Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks—Part I: Gaussian Case

Alejandro Ribeiro, *Student Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—We study deterministic mean-location parameter estimation when only quantized versions of the original observations are available, due to bandwidth constraints. When the dynamic range of the parameter is small or comparable with the noise variance, we introduce a class of maximum-likelihood estimators that require transmitting just one bit per sensor to achieve an estimation variance close to that of the (clairvoyant) sample mean estimator. When the dynamic range is comparable or larger than the noise standard deviation, we show that an optimum quantization step exists to achieve the best possible variance for a given bandwidth constraint. We will also establish that in certain cases the sample mean estimator formed by quantized observations is preferable for complexity reasons. We finally touch upon algorithm implementation issues and guarantee that all the numerical maximizations required by the proposed estimators are concave.**

***Index Terms*—Parameter estimation, wireless sensor networks.**

I. Introduction

Wireless sensor networks (WSNs) comprise a large number of geographically distributed nodes characterized by low power constraints and limited computation capability. However, with sensor collaboration, potentially powerful networks can be constructed to monitor and control environments [9]. While a number of works address sensor collaboration for distributed detection (see, e.g., [22], [23], and references therein), the equally challenging problem of distributed estimation has not received much attention. In distributed estimation for WSN each sensor has available a subset of the observations that must be either transmitted to a central node (WSN with a fusion center) or shared among nodes (ad hoc WSN). Various design and implementation issues were pursued in early works [2], [7], [10]. More recently, a great deal of attention has focused on decentralized algorithms exploiting spatial correlation to reduce transmission requirements [3], [4], [6], [14], [17], [19].

A related problem for both types of WSN (ad hoc, or with a fusion center) is that bandwidth is limited, necessitating the estimator to be formed using quantized versions of the original observations. In this setup, quantization becomes an integral part of the estimation process, since one may think of quantization as a means of constructing *binary observations*. We then deal with *parameter estimation given a set of binary observations*, as opposed to the (clairvoyant) parameter estimation based on unquantized (analog) measurements. When the noise probability density function (pdf) is known, transmitting a single bit per sensor can lead to minimal loss in the estimator variance compared with the clairvoyant estimator [1], [15]. Various algorithms have been introduced in [15] to reduce the worst case variance increase in certain setups. Alternatively, when the noise pdf is unknown, universal (pdf-unaware) estimators based on quantized sensor data have been introduced recently [11]–[13].

Our focus here is on bandwidth-constrained distributed mean-location parameter estimation in additive white Gaussian noise (AWGN). We seek maximum-likelihood estimators (MLEs) and benchmark their variances with the Cramer–Rao lower bound (CRLB) that, at least asymptotically, is achieved by the MLE. We will show that the deciding factor in the choice of the estimator is the signal-to-noise ratio (SNR), defined here as the dynamic range of the parameter square over the observation noise variance.

Our approach is motivated by the observation that an estimator based on the transmission of a single binary observation per sensor can have variance as small as $\pi/2$ times that of the clairvoyant sample mean estimator (Section III). This result was derived first in [15] and is included here as a motivational starting point. By noting that this excellent performance can only be achieved under careful design choices, we introduce a class of estimators that minimize the average variance over a given weight function, establishing that *in the low-to-medium SNR range this class of MLE performs close to the clairvoyant estimator's variance* (Section III). We then turn our attention to the high SNR regime, and show that a quantization step close to the noise's standard deviation is nearly optimal in the sense of minimizing a properly defined per-bit CRLB (Section V), establishing a second result, on the *optimal number of bits per sensor to be transmitted*. The sample mean estimator based on quantized observations is subsequently analyzed to show that at high SNR even a simple-minded estimator requires transmission of only a small number of extra bits than the MLE. This allows us to establish analytically that *bandwidth-constrained distributed estimation is not a relevant problem in high SNR scenarios*. For such cases, we advocate using the sample mean estimator based on the quantized observations for its low com-

Manuscript received July 14, 2004; revised December 22, 2004. Work in this paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government. Part of the results in this paper appeared in [20]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yucel Altunbasak.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: aribeiro@ece.umn.edu; georgios@ece.umn.edu).

Digital Object Identifier 10.1109/TSP.2005.863009

plexity (Section VI). The last conclusion of the present paper is that *numerical maximization required by our MLE can be posed as a convex optimization problem*, thus ensuring convergence of e.g., Newton-type iterative algorithms. We finally present numerical results in Section VII and conclude the paper in Section VIII.

II. Problem Statement

This paper considers the problem of estimating a deterministic scalar parameter θ in the presence of zero-mean AWGN

$$x(n) = \theta + w(n), \quad n = 0, 1, \ldots, N-1 \tag{1}$$

where $w(n) \sim \mathcal{N}(0, \sigma^2)$, and n is the sensor index. Throughout, we will use $p(w) := 1/(\sqrt{2\pi}\sigma)\exp[-w^2/(2\sigma^2)]$ to denote the noise pdf.

If all the observations $\{x(n)\}_{n=0}^{N-1}$ were available, the MLE of θ would be the Sample Mean Estimator, $\bar{x} = N^{-1}\sum_{n=0}^{N-1} x(n)$. Rightfully, this can be regarded as a clairvoyant estimator for the bandwidth constrained problem, whose variance is known to be [8, p. 30]

$$\mathrm{var}(\bar{x}) = \frac{\sigma^2}{N}. \tag{2}$$

Due to bandwidth limitations, however, the observations $x(n)$ have to be quantized and estimation can only be based on these quantized values. To this end, we will henceforth think of quantization as the construction of a set of indicator variables (that will be referred to, as binary observations)

$$b_k(n) = \mathbf{1}\{x(n) \in (\tau_k, +\infty)\}, \quad k \in \mathbf{Z} \tag{3}$$

where τ_k is a threshold defining $b_k(n)$, $\mathbf{Z}$ denotes the set of integers, and k is used to index the set of binary observations constructed from the observation $x(n)$. The bandwidth constraint manifests itself in dictating estimation of θ to be based on the binary observations $\{b_k(n), k \in \mathbf{Z}\}_{n=0}^{N-1}$. The goal of this paper is twofold: 1) *to develop the MLE for estimating θ given a set of binary observations*, and 2) *to study the associated CRLB*—a bound that is achieved by the MLE as $N \to \infty$.

Instrumental to the ensuing derivations is the fact that each $b_k(n)$ in (3) is a Bernoulli random variable with parameter

$$q_k(\theta) := \Pr\{b_k(n) = 1\} = F(\tau_k - \theta), \quad k \in \mathbf{Z} \tag{4}$$

where $F(x) := 1/(\sqrt{2\pi}\sigma)\int_x^{+\infty}\exp(-u^2/2\sigma^2)du$ is the complementary cumulative distribution function (CDF) of $w(n)$.

The problem under consideration bears similarities and differences with quantization. On the one hand, for a fixed n the set of binary observations $\{b_k(n), k \in \mathbf{Z}\}$ specifies uniquely the quantized value of $x(n)$ to one of the prespecified levels $\{\tau_k, k \in \mathbf{Z}\}$. On the other hand, different from quantization in which the goal is to *reconstruct* $x(n)$ (and the optimum solution is known to be given by Lloyd's quantizer [18, p. 108]); our goal here is to *estimate* θ.

III. MLE Based on Binary Observations: Common Thresholds

Let us consider the most stringent bandwidth constraint, requiring sensors to transmit one bit per $x(n)$ observation. And as a simple first approach, let every sensor use the same threshold τ_c to form

$$b(n) = \mathbf{1}\{x(n) \in (\tau_c, +\infty)\}, \quad n = 0, 1, \ldots, N-1. \tag{5}$$

Dropping the subscript k, we let $\mathbf{b} := [b(0), \ldots, b(N-1)]^T$, and denote as $q(\theta)$ the parameter of these Bernoulli variables. We are now ready to derive the MLE and the pertinent CRLB.

Proposition 1: [15] The MLE $\hat{\theta}$ based on the vector of binary observations $\mathbf{b}$ is given by

$$\hat{\theta} = \tau_c - F^{-1}\left(\frac{1}{N}\sum_{n=0}^{N-1} b(n)\right). \tag{6}$$

Furthermore, the CRLB for any unbiased estimator $\hat{\theta}$ based on $\mathbf{b}$ is given by

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{N}\left[\frac{p^2(\tau_c-\theta)}{F(\tau_c-\theta)\left[1-F(\tau_c-\theta)\right]}\right]^{-1} := B(\theta). \tag{7}$$

Proof: Due to the noise independence, the pdf of $\mathbf{b}$ is $p(\mathbf{b}, \theta) = \prod_{n=0}^{N-1}[q(\theta)]^{b(n)}[1-q(\theta)]^{1-b(n)}$. Taking logarithm yields the log-likelihood

$$L(\theta) = \sum_{n=0}^{N-1} b(n)\ln(q(\theta)) + (1-b(n))\ln(1-q(\theta)) \tag{8}$$

whose second derivative with respect to θ is

$$\begin{aligned}\ddot{L}(\theta) = &\sum_{n=0}^{N-1} b(n)\left[-\frac{p^2(\tau_c-\theta)}{q^2(\theta)} + \frac{\dot{p}(\tau_c-\theta)}{q(\theta)}\right] \\ &+ \sum_{n=0}^{N-1}[1-b(n)]\left[-\frac{p^2(\tau_c-\theta)}{[1-q(\theta)]^2} - \frac{\dot{p}(\tau_c-\theta)}{1-q(\theta)}\right]\end{aligned} \tag{9}$$

for which we used that $\partial q(\theta)/\partial\theta = p(\tau_c - \theta)$, and introduced the definition $\dot{p}(\theta) := \partial p(\theta)/\partial\theta$.

Since for a Bernoulli variable $\mathrm{E}[b(n)] = q(\theta)$, the CRLB in (7) follows after taking the negative inverse of $\mathrm{E}[\ddot{L}(\theta)]$. The MLE can be found either by maximizing (8), or simply after recalling that the MLE of $q(\theta)$ is

$$\widehat{q} = \frac{1}{N}\sum_{n=0}^{N-1} b(n), \tag{10}$$

and using the invariance of MLE [cf. (4) and (10)]. □

Proposition 1 asserts that θ can be consistently estimated from a single binary observation per sensor, with variance as small as $B(\theta)$. Minimizing the latter over θ reveals that $B_{\min}$ is achieved when $\tau_c = \theta$ and is given by

$$B_{\min} = \frac{2\pi\sigma^2}{4N} \approx 1.57\frac{\sigma^2}{N}. \tag{11}$$

In words, if we place τ_c optimally, the variance increases only by a factor of $\pi/2$ with respect to the clairvoyant estimator $\bar{x}$ that

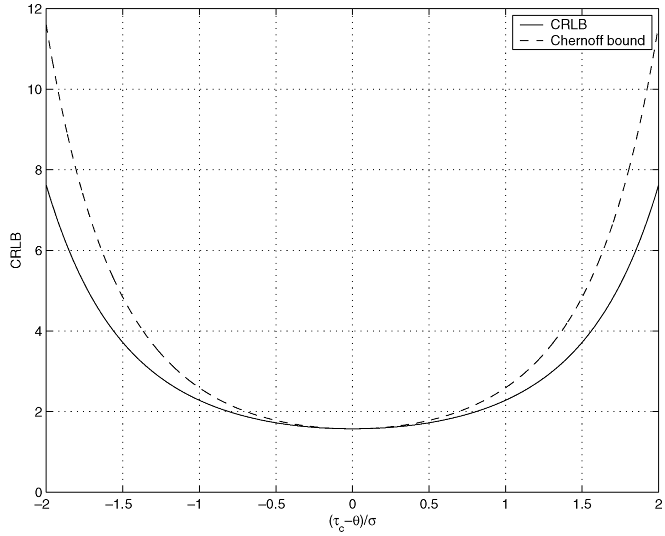

Fig. 1. CRLB and Chernoff bound in (13) as a function of the distance between τ_c and θ measured in AWGN standard deviation (σ) units.

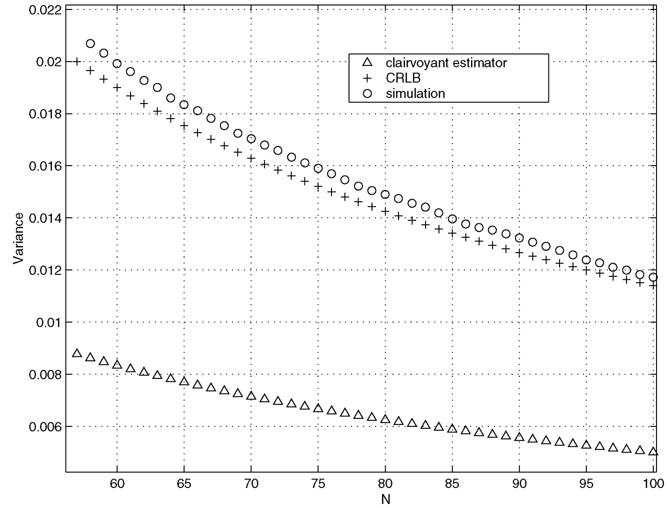

Fig. 2. MLE in (6) based on binary observations performs close to the clairvoyant sample mean estimator when θ is close to the threshold defining the binary observation ($\sigma = 1$, $\tau_c = 0$, and $\theta = 1$).

relies on unquantized observations. Using the (tight) Chernoff bound for the complementary CDF

$$F(\tau_c - \theta)\left[1 - F(\tau_c - \theta)\right] \leq \frac{1}{4} e^{-\frac{(\tau_c-\theta)^2}{2\sigma^2}} \tag{12}$$

based on which a simple bound on $B(\theta)$ can be obtained

$$B(\theta) \leq \frac{\pi\sigma^2}{2N} e^{+\frac{1}{2}\left[\frac{(\tau_c-\theta)}{\sigma}\right]^2}. \tag{13}$$

Fig. 1 depicts $B(\theta)$ and its Chernoff bound, from where it becomes apparent that for $|\tau_c - \theta|/\sigma \leq 1$ the increase in variance relative to (2) will be around 2 [cf. (7) and (13)]. Roughly speaking, to achieve a variance close to $\mathrm{var}(\bar{x})$ in (2), it suffices to place τ_c "σ-close" to θ. Fig. 2 shows a simulation where we have chosen $\tau_c = \theta + \sigma$, to verify that the penalty is, indeed, small.

Accounting for the dependence of $\mathrm{var}(\hat{\theta})$ on τ, σ and the unknown θ, one can envision an iterative algorithm in which the threshold is iteratively adjusted over time. Call $\tau_c^{(j)}$ the threshold used at time j, and $\hat{\theta}^{(j)}$ the corresponding estimate obtained as in (6). Having this estimate, we can now set $\tau_c^{(j+1)} = \hat{\theta}^{(j)}$, for subsequent estimates not only benefit from the increased number of observations but also from improved binary observations. Such an iterative algorithm fits rather nicely to, e.g., a target tracking application.

IV. MLE Based on Binary Observations: Nonidentical Thresholds

The variance of the estimator introduced in Section III will be close to $\mathrm{var}(\bar{x})$ whenever the actual parameter θ is close to the threshold τ in standard deviation (σ) units. This can be guaranteed when the possible values of θ are restricted to an interval of size comparable to σ; or in other words, when the *dynamic range of* θ *is in the order of* σ. When the dynamic range of θ is large relative to σ, we pursue a different approach using binary observations $b_k(n)$, generated from different regions $(\tau_k, +\infty)$ in order to assure that there will always be a threshold τ_k close to the true parameter. Consider, for each n, the set of binary measurements defined by (3) and to maintain the bandwidth constraint, let each sensor transmit *only one* out of this set of binary observations.

Let N_k be the total number of sensors transmitting binary observations based on the threshold τ_k, and define $\rho_k := N_k/N$ as the corresponding fraction of sensors. We further suppose that the index k_n chosen by sensor n, is known at the destination (the fusion center or peer sensors in an ad hoc WSN). Algorithmically, we can summarize our approach in three steps.

[S1] Define a set of thresholds $\boldsymbol{\tau} = \{\tau_k, k \in \mathbf{Z}\}$ and associated frequencies $\boldsymbol{\rho} = \{\rho_k, k \in \mathbf{Z}\}$.

[S2] Assign the index k_n to sensor n; i.e., sensor n generates the binary observation $b_{k_n}(n)$ using the threshold τ_{k_n}. Define $\mathbf{b} := [b_{k_0}(0), \ldots, b_{k_{N-1}}(N-1)]^T$.

[S3] Transmit the corresponding binary observations to find the MLE as we describe next.

Similar to (8), the log-likelihood function is given by

$$L(\theta) = \sum_{n=0}^{N-1} b_{k_n}(n) \ln\left(q_{k_n}(\theta)\right) + \left(1 - b_{k_n}(n)\right) \ln\left(1 - q_{k_n}(\theta)\right) \tag{14}$$

from where we can define the MLE of θ given the $\{b_{k_n}\}_{n=0}^{N-1}$

$$\hat{\theta} = \arg\max_\theta \left\{L(\theta)\right\}. \tag{15}$$

As $\hat{\theta}$ in (15) cannot be found in closed-form, we resort to a numerical search, such as Newton's algorithm that is based on the iteration

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} - \frac{\dot{L}\left(\hat{\theta}^{(i)}\right)}{\ddot{L}\left(\hat{\theta}^{(i)}\right)} \tag{16}$$

where $\dot{L}(\theta) := \partial L(\theta)/\partial\theta$, and $\ddot{L}(\theta) := \partial^2 L(\theta)/\partial\theta^2$ are the first and second derivatives of the log-likelihood function that we compute explicitly in (58) and (59) of Appendix A. Albeit numerically found, the MLE in (15) is guaranteed to converge to the global optimum of $L(\theta)$ thanks to the following property:

Proposition 2: The MLE problem (14), (15) is convex on θ.

Proof: The Gaussian pdf $p(x)$ is log-concave [5, p. 104]; furthermore, the regions $\mathcal{R}_k := (\tau_k, +\infty)$ and $\mathcal{R}_k^{(c)}$ are half-lines, and accordingly convex sets. To complete the proof just note that $q_k(\theta)$ and $1-q_k(\theta)$ are integrals of a log-concave function $(p(x))$ over convex sets ($\mathcal{R}_k$ and $\mathcal{R}_k^{(c)}$ respectively); thus, they are log-concave and their logarithms are concave. Given that summation preserves concavity, we infer that $L(\theta)$ is a concave function of θ. □

Although numerical MLE problems are typically difficult to solve, due to local minima requiring complicated search algorithms, this is not the case here. The concavity of $L(\theta)$ guarantees convergence of the Newton iteration (16) to the global optimum, regardless of initialization.

The CRLB for this problem follows from the expected value of $\ddot{L}(\theta)$ and is stated in the following proposition.

Proposition 3: The CRLB for any unbiased estimator $\hat{\theta}$ based on $\mathbf{b}$ is

$$B(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) = \frac{1}{N}\left[\sum_k \frac{\rho_k p^2(\tau_k - \theta)}{F(\tau_k - \theta)\left[1 - F(\tau_k - \theta)\right]}\right]^{-1} := \frac{1}{N} S^{-1}(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}). \quad (17)$$

Proof: See Appendix A.

Since the CRLB in (17) depends on the *design* parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$, Proposition 3 reveals that using nonidentical thresholds across sensors provides an additional degree of freedom. This is precisely what we were looking for in order to overcome the limitations of the estimator introduced in Section III. In the ensuing subsection, we will delve on the selection of $(\boldsymbol{\tau}, \boldsymbol{\rho})$.

A. Selecting the Parameters $(\boldsymbol{\tau}, \boldsymbol{\rho})$

Since the CRLB depends also on θ, the selection of $(\boldsymbol{\tau}, \boldsymbol{\rho})$ depends not only on the estimator variance for a specific value of θ, but also on how confident we are that the actual parameter will take on this value. To incorporate this confidence we introduce a weighting function, $W(\theta)$, which accounts for the relative importance of different values of θ. For instance, if we know *a priori* that $\theta \in (\Theta_1, \Theta_2)$, we can choose $W(\theta) = u(\theta - \Theta_1) - u(\theta - \Theta_2)$, where $u(\cdot)$ is the unit step function.

Given this weighting function, a reasonable performance indicator is the weighted variance

$$\mathcal{C}_W := \int_{-\infty}^{+\infty} W(\theta) \mathrm{var}(\hat{\theta}) d\theta. \quad (18)$$

Although we do not have an expression for the variance of the MLE in (15) but only the CRLB (17), we know that the MLE will approach this bound as $N \to \infty$. Consequently, selecting the best possible $(\boldsymbol{\tau}, \boldsymbol{\rho})$ for a prescribed $W(\theta)$ amounts to finding the set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ that *minimizes the weighted asymptotic variance* given by the weighted CRLB [c.f (17) and (18)]

$$\lim_{N \to +\infty} N\mathcal{C}_W = N\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho}) := N \int_{-\infty}^{+\infty} W(\theta) B(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) d\theta = \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})} d\theta. \quad (19)$$

Thus, the optimum set $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$, should be selected as the solution to the problem

$$(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*) = \arg\min_{(\boldsymbol{\tau}, \boldsymbol{\rho})} \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})} d\theta, \quad \text{s.t.} \sum_k \rho_k = 1, \quad \rho_k \geq 0 \; \forall k. \quad (20)$$

Solving (20) is complex, but through a proper relaxation we have been able to obtain the following insightful theorem.

Theorem 1: Assume that $\int_{-\infty}^{+\infty} W^{1/2}(\theta) d\theta < \infty$. Then, the weighted CRLB of any estimator $\hat{\theta}$ based on binary observations must satisfy

$$\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho}) \geq \mathcal{B}_{\min} := \frac{1}{N} \frac{\left[\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta) d\theta\right]^2}{\int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)[1-F(u)]} du}. \quad (21)$$

Furthermore, the bound is attained if and only if there exist a set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ such that

$$S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) = \mathcal{K} W^{\frac{1}{2}}(\theta), \quad \mathcal{K} := \frac{\int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)[1-F(u)]} du}{\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta) d\theta}. \quad (22)$$

Proof: See Appendix B.

Note that the claims of Theorem 1, are reminiscent of Cramer-Rao's Theorem in the sense that (21) establishes a bound, and (22) offers a condition for this bound to be attained.

To gain intuition on the performance limit dictated by Theorem 1, let us specialize (21) to a Gaussian-shaped $W(\theta)$, with variance σ_θ^2. In this case, the numerator in (21) becomes

$$\left[\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta) d\theta\right]^2 = 2\sqrt{2\pi}\sigma_\theta. \quad (23)$$

The denominator in (21) that depends on the noise distribution cannot be integrated in closed form, but we can resort to the following numerical approximation

$$\int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)\left[1 - F(u)\right]} du \approx \frac{1.81}{\sigma}. \quad (24)$$

Substituting (24) and (23) in (21), we finally obtain

$$\mathcal{B}_{\min}^{GG} \approx 2.77 \frac{\sigma_\theta \sigma}{N} = 2.77 \frac{\sigma_\theta}{\sigma} \left(\frac{\sigma^2}{N}\right). \quad (25)$$

Perhaps as we should have expected, the best possible weighted variance for any estimator based on a single binary observation per sensor can only be close to the clairvoyant variance in (2) when $\sigma_\theta \approx \sigma$—a condition valid in low to medium SNR scenarios. When the SNR is high ($\sigma_\theta \gg \sigma$), the performance gap between (2) and (25) is significant and a different approach should be pursued.

A similar derivation leads to an analogous expression for a uniform weight function $W(\theta) = u(\theta - \Theta_1) - u(\theta - \Theta_2)$

$$\mathcal{B}_{\min}^{GU} \approx 0.55 \frac{|\Theta_2 - \Theta_1|}{\sigma} \left(\frac{\sigma^2}{N}\right). \quad (26)$$

Equation (26) similarly allow us to infer that the variance of any estimator based on a single binary observation per sensor can only be close to the clairvoyant variance in (2) when $|\Theta_2 - \Theta_1| \approx \sigma$, which corresponds to a low to medium SNR.

Regarding the achievability of the bound in (21), note that although we cannot assure that there always exists a set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ such that $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) = \mathcal{K} W^{1/2}(\theta)$, we can adopt as a relaxed optimal solution the set $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ that minimizes the distance between $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ and $\mathcal{K} W^{1/2}(\theta)$

$$(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger) = \arg\min_{(\boldsymbol{\tau}, \boldsymbol{\rho})} \left\| \mathcal{K} W^{\frac{1}{2}}(\theta) - \sum_k \frac{\rho_k p^2(\tau_k - \theta)}{F(\tau_k - \theta)\left[1 - F(\tau_k - \theta)\right]} \right\| \quad \text{s.t.} \quad \rho_k > 0. \quad (27)$$

The norm measuring the distance can be any norm in the space of functions. Notwithstanding, we find it convenient to work with the L_2 norm.

It is fair to emphasize that $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ obtained as the solution of (27) will in general be different from the optimum $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$ obtained as the solution of (20). Nonetheless (27) offers a more tractable formulation that can be easily solved by methods we outline in Subsection IV-C, and test in Section VII. It will turn out that solving (27) numerically yields a small minimum distance, illustrating that the estimator (15) based on $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ is nearly optimal.

Remark 1: The use of a weight function in our deterministic parameter estimation problem is motivated by maximum *a posteriori* (MAP) estimation principles which apply to random parameters. Viewing our deterministic parameter θ as a random one with prior distribution $W(\theta)$, the log-distribution after observing the vector of binary observations is given by

$$L_{\text{MAP}}(\theta) = L(\theta) + \ln\left[W(\theta)\right] \quad (28)$$

with $L(\theta)$ given by (14). The MAP estimator is defined as $\hat{\theta}_{\text{MAP}} = \arg\max[L_{\text{MAP}}(\theta)]$.

Note that being $L(\theta) \propto N$, it holds that $L_{\text{MAP}}(\theta) \to L(\theta)$ when $N \to \infty$, and accordingly both estimators coincide asymptotically. In particular, the average variance of the MAP estimator converges to the average variance of the MLE, and minimization of $\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho})$ as defined in (19) yields the asymptotically optimum MAP estimator (as well as the asymptotically optimal MLE).

Also worth mentioning is that if the prior distribution $W(\theta)$ is log-concave then the likelihood in (28) is concave. This is the case for many distributions including the Gaussian and the uniform one.

B. An Achievable Upper Bound on $\mathcal{B}_W(\boldsymbol{\tau}, \boldsymbol{\rho})$

To explore whether we can approach the bound in (21), we introduce the following Chernoff bound [cf. (12) and (17)]

$$B(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) \le \frac{1}{N} \left[\frac{4}{\sqrt{2\pi}\sigma} \sum_k \rho_k p(\tau_k - \theta) \right]^{-1} := \frac{1}{N} T^{-1}(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}). \quad (29)$$

Being a superposition of shifted Gaussian bells with variance σ, the bound $T(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ is easier to manipulate than $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ in (17). However, the major implication of (29) is that by adjusting the spacing $\tau_k - \tau_{k-1} := \tau = \sigma$, the set of functions $\mathcal{G} := \{p(\theta - k\tau), k \in \mathbf{Z}\}$ becomes a *Gabor basis* [16, Ch. 6]. Therefore, $T(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$ can be thought of as a Gabor expansion with coefficients $\boldsymbol{\rho}$, and thus capable of approximating $W^{1/2}(\theta)$ with arbitrary accuracy.

Theorem 2: Define the Gabor basis $\mathcal{G} := \{p(\theta - k\tau), k \in \mathbf{Z}\}$ with $\tau = \sigma$, and consider the coefficients $\mathbf{c} = \{c_k, k \in \mathbf{Z}\}$ of the Gabor expansion of $W^{1/2}(\theta)$ given by

$$\mathbf{c} = \arg\min_{\{c_k\}} \left\| W^{\frac{1}{2}}(\theta) - \sum_k c_k p(k\tau - \theta) \right\|. \quad (30)$$

Also, assume that $\int_{-\infty}^{+\infty} W^{1/2}(\theta) d\theta < \infty$. If $c_k \ge 0 \; \forall k$, then the weighted CRLB of the estimator based on the binary observations defined by $\boldsymbol{\tau}^\ddagger := \{\tau_k = k\tau, k \in \mathbf{Z}\}$, and $\boldsymbol{\rho}^\ddagger := \{\rho_k = c_k / (\sum_k c_k), k \in \mathbf{Z}\}$ is bounded by

$$\mathcal{B}_W(\boldsymbol{\tau}^\ddagger, \boldsymbol{\rho}^\ddagger) \le \mathcal{B}_{\max} := \frac{\sqrt{2\pi}\sigma}{4N} \left[\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta) d\theta \right]^2. \quad (31)$$

Proof: See Appendix B.

Corollary 1: If $W(\theta)$ is Gaussian-shaped with $\sigma_\theta \ge \sigma/\sqrt{2}$, then the weighted CRLB of the estimator based on binary observations constructed form the set $(\boldsymbol{\tau}^\ddagger, \boldsymbol{\rho}^\ddagger)$ as in Theorem 2 is bounded by

$$\mathcal{B}_W(\boldsymbol{\tau}^\ddagger, \boldsymbol{\rho}^\ddagger) \le \mathcal{B}_{\max} = \pi \frac{\sigma \sigma_\theta}{N}. \quad (32)$$

Proof: The coefficients of the Gabor transform of $W^{1/2}(\theta)$ using the basis $\mathcal{G}$ are all positive [16, Chap. 6]; and the integral of $W^{1/2}(\theta)$ is given by (23). □

Note that Theorem 2 is weaker than Theorem 1 in the sense that the former asserts an *upper* bound on the asymptotic variance while the latter claims a *lower* bound. On the other hand, Theorem 1 is weaker because it claims the *existence* of the lower bound, while Theorem 2 claims the *achievability* of the upper bound.

Perhaps more important is that comparison of (32) with (25) implies that

$$\mathcal{B}_{\max} \approx 1.14 \mathcal{B}_{\min} \quad (33)$$

that is, the gap between the lower and upper bound is small. And, as we wanted to prove, the solution of (27) should give an estimator whose CRLB is close to (within 14%) the bound $\mathcal{B}_{\min}$ in (21).

Even though, it is possible by reducing the distance between thresholds (or using nonuniform spacings) to further reduce the variance, Theorem 2 asserts that this reduction will be no greater than 14% relative to a uniform spacing with $\tau = \sigma$. Consequently, a threshold spacing $\tau = \sigma$ approximates the best performance in (25) reasonably well. Moreover, numerical results in Section VII will justify that a spacing $\tau \approx 2\sigma$ is good enough for practical purposes. Note that this result is somewhat counterintuitive since we tend to think that reducing the distance between thresholds would improve the estimator. However, the truth is that with a uniform spacing $\tau = \sigma$ (or $\tau = 2\sigma$ as numerical results illustrate) there is no need for increasing the number of thresholds any further.

C. Algorithmic Implementation

Theorem 1 led to the definition of a near-optimal set $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ given by the solution of the infinite-dimensional least-squares problem in (27). Furthermore, Theorem 2 reinforced the usefulness of this near-optimal solution as we proved that the CRLB for this $(\boldsymbol{\tau}^\dagger, \boldsymbol{\rho}^\dagger)$ cannot be very far from the optimal $(\boldsymbol{\tau}^*, \boldsymbol{\rho}^*)$ defined by (20). In the present subsection, we will analyze the numerical implementation of (27).

A by-product of Theorem 2 is that a uniform threshold spacing $\tau_{k+1} - \tau_k := \tau > \sigma$ captures most of the optimality, and accordingly we begin the numerical implementation by defining a threshold spacing $\tau \geq \sigma$. This reduces the degrees of freedom by one, simplifying the numerical implementation to that of finding the set of corresponding frequencies ρ_k.

The first step is to obtain a finite dimensional problem by discretizing the functions in (27)

$$\boldsymbol{\rho}^* = \arg\min_{\boldsymbol{\rho}} \|\mathbf{s} - \mathbf{P}\boldsymbol{\rho}\| \qquad \text{s.t.} \quad \boldsymbol{\rho} \succeq \mathbf{0} \tag{34}$$

where $\mathbf{s} := [S(\theta_0), \ldots, S(\theta_M)]^T$, $\boldsymbol{\rho} := [\rho_1, \ldots, \rho_L]^T$, $\succeq$ denotes element-wise inequality ($\rho_l \geq 0\, \forall l$); M controls the discretization step, L is the number of thresholds whose frequencies are large enough to be considered of interest; and the matrix $\mathbf{P}$, has entries given by

$$[\mathbf{P}]_{ij} = \frac{p^2(\tau_i - \theta_j)}{F(\tau_i - \theta_j)\,[1 - F(\tau_i - \theta_j)]}. \tag{35}$$

Discretization introduces numerical errors that can be controlled by choosing a small enough step in the (numerical) evaluation of the integrals. However, this discretization alters the implicit constraint that $\sum_k \rho_k = 1$, which was enacted by the normalization constant $\mathcal{K}$ in (22). Once the integral is discretized this normalization no longer holds and we have to make the following constraint explicit:

$$\boldsymbol{\rho}^* = \arg\min_{\boldsymbol{\rho}} \|\mathbf{s} - \mathbf{F}\boldsymbol{\rho}\| \qquad \text{s.t. } \boldsymbol{\rho} \succeq \mathbf{0},\ \boldsymbol{\rho}^T\mathbf{1} = 1. \tag{36}$$

Note that the constrained least squares problem in (36) is convex, since the objective is convex (norms are convex) and the constraints are linear. Moreover, (36) can be transformed to a second-order cone program (SOCP) after introducing the auxiliary variable t to obtain

$$\boldsymbol{\rho}^* = \arg\min_{(t,\boldsymbol{\rho})} t \qquad \text{s.t.} \quad \|\mathbf{s} - \mathbf{F}\boldsymbol{\rho}\| \leq t, \quad \boldsymbol{\rho} \succeq \mathbf{0}, \quad \boldsymbol{\rho}^T\mathbf{1} = 1. \tag{37}$$

It is known that a SOCP can be efficiently solved with standard convex optimization packages [21]. The implementation of this design is illustrated in Section VII for a pair of different weighting functions $W(\theta)$.

V. Relaxing the Bandwidth Constraint

The variances of the estimators in Sections III and IV are close to $\text{var}(\bar{x})$ either when the parameter's range is small, or, in the order of the noise variance. Formally, if for a Gaussian weight function we define the SNR as $\gamma := \sigma_\theta^2/\sigma^2$, the variance of the estimator in (15) is [cf. (2) and (25)]

$$\mathcal{B}_{\min} = 2.77\sqrt{\gamma}\,\text{var}(\bar{x}). \tag{38}$$

If we let N_{sm} be the number of observations required by $\bar{x}$ to achieve the same variance of the estimator in (15), we can see that the number N of binary observations must increase by a factor $N/N_{sm} = 2.77\sqrt{\gamma}$. It is clear that in high-γ scenarios, we need a different approach motivating the *relaxation of the bandwidth constraint* pursued in this section.

Specifically, using a sequence of thresholds $\boldsymbol{\tau} := \{\tau_k, k \in \mathbf{Z}\}$, we will rely on multiple binary observations per sensor, $\mathbf{b}(n) := \{b_k(n), k \in \mathbf{Z}\}$, with corresponding Bernoulli parameters $\mathbf{q} := \{q_k = \Pr\{x(n) > \tau_k\}, k \in \mathbf{Z}\}$. Without loss of generality, we will assume that $\tau_{k_1} < \tau_{k_2}$, when $k_1 < k_2$. The entries of $\mathbf{b}(n)$ are not independent, since $x(n)$ cannot be at the same time smaller than τ_{k_1} and larger than τ_{k_2} for $k_1 < k_2$; hence, $\mathbf{b}$ can only take on realizations

$$\boldsymbol{\beta}_l = \{\beta_k, k \in Z | y_k = 1 \text{ for } k \leq l,\ y_k = 0 \text{ for } k > l\}. \tag{39}$$

The realization $\mathbf{b}(n) = \boldsymbol{\beta}_l$ corresponds to the event $\{x(n) \in (\tau_l, \tau_{l+1})\}$, which re-iterates our earlier comment that creating multiple binary observations is just a different way of looking at quantization.

We now express the distribution of $\mathbf{b}(n)$ in terms of θ, and from there obtain the per-sensor log-likelihood as

$$L_n(\theta) = \sum_{k=-\infty}^{+\infty} \delta\left[\boldsymbol{\beta}_k - \mathbf{b}(n)\right] \ln\left[q_{k+1}(\theta) - q_k(\theta)\right], \tag{40}$$

where $\delta[\boldsymbol{\beta}_k - \mathbf{b}(n)] := 1$ if $\boldsymbol{\beta}_k = \mathbf{b}(n)$, and 0 otherwise. Independence across sensors implies,

$$L(\theta) = \sum_{n=0}^{N-1} L_n(\theta) \tag{41}$$

and yields the MLE of θ given $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ as

$$\hat{\theta} = \arg\max_\theta \{L(\theta)\}. \tag{42}$$

Two important features of $\hat{\theta}$ in (42) are summarized next.

Proposition 4:

a) The log-likelihood (41) is a concave function of θ.

b) The CRLB of any unbiased estimator of θ based on $\{\mathbf{b}(n)\}_{n=0}^{N-1}$ is given by

$$B(\theta) = \frac{1}{N}\left[\sum_{k=-\infty}^{+\infty} \frac{[p(\tau_{k+1}-\theta) - p(\tau_k - \theta)]^2}{F(\tau_{k+1}-\theta) - F(\tau_k - \theta)}\right]^{-1}. \tag{43}$$

Proof: See Appendix C.

The concavity of $L(\theta)$ in (41) asserted by Proposition 4-a) implies existence of a reliable numerical implementation of $\hat{\theta}$ in (42). To understand Proposition 4-b) notice that for an infinite set of equally spaced thresholds (with spacing $\tau := \tau_{k+1} - \tau_k$), $B(\theta)$ in (43) is periodic with period τ. Fig. 3 depicts $B(\theta)$ parameterized by τ/σ, along with the maximum and minimum values of $B(\theta)$ as functions of τ/σ. Note that for a given τ the worst and best variances are almost equal for $\tau \leq 2\sigma$, being for all practical purposes constant when $\tau \leq \sigma$. More important, when $\tau \leq \sigma$, $B(\theta)$ is almost equal to the clairvoyant estimator's variance.

We now turn our attention to designing a transmission scheme for the infinite number of binary observations per sensor. This can be done by noting that if $b_k(n) = 1$, then $b_{k'}(n) = 1$ for $k' < k$; and likewise if $b_k(n) = 0$, then $b_{k'}(n) = 0$ for $k' > k$. Accordingly, each binary observation transmitted provides information about half of the thresholds, and the required number of bits N_t to be transmitted per sensor grows logarithmically with the allowable parameter range. The actual value of N_t will

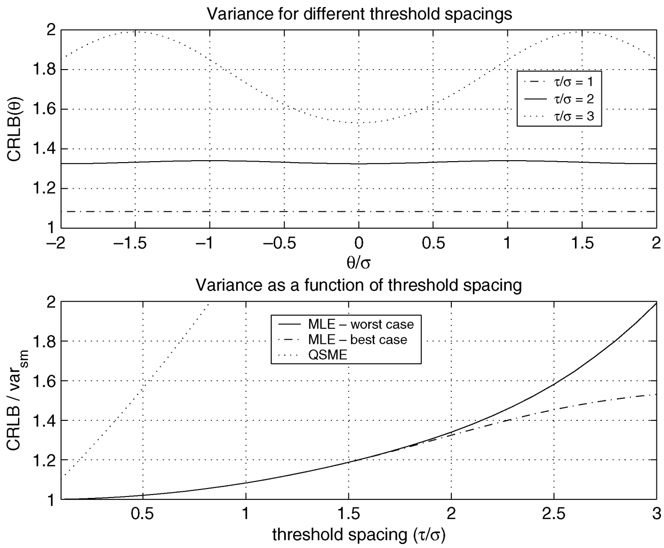

Fig. 3. Variance of the estimator relying on the whole sequence of binary observations. The room for improved performance once $\tau < \sigma$ is small.

depend on the parameter's range; e.g., for $\theta \in [-U, U]$ it will be $N_t \approx \log_2[(\sigma + 2U)/\tau]$. When the prior knowledge about θ dictates a Gaussian prior $W(\theta)$ the result can be summarized in the following proposition.

Proposition 5: When $W(\theta)$ is a Gaussian bell with variance σ_θ^2, the infinite set of binary observations $\mathbf{b}(n)$ can be transmitted using N_t bits satisfying

$$\mathrm{E}(N_t) < 3 + \left[\log_2 Q^{-1}\left(\frac{1}{4}\right) + \frac{1}{2}\log_2\left(\frac{\sigma_\theta^2 + \sigma^2}{\tau^2}\right)\right]_+ \tag{44}$$

where $\tau := \tau_{k+1} - \tau_k \ \forall k$, $Q(x) := 1/(\sqrt{2\pi})\int_x^{+\infty} \exp(-u^2/2)du$, and $[x]_+ := \max(0, x)$.

Proof: See Appendix D.

Combining Propositions 4-b) and 5 yields a benchmark on the performance of estimators based on binary observations. For a given bandwidth constraint, we determine τ from (44), and from there the benchmark variance from (43).

Note that (44) can also be written using a slightly more intuitive expression in terms of γ

$$\mathrm{E}(N_t) < 2.43 + \frac{1}{2}\log_2(1+\gamma) + \log_2\left(\frac{\sigma}{\tau}\right) \tag{45}$$

where we substituted the constants in (44) by their explicit values, and assumed for simplicity that the argument inside the $[\cdot]_+$ operator is positive (valid if $\tau^2 < 0.45(\sigma_\theta^2 + \sigma^2)$). The first logarithmic term in (45) can be viewed as quantifying the information that each observation $x(n)$ carries about the underlying parameter, while the second logarithmic term can be thought of as quantifying our confidence on the observations. By decreasing τ beyond σ we are adding bits to the quantization of $x(n)$ reflecting our belief that there is more information to be extracted from it. In the next section, we will see that it makes sense to set $\tau = \sigma$, in which case N_t reduces to

$$\mathrm{E}(N_t) < 2.43 + \frac{1}{2}\log_2(1+\gamma). \tag{46}$$

Equation (46) is valid, when $\gamma \geq 1.20$, in which case it is a tight bound in the expected value of transmitted bits. When $\gamma < 1.20$, the bound reduces to $\mathrm{E}(N_t) < 3$ which is too loose to be of practical interest. However, remember that for this low SNR scenario we advocate the estimator introduced in Section IV and this limitation of (46) is not a concern.

A. Optimum Threshold Spacing

Estimation problems are usually posed for a given number of measurements; but for bandwidth-constrained problems, a more meaningful formulation is to prescribe the total number of available bits, N_b. That is, given the channel (bandwidth, SNR, and time) we are allowed to transmit up to N_b bits that have to be allocated among the observations. Fine quantization implies a small per-observation variance, but also a small number of observations N; while coarse quantization increases the variance per observation but allows for a larger N.

A convenient metric for a bandwidth-constrained estimation problem is the following:

Definition 1: Suppose that for a given estimator based on binary observations, the transmission of binary observations requires an average of $\bar{N}_t$ bits. Define the per-bit worst-case CRLB as

$$C_b = \bar{N}_t \max_\theta \{B(\theta)\}. \tag{47}$$

For a bandwidth constraint N_b, the variance will be bounded by $\mathrm{var}(\hat{\theta}) \geq C_b/N_b$.

Applying Definition 1 to the CRLB in (43), we deduce that C_b is a function of the spacing τ

$$C_b(\tau) = \bar{N}_t(\tau) \max_\theta \{B(\theta, \tau)\} \tag{48}$$

what raises the question about the existence of an optimum threshold spacing $\tau^*(\gamma)$

$$\tau^* = \arg\min_\tau \{C_b(\tau)\}. \tag{49}$$

For this question to be meaningful, τ^* should be neither zero nor infinity, which is true for the problem considered in the current section.

Proposition 6: For a Gaussian-shaped $W(\theta)$, the optimum threshold spacing τ^* in (49) is finite and different from zero.

Proof: When $\tau \to 0$, $C_b(\tau) \to +\infty$ because $\bar{N}_t \to +\infty$ while $B(\theta)$ is bounded; furthermore when $\tau \to +\infty$, $C_b(\tau) \to +\infty$ because $B(\theta) \to +\infty$ faster that $\bar{N}_t \to 0$ (exponentially versus logarithmically). As $C_b(\tau)$ is continuous and approaches ∞ in both extremes, it must have a minimum. □

By taking into account the bandwidth constraint, we proved the existence of an optimum quantization step τ^* that minimizes C_b for a given γ; and a corresponding *optimum number of bits per observation*. Fig. 4 depicts $\tau^*(\gamma)$. It is apparent from these curves, that the optimum value is quite insensitive to variations of γ. When γ varies from 0 db to 50 db (a 10^5 range) τ^* moves from 2σ to σ. Furthermore, the curves $C_b(\tau)$ are very flat around the optimum, implying that we can adopt $\tau = \sigma$ as a working compromise for the optimum threshold spacing (i.e., quantization step).

VI. Quantized Sample Mean Estimator

It is interesting to compare the MLE estimator in (42) with the low complexity quantized sample mean estimator (QSME).

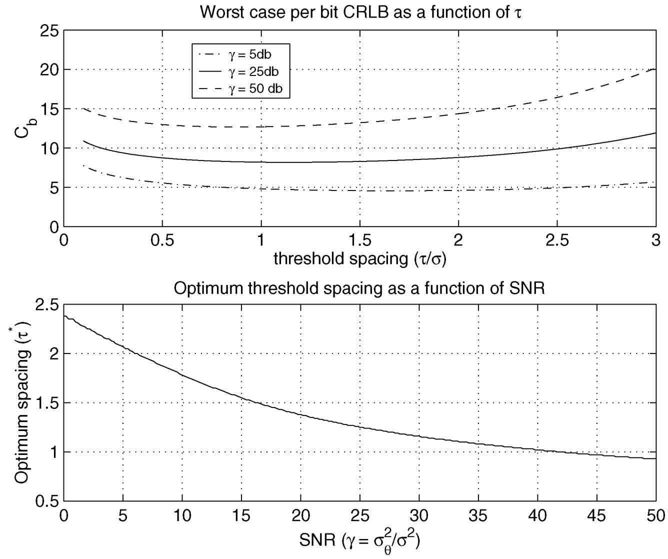

Fig. 4. Variation of the threshold spacing that minimizes the worst-case per-bit CRLB with the SNR. $C_b(\tau)$ is very flat around the optimum and τ^* has a small change when the SNR moves over a range of 50 db.

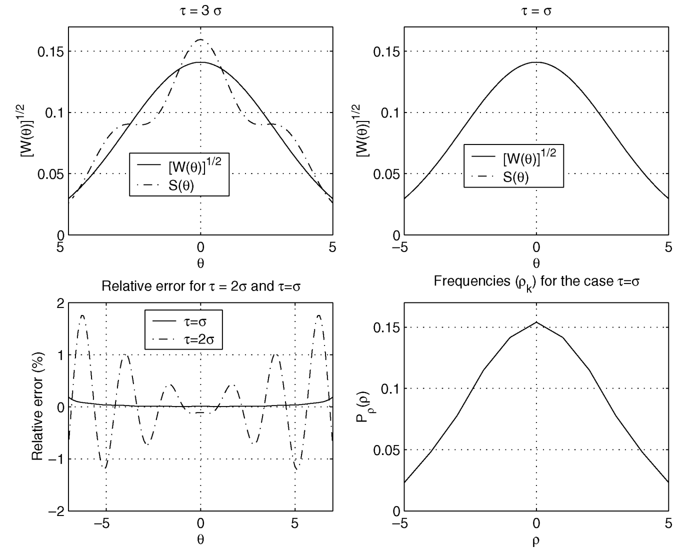

Fig. 5. Gaussian noise and Gaussian-shaped weight function. Although a threshold spacing $\tau = \sigma$ reduces the approximation error to almost zero, a spacing $\tau = 2\sigma$ is good enough in practice ($\sigma = 1$, and $\sigma_\theta = 2$).

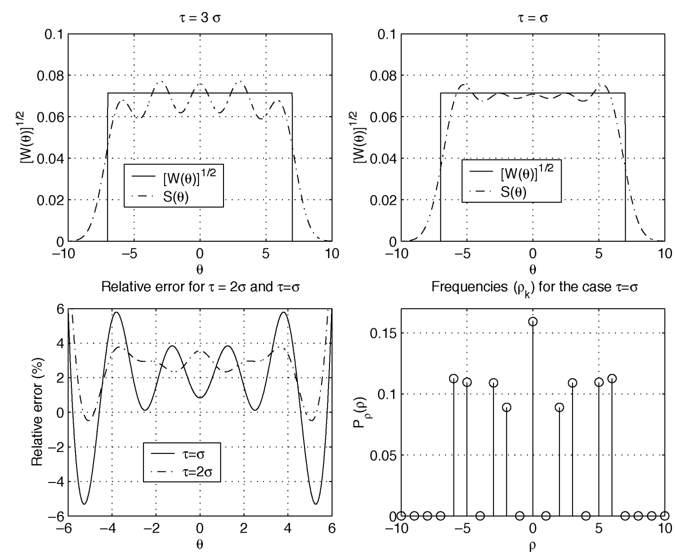

Fig. 6. Gaussian noise and uniform weight function. A threshold spacing $\tau = \sigma$ has smaller MSE but a spacing $\tau = 2\sigma$ is better in most of the nonzero probability interval ($\sigma = 1$, and prior U[−7,7]).

Consider the observations $\{x(n)\}_{n=0}^{N-1}$ and quantize them with a uniform quantizer at resolution τ to obtain

$$x_Q(n) = \tau \text{ round } \left[\frac{x(n)}{\tau}\right] \tag{50}$$

where x_Q denotes the quantized observations and round(x) is the integer closest to x.

The QSME is just the sample mean of the quantized observations

$$\bar{x}_Q(n) := \frac{1}{N}\sum_{n=0}^{N-1} x_Q(n) \tag{51}$$

which is a desirable estimator if one just ignores the bandwidth constraint. Interestingly, this simple estimator is not very far from the MLE in (42) as stated in the following proposition.

Proposition 7: The variance of the QSME in (51) is bounded by

$$\mathrm{E}\left[(\bar{x}_Q - \theta)^2\right] \leq \left(1 + \frac{\tau}{\sigma} + \frac{\tau^2}{4\sigma^2}\right)\frac{\sigma^2}{N}. \tag{52}$$

Proof: See Appendix E.

Note that since $\bar{x}_Q$ is biased, the pertinent performance metric is the Mean Square Error (MSE), not the variance. Fig. 3 shows that the MSE of the MLE for a threshold spacing $\tau = 2\sigma$ is roughly comparable to the MSE of the QSME for a spacing $\tau = \sigma/2$.

For low SNR problems in which the cost of each sequence of binary observations is just a few bits, adding two more bits in order to use the QSME offers a rather poor solution. Meanwhile, when the SNR is high, the addition of two bits to a long sequence carries a small relative increase in the bandwidth requirement. While the break point depends on the desired complexity–performance tradeoff, it is clear that when the SNR is high the bandwidth-constrained estimation problem is of little interest since even a “simple-minded” estimator performs close to the optimum MLE in (42). The effort in finding efficient bandwidth-constrained distributed estimation algorithms should, thus, be focused on low SNR scenarios.

VII. Numerical Results

We implement here the estimator introduced in Section IV, for which there are two aspects we want to study; the design of $(\boldsymbol{\tau}, \boldsymbol{\rho})$ by numerically solving (37); and the implementation of the estimator itself. Recall that with thresholds spaced by less than σ, the room for increasing performance is limited, and thus we are also interested in studying the effect the spacing τ has on the average estimation variance.

A. Designing $(\boldsymbol{\tau}, \boldsymbol{\rho})$

For a given threshold spacing τ, the set of frequencies $\boldsymbol{\rho}$ is obtained as the solution of the SOCP in (37). Figs. 5 and 6 show

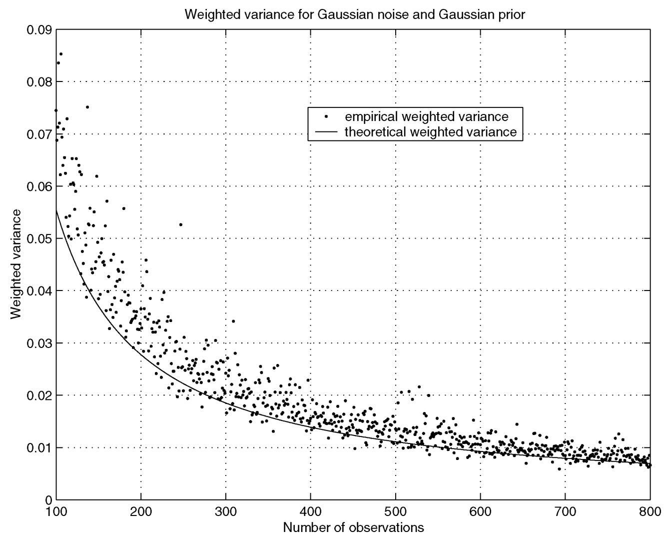

Fig. 7. Gaussian noise and Gaussian weight function. With a threshold spacing $\tau = 2\sigma$ we achieve a good approximation to the minimum asymptotic average variance ($\sigma = 1$, $\tau = 2$, and $\sigma_\theta = 2$).

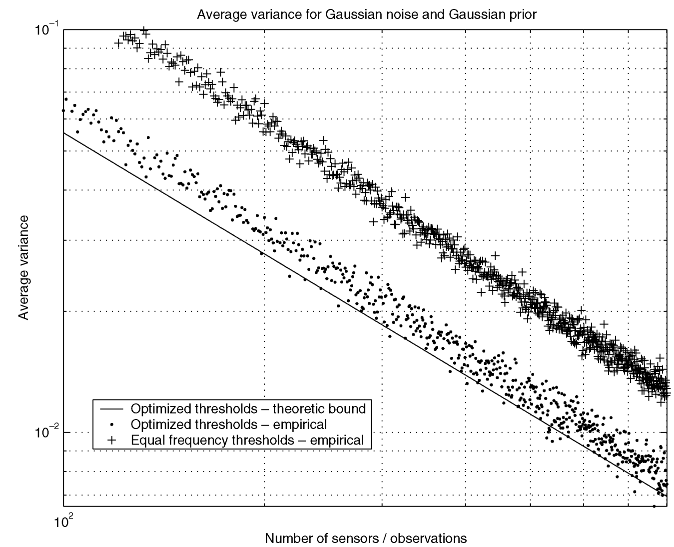

Fig. 8. Average variance of the optimum set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ found as the solution of (37), yields a noticeable advantage over the use of equispaced equal frequency thresholds as defined by (55) ($\sigma = 1$, $\tau = 2$, and $\sigma_\theta = 2$).

the result of computing $\boldsymbol{\rho}$ for the case of Gaussian and uniform weighting functions, respectively. In both cases, it is apparent that a threshold spacing $\tau = 2\sigma$ suffices to achieve a small MSE.

This is even more clear in the uniform case where reducing the spacing results in nulling some of the ρ_k. Particularly interesting are the error curves depicting the difference between $W^{1/2}(\theta)$ and $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$. When the threshold spacing is reduced from $\tau = 2\sigma$ to $\tau = \sigma$ the error is almost unchanged. Although we have not established this analytically, it appears that *choosing the thresholds with a spacing smaller than* 2σ *is of no practical value*.

B. Estimation With 1 bit per Sensor

The estimation problem itself is solved using Newton's algorithm based on the iteration (16). The results are shown on Fig. 7 for a Gaussian weight function with 2σ spacing between thresholds. For each value of N, the experiment is repeated 200 times, and the average variance is plotted against the theoretical threshold which reasonably predicts its value. This also reinforces the observation that a threshold spacing $\tau = 2\sigma$, is good enough for practical purposes.

C. Comparison With Deterministic Control Signals

The exponential increase of the CRLB in (7) was first observed in [15]. In particular, if $\theta \in [-\Theta, \Theta]$, and we define $\Delta = \Theta/\sigma$ the worst-case CRLB is [cf. (7) with $\tau_c = 0$],

$$B_{\max} = \frac{2\pi\sigma^2}{N} Q(\Delta)\,[1 - Q(\Delta)]\, e^{\Delta^2} \tag{53}$$

whose growth is approximately exponential in Δ; which can be interpreted as the parameter range measured in standard deviation units.

As noted before, while this is satisfactory for $\Delta \approx 1$, a different approach is needed for $\Delta > 1$. To alleviate this problem, adding control signals to the original observations was advocated by [15]. Though different classes of control signals were proposed in [15], particularly related to the present work are deterministic control signals of the form

$$x(n) = \theta + w(n) + s(n) \tag{54}$$

where $s(n)$ is a known periodic waveform, chosen as to minimize the worst case CRLB in $[-\Theta, \Theta]$. To this end it suffices to use a K-periodic sawtooth waveform $s(n) = 2\Theta/(K-1)[-(K-1)/2 + n \bmod K]$ with appropriately chosen period K.

Such control signals can be seen to be equivalent to the use of multiple equispaced thresholds with equal frequency,

$$\tau_k = \frac{2\Theta}{K-1}k, \quad \rho_k = \frac{1}{K} \tag{55}$$

for $k = 1, \ldots, K$. It can be shown that for large enough Θ, the maximum CRLB is minimized by the (min–max optimum) threshold spacing $\tau_k - \tau_{k-1} = 2.59\sigma$.

The difference with the approach in Section IV is, of course, that the thresholds are optimized after averaging over a certain weight function.

To illustrate a case where our approach could be of interest consider a parameter $\theta \in (-\infty, \infty)$ and a Gaussian weight function with variance σ_θ. While in this case the range is not limited, we can consider for practical purposes that $\theta \in [-3\sigma_\theta, 3\sigma_\theta]$ and compare the performance of the estimator defined by the set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ given in (55), with the optimum set $(\boldsymbol{\tau}, \boldsymbol{\rho})$ found as the solution of (37). An example comparison is depicted in Fig. 8, from where we observe a noticeable difference favoring the second approach.

This difference is just a manifestation of optimization options, with each option being applicable in different situations.

VIII. Concluding Remarks

We have considered bandwidth constrained distributed estimation for WSN. Under the strict bandwidth constraint of just 1 bit per sensor, we introduced a class of MLE's that attain a variance close to the sample mean estimator's variance when the noise variance is comparable with the dynamic range of the parameter to be estimated. This class of estimators is well suited for low-to-medium SNR problems.

Relaxing the bandwidth constraint, we also constructed the best possible estimator for a given total number of bits. This led us to the per bit CRLB which revealed a tradeoff between reducing the quantization step and making room for transmitting more independent observations. A general rule of thumb is that selecting a *quantization step equal to the noise variance is good enough for most practical situations*.

Finally, by comparing this last MLE with the QSME, we deduced that *for high SNR problems even the least complex scheme performs close to the optimum*. Consequently, bandwidth-constrained distributed estimation is not a relevant problem in such cases and the QSME should be used for its low complexity. We have also studied implementation issues, and established that *all the MLE problems we considered are convex*. Consequently, they can be efficiently solved and their numerical convergence is assured.

Though the noise was assumed Gaussian throughout the paper, it is worth noting that Theorem 1 and Propositions 1, 3, and 4-b) are valid for any noise distribution; Propositions 2 and 4-a) only require a log-concave distribution.

Appendix A
Proof of Proposition 3

Let us first note that from the form of $L(\theta)$ in (14), the MLE estimators of q_k

$$\widehat{q_k} = \frac{1}{N_k} \sum_{n=0}^{N-1} b_{k_n} \delta\left(k_{n_k} - k\right) \tag{56}$$

are sufficient statistics for this problem, and, (14) reduces to

$$L(\theta) = \sum_k N_k \widehat{q_k} \ln\left(q_k(\theta)\right) + N_k\left(1 - \widehat{q_k}\right) \ln\left(1 - q_k(\theta)\right) \tag{57}$$

from where we can obtain the first

$$\dot{L}(\theta) = \sum_k N_k \widehat{q_k} \frac{p(\tau_k - \theta)}{F(\tau_k - \theta)} - N_k\left(1 - \widehat{q_k}\right) \frac{p(\tau_k - \theta)}{1 - F(\tau_k - \theta)} \tag{58}$$

and second derivative

$$\ddot{L}(\theta) = \sum_k N_k \widehat{q_k} \left[-\frac{p^2(\tau_k - \theta)}{F^2(\tau_k - \theta)} + \frac{\dot{p}(\tau_k - \theta)}{F(\tau_k - \theta)}\right] - N_k\left(1 - \widehat{q_k}\right)\left[\frac{p^2(\tau_k - \theta)}{\left[1 - F(\tau_k - \theta)\right]^2} + \frac{\dot{p}(\tau_k - \theta)}{1 - F(\tau_k - \theta)}\right] \tag{59}$$

required by Newton's iteration in (16). In deriving (58) and (59) we used that $q_k(\theta) = F(\tau_k - \theta)$, and $\partial q_k(\theta)/\partial\theta = p(\tau_k - \theta)$. As we defined before, $\dot{p}(u) := \partial p(u)/\partial u = -(u/\sigma^2)p(u)$.

To obtain the CRLB, we simply take the expected value of (59) with respect to $\mathbf{b}$. Since $\widehat{q_k}$ is unbiased $(E(\widehat{q_k}) = q_k)$, the terms involving $\dot{p}(\tau_k - \theta)$ disappear and (17) follows. □

Appendix B
Proof of Theorems 1 and 2

We begin by introducing a lemma required by the proofs of both theorems.

Lemma 2: If $\int_{-\infty}^{+\infty} W^{1/2}(\theta)d\theta < \infty$ the solution of the variational problem

$$\arg\min_S \mathcal{F}(S) := \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta)} d\theta \qquad \text{s.t.} \quad \int_{-\infty}^{+\infty} S(\theta) d\theta = \mathcal{I} \tag{60}$$

is given by the function

$$S^*(\theta) = \frac{\mathcal{I}}{\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta)} W^{\frac{1}{2}}(\theta). \tag{61}$$

The corresponding minimum value is given by

$$\mathcal{F}(S^*) = \frac{\left[\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta) d\theta\right]^2}{\mathcal{I}}. \tag{62}$$

Proof: Introducing a multiplier λ and considering a variation $\delta S(\theta)$ we obtain

$$\delta\mathcal{F} = \int_{-\infty}^{+\infty} \left[\frac{-W(\theta)}{S^2(\theta)} - \lambda\right] \delta S(\theta) d\theta \tag{63}$$

which after setting the variation to zero yields

$$S(\theta) = \frac{-1}{\lambda} \left[W(\theta)\right]^{\frac{1}{2}}. \tag{64}$$

The latter is equivalent to (61), and the multiplier λ can be obtained from the constraint as

$$-\frac{1}{\lambda} = \frac{\mathcal{I}}{\int_{-\infty}^{+\infty} \left[W(\theta)\right]^{\frac{1}{2}}}. \tag{65}$$

The optimum value of the functional is found after substituting (64) into $\mathcal{F}(S)$ to yield (62). □

1) Theorem 1: Notice that $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$, considered as a function of θ, cannot vary over the whole space of functions, but over a restricted class dictated by (17), subject to the constraint $\sum_k \rho_k = 1$. Minimizing (20) is complex, but as pointed out before it can be relaxed to a simpler problem by translating the constraint over $\boldsymbol{\rho}$ into a constraint over $S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho})$. To this end, consider

$$\int_{-\infty}^{+\infty} S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) d\theta = \int_{-\infty}^{+\infty} \sum_k \frac{\rho_k p^2(\tau_k - \theta)}{F(\tau_k - \theta)\left[1 - F(\tau_k - \theta)\right]} d\theta \tag{66}$$

and interchange summation with integration to obtain

$$\int_{-\infty}^{+\infty} S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) d\theta = \sum_k \rho_k \int_{-\infty}^{+\infty} \frac{p^2(\tau_k - \theta)}{F(\tau_k - \theta)\left[1 - F(\tau_k - \theta)\right]} d\theta. \tag{67}$$

But now, note that the integrals inside the summation are all equal since they contain shifted versions of the same function. Moreover, recalling that $\sum_k \rho_k = 1$, we obtain

$$\int_{-\infty}^{+\infty} S(\theta, \boldsymbol{\tau}, \boldsymbol{\rho}) d\theta = \int_{-\infty}^{+\infty} \frac{p^2(u)}{F(u)\left[1 - F(u)\right]} du := \mathcal{I}_1. \tag{68}$$

Thus, we can relax (20) to

$$\arg\min_S \frac{1}{N} \int_{-\infty}^{+\infty} \frac{W(\theta)}{S(\theta)} d\theta, \qquad \text{s.t.} \quad \int_{-\infty}^{+\infty} S(\theta) d\theta = \mathcal{I}_1. \tag{69}$$

It is important to stress that (20) and (69) are not equivalent because the functions $\{S(\theta,\boldsymbol{\tau},\boldsymbol{\rho})\}$ subject to $\sum_k \rho_k = 1$ are a *subset* of the functions $\{S(\theta)\}$ subject to $\int_{-\infty}^{+\infty} S(\theta)d\theta = \mathcal{I}_1$. However, the very condition of being a subset dictates that the solution of (69) is a lower bound on the solution of (20).

Invoking Lemma 1 to solve the variational problem in (69), we obtain the bound (21), and requiring $S(\theta,\boldsymbol{\tau},\boldsymbol{\rho}) = S^*(\theta)$ renders the condition for this bound achievable. □

2) Theorem 2: Starting from (29) we can rephrase the problem of finding the optimum set $(\boldsymbol{\tau},\boldsymbol{\rho})$, as that of finding the set that minimizes the average of the bound (29) over the weighting function $W(\theta)$

$$\mathcal{B}_W(\boldsymbol{\tau},\boldsymbol{\rho}) \leq \frac{1}{N}\int_{-\infty}^{+\infty} \frac{W(\theta)}{T(\theta,\boldsymbol{\tau},\boldsymbol{\rho})}d\theta. \tag{70}$$

As before, we relax the problem to the minimization of a functional $\mathcal{F}(T)$ subject to a constraint on the integral of $T(\theta,\boldsymbol{\tau},\boldsymbol{\rho})$

$$\int_{-\infty}^{+\infty} T(\theta,\boldsymbol{\tau},\boldsymbol{\rho})d\theta = \int_{-\infty}^{+\infty} \frac{4}{\sqrt{2\pi}\sigma}\sum_k \rho_k p(\tau_k-\theta)d\theta = \frac{4}{\sqrt{2\pi}\sigma}. \tag{71}$$

By Lemma 1, the minimum bound is achieved when

$$T^*(\theta) = \frac{4}{\sqrt{2\pi}\sigma}\frac{W^{\frac{1}{2}}(\theta)}{\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta)d\theta} \tag{72}$$

and the function that achieves this minimum is

$$\mathcal{F}(T^*) = \frac{\sqrt{2\pi}\sigma}{4}\left[\int_{-\infty}^{+\infty} W^{\frac{1}{2}}(\theta)d\theta\right]^2. \tag{73}$$

Finally, and this time different from the proof of Theorem 1, we can go a step forward. According to the hypothesis, the coefficients $\mathbf{c} = \{c_k, k\in\mathbf{Z}\}$ of the Gabor expansion of $W^{1/2}(\theta)$ over the basis $\mathcal{G} := \{p(\theta - k\tau), k\in\mathbf{Z}\}$ are all positive. Thus, by setting $\boldsymbol{\tau}^\ddagger := \{\tau_k = k\tau, k\in\mathbf{Z}\}$, and $\boldsymbol{\rho}^\ddagger := \{\rho_k = c_k/(\sum_k c_k), k\in\mathbf{Z}\}$ we have that

$$T^*(\theta) = T(\theta,\boldsymbol{\tau}^\ddagger,\boldsymbol{\rho}^\ddagger). \tag{74}$$

Q.E.D. □

Appendix C
Proof of Proposition 4

To compute the CRLB, recall that $q_k(\theta) = F(\tau_k - \theta)$ and differentiate (40) to obtain

$$\frac{\partial L_n(\theta)}{\partial\theta} = \sum_{k=-\infty}^{+\infty}\delta\left[\mathbf{y}_k - \mathbf{b}(n)\right]\left[\frac{[p(\tau_{k+1}-\theta)-p(\tau_k-\theta)]^2}{[F(\tau_{k+1}-\theta)-F(\tau_k-\theta)]^2} - \frac{[\dot{p}(\tau_{k+1}-\theta)-\dot{p}(\tau_k-\theta)]}{F(\tau_{k+1}-\theta)-F(\tau_k-\theta)}\right] \tag{75}$$

where $\dot{p}(u) := \partial p(u)/\partial u$. On the other hand, note that

$$\mathrm{E}\left[\delta\left(\mathbf{y}_k - \mathbf{b}(n)\right)\right] = F(\tau_{k+1}-\theta) - F(\tau_k-\theta). \tag{76}$$

Averaging (75), yields the per-observation CRLB

$$[B_n(\theta)]^{-1} = \sum_{k=-\infty}^{+\infty}\frac{[p(\tau_{k+1}-\theta)-p(\tau_k-\theta)]^2}{F(\tau_{k+1}-\theta)-F(\tau_k-\theta)} - \sum_{k=-\infty}^{+\infty}[\dot{p}(\tau_{k+1}-\theta)-\dot{p}(\tau_k-\theta)]. \tag{77}$$

Finally, note that the second summation is equal to zero from where (43) follows readily.

To prove that $L_n(\theta)$ is concave, note that we can write

$$[q_{k+1}(\theta) - q_k(\theta)] = \int_{\tau_k-\theta}^{\tau_{k+1}-\theta} p(u)du = \int_{(\tau_k,\tau_{k+1})} p(u-\theta)du. \tag{78}$$

Since $[q_{k+1}(\theta) - q_k(\theta)]$ is the integral of a log-concave function $(p(u))$ over a convex region (the segment (τ_k,τ_{k+1})) it is log-concave. $L_n(\theta)$ is concave because it is a weighted sum of logarithms of log-concave functions. □

Appendix D
Proof of Proposition 5

We start by proving a property of the Gaussian Complementary CDF, stated in the following lemma.

Lemma 2: If the positive argument function $f(u) : \mathbf{R}^+ \to \mathbf{R}$, solves

$$F(u) = 2F\left(u + f(u)\right) \tag{79}$$

then, $f(u) < f(0), \forall u > 0$.

Proof: Note that $f(u) > 0$, and take derivatives in both sides of (79) to obtain the derivative of $f(u)$

$$\dot{f}(u) := \frac{df(u)}{du} = \frac{1}{2}e^{-\frac{f(u)[f(u)-2u]}{2\sigma^2}} - 1. \tag{80}$$

Now note that $f(0) = F^{-1}(1/4) \approx 0.67\sigma$, and consequently $\dot{f}(0) \approx -0.37 < 0$ implying that $f(u)$ is decreasing around 0. Supposing that $f(0)$ is reached at other values, let $v > 0$ denote the smallest real number such that $f(v) = f(0)$. From (80) we see that $\dot{f}(u)$ is decreasing in u; so it must be that $\dot{f}(v) < f(u) < 0$, implying that $f(v)$ is decreasing around v. This is a contradiction since $f(u)$ is continuous. □

To prove Proposition 5. we introduce a transmission scheme requiring the number of bits given by (44). Begin by noting that the unconditional distribution of $x(n)$ is Gaussian

$$x(n) \sim \mathcal{N}(\mu,\sigma_x) \sim \mathcal{N}\left(\mu, \sqrt{\sigma_\theta^2 + \sigma^2}\right) \tag{81}$$

and assume without loss of generality that $\mu = 0$.

Given (81), construct the sequence $\Gamma = \{\gamma_j\}_{k=0}^{\infty}$ whose elements satisfy

$$\Pr\left\{x(n) > \gamma_j\right\} = 2^{-(j+1)}. \tag{82}$$

That is, γ_0 is the median of the distribution (i.e., $\gamma_0 = \mu = 0$), γ_1 is such that $(1/4)^{th}$ of the probability lies to its right, γ_2 such that $(1/8)^{th}$, and so on.

Remark 2: Application of Lemma 2 to the definition of Γ allows us to conclude that $\gamma_k - \gamma_{k-1} \leq \gamma_1$.

We point two consequences of the definition of the sequence Γ: first, the probability that $x(n)$ lies between γ_{j-1} and γ_j *given* that $x(n) > 0$ is

$$\Pr\{\gamma_{j-1} < x(n) < \gamma_j\} = 2^{-j} \tag{83}$$

and second in any given interval $(\gamma_j, \gamma j+1)$ the number $N(\gamma_j)$ of equally spaced thresholds τ_k is bounded by

$$N(\gamma_j) < \left\lceil \frac{\gamma_j - \gamma_{j-1}}{\tau} \right\rceil. \tag{84}$$

We can now introduce the transmission scheme as follows:

[S1] If $x(n) \geq 0$ then transmit 1; and if $x(n) < 0$ transmit 0 and change the sign of $x(n)$.

[S2] Transmit 1 if $x(n) > \gamma_j$, and 0 otherwise. Start this process at $j = 1$, and repeat it until the first j for which $x(n) < \gamma_j$. This confines $x(n)$ to the interval (γ_{j-1}, γ_j) that contains a total number of $N(\gamma_j)$ thresholds bounded by (84).

[S3] Enumerate the binary observations $b_k(n)$ over the interval (γ_{j-1}, γ_j) from 1 to $N(\gamma_j)$ and transmit them as follows.

- [S3a] Set $K = \lceil N(\gamma_j)/2 \rceil$, and transmit $b_K(n)$.
- [S3b] Set $K = \lceil K + K/2 \rceil$ if $b_K(n) = 1$, or $K = \lceil K - K/2 \rceil$ if $b_K(n) = 0$.
- [S3c] Repeat **[S3a]**–**[S3b]** until $\tau_K - x(n) < \tau$.

Despite the involved description, the scheme is actually very simple. In steps [S1] and [S2] we divide the line in segments so that the probability of finding $x(n)$ in any of them is 1/2 the probability of finding it in the previous one. After this is completed, we switch back to the binary observations and transmit them using pretty much the same scheme. We start with the binary observation that lies (approximately) in the middle of the interval, and then depending on the value of the binary observation we transmit the observation closest to the first quarter or the third quarter and so on.

Although not strictly needed for the proof we elaborate on two issues. On the one hand, note that the thresholds γ_j are not being used to *define* binary observations, but instead to *transmit* the observations defined by the thresholds τ_k. On the other hand, observe that the rationale for steps [S3a]–[S3c] is that over the interval (γ_{j-1}, γ_j) the pdf of $x(n)$ is more or less constant; i.e., $x(n)$ is approximately uniformly distributed when the event $\{x(n) \in (\gamma_{j-1}, \gamma_j)\}$ is given. This justifies the way the transmission is designed since we expect the probabilities of finding $x(n)$ to the right or the left of the middle threshold τ_K to be equal once we know that $x(n) \in (\gamma_{j-1}, \gamma_j)$.

Now, we compute the expected value of the transmitted number of bits. First, note that the number of bits required in steps [S3a]–[S3c] can be bounded using (84)

$$l_j^{(1)} = \lceil \log_2(N(\gamma_j) - 1\rceil_+ < \left[\log_2 \frac{\gamma_{j+1} - \gamma_j}{\tau}\right]_+ \tag{85}$$

where the -1 comes from the stopping criterion (the last binary observation does not need to be transmitted), and the $+$ subscripts are because $l_j^{(1)}$ cannot be negative.

Second, note that if $x(n) \in (\gamma_{j-1}, \gamma_j)$, then transmitting the sequence Γ requires

$$l_j^{(2)} = j + 1 \tag{86}$$

bits. Combining the number of bits required when $x(n) \in (\gamma_{j-1}, \gamma_j)$ given by (85) and (85), with the probability that $x(n)$ belongs to this interval, we find the expected value of N_t as

$$\begin{aligned} \mathrm{E}(N_t) &= \sum_{j=1}^{+\infty} \left(l_j^{(1)} + l_j^{(2)}\right) 2^{-j} \\ &< \sum_{j=1}^{+\infty} \left[\left(\log_2 \frac{\gamma_{j+1} - \gamma_j}{\tau}\right)_+ + j + 1\right] \frac{1}{2^j}. \end{aligned} \tag{87}$$

To complete the proof, invoke Remark 2 and note that $\gamma_1 = F^{-1}[1/4] = \sigma_x Q^{-1}[1/4]$ to reduce (87) to

$$\mathrm{E}(N_t) < \sum_{j=1}^{+\infty} \frac{1}{2^j} \left[1 + \left(\log_2 Q^{-1}\left[\frac{1}{4}\right] \frac{\sigma_x}{\tau}\right)_+\right] + \sum_{j=1}^{+\infty} \frac{j}{2^j}. \tag{88}$$

Substituting the values of the geometric series ($\sum_{j=1}^{+\infty} 2^{-j} = 1$ and $\sum_{j=1}^{+\infty} j 2^{-j} = 2$), and remembering that $\sigma_x = \sqrt{\sigma_\theta^2 + \sigma^2}$, (44) follows. □

Appendix E
Proof of Proposition 7

Let $b(\theta)$ denote the bias of the QSME,

$$b(\theta) := \mathrm{E}(\bar{x}_Q - \theta) = \mathrm{E}(\bar{x}_Q - \bar{x}), \tag{89}$$

where the second equality follows from the unbiasedness of $\bar{x}$. We can write the MSE in terms of the bias

$$\begin{aligned} &\mathrm{E}\left[(\bar{x}_Q - \theta)^2\right] \\ &\quad = \mathrm{E}\left[\left(\bar{x}_Q - \theta - b(\theta)\right)^2\right] + b^2(\theta) \\ &\quad = \mathrm{E}\left[\left((\bar{x}_Q - \bar{x} - b(\theta)) - (\theta - \bar{x})\right)^2\right] + b^2(\theta) \end{aligned} \tag{90}$$

where in the second equality we added and subtracted $\bar{x}$. The important point is that based on (89) the two variables $(\bar{x}_Q - \bar{x} - b(\theta))$ and $(\theta - \bar{x})$ are zero-mean; hence, we can apply the triangle inequality

$$\begin{aligned} &\mathrm{E}\left[(\bar{x}_Q - \theta)^2\right] \\ &\quad \leq \mathrm{E}\left[\left(\bar{x}_Q - \bar{x} - b(\theta)\right)^2\right] + b^2(\theta) + \mathrm{E}\left[(\theta - \bar{x})^2\right] \\ &\qquad + 2\sqrt{\mathrm{E}\left[\left(\bar{x}_Q - \bar{x} - b(\theta)\right)^2\right] \mathrm{E}\left[(\theta - \bar{x})^2\right]} \\ &\quad \leq \mathrm{E}\left[(\bar{x}_Q - \bar{x})^2\right] + \mathrm{E}\left[(\theta - \bar{x})^2\right] \\ &\qquad + 2\sqrt{\mathrm{E}\left[\left(\bar{x}_Q - \bar{x} - b(\theta)\right)^2\right] \mathrm{E}\left[(\theta - \bar{x})^2\right]} \\ &\quad = \mathrm{E}\left[(\bar{x}_Q - \bar{x})^2\right] + \mathrm{var}(\bar{x}) \\ &\qquad + 2\sqrt{\mathrm{E}\left[\left(\bar{x}_Q - \bar{x} - b(\theta)\right)^2\right] \mathrm{var}(\bar{x})}. \end{aligned} \tag{91}$$

Finally, note that since the quantization error is absolutely bounded by $\tau/2$ we have

$$\begin{aligned}\mathrm{E}\left[(\bar{x}_Q-\bar{x}-b(\theta))^2\right] &\leq \mathrm{E}\left[(\bar{x}_Q-\bar{x})^2\right]\\ &= \frac{1}{N^2}\sum_{n=0}^{N-1}\mathrm{E}\left[(x(n)-x_Q(n))^2\right]\\ &\leq \frac{\tau^2}{4N}. \end{aligned} \tag{92}$$

Substituting (92) and (2) into (91), we obtain

$$\mathrm{E}\left[(\bar{x}_Q-\theta)^2\right] \leq \frac{\sigma^2}{N}+\frac{\tau^2}{4N}+2\sqrt{\frac{\sigma^2}{N}\frac{\tau^2}{4N}} \tag{93}$$

which after simplifying establishes (52). □

Acknowledgment

The authors would like to thank Prof. Z.-Q. Luo of the University of Minnesota for his valuable feedback and for providing preprints [11]–[13].

References

[1] M. Abdallah and H. Papadopoulos, "Sequential signal encoding and estimation for distributed sensor networks," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Salt Lake City, UT, May 2001, pp. 2577–2580.

[2] E. Ayanoglu, "On optimal quantization of noisy sources," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1450–1452, Nov. 1990.

[3] B. Beferull-Lozano, R. L. Konsbruck, and M. Vetterli, "Rate-distortion problem for physics based distributed sensing," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, Montreal, QC, Canada, May 2004, pp. 913–916.

[4] D. Blatt and A. Hero, "Distributed maximum likelihood estimation for sensor networks," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, Montreal, QC, Canada, May 2004, pp. 929–932.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[6] E. Ertin, R. Moses, and L. Potter, "Network parameter estimation with detection failures," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Montreal, QC, Canada, May 2004, pp. 273–276.

[7] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[9] S. Kumar, F. Zao, and D. Shepherd, Eds., *IEEE Signal Process. Mag. (*Special Issue on Collaborative Information Processing*)*, Mar. 2002, vol. 19.

[10] W. Lam and A. Reibman, "Quantizer design for decentralized systems with communication constraints," *IEEE Trans. Commun.*, vol. 41, pp. 1602–1605, Aug. 1993.

[11] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.

[12] ——, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 735–744, Apr. 2005.

[13] Z.-Q. Luo and J.-J. Xiao, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[14] R. D. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2245–2253, Aug. 2002.

[15] 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, no. 3, pp. 978–1002, Mar. 2001.

[16] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[17] 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.

[18] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill Higher Education, 2001.

[19] M. G. Rabbat and R. D. Nowak, "Decentralized source localization and tracking," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 3, Montreal, QC, Canada, May 2004, pp. 921–924.

[20] A. Ribeiro and G. B. Giannakis, "Distributed estimation in Gaussian noise for bandwidth-constrained wireless sensor networks," in *Proc. 38th Asilomar Conf. Signals, Systems, Computers*, vol. 2, Monterrey, CA, Nov. 7–10, 2004, pp. 1407–1411.

[21] J. F. Sturm. Using Sedumi 1.02, a Matlab Toolbox for Optimization Over Symmetric Cones. [Online]. Available: http://fewcal.kub.nl/sturm/software/sedumi.html

[22] Y. Sung, L. Tong, and A. Swami, "Asymptotic locally optimal detector for large-scale sensor networks under the Poisson regime," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Montreal, QC, Canada, May 2004, pp. 1077–1080.

[23] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

Alejandro Ribeiro (S'05) received the B.Sc. degree in electrical engineering from Universidad de la Republica Oriental del Uruguay, Montevideo, Uruguay, in 1998. Since May 2003, he has been working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

From 1998 to 2003, he was a member of the Technical Staff at Bellsouth Montevideo, Uruguay. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on wireless cooperative communications, random access, wireless ad hoc and sensor networks.

Mr. Ribeiro is a Fulbright Scholar.

Georgios B. Giannakis (F'97) received the Diploma in Electrical Engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 200 journal papers, 350 conference papers, and two edited books. Current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space–time coding, multicarrier, ultra-wide-band wireless communication systems, cross-layer designs, and sensor networks.

Dr. Giannakis is the (co-)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (in 1992, 1998, 2000, 2001, 2003, and 2004). He also received the SP Society's Technical Achievement Award in 2000. He served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as member of the SP Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a Member of the IEEE Fellows Election Committee. He has also served as a Member of the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.