Compressive Wireless Sensing

Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Robert Nowak *
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
bajwa@cae.wisc.edu, jhaupt@cae.wisc.edu, akbar@engr.wisc.edu, nowak@engr.wisc.edu

ABSTRACT

Compressive Sampling is an emerging theory that is based on the fact that a relatively small number of random projections of a signal can contain most of its salient information. In this paper, we introduce the concept of Compressive Wireless Sensing for sensor networks in which a fusion center retrieves signal field information from an ensemble of spatially distributed sensor nodes. Energy and bandwidth are scarce resources in sensor networks and the relevant metrics of interest in our context are 1) the latency involved in information retrieval; and 2) the associated power-distortion trade-off. It is generally recognized that given sufficient prior knowledge about the sensed data (e.g., statistical characterization, homogeneity etc.), there exist schemes that have very favorable power-distortion-latency trade-offs. We propose a distributed matched source-channel communication scheme, based in part on recent results in compressive sampling theory, for estimation of sensed data at the fusion center and analyze, as a function of number of sensor nodes, the trade-offs between power, distortion and latency. Compressive wireless sensing is a universal scheme in the sense that it requires no prior knowledge about the sensed data. This universality, however, comes at the cost of optimality (in terms of a less favorable power-distortion-latency trade-off) and we quantify this cost relative to the case when sufficient prior information about the sensed data is assumed.

Categories and Subject Descriptors

E.4 [**Data**]: Coding and Information Theory—*Data compaction and compression, Formal models of communication*; H.1.1 [**Models and Principles**]: Systems and Information Theory—*General systems theory, Information theory*

*This work was supported in part by the National Science Foundation under Grants No. CCR-0310889 and CNS-0519824.

IPSN'06, April 19–21, 2006, Nashville, Tennessee, USA.

General Terms

Algorithms, Design, Performance, Theory

Keywords

Wireless sensor networks, compressive sampling, uncoded communications

1. INRODUCTION

Sensor networking is an emerging technology that promises an unprecedented ability to monitor the physical world via a spatially distributed network of small and inexpensive wireless sensor nodes that have the ability to self-organize into a well-connected network. A typical wireless sensor network, as shown in Fig. 1, consists of a large number of sensor nodes, spatially distributed over a region of interest, that observe some (noisy) data. In many applications, a distant fusion center (FC) retrieves relevant field information from the sensor nodes. Energy and bandwidth are scarce resources in such networks since communication from the sensor nodes to FC generally takes place over a power and bandwidth constrained wireless channel. Consequently, a major challenge in the design of sensor networks is developing schemes that extract relevant information about the sensed data (sensor field) at a desired fidelity at FC with least consumption of network resources. In this regard, the relevant metrics of interest are 1) the latency (or alternatively, bandwidth) involved in information retrieval; and 2) the associated power-distortion trade-off: the power P_{tot} consumed by the sensor network in delivering relevant information to FC at the desired distortion D.

In this paper, we introduce the concept of *Compressive Wireless Sensing* (CWS) for energy efficient estimation (at FC) of sensor data that contain some sort of *structural regularity*. CWS is based on a distributed matched source-channel communication architecture and is inspired by recent results in wireless communications [4, 5, 10, 1, 12] and compressive sampling theory [2, 3, 8], and rests on the fact that a relatively small number of random projections of a signal can contain most of its salient information. CWS, in essence, is a completely decentralized scheme for delivering random projections of the sensor network data to FC in a distributed and energy efficient manner and under the right conditions, FC can recover a *good* approximation of the data from these random projections. Three distinct features of CWS are: 1) processing and communications are combined into one distributed operation; 2) it requires almost no in-network processing and communications; and 3) consistent

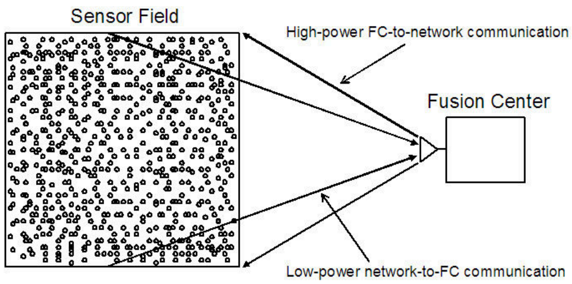

Figure 1: Sensor network with a fusion center (FC). Circles denote sensor nodes. FC can communicate to the network at a very high power whereas communication channel from the network to FC is power and bandwidth constrained.

field estimation is possible ($D \searrow 0$ as node density increases), even if little or no prior knowledge about the sensed data is assumed, while P_{tot} grows at most sub-linearly with the number of nodes in the network. Thus, CWS provides a universal and efficient approach to distributed estimation of sensor network data without putting strict constraints on the underlying structure of sensed data.

1.1 Relationship to Previous Work

It is generally recognized that given sufficient prior knowledge about the sensed data (e.g., statistical/topological characterization or homogeneity of the sensor network data), there exist schemes that have very favorable power-distortion-latency (-bandwidth) trade-offs (see, e.g., [4, 5, 1, 6]). CWS, however, is a universal scheme in the sense that it requires no prior knowledge about the sensed data. Nevertheless, this universality comes at the cost of optimality (in terms of a less favorable power-distortion-latency trade-off). For example, assuming no prior knowledge about the sensed data, the theoretical analysis of CWS in Section 4 yields a power-distortion-latency trade-off of the form[1]

$$D \sim {P_{tot}}^{-2\alpha/(2\alpha+1)} \sim L^{-2\alpha/(2\alpha+1)} \quad (1)$$

where $\alpha > 0$ (which need not be known to the network itself) quantifies the structural regularity of the sensor network data (cf. Section 2). Note that this relation does not mean that a fixed number of sensor nodes using more power and/or latency can provide more accuracy. Rather, distortion (D), power consumption (P_{tot}) and latency (L) are functions of the number of nodes, and the above relation indicates how the three performance metrics behave as the density of nodes increases. On the other hand, assuming sufficient prior knowledge about the sensed data, we show in Section 3 that there exists an efficient distributed estimation scheme that achieves the distortion scaling of an ideal centralized estimator and has a power-distortion-latency trade-off of the form

$$D \sim {P_{tot}}^{-2\alpha} \sim L^{-2\alpha} \quad (2)$$

Thus, in essence, this paper identifies a trade-off between universality and optimality: CWS is universal for a broad class of sensor fields but cannot reach the optimality of (2), whereas an optimal distributed scheme, such as the one presented in Section 3, can fail miserably under false prior information (cf. Sections 4 and 5) and therefore, can never be universal. CWS is, therefore, primarily a framework for sensor networks having either little prior knowledge about the sensed field or low confidence level about the accuracy of the available knowledge.

Finally, most previous works in the area of sensor data estimation have focused on multihop communication schemes and in-network data processing and compression [13, 9, 11, 15]. This requires a significant level of network infrastructure, and the theoretical approaches in the works above generally assume this infrastructure as given. Our approach, in contrast to previous methods, eliminates the need for in-network communications and processing and instead requires phase synchronization among nodes, which imposes a relatively small burden on network resources and can be achieved by employing the distributed synchronization scheme described in [12]. Thus, our proposed wireless sensing system is perhaps more accurately viewed as a *sensor ensemble* which is appropriately queried by an *information retriever* (FC) to elicit the desired information about the sensed data.

1.2 Organization

The rest of this paper is organized as follows. In Section 2 we formally describe the problem considered in this paper and develop the basic communication architecture of our scheme. Section 3 describes an energy efficient distributed estimation scheme that, under the assumption of sufficient prior knowledge about the sensed data, achieves the distortion scaling of an ideal centralized estimator. In Section 4, we introduce the concept of CWS and analyze, as a function of number of sensor nodes, the associated trade-offs between power, distortion and latency. In Section 5, we make a comparison between CWS and the distributed scheme of Section 3 using numerical results and show the basic trade-off between universality and optimality. Finally, we summarize our results and present concluding remarks in Section 6.

2. PROBLEM FORMULATION AND APPROACH

In this section, we give an overview to the problem and approach considered in this paper. In the following sections, we shall elaborate on the technical details of the scheme. To begin, consider a wireless sensor network with n nodes where each node takes a noisy sample of the form

$$x_j = x_j^* + w_j\,, \quad j = 1, \ldots, n \quad (3)$$

and the errors $\{w_j\}_{j=1}^n$ are independent, zero-mean Gaussian random variables with variance σ_w^2. We can consider this data as a vector $x \in \mathbb{R}^n$ such that $x = x^* + w$, where $x^* \in \mathbb{R}^n$ is the noiseless data vector and $w \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_n)$. We further assume that $|x_j^*| \leq B$, $j = 1, \ldots, n$, for some known constant $B > 0$, which is determined by the sensing range of the sensors.

It is a well known fact in the field of data compression, evidenced by the success of familiar compression standards such as JPEG, MPEG and MP3, that data in real world often contain redundancies. Moreover, data collected at nearby nodes in a dense sensor network is expected to be highly correlated [14]. Therefore, it is quite reasonable to assume that x^* is *compressible* in the sense that it is *well-*

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

approximated by a linear combination of k vectors taken from an orthonormal basis of $\mathbb{R}^n$ (e.g., smooth signals tend to be compressible in the Fourier basis and piecewise smooth signals tend to be compressible in a wavelet/wedgelet basis). More precisely, let $\mathbf{\Psi} \triangleq \{\psi_i\}_{i=1}^n$ be an orthonormal basis of $\mathbb{R}^n$. Denote by $\theta_i = \psi_i^T x^*$ (projection of x^* onto ψ_i) the coefficients of x^* in this new basis. Relabel these coefficients so that

$$|\theta_1| \geq |\theta_2| \geq \cdots \geq |\theta_n| \tag{4}$$

The *best k-term approximation* of x^* in terms of $\mathbf{\Psi}$ is given by

$$x^{*(k)} = \sum_{i=1}^{k} \theta_i \, \psi_i \tag{5}$$

and we say that x^* is α-compressible in $\mathbf{\Psi}$ (or that $\mathbf{\Psi}$ is the compressing basis of x^*) if the *average squared-error* behaves like

$$\frac{\left\| x^* - x^{*(k)} \right\|^2}{n} \triangleq \frac{1}{n} \sum_{j=1}^{n} \left(x_j^* - x_j^{*(k)} \right)^2 = O\left(k^{-2\alpha}\right) \tag{6}$$

for some $\alpha > 0$, where the parameter α governs the degree to which x^* is compressible with respect to $\mathbf{\Psi}$. Note that the ordering of coefficients in (4) may be a function of the underlying signal x^* and in such cases, could never be known a priori.

Given x, the goal of the sensor network is to compute a reconstruction $\widehat{x}$ of the noiseless data vector x^* at the FC with a small latency (L) and expected squared-error, $D = \mathbb{E}\left[\frac{1}{n}\left\|\widehat{x} - x^*\right\|^2\right]$, while at the same time consuming minimal amount of total power P_{tot}.

Before proceeding further, we shall make the following assumptions concerning the communications from the sensor network to the FC:

1. Each sensor is equipped with a single isotropic antenna.
2. The sensors are constrained to a maximum individual transmit power of P.
3. The sensors communicate with FC over a narrowband Additive White Gaussian Noise (AWGN) wireless channel of bandwidth W Hz at some carrier frequency f_c, where $f_c \gg W$, and each channel use is characterized by transmission over a period of $T = 1/W$ seconds.
4. Each sensor has a local oscillator synchronized to the carrier frequency f_c and the network is fully phase synchronized in the sense that the sensor transmissions arrive at the FC in a phase coherent fashion. This may be achieved by employing the distributed synchronization scheme described in [12].
5. Let d_j, $j = 1, \ldots, n$, be the distance between the sensor at location j and FC. The FC is assumed to be far away from the sensor network so that $d_1 \approx \cdots \approx d_n \approx d$ and therefore, the path losses of all nodes are identical.
6. There is no multipath fading, which would indeed be the case in many remote sensing applications with static sensor nodes that have a line-of-sight connection to the FC.

Ideal Centralized Estimation: Let us first consider an ideal centralized estimator in which the sensor measurements $\{x_j\}_{j=1}^n$ are assumed to be available at the FC noise-free. The distortion scaling of this estimator would serve as a benchmark for assessing the distortion related performances of the distributed schemes presented in Sections 3 and 4.

Given x, a centralized estimator $\widehat{x}_{cen}$ at the FC can be easily constructed by projecting x onto the first k elements of $\mathbf{\Psi}$

$$\begin{aligned} \widehat{x}_{cen} &= \sum_{i=1}^{k} \left(\psi_i^T x\right) \psi_i \\ &= x^{*(k)} + \sum_{i=1}^{k} \left(\psi_i^T w\right) \psi_i \end{aligned} \tag{7}$$

which results in a bias/variance trade-off

$$D_{cen} \;=\; \mathbb{E}\left[\frac{1}{n}\left\|\widehat{x}_{cen} - x^*\right\|^2\right] \;\preceq\; k^{-2\alpha} + \left(\frac{k}{n}\right)\sigma_w^2 \tag{8}$$

where the first term is the squared bias and the second is the variance. The minimum is attained by setting $k \sim n^{1/(2\alpha+1)}$, resulting in

$$D_{cen} \;\preceq\; n^{-2\alpha/(2\alpha+1)} \tag{9}$$

Next, we present a communication architecture for computing projections of sensor network data onto any normalized vector in $\mathbb{R}^n$, which would act as a basic building block of our proposed scheme.

2.1 Distributed Projections in Wireless Sensor Networks

In this section, we develop the basic communication architecture that acts as a building block of CWS. At the heart of our approach is an energy efficient, distributed method of computing projections of the sensor network data onto any normalized vector in $\mathbb{R}^n$ by exploiting the spatial averaging inherent in a multiple access channel (MAC). To begin, we first define the notion of a *Sparsity Map.*

Definition 1. Let $q \in \mathbb{R}^n$ and $S_p : \mathbb{R}^n \to \mathrm{P}(\{1, \ldots, n\})$, where $\mathrm{P}(X)$ means power set of X. We call S_p the sparsity map of q if $S_p(q) = \{j \in \{1, \ldots, n\} : q_j \neq 0\}$ and $|S_p(q)|$ is a counting measure on $S_p(q)$.

Now, let $\varphi \in \mathbb{R}^n$, where $\left\|\varphi\right\|^2 = 1$, and $\upsilon = \varphi^T x^* = \sum_{j=1}^n \varphi_j x_j^*$ be the projection of x^* onto φ. Using the notion of sparsity map, denote $|S_p(\varphi)| = n_\varphi$. Since $\left\|\varphi\right\|^2 = 1$, this implies $|\varphi_j|^2 \approx \left\|\varphi\right\|^2/n_\varphi = 1/n_\varphi \;\forall\; j \in S_p(\varphi)$. Given x, we assume that the goal of the sensor network is to compute an estimate ($\widehat{\upsilon}$) of υ at the FC. One possibility is to nominate a clusterhead in the network and then, assuming all the sensor nodes know φ and have constructed routes which form a spanning tree through the network to the clusterhead, sensor nodes locally compute $\varphi_j x_j$ and aggregate these values up the tree to obtain $\widehat{\upsilon} = \sum_{j=1}^n \varphi_j x_j$ at the clusterhead. However, even if we ignore the communication cost of delivering $\widehat{\upsilon}$ from the cluterhead to the FC, it is easy to see that this scheme requires at least n transmissions.

Another, more promising, alternative is to exploit recent results concerning uncoded coherent transmission schemes

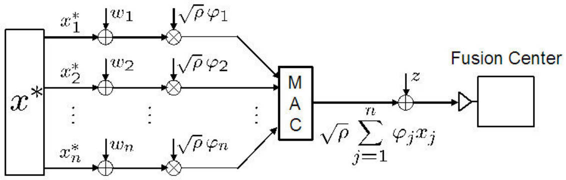

Figure 2: A distributed communication architecture for computing projections of sensor network data at the fusion center.

[4, 5, 10, 1]. The proposed distributed communication architecture, illustrated in Fig. 2, involves phase-coherent, low-power, analog transmission of weighted sample values directly from the nodes in the network to the FC via the narrowband AWGN network-to-FC communication channel. To begin with, assume all the nodes in the network have knowledge of φ. Practical schemes of how the sensor network might achieve this would be discussed in Section 3.2. Each node multiplies its measurement x_j with $(\sqrt{\rho}\,\varphi_j)$ to obtain $m_j = \sqrt{\rho}\,\varphi_j x_j$, where $\rho > 0$ is a scaling factor used to satisfy sensors' transmit power constraint P, and all the nodes coherently tranmist their respective m_j's in an analog fashion over the network-to-FC communication channel. Clearly, $\mathbb{E}[|m_j|^2] \le \rho\,(B^2 + \sigma_w^2)/n_\varphi$ if $j \in S_p(x^*) \cap S_p(\varphi)$ and $\mathbb{E}[|m_j|^2] = \rho\,\sigma_w^2/n_\varphi$ if $j \in S_p(x^*)^c \cap S_p(\varphi)$. Thus, $\mathbb{E}[|m_j|^2] \le \rho\,(B^2 + \sigma_w^2)/n_\varphi \;\forall\; j \in S_p(\varphi)$ and $m_j \equiv 0$ if $j \notin S_p(\varphi)$. Hence, the average transmission power for each sensor ($\in S_p(\varphi)$) is given by

$$P_j \le \rho\,(B^2 + \sigma_w^2)/n_\varphi \qquad (10)$$

and to satisfy the individual sensor transmit power constraint, we need to take $\rho = (n_\varphi \lambda P)/(B^2 + \sigma_w^2)$ for $0 < \lambda \le 1$, resulting in $P_j \le \lambda P \;(\le P)$.

Because of the coherent transmission by the sensor nodes, the network-to-FC communication channel is effectively transformed into an AWGN MAC channel and the received signal at the FC is given by

$$\begin{aligned} r &= \sum_{j=1}^{n} m_j + z \;=\; \sqrt{\rho}\,\sum_{j=1}^{n}\varphi_j x_j + z \\ &= \sqrt{\rho}\,\varphi^T(x^* + w) + z \;=\; \sqrt{\rho}\,(\upsilon + \widetilde{w}) + z \end{aligned} \qquad (11)$$

where $z \sim \mathcal{N}(0, \sigma_z^2)$ is the channel additive white Gaussian noise and $\widetilde{w} \sim \mathcal{N}(0, \sigma_w^2)$. Strictly speaking, the received signal from each node, m_j, in the above expression should be scaled by an attenuation constant, $a_j \in (0, 1)$, that depends on the distance d_j between the node and FC and the path loss exponent. However, under the assumption of identical path losses, the a_j's are nearly the same and we ignore this uniform attenuation since it will uniformly increase the required power per node by a constant factor to attain a desired distortion.

In essence, the above setup corresponds to obtaining a noisy projection of x onto φ at the FC that is scaled by $\sqrt{\rho}$ and given r, the FC can easily estimate υ as $\widehat{\upsilon} = r/\sqrt{\rho}$ and the resulting distortion is given by

$$\begin{aligned} D_\upsilon &= \mathbb{E}\left[|\widehat{\upsilon} - \upsilon|^2\right] \;=\; \sigma_w^2 + \frac{\sigma_z^2}{\rho} \\ &= \sigma_w^2 \;+\; \frac{\sigma_z^2(B^2 + \sigma_w^2)}{n_\varphi \lambda P} \end{aligned} \qquad (12)$$

where the first term in the above expression is due to the measurement noise and the second term is due to the communication noise and the key question becomes: What is the necessary and sufficient value of λ (and correspondingly of ρ) to make the distortion in (12) as small as possible? This question is answered in the proof of the following theorem.

THEOREM 1. *Given the observation model of* (3), *it is possible to obtain an estimate* ($\widehat{\upsilon}$) *of the projection of sensor network data onto any normalized vector in* $\mathbb{R}^n$, *such that* $D_\upsilon \sim \sigma_w^2$, *by using only a fixed amount of total power,* $P_\upsilon = O(1)$, *independent of the number of nodes in the network and the structure of the vector on which data is projected.*[2]

PROOF. To prove this theorem, observe that the first term in (12) is unaffected by the proposed communication scheme and the second term decays as $1/\lambda$. For fastest distortion reduction, both the terms in (12) must be of the same order. That is,

$$\sigma_w^2 \;\sim\; \frac{\sigma_z^2(B^2 + \sigma_w^2)}{n_\varphi \lambda P} \;\Longleftrightarrow\; \lambda \;\sim\; \frac{\sigma_z^2(B^2 + \sigma_w^2)}{n_\varphi \sigma_w^2 P} \qquad (13)$$

Hence, the necessary and sufficient λ to obtain the optimal distortion should be chosen as

$$\lambda \;\sim\; \frac{\sigma_z^2(B^2 + \sigma_w^2)}{n_\varphi \sigma_w^2 P} \;\sim\; \frac{1}{n_\varphi} \qquad (14)$$

and from (12), this would result in $D_\upsilon \sim \sigma_w^2$. Moreover, since a total of n_φ nodes used the communication channel during this distributed projection[3], the necessary and sufficient total power (P_υ) involved in obtaining $\widehat{\upsilon}$ at the FC would behave as

$$P_\upsilon \;=\; \sum_{j=1}^{n} P_j \;\le\; n_\varphi\,(\lambda P) \;\sim\; \frac{\sigma_z^2(B^2 + \sigma_w^2)}{\sigma_w^2} \;=\; O(1) \quad (15)$$

This completes the proof. □

Remark 1. Given the observation model of (3), it is easy to see that $D_\upsilon \sim \sigma_w^2$ is the best that any (centralized or distributed) scheme can hope to achieve in terms of distortion and Theorem 1 shows that our distributed scheme can achieve that by using only a fixed amount of power.

3. DISTRIBUTED ESTIMATION FROM NOISY PROJECTIONS

In this section, using the communication architecture presented in Section 2.1 as a basic building block, we present a completely decentralized scheme for efficient estimation of sensor network data at the FC. The underlying assumption is that the sensor nodes not only have a complete knowledge of the basis in which x^* is compressible but also the precise

[2]With a slight abuse of notation, $\sim$ here implies that both quantities are '*of the same order*'

[3]Recall that m_j, and thus P_j, would be equal to zero for $j \notin S_p(\varphi)$

knowledge of the ordering of its coefficients in the compressing basis, as in (4). Under this assumption, we analyze the power-distortion-latency trade-offs in this scheme as a function of number of sensor nodes and show that the proposed distributed scheme can achieve the optimal centralized distortion scaling of (9).

To begin with, let $\mathbf{\Psi} \triangleq \{\psi_i\}_{i=1}^n$ be an orthonormal basis of $\mathbb{R}^n$ such that $\left\|x^* - x^{*(k)}\right\|^2/n = O(k^{-2\alpha})$, where $x^{*(k)} = \sum_{i=1}^k \theta_i\,\psi_i$ (perhaps after re-labeling the indices i) and each coefficient θ_i is computed as a projection of the form $\theta_i = \psi_i^T x^* = \sum_{j=1}^n \psi_{ij} x_j^*$. The sensor network can compute k projections of x onto $\{\psi_i\}_{i=1}^k$ by employing the scheme of Section 2.1 in k consecutive channel uses. Thus, at the end of k channel uses, each one corresponding to a projection of x onto an element of $\mathbf{\Psi}$, FC has access to the estimates of k projection coefficients given by

$$\widehat{\theta}_i \;=\; r_i/\sqrt{\rho_i} \;=\; \theta_i + \psi_i^T w + z_i/\sqrt{\rho_i}\;, \quad i = 1,\ldots,k \quad (16)$$

where $z_i \sim \mathcal{N}(0, \sigma_z^2)$ is the MAC AWGN corresponding to i-th channel use and $\rho_i = (n_{\psi_i}\lambda_i P)/(B^2+\sigma_w^2)\,,\ n_{\psi_i} = |S_p(\psi_i)|$ and $0 < \lambda_i \leq 1$; resulting in $D_{\theta_i} = \mathbb{E}\,[|\widehat{\theta}_i - \theta_i|^2] = \sigma_w^2 + \sigma_z^2/\rho_i$. From these k projection coefficients, FC can easily estimate x^* as

$$\begin{aligned}\widehat{x} \;&=\; \sum_{i=1}^k \widehat{\theta}_i\psi_i \;=\; x^{*(k)} \;+\; \sum_{i=1}^k \left(\psi_i^T w + z_i/\sqrt{\rho_i}\right)\psi_i \\ &=\; \widehat{x}_{cen} \;+\; \sum_{i=1}^k (z_i/\sqrt{\rho_i})\,\psi_i = \widehat{x}_{cen} \;+\; \widetilde{z} \qquad (17)\end{aligned}$$

where $\widetilde{z} \sim \mathcal{N}\left(0, \mathbf{diag}\,(\sigma_z^2/\rho_1,\ldots,\sigma_z^2/\rho_k)\right)$ by virtue of the fact that z_i is independent of z_j for $i \neq j$, and the resulting distortion is given by

$$\begin{aligned}D \;&=\; \mathbb{E}\left[\frac{1}{n}\left\|\widehat{x} - x^*\right\|^2\right] \;=\; D_{cen} \;+\; \frac{1}{n}\sum_{i=1}^k \frac{\sigma_z^2}{\rho_i} \\ &\preceq\; k^{-2\alpha} \;+\; \left(\frac{k}{n}\right)\sigma_w^2 \;+\; \frac{1}{n}\sum_{i=1}^k \frac{\sigma_z^2}{\rho_i} \qquad (18)\end{aligned}$$

where the first two terms correspond to D_{cen} and the last term is the distortion induced by k noisy MAC communications. The above relation governs the interplay between D, n, k, α and λ_i's. For fastest distortion reduction, all three terms in (18) must scale (as a function of n) at the same rate. That is,

$$k^{-2\alpha} \;\sim\; \left(\frac{k}{n}\right)\sigma_w^2 \;\sim\; \frac{1}{n}\sum_{i=1}^k \frac{\sigma_z^2}{\rho_i} \qquad (19)$$

Analyzing the above expression shows that k must be chosen, independently of $\{\rho_i\}_{i=1}^k$, as $k \sim n^{1/(2\alpha+1)}$ and the corresponding distortion at FC would scale as

$$D \;\preceq\; n^{-2\alpha/(2\alpha+1)} \qquad (20)$$

that has the same scaling behavior as D_{cen}. Moreover, since a total of n_{ψ_i} nodes communicated during the i-th MAC transmission, the total power consumed by the sensor network during the entire reconstruction process is given by

$$P_{tot} \;=\; \sum_{i=1}^k P_{\theta_i} \;\leq\; \sum_{i=1}^k n_{\psi_i}\,(\lambda_i P) \qquad (21)$$

Let us call $\sum_{i=1}^k n_{\psi_i}\lambda_i = \Gamma$, then $P_{tot} \leq P\,\Gamma$ and the only question that remains to be answered is how to choose λ_i's so that Γ is minimized, which in turns minimizes P_{tot}. The answer to this question lies in the following theorem.

THEOREM 2. *Using the above distributed scheme for estimation of x^* and given the observation model of* (3), *the final distortion at the FC scales as given in* (20) *if and only if*

$$\Gamma \;\succeq\; n^{-2\alpha/(2\alpha+1)}$$

Moreover,

$$\lambda_i \;=\; \sigma_z^2(B^2+\sigma_w^2)/(n_{\psi_i}\sigma_w^2 P) \;\sim\; 1/n_{\psi_i}\;, \quad i = 1,\ldots,k$$

is the only set of λ_i's that achieves the lower bound for Γ in the sense that

$$\Gamma \;\sim\; n^{-2\alpha/(2\alpha+1)}$$

PROOF. The proof of this theorem is given in the Appendix. □

3.1 Power-Distortion-Latency Trade-offs

In this section, we present the power-distortion-latency trade-offs involved in the proposed distributed estimation scheme. Recall that in order to achieve the optimal distortion scaling

$$D \;\preceq\; n^{-2\alpha/(2\alpha+1)}\,, \qquad (22)$$

the network had to employ $k = n^{1/(2\alpha+1)}$ MAC transmissions, each one corresponding to a projection of x onto an element of $\mathbf{\Psi}$, and under the assumption that the k projections shared the channel via Time Division Multiple Access (TDMA), we get the following relation for the latency L involved in information retrieval from the network[4]

$$L \;\sim\; n^{1/(2\alpha+1)} \qquad (23)$$

Moreover, if we take $\lambda_i \sim 1/n_{\psi_i}$, $i = 1,\ldots,k$, then from (21) and Theorem 2 we get the following relation for the total power P_{tot} consumed by the network in information retrieval

$$P_{tot} \;\leq\; k\;\frac{\sigma_z^2(B^2+\sigma_w^2)}{\sigma_w^2} \;\preceq\; n^{1/(2\alpha+1)} \qquad (24)$$

Hence, given the observation model of (3) and assuming that the sensor network has sufficient prior knowledge about the underlying signal structure (*i.e.*, compressing basis of x^* and the ordering of its coefficients in that basis), the proposed distributed estimation scheme can achieve the optimal centralized distortion scaling of (9) and from (22), (23) and (24), the associated power-disortion-latency trade-off is given by

$$D \;\sim\; {P_{tot}}^{-2\alpha} \;\sim\; L^{-2\alpha} \qquad (25)$$

3.2 Communicating the Compressing Basis to the Network

Assuming the designer of the sensor network has knowledge of the compressing basis $\mathbf{\Psi}$ of x^*, we try to address the issue of how to communicate the compressing basis (or

[4]The projections may equally well share the channel via Frequency Division Multiple Access (FDMA) and that would translate the latency requirements into the bandwidth requirements.

a subset of it) to the sensor nodes – an assumption inherent to the optimality of above scheme. Pre-storage of this information in the sensor nodes is not a viable option because of possible node failures, changes in the structure of sensed data etc. Moreover, pre-storage of the entire compressing basis or a subset of it, $\{\psi_i\}_{i=1}^{l}$, where $1 \leq l \leq n$, in each sensor node would require at least $O(n)$ bits per sensor for storage which might not always be feasible in large scale sensor networks. Even pre-storage of only corresponding entries of the k basis elements, $\{\psi_{i,j}\}_{i=1}^{k}$, in the j-th sensor node would still require at least $O\left(n^{1/(2\alpha+1)}\right)$ bits per sensor for optimal distortion scaling.

Another, more feasible but not always practical, approach to this problem is that the FC transmits this information to the sensor nodes (over a separate FC-to-network communication channel) before the start of each projection. For the case of ψ_i that has some sort of regularity in its structure so that it does not require addressing each node individually (e.g. $\psi_i = [\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}]^T$), this can be readily achieved by broadcasting a few command signals from the FC to the nodes. However, depending upon the structure of the normalized vector, this approach may require the FC to be able to address each sensor individually which again might not be practical in large scale sensor networks. However, we will show in the next section that among many other things, compressive wireless sensing scheme can easily work around this problem.

4. COMPRESSIVE WIRELESS SENSING

In Section 3, we proposed an efficient distributed estimation scheme that achieves the optimal centralized distortion scaling of (9) under the assumption that the network (nodes and/or FC) has sufficient knowledge about the basis in which x^* is compressible. Generally speaking, however, even if the destination knows the basis in which x^* is compressible, it is quite likely that it will not know ahead of time the precise ordering of the coefficients of x^* in this basis. As an example, consider the following simple scenario. Suppose x^* is a spatially non-sparse vector of length n ($|S_p(x^*)| = n$) with only one non-zero coefficient of amplitude $\sqrt{n}$ in some transform basis $\mathbf{\Psi} \triangleq \{\psi_i\}_{i=1}^{n}$ such that $\|x^*\|^2/n = 1$ (*i.e.*, x^* is super sparse in $\mathbf{\Psi}$). This is an example of the case where we know the basis in which x^* is compressible but do not know the ordering of the coeffiecients. One naive approach to this problem is to require each sensor to digitally transmit its measurement to the destination, where the reconstruction is then performed. Alternatively, all the sensors might collaboratively process their measurements to reconstruct x^* in-network and then transmit the result to the destination. Both approaches, however, while providing us with consistent estimates, would require P_{tot} and L to be at least $\sim n$.

Another approach to this problem could be to use the distributed scheme described in Section 3. However, since the network does not have a precise knowledge of the ordering of coefficients, it would have to resort to random transform domain sampling where the network computes a distributed projection of the data onto ψ_i and i is selected uniformly at random from the set $\{1, \ldots, n\}$. Ignoring the distortion due to the measurement noise, the squared reconstruction error would be 0 at the FC if the spike in $\mathbf{\Psi}$ domain corresponds to ψ_i and 1 otherwise and the probability of not finding the spike in k trials is $\left(1 - \frac{1}{n}\right)^k$, giving an average squared-error of $\left(1 - \frac{1}{n}\right)^k \cdot 1 + (k/n) \cdot 0 = \left(1 - \frac{1}{n}\right)^k$. If n is large, we can approximate this by $D = \left(1 - \frac{1}{n}\right)^k \approx e^{-k/n}$. Therefore, for any $k < n$, we have $D \not\to 0$ as $n \to \infty$ while P_{tot} and $L \sim k$. This simple example shows that the power-distortion-latency tradeoff of (25) might break down if the network does not have enough prior knowledge about the sensed signal

Another more general, and perhaps relevant, example is a situation in which the signal is piecewise constant. Signals of this type do lie in a low-dimensional subspace of the wavelet domain, but precisely which subspace depends on the locations of the changepoints in the signal, which of course is unlikely to be known a priori. Broadly speaking, any signal that is generally smooth apart from some localized sharp changes or edges will essentially lie in a low-dimensional subspace of a multiresolution basis such as wavelets or curvelets, but the subspace will be function-dependent and thus, preclude the use of methods, like the one of the previous section, that require prior specification of the basis functions to be used in the projection process. This is where the *universality* of compressive wireless sensing scheme, presented in this section, comes into play. As we shall see, compressive wireless sensing provides us with a consistent estimation scheme ($D \searrow 0$ as node density increases), even if little or no prior knowledge about the sensed data is assumed, while P_{tot} grows at most sub-linearly with the number of nodes in the network.

Recall that if $\upsilon = \varphi^T x^* = \sum_{j=1}^{n} \varphi_j x_j^*$ is the projection of x^* onto a normalized vector $\varphi \in \mathbb{R}^n$ then using the communication architecture described in Section 2.1 and consuming only $O(1)$ amount of power, the FC can obtain an estimate of υ given by

$$\widehat{\upsilon} = \varphi^T \, (x^* + w) + \widetilde{z} \tag{26}$$

where $\widetilde{z} \sim \mathcal{N}(0, \sigma_z^2/\rho)$ is the scaled AWGN communication noise and $\sigma_z^2/\rho \sim \sigma_w^2$ (cf. Theorem 1). Now, instead of projecting the sensor network data onto a subset of a deterministic basis of $\mathbb{R}^n$, in Compressive Wireless Sensing (CWS), the FC tries to reconstruct x^* from *noisy random projections* of the sensor network data. Specifically, let $\{\phi_i \in \mathbb{R}^n\}_{i=1}^{n}$ be an independent and identically distributed (i.i.d) sequence of Rademacher random vectors *i.e.*, $\{\phi_{i,j}\}_{j=1}^{n} = \pm 1/\sqrt{n}$, each with probability 1/2, and the FC tries to reconstruct x^* by projecting x onto k of these random vectors. Because the entries of the projection vector ϕ_i are generated at random, observations of this form are called *(noisy) random projections* of the signal. An important consequence of using random Rademacher vectors is that each sensor can locally draw the elements of the random vectors $\{\phi_i\}_{i=1}^{k}$ in an efficient manner by using the seed of a pseudo-random generator and its (network) address. Similarly, given the seed values and the number of nodes in the network, the destination can easily reconstruct the vectors $\{\phi_i\}_{i=1}^{k}$. Therefore, in CWS the FC does not need to convey any information to the sensor nodes regarding the projection vectors.

After employing k random projections, the observations at the FC take the form of

$$\begin{aligned} y_i &= \sum_{j=1}^{n} \phi_{ij}(x_j^* + w_j) + \widetilde{z}_i \\ &= \phi_i^T (x^* + w) + \widetilde{z}_i \,, \; i = 1, \ldots, k, \end{aligned} \tag{27}$$

where $w = [w_1 \dots w_n]^T$, and $\{w_j\}_{j=1}^n$ and $\{\widetilde{z}_i\}_{i=1}^k$ are i.i.d. zero-mean Gaussian random variables, independent of $\{\phi_{i,j}\}$, with variances σ_w^2 and $\sigma_z^2/\rho_i \sim \sigma_w^2$ respectively. Notice that the observations above are equivalent (in distribution) to observations of the form

$$y_i = \phi_i^T x^* + \eta_i\ ,\ i = 1, \dots, k, \tag{28}$$

where $\{\eta_i\}$ are i.i.d zero-mean Gaussian random variables *independent of* $\{\phi_{i,j}\}$ with variance $\sigma^2 \sim \sigma_w^2$ (since $\mathbb{E}[|\widetilde{z}_i|^2] \sim \sigma_w^2\ \forall\ i$). This result follows directly from [8], where the equivalence of $\{\phi_i^T w + \widetilde{z}_i\}$ and $\{\eta_i\}$ (in distribution) and the independence of $\{\eta_i\}$ and $\{\phi_{i,j}\}$ is proved.

Given a countable collection $\mathcal{X}$ of candidate reconstruction functions, such that each $\widetilde{x} \in \mathcal{X}$ satisfies $|\widetilde{x}_j| \le B$ for all entries $j = 1, \dots, n$, the CWS estimate of x^*, $\widehat{x}_k$, is obtained as a solution of

$$\widehat{x}_k\ =\ \arg\min_{\widetilde{x}\in\mathcal{X}} \left\{ \widehat{R}(\widetilde{x}) + \frac{c(\widetilde{x}) \log 2}{k\epsilon} \right\} \tag{29}$$

where $c(\widetilde{x})$ is a non-negative number assigned to each $\widetilde{x} \in \mathcal{X}$ such that $\sum_{\widetilde{x}\in\mathcal{X}} 2^{-c(\widetilde{x})} \le 1$, $\epsilon > 0$ is a constant that depends on the function bound B and the noise variance σ^2 as described in [8], and $\widehat{R}(\widetilde{x})$ is the empirical risk defined as

$$\widehat{R}(\widetilde{x}) = \frac{1}{k} \sum_{i=1}^{k} \left(y_i - \sum_{j=1}^{n} \phi_{i,j} \widetilde{x}_j \right)^2 . \tag{30}$$

If we assume that we can find a basis in which the signal x^* is α-compressible, then we can use this compressing basis in the reconstruction and define $c(\widetilde{x})$ in terms of it. Thus, the optimization problem becomes

$$\widehat{\theta}_k = \arg\min_{\theta\in\Theta} \left\{ \|y - \Phi^T T\theta\|_2^2 + \frac{2\log(2)\log(n)}{\epsilon} \|\theta\|_0 \right\} \tag{31}$$

where θ is the representation of $\widetilde{x}$ in the compressing basis, Φ^T is the transpose of the $n \times k$ matrix of projection vector elements, and T is the transform that takes $\widetilde{x} \in \mathcal{X}$ to the compressing domain such that $\widetilde{x} = T\theta$. As shown in [8], for α-compressible x^*, such an estimate would satisfy

$$D = E\left[\frac{\|\widehat{x}_k - x^*\|^2}{n} \right] \preceq \left(\frac{k}{\log n} \right)^{-2\alpha/(2\alpha+1)}, \tag{32}$$

while if x^* is truly sparse (has only m nonzero coefficients in the compressing basis), then

$$D = E\left[\frac{\|\widehat{x}_k - x^*\|^2}{n} \right] \preceq \left(\frac{k}{m \log n} \right)^{-1} \tag{33}$$

4.1 Power-Distortion-Latency Trade-offs

Recall that in order to achieve the optimal distortion scaling of (32) and (33), the network had to employ k MAC transmissions, each one corresponding to a projection of x onto a random vector, and consuming $O(1)$ power. Therefore, the latency L and total power P_{tot} involved in information retrieval from the network is given by

$$L\ \sim\ k \tag{34}$$
$$P_{tot}\ \sim\ k \tag{35}$$

Therefore, ignoring *log* factors, the power-distortion-latency trade-off for the case when x^* is α-compressible is given by

$$D\ \sim\ {P_{tot}}^{-2\alpha/(2\alpha+1)}\ \sim\ L^{-2\alpha/(2\alpha+1)}, \tag{36}$$

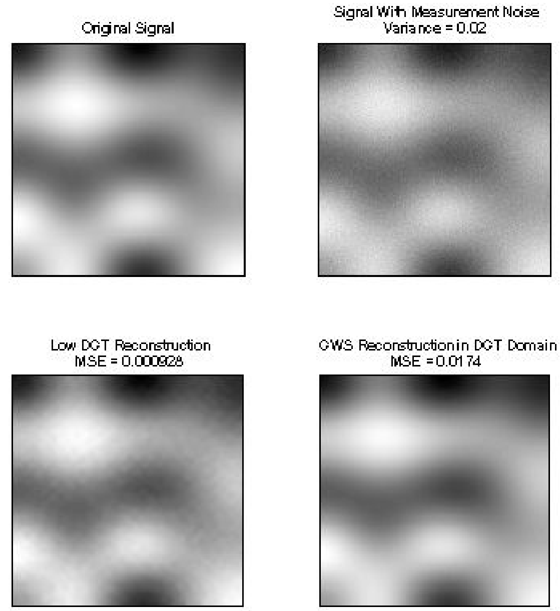

Figure 3: A truly sparse signal in the DCT domain. Measuring directly in the DCT domain leads to a better reconstruction, but CWS also yields a consistent estimate.

and by

$$D\ \sim\ {P_{tot}}^{-1}\ \sim\ L^{-1} \tag{37}$$

when it is truly sparse.

Comparing these trade-offs with the one achievable using the estimation scheme of Section 3 yields some interesting insight. Regardless of the compressibility of x^*, the best P-D trade-off that one can hope to achieve using CWS is $D \sim {P_{tot}}^{-1}$. On the other hand, if enough knowledge about the compressing basis of x^* is available a priori, one can employ the scheme of Section 3 and do much better, $D \sim {P_{tot}}^{-2\alpha}$. Therefore, given sufficient prior knowledge about the signal, CWS can be far from optimal but under circumstances where there is little or no knowledge available about x^*, CWS should be the estimation scheme of choice as discussed at the start of this section.

5. NUMERICAL RESULTS

In this section we present some numerical results to demonstrate the tradeoff between the universality of compressive wireless sensing and the optimality of sampling in the relevant subspace of a sparse signal, assuming that the relevant subspace is known a priori. For all examples in this section, the signal components are scaled to take values in the range $\pm B$, where $B = 2$. Further, each sensor measurement is contaminated with zero-mean additive Gaussian measurement noise with variance $\sigma_w^2 = 0.02$ and ρ_i is chosen so that each projection has zero-mean additive Gaussian communication noise with effective variance of $\sigma_z^2/\rho_i = 0.02$.

The original signals are of size $256 \times 256 = 65536$ pixels, and the reconstruction is performed using $k = 1600$ projections which are either random in the CWS case or specified elements of a given basis in the "assumed subspace" case. For each example, the "assumed subspace" is taken to be

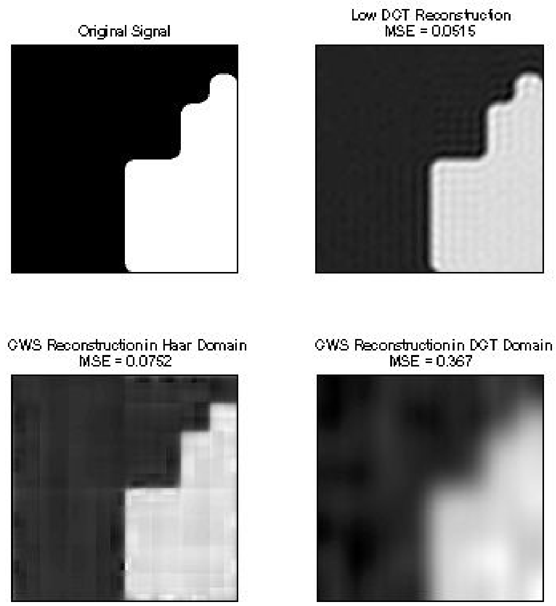

Figure 4: A signal that is approximately sparse in the Haar Wavelet domain. The low DCT and CWS Wavelet reconstructions have the same asymptotic distortion rate.

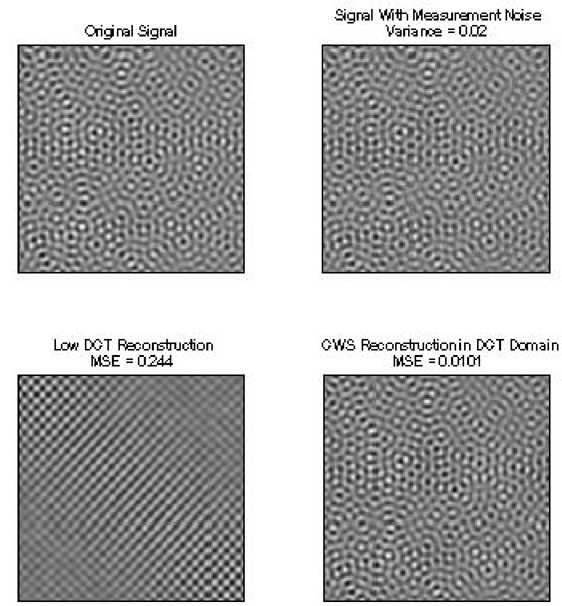

Figure 5: Another signal that is sparse in the DCT domain, but only part of the signal energy is in the directly observed frequencies. CWS performs much better than direct frequency sampling, illustrating the universality of CWS and the cost of directly measuring an incorrect subspace.

a low frequency segment of the Discrete Cosine Transform (DCT) domain. Specifically, the lowest $\sqrt{k} = 40$ coefficients in each dimension of the DCT domain are measured, and the reconstruction is carried out by using these measured coefficients and setting unmeasured coefficients to zero.

The first example, shown in Fig. 3, is a signal that consists of 25 nonzero low-frequency components in the DCT Domain (the lowest five coefficients in each dimension are nonzero). The original signal and the signal corrupted with noise are shown in the top row. The bottom row shows two reconstructions - one is the estimate obtained by measuring the k lowest frequency components in the DCT domain, and the other is a reconstruction obtained from k random projections using the DCT basis as the reconstruction basis. Direct sampling of the DCT domain leads to an estimate with lower MSE, as expected, but the CWS reconstruction estimate is also consistent with the original image and exhibits MSE below the measurement noise variance.

The second example, shown in Fig. 4, is a piecewise constant image with a boundary that is approximately sparse in the Haar Wavelet Domain. The original image is shown along with the lowest frequency DCT reconstruction in the top row. The bottom row shows two reconstructions obtained from random projections. The interesting point to note here is that the *same* set of k random projections can be used to obtain several reconstructions of the signal, simply by using different bases in the reconstruction algorithm. In this case, CWS gives consistent estimates of the actual signal using two different domains (Haar and DCT).

For this example, notice that the MSE of CWS-Haar reconstruction is comparable to that of the "assumed subspace" reconstruction. This is because the "assumed subspace" case achieves $D \sim {P_{tot}}^{-2\alpha}$, and for piecewise constant signals represented in the DCT domain, $2\alpha = 1/2$, so $D \sim {P_{tot}}^{-1/2}$. On the other hand, CWS yields a power-distortion relation of $D \sim {P_{tot}}^{-2\alpha/(2\alpha+1)}$. But the approximation error exponent for piecewise constant functions represented in a wavelet basis is $2\alpha = 1$, so $D \sim {P_{tot}}^{-1/2}$ in this case as well.

The last example, shown in Fig. 5, illustrates the universality of CWS. In this case the signal of interest is sparse in a low-frequency subspace of the DCT domain, but only a portion of this subspace is contained in the lowest k frequency components that are directly observed. This situation might arise when the sensors are being used to estimate a communication signal, but the frequency band in which the signal is present is not completely known a priori. The original signal and the signal with measurement noise are shown in the top row. The bottom row shows two estimates, one using the low-frequency DCT measurements and the other obtained from random projections using the DCT domain for reconstruction. The "assumed subspace" approach fails in this case because the subspace being measured does not contain a sufficient amount of the signal energy, while CWS is able to identify the actual subspace and produce an estimate with MSE lower than the measurement noise variance.

6. CONCLUSIONS

In this paper, we have introduced and analyzed the concept of Compressive Wireless Sensing for energy efficient estimation (at FC) of sensor data that is compressible in some basis of $\mathbb{R}^n$ and analyzed, as a function of the number of sensor nodes, the associated power-distortion-latency tradeoffs. CWS is a universal scheme in the sense that it provides us with a consistent field estimation ($D \searrow 0$ as node density increases), even if little or no prior knowledge about the sensed data is assumed, while P_{tot} grows at most sub-linearly with the number of nodes in the network. This universality, however, does come at the cost of optimality in terms of a less favorable power-disortion-latency trade-off which is a direct consequence of not having sufficient prior knowledge about sensed data, forcing us to probe the entire n-dimensional space using random projections instead of focusing our energy on the subspace of interest. Nevertheless, because of this precise reason, CWS has the ability to capture part of signal under all circumstances, whereas projecting the sensor network data onto some subspace, when not enough information is available, can result in a distortion much greater than the one achievable by CWS, as evidenced by the results of Section 5. Therefore, we contend that CWS should be the estimation scheme of choice in cases when either little prior knowledge about the sensed field is available or confidence level about the accuracy of the available knowledge is low.

7. REFERENCES

[1] W. U. Bajwa, A. M. Sayeed, and R. Nowak. Matched source-channel communication for field estimation in wireless sensor networks. In *Proc. 4th Intl. Conf. on Information Processing in Sensor Networks (IPSN '05)*, pages 332–339, Los Angeles, CA, Apr. 2005.

[2] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Submitted, Jun. 2004.

[3] D. L. Donoho. Compressed sensing. Manuscript, Sep. 2004.

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

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

[6] M. Gastpar and M. Vetterli. Power, spatio-temporal bandwidth, and distortion in large sensor networks. *IEEE J. Select. Areas Commun.*, 23(4):745–754, Apr. 2005.

[7] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities.* Cambridge University Press, 1967.

[8] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. Submitted to IEEE Trans. Information Theory, Mar. 2005.

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

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

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

[12] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac. Scalable feedback control for distributed beamforming in sensor networks. In *Proc. Int. Symp. Info. Th. (ISIT'05)*, Sep. 2005.

[13] R. Nowak, U. Mitra, and R. Willett. Estimating inhomogeneous fields using wireless sensor networks. *IEEE J. Select. Areas Commun.*, 22(6):999–1006, 2004.

[14] S. S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed compression in a dense microsensor network. *IEEE Signal Processing Magazine*, 19(2):51–60, Mar. 2002.

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

APPENDIX

A. PROOF OF THEOREM 2

Recall that $D \preceq n^{-2\alpha/(2\alpha+1)}$ requires $k^{-2\alpha} \sim \left(\frac{k}{n}\right)\sigma_w^2 \sim \frac{1}{n}\sum_{i=1}^{k}\frac{\sigma_z^2}{\rho_i}$, resulting in $k \sim n^{1/(2\alpha+1)}$ and

$$k\,\sigma_w^2 \;\sim\; \sum_{i=1}^{k}\sigma_z^2/\rho_i \;\Longleftrightarrow\; \sum_{i=1}^{k}\frac{1}{n_{\psi_i}\lambda_i} \;\sim\; k\,\frac{\sigma_w^2 P}{\sigma_z^2(B^2+\sigma_w^2)} \tag{38}$$

Therefore, the statement of the theorem can be proved by finding a solution of the following optimization problem

$$\begin{aligned} \min \qquad & \Gamma \;=\; \sum_{i=1}^{k} n_{\psi_i}\lambda_i \\ \text{s.t.} \qquad & \sum_{i=1}^{k}\frac{1}{n_{\psi_i}\lambda_i} \;=\; k\,\frac{\sigma_w^2 P}{\sigma_z^2(B^2+\sigma_w^2)} \end{aligned}$$

From arithmetic-geometric-harmonic means inequality [7], we have that

$$\frac{1}{k}\sum_{i=1}^{k} n_{\psi_i}\lambda_i \;\geq\; \frac{k}{\sum_{i=1}^{k}\frac{1}{n_{\psi_i}\lambda_i}} \tag{39}$$

and since $\sum_{i=1}^{k}\frac{1}{n_{\psi_i}\lambda_i}$ is constrained to be $k\,\frac{\sigma_w^2 P}{\sigma_z^2(B^2+\sigma_w^2)}$, we get

$$\Gamma \;=\; \sum_{i=1}^{k} n_{\psi_i}\lambda_i \;\geq\; k\,\frac{\sigma_z^2(B^2+\sigma_w^2)}{\sigma_w^2 P} \tag{40}$$

Moreover, the inequality in (39) reduces to an equality if and only if [7]

$$n_{\psi_1}\lambda_1 = \cdots = n_{\psi_k}\lambda_k \;=\; \frac{\sigma_z^2(B^2+\sigma_w^2)}{\sigma_w^2 P} \tag{41}$$

Thus, by putting $k = n^{-2\alpha/(2\alpha+1)}$ in (40), we get the first part of theorem and (41) implies that $\{\lambda_i = \sigma_z^2(B^2+\sigma_w^2)/(n_{\psi_i}\sigma_w^2 P) \sim 1/n_{\psi_i}\}_{i=1}^{k}$ is the only set of λ_i's that achieves the lower bound for Γ.