# Backcasting: Adaptive Sampling for Sensor Networks

Rebecca Willett, Aline Martin, and Robert Nowak *

**Abstract**

Wireless sensor networks provide an attractive approach to spatially monitoring environments. Wireless technology makes these systems relatively flexible, but also places heavy demands on energy consumption for communications. This raises a fundamental trade-off: using higher densities of sensors provides more measurements, higher resolution and better accuracy, but requires more communications and processing. This paper proposes a new approach, called "backcasting," which can significantly reduce communications and energy consumption while maintaining high accuracy. Backcasting operates by first having a small subset of the wireless sensors communicate their information to a fusion center. This provides an initial estimate of the environment being sensed, and guides the allocation of additional network resources. Specifically, the fusion center backcasts information based on the initial estimate to the network at large, selectively activating additional sensor nodes in order to achieve a target error level. The key idea is that the initial estimate can detect correlations in the environment, indicating that many sensors may not need to be activated by the fusion center. Thus, adaptive sampling can save energy compared to dense, non-adaptive sampling. This method is theoretically analyzed in the context of field estimation and it is shown that the energy savings can be quite significant compared to conventional approaches. For example, when sensing a piecewise smooth field with an array of $100 \times 100$ sensors, adaptive sampling can reduce the energy consumption by roughly a factor of $10$ while providing the same accuracy achievable if all sensors were activated.

## 1 Adaptive Sampling

Wireless networks of spatially distributed sensors are emerging as a fundamental new tool for monitoring environments. The longevity of such networks depends crucially on the wise and frugal use of energy. High spatial densities of sensors are desirable for achieving high resolution and accurate estimates of the environmental conditions, but high densities also place heavy demands on bandwidth and energy consumption for communication. This paper proposes an adaptive two-step approach that can significantly reduce energy consumption while maintaining high accuracy. In the first step, called the *preview* step, an initial estimate of the environment is formed using a subset of the sensor nodes. Based on the initial estimate, additional sensors are selectively activated in the second step, called the *refinement* step, to improve the accuracy of the preview estimate. The key idea is that the initial estimate can detect correlations in the environment, indicating that many sensors may not be required to achieve a desired level of accuracy. Thus, adaptive sampling can save energy compared to dense, non-adaptive sampling, and can be used to calculate accurate field estimates while only activating a fraction of the available sensors. We refer to the overall process as backcasting to emphasize the role of feedback in network resource allocation.

We propose a general technique that could prove useful in a variety of sensor network applications. The practicality and benefits of adaptive sampling will depend on the specific application. The basic idea

*R. Nowak (nowak@engr.wisc.edu) was supported by the National Science Foundation, grants CCR-0310889 and ANI-0099148, the Office of Naval Research grant N00014-00-1-0390, and the State of Texas ATP, grant 003604-0064-2001. R. Willett (willett@rice.edu), A. Martin (alinemartin@wisc.edu), and R. Nowak are with the Departments of Electrical and Computer Engineering at the University of Wisconsin-Madison and Rice University.

might be implemented as follows. Consider a sensor network with $n$ sensors distributed uniformly over a square meter region. The sensors collect data reflective of environmental conditions in a local area about each sensor. At one extreme, the environmental conditions are constant across the region, and at the other, the conditions vary rapidly from point to point. Assume that each measurement contains a small amount of zero-mean noise. Given all $n$ data, the best estimate of the field has an expected mean squared error (MSE) that behaves like $n^{-\nu}$ for some $0 < \nu \leq 1$. For example, the best case, $\nu = 1$, occurs when the environmental conditions are constant (apart from random noise). In practice, one doesn't know whether the environmental conditions are nearly constant or rapidly varying, so it is difficult to predict how the MSE will behave. However, suppose that a subset of sensors, say $pn$, $0 < p < 1$, is called upon to make an initial estimate. Information from these sensors is transmitted to a fusion center, at a lower energy and bandwidth cost than transmitting from all $n$ sensors. The MSE of this estimate will behave like $(pn)^{-\nu}$. This estimate might be "good enough" if $\nu$ is close to one and $p$ isn't too small. On the other hand, since $\nu$ is unknown *a priori* and because the field may be inhomogeneous, the fusion center may determine that more information is needed from additional sensor nodes. To obtain additional information, the fusion center can backcast information about the initial estimate to the network. Based on this information, additional sensors in the network may then be activated as necessary and participate in the process to improve the estimate of the environment being sensed. For example, the information transmitted from the fusion center could be the initial estimate or it may simply be a "wake-up" signal to certain additional sensors that will then activate and forward their information. This procedure can reduce the total number of communications and the number of sensors that are activated, thereby reducing the energy consumption of the network.

To demonstrate the considerable potential of adaptive sampling, we focus on one well-studied application, the estimation of a spatial field. In previous work, we proposed and developed a field estimation algorithm based on multiscale partitioning methods [1, 2]. The algorithms are quite practical and map nicely onto a sensor network architecture. Moreover, we demonstrated theoretically that our methods nearly achieve the best-case MSE/Energy trade-offs. However, there was an implicit assumption in our earlier results: only one-way communication from the sensor nodes to a desired destination was allowed. If one changes that crucial assumption by allowing for two-way communication (to and from the fusion center), then the basic trade-offs between the accuracy and energy consumption change in a dramatic way. In fact, it is possible to significantly decrease the total and per-sensor energy consumption of the network while maintaining a low MSE using adaptive sampling.

## 2 Wireless Sensing of 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 trade-off 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 is smooth or piecewise smooth. That is, the field varies smoothly from point to point, with the possible exception of a boundary or edge between two smooth regions (like the case depicted in Figure 1). Each sensor makes a measurement of the field at its location which is contaminated with a zero-mean Gaussian noise. The error can model a variety of potential uncertainties including sensor noise, small environmental disturbances, and quantization effects. It is known that under these assumptions the expected MSE cannot, in general, decay faster than $O(n^{-\nu})$, for some $0 < \nu \leq 1$ that depends on the smoothness of the regions and whether or not boundaries are present in the field [3–5]. That is, no estimator (based on centralized or distributed processing) can exceed this convergence speed-limit.

To quantify the energy required to compute and transmit a field estimate of this accuracy, we assume a simple multihop communication protocol. We calculate the total energy consumption $\mathcal{E}(n)$ according to:

$$\mathcal{E}(n) \;=\; b(n) \times h(n) \times e(n), \tag{1}$$

where $b(n)$ is the number of bits that must be transmitted, $h(n)$ is the average number of hops each bit must traverse, and $e(n)$ is the energy required to send one bit over one hop. Note all quantities depend on $n$, the number of sensors in the network. As the density of sensors increases, both $b(n)$ and $h(n)$ generally increase, while $e(n)$ tends to decrease due to the decrease in distance between sensor nodes. According to this measure of energy, even simple calculations require communications that consume at least $O(e(n) \cdot n)$ units of energy. For example, consider computing the average values between the nearest neighboring sensors in the network. This requires the transmission of $O(n)$ bits, each over one hop. Such a trivial operation appears to be necessary for almost any imaginable field estimation process. Thus, the total energy required to compute and transmit the field estimate is at least $O(e(n) \cdot n)$, or $O(e(n))$ units of energy per sensor. Note that, here and throughout the paper, we assume that local processing and computation at each sensor requires a negligible amount of energy in comparison to sensor activation and communications.

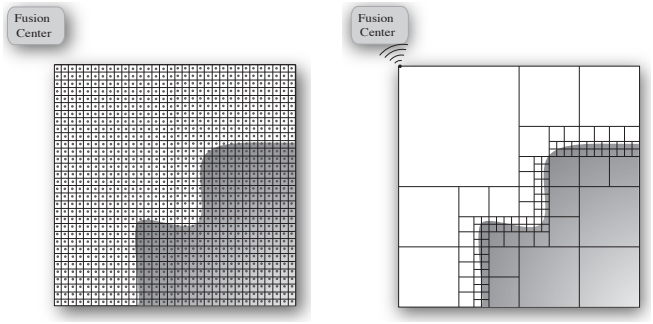
The expected MSE decay is at most $D(n) = O(n^{-\nu})$ and the energy consumption is at least $\mathcal{E}(n) = O(e(n) \cdot n)$. In our earlier work [1, 2] we showed that hierarchical, multiscale algorithms can nearly achieve the best MSE using minimal energy. These algorithms employed only one-way multihop communication from the sensor nodes to a fusion center. In this paper, it is shown that the proposed two-step estimation method can be used to improve on our earlier results. Specifically, it is possible to nearly achieve the best MSE, and the expected energy consumption is reduced to $O(e(n) \cdot n^{3/4})$, or $O(e(n) \cdot n^{-1/4})$ per sensor. This is possible through data-adaptive sampling of the field, as opposed to dense, non-adaptive sampling. The reduction in energy requirements can significantly extend the lifetime of a sensor network. For example, when sensing a piecewise smooth field with an array of $100 \times 100$ sensors, adaptive sampling can reduce the energy consumption by roughly a factor of $10$ while providing the same accuracy achievable if all sensors were activated.

## 3 Hierarchical Field Estimation

In our previous analysis of the accuracy-energy tradeoff in sensor networks, we proposed a hierarchical approach to estimation and communication based on platelets [1]. Our adaptive sampling approach will also be based on platelet-based estimation, and therefore we review it here. The sensor measurements can be viewed as sampling the field over a partition of $n$ nested sub-squares of sidelength $1/\sqrt{n}$, as shown in Figure 1(a). In principle, this initial partition can be generated by a *recursive dyadic partition* (RDP) as follows. First divide the domain into four sub-squares of equal size. Repeat this process again on each sub-square. Repeat this $1/2 \log_2 n$ times. This gives rise to a *complete* RDP of resolution $1/\sqrt{n}$ (the square partition of the sensing domain shown in Figure 1(a)). The RDP process can be represented with a quadtree structure. The quadtree can be pruned back to produce an RDP with non-uniform resolution, as shown in Figure 1(b).

For each RDP, there is an associated quadtree structure (generally of non-uniform depth corresponding to the non-uniform resolution of most RDPs). The leafs of each quadtree represent dyadic (sidelength equal to a 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 sensors in each square serves as a "clusterhead," which will assimilate information from the other sensors in the square, as shown in Figure 2.

Let $P$ denote a certain RDP and define the estimator of the field on each square of the partition to be the least-squares fit of a planar model (*e.g.,* platelet) to the measurements in the square. With this in mind,




(a) (b)

Figure 1: Recursive dyadic partitions. (a) A wireless sensor network sampling a two-dimensional field and a complete recursive dyadic partition of the field. Dots indicate sensor locations and squares indicate the extent of each sensor's measurement. (b) A pruned recursive dyadic partition of the field in (a).

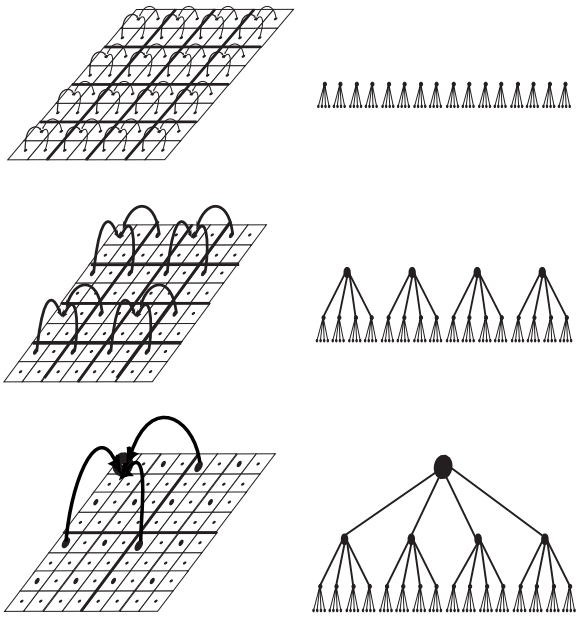


Figure 2: Hierarchical communication in a wireless sensor network. On the top, the first level of communication involves clusters of four individual sensors transmitting their measurements to their clusterhead. In the center, the next level of communication is depicted, in which each clusterhead from the previous level communicates with its clusterhead (using a multihop protocol). On the bottom, the final level of communication is depicted, in which the clusterheads from the previous level send their estimates to a fusion center (upper left-hand corner). To the right of each field, the corresponding tree is depicted.

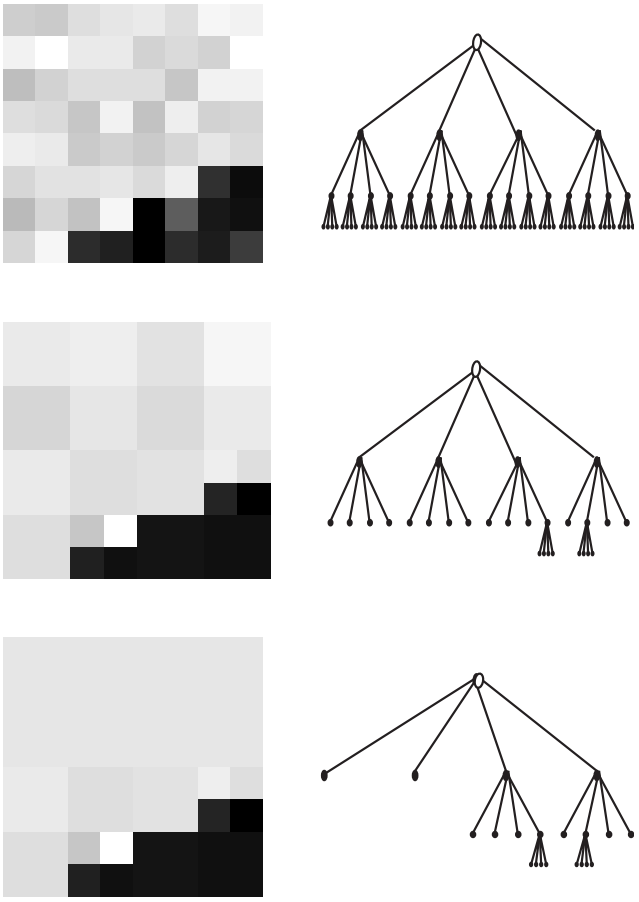


Figure 3: Hierarchical field estimation. On the top, an array of noisy field measurements is shown, and to its right the corresponding unpruned RDP. In the center, the field estimate after one level of pruning is shown with its corresponding pruned tree. On the bottom, the final field estimate is show with the final tree.

a *complexity penalized* estimator is defined as follows. Let $\widetilde{f}(P)$ denote a model of the field (based on the least-squares platelet fits on each square of $P$). The empirical measure of performance is the sum-of-squared errors between $\widetilde{f}(P)$ and the data $x = \{x_{i,j}\}$:

$$R(\widetilde{f}(P), x) \;\; = \;\; \sum_{i,j=1}^{\sqrt{n}} \left( \widetilde{f}_{i,j}(P) - x_{i,j} \right)^2 \tag{2}$$

where $i, j$ denotes the location in the field. This measure is proportional to the negative log likelihood given our assumption of zero-mean Gaussian measurement error in each sensor; *i.e.,*

$$x_{i,j} \sim \mathcal{N}(f_{i,j}, \sigma^2),$$

where $f$ is the true field and $\sigma^2$ is the noise variance of each sensor measurement.

For fixed partition $P$, the choice of $\widetilde{f}(P)$ that minimizes $R(\widetilde{f}(P), x)$ is simply given by the least-squares fits on each square, as discussed above. Now define the complexity penalized estimator as

$$\widehat{f}_n \;=\; \arg \min_{\widetilde{f}(P): P \in \mathcal{P}_n} \; R(\widetilde{f}(P), x) \;+\; \sigma^2 (\log n) |P|, \tag{3}$$

where $\mathcal{P}$ denotes all possible partitions (prunings), $|P|$ denotes the number of squares in the partition $P$. It is well known that the optimization in (3) can be solved using a bottom-up tree pruning algorithm [4, 6, 7]. The hierarchy of clusterheads facilitates this process in the sensor network. At each level of the hierarchy, the clusterhead receives the fits 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 least-squares estimate on the larger square associated with the clusterhead. An example of the estimation and pruning process is depicted in Figure 3.

The performance of the proposed estimators can be studied in terms of (minimax) lower bounds and upper bounds on the MSEs. Upper bounds on the error can be established using the Li-Barron bound [8] and Nowak and Kolaczyk's generalization of this bound [9]. Specifically,

$$\text{MSE}(\widehat{f}_n, f) = \min_{\widetilde{f}(P): P \in \mathcal{P}_n} \frac{2}{n} R(\widetilde{f}(P), f) + \frac{4\sigma^2 |P| \log n}{n}, \tag{4}$$

where the first term is the approximation error and the second term is the estimation error.

In this paper, we focus on one class of piecewise smooth fields. We assume that the field is twice-continuously differentiable everywhere except near possible boundaries (or "edges") where the field changes sharply. Boundaries, if they indeed are present, are assumed to be 1-d curves with box-counting dimension one in the 2-d plane of the field. For this class of functions, it is known that the minimax MSE rate is bounded below by $O(n^{-1/2})$ in the presences of a boundary and $O(n^{-2/3})$ in the absence of a boundary. If it is assumed that the boundary is present and a $C^2$ curve, then the minimax MSE rate is bounded below by $O(n^{-2/3})$. It can be shown that solving the optimization in (3) yields a partition which balances the approximation error and estimation error terms in the bound on the MSE in (4), resulting in a final bound of $O((\log n/n)^{-1/2})$ in the presence of a boundary and $O((\log n/n)^{-2/3})$ in the absence of a boundary. The error decays more slowly in the presence of a boundary because the $O(\sqrt{n})$ sensors near the boundary cannot aggregate their measurements with those of their neighbors.

The energy required to form $\widehat{f}_n$ can be studied in terms of how the number of sensors affects per sensor energy consumption. To illustrate, let us begin with the most simple case, a constant field. Since the field is constant, with high probability the measurements will be aggregated at each level of the hierarchy [2]. This means that at scale $j$ (where $j = 0$ is the finest scale) $O(n/4^j)$ bits must be transmitted to the clusterheads

at the next scale. Because clusterheads are farther apart at coarser scales, each such transmission would require $O(2^j)$ hops. Therefore, the total energy required for the estimation process is

$$\begin{aligned}\mathcal{E} &= O\left(e(n)\sum_{j=0}^{\log_4 n-1} 2^j n/4^j\right)\\ &= O\left(e(n)\cdot n\right).\end{aligned}$$

Now consider the more complex case of a piecewise smooth field. If the field being estimated consists of twice-continuously differentiable (denoted $C^2$) smooth regions separated by boundaries, then at scale $j$, $O(n/4^j + \sqrt{n})$ bits need to be transmitted up the hierarchy. The $\sqrt{n}$ term results from the fact that $O(\sqrt{n})$ sensors sense the boundaries and therefore cannot aggregate their measurements with their neighbors. Thus these boundary measurements need to be passed up the hierarchy at every level. Each transmission requires $O(2^j)$ hops, and so the total energy required for this estimation is again

$$\begin{aligned}\mathcal{E} &= O\left(e(n)\sum_{j=0}^{\log_4 n-1} 2^j (n/4^j + \sqrt{n})\right)\\ &= O\left(e(n)\cdot (n + 2^{\log_4 n}\sqrt{n})\right)\\ &= O\left(e(n)\cdot n\right).\end{aligned}$$

The analysis above shows that the communication energy cost of our previous algorithms is roughly the same, whether the field is very simple or relatively complex in nature. Adaptive sampling, as we will show, allows us to achieve the same high accuracy, but reduces the energy cost in two ways. First, by adaptively sampling the field, fewer sensors are activated (sensor activation can be a huge energy drain; keeping sensors asleep is highly desirable). Second, since fewer measurements are collected, the number of communications required to achieve an accurate estimate is reduced.

# 4 Backcasting

Detailed derivations of the theoretical results described in this section are found in the Appendix. We assume the field being sensed is supported on one square meter (referred to as the unit square), and that a total of $n$ wireless sensors have been deployed uniformly over the square (but not necessarily activated). The steps involved in each stage are detailed below:

**Preview:** The goal of this stage is to provide a coarse estimate of the field being sensed. Specifically, the field is estimated up to a resolution of $n^{-1/2}$ square meters. The unit square is partitioned into $n^{1/2}$ "subsquares", each of sidelength $n^{-1/4}$. In each subsquare, $n^{1/4}$ sensors (of the total $n^{1/2}$ available in the subsquare) are activated, one of which is designated as the "clusterhead" for the subsquare. This provides $n^{1/4}$ measurements per subsquare, as demonstrated in Figure 4(a). Also, the activated sensors can be aligned, spaced about $n^{-1/2}$ meters apart, to form a multihop communication path from one side of the subsquare to another. Thus, $n^{3/4}$ "preview" sensors are activated in total. These sensors are used to form an initial estimate of the field using a hierarchical platelet estimation method described above, as shown in Figure 4(b). The estimate generated by these preview sensors is passed to the fusion center. With high probability, this partition will be composed of small squares in regions close to any boundary present in the field and larger squares in smoother regions of the field. This stage requires $O(e(n)\cdot n^{3/4})$ units of energy, and $O(n^{3/4})$ sensors must be on to facilitate the multihop communications.

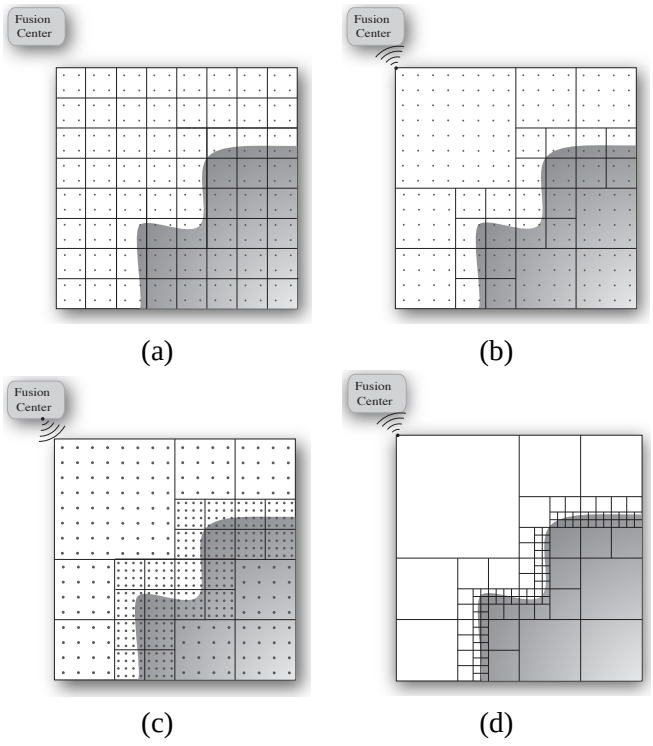



Figure 4: Backcasting procedure. (a) Initial, low resolution RDP. (b) Preview. (c) Adaptive sensor activation. (d) Estimation refinement.

**Refinement:** The fusion center sends an "activation" message to the preview clusterheads in only the smallest squares of the preview partition (*i.e.*, those remaining subsquares that were not aggregated due to the presence of a boundary). This communication requires $O(e(n) \cdot n^{3/4})$ units of energy. Each such clusterhead then activates all $n^{1/2}$ sensors within its subsquare in order to fine-tune the partition in boundary-containing regions, as shown in Figure 4(c). The activation requires $O(e(n) \cdot n^{1/2})$ energy per subsquare. Assuming that the boundary is a $C^2$ curve, only $O(n^{1/4})$ preview subsquares will contain the boundary and require fine-tuning. Thus, the total energy consumed adaptively activating more sensors is $O(e(n) \cdot n^{3/4})$.

Each of the $O(n^{1/4})$ smallest preview subsquares now contains $n^{1/2}$ activated sensors sampling the underlying field. These sensors are used to generate a refined estimate of the field in these subsquares, using a local recursive dyadic partitioning scheme, as shown in Figure 4(d). This requires a total $O(e(n) \cdot n^{3/4})$ units of energy. The clusterheads must then communicate the refined estimates to the fusion center, which requires $O(e(n) \cdot n)$ units of energy using multihop communications via the $n^{3/4}$ sensors activated in the preview step. The energy requirement for this last step can be significantly reduced when the boundary is smooth. Specifically, instead of transmitting the full fine-tuned partition to the fusion center, it is possible to simply fit a line to the boundary estimate and transmit the boundary and platelet coefficients. This reduces the energy expenditure to $O(e(n) \cdot n^{3/4})$ units, since only a few parameters must be communicated from the clusterheads to the fusion center.

The following theorem summarizes the performance of the proposed method.

**Theorem:** *Assume there is a wireless sensor network of $n$ sensors arranged on a uniform grid over a field composed of smooth (twice continuously differentiable) regions separated by (twice continuously differentiable) boundaries. Then, using the hierarchical adaptive sampling method described above, the field can*

| Field | $n$ | MSE$_n$ | MSE$_p$ | MSE$_r$ | Energy Savings |
|---|---|---|---|---|---|
| Boundary | $256 \times 256$ | $6.26 \times 10^{-4}$ | $6.06 \times 10^{-3}$ | $1.32 \times 10^{-3}$ | 0.2031 |
| Boundary | $64 \times 64$ | $1.09 \times 10^{-3}$ | $1.47 \times 10^{-2}$ | $2.48 \times 10^{-3}$ | 0.6875 |
| No Boundary | $256 \times 256$ | $3.95 \times 10^{-5}$ | $1.94 \times 10^{-4}$ | $1.94 \times 10^{-4}$ | 0.0625 |
| No Boundary | $64 \times 64$ | $2.17 \times 10^{-4}$ | $2.17 \times 10^{-4}$ | $2.17 \times 10^{-4}$ | 0.3125 |

Table 1: Error and energy savings for wireless sensor networks of various sizes with and without adaptive sampling.

*be estimated with an MSE of* $D(n) = O((\log^2 n/n)^{1/2})$ *using an average of* $\mathcal{E}(n) = O(e(n) \cdot n^{3/4})$ *units of energy.*
This result is derived in the Appendix.

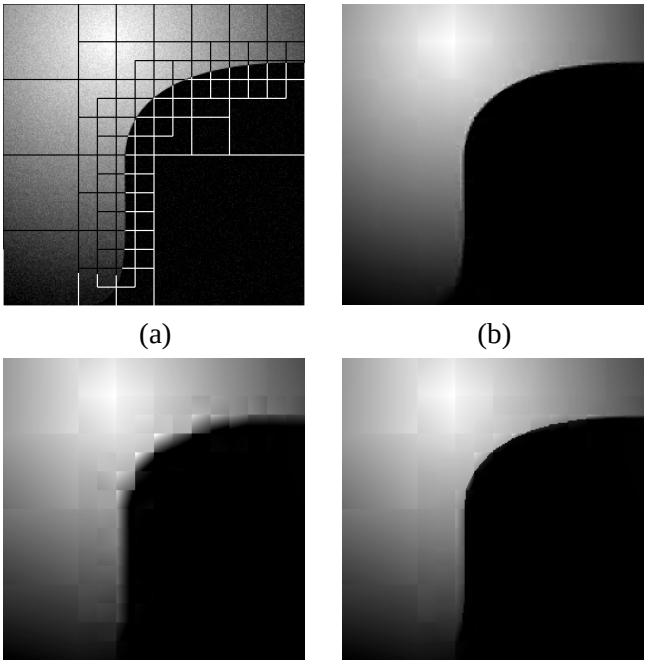
The MSE above is the same as that which we achieved for a piecewise smooth field using all $n$ sensors and $O(e(n) \cdot n)$ units of energy [2]. Thus, using adaptive sampling we reduce the amount of energy required by a factor of $n^{1/4}$.

This has important implications for the deployment of a practical system. Using the proposed method, a system developer should be able to determine how many sensors to engage in order to achieve a desired accuracy. It is known in advance that the presence of a boundary will dominate the error, but the location of the boundary is unknown. Thus the sensors should be distributed evenly across the field. If no boundary is present, then the subset of sensors activated in the preview stage is sufficient to achieve the target error. If a boundary is present, then the preview stage will determine the approximate location of the boundary and adaptively activate more sensors in order to achieve the target error. To illustrate the significance of this gain, we consider estimating a field with $n = 10,000$ sensors. Then the total energy expenditure with adaptive sampling is $O(e(n) \cdot 10^3)$ energy units instead of the $O(e(n) \cdot 10^4)$ energy units required without adaptive sampling. Similar gains are seen in terms of activation energy. Thus, a system of $10,000$ sensors, each equipped with batteries to provide one year of continuous operation, could remain operational up to 10 years using adaptive sampling.

# 5 Experimental Results

We simulated the proposed method in four different situations presented below (Figure 5 through Figure 8). In each case, we show the preview partition superimposed over the noisy field, the preview estimate, and the final refined estimate. Also, for comparison we show the field estimate obtained using all $n$ sensors. The cases depicted in Figures 5 and 6 involve fields with a boundary. Notice that the visual quality and the MSEs of the adaptive sampling estimate are comparable to those obtained using all $n$ sensors. In Figures 7 and 8, the field does not contain a boundary and the preview and refined estimates are identical (no additional nodes are activated after the preview stage, as expected).

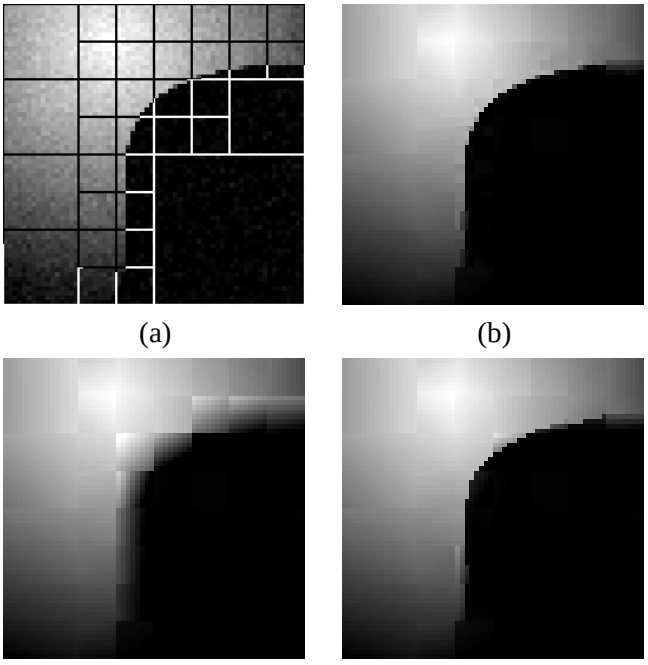
The MSEs for each case are presented Table 1. The MSE of the preview estimate is denoted MSE$_p$, the final refined estimate's MSE is denoted MSE$_r$, and the MSE of the estimator based on all $n$ sensors is denoted MSE$_n$. In the cases containing boundaries, we see that MSE$_r$ is reduced by roughly a factor of $O(n^{1/4})$ compared to MSE$_p$, as theoretically predicted in the appendix. This is due to the increase in spatial resolution near the boundary after the refinement step. An empirical calculation of the energy reduction, compared to activating all $n$ sensors, is also listed in Table 1. The energy reduction is simply the fraction of the number of activated sensors relative to the total number $n$.



(a) (b)

(c) (d)

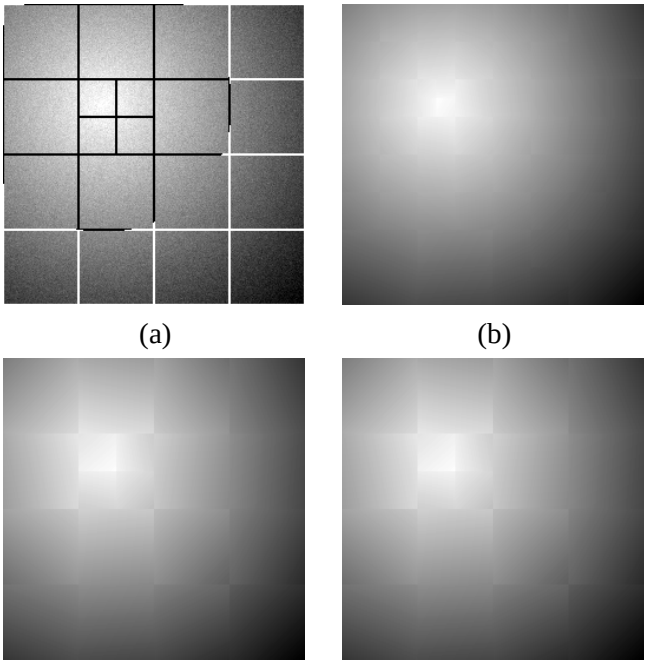
Figure 5: Estimation of a piecewise smooth field with $\sigma^2 = 1/10$ and $256 \times 256$ sensors. (a) Preview partition superimposed over noisy field. (b) Estimate using all $n$ sensors without adaptive sampling (MSE $= 6.26 \times 10^{-4}$). (c) Preview estimate (MSE $= 6.06 \times 10^{-3}$). (d) Final estimate (MSE $= 1.32 \times 10^{-3}$).



(a) (b)

(c) (d)

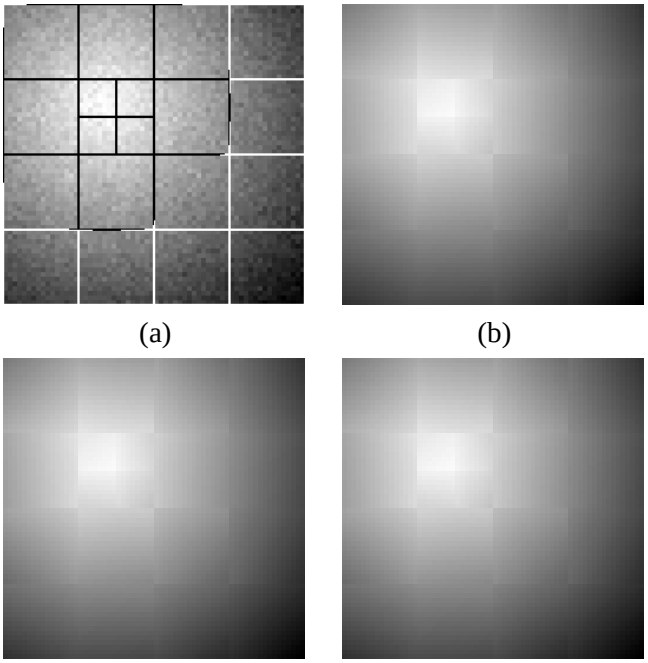
Figure 6: Estimation of a piecewise smooth field with $\sigma^2 = 1/10$ and $64 \times 64$ sensors. (a) Preview partition superimposed over noisy field. (b) Estimate using all $n$ sensors without adaptive sampling (MSE $= 1.09 \times 10^{-3}$). (c) Preview estimate (MSE $= 1.47 \times 10^{-2}$). (d) Final estimate (MSE $= 2.48 \times 10^{-3}$).



(a) (b)

(c) (d)

Figure 7: Estimation of a smooth field with $\sigma^2 = 1/10$ and $256 \times 256$ sensors. (a) Preview partition superimposed over noisy field. (b) Estimate using all $n$ sensors without adaptive sampling (MSE $= 3.95 \times 10^{-5}$). (c) Preview estimate (MSE $= 1.94 \times 10^{-4}$). (d) Final estimate (MSE $= 1.94 \times 10^{-4}$).



(a) (b)

(c) (d)

Figure 8: Estimation of a smooth field with $\sigma^2 = 1/10$ and $64 \times 64$ sensors. (a) Preview partition superimposed over noisy field. (b) Estimate using all $n$ sensors without adaptive sampling (MSE $= 2.17 \times 10^{-4}$). (c) Preview estimate (MSE $= 2.17 \times 10^{-4}$). (d) Final estimate (MSE $= 2.17 \times 10^{-4}$).

# 6 Conclusions

This paper proposes an adaptive approach to energy conservation in sensor networks. This method operates by first having a small subset of the wireless sensors communicate their information to a fusion center. This provides an initial estimate of the environment being sensed, and guides the allocation of additional network resources. The basic idea is potentially applicable to a wide variety of sensor networking problems, but to demonstrate its potential we focused on one particular case.

We analyzed the potential of adaptive sampling in the context of field estimation. We assume the field being sensed is supported on one square meter, and that a total of $n$ wireless sensors were deployed uniformly over the square. Adaptive sampling, in this case, involves a two step process: 1. $n^{3/4}$ sensors are activated to produce a coarse estimate of the field at the fusion center; 2. based on the coarse estimate, the fusion center determines which regions of the field may contain boundaries or sharply varying behavior, and activates up to $O(n^{3/4})$ additional sensors in those regions; the additional sensors provide a finer resolution estimates in those regions and the refined estimates are communicated to the fusion center. Theoretical analysis demonstrates that the final estimate has an MSE of $D(n) = O(n^{-1/2})$, and each stage requires $O(n^{3/4})$ communication hops. The MSE is of the same order that is achievable using all $n$ sensors and $O(n)$ communication hops. Thus, adaptive sampling provides the same level of overall accuracy, but requires a factor of roughly $n^{1/4}$ fewer activated nodes and communication hops. The resulting savings in energy consumption can be substantial. Simulated experiments confirm the theoretical predictions.

Note that in the proposed scheme, the sensors used in the preview stage must always remain activated and clusterheads expend more energy communicating than other sensors. This energy bottleneck potentially limits the lifetime of the system. The effect can be alleviated, however, by periodically changing which sensors are designated preview sensors and clusterheads and evenly distributing energy consumption among all the sensors in the network.

Also note that our assumption that the sensors lie on a uniform grid is merely a convenient mechanism for analysis and not a necessary component of the proposed method; it is sufficient to assume that locations of the sensors are distributed uniformly at random. This is relevant to many applications in which the sensors may be placed randomly throughout a region. In such scenarios, the network must incur an initial energy expenditure setting up the communication hierarchy, and it may be necessary to activate additional sensors in some regions to facilitate multihop communications, depending on the strength of the transmitters. However, sensor placement does not affect the overall performance of the network as the number of sensors grows large.

The error and energy analysis of each stage of the proposed procedure is detailed in this appendix. For the purpose of this discussion, we assume $n$ sensors arranged on a $\sqrt{n} \times \sqrt{n}$ uniform grid over a square meter, and that the field being sensed is a $C^2$-smooth surface which contains a boundary. (Estimation of a field not containing a boundary will be discussed at the conclusion of this appendix.) We assume that any boundary present is a $C^2$ curve, and thus would affect $O(n^{1/2})$ of the sensors. For simplicity of presentation, we omit log terms in the following analysis.

# A Preview

The goal of this stage is to provide a coarse estimate of the field being sensed. Specifically, the field is estimated up to a resolution of $n^{-\beta}$ square meters, for some $0 < \beta < 1$. (In the body of the paper, we used $\beta = 1/2$. We will demonstrate that this is the optimal choice later in this section.) The unit square is partitioned into $n^{\beta}$ smaller "subsquares", each of sidelength $n^{-\beta/2}$. In each subsquare, $n^{\gamma}$ ($\gamma \leq 1 - \beta$) sensors are activated, one of which is designated as the "clusterhead" for the subsquare. This provides $n^{\gamma}$ measurements per subsquare. Also, the activated sensors are aligned, spaced about $n^{-1/2}$ meters apart, to

form a multihop communication path from one side of the subsquare to another. Thus, $n^{\beta+\gamma}$ sensors are activated in this stage and are called *preview sensors*.

The hierarchical estimator described above will aggregate squares away from the boundary up to some level of the estimation determined by the smoothness of the surface. For example, measurements of a constant or linear field would be aggregated until the partition consisted of one large square, while measurements of a less smooth field would be aggregated until the underlying field is too nonlinear for aggregation. Measurements from sensors near the boundary, however, would seldom be aggregated with their neighbors and would continue to be transmitted up the hierarchy to the fusion center.

The approximation error of platelets on a $C^2$ surface decays like $1/m^2$, where $m$ is the number of platelets, and the estimation error decays like $m/n^{\beta+\gamma}$. Note that these errors correspond to the terms of the minimax upper bound on the MSE in (4), and that this bound is minimized by choosing $m \sim n^{(\beta+\gamma)/3}$; *i.e.,* the RDP associated with the preview sensors will be pruned in regions away from the boundary until approximately $m \sim n^{(\beta+\gamma)/3}$ regions remain in the partition, resulting in an MSE in these regions of $1/m^2 \sim n^{-2(\beta+\gamma)/3}$. Since the field estimate for these regions will not be updated in the refinement stage, we must choose $\beta$ and $\gamma$ such that $\beta + \gamma = 3/4$ to ensure that the associated MSE decays like $n^{-1/2}$. This rate only holds for regions away from the boundary; the error in boundary regions at this stage is higher. To see this, note that the boundary passes through $O(n^{\beta/2})$ subsquares, which cover a total area of $O(n^{-\beta/2})$, and that a platelet fit to the measurements in one of these subsquares exhibits a bias of $O(1)$ because of the probable presence of a boundary. Thus these regions contribute a total of $O(n^{-\beta/2})$ to the MSE at this stage, which dominates the error in the smooth regions of the field. The effect of the refinement stage is to drive the error in the boundary regions down to the target rate of $O(n^{-1/2})$, as detailed in the next section.

Note that it is possible that some subsquares near the boundary could be aggregated erroneously with their neighbors. For large enough $n$, an erroneously pruned subsquare is composed of two smooth regions, and because the subsquare was pruned, we know that the area of one of the two smooth regions is small with respect to the total area of the subsquare. However, worst-case analysis reveals that the bias induced by this area can force the error to decay more slowly than the desired $n^{-1/2}$. We propose a technique that overcomes this difficulty while requiring only minimal extra energy. Note that erroneous prunings occur when the boundary of the two regions is closely aligned with the uniform RDP just one level of resolution coarser than the unpruned preview RDP, *i.e.,* the uniform RDP containing $n^{\beta}/4$ subsquares. The basic idea is then to perform two preview stages – one as described above, and the other on a version of the partition which is shifted by one subsquare of length $n^{-\beta/2}$ in each coordinate. This ensures that the boundary is detected with high probability in one of the two preview stages. A cluster erroneously pruned in one of the preview stages would not be pruned in the other preview stage. In the refinement stage we refine any of the preview clusters that are left unpruned in either the first or second preview steps.

Next we bound the energy required to generate the preview partition. There are two components to the energy expenditure in the preview stage: (1) energy required to form initial platelet fits to the $n^{\gamma}$ measurements in each of the $n^{\beta}$ preview stage subsquares, (2) energy required to perform hierarchical platelet estimation using the initial $n^{\beta}$ platelet fits. To form the initial platelet fits in the preview stage subsquares, three sufficient statistics per subsquare must be calculated. Each sufficient statistic is a weighted average of the $n^{\gamma}$ measurements, and so all initial platelet fits can be computed with $O(e(n) \cdot n^{\gamma+\beta}) = O(e(n) \cdot n^{3/4})$ units of energy.

To analyze the energy consumed while performing hierarchical platelet estimation in the preview stage, let $j$ index the scale of the partition; a square at scale $j$ has sidelength $2^j n^{-\beta/2}$. Thus, $j = 0$ corresponds to the smallest possible square in the preview partition, which has a sidelength of $n^{-\beta/2}$, and $j = J \equiv \beta \log_4 n$ represents the largest square in the partition, *i.e.,* the entire field. At a given scale, $j$, some clusterheads fit a platelet to all the measurements in the cluster and only need to transmit three sufficient statistics to encode the platelet. Other clusterheads at scale $j$ cannot fit a platelet to the measurements without incurring excess error; in these cases the multiple-platelet estimate must be transmitted. While each clusterhead at

a given scale only needs to transmit either a single platelet fit or a multiple-platelet estimate, we assume for simplicity of presentation that both are transmitted at every scale to upper bound the required energy. At scale $j$, $O(n^\beta/4^j + n^{1/4} + n^{\beta/2})$ bits need to be transmitted to the next level in the hierarchy, where the first term is the number of possible aggregated squares, the second term is the number of regions in the pruned preview stage RDP in regions away from the boundary, and the third term is the number of boundary squares at scale $j$. (Recall from earlier in this section that the preview stage RDP is pruned to have $O(m) = O(n^{(\beta+\gamma)/3}) = O(n^{1/4})$ regions.)

Next note that these bits have to be transmitted a distance of one sidelength, $2^j n^{-\beta/2}$ meters. Each such transmission requires $2^j n^{-\beta/2}/n^{-1/2} = 2^j n^{(1-\beta)/2}$ hops. From here, the energy consumed in the preview stage, $\mathcal{E}_p$, can be expressed as

$$\begin{aligned}
\mathcal{E}_p = \; & O\Big(e(n)\cdot n^{3/4} \\
& \quad + e(n)\cdot \sum_{j=0}^{J-1} 2^j n^{(1-\beta)/2}\left(\frac{n^\beta}{4^j} + n^{1/4} + n^{\beta/2}\right)\Big) \\
= \; & O\Big(e(n)\cdot n^{3/4} \\
& \quad + e(n)\cdot n^{(1-\beta)/2}\sum_{j=0}^{J-1}\left(\frac{n^\beta}{2^j} + \frac{n^{3/4}}{2^{J-j}} + \frac{n^\beta}{2^{J-j}}\right)\Big) \\
= \; & O\left(e(n)\cdot\left(n^{3/4} + n^{(1+\beta)/2}\right)\right).
\end{aligned} \tag{5}$$

Note that for $\beta = 1/2$, $\mathcal{E}_p = O(e(n)\cdot n^{3/4})$; we will show later in the appendix that this is the optimal choice for $\beta$. Note that $\beta = 1/2$ implies $\gamma = 1/4$, which ensures enough active sensors in the preview stage to facilitate multihop communication.

# B Refinement

In this stage of the estimation procedure, the fusion center receives the preview estimate of the field from the top clusterhead and then determines which, if any, additional sensors need to be activated in order to better estimate the boundary. It is assumed that a boundary may pass through every square of sidelength $n^{-\beta/2}$ in the preview partition (i.e., those squares that were not aggregated in the preview stage). The fusion center then sends a message to the clusterhead in each such square to request additional measurements. As there are $O(n^{\beta/2})$ boundary squares in the preview partition, the message will initially contain $O(n^{\beta/2})$ bits which must be split into four and transmitted to the next four clusterheads in the hierarchy, which requires $O(1/2\; n^{1/2})$ hops. Each time the message is transmitted another level down in the hierarchy, the size of the message is quartered and the number of hops is halved. This results in a total energy expenditure of

$$\begin{aligned}
\mathcal{E}_{r_1} &= O\left(e(n)\cdot\sum_{j=0}^{J-1}\frac{n^{\beta/2}}{4^j}\frac{n^{1/2}}{2^j}\right) \\
&\le O\left(e(n)\cdot n^{(1+\beta)/2}\right).
\end{aligned} \tag{6}$$

Each of these clusterheads in the boundary squares then sends an activation signal to all $n^{1-\beta}$ sensors within its square. This activation process can also be performed in a coarse-to-fine fashion, analogous to the way the fusion center communicated to the clusterheads of the preview partition. Specifically, at the top level the wake-up message must travel to four sensors via $O(n^{(1-\beta)/2})$ hops. At each successive message, the

number of signals transmitted is quadrupled and the distance is halved, so at scale $j$, the energy expenditure for one subsquare is $e(n) \cdot 4^j/2^j \cdot n^{(1-\beta)/2}$. This must be done for each of the $O(n^{\beta/2})$ boundary squares. This results in a total energy expenditure of

$$\begin{aligned}\mathcal{E}_{r_2} &= O\left(e(n) \cdot n^{\beta/2} \sum_{j=0}^{(1-\beta)\log_4 n - 1} 2^j n^{(1-\beta)/2}\right) \\ &= O\left(e(n) \cdot n^{(1+\beta)/2}\right).\end{aligned}$$

Thus, the energy required for the activation signaling process is $O(e(n) \cdot n^{(1+\beta)/2})$. Since there are $O(n^{\beta/2})$ boundary squares in the preview partition, and $O(n^{1-\beta})$ sensors are activated in each one, a total of $O(n^{1-\beta/2})$ additional sensors are activated.

The per-sensor MSE for these squares decays like the target rate, $n^{-1/2}$. To see this, first note that in each of the $O(n^{\beta/2})$ subsquares being refined, the boundary passes near $O(n^{(1-\beta)/2})$ of the $O(n^{1-\beta})$ sensors. Each of these sensors near the boundary induces a bias of $O(1)$ and covers an area of $1/n$. Thus the total per-sensor MSE incurred by sensors near the boundary is $O(n^{\beta/2} \cdot n^{(1-\beta)/2} \cdot 1/n) = O(n^{-1/2})$. Next note that in each of the $O(n^{\beta/2})$ subsquares being refined, $O(n^{1-\beta})$ sensors are away from the boundary on a $C^2$ surface. As in the preview stage, the approximation error and estimation error are balanced by fitting $O(n^{(1-\beta)/3})$ platelets to these measurements, resulting in a per-sensor error of $O(n^{-2(1-\beta)/3})$ for this subsquare. Since there are $n^{1-\beta}$ sensors in the subsquare, and each covers an area of $1/n$, the total per-sensor MSE for the field incurred by these sensors is $O(n^{\beta/2}) \cdot O(n^{-2(1-\beta)/3}) \cdot n^{1-\beta} \cdot 1/n = O(n^{-2/3+\beta/6})$, which is $O(n^{-7/12})$, lower than the target rate of $n^{-1/2}$, for $\beta = 1/2$.

To calculate the energy required to refine the preview estimate of the field, first note that the refinement energy per square is proportional to the number of measurements taken in that square, as was the case in the preview stage. This step thus requires $O(e(n) \cdot n^{1-\beta})$ units of energy per boundary square, and since there are $O(n^{\beta/2})$ such squares,

$$\mathcal{E}_{r_3} = O\left(e(n) \cdot n^{1-\beta/2}\right). \tag{7}$$

These estimates are transmitted to the fusion center. Since each of the $O(n^{\beta/2})$ preview squares is refined to produced a partition with $O(n^{(1-\beta)/2})$ boundary regions and $O(n^{(1-\beta)/3})$ non-boundary regions, a total of $O(n^{(1-\beta)/2})$ bits per preview square need to be transmitted an average of $O(n^{1/2})$ hops for a total energy expenditure of $O(e(n) \cdot n)$. This can be significantly reduced by simply fitting a line to the boundary estimate generated by the partition and transmitting the boundary and platelet coefficients. Assuming the boundary is a $C^2$ curve, this approximation does not increase the error of the estimate. In this case, a total of $O(n^{\beta/2})$ bits need to be transmitted an average of $O(n^{1/2})$ hops for a total energy expenditure of

$$\mathcal{E}_{r_4} = O\left(e(n) \cdot n^{(1+\beta)/2}\right). \tag{8}$$

# C Discussion

Summing the energy from each stage, we obtain a total energy of

$$\begin{aligned}\mathcal{E} &= \mathcal{E}_p + \sum_{i=1}^{4} \mathcal{E}_{r_i} \\ &= O\left(e(n)\left(n^{3/4} + n^{(1+\beta)/2} + n^{1-\beta/2}\right)\right).\end{aligned} \tag{9}$$

Balancing terms, we find that $\beta = 1/2$ minimizes the total energy consumption, yielding a final energy of

$$\mathcal{E} = O\left(e(n) \cdot n^{3/4}\right). \tag{10}$$

Equation (10) tells us that if we set our target MSE decay rate to $n^{-1/2}$, then estimating the field using adaptive sampling will cost $O(e(n) \cdot n^{3/4})$ units of energy instead of the $O(e(n) \cdot n)$ units required by the hierarchical estimation method with non-adaptive sampling. Furthermore, note that only $O(n^{3/4})$ sensors needed to be activated for the entire procedure.

# D Absence of a Boundary

Note that the above analysis is also valid for the case where there is no boundary, and the sensor network is estimating a smoothly varying field. In this case, $\mathcal{E}_p$ would still be $O(e(n) \cdot n^{3/4})$. However, with high probability all of the resulting preview squares would be the result of measurement aggregation and hence the fusion center would not perform refinement. Thus the target error rate of $O(n^{-1/2})$ can be reached with $O(n^{3/4})$ sensors awake and $O(e(n) \cdot n^{3/4})$ units of energy.

# References


[1] R. Nowak, U. Mitra, and R. Willett. Estimating inhomogeneous fields using wireless sensor networks, 2003. To appear in *IEEE Journal on Selected Areas in Communications*.

[2] R. Nowak and U. Mitra. Boundary estimation in sensor networks: Theory and methods. In *2nd International Workshop on Information Processing in Sensor Networks*, volume 20, Palo Alto, CA, April 22-23 2003.

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

[4] D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Ann. Statist.*, 27:859 – 897, 1999.

[5] E. Candès and D. Donoho. Curvelets: A surprisingly effective nonadaptive representation for objects with edges, To appear in *Curves and Surfaces*, L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN.

[6] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1983.

[7] R. Willett and R. Nowak. Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *IEEE Transactions on Medical Imaging*, 22(3), 2003.

[8] Q. Li and A. Barron. *Advances in Neural Information Processing Systems 12*, chapter Mixture Density Estimation. MIT Press, 2000.

[9] E. Kolaczyk and R. Nowak. Multiscale likelihood analysis and complexity penalized estimation. To appear in Annals of Statistics. Available at http://www.ece.wisc.edu/∼nowak/msla.pdf, 2004.