Linear Coherent Decentralized Estimation

Jin-Jun Xiao, *Member, IEEE*, Shuguang Cui, *Member, IEEE*, Zhi-Quan Luo, *Fellow, IEEE*, and Andrea J. Goldsmith, *Fellow, IEEE*

***Abstract*—We consider the distributed estimation of an unknown vector signal in a resource constrained sensor network with a fusion center. Due to power and bandwidth limitations, each sensor compresses its data in order to minimize the amount of information that needs to be communicated to the fusion center. In this context, we study the linear decentralized estimation of the source vector, where each sensor linearly encodes its observations and the fusion center also applies a linear mapping to estimate the unknown vector signal based on the received messages. We adopt the mean squared error (MSE) as the performance criterion. When the channels between sensors and the fusion center are orthogonal, it has been shown previously that the complexity of designing the optimal encoding matrices is NP-hard in general. In this paper, we study the optimal linear decentralized estimation when the multiple access channel (MAC) is coherent. For the case when the source and observations are scalars, we derive the optimal power scheduling via convex optimization and show that it admits a simple distributed implementation. Simulations show that the proposed power scheduling improves the MSE performance by a large margin when compared to the uniform power scheduling. We also show that under a finite network power budget, the asymptotic MSE performance (when the total number of sensors is large) critically depends on the multiple access scheme. For the case when the source and observations are vectors, we study the optimal linear decentralized estimation under both bandwidth and power constraints. We show that when the MAC between sensors and the fusion center is noiseless, the resulting problem has a closed-form solution (which is in sharp contrast to the orthogonal MAC case), while in the noisy MAC case, the problem can be efficiently solved by semidefinite programming (SDP).**

***Index Terms*—Convex optimization, distributed estimation, energy efficiency, linear source-channel coding, multiple access channel (MAC).**

Manuscript received December 10, 2006; revised June 4, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aleksandar Dogandzic. The work of J.-J. Xiao and Z.-Q. Luo was supported in part by the National Science Foundation by Grant DMS-0610037, and by the USDOD ARMY by Grant W911NF-05-1-0567. The work of S. Cui was supported in part by the University of Arizona Foundation. The work of A. J. Goldsmith was supported in part by the National Science Foundation by Grant CCR-0325639-001. This work was presented in part at the IEEE Global Conference on Communications, San Francisco, CA, November 2006.

J.-J. Xiao and Z.-Q. Luo are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 (e-mail: xiao@ece.umn.edu; luozq@ece.umn.edu).

S. Cui is with the Wireless Systems and Collaborative Communication Lab, Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: cui@ece.arizona.edu).

A. J. Goldsmith is with the Wireless System Lab, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.906762

I. Introduction

CONSIDER a distributed sensor network where observation data is collected at different sensors and transmitted, possibly after compression and encoding, to a fusion center (see Fig. 1). The fusion center (e.g., an unmanned aerial vehicle) aggregates the data for a specific signal processing task, such as target detection or parameter estimation. Local data compression is effectively achieved when each sensor sends to the fusion center only a short summary of its data. Upon receiving the sensor messages, the fusion center combines them according to a fusion rule to generate the final estimate. In this framework, the traditional centralized solution corresponds to the case where all raw data is transmitted to the fusion center without data compression or channel distortion. However, if communication is costly, as is the case in some wireless sensor networks, there can be a significant power-saving advantage if less information is transmitted to the fusion center without degrading the overall performance. We may thus pose the following question: Given a fixed bandwidth and power budget, how should we encode the local messages, transmit the signal, and define the final fusion rule in order to maximize the overall system performance? In this paper, we devote to studying the optimal linear encoding of local observations under both power and bandwidth constraints.

There are at least two ways to model the finite bandwidth constraint. The first one is to directly limit the number of binary bits that each sensor can send to the fusion center per observation period. This bandwidth measure is natural from the digital communication point of view and was adopted widely in previous studies on communication complexity [1], [2], distributed optimization [3], distributed control [4], [5], and distributed signal processing including detection [6], [7], estimation [8]–[11], and tracking [12]. The second one is to limit the number of real-valued messages that can be sent from each sensor to the fusion center per observation period, which is directly proportional to the physical frequency bandwidth in the system. The two measures of bandwidth constraint are fundamentally related to each other, while the second one is more suited for analog transmission schemes such as amplify and forward [13], [14]. In this paper, we adopt the second bandwidth measure and consider the analog transmission of the real-valued sensor messages.

Another important property of many wireless sensor networks is their stringent energy constraint. In such networks, sensors have only small-size batteries whose replacement can be costly if not impossible. Thus, sensor network operations must be energy efficient in order to maximize network lifetime. Recently, many new results have appeared in the sensor network literature with a focus on energy-efficient distributed data fusion. From an information-theoretic perspective, [15], [17]–[20] investigate the mean squared estimation error performance versus transmit power for the quadratic CEO problem with a coherent multiple access channel (MAC). Notably, it is shown in [15] and [16] that if the sensor statistics are Gaussian, a simple uncoded (analog-and-forwarding) scheme

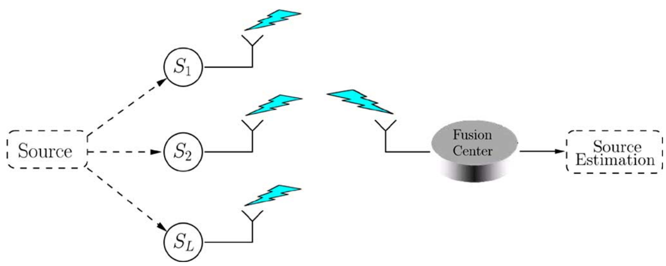


Fig. 1. Model of a wireless sensor network with a fusion center.

dramatically outperforms the separate source-channel coding approach and leads to an optimal scaling behavior. The uncoded communication scheme is further proved to preserve the optimal scaling law in [19] for sensor networks with node statistics satisfying a certain mean condition, while the source-channel matching result is extended to more general homogeneous signal fields in [20]. If the sensor measurements are not continuous but in a finite alphabet, type-based transmission schemes are proposed in [21]–[23]. The many-to-one transport capacity and compressibility are investigated for dense joint estimation sensor networks in [24]. When a full coordination among sensors is unavailable and the underlying communication links are not reliable, the distributed estimation problem is investigated in [25], where an information-theoretic achievable rate-distortion region is elegantly derived. The work in [26] studies the in-network processing approaches based on a hierarchical data-handling and communication architecture for the estimation of field sources. In addition, by assuming only local sensor information exchange, [27] proposes a distributed algorithm for reaching network-wide consensus.

In this paper, we consider the joint estimation of a vector source by a sensor network with a fusion center. Both bandwidth and power constraints are imposed on the transmitted signals from local sensors. We limit our discussion to the class of linear decentralized estimation schemes where the compression functions at local sensors and the fusion function at the fusion center are all linear. As a result of the bandwidth constraint, each sensor transmits to the fusion center a fixed number of real-valued messages per observation. The power constraint limits the strength of the transmitted signals. Under a mean squared error (MSE) criterion, we design the optimal linear decentralized estimation schemes based on the channel states and the second-order statistics of the source/observation.

For the case of orthogonal channel usage between sensors and the fusion center, the linear decentralized estimation problem has been first studied in [10] with a general sensor observation model. Neglecting the communication channel noise, the authors of [10] proposed a Gauss-Seidel type iterative algorithm to compute the linear message functions (whose convergence to the global optimum is not guaranteed). The complexity of linear decentralized estimation design has been recently studied in [28]. It has been shown therein that surprisingly, in the orthogonal channel case, even for the linear sensor observation model, the complexity of designing the optimal linear decentralized estimation is NP-hard in general. Because of this, efficient Gauss-Seidel algorithms have been proposed in [29] to determine the suboptimal designs by exploring the classic tools of canonical component analysis.

We investigate the optimal design of linear decentralized estimation without assuming orthogonal channel usage. As a result, messages transmitted from different sensors may cause interference to each other. We assume that there is perfect synchronization between sensors and the fusion center so that the transmitted messages from local sensors can be coherently combined at the fusion center. With such an assumption, one key design consideration at local sensors and the fusion center is how to jointly process the sensed and received information in a constructive way. In this paper, we first give a general problem formulation of linear coherent decentralized estimation, and then solve it for two cases: i) Source/observations are scalars; and ii) source/observations are vectors. For the case of scalar source/observations, we focus on the power allocation among sensors. We give the optimal power scheduling via convex optimization techniques and derive an elegant relationship between the best achievable distortion and the total sensor transmit power. When source/observations are vectors, we consider the optimal design of encoding/decoding matrices in two scenarios: i) By neglecting the channel noise, we show that the resulting problem has a closed-form solution; and ii) When channel noise is present, the optimal design of linear decentralized estimation subject to transmit power and bandwidth constraints can be efficiently solved with semidefinite programming (SDP) combined with certain relaxation techniques. To solve the optimal encoding matrices, we assume that the fusion center has full knowledge of the sensor observation model and the channel states between sensors and the fusion center. Each sensor can then obtain their individual encoding functions based on their local information and limited feedback from the fusion center. We also study the asymptotic MSE performance when the number of sensors is large. Under a finite total power budget, we show that the best achievable MSE is bounded away from zero for orthogonal MAC, but scales with $1/L$ when the MAC is coherent, where L is the total number of sensors.

Organization of the Paper: The rest of the paper is organized as follows. In Section II, we give the general problem formulation of linear decentralized estimation. The corresponding MSE performance and power consumption are derived in terms of the local sensor encoding matrices. In Section III, we study the case when both source and observations are scalars, and give the corresponding optimal power scheduling. We also give the asymptotic analysis of the MSE performance with both coherent and orthogonal MACs. Sections IV and V discuss the case when both source and observations are vectors. In Section IV, we study the optimal linear decentralized estimation under bandwidth constraints without considering channel noises. A close-form solution is obtained. In Section V, we consider the noisy channel case, and study the general linear decentralized estimation under both power and bandwidth constraints. Section VI gives the summary and concludes the paper.

Notations: Throughout this paper we adopt the following notations. A lower case letter denotes a scalar, a boldface/lowercase letter denotes a vector, and a boldface/uppercase letter denotes a matrix. For a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\lambda_i(\mathbf{A})$ denotes its ith eigenvalue assuming $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \ldots \geq \lambda_n(\mathbf{A})$. In addition, $\mathrm{Tr}(\mathbf{A})$, $\mathbf{A}^T$ and $\mathbf{A}^\dagger$ denote the trace, transpose and pseudoinverse of $\mathbf{A}$ respectively. The letter $\mathbf{I}_n$ denotes an identity matrix of size $n \times n$. For two matrices $\mathbf{B}$ and $\mathbf{C}$, the relation $\mathbf{B} \succeq \mathbf{C}$ means that $\mathbf{B} - \mathbf{C}$ is positive semi-definite.

II. Problem Formulation

Suppose there are L sensors, each making observations on a common unknown random vector signal $\mathbf{s} = [s_1, s_2, \ldots, s_p]^T$. We assume that $\mathbf{s}$ has zero mean and covariance matrix $\mathbf{C}_\mathbf{s}$.

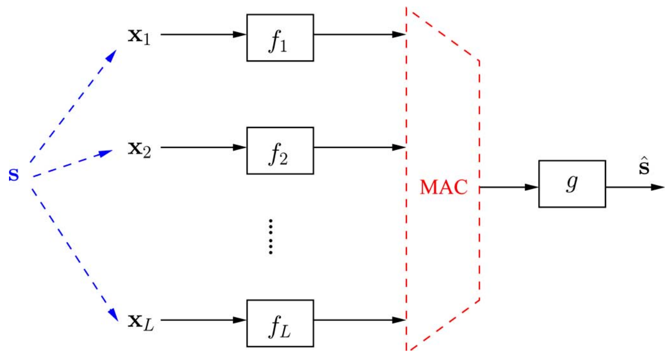


Fig. 2. Decentralized estimation scheme in a wireless sensor network with a fusion center.

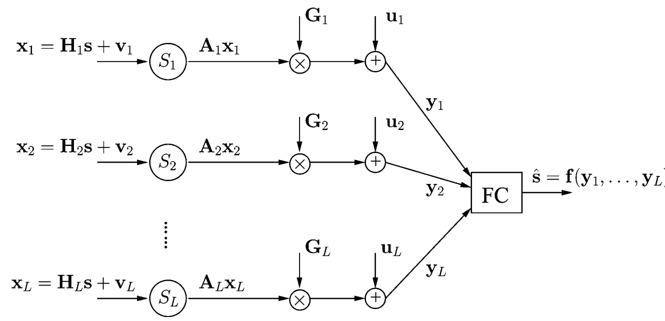


Fig. 3. Linear decentralized estimation with orthogonal MAC.

Sensor observations $\mathbf{x}_i$ in general have a conditional distribution based on $\mathbf{s}$: $P(\mathbf{x}_i|\mathbf{s})$.

The L sensors are coordinated by a fusion center to jointly estimate the random signal $\mathbf{s}$. For most wireless sensor networks, sensors only have limited battery power and limited communication capability. Therefore, local data compression may be needed at each sensor to reduce the communication requirement between sensors and the fusion center. We assume that sensors are distributed and there is no inter-sensor communication. Each sensor compresses its observations $\mathbf{x}_i$ to $\mathbf{m}_i$ by a mapping $f_i : \mathbf{x}_i \to \mathbf{m}_i$. The message $\mathbf{m}_i(\mathbf{x}_i)$ are transmitted to the fusion center through a multiple access channel (MAC). The role of the fusion center is to generate the final estimate of $\mathbf{s}$ based on the received messages, where the estimate is denoted by another mapping g (see Fig. 2).

Throughout this work, we shall only focus on a linear decentralized estimation where both observation models and encoding/decoding mappings are linear. The linear coding strategy is motivated by the fact that when a memoryless Gaussian source is transmitted through an AWGN channel, linear amplify-and-forward coding achieves the best power-distortion tradeoff. The optimality of linear coding has been extended to the estimation of a Gaussian source over an AWGN MAC [15], [16]. Another reason we limit ourselves to linear sensor observation models and linear coding is for the ease of analysis. With such assumptions, we can avoid unnecessary technical complications but still be able capture the fundamental power-distortion tradeoff in distributed estimation.

Specifically, we assume the sensor observations $\mathbf{x}_i \in \mathbb{R}^{\ell_i \times 1}$ are the linear combination of $\mathbf{s}$ corrupted by additive noises and can be described as

$$\mathbf{x}_i = \mathbf{H}_i\mathbf{s} + \mathbf{v}_i, \quad 1 \le i \le L \tag{1}$$

where $\mathbf{H}_i \in \mathbb{R}^{\ell_i \times p}$ are observation matrices; Noise $\mathbf{v}_i \in \mathbb{R}^{\ell_i \times 1}$ are spatially uncorrelated among different sensors, and each $\mathbf{v}_i$ has zero mean and covariance matrix $\mathbf{C}_{\mathbf{v}_i}$. Without loss of generality, we can assume that $\mathbf{C}_\mathbf{s} = \mathbf{I}_p$ and $\mathbf{C}_{\mathbf{v}_i} = \mathbf{I}_{\ell_i}$. Otherwise, we can introduce $\mathbf{s}^{(1)} = \mathbf{C}_\mathbf{s}^{-(1/2)}\mathbf{s}$, $\mathbf{v}_i^{(1)} = \mathbf{C}_{\mathbf{v}_i}^{-(1/2)}\mathbf{v}_i$, $\mathbf{x}_i^{(1)} = \mathbf{C}_{\mathbf{v}_i}^{-(1/2)}\mathbf{x}_i$, and $\mathbf{H}_i^{(1)} = \mathbf{C}_{\mathbf{v}_i}^{-(1/2)}\mathbf{H}_i\mathbf{C}_\mathbf{s}^{1/2}$. Then we obtain an equivalent model $\mathbf{x}_i^{(1)} = \mathbf{H}_i^{(1)}\mathbf{s}^{(1)} + \mathbf{v}_i^{(1)}$ in which $\mathbf{C}_{\mathbf{s}^{(1)}} = \mathbf{I}_p$, $\mathbf{C}_{\mathbf{v}_i^{(1)}} = \mathbf{I}_{\ell_i}$.

The linear design of the estimation scheme in which the message functions f_i and the fusion function g are linear is motivated by the results derived in [14], [15], where it has been proved that in the so-called Gaussian sensor network (in which source, observations, and MAC noises are all Gaussian), linear source-channel coding strategies can significantly outperform the traditional separate source-channel coding.

Depending on the different multiple access schemes, we investigate two cases for the MAC between sensors and fusion center: orthogonal and coherent. For the case of orthogonal MAC, we assume that sensors have their independent noninterfering channels to the fusion center. This can be realized, e.g., by orthogonal time/frequency/code division multiple access (TDMA/FDMA/CDMA). As for the coherent MAC, we allow all sensors transmit simultaneously by using, e.g., the same time slot or frequency band. Assuming perfect synchronization between sensors and the fusion center, we can ensure that transmitted signals from all sensors can reach the fusion center as a coherent sum.

In the linear mapping of vector observations, one immediate question is about how many real messages to which each observation $\mathbf{x}_i$ shall be compressed. This is determined by the degrees of freedom of the channel from sensor i to the fusion center. Assume that for each observation time snapshot, sensor i can transmit q_i real messages to the fusion center, which is potentially decided by the channel bandwidth. With such an assumption, the message functions can be presented as

$$\mathbf{m}_i(\mathbf{x}_i) = \mathbf{A}_i\mathbf{x}_i, \quad \text{where} \quad \mathbf{A}_i \in \mathbb{R}^{q_i \times \ell_i}.$$

Based on the received messages, the fusion center then generates an estimate $\hat{\mathbf{s}}$ to minimize the MSE $D \stackrel{\text{def}}{=} \mathrm{Tr}(\mathbf{D})$, where $\mathbf{D} \stackrel{\text{def}}{=} E[(\mathbf{s}-\hat{\mathbf{s}})(\mathbf{s}-\hat{\mathbf{s}})^T]$. We describe the remaining part of the linear decentralized estimation for two MAC cases as follows.

- *Orthogonal MAC*. In this case, the L sensors transmit their observations to the fusion center via L orthogonal channels (see Fig. 3). For channel i, the received signal can be written as

$$\begin{aligned}\mathbf{y}_i &= \mathbf{G}_i\mathbf{m}_i + \mathbf{u}_i = \mathbf{G}_i\mathbf{A}_i\mathbf{x}_i + \mathbf{u}_i \\ &= \mathbf{G}_i\mathbf{A}_i\mathbf{H}_i\mathbf{s} + \mathbf{G}_i\mathbf{A}_i\mathbf{v}_i + \mathbf{u}_i \\ &\stackrel{\text{def}}{=} \mathbf{B}_i\mathbf{H}_i\mathbf{s} + \mathbf{B}_i\mathbf{v}_i + \mathbf{u}_i\end{aligned} \tag{2}$$

  where $\mathbf{G}_i \in \mathbb{R}^{q_i \times q_i}$ are the channel matrix from sensor i to the fusion center, $\mathbf{B}_i \stackrel{\text{def}}{=} \mathbf{G}_i\mathbf{A}_i$, and $\mathbf{u}_i \in \mathbb{R}^{q_i \times 1}$ is the additive channel noise with covariance matrix $\mathbf{C}_{\mathbf{u}_i}$. Without loss of generality, we can assume $\mathbf{C}_{\mathbf{u}_i} = \mathbf{I}_{q_i}$. Otherwise we can absorb $\mathbf{C}_{\mathbf{u}_i}^{-(1/2)}$ into the channel gain matrix $\mathbf{G}_i$ to obtain an identity channel noise covariance matrix, which is the so-called noise whitening.

It is easy to see that the linear MMSE estimator of $\mathbf{s}$ based on $\{\mathbf{y}_i : 1 \le i \le L\}$ takes the form

$$\hat{\mathbf{s}} = \left\{ \mathbf{I}_p + \sum_{i=1}^{L} \mathbf{H}_i^T \mathbf{B}_i^T \left(\mathbf{B}_i \mathbf{B}_i^T + \mathbf{I}_{q_i} \right)^{-1} \mathbf{B}_i \mathbf{H}_i \right\}^{-1} \times \sum_{i=1}^{L} \mathbf{H}_i^T \mathbf{B}_i^T \left(\mathbf{B}_i \mathbf{B}_i^T + \mathbf{I}_{q_i} \right)^{-1} \mathbf{y}_i$$

and has an MSE matrix $\mathbf{D}$ satisfying (see, e.g., [30, Theorem 12.1])

$$\mathbf{D}^{-1} = \mathbf{I}_p + \sum_{i=1}^{L} \mathbf{H}_i^T \mathbf{B}_i^T \left(\mathbf{B}_i \mathbf{B}_i^T + \mathbf{I}_q \right)^{-1} \mathbf{B}_i \mathbf{H}_i. \quad (3)$$

We note that for the orthogonal MAC case, the bandwidth constraint q_i can be different over sensors.

- *Coherent MAC*. Another case is that all L sensors transmit simultaneously. The transmitted signals from all sensors reach the fusion center as a coherent sum under the assumption of perfect synchronization between sensors and the fusion center.[1] In this case, we assume that each sensor transmits the same number of real messages, i.e., $q_i = q$ for all i. The received signal $\mathbf{y}$ at the fusion center can be expressed as (see Fig. 4)

$$\mathbf{y} = \sum_{i=1}^{L} \mathbf{G}_i \mathbf{m}_i + \mathbf{u} = \sum_{i=1}^{L} \mathbf{G}_i \mathbf{A}_i \mathbf{x}_i + \mathbf{u} \quad (4)$$

where $\mathbf{G}_i \in \mathbb{R}^{q \times q}$ are the channel matrix from sensor i to the fusion center, and $\mathbf{u} \in \mathbb{R}^{q \times 1}$ is the additive channel noise with covariance matrix $\mathbf{C_u}$. Again, without loss of generality, we can assume $\mathbf{C_u} = \mathbf{I}_q$.

Notice that $\mathbf{B}_i \stackrel{\text{def}}{=} \mathbf{G}_i \mathbf{A}_i$, and we further introduce vector notations

$$\mathbf{H} = \left[\mathbf{H}_1^T, \mathbf{H}_2^T, \ldots, \mathbf{H}_L^T \right]^T, \quad \mathbf{H} \in \mathbb{R}^{\ell \times p}$$
$$\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_L], \quad \mathbf{B} \in \mathbb{R}^{q \times \ell}$$
$$\mathbf{v} = \left[\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_L^T \right]^T, \quad \mathbf{v} \in \mathbb{R}^{\ell \times 1}$$

with $\ell \stackrel{\text{def}}{=} \ell_1 + \ell_2 + \cdots + \ell_L$. As such, we have

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^{L} \mathbf{G}_i \mathbf{A}_i \mathbf{H}_i \mathbf{s} + \sum_{i=1}^{L} \mathbf{G}_i \mathbf{A}_i \mathbf{v}_i + \mathbf{u} \\ &= \sum_{i=1}^{L} \mathbf{B}_i \mathbf{H}_i \mathbf{s} + \sum_{i=1}^{L} \mathbf{B}_i \mathbf{v}_i + \mathbf{u} \\ &= \mathbf{B}\mathbf{H}\mathbf{s} + \mathbf{B}\mathbf{v} + \mathbf{u}. \end{aligned} \quad (5)$$

Given $\mathbf{y}$, the fusion center then generates an estimate $\hat{\mathbf{s}}$. Specifically, the linear MMSE estimator in terms of $\mathbf{H}$ and $\mathbf{B}$ is $\hat{\mathbf{s}} = \left[\mathbf{I}_p + \mathbf{H}^T \mathbf{B}^T (\mathbf{B}\mathbf{B}^T + \mathbf{I}_q)^{-1} \mathbf{B}\mathbf{H} \right]^{-1} \mathbf{H}^T \mathbf{B}^T (\mathbf{B}\mathbf{B}^T + \mathbf{I}_q)^{-1} \mathbf{y}$, which is linear and achieves an MSE matrix $\mathbf{D}$ given as

$$\mathbf{D}^{-1} = \mathbf{I}_p + \mathbf{H}^T \mathbf{B}^T (\mathbf{B}\mathbf{B}^T + \mathbf{I}_q)^{-1} \mathbf{B}\mathbf{H}. \quad (6)$$

[1]Note that for orthogonal MAC, we only need to assume pair-wise synchronization between each sensor and the fusion center, and synchronization among sensor nodes is not required.

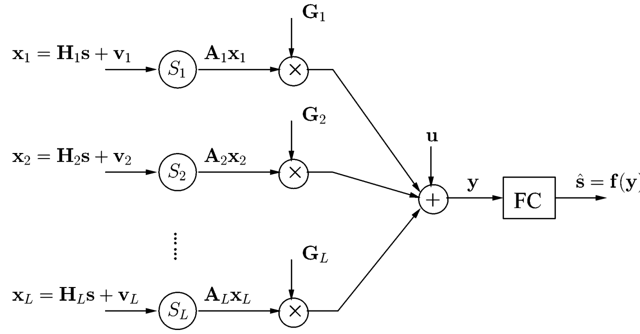


Fig. 4. Linear decentralized estimation with coherent MAC.

Remark 1: For both orthogonal and coherent MAC, we have assumed their noise covariance matrices to be unitary. Such an assumption is based on the following system setup. Suppose for the case of orthogonal MAC, the frequency bandwidth of the channel from each sensor to the fusion center is B_s, then for the coherent MAC, all sensors use the same frequency band (also with a frequency bandwidth B_s) to transmit their observations simultaneously. The sampling rate at the channel output is the same for both MAC, which is the source symbol rate. Thus, the noise power in the coherent MAC is the same as that in each subchannel of the orthogonal MAC.

We in addition assume in the rest of the paper that $\ell \ge p$ and rank$(\mathbf{H}) = p$. This corresponds to the case when every component of the source $\mathbf{s}$ is observed by at least one sensor.

Remark 2: Note that when all sensor observations $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_L^T]^T$ are directly available to the fusion center, a centralized estimator $\hat{\mathbf{s}}_0 = (\mathbf{I}_p + \mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ achieves an MSE

$$D_0 = \mathrm{Tr}\left[(\mathbf{I}_p + \mathbf{H}^T \mathbf{H})^{\dagger} \right] \quad (7)$$

which is the performance benchmark for all decentralized estimators with nonideal channel models.

The bandwidth constraints lead to a dimensionality condition on $\mathbf{A}_i$, i.e., $\mathbf{A}_i \in \mathbb{R}^{q_i \times \ell_i}$. Suppose in addition, the transmit power constraint at sensor i is P_i. We have the following constraint on $\mathbf{A}_i$:

$$\begin{aligned} \mathrm{Tr}\left\{ E\left(\mathbf{m}_i \mathbf{m}_i^T \right) \right\} &= \mathrm{Tr}\left\{ E\left[\mathbf{A}_i \mathbf{x}_i \mathbf{x}_i^T \mathbf{A}_i^T \right] \right\} \\ &= \mathrm{Tr}\left\{ E\left[\mathbf{A}_i (\mathbf{H}_i \mathbf{s} + \mathbf{v}_i)(\mathbf{H}_i \mathbf{s} + \mathbf{v}_i)^T \mathbf{A}_i^T \right] \right\} \\ &= \mathrm{Tr}\left[\mathbf{A}_i \left(\mathbf{H}_i \mathbf{H}_i^T + \mathbf{I}_{\ell_i} \right) \mathbf{A}_i^T \right] \le P_i. \end{aligned} \quad (8)$$

Thus, to design the optimal linear decentralized estimation scheme, we shall solve the optimal encoding matrices $\mathbf{A}_i : 1 \le i \le L$ subject to power and bandwidth constraints such that $D = \mathrm{Tr}(\mathbf{D})$ is minimized. This leads to the following optimization problem:

$$\begin{aligned} \min_{\mathbf{A}_i : 1 \le i \le L} \quad & \mathrm{Tr}(\mathbf{D}) \\ \text{s.t.} \quad & \mathbf{D} \text{ satisfies (3) or (6)} \\ & \mathbf{A}_i \in \mathbb{R}^{q_i \times \ell_i}, \quad 1 \le i \le L \\ & \mathrm{Tr}\left[\mathbf{A}_i \left(\mathbf{H}_i \mathbf{H}_i^T + \mathbf{I}_{\ell_i} \right) \mathbf{A}_i^T \right] \le P_i \\ & 1 \le i \le L. \end{aligned} \quad (9)$$

TABLE I
SUMMARY OF PARAMETERS

$\mathbf{s}$	The source vector to be estimated
p	Dimension of the source vector $\mathbf{s}$
L	Total number of sensors
$\mathbf{x}_i$	Observation of sensor i
ℓ_i	Dimension of $\mathbf{x}_i$
ℓ	Total number of observations from all sensors, *i.e.*, $\ell = \ell_1 + \ell_2 + \ldots + \ell_L$
q_i	Degrees of freedom of the channel from sensor i to the fusion center in the case of orthogonal multiple access
q	Degrees of freedom of the coherent multiple access channel

To solve the optimization problem (9), we assume that the fusion center has the knowledge of sensor observation models and channel states between sensors to the fusion center: $\{(\mathbf{H}_i, \mathbf{G}_i) : i = 1, 2, \ldots, L)\}$. This assumption is reasonable in cases where the network condition and the signal observation model change slowly in a quasi-static manner. Thus, once $\{(\mathbf{H}_i, \mathbf{G}_i) : i = 1, 2, \ldots, L)\}$ are acquired by the fusion center, they can be used for a reasonably long period of time.

When the MAC is orthogonal, it has been shown in [28, Theorem 1] that the complexity of designing the optimal linear encoding matrices $\mathbf{A}_i$ is NP-hard (in L) when $\ell_i \geq 2 : 1 \leq i \leq L$, even for the special case of $q_i = 1$ and $\mathbf{u}_i = 0$ for all i (i.e., channels are noiseless). In this paper, we will focus on the linear decentralized estimation under coherent MAC. The main work can be outlined as follows:

- *Scalar case*. We first consider a special case where source and observations are all scalars. As such, the design of linear encoding matrices is reduced to the selection of linear scaling factors. We discuss the optimal power allocation among sensors and derive the optimal tradeoff between MSE and the total transmit power. Interestingly, we show that turning off sensors is not necessary in the coherent MAC, which is in sharp contrast to the orthogonal MAC case, where turning off sensors is proved to conserve energy [31], [32]. In addition, we show that in term of L, the MSEs with the two multiple access schemes have significantly different asymptotic behaviors.
- *Vector case*. In this case, we aim for designing the optimal coding matrices $\mathbf{A}_i$s to minimize the MSE subject to both power and bandwidth constraints. We first ignore the channel noise $\mathbf{u}$. The purpose of this study is to illustrate how the bandwidth constraint q affects the achieved MSE. Surprisingly, unlike the orthogonal channel case, the optimal coding matrices for the coherent MAC have analytical forms. Secondly we consider the noisy MAC channel and impose power constraints on transmitted messages $\mathbf{m}_i$s. We formulate the optimization of $\mathbf{A}_i$s as an SDP problem, and illustrate by numerical results how the bandwidth and power constraints affect the achieved MSE.

For convenience, we summarize the parameters defined in this section in Table I.

III. SCALAR CASE

In this section, we analyze the linear decentralized estimation when source and observations are scalars. We first give the optimal power allocation for the case of coherent MAC, and then compare its performance to the case of orthogonal MAC.

When source and observations are scalars (i.e., $p = 1$ and $\ell_i = 1$ for all i), the observation model in (1) is reduced to

$$x_i = h_i s + v_i, \quad 1 \leq i \leq L \tag{10}$$

where both s and v_i are assumed to have zero mean and unitary variance, but otherwise unknown. The noise v_i is also assumed to be spatially independent over i. Following (7), the benchmark MSE D_0 of estimating s based on x_is satisfies

$$\frac{1}{D_0} = 1 + \sum_{i=1}^{L} h_i^2. \tag{11}$$

Suppose the corresponding analog forwarding encoder is given by $m_i(x_i) = a_i x_i$, where a_i is the scaling factor to be designed. The resulting average transmit power of sensor i is

$$P_i = E\left(m_i^2\right) = a_i^2\left(h_i^2 + 1\right). \tag{12}$$

After the amplification, m_is are transmitted to the fusion center with a coherent MAC. In light of (4), the received signal at the fusion center is

$$\begin{aligned} y &= \sum_{i=1}^{L} g_i m_i + u = \sum_{i=1}^{L} g_i a_i x_i + u \\ &= \sum_{i=1}^{L} g_i a_i h_i s + \sum_{i=1}^{L} g_i a_i v_i + u \end{aligned} \tag{13}$$

where g_i is the channel gain of sensor i to the fusion center, and u is the channel noise. Similarly, u is assumed to have zero mean and unitary variance. The linear MMSE estimator of s from y is $\hat{s} = (E(sy)/E(y^2))y$, which has distortion D satisfying

$$\frac{1}{D} = 1 + \left(1 + \sum_{i=1}^{L} g_i^2 a_i^2\right)^{-1} \left(\sum_{i=1}^{L} g_i h_i a_i\right)^2. \tag{14}$$

A. Uniform Power Scheduling

Suppose all sensors use the same transmit power, that is, $P_i = P/L$ where P is the total transmit power. Therefore, from (12) we get $a_i = \sqrt{P/(L(1+h_i^2))}$. Let $D_u(P)$ denote the achieved distortion with uniform power scheduling. From (14), $D_u(P)$ satisfies

$$\frac{1}{D_u(P)} = 1 + \left(\frac{L}{P} + \sum_{i=1}^{L} \frac{g_i^2}{1+h_i^2}\right)^{-1} \left(\sum_{i=1}^{L} h_i \sqrt{\frac{g_i^2}{1+h_i^2}}\right)^2.$$

We see that when $P \to \infty$, $D_u(P)$ monotonically decreases with P and

$$\begin{aligned} \frac{1}{D_u(\infty)} &= 1 + \left(\sum_{i=1}^{L} \frac{g_i^2}{1+h_i^2}\right)^{-1} \left(\sum_{i=1}^{L} h_i \sqrt{\frac{g_i^2}{1+h_i^2}}\right)^2 \\ &\leq 1 + \sum_{i=1}^{L} h_i^2 = \frac{1}{D_0} \end{aligned}$$

where in the last inequality, we applied the Cauchy-Schwartz inequality, and the equal sign holds if and only if there exists $\beta > 0$ such that

$$\frac{g_i^2}{(1+h_i^2)\,h_i^2} = \beta, \quad \text{for all } 1 \le i \le L. \tag{15}$$

We summarize the above result in the following theorem.

Theorem 3: Let $D_u(P)$ denote the MSE achieved by analog forwarding where each of the L sensors uses exactly the same transmit power of P/L. Then, for every finite P

$$D_u(P) > D_u(\infty) \ge D_0.$$

Moreover, even with infinite transmit power at every sensor, $D_u(\infty)$ is always strictly larger than D_0 unless condition (15) holds.

B. Optimal Power Scheduling

Now we consider an optimal power allocation strategy whereby transmit power is optimally scheduled among sensors to achieve the best estimation performance. Let $D(P_1, P_2, \ldots, P_L)$ denote the distortion achieved by assigning P_i to sensor i. We study the following MMSE estimation problem under a total power constraint:

$$\begin{aligned} \min_{P_i:1\le i\le L} \quad & D(P_1, P_2, \ldots, P_L) \\ \text{s.t.} \quad & \sum_{i=1}^{L} P_i \le P. \end{aligned} \tag{16}$$

We can write the above problem in terms of the coding factors a_i, and obtain the following problem:

$$\begin{aligned} \max_{a_i:1\le i\le L} \quad & 1 + \left(1 + \sum_{i=1}^{L} g_i^2 a_i^2\right)^{-1} \left(\sum_{i=1}^{L} g_i h_i a_i\right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^{L} \left(h_i^2 + 1\right) a_i^2 \le P. \end{aligned} \tag{17}$$

It is easy to see that the above problem is not convex in $\{a_i : 1 \le i \le L\}$. However, we can transform it into an equivalent convex form that is efficiently solvable, for which the detail is given in the Appendix. We summarize the obtained result in the following theorem.

Theorem 4: For the linear decentralized estimation of s based on the observation model in (10) and the MAC model in (13), assuming a total transmit power constraint P imposed on all sensors, the best achievable distortion satisfies

$$\frac{1}{D} = 1 + \sum_{i=1}^{L} \frac{h_i^2}{1 + (1+h_i^2)/(g_i^2 P)}. \tag{18}$$

The optimal power allocation achieving the above power-distortion tradeoff is

$$P_i^{\text{opt}} = c_i P, \quad i = 1, \cdots, L \tag{19}$$

where

$$c_i = c\frac{g_i^2 h_i^2 \left(h_i^2+1\right)}{\left(h_i^2+1+g_i^2 P\right)^2}, \quad c = \left(\sum_{i=1}^{L} \frac{g_i^2 h_i^2 \left(h_i^2+1\right)}{\left(h_i^2+1+g_i^2 P\right)^2}\right)^{-1}.$$

Remark 5: Formula (19) is intuitively appealing as it indicates that the optimal power scheduling admits a distributed implementation. First, the fusion center broadcast the constant c and P (which are both universal over i) to the local sensors. Then, the local sensors use c, P, and two local parameters g_i^2, h_i^2 to determine their individual transmit power. It is also interesting to note that no sensors are turned off unless for trivial cases where either $h_i = 0$ or $g_i = 0$. This is different from the optimal power scheduling for the orthogonal MAC, where sensors with poor observation quality or small channel gains are turned off to conserve energy [32].

Remark 6 (Asymptotic Analysis): It is interesting to analyze the asymptotic behavior of D when $P \to \infty$ or when $L \to \infty$.

- When P gets larger but L is fixed, we have that D approaches D_0 as

$$\begin{aligned} \frac{1}{D_0} - \frac{1}{D} &= \sum_{i=1}^{L} \frac{h_i^2\left(1+h_i^2\right)/\left(g_i^2 P\right)}{1+(1+h_i^2)/(g_i^2 P)} \\ &\approx \frac{1}{P}\sum_{i=1}^{L} \frac{h_i^2\left(1+h_i^2\right)}{g_i^2}. \end{aligned}$$

  Therefore, $D - D_0 \approx (D_0^2/P)(\sum_{i=1}^{L}(h_i^2(1+h_i^2)/g_i^2))$. This implies that with fixed number of sensors, when $P \to \infty$, D approaches the benchmark MSE D_0 with a gap $D - D_0$ proportional to $1/P$. In this case, D does not go to zero since its performance is limited by the finite number of sensor observations.
- When L is large but P is fixed, assuming $\{g_i^2, h_i^2 : 1 \le i \le L\}$ are i.i.d. over i, we obtain that

$$\begin{aligned} D &= \left(1 + \sum_{i=1}^{L} \frac{h_i^2}{1+(1+h_i^2)/(g_i^2 P)}\right)^{-1} \\ &\approx \frac{1}{L}\left\{E\left[\frac{h_i^2}{1+(1+h_i^2)/(g_i^2 P)}\right]\right\}^{-1}. \end{aligned} \tag{20}$$

  This implies that surprisingly, the MSE of the coherent MAC decreases in the order of $1/L$ even though the total transmit power P is finite. Such an asymptotic behavior however does not hold for the orthogonal MAC (see Section III-C).

Theorem 4 implies that with optimal power scheduling over a finite number of sensors, the analog forwarding achieves an MSE D converging to D_0 as the total transmit power $P \to \infty$. This is in contrast to Theorem 3 that reveals that the uniform power scheduling typically achieves an MSE that is strictly larger than D_0, even when $P \to \infty$ (unless condition (15) holds). In most inhomogeneous sensing environment, the margin between $D_u(\infty)$ and D_0 is significant. In Fig. 5, we plot the curves of MSE D versus the total transmit power P under both uniform and optimal power schedules where the total number of sensors $L = 10$. Note that the power P is taken relative to the channel noise power. Since we assume that the channel noise has unitary variance, we thus label the transmit

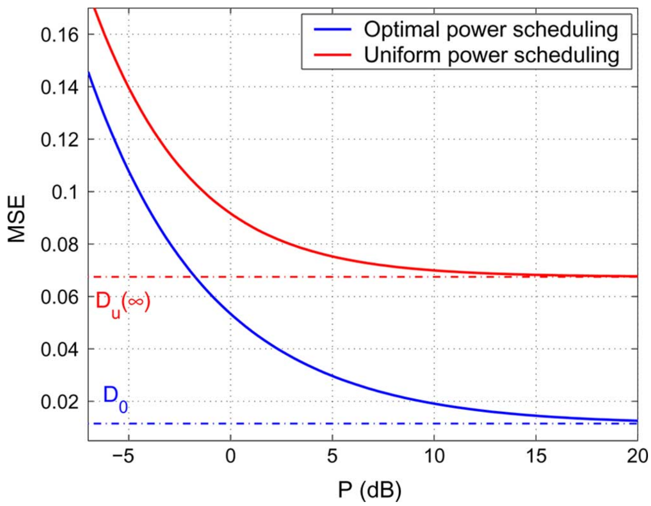


Fig. 5. Uniform power scheduling versus optimal power scheduling when $P \to \infty$. Note that the power P is taken relative to the channel noise power. Since we assume that the channel noise has unitary variance, we thus label the transmit power in the unit of dB.

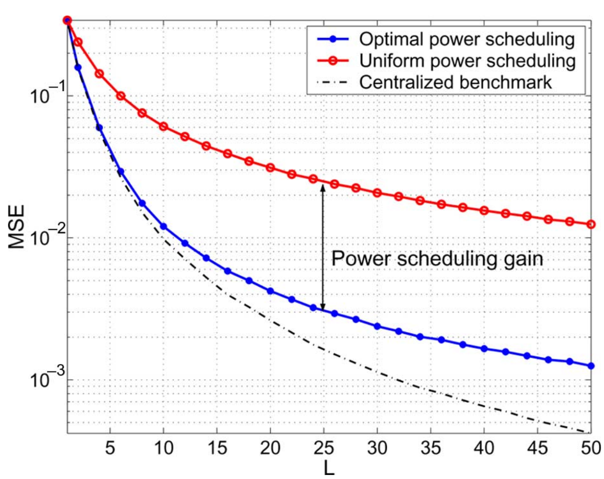


Fig. 6. Uniform power scheduling versus optimal power scheduling when $L \to \infty$.

power in the unit of dB. In the simulation, sensor observation noise variances are taken from a Chi-squared distribution with degree 1, and channel gains g_i are taken as $c_g \cdot d^{-3.5}$ where d is uniformly taken from the real interval [1, 10], and c_g is a normalization constant to make $E(g_i) = 1$. In the figure, the simulated MSE is averaged over 1000 realizations of $\{\sigma_i^2, g_i : i = 1, 2, \ldots, L\}$, and is actually the expected MSE. We can see that when P increases, D values converge to two different limits that are D_0 and $D_u(\infty)$, respectively. As can be seen, there is a large gap between $D_u(\infty)$ and D_0. Fig. 6 plots the curves comparing the achieved MSE when L increases with the total power being constant: $P = 16$ dB. A significant performance gain from the optimal power scheduling can be observed from the gap between the two MSE curves. In addition, the performance gain becomes more significant as the total number of sensors L increases.

C. Compared to Orthogonal MAC

In this section, we compare the MSE performance between orthogonal MAC and coherent MAC. When multiple access between sensors and the fusion center is orthogonal and signals are scalars, in light of (2), the received signals at the fusion center can be represented as

$$y_i = g_i m_i + u_i = g_i a_i h_i s + g_i a_i v_i + u_i, \quad 1 \le i \le L$$

where u_i is the channel noise with zero mean and unitary variance.

From $\{y_i : 1 \le i \le L\}$, we can estimate s by a linear MMSE estimator and achieve an MSE D satisfying [3]

$$\frac{1}{D} = 1 + \sum_{i=1}^{L} \frac{g_i^2 a_i^2 h_i^2}{1 + g_i^2 a_i^2} = 1 + \sum_{i=1}^{L} \frac{h_i^2}{1 + (1 + h_i^2) / (g_i^2 P_i)} \tag{21}$$

where in the second equality we used the fact that $P_i = a_i^2(1 + h_i^2)$ [see (12)]. It is interesting to see that (18) and (21) are almost identical except that in each term of the right-hand side (RHS) sum, one is P and the other is P_i. This reveals that the coherent MAC with optimal power scheduling has the same MSE performance with the orthogonal MAC in which each sensor uses the total transmit power P. Such a fact leads to a significant difference for the asymptotic performance (in terms of L) between these two access schemes. We shall also comment that for the case of orthogonal MAC, although the designing the optimal linear decentralized estimation is NP-hard when observations are vectors [28], it can be reformulated into an equivalent convex form for the scalar case, and the solution is given in [32].

Let P be fixed but $L \to \infty$. For the orthogonal MAC with uniform power allocation $P_i = P/L$, under the assumption that $\{g_i^2, h_i^2 : 1 \le i \le L\}$ are i.i.d. over i, it follows from (21) that

$$\begin{aligned} \frac{1}{D} &= 1 + \sum_{i=1}^{L} \frac{h_i^2}{1 + L(1 + h_i^2) / (g_i^2 P)} \\ &= 1 + \frac{1}{L} \sum_{i=1}^{L} \frac{h_i^2}{1/L + (1 + h_i^2) / (g_i^2 P)} \\ &\to 1 + E\left[\frac{g_i^2 h_i^2 P}{1 + h_i^2}\right]. \end{aligned}$$

This implies that for orthogonal MAC, the achievable MSE is finite even though L is allowed to go to infinity. This is in sharp contrast to the asymptotic performance of the coherent MAC given in (20), which improves as $1/L$. The optimal power scheduling of the orthogonal MAC for the case of scalar source and observations is given in [32]. With optimal power scheduling, there is performance gain but the asymptotic performance still holds.

We see that for orthogonal MAC with a finite amount of total transmit power P, the overall MSE D does not decrease to zero even if L approaches infinity. This is a consequence of using orthogonal links from the sensors to the fusion center, which leads to L different channel noises (i.e., $u_i : 1 \le i \le L$) such that the corruption of channel noise cannot be eliminated even when L goes to infinity. However, in the coherent MAC, only one channel noise (i.e., u) is generated per transmission. Thus, as a result of the coherent combination, the signal to noise ratio at the received message scales with L due to the correlation among transmitted messages, even though when the total transmit power is finite.

The comparison of the achieved MSE for both orthogonal and coherent MAC is plotted in Fig. 7, in which we take the total power $P = 16$ dB (relative to channel noise variance).

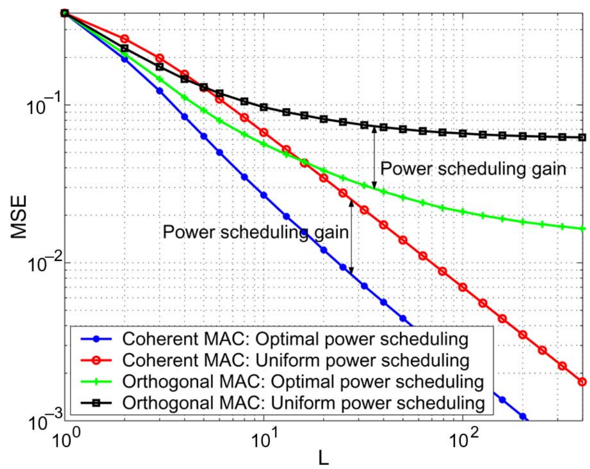


Fig. 7. MSE performance comparison between orthogonal and coherent MACs.

We see that with this finite total transmit power, the MSE of the coherent MAC case does decrease to zero for both uniform and optimal power scheduling when L increases. In addition, both MSE curves show an asymptotic behavior of $1/L$. However, the MSE for the orthogonal MAC behaves differently. As can be seen from Fig. 7, for the uniform power scheduling, the MSE of the orthogonal MAC goes to a finite level bounded below, while for the optimal power scheduling, there is a power scheduling gain. Such a gain is obtained by assigning most of the power to sensors with good observation and channel qualities (see details in [32]).

IV. Vector Case: Noiseless Channels

Now we focus on the vector case. In this section, we study the design of linear decentralized estimation by idealizing the communication link from sensors to the fusion center, or equivalently, assigning $\mathbf{u} = \mathbf{0}$ in (5). The main motivation here is to treat the decentralized estimation from a linear compression perspective, and observe how the bandwidth alone (which represents the number of linearly encoded messages) affects the estimation performance.

With the assumption of noiseless coherent MAC, the received signal has the form [see (5)]

$$\mathbf{y} = \mathbf{B}\mathbf{H}\mathbf{s} + \mathbf{B}\mathbf{v}.$$

We thus obtain that the linear MMSE estimator of $\mathbf{s}$ is $\hat{\mathbf{s}} = \left(\mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{y}$, and the MSE $\mathbf{D}$ satisfies

$$\mathbf{D}^{-1} = \mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}\mathbf{H} \tag{22}$$

where $(\mathbf{B}\mathbf{B}^T)^\dagger$ stands for the pseudoinverse of $\mathbf{B}\mathbf{B}^T$. Fixing the bandwidth (to be q) introduces a dimensionality constraint on each encoding matrix $\mathbf{A}_i$. We assume that channels from sensor i to the fusion center are nondegenerate, i.e., $\mathbf{G}_i$ is invertible for all i. Since $\mathbf{B}_i = \mathbf{G}_i\mathbf{A}_i$, solving $\mathbf{A}_i$ is then equivalent to solving $\mathbf{B}_i$. We therefore construct the following problem from which we can solve the optimal $\mathbf{A}_i$:

$$\begin{aligned} \min_{\mathbf{B}_i:1\le i\le L} \quad & \mathrm{Tr}(\mathbf{D}) \\ \text{s.t.} \quad & \mathbf{D}^{-1} = \mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}\mathbf{H} \\ & \mathbf{B} \in \mathbb{R}^{q\times\ell}. \end{aligned} \tag{23}$$

This problem has been previously discussed in [10], [29] in similar forms. In the following, we solve (23) using the property of projection mapping and the Cauchy's Interlacing Theorem [33], by noticing that $\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}$ is a projection matrix. We perform the singular value decomposition (SVD) of $\mathbf{B}$ and obtain

$$\mathbf{B} = \mathbf{U}_B\Lambda_B\mathbf{V}_B \tag{24}$$

where $\mathbf{U}_B$ and $\mathbf{V}_B$ are unitary matrices and Λ_B is diagonal. Assume $r_{\boldsymbol{B}} = \mathrm{rank}(\mathbf{B}) \le q$ since $\mathbf{B} \in \mathbb{R}^{q\times\ell}$, then Λ_B has the first $r_{\boldsymbol{B}}$ diagonal entries nonzero. We can calculate that

$$\begin{aligned} \mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B} &= \mathbf{V}_B^T\Lambda_B^T\mathbf{U}_B^T\left(\mathbf{U}_B\Lambda_B\Lambda_B^T\mathbf{U}_B^T\right)^\dagger\mathbf{U}_B\Lambda_B\mathbf{V}_B \\ &= \mathbf{V}_B^T\Lambda_B^T\left(\Lambda_B\Lambda_B^T\right)^\dagger\Lambda_B\mathbf{V}_B \\ &= \mathbf{V}_B^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\mathbf{V}_B. \end{aligned} \tag{25}$$

We proceed by performing the SVD of $\mathbf{H}:\mathbf{H} = \mathbf{U}_H\Lambda_H\mathbf{V}_H$. Combining it with (25) we obtain

$$\begin{aligned} &\mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}\mathbf{H} \\ &\quad = \mathbf{V}_H^T\Lambda_H^T\mathbf{U}_H^T\mathbf{V}_B^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\mathbf{V}_B\mathbf{U}_H\Lambda_H\mathbf{V}_H. \end{aligned} \tag{26}$$

Thus, it follows from (22) and (26) that

$$\begin{aligned} \mathbf{V}_H\mathbf{D}^{-1}\mathbf{V}_H^T &= \mathbf{V}_H\left(\mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^\dagger\mathbf{B}\mathbf{H}\right)\mathbf{V}_H^T \\ &= \mathbf{V}_H\Bigg(\mathbf{I}_p + \mathbf{V}_H^T\Lambda_H^T\mathbf{U}_H^T\mathbf{V}_B^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix} \\ &\qquad \times\mathbf{V}_B\mathbf{U}_H\Lambda_H\mathbf{V}_H\Bigg)\mathbf{V}_H^T \\ &= \mathbf{I}_p + \Lambda_H^T\mathbf{U}_H^T\mathbf{V}_B^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\mathbf{V}_B\mathbf{U}_H\Lambda_H. \end{aligned} \tag{27}$$

Introduce a unitary matrix $\mathbf{W} \overset{\text{def}}{=} \mathbf{V}_B\mathbf{U}_H$. Also let $\lambda_j(\mathbf{A})$ denote the jth smallest eigenvalues of a symmetric matrix $\mathbf{A}$. Then it is easy to see from (27) that for any $1 \le j \le r_{\boldsymbol{B}}$,

$$\begin{aligned} \lambda_j(\mathbf{D}^{-1}) &= \lambda_j\left(\mathbf{V}_H\mathbf{D}^{-1}\mathbf{V}_H^T\right) \\ &= \lambda_j\left(\mathbf{I}_p + \Lambda_H^T\mathbf{W}^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\mathbf{W}\Lambda_H\right) \\ &= \lambda_j\left(\mathbf{I}_\ell + \begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T\begin{bmatrix}\mathbf{I}_{r_{\boldsymbol{B}}} & 0\\ 0 & 0\end{bmatrix}\right) \\ &= \lambda_j\left[\mathbf{I}_{r_{\boldsymbol{B}}} + \left(\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T\right)_{(1:r_{\boldsymbol{B}},1:r_{\boldsymbol{B}})}\right] \end{aligned} \tag{28}$$

where $(\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T)_{(1:r_{\boldsymbol{B}},1:r_{\boldsymbol{B}})}$ denotes the upper $r_{\boldsymbol{B}} \times r_{\boldsymbol{B}}$ block of $\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T$. Due to the Cauchy's Interlacing Theorem for eigenvalues of a subblock matrix [33, Corollary III.1.5], we obtain that

$$\lambda_j\left[(\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T)_{r_{\boldsymbol{B}}\times r_{\boldsymbol{B}}}\right] \leq \lambda_j\left(\Lambda_H\Lambda_H^T\right) = \lambda_j(\mathbf{H}\mathbf{H}^T) \quad \text{for all } 1 \leq j \leq r_{\boldsymbol{B}} \tag{29}$$

noticing that $\mathbf{H}$ has an SVD: $\mathbf{H} = \mathbf{U}_H\Lambda_H\mathbf{V}_H$.

Therefore, we obtain from (28) and (29) that

$$\begin{aligned}\lambda_j(\mathbf{D}^{-1}) &= \lambda_j\left[\mathbf{I}_{r_{\boldsymbol{B}}} + (\mathbf{W}\Lambda_H\Lambda_H^T\mathbf{W}^T)_{(1:r_{\boldsymbol{B}},1:r_{\boldsymbol{B}})}\right] \\ &\leq 1 + \lambda_j(\mathbf{H}\mathbf{H}^T) \quad \text{when } 1 \leq j \leq r_{\boldsymbol{B}}\end{aligned}$$

and $\lambda_j(\mathbf{D}^{-1}) = 1$, when $r_{\boldsymbol{B}} < j \leq p$. Therefore, the distortion

$$D = \mathrm{Tr}(\mathbf{D}) = \sum_{j=1}^{p}\lambda_j(\mathbf{D}) \geq \sum_{j=1}^{r_{\boldsymbol{B}}}\frac{1}{1+\lambda_j(\mathbf{H}\mathbf{H}^T)} + (p - r_{\boldsymbol{B}}). \tag{30}$$

In the following, we choose the optimal $\mathbf{B}$ such that the distortion D is minimized. This is equivalent to choosing appropriate $\mathbf{V_B}$, $\mathbf{U_B}$, and $\Lambda_\mathbf{B}$ for the SVD of $\mathbf{B} = \mathbf{U}_B\Lambda_B\mathbf{V}_B$. Notice that $r_{\boldsymbol{B}} \overset{\mathrm{def}}{=} \mathrm{rank}(\mathbf{B}) \leq q$ given $\mathbf{B} \in \mathbb{R}^{q\times\ell}$. Thus, it follows from (30) that the optimal solutions are obtained when $r_{\boldsymbol{B}} = q$ and the equality holds in the last inequality of (30). First, according to (27), the condition $r_{\boldsymbol{B}} = q$ is equivalent to taking $\Lambda_B \in \mathbb{R}^{q\times\ell}$ in the following form

$$\Lambda_\mathbf{B} = \left[\Lambda_{q\times q}, \mathbf{0}_{q\times(\ell-q)}\right] \tag{31}$$

where $\Lambda_{q\times q}$ is any positive diagonal matrix with size q. Secondly, to make the equality hold in the last inequality of (30), we need the equality in (29) to hold with $r_{\boldsymbol{B}} = q$. This requires $\mathbf{W} = \mathbf{I}_q$, which leads to $\mathbf{V}_B = \mathbf{U}_H^T$ since $\mathbf{W} = \mathbf{V}_B\mathbf{U}_H$. In addition, we see that $\mathbf{U_B}$ can be any unitary matrix since it has on impacts on D.

As a summary, we establish the following theorem.

Theorem 7: For the linear decentralized estimation of a vector signal $\mathbf{s}$ when observations are given as in (1) and the MAC between sensors and the fusion center is noiseless but subject to a bandwidth constraint q, the optimal encoding matrices $\mathbf{A}_i^{\mathrm{opt}} = \mathbf{G}_i^{-1}\mathbf{B}_i^{\mathrm{opt}} : 1 \leq i \leq L$, where $\mathbf{B}_i^{\mathrm{opt}}$ is given as follows.

- First we give $\mathbf{B}^{\mathrm{opt}} = \mathbf{U_B}\Lambda_B\mathbf{U}_\mathbf{H}^T$, where $\mathbf{U_H} \in \mathbb{R}^{\ell\times\ell}$ is the left eigenspace of $\mathbf{H}$, $\mathbf{U_B}$ is any unitary matrix of size q, and Λ_B is as given in (31).
- After obtaining $\mathbf{B}^{\mathrm{opt}}$, we can get $\mathbf{B}_i^{\mathrm{opt}}$ from $\mathbf{B}^{\mathrm{opt}} = (\mathbf{B}_1^{\mathrm{opt}}, \mathbf{B}_2^{\mathrm{opt}}, \ldots, \mathbf{B}_L^{\mathrm{opt}})$.

The achieved MSE $D = \mathrm{Tr}(\mathbf{D})$, as a function of q, can be represented as

$$D(q) = \begin{cases} p - q + \sum_{i=1}^{q}\frac{1}{1+\lambda_i(\mathbf{H}^T\mathbf{H})}, & q < p \\ \sum_{i=1}^{p}\frac{1}{1+\lambda_i(\mathbf{H}^T\mathbf{H})} = D_0, & q \geq p \end{cases}$$

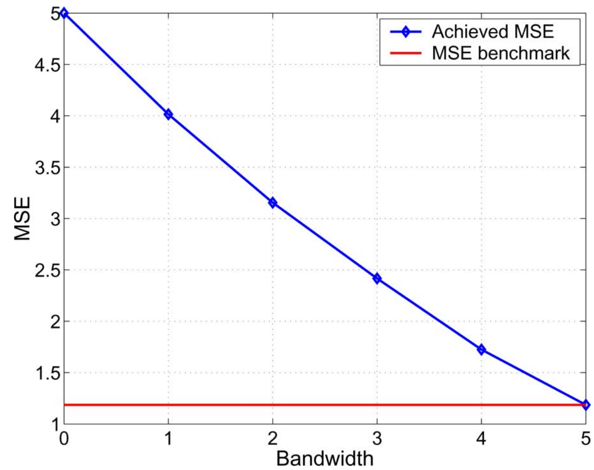


Fig. 8. MSE versus bandwidth q (noiseless channel case).

where $\lambda_i(\mathbf{H}^T\mathbf{H}) : 1 \leq i \leq p$ are eigenvalues of $\mathbf{H}^T\mathbf{H}$ with a decreasing order.

From the above theorem we can see that essentially, the optimal $\mathbf{B}$ should be full row-rank and can be obtained by matching the right eigenspace of $\mathbf{B}$ with the left eigenspace of $\mathbf{H}$ and choosing the left eigenspace of $\mathbf{B}$ arbitrarily. Another observation from Theorem 7 is that when the bandwidth q reaches p, which is the number of components in the unknown signal $\mathbf{s}$, the achieved MSE by optimally designed linear decentralized estimation obtains the centralized benchmark. Increasing the bandwidth q further does not improve the MSE performance. This demonstrates the relationship between the channel bandwidth and MSE performance, i.e., from the channel bandwidth point of view, q coherent transmission is enough for the best MSE performance.

A numerical plot of the MSE versus number of transmissions (i.e., q) is given in Fig. 8, in which we take $p = 5$, $L = 8$, and assume that each sensor makes only one observation (i.e., $\ell_i = 1$) and each entry of $\mathbf{H}_i$ has a complex Gaussian distribution. The simulated MSE is averaged over 500 realizations of $\{\mathbf{H}_i : i = 1, 2, \ldots, L\}$. Note that the noiseless channel case that we considered in Section III corresponds to an infinite transmit power in this general setup. From the figure we see that once the bandwidth q reaches $p = 5$, the MSE reaches the benchmark $D_0 = \mathrm{Tr}[(\mathbf{I}_p + \mathbf{H}^T\mathbf{H})^{-1}]$ (see (7)). Increasing q further does not help improve D.

V. Vector Case: Noisy Channels

In this section we study the case of noisy coherent MAC. With channel noise, signal power becomes relevant in determining the fundamental performance of wireless links. Therefore, in addition to the bandwidth constraint q, we add a transmit power constraint P_i on the messages transmitted from each sensor. This leads to the constraint on the encoding matrices $\mathbf{A}_i$ given in (8).

Throughout this section we need a technical assumption that there is no intersymbol interference between message transmissions over each link, and the channel gain remains constant during each observation period, which results in

$$\mathbf{G}_i = g_i\mathbf{I}_q, \quad 1 \leq i \leq L. \tag{32}$$

Based on the generic problem in (9), the optimal $\mathbf{A}_i$ can be obtained by solving the following problem:

$$\begin{aligned}\min_{\mathbf{A}_i,\mathbf{B}_i:1\le i\le L} \quad & \mathrm{Tr}(\mathbf{D})\\ \text{s.t.} \quad & \mathbf{D}^{-1} = \mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{I}_q+\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{H}\\ & \mathrm{Tr}\left[\mathbf{A}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}_{\ell_i}\right)\mathbf{A}_i^T\right] \le P_i\\ & \mathbf{B}_i = \mathbf{G}_i\mathbf{A}_i, \quad 1\le i\le L\\ & \mathbf{B} = [\mathbf{B}_1,\mathbf{B}_2,\ldots,\mathbf{B}_L] \in \mathbb{R}^{q\times\ell}. \end{aligned} \tag{33}$$

Problem (33) is not convex over $\{\mathbf{A}_i,\mathbf{B}_i : 1\le i\le L\}$, due to the nonconvex property of the first constraint. To obtain an efficient solution, we will transform (33) into an SDP problem with certain relaxation. At our first step, we will apply matrix inversion lemma to transform the nonconvex constraint, and obtain

$$(\mathbf{I}_\ell + \mathbf{B}^T\mathbf{B})^{-1} = \mathbf{I}_\ell - \mathbf{B}^T(\mathbf{I}_q + \mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B}.$$

Therefore

$$\begin{aligned}\mathbf{D}^{-1} &= \mathbf{I}_p + \mathbf{H}^T\mathbf{B}^T(\mathbf{I}_q+\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{H}\\ &= \mathbf{I}_p + \mathbf{H}^T\left(\mathbf{I}_\ell - (\mathbf{I}_\ell+\mathbf{B}^T\mathbf{B})^{-1}\right)\mathbf{H}\\ &= \mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{H}^T(\mathbf{I}_\ell+\mathbf{Q})^{-1}\mathbf{H}\end{aligned}$$

where we introduce a new positive semi-definite matrix $\mathbf{Q} \stackrel{\mathrm{def}}{=} \mathbf{B}^T\mathbf{B}$. For $1\le i\le L$, denote the ith diagonal block of $\mathbf{Q}$ as $\mathbf{Q}_i \in \mathbb{R}^{\ell_i\times\ell_i}$. Then $\mathbf{Q}_i = \mathbf{B}_i^T\mathbf{B}_i$.

We will solve (33) in terms of the new variable $\mathbf{Q}$. To ease the technical analysis, we need the assumption on $\mathbf{G}_i$ in (32). Under such an assumption, the power constraints (8) can be recast as

$$\begin{aligned}\mathrm{Tr}\left[\mathbf{A}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}\right)\mathbf{A}_i^T\right] &= \mathrm{Tr}\left[\mathbf{G}_i^{-1}\mathbf{B}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}\right)\mathbf{B}_i^T\mathbf{G}_i^{-T}\right]\\ &= g_i^{-2}\mathrm{Tr}\left[\mathbf{B}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}\right)\mathbf{B}_i^T\right]\\ &= g_i^{-2}\mathrm{Tr}\left[\mathbf{Q}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}\right)\right] \le P_i.\end{aligned}$$

Or equivalently, $\mathrm{Tr}[\mathbf{Q}_i(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I})] \le g_i^2 P_i$.

Notice that we can assume $\ell \ge q$, since otherwise, the number of coherent transmissions is larger than total number of sensor observations, which is in general not necessary.[2] Therefore, $\mathbf{B}$ is a fat matrix, the constraint $\mathbf{Q} \stackrel{\mathrm{def}}{=} \mathbf{B}^T\mathbf{B}$ is equivalent to $\mathrm{rank}(\mathbf{Q}) = q$. Finally, in terms of $\mathbf{Q}$, we can transform (33) to the following problem:

$$\begin{aligned}\min_{\mathbf{Q}} \quad & \mathrm{Tr}(\mathbf{D})\\ \text{s.t.} \quad & \mathbf{D}^{-1} = \mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{H}^T(\mathbf{I}_\ell+\mathbf{Q})^{-1}\mathbf{H}\\ & \mathrm{Tr}\left[\mathbf{Q}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}_p\right)\right] \le g_i^2P_i, \quad 1\le i\le L\\ & \mathbf{Q}\succeq 0, \quad \mathrm{rank}(\mathbf{Q}) = q.\end{aligned}$$

[2]In fact, if $\ell > q$, we can see that the rank constraint $\mathrm{rank}(\mathbf{Q}) = q$ is not needed anymore. In that case, we can simply solve an optimal positive semidefinite matrix $\mathbf{Q}^{\mathrm{opt}} \in \mathbb{R}^{\ell\times\ell}$. From $\mathbf{Q}^{\mathrm{opt}}$ we can see that any $\mathbf{B}^{\mathrm{opt}} = \mathbf{U}^T[\mathbf{Q}^{\mathrm{opt}}]^{1/2}$ with $\mathbf{U} \in \mathbb{R}^{\ell\times q}$ and $\mathbf{U}\mathbf{U}^T = \mathbf{I}_\ell$ is a feasible solution. The equation $\mathbf{U}\mathbf{U}^T = \mathbf{I}_\ell$ has infinite number of solutions when $\ell \le q$.

The first constraint on $\mathbf{D}$ can be changed to

$$\mathbf{D}^{-1} \preceq \mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{H}^T(\mathbf{I}_\ell+\mathbf{Q})^{-1}\mathbf{H}$$

since at the optimal solution, the equality holds, which can be proved by complementary slackness theorem [34]. Further applying the Schur's complement [34], we see that this constraint is equivalent to

$$\begin{bmatrix}\mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{H}^T(\mathbf{I}_\ell+\mathbf{Q})^{-1}\mathbf{H} & \mathbf{I}_p\\ \mathbf{I}_p & \mathbf{D}\end{bmatrix} \succeq 0. \tag{34}$$

Introducing a positive semidefinite matrix $\mathbf{R}$ such that $\mathbf{R} \succeq \mathbf{H}^T(\mathbf{I}_\ell+\mathbf{Q})^{-1}\mathbf{H}$, by Schur's complement, we obtain that the above constraint can be written in the form of two convex constraints:

$$\begin{bmatrix}\mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{R} & \mathbf{I}_p\\ \mathbf{I}_p & \mathbf{D}\end{bmatrix} \succeq 0, \quad \begin{bmatrix}\mathbf{R} & \mathbf{H}^T\\ \mathbf{H} & \mathbf{I}_\ell+\mathbf{Q}\end{bmatrix} \succeq 0.$$

Therefore, we eventually reach the following problem:

$$\begin{aligned}\min_{\{\mathbf{Q},\mathbf{R}\}} \quad & \mathrm{Tr}(\mathbf{D})\\ \text{s.t.} \quad & \begin{bmatrix}\mathbf{I}_p + \mathbf{H}^T\mathbf{H} - \mathbf{R} & \mathbf{I}_p\\ \mathbf{I}_p & \mathbf{D}\end{bmatrix} \succeq 0\\ & \begin{bmatrix}\mathbf{R} & \mathbf{H}^T\\ \mathbf{H} & \mathbf{I}_\ell+\mathbf{Q}\end{bmatrix} \succeq 0\\ & \mathrm{Tr}\left[\mathbf{Q}_i\left(\mathbf{H}_i\mathbf{H}_i^T+\mathbf{I}_p\right)\right] \le g_i^2P_i, \quad 1\le i\le L\\ & \mathbf{Q}\succeq 0, \quad \mathrm{rank}(\mathbf{Q}) = q.\end{aligned} \tag{35}$$

The objective function and constraints in (35) are convex except for the last rank constraint, which is in general nonconvex [34]. However, compared to the original problem in (33), (35) can be much more efficiently solved by numerical approximations. In the following, we will remove the rank constraint and solve (35) accordingly in a relaxed way. With such a relaxation, the obtained solution is suboptimal. However, we will show through numerical examples that in fact, the optimal $\mathbf{Q}$ has most of its spectrum distributed over the first few eigenmodes. By solving the relaxed SDP problem, we obtain a $\mathbf{Q}$ that may have rank larger than q.[3] One natural way of obtaining the solution to the original problem from such a $\mathbf{Q}$ is to perform the eigendecomposition:

$$\mathbf{Q} = \sum_{i=1}^{\ell} \lambda_i(\mathbf{Q})\mathbf{v}_{\mathbf{Q},i}\mathbf{v}_{\mathbf{Q},i}^T$$

where $\mathbf{v}_{\mathbf{Q},i}$ is the eigenvector of $\mathbf{Q}$ corresponds to $\lambda_i(\mathbf{Q})$. Taking the largest q principle eigencomponents of $\mathbf{Q}$, we obtain a solution for $\mathbf{B}$ as follows:

$$\mathbf{B} = \left[\sqrt{\lambda_1(\mathbf{Q})}\mathbf{v}_{\mathbf{Q},1}, \sqrt{\lambda_2(\mathbf{Q})}\mathbf{v}_{\mathbf{Q},2}, \ldots, \sqrt{\lambda_q(\mathbf{Q})}\mathbf{v}_{\mathbf{Q},q}\right]^T$$

[3]However, it can be observed from the numerical examples that the optimal solution of Q has most of its spectrum distributed at the first q eigenvalues.

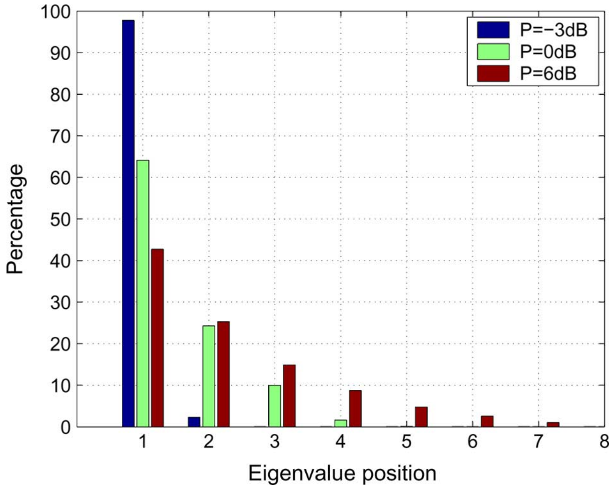


Fig. 9. Eigenvalue distribution of $\mathbf{Q}^{\rm opt}$ ($L = 8, \ell_i = 1$).

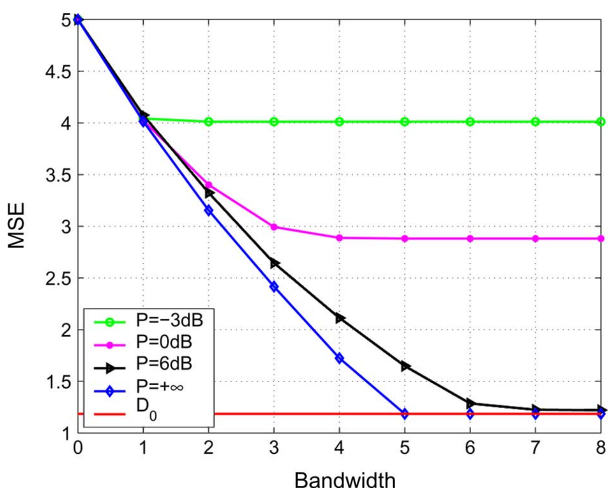


Fig. 10. MSE versus bandwidth q with different power budgets ($L = 8$, $\ell_i = 1$).

which is a $q\times\ell$ matrix. From $\mathbf{B}$, we can obtain $\mathbf{B}_i$ and eventually get the encoding matrices $\mathbf{A}_i$, similar to what is presented in Theorem 7.

SDP is a special class of convex optimization problem, and therefore enjoys all the advantages of convexity. There are well-developed numerical methods to solve a general convex optimization problem, among which the most well-known one is the interior point method. In the numerical example, we adopt an optimization toolbox: Self-Dual-Minimization (SeDuMi) [35] to solve the SDP formulated in (35) (after relaxing the last rank constraint). SeDuMi is a software package that solves optimization problems over symmetric cones using the primal-dual interior-point methods.

In the simulation, we choose each entry of $\mathbf{H}_i$ to be complex Gaussian with unit variance and $g_i = 1, i = 1, \cdots, L$ is the same as the scalar case, i.e., g_i are taken as $c_g \cdot d^{-3.5}$ where d is uniformly taken from the real interval [1, 10], and c_g is a normalization constant to make $E(g_i) = 1$. For the transmit power constraints, we take different power levels: $P_i = -3$, 0, 6 dB for all i. The simulated MSE is averaged over 500 realizations of $\{\mathbf{H}_i, \mathbf{G}_i = g_i\mathbf{I}_{q_i} : i = 1, 2, \ldots, L\}$. Note that the noiseless channel case that we considered in Section III corresponds to an infinite transmit power in this general setup.

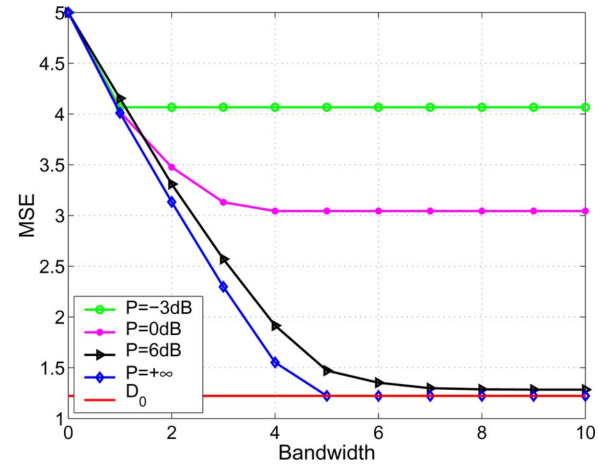


Fig. 11. MSE versus bandwidth q with different power budgets ($L = 5$, $\ell_i = 2$).

In the first example, we take $p = 5$, $L = 8$, and $\ell_i = 1$ for all i. A numerical plot of the eigenvalue distribution of $\mathbf{Q}$ solved from the SDP is given in Fig. 9, in which the vertical axis represents the percentage of each eigenvalue against the total spectrum (sum of all eigenvalues) of $\mathbf{Q}^{\rm opt}$. We see that although the rank constraint of $\mathbf{Q}$ has been relaxed, the optimal solution $\mathbf{Q}^{\rm opt}$ only has the first few eigenvalues of significant contribution. The achieved MSE is plotted in Fig. 10. Both Fig. 9 and Fig. 10 imply that in the noisy channel case, increasing the bandwidth (q value) above a certain threshold does not improve D, where the threshold is jointly decided by the power constraint and number of components in the source (i.e., p). Another example of the achieved MSE for $p = 5$, $L = 5$, and $\ell_i = 2$ is given in Fig. 11.

VI. Conclusion

Motivated by the optimality of uncoded transmission of a Gaussian source through an AWGN channel, we have considered the linear decentralized estimation of an unknown signal by a sensor network with a fusion center. By assuming nonorthogonal channel usage and coherent combination of sensor messages at the fusion center, we have designed the optimal linear decentralized estimation scheme subject to bandwidth and/or power constraints, for the cases where source and observations are scalars or vectors.

For the scalar case, we have considered the optimal power allocation among sensors. We have derived the optimal power scheduling that optimizes the achieved MSE under a total power constraint. It is shown that the optimal transmit power for each sensor can be determined in a distributed manner by using local observation signal to noise ratio and individual channel gain, provided that the fusion center broadcast two appropriate constants (P and c; see Remark 5). Simulations show that the proposed power scheduling strategy significantly improves the mean squared error performance when compared to the uniform power scheduling. We have also shown that the MSE performance has significantly different asymptotic behaviors when L is large for orthogonal and coherent MACs.

For the vector case, we have studied the linear decentralized estimation by designing the optimal encoding matrices. We have shown that under a bandwidth constraint, the optimal encoding matrices have a closed-form solution when neglecting channel noise. This is in contrast to the case of orthogonal channel usage, where the complexity of designing the optimal linear decentralized estimation is NP-hard in general [28]. By taking the channel noise into consideration, we can efficiently solve the problem through SDP under both bandwidth and power constraints.

Throughout this work, we have assumed that $\{(\mathbf{H}_i, \mathbf{G}_i)\}_{i=1}^L$ remain constant over each observation period. This assumption is reasonable when the network condition changes slowly (e.g., in a quasi-static manner). In such a case, the fusion center can first collect the values of $\{(\mathbf{H}_i, \mathbf{G}_i)\}_{i=1}^L$ (via training sequences), and then solve for the optimal encoding matrices. Concerning the feedback, for the scalar case, only two universal parameters need to be broadcasted from the fusion center to all sensors. Each sensor can then determine its coding factor by using the local parameters. For the vector case, our current scheme requires the feedback of $\mathbf{A}_i$. Thus, a distributed implementation which does not require heavy communication load for feedback is desirable. In addition, we have assumed perfect synchronization between sensors and the fusion center. The effect on the system performance due to nonideal synchronization, nonideal fusion center feedback, or partial knowledge of channel states and sensing models, is worth further investigation.

Appendix
Solving the Optimal Power Scheduling

Since (17) is not a convex problem, we will solve an alternative but equivalent problem to obtain the optimal power allocation. Suppose the optimal solution of (16) takes the form $D^{\mathrm{opt}}(P)$, then it should be a monotonically decreasing function of P. Now we consider an alternative case of minimizing the total power consumption while meeting a given MSE constraint. This gives the following problem:

$$\begin{aligned} \min_{P_i} \quad & P = \sum_{i=1}^{L} P_i \\ \text{s.t.} \quad & D(P_1, P_2, \ldots, P_L) \le D. \end{aligned} \tag{36}$$

Suppose the optimal solution of this problem is $P^{\mathrm{opt}}(D)$. It is easy to see that $D^{\mathrm{opt}}(P)$ and $P^{\mathrm{opt}}(D)$ are inverses of each other. Instead of solving (16), we propose to solve (36), since the latter can be transformed into a convex problem. This is shown in the sequel.

In terms of the coding factor a_is, we obtain that (36) is equivalent to the following problem:

$$\begin{aligned} \min_{a_1, a_2, \ldots, a_L} \quad & \sum_{i=1}^{L} \left(1 + h_i^2\right) a_i^2 \\ \text{s.t.} \quad & \left(1 + \sum_{i=1}^{L} g_i^2 a_i^2\right)^{-1} \left(\sum_{i=1}^{L} g_i h_i a_i\right)^2 \ge D^{-1} - 1. \end{aligned}$$

The above problem is not convex in terms of $\{a_i\}_{i=1}^L$. We introduce a slack variable $t = \sum_{i=1}^{L} g_i h_i a_i$ and it becomes

$$\begin{aligned} \min_{a_1, a_2, \ldots, a_L; t} \quad & \sum_{i=1}^{L} \left(1 + h_i^2\right) a_i^2 \\ \text{s.t.} \quad & 1 + \sum_{i=1}^{L} g_i^2 a_i^2 \le (D^{-1} - 1)^{-1} t^2 \\ & \sum_{i=1}^{L} g_i h_i a_i - t = 0. \end{aligned} \tag{37}$$

This problem is convex over the decision variables $\{a_1, a_2, \ldots, a_L; t\}$. In fact, it is a second-order cone programming (SOCP) that is efficiently solvable by interior point methods [34]. However, we show below that it can actually be solved analytically.

The Lagrangian function for (37) is

$$\begin{aligned} G(q, t; \nu, \mu) = & \sum_{i=1}^{L} \left(1 + h_i^2\right) a_i^2 \\ & + \nu \left(1 + \sum_{i=1}^{L} g_i^2 a_i^2 - (D^{-1} - 1)^{-1} t^2\right) \\ & + \mu \left(t - \sum_{i=1}^{L} g_i h_i a_i\right) \end{aligned}$$

where $\nu \ge 0$ and $\mu \in \mathbb{R}$.

From the Lagrangian function we can derive the following Karush-Kuhn-Tucker (KKT) conditions [34]:

$$\begin{aligned} & \frac{\partial G}{\partial t} = -2\nu (D^{-1} - 1)^{-1} t + \mu = 0 \\ & \frac{\partial G}{\partial a_i} = 2\left(1 + h_i^2 + \nu g_i^2\right) a_i - g_i h_i \mu = 0, \quad 1 \le i \le L \\ & \nu \left(1 + \sum_{i=1}^{L} g_i^2 a_i^2 - (D^{-1} - 1)^{-1} t^2\right) = 0 \\ & \sum_{i=1}^{L} g_i h_i a_i - t = 0. \end{aligned}$$

To solve the above system, we first obtain from the second KKT condition that

$$a_i = \frac{g_i h_i \mu}{2\left(1 + h_i^2 + \nu g_i^2\right)}. \tag{38}$$

Plugging (38) and $t = \sum_{i=1}^{L} g_i h_i a_i$ into the first KKT condition we get that ν is the root of the following equation (in ν):

$$\sum_{i=1}^{L} \frac{\nu g_i^2 h_i^2}{1 + h_i^2 + \nu g_i^2} = D^{-1} - 1. \tag{39}$$

Notice that the function on the left side is an increasing function in ν. Next we solve for μ. From (39) we see that $\nu = 0$ only for a trivial case when $D = 1$ (since we have assumed that the source s has a unitary variance). Thus we can assume $\nu \neq 0$. Again it follows from the first KKT condition that $t = ((D^{-1}-1)\mu)/2\nu$. We plug it altogether with (38) into the third KKT condition, and obtain that

$$\mu = 2\left(\frac{D^{-1}-1}{\nu^2} - \sum_{i=1}^{L} \frac{g_i^4 h_i^2}{\left(1+h_i^2+\nu g_i^2\right)^2}\right)^{-1/2}. \tag{40}$$

Thus it follows from (38) that the optimal power scheduling, in terms of ν and μ, is

$$P_i = a_i^2\left(1+h_i^2\right) = \frac{\mu^2}{4}\frac{g_i^2 h_i^2\left(1+h_i^2\right)}{\left(1+h_i^2+\nu g_i^2\right)^2}. \tag{41}$$

Interestingly, we can verify that

$$\begin{aligned} P &= \sum_{i=1}^{L} P_i = \frac{\mu^2}{4}\sum_{i=1}^{L}\frac{h_i^2\left(1+h_i^2\right)g_i^2}{\left(1+h_i^2+\nu g_i^2\right)^2} \\ &\overset{(a)}{=} \left(\frac{\sum_{i=1}^{L}\frac{\nu g_i^2 h_i^2}{1+h_i^2+\nu g_i^2}}{\nu^2} - \sum_{i=1}^{L}\frac{h_i^2 g_i^4}{\left(1+h_i^2+\nu g_i^2\right)^2}\right)^{-1} \\ &\quad \times \sum_{i=1}^{L}\frac{h_i^2\left(1+h_i^2\right)g_i^2}{\left(1+h_i^2+\nu g_i^2\right)^2} \\ &\overset{(b)}{=} \nu \end{aligned}$$

where (a) is due to (39), (40), and (b) follows from some direct calculation. Therefore, $\nu = P$. By (39), we obtain that at the optimal solution, D and P satisfy

$$\frac{1}{D} = 1 + \sum_{i=1}^{L}\frac{h_i^2}{1+\left(1+h_i^2\right)/\left(g_i^2 P\right)}. \tag{42}$$

Therefore, the minimum amount of power $P^{\text{opt}}(D)$ (which is the solution to (36) of achieving a given distortion D) satisfies (42), and the optimal power allocation achieving such a minimum total power is given by (39)–(41).

Since $P^{\text{opt}}(D)$ and $D^{\text{opt}}(P)$ are inverses of each other [where $D^{\text{opt}}(P)$ is the solution to (16) of achieving the minimum D under the total power constraint P], we obtain that $D^{\text{opt}}(P)$ is given by (42) as well. The corresponding optimal power allocation (19) can be calculated from (41), using the fact that $\nu = P$ and $\sum_{i=1}^{L} P_i = P$.

REFERENCES

[1] A. C. Yao, "Some complexity questions related to distributed computing," in *Proc. 11th STOC*, 1979, pp. 209–213.

[2] Z.-Q. Luo, "Communication complexity of some problems in distributed computation," Ph.D. dissertation, Tech. Rep. LIDS-TH-1909, Lab. Inf. Decision Syst., MIT. Cambridge, MA, 1989.

[3] J. N. Tsitsiklis and Z.-Q. Luo, "Communication complexity of convex optimization," *J. Complex.*, vol. 3, pp. 231–243, 1987.

[4] D. A. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Trans. Autom. Control*, vol. 30, no. 5, pp. 418–425, May 1985.

[5] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.

[6] J. N. Tstsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993, JAI Press.

[7] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[8] J. A. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.

[9] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. Commun.*, vol. 48, pp. 181–184, Feb. 2000.

[10] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion Part VI: Sensor data compression," in *Proc. Int. Conf. Inf. Fusion*, Queensland, Australia, 2003, pp. 221–228.

[11] J.-J. Xiao, A. Ribeiro, G. B. Giannakis, and Z.-Q. Luo, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.

[12] A. S. Willsky, M. Bello, D. A. Castanon, B. C. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Trans. Autom. Control*, vol. 27, no. 4, pp. 799–813, Aug. 1982.

[13] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, pp. 558–567, Oct. 1965.

[14] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1147–1158, May 2003.

[15] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science*. New York: Springer, Apr. 2003, vol. 2634, pp. 162–177.

[16] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian sensor network," presented at the 2007 Inf. Theory Appl. Workshop, San Diego, CA, Jan.–Feb. 2007.

[17] A. D. Sarwate and M. Gastpar, "Fading observation alignment via feedback," in *Proc. 4th Int. Symp. Inf. Process. Sensor Networks*, Los Angeles, CA, Apr. 2005, pp. 317–323.

[18] K. Eswaran and M. Gastpar, "On the quadratic AWGN CEO problem and non-Gaussian sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 219–223.

[19] K. Liu, H. El-Gamal, and A. Sayeed, "On optimal parametric field estimation in sensor networks," in *Proc. IEEE/SP 13th Workshop on Statist. Signal Process.*, Jul. 2005, pp. 1170–1175.

[20] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor network," in *Proc. 4th Int. Symp. Inf. Process. Sens. Netw.*, Los Angeles, CA, Apr. 2005, pp. 332–339.

[21] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access in sensor networks," in *IEEE Workshop on Signal Process. Adv. Wireless Commun.*, New York, Jun. 2005, pp. 1018–1022.

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

[23] K. Liu and A. Sayeed, "Optimal distributed detection strategies for wireless sensor networks," presented at the 43rd Allerton Conf. Commun., Contr., and Comput., Monticello, IL, Oct. 2004.

[24] 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 Int. Symp. Inf. Process. Sens. Netw.*, Palo Alto, CA, Apr. 2003, pp. 1–16.

[25] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 765–775, Apr. 2005.

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

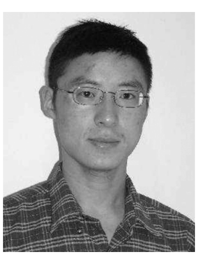
[27] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Process. Sens. Netw.*, Los Angeles, CA, Apr. 2005, pp. 63–70.

[28] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 1441–1445.
[29] I. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.* , vol. 55, no. 8, pp. 4284–4299, Aug. 2007 .
[30] S. M. Kay*, Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[31] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 413–422, Feb. 2006.
[32] S. Cui, J. Xiao, A. J. Goldsmith, Z. Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.* vol. 55, no. 9, pp. 4683–4695, Sep. 2007 [Online]. Available: http://www.ece.arizona.edu/~cui/publication.html
[33] R. Bhatia*, Matrix Analysis*. New York: Springer-Verlag, 1991, vol. 169, Graduate Texts in Mathematics.
[34] S. Boyd and L. Vandenberghe*, Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[35] J. F. Sturm, Using SeDuMi 1.02, a Matlab Toolbox for Optimization Over Symmetric Cones [Online]. Available: http://www.optimization-online.org/DB_HTML/2001/10/395.html

Zhi-Quan Luo (F'07) received the B.Sc. degree in applied mathematics in 1984 from Peking University, Beijing, China. Subsequently, he was selected by a joint committee of the American Mathematical Society and the Society of Industrial and Applied Mathematics to pursue Ph.D. doctoral studies in the United States. After a one-year intensive training course in mathematics and English at the Nankai Institute of Mathematics, Tianjin, China, he entered the Operations Research Center and the Department of Electrical Engineering and Computer Science of the Massachusetts Institute of Technology (MIT), Cambridge, where he received the Ph.D. degree in operations research in 1989.

From 1989 to 2003, he held a faculty position with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, where he eventually became the Department Head and held a Canada Research Chair in Information Processing. Since April 2003, he has been with the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities, as a Full Professor and holds an endowed ADC Chair in digital technology. His research interests include the union of optimization algorithms, data communication, and signal processing.

Prof. Luo serves on the IEEE Signal Processing Society Technical Committees on Signal Processing Theory and Methods (SPTM), and on the Signal Processing for Communications (SPCOM). He is a corecipient of the 2004 IEEE Signal Processing Society's Best Paper Award, and has held editorial positions for several international journals including the *Journal of Optimization Theory and Applications, Mathematics of Computation*, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He currently serves on the editorial boards for a number of international journals including *SIAM Journal on Optimization, Mathematical Programming* and *Mathematics of Operations Research*.



Jin-Jun Xiao (S'04–M'07) received the B.Sc. degree in applied mathematics from Jilin University, Changchun, China, in 1997. He received the M.Sc. degree in mathematics and the Ph.D. degree in electrical engineering from the University of Minnesota, Twin Cities, in 2003 and 2006 respectively.

He is currently a Postdoctoral Fellow with Washington University, St. Louis, MO. In summer 2002, he was with Symbol Technologies, Holtsville, NY, where he worked on barcode signal processing. His current research interests are in distributed signal processing, wireless communications, multiuser information theory, and their application to sensor networks.

Shuguang Cui (S'99–M'05) received the B.Eng. degree in radio engineering with the highest distinction from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, the M.Eng. degree in electrical engineering from McMaster University, Hamilton, Canada, in 2000, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2005.

From 1997 to 1998, he was with Hewlett-Packard, Beijing, P. R. China, as a system engineer. In summer 2003, he worked at National Semiconductor, Santa Clara, CA, on the ZigBee project. His current research interests include cross-layer energy minimization for low-power sensor networks, hardware and system synergies for high-performance wireless radios, statistical signal processing, and general communication theories. He is currently an Assistant Professor of Electrical and Computer Engineering, University of Arizona, Tucson.

Dr. Cui received an NSERC graduate fellowship from the National Science and Engineering Research Council of Canada and the Canadian Wireless Telecommunications Association (CWTA).

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California, Berkeley.

She is a Professor of Electrical Engineering at Stanford University, Stanford, CA, and was previously an Assistant Professor of Electrical Engineering at the California Institute of Technology (Caltech), Pasadena. She has also held industry positions with Maxim Technologies and AT&T Bell Laboratories, and is currently on leave from Stanford University as cofounder and CTO of a wireless startup company. Her research includes work on capacity of wireless channels and networks, wireless communication theory, energy-constrained wireless communications, wireless communications for distributed control, and cross-layer design of wireless networks. She is author of the book *Wireless Communications* and coauthor of the book *MIMO Wireless Communications*, (Cambridge, U.K.: Cambridge University Press).

Dr. Goldsmith is a Fellow of Stanford University, and currently holds Stanford's Bredt Faculty Development Scholar Chair. She has received several awards for her research, including the National Academy of Engineering Gilbreth Lectureship, the Alfred P. Sloan Fellowship, the Stanford Terman Fellowship, the National Science Foundation CAREER Development Award, and the Office of Naval Research Young Investigator Award. She was also a corecipient of the 2005 IEEE Communications Society and Information Theory Society joint paper award. She currently serves as Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and as Editor of the journals *Foundations and Trends in Communications and Information Theory* and *Foundations and Trends in Networks*. She was previously an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and for the IEEE *Wireless Communications Magazine*, and has served as Guest Editor for several IEEE journal and magazine Special Issues. She is active in committees and conference organization for the IEEE Information Theory and Communications Societies, and is currently serving as technical program co-chair for ISIT 2007. She is an elected member of the Board of Governers for both societies, a distinguished lecturer for the IEEE Communications Society, and the second vice-president and student committee chair of the IEEE Information Theory Society.