Matched Source-Channel Communication for Field Estimation in Wireless Sensor Networks

Waheed Bajwa
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
Email: bajwa@cae.wisc.edu

Akbar Sayeed
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
Email: akbar@engr.wisc.edu

Robert Nowak
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
Email: nowak@engr.wisc.edu

***Abstract*— Sensing, processing and communication must be jointly optimized for efficient operation of resource-limited wireless sensor networks. We propose a novel source-channel matching approach for distributed field estimation that naturally integrates these basic operations and facilitates a unified analysis of the impact of key parameters (number of nodes, power, field complexity) on estimation accuracy. At the heart of our approach is a distributed source-channel communication architecture that matches the spatial scale of field coherence with the spatial scale of node synchronization for phase-coherent communication: the sensor field is uniformly partitioned into multiple cells and the nodes in each cell coherently communicate simple statistics of their measurements to the destination via a dedicated noisy multiple access channel (MAC). Essentially, the optimal field estimate in each cell is implicitly computed at the destination via the coherent spatial averaging inherent in the MAC, resulting in optimal power-distortion scaling with the number of nodes. In general, smoother fields demand lower per-node power but require node synchronization over larger scales for optimal estimation. In particular, optimal mean-square distortion scaling can be achieved with sub-linear power scaling. Our results also reveal a remarkable power-density tradeoff inherent in our approach: increasing the sensor density reduces the total power required to achieve a desired distortion. A direct consequence is that consistent field estimation is possible, in principle, even with vanishing total power in the limit of high sensor density.**

I. Introduction

Sensing, processing and communication are the basic operations performed by wireless sensor networks. Due to the limited nature of valuable resources (computation, power, bandwidth) it is generally agreed that these operations should be jointly optimized in order to deliver information of the highest accuracy for a given resource allocation. This goal can be viewed as a generalization of rate-distortion theory, wherein the transmission rate is replaced by a more general set of resource constraints. In practice, however, it has been difficult to find satisfactory approaches toward this goal. While much of the research to date has focused on optimizing the basic sensor network operations separately, recent results on distributed estimation or detection of a single source indicate that joint optimization through a form of source-channel matching, facilitated by limited local node cooperation, can result in dramatic power savings [1], [2], [3], [4], [5].

In this paper, we propose a distributed architecture for matched source-channel communication for optimized sensing and estimation of a homogeneous spatial field using a wireless network of sensors. The architecture is illustrated in Fig. 1. A network with n nodes is uniformly partitioned into m cells. Local sufficient statistics from each cell are coherently communicated to a distant destination using dedicated multiple access channels (MAC). Our approach unifies the operations of sensing, processing and communication into a single

This research was supported in part by the NSF under grants CCR-0113385 and CCR-0310889, and by the ONR under grant N00014-00-1-0966.

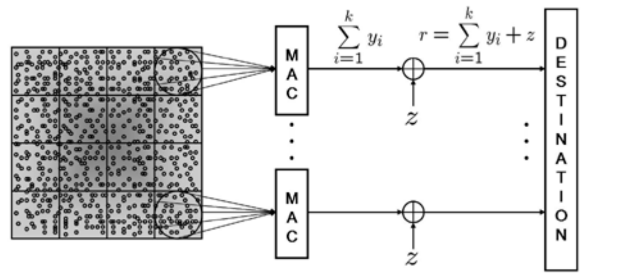

Fig. 1. A distributed architecture for homogeneous field estimation with wireless sensors. A network with n nodes is uniformly partitioned into m cells, with $k = n/m$ nodes in each cell. The size of each cell is matched to the field smoothness, and the nodes in each cell communicate coherently to the destination via a dedicated multiple access channel (MAC).

scheme that allows us to characterize the relationships and tradeoffs among these operations and reveals fundamental power-distortion scaling laws that govern the quality of the field estimate delivered to a remote destination. Three remarkable features of our scheme are: 1) processing and communications are combined into one operation; 2) it requires almost no collaboration among sensing nodes; and 3) consistent field estimation is possible as node density increases even if the total network power consumption tends to zero. Our approach involves a novel combination of existing results from approximation theory [6], statistics [7], and wireless communications [1], [2], [3], [4], [5], [8]. The decay rate of the squared L_2-distortion at the destination depends on three quantities; n, the number of wireless sensing nodes in the network, P, the communication power expended by each node, and $\alpha > 0$, a parameter that quantifies the smoothness of the underlying field. Under our optimal design, the distortion obeys the equation

$$D(n, P, \alpha) \leq C \left\{ (Pn)^{-2\alpha/d} + P \right\}. \tag{1}$$

where d is the dimension of the field and $C > 0$ is a constant independent of n and P and dependent only on measurement and communication channel noise levels and on the accuracy of Taylor approximations to f. As functions of P, the first and second terms in the expression above are strictly monotonically decreasing and increasing, respectively. Thus, the minimum distortion is achieved by the calibration $P = n^{-2\alpha/(2\alpha+d)}$. Note that in this case the distortion and the power expended by each node decrease at the same polynomial rate as the density of the network increases.

From an architectural and protocol viewpoint, our approach represents a departure from existing methodologies which emphasize the networking aspects, such as multi-hop routing of information. Our proposed wireless sensing system perhaps is more accurately viewed as a *sensor ensemble* which is appropriately queried by an *information transponder* (destination or fusion center) to elicit the desired information about the signal field. The nodes in different cells in Fig. 1 form *coherent sub-ensembles* in that they communicate the local field information in a phase-coherent fashion to the transponder. This locally coherent response from the sensor sub-ensembles simultaneously reduces the impact of the two key sources of error: measurement noise and communication noise. The resulting optimal power-distortion scaling laws reveal a fundamental *power-density tradeoff*: increasing the sensor density increases the cardinality of each coherent sub-ensemble thereby dramatically reducing the network power required to attain a target distortion.

A. An Overview of the Proposed Scheme

Here we give a short introduction to the proposed method. In the following sections we elaborate on the technical details of the scheme. To begin, define the class $C^\alpha(B)$ to consist of all real-valued functions f on $[0,1]^d$ that possess continuous partial derivatives up to order α and satisfy

$$\forall\ z, x \in [0,1]^d: \quad |f(z) - P_x(z)| \le B\|z-x\|^\alpha, \qquad (2)$$

where $B, \alpha > 0$, and $P_x(\cdot)$ denotes the Taylor polynomial of order $\alpha - 1$ for f at point x. Assume that the field being sensed is some $f \in C^\alpha(B)$. The goal of our scheme is to deliver $\widehat{f}$, a low-distortion estimate of the field f, to a remote destination. The n sensor nodes are at distinct and known locations, $\xi_1, \ldots, \xi_n$, that are roughly uniformly distributed over $[0,1]^d$. For our purposes we will assume that every square region S of volume $V \gg 1/n$ contains $n_S \asymp nV$ sensors[1] and that distance between sensors behaves like $n^{-1/d}$. For example, this condition is satisfied if the sensors are arranged on a regular lattice, but less structured sensor layouts are admissible (e.g., small bounded perturbations of a regular lattice). The condition is necessary to insure that the resulting system of estimation equations is full rank.

Each node takes a noisy sample of the form

$$x_i = f(\xi_i) + w_i \qquad (3)$$

where the errors w_i, $i = 1, \ldots, n$, are independent, zero-mean random variables with variance σ_w^2. If the destination has direct access to the measurements (noise-free communication), the well-known (minimax) optimal estimation error is

$$\inf_{\widehat{f}} \sup_{f} E\{\|f - \widehat{f}\|_{L_2}^2\} \ \asymp\ n^{-2\alpha/(2\alpha+d)}$$

in this case [9]. Moreover, a practical field estimator can be easily constructed by partitioning $[0,1]^d$ into $m = n^{d/(2\alpha+d)}$ (square) cells of volume $1/m$ and computing the least squares fit of an order $p = \alpha - 1$ polynomial to the noisy samples in each such cell [9]. This results in a bias/variance trade-off

$$E\{\|f - \widehat{f}\|_{L_2}^2\} \ \preceq\ m^{-2\alpha/d} + m/n \qquad (4)$$

where the first term is the squared bias and the second is the variance. The minimum is attained by setting $m = n^{d/(2\alpha+d)}$, which produces the minimax rate above. Finally, note that in the absence of measurement noise ($\sigma_w^2 = 0$) only the bias term appears and we can write

$$\|f - \widehat{f}\|_{L_2}^2 \ \preceq\ m^{-2\alpha/d}.$$

[1]We write $a_n \preceq b_n$ when $a_n = O(b_n)$ and $a_n \asymp b_n$ if both $a_n \preceq b_n$ and $b_n \preceq a_n$.

There are precisely $\ell = \binom{p+d}{p}$ sufficient statistics per cell. These statistics are simple linear combinations of the measurements. Therefore, the estimate can be delivered to the destination (modulo a simple linear reconstruction process to be carried out there) by transmitting the $m\ell$ sufficient statistics. A very important fact is that the sufficient statistics for each least squares fit can be computed by weighted averages of the samples, where the weighting factor applied to each sample depends only on the location of the corresponding sensor (see the Appendix for details). Thus, determination of the weighting factors requires no collaboration between nodes.

The remaining issue is the communication protocol for transporting the local sufficient statistics to a distant destination. One possibility is to nominate a local clusterhead in each cell, which receives weighted sample values from the other nodes in its cell and then computes, encodes and transmits the sufficient statistics to the receiver. Another, more promising, alternative is to exploit recent results concerning uncoded coherent transmission schemes [1], [3], [4]. The proposed distributed communication architecture, illustrated in Fig. 1, is matched locally to the estimator partition: the nodes in each cell coherently communicate to the destination via a dedicated additive white Gaussian noise (AWGN) MAC connecting the n/m nodes to the destination. This approach involves phase-coherent, low-power, analog transmissions of weighted sample values directly from the nodes in each cell to the destination. The destination receives the coherent transmissions and the required averaging is implicitly computed as a result of the spatial averaging in the MAC. Coherent transmission from each cell provides an $\frac{n}{m}$-fold *power amplification*: the total received power at the destination ($\frac{Pn^2}{m}$) is $\frac{n}{m}$ times the total transmitted power (Pn). The resulting communication error scales like $\frac{m}{\sqrt{Pn}}$, resulting in a total squared L_2 error at the destination

$$E\{\|f - \widehat{f}\|_{L_2}^2\} \ \preceq\ m^{-2\alpha/d} + \frac{m}{n} + \frac{m^2}{Pn^2},$$

where the last two terms are variances due to the measurement and communication noises, respectively. These are both increasing functions of m, so the overall minimum distortion is achieved when they are of equal size; thus we have the calibration $P = m/n$ giving us the relation in (1) above in which optimal decay rate is achieved with vanishing power per node. In the absence of measurement noise the squared L_2 error is

$$E\{\|f - \widehat{f}\|_{L_2}^2\} \ \preceq\ m^{-2\alpha/d} + \frac{m^2}{Pn^2},$$

and can decay at a faster optimal rate approaching $n^{-2\alpha/(\alpha+d)}$ if the power per node P is non-vanishing.[2]

B. Relationship to Previous Work

First, let us comment on the field model used in our approach. We assume that the field f is an unknown deterministic function with α continuous derivatives. This can be relaxed to the less restrictive assumption that f is Hölder-α smooth, for any $\alpha > 0$. In such cases, our estimator would be based on piecewise polynomial fits of order $\lceil \alpha \rceil - 1$. We assume our sensors take samples of the field contaminated by noise. An alternative perspective, that is commonly used in related work [10], [11], [12], [13], is that f is a realization of a stationary (often bandlimited) random field with some known correlation function. The decay characteristics of the correlation (e.g., the rate of decay) play a role analogous to that of α in our paper. Both field models express a notion of smoothness in the field. The choice

[2]We are implicitly assuming that the total network power cannot scale faster than linearly with the number of nodes.

between the deterministic or stochastic model is mostly a matter of taste and mathematical convenience in the case of homogeneous fields; fundamentally, both models embody the same smoothness characteristics. However, the deterministic formulation can be more readily generalized to include inhomogeneities, such as boundaries [8].

Secondly, field estimation using wireless sensor networks generally requires a combination of *in-network communications* to compute local sufficient statistics and *out-of-network communications* to transport the sufficient statistics to a usually distant destination. Most previous works in this area have focused on multihop communication schemes and in-network data processing and compression [8], [10], [11], [13]. This requires a significant level of network infrastructure, and the theoretical approaches in the works above generally assume this infrastructure as given. However, in practice the operation and maintenance of the network is often the most challenging issue and imposes a significant burden on network resources. Our new approach, in contrast to previous methods, eliminates the need for in-network communications and processing and instead requires local phase synchronization among nodes (which imposes a relatively small burden on network resources). Furthermore, coherent communication of local sufficient statistics to the destination results in dramatic power savings in out-of-network communications. For instance, if multi-hop communication protocols are used for in-network processing, as in our previous related work [8], then local sufficient statistics are transported to a cluster head in the network and subsequently transmitted to a remote destination. The final transmission out of the network requires a significant amount of power. In fact, the total power required to achieve a target distortion level at the destination would far exceed the power required by our new approach (see Section III). In our case, coherent communication from each cell requires a *monotonically decreasing* P with n in order to attain the same distortion scaling. As a result, the out-of-network power savings due to our approach grow unbounded with the sensor density (number of nodes). Thus, we conclude that our new approach results in very significant power savings compared to existing methods by eliminating the need for in-network communications and by employing locally phase-coherent transmissions for out-of-network communications.

Finally, the notion of distributed source-channel matching can be thought of as a form of joint source-channel coding in sensor networks. The importance of source-channel matching for estimation of multiple sources has also been noted in [3]. An important implication of the proposed matched source-channel communication architecture is that joint source-channel coding is inherently local: the size of each cell, n/m, defines the optimal scale of local phase synchronization across nodes.

II. Optimal Estimation in a Centralized System

We first consider the structure of the optimal centralized estimator in which the sensor measurements are available at the destination, with no added cost or noise due to communications. The distortion scaling of the centralized estimator serves as a benchmark for assessing the performance of the distributed estimation algorithms. For homogenously smooth signal fields, optimal asymptotic distortion scaling can be achieved by partitioning the signal field into m uniform cells, $Q_1, \ldots, Q_m$, each with $k = n/m$ nodes. For a $C^\alpha(B)$ field, the optimal signal field estimate, $\widehat{f}$, is a piecewise polynomial fit to the data in each cell.

To illustrate the idea, let us consider the situation when $\alpha = 1$, in which case the field estimator is constant on each cell of the partition. The value in cell j is obtained by averaging the measurements in that cell:

$$\hat{s}_j = \frac{1}{k} \sum_{i:\xi_i \in Q_j} x_i = \frac{1}{k} \sum_{i:\xi_i \in Q_j} s_i + \frac{1}{k} \sum_{i=1}^{k} w_i \tag{5}$$

where $s_i = f(\xi_i)$ and the sum is over the k measurements in cell Q_j. The resulting field estimate is given by

$$\widehat{f}(\xi) \;=\; \sum_{j=1}^{m} \hat{s}_j 1_{\xi \in Q_j} \tag{6}$$

where $1_{\xi \in Q_j}$ is 1 if ξ is in Q_j and zero otherwise. The squared bias of $\widehat{f}$ is

$$\begin{aligned} \|f - E\{\widehat{f}\}\|_{L_2}^2 \;&=\; \int (f - E\{\widehat{f}\})^2 \\ &=\; \int (f - \bar{f})^2 \;\le\; Bm^{-2/d} \end{aligned}$$

where $\bar{f}$ is the piecewise constant field obtained by averaging f on each cell of the partition and the inequality itself follows from a simple Taylor series argument. The variance of $\widehat{f}$ is proportional to the ratio of the number of its degrees of freedom, m, to the number of samples, n. Thus, we have the following upper bound on the distortion achievable with a centralized approach:

$$D_{cen} \;\equiv\; E\{\|f - \widehat{f}\|_{L_2}^2\} \;\preceq\; m^{-2/d} + m/n \; .$$

In the general case, the squared bias behaves like $m^{-2\alpha/d}$ (which follows from a higher order Taylor series remainder) and the variance remains proportional to m/n, leading to expression (4); see the Appendix for details on the construction of the estimator. While the expressions above ignore constant factors, and are helpful for studying scaling behaviors, the SNR is an important constant to bear in mind in practice. The measurement SNR is $\rho_{\text{meas}} = \sigma_s^2/\sigma_w^2$, where σ_s is the field signal strength (average amplitude) and σ_w is the standard deviation of the measurement noise.

$$D_{cen} \;\preceq\; \frac{1}{m^{2\alpha/d}} + \frac{1}{\rho_{\text{meas}}} \frac{m}{n} \; . \tag{7}$$

For optimal distortion scaling, the bias and variance should be reduced at the same rate

$$\frac{1}{m^{2\alpha/d}} = \frac{1}{\rho_{\text{meas}}} \frac{m}{n} \;\Longleftrightarrow\; m_{\text{opt}} = \rho_{\text{meas}}^{\frac{d}{2\alpha+d}} n^{\frac{d}{2\alpha+d}} \tag{8}$$

Thus, for optimal distortion scaling, $m \asymp n^{\frac{d}{2\alpha+d}}$ and the constant is proportional to the measurement SNR. The resulting distortion scales as

$$\sup_{f \in C^\alpha(B)} E\{\|f - \widehat{f}\|_{L_2}^2\} \;\asymp\; n^{-\frac{2\alpha}{2\alpha+d}} \tag{9}$$

since our upper bound matches the minimax lower bound.

III. Distributed Estimation via Noisy Communications

We now analyze the distortion when the local sufficient statistics from each cell are communicated to the destination over a noisy channel. Our goal is to characterize the power-distortion scaling laws that relate the final distortion to the transmit power as a function of the number of nodes. We again illustrate the key ideas for $\alpha = 1$ for which a constant signal estimate suffices in each cell. Details on the general case are provided in the Appendix. The MAC connecting each cell to the destination in Fig. 1 forms the basic building block of the overall architecture; we assume that the MACs for different cells are non-interfering, which may be achieved via time-division or frequency-division multiplexing over the available bandwidth. The sensor measurements in cell j are given by $x_{i,j} = s_{i,j} + w_{i,j}$, $i = 1, \cdots, k$. We will focus our attention on an arbitrary cell for the moment, and thus we suppress the subscript j in the following analysis. Each node communicates an amplified version of its local measurement to the destination

$$y_i = \sqrt{P} x_i \tag{10}$$

where $P = P_o/(\sigma_s^2 + \sigma_w^2)$ and P_o is the transmit power per node.[3] The signal received at the destination is given by

$$\begin{aligned} r &= \sum_{i=1}^{k} y_i + z = \sqrt{P} \sum_{i=1}^{k} x_i + z \\ &= \sqrt{P} \sum_{i=1}^{k} s_i + \sqrt{P} \sum_{i=1}^{k} w_i + z \end{aligned} \tag{11}$$

where $z \sim N(0, \sigma_z^2)$ is the AWGN in the MAC. Strictly speaking, the received signal from each node, y_i, in (11) should be scaled with an attenuation constant, $a_i \in (0,1)$, that depends on the distance between the node and destination and the path loss exponent. However, for simplicity we are assuming that the destination is far enough so that all the distances, and hence the a_i's, are nearly the same. We ignore this uniform attenuation since it will uniformly increase the required power per node by a constant factor to attain a desired distortion. When the attenuations for different nodes are significantly different, it would require non-uniform power allocation across nodes, and a detailed discussion of this issue will be reported elsewhere.

Inspired by the structure of the centralized estimator in (5), the estimate at the destination is formed from the received signal r as

$$\hat{s}_{des} = \frac{1}{k\sqrt{P}} r = \hat{s}_{cen} + \frac{z}{\sqrt{P} k} \tag{12}$$

where $\hat{s}_{cen}$ denotes the centralized estimate in (5), and the corresponding distortion is given by

$$\begin{aligned} D_{des} &= D_{cen} + \frac{\sigma_z^2}{Pk^2} = D_{cen} + D_{com} \\ &\preceq \frac{1}{m^{2/d}} + \frac{1}{\rho_{\text{meas}} k} + \frac{1}{\rho_{\text{com}} P k^2} \end{aligned} \tag{13}$$

where $\rho_{\text{com}} = \sigma_s^2/\sigma_z^2$ is the communication[4] SNR, the first two terms correspond to D_{cen} and the last term corresponds to D_{com}, the distortion introduced due to noisy MAC communication. If we consider the general $C^\alpha(B)$ setting, then only the bias term is effected in the expression above, resulting in

$$D_{des} \preceq \frac{1}{m^{2\alpha/d}} + \frac{1}{\rho_{\text{meas}} k} + \frac{1}{\rho_{\text{com}} P k^2}$$

The above relation governs the interplay between D, P, n and α. The $1/k^2$ factor in D_{com} is due to phase-coherent transmission from each node in the cell (coherent beamforming) that results in a k-fold power amplification: the total received power Pk^2 is k times the total transmit power Pk.

Measurement-Limited Regime. For fastest distortion reduction with n, all three terms in (13) must scale at the same rate. When there is significant measurement noise, the distortion scaling rate is limited by the variance term in D_{cen}. Thus, for optimal distortion scaling P must scale as

$$P_{opt} \propto \frac{1}{k} = \frac{n}{m} , \tag{14}$$

resulting in $P_{opt} \asymp D_{des} \asymp n^{-\frac{2\alpha}{2\alpha+d}}$, when $k = n/m$ is selected to balance the estimation bias and variance terms. By equating the three terms in (13), with the first two calibrated for optimal centralized scaling via (8), the optimal power scaling should behave as

$$P_{opt} = \frac{\rho_{\text{meas}}^{\frac{2\alpha+2d}{2\alpha+d}}}{\rho_{\text{com}}} n^{-\frac{2\alpha}{2\alpha+d}} . \tag{15}$$

The constant monotonically increases in ρ_{meas} and decreases in ρ_{com} which makes intuitive sense: high SNR measurements offer higher accuracy and it requires more power to deliver this accuracy, and lower communication SNR requires more power allocation. Thus, in the measurement limited case, optimal distortion is achieved when the total network power, nP_{opt}, scales sub-linearly with n; $P_{opt} \downarrow 0$ as $n \uparrow \infty$. Essentially, the k-fold coherent beamforming gain cannot be fully exploited due to measurement noise and, thus, optimal distortion scaling is achieved with vanishing power per node.

Communication-Limited Regime. Now suppose that we have noise-free measurements; $\sigma_w^2 = 0$. In this case, the variance term in D_{cen} in (13) disappears and the distortion can be reduced at a faster $1/k^2$ rate with a constant, non-vanishing power per node $P > 0$. The optimal calibration for m is

$$m = (\rho_{\text{com}} P)^{\frac{d}{2\alpha+2d}} n^{\frac{d}{\alpha+d}} \tag{16}$$

and the corresponding optimal distortion scaling is

$$D_{des} \propto 1/k^2 \asymp n^{-\frac{2\alpha}{\alpha+d}} . \tag{17}$$

In this case, the distortion scaling is limited by the communication noise, and a faster distortion scaling ($1/k^2$ versus $1/k$) can be achieved compared to the measurement-limited case. The faster distortion reduction is attained by fully exploiting the coherent MAC power amplification and allowing the total network power to scale linearly with n.

Bias-Limited Regime. The above distortion scaling in the communication-limited regime is optimal under the constraint that the power per node, P, cannot increase with n. If we place no constraints on power and assume noise-free measurements, then choosing $m = n$ and $P \asymp n^{2\alpha/d}$ (super-linear network power scaling) results in the optimal centralized bias-limited distortion scaling $D_{des} \asymp n^{-2\alpha/d}$.

A. Impact of Imperfect Phase Synchronization

The above analysis assumes perfect phase synchronization within each cell. The impact of imperfect phase synchronization can be incorporated by assuming a beamforming gain of k^β, $\beta \in [0,1]$, where $\beta = 0$ corresponds to incoherent communication (no beamforming gain) and $\beta = 1$ corresponds to perfectly phase-coherent communication (maximum beamforming gain of k). The corresponding distortion due to communication noise scales as $D_{com} \asymp \frac{1}{Pk^{1+\beta}}$ and the power allocation required to achieve the optimal (centralized) distortion scaling in the measurement-limited regime is

$$P = \frac{\rho_{\text{meas}}^{\frac{2\alpha+d(1+\beta)}{2\alpha+d}}}{\rho_{\text{com}}} n^{-\frac{2\alpha\beta}{2\alpha+d}} \tag{18}$$

which, for $\beta < 1$, is higher than (15) corresponding to $\beta = 1$. In particular, $\beta = 0$ corresponds to the situation in which a clusterhead

[3] For simplicity of notation, we treat P as the per-node power, ignoring the scaling factor since it does not impact the scaling behavior.

[4] The communication SNR is measured relative to the strength of the sensed signal. For each cell, the actual transmit communication SNR is $\rho_{\text{com}} Pk$ and the received communication SNR is $\rho_{\text{com}} Pk^2$.

in each cell communicates the local sufficient statistics to the destination using the total transmit power Pk available for each cell. There is no beamforming gain in this case since only the clusterhead transmits the local statistics from the cell rather than all nodes in the cell transmitting it in a phase-coherent fashion. As a result, in this case a constant non-vanishing power $P = \rho_{\text{meas}}^{\frac{2\alpha+d(1+\beta)}{2\alpha+d}} / \rho_{\text{com}}$ is needed to achieve the optimal distortion scaling. Thus, the power savings due to phase-coherent transmission in our approach grow unbounded as compared to existing approaches based on in-network processing in which a clusterhead communicates the local statistics to the destination.

B. Feasible Power Scaling for Consistent Estimation

So far we have considered optimal power scaling that results in minimum (centralized) distortion scaling in the final estimate. Even if the available per-node power is less than the optimal, D_{des} can still be driven to zero, albeit at a slower, sub-optimal rate. We now characterize the minimum power allocation that guarantees that D_{des} goes to zero in the limit of large number of nodes. In this case, the partition scaling (growth rate of the number of cells) has to be adapted to the power scaling to guarantee a consistent final estimate.

Consider the measurement-limited regime first. Let $P(n) = 1/n^\gamma$ for $\gamma \geq \gamma_{opt}$, where $\gamma_{opt} = 2\alpha/(2\alpha + d)$ denotes the optimal power scaling, and let $m = n^\delta$ for $\delta \in (0, 1)$. Ignoring constants, the distortion at the destination scales as

$$D_{des} \preceq \frac{1}{n^{2\alpha\delta/d}} + \frac{1}{n^{1-\delta}} + \frac{1}{n^{2(1-\delta)-\gamma}}. \quad (19)$$

Note that choosing $\gamma = \gamma_{opt}$ and $\delta = \delta_{opt} = d/(2\alpha + d)$ results in $\gamma_{opt} = (1 - \delta_{opt})$ and all three terms in (19) decay at the same optimal rate. If the partition scaling is kept fixed at $\delta = \delta_{opt}$, then it follows from (19) that as long as $\gamma < 2(1 - \delta_{opt}) = 2\gamma_{opt}$ the distortion goes to zero at the rate

$$D_{des} \asymp D_{com} \asymp n^{\gamma - 2\gamma_{opt}} \asymp n^{\gamma - 4\alpha/(2\alpha+d)} . \quad (20)$$

On the other hand, for a given $\gamma > \gamma_{opt} = 1 - \delta_{opt}$, the distortion scaling will be limited by the last term, D_{com}, in (19) and the partition (i.e., choice of δ) can be adapted to guarantee that the slower of the first two terms decays at least as fast as the last term: $\min(2\alpha\delta/d, 1-\delta) \geq 2(1-\delta)-\gamma$. It is clear that the second (variance) term always decays faster than the last term (when $\gamma > \gamma_{opt}$) so δ should be chosen to match the first (bias) and third (com) terms:

$$\delta = \frac{(2 - \gamma)d}{2(\alpha + d)} . \quad (21)$$

With such matching, as long as $\gamma \in [\gamma_{opt}, 2)$, distortion goes to zero at the rate

$$D_{des} \asymp D_{com} \asymp n^{-\alpha(2-\gamma)/(\alpha+d)} \quad (22)$$

in the measurement-limited regime. Note that any $\gamma > \gamma_{opt}$ corresponds to sub-optimal (less than optimal) power per node and the resulting D_{des} goes to zero at the above slower rate. In particular, D_{des} can be driven to zero asymptotically even if the per-node power $P(n)$ decays just a little slower than $1/n^2$ ($\gamma = 2$; cut-off power allocation). This is remarkable since it shows that, in principle, consistent field estimation is possible in the limit of a large number of nodes even if the total network power $P_{tot} = nP(n)$ goes to zero! This is due to the fact that as $\gamma \to 2$, $\delta \to 0$ in (21) and hence $k = n^{1-\delta} \to n$ yielding the highest possible coherent beamforming gain.

Clearly, $\gamma < \gamma_{opt}$ is wasteful in the measurement-limited regime since in this case D_{com} will decay faster than the dominating variance term in (19). However, such higher power allocation is beneficial in the communication-limited and bias-limited regimes in which the variance (second) term disappears. In the absence of measurement noise, for any $\gamma \in [-2\alpha/d, 2)$, δ can be chosen to match the bias and communication terms as in (21) and the resulting distortion scales as (22). In particular, $\gamma = 0$ corresponds to the optimal power allocation in the communication-limited regime (linear network power scaling), and $\gamma = -2\alpha/d$ corresponds to the optimal power allocation in the bias-limited regime (super-linear network power scaling).

C. Optimality of the Power-Distortion Scaling Laws

We now compare the performance of the proposed coherent uncoded communication scheme to that of an ideal coded communication strategy to argue the power-distortion optimality of the proposed distributed estimation architecture. By "optimal" we mean that our scheme requires the least power allocation, as a function of the number of nodes, to achieve the optimal centralized distortion scaling (i.e., the distortion achievable when the destination has direct access to the measurements). In the ideal coded strategy we assume that the nodes can fully cooperate. In this case, each node in cell j knows the optimal (centralized) estimate $\hat{s}_j$ given by (5) for the $\alpha = 1$ case. More generally, $\hat{s}_j$ is a (finite dimensional) vector of the polynomial coefficients providing a least squares fit to the measurements in cell j. For the sake of exposition, we discuss the scalar case, since extensions to the vector case are straightforward.

First of all, note that the optimal power-distortion scaling achieved under the assumption that the nodes can fully cooperate (as in the ideal coded strategy) serves as a performance benchmark for any strategy in which the nodes cannot exchange data (as in the uncoded strategy). With full node cooperation, the classical source-channel separation principle applies: each node in the cell identically encodes the optimal estimate $\hat{s}_j$ and coherently transmits the identically coded bits over the MAC, effectively transforming the distributed MAC into a classical point-to-point AWGN channel. The capacity of the effective AWGN channel is

$$C(P, k) = \frac{1}{2} \log \left(1 + \frac{Pk^2}{\sigma_z^2} \right) \text{bits/channel use} \quad (23)$$

Note that the capacity of a point-to-point AWGN channel is a monotonic function of the received SNR and coherent beamforming maximizes the received SNR, and hence capacity, for a given transmit power as in (23). The problem is now reduced to the classical rate-distortion problem in which source coding and channel coding can be done independently (in the absence of latency constraints). If the common signal component in $\hat{s}_j$ is modeled as zero-mean i.i.d. Gaussian random process (over time) with variance σ_s^2 and the measurement noise is assumed to be Gaussian then the lowest bit rate needed to encode $\hat{s}_j$ with a target distortion $D_{coded} < \sigma_s^2 + \sigma_w^2/k$ is given by the rate-distortion function [14]

$$R_o(D_{coded}) = \frac{1}{2} \log \left(\frac{\sigma_s^2 + \sigma_w^2/k}{D_{coded}} \right) \text{ bits/channel use,} \quad (24)$$

where we have implicitly assumed a single channel use for each temporal field sample. It is well-known that for any source with the same variance $R(D_{coded}) \leq R_o(D_{coded})$. Thus, solving $R_o(D_{coded}) = C(P, k)$ provides an upper bound on the achievable D_{coded} for a given power P. The final distortion at the destination is given by

$$\begin{aligned} D_{des,coded} &= D_{cen} + D_{coded} \\ &\preceq \sigma_s^2 \left(\frac{1}{m^{2/d}} + \frac{1}{\rho_{\text{meas}} k} \right) + \frac{\sigma_s^2 + \frac{\sigma_w^2}{k}}{1 + \frac{Pk^2}{\sigma_z^2}} \\ &\preceq m^{-2/d} + \frac{1}{k} + \frac{1}{Pk^2} + \frac{1}{Pk^3} \end{aligned} \tag{25}$$

which has identical scaling behavior as the uncoded scheme (see (13) and (14)) since the dominant term due to communication errors is the $1/Pk^2$ term. Thus, the ideal coded strategy achieves the optimal power-distortion scaling (25) under the assumption of full node cooperation, while uncoded coherent beamforming achieves the same optimal power-distortion scaling without data exchange between the nodes. We refer the readers to [1] for a different argument for the optimality of uncoded communication in the single-source problem (the signal in each cell in our formulation can effectively be viewed as a single source).

D. Power-Density Tradeoff

The power-distortion scaling laws achieved by our matched source-channel communication strategy reveal a conservation relation

$$PD_{des} \propto \frac{1}{k^2} \tag{26}$$

which follows from the fact that, in all regimes, the total distortion scaling equals the scaling in D_{com}. Recall the feasible power allocations for consistent estimation discussed in Section III-B. The total network power needed to achieve a given target distortion, $D_{des} = D_o$, for a given partition scaling, $m = n^{\delta}$, satisfies

$$P_{tot} = nP \propto \frac{n}{D_o k^2} = \frac{n^{2\delta - 1}}{D_o} \tag{27}$$

where n denotes the number of nodes needed to achieve D_o for the given δ. As discussed in Section III-B, a smaller δ (faster scaling of k) requires a larger number of nodes n (higher sensor density) to attain a target distortion, albeit with lower power per node P. The key question is whether choosing a smaller δ also results in a lower total power consumption P_{tot}? Let $\delta_2 < \delta_1$ denote two partition scalings and let $n_2 > n_1$ denote the corresponding number of nodes needed to attain the target distortion D_o. It follows form (27) that the total network powers consumed in the two cases are related by

$$\frac{P_{tot}(n_2)}{P_{tot}(n_1)} = \frac{n_2 k^2(n_1)}{n_1 k^2(n_2)} = \frac{n_2 n_1^{2(1-\delta_1)}}{n_1 n_2^{2(1-\delta_2)}} \propto \frac{n_1}{n_2} \tag{28}$$

where we have used the fact that $n_1^{\delta_1} \propto n_2^{\delta_2}$ since the bias terms (see (19)) should be of the same order in the two cases to yield the same target distortion. The relation (28) shows that increasing the sensor density n by a factor of N reduces the total required network power by a factor of N to attain a given target distortion. This reveals a remarkable *power-density tradeoff* inherent in our approach: *increasing the sensor density increases the cardinality of coherent sub-ensembles and reduces the total network power required to achieve a target distortion.* A direct consequence is that consistent estimation is possible, in principle, even with vanishing total network power by increasing the sensor density. This is because the beamforming gain k (the cardinality of the coherent sub-ensembles) increases monotonically with n.

IV. Discussion and Numerical Results

The essence of our matched source-channel communication architecture is to match the spatial scale of signal field coherence to the spatial scale of phase-coherent communication. The resulting distributed source-channel matching is effected locally and independently in each cell of the network partition in Fig. 1. The network partition is dictated by classical estimation-theoretic considerations that do not depend on the decentralized nature of the problem. Furthermore, the growth rate of the number of cells, $m \asymp n^{d/(2\alpha+d)}$, with n determines the optimal bias-variance tradeoff in the estimation process, $D_{cen} \asymp m/n \asymp n^{-2\alpha/(2\alpha+d)}$: the smoother the field (larger α), the slower the growth rate of the number of cells and the lower the achievable L_2 distortion. From a purely communication theoretic viewpoint, the number of nodes per cell, $k = n/m \asymp n^{2\alpha/(2\alpha+d)}$, is the key quantity: phase coherent transmission by the nodes (in each cell) in the direction of destination results in a k fold power amplification that directly impacts the distortion introduced by the noisy MAC. The local matching coupled with uncoded analog transmission naturally integrates computation and communication: local sufficient statistics are directly available at the destination due to spatial averaging inherent in the MAC. The spatial averaging simultaneously reduces the impact of the two key sources of error: measurement noise and communication noise. The optimal power-distortion scaling laws are obtained by balancing the estimation and communication components of the L_2 distortion. The power-distortion scaling laws are optimal in the sense that they determine the minimum power scaling needed under noisy communications to attain the optimal centralized distortion scaling (under noise-free communications).

Our analysis identifies three distinct optimal power-distortion scaling regimes: 1) measurement-limited, 2) communication-limited (no measurement noise, but linear constraint on total network power scaling) and 3) bias-limited (no measurement noise and no power constraints). In the measurement-limited regime, optimal distortion ($D_{des} \asymp n^{-2\alpha/(2\alpha+d)}$) is achieved with sub-linear network power scaling ($P \asymp n^{-2\alpha/(2\alpha+d)} \downarrow 0$); in the communication-limited regime, a faster distortion scaling ($D_{des} \asymp n^{-2\alpha/(\alpha+d)}$) is achieved with linear network power scaling ($P > 0$); and in the bias-limited regime, the fastest distortion scaling ($D_{des} \asymp n^{-2\alpha/d}$) is achieved with super-linear network power scaling ($P \asymp n^{2\alpha/d} \uparrow \infty$). Furthermore, consistent estimation is achievable in the measurement-limited regime ($D_{des} \downarrow 0$) as long as P goes to zero a little slower than $1/n^2$ (cut-off power scaling). These power scaling regimes are illustrated in Fig. 2 and the corresponding distortion scaling is illustrated in Fig. 3 for $\alpha = 1$ and $d = 2$. For the sake of illustration, these plots assume that the constants of proportionality in the scaling relations are unity; the actual constants depend on ρ_{meas} and ρ_{com}, as calculated earlier.

Figures 2 and 3 also illustrate the power-density tradeoff. For example, suppose we want to attain a target distortion of 0.1. With optimal distortion scaling in the measurement-limited regime (solid curve in Fig. 3), the desired distortion can be attained with $n = 100$ nodes, consuming a total network power $P_{tot} = nP = 100 * 10^{-1} = 10$, as calculated from Fig. 2. On the other hand, if we operate at a sub-optimal distortion scaling (the third dashed-dot feasible curve in Fig. 3), we can attain the desired distortion with $n = 1000$ nodes consuming a total network power $P_{tot} = nP = 1000 * 10^{-3} = 1$. Thus, as predicted by (28), increasing the sensor density by a factor of 10 reduces the total network consumption by a factor of 10.

While the power pooling gain increases with k, the volume occu-

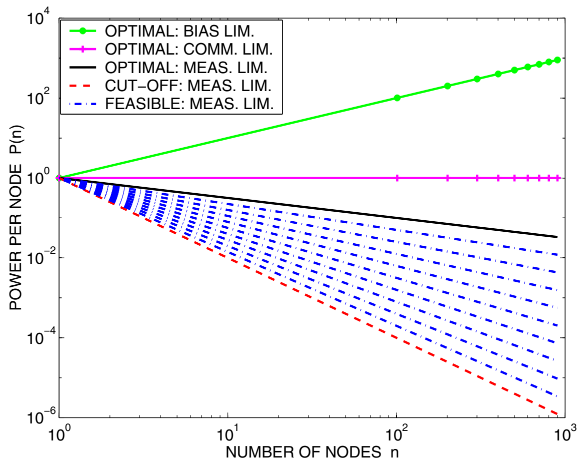

Fig. 2. Different scaling regimes for P. $\alpha = 1$, $d = 2$.

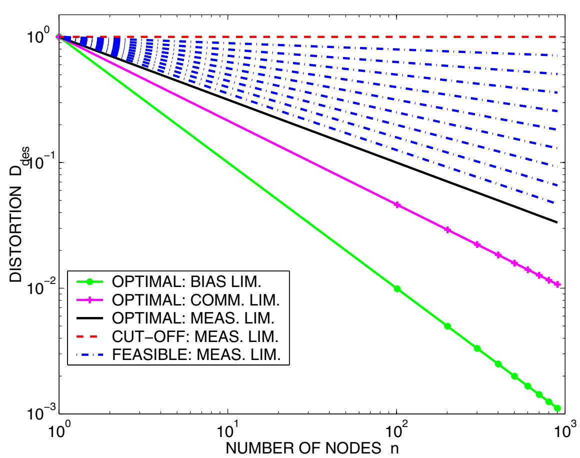

Fig. 3. Different scaling regimes for D_{des}. $\alpha = 1$, $d = 2$.

pied by the nodes in each cell, k/n, decreases monotonically with n. Thus, local node synchronization for phase-coherent transmission is required on smaller and smaller spatial scales as the sensor density increases. On the other hand, the number of cells, m, represents the degrees of freedom in the signal field at a given sampling density: a finite number of parameters (sufficient statistics), determined by α, need to be communicated from each cell to the destination. Under our assumption of dedicated, non-interfering MACs from each cell to the destination, m also represents the bandwidth or latency required for communicating the sufficient statistics.

V. Conclusions and Future Work

This paper develops a new approach to field estimation using wireless sensor networks that combines the operations of processing and communication in a power efficient manner. Our approach is based on a distributed source-channel communication architecture that matches the spatial scale of field coherence with the spatial scale of node synchronization for coherent communication. The averaging required in the optimal field estimator is implicitly computed at the destination via the coherent spatial averaging inherent in the MAC, resulting in optimal power-distortion scaling with the number of nodes. In particular, optimal mean-square distortion scaling can be achieved with sub-linear network power scaling. Our results also reveal a remarkable *power-density tradeoff* due to coherent communication: increasing the sensor density reduces the total network power required to achieve a desired distortion. A direct consequence is that consistent field estimation is possible, in principle, even with vanishing total network power in the limit of high sensor density. Our future work includes extensions to inhomogeneous fields.

Appendix

It follows from the definition of $C^{\alpha}(B)$ in (2) that the remainder of an order $\alpha - 1$ Taylor approximation to a function $f \in C^{\alpha}(B)$ on a cell of sidelength $m^{-1/d}$ is $O(m^{-\alpha/d})$. Now consider the form of the Taylor polynomial, which can be re-written as

$$f_p(\xi) = \sum_{j=0}^{p} \sum_{\sum q_k = j} a_{q_1,\ldots,q_d} \prod_{k=1}^{d} (\xi(k) - \nu(k))^{q_k}$$

where the second sum ranges over all non-negative integers $(q_1, \ldots, q_d)$ such that $\sum_k q_k = j$ and the coefficients $a_{q_1,\ldots,q_d}$ are determined by the partial derivatives of f. Note that the Taylor polynomial is a linear function in the coefficients. Our estimator must determine the least squares fit of the coefficients to the sensor data in each cell. Consider an arbitrary cell Q. Without loss of generality, we may assume that one corner is the point $\nu = (0, \ldots, 0)$. The least squares fit is given by the coefficients that minimize

$$\sum_{i=1}^{n_Q} \left(x_i - \sum_{j=0}^{p} \sum_{\sum q_k = j} a_{q_1,\ldots,q_d} \prod_{k=1}^{d} \xi_i^{q_k}(k) \right)^2$$

where n_Q is the number of sensors in the cell, x_i is the sample at sensor i, and $\xi_i(k)$ is the k-th location coordinate of sensor i. Thus, the sufficient statistics for the least squares fit are given by sums of the form

$$\theta = \sum_{i=1}^{n_Q} \left[x_i \prod_{k=1}^{d} \xi_i^{q_k}(k) \right]$$

Note that in the special case of $\alpha = 1$, the order of the Taylor polynomial is zero (consequently $q_k = 0$) and the single sufficient statistic is the sum of the observations. If $\alpha = 2$ and $d = 2$, then there are three sufficient statistics in each cell: $\theta_1 = \sum_{i=1}^{n_Q} x_i$, $\theta_2 = \sum_{i=1}^{n_Q} x_i \xi_i(1)$, and $\theta_3 = \sum_{i=1}^{n_Q} x_i \xi_i(2)$. In general, the "weight" on the x_i term, $\prod_{k=1}^{d} \xi_i^{q_k}(k)$, is a function of the location of sensor i, and hence no cooperation is required to compute the weights.

At the destination, a linear operation (depending on the locations of sensors in each cell) transforms the sufficient statistics into Taylor polynomial coefficients. For example, in the $\alpha = 2$, $d = 2$ case there is a 3×3 matrix T that maps $(\theta_1, \theta_2, \theta_3) \mapsto (a_1, a_2, a_3)$ and the resulting field estimate has the form

$$\widehat{f}(\xi) = \sum_{j=1}^{m} \left[a_{1,j} + a_{2,j}\xi(1) + a_{3,j}\xi(2) \right] 1_{\xi \in Q_j}$$

where the additional subscript j appearing on each coefficient indicates the set of coefficients associated with each cell. The largest eigenvalue of T upper bounds the amplification of errors (estimation and communication) in the reconstruction process; at most the variance is increased by a constant factor. Specifically, consider the field reconstruction in one cell. Let $\mathbf{\Theta}_o$ denote the three dimensional vector of sufficient statistics in the absence of measurement noise. The estimate of $\mathbf{\Theta}_o$ at the destination takes the form

$$\hat{\mathbf{\Theta}}_{des} = \mathbf{\Theta}_o + \tilde{\mathbf{w}} + \frac{\mathbf{z}}{\sqrt{P} n_Q}$$

where $\tilde{\mathbf{w}}$ is a linear transformation of measurement noise and $\mathbf{z}$ is the communication noise for three channel uses. The reconstructed signal at the destination is

$$\widehat{f}(\xi) = \langle \mathbf{t}_1 + \xi(1)\mathbf{t}_2 + \xi(2)\mathbf{t}_3, \hat{\mathbf{\Theta}}_{des} \rangle$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\mathbf{t}_i$ is the i-th row of T. We can also consider the reconstruction at the locations of the nodes in the network $\hat{s}_{des,i} = \widehat{f}(\xi_i)$ and evaluate the *per-node* distortion instead of the L_2 distortion. Both distortions are of the same order, but the per-node distortion calculation explicitly reveals the role of n_Q, the number of sensors in a cell:

$$\begin{aligned} D_{des} &= \frac{1}{n_Q} \sum_{i=1}^{n_Q} E[(s_i - \hat{s}_{des,i})^2] \\ &= D_{cen} + \frac{1}{P n_Q^3} \sum_{i=1}^{n_Q} E[\langle \mathbf{v}_i, \mathbf{z} \rangle^2] \\ &\preceq \frac{1}{m^{2\alpha/d}} + \frac{1}{n_Q} + \frac{1}{P n_Q^2} \end{aligned}$$

where $\mathbf{v}_i = \mathbf{t}_1 + \xi_i(1)\mathbf{t}_2 + \xi_i(2)\mathbf{t}_3$ and the second term in the second equality represents D_{com}. If the largest eigenvalue of T is bounded, the variance term in $D_{cen} \preceq 1/n_Q$ and $\sum_{i=1}^{n_Q} E[\langle \mathbf{v}_i, \mathbf{z} \rangle^2] \preceq n_Q$, resulting in the last relation which exhibits same scaling as in (13). Similarly, this argument can be extended to general $C^\alpha(B)$ fields to show similar scaling behavior.

References

[1] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Proc. 2nd Intl. Workshop on Information Processing in Sensor Networks (IPSN'03)*, Apr. 2003, pp. 162–177.

[2] A. M. Sayeed, "A statistical signal modeling framework for wireless sensor networks," *UW Tech. Rep. ECE-1-04*, Feb. 2004.

[3] M. Gastpar and M. Vetterli, "Power-bandwidth-distortion scaling laws for sensor networks," in *Proc. 3rd Intl. Symp. on Information Processing in Sensor Networks (IPSN'04)*, Apr. 2004, pp. 320–329.

[4] K. Liu and A. M. Sayeed, "Optimal distributed detection strategies for wireless sensor networks," in *Proc. 42nd Annual Allerton Conference on Commun., Control and Comp.*, Oct. 2004.

[5] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of Gaussian space-time sources in wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1026–1036, 2004.

[6] R. DeVore, "Nonlinear approximation," *Acta Numerica*, pp. 51–150, 1998.

[7] E. Kolaczyk and R. Nowak, "Multiscale likelihood analysis and complexity penalized estimation," *Annals of Statistics*, vol. 32, no. 2, pp. 500–527, 2004.

[8] R. Nowak, U. Mitra, and R. Willett, "Estimating inhomogeneous fields using wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 999–1006, 2004.

[9] A. P. Korostelev and A. B. Tsybakov, *Minimax theory of image reconstruction*. New York: Springer-Verlag, 1993.

[10] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A bit-conservation principle," in *Proc. 2nd Intl. Workshop on Information Processing in Sensor Networks (IPSN'03)*, Apr. 2003.

[11] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data," in *Proc. 2nd Intl. Workshop on Information Processing in Sensor Networks (IPSN'03)*, Apr. 2003, pp. 1–16.

[12] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor networks," *IEEE Sig. Proc. Mag.*, vol. 19, pp. 51–60, 2002.

[13] S. D. Servetto, "On the feasibility of large-scale wireless sensor networks," in *Proc. 40th Annual Allerton Conference on Commun., Control and Comp.*, 2002.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.