

Estimating Inhomogeneous Fields Using Wireless Sensor Networks

Robert Nowak, *Senior Member, IEEE*, Urbashi Mitra, *Senior Member, IEEE*, and Rebecca Willett, *Student Member, IEEE*

Abstract—Sensor networks have emerged as a fundamentally new tool for monitoring spatial phenomena. This paper describes a theory and methodology for estimating inhomogeneous, two-dimensional fields using wireless sensor networks. Inhomogeneous fields are composed of two or more homogeneous (smoothly varying) regions separated by boundaries. The boundaries, which correspond to abrupt spatial changes in the field, are nonparametric one-dimensional curves. The sensors make noisy measurements of the field, and the goal is to obtain an accurate estimate of the field at some desired destination (typically remote from the sensor network). The presence of boundaries makes this problem especially challenging. There are two key questions: 1) Given n sensors, how accurately can the field be estimated?; 2) How much energy will be consumed by the communications required to obtain an accurate estimate at the destination? Theoretical upper and lower bounds on the estimation error and energy consumption are given. A practical strategy for estimation and communication is presented. The strategy, based on a hierarchical data-handling and communication architecture, provides a near-optimal balance of accuracy and energy consumption.

Index Terms—Distributed estimation, multiresolution analysis, sensor networks, wavelets.

I. INTRODUCTION

SENSOR networks have emerged as a fundamentally new tool for monitoring inaccessible environments such as non-destructive evaluation of buildings and structures, contaminant tracking in the environment, habitat monitoring in the jungle, and surveillance in military zones. These ad hoc networks are envisioned to be a collection of embedded sensors, actuators, and processors. We shall assume that communication between sensors is done in a wireless fashion. Sensor networks are distinguished from more classical networks due to strict limitations on energy consumption, the density of nodes, the simplicity of the processing power of nodes, and possibly high environmental dynamics. An important problem in sensor networking applications is the estimation of spatially varying processes or fields. Consider a network sensing a field composed of two or

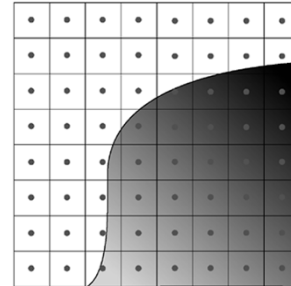


Fig. 1. A wireless sensor network sampling a two-dimensional (2-D) field. Dots indicate sensor locations and squares indicate the extent of each sensor's measurement.

more regions of distinct behavior (e.g., differing mean values for the sensor measurements). An example of such a field is depicted in Fig. 1. In this application, the goal of the sensor network is to sense the field, construct an estimate of the field, and communicate the estimate to a desired (typically remote) destination. In contrast to prior work, we drop the assumption of homogeneity and instead consider inhomogeneous fields that are composed of two or more homogeneous (smoothly varying) regions separated by smooth boundaries. While the proposed method exhibits near-optimal theoretical properties and is practical for both homogeneous and inhomogeneous fields, we focus in this paper on inhomogeneous fields. This is a particularly challenging problem in sensor networks because boundary detection requires significant system resources. Furthermore, abrupt changes in the field are often of greatest scientific interest and, therefore, we view the inhomogeneous case as especially relevant to applications.

A. Wireless Sensing of Inhomogeneous Fields

There are two fundamental limitations in the estimation problem. First, the accuracy of a field estimate is limited by the spatial density of sensors in the network and by the amount of noise associated with the measurement process. Second, energy constraints may limit the complexity of the estimate that is computed and ultimately communicated to a desired destination. The tradeoff between accuracy and energy consumption can be characterized as follows.

Assume that n sensor nodes are arranged on an $\sqrt{n} \times \sqrt{n}$ square lattice (assuming a planar, square field). Suppose that the field being sensed consists of two homogeneous regions separated by a one-dimensional (1-D) boundary (like the case depicted in Fig. 1). To be specific, let us assume that the field consists of two smoothly varying regions separated by a smooth boundary. Under these assumptions, most of the nodes will be in smooth regions away from the boundary, and only $O(\sqrt{n})$

Manuscript received July 15, 2003; revised February 1, 2004. The work of R. Nowak was supported in part by the National Science Foundation under Grant CCR-0310889 and Grant ANI-0099148, in part by the Office of Naval Research under Grant N00014-00-1-0390, and in part by the State of Texas ATP under Grant 003604-0064-2001. The work of U. Mitra was supported in part by the Texas Instruments Visiting Professorship and in part by the University of Southern California, Los Angeles, CA.

R. Nowak and R. Willett are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53707 USA and also with Rice University, Houston, TX 77001 USA (e-mail: nowak@engr.wisc.edu; willett@rice.edu).

U. Mitra is with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90007 USA (e-mail: ubli@usc.edu).

Digital Object Identifier 10.1109/JSAC.2004.830893

nodes lie next to the boundary. Each sensor node makes a (noisy) measurement of the field. Specifically, each sensor makes a measurement of the field at its location which is contaminated with a zero-mean Gaussian distributed error. Assume that the noises are independent and identically distributed at each sensor. It is known that, under the assumptions on the class of fields above and the white noise model, the mean-square error (MSE) cannot, in general, decay at a rate faster than cn^ν , for some $0 < \nu \leq 1$ and some $c > 0$ independent of n that depends on the smoothness of the regions and the boundary [1]–[3]. That is, no estimator (based on centralized or distributed processing) can exceed this convergence speed limit.

To quantify the total energy required to compute and transmit a field estimate of this accuracy, we assume a $1/r$ energy decay behavior, where r is the distance from the transmitter (in light of the $2 - d$ nature of our formulation). Other measures are possible (e.g., $1/r^\alpha$ for some $\alpha \geq 1$); see [4] for examples. Note that even simple calculations require communications that consume at least $O(\sqrt{n})$ bit-meters of energy. For example, consider computing the average values between the nearest neighboring nodes in the network. This requires the transmission of $O(n)$ bits (a certain number for each node in the network) over a distance of $O(1/\sqrt{n})$ meters (the distance between nearest neighbors). Such a trivial operation is necessary for almost any imaginable field estimation process. Thus, the total energy required to compute and transmit the field estimate is at best $O(\sqrt{n})$. Note that, here and throughout the paper, we assume that local computations require a negligible amount of energy in comparison to communications. Combining this lower bound on the energy consumption with the lower bound on the MSE decay rate above yields a “best-case” tradeoff between accuracy (MSE) and energy (\mathcal{E}) of the form $\text{MSE} \propto 1/\mathcal{E}^{2\nu}$. It is important to note that this relation should not be interpreted to mean that a fixed number of sensor nodes using more energy can provide more accuracy. Rather, both the MSE and the energy consumption are functions of the number of sensor nodes, and the above relation indicates how the accuracy and energy consumption behave as the density of nodes increases. Moreover, note that the MSE cannot decay at a rate faster than n^{-1} , the parametric rate. Therefore, the very best tradeoff between MSE and energy is $\text{MSE} \propto 1/\mathcal{E}^2$, and this is generally unachievable except under very restrictive and unrealistic assumptions on the nature of the field.

This paper explores these basic tradeoffs between MSE and energy consumption as functions of node density. We propose and develop field estimation algorithms based on multiscale partitioning methods. The algorithms are quite practical and map nicely onto a sensor network architecture. Moreover, we demonstrate theoretically that our methods nearly achieve the best-case MSE/energy tradeoffs discussed above. The theory hinges on an application of our extension [5] of the Li–Barron bound for complexity regularized model selection [6] to bound the MSE and to bound the expected energy consumption. Since our methods (nearly) achieve the optimal tradeoffs, no other schemes can be devised that will (asymptotically) perform significantly better under the stated assumptions. Simulation experiments verify the predicted theoretical performance of our methods. The algorithms and theoretical analysis in this paper expand upon our previous work [7], which considered only piecewise constant fields.

B. Related Work

Due to the recent emergence of sensor network research, there is limited literature concerning field estimation via wireless networks. We are aware of several lines of investigation. First, several authors [8]–[11] have considered the problem of estimating a field via wireless sensor networks. In those works, rate-distortion analysis is employed to study essentially the same tradeoff between accuracy and communication costs as discussed above. However, the class of fields considered by those authors differs strikingly from our investigation. Rather than consider inhomogeneous fields as we do here, they considered classes of homogeneous processes without boundaries (e.g., stationary Gaussian fields). While possibly adequate for some applications, we feel that boundaries corresponding to abrupt spatial changes in the field may occur quite frequently in many envisioned application domains. Moreover, abrupt changes in the field are likely to be the features most critical to the user. The theory and methods in [8]–[10] do not address such situations. Second, to address the important problem of boundary detection, several proposals have been made to employ classical image processing tools in sensor network algorithms. In particular, the work of [12] has led to simple and communication-efficient strategies for detecting field boundaries using wireless sensor networks. Similar “image-processing” influenced schemes are considered in [13]. However, like most classical image edge detection methods, it is difficult to quantify the accuracy of these strategies.

We also mention the close connection between the hierarchical estimation and communication architecture employed in our methods and the data collection algorithm proposed in [14], which also exploits a hierarchy to reduce communication costs. That algorithm is based on hierarchical compression scheme where clusterheads aggregate measurements from child nodes to pass to the next layer of hierarchy. Very similar strategies underlie our methods and are critical for achieving near-optimal tradeoffs between accuracy and communication. Thus, we view our work as theoretically supporting the use of such hierarchies in sensor network architectures.

C. Organization

This paper is organized as follows. In Section II, we define the problem and develop the basic algorithmic and communication structure of our field estimators. In Section III, we study the theoretical properties of the estimators. In particular, we develop bounds on the mean-squared estimation error and on the communication costs of the proposed methods. Simulation results are provided in Section IV. Section V draws final conclusions and discusses ongoing research. The Appendix contains key derivations and proofs.

II. PROBLEM FORMULATION AND APPROACH

Assume a 2-D field, denoted by f , is being sensed and that it is composed of smooth regions of Hölder- α regularity in \mathbb{R}^2 , $\alpha \in (0, 2]$, with smooth contours or *boundaries* that are Hölder- β regular in \mathbb{R}^1 , $\beta \in (0, 2]$. As special cases, if α and β are integer valued, then these cases include fields in which smooth regions are C^α (α continuous derivatives in \mathbb{R}^2) and the boundaries are C^β (curves that behave locally like C^β 1-D functions). The domain of the field is assumed to be the unit square $[0, 1]^2$.

Furthermore, the range of the field is assumed to be within $[-R, R]$, for some constant $R > 0$. Extensions of our theory and methods to three-dimensional fields are straightforward, and we will focus on 2-D fields and sensor networks to illustrate the salient ideas.

The sensor network is a collection of n wireless nodes arranged on a uniform, square lattice in $[0, 1]^2$. Each node measures the field at its position. Denote the sensor measurements by $x \equiv \{x_{i,j}\}$, where the indexes range $1 \leq i, j \leq \sqrt{n}$. The indices indicate the location of the corresponding node which is at the center of the square $S_{i,j} = [(i-1)/\sqrt{n}, i/\sqrt{n}] \times [(j-1)/\sqrt{n}, j/\sqrt{n}]$. Let $f_{i,j} = n \int_{S_{i,j}} f$, the average value of the field over the square $S_{i,j}$. The array $\{f_{i,j}\}$ can be viewed as a “pixelized” version of the original field. Assume that the measurement at each node has the form

$$x_{i,j} = f_{i,j} + n_{i,j} \quad (1)$$

where $n_{i,j}$ is an observation noise that could encompass ambient noise in the environment, electronic transducer noise sources, and possibly quantization errors. The noises $\{n_{i,j}\}$ are assumed to be independently and identically Gaussian distributed with mean zero and variance σ^2 . Thus, we have $x_{i,j} \sim \mathcal{N}(f_{i,j}, \sigma^2)$, a realization of a Gaussian random variable with mean $f_{i,j}$ and variance σ^2 . The goal of the sensor network is to compute an estimate of the field f given the noisy measurements x and to transmit this estimate to a desired location (assumed to be remote from the sensor network). The basic problem is illustrated in Fig. 1.

In our hierarchical approach to estimation and communication, the sensor measurements can be viewed as sampling the field over a partition of n subsquares of sidelength $1/\sqrt{n}$, as shown in Fig. 1. In principle, this initial partition can be generated by a *recursive dyadic partition* (RDP) as follows. First, divide the domain into four subsquares of equal size. Repeat this process again on each subsquare. Repeat this $1/2 \log_2 n = J$ times. This gives rise to a *complete* RDP of resolution $1/\sqrt{n}$ (the rectangular partition of the sensing domain shown in Fig. 1). The RDP process can be represented with a quadtree structure. The quadtree can be pruned back to produce an RDP with nonuniform resolution.

Let \mathcal{P}_n denote the set of all RDPs, including the initial complete RDP and all possible prunings. For each RDP $P \in \mathcal{P}_n$, there is an associated quadtree structure (generally of nonuniform depth corresponding to the nonuniform resolution of most RDPs). The leafs of each quadtree represent dyadic (sidelength equal to a negative power of two square regions of the associated partition). For a given RDP and quadtree, each sensor node belongs to a certain dyadic square. We consider these squares “clusters” and assume that one of the nodes in each square serves as a “clusterhead,” which will assimilate information from the other nodes in the square. Notice that if one considers all RDPs in \mathcal{P}_n , then each sensor node actually belongs to a nested hierarchy of $1/2 \log_2 n$ dyadic squares of sidelengths $1/\sqrt{n}, 2/\sqrt{n}, 4/\sqrt{n}, \dots, 1$, respectively. Thus, we have a hierarchy of clusters and clusterheads.

Consider a certain RDP $P \in \mathcal{P}_n$. Define the estimator of the field as follows. On each square of the partition, we consider two possible estimators of the field in that region. One estimator is simply the least squares fitting of a constant on the corresponding square, in other words the mean of the mea-

surements in the square. A discrete collection of such models is generated by quantizing the dynamic range $[-R, R]$ of the field to $n^{1/4}$ levels (in other words using $1/4 \log n$ bits to represent the mean value). This choice for the quantization is due to the fact that ultimately the MSE of the estimator is $O(n^{-1/2})$ (as shown in the next section) and this choice puts the squared quantization error at that level. The second estimator is a least squares fitting of a piecewise linear model known as a *platelet* [15]. A platelet is formed by splitting a dyadic square into two wedge-shape regions (defined by a line connecting two points on the boundary of the square) and fitting a planar surface to the field on each wedge. Each platelet consists of six parameters, which describe the two surfaces, and a pair of points on the boundary of the square, which describe the wedge. A discrete collection of platelets are generated by discretizing the surface parameters to $1/3 \log n$ bits each and selecting the boundary points from a finite collection of points spaced evenly $\delta = n^{-2/3}$ units apart along the boundary of the square. Again, these choices of quantization are made so that the squared quantization error matches the overall MSE, which is shown to be $O(n^{-2/3})$ in the next section. The platelet estimator is determined by selecting the element of this collection that minimizes the sum of squared errors between it and the data. In summary, the two types of estimators employed at each leaf of the RDP (denoted by P) are the following.

- Estimator 1, denoted by $\hat{f}^h(P)$: least squares fit of a constant to the data in each square.
- Estimator 2, denoted by $\hat{f}^p(P)$: least squares fit of a platelet to the data in each square.

Estimator 1 is very similar to a Haar wavelet estimator and, therefore, we will call it the “Haar estimator.” Estimator 2 will be called the “platelet estimator.” The remaining issue is the selection of the pruned RDP. The complete (unpruned) RDP will, of course, provide a perfect fit to the data (i.e., $\hat{f}^h(P) = \hat{f}^p(P) = x$). This choice of partition is undesirable for two reasons. First, the data are noisy and fitting constants or platelets over larger squares will affect an averaging that reduces the noise. Second, the unprocessed data x will require the maximum amount of energy to transmit to the destination. It turns out that properly penalizing partitions based on the number of squares (or leafs in the corresponding tree) provides both a more accurate estimate of the field and a compressed representation of the data more suitable for transmission to the destination.

The *complexity penalized* estimator is defined as follows. Let $f^m(P)$ denote a model of the field (based either on constant ($m = h$) or platelet ($m = p$) models on each square of P). The empirical measure of performance is the sum-of-squared errors between $f^m(P)$ and the data $x = \{x_{i,j}\}$.

$$R(f^m(P), x) = \sum_{i,j=1}^{\sqrt{n}} (f_{i,j}^m(P) - x_{i,j})^2. \quad (2)$$

For a fixed partition, P , the choice of $f^m(P)$ that minimizes $R(f^m(P), x)$ is simply given by the least squares fits on each square, as discussed above. For $m = h$ or $m = p$, define the complexity penalized estimator

$$\hat{f}_n^m = \arg \min_{\substack{f^m(P): \\ P \in \mathcal{P}_n}} R(f^m(P), x) + 2\sigma^2 g^m(n) |P| \quad (3)$$

where $|P|$ denotes the number of (square or wedge-shaped) regions in the partition P , and $g^m(n)$ is a certain monotonically

increasing function of n that discourages unnecessarily high-resolution partitions (appropriate choices of $g^m(n)$ will be discussed in Section III). It is well known that the optimization in (3) can be solved using a bottom-up tree pruning algorithm [2], [15], [16]. This is possible because both the sum-of-squared errors and the penalty are additive functions and, therefore, the squared error plus penalty cost can be separated into terms associated with each individual square of the candidate partition.

The hierarchy of clusterheads facilitates this process in the sensor network. First consider the case for the Haar estimator \hat{f}^h . At each level of the hierarchy, the clusterhead receives the means from the four clusterheads below it, and compares the total cost of these estimates (sum-of-squared errors plus penalties) with the total cost of the mean estimate on the larger square associated with the clusterhead. Implementation of platelet approximations is more challenging in a decentralized sensor network structure. This is due to the fact that the optimization involved requires that all data be passed from the bottom up to the root of the tree structure [15]. More specifically, all the data in each square are needed to select the best wedge split. This may require a prohibitive amount of communication in large networks. Therefore, we advocate the following approximate strategy. Rather than passing all data from finer scales to coarser scales, we will instead pass only the optimized platelet fits from below. For example, each parent node will receive the best platelet fits from its four children nodes, and select an optimal platelet fit for itself based on these, rather than on the complete data from the four children. In the simulation section, it is demonstrated that this approximate strategy works about as well as the optimal one. The key advantage of the approximate strategy is that now only a handful of summary statistics need to be passed up the tree structure, and the communication requirements for \hat{f}^p are on the same order as those for \hat{f}^h .

The energy required for the above process is the sum of two terms: 1) the in-network communication required to obtain the estimate at the final clusterhead and 2) the out-of-network cost of transmitting the estimate to the destination. Note that the in-network communication only requires the transmission of data and/or sufficient statistics over short distances within the sensor network, while out-of-network communication is typically over a much greater distance and may dominate the overall cost of communication. The energy requirements will be discussed in detail in Section III.

III. THEORETICAL PERFORMANCE

In Section I, we described a fundamental tradeoff between accuracy (MSE) and energy (\mathcal{E}) of the form $\text{MSE} \propto 1/\mathcal{E}^{2\nu}$. In this section, we demonstrate that the hierarchical methods proposed above nearly achieve this optimal tradeoff for the cases of $\nu = 1/2$ and $\nu = 2/3$, which correspond to two different levels of assumed smoothness in the field and boundaries. The theory hinges on an application of our extension [5] of the Li–Barron bound for complexity regularized model selection [6] to upper bound the MSE of the estimators and to bound the expected energy consumed by communications to compute the estimates. Since our methods (nearly) achieve the optimal tradeoffs, no other schemes can be devised that will (asymptotically) perform significantly better.

Assume that the 2-D field denoted by f is composed of smooth regions of Hölder- α regularity in \mathbb{R}^2 , $\alpha \in (0, 2]$, with smooth boundaries that are Hölder- β regular in \mathbb{R}^1 , $\beta \in (0, 2]$. We will focus on two classes.

- $C1$: Hölder-1 regular smooth regions separated by Hölder-1 regular boundaries.
- $C2$: Hölder-2 regular smooth regions separated by Hölder-2 regular boundaries.

The class $C2$ is a subset of $C1$ and places stronger assumptions on the smoothness than class $C1$ and, therefore, one expects better performance (in terms of MSE and energy consumption) is possible in class $C2$, which turns out to be true. Stronger smoothness assumptions are, of course, possible; say Hölder- α , $\alpha > 2$. However, the algorithms we proposed above are not capable of taking advantage of such additional assumptions; polynomial approximations to the surfaces and boundaries would be necessary for smoother functions, and we do not pursue such possibilities in this paper. We will show that our algorithms are nearly optimal for classes $C1$ and $C2$, achieving the tradeoffs mentioned in the paragraph above. If one were to consider even smoother conditions, then the best tradeoff for which one could aim is $\text{MSE} \propto 1/\mathcal{E}^2$ since the MSE cannot decay at a rate faster than n^{-1} , the parametric rate and the communication energy must be at best $O(\sqrt{n})$ bit-meters, as discussed in Section I.

A. MSE Performance

The performance of the proposed estimators can be studied in terms of (minimax) lower bounds and upper bounds on the MSEs. The minimax error is the error incurred by the best possible (Haar or platelet) estimator on the hardest possible field in class $C1$ or $C2$. The MSE is defined as

$$\text{MSE}(\hat{f}^m) = \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\hat{f}^m(i,j) - f(i,j))^2]. \quad (4)$$

First, consider the $C1$ class. Taking $g^h(n) \approx (1/4) \log n$, let \tilde{f} denote any estimator based on the data x . Then, for n sufficiently large

$$\begin{aligned} c_1 n^{-1/2} &\leq \inf_{\tilde{f}} \sup_{f' \in C1} \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\tilde{f}(i,j) - f'(i,j))^2] \\ &\leq \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\hat{f}^h(i,j) - f(i,j))^2] \\ &= O\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \end{aligned} \quad (5)$$

for some $c_1 > 0$ independent of n . Similarly, for class $C2$ take $g^p(n) \approx (4/3) \log n$. Then

$$\begin{aligned} c_2 n^{-2/3} &\leq \inf_{\tilde{f}} \sup_{f' \in C2} \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\tilde{f}(i,j) - f'(i,j))^2] \\ &\leq \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\hat{f}^p(i,j) - f(i,j))^2] \\ &= O\left(\left(\frac{\log n}{n}\right)^{2/3}\right) \end{aligned} \quad (6)$$

for some $c_2 > 0$ independent of n . The minimax lower bounds follow from fairly standard “hypercubical” type arguments that recast the field estimation problem into a simpler series of independent hypothesis tests, for which the error rates are easy to determine [1], [2].

The upper bounds are based on a combination of approximation theory (from the field of wavelets and harmonic analysis) coupled with upper bounds on the variance of estimators based on such approximations. We begin by recalling a fundamental upper bound on expected error of complexity penalized estimators. This particular bound was originally developed for mixture density modeling [6], and we later extended it to more general settings [5]. Here, we state a specialized version of the bound, tailored to the estimator proposed in (3).

Let \mathcal{F}_n^m denote the set of all possible models of the field. This set contains either Haar approximations ($m = h$) or platelet approximations ($m = p$), as described in Section II. Recall that the parameters of the models are quantized, so the set \mathcal{F}_n^m consists of a finite number of models. Let \tilde{f} denote a generic element of \mathcal{F}_n^m , and recall that the true field is denoted by f and our estimators are denoted by \hat{f}^m . Assume that $g^m(n)$ satisfies the summability condition (Kraft inequality)

$$\sum_{\tilde{f}(P) \in \mathcal{F}_n^m} e^{-g^m(n)|P|} \leq 1 \quad (7)$$

where $|P|$ denotes the number of (square- or wedge-shaped) leaves in the partition associated with $\tilde{f}(P)$. It is shown in the Appendix that $g^h(n) \approx 1/4 \log n$ and $g^p(n) \approx 4/3 \log n$ satisfy (7). Then, we can apply Theorem 7 in [5] to upper bound the MSE of the estimator \hat{f}_n^m as follows:

$$\begin{aligned} \text{MSE}(\hat{f}_n^m) &= \frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E[(\hat{f}_n^m(i,j) - f(i,j))^2] \\ &\leq \min_{\tilde{f}(P) \in \mathcal{F}_n^m} \left\{ \frac{2}{n} \sum_{i,j=1}^{\sqrt{n}} (\tilde{f}_{i,j} - f_{i,j})^2 + \frac{8\sigma^2 g^m(n)}{n} |P| \right\} \quad (8) \end{aligned}$$

The upper bound involves two terms. The first term, $2 \sum_{i,j=1}^{\sqrt{n}} (\tilde{f}_{i,j} - f_{i,j})^2$, is a bound on the bias or approximation error. The second term, $8\sigma^2 g^m(n)|P|$, is a bound on the variance or estimation error. The bias term, which measures the squared error between the best possible model in \mathcal{F}_n^m and the true field, can be easily bounded. Under the assumption that the underlying field belongs to class $C1$ or $C2$, well-known results from approximation theory [2]–[15] can be applied. In particular, if $|P| = M$ (corresponding to a partition with M squares or wedges), then for class $C1$, using either the Haar or platelet estimator, the bias term is $O(M^{-1})$ and for class $C2$, using the platelet estimator, it is $O(M^{-2})$. Then, by equating the bias and variance terms and solving for M in terms of n leads to the MSE upper bounds stated above. Also, note that the Haar estimator has the same MSE bound for both classes $C1$ and $C2$ because it cannot exploit the additional field smoothness in $C2$. Further details are given in the Appendix.

B. Expected Communication Requirements

Hierarchical networks of sensors, such as those described earlier, allow field estimates to be calculated and transmitted with

a significant reduction in energy. In this section, we will analyze the energy consumption associated with the field estimation methods described earlier. We assume that a certain received energy level E_0 is required for reliable communication, and that the transmitted energy decays like $1/r$, where r is the distance from the transmitter. Therefore, the energy required to reliably send 1 bit over r meters grows linearly in r . Other models for energy decay can be considered and incorporated into our framework [4], but we will not pursue this further here.

Without the hierarchical structure of the Haar and platelet estimation methods, one might naively transmit each sensor’s measurements to a remote processor at a cost of $\mathcal{E} \approx kn$ bit-meters, assuming that a constant number of bits k is transmitted per measurement. On the other hand, we argued in Section I that \mathcal{E} is as best $\geq O(\sqrt{n})$. Here, we show that the Haar and platelet estimators can be constructed and communicated with $O(\sqrt{n})$ bit-meters of expected communication.

The communication requirements can be split into two components. The “in-network” cost of computing the optimal estimator, and the “out-of-network” cost of transmitting the estimate to a remote destination. The out-of-network cost is proportional to the number of leaves in the partition associated with the optimal estimator. The size of the partition associated with the optimal estimator can be deduced from the MSE bound as follows. Suppose the optimal partition consists of M (square- or wedge-shaped) regions. Then, the variance is at least $O(M/n)$ since we have $O(M)$ free parameters to estimate ($O(1)$ parameters per region). Recall that the MSE is less than or equal to $O((\log n/n)^\nu)$, $\nu = 1/2$ for the Haar estimator, and $\nu = 2/3$ for the platelet estimator. Since the variance must be less than the MSE, it follows that $E[M] = O(n^{1-\nu} \log^\nu n)$. That is, for the Haar estimator $E[M] = O((n \log n)^{1/2})$ and for the platelet estimator $E[M] = O((n \log^2 n)^{1/3})$.

In the case of the Haar estimator \hat{f}^h , it turns out that the expected in-network communication costs are also proportional to the size of the partition associated with the optimal estimator. This is easy to see as follows. At the first stage of the hierarchy, $O(n)$ values must be transmitted over a distance of $O(n^{-1/2})$ meters [each group of four nodes send their measurements to the nearest cluster head, $O(n^{-1/2})$ meters away], requiring $O(\sqrt{n})$ bit-meters of energy. At each subsequent stage in the hierarchy, we expect $O((n \log n)^{1/2})$ squares will remain. Therefore, ignoring the log factor, $O(\sqrt{n})$ values must be transmitted to clusterheads which are $O(2^j/\sqrt{n})$, $j = 1, \dots, 1/2 \log n$ meters away. Thus, in total, the energy required for computing \hat{f}^h is $O(\sqrt{n})$ bit-meters.

The expected in-network costs for the platelet estimator, \hat{f}^p , are higher. As pointed out in Section II, the optimization involved requires that all data be passed from the bottom up to the root of the tree structure [15]. This leads to an in-network energy requirement of $O(n)$ bit-meters. However, the approximate strategy, based on passing only the optimized platelet fits up through the hierarchy, has the same energy requirements as the Haar estimator. In this case, the first stage requires $O(n^{1/2})$ bit-meters and subsequent stages involve passing $O(\sqrt{n})$ parameter values to clusterheads.

In conclusion, the expected in-network energy cost of $O(n^{1/2})$ and out-of-network energy cost of less than $O(n^{1/2})$

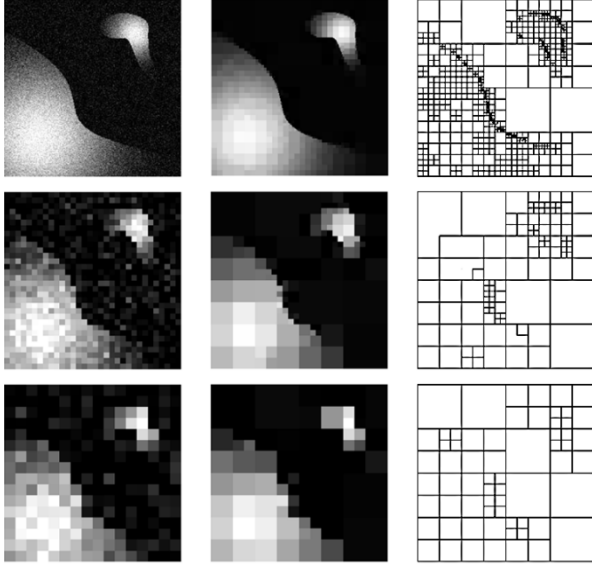


Fig. 2. Effect of sensor network density (resolution) on field estimation with the Haar estimator. Column 1 is the noisy set of measurements, column 2 is the estimated field, and column 3 is the associated partition. In Row 1, 65 536 sensors are used for form a partition with 889 regions, resulting in an MSE of $9.85e-2$. In row 2, 1024 sensors are used to form a partition with 127 regions, resulting in an MSE of $4.86e-1$. In row 3, 256 sensors are used to form a partition with 64 regions, resulting in an MSE of $7.26e-1$.

bit-meters combined with the upper and lower bounds on the MSEs produce the stated tradeoffs at the beginning of this section.

IV. SIMULATION EXPERIMENTS

The theoretical properties of the proposed multiscale estimators are supported in simulation experiments. We first simulated observations of a sample field for $n = 16^2$, 32^2 , and 256^2 , and compared the performance of the Haar and platelet estimators. The noise variance is $\sigma^2 = 1$, the intensity of the field has values ranging from zero to ten, and all estimates were computed using the theoretical penalties developed in Section I. These results were compared with the approximate platelet estimator, which as discussed above has a communication cost comparable to the Haar estimator. Fig. 2 displays the Haar estimates for the three different network sizes. The partitions displayed in the third column show that the size of the partition scales with the square root of the number of sensors, as predicted. The estimates in Fig. 3 show that the platelet method results in smoother estimates with more sharply defined edges on significantly small partitions. Again, the partition sizes grow with the cube root of the size of the network, as predicted by the theory above. The partition images also demonstrate the impact of the wedge fits on accurate boundary estimation. Compare these partitions with those generated by the Haar analysis. Because of the fine partition near the boundaries in the Haar estimate, estimates of the values of the field in those regions will have a large variance. The wedge-shaped blocks, in contrast, are generally larger, and so estimates of those regions are made using more sensors, thus lowering the estimator variance. Fig. 4 shows that approximate platelet estimator, calculated using only the estimates from the next finer scale instead of all the observations, vary only slightly from the optimal platelet estimator in Fig. 3. Thus, the platelet

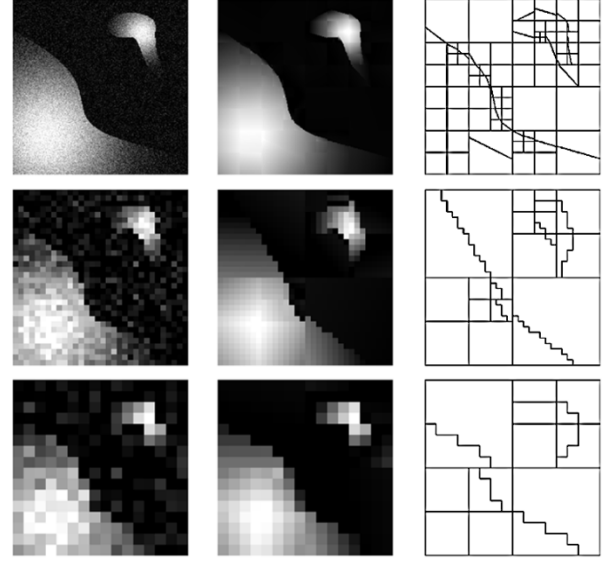


Fig. 3. Effect of sensor network density (resolution) on field estimation with platelets. Column 1 is the noisy set of measurements, column 2 is the estimated boundary, and column 3 is the associated partition. In row 1, 65 536 sensors are used for form a partition with 105 regions, resulting in an MSE of $3.64e-2$. In row 2, 1024 sensors are used to form a partition with 25 regions, resulting in an MSE of $3.33e-1$. In row 3, 256 sensors are used to form a partition with 16 regions, resulting in an MSE of $3.19e-1$.

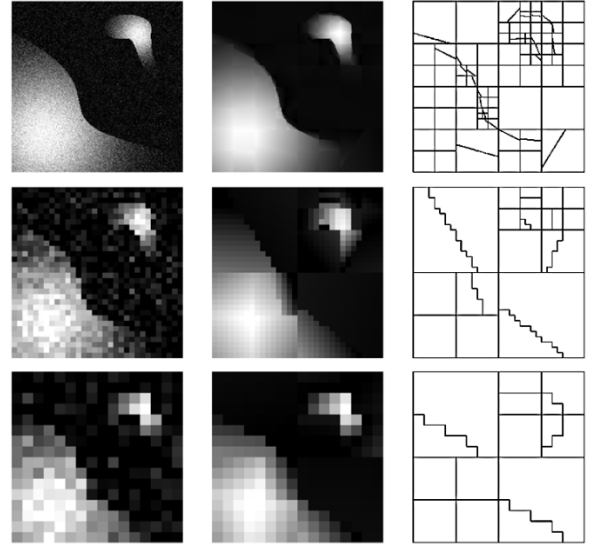


Fig. 4. Effect of sensor network density (resolution) on field estimation with platelets without passing all the data from fine to coarse levels. Column 1 is the noisy set of measurements, column 2 is the estimated boundary, and column 3 is the associated partition. In row 1, 65 536 sensors are used for form a partition with 107 regions, resulting in an MSE of $7.84e-2$. In row 2, 1024 sensors are used to form a partition with 23 regions, resulting in an MSE of $4.70e-1$. In row 3, 256 sensors are used to form a partition with 15 regions, resulting in an MSE of $3.71e-1$.

approximations should be an effective tool even when in-network communication costs prohibit extensive data passing.

Finally, Fig. 5 shows Haar, platelet, and approximate platelet estimates for a network with $n = 64^2$ sensors. The estimates in column 1 were calculated from data with a noise variance of 1, while the estimates in column 2 were calculated from data with a noise variance of 10. As predicted, weighting the penalization by the variance of the noise produces good estimates at different SNRs.

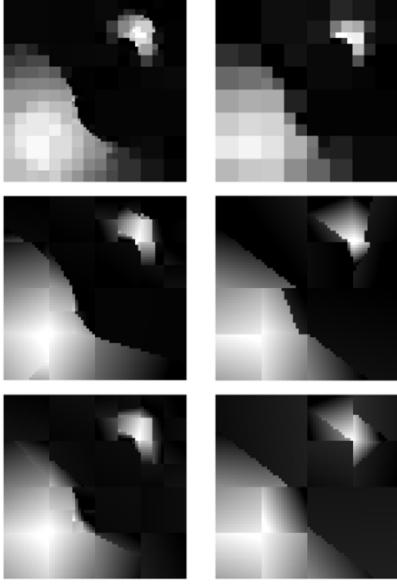


Fig. 5. Effect of noise variance on Haar and platelet estimates. Estimates in the first column were made using noisy data with variance 1. Estimates in the second column were made using noisy data with variance 10. In row 1, Haar estimates with 4096 sensors resulted in MSEs of $2.78e-1$ (for $\sigma^2 = 1$) and $1.03e+0$ (for $\sigma^2 = 10$). In row 2, platelet estimates with 4096 sensors resulted in MSEs of $1.73e-1$ (for $\sigma^2 = 1$) and $5.18e-1$ (for $\sigma^2 = 10$). In row 3, approximate platelet estimates with 4096 sensors resulted in MSEs of $2.34e-1$ (for $\sigma^2 = 1$) and $9.75e-1$ (for $\sigma^2 = 10$).

V. CONCLUSION

In this paper, we have proposed a method for boundary estimation in sensor networks. The boundary estimate is determined via complexity regularization of a hierarchical tree-based estimation method. We demonstrated theoretically that our method nearly achieves the optimal tradeoff $\text{MSE} \propto 1/\mathcal{E}^{2\nu}$, which shows that no other scheme can be devised that will (asymptotically) perform significantly better under the stated assumptions. Simulation experiments agreed very well with the theoretical predictions.

APPENDIX I

ESTIMATOR PENALIZATION

We penalize the platelet estimates according to the length of a prefix code which can uniquely describe each possible platelet estimate (i.e., codes which satisfy the Kraft inequality). These codelengths are proportional to the size the partition associated with each model and, thus, penalization leads to estimates that favor smaller partitions. In particular

$$g^p(n) \equiv 2/3 \log_e 2 + 4/3 \log_e n. \quad (9)$$

To see how this is derived, consider constructing a unique code for every $\tilde{f} \in \mathcal{F}_n^p$. If \tilde{f} has k leafs, then k can be decomposed into $3m + 1$ square-shaped leafs plus some number of wedge-shaped leafs, where m is the number of dyadic splits in the partition and the number of wedge-shaped leafs is bounded above by $3m + 1$. The total number of k leafs, then, is bounded above by $2(3m + 1)$. The quadtree representing the RDP of the data can be encoded by representing each of the $4m + 1 < 2k/3$ nodes with a 0 for an internal node and a 1 for a leaf node. This can easily be verified with an inductive argument. Each of these dyadic squares can then be split into two wedge-shaped blocks.

We can allow as few as $n^{2/3}$ possible wedgelet splits per dyadic block without affecting the error decay rate. The splits can then be encoded if we allot $1/3 \log_2 n$ bits to each wedge-shaped block. Finally, the three coefficients of a planar fit on one of the wedge-shaped blocks can be uniformly quantized to one of $n^{1/3}$ levels without changing the error decay rate, and so all three coefficients can be encoded with a total of $\log_2 n$ bits. Thus, the necessary codelength for an estimate with k leafs in the partition requires $k(2/3 + 4/3 \log_2 n)$ bits.

From the Kraft inequality, we know that the existence of this uniquely decodable scheme guarantees that

$$\sum_{\tilde{f} \in \mathcal{F}_n^p} 2^{-k(2/3 + 4/3 \log_2 n)} \leq 1.$$

Therefore, if $g^p(n) = k(2/3 \log_e 2 + 4/3 \log_e n)$, then

$$\begin{aligned} \sum_{\tilde{f} \in \mathcal{F}_n^p} e^{-g^p(n)} &= \sum_{\tilde{f} \in \mathcal{F}_n^p} e^{-\log_e(2)k(2/3 + 4/3 \log_2 n)} \\ &= \sum_{\tilde{f} \in \mathcal{F}_n^p} 2^{-k(2/3 + 4/3 \log_2 n)} \\ &\leq 1 \end{aligned}$$

as desired.

By a similar argument, Haar estimates should be penalized as

$$g^h(n) \equiv 4/3 \log_e 2 + 1/4 \log_e n. \quad (10)$$

The analysis follows as in the platelet case, except that: 1) the wedgelet locations do not need to be encoded; 2) only one Haar coefficient is needed in place of the three platelet coefficients; and 3) $n^{1/4}$ quantization levels are required to guarantee near-optimal error decay rates. This means that the necessary codelength for an estimate with k leafs in the partition requires $k(4/3 + 1/4 \log_2 n)$ bits, which results in the above penalization.

APPENDIX II

RATE OF MSE DECAY

First, consider the case for the Haar estimator \hat{f}^h . Consider a complete RDP with M^2 squares of sidelength $1/M$. It is known that if the boundary is a Lipschitz function, or more generally has a box counting dimension of 1 (the case for both classes $C1$ and $C2$), then the boundary passes through $\ell \leq CM$ of the squares, for some constant $C > 0$ [2], [15]. Furthermore, there exists a pruned RDP with at most $C'M$ leafs, where $C' = 8(C + 2)$, that includes the above ℓ squares of sidelength $1/M$ that contain the boundary [2], [15].

Now, consider the upper bound (8), which as stated earlier follows as from an application of Theorem 7 in [5]

$$\begin{aligned} &\frac{1}{n} \sum_{i,j=1}^{\sqrt{n}} E \left[\left(\hat{f}_{i,j}^h - f_{i,j} \right)^2 \right] \\ &\leq \min_{\tilde{f} \in \mathcal{F}_n^h} \frac{1}{n} \left\{ 2 \sum_{i,j=1}^{\sqrt{n}} (\tilde{f}_{i,j} - f_{i,j})^2 + 8\sigma^2 g^h(n) |\tilde{f}| \right\} \\ &\leq \min_{\tilde{f} \in \mathcal{F}_n^h} 2 \int_{[0,1]^2} (\tilde{f} - f)^2 + 8\sigma^2 \frac{1}{4} \frac{\log n}{n} C' M \end{aligned}$$

where the discretized squared error is bounded by the corresponding continuous counterpart. The squared error $\int_{[0,1]^2} (\hat{f} - f)^2 \leq (K_1/M) + (K_2/\sqrt{n})$, where the first term is due to the error between the $1/M$ resolution partition along the boundary, and the $1/\sqrt{n}$ term is due to the quantization error overall. Thus, the MSE behaves like

$$\text{MSE} = O(1/M) + O(1/\sqrt{n}) + O\left(M \frac{\log n}{n}\right).$$

Taking $M = \sqrt{n/(\log n)}$ produces the desired result: $\text{MSE} = O(\sqrt{\log n/n})$.

The platelet case is handled in an analogous manner. The only significant change is that the squared error $\int_{[0,1]^2} (\hat{f} - f)^2 \leq (K_1/M^2) + (K_2/n^{2/3})$, where the first term is due to the error between the plate approximation in the smooth regions and the wedge approximation to the boundary, and the $n^{-2/3}$ term is due to the quantization error overall. Taking $M = \sqrt[3]{n/(\log n)}$ produces the desired result: $\text{MSE} = O((\log n/n)^{2/3})$.

REFERENCES

- [1] A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction*. New York: Springer-Verlag, 1993.
- [2] D. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, pp. 859–897, 1999.
- [3] E. J. Candes and D. L. Donoho *et al.*, "Curvelets—A surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surfaces*, I. L. Schumaker *et al.*, Eds. Nashville, TN: Vanderbilt Univ. Press, 1999.
- [4] R. Willett, A. Martin, and R. Nowak, "Backcasting: A new approach to energy conservation in sensor networks," presented at the 3rd. Int. Symp. Information Processing Sensor Networks (IPSN'04), Berkeley, CA, Apr. 2004.
- [5] E. Kolaczyk and R. Nowak, (2002) Multiscale likelihood analysis and complexity penalized estimation. *Annals Statistics* [Online]. Available: www.ece.wisc.edu/nowak/pubs.html
- [6] Q. Li and A. Barron, "Mixture density estimation," in *Advances in Neural Information Processing Systems 12*, S.olla, T. Leen, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2000.
- [7] R. Nowak and U. Mitra, "Boundary estimation in sensor networks: Theory and methods," in *Information Processing in Sensor Networks*. New York: Springer-Verlag, 2003, pp. 80–95.
- [8] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," presented at the 8th ACM Int. Conf. Mobile Computing and Networking (Mobicom), Atlanta, GA, 2002, pp. 140–147.
- [9] D. Marco, E. Duarte-Melo, M. Liu, and D. Neuhoff, "On the many-to-one transport capacity of dense wireless sensor networks and the compressibility of its data," in *Information Processing in Sensor Networks*. New York: Springer-Verlag, 2003, pp. 1–16.
- [10] S. D. Servetto, "Sensing Lena—Massively distributed compression of sensor images," in *Proc. IEEE Int. Conf. Image Processing (ICIP'03)*, Sept. 2003, pp. 613–616.
- [11] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (discus)," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.
- [12] K. Chintalapudi and R. Govindan, "Localized edge detection in sensor fields," *Ad-Hoc Networks J.*, vol. 1, pp. 273–291, Sept. 2003.
- [13] D. Devaguptapu and B. Krishnamachari, "Applications of localized image processing techniques in wireless sensor networks," in *Proc. SPIE Aerospace'03*, Orlando, FL, Apr. 2003, pp. 247–256.
- [14] D. Ganesan, B. Greenstein, D. Perelyubskiy, D. Estrin, and J. Heidemann, "An evaluation of multi-resolution search and storage in resource-constrained sensor networks," in *Proc. 1st ACM Conf. Embedded Networked Sensor Systems (SenSys)*, Los Angeles, CA, 2003, pp. 89–102.
- [15] R. Willett and R. Nowak, "Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, pp. 332–350, Mar. 2003.
- [16] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1983.

Robert Nowak (S'88–M'89–SM'04) received the B.S. (with highest distinction), M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison, in 1990, 1992, and 1995, respectively.

He spent several summers with General Electric Medical Systems' Applied Science Laboratory, Waukesha, WI, where he received the General Electric Genius of Invention Award and a U.S. patent for his work in 3-D computed tomography. From 1996 to 1999, he was an Assistant Professor at Michigan State University, East Lansing. From 1999 to 2003, he held Assistant and Associate Professor positions at Rice University, Houston, TX. He also held a visiting position at INRIA, Sophia-Antipolis, France, in 2001. He is now an Associate Professor at the University of Wisconsin-Madison. His research interests include statistical signal and image processing, wavelets and multiscale analysis, computational and applied mathematics, and applications in biomedicine, learning systems, and communication networks.

Dr. Nowak received the National Science Foundation CAREER Award in 1997, the Army Research Office Young Investigator Program Award in 1999, the Office of Naval Research Young Investigator Program Award in 2000, and the IEEE Signal Processing Society Young Author Best Paper Award in 2000.

Urbashi Mitra (S'88–M'88–SM'04) received the B.S. and M.S. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1987 and 1989, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1994.

From 1989 to 1990, she worked as a Member of Technical Staff at Bellcore, Red Bank, NJ. From 1994 to 2000, she was an Assistant Professor in the Department of Electrical Engineering, The Ohio State University (OSU), Columbus. She has been an Associate Professor in the Department of Electrical Engineering, University of Southern California, Los Angeles, since 2000. She has had visiting appointments at the Eurecom Institute, Sophia Antipolis, France, Rice University, Houston, TX, and Stanford University, Stanford, CA.

Dr. Mitra is the recipient of the Texas Instruments Visiting Professor Fellowship, the 2001 Okawa Foundation Award, the 2000 Lumley Award for Research (OSU College of Engineering), the 1997 MacQuigg Award for Teaching (OSU College of Engineering), the 1996 National Science Foundation (NSF) CAREER Award, the 1994 NSF International Postdoctoral Fellowship, the 1998 Lockheed Leadership Fellowship, and the 1987 California Microelectronics Fellowship. She Co-Chaired the IEEE Communication Theory Symposium at ICC 2003 in Anchorage, AK, with S. McLaughlin, Georgia Institute of Technology, Atlanta. She was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1996 to 2001. She is currently a Member of the IEEE Information Theory Board of Governors.

Rebecca Willett (S'01) received the B.S.E. degree (with highest distinction) in electrical and computer engineering and computer science from Duke University, Durham, NC, in 2000 and the M.S. degree in electrical and computer engineering from Rice University, Houston, TX, in 2002. She is currently working toward the Ph.D. degree in electrical engineering from Rice University.

She has held visiting positions with General Electric Medical Systems' Applied Science Laboratory, Waukesha, WI, in 2002 and INRIA, Sophia-Antipolis, France, in 2003. She is an Honorary Fellow at the University of Wisconsin-Madison. Her research interests include statistical signal and image processing, wavelets and multiscale analysis, machine learning, and applications in biomedicine, and astronomy.

Ms. Willett received a National Science Foundation Graduate Research Fellowship in 2000.