Jin-Jun Xiao, Alejandro Ribeiro, Zhi-Quan Luo, and Georgios B. Giannakis

Distributed Compression-Estimation Using Wireless Sensor Networks

[The design goals of performance, bandwidth efficiency, scalability, and robustness]

A wireless sensor network (WSN) consists of a large number of spatially distributed signal processing devices (nodes), each with finite battery lifetime and thus limited computing and communication capabilities. When properly programmed and networked, nodes in a WSN can cooperate to perform advanced signal processing tasks with unprecedented robustness and versatility, thus making WSN an attractive low-cost technology for a wide range of remote sensing and environmental monitoring applications [1], [32].

Prolonging the lifetime of a WSN is important for both commercial and tactical applications. With nonrechargeable batteries, this requirement places stringent energy constraints on the design of all WSN operations. Energy limitation is one of the major differences between a WSN and other wireless networks such as wireless local area networks, where energy efficiency is of a lesser concern. Also, WSNs are often self-configured networks with little or no pre-established infrastructure as well as a topology that can change dynamically. Moreover, there may be physical obstacles in the network environment that can degrade considerably the wireless links among sensors. All these present formidable challenges to the design of communication, networking, and local signal processing algorithms performed by a WSN. In this article, we focus on distributed estimation tasks performed by a WSN under energy and bandwidth constraints.

Since data are collected by sensors at geographically distinct locations, estimation using a WSN requires not only local information processing but also intersensor communications. The latter brings in a wireless communication and networking aspect of the problem that is absent from the traditional centralized estimation framework. In fact, a major challenge in WSN research is the integrated design of local signal processing operations and strategies for intersensor communication and networking so as to strike a desirable tradeoff among energy efficiency, simplicity, and overall system performance. For instance, to maximize battery lifetime and reduce communication bandwidth, it is essential for each sensor to locally compress its observed data so that only low rate intersensor communication is required. This motivates joint design of the compression-estimation module per sensor.

Designing distributed compression-estimation algorithms in the context of a WSN differs from the traditional centralized framework in several important aspects.

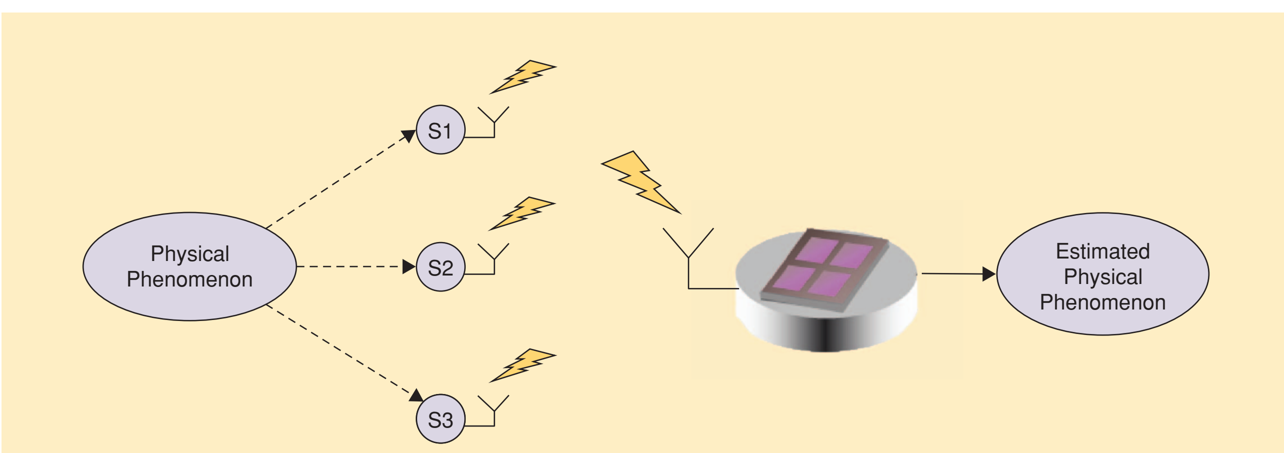
- Constraints on sensor cost, bandwidth, and energy budget dictate that *low* quality sensor observations may have to be aggressively quantized, e.g., down to a few bits per sample per node. Thus, estimators must be developed based on severely quantized versions of very noisy observations.
- Obtaining the complete signal models for a large number of sensors may be impractical, particularly in dynamic sensing environments. This preempts application of optimum estimation algorithms and motivates distributed estimators based on partially known or unknown data/noise models.
- Sensors may enter or leave the network dynamically, resulting in unpredictable changes in network size and topology. Thus, to ensure robust operation, compression-estimation algorithms for WSNs have to work with limited (or no) knowledge of the network topology and/or size.
- Local compression at a sensor node depends not only on the quality of sensor observation, but also on the quality of the wireless communication channel(s) from the node.

In addition, the design of distributed algorithms should be coupled with the underlying WSN topology. We consider two popular WSN deployments characterized by the presence or absence of a fusion center (FC).

- When an FC is present, there is no intersensor communication; communication is only between sensors and the FC. The FC collects locally processed data and produces a final estimate; see Figure 1.
- In ad hoc WSNs, there is no FC. The network itself is responsible for processing the collected information, and to this end, sensors communicate with each other through the shared wireless medium; see Figure 2.

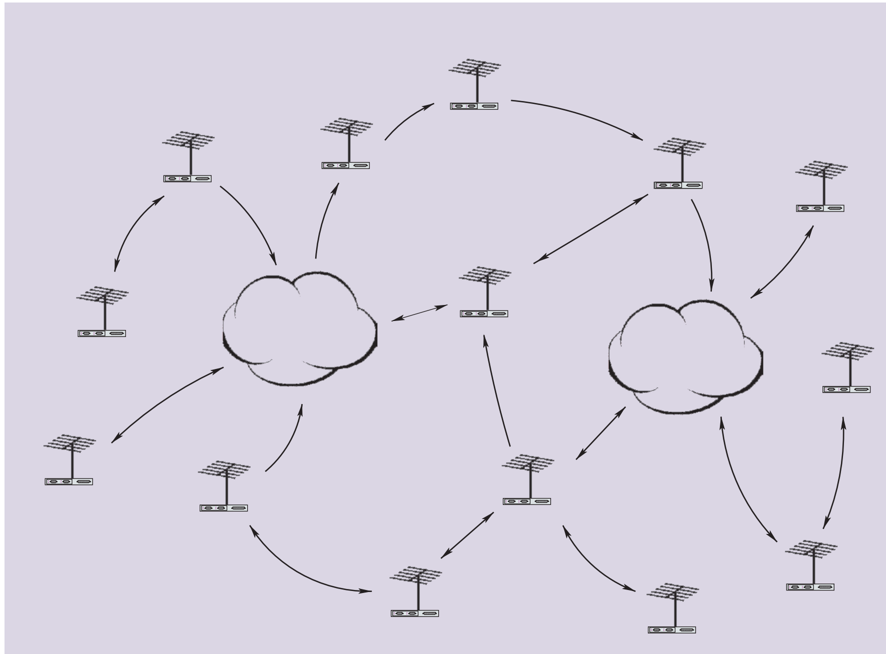
Hybrids are also possible in which the WSN is partitioned into clusters possibly with a hierarchical structure. Each cluster has a local FC generating intermediate estimates, which in turn are combined to obtain a final estimate.

The focus of this article is on distributed compression and estimation using WSNs in which the main design goals are performance, bandwidth efficiency, scalability, and robustness to changes in the network or environment. (The distributed detection in WSNs is discussed in [10].) We first pursue deterministic parameter estimators and study the intertwining tasks of quantization and estimation in: i) low signal-to-noise ratio (SNR) situations where the noise standard deviation is in the order of the parameter's dynamic range and ii) universal



[FIG1] A WSN topology with an FC.

estimation where the sensor data and noise model are unknown. The ultimate objective is to understand how the signal processing capability of a WSN scales up with its size and to develop robust distributed signal processing algorithms and protocols with low bandwidth requirements and optimal performance. We will see that in the low SNR regime, universal distributed estimators not only exist but also achieve performance close to that of estimators based on the original (nonquantized) observations. Moreover, since network resources (e.g., power and bandwidth) are scarce, their optimal allocation and scheduling can lead to significant savings.

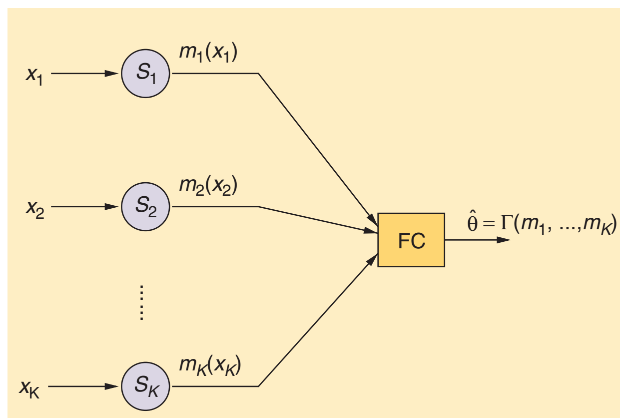


[FIG2] Ad hoc WSN.

The techniques and basic results that are derived for the parameter estimation paradigms outlined first are later extended to more general and practical signal models. A Bayesian estimation framework is laid out along with an application to state estimation of dynamical stochastic processes. The final part of the article addresses several issues pertaining to WSNs with an FC from an information theoretic point of view. These properties not only offer performance benchmarks for distributed signal processing but also provide general guidelines for algorithmic designs.

DISTRIBUTED ESTIMATION FRAMEWORK

Let us consider a generic distributed estimation problem using a WSN with an FC. Our goal is to estimate a $p \times 1$ vector parameter $\boldsymbol{\theta} \in \mathbb{R}^p$ from K independent scalar observations x_k collected by as many distributed sensors, as depicted in Figure 3. The observations obey the model



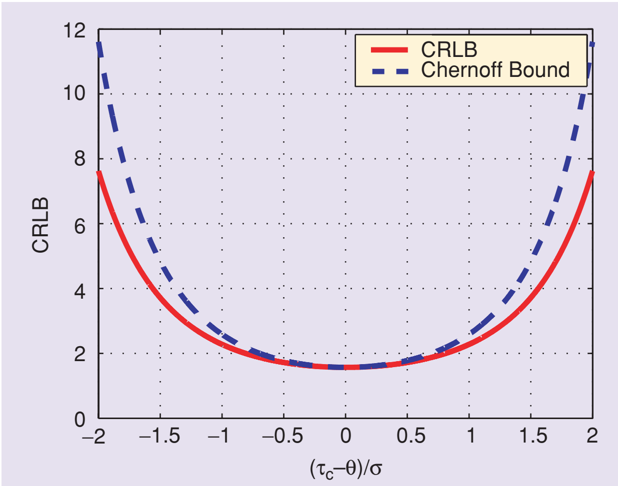
[FIG3] Distributed estimation setup.

$$x_k = \phi_k(\theta) + w_k, \quad k = 1, \ldots, K, \tag{1}$$

where $\phi_k : \mathbb{R}^p \to \mathbb{R}$ is generally a nonlinear function and the noise terms w_k, $k = 1, \ldots, K$ are zero-mean independent random variables with variance $\sigma_k^2 := E(w_k^2)$. Let $p_k(w)$ be the probability density function (pdf) of w_k and $F_k(w) := \int_w^\infty p_k(u)du$ denote the corresponding complementary cumulative distribution function (ccdf). Although our focus is on the parameter estimation problem in (1), the methods here can be extended to nonparametric models as well. Interested readers are also referred to [35], which discusses robust nonparametric methods using distributed learning.

Distributed estimation using a WSN entails a local compression stage in which sensors perform local quantization of their observations to obtain finite-rate messages $m_k(x_k)$. These messages are then sent to the FC where a final estimate $\hat{\boldsymbol{\theta}} = \Gamma(m_1, \ldots, m_K)$ is generated.

If infinite bandwidth were available, each sensor could send its analog-amplitude observation x_k to the FC corresponding to the setup discussed in the previous paragraph with $m_k(x_k) = x_k$. Upon receiving these real-valued messages, the FC can use any of a number of estimation techniques, depending on the extent of its prior knowledge about the pdf $p_k(w)$, to generate an optimal (in some statistical sense) estimate $\hat{\boldsymbol{\theta}}_0 = \Gamma_0(x_1, \ldots, x_K)$. If, for example, $\boldsymbol{\theta}$ is scalar (denoted by θ) and we consider the simple signal model $x_k = \theta + w_k$, a popular approach is to compute the best linear unbiased estimator (BLUE) $\hat{\theta}_0 = \hat{\theta}_{\text{BLUE}} := (\sum_{k=1}^K x_k/\sigma_k^2)/(\sum_{k=1}^K 1/\sigma_k^2)$ whose mean-square error (MSE) [24]



[FIG4] Performance penalty of a 1-b estimator with respect to the sample mean estimator. When the parameter's dynamic range is on the order of the observation noise variance, the ratio var($\hat{\theta}$)/var($\hat{\theta}_0$) is not large.

$$E[(\hat{\theta}-\theta)^2] = \left(\sum_{k=1}^{K}\frac{1}{\sigma_k^2}\right)^{-1}, \tag{2}$$

is minimum among all linear unbiased estimators. If furthermore, the noise at sensor k adheres to a Gaussian pdf $\mathcal{N}(0, \sigma_k^2)$, then $\hat{\theta}_{\text{BLUE}}$ is the minimum variance unbiased estimator (MVUE) that minimizes the estimator variance var $(\hat{\theta})$ for all values of θ. In the particular case $\sigma_k^2 = \sigma^2$ for all $k = 1, 2, \ldots, K$, we obtain the sample mean estimator $\bar{x} := (1/K)\sum_{k=1}^{K} x_k$ whose MSE is known to be σ^2/K.

Having each sensor send the analog-amplitude x_k to the FC may violate the severe bandwidth and power constraints that sensors are envisioned to obey. In such cases, it may be preferable to let each sensor transmit a quantized version of x_k to the FC in the form a finite rate message $m_k(x_k)$, to enable forming at the FC the estimator $\hat{\boldsymbol{\theta}} = \Gamma(\hat{m}_1, \ldots, \hat{m}_K)$ based on the received messages $\hat{m}_1, \ldots, \hat{m}_K$ (which are versions of $m_1(x_1), \ldots, m(x_k)$ corrupted by the noisy channel). Naturally, the MSE performance of $\hat{\boldsymbol{\theta}} = \Gamma(\hat{m}_1, \ldots, \hat{m}_K)$ is in general inferior to that of $\hat{\boldsymbol{\theta}}_0 = \Gamma_0(x_1, \ldots, x_K)$ due to quantization- and channel-induced errors. Even though the optimal centralized estimator $\hat{\boldsymbol{\theta}}_0 = \Gamma_0(x_1, \ldots, x_K)$ based on analog-amplitude observations may be impractical in a WSN context, it serves as a useful clairvoyant benchmark to evaluate the performance of distributed estimators $\hat{\boldsymbol{\theta}}$.

Our goal in the remaining sections is to

i) derive efficient local quantization schemes $m_k(x_k)$ and distributed estimators $\hat{\theta} = \Gamma(m_1, \ldots, m_K)$ under energy and bandwidth constraints

ii) benchmark their MSE performance and quantify the performance loss when compared to the centralized clairvoyant estimators $\hat{\theta}_0$

iii) ensure low-complexity $\{\Gamma, m_k : k = 1, \ldots, K\}$ alternatives

iv) design adaptive resource (e.g., power and bandwidth) allocation and scheduling strategies to improve the overall network performance.

DISTRIBUTED ESTIMATORS

In this section, we will present distributed estimators for: i) known univariate noise pdfs, ii) known noise pdfs with a finite number of unknown parameters, iii) completely unknown noise pdfs, and iv) generalizations to multivariate and possibly correlated pdfs. Even though the estimators will turn out to require minimal communication overhead from the sensors to the FC, they will exhibit essentially identical MSE performance and comparable complexity with the corresponding clairvoyant estimators.

COMPLETELY KNOWN PDF

Let us start by considering the signal model

$$x_k = \theta + w_k \tag{3}$$

when the noise pdf $p_k(w) = p(w)$ for all k and $p(w)$ is known. Albeit simple, this model will illustrate basic properties that carry over to more pragmatic models we will consider later. For simplicity, we will impose a rate constraint of one binary bit per sensor sample, but our results can be easily extended to any fixed number of bits per sensor sample. For binary messages (i.e., $L_k = 1$), we can consider the halfline $B_c := (\tau_c, \infty) \in \mathbb{R}$ and define the message functions as $m_k(x_k) = \mathbf{1}\{x_k \in (\tau_c, \infty)\}$ indicating whether the observation x_k belongs to B_c or not. Moreover, we make the simplifying assumption that the channels from the sensors to the FC are ideal, so that $\hat{m}_k(x_k) = m_k(x_k)$, for all k.

Given that the noise is i.i.d. and the noise ccdf $F_w(w)$ is known, it is easy to find the maximum likelihood estimator (MLE) $\hat{\theta}_{\text{MLE}} = \Gamma_{\text{MLE}}(m_1, \ldots, m_K)$ [6]. Indeed, since m_k is an indicator variable, it is Bernoulli distributed with parameter given by the probability $q := \Pr\{x_k \in B_c\} = F_w(\tau_c - \theta)$. As q and θ are related by a one-to-one function and the MLE of q is $\hat{q} = (1/K)\sum_{k=1}^{K} m_k$, we deduce from the invariance property of MLEs the closed-form expression [34], [38]

$$\hat{\theta}_{\text{MLE}} = \tau_c - F_w^{-1}\left(\frac{1}{K}\sum_{k=1}^{-K} m_k\right). \tag{4}$$

Although m_k is a discontinuous function of x_k, $\hat{\theta}_{\text{MLE}}$ is an estimator whose computational cost is in the order of the optimal clairvoyant estimators such as the sample mean estimator $\bar{x}$ in (2).

The Cramer-Rao lower bound (CRLB) for estimating θ based on $\{m_k\}_{k=1}^{K}$ provides a performance limit for the variance of any estimator $\hat{\theta} = \Gamma(m_1, \ldots, m_K)$ and it is achieved by $\hat{\theta}_{\text{MLE}}$ for K sufficiently large. For our problem, the CRLB is given by [38], $\mathcal{B}(\theta) := F_w(\tau_c - \theta)[1 - F_w(\tau_c - \theta)]/p^2(\tau_c - \theta)$, from which we infer that the ultimate performance limit is determined by the distance between τ_c and θ. For the particular case of Gaussian noise, we can define $\Delta_c := (\tau_c - \theta)/\sigma$ as the (σ)-distance between the parameter θ and the threshold τ_c measured in standard deviation units and let $Q(v)$ denote the standardized Gaussian ccdf. Since the noise is Gaussian, the sample mean estimator $\bar{x}$ is the MVUE with variance var $(\hat{\theta}_0) = \sigma^2/K$.

Compared with this benchmark estimator, the one in (4) incurs loss measured by the ratio $\mathcal{B}(\theta)/\mathrm{var}(\hat{\theta}_0) = (2\pi)Q(\Delta_c)[1-Q(\Delta_c)]/e^{-\Delta_c^2} \le (\pi/2)e^{\Delta_c^2/2}$, which we depict in Figure 4 versus Δ_c [38], [39].

Figure 4 reveals something unexpected: relying on a single bit per x_k, the estimator in (4) may exhibit only $\pi/2$ times higher variance compared to the clairvoyant $\hat{\theta}_0 = \bar{x}$ that relies on the nonquantized data x_k. But this minimal loss in performance corresponds to the ideal choice $\Delta_c = 0$, which implies $\tau_c = \theta$ and requires perfect knowledge of the unknown θ for selecting the quantization threshold τ_c. How do we select τ_c and how much do we lose when the unknown θ lies anywhere in $(-\infty, \infty)$, or when θ lies in $[\Theta_1, \Theta_2]$, with Θ_1, Θ_2 finite and known a priori. Intuition suggests selecting the threshold as close as possible to the parameter. This can be realized with an iterative estimator $\hat{\theta}^{(i)}$, which can be formed as in (4), using $\tau_c^{(i)} = \hat{\theta}^{(i-1)}$, the parameter estimate from the previous $(i-1)$st iteration. As we will see later, this iterative threshold placement matches nicely with state estimation of dynamical processes based on binary observations.

But in the batch formulation considered herein, selecting τ_c is challenging; a closer look at $\mathcal{B}(\theta)$ confirms that the loss can be huge if $\tau_c - \theta \gg 0$. The implication of the latter is twofold: i) since the loss shows up in the CRLB, the potentially high variance of estimators based on quantized observations is inherent to the possibly severe bandwidth limitations of the problem itself and is not unique to a particular estimator; ii) how successful the τ_c selection is depends on the dynamic range $|\Theta_1 - \Theta_2|$ that makes sense because the latter affects the error due to the quantization of x_k to m_k. In fact, two sources of error are present in joint quantization-estimation problems: quantization and noise.

To account for both, the proper figure of merit for estimators based on binary observations is the quantization SNR (Q-SNR) that we define as [39]

$$\gamma := \frac{|\Theta_1 - \Theta_2|^2}{\sigma^2}. \tag{5}$$

Notice that contrary to common wisdom, the smaller Q-SNR is, the easier it becomes to select τ_c judiciously. Furthermore, the variance increase in $\mathcal{B}(\theta)$ relative to the variance of the clairvoyant $\hat{\theta}_0$ is smaller, for a given σ. This is because as the Q-SNR increases the problem becomes more difficult in general, but the rate at which the estimation variance increases is smaller for the CRLB in Figure 4 than for var $(\hat{\theta}_0) = \sigma^2/K$.

KNOWN NOISE PDF WITH UNKNOWN PARAMETERS

The estimator in (4) requires perfect knowledge of the noise pdf $p_w(w)$, which may not always be available. Here we suppose that $p_w(w) = p_w(w;\psi)$ is known and depends on the parameter vector $\psi \in \mathbb{R}^{L\times 1}$, which is unknown. Consider, for example, the case frequently encountered in practice in which the noise pdf is known (say Gaussian) except for its variance $E(w_k^2) = \sigma^2$. Note that the problem of estimating θ when the noise pdf is $p_w(w;\sigma)$ can be addressed by writing $x_k = \theta + \sigma v_k$ with $E(v_k^2) = 1$ and estimating θ while viewing σ as a nuisance parameter [39].

If we define a single quantization region as in "Distributed Estimators," different combinations (θ, σ) lead to sets of messages $\{m_k\}_{k=1}^K$ with identical probabilities. To avoid this ambiguity problem, we define two regions $B_j := (\tau_j, \infty)$, $j = 1, 2$ with $\tau_1 < \tau_2$ and let half the sensors use B_1 to construct their binary observations and the remaining half use B_2. Accordingly, the messages are defined as $m_k := \mathbf{1}\{x_k \in (\tau_1, \infty)\}$ for $k \in [1, K/2]$ and $m_k := \mathbf{1}\{x_k \in (\tau_2, \infty)\}$ for $k \in [K/2+1, K]$.

As well as in the previous subsection, the m_k messages are Bernoulli with parameters $q_j := \Pr\{x_k \in B_j\} = F_v[(\tau_j - \theta)/\sigma]$ depending on whether S_k uses threshold τ_1 or τ_2 to construct m_k. These expressions for the Bernoulli parameters imply that (θ, σ) and (q_1, q_2) are related by the nonlinear 2×2 mapping $[q_1, q_2]^T = F_v[(\tau_1 - \theta)/\sigma], F_v[(\tau_2 - \theta)/\sigma]^T$ that can be inverted to express (θ, σ) in terms of (q_1, q_2). This, plus the invariance property of MLEs leads to [39]

$$\hat{\theta}_{\mathrm{MLE}} = \frac{F_v^{-1}(\hat{q}_2)\tau_1 - F_v^{-1}(\hat{q}_1)\tau_2}{F_v^{-1}(\hat{q}_2) - F_v^{-1}(\hat{q}_1)}, \tag{6}$$

where the MLEs of q_1, q_2 can be found as $\hat{q}_1 = (2/K)\sum_{k=1}^{K/2} m_k$ and $\hat{q}_2 = (2/K)\sum_{k=K/2+1}^{K} m_k$. Likewise, we can obtain the MLE of σ if we are interested in the noise power.

Upon defining the σ-distances $\Delta_j = (\tau_j - \theta)/\sigma$ and the ratios $\mathcal{B}_j(\theta) := F_v(\Delta_j)[1 - F_v(\Delta_j)]/p^2(\Delta_j)$, the CRLB can be written as $\mathcal{B}(\theta)/(\sigma^2/K) = 2/(\Delta_2 - \Delta_1)[(\Delta_2/\Delta_1)B_1(\theta) + (\Delta_1/\Delta_2)B_2(\theta)]$. Interestingly, $\mathcal{B}(\theta)$ is a linear combination of $\mathcal{B}_1(\theta)$, $\mathcal{B}_2(\theta)$ that are identical to the ratio depicted in Figure 4. This establishes that the variance penalty with respect to the clairvoyant sample mean estimator is still a relatively small factor when the Q-SNR takes small-to-medium values [39].

The approach here can be generalized to noise pdfs that depend on L parameters by defining $L+1$ regions $B_l := (\tau_l, \infty)$ and dividing the sensors in $L+1$ groups so that the lth group constructs their binary observations as $m_k = \mathbf{1}\{x_k \in B_l\}$. (See Figure 5.) This applies when the noise adheres to, e.g., a Gaussian mixture pdf. In all these cases we find that a low complexity MLE can be constructed in closed form by invoking the invariance property of MLEs. The associated normalized penalty $\mathrm{var}(\hat{\theta}_{\mathrm{MLE}})/\mathrm{var}(\hat{\theta}_0)$ is small when γ is. Even when the noise pdf is completely unknown we can develop nonparametric universal estimators sharing the latter property as we show later.

VECTOR PARAMETERS IN COLORED GAUSSIAN NOISE

Results presented earlier can be extended to the vector signal model (1). For illustrative purposes, let us assume that the noise pdf $p_k(w)$ and the corresponding ccdf $F_k(w)$ are known but may change from sensor to sensor and recall that we denote the noise power as $E(w_k^2) = \sigma_k^2$.

As before, we define 1-b messages $m_k := \mathbf{1}\{x_k \in (\tau_k, \infty)\}$, and note that m_k is Bernoulli distributed with parameter $q_k := \Pr\{x_k \in (\tau_k, \infty)\} = F_k[\tau_k - \phi_k(\theta)]$. Defining the log-likelihood function

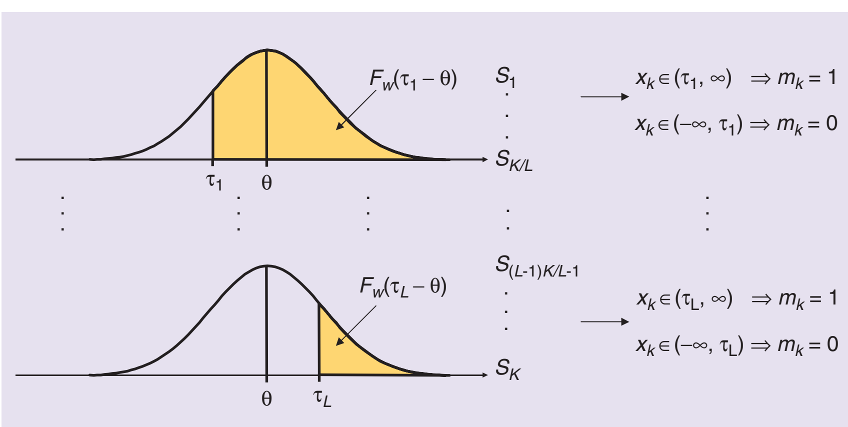
$$L(\boldsymbol{\theta}) := \sum_{k=0}^{K-1} m_k \ln q_k + (1 - m_k) \ln(1 - q_k), \qquad (7)$$

we can find the MLE of $\hat{\boldsymbol{\theta}}$ based on observations $\{m_k\}_{k=0}^{K-1}$ as $\hat{\boldsymbol{\theta}}_{\text{MLE}} := \arg\max_{\hat{\theta}}\{L(\boldsymbol{\theta})\}$ [39].

The search for $\hat{\boldsymbol{\theta}}$ can be challenging due to the multimodal nature of $L(\boldsymbol{\theta})$ as well as the numerical difficulties caused by q_k values being close to zero. However, if a1) the noise pdfs $p_k(w)$ are log-concave, and a2) the functions $\phi_k(\boldsymbol{\theta})$ are linear, then the likelihood $L(\boldsymbol{\theta})$ becomes a concave function of $\boldsymbol{\theta}$. The concavity of $L(\boldsymbol{\theta})$ implies that computationally efficient search algorithms e.g., interior point methods, are guaranteed to converge to the global maximum $\hat{\boldsymbol{\theta}}_{\text{MLE}}$. Note that a1) is satisfied by common noise pdfs, including the multivariate Gaussian, uniform in a convex set, as well as generalized Gaussian [7, p. 104]; while a2) is typical in parameter estimation. Moreover, even when a2) is not satisfied, linearizing $\phi_k(\boldsymbol{\theta})$ using Taylor's expansion is a common first step, typical in, e.g., parameter tracking applications.

To quantify the performance penalty in the vector case, we define the equivalent noise powers $\rho_k^2 := F_k(\tau_k - \phi_k(\boldsymbol{\theta}))[1 - F_k(\tau_k - \phi_k(\boldsymbol{\theta}))]/p^2[\tau_k - \phi_k(\boldsymbol{\theta})]$, and consider two signal models according to (1) with the noise powers given by $\sigma := [\sigma_0^2, \ldots, \sigma_{K-1}^2]^T$ and $\rho := [\rho_0, \ldots, \rho_{K-1}]^T$, respectively. It can be shown that the CRLB $\mathcal{B}_x(\boldsymbol{\theta}; \boldsymbol{\rho})$ when estimating $\boldsymbol{\theta}$ based on $\{x_k\}_{k=1}^K$ with noise powers given by $\boldsymbol{\rho}$, coincides with the CRLB $\mathcal{B}_m(\boldsymbol{\theta}; \boldsymbol{\sigma})$ associated with the estimation of θ based on $\{m_k\}_{k=0}^{K-1}$ when the noise powers are the components of $\boldsymbol{\sigma}$ [38], [39]. Equivalently, it follows that performance of a centralized estimator when the noises have variance ρ_k^2 coincides with the performance of a single-bit distributed estimator when the noise variances are ρ_k^2 with the ratio ρ_k^2/σ_k^2 characterized as in Figure 4.

Even though we considered scalar observations so far, the results generalize to vector observations $\mathrm{x}_k = \phi_k(\boldsymbol{\theta}) + \mathrm{w}_k$ as long as the components of w_k are independent (e.g., white Gaussian noise). If w_k is Gaussian but colored, the approach described here can also be used after local prewhitening [39].



[FIG5] When the noise pdf $p_w(w; \psi)$ is known but depends on L unknown parameters we divide the sensors in L groups each using a different threshold τ_l to construct the binary message m_k. Note that the messages are Bernoulli distributed with parameters $q_l := \Pr\{x_k \in (\tau_l, \infty)\} = F_w(\tau_l - \theta)$ when sensor S_k uses the threshold τ_l.

UNIVERSAL APPROACHES

As shown earlier, optimal distributed estimators depend on the parametric model and the noise pdfs. In certain cases though, characterizing the exact sensor observation distributions for a large number of sensors may be impossible, especially in a dynamic sensing environment. Such applications motivate universal distributed estimators that are independent of the noise or parameter distributions, under either bandwidth or energy constraints.

ESTIMATION IN A HOMOGENEOUS ENVIRONMENT

Again, let us consider a WSN with an FC and the signal model (3) where w_k are spatially uncorrelated with zero mean but otherwise unknown. For the moment, let us also assume that all wireless channels are orthogonal and distortionless. As discussed earlier, if the sensors could communicate their real-valued observations to the FC error free, then the sample mean estimator achieves an MSE performance of σ^2/K, implying that the WSN has *an estimation capability that scales linearly with network size K*. We have seen that under a rate constraint of one binary bit per sensor sample, the same $O(1/K)$ scaling law remains valid when the noise pdf is completely or partially known. Surprisingly, this scaling law can even be achieved by universal distributed estimators, as we explain below.

The idea is to represent sensor observations in binary form and quantize them to different bit positions across sensors. Specifically, we can have $1/2$ of the sensors quantize their observations to the first most significant bit (MSB), $1/4$ of the sensors quantize their observations to the second MSB, $1/8$ of the sensors quantize their observations to the third MSB, and so on [27]. The resulting bits are then used as the 1-b messages for individual sensors. The FC simply averages the received 1-b messages to generate an estimate of ϕ. Clearly, this distributed estimator is universal as it is completely independent of the noise pdf. Assume all observations x_k are bounded in an interval $[-W, W]$ and they are conditionally independent given θ. Then, the mean of these message functions $(m_1 + m_2 + \cdots + m_K)/K$ estimates θ with an MSE upper bounded by W^2/K. Notice that this estimation scheme assigns more sensors to estimate the first MSB of θ than any other bit. This is intuitively reasonable since getting the first MSB of θ right has the highest impact on minimizing the final MSE.

One limitation of the aforementioned strategy is that it requires the use of an FC and knowledge of network size K to specify which sensor should quantize its observation to which bit. Moreover, the resultant estimator is nonisotropic in the sense that sensors quantize their observations to possibly different MSBs. This is difficult to implement in an ad hoc sensor network where there is little or no coordination among sensors. For such networks we can use the following *probabilistic* estimation scheme [28]:

- With each new sample x_k, sensor k flips a coin and, with probability $1/2$, quantizes x_k to the first MSB, with probability $1/4$ quantizes x_k to the second MSB, and so on. The quantization outcome is sent to all its neighbors.
- Messages are communicated among sensors via an underlying WSN protocol. Each sensor recursively computes the average of all received binary messages that are distinct (determined by, say, the sender's ID), and uses it as an estimator of θ.

Intuitively, with the aforementioned coin flipping per sensor, there will be roughly $1/2$ of the sensors in the network quantizing their observations to the first MSB, about $1/4$ of the sensors in the network quantizing their observations to the second MSB, and so on. Thus, this probabilistic estimation scheme should closely approximate the MSE performance of the previous nonprobabilistic one. This is indeed the case. Assume that each message has a header containing the sender's ID and eventually arrives at its destination without error. Then each node in the WSN produces an unbiased estimate of θ with an MSE of at most $4W^2/(K_1+1)$, where K_1 denotes the number of distinct messages received by this sensor. Notice that this probabilistic estimator is *isotropic* and *robust* in the sense that all sensors operate identically and independently and remain oblivious to possible changes in the network size or the noise pdf. This probabilistic distributed estimator can also be adapted for distributed detection [48].

The performance of a universal estimator is characterized by the worst case MSE over all possible distributions of the observations x_k with support $[-W, W]$. Given the binary nature of messages, the message functions must take the form $m_k(x_k) = \mathbf{1}\{x_k \in S_k\}$, indicating whether the observation x_k belongs to S_k or not, where S_k is a subset of $\mathbb{R}$. The design of an optimal 1-bit universal estimator is then to choose $\{S_1, \ldots, S_K; \Gamma\}$ such that $\max_{p_k(w),\theta} E(|\theta - \hat{\theta}|^2)$ is minimized. An example in [53] shows that $\max_{p_k(w),\theta} E(|\theta - \hat{\theta}|^2) \geq W^2/(4K)$. Thus the best achievable MSE for single-bit universal estimators is $W^2/(4K)$, which implies that performance of the universal estimators in [27] and [28] is within a constant factor of 4 to being optimal.

Beyond the simple 1-b per sensor observation, universal estimators can be derived for any fixed rate, and channel distortions can also be accounted for [27].

ESTIMATION IN INHOMOGENEOUS ENVIRONMENTS

In an inhomogeneous sensing environment, different sensors may have different quality of observations due to the fact that sensors closer to the target may have a higher local SNR than those farther away. While characterizing the pdf of sensor observations is difficult in practice, it is often possible for each sensor to characterize its local SNR. This can be accomplished by comparing the received signal power with and without the presence of the signal of interest θ.

In such inhomogeneous environments, it is no longer reasonable to insist on having each sensor transmit identical number of bits to the FC. Intuitively, sensors with higher local SNRs should send more bits to the FC and weigh these bits more than those from sensors with lower SNRs. To this end, we can let each sensor compress its observation to a discrete message with length proportional to the logarithm of its local SNR and then transmit the resulting message to the FC. The final estimate of the unknown parameter is computed at the FC by combining the received bits according to a universal fusion rule. The following distributed estimator in an inhomogeneous sensing environment is proposed in [51].

- At sensor k, choose

$$L_k = \left\lceil \log \frac{W}{\sigma_k} \right\rceil, \tag{8}$$

  and take m_k to be the first L_k bits of the binary expansion of $(W + x_k)/2W \in [0, 1]$.
- The final estimator at the FC is

$$\hat{\theta} = \left(\sum_{k=1}^{K} 2^{2L_k} \right)^{-1} \sum_{k=1}^{K} 2^{2L_k} W(2m_k - 1). \tag{9}$$

To form the estimator in (9), each sensor only needs to know its own noise variance to determine the number of bits L_k. The final fusion (9) is completely determined by the received messages. Thus, such an estimation scheme is totally distributed and easily implemented in a WSN. As expected, higher quality sensors with smaller noise variance send more bits and their messages carry more weight at the final fusion process. Notice that $\hat{\theta}$ in (9) is unbiased, i.e., $E(\hat{\theta}) = \theta$, with MSE

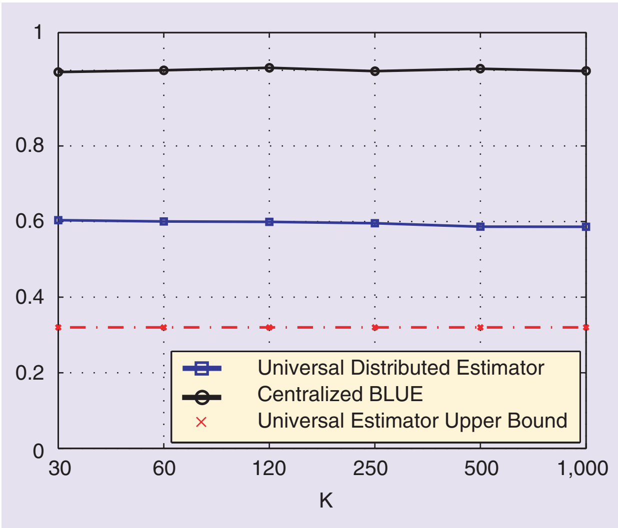
$$E(\hat{\theta} - \theta)^2 < \frac{25}{8} \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)^{-1},$$

which is optimal (up to a factor of 3.125) when compared to the centralized BLUE estimator.

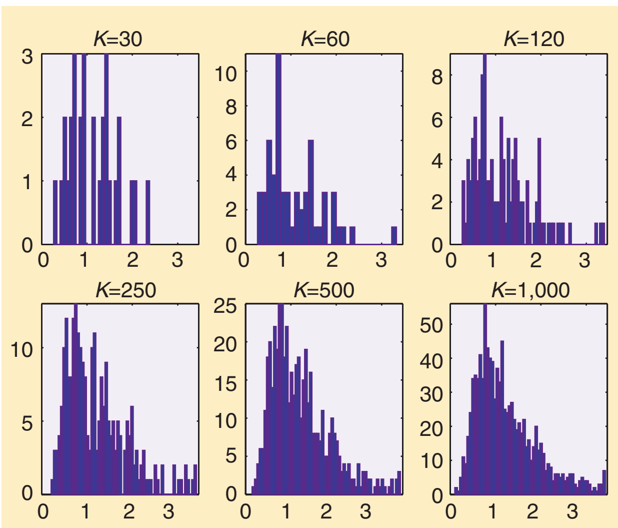
An example from [51] comparing on the basis of MSE the universal estimator in (9) with the centralized BLUE estimator is shown in Figure 6. The asymptotic efficiency is defined as

$$\text{asymptotic efficiency} = \frac{1}{\text{MSE} \cdot \sum_{k=1}^{K} 1/\sigma_k^2}.$$

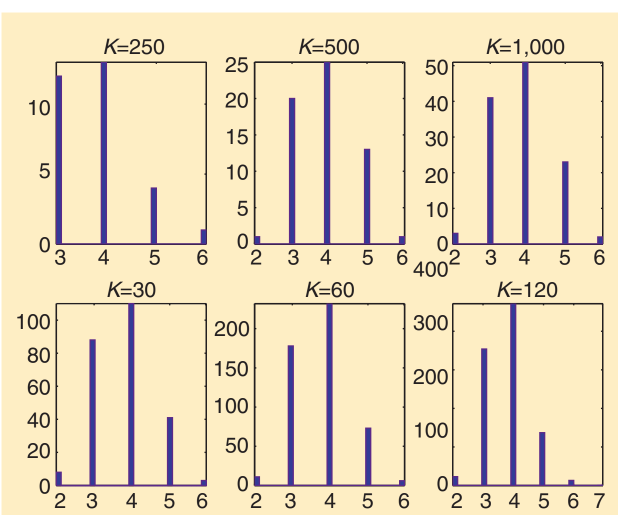
Clearly, the larger the asymptotic efficiency, the more efficient is the estimation scheme. In all the simulation runs, we take $\theta = 1$, $W = 9$. Sensor noises are uniformly distributed with standard deviations shown in Figure 7. The distributions of the number of bits transmitted by all K sensors are plotted in Figure 8. This shows that the universal estimator in (9) requires a surprisingly low communication overhead (about 3.8 b per sample on average) and achieves essentially the same order of MSE as the centralized BLUE estimator.



[FIG6] MSE performance of the universal distributed estimator.



[FIG7] Distribution of sensor noise standard deviations.



[FIG8] Distribution of the number of bits transmitted by local sensors.

ENERGY MINIMIZING ESTIMATION

The estimation schemes so far rely on the idea of adapting the bit allocation depending on the observation SNR. For the purpose of energy efficiency (which has obviously been a design criterion for almost all aspects of WSN design; see [2], [8], [25], and [36]), a sensor should choose a message length of $L_k = 0$ if the quality of its channel to the FC is very poor, even if the quality of its observation is high. Thus, to maximize energy savings, it is necessary to adapt the message length L_k not only based on the local observation SNRs but also based on the intended channel quality. The work in [54] examined an energy minimizing estimation problem by modeling the wireless links between sensors and the FC as additive white Gaussian noise (AWGN) channels with known path-gains g_k. Sensors adopt uncoded quadrature amplitude modulation (QAM) for the quantized bits. Energy models for uncoded M-QAM transmissions are available in [13], [14], and [21]. If sensor k sends L_k bits with QAM of constellation size 2^{L_k} at a bit error probability p_b^k, then the total amount of required transmission energy is given by

$$E_k = \frac{c_k}{g_k}\left(\ln\frac{2}{p_b^k}\right)\left(2^{L_k} - 1\right),$$

where c_k is a system constant. To achieve a target distortion D_0 and minimize the total sensor transmission power, the sensor scheduling problem can be formulated as a convex program, and the optimal value of L_k can be derived in terms of $\{\sigma_k^2, g_k\}$ as [54]

$$L_k^{opt} = \log\left(1 + \frac{W}{\sigma_k}\sqrt{(\eta_0 g_k - 1)^+}\right), \qquad (10)$$

where η_0 is a universal constant decided jointly by the target MSE, sensor noise levels, and channel gains.

The message length in (10) is intuitively appealing as it indicates that the message length should be proportional to the logarithm of the local SNR scaled by the channel path gain. This is in the same spirit as the message length formula in (8) when the channels are ideal, although the latter was derived from a different perspective. Also notice that when $\eta_0 g_k \le 1$, we have $L_k = 0$, and therefore $P_k = 0$. Since g_k is the channel gain, this implies that when the channel quality for sensor k is worse than the threshold η_0, we should discard its observation to save energy. Such a strategy of discarding observations for the purpose of energy saving has been proposed in the context of censoring sensors [37].

To obtain the desired quantization and transmit power levels, we have assumed in this article that the fusion center knows $\{(\sigma_k^2, g_k) : k = 1, 2, \ldots, K\}$. This assumption is reasonable in cases where the network condition and the signal being estimated change slowly in a quasi-static manner. Thus, once $\{(\sigma_k^2, g_k)\}$ are acquired by the fusion center, they can be used for a reasonably long period of time. Also, our approach can be generalized to the estimation of a memoryless discrete-time random process

$\theta(t)$. Due to the temporal memoryless property of the source and sensor observations, we can impose sample-by-sample estimation without significant estimation performance loss but obtain important features such as easy implementation and no coding and estimation delay.

We now present some simulation results from [54]. In the simulations, the parameters are chosen as $K = 1,000$, $\sigma_k^2 = 1$ for all k, and the channel path loss coefficients $a_k = g_k^{-1} = d_k^\alpha$ with $d_k \in [1, 10]$ and $\alpha \in [2, 6]$. Figure 9 (a) illustrates that as the WSN heterogeneity increases, a large number of sensors with low channel gains or low-SNR observations will transmit nothing (i.e., $L_k = 0$). Figure 9(b) also reveals major energy savings compared to uniform quantization or uniform power scheduling. Besides adaptive quantization, other interesting strategies such as protecting different bits with different bit error rates have been discussed in [26].

JOINT ESTIMATION OF A VECTOR SOURCE

The universal distributed estimation can be extended to the general signal model (1) with vector observations. For illustrative purposes, we considered observation model $\mathrm{x}_k = \mathrm{H}_k\theta + \mathrm{w}_k$, where H_k is a matrix with dimension (r_k, p). We assume that noise w_k has zero mean and covariance matrix $\mathbf{C}_k$ but otherwise unknown. Noises n_k are spatially uncorrelated across sensors. Without loss of generality, the source covariance matrix $E(\boldsymbol{\theta}\boldsymbol{\theta}^T)$ is assumed to be identity.

It is possible to extend the universal estimators described earlier to this vector model [29]. There are two main steps in this extension. First, at each sensor, the dimension of x_k can be reduced by adopting the dimensionality reduction strategy proposed in [30]. It turns out that to perform the centralized BLUE estimator, each sensor only needs to send to the FC a number real messages equal to rank $(\mathrm{H}_k^T\mathrm{C}_k^{-1}\mathrm{H}_k)$. After reducing the dimension of $\mathbf{x}_k$, a universal quantization is performed on each component, with the number of bits jointly determined by the pair of local matrices $(\mathbf{C}_k, \mathrm{H}_k)$. In particular, to ensure a factor of 2 away from the performance of the centralized BLUE, the number of bits that must be sent from each sensor to the FC is on average no more than $L_k = (1/2)\log\det(\mathrm{I} + \mathrm{H}_k^T C_k^{-1}\mathrm{H}_k)$ binary bits.

The quantity $(1/2)\log\det(\mathrm{I} + \mathrm{H}_k^T C_k^{-1}\mathrm{H}_k)$ coincides with Shannon's capacity of a "virtual AWGN channel" from nature to sensor k with channel matrix given by H_k, noise covariance matrix $\mathbf{C}_k$, and input power given by identity matrix. The fact that L_k represents channel capacity shows nicely that the message length is decided by the number of "useful" bits contained in $\mathrm{x}_k = \mathrm{H}_k\theta + \mathrm{w}_k$. Furthermore, this message length function is reminiscent of that in (8) for the scalar case.

This vector source estimation problem has also been studied in the context of linear decentralized estimation in [31], [43], [55], and [56] for the purpose of dimensionality reduction and power control under both orthogonal and nonorthogonal multiple access.

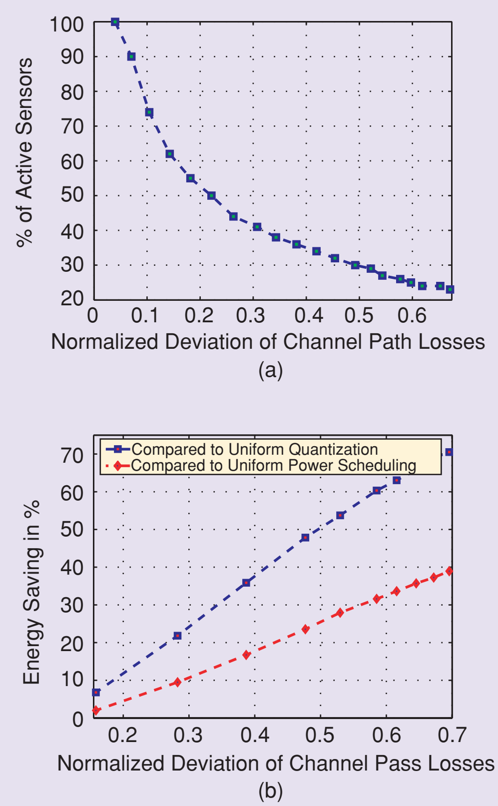
BAYESIAN ESTIMATION OF RANDOM SIGNALS

When knowledge about the parameter of interest is available in the form of a prior distribution $p(\boldsymbol{\theta})$, we can pose our distributed estimation problem in a Bayesian framework. Consider the signal model in (1) and define the messages as $m_k := \mathbf{1}\{x_k \in (\tau_k, \infty)\}$. Letting $\mathrm{m}_{1:K} := [m_1, \ldots, m_K]$ denote the message sequence, the minimum mean-square error (MMSE) estimator can be found as the conditional mean of the posterior distribution $\hat{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathrm{m}_{1:K}]$ with $p[\boldsymbol{\theta}|\mathrm{m}_{1:K}] = p(\mathrm{m}_{1:K}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathrm{m}_{1:K})$ obtained through Bayes' theorem.

Since computing the conditional expectation requires prohibitively expensive numerical integrations, we consider the maximum a posteriori (MAP) estimator $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \arg\max p[\boldsymbol{\theta}|\mathrm{m}_{1:K}]$ that requires numerical maximization instead of a numerical integration [17]. Given that $p(\mathrm{m}_{1:K})$ is a constant and the logarithm is a monotonically increasing function, $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}$ can be found as

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \arg\max\left\{\log[p(\mathrm{m}_{1:K}|\boldsymbol{\theta})] + \log[p(\boldsymbol{\theta})]\right\}, \qquad (11)$$

where $\log[p(\mathrm{m}_{1:K}|\boldsymbol{\theta})]$ coincides with the log-likelihood function in (7). If the noise pdf is log-concave, it can be proved that $\log[p(\mathrm{m}_{1:K}|\boldsymbol{\theta})]$ is a concave function of $\boldsymbol{\theta}$ [17], a result



[FIG9] (a) Number of active sensors decreases as the channel path losses become more heterogeneous. (b) Power savings compared to uniform power scheduling or uniform quantization.

that we already mentioned. Furthermore, if the prior $p(\boldsymbol{\theta})$ is also log-concave, then $\log[p(\boldsymbol{\theta})]$ is concave by definition and, thus, $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}}$ can be found as the maximum of a concave function.

Since $\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} \rightarrow \hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ for K sufficiently large, performance of the MAP estimator approaches that of the MLE we discussed earlier. In fact, it turns out that the penalty paid by the MAP estimator in (11) relative to the clairvoyant MAP based on analog-amplitude observations is smaller than the penalty paid by the respective MLEs [17].

DISTRIBUTED KALMAN FILTERING

Consider an ad hoc WSN deployed to estimate the state of a dynamic stochastic process. Let n denote the time index, $\mathbf{x}(n) \in \mathbb{R}^{p\times 1}$ the state at time n, $v(n,k) \in \mathbb{R}$ the scalar observation of sensor S_k at time n, and consider the following state-observation model

$$\begin{aligned} \mathbf{x}(n) &= \mathbf{A}(n)\mathbf{x}(n-1) + \mathbf{u}(n) \\ y(n,k) &= \mathbf{h}^T(n,k)\mathbf{x}(n) + v(n,k), \end{aligned} \tag{12}$$

where the matrix $\mathbf{A}(n) \in \mathbb{R}^{p\times p}$, the vector $\mathbf{h}(n,k) \in \mathbb{R}^{p\times 1}$, the driving input $\mathbf{u}(n)$ is normally distributed with zero mean and variance $\mathbf{C}_u(n)$ and the observation noise $v(n,k)$ is zero-mean AWGN and independent across sensors with variance $\sigma_v^2(n,k)$. Supposing that $\mathbf{A}(n)$, $\mathbf{C}_u(n)$, $\mathbf{h}(n,k)$, and $\sigma_v^2(n,k)$ are available for all n, k, the goal of the WSN is for each sensor S_k to form an estimate of $\mathbf{x}(n)$.

Without loss of generality, we assume that sensors broadcast their data in a time division multiple access (TDMA) fashion with $k(n)$ indexing the sensor scheduled at the n th time slot; for simplicity we denote $S_{k(n)} = S(n)$. If we had infinite bandwidth available, the sensor $S(n)$ scheduled for transmission would communicate its observation $y(n,k(n)) = y(n)$ to all other sensors. Having the entire set of observations $\mathbf{y}_{0:n} := [y(0),\ldots,y(n)]^T$ available, each sensor would then be able to obtain the MMSE estimate $\hat{\mathbf{x}}(n|n) := E[\mathbf{x}(n)|\mathbf{y}_{0:n}]$ and its corresponding error covariance matrix $\mathbf{M}(n|n) := E[(\hat{\mathbf{x}}(n|n) - \mathbf{x}(n))(\hat{\mathbf{x}}(n|n) - \mathbf{x}(n))^T]$ by means of Kalman filtering (KF) iterations, each of which includes a prediction step and a correction step [24, Chap. 13].

Supposing that $\hat{\mathbf{x}}(n-1|n-1)$ and $\mathbf{M}(n-1|n-1)$ are available at time n, it follows from the linear model in (12) that the *predicted* estimate $\hat{\mathbf{x}}(n|n-1)$ and its corresponding covariance matrix $\mathbf{M}(n|n-1)$ are given by

$$\begin{aligned} \hat{\mathbf{x}}(n|n-1) &= \mathbf{A}(n)\hat{\mathbf{x}}(n-1|n-1) \\ \mathbf{M}(n|n-1) &= \mathbf{A}(n)\mathbf{M}(n-1|n-1)\mathbf{A}^T(n) + \mathbf{C}_u(n). \end{aligned} \tag{13}$$

Following this prediction step we use the innovation sequence $\tilde{y}(n) := [y(n) - \mathbf{h}^T(n)\hat{\mathbf{x}}(n|n-1)]$ to obtain the *corrected* estimate $\hat{\mathbf{x}}(n|n)$ using the well-known KF correction; see e.g., [24 Sec. 13.6]. The innovation $\tilde{y}(n)$ represents the information about the state contained in the current observation that cannot be predicted from past observations.

SIGN OF INNOVATIONS-KF

We wish to derive a distributed KF whereby observations made at each sensor are used to update state estimates at all sensors. Our goal though is to ensure that the required exchange of information among sensors entails low-communication overhead. To this end, we use as messages $m(n)$ the sign of the innovation (SOI):

$$m(n) := \operatorname{sign}[\tilde{y}(n)] = \operatorname{sign}[y(n) - \tilde{y}(n|n-1)]. \tag{14}$$

Note that quantizing $y(n)$ to the SOI $m(n)$ only alters the observation model and consequently the prediction step for the SOI-KF coincides with the prediction step for the clairvoyant KF. The sign nonlinearity, though, implies that $p[\mathbf{x}(n)|\mathbf{m}_{0:n-1}]$ is non-Gaussian and computation of the exact MMSE estimate requires (computationally expensive) numerical integrations and (memory intensive) propagation of the posterior pdf. However, based on customary simplifications made in nonlinear filtering, we can approximate the MMSE with the following correction recursions [40]:

$$\begin{aligned} \hat{\mathbf{x}}(n|n) &= \hat{\mathbf{x}}(n|n-1) + m(n)\frac{(\sqrt{2/\pi})\mathbf{M}(n|n-1)\mathbf{h}(n)}{\sqrt{\mathbf{h}(n)^T\mathbf{M}(n|n-1)\mathbf{h}(n) + \sigma_v^2}} \\ \mathbf{M}(n|n) &= \mathbf{M}(n|n-1) - \frac{(2/\pi)\mathbf{M}(n|n-1)\mathbf{h}(n)\mathbf{h}(n)^T\mathbf{M}(n|n-1)}{\mathbf{h}(n)^T\mathbf{M}(n|n-1)\mathbf{h}(n) + \sigma_v^2}. \end{aligned} \tag{15}$$

Even at a minimal communication cost, the SOI-KF is strikingly similar to the clairvoyant KF. The covariance updates in particular are identical except for the $2/\pi$ factor in (15). We emphasize that (15) is not the result of proposing a KF-like recursion based on a priori heuristics. On the contrary, the SOI-KF implements MMSE estimation based on the SOI in (14) whose form ends up being, a posteriori, reminiscent of the KF.

While the MSE corrections of the KF and SOI-KF are similar, the estimate updates for $\hat{\mathbf{x}}(n|n)$ appear to be quite different. However, it is possible to express the SOI-KF corrector in (15) in a form that exemplifies its link with the KF corrector. Indeed, if we define the SOI-KF innovation sequence as $\tilde{m}(n|n-1) := \sqrt{(2/\pi)E[\tilde{y}^2(n|n-1)]}m(n)$, it is not difficult to show that the SOI-EKF correction can be written as

$$\begin{aligned} \hat{\mathbf{x}}(n|n) &= \mathbf{x}(n|n-1) \\ &\quad + \frac{\mathbf{M}(n|n-1)\mathbf{h}(n)}{\mathbf{h}^T(n)\mathbf{M}(n|n-1)\mathbf{h}(n) + \sigma_v^2(n)}\,\tilde{m}(n|n-1), \end{aligned} \tag{16}$$

which is identical to the KF update if we replace $\tilde{m}(n|n-1)$ with the innovation $\tilde{y}(n|n-1) = y(n) - \hat{y}(n|n-1)$. Moreover, note that the units of $\tilde{m}(n|n-1)$ and $\tilde{y}(n|n-1)$ are the same, and that $E[\tilde{m}(n|n-1)] = E[\tilde{y}(n|n-1)] = 0$. Even more interesting, by definition it holds that $E[\tilde{m}^2(n|n-1)] = (2/\pi)E[\tilde{y}^2(n|n-1)]$, which explains the relationship between the covariance corrections for the KF and for the SOI-KF in

(15). The difference between the SOI-KF and KF corrections is that in the SOI-KF the magnitude of the correction at each step is determined by the magnitude of $E[\tilde{m}^2(n|n-1)]$, and it is the same regardless of how large or small the actual innovation $\tilde{m}(n|n-1)$ is.

SOI-KF IMPLEMENTATION AND MSE PERFORMANCE

Implementation of the SOI-KF requires running two separate algorithms. The observation-transmission algorithm is run by the sensors as dictated by the scheduling algorithm and starts by collecting the observation $y(n,k) \leftrightarrow y(n)$. The sensor then computes the state and observation predictions $\hat{\mathrm{x}}(n|n-1)$ and $y(n|n-1) = \mathrm{h}^T(n)\hat{\mathrm{x}}(n|n-1)$. Based on $y(n|n-1) = \mathrm{h}^T(n)\hat{\mathrm{x}}(n|n-1)$, it obtains the SOI as in (14) and broadcasts it to all other sensors as the message $m(n)$. The reception-estimation algorithm is continuously run by all sensors to track $\mathrm{x}(n)$ and is identical to a KF algorithm except for the (minor) differences in the update equations. At each time slot, the state prediction is computed using (13) and after receiving the SOI $m(n)$ the corrected estimate is obtained using (15).

MSE performance of the SOI-KF can be related with that of the KF by defining an equivalent system that is identical to the model in (12) except that the observation noise power at time n is $(\pi/2)\sigma_v^2(n)$. It turns out that the steady-state MSE of the clairvoyant KF run on this equivalent system basically coincides with the steady-state MSE of a SOI-KF run on the original system [40]. In other words, the MSE increase when using the SOI-KF is as much as the KF would incur when applied to a model with $\pi/2$ higher observation noise variance.

While we presented SOI-KF for scalar observations, generalizations are available to vector observations and colored noise after prewhitening [40].

TARGET TRACKING WITH SOI-EKF

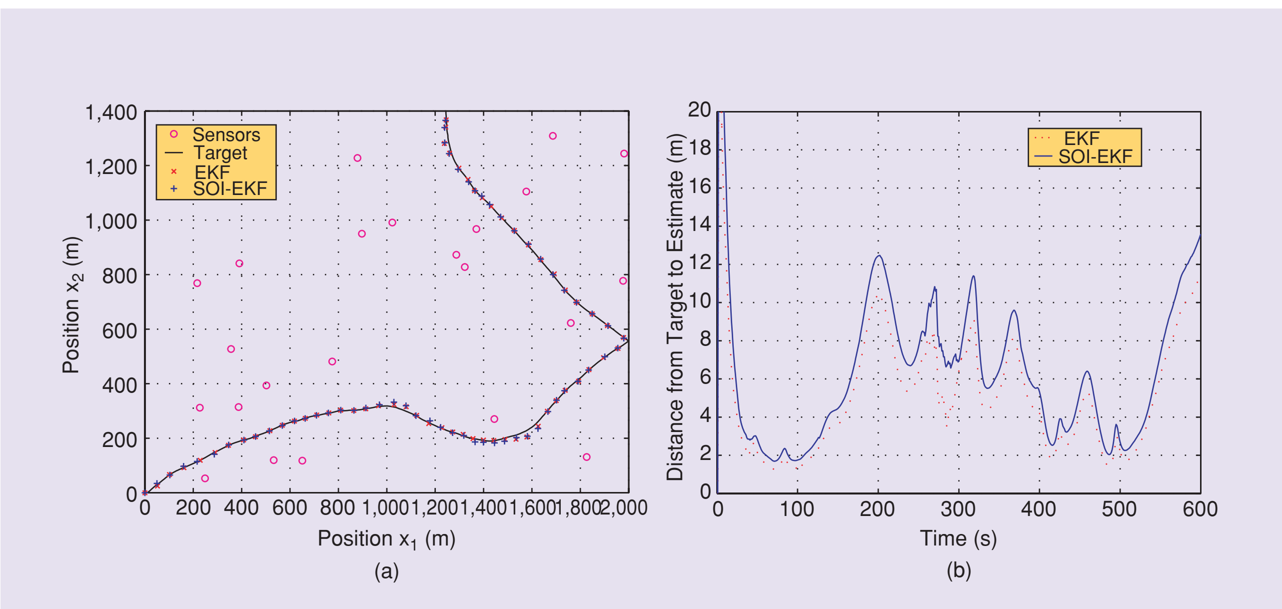
Target tracking based on distance-only measurements is a typical problem in bandwidth-constrained distributed estimation with WSNs (see e.g., [15] and [16]) for which an extended SOI-KF to nonlinear models appears to be particularly attractive. Consider K sensors randomly and uniformly deployed in a square region of $2L \times 2L$ meters and suppose that sensor positions $\{\mathrm{x}^k\}_{k=1}^K$ are known.

The WSN is deployed to track the position $\mathrm{x}(n) := [x_1(n), x_2(n)]^T$ of a target, whose state model accounts for $\mathrm{x}(n)$ and the velocity $\mathrm{v}(n) := [v_1(n), v_2(n)]^T$ but not for the acceleration that is modeled as a random quantity. Under these assumptions, we obtain the state equation [22]

$$\begin{aligned} \mathrm{x}(n) &= \mathrm{x}(n-1) + T_s\mathrm{v}(n-1) + (T_s^2/2)\mathrm{u}(n) \\ \mathrm{v}(n) &= \mathrm{v}(n-1) + T_s\mathrm{u}(n), \end{aligned} \tag{17}$$

where T_s is the sampling period and the random vector $\mathrm{u}(n) \in \mathbb{R}^2$ is zero-mean white Gaussian; i.e., $p(\mathrm{u}(n)) = \mathcal{N}(\mathrm{u}(n); \mathbf{0}; \sigma_u^2\mathbf{I})$. The sensors gather information about their distance to the target by measuring the received power of a pilot signal following the path-loss model $y_k(n) = \alpha \log \|\mathrm{x}(n) - \mathrm{x}^k\| + v(n)$ with $\alpha \geq 2$ a constant, $\|\mathrm{x}(n) - \mathrm{x}^k\|$ denoting the distance between the target and S_k, and $v(n)$ the observation noise with pdf $p(v(n)) = \mathcal{N}(v(n); 0; \sigma_v^2)$.

Mimicking an extended (E)KF approach, we linearize this observation model in a neighborhood of $\hat{\mathrm{x}}(n|n-1)$ to obtain an approximate observation model that along with the state evolution in (17) is of the form (12). We can now use the SOI-KF to track the target's position $\mathrm{x}(n)$, which offers a version of EKF with low communication cost. The results of simulating this tracker (that we abbreviate as SOI-EKF) are depicted in Figure



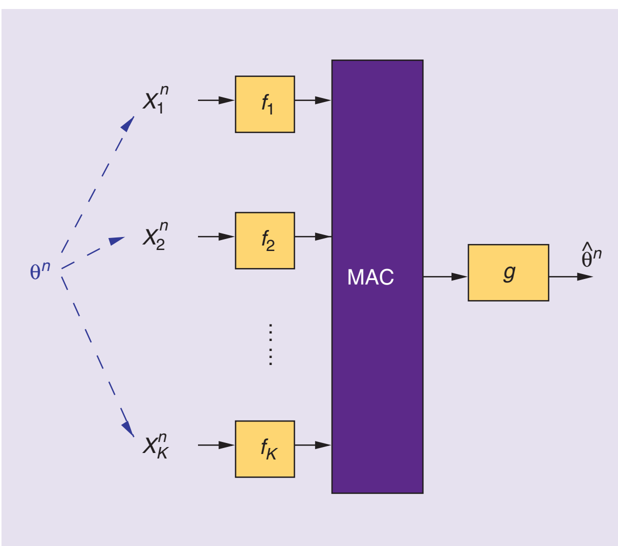
[FIG10] Target tracking with EKF and SOI-EKF.

10, where we see that the SOI-EKF succeeds in tracking the target with position errors less than 10 m. While this accuracy is just a result of the specific experiment, the important point here is that the clairvoyant EKF and the SOI-EKF yield almost identical performance even when the former relies on analog-amplitude observations and the SOI-EKF on the transmission of a single bit per sensor.

INFORMATION THEORETIC PERSPECTIVES

In this section, we study the WSN illustrated in Figure 1 by approaching the source acquisition, data communication, and final fusion processes from an information theoretic point of view.

As depicted in Figure 11, the source parameter of interest is modeled by a discrete-time memoryless random process $\{\theta(t) : 1 \leq t < \infty\}$. Sensor observations are denoted by $\{X_k(t) : k = 1, 2, \ldots, K\}$, and their joint conditional distribution (given the source $\theta(t)$) is known. A general coding scheme with block length n can be described as follows. First, sensor observations $X_k^n := \{X_k(t) : 1 \leq t \leq n\}$ are encoded in a distributed fashion (f_k denotes the encoder of the kth sensor). Then a single decoder g decodes $\theta^n := \{\theta(t) : 1 \leq t \leq n\}$ based on the received information from the distributed encoders. In what follows, we will discuss two cases where the encoders are designed under either rate or cost constraints.



[FIG11] A coding scheme in a WSN with an FC.

SOURCE CODING UNDER RATE CONSTRAINTS

When rate constraints are imposed on encoded messages, we obtain a source-coding problem in which the goal is to characterize the rate-distortion region $\mathcal{R}(D)$. The latter consists of all rate-tuples $R = (R_1, R_2, \ldots, R_K)$ that allow for the reconstruction of the source θ within certain distortion level D, when the sensor observations are encoded at a rate not exceeding R_k per sensor $k \in [1, K]$. This is the so-called CEO problem that was first introduced in [5] and subsequently studied in [9], [33], and [46]. A natural source coding scheme can be described as follows. Each sensor encoder first quantizes its observation, these quantized processes are then losslessly transmitted to the decoder using the random binning scheme [11]. The decoder uses these quantized processes to form its reproductions. This leads to the Berger-Tung inner region [4], [45]. Except for inner and outer bounds on $\mathcal{R}(D)$ derived in [4] and [45], and the quadratic Gaussian case addressed in [33], the CEO problem remains open to this date.

The $\mathcal{R}(D)$ region serves as a performance benchmark for distributed estimation under bandwidth constraints. In the Gaussian quadratic CEO problem, [33] derived an asymptotic total rate distortion function of the form $D \approx \sigma^2/(2R_\Sigma)$ when both K and R_Σ are large, where R_Σ is the total rate and σ^2 is the sensor noise variance. For the special case of 1 b per sensor sample, the total communication rate $R_\Sigma = K$, and thus the best achievable MSE performance dictated by rate distortion theory is no less than $\sigma^2/(2K)$. Recall that the distributed estimators achieve an asymptotic MSE of $(\pi\sigma^2)/(2K)$ when the threshold of local message can be taken close to θ. This MSE is a factor π away from the performance limit predicted by the rate-distortion function. Also, the universal estimators in "Universal Approaches" exhibit MSE of W^2/K that has the correct asymptotic behavior with respect to the network size K. This implies that the simple distributed estimators all have the optimal scaling behavior in terms of network size K. In contrast, the information theoretic schemes suggested by [4], [33], and [45] require complete knowledge of source/observation distribution, ions as well as long block lengths to achieve the optimal MSE performance predicted by rate distortion.

In the inhomogeneous case where local sensor SNRs are not identical, specifying the optimal rate allocation minimizing the sum rate R_Σ in the rate distortion region is also an interesting problem. It is well-known that the optimal rate allocation point that attains the sum rate distortion function is not unique and is actually a polymatroid with $L!$ vertices [9]. The optimal rate allocation region can be found through optimal Gaussian test channels and is given in [9] for the case of scalar source and observations. Vector sources and vector observations have been studied in [30] and [41]. The optimal rate allocation strategy is equivalent to searching for optimal covariance matrices in AWGN test channels and can be interpreted as a distributed Karhunen-Loève transform [19] problem. The rate allocation problem with distributed Karhunen-Loève transform is nonconvex, but suboptimal coordinate descent algorithms for optimizing individual sensor local covariance matrix have been proposed in [19] and [42]. Recently, [52] reformulated the original problem in an equivalent convex form that is efficiently solvable by interior point methods [7]. A lower bound of the sum rate distortion function of the Gaussian multiterminal source coding has also been proposed in [50] by considering the joint compression of correlated Gaussian sources under individual distortion criteria.

Figure 12 plots three rate distortion curves: i) a lower bound of the sum rate distortion function assuming full cooperation among the encoders; ii) an upper bound of the sum rate distortion function using the so-called EC (estimation-first compression-second) scheme introduced in [42]; and iii) the optimal sum rate distortion function that is calculated by solving the convex problem formulated in [52] with the numerical MAXDET routine [47]. The detail of the simulation setup is referred to [52].

ON THE OPTIMAL COST-DISTORTION TRADEOFF

When channels from sensors to the FC are noisy, we need to introduce cost constraints on the transmitted symbols from each individual sensor. Such constraints may include power constraints as a special case. In this way, it is possible to cast distributed estimation as a source-channel communication problem. The fundamental objective of the latter is to determine the optimal tradeoff between cost and distortion in an information theoretic sense regardless of complexity and delay.

In several important cases, source coding and channel coding can be separated without performance loss. For example, in a point-to-point link, source and channel coding can be performed separately without performance degradation if the source and channel are both discrete and memoryless [44, Theorem 21]. This source-channel separation theorem is quite appealing from a practical standpoint since it implies that source coding can be performed without channel knowledge and similarly for channel encoding. Unfortunately, the separation theorem does not extend to general links [11, Chapter 14]. An interesting counterexample can be found in [12] for lossless transmission of correlated sources through an interfering (nonorthogonal) multiple access channel. In this case, separating source from channel coding is suboptimal (see also [20]).

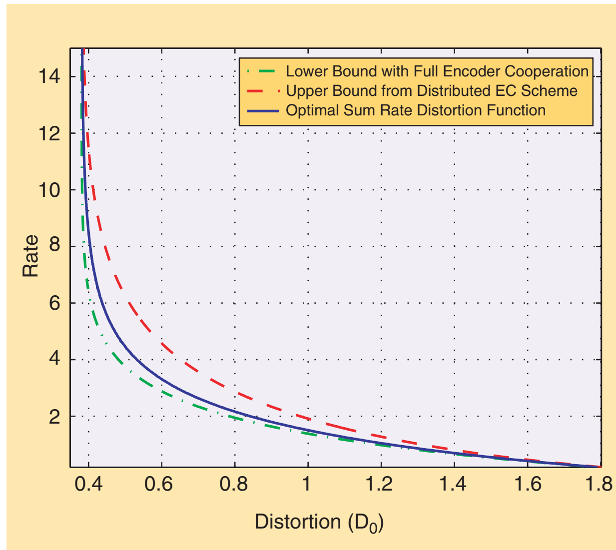
However, for the sensor network in Figure 1, if the intersensor interference is resolved by reservation-based orthogonal protocols (e.g., TDMA or FDMA) and local sensors have noninterfering channels to the FC, it turns out that the optimal tradeoff between cost and distortion can be achieved by separate source and channel coding [49]. Proving the separation theorem in this case entails a *multiple-letter characterization* of the rate-distortion region. By combining this multiletter representation of $\mathcal{R}(D)$ with a continuity property under orthogonal multiaccess, a cost distortion pair (Γ, D) is achievable if and only if

$$\mathcal{C}(\Gamma) \cap \mathcal{R}(D) \neq \emptyset,$$

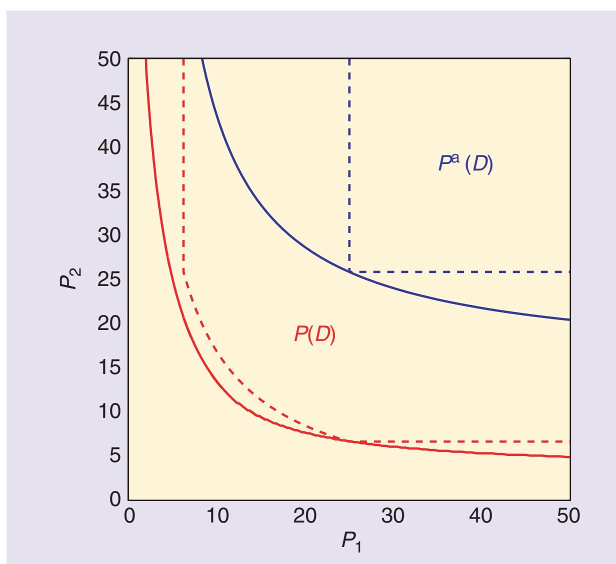
where $\mathcal{C}(\Gamma)$ is the multiaccess channel capacity. This work extends the results of [3] and [23] for the lossless transmission of correlated sources from finite alphabets through an orthogonal multiple access channel to the rate distortion case for the WSN in Figure 1.

EXAMPLE (GAUSSIAN SENSOR NETWORKS)

For the special case of estimating a Gaussian source using MSE as distortion measure, we can compare the power distortion region achieved by the separation principle with those achieved by joint source-channel coding strategies. In particular, it is shown in [8] that when each sensor has a fixed power budget and sensors are accessing the channel synchronously, the distortion achieved by optimal separate source and channel coding decreases at a rate $1/\log K$, while for a simple "analog" uncoded transmissions, the MSE decreases like $1/K$, which is much faster. However, since the separation theorem holds with orthogonal access, the optimal tradeoff between total sensor power and overall



[FIG12] The optimal sum rate distortion function compared to an EC upper bound and a full-encoder-cooperation lower bound.



[FIG13] Comparison of power-distortion region achieved by the separate source and channel coding and the uncoded transmission (denoted by $\mathcal{P}(D)$ and $\mathcal{P}^a(D)$) respectively) when the total number of sensors $K = 2$. The regions with dashed boundaries correspond to one specific pair of Gaussian test channels with noise variances $\delta_1^2 = 0.53$, $\delta_2^2 = 0.38$. Exhausting all feasible pairs of (δ_1^2, δ_2^2) gives the complete regions.

distortion is achieved by separate source and channel coding. As a result, the "digital" strategy outperforms the "analog" uncoded transmission strategy. This result should be contrasted to the case of nonorthogonal multiple access for which the "analog" uncoded transmission strategy is known to significantly outperform the digital approach of separate source and channel coding. An example of the achieved power distortion regions for the separation principle and uncoded analog transmission is depicted in Figure 13.

[OPTIMAL DISTRIBUTED ESTIMATORS DEPEND ON THE PARAMETRIC MODEL AND THE NOISE PDFS.]

CLOSING REMARKS

This article provided an overview of distributed estimation-compression problems encountered with WSNs. A general formulation of distributed compression-estimation under rate constraints was introduced, pertinent signal processing algorithms were developed, and emerging tradeoffs were delineated from an information theoretic perspective. Specifically, we designed rate-constrained distributed estimators for various signal models with variable knowledge of the underlying data distributions. We proved theoretically, and corroborated with examples, that when the noise distributions are either completely known, partially known or completely unknown, distributed estimation is possible with minimal bandwidth requirements which can achieve the same order of MSE performance as the corresponding centralized clairvoyant estimators. A distributed state estimation problem in the context of WSN has also been considered when there is prior information about the parameter of interest using the sign of innovations. For WSNs operating in inhomogeneous environments, we presented resource allocation and sensor scheduling algorithms that can result in considerable cost savings and MSE improvement.

We have not considered the interaction of routing with our distributed compression-estimation framework. This and further cross-layer optimized protocols accounting for all layers in the stack is worth further investigation and is expected to improve the overall design of distributed estimators using WSNs.

AUTHORS

Jin-Jun Xiao (xiao@ece.umn.edu) received the B.Sc. degree in applied mathematics from Jilin University, Changchun, China, in 1997 and the M.Sc. degree in mathematics from the University of Minnesota, Twin Cities, in 2003, where he is currently pursuing the Ph.D. degree in electrical engineering. In summer 2002, he worked on barcode signal processing at Symbol Technologies, Holtsville, New York. His current research interests are in distributed signal processing, multiuser information theory, and their application to wireless sensor networks. He is a Student Member of the IEEE.

Alejandro Ribeiro (aribeiro@ece.umn.edu) received his B.Sc. degree in electrical engineering from Universidad de la Republica Oriental del Uruguay, Montevideo, in 1998. He is currently working towards his Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, where he received his M.Sc. degree in electrical engineering in 2005. From 1998 to 2003 he was a member of the technical staff at Bellsouth Montevideo. 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, and distributed signal processing. He is a Fulbright Scholar.

Zhi-Quan (Tom) Luo (luozq@ece.umn.edu) received the B.Sc. degree in mathematics from Peking University, China, in 1984 and a Ph.D. degree in operations research from the Massachusetts Institute of Technology in 1989. In 1989, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, where he became a professor in 1998 and held the Canada Research Chair in Information Processing since 2001. Since 2003, he has been a professor with the Department of Electrical and Computer Engineering at the University of Minnesota and holds an ADC Chair in Digital Technology. His research interests lie in the union of large-scale optimization, signal processing, data communications, and information theory. He is a member of SIAM and MPS. He is an associate editor for several international journals including *SIAM Journal on Optimization*, *Mathematical Programming, Mathematics of Computation*, and *Mathematics of Operations Research*. He is a Senior Member of the IEEE.

Georgios B. Giannakis (georgios@ece.umn.edu) received his B.Sc. in 1981 from the National Technical University of Athens, Greece and his M.Sc. and Ph.D. in electrical engineering in 1983 and 1986 from the University of Southern California. Since 1999 he has been a professor with the Department of Electrical and Computer Engineering at the University of Minnesota, where he holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, subjects on which he has published more than 250 journal papers, 400 conference papers, and two edited books. Current research focuses on complex-field and space-time coding, multicarrier, ultrawide band radios, cross-layer designs and wireless sensor networks. He is the (co-) recipient of six best paper awards from the IEEE Signal Processing and Communications Societies and also received the IEEE Signal Processing Society's and EURASIP's Technical Achievement Awards in 2000 and 2005. He is a Fellow of the IEEE.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, pp. 393–422, Mar. 2002.

[2] J.N. Al-Karaki and A.E. Kamal, "Routing techniques in wireless sensor networks: A survey," *IEEE Trans. Wireless Commun.,* vol. 11, no. 2, pp. 38–45, Mar.–Apr. 2004.

[3] J. Barros and S.D. Servetto, "Network information flow with correlated sources," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 155–170, Jan. 2006.

[4] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications,* vol. 229 (CISM Courses and Lectures), G. Longo, Ed. New York: Springer-Verlag, 1978, pp. 171–231.

[5] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

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

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[8] M. Cardei and J. Wu, "Energy-efficient coverage problems in wireless ad hoc sensor networks," *Comput. Commun. J.*, vol. 29, no. 4, pp. 413–420, Feb. 2006.

[9] J. Chen, X. Zhang, T. Berger, and S.B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Select Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.

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

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

[12] T.M. Cover, A.A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

[13] S. Cui, "Cross-layer optimization in energy-constrained networks," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., CA, 2005.

[14] S. Cui, A.J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimizatoin," *IEEE Wireless Commun*., vol. 4, no. 5, pp. 2349–2360, Sept. 2005.

[15] P. Djuric, J. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. Bugallo, and J. Miguez, "Particle filtering," *IEEE Trans. Signal Processing,* vol. 20, pp. 19–38, Sept. 2003.

[16] P. Djuric, M. Vemula, and M. Bugallo, "Tracking with particle filtering in tertiary wireless sensor networks," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, Philadelphia, Mar. 2005, vol. 4, pp. 757–760.

[17] S. Faisal, A. Ribeiro, and G.B. Giannakis, "Bandwidth-constrained MAP estimation for wireless sensor networks," in *Proc. 39th Asilomar Conf. Signals, Systems, Computers*, Oct.–Nov. 2005, pp. 215–219.

[18] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science,* vol. 2634, pp. 162–177. New York: Springer-Verlag, Apr. 2003.

[19] M. Gastpar, P.L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loéve transform," *IEEE Trans. Inform. Theory*, to be published.

[20] M. Gastpar, M. Vetterli, and P.L. Dragotti, "Sensing reality and communicating bits: A dangerous liaison," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 70–83, 2006.

[21] A.J. Goldsmith and S.B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *IEEE Wireless Commun. Mag.,* vol. 9, no. 4, pp. 8–27, Aug. 2002.

[22] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Processing,* vol. 50, pp. 425–437, Feb. 2002.

[23] T.S. Han, "Cover-Slepian-Wolf theorem for a network of channels," *Inform. Control,* vol. 47, no. 1, pp 67–83, 1980.

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

[25] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Trans. Parallel Distrib. Syst.,* vol. 13, pp. 924–935, Sept. 2002.

[26] X. Luo and G.B. Giannakis, "Energy-constrained optimal quantization for wireless sensor networks," *IEEE 1st Conf. Sensor Ad Hoc Commun. Networks*, Santa Clara, California, Oct. 2004, pp. 272–278.

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

[28] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Select. Areas Commun.,* vol. 23, no. 4, pp. 735–744, Apr. 2005.

[29] Z.-Q. Luo and J.-J. Xiao, "Universal decentralized estimation in a bandwidth constrained sensor network," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, Philadelphia, PA, Mar. 2005, pp. 829–832.

[30] Z.-Q. Luo and J.N. Tsitsiklis, "Data fusion with minimal communication," *IEEE Trans. Inform. Theory,* vol. 40, no. 5, pp. 1551–1563, Sept. 1994.

[31] Z.-Q. Luo, G.B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 1441–1445.

[32] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. 1st ACM Int. Workshop Wireless Sensor Networks Applications*, Atlanta, Georgia, 2002, vol. 3, pp. 88–97.

[33] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

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

[35] J.B. Predd, S.R. Kulkarni, and H.V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 56–69, 2006.

[36] J.M. Rabaey, M.J. Ammer, J.L. da Silva, D. Patel, and S. Roundy, "PicoRadio supports ad hoc ultra-low power wireless networking," *IEEE Computer*, vol. 33, no. 7, pp. 42–48, July 2000.

[37] 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, no. 2, pp. 554–568, Apr. 1996.

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

[39] A. Ribeiro and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, part II: unknown pdf," *IEEE Trans. Signal Processing*, to be published.

[40] A. Ribeiro, G.B. Giannakis, and S. Roumeliotis, "SOI-KF: distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Trans. Signal Processing*, to be published.

[41] I.D. Schizas, G.B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Oct.–Nov. 2005, pp. 876–885.

[42] I.D. Schizas, G.B. Giannakis, and N. Jindal, "Distortion-Rate analysis for distributed estimation with wireless sensor networks," in *Proc. 43th Allerton Conf.*, Sept. 2005, pp. 1029–1033.

[43] I.D. Schizas, G.B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," submitted for publication.

[44] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July and Oct. 1948.

[45] S.Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School Electric. Eng., Cornell Univ., Ithaca, NY, May 1978.

[46] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1549–1559, Sept. 1997.

[47] S.-P. Wu, L. Vandenberghe, and S. Boyd, "MAXDET: Software for determinant maximization problems" [Online]. Available://www.stanford.edu/~boyd /MAXDET.html

[48] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth constrained sensor network,"*IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2617–2624, Aug. 2005.

[49] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channel communication under orthogonal multiple access," in *Proc. 43th Allerton Conf.*, Sept. 2005 pp. 1247–1256.

[50] J.-J. Xiao and Z.-Q. Luo, "Compression of correlated Gaussian sources under individual distortion criteria," in *Proc. 43th Allerton Conf.*, Monticello, Illinois, Sept. 2005 pp. 438–447.

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

[52] J.-J. Xiao and Z.-Q. Luo, "Optimal rate allocation for the vector Gaussian CEO problem," in *Proc. 1st IEEE Int. Workshop Computational Advances Multisensor Adaptive Processing*, Puerto Vallarta, Mexico, Dec. 2005, pp. 56–59.

[53] J.-J. Xiao, Z.-Q. Luo, and G.B. Giannakis, "Performance bounds for the rate-constrained universal decentralized estimators in sensor networks," in *Proc. IEEE Workshop Signal Processing Advances Wireless Communications*, Jun. 2005, pp. 126–130.

[54] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A.J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 413–422, Feb. 2006.

[55] J.-J. Xiao, Z.-Q. Luo, S. Cui, and A.J. Goldsmith, "Linear coherent decentralized estimation," submitted for publication.

[56] K. Zhang, X.R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion part VI: Sensor data compression," in *Proc. Int. Conf. Information Fusion*, Queensland, Australia, 2003, pp. 221–228.

SP