Joint Source-Channel Communication for Distributed Estimation in Sensor Networks

Waheed U. Bajwa, *Student Member, IEEE,* Jarvis D. Haupt, *Student Member, IEEE,*
Akbar M. Sayeed, *Senior Member, IEEE,* and Robert D. Nowak, *Senior Member, IEEE*

***Abstract*— Power and bandwidth are scarce resources in dense wireless sensor networks and it is widely recognized that joint optimization of the operations of sensing, processing and communication can result in significant savings in the use of network resources. In this paper, a distributed joint source-channel communication architecture is proposed for energy-efficient estimation of sensor field data at a distant destination and the corresponding relationships between power, distortion, and latency are analyzed as a function of number of sensor nodes. The approach is applicable to a broad class of sensed signal fields and is based on distributed computation of appropriately chosen projections of sensor data at the destination – phase-coherent transmissions from the sensor nodes enable exploitation of the distributed beamforming gain for energy efficiency. Random projections are used when little or no prior knowledge is available about the signal field. Distinct features of the proposed scheme include: 1) processing and communication are combined into one distributed projection operation; 2) it virtually eliminates the need for in-network processing and communication; 3) given sufficient prior knowledge about the sensed data, consistent estimation is possible with increasing sensor density even with vanishing total network power; and 4) consistent signal estimation is possible with power and latency requirements growing at most sub-linearly with the number of sensor nodes even when little or no prior knowledge about the sensed data is assumed at the sensor nodes.**

***Index Terms*— Compressive sampling, distributed beamforming, scaling laws, sensor networks, source-channel communication, sparse signals**

I. Introduction

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 devices that have the ability to self-organize into a well-connected network. A typical wireless sensor network (WSN), as shown in Fig. 1, consists of a large number of wireless sensor nodes, spatially distributed over a region of interest, that can sense (and potentially actuate) the physical environment in a variety of modalities, including acoustic, seismic, thermal, and infrared. A wide range of applications of sensor networks are being envisioned in a number of areas, including geographical monitoring (e.g., habitat monitoring, precision agriculture), industrial control (e.g., in a power plant or a submarine), business management (e.g., inventory tracking with radio frequency identification tags), homeland security (e.g., tracking and classifying moving targets) and health care (e.g., patient monitoring, personalized drug delivery).

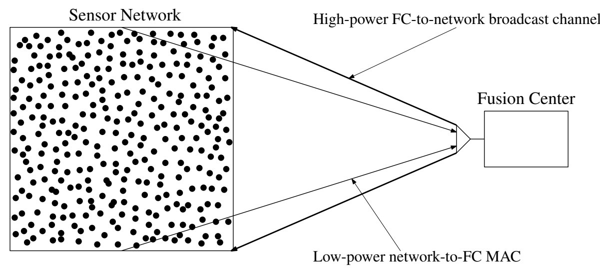


Fig. 1. Sensor network with a fusion center (FC). Black dots denote sensor nodes. FC can communicate to the network over a high-power broadcast channel whereas the multiple-access channel (MAC) from the network to the FC is power constrained.

The essential task in many applications of sensor networks is to extract relevant information about the sensed data and deliver it with a desired fidelity to a (usually) distant destination, termed as the fusion center (FC). The overall goal in the design of sensor networks is to execute this task with least consumption of network resources – energy and bandwidth being the most limited resources, typically. In this regard, the relevant metrics of interest are: (i) the average total network power consumption P_{tot} for estimating a snapshot of the signal field; (ii) the distortion D in the estimate; and (iii) the latency L incurred in obtaining the estimate (defined as the number of network-to-FC channel uses per snapshot). It is also generally recognized that jointly optimizing the operations of sensing, processing and communication can lead to very energy efficient operation of sensor networks.

In this paper, we propose a distributed joint source-channel communication architecture for energy efficient estimation of sensor field data at the FC. Under mild assumptions on the *spatial smoothness* of the signal field (cf. Section II), we analyze the corresponding relationships between power, distortion, and latency as well as their scaling behavior with the number of sensor nodes. Our approach is inspired by recent results in wireless communications [1]–[3] and represents a new, non-traditional attack on the problem of sensing, processing and communication in distributed wireless sensing systems. Rather than digitally encoding and transmitting samples from individual sensors, we consider an alternate encoding paradigm

This work was supported in part by the National Science Foundation under Grants CCF-0431088, CCR-0350213, CNS-0519824 and ECS-0529381. This paper was presented in part at the Fifth International Conference on Information Processing in Sensor Networks, Nashville, TN, April 2006 and at IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, May 2006.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (E-mails: bajwa@cae.wisc.edu, jdhaupt@wisc.edu, {akbar, nowak}@engr.wisc.edu).

based on the projections of samples from many sensors onto appropriate spatial basis functions (e.g., local polynomials, wavelets). The joint source-channel communication architecture at the heart of our approach is an energy efficient method for communicating such projections to the FC – the projections are communicated in a phase-coherent fashion over the network-to-FC multiple-access channel (MAC). This architecture was first proposed and analyzed in [2] in the context of spatially homogeneous signal fields. This paper generalizes the approach to a broader class of signals classified as either *compressible* or *sparse* (see Section II).

The power of the proposed approach is that, in principle, one can choose to acquire samples in the domain of any basis that is particularly well-suited to the spatial structure of the signal field being sensed (e.g., smooth signals tend to be well-approximated in the Fourier basis and wavelet bases tend to be well-suited for the approximation of piecewise smooth signals [4]). Thus, if one has reasonable prior knowledge about the signal (e.g., spatial statistics or smoothness characteristics of the sensed field), then each sensing operation maximizes the potential gain in information per sample. More generally, however, we may have little prior knowledge about the sensed field. And, in some applications, the physical phenomenon of interest may contain time-varying spatial edges or boundaries that separate very different physical behaviors in the measured signal field (e.g., an oceanic oil spill, limited spatial distributions of hazardous biochemical agents). To handle such scenarios, we introduce the concept of *compressive wireless sensing* (CWS) in the later part of the paper that is inspired by recent results in compressive sampling theory [5]–[7] and fits perfectly into our proposed source-channel communication architecture.

The key idea in CWS is that neither the sensor nodes nor the FC need to know/specify the optimal basis elements in advance, and rests on the fact that a relatively small number of random projections of a compressible or sparse signal contain most of its salient information. Thus, in essence, CWS is a universal scheme based on delivering random projections of the sensor data to the FC in an efficient manner. Under the right conditions, the FC can recover a good approximation of the data from these random projections. Nevertheless, this *universality* comes at the cost of a less favorable power-distortion-latency relationship that is a direct consequence of not exploiting prior knowledge of the signal field in the choice of projections that are communicated to the FC. This trade-off between universality and prior knowledge in CWS is quantified in Section VI.

A. Relationship to Previous Work

First, let us comment on the signal model being used in this paper. We assume that the physical phenomenon under observation is characterized by an unknown but deterministic sequence of vectors in $\mathbb{R}^n$, where each vector in the sequence is α-compressible or M-sparse in some orthonormal basis of $\mathbb{R}^n$ (see Section II). Alternative assumptions that are commonly used in previous work are that the signal field is either a realization of a stationary (often bandlimited) random field with some known correlation function [8]–[11], or it is fully described by a certain number of degrees of freedom (often less than n) that are random in nature [3], [12]. All of these signal models, however, express a notion of smoothness or complexity in the signal field, and the decay characteristics of the correlation function (e.g., the rate of decay) or the number of degrees of freedom (DoF) in the field play a role analogous to that of α and M in this work. Essentially, the choice between a deterministic or a stochastic model is mostly a matter of taste and mathematical convenience, the latter being more prevalent when it comes to information-theoretic analysis of the problem (also, see [13] and the discussion therein). However, the deterministic formulation can be more readily generalized to include inhomogeneities, such as boundaries, in the signal field [14].

Secondly, it is generally recognized that the basic operations of sensing, processing (computation), and communication in sensor networks are interdependent and, in general, they must be jointly optimized to attain optimal trade-offs between power, distortion and latency. This joint optimization may be viewed as a form of distributed joint source-channel communication (or coding), involving both estimation (compression) and communication. Despite the need for optimized joint source-channel communication, our fundamental understanding of this complex problem is very limited, owing in part to the absence of a well-developed network information theory [15]. As a result, a majority of research efforts have tried to address either the compression or the communication aspects of the problem. Recent results on joint source-channel communication for distributed estimation or detection of sources in sensor networks [1]–[3], [12], [16]–[18], although relatively few, are rather promising and indicate that limited node cooperation can sometimes greatly facilitate optimized source-channel communication and result in significant energy savings that more than offset the cost of cooperation. Essentially, for a given signal field, the structure of the optimal estimator dictates the structure of the corresponding communication architecture. To the best of our knowledge, the most comprehensive treatment of this problem to date (in the context of WSNs) has been carried out by Gastpar and Vetterli in [3] (see also [12]). While some of our work is inspired by and similar in spirit to [3], Gastpar and Vetterli have primarily studied the case of finite number of independent sources that is analogous to that of an M-sparse signal, albeit assuming Gaussian DoF and multiple FCs. Moreover, the number of DoF in [3] is assumed to be fixed and does not scale with the number of nodes in the network. Our work, in contrast, not only extends the results of [3] to the case when the number of DoF of an M-sparse signal scales with n, but also applies to a broader class of signal fields and gives new insights into the power-distortion-latency relationships for both compressible and sparse signals (cf. Section V). Furthermore, we also present extensions of our methodology to situations in which very limited prior information about the signal field is available.

Thirdly, in the context of compressive sampling theory [5]–[7], while the idea of using random projections for the estimation of sensor network data has recently received some attention in the sensor networking community, the focus has primarily been on the compression or estimation aspects of the problem (see, e.g., [7], [19]–[21]), and this paper is the first to carefully investigate the potential of using random projections from a source-channel communication perspective (cf. Section VI).

Finally, from an architectural and protocol viewpoint, most existing works in the area of sensor data estimation emphasize the networking aspects by focusing on multi-hop communication schemes and in-network data processing and compression (see, e.g., [8], [10], [11], [14]). This typically requires a significant level of networking infrastructure (e.g., routing algorithms), and existing works generally assume this infrastructure as given. Our approach, in contrast to these methods, eliminates the need for in-network communications and processing, and instead requires phase synchronization among nodes that imposes a relatively small burden on network resources and can be achieved, in principle, by employing distributed synchronization/beamforming schemes, such as those described in [22], [23]. Although we use the common term '*sensor network*' to refer to such systems, the systems we envision often act less like networks and more like coherent ensembles of sensors and thus, our proposed wireless sensing system is perhaps more accurately termed a '*sensor ensemble*' that is appropriately queried by an '*information retriever*' (FC) to acquire the desired information about the sensed data.

B. Notational Convention

We establish scaling relationships between different quantities that are denoted by the symbols $\preceq$, $\asymp$ and $\sim$ (read as '*big-oh*', '*asymptotically equivalent*' and '*of-the-order of*' respectively). Specifically, if $f(n), g(n)$ and $h(n)$ are positive-valued functions of $n \in \mathbb{N}$, then we write $f \preceq g$ if there exists a constant $C > 0$ such that $f(n) \leq C\, g(n)\ \forall\ n \in \mathbb{N}$, $f \asymp g$ if $f \preceq g$ and $g \preceq f$ and $f \sim g$ if $f \preceq h$ and $g \preceq h$. Sometimes, we also use the more standard notation $f = \mathcal{O}(g)$ for both *big-oh* and *asymptotically equivalent* scaling relations. Finally, we use $|A|$ to denote the cardinality of a finite set A and $\triangleq$ to mean '*equality by-virtue-of definition*'.

C. Organization

The rest of this paper is organized as follows. In Section II, we describe the system model and associated assumptions on the signal field and the communication channel. In particular, in Section II-A, we formalize the notions of compressible and sparse signals. In Section III, we review the optimal distortion scaling benchmarks for compressible and sparse signals under the assumption that the sensor measurements are available to the FC without any added cost or noise due to communications. In Section IV, we develop the basic building block in our source-channel communication architecture for computing and communicating projections of the sensor field data to the FC. Using this basic building block, we describe and analyze an energy efficient distributed estimation scheme in Section V that achieves the distortion scaling benchmarks of Section III for both compressible and sparse signals under the assumption of sufficient prior knowledge about the compressing (and sparse) basis. In Section VI, we introduce the concept of CWS for the case when sufficient prior knowledge about the compressing/sparse basis is not available and analyze the associated power-distortion-latency scaling laws. Up to this point, we operate under the assumptions that the network is fully synchronized and transmissions from the sensor nodes do not undergo fading. We relax these assumptions in Section VII and study the impact of fading and imperfect phase synchronization on the scaling laws obtained in Sections IV, V and VI. Finally, we present some simulation results in Section VIII to illustrate the proposed methodologies and concluding remarks are provided in Section IX.

II. System Model and Assumptions

We begin by considering a WSN with n nodes observing some physical phenomenon in space and discrete-time[1], where each node takes a noisy sample at time index k of the form

$$x_j^k \;=\; s_j^k + w_j^k, \quad j = 1, \ldots, n,\ k \in \mathbb{N}, \tag{1}$$

and the noiseless samples $\left\{s_j^k,\ k \in \mathbb{N}\right\}$ at each sensor correspond to a deterministic *but unknown* sequence in $\mathbb{R}$. We further assume that $|s_j^k| \leq B$ $(\forall\ j = 1, \ldots, n,\ k \in \mathbb{N})$ for some known constant $B > 0$ that is determined by the sensing range of the sensors, and the measurement errors $\left\{w_j^k\right\}$ are zero-mean Gaussian random variables with variance σ_w^2 that are independent and identically distributed (i.i.d.) across space and time.

Notice that the observed data $\left\{x_j^k = s_j^k + w_j^k\right\}_{j=1}^n$ at time k can be considered as a vector $\boldsymbol{x}^k \in \mathbb{R}^n$ such that $\boldsymbol{x}^k = \boldsymbol{s}^k + \boldsymbol{w}^k$, where $\boldsymbol{s}^k \in \mathbb{R}^n$ is the noiseless data vector and $\boldsymbol{w}^k \sim \mathcal{N}(\boldsymbol{0}_{n\times 1}, \sigma_w^2 \boldsymbol{I}_{n\times n})$ is the measurement noise vector. Therefore, the physical phenomenon under observation can be characterized by the deterministic but unknown sequence of n-dimensional vectors

$$\boldsymbol{S} \;\triangleq\; \left\{\boldsymbol{s}^k\right\}_{k\in N} = \left\{\boldsymbol{s}^1, \boldsymbol{s}^2, \ldots\right\}. \tag{2}$$

Furthermore, we assume no dependence between different time snapshots of the physical phenomenon. Note that if we were to model $\boldsymbol{S}$ as a stochastic signal, this would be equivalent to saying that $\boldsymbol{S}$ is a discrete (vector-valued) memoryless source.

A. Sensor Data Model

It is a well-known fact in the field of transform coding that real-world signals can often be efficiently approximated and encoded in terms of Fourier, wavelet or other related transform representations [13], [24]–[27]. For example, smooth signals can be accurately approximated using a truncated Fourier or wavelet series, and signals and images of bounded variation can be represented very well in terms of a relatively small number of wavelet coefficients [4], [6], [28]. Indeed, features such as smoothness and bounded variation are found in images, video, audio, and various other types of data, as evident

[1]The discrete-time model is an abstraction of the fact that the field is being temporally sampled at some rate of T_s seconds that depends upon the physics of the observed phenomenon.

from the success of familiar compression standards such as JPEG, MPEG and MP3 that are based on Fourier and wavelet transforms.

We take the transform coding point of view in modeling the signal observed by the sensor nodes. Specifically, we assume that the physical phenomenon described by $\boldsymbol{S}$ is (deterministic and) spatially compressible in the sense that each noiseless snapshot $\boldsymbol{s}^k$ is well-approximated by a linear combination of m vectors taken from an orthonormal basis of $\mathbb{R}^n$. We formalize this notion in the following definition.

Definition 1 (Compressible Signals): Let $\boldsymbol{\Psi} \triangleq \{\boldsymbol{\psi}_i\}_{i=1}^n$ be an orthonormal basis of $\mathbb{R}^n$. Denote the coefficients of $\boldsymbol{s}^k$ in this basis (inner products between $\boldsymbol{s}^k$ and the basis vectors $\boldsymbol{\psi}_i$) by $\theta_i^k \triangleq \boldsymbol{\psi}_i{}^T \boldsymbol{s}^k = \sum_{j=1}^n \psi_{ij} s_j^k$, where $(\cdot)^T$ represents the transpose operation. Re-index these coefficients and the corresponding basis vectors so that

$$|\theta_1^k| \geq |\theta_2^k| \geq \cdots \geq |\theta_n^k|. \tag{3}$$

The *best m-term approximation* of $\boldsymbol{s}^k$ in terms of $\boldsymbol{\Psi}$ is given by

$$\boldsymbol{s}^{k,(m)} \triangleq \sum_{i=1}^m \theta_i^k \, \boldsymbol{\psi}_i \,, \tag{4}$$

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

$$\begin{aligned} \frac{1}{n} \left\| \boldsymbol{s}^k - \boldsymbol{s}^{k,(m)} \right\|_2^2 &\triangleq \frac{1}{n} \sum_{j=1}^n \left(s_j^k - s_j^{k,(m)} \right)^2 \\ &\leq C_o \, m^{-2\,\alpha}, \quad k \in \mathbb{N}, \end{aligned} \tag{5}$$

for some constants $C_o > 0$ and $\alpha \geq 1/2$, where the parameter α governs the degree to which $\boldsymbol{S}$ is compressible with respect to $\boldsymbol{\Psi}$.

In addition, we will also consider the special case where, instead of being merely compressible, $\boldsymbol{S}$ is spatially sparse in the sense that each noiseless temporal sample $\boldsymbol{s}^k$ can be fully described by a few $\boldsymbol{\Psi}$-coefficients. We formalize this notion as follows.

Definition 2 (Sparse Signals): We say that $\boldsymbol{S}$ is M-sparse in $\boldsymbol{\Psi}$ (or that $\boldsymbol{\Psi}$ is the M-sparse basis of $\boldsymbol{S}$) if

$$\boldsymbol{s}^k = \sum_{i \in \mathcal{I}^k} \theta_i^k \, \boldsymbol{\psi}_i \,, \quad k \in \mathbb{N}, \tag{6}$$

where $\mathcal{I}^k \subset \{1, 2, \ldots, n\}$, $k \in \mathbb{N}$, and $\max\limits_k |\mathcal{I}^k| \leq M < n$, i.e., each noiseless data vector $\boldsymbol{s}^k$ has at most $M < n$ non-zero coefficients corresponding to some basis $\boldsymbol{\Psi}$ of $\mathbb{R}^n$.

Remark 1: An equivalent definition of compressibility or sparsity may be defined by assuming that, for some $0 < p \leq 1$ and some $R = R(n) > 0$, the $\boldsymbol{\Psi}$-coefficients of $\boldsymbol{s}^k$ belong to an ℓ_p ball of radius R [6], [29], [30], i.e.,

$$\left(\sum_{j=1}^n |\theta_j^k|^p \right)^{1/p} \leq R, \quad k \in \mathbb{N}\,. \tag{7}$$

To see that this is indeed an equivalent definition, first note that (7) can hold only if the cardinality of the set $\{\theta_j^k : |\theta_j^k| > 1/N,\, N \in \mathbb{N},\, j = 1, 2, \ldots, n\}$ is upper bounded by $R\,N^{1/p}$ [30], [31]. Hence, the ℓ_p constraint of (7) in turn requires that the j-th largest (and re-indexed according to magnitude) coefficient θ_j^k is smaller than or equal to $R\,j^{-1/p}$, resulting in

$$\begin{aligned} \left\| \boldsymbol{s}^k - \boldsymbol{s}^{k,(m)} \right\|_2^2 &= \sum_{j=m+1}^n |\theta_j^k|^2 \\ &\leq C_p \, R^2 \, m^{1-2/p}, \quad k \in \mathbb{N}, \end{aligned} \tag{8}$$

for some constant C_p that depends only on p [6], [29], [30]. Thus, our definition of compressible signals is equivalent to assuming that the ordered $\boldsymbol{\Psi}$-coefficients of each noiseless data vector $\boldsymbol{s}^k$ exhibit a power law decay

$$|\theta_j^k| \leq R\,j^{-1/p}, \quad j = 1, \ldots, n,\ k \in \mathbb{N}, \tag{9}$$

where $1/p = \alpha + 1/2$ and $R = \sqrt{\frac{n\,C_o}{C_p}}$ in our case (cf. (5), (8)). Indeed, power law decays like this arise quite commonly in nature and we refer the readers to [6], [13], [29], [31] for some of those instances. Finally, with regard to the notion of sparsity, note that the ℓ_p constraint of (7) simply reduces to measuring the number of non-zero $\boldsymbol{\Psi}$-coefficients as $p \to 0$ and thus, corresponds to our definition of sparse signals with $R = M$.[2]

Remark 2: The above sensor data model can be relaxed to allow temporal dependence between time snapshots of the physical phenomenon by assuming *spatio-temporal* compressibility (or sparsity) of the source signal in an appropriate space-time basis. While a detailed analysis of this setup is beyond the scope of this paper, some of the techniques presented in this paper can be extended to incorporate this scenario.

Remark 3: Note that while this paper is not concerned with the issue of sensor placement (sampling) in the signal field, the choice of a good compressing basis is inherently coupled with the sensors' locations within the WSN. For example, while Fourier basis would suffice as a compressing basis for a sensor network observing a smooth signal field in which sensors are placed on a uniform grid, random (irregular) placement of sensors within the same field may warrant the use of an irregular wavelet transform as the appropriate compressing basis [32].

B. Communication Setup

Given the observation vector $\boldsymbol{x}^k$ at time k, the aim of the sensor nodes (and the network as a whole) is to communicate a reliable-enough estimate $\widehat{\boldsymbol{s}}^k$ of the noiseless data vector $\boldsymbol{s}^k$ to a distant FC, where the reliability is measured in terms of the mean-squared error (MSE). Before proceeding further, however, we shall make the following assumptions concerning communications between the sensor nodes and the FC:

1) Each sensor and the FC are equipped with a single omni-directional antenna and sensors communicate to the FC over a narrowband additive white Gaussian noise (AWGN) multiple-access channel (MAC), where each channel use is characterized by transmission over a

[2] For an M-sparse signal, no particular decay structure is assumed for the M non-zero coefficients of $\boldsymbol{s}^k$ in $\boldsymbol{\Psi}$.

period of T_c seconds. Furthermore, the FC can communicate to the sensor nodes over an essentially noise-free broadcast channel.

2) Transmissions from the sensor nodes to the FC do not suffer any fading [33]–[35], which would indeed be the case in many remote sensing applications, such as desert border monitoring, with little or no scatterers in the surrounding environment and static sensor nodes having a strong line-of-sight connection to the FC [36].
3) Each sensor knows its distance from the FC and thus, can calculate the channel path gain $\sqrt{h_j}$ given by [33]–[35]

$$\sqrt{h_j} \;\triangleq\; \frac{1}{{d_j}^{\zeta/2}}\,, \quad j = 1, 2, \ldots, n, \tag{10}$$

where $1 \le d_j \le d_u < \infty$ is the distance between the sensor at location j and the FC, and $\zeta \ge 2$ is the path-loss exponent [36], [37]. In principle, even when the distances and/or path loss exponent are unknown, these channel gains could be estimated at the FC using received signal strength and communicated back to the sensors during network initialization.
4) The network is fully synchronized with the FC in the following sense [34], [35]: (i) *Carrier Synchronization*: All sensors have a local oscillator synchronized to the receiver carrier frequency; (ii) *Time Synchronization*: For each channel use, the relative timing error between sensors' transmissions is much smaller than the channel symbol duration T_c; and (iii) *Phase Synchronization*: Sensors' transmissions arrive at the FC in a phase coherent fashion, which can be achieved by employing the distributed phase synchronization schemes described in [22], [23].
5) Sensor transmissions are constrained to a sum transmit power of P per channel use. Specifically, let y_j be the transmission of sensor j in any channel use. Then, it is required that

$$\sum_{j=1}^{n} \mathbb{E}\left[|y_j|^2\right] \;\le\; P. \tag{11}$$

6) The network is allowed L network-to-FC channel uses per source observation, which we term as the latency of the system. If, for example, these L channel uses were to be employed using time division multiple access (TDMA) then this would require that the temporal sampling time $T_s \ge L\,T_c$; hence, the term latency. In a system with no bandwidth constraints, this could also be interpreted as the effective bandwidth of the network-to-FC MAC.

Given this communication setup, an estimation scheme corresponds to designing n source-channel encoders $(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n)$ – one for each sensor node, and the decoder $\boldsymbol{G}$ for the FC such that at each time instant k, given the observations $\left\{x_j^{\kappa}\right\}_{\kappa=1}^{k}$ up to time k at node j, the encoders generate an L-tuple $\boldsymbol{y}_j^k \triangleq \boldsymbol{F}_j\left(\left\{x_j^{\kappa}\right\}_{\kappa=1}^{k}\right) = \left(y_{j1}^k, \ldots, y_{jL}^k\right)^T$ corresponding to L-channel uses per source observation (that also satisfy the power constraint of (11)). And at

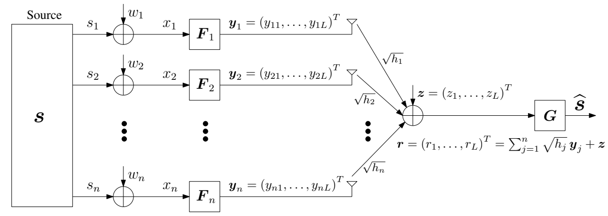


Fig. 2. L-channel use snapshot of the sensor network per source observation. The superscript corresponding to the time index has been dropped in the figure to simplify notation.

the end of the L-th channel use, the decoder $\boldsymbol{G}$ produces an estimate $\widehat{\boldsymbol{s}}^k$ of the noiseless data vector $\boldsymbol{s}^k$ given by $\widehat{\boldsymbol{s}}^k \triangleq \boldsymbol{G}\left(\{\boldsymbol{r}^{\kappa}\}_{\kappa=1}^{k}\right)$, where $\boldsymbol{r}^{\kappa} = \sum_{j=1}^{n} \sqrt{h_j}\,\boldsymbol{y}_j^{\kappa} + \boldsymbol{z}^{\kappa}$ and $\boldsymbol{z}^{\kappa} \sim \mathcal{N}(\boldsymbol{0}_{L\times 1}, \sigma_z^2 \boldsymbol{I}_{L\times L})$ is the MAC AWGN vector corresponding to the L-channel uses at time instant κ (see Fig. 2), and the goal of the sensor network is to minimize (i) the average total network power consumption per source observation

$$P_{\text{tot}} \;\triangleq\; \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sum_{j=1}^{n} \mathbb{E}\left[\left|y_{j\ell}^k\right|^2\right]; \tag{12}$$

(ii) the mean-squared error distortion measure

$$D \;\triangleq\; \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}^k\right\|_2^2\right]; \tag{13}$$

and (iii) the latency L (# of channel uses per source observation) of the system.[3] Thus, for a fixed number of sensor nodes n, the performance of any estimation scheme is characterized by the triplet $(P_{\text{tot}}(n), D(n), L(n))$ and rather than obtaining an exact expression for this triplet, our goal would be to analyze how do these three quantities scale with n for a given scheme. Moreover, minimization of all three quantities in the triplet is sometimes a conflicting requirement and there is often a trade-off involved between minimizing P_{tot}, D and L, and we shall also be analyzing this power-distortion-latency trade-off as a function of n.

Remark 4: Notice that implicit in this formulation is the fact that no collaboration among the sensor nodes is allowed for the purposes of signal estimation, i.e., encoder $\boldsymbol{F}_j$ does not have access to the inputs of any sensor other than sensor j.

Remark 5: Note that while stating the performance metrics of power and latency, we have ignored the cost of initializing the sensor network (primarily corresponding to the cost of channel gain estimation/phase synchronization algorithms under the current communication setup and the cost of initial route/topology discovery algorithms under the more traditional multi-hop communication setups). This is because the average cost of this initialization (over time) tends to zero as k – the time scale of the network operation – tends to infinity. Of course, in practice, a one-time initialization may not suffice and these procedures may have to be repeated from time to

[3]Notice that with the distortion metric as defined in (13), the MSE of any arbitrary length signal can at worst be a constant since $\lim\limits_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k\right\|_2^2\right] \le B^2$.

time, but we will assume that the corresponding costs are negligible compared to the routine sensing and communication operations.

III. Optimal Distortion Scaling in a Centralized System

In this section, we consider a system in which the sensor measurements $\left\{x_j^k\right\}_{j=1}^n$ at each time instant k are assumed to be available at the FC with no added cost or noise due to communications, and we review the corresponding classical estimation theory results (see, e.g., [13], [38], [39]). Note that such a system corresponds to a sensor network with a noise-free network-to-FC MAC and thus, the optimal distortion scaling achievable under this *centralized* setting serves as a benchmark for assessing the distortion related performance of any scheme under the original setup.

A. Compressible Signals

Given the observation vector $\boldsymbol{x}^k$ at the FC, an optimal centralized estimator for an α-compressible signal can be easily constructed by projecting $\boldsymbol{x}^k$ onto the m basis vectors of $\boldsymbol{\Psi}$ corresponding to m largest (in the absolute sense) $\boldsymbol{\Psi}$-coefficients of $\boldsymbol{s}^k$ (see, e.g., [13]), i.e., if $\boldsymbol{\Psi}_m^k$ is the $n \times m$ matrix of those basis vectors, where the superscript k indicates that the re-indexing in (3) may be a function of the time index k, then $\boldsymbol{s}^k$ can be estimated as

$$\begin{aligned} \widehat{\boldsymbol{s}}_{\text{cen}}^k &\triangleq \boldsymbol{\Psi}_m^k \left({\boldsymbol{\Psi}_m^k}^T \boldsymbol{x}^k\right) \\ &= \boldsymbol{s}^{k,(m)} + \boldsymbol{\Psi}_m^k \left({\boldsymbol{\Psi}_m^k}^T \boldsymbol{w}^k\right), \end{aligned} \tag{14}$$

which results in

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}_{\text{cen}}^k\right\|_2^2\right] &= \frac{1}{n}\left\|\boldsymbol{s}^k - \boldsymbol{s}^{k,(m)}\right\|_2^2 + \\ &\qquad \frac{1}{n}\mathbb{E}\left[\left\|\boldsymbol{\Psi}_m^k \left({\boldsymbol{\Psi}_m^k}^T \boldsymbol{w}^k\right)\right\|_2^2\right] \qquad (15) \\ &\le C_o\, m^{-2\alpha} + \left(\frac{m}{n}\right)\sigma_w^2 . \qquad (16) \end{aligned}$$

Furthermore, from (15), we also have the trivial lower bound of

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}_{\text{cen}}^k\right\|_2^2\right] &\ge \frac{1}{n}\mathbb{E}\left[\left\|\boldsymbol{\Psi}_m^k \left({\boldsymbol{\Psi}_m^k}^T \boldsymbol{w}^k\right)\right\|_2^2\right] \\ &= \left(\frac{m}{n}\right)\sigma_w^2 \end{aligned} \tag{17}$$

and combining the upper and lower bounds of (16) and (17), we obtain

$$\left(\frac{m}{n}\right)\sigma_w^2 \le \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}_{\text{cen}}^k\right\|_2^2\right] \le C_o\, m^{-2\alpha} + \left(\frac{m}{n}\right)\sigma_w^2 . \tag{18}$$

From this expression, we see that the choice of m affects the classic bias-variance trade-off [39]: increasing m causes the bound $C_o\, m^{-2\alpha}$ on the approximation error $\frac{1}{n}\left\|\boldsymbol{s}^k - \boldsymbol{s}^{k,(m)}\right\|_2^2$ (the squared "bias") to decrease, but causes the stochastic component of the error due to the measurement noise $\frac{1}{n}\mathbb{E}\left[\left\|\boldsymbol{\Psi}_m^k \left({\boldsymbol{\Psi}_m^k}^T \boldsymbol{w}^k\right)\right\|_2^2\right] = \left(\frac{m}{n}\right)\sigma_w^2$ (the "variance") to increase. The upper bound is tight, in the sense that there exist signals for which the upper bound is achieved, and in such cases the upper bound is minimized (by choice of m) by making the approximation error and the stochastic component of the error scale at the same rate, i.e.,

$$m^{-2\alpha} \asymp \frac{m}{n} \iff m \asymp n^{1/(2\alpha+1)}, \tag{19}$$

resulting in the following expression for optimal distortion scaling of an α-compressible signal in a centralized system[4]

$$D_{\text{cen}}^* = \lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}_{\text{cen}}^k\right\|_2^2\right] \asymp n^{-2\alpha/(2\alpha+1)}. \tag{20}$$

B. Sparse Signals

Similar to a compressible signal, an optimal centralized estimator for an M-sparse signal corresponds to projecting the observation vector onto the M basis vectors of $\boldsymbol{\Psi}$ corresponding to M non-zero $\boldsymbol{\Psi}$-coefficients of $\boldsymbol{s}^k$ (see, e.g., [38]), i.e., if $\boldsymbol{\Psi}_M^k$ is the $n \times M$ matrix of those basis vectors, then $\boldsymbol{s}^k$ can be estimated as

$$\begin{aligned} \widehat{\boldsymbol{s}}_{\text{cen}}^k &\triangleq \boldsymbol{\Psi}_M^k \left({\boldsymbol{\Psi}_M^k}^T \boldsymbol{x}^k\right) \\ &= \boldsymbol{s}^k + \boldsymbol{\Psi}_M^k \left({\boldsymbol{\Psi}_M^k}^T \boldsymbol{w}^k\right), \end{aligned} \tag{21}$$

which results in the usual parametric rate

$$\begin{aligned} \mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k - \widehat{\boldsymbol{s}}_{\text{cen}}^k\right\|_2^2\right] &= \frac{1}{n}\mathbb{E}\left[\left\|\boldsymbol{\Psi}_M^k \left({\boldsymbol{\Psi}_M^k}^T \boldsymbol{w}^k\right)\right\|_2^2\right] \\ &= \left(\frac{M}{n}\right)\sigma_w^2, \end{aligned} \tag{22}$$

resulting in the following expression for optimal distortion scaling of an M-sparse signal in a centralized system

$$D_{\text{cen}}^* = \left(\frac{M}{n}\right)\sigma_w^2 \asymp \frac{M}{n}. \tag{23}$$

Note that it might very well be that the number of DoF of an M-sparse signal scales with the number of nodes n in the network. For example, two-dimensional piecewise constant fields with one-dimensional boundaries separating constant regions can be compressed using the discrete wavelet transform and have $M \asymp n^{1/2}\log(n)$ non-zero wavelet coefficients [14]. Therefore, we model M as $M \asymp n^{\mu}$, where $0 \le \mu < 1$ and hence, the inclusion of M in the scaling relation in (23).

Remark 6: Note that the optimal distortion scaling relations of (20) and (23) for compressible and sparse signals have been obtained under the assumption that the FC has precise knowledge of the ordering of coefficients of $\boldsymbol{s}^k$ in the compressing basis (indices of non-zero coefficients of $\boldsymbol{s}^k$ in the sparse basis). This is not necessarily a problem in a centralized setting and in cases where this information is not available, coefficient thresholding methods can be used to automatically select the appropriate basis elements from the noisy data, and these methods obey error bounds that are within a constant or logarithmic factor of the ones given above (see, e.g., [40], [41]).

[4] $*$ in D_{cen}^* refers to the fact that this is the *optimal* centralized distortion scaling.

IV. Distributed Projections in Wireless Sensor Networks

In this section, we develop the basic communication architecture that acts as a building block of our proposed estimation scheme. As evident from the previous section, each DoF of a compressible or sparse signal corresponds to projection of sensor network data onto an n-dimensional vector in $\mathbb{R}^n$ and at the heart of our approach is a distributed method of communicating such projections to the FC in a power efficient manner by exploiting the spatial averaging inherent in an AWGN MAC.

To begin, assume that the goal of the sensor network is to obtain an estimate of the projection of noiseless sensor data, corresponding to each observation of the physical phenomenon, onto a vector in $\mathbb{R}^n$ at the FC. That is, let us suppose that at each time instant k, we are interested in obtaining an estimate $\widehat{\upsilon}^k$ of

$$\upsilon^k \triangleq \boldsymbol{\varphi}^T \boldsymbol{s}^k = \sum_{j=1}^{n} \varphi_j s_j^k, \tag{24}$$

where $\boldsymbol{\varphi} \in \mathbb{R}^n$. One possibility for realizing this goal is to nominate a clusterhead in the network and then, assuming all the sensor nodes know their respective φ_j's and have constructed routes which form a spanning tree through the network to the clusterhead, each sensor node can locally compute $\varphi_j x_j^k = \varphi_j(s_j^k + w_j^k)$ and these values can be aggregated up the tree to obtain $\widehat{\upsilon}^k = \sum_{j=1}^n \varphi_j x_j^k$ at the clusterhead, which can then encode and transmit this estimate to the FC. However, even if we ignore the communication cost of delivering $\widehat{\upsilon}^k$ from the clusterhead to the FC, it is easy to check that such a scheme requires at least n transmissions. For a similar reason, gossip algorithms such as the ones described in [42], [43], while known for their robustness in the face of changing network topology, might not be the schemes of first choice for these types of applications.

Another, more promising, alternative is to exploit recent results concerning uncoded (analog) coherent transmission schemes in WSNs [1]–[3], [16]. The proposed distributed joint source-channel communication architecture requires only one channel use per source observation ($L_\upsilon = 1$) and is based on the notion of so-called "matched source-channel communication" [2], [3]: the structure of the network communication architecture should "match" the structure of the optimal estimator. Under the current setup, this essentially involves phase-coherent, low-power, analog transmission of appropriately weighted sample values directly from the nodes in the network to the FC via the AWGN network-to-FC MAC and the required projection is implicitly computed at the FC as a result of the spatial averaging in the MAC. In light of the communication setup of Section II, full characterization of this architecture essentially entails characterization of the corresponding *scalar-output* source-channel encoders $(F_1, \dots, F_n)$ at the sensor nodes and the *scalar-input* decoder G at the FC, where scalar nature of the encoders and the decoder is owing to the fact that (by construction) $L_\upsilon = 1$ in this scenario.

To begin with, each sensor encoder F_j in this architecture corresponds to simply multiplying the sensor measurement x_j^k with $\left(\sqrt{\frac{\rho}{h_j}}\,\varphi_j\right)$ to obtain[5]

$$y_j^k \triangleq F_j\left(\{x_j^\kappa\}_{\kappa=1}^k\right) = \sqrt{\frac{\rho}{h_j}}\,\varphi_j x_j^k, \quad j = 1, \dots, n, \tag{25}$$

where $\rho > 0$ is a scaling factor used to satisfy sensors' sum transmit power constraint P, and all the nodes coherently transmit their respective y_j^k's in an analog fashion over the network-to-FC MAC. Under the synchronization assumption of Section II and the additive nature of an AWGN MAC, the corresponding received signal at the FC is given by

$$\begin{aligned} r^k &= \sum_{j=1}^{n} \sqrt{h_j}\, y_j^k + z^k = \sqrt{\rho} \sum_{j=1}^{n} \varphi_j x_j^k + z^k \\ &= \sqrt{\rho}\, \upsilon^k + \left(\sqrt{\rho}\, \boldsymbol{\varphi}^T \boldsymbol{w}^k + z^k\right), \end{aligned} \tag{26}$$

where $z^k \sim \mathcal{N}(0, \sigma_z^2)$ is the MAC AWGN at time k (independent of $\boldsymbol{w}^k$). In essence, the encoders $(F_1, \dots, F_n)$ correspond to delivering to the FC a noisy projection of $\boldsymbol{s}^k$ onto $\boldsymbol{\varphi}$ that is scaled by $\sqrt{\rho}$ (cf. (26)). Given r^k, the decoder G corresponds to a simple re-scaling of the received signal, i.e.,

$$\begin{aligned} \widehat{\upsilon}^k &\triangleq G\left(\{r^\kappa\}_{\kappa=1}^k\right) = \frac{r^k}{\sqrt{\rho}} \\ &= \upsilon^k + \boldsymbol{\varphi}^T \boldsymbol{w}^k + \frac{z^k}{\sqrt{\rho}}. \end{aligned} \tag{27}$$

We are now ready to characterize the power-distortion-latency triplet $(P_{\text{tot},\upsilon}, D_\upsilon, L_\upsilon)$ of the proposed joint source-channel communication architecture for computing distributed projections in WSNs.[6]

Theorem 1: Let $\boldsymbol{\varphi} \in \mathbb{R}^n$ and let $\upsilon^k = \boldsymbol{\varphi}^T \boldsymbol{s}^k$. Given the sensor network model of Section II, the joint source-channel communication scheme described by the encoders in (25) and the decoder in (27) can achieve the following end-to-end distortion by employing only one channel use per source observation

$$\begin{aligned} D_\upsilon &\triangleq \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}\left[\left|\upsilon^k - \widehat{\upsilon}^k\right|^2\right] \\ &= \sigma_w^2 \|\boldsymbol{\varphi}\|_2^2 + \left(\frac{\sigma_z^2\, {d_u}^\zeta \left(B^2 + \sigma_w^2\right)}{\lambda P}\right) \|\boldsymbol{\varphi}\|_2^2, \end{aligned} \tag{28}$$

where $\widehat{\upsilon}^k$ is the estimate of υ^k at the FC, σ_w^2 is the measurement noise variance, σ_z^2 is the channel noise variance, B is the bound on $|s_j^k|$, d_u is the bound on the maximum distance between the sensor nodes and the FC, ζ is the path-loss exponent, P is the sum transmit power constraint per channel use and $\lambda = \lambda(n) \in (0, 1]$ is a design parameter used to control total network power consumption. Moreover,

[5] Practical schemes of how each sensor encoder might get access to its respective φ_j is discussed in Section V-C.

[6] $(P_{\text{tot},\upsilon}, D_\upsilon, L_\upsilon)$ triplet here corresponds to power, distortion and latency of the projection coefficient as opposed to (P_{tot}, D, L) in Section II that corresponds to power, distortion and latency required to estimate the entire signal.

the total network power consumption per source observation associated with achieving this distortion is given by

$$\lambda P \left(\frac{\sigma_w^2}{{d_u}^{\zeta} \left(B^2 + \sigma_w^2 \right)} \right) \;\le\; P_{\text{tot},\upsilon} \;\triangleq\; \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{n} \mathbb{E}\left[\left| y_j^k \right|^2 \right] \;\le\; \lambda P. \quad (29)$$

Proof: To establish this theorem, first observe that (27) implies that $\forall\ k \in \mathbb{N}$

$$\mathbb{E}\left[\left| \upsilon^k - \widehat{\upsilon}^k \right|^2 \right] \;=\; \mathbb{E}\left[\left| \boldsymbol{\varphi}^T \boldsymbol{w}^k + \frac{z^k}{\sqrt{\rho}} \right|^2 \right] \;=\; \sigma_w^2 \, \|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_z^2}{\rho}\,, \quad (30)$$

resulting in the following expression for the projection coefficient MSE

$$D_\upsilon \;=\; \sigma_w^2 \, \|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_z^2}{\rho}\,. \quad (31)$$

As for obtaining an expression for $P_{\text{tot},\upsilon}$, note that (25) implies that $\forall\ k \in \mathbb{N}$,

$$\begin{aligned} \rho\ \sigma_w^2 \sum_{j=1}^{n} |\varphi_j|^2 \;&\le\; \sum_{j=1}^{n} \mathbb{E}\left[\left| y_j^k \right|^2 \right] \\ &=\; \sum_{j=1}^{n} \mathbb{E}\left[\frac{\rho}{h_j} \left(s_j^k + w_j^k \right)^2 |\varphi_j|^2 \right] \\ &\le\; \rho \, {d_u}^{\zeta} \left(B^2 + \sigma_w^2 \right) \sum_{j=1}^{n} |\varphi_j|^2, \end{aligned} \quad (32)$$

and thus,

$$\rho \;=\; \lambda \, P \left(\frac{1}{{d_u}^{\zeta} \left(B^2 + \sigma_w^2 \right) \, \|\boldsymbol{\varphi}\|_2^2} \right) \quad (33)$$

would suffice to satisfy the sum transmit power constraint of (11), where $\lambda = \lambda(n) \in (0, 1]$ is a power scaling factor to be used by the designer of a WSN to control total network power consumption. This in turn results in the following expression for total network power consumption per source observation

$$\lambda P \left(\frac{\sigma_w^2}{{d_u}^{\zeta} \left(B^2 + \sigma_w^2 \right)} \right) \;\le\; P_{\text{tot},\upsilon} \;\le\; \lambda P. \quad (34)$$

Finally, to complete the proof of the theorem, we substitute in (31) the value of ρ from (33) to obtain (28). ■

Notice that the projection coefficient distortion D_υ achieved by the proposed joint source-channel communication architecture has been expressed in terms of two separate contributions (cf. (28), (31)), the first of which is independent of the proposed communication scheme. This term is solely due to the noisy observation process ($\sigma_w^2 \neq 0$) and scales like $\|\boldsymbol{\varphi}\|_2^2$. The second contribution is primarily due to the noisy communication channel and scales like $\|\boldsymbol{\varphi}\|_2^2/\lambda$. Moreover, given the observation model of Section II, it is easy to check that $D_\upsilon^* \asymp \|\boldsymbol{\varphi}\|_2^2$ is the best that any (centralized or distributed) scheme can hope to achieve in terms of an order relation for distortion scaling [38]. Therefore, for optimal distortion scaling, it is sufficient that the second term in (31) also scales like $\|\boldsymbol{\varphi}\|_2^2$ and hence, $\lambda = \mathcal{O}(1)$ would suffice to ensure that

$$D_\upsilon \;\asymp\; \|\boldsymbol{\varphi}\|_2^2 \;\asymp\; D_\upsilon^*\,. \quad (35)$$

Consequently, the total network power consumption associated with achieving this optimal distortion scaling would be given by $P_{\text{tot},\upsilon} \;=\; \mathcal{O}(1)$ (cf. (34)). We summarize this insight as follows.

Corollary 1: Let $\boldsymbol{\varphi} \in \mathbb{R}^n$ and let $\upsilon^k = \boldsymbol{\varphi}^T \boldsymbol{s}^k$. Given the sensor network model of Section II and assuming that the system parameters $(B, \sigma_w^2, \sigma_z^2, d_u, \zeta, P)$ do not vary with the number of nodes n in the network, the joint source-channel communication scheme described by the encoders in (25) and the decoder in (27) can obtain an estimate $\widehat{\upsilon}^k$ of υ^k at the FC, such that $D_\upsilon \asymp \|\boldsymbol{\varphi}\|_2^2 \asymp D_\upsilon^*$, by employing only one channel use per source observation, $L_\upsilon = 1$, and using a fixed amount of total network power, $P_{\text{tot},\upsilon} \;=\; \mathcal{O}(1)$.

Observation 1: While the original problem has been setup under a *fixed* sum transmit power constraint P, one of the significant implications of the preceding analysis is that even if one allows P to grow with the number of nodes in the network – say, e.g., $P = \mathcal{O}(n)$ – one cannot improve on the distortion scaling law of $\mathcal{O}\left(\|\boldsymbol{\varphi}\|_2^2\right)$. In other words, when it comes to estimating a single projection coefficient in the presence of noise, using more than a fixed amount of total power per channel use is wasteful as the distortion due to the measurement noise (first term in (31)) is the limiting factor in the overall distortion scaling.

Observation 2: Even though the joint source-channel communication architecture described in this section is meant to be a building block for the signal estimation scheme, the architecture is important in its own right too. Often times, for example, rather than obtaining an estimate of the noiseless sensor data at the FC, the designer of a WSN is merely interested in obtaining the estimates of a few of its linear summary statistics. And, given that any linear summary statistic is nothing but the projection of noiseless sensor data onto a vector in $\mathbb{R}^n$, preceding analysis implies that one can obtain such linear summary statistics at the FC with minimal distortion (and latency) and consumption of only a small amount of total network power.

Example 1 (Sensor Data Average): To illustrate the idea further, consider a specific case where the designer of a WSN is interested in obtaining an estimate of the average $\bar{s}^k = \frac{1}{n} \sum_{j=1}^{n} s_j^k$ of noiseless sensor data at each time instant k. This would correspond to the projection vector being given by $\boldsymbol{\varphi} = (1/n, \ldots, 1/n)^T$ and thus, using the communication architecture described in this section, an estimate of $\bar{s}^k$ can be obtained at the FC such that $D_{\bar{s}} \;\asymp\; 1/n \;\asymp\; D_{\bar{s}}^*$ (the parametric rate), $L_{\bar{s}} \;=\; 1$ and $P_{\text{tot},\bar{s}} \;=\; \mathcal{O}(1)$.

V. Distributed Estimation from Noisy Projections: Known Subspace

In this section, we build upon the joint source-channel communication architecture of Section IV and using it as a basic building block, present a completely decentralized scheme for efficient estimation of sensor network data at the FC. The analysis in this section is carried out under the assumption that the designer of the WSN has complete knowledge of the basis in which $\boldsymbol{S}$ is compressible (or sparse) as well as precise knowledge of the ordering of its coefficients in the compressing basis (indices of non-zero coefficients in the sparse basis) at each time instant k. We refer to this scenario as the '*known subspace*' case and, under this assumption, analyze the corresponding power-distortion-latency scaling laws of the proposed scheme as a function of number of sensor nodes in the network. As to the question of whether the known subspace assumption is a reasonable one, the answer depends entirely on the underlying physical phenomenon. For example, if the signal is smooth or bandlimited, then the Fourier or wavelet coefficients can be ordered (or partially ordered) from low frequency/resolution to high frequency/resolution. Alternatively, if the physical phenomenon under observation happened to be spatially Hölder smooth at each time instant k, then it would be quite reasonable to treat the resulting sensor network data under the known subspace category (see, e.g., [2], [44]).

A. Estimation of Compressible Signals

To begin with, let $\boldsymbol{\Psi} = \{\boldsymbol{\psi}_i\}_{i=1}^n$ be the compressing basis of $\boldsymbol{S}$ such that $\frac{1}{n}\left\|\boldsymbol{s}^k - \boldsymbol{s}^{k,(m)}\right\|_2^2 = \mathcal{O}(m^{-2\alpha})\ \forall\ k \in \mathbb{N}$. In Section IV, we showed that using the communication scheme described by the encoders in (25) and the decoder in (27), one projection per snapshot can be efficiently communicated to the FC by employing only one channel use ($L_v = 1$). By a simple extension of the encoders/decoder structure of Section IV, however, the network can equally well communicate $L\ (>1)$ projections per snapshot in L consecutive channel uses (one channel use *per projection* per snapshot). Essentially, at each time instant k, the L-tuples generated by the encoders $\boldsymbol{F}_j$ are given by (cf. Section II, Fig. 2)

$$\begin{aligned} \boldsymbol{y}_j^k &= \boldsymbol{F}_j\left(\{x_j^\kappa\}_{\kappa=1}^k\right) = \left(y_{j1}^k, \ldots, y_{jL}^k\right)^T \\ &= \sqrt{\frac{\rho}{h_j}}\left(\psi_{1j}\, x_j^k, \ldots, \psi_{Lj}\, x_j^k\right)^T, \quad j = 1, \ldots, n, \end{aligned} \tag{36}$$

where $\rho = (\lambda P)/({d_u}^\zeta(B^2+\sigma_w^2))$, and at the end of the L-th channel use, the received signal at the input of the decoder $\boldsymbol{G}$ is given by

$$\begin{aligned} \boldsymbol{r}^k &= \sum_{j=1}^n \sqrt{h_j}\,\boldsymbol{y}_j^k + \boldsymbol{z}^k \\ &= \sqrt{\rho}\left(\sum_{j=1}^n \psi_{1j}\, x_j^k, \ldots, \sum_{j=1}^n \psi_{Lj}\, x_j^k\right)^T + \boldsymbol{z}^k \\ &= \sqrt{\rho}\,\boldsymbol{\theta}_L^k + \left(\sqrt{\rho}\,{\boldsymbol{\Psi}_L^k}^T\boldsymbol{w}^k + \boldsymbol{z}^k\right), \end{aligned} \tag{37}$$

where $\boldsymbol{\Psi}_L^k$ is the $n \times L$ matrix of the basis vectors corresponding to L largest (in magnitude) $\boldsymbol{\Psi}$-coefficients of $\boldsymbol{s}^k$, $\boldsymbol{\theta}_L^k \triangleq (\theta_1^k, \ldots, \theta_L^k)^T = {\boldsymbol{\Psi}_L^k}^T\boldsymbol{s}^k$ and $\boldsymbol{z}^k \sim \mathcal{N}(\boldsymbol{0}_{L\times 1}, \sigma_z^2\boldsymbol{I}_{L\times L})$ is the MAC AWGN vector (independent of $\boldsymbol{w}^k$). Thus, at the end of the L-th channel use, the decoder has access to L scaled, noisy projections of $\boldsymbol{s}^k$ onto L distinct elements of $\boldsymbol{\Psi}$ and, using these noisy projections, it produces an estimate of the noiseless data vector $\boldsymbol{s}^k$ given by

$$\begin{aligned} \widehat{\boldsymbol{s}}^k &= \boldsymbol{G}\left(\{\boldsymbol{r}^\kappa\}_{\kappa=1}^k\right) = \boldsymbol{\Psi}_L^k\left(\frac{\boldsymbol{r}^k}{\sqrt{\rho}}\right) \\ &= \boldsymbol{s}^{k,(L)} + \boldsymbol{\Psi}_L^k\left({\boldsymbol{\Psi}_L^k}^T\boldsymbol{w}^k\right) + \frac{\boldsymbol{\Psi}_L^k\,\boldsymbol{z}^k}{\sqrt{\rho}}. \end{aligned} \tag{38}$$

Notice the intuitively pleasing similarity between $\widehat{\boldsymbol{s}}^k$ and $\widehat{\boldsymbol{s}}_{\text{cen}}^k$ (cf. (14), (38)): the first two terms in the above expression correspond identically to the centralized estimate of a compressible signal (with m replaced by L) and the last term is introduced due to the noisy MAC communication. Consequently, this results in the following expression for distortion of a compressible signal at the FC

$$\begin{aligned} \left(\frac{L}{n}\right)\sigma_w^2 + \left(\frac{L}{n}\right)\left(\frac{\sigma_z^2\,{d_u}^\zeta\left(B^2+\sigma_w^2\right)}{\lambda P}\right) &\le \\ D \le C_o\,L^{-2\alpha} + \left(\frac{L}{n}\right)\sigma_w^2 &+ \\ \left(\frac{L}{n}\right)\left(\frac{\sigma_z^2\,{d_u}^\zeta\left(B^2+\sigma_w^2\right)}{\lambda P}\right)&. \end{aligned} \tag{39}$$

Finally, simple manipulations along the lines of the ones in Section IV result in the following expression for total network power consumption

$$\lambda\,L\,P\left(\frac{\sigma_w^2}{{d_u}^\zeta\left(B^2+\sigma_w^2\right)}\right) \le P_{\text{tot}} \le \lambda\,L\,P. \tag{40}$$

The above two expressions essentially govern the interplay between P_{tot}, D and L of the proposed distributed estimation scheme and in the sequel, we shall analyze this interplay in further details.

1) Minimum Power and Latency for Optimal Distortion Scaling: Similar to the case of distortion scaling in the centralized setting, (39) shows that the choice of number of projections per snapshot in the distributed setting also results in a bias-variance trade-off: increasing L causes the bound $C_o\,L^{-2\alpha}$ on the approximation error $\frac{1}{n}\left\|\boldsymbol{s}^k - \boldsymbol{s}^{k,(L)}\right\|_2^2$ to decrease, but causes the stochastic components of the error due to the measurement noise $\frac{1}{n}\,\mathbb{E}\left[\left\|\boldsymbol{\Psi}_L^k\left({\boldsymbol{\Psi}_L^k}^T\boldsymbol{w}^k\right)\right\|_2^2\right] = \left(\frac{L}{n}\right)\sigma_w^2$ and the communication noise $\frac{1}{n}\,\mathbb{E}\left[\left\|\boldsymbol{\Psi}_L^k\,\boldsymbol{z}^k/\sqrt{\rho}\right\|_2^2\right] = \left(\frac{L}{n}\right)\left(\frac{\sigma_z^2\,{d_u}^\zeta\left(B^2+\sigma_w^2\right)}{\lambda P}\right)$ to increase. Consequently, the tightest upper bound scaling in (39) is attained by making the approximation error, the measurement noise error and the communication noise error scale (as a function of n) at the same rate. That is, assuming that the system parameters $(C_o, B, \sigma_w^2, \sigma_z^2, d_u, \zeta, P)$ do not depend on n,

$$L^{-2\alpha} \asymp \frac{L}{n} \asymp \frac{L}{\lambda n}, \tag{41}$$

implying that L must be chosen, independently of λ, as

$$L \asymp n^{1/(2\alpha+1)}, \tag{42}$$

which in turn requires that $\lambda = \mathcal{O}(1)$, resulting in the following expression for optimal distortion scaling

$$D^* \asymp L^{-2\alpha} \asymp n^{-2\alpha/(2\alpha+1)} \tag{43}$$

that has the same scaling behavior as that of D^*_{cen} (cf. (20)). Moreover, the total network power consumption associated with achieving this optimal distortion scaling is given by (see (40))

$$P_{\text{tot}} \asymp L \asymp n^{1/(2\alpha+1)}. \tag{44}$$

Combining (42), (43) and (44), we can also compactly characterize the relationship between optimal distortion scaling and the associated power and latency requirements in terms of the following expression

$$D^* \sim P_{\text{tot}}{}^{-2\alpha} \sim L^{-2\alpha}. \tag{45}$$

Note that this expression does not mean that a WSN with fixed number of sensor nodes using more power and/or latency can provide better accuracy. Rather, power, distortion and latency are functions of the number of nodes and the above relation indicates how the three performance metrics behave with respect to each other as the density of nodes increases.

Remark 7: Equation (44) shows that the total network power requirement of our proposed scheme for optimal distortion scaling $(D^* \asymp n^{-2\alpha/(2\alpha+1)})$ is given by $P_{\text{tot}} \asymp n^{1/(2\alpha+1)}$. A natural question is: *How good is this scheme in terms of power scaling?* While a comparison with all conceivable schemes does not seem possible, in order to give an idea of the performance of our proposed scheme we compare it to a setup where all the nodes in the network noiselessly communicate their measurements to a designated cluster of $1 \leq \tilde{n} \leq n$ nodes. Each node in the cluster computes the required L projections of the measurement data for each snapshot and then all the $\tilde{n}$ nodes coherently transmit these (identical) projections to the FC over the MAC; in this case, the $\tilde{n} \times 1$ MAC is effectively transformed into a point-to-point AWGN channel with an $\tilde{n}$-fold power-pooling (beamforming) gain. One extreme, $\tilde{n} = 1$, corresponds to a single clusterhead (no beamforming gain), whereas the other extreme, $\tilde{n} = n$, corresponds to maximum beamforming gain. Note that in our proposed scheme, nodes transmit coherently (and hence benefit from power-pooling) but there is no data exchange between them. An exact comparison of our scheme with the above setup involving in-network data exchange is beyond the scope of this paper since quantifying the cost of required in-network communication is challenging and requires making additional assumptions. Thus, we ignore the cost of in-network communication and provide a comparison just based on the cost of communicating the projections to the FC – though, in general, we expect the in-network cost to increase with the size $\tilde{n}$ of the cluster. Under this assumption, the analysis in Appendix I shows that our scheme requires less communication power compared to the $\tilde{n} = 1$ case, whereas it requires more power compared to the $\tilde{n} = n$ case. In particular, the power scaling achieved by our proposed scheme (for optimal distortion scaling) is identical to that in the case when there are $\tilde{n} = \frac{n}{L} \asymp n^{2\alpha/(2\alpha+1)}$ nodes in the designated cluster to coherently communicate the required $L \asymp n^{1/(2\alpha+1)}$ projection coefficients to the FC. Note that since $n^{2\alpha/(2\alpha+1)} \nearrow n$ for highly compressible signals $(\alpha \gg 1)$, the performance of our proposed estimation scheme in this case approaches that of the $\tilde{n} = n$ extreme, without incurring any overhead of in-network communication.

2) Power-Distortion-Latency Scaling Laws for Consistent Estimation: Preceding analysis shows that in order to achieve the optimal centralized distortion scaling $n^{-2\alpha/(2\alpha+1)}$, the network must expend power P_{tot} and incur latency L that scale (with n) at a sublinear rate of $n^{1/(2\alpha+1)}$. This may pose a bottleneck in deploying dense WSNs for certain types of applications that might require extended battery life or faster temporal sampling of the physical phenomenon. Cursory analysis of (39) and (40), however, shows that it is possible to lower these power and latency requirements at the expense of *sub-optimal* distortion scaling, and for the remainder of this subsection, we shall be analyzing these power-distortion-latency scaling regimes.

Notice that under the assumption of system parameters $(C_o, B, \sigma_w^2, \sigma_z^2, d_u, \zeta, P)$ not varying with n, L and λ are the only two quantities that bear upon the required network power and achievable distortion of the estimation scheme (see (39), (40)). Therefore, we begin by treating L (effective number of projections per snapshot) as an independent variable and model its scaling behavior as $L \asymp n^\beta$ for $\beta \in (0, 1)$, while we model the scaling behavior of λ as $\lambda \asymp n^{-\delta}$ for $\delta \in [0, \infty)$ (recall, $0 < \lambda \leq 1$).[7] Note that $\beta = \beta^* \triangleq 1/(2\alpha+1)$ has already been solved previously (resulting in $\delta = \delta^* \triangleq 0$) and corresponds to the optimal distortion scaling of (43).

Bias-Limited Regime. Recall that $L \asymp n^{1/(2\alpha+1)}$ is the critical scaling of the number of projections at which point the distortion component due to the approximation error scales at the same rate as the distortion component due to the measurement noise (cf. (41), (42)). If, however, we let L $(\asymp n^\beta)$ scale at a rate such that $\beta < \beta^*$, then the first term in the upper bound in (39) that is due to the approximation error (bias term) starts to dominate the second term that is due to the measurement noise and, ignoring constants, the resulting distortion at the FC scales as

$$n^{-1+\beta+\delta} \preceq D \preceq n^{-2\alpha\beta} + n^{-1+\beta+\delta}, \tag{46}$$

and the corresponding choice of *optimal* δ is given by

$$\delta = 1 - (2\alpha+1)\beta, \tag{47}$$

where optimal here refers to the fact that (i) $\delta < 1-(2\alpha+1)\beta$ is wasteful of power since distortion component due to the approximation error (first term in the upper bound in (46)) in that case decays slower than the distortion component to the communication noise (second term in the upper bound in (46)); and (ii) $\delta > 1-(2\alpha+1)\beta$ is wasteful of projections (i.e., latency) since distortion component due to the approximation error in that case decays faster than the distortion component

[7] There is nothing particular about choosing L as the independent variable except that it makes the analysis slightly easier. Nevertheless, we might as well start off by treating λ as the independent variable and reach the same conclusions.

due to the communication noise. With this balancing of L and λ, distortion goes to zero at the rate

$$D \asymp n^{-2\alpha\beta}, \tag{48}$$

as long as the chosen $\beta \in \big(0, 1/(2\alpha+1)\big)$, and the corresponding total network power consumption is given by (cf. (40))

$$P_{\text{tot}} \asymp n^{(2\alpha+2)\beta-1}, \tag{49}$$

resulting in the following expression for power-distortion-latency scaling relationship in the bias-limited regime

$$D \sim P_{\text{tot}}^{\frac{-2\alpha\beta}{(2\alpha+2)\beta-1}} \sim L^{-2\alpha}. \tag{50}$$

Variance-Limited Regime. On the other hand, if we let L scale at a rate such that $\beta > \beta^*$, then the second term in the upper bound in (39) that is due to the measurement noise (variance term) starts to dominate the bias term and the resulting distortion at the FC scales as

$$D \asymp n^{-1+\beta} + n^{-1+\beta+\delta}, \tag{51}$$

and the corresponding choice of optimal δ is given by $\delta = 0$ $(= \delta^*)$. This implies that as long as the chosen $\beta \in \big(1/(2\alpha+1), 1\big)$, distortion in the variance-limited regime goes to zero at the rate

$$D \asymp n^{-1+\beta}, \tag{52}$$

and the corresponding total network power consumption is given by

$$P_{\text{tot}} \asymp n^{\beta}, \tag{53}$$

resulting in the following expression for power-distortion-latency scaling relationship in the variance-limited regime

$$D \sim P_{\text{tot}}^{-\frac{1}{\beta}+1} \sim L^{-\frac{1}{\beta}+1}. \tag{54}$$

Notice that as $\beta \to \beta^*$, both (50) and (54) collapse to the power-distortion-latency scaling relationship of (45), indicating that the optimal distortion scaling D^* corresponds to the transition point between the bias-limited and variance-limited regimes. Thus, (50) and (54) completely characterize the power-distortion-latency scaling relationship of the proposed distributed estimation scheme for a compressible signal in the known subspace case. This scaling relationship is also illustrated in Fig. 3, where the scaling exponents of P_{tot} and D are plotted against $\beta \in (0, 1)$ (the chosen scaling exponent of L) for different values of α.

Observation 3: Analysis of (50), (54) and Fig. 3 shows that (i) any distortion scaling that is achievable in the variance-limited regime is also achievable in the bias-limited regime; and (ii) scaling of P_{tot} in the variance-limited regime is uniformly worse than in the bias-limited regime. This implies that any WSN observing an α-compressible signal in the known subspace case should be operated only either in the bias-limited regime or at the optimal distortion scaling point, i.e., $\beta \in (0, 1/(2\alpha+1)]$. Thus, given α and a target distortion scaling of $D \asymp n^{-\gamma}$, $0 < \gamma \leq 2\alpha/(2\alpha+1)$, the number of projections computed by the WSN per snapshot needs to be scaled as $L \asymp n^{\beta}$, where $\beta = \gamma/2\alpha$ (cf. (50)), and

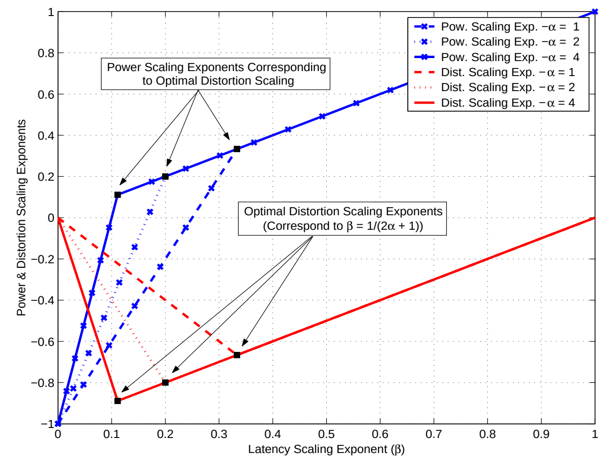


Fig. 3. Power-distortion-latency scaling relationship of compressible signals in the known subspace case. The scaling exponents of P_{tot} and D are plotted against $\beta \in (0, 1)$ for different values of α. The filled black square on each curve corresponds to the operating point for optimal distortion scaling ($\beta = \beta^*$), with bias-limited and variance-limited regimes corresponding to the curve on its left and right side, respectively.

the corresponding total network power consumption would be given by (49).[8]

Observation 4: Another implication of the analysis carried out in this section is that the more compressible a signal is in a particular basis (i.e., the higher the value of α), the easier it is to estimate that signal in the bias-limited regime/at the optimal distortion scaling point (easier in terms of an improved power-distortion-latency relationship).[9]

Observation 5: One of the most significant implication of the preceding analysis is that, while operating in the bias-limited regime, if β is chosen to be such that $\beta < 1/(2\alpha+2)$ then the scaling exponent of P_{tot} would be negative (cf. (49), Fig. 3). This is remarkable since it shows that, in principle, consistent signal estimation is possible ($D \searrow 0$ as $n \to \infty$) even if the total network power consumption P_{tot} goes to zero!

3) Power-Density Trade-off: Viewed in a different way, Observation 5 also reveals a remarkable power-density trade-off inherent in our approach: *increasing the sensor density, while keeping the latency requirements the same, reduces the total network power consumption required to achieve a target distortion level*. This essentially follows from the fact that the power-distortion scaling law in the bias-limited regime (including the optimal distortion scaling point) follows a conservation relation given by (cf. (48), (49))

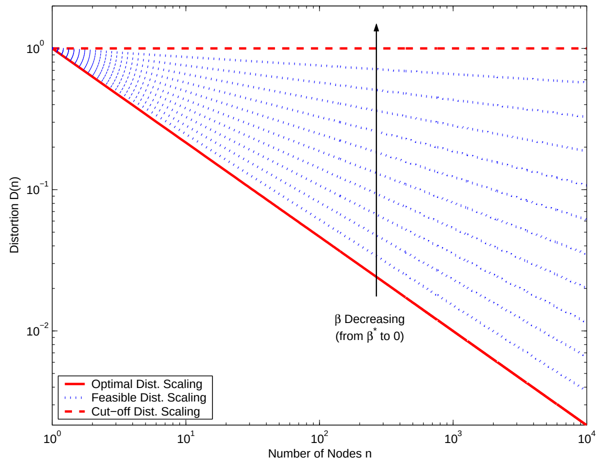
$$P_{\text{tot}} D \asymp n^{2\beta-1}. \tag{55}$$

Specifically, let $\beta_2 < \beta_1$ denote two latency scalings in the bias-limited regime and let $n_2 > n_1$ denote the corresponding number of nodes needed to achieve a target distortion level $D(n_1, \beta_1) = D(n_2, \beta_2) = D_o$. Then, we have from (48) that

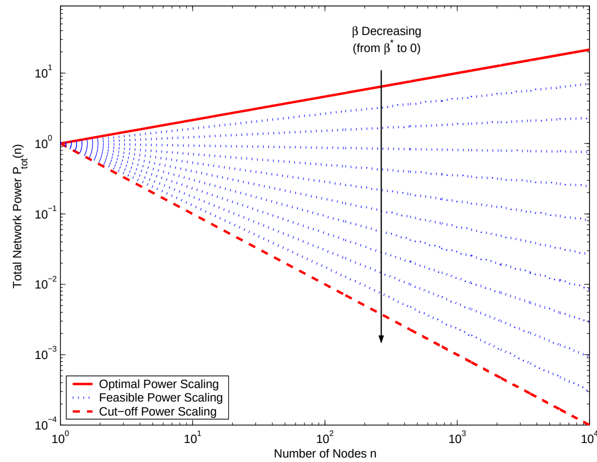
$$D(n_1, \beta_1) = D(n_2, \beta_2) \quad \Rightarrow \quad n_1^{-2\alpha\beta_1} = n_2^{-2\alpha\beta_2}, \tag{56}$$

[8] The designer of a WSN could also reverse the roles of D and L by specifying a target latency scaling and obtaining the corresponding distortion (and power) scaling expression.

[9] Note that the power-distortion-latency scaling in the variance-limited regime is independent of α (cf. (54)).



(a) Different scaling curves for D



(b) Different scaling curves for P_{tot}

Fig. 4. Power-Density trade-off for compressible signals in the known subspace case. Various power and distortion scaling curves, each one corresponding to a different value of β, are plotted on a log-log scale against the number of nodes for $\alpha = 1$. The dashed curves are the cut-off scalings for consistent signal estimation (corresponding to $\beta \searrow 0$).

and therefore, the corresponding latency requirements are rather trivially related to each other as

$$\frac{L_2(n_2, \beta_2)}{L_1(n_1, \beta_1)} = \frac{n_2^{\beta_2}}{n_1^{\beta_1}} = 1. \tag{57}$$

Moreover, it follows from (55) that the total network power consumptions in the two cases are related by

$$\frac{P_{\text{tot}}(n_2, \beta_2)}{P_{\text{tot}}(n_1, \beta_1)} = \frac{n_2^{2\beta_2 - 1}}{n_1^{2\beta_1 - 1}} = \frac{n_1}{n_2}, \tag{58}$$

where we have used the fact that (56) implies that $n_1^{2\beta_1} = n_2^{2\beta_2}$. Relations (57) and (58) show that increasing the sensor density by a factor of N, while keeping the number of projections (per snapshot) communicated by the network to the FC the same, reduces the total network power required to attain a given target distortion by a factor of N.

This power-density trade-off is also illustrated in Fig. 4, where various power and distortion scaling curves (corresponding to different values of β) are plotted on a log-log scale against the number of nodes for $\alpha = 1$. For the sake of illustration, these plots assume that the constants of proportionality in the scaling relations are unity. In order to illustrate the power-density trade-off, suppose that we want to attain a target distortion of $D_o = 0.02$. With optimal distortion scaling (solid curve in Fig. 4(a) corresponding to $\beta_1 = \beta^* = 1/3$), the desired distortion can be attained with $n_1 \approx 353$ nodes, consuming a total network power $P_{\text{tot}}(n_1, \beta_1) \approx 7.07$, as calculated from the solid curve in Fig. 4(b). On the other hand, however, if we operate on a sub-optimal distortion scaling curve (say, e.g., the third dotted feasible curve from the bottom in Fig. 4(a) corresponding to $\beta_2 = 8/33$), we would attain the desired distortion of $D_o = 0.02$ with $n_2 \approx 3192$ nodes – roughly a factor of 9 increase in sensor density – but would only consume a total network power of $P_{\text{tot}}(n_2, , \beta_2) \approx 0.78$, as calculated from the third dotted feasible curve from the top in Fig. 4(b). Thus, as predicted, increasing the sensor density roughly by a factor of 9 reduces the total network power consumption by a factor of 9, while the latency requirements stay the same ($n_1^{\beta_1} = n_2^{\beta_2}$).

B. Estimation of Sparse Signals

The analysis for the estimation of an M-sparse signal in the known subspace case using the joint source-channel communication architecture of Section IV can be carried out along almost the same lines as for compressible signals in Section V-A, with the only obvious difference being that $\boldsymbol{S}$ now lies exactly in an M-dimensional subspace of $\mathbb{R}^n$ and therefore, L has to be taken exactly equal to M, i.e., is no longer a variable parameter in the hands of the designer of a WSN. This results in the following expressions for the end-to-end distortion at the FC and the corresponding total network power consumption and system latency

$$D = \left(\frac{M}{n}\right)\sigma_w^2 + \left(\frac{M}{n}\right)\left(\frac{\sigma_z^2\, {d_u}^{\zeta}\left(B^2 + \sigma_w^2\right)}{\lambda P}\right), \tag{59}$$

$$\lambda M P\left(\frac{\sigma_w^2}{{d_u}^{\zeta}\left(B^2 + \sigma_w^2\right)}\right) \leq P_{\text{tot}} \leq \lambda M P, \tag{60}$$

$$L = M. \tag{61}$$

Unlike the compressible signal case, however, the only controllable parameter in this case is the power scaling factor λ, modeled as $\lambda \asymp n^{-\delta}$ for $\delta \in [0, \infty)$,[10] and in the sequel, we analyze the effect of various scaling behaviors of λ on the power-distortion-latency scaling relationship of the proposed estimation scheme.

1) Power-Distortion-Latency Scaling Laws for Consistent Estimation: We start out by first analyzing the optimal distortion scaling that is achievable for an M-sparse signal. Notice

[10] Recall that while we do model the scaling behavior of M as $M \asymp n^{\mu}$, $0 \leq \mu < 1$, the choice of μ is not in our hands in this case and depends upon the underlying physical phenomenon. Essentially, μ here plays the role analogous to that of α in the compressible case.

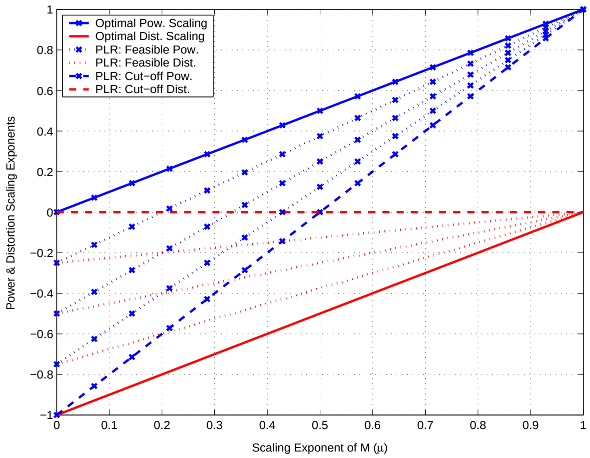


Fig. 5. Power-distortion-latency scaling relationship of sparse signals in the known subspace case. The scaling exponents of P_{tot} and D are plotted against the scaling exponent of M ($\asymp n^{\mu}$) for $0 \leq \mu < 1$, while the scaling exponent of L is the same as that of M. The solid curves correspond to the optimal distortion scaling exponents and the corresponding total network power scaling exponents, while the dotted curves correspond to various power-limited regime (PLR) scalings that result in consistent signal estimation. The dashed curves are the cut-off scalings for consistent signal estimation.

that for fastest distortion reduction, the first term (due to the measurement noise) in (59) should scale at the same rate as the second term (due to the communication noise). This in turn requires that $\lambda = \mathcal{O}(1)$ (or $\delta = \delta^* \triangleq 0$), resulting in the following expression for optimal distortion scaling

$$D^* \asymp \frac{M}{n} \asymp n^{-1+\mu}, \tag{62}$$

which has the same scaling behavior as that of D^*_{cen} (cf. (23)). Moreover, the total network power consumption associated with achieving this optimal distortion scaling is given by

$$P_{\text{tot}} \asymp L = M \asymp n^{\mu}. \tag{63}$$

Equations (61), (62) and (63) can also be combined together for $\mu \in (0,1)$ to express the relationship between optimal distortion scaling and the corresponding power and latency requirements in terms of the following expression

$$D^* \sim P_{\text{tot}}^{-\frac{1}{\mu}+1} \sim L^{-\frac{1}{\mu}+1}. \tag{64}$$

Notice that the above relationship has the same form as that of (54) (the power-distortion-latency relationship of a compressible signal in the variance-limited regime) which is precisely what one would expect since there is no bias related distortion component for a sparse signal (cf. (59)).

Remark 8: Equations (61), (62), and (63) show that for the case of $\mu = 0$, i.e., $M = \mathcal{O}(1)$, optimal distortion scaling $D^* \asymp n^{-1}$ can be obtained by consuming only a fixed amount of total network power and incurring a fixed latency, i.e., $P_{\text{tot}} = L = \mathcal{O}(1)$. This result is similar in spirit to the one obtained in [3] that primarily studies the case analogous to that of a sparse signal with non-scaling DoF ($M = \mathcal{O}(1)$), albeit assuming Gaussian sources and multiple FCs (see Theorems 1 and 3 therein).

Power-limited Regime. On the other hand, if we take $\delta > 0$ then the distortion component due to the communication noise (second term in (59)) starts to dominate the distortion component due to the measurement noise (first term in (59)) and, ignoring the constant parameters, the resulting distortion at the FC scales as

$$D \asymp n^{-1+\mu+\delta} \tag{65}$$

in the power-limited regime. This implies that as long as $\delta \in (0, 1-\mu)$, distortion can still be driven to zero, albeit at a slower, *sub-optimal* rate of $n^{-1+\mu+\delta}$ ($\succ D^*$). In particular, this means that D can be asymptotically driven to zero even if the total network power P_{tot} ($\asymp n^{\mu-\delta}$) scales just a little faster than $n^{2\mu-1}$ (cf. (60)). This observation is similar in spirit to the one made for compressible signals since it shows that, in principle, consistent signal estimation is possible in the limit of a large number of nodes for $\mu \in [0, 1/2]$ (i.e., the number of DoF M scaling at most as fast as $\sqrt{n}$) even if the total network power P_{tot} goes to zero. Finally, this power-distortion-latency scaling relationship in the power-limited regime can be expressed as

$$D \sim P_{\text{tot}}^{\frac{-1+\mu+\delta}{\mu-\delta}} \sim L^{\frac{-1+\delta}{\mu}+1}. \tag{66}$$

2) Discussion: While quite similar in spirit, there are still some key differences between the power-distortion-latency scaling laws of the proposed estimation scheme for compressible and sparse signals. To begin with, unlike for compressible signals, the latency scaling requirements for sparse signals are dictated by the underlying physical phenomenon ($L = M \asymp n^{\mu}$) and cannot be traded-off for power and/or distortion without making further assumptions on the decay characteristics of the M non-zero coefficients of $\boldsymbol{S}$. Secondly, the scenario of consistent signal estimation of sparse signals with decaying total network power consumption exists if and only if the number of DoF M scales at a rate less than or equal to $\sqrt{n}$, i.e., $\mu \leq 1/2$ (see Fig. 5).[11] And finally, as a flip side to this observation, the power-density trade-off for sparse signals exists only when $0 \leq \mu < 1/2$, happens to be a function of μ and is not as pronounced for $0 < \mu < 1/2$. Specifically, for an increase in the sensor density by a factor of N, the total network power consumption requirements can only be reduced by a factor of $N^{1-2\mu}$, $0 \leq \mu < 1/2$, in order to attain the same target distortion for a sparse signal (cf. (65)).

C. Communicating the Projection Vectors to the Network

Recall that an implicit requirement for employing the proposed distributed estimation scheme in the known subspace case is that the sensor encoders have access to the respective projection vectors' elements at each time instant k (cf. (36)). In this subsection, we address the issue of how one might communicate this information to the sensor nodes. One viable option in this regard could be the pre-storage of relevant information in each sensor node. However, pre-storage of the entire compressing (sparse) basis $\boldsymbol{\Psi}$ or a subset of it, $\{\psi_i\}_{i=1}^{L}$, where $1 \leq L \leq n$, in each sensor node is not feasible in large-scale WSNs since this would require at least n storage elements per sensor node, and a better alternative is to store

[11] Recall that for compressible signals, this observation holds true for all α.

only the corresponding *non-zero* elements of the L projection vectors, $\{\psi_{ij} : \psi_{ij} \neq 0\}_{i=1}^{L}$, in the j-th sensor node. In the context of [2], for example, this would mean having only $\mathcal{O}(1)$ storage elements per sensor node, since the structure of the proposed projection vectors in [2] is such that the cardinality of the set $\{\psi_{ij} : \psi_{ij} \neq 0\}_{i=1}^{L}$ is identically equal to one $\forall\, j = 1, \dots, n$. Other instances when pre-storage might be a feasible option could be, for example, when the projection vectors' elements come from an analytical expression. Pre-storage, however, suffers from the drawback that sensor nodes pre-stored with one compressing (sparse) basis vectors might not be readily deployable in signal fields compressible (sparse) in some other basis.

Another more feasible, but not always practical, approach to the communication of projection vectors to the network could be that the FC transmits this information over the FC-to-network broadcast channel at either the start of the estimation process or at the start of each network-to-FC channel use. For the case of basis $\boldsymbol{\Psi}$ whose vectors have some sort of spatial regularity in their structure such that they do not require addressing each sensor node individually (e.g., vectors describable by a few parameters such as in [2]), this could be readily accomplished by broadcasting a few command signals from the FC to the network. One could also increase the addressing resolution of the FC by equipping it with multiple transmit antennas and using some of the techniques described in [45]. However, depending upon the structure of the compressing (sparse) basis, this approach may require the FC to be able to address each sensor node individually which may or may not be practical in large-scale dense WSNs. We will show in Section VI, however, that one benefit of *compressive wireless sensing* is a straightforward treatment of this issue.

VI. Distributed Estimation from Noisy Projections: Unknown Subspace

In Section V, we proposed an efficient distributed estimation scheme that achieves the optimal centralized distortion scaling D^*_{cen} for both compressible and sparse signals under the assumption that the WSN has complete knowledge of the basis in which $\boldsymbol{S}$ is compressible or sparse. Generally speaking, however, even if the basis in which $\boldsymbol{S}$ is compressible (sparse) is known, it is quite likely that the precise ordering of its coefficients (indices of its non-zero coefficients) in that basis at each time instant k might not be known ahead of time – a scenario that we refer to as the '*unknown subspace*' or '*adaptive subspace*' case. As an example, consider the following simple case. Suppose $\boldsymbol{S}$ is very sparse in some basis $\boldsymbol{\Psi} = \{\boldsymbol{\psi}_i\}_{i=1}^{n}$ such that each temporal sample $\boldsymbol{s}^k$ has only one non-zero coefficient of amplitude $\sqrt{n}\,B$ corresponding to *some* element $\boldsymbol{\psi}_i$ of $\boldsymbol{\Psi}$ and i is drawn at random from the set $\{1, 2, \dots, n\}$. This is an example of the case where we know the basis in which $\boldsymbol{S}$ is sparse but do not know the indices of its non-zero coefficients in that basis.

One naïve approach to this problem would be to use the distributed estimation scheme described in Section V. However, since the network does not have a precise knowledge of the index of the true basis vector, it would need to be determined by trial and error (e.g., deterministically or randomly selecting basis vectors in some fashion). As an illustration, consider a randomized selection process: the network computes the projection of the sensor data onto $\boldsymbol{\psi}_i$ and i is selected L times uniformly at random (without replacement) from the set $\{1, 2, \dots, n\}$. Ignoring the distortion due to the measurement and communication noise, the squared distortion error would be 0 at the FC if the spike in the $\boldsymbol{\Psi}$ domain corresponds to one of the uniformly picked $\boldsymbol{\psi}_i$'s and B^2 otherwise, and the probability of not finding the spike in L trials is $\prod_{i=0}^{L-1}\left(1 - \frac{1}{n-i}\right)$. If n is large enough and $L \ll n$, we can approximate the resulting distortion by $D \approx \left(1 - \frac{1}{n}\right)^L B^2 \approx e^{-L/n} B^2 \to B^2$ as $n \to \infty$, i.e., equivalent to the MSE that is achievable even without any information.

Another more general, and perhaps relevant, example is a situation in which the signal field is spatially piecewise smooth. 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 change points in the signal, which of course are unlikely to be known a priori. Broadly speaking, any signal that is generally spatially 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 a function of the time index k and thus, will preclude the use of methods like the one in Section V that require prior specification of the basis vectors to be used in the projection process. This is where the *universality* of compressive wireless sensing (CWS) scheme, presented in this section, comes into play. As we shall see, CWS 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} and L grow at most sub-linearly with the number of nodes in the network.

A. Compressive Wireless Sensing

Recall that if $\upsilon^k = \boldsymbol{\varphi}^T \boldsymbol{s}^k = \sum_{j=1}^{n} \varphi_j s_j^k$ is the projection of $\boldsymbol{s}^k$ onto a vector $\boldsymbol{\varphi} \in \mathbb{R}^n$ then, using the communication architecture described in Section IV and consuming only $\mathcal{O}(1)$ amount of power, the FC can obtain an estimate of υ^k in one channel use that is given by

$$\widehat{\upsilon}^k \;=\; \upsilon^k + \boldsymbol{\varphi}^T \boldsymbol{w}^k + \widetilde{z}^k, \tag{67}$$

where $\widetilde{z}^k \sim \mathcal{N}(0, \sigma_z^2/\rho)$ is the scaled MAC AWGN (cf. (27)). The basic idea behind CWS is that instead of projecting the sensor network data onto a subset of a deterministic basis of $\mathbb{R}^n$, the FC tries to reconstruct $\boldsymbol{s}^k$ from *random projections* of the sensor network data. Specifically, let $\{\boldsymbol{\phi}_i \in \mathbb{R}^n\}_{i=1}^{n}$ be an i.i.d. sequence of (normalized) Rademacher random vectors, i.e., $\{\phi_{ij}\}_{j=1}^{n} = \pm 1/\sqrt{n}$, each with probability $1/2$, and the FC tries to reconstruct $\boldsymbol{s}^k$ by projecting $\boldsymbol{x}^k$ onto L of these random vectors.[12] Because the entries of each projection vector $\boldsymbol{\phi}_i$ are generated at random, observations of this form are called random projections of the signal.

[12]The L Rademacher vectors are to be generated independently at each time instant k. However, we omit the superscript corresponding to the time index to simplify the notation.

Remark 9: An important consequence of using Rademacher random vectors for projection purposes is that each sensor can locally draw the elements of the projection vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$ in an efficient manner by simply using its network address as the seed of a pseudo-random number generator (see, e.g., [46]). Moreover, given these seed values and the number of nodes in the network, the FC can also easily reconstruct the vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$. Therefore, in addition to being a universal estimation scheme, CWS has an added advantage that no extended information concerning the projection vectors needs to be communicated to (or stored inside) the sensor nodes (cf. Section V-C).

After employing L Rademacher projections, the corresponding projection estimates at the FC are given by

$$\widehat{v}_i^k \;=\; \boldsymbol{\phi}_i^T \boldsymbol{s}^k + \boldsymbol{\phi}_i^T \boldsymbol{w}^k + \widetilde{z}_i^k, \quad i = 1, \ldots, L, \tag{68}$$

where $\boldsymbol{w}^k = (w_1^k, \ldots, w_n^k)^T$, and $\{w_j^k\}_{j=1}^n$ and $\{\widetilde{z}_i^k\}_{i=1}^L$ are i.i.d. zero-mean Gaussian random variables, independent of each other and $\{\phi_{ij}\}$, with variances σ_w^2 and σ_z^2/ρ, respectively. The reconstruction process can be described as follows – let $\mathcal{S}_q$ denote a countable collection of candidate reconstruction vectors such that

$$\mathcal{S}_q \;\subset\; \{\boldsymbol{y} \in \mathbb{R}^n : |y_j| \leq B,\ j = 1, \ldots, n\}, \tag{69}$$

and define a CWS estimate $\widehat{\boldsymbol{s}}^k$ as

$$\widehat{\boldsymbol{s}}^k \;=\; \arg\min_{\boldsymbol{s} \in \mathcal{S}_q} \left\{ \widehat{R}(\boldsymbol{s}) + \frac{c(\boldsymbol{s}) \log(2)}{L\epsilon} \right\}. \tag{70}$$

The first term in the objective function is the empirical risk, defined as

$$\widehat{R}(\boldsymbol{s}) \;=\; \frac{1}{L} \sum_{i=1}^{L} \left(\widehat{v}_i^k - \boldsymbol{\phi}_i^T \boldsymbol{s} \right)^2, \tag{71}$$

which measures the average (Euclidean) distance between the observations $\{\widehat{v}_i^k\}_{i=1}^L$ and the projections of a given candidate vector $\boldsymbol{s}$ onto the corresponding Rademacher vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$. The quantity $c(\boldsymbol{s})$ in the second term is a non-negative number assigned to each candidate vector in $\mathcal{S}_q$ such that $\sum_{\boldsymbol{s} \in \mathcal{S}_q} 2^{-c(\boldsymbol{s})} \leq 1$, and is designed to penalize candidate vectors proportional to their complexity (see (76)). Finally, $\epsilon > 0$ is a constant (independent of L and n) that controls the relative contribution of the complexity term to the objective function. In the context of Theorem 1 in [7], $\sigma_w^2 = 0$ and so ϵ depends only the sample bound B and the noise variance σ_z^2/ρ.

In order to apply the results of [7] to the observation model (68), the effect of the *projected noise* terms $\{\boldsymbol{\phi}_i^T \boldsymbol{w}^k\}_{i=1}^L$ needs to be determined. First, suppose that the projection vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$ were mutually orthogonal. In that case, it is easy to see that the projected noises are equivalent (in distribution) to i.i.d. zero-mean Gaussian noises with variance σ_w^2. In addition, note that $\{\boldsymbol{\phi}_i^T \boldsymbol{w}^k\}$ and $\{\phi_{ij}\}$ are independent. To see this, notice that for any fixed vector $\boldsymbol{g} \in \mathbb{R}^n$, the joint characteristic function of $\boldsymbol{\phi}_i^T \boldsymbol{w}^k$ and $\boldsymbol{\phi}_i^T \boldsymbol{g}$ can be factored into the product of the individual characteristic functions, i.e.,

$$\mathbb{E}\left[e^{j\nu_1 \boldsymbol{\phi}_i^T \boldsymbol{w}^k + j\nu_2 \boldsymbol{\phi}_i^T \boldsymbol{g}} \right] \;=\; \mathbb{E}\left[e^{j\nu_1 \boldsymbol{\phi}_i^T \boldsymbol{w}^k} \right] \cdot \mathbb{E}\left[e^{j\nu_2 \boldsymbol{\phi}_i^T \boldsymbol{g}} \right], \tag{72}$$

and taking $\boldsymbol{g}$ to be a vector that has one at location 'j' and zero at all other locations establishes the independence of $\{\boldsymbol{\phi}_i^T \boldsymbol{w}^k\}$ and $\{\phi_{ij}\}$. In this case, if we pick

$$\rho \;=\; \frac{P}{{d_u}^{\zeta} \left(B^2 + \sigma_w^2\right)} \tag{73}$$

then the observations in (68) would be equivalent (in distribution) to observations of the form

$$\widehat{v}_i^k \;=\; \boldsymbol{\phi}_i^T \boldsymbol{s}^k + \eta_i^k, \quad i = 1, \ldots, L, \tag{74}$$

where $\{\eta_i^k\}_{i=1}^L$ are i.i.d. zero-mean Gaussian random variables – *independent* of $\{\phi_{ij}\}$ – with variance $\sigma^2 = \sigma_w^2 + \sigma_z^2\, {d_u}^{\zeta} \left(B^2 + \sigma_w^2\right)/P$, and the results of Theorem 1 in [7] can be applied directly.

On the other hand, our model only assumes that the vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$ are mutually orthogonal in expectation; hence, the projected noise is colored – if $\boldsymbol{\Phi}$ is the $L \times n$ matrix whose rows are $\{\boldsymbol{\phi}_i^T\}_{i=1}^L$ then, given $\boldsymbol{\Phi}$, the projected noise vector $\boldsymbol{\Phi}\boldsymbol{w}^k$ is a zero-mean Gaussian vector with covariance matrix $\boldsymbol{\Phi}\boldsymbol{\Phi}^T \sigma_w^2$. Without loss of generality, however, we can assume that the projected noise $\{\boldsymbol{\phi}_i^T \boldsymbol{w}^k\}_{i=1}^L$ behaves approximately like white Gaussian noise and consequently, use the observation model of (74) for further analysis. This approximation is motivated by the asymptotic results presented in Section IV-B of [29] which show that the extreme eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ are almost surely (a.s.) contained in the interval $\left[(1-\sqrt{c})^2, (1+\sqrt{c})^2\right]$ in the limit as $L, n \to \infty$ with $L/n \to c$. Since we assume L grows sublinearly with n so $L/n \to 0$ in our case and consequently, all the eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ tend to 1 a.s. In other words, $\{\boldsymbol{\phi}_i\}_{i=1}^L$ become mutually orthogonal asymptotically, and the degree of coloring becomes negligible for large values of n (this approximation is also shown to work well in practice – see Section VIII).

Explicit bounds on the reconstruction error using the CWS estimate of (70) can then be obtained by first assuming that we can find a basis $\boldsymbol{\Psi}$ at the FC in which the signal $\boldsymbol{S}$ is α-compressible and then, using this compressing basis in the reconstruction process by defining $\mathcal{S}_q$ and $c(\boldsymbol{s})$ in terms of $\boldsymbol{\Psi}$. Specifically, let

$$\begin{aligned} \boldsymbol{\Theta}_q \;\triangleq\; \Big\{ &\boldsymbol{\theta} \in \mathbb{R}^n : |(\boldsymbol{\Psi}\boldsymbol{\theta})_j| \leq B,\ \theta_j \text{ uniformly} \\ &\text{quantized to } n^q \text{ levels},\ j = 1, \ldots, n \Big\} \end{aligned} \tag{75}$$

be a set of quantized candidate solutions in the transform domain $\boldsymbol{\Psi}$, so that $\mathcal{S}_q = \{\boldsymbol{s} \in \mathbb{R}^n : \boldsymbol{s} = \boldsymbol{\Psi}\boldsymbol{\theta},\ \boldsymbol{\theta} \in \boldsymbol{\Theta}_q\}$. Furthermore, let the penalty term $c(\boldsymbol{s}) = c(\boldsymbol{\theta})$ be

$$c(\boldsymbol{\theta}) \;\triangleq\; (1+q)\log(n)\|\boldsymbol{\theta}\|_0, \tag{76}$$

where $\|\cdot\|_0$ counts the number of non-zero elements in a vector. Then, the optimization problem becomes

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^k \;=\; \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_q} \Big\{ &\left\| \widehat{\boldsymbol{v}}_L^k - \boldsymbol{\Phi}_L^T \boldsymbol{\Psi}\boldsymbol{\theta} \right\|_2^2 + \\ &\frac{(1+q)\log(2)\log(n)}{\epsilon} \|\boldsymbol{\theta}\|_0 \Big\}, \end{aligned} \tag{77}$$

where $\widehat{\boldsymbol{v}}_L^k \triangleq (\widehat{v}_1^k, \ldots, \widehat{v}_L^k)^T$, $\boldsymbol{\Phi}_L$ is the $n \times L$ matrix of Rademacher projection vectors $\{\boldsymbol{\phi}_i\}_{i=1}^L$, and $\widehat{\boldsymbol{\theta}}^k$ is the estimate of the representation of $\boldsymbol{s}^k$ in the compressing basis $\boldsymbol{\Psi}$, i.e., $\widehat{\boldsymbol{s}}^k \triangleq \boldsymbol{\Psi}\widehat{\boldsymbol{\theta}}^k$. As shown in [7], for an α-compressible $\boldsymbol{S}$, such an estimate would satisfy[13]

$$D \preceq \left(\frac{L}{\log(n)}\right)^{-2\alpha/(2\alpha+1)}, \tag{78}$$

while, for an M-sparse signal, this would result in

$$D \preceq \left(\frac{L}{M\log(n)}\right)^{-1}. \tag{79}$$

B. Power-Distortion-Latency Scaling Laws

Recall that in order to achieve the distortion scaling of (78) and (79), the network had to employ L network-to-FC MAC uses per source observation, each one corresponding to a projection of $\boldsymbol{s}^k$ onto a random Rademacher vector. And while the projection vectors in this case happened to be random as opposed to the analysis carried out in Section IV, it is a simple exercise to show that with the scaling factor ρ as given in (73), each projection of the sensor network data onto a Rademacher vector still consumes only $\mathcal{O}(1)$ amount of power. Therefore, power-distortion-latency scaling relationship of the CWS scheme for the case when $\boldsymbol{S}$ is α-compressible can be given by

$$D \sim \left(\frac{P_{\text{tot}}}{\log(n)}\right)^{-2\alpha/(2\alpha+1)} \sim \left(\frac{L}{\log(n)}\right)^{-2\alpha/(2\alpha+1)}, \tag{80}$$

while for an M-sparse signal with M and L scaling as $M \asymp n^{\mu}$, $0 \leq \mu < 1$, and $L \asymp n^{\beta}$, $0 < \beta < 1$, it can be given by

$$D \sim \log(n)\, {P_{\text{tot}}}^{-1+\frac{\mu}{\beta}} \sim \log(n)\, L^{-1+\frac{\mu}{\beta}}. \tag{81}$$

Comparison of these power-distortion-latency relationships with the ones achievable in Section V yields an interesting insight: regardless of the compressibility (sparsity) of $\boldsymbol{S}$, if there is enough prior knowledge about the underlying physical phenomenon, the distortion achievable under CWS would always be greater than the one achievable in the known subspace case, when using the same amount of power and latency *and* identical reconstruction basis. As an example, whereas one can obtain a distortion scaling of $D \asymp n^{-2\alpha/(2\alpha+1)}$ by employing $(P_{\text{tot}} \asymp)L \asymp n^{1/(2\alpha+1)}$ projections for the estimation of an α-compressible signal in the known subspace case, the distortion scaling in the unknown/adaptive subspace case, when using the same number of projections (and ignoring the $\log(n)$ factor), can only be given by $D \preceq n^{-2\alpha/(2\alpha+1)^2}$ – a significantly slower decay (cf. (45), (80)).

On the other hand, by virtue of a toy example, we have already seen at the start of this section what could happen to the distortion scaling in the known subspace case if the known subspace assumption is false and that is where the universality of CWS comes into play: given sufficient prior knowledge about the underlying signal field, CWS can be far from optimal but under circumstances where there is *little* or *no* knowledge available about the signal field, CWS should be the estimation scheme of choice.

VII. Impact of Fading and Imperfect Phase Synchronization

The joint source-channel communication architecture presented in Section IV for computing distributed projections in WSNs (and extended to estimation of compressible/sparse signals in Sections V and VI) is analyzed under the assumptions that the network is fully synchronized and transmissions from the sensor nodes do not undergo fading. This assumption may not hold in practice for sensor network deployments in scattering environments and due to drifts in phases of sensor oscillators. Therefore, we relax these assumptions in this section and study the impact of fading and imperfect phase synchronization on the previously obtained scaling laws. In particular, we establish that (i) the power-distortion-latency laws of Sections IV and V continue to hold as long as the random channel gains of received signals at the FC (due to fading and phase synchronization errors) have a non-zero mean; and (ii) the CWS scaling laws continue to hold as long as the mean of these random channel gains is not too small.

A. Distributed Projections in Wireless Sensor Networks

We begin by analyzing the impact of fading and imperfect phase synchronization on the power-distortion-latency scaling law of the proposed communication scheme for computing distributed projections (cf. Theorem 1). This is accomplished by assuming that the communication scheme is still described by the encoders in (25) but, as a result of narrowband fading and phase synchronization errors, each sensor's transmitted signal is received at the FC after multiplication by a random channel gain $\gamma_j^k \triangleq g_j^k \cos(\Delta_j^k)$, $j = 1, 2, \ldots, n$, where the random variables $\{g_j^k\}$ and $\{\Delta_j^k\}$ are i.i.d. (across sensors), that are also assumed to be independent of each other [36], [37].[14] Note that $\{g_j^k\}$ are non-negative valued random variables – typically modeled as Rayleigh, Rician or log-normal distributed – and correspond to random fading envelopes of received signals at the FC, whereas $\{\Delta_j^k\}$ model the combined effect of random phase-shifts due to multi-path scattering and imperfect phase synchronization between the sensors and the FC. We assume that the precise values and distributions of these random variables are not available to the sensors or the FC, but their means are known at the FC.

Consequently, as a result of fading and imperfect phase synchronization, the FC receives

$$\begin{aligned} r^k &= \sum_{j=1}^{n} \sqrt{h_j}\,\gamma_j^k y_j^k + z^k = \sqrt{\rho}\sum_{j=1}^{n} \varphi_j \gamma_j^k x_j^k + z^k \\ &= \sqrt{\rho}\,\boldsymbol{\varphi}^T\big(\boldsymbol{\gamma}^k \odot \boldsymbol{s}^k\big) + \Big(\sqrt{\rho}\,\boldsymbol{\varphi}^T\big(\boldsymbol{\gamma}^k \odot \boldsymbol{w}^k\big) + z^k\Big), \end{aligned} \tag{82}$$

where $\boldsymbol{\gamma}^k \triangleq (\gamma_1^k, \ldots, \gamma_n^k)^T$, $\odot$ represents a Hadamard product (element-wise multiplication), and ρ is still given by the

[13]The stated results hold for all $q \geq 1$; the explicit dependencies of the leading constants on the quantization parameter q are derived in [7].

[14]Recall that we are doing real-signaling; the random channel gains are, therefore, given by $\gamma_j^k = g_j^k \cos(\Delta_j^k)$ instead of $\gamma_j^k = g_j^k e^{\jmath \Delta_j^k}$.

expression in (33). Clearly, this coincides with the received signal in (26) if and only if $\boldsymbol{\gamma}^k = \mathbf{1}$, where $\mathbf{1}$ denotes a vector of all ones. However, by a slight modification of the decoder in (27), it can be shown that the scaling law established in Theorem 1 is still achievable as long as the distribution of random channel gains is such that the network remains at least "barely synchronized" in the sense that $\mathbb{E}[\gamma_j^k] \triangleq \overline{\gamma} \neq 0$. This condition would be satisfied, e.g., if $\Delta_j^k \sim \mathfrak{unif}[-\pi+\epsilon, \pi-\epsilon]$ for any $\epsilon > 0$. The modified decoder G in this scenario is given by

$$\begin{aligned} \widehat{\upsilon}^k \;&=\; G\left(r^k\right) \;=\; \frac{r^k}{\overline{\gamma}\sqrt{\rho}} \\ &=\; \boldsymbol{\varphi}^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{s}^k\right) + \boldsymbol{\varphi}^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{w}^k\right) + \frac{z^k}{\overline{\gamma}\sqrt{\rho}}\,, \end{aligned} \tag{83}$$

where $\widetilde{\boldsymbol{\gamma}}^k \triangleq (\frac{\gamma_1^k}{\overline{\gamma}}, \ldots, \frac{\gamma_n^k}{\overline{\gamma}})^T$, and the achievable distortion using this modified decoder can be characterized by the following result.

Theorem 2: Let $\boldsymbol{\varphi} \in \mathbb{R}^n$ and let $\upsilon^k = \boldsymbol{\varphi}^T \boldsymbol{s}^k$. Suppose that the random channel gains $\{\gamma_j^k\}$ due to fading and imperfect phase synchronization are i.i.d across sensors and have a non-zero mean ($\mathbb{E}[\gamma_j^k] = \overline{\gamma} \neq 0$).[15] Then, given the sensor network model of Section II, the joint source-channel communication scheme described by the encoders in (25) and the modified decoder in (83) can achieve the following end-to-end distortion by employing only one channel use per source observation

$$\begin{aligned} D_\upsilon \;\leq\; &\left(\frac{\sigma_w^2\,\overline{\overline{\gamma}} + B^2\left(\overline{\overline{\gamma}} - \overline{\gamma}^2\right)}{\overline{\gamma}^2}\right)\|\boldsymbol{\varphi}\|_2^2 \;+ \\ &\qquad\left(\frac{\sigma_z^2\, {d_u}^{\zeta}\left(B^2 + \sigma_w^2\right)}{\lambda\overline{\gamma}^2 P}\right)\|\boldsymbol{\varphi}\|_2^2\,, \end{aligned} \tag{84}$$

where $\widehat{\upsilon}^k$ is the estimate of υ^k at the FC and $\overline{\overline{\gamma}} \triangleq \mathbb{E}[|\gamma_j^k|^2] \leq 1$ by the law of conservation of energy.

Proof: To establish this theorem, note that (83) implies that $\forall\ k \in \mathbb{N}$

$$\begin{aligned} \mathbb{E}\left[\left|\upsilon^k - \widehat{\upsilon}^k\right|^2\right] \;&=\; \mathbb{E}\Bigg[\bigg|\left(\boldsymbol{\varphi}^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{s}^k\right) - \boldsymbol{\varphi}^T \boldsymbol{s}^k\right) + \\ &\qquad\qquad \boldsymbol{\varphi}^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{w}^k\right) + \frac{z^k}{\overline{\gamma}\sqrt{\rho}}\bigg|^2\Bigg] \\ &\overset{(a)}{=}\; \mathbb{E}\left[\left|\boldsymbol{\varphi}^T\left(\left(\widetilde{\boldsymbol{\gamma}}^k - \mathbf{1}\right) \odot \boldsymbol{s}^k\right)\right|^2\right] \;+\; \frac{\sigma_w^2\,\overline{\overline{\gamma}}}{\overline{\gamma}^2}\|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_z^2}{\overline{\gamma}^2\rho} \\ &=\; \sum_{j=1}^{n}\left(\varphi_j s_j^k\right)^2 \mathbb{E}\left[\left(\frac{\gamma_j^k}{\overline{\gamma}} - 1\right)^2\right] \;+\; \frac{\sigma_w^2\,\overline{\overline{\gamma}}}{\overline{\gamma}^2}\|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_z^2}{\overline{\gamma}^2\rho} \\ &\overset{(b)}{\leq}\; \frac{B^2\left(\overline{\overline{\gamma}} - \overline{\gamma}^2\right)}{\overline{\gamma}^2}\|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_w^2\,\overline{\overline{\gamma}}}{\overline{\gamma}^2}\|\boldsymbol{\varphi}\|_2^2 \;+\; \frac{\sigma_z^2}{\overline{\gamma}^2\rho}\,, \end{aligned} \tag{85}$$

where (a) essentially follows from the fact that the random channel gain vector $\boldsymbol{\gamma}^k$ is independent of the zero-mean measurement noise vector $\boldsymbol{w}^k$ and zero-mean communication noise z^k, and (b) primarily follows from the fact that $|s_j^k| \leq B$. Finally, to complete the proof of the theorem, we substitute in (85) the value of ρ from (33) and take the limit in k to obtain (84). ∎

Remark 10: A corresponding lower bound on the projection coefficient distortion D_υ under the modified decoder of (83) is given by (28). This follows trivially from (a) in (85) and the fact that $\overline{\gamma}^2 \leq \overline{\overline{\gamma}} \leq 1$.

Remark 11: Since the structure of source-channel encoders $(F_1, \ldots, F_n)$ remains unchanged under fading and imperfect phase synchronization, the total network power consumption associated with achieving the distortion in (84) is still given by (34).

Notice that even under the effects of fading and imperfect synchronization, the projection coefficient distortion D_υ achieved by the proposed joint source-channel communication architecture (using the modified decoder of (83)) is given by a sum of two separate terms, the first of which scales like $\|\boldsymbol{\varphi}\|_2^2$, while the second term that is primarily due to the noisy communication channel scales like $\|\boldsymbol{\varphi}\|_2^2/\lambda$ (cf. (84), Remark 10). Comparing this observation with the scaling law established in Section IV shows that Theorem 2 describes the same distortion scaling behavior as Theorem 1, with the only difference being that the scaling constants are now different (they depend upon the second-order statistics of channel gains). In particular, $L_\upsilon = 1$ and $P_{\text{tot},\upsilon} = \mathcal{O}(1)$ is still sufficient to ensure that $D_\upsilon \asymp \|\boldsymbol{\varphi}\|_2^2 \asymp D_\upsilon^*$, as long as $\overline{\gamma} \neq 0$ (cf. Corollary 1).

B. Distributed Estimation from Noisy Projections: Known Subspace

Similar to the case of estimation of a single projection coefficient under the effects of fading and imperfect phase synchronization, it is a simple exercise to show that by using the joint source-channel communication scheme described by the encoders in (36) and under a slightly modified decoder $\boldsymbol{G}$ given by

$$\widehat{\boldsymbol{s}}^k \;=\; \boldsymbol{G}\left(\boldsymbol{r}^k\right) \;=\; \boldsymbol{\Psi}_L^k\left(\frac{\boldsymbol{r}^k}{\overline{\gamma}\sqrt{\rho}}\right), \tag{86}$$

the end-to-end distortion of an α-compressible signal at the FC in the presence of fading and phase synchronization errors can be upper bounded by

$$\begin{aligned} D \;\leq\; C_o\, L^{-2\alpha} \;+\; &\left(\frac{L}{n}\right)\left(\frac{\sigma_w^2\,\overline{\overline{\gamma}} + B^2\left(\overline{\overline{\gamma}} - \overline{\gamma}^2\right)}{\overline{\gamma}^2}\right) \;+ \\ &\left(\frac{L}{n}\right)\left(\frac{\sigma_z^2\, {d_u}^{\zeta}\left(B^2 + \sigma_w^2\right)}{\lambda\overline{\gamma}^2 P}\right) \end{aligned} \tag{87}$$

and lower bounded by the expression in the lower bound of (39), as long as $\{\gamma_j^k\}$ are i.i.d. across sensors and $\overline{\gamma} \neq 0$. Ignoring constants, this implies that the resulting distortion of an α-compressible signal in this scenario still scales as

$$\left(\frac{L}{n}\right) \;+\; \left(\frac{L}{\lambda n}\right) \;\preceq\; D \;\preceq\; L^{-2\alpha} \;+\; \left(\frac{L}{n}\right) \;+\; \left(\frac{L}{\lambda n}\right), \tag{88}$$

[15] At the expense of some extra notation, the scaling laws stated in this section can be obtained even when $\{\gamma_j^k\}$ are not identically distributed (as long as they are independent across sensors and have non-zero means). For the sake of this exposition, however, and because it suffices to illustrate the principles, we focus only on the i.i.d. case.

i.e., has the same scaling behavior as that of D in (39). Similarly, it can be shown that using the modified decoder of (86) (with L replaced by M), the end-to-end distortion of an M-sparse signal in this scenario would scale as (cf. (59))

$$D \asymp \left(\frac{M}{n}\right) + \left(\frac{M}{\lambda n}\right). \tag{89}$$

Moreover, given that the structure of source-channel encoders $(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n)$ remains unchanged under fading and imperfect phase synchronization, the total network power consumption per source observation associated with achieving these distortion scalings for compressible and sparse signals would still be given by (40) and (60), respectively. Comparing these results with the ones obtained in Section V show that the previously established power-distortion-latency scaling laws for estimation of compressible and sparse signals in the known subspace case continue to hold under the effects of fading and imperfect phase synchronization, provided $\{\gamma_j^k\}$ have a non-zero mean and the FC uses the modified decoder of (86).

Remark 12: Note that these results are similar in spirit to some of the earlier results obtained in the context of joint source-channel communication for distributed estimation of sources – see, e.g., [3], [18], [47], [48]. In particular, those results also indicate that fading (and/or imperfect synchronization) tends to have no effect on the distortion scaling as long as the random channel gains have non-zero means.

C. Compressive Wireless Sensing

In the presence of phase synchronization errors only (no fading, i.e., $\gamma_j^k = \cos(\Delta_j^k)$ only), CWS observations are given by ($\forall\ i = 1, \ldots, L$)

$$\widehat{v}_i^k \;=\; \boldsymbol{\phi}_i^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{s}^k\right) + \boldsymbol{\phi}_i^T\left(\widetilde{\boldsymbol{\gamma}}^k \odot \boldsymbol{w}^k\right) + \frac{z_i^k}{\overline{\gamma}\sqrt{\rho}}, \tag{90}$$

where $\widetilde{\boldsymbol{\gamma}}^k = (\frac{\gamma_1^k}{\overline{\gamma}}, \ldots, \frac{\gamma_n^k}{\overline{\gamma}})^T$ (see (68)). Defining the vector $\widetilde{\boldsymbol{\gamma}}^k$ as $\widetilde{\boldsymbol{\gamma}}^k \triangleq \boldsymbol{1} + \boldsymbol{\delta}^k$ and substituting into the above gives

$$\widehat{v}_i^k \;=\; \boldsymbol{\phi}_i^T \boldsymbol{s}^k + \boldsymbol{\phi}_i^T \boldsymbol{w}^k + \boldsymbol{\phi}_i^T\left(\boldsymbol{\delta}^k \odot (\boldsymbol{s}^k + \boldsymbol{w}^k)\right) + \frac{z_i^k}{\overline{\gamma}\sqrt{\rho}}, \tag{91}$$

where $\boldsymbol{\delta}^k$ is a zero-mean random vector with i.i.d. entries given by $\delta_j^k = \widetilde{\gamma}_j^k - 1,\ j = 1, \ldots, n$. Comparing this with (68), we see that the net effect of phase synchronization errors is the introduction of a new noise-like term of the form $\boldsymbol{\phi}_i^T\left(\boldsymbol{\delta}^k \odot (\boldsymbol{s}^k + \boldsymbol{w}^k)\right)$. Foregoing a rigorous theoretical analysis of the effects of this contribution, we instead assume (by the Central Limit Theorem) that it is approximately Gaussian distributed, in which case it can be treated like the projected noise $\{\boldsymbol{\phi}_i^T \boldsymbol{w}^k\}$, as in Section VI-A. Further, assuming that $\Delta_j^k \overset{\text{i.i.d.}}{\sim} \texttt{unif}[-b, b]$ and that b is small, we can use a one-term Taylor series approximation of the variance of the new zero-mean noise contribution. The result is that each CWS observation is again given by (74) but the equivalent noise variance is given by $\sigma^2 = \sigma_w^2 + \sigma_z^2\, d_u{}^{\zeta}\left(B^2 + \sigma_w^2\right)/(\overline{\gamma}^2 P) + (B^2 + \sigma_w^2) b^4/45$, the last term in the expression being the contribution of the new phase synchronization error term.

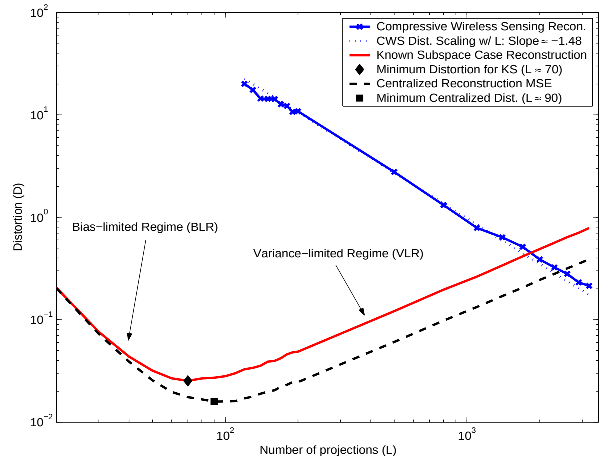


Fig. 6. Distortion scaling of a fixed length α-compressible signal as a function of number of projections L under both known and unknown subspace assumptions (log-log scale): number of sensor nodes n = 8192; α = 1 (in Haar basis); baseline MSE (σ_w^2) = 1; measurement SNR = 20 dB; received communication SNR per projection = 0 dB.

More generally, if we also define the fading envelope of each sensor's transmission as $g_j^k \triangleq 1 + \epsilon_j^k$, then the overall random channel gain of each received signal becomes $\gamma_j^k = g_j^k \cos(\Delta_j^k) = (1 + \epsilon_j^k)(1 + \delta_j^k) = 1 + \epsilon_j^k + \delta_j^k + \epsilon_j^k \delta_j^k$. The net result is a new noise-like term of the form $\boldsymbol{\phi}_i^T\left((\boldsymbol{\epsilon}^k + \boldsymbol{\delta}^k + \boldsymbol{\epsilon}^k \boldsymbol{\delta}^k) \odot (\boldsymbol{s}^k + \boldsymbol{w}^k)\right)$. With appropriate modeling of the ϵ_j^k terms, the additional variance due to this contribution can also be computed and the optimization problem in (77) can be updated accordingly. This approach was used in the simulations and appears to work well in practice for a range of phase synchronization errors, with or without mild fading, as shown in Fig. 10.

VIII. Simulation Results

In this section, we present a few simulation results to numerically demonstrate some of the power-distortion-latency relationships of our scheme under both known and unknown/adaptive subspace assumptions. All signals discussed in this section are contaminated with zero-mean additive white Gaussian measurement noise of variance $\sigma_w^2 = 1$, i.e., the baseline MSE of all signals is taken to be 1. Moreover, the measurement SNR of all signals, defined to be $\left(\|\boldsymbol{s}^k\|_2^2/n\right)/\sigma_w^2$, is given by $\text{SNR}_{\text{meas}} = 20$ dB, and the received communication SNR for each projection, defined to be ρ/σ_z^2, is given by $\text{SNR}_{\text{comm}} = 0$ dB (unless otherwise stated).

The first simulation result, corresponding to Fig. 6, illustrates the distortion scaling D of a spatially piecewise smooth signal field with the number of projections L using both CWS and known subspace case reconstructions, where the signal field is sampled by $n = 8192$ sensor nodes in a noisy manner. Such signals tend to be compressible in the Haar domain with $\alpha = 1$ and this value of α was also verified numerically. For the purposes of known subspace reconstruction, the observation vector is projected onto L Haar basis elements corresponding to L largest coefficients of the noiseless vector

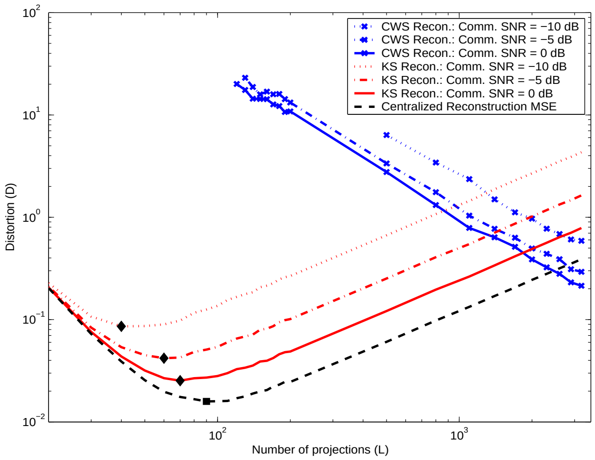


Fig. 7. Distortion scaling of a fixed length α-compressible signal as a function of number of projections L for various values of received communication SNR per projection under both known and unknown subspace assumptions (log-log scale): number of sensor nodes $n = 8192$; $\alpha = 1$ (in Haar basis); baseline MSE $(\sigma_w^2) = 1$; measurement SNR = 20 dB.

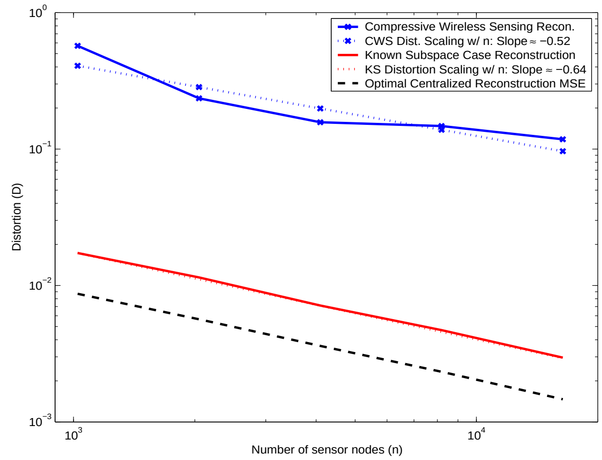


Fig. 8. Distortion scaling of an M-sparse signal as a function of number of sensor nodes n under both known and unknown subspace assumptions (log-log scale): number of non-zero coefficients $M \asymp n^{1/3}$ (in Haar basis); baseline MSE $(\sigma_w^2) = 1$; measurement SNR = 20 dB; received communication SNR per projection = 0 dB; number of projections – Known subspace case reconstruction: $L = M \asymp n^{1/3}$, CWS reconstruction: $L \asymp \log(n)\, n^{1/2} M \asymp \log(n)\, n^{5/6}$.

using the scheme described in Section V, while for the case of CWS reconstruction, the observation vector is projected onto L random Rademacher vectors. The resultant reconstruction MSEs are shown in the figure using solid curves (on a log-log scale), while the dotted curve and dashed curve in the figure correspond to linear fit of CWS distortion curve and reconstruction MSE in a centralized setting (i.e., $\sigma_z^2 = 0$), respectively. Finally, the total network power consumption for both CWS and known subspace case reconstructions is given by $P_{\text{tot}} \asymp L$, owing to the fact that we have chosen $\lambda = \mathcal{O}(1)$ in this simulation.

As predicted by the theory, distortion curve for known subspace reconstruction in Fig. 6 hits its minimum at a point where the distortion due to the approximation error is balanced by the distortion due to the observation and communication noise, and starts to rise after $L \approx 70$ projections since each subsequent projection contributes only a small amount of signal but a larger amount of noise. Note that minimum distortion in the centralized setting is attained for $L \approx 90$ projections. This is because distortion scaling constants in the known subspace case depend upon σ_w^2 and σ_z^2/ρ (see (39)), while $\sigma_z^2 = 0$ in the centralized case. For the case of CWS, distortion scaling follows a slope of -1.48 that turns out to be better than the expected value of $-2\alpha/(2\alpha+1) = -2/3$ (see (78)). This, however, does not contradict the results reported in Section VI since we only have upper bounds for distortion scaling in the CWS case. Finally, Fig. 7 illustrates the fact that varying the received communication SNR per projection has no effect on the scaling behavior of known subspace and CWS reconstruction MSEs (except a change in the scaling constants).

The second simulation result, corresponding to Fig. 8, illustrates the distortion scaling D of an M-sparse signal field with the number of sensor nodes using both CWS and known subspace case reconstructions, where we also scale the number of DoF in the signal as $M \asymp n^{\mu} = n^{1/3}$ in the Haar basis. For the purposes of known subspace case reconstruction, the observation vector is projected onto $L = M$ Haar basis elements corresponding to the M non-zero coefficients of the noiseless vector using the scheme described in Section V, while for the case of CWS reconstruction, the observation vector is projected onto $L \asymp \log(n)\, n^{1/2} M \asymp \log(n)\, n^{5/6}$ random Rademacher vectors. The resultant reconstruction MSEs are shown in the figure using solid curves (on a log-log scale), while the dotted and dashed curves in the figure correspond to linear fit of known subspace/CWS distortion curves and reconstruction MSE in a centralized setting, respectively. Finally, the total network power consumption for CWS and known subspace case reconstructions is given by $P_{\text{tot}} \asymp \log(n)\, n^{5/6}$ and $P_{\text{tot}} \asymp n^{1/3}$, respectively, owing to the fact that we have chosen $\lambda = \mathcal{O}(1)$ in this simulation.

As predicted by the theory, the distortion scaling curve for known subspace reconstruction in this case tends to follow a slope of $-1+\mu \approx -0.64$ (see (62)). Similarly, the distortion scaling curve for CWS reconstruction in this case can be expressed as $D \asymp M \log(n)/L \approx n^{-0.52}$ – again in accordance with the theory (see (79)). Finally, Fig. 9 and Fig. 10 illustrate the robustness of our proposed scheme to a range of phase synchronization errors, with or without fading, under both known and unknown/adaptive subspace assumptions.

IX. Conclusion

In this paper, we have presented a distributed joint source-channel communication architecture for estimation of sensor network data at the FC and analyzed the corresponding power-distortion-latency relationships as a function of the number of sensor nodes. Our approach is built on distributed computation of appropriately chosen projections of the sensor data at the fusion center. Phase-coherent transmissions from the sensors enable exploitation of the distributed beamforming gain for dramatic reductions in power consumption. A few distinct features of our approach are: 1) processing and communication

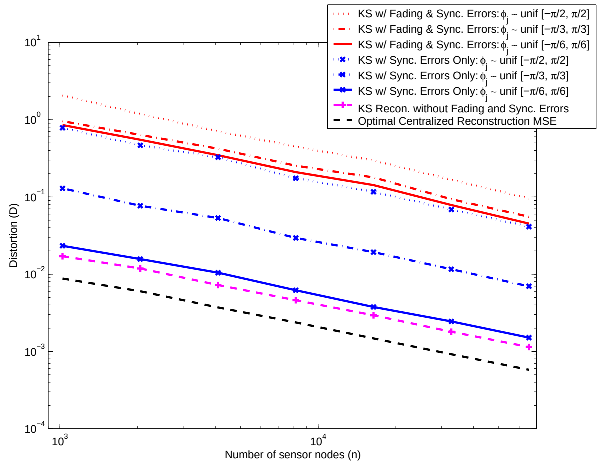


Fig. 9. Distortion scaling of an M-sparse signal as a function of number of sensor nodes n under the effects of fading and phase synchronization errors (Known subspace case reconstruction only): number of non-zero coefficients $M \asymp n^{1/3}$ (in Haar basis); baseline MSE (σ_w^2) = 1; measurement SNR = 20 dB; received communication SNR per projection = 0 dB; fading envelope: Rayleigh distributed; number of projections $L = M \asymp n^{1/3}$.

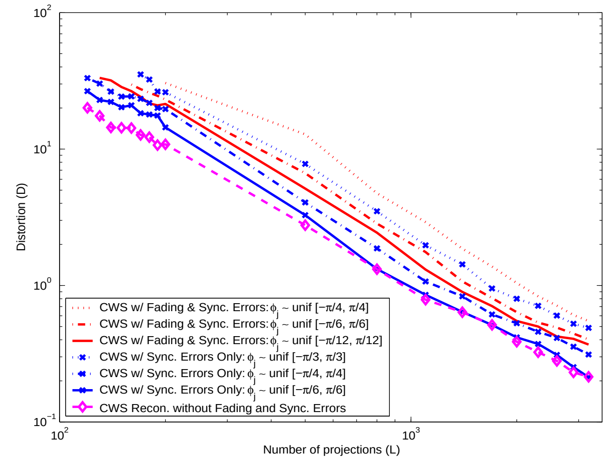


Fig. 10. Distortion scaling of a fixed length α-compressible signal as a function of number of projections L under the effects of fading and phase synchronization errors (CWS reconstruction only): number of sensor nodes n = 8192, α = 1 (in Haar basis), baseline MSE (σ_w^2) = 1, measurement SNR = 20 dB, received communication SNR per projection = 0 dB; fading envelope: Rician distributed (K-factor of 7.5).

are combined into one distributed projection operation; 2) it requires almost no in-network processing and communication; and 3) given sufficient prior knowledge about the sensed data, asymptotically consistent signal estimation is possible even if the total network power consumption goes to zero.

In addition, we have also introduced and analyzed a universal estimation scheme – compressive wireless sensing (CWS) – that provides asymptotically consistent signal estimates, even if little or no prior knowledge about the sensed data is assumed. Furthermore, power and latency requirements in CWS grow at most sub-linearly with the number of nodes in the network. This universality, however, comes at the cost of less favorable power-distortion-latency relationship: the absence of sufficient prior knowledge about the signal field leads to probing the entire n-dimensional space using random projections instead of focusing on the subspace of interest. However, for precisely the same reason, CWS has the ability to capture part of signal under all circumstances and does not require reprogramming of the network for different sensing scenarios – different hypotheses on the signal field structure can be tested at the fusion center via the reconstruction algorithms. Furthermore, projecting the sensor network data onto a fixed subspace may result in a distortion much greater than the one achievable by CWS if prior information about the signal field is inaccurate. 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.

Appendix I
In-Network Collaboration: Power-Distortion Trade-off Revisited

Analysis in Sections V-A and V-B shows that $\lambda = \mathcal{O}(1)$ is necessary for optimal distortion scaling in estimation of compressible and sparse signals, resulting in $P_{\text{tot}} \asymp n^{1/(2\alpha+1)}$ and $P_{\text{tot}} \asymp M$, respectively. In this appendix we partially address the question: *How good is the power-distortion scaling of our proposed scheme?* While a comparison with all conceivable distributed estimation schemes does not seem possible, we compare the performance of the proposed scheme (which does not require data exchange between nodes) to a more favorable and idealized setup where the nodes in the network can communicate their observations in an error-free manner to a designated cluster of $1 \leq \tilde{n} \leq n$ nodes. We do not make any assumptions on the nature of in-network communication and also ignore the incurred cost on energy consumption (since quantifying this cost requires making additional system-specific assumptions). Thus, our performance comparison is solely based on the power required for network-to-FC communication to achieve optimal distortion scaling. Note that $\tilde{n} = 1$ corresponds to all nodes routing their measurements to a single clusterhead in the network (using perhaps multi-hop communications), while $\tilde{n} = n$ corresponds to all the nodes in the network noiselessly sharing their data with each other (using perhaps gossip algorithms). Once $\tilde{n}$ nodes in the network have access to the entire observation vector $\boldsymbol{x}^k$ following each snapshot, they compute the required L projection coefficients (with respect to a given basis) and then coherently transmit the resulting projection coefficients to the FC using a sum transmit power of P per channel use. This effectively transforms the cluster-to-FC MAC into a point-to-point AWGN channel with $\tilde{n}$-fold power-pooling gain due to coherent beamforming of identical data.

We focus on estimation of α-compressible signals. Specifically, we assume that $\tilde{n}$ nodes in the designated cluster have access to identical estimates of the required L $\boldsymbol{\Psi}$-coefficients at the end of the data-exchange stage, i.e.,

$$\hat{\theta}_\ell^k \;=\; \boldsymbol{\psi}_\ell^T \boldsymbol{x}^k \;=\; \theta_\ell^k + \boldsymbol{\psi}_\ell^T \boldsymbol{w}^k, \quad \ell = 1, \dots, L. \tag{92}$$

By a simple extension of the encoder structure of Section V, the transmitting cluster of $\tilde{n}$ nodes coherently beamforms these

L projection coefficients per snapshot in L consecutive channel uses as follows

$$\boldsymbol{y}_j^k \;=\; \boldsymbol{F}_j\left(\{\hat{\theta}_\ell^k\}_{\ell=1}^L\right) \;=\; \left(y_{j1}^k,\ldots,y_{jL}^k\right)^T$$
$$=\; \frac{1}{\sqrt{h_j}}\left(\sqrt{\rho_1}\,\hat{\theta}_1^k,\ldots,\sqrt{\rho_L}\,\hat{\theta}_L^k\right)^T,\quad j=1,\ldots,\tilde{n}, \tag{93}$$

where $\{\rho_\ell\}_{\ell=1}^L$ are scaling factors used to satisfy the sum power constraint P in each of the L channel uses. At the end of the L-th channel use, the received signal at the input of the decoder $\boldsymbol{G}$ is given by

$$\boldsymbol{r}^k \;=\; \sum_{j=1}^{\tilde{n}}\sqrt{h_j}\,\boldsymbol{y}_j^k \;+\; \boldsymbol{z}^k$$
$$=\; \tilde{n}\left(\sqrt{\rho_1}\,\hat{\theta}_1^k,\ldots,\sqrt{\rho_L}\,\hat{\theta}_L^k\right)^T \;+\; \boldsymbol{z}^k$$
$$=\; \tilde{n}\,\boldsymbol{\Gamma}\,\boldsymbol{\theta}_L^k \;+\; \tilde{n}\,\boldsymbol{\Gamma}\left({\boldsymbol{\Psi}_L^k}^T\boldsymbol{w}^k\right) \;+\; \boldsymbol{z}^k, \tag{94}$$

where $\boldsymbol{\Gamma}\triangleq\textbf{diag}(\sqrt{\rho_1},\ldots,\sqrt{\rho_L})$, $\boldsymbol{\theta}_L^k=(\theta_1^k,\ldots,\theta_L^k)^T$, $\boldsymbol{z}^k\sim\mathcal{N}(\boldsymbol{0}_{L\times 1},\sigma_z^2\boldsymbol{I}_{L\times L})$ is an AWGN vector, and $\tilde{n}$ is the power-pooling gain of identical coherent transmissions from $\tilde{n}$ nodes. An estimate of the noiseless data vector can be formed at the FC as

$$\widehat{\boldsymbol{s}}^k \;=\; \boldsymbol{G}\left(\boldsymbol{r}^k\right) \;=\; \boldsymbol{\Psi}_L^k\left(\frac{\boldsymbol{\Gamma}^{-1}\boldsymbol{r}^k}{\tilde{n}}\right)$$
$$=\; \boldsymbol{s}^{k,(L)} \;+\; \boldsymbol{\Psi}_L^k\left({\boldsymbol{\Psi}_L^k}^T\boldsymbol{w}^k\right) \;+\; \boldsymbol{\Psi}_L^k\left(\frac{\boldsymbol{\Gamma}^{-1}\boldsymbol{z}^k}{\tilde{n}}\right). \tag{95}$$

As for fixing the values of $\{\rho_\ell\}$, note that (93) implies that $\forall\;\ell=1,\ldots,L$,

$$\sum_{j=1}^{\tilde{n}}\mathbb{E}\left[\left|y_{j\ell}^k\right|^2\right] \;=\; \sum_{j=1}^{\tilde{n}}\mathbb{E}\left[\frac{\rho_\ell}{h_j}\left|\hat{\theta}_\ell^k\right|^2\right]$$
$$=\; \mathbb{E}\left[\left|\theta_\ell^k+\boldsymbol{\psi}_\ell^T\boldsymbol{w}^k\right|^2\right]\sum_{j=1}^{\tilde{n}}\frac{\rho_\ell}{h_j}$$
$$\leq\; \tilde{n}\,\rho_\ell\,{d_u}^\zeta\left(\frac{n\,C_o}{C_p}\ell^{-2\alpha-1}+\sigma_w^2\right), \tag{96}$$

where the upper bound essentially follows from the fact that the squared magnitudes of the ordered coefficients $\{\theta_\ell^k\}$ in the case of compressible signals are bounded as $|\theta_\ell^k|^2\leq\frac{n\,C_o}{C_p}\ell^{-2\alpha-1}$ (see (9)). This implies that

$$\rho_\ell \;=\; \frac{\lambda\,P}{\tilde{n}\,{d_u}^\zeta\left(n\,\widetilde{C}_o\,\ell^{-2\alpha-1}+\sigma_w^2\right)},\quad \ell=1,\ldots,L, \tag{97}$$

would suffice to satisfy the sum power constraint of P for each of the L channel uses, where $\widetilde{C}_o\triangleq C_o/C_p$ and $\lambda\in(0,1]$ is again the power scaling factor for controlling total network power consumption. We are now ready to state the distortion achievable for an α-compressible signal under the assumption of in-network collaboration.

Theorem 3: Given the sensor network model of Section II for an α-compressible signal and under the assumption of in-network collaboration enabling $\tilde{n}$ nodes in the network to have access to the entire observation vector $\boldsymbol{x}^k$ at each time instant k, the beamforming strategy described by the encoders in (93) and the decoder in (95) can achieve the following end-to-end distortion by employing L channel uses per source observation

$$\left(\frac{L}{n}\right)\sigma_w^2 \;+\; \left(\frac{1}{\tilde{n}}\right)\left(\frac{\sigma_z^2\,\widetilde{C}_o}{\lambda\,P}\right) \;\leq\; D \;\leq\; C_o\,L^{-2\alpha} \;+$$
$$\left(\frac{L}{n}\right)\sigma_w^2\left(1+\sigma_z^2\,{d_u}^\zeta\right) \;+\; \left(\frac{1}{\tilde{n}}\right)\left(\frac{2\,\sigma_z^2\,{d_u}^\zeta\,\widetilde{C}_o}{\lambda\,P}\right). \tag{98}$$

Proof: To establish this theorem, first observe that (95) implies that $\forall\;k\in\mathbb{N}$

$$\mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k-\widehat{\boldsymbol{s}}^k\right\|_2^2\right]$$
$$\leq\; C_o\,L^{-2\alpha} \;+\; \left(\frac{L}{n}\right)\sigma_w^2 \;+\; \left(\frac{1}{n}\right)\sum_{\ell=1}^L\frac{\sigma_z^2}{\tilde{n}^2\,\rho_\ell}$$
$$=\; C_o\,L^{-2\alpha} \;+\; \left(\frac{L}{n}\right)\sigma_w^2 \;+\; \left(\frac{\sigma_z^2\,{d_u}^\zeta\,\widetilde{C}_o}{\tilde{n}\,\lambda\,P}\right)\cdot$$
$$\sum_{\ell=1}^L\ell^{-2\alpha-1} \;+\; \left(\frac{L}{n\,\tilde{n}}\right)\sigma_z^2\,{d_u}^\zeta\,\sigma_w^2$$
$$\leq\; C_o\,L^{-2\alpha} \;+\; \left(\frac{L}{n}\right)\sigma_w^2\left(1+\sigma_z^2\,{d_u}^\zeta\right) \;+$$
$$\left(\frac{1}{\tilde{n}}\right)\left(\frac{2\,\sigma_z^2\,{d_u}^\zeta\,\widetilde{C}_o}{\lambda\,P}\right), \tag{99}$$

where the last inequality follows from the fact that

$$\sum_{\ell=1}^L\ell^{-2\alpha-1} \;\leq\; 1+\int_1^L x^{-2\alpha-1}dx \;\leq\; 2 \tag{100}$$

and $\frac{L}{n\,\tilde{n}}\leq\frac{L}{n}$. Furthermore, from (95), we also have a lower bound of

$$\mathbb{E}\left[\frac{1}{n}\left\|\boldsymbol{s}^k-\widehat{\boldsymbol{s}}^k\right\|_2^2\right] \;\geq\; \left(\frac{L}{n}\right)\sigma_w^2 \;+\; \left(\frac{1}{n}\right)\sum_{\ell=1}^L\frac{\sigma_z^2}{\tilde{n}^2\,\rho_\ell}$$
$$\geq\; \left(\frac{L}{n}\right)\sigma_w^2 \;+\; \left(\frac{1}{\tilde{n}}\right)\left(\frac{\sigma_z^2\,\widetilde{C}_o}{\lambda\,P}\right). \tag{101}$$

Finally, combining the upper and lower bounds of (99) and (101), and taking the limit in k yields (98), thus completing the proof. ■

Remark 13: Under the assumption of $\tilde{n}$ nodes coherently transmitting the identical data, the cluster-to-FC MAC is effectively transformed into a point-to-point multiple-input single-output AWGN channel. Consequently, while the distortion expression in (98) has been obtained corresponding to an analog beamforming strategy of (93), a similar expression for distortion can be obtained by appropriately transforming the compressible source model into a stochastic one, and employing standard rate-distortion and capacity-cost analysis (i.e., by employing "digital" beamforming).

Remark 14: Note that the last term in the upper and lower bounds in (98) corresponds to the distortion component due to the noisy communication channel. The factor of $\tilde{n}$ in that term corresponds to the power-pooling gain due to coherent transmission of identical data: the greater the number of nodes coherently beamforming the identical data, the greater the power-pooling gain. Comparing this communication noise term to the last term in the upper and lower bounds in (39) shows that, in terms of scaling, the performance of the proposed estimation scheme of Section V is equivalent to that of an in-network collaboration based system that has a beamforming cluster of $\tilde{n} = \frac{n}{L}$ nodes.

Analysis of (98) reveals that for optimal distortion scaling under the in-network collaboration assumption, $L \asymp n^{1/(2\alpha+1)}$ and the distortion component due to the communication noise should also scale at least as $\frac{L}{n} \asymp n^{-2\alpha/(2\alpha+1)}$. Consequently, this implies that as long as the extent of in-network collaboration is such that $\tilde{n} < n^{2\alpha/(2\alpha+1)}$, one cannot achieve the optimal distortion scaling under a fixed transmit power constraint of P: the power constraint itself needs to be scaled up as $P \asymp n^{2\alpha/(2\alpha+1)}/\tilde{n}$ to achieve optimal distortion scaling. On the other hand, if the extent of in-network collaboration is such that $\tilde{n} > n^{2\alpha/(2\alpha+1)}$ then, in fact, λ need not be given by $\lambda = \mathcal{O}(1)$. Rather, in that situation, it can be scaled down as $\lambda \asymp n^{2\alpha/(2\alpha+1)}/\tilde{n}$. Going back to the two extremes of $\tilde{n} = 1$ and $\tilde{n} = n$, this means that for the case of a single clusterhead in the network, we have $P_{\text{tot}} = \mathcal{O}(\lambda\, L\, P) = \mathcal{O}(n)$ and for the case where all the nodes in the network act as a big clusterhead, we have $P_{\text{tot}} = \mathcal{O}(\lambda\, L\, P) = \mathcal{O}(1)$. Essentially, as the cardinality of the beamforming cluster $\tilde{n}$ scales up as $1 \nearrow n$, the total network power scales down from $\mathcal{O}(n)$ to $\mathcal{O}(1)$. Remarkably, the proposed estimation scheme of Section V achieves the performance equivalent to that of a cluster with $\tilde{n} = n^{2\alpha/(2\alpha+1)}$ nodes, without requiring any in-network collaboration. Furthermore, while we have ignored the cost of in-network communication, we expect that it will increase monotonically with increase in the size of the beamforming cluster $\tilde{n}$.

REFERENCES

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

[2] 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 Inform. Processing in Sensor Networks (IPSN'05)*, Los Angeles, CA, Apr. 2005, pp. 332–339.

[3] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.

[4] M. Vetterli, "Wavelets, approximation, and compression," *IEEE Signal Processing Mag.*, vol. 18, no. 5, pp. 59–73, Sept. 2001.

[5] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[7] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

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

[9] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 51–60, Mar. 2002.

[10] 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. Conf. on Inform. Processing in Sensor Networks (IPSN'03)*, Apr. 2003, pp. 1–16.

[11] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A bit-conservation principle," in *Proc. 2nd Intl. Conf. on Inform. Processing in Sensor Networks (IPSN'03)*, Apr. 2003, pp. 17–31.

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

[13] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2435–2476, Oct. 1998.

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

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, Inc., 1991.

[16] 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.

[17] S. Cui, J. Xiao, A. Goldsmith, Z. Q. Luo, , and H. V. Poor, "Energy-efficient joint estimation in sensor networks: Analog vs. digital," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Sig. Proc. (ICASSP'05)*, Philadelphia, PA, Mar. 2005, pp. 745–748.

[18] G. Mergen and L. Tong, "Type-based estimation over multiaccess channels," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 613–626, Feb. 2006.

[19] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," pre-print. [Online]. Available: http://www.dsp.ece.rice.edu/cs/DCS112005.pdf

[20] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *Proc. 5th Intl. Conf. on Inform. Processing in Sensor Networks (IPSN'06)*, Nashville, TN, Apr. 2006, pp. 177–185.

[21] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. 5th Intl. Conf. on Inform. Processing in Sensor Networks (IPSN'06)*, Nashville, TN, Apr. 2006, pp. 51–59.

[22] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Scalable feedback control for distributed beamforming in sensor networks," in *Proc. Intl. Symp. Inform. Theory (ISIT'05)*, Sept. 2005, pp. 137–141.

[23] ——, "Distributed transmit beamforming using feedback control," pre-print. [Online]. Available: http://www.ece.ucsb.edu/Faculty/Madhow/Publications/mudumbai_fback.pdf

[24] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 719–746, Mar. 1992.

[25] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1999.

[26] V. K. Goyal, *Single and Multiple Description Transform Coding with Bases and Frames*. Philadelphia, PA: SIAM, 2001.

[27] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1895–1921, July 2002.

[28] R. A. Devore, "Nonlinear approximation," in *Acta Numerica*, A. Iserles, Ed. Cambridge, U.K.: Cambridge University Press, 1998, vol. 7, pp. 51–150.

[29] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[30] A. Cohen, W. Dahmen, and R. A. Devore, "Compressed sensing and best k-term approximation," pre-print. [Online]. Available: http://www.math.sc.edu/∼devore/publications/CDDSensing_6.pdf

[31] D. L. Donoho, "Sparse components of images and optimal atomic decompositions," in *Constructive Approximation*. New York: Springer, Jan. 2001, vol. 17, no. 3, pp. 353–382.

[32] R. Wagner, H. Choi, R. Baraniuk, and V. Delouille, "Distributed wavelet transform for irregular sensor network grids," in *Proc. 13th IEEE/SP Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005, pp. 1196–1201.

[33] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.

[34] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.

[35] H. El-Gamal, "On the scaling laws of dense wireless sensor networks: The data gathering channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 1229–1234, Mar. 2005.

[36] A. Goldsmith, *Wireless Communications*. New York: Cambridge University Press, 2005.

[37] T. S. Rappaport, *Wireless Communications: Principles and Practice*. New Jersey: Prentice-Hall, 2002.

[38] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1993.

[39] L. Wasserman, *All of Nonparametric Statistics*, ser. Springer Texts in Statistics. New York: Springer-Verlag, 2006.

[40] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," in *Biometrika*. Oxford, U.K.: Oxford University Press, 1994, vol. 81, no. 3, pp. 425–455.

[41] ——, "Ideal denoising in an orthonormal basis chosen from a library of bases," in *C. R. Acad. Sci., Ser. I*, Paris, 1994, vol. 319, pp. 1317–1322.

[42] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proc. IEEE Infocom*, vol. 3, Miami, FL, Mar. 2005, pp. 1653–1664.

[43] A. G. Dimakis, A. D. Sarwate, and M. Wainwright, "Geographic gossip : Efficient aggregation for sensor networks," in *Proc. 5th Intl. Conf. on Inform. Processing in Sensor Networks (IPSN'06)*, Nashville, TN, Apr. 2006, pp. 69–76.

[44] A. P. Korostelev and A. B. Tsybakov, *Minimax Theory of Image Reconstruction*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, 1993, vol. 82.

[45] T. Sivanadyan and A. M. Sayeed, "Active wireless sensing for rapid information retrieval in sensor networks," in *Proc. 5th Intl. Conf. on Inform. Processing in Sensor Networks (IPSN'06)*, Nashville, TN, Apr. 2006, pp. 85–92.

[46] P. L'Ecuyer, "Random number generation," in *Handbook of Computational Statistics: Concepts and Methods*, J. E. Gentle, W. Härdle, and Y. Mori, Eds. Springer-Verlag, 2004, ch. II.2, pp. 35–70.

[47] K. Liu, H. El-Gamal, and A. M. Sayeed, "On optimal parametric field estimation in sensor networks," in *Proc. 13th IEEE/SP Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005, pp. 1170–1175.

[48] P. Gao and C. Tepedelenlioglu, "Practical issues in parameter estimation over fading channels with TBMA wireless sensor networks," in *Proc. 49th IEEE Global Telecommunications Conference (GLOBECOM'06)*, San Francisco, CA, Nov. 2006.

Waheed U. Bajwa received the B.E. degree in electrical engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan in 2001 and the M.S. degree in electrical engineering from the University of Wisconsin-Madison in 2005. He was an Intern in the Islamabad office of Communications Enabling Technologies, Irvine, CA (now Quartics) during 2000-2001, a Design Engineer in Communications Enabling Technologies from 2001-2003, and an R&D Intern in the RF and Photonics Lab of General Electric Global Research Center, Niskayuna, NY during the summer of 2006. He is currently pursuing the Ph.D. degree in electrical engineering (mathematics minor) from the University of Wisconsin-Madison under the supervision of Prof. Akbar Sayeed and Prof. Robert Nowak. His research interests include wireless communications, statistical signal processing, information theory and statistical learning theory.

Mr. Bajwa received the Morgridge Distinguished Graduate Fellowship from the University of Wisconsin-Madison in 2003, and Best in Academics Gold Medal and President's Gold Medal in Electrical Engineering from the National University of Sciences and Technology in 2001. He was also a Junior NUST Student of the Year in 2000.

Jarvis D. Haupt received the B.S. degree in Electrical Engineering (Computer Engineering option) with a second major in Mathematics, and the M.S. degree in Electrical Engineering in 2002 and 2003, respectively, from the University of Wisconsin-Madison. He is currently pursuing the Ph.D. degree in Electrical Engineering (Mathematics minor) from the University of Wisconsin-Madison under the direction of Professor Robert Nowak. His research interests include statistical signal processing and statistical learning theory.

In the summers of 2004 and 2005, Mr. Haupt was a Hardware Research and Development Intern, examining network routing and bandwidth optimizations for Cray, Inc. He was a teaching assistant for two semesters and received honorable mention for the Gerald Holdridge Teaching Award for his work during the 2003-2004 academic year. He was also the co-chair of the University of Wisconsin College of Engineering Teaching Improvement Program during the 2004-2005 academic year.

Akbar M. Sayeed received the B.S. degree (1991) from the University of Wisconsin-Madison, and the M.S. (1993) and Ph.D. (1996) degrees from the University of Illinois at Urbana-Champaign, all in Electrical and Computer Engineering. He was a postdoctoral fellow at Rice University (1996-1997) and he has been with the University of Wisconsin-Madison since 1997 where he is currently Associate Professor of Electrical and Computer Engineering. His current research interests include wireless communications, multi-dimensional communication theory, statistical signal processing, information theory, and applications in wireless networks.

Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and a UW Grainger Junior Faculty Fellowship (2003). He served as an Associate Editor for the IEEE Signal Processing Letters (1999-2002), and is currently serving on the signal processing for communications technical committee of the IEEE Signal Processing Society.

Robert D. Nowak received the B.S. (with highest distinction), M.S., and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1990, 1992, and 1995, respectively. He was a Postdoctoral Fellow at Rice University in 1995-1996, an Assistant Professor at Michigan State University from 1996-1999, held Assistant and Associate Professor positions at Rice University from 1999-2003, and was a Visiting Professor at INRIA in 2001. Dr. Nowak is now the McFarland-Bascom Professor of Engineering at the University of Wisconsin-Madison. He has served as an Associate Editor for the IEEE Transactions on Image Processing, and is currently an Associate Editor for the ACM Transactions on Sensor Networks and the Secretary of the SIAM Activity Group on Imaging Science. He has also served as a Technical Program Chair for the IEEE Statistical Signal Processing Workshop and the IEEE/ACM International Symposium on Information Processing in Sensor Networks.

Dr. Nowak received the General Electric Genius of Invention Award in 1993, the National Science Foundation CAREER Award in 1997, the Army Research Office Young Investigator Program Award in 1999, the Office of Naval Research Young Investigator Program Award in 2000, and IEEE Signal Processing Society Young Author Best Paper Award in 2000. His research interests include statistical signal processing, machine learning, imaging and network science, and applications in communications, bio/medical imaging, and genomics.