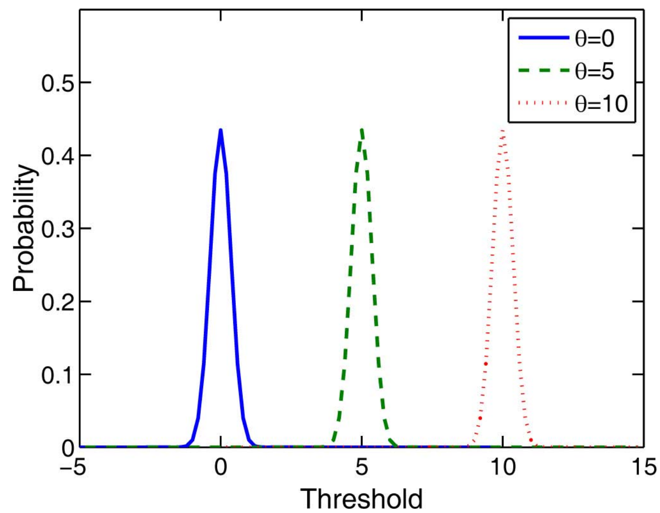


Fig. 4. Asymptotic distribution of τ under different values of θ.

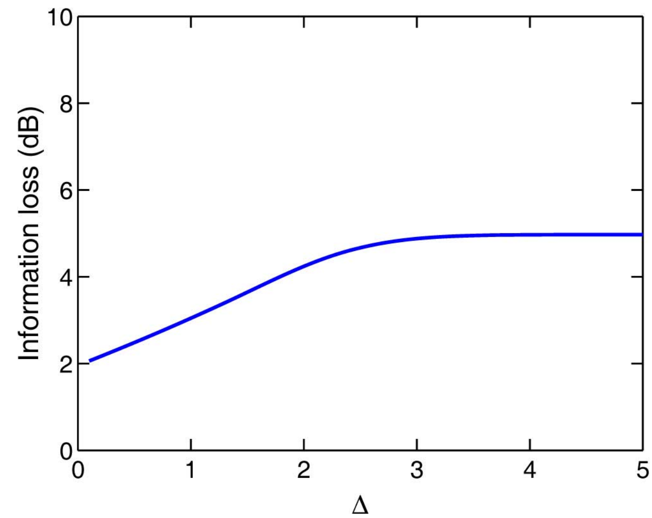


Fig. 5. AQ-FS: asymptotic information loss versus Δ.

where $\mathbf{g}$ is defined as

$$\mathbf{g} \triangleq [G(-k\Delta;\theta) \quad G((-k+1)\Delta;\theta) \quad \ldots \quad G((k+1)\Delta;\theta)]^T.$$

Proof: See Appendix B. ■

To better evaluate the performance of the proposed AQ approach, we borrow from [14] the concept of information loss, which is defined as the ratio (in decibels) of the CRB for the AQ scheme to the CRB for the clairvoyant estimator using unquantized data [see (6)]

$$\gamma \triangleq 10\log_{10}\frac{\mathrm{CRB}_{\mathrm{AQ}}(\theta)}{\mathrm{CRB}_{\mathrm{NQ}}(\theta)}. \quad (33)$$

The asymptotic information loss of the AQ-FS scheme, which is defined as the information loss asymptotically incurred by using AQ-FS for quantization, can be obtained by replacing the numerator in (33) with the asymptotic CRB of (32). Fig. 5 depicts the asymptotic information loss as a function of the step size Δ, where we set $\sigma_w = 1$ for the Gaussian noise. As we can see, a smaller Δ helps reduce the information loss. Also, additional performance degradation incurred by increasing Δ is mild: even with $\Delta = 5\sigma_w$, the loss is within 5 dB.

The above results suggest us to choose a small step size Δ as long as the thresholds $\{\tau_n\}$ can reach the convergence. However, in scenarios with a limited number of sensors, if Δ is chosen too small, the thresholds are kept in the catching-up phase and never attain the settling phase (see Section III-B). This, as will be shown in Section VI, results in a sharp performance degradation. Since a larger Δ incurs a mild information loss, it is safer to use a larger Δ than a smaller Δ when no information of θ is available.

C. CRB of AQ-VS

As observed from (24), we expect a better performance when more thresholds are set close to the unknown parameter θ. The AQ-VS scheme adjusts the step size adaptively according to previous two quantized data, which is able to have the thresholds $\{\tau_n\}$ come more quickly around θ and keep a lower granular noise level [see Fig. 2(b)]. Hence it generally provides a performance improvement relative to the AQ-FS scheme. However, a further analysis on the AQ-VS scheme is impeded by finding the exact distributions of $\{\tau_n\}$, i.e., $\{P(\tau_n;\theta)\}$. This is because the thresholds $\{\tau_n\}$, adjusted by a variable step size, take on much more possible values than those in the AQ-FS scheme. Experimental results show that the number of possible values for τ_n increases exponentially with n. The exact computation of $P(\tau_n;\theta)$ is, therefore, cumbersome, especially when the number of sensors N is large. Nevertheless, (24) can still be evaluated numerically by Monte Carlo integration.

D. CRB of AQ-ML

The AQ-ML scheme computes the ML estimate of θ as the threshold. Since the ML estimator is a nonlinear function, as the AQ-VS scheme, the threshold τ_n is a discrete random variable with the number of possible values for τ_n increases exponentially with n. Specifically, sensor n has 2^{n-1} possible threshold values, with each value chosen with a certain probability. Hence, we encounter the same problem as that in AQ-VS, and numerical integration can be employed to evaluate (24).

To circumvent the difficulty in computing the exact $P(\tau_n;\theta)$, we examine the asymptotic performance, which offers additional insight into the AQ-ML scheme. We have the following results.

Proposition 2: For Gaussian sensor noise w_n with zero-mean and variance σ_w^2, the CRB of the AQ-ML scheme converges to $\pi/2$ times that of the clairvoyant estimator (6) as N increases, i.e.,

$$\mathrm{CRB}_{\text{AQ-ML}}(\theta) \rightarrow \frac{\pi\sigma_w^2}{2N} \approx 1.57\frac{\sigma_w^2}{N}. \quad (34)$$

Proof: See Appendix C. ■

From Proposition 2, we see that the information loss of the AQ-ML scheme asymptotically achieves $10\log_{10}(\pi/2)$, which is also attained by the best 1-bit quantizer, i.e., the FQ approach with an optimum (although practically infeasible) choice of threshold $\tau = \theta$ (see Section II). This indicates that our AQ

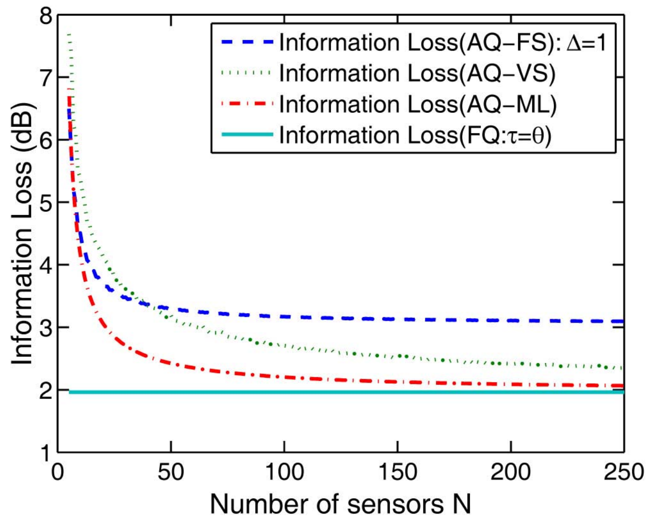


Fig. 6. Information loss of the three AQ schemes and the FQ approach with the optimum threshold $\tau_{\rm opt} = \theta$.

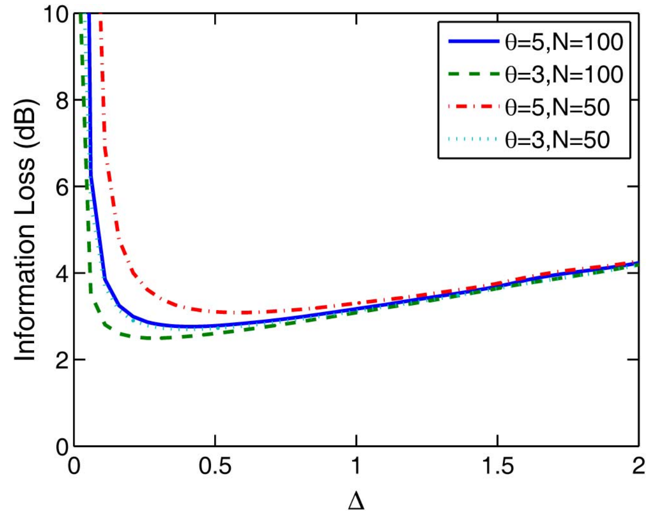


Fig. 7. AQ-FS: Information loss versus Δ.

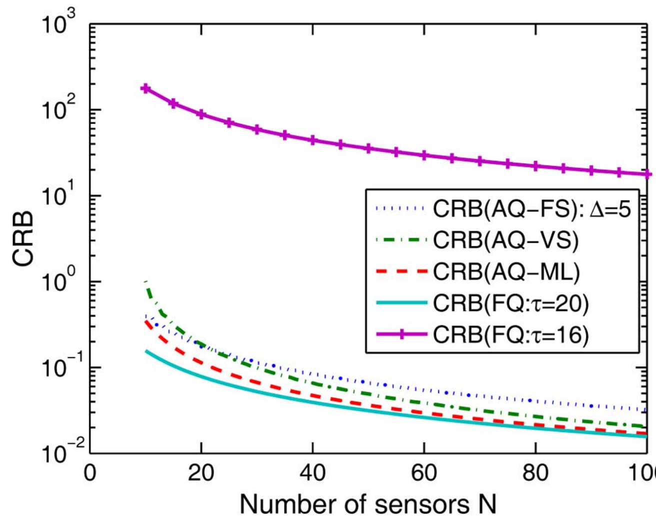


Fig. 8. CRBs of the three AQ schemes and FQ approach with different thresholds when $\theta = 20$.

scheme adaptively finds the best threshold by learning from prior transmissions.

VI. Simulation Results

In this section, we illustrate the performance of our proposed three AQ schemes. The noise $\{w_n\}$ are i.i.d. Gaussian random variables with zero mean and variance $\sigma_w^2 = 1$ throughout the following examples.

A. *Information Loss Relative to the Clairvoyant Estimator*

We first examine the information loss [defined in (33)] of the three AQ schemes relative to the clairvoyant estimator using unquantized data. We set $\theta = 5$, $\Delta = 1$ for the AQ-FS and $K = 1.2$ for the AQ-VS. Fig. 6 shows the information loss of the three AQ schemes as a function of the number of sensors N. It can be seen that for all three AQ schemes, the information loss decreases with an increasing N. This is because the AQ schemes benefit from the previous transmissions by adaptively choosing a proper quantization threshold. Also, we observe that the information loss of the AQ-ML scheme approaches that of FQ with the optimum threshold, i.e., $\tau = \theta$, which corroborates our previous claim in Proposition 2.

The effect of the choice of the step size Δ for the AQ-FS scheme is next investigated. Fig. 7 shows the information loss of the AQ-FS scheme under different choices of Δ. It is seen that the optimal choice of Δ is related to the unknown parameter θ and the number of sensors N. This can be intuitively justified because a larger θ (for a fixed N) requires a larger step size to move up quickly close to the unknown parameter; likewise, a smaller N (for a fixed θ) requires a larger Δ to achieve the same effect. Another observation made on the figure is that, around the optimal value of Δ, the performance degrades significantly for smaller Δ, while it remains fairly flat for larger Δ. This suggests that it is safer to use a larger Δ than a smaller Δ when no information of θ is available.

B. *Comparison to the FQ Approach*

Fig. 8 shows the CRBs of the three AQ schemes and the FQ approach. For the FQ approach, $\tau = \theta$ is the optimum choice. When $\tau = \theta = 20$, the performance of FQ achieves the best among all one-bit estimators. However, as we also see from the figure, the FQ approach is very sensitive to the value of τ; as the threshold τ becomes farther apart from θ (even not too far apart), the performance of the FQ degrades significantly. Since θ to be estimated is unknown, the choice of τ is always a tricky issue. Our proposed AQ schemes do not have the above problem. In particular, the performance of the AQ-ML scheme approaches that of the FQ with the optimum threshold ($\tau = \theta$) without knowing the true θ.

C. *Comparison to the RG Algorithm [15]*

We compare our proposed AQ schemes with the multiple thresholding algorithm (denoted "RG") introduced in [15, Section IV]. For the RG, the knowledge of the pdf of the unknown θ, $p_\theta(x)$, also called the weighting function therein, is required

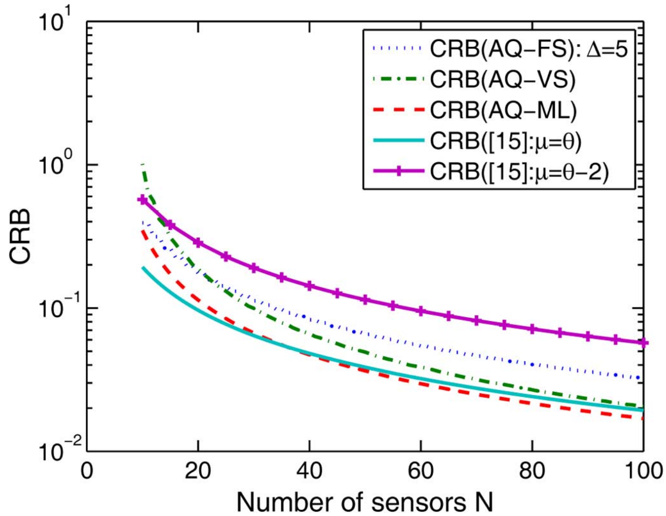


Fig. 9. CRBs of the three AQ schemes and RG approach with different means when $\theta = 20$.

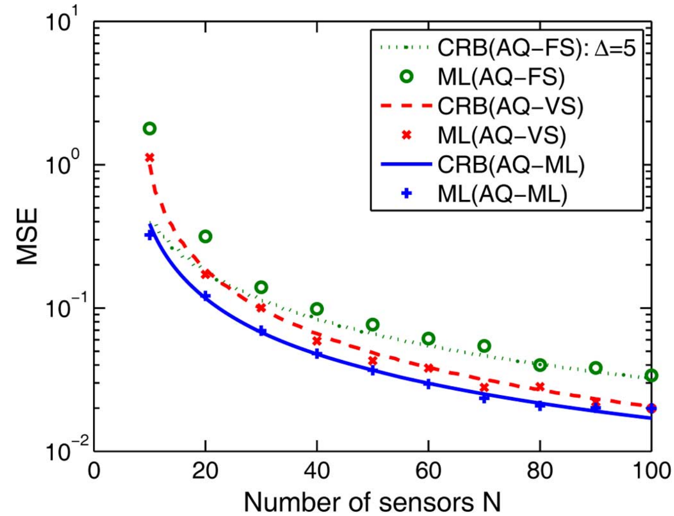


Fig. 10. MSEs of the MLEs for the three AQ schemes when $\theta = 20$.

to determine the frequencies $\{\rho_k\}$, with which the kth sensor's threshold τ_k is assigned. Here, we assume $p_\theta(x)$ is Gaussian (also used in [15]). To examine the impact of the accuracy of the prior knowledge on the performance of the RG, we consider two difference cases for $p_\theta(x)$, namely, $p_\theta(x) \sim \mathcal{N}(\theta, 1)$ and $p_\theta(x) \sim \mathcal{N}(\theta - 2, 1)$, respectively. The first choice represents an ideal though impractical case since θ is unknown in general. The second choice reflects the fact that, in practice, the prior knowledge (the mean $p_\theta(x)$) might be off from the unknown θ, and we use this choice to determine how sensitive the RG is to such mismatch. The set of thresholds $\{\tau_k\}$ used in the RG algorithm are distributed with uniform spacing $\tau_{k+1} - \tau_k = 2$ (note that smaller spacings result in similar performance).

Fig. 9 shows the CRBs of the proposed AQ schemes and the RG with $\mu = \theta$ and $\mu = \theta - 2$, respectively. From the figure, we see that, for the RG approach, a slight deviation of the mean from θ causes a performance degradation, which indicates that the RG approach is a bit sensitive to the knowledge of the prior pdf of θ. To obtain a good performance, an accurate knowledge of its mean is required. We also observe that all three AQ schemes present performance advantages over the RG when $\mu = \theta - 2$. In particular, without any prior information of the unknown parameter, the AQ-ML scheme even achieves a better performance than RG with $\mu = \theta$ when N is not too small.

D. MLE

The mean square errors of the maximum likelihood estimators for the three AQ schemes are included and compared with the corresponding CRB in Fig. 10. It is observed that the MSEs approach the CRB asymptotically with an increasing N.

VII. Concluding Remarks

An adaptive quantization approach was introduced for distributed estimation in bandwidth/power constrained WSNs. Through a sequential transmission strategy, each sensor designs its quantizer by learning from prior transmissions from other sensors. Three adaptive AQ schemes were proposed and their corresponding MLEs developed. Asymptotic CRB analysis shows that, without any prior knowledge of θ, the proposed 1-bit AQ schemes are able to achieve a consistent performance with a mild performance degradation as compared with the clairvoyant estimator using unquantized observations. In particular, for the ML-based AQ scheme, the information loss is as low as 1.96 dB, which is equivalent to that of the best 1-bit quantizer using the theoretically optimum but practically infeasible threshold. Extensive simulation results have been presented to illustrate the effectiveness of the AQ schemes and the advantages over existing techniques.

While we considered only the 1-bit (per sample) quantization case, our AQ approach can be extended for multibit quantization. Consider, for example, AQ-FS. Instead of using a 1-bit quantizer to just take the sign of the difference between the current observation x_n and quantization threshold τ_n, a multibit quantizer (either uniform and nonuniform) can be used to quantize $x_n - \tau_n$ and provide finer adjustment of the subsequent quantization threshold. This multibit AQ-FS effectively uses a number of step sizes (as opposed to a fixed step size in 1-bit AQ-FS) determined by the number of bits used for quantization. Extensions of AQ-VS and AQ-ML are also possible and will be reported elsewhere.

It should be noted that the effectiveness of our proposed schemes is based on the fact that the closer the quantization thresholds come to the unknown parameter to be estimated, the better performance the estimator achieves. This fact is true for the additive noise model as considered in this paper with several noise distributions including Gaussian, Laplacian, Cauchy, etc. While this model itself covers a range of important applications, AQ is a rather general approach for solving other distributed estimation problems. For these problems, the relationship between the optimal threshold and the unknown parameter may no longer be identical. In such cases, the principle of AQ still applies. In particular, we can first consider FQ in which each sensor employs one or a set of common quantization thresholds, depending on the problem. We can then find the relationship between the optimal quantization threshold(s) and the unknown parameter through the CRB analysis. This optimum quantizer is again practically infeasible due to its dependence on the unknown parameter. It can be replaced by an adaptive solution that sequentially updates the threshold(s) such that it approaches the optimum one by adaptive learning.

APPENDIX A
THE STEADY STATE PROBABILITY VECTOR $\mathbf{p}$

Considering the transitions between $\{\tau_{2n}\}$ and $\{\tau_{2n+1}\}$ as $n \to \infty$, we can easily verify that

$$\begin{aligned} P_{e,2j+1} &= P_{o,2j} P(\tau_{2n} = (2j+1)\Delta | \tau_{2n-1} = 2j\Delta) \\ &\quad + P_{o,2j+2} P(\tau_{2n} = (2j+1)\Delta | \tau_{2n-1} = (2j+2)\Delta) \\ P_{o,2j} &= P_{e,2j-1} P(\tau_{2n+1} = 2j\Delta | \tau_{2n} = (2j-1)\Delta) \\ &\quad + P_{e,2j+1} P(\tau_{2n+1} = 2j\Delta | \tau_{2n} = (2j+1)\Delta) \end{aligned}$$

where the transition probability is given by (28) and is independent of n. Since k is chosen large enough to include all typical states, it is reasonable to assume that the probabilities $\{P_{e,\pm 2k_1+1}\}$ and $\{P_{o,\pm 2k_1}\}$ are zero for $k_1 > k$. With the two-sided bounding, we can rewrite the above equations in matrix form as

$$\mathbf{p} = \mathbf{T}\mathbf{p} \tag{35}$$

where $\mathbf{T}$ is the transition matrix given in (36), with $t_{i,j} \triangleq P(\tau_{n+1} = j\Delta | \tau_n = i\Delta)$. See (36) at the bottom of the page.

We see that $\mathbf{T}$ is a sparse matrix and its nonzero entries can be predetermined for specified values of σ_w and Δ. The special structure of $\mathbf{T}$ enables us to recursively express the entries of $\mathbf{p}$ in terms of its first entry $P_{o,-2k}$. For example, we have

$$\begin{aligned} P_{e,-2k+1} &= \frac{P_{o,-2k}}{t_{-2k+1,-2k}} \\ P_{o,-2k+2} &= \left(\frac{1}{t_{-2k+1,-2k}} - t_{-2k,-2k+1} \right) \\ &\quad \times \frac{P_{o,-2k}}{t_{-2k+2,-2k+1}}. \end{aligned}$$

Using the unit probability constraints

$$\begin{aligned} \sum_{i=0}^{2k} P_{o,-2k+2i} &= 1, \\ \sum_{i=0}^{2k} P_{e,-2k+2i+1} &= 1 \end{aligned} \tag{37}$$

the vector $\mathbf{p}$ can be readily solved. In fact, one of the constraints can help to solve $\mathbf{p}$, and it turned out that the other constraint is satisfied automatically.

APPENDIX B
PROOF OF PROPOSITION 1

We express (24) as the summation of the following two terms:

$$\begin{aligned} J_{\text{AQ-FS}}(\theta) &= \sum_{n=1}^{N} \int P(\tau_n;\theta) G(\tau_n;\theta) d\tau_n \\ &= \sum_{n=1}^{N_c} \int P(\tau_n;\theta) G(\tau_n;\theta) d\tau_n \\ &\quad + \sum_{n=N_c+1}^{N} \int P(\tau_n;\theta) G(\tau_n;\theta) d\tau_n \\ &\triangleq J_1 + J_2 \end{aligned} \tag{38}$$

where N_c is chosen to ensure the distribution of $P(\tau_n;\theta)$, for $n > N_c$, reaches its convergence or has a negligible difference from the steady-state probability vector $\mathbf{p}$. We have

$$N_c \frac{2}{\pi\sigma_w^2} > J_1 > 0 \tag{39}$$

and

$$J_2 \overset{(a)}{=} \frac{1}{2}(N - N_c)\mathbf{g}^T\mathbf{p} \tag{40}$$

where (a) comes from the fact that $P(\tau_n;\theta)$ is discrete and invariant for $n > N_c$. Therefore, the CRB is upper and lower bounded as

$$\frac{N}{N-N_c}\frac{2}{N\mathbf{g}^T\mathbf{p}} > \text{CRB}_{\text{AQ-FS}}(\theta) > \frac{N}{N+N_c\xi}\frac{2}{N\mathbf{g}^T\mathbf{p}} \tag{41}$$

where $\xi \triangleq (2/(\pi\sigma_w^2)) - (\mathbf{g}^T\mathbf{p}/2)$.

Since $N \gg N_c$, the upper bound and the lower bound both approach $(2/(N\mathbf{g}^T\mathbf{p}))$. The proof is completed here.

APPENDIX C
PROOF OF PROPOSITION 2

Note that sensor m computes its threshold τ_m as

$$\begin{aligned} \tau_m &= \arg\max_\theta \log P(b_1, \ldots, b_{m-1};\theta) \\ &= \arg\max_\theta \sum_{k=1}^{m-1} \left\{ \left(\frac{1+b_k}{2}\right) \ln[F_w(\tau_k - \theta)] \right. \\ &\qquad \left. + \left(\frac{1-b_k}{2}\right) \ln[1 - F_w(\tau_k - \theta)] \right\}. \end{aligned} \tag{42}$$

$$\mathbf{T} = \begin{bmatrix} 0 & t_{-2k+1,-2k} & 0 & \cdots & & & 0 \\ t_{-2k,-2k+1} & 0 & t_{-2k+2,-2k+1} & 0 & \cdots & & 0 \\ 0 & t_{-2k+1,-2k+2} & 0 & t_{-2k+3,-2k+2} & 0 & \cdots & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & \cdots & & & t_{2k-1,2k} & 0 & t_{2k+1,2k} \\ 0 & \cdots & & & & t_{2k,2k+1} & 0 \end{bmatrix} \tag{36}$$

It can be easily verified that the log-likelihood function $\log P(b_1,\ldots,b_{m-1};\theta)$ satisfies the "regularity" conditions and, hence, for large data records (i.e., m is large), the ML estimate τ_m is consistent [27], [29]. Consequently, for any small $\epsilon > 0$ and $\varepsilon > 0$, we can find a sufficiently large m such that

$$P(|\tau_n-\theta|<\epsilon) > 1-\varepsilon \quad n\geq m. \tag{43}$$

Considering (24), we express $J_{\text{AQ-ML}}(\theta)$ as the summation of the following two terms:

$$\begin{aligned} J_{\text{AQ-ML}}(\theta) &= \sum_{n=1}^{m-1}\int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n \\ &\quad + \sum_{n=m}^{N}\int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n \\ &\triangleq J_1 + J_2 \end{aligned} \tag{44}$$

where m is chosen to satisfy (43). By utilizing the properties of the function $G(\tau_n;\theta)$, the first term and the second term of (44) can be bounded as follows, respectively:

$$\begin{aligned} 0 < J_1(\theta) &\triangleq \sum_{n=1}^{m-1}\int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n \\ &< \sum_{n=1}^{m-1} G(\tau_n=\theta;\theta) = \frac{2(m-1)}{\pi\sigma_w^2} \end{aligned} \tag{45}$$

$$J_2(\theta) < \frac{2(N-m+1)}{\pi\sigma_w^2}. \tag{46}$$

On the other hand, we have

$$\begin{aligned} J_2(\theta) &\triangleq \sum_{n=m}^{N}\int P(\tau_n;\theta)G(\tau_n;\theta)d\tau_n \\ &= \sum_{n=m}^{N}\left[\int P(|\tau_n-\theta|<\epsilon)G(\tau_n;\theta)d\tau_n \right. \\ &\qquad \left. + \int P(|\tau_n-\theta|\geq\epsilon)G(\tau_n;\theta)d\tau_n\right] \\ &> \sum_{n=m}^{N}\int P(|\tau_n-\theta|<\epsilon)G(\tau_n;\theta)d\tau_n \\ &> \sum_{n=m}^{N}(1-\varepsilon)G(\tau_n=\theta-\epsilon;\theta) \\ &\overset{(a)}{>} (N-m+1)(1-\varepsilon)\frac{4}{\sqrt{2\pi}\sigma_w}p_w(\epsilon) \\ &\overset{(b)}{>} (N-m+1)\frac{2}{\pi\sigma_w^2}(1-\varepsilon)\left(1-\frac{\epsilon^2}{2\sigma_w^2}\right) \end{aligned} \tag{47}$$

where (a) comes from the Chernoff bound $F_w(x)(1-F_w(x)) \leq (1/4)e^{-(x^2/2\sigma_w^2)}$ and (b) follows from the Taylor expansion. Since $(2(m-1)/(\pi\sigma_w^2)) > J_1(\theta) > 0$, we can further write

$$J_1(\theta) = \frac{2(m-1)\eta}{\pi\sigma_w^2} \tag{48}$$

where $0<\eta<1$.

Combining (44)–(48), we therefore have

$$\begin{aligned} \frac{2N}{\pi\sigma_w^2} > J_{\text{AQ-ML}}(\theta) &> (N-m+1)\frac{2}{\pi\sigma_w^2}\xi + \frac{2(m-1)\eta}{\pi\sigma_w^2} \\ &= \frac{2N\xi}{\pi\sigma_w^2} - \frac{2(m-1)\eta'}{\pi\sigma_w^2} \end{aligned} \tag{49}$$

where $\xi \triangleq (1-\varepsilon)\left(1-\epsilon^2/(2\sigma_w^2)\right)$ and $\eta' \triangleq \xi-\eta$. The CRB is, therefore, lower bounded and upper bounded by

$$\begin{aligned} \frac{\pi\sigma_w^2}{2N} < \text{CRB}_{\text{AQ-ML}}(\theta) &< \frac{\pi\sigma_w^2}{2}\frac{1}{N\xi-(m-1)\eta'} \\ &= \frac{\pi\sigma_w^2}{2N}\frac{1}{\xi-\frac{(m-1)\eta'}{N}}. \end{aligned} \tag{50}$$

Considering $N \gg m$ and m is sufficiently large to ensure $\epsilon\to 0$ and $\varepsilon \to 0$, i.e., $\xi \to 1$, we have

$$\text{CRB}_{\text{AQ-ML}}(\theta) \to \frac{\pi\sigma_w^2}{2N} \approx 1.57\frac{\sigma_w^2}{N}. \tag{51}$$

The proof is completed here.

References

[1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, pp. 17–29, Mar. 2002.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, pp. 102–114, Aug. 2002.

[3] R. R. Tenney and N. R. Sandell, Jr, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, pp. 501–510, Jul. 1981.

[4] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, pp. 554–568, Apr. 1996.

[5] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Process.*, vol. 53, pp. 2617–2624, Aug. 2005.

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

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

[8] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 51–60, Mar. 2002.

[9] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. 2005 IEEE Int. Symp. Inf.*, Sep. 2005.

[10] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 53, pp. 1631–1639, May 2005.

[11] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1456–1459, Jul. 1993.

[12] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. Commun.*, vol. 48, pp. 181–184, Feb. 2000.

[13] W. M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Trans. Commun.*, vol. 41, pp. 1602–1605, Nov. 1993.

[14] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, pp. 978–1002, Mar. 2001.

[15] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part I: Gaussian PDF," *IEEE Trans. Signal Process.*, vol. 54, pp. 1131–1143, Mar. 2006.

[16] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2210–2219, Jun. 2005.

[17] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part II: Unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, pp. 2784–2796, Jul. 2006.

[18] Z. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 735–744, Apr. 2005.

[19] Z. Luo and J. Xiao, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3564–3575, Oct. 2005.

[20] P. Venkitasubramaniam, L. Tong, and A. Swami, "Quantization for maximin ARE in distributed estimation," *IEEE Trans. Signal Process.*, vol. 55, pp. 3596–3605, Jul. 2007.

[21] Y. Huang and Y. Hua, "Multi-hop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, pp. 1004–1007, Dec. 2007.

[22] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and dynamic sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, Feb. 2006.

[23] A. Ribeiro and G. B. Giannakis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Trans. Signal Process.*, vol. 54, pp. 4782–4795, Dec. 2006.

[24] H. Li and J. Fang, "Distributed adaptive quantization and estimation for wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, pp. 669–672, Oct. 2007.

[25] B. Chen, L. Tong, and P. K. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 16–26, Jul. 2006.

[26] D. A. Pierre, *Optimization Theory With Applications*. New York: Wiley, 1969.

[27] S. M. Kay, *Fundamentals of Statistical Signal Process.: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[28] T. L. Fine, "The response of a particular nonlinear system with feedback to each of two random processes," *IEEE Trans. Inf. Theory*, vol. IT-14, pp. 255–264, Mar. 1968.

[29] M. J. Crowder, "Maximum likelihood estimation for dependent observations," *J. Roy. Statist. Soc. B (Methodological)*, vol. 38, no. 1, pp. 45–53, 1976.

Jun Fang (M'08) received the B.Sc. and M.Sc. degrees in electrical engineering from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree in electrical engineering from National University of Singapore, Singapore, in 2006.

During 2006, he was with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, as a Postdoctoral Research Associate. Currently, he is a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ. His research interests include statistical signal processing, wireless communications, and distributed estimation and detection with their applications on wireless sensor networks.

Hongbin Li (M'99) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory, Rome, NY, in the summers of 2003 and 2004. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is an Associate Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He is a member of the Sensor Array and Multichannel Technical Committee of the IEEE Signal Processing Society. He is or has been an Editor or Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was a Guest Editor for the *EURASIP Journal on Advances in Signal Processing* Special Issue on Distributed Signal Processing Techniques for Wireless Sensor Networks.