Linear Estimation of Correlated Data in Wireless Sensor Networks with Optimum Power Allocation and Analog Modulation

Israfil Bahceci and Amir K. Khandani

***Abstract*—In this paper, we study the energy-efficient distributed estimation problem for a wireless sensor network where a physical phenomena that produces correlated data is sensed by a set of spatially distributed sensor nodes and the resulting noisy observations are transmitted to a fusion center via noise-corrupted channels. We assume a Gaussian network model where (i) the data samples being sensed at different sensors have a correlated Gaussian distribution and the correlation matrix is known at the fusion center, (ii) the links between the local sensors and the fusion center are subject to fading and additive white Gaussian noise (AWGN), and the fading gains are known at the fusion center, and (iii) the central node uses the squared error distortion metric. We consider two different distortion criteria: (i) individual distortion constraints at each node, and (ii) average mean square error distortion constraint across the network. We determine the achievable power-distortion regions under each distortion constraint. Taking the delay constraint into account, we investigate the performance of an uncoded transmission strategy where the noisy observations are only scaled and transmitted to the fusion center. At the fusion center, two different estimators are considered: (i) the best linear unbiased estimator (BLUE) that does not require knowledge of the correlation matrix, and (ii) the minimum mean-square error (MMSE) estimator that exploits the correlations. For each estimation method, we determine the optimal power allocation that results in a minimum total transmission power while satisfying some distortion level for the estimate (under both distortion criteria). The numerical comparisons between the two schemes indicate that the MMSE estimator requires less power to attain the same distortion provided by the BLUE and this performance gap becomes more dramatic as correlations between the observations increase. Furthermore, comparisons between power-distortion region achieved by the theoretically optimum system and that achieved by the uncoded system indicate that the performance gap between the two systems becomes small for low levels of correlation between the sensor observations. If observations at all sensor nodes are uncorrelated, the uncoded system with MMSE estimator attains the theoretically optimum system performance.**

***Index Terms*—Distributed estimation, wireless sensor network, power-distortion region, MMSE, BLUE.**

Paper approved by M. Skoglund, the Editor for Source/Channel Coding of the IEEE Communications Society. Manuscript received July 3, 2006; revised June 25, 2007.

I. Bahceci is with Advanced Wireless Access Research, Nortel Networks, Ottawa, ON K2H8E9 Canada (e-mail: bahcecis@nortel.com).

A. K. Khandani is with the Department of Elect. & Comp. Eng., University of Waterloo, Waterloo, ON N2L3G1 Canada (e-mail: khandani@cst.uwaterloo.ca).

Digital Object Identifier 10.1109/TCOMM.2008.060396.

I. INTRODUCTION

WIRELESS sensor networking (WSN) is an emerging technology in many application areas including environment monitoring, health, security and surveillance, and robotic exploration [1]. Networks of sensor systems allow for many distributed processing and cooperative communication techniques including distributed data compression [2], tracking and classification [3], and distributed detection [4], [5] and distributed estimation [6]. In this paper, we focus on an estimation problem where each sensor sends its observation to a fusion center where a global estimation is made. Because of hard energy limitations, a significant research problem is to develop schemes that minimize the transmission energy while satisfying a certain distortion level.

Various approaches can be followed to solve this problem. For instance, one can digitize all observed data at the local sensors using distributed compression/coding algorithms and transmit the digitized data to the fusion center. This is suggested by the aforementioned source-channel separation theorem of Shannon [7]. Here, the problem falls into the area of multiterminal source coding where the main issue is to characterize the rate-distortion region. More specifically, the goal is to determine all rate vectors $(R_1, \ldots, R_K)$ at which the source samples at all K sensors can be encoded separately and then decoded jointly at the fusion center attaining a prespecified distortion level $(D_1, \ldots, D_K)$. Coding for multiterminal source-channel communications and the associated rate-distortion region is extensively studied in the literature and several partial solutions are reported to date [8–15]. In [8], Slepian and Wolf consider lossless coding of discrete memoryless correlated sources and show that it is possible to attain the rate-distortion region of joint encoding/decoding by a scheme in which the sources are encoded separately while decoded jointly. In [9], Wyner and Ziv initiate the research on lossy source coding where they determine the rate-distortion region for source coding with side information. Berger [10] and Tung [11] determined tight inner bounds to the rate distortion region for a finite-alphabet source. Among other results, for a two-terminal Gaussian source coding case, tight inner and outer bounds on the rate-distortion region are reported by Oohama [13] and Zamir and Berger [15]; and recently, a complete solution is obtained by Wagner et. al. [16]. Recent work on multiterminal source coding problem with more general cases subsuming many special cases include [17] by Wagner et. al. and [18] by Servetto where tighter inner and outer bounds are provided. An important class of the multi-terminal source/channel coding problem is the Chief Executive Officer (CEO) problem where a set of separate agents make noisy observations of a common source and transmit a summary (encoded version) to the CEO where the final decision (e.g., estimation) is made [19–23]. In addition to the theoretical results mentioned above, several distributed estimation schemes have also been proposed, for example [6,24–26].

The separate source-channel coding generally incurs large delays. However in some cases, such as point-to-point communication over an additive white Gaussian noise channel, it is possible to use a simple amplify-and-forward approach [27–29] to achieve optimum performance. For multiterminal source-channel communications, however, it is not clear whether this approach performs well [23], [30]. Although in some special cases, an optimal quantization followed by a Slepian-Wolf coding achieves optimality [15], since the information in sensor networks is, in general, delay sensitive, and because of the bandwidth and energy constraints, one requires simple coding/processing methods, rather than relying on the source-channel separation theorem. In [31], Cui et al. study an uncoded analog transmission method for estimation in sensor networks where it is assumed that the noisy version of *the same* signal is observed at the local nodes and single-side band analog modulation is employed to transmit the real-valued observations to the fusion center over orthogonal channels. This analog approach is simple since it relies on an amplify-and-forward technique. Gastpar and Vetterli [30] study a similar problem for a network where sensors communicate over a multiaccess channel to a central node. They show that, for a Gaussian source, the decay rate of the distortion attained by uncoded transmission is larger than that attained by the best coding scheme based on the source-channel separation theorem. As the number of sensor nodes increases to infinity, this system attains optimum performance.

In sensor networks, the sensor observations are likely to vary from one sensor to the other. However, especially for dense sensor networks, the observations might be strongly correlated [32], [33]. Gastpar and Vetterli [34] study a similar network monitoring an L dimensional Gaussian source and determine lower bounds for the end-to-end distortion. Nowak and Sayeed [35–37] recently investigated the estimation of a two-dimensional piece-wise linear spatial field. In this paper, we also study the estimation problem for a network where the sensor nodes observe spatially correlated data with an emphasis on linear transmission and estimation along with optimum power allocation. This problem is similar to the Quadratic Gaussian CEO problem e.g., [20,21,23,34], with the exception that, in our model, the agents (sensor nodes) observe independently corrupted samples of a spatially correlated data field. We assume that at each observation instant, the source samples are jointly Gaussian, and the noise corrupting these samples are spatially independent additive white Gaussian noise (AWGN). The transmissions from sensor nodes to the fusion center take place over orthogonal channels (e.g., by TDMA/FDMA) that are subject to fading and additive channel noise which is also Gaussian. Using the signals gathered from all sensor nodes, the fusion center estimates the source vector according to some distortion criterion. In this paper, we study two different measures to characterize the distortion: (i) individual mean squared error for estimation quality of the signal at each sensor node, and (ii) the mean-squared error distortion averaged across the sensor nodes. For each distortion measure, we determine the achievable power-distortion pairs that can be attained by the theoretically optimum system. Taking the delay constraint into account, we next study the amplify-and-forward-based uncoded transmission strategy for this network. This transmission approach can be imagined as a *rate-1 joint source-channel coding with separate encoders* at different sensor nodes and the estimation can be viewed as a *joint decoding* at the fusion center. The estimation method depends critically on the information available to the fusion center about the statistics of the source. We consider two situations: (i) where there is no statistical knowledge of the source, and (ii) where the autocorrelation matrix of the source vector is known at the fusion center, and therefore, the optimum estimator is the minimum mean-squared error estimator (MMSE). In situation (i), we study the best linear unbiased estimator (BLUE) that does not require any statistical information. For both cases, we determine the optimum power allocation schemes that minimize the total power required to satisfy a certain distortion level. We note that the estimation problem considered in [31] is different from the one we consider since we study the estimation of a spatially correlated field, while [31] considers the estimation of a common source. As we see in Section III, the solutions to these two problems differ significantly from each other.

In Section II of this paper, we describe the sensor network model and specify the parameters for analog transmission. We study in Section III the estimation based on the uncoded analog forwarding strategy where we solve the optimum power allocation problems for various schemes. In Section IV, we present numerical examples for the optimum power allocation and compare the performance between various schemes. Finally, in Section V, we summarize the results and list future directions.

II. System Model

Assume that there are K sensors and the observation at the k^{th} sensor at time t, $x_k(t)$, $k = 1, \ldots, K$, is a random signal given by

$$x_k(t) = \theta_k(t) + w_k(t),\ t = 1, 2, \ldots \tag{1}$$

where $\theta_k(t)$ is the value of the observed field and $w_k(t) \sim \mathcal{C}(0, \sigma_k^2)$ is the additive white Gaussian (AWGN) noise at node k. Let $\boldsymbol{\theta}(t) = [\theta_1(t), \ldots, \theta_k(t)]$ and $\mathbf{w}(t) = [w_1(t), \ldots, w_K(t)]$. We assume that $\boldsymbol{\theta}(t)$ is an independently and identically distributed Gaussian vector whose autocorrelation matrix is given by (for brevity, we drop the time parameter)

$$\mathbf{R}_{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\boldsymbol{\theta}^H\}.$$

The transmitted signal from sensor node k, $k = 1, \ldots, K$, is given by

$$y_k(t) = \sqrt{\alpha_k} x_k(t)$$

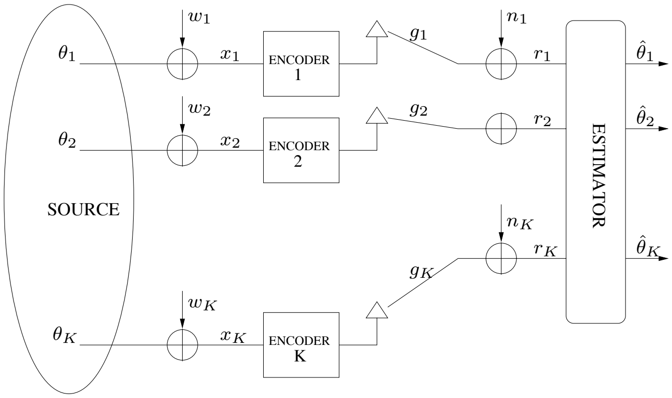


Fig. 1. Network model. Encoder k, $k = 1, \ldots, K$, corresponds to an amplify-and-forward scheme for the uncoded transmission strategy. With source-channel separation theorem, Encoder k collects a long sequence of $x_k(t)$ and generates another sequence $y_k(t)$, $t = 1, 2, \ldots$, in which case the estimator may experience a large delay.

where α_k is power scaling parameter. Thus, the average transmit power at node k is given by

$$E\{y_k^2(t)\} = \alpha_k \left([\mathbf{R}_{\boldsymbol{\theta}}]_{k,k} + \sigma_k^2\right)$$

where $E\{\cdot\}$ denotes the expectation operator and $[\mathbf{A}]_{i,j}$ denotes $(i,j)^{\text{th}}$ entry of the matrix $\mathbf{A}$. We assume that some form of orthogonal multiple access technique such as time or frequency division multiple access (TDMA/FDMA) can be employed to realize the access to the fusion center [23], [31]. Assuming a Rayleigh flat fading channel with a gain factor of g_k between the k^{th} node and the fusion center[†], we can express the received signal as

$$r_k(t) = \sqrt{\alpha_k g_k}\theta_k(t) + \sqrt{\alpha_k g_k} w_k(t) + n_k(t)$$

where $n_k(t) \sim \mathcal{C}(0, \xi_k^2)$ denotes the additive white Gaussian channel noise for the transmission from the k^{th} node. In vector form, we have the input-output relation (we drop the time parameter)

$$\mathbf{r} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v} \tag{2}$$

where

$$\begin{aligned} \mathbf{r} &= [r_1, \ldots, r_K]^T \text{and } \boldsymbol{\theta} = [\theta_1, \ldots, \theta_K], \\ \mathbf{H} &= \text{diag}(\sqrt{\alpha_1 g_1}, \ldots, \sqrt{\alpha_K g_K}), \\ \mathbf{v} &= (\sqrt{\alpha_1 g_1} w_1 + n_1, \ldots, \sqrt{\alpha_K g_K} w_K + n_K) \end{aligned}$$

and diag$(\cdot)$ denotes a diagonal matrix formed from its vector argument. Since $\mathbf{w} = [w_1, \ldots, w_K]$ and $\mathbf{n} = [n_1, \ldots, n_K]$ are independently distributed vectors with independently distributed entries, $\mathbf{v}$ is also a Gaussian vector whose covariance matrix is given by

$$\mathbf{R}_{\mathbf{v}} = \text{diag}(\alpha_1 g_1 \sigma_1^2 + \xi_1^2, \ldots, \alpha_K g_K \sigma_K^2 + \xi_K^2)$$

Let $\hat{\boldsymbol{\theta}} = f(\boldsymbol{\theta})$ denote any estimate of $\boldsymbol{\theta}$. The error covariance matrix is defined by

$$\mathbf{R}_{\boldsymbol{\epsilon}} = E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}.$$

The Cramer-Rao bound on $\mathbf{R}_{\boldsymbol{\epsilon}}$ for the signal model in (2) is given by [39]

$$\mathbf{R}_{\boldsymbol{\epsilon}} \succeq (\mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1}, \tag{3}$$

which can be attained by the linear minimum mean square error estimator (MMSE) for a Gaussian signal model[‡]. The Cramer-Rao bound in (3) specifies the best error-covariance matrix attainable by any estimator for the prescribed signal model in (2). Note that the k^{th} diagonal entry of the error covariance matrix, $[\mathbf{R}_{\boldsymbol{\epsilon}}]_{k,k}$, is the squared error distortion at node k. In this paper, we consider two distortion measures: (i) an individual distortion measure for each node

$$\mathbf{d} = [d_1, \ldots, d_K] \tag{4}$$

where $d_k = E\{|\theta_k - \hat{\theta}_k|^2\} = [\mathbf{R}_{\boldsymbol{\epsilon}}]_{k,k}$, and (ii) an average distortion measure

$$\begin{aligned} d_s &= \frac{1}{K} E\left\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\} = \frac{1}{K} \text{tr}\left(E\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\}\right) \\ &= \frac{1}{K} \text{tr}(\mathbf{R}_{\boldsymbol{\epsilon}}) \end{aligned} \tag{5}$$

where tr$(\cdot)$ denotes the trace operator. This latter distortion is a measure of average mean squared-error in the estimation across all sensor nodes. These two distortion measures can be employed depending on the sensor network application.

The amplify-and-forward strategy described above can be imagined as a single-letter rate-1 joint source channel code. In general, longer source and channel codes may be required to achieve better performance. The theoretically optimum performance bounds for achievable power-distortion pairs in such a system can be derived using Shannon rate-distortion bounds and channel capacity formula [28], which is obtained in Appendix A.

III. Uncoded Analog Transmission

In this section, we study a rate-1 joint source channel code, the amplify-and-forward strategy along with the minimum power consumption and optimum power allocation under some distortion constraint. The optimal estimator differs depending on the availability of the source statistics at the receiver. We study the best linear unbiased estimator (BLUE) and the minimum mean square error (MMSE) estimator.

A. Analog Transmission with Best Linear Unbiased Estimation

Assume that the source statistics, e.g., the correlation matrix, is not available to the receiver. Since MMSE estimator requires knowledge of the autocorrelation matrix $\mathbf{R}_{\boldsymbol{\theta}}$, it can not be employed. In this case, the best linear unbiased estimator (BLUE) is the optimal choice and it is given by [39]

$$\hat{\boldsymbol{\theta}} = \left[\mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H}\right]^{-1} \mathbf{H} \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{y}.$$

The mean-squared error for this estimator can be obtained as

$$\begin{aligned} \mathbf{R}_{\boldsymbol{\epsilon}} &= \left[\mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H}\right]^{-1} \\ &= \text{diag}\{\sigma_1^2 + \frac{\xi_1^2}{\alpha_1 g_1}, \ldots, \sigma_K^2 + \frac{\xi_K^2}{\alpha_K g_K}\} \end{aligned}$$

[†]The probability density is given by $f_{\mathbf{g}_k}(g) = 2ge^{-g^2}$.

[‡]The matrix inequality $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) denotes that $\mathbf{A} - \mathbf{B}$ is a nonnegative (positive) definite matrix.

We will solve the optimum power allocation problem for the two distortion criteria discussed above.

1) Individual Power Constraint: Let us denote the average power of the observed signal at node k by $W_k^2 = [\mathbf{R}_{\boldsymbol{\theta}}]_{k,k} + \sigma_k^2$. Then, the transmit power at node k is $P_k = W_k^2 \alpha_k$ (α_k will be determined soon), and the minimum-energy power allocation problem based on the BLUE can be expressed as

$$\begin{array}{cc} \min & \sum_{k=1}^K W_k^2 \alpha_k \\ \text{s.t.} & \sigma_k^2 + \frac{\xi_k^2}{g_k \alpha_k} \leq d_k,\ k = 1, \ldots, K \end{array}$$

for $\alpha_k \geq 0,\ k = 1, \ldots, K$. By defining $r_k = \sigma_k^2 + \frac{\xi_k^2}{g_k \alpha_k}$, we obtain a convex optimization problem over r_k:

$$\begin{array}{cc} \min & \sum_{k=1}^K W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} \\ \text{s.t.} & r_k \leq d_k,\ r_k \geq \sigma_k^2,\ k = 1, \ldots, K. \end{array}$$

for $r_k \geq 0,\ k = 1, \ldots, K$. (Note that the objective function and the constraints are all convex.) The Lagrangian cost function is given by

$$J(\lambda_1, \ldots, \lambda_K, \mu_1, \ldots, \mu_K, \mathbf{r}) = \sum_{k=1}^K W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} + \sum_{k=1}^K \lambda_k(\sigma_k^2 - r_k) + \sum_{k=1}^K \mu_k (r_k - d_k)$$

and the KKT conditions follow as

$$-W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)^2} - \lambda_k + \mu_k = 0 \quad (6)$$

$$\mu_k (r_k - d_k) = 0,\ \lambda_k(-r_k + \sigma_k^2) = 0 \quad (7)$$

$$r_k - d_k \leq 0,\ -r_k + \sigma_k^2 \leq 0 \quad (8)$$

for $k = 1, \ldots, K$. Since $-r_k + \sigma_k^2$ has to be strictly negative (otherwise, the total power becomes infinity), we have $\lambda_k = 0$, $k = 1, \ldots, K$. By simple algebra, we find the optimal power allocation coefficients

$$\alpha_k = \frac{\xi_k^2}{g_k(d_k - \sigma_k^2)} \quad (9)$$

Note that d_k lies in the interval (σ_k^2, ∞) because of observation noise.

2) Average Distortion Constraint: The minimum-energy power allocation problem for this case can be expressed as

$$\begin{array}{cc} \min & \sum_{k=1}^K W_k^2 \alpha_k \\ \text{s.t.} & \frac{1}{K} \sum_{k=1}^K \sigma_k^2 + \frac{\xi_k^2}{g_k \alpha_k} \leq d_s \end{array}$$

for $\alpha_k \geq 0,\ k = 1, \ldots, K$. Using $r_k = \sigma_k^2 + \frac{\xi_k^2}{g_k \alpha_k}$ we have the equivalent problem

$$\begin{array}{cc} \min & \sum_{k=1}^K W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} \\ \text{s.t.} & \frac{1}{K} \sum_{k=1}^K r_k \leq d_s,\ r_k \geq \sigma_k^2, k = 1, \ldots, K \end{array}$$

for $r_k \geq 0,\ k = 1, \ldots, K$. The Lagrangian for this problem is given by

$$J(\lambda_1, \ldots, \lambda_K, \mu, \mathbf{r}) = \sum_{k=1}^K W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)} + \sum_{k=1}^K \lambda_k(\sigma_k^2 - r_k) + \mu \left(\frac{1}{K} \sum_{k=1}^K r_k - d_s \right)$$

and the KKT conditions follow as

$$-W_k^2 \frac{\xi_k^2}{g_k(r_k - \sigma_k^2)^2} - \lambda_k + \frac{\mu}{K} = 0 \quad (10)$$

$$\mu \left(\frac{1}{K} \sum_{k=1}^K r_k - d_s \right) = 0,\ \lambda_k(-r_k + \sigma_k^2) = 0 \quad (11)$$

$$\frac{1}{K} \sum_{k=1}^K r_k - d_s \leq 0,\ -r_k + \sigma_k^2 \leq 0 \quad (12)$$

for $k = 1, \ldots, K$. As in the previous case, $-r_k + \sigma_k^2$ has to be strictly negative, so $\lambda_k = 0$. After some manipulations, we have

$$\alpha_k^{\text{opt}} = \frac{\sqrt{\frac{\xi_k^2}{W_k^2 g_k}} \sum_{k=1}^K \sqrt{\frac{W_k^2 \xi_k^2}{g_k}}}{K d_s - \sum_{k=1}^K \sigma_k^2} \quad (13)$$

We note that the estimation problem described here is different from the one considered in [31]. In [31], the data observed at each sensor node is assumed to be exactly the same while here we consider spatially varying data. As a result, the optimized power allocations for these two problems have different solutions. In [31], the optimal power allocation might result in turning off some of the sensors, while in the case of spatially varying data, each sensor has to transmit its observation with power proportional to the inverse of square root of the channel SNR, where $SNR = W_k^2 g_k / \xi_k^2$.

B. Analog Transmission with Minimum Mean Square Error Estimation

We now assume that $\mathbf{R}_\theta$ is known at the receiver and therefore, we can use MMSE estimator.

1) Individual Distortion Constraint: First, we assume that the estimation error for the sample observed at node k is constrained to be no more than d_k, $k = 1, \ldots, K$. The MMSE estimation for $\boldsymbol{\theta}$ in (1) is given by [39]

$$\hat{\boldsymbol{\theta}} = \mathbf{R}_{\boldsymbol{\theta}} \mathbf{H}^T (\mathbf{H} \mathbf{R}_{\boldsymbol{\theta}} \mathbf{H}^T + \mathbf{R}_{\mathbf{v}})^{-1} \mathbf{y}$$

and the minimum mean-squared error covariance matrix for this estimator is given by

$$\begin{aligned} \mathbf{R}_{\boldsymbol{\epsilon}} &= \mathbf{R}_{\boldsymbol{\theta}} - \mathbf{R}_{\boldsymbol{\theta}} \mathbf{H}^T (\mathbf{H} \mathbf{R}_{\boldsymbol{\theta}} \mathbf{H}^T + \mathbf{R}_{\mathbf{v}})^{-1} \mathbf{H} \mathbf{R}_{\boldsymbol{\theta}} \quad (14) \\ &= \left(\mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H} + \mathbf{R}_{\boldsymbol{\theta}}^{-1} \right)^{-1} \end{aligned}$$

where $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, and the second equality follows by using the Matrix Inversion Lemma. We can express the power optimization problem as follows:

$$\begin{array}{cc} \min & \sum_{k=1}^K W_k^2 \alpha_k \\ \text{s.t.} & [\mathbf{R}_{\boldsymbol{\epsilon}}]_{k,k} \leq d_k,\ k = 1, \ldots, K. \end{array} \quad (15)$$

The optimization in (15) finds the power gain allocations that result in minimum total transmit power such that a maximum distortion level of d_k is allowed for node k, $k = 1, \ldots, K$. Let us define the $K \times 1$ vector $\mathbf{e}_j = [e_1, \ldots, e_K]$, $j = 1, \ldots, K$, such that $e_k = 1$ for $k = j$ and $e_k = 0$ for $k \neq j, k = 1, \ldots, K$. Then, $(i,j)^{\text{th}}$ entry of $\mathbf{A}$ can be expressed as $[\mathbf{A}]_{i,j} = \mathbf{e}_i^T \mathbf{A} \mathbf{e}_j$. Thus, we can rewrite (15) as

$$\begin{array}{ll} \min & \sum_{k=1}^{K} W_k^2 \alpha_k \\ \text{s.t.} & \mathbf{e}_k^T \left[(\mathbf{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \right] \mathbf{e}_k \le d_k, \; k = 1, \ldots, K \end{array}$$

where $\mathbf{\Gamma} = \mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H} = \text{diag}\left(\frac{\alpha_1 g_1}{\alpha_1 g_1 \sigma_1^2 + \xi_1^2}, \ldots, \frac{\alpha_K g_K}{\alpha_K g_K \sigma_K^2 + \xi_K^2} \right)$. Let us define

$$r_k = \frac{\alpha_k g_k}{\alpha_k g_k \sigma_k^2 + \xi_k^2}$$

and make a change of variable to obtain an equivalent optimization problem over $r_1, \ldots, r_K$:

$$\begin{array}{ll} \min & \sum_{k=1}^{K} \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{r_k}{1 - r_k \sigma_k^2} \right) \\ \text{s.t.} & \mathbf{e}_k^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{e}_k \le d_k, \; 0 \le r_k < \frac{1}{\sigma_k^2}, \; k = 1, \ldots, K \end{array} \tag{16}$$

where $\mathbf{R} = \text{diag}\{r_1, \ldots, r_k\}$.

Observing that $\mathbf{R}$ and $\mathbf{R}_{\boldsymbol{\theta}}^{-1}$ are both symmetric positive definite matrices, their sum is also a symmetric positive definite matrix. Therefore, all K constraints in (16) defines a convex set over $r_1, \ldots, r_K$ [40]. Since the objective function is also a convex function, the optimization problem in (16) is convex and it can be solved by Lagrange multipliers method. The Lagrangian cost function is given by

$$J(\lambda_1, \ldots, \lambda_K, \mathbf{R}) = \sum_{k=1}^{K} \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{r_k}{1 - r_k \sigma_k^2} \right) + \sum_{l=1}^{K} \lambda_l \left(\mathbf{e}_l^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{e}_l - d_l \right) + \sum_{k=1}^{K} \mu_k (r_k - \frac{1}{\sigma_k^2}) - \sum_{k=1}^{K} \gamma_k r_k.$$

Using $\frac{\partial \mathbf{X}^{-1}}{\partial [\mathbf{X}]_{r,s}} = -\mathbf{X}^{-1} \mathbf{E}_{rs} \mathbf{X}^{-1}$ and $\mathbf{e}_r \mathbf{e}_s^T = \mathbf{E}_{rs}$, we obtain the gradient of the Lagrangian as

$$\frac{\partial J(\lambda_1, \ldots, \lambda_K, \mathbf{R})}{\partial r_k} = \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{1}{1 - r_k \sigma_k^2} \right)^2 - \sum_{l=1}^{K} \lambda_l \left[\mathbf{e}_l^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{e}_k \right]^2 + \mu_k - \gamma_k \tag{17}$$

Thus, we have the following KKT conditions:

$$W_k^2 \frac{\xi_k^2}{g_k} \left(\frac{1}{1 - r_k \sigma_k^2} \right)^2 - \sum_{l=1}^{K} \lambda_l \left[\mathbf{e}_l^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{e}_k \right]^2 + \mu_k - \gamma_k = 0, \tag{18}$$

$$\mathbf{e}_k^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \mathbf{e}_k = d_k, \; \mu_k (r_k - 1/\sigma_k^2) = 0, \; \gamma_k (-r_k) = 0 \tag{19}$$

$$r_k - 1/\sigma_k^2 \le 0, \; r_k \ge 0 \tag{20}$$

for $k = 1, \ldots, K$. A closed form expression for this problem is not tractable, but we can resort to numerical techniques to solve for $r_1, \ldots, r_K$. A simple and straightforward solution exists for the case where local observations are independent, i.e., $\mathbf{R}_{\boldsymbol{\theta}}$ is diagonal. For a general correlation model, we present several numerical results in Section IV.

Example 1: *Independent Observations:* Let us assume that the source samples at the sensor nodes are mutually independent from each other, i.e., $\mathbf{R}_{\boldsymbol{\theta}} = \text{diag}(\chi_1^2, \ldots, \chi_K^2)$, where χ_k^2 is the variance of the source sample at node k, $k = 1, \ldots, K$. After some manipulations, the KKT conditions in (18)-(20) can be simplified to

$$W_k^2 \frac{\xi_k^2}{g_k} \left(\frac{1}{1 - r_k \sigma_k^2} \right)^2 = \frac{\lambda_k}{(r_k + \chi_k^{-2})^2}, \; k = 1, \ldots, K \tag{21}$$

$$\frac{1}{r_k + \chi_k^{-2}} = d_k, \; k = 1, \ldots, K. \tag{22}$$

From (22), we have

$$r_k^{\text{opt}} = \left(\frac{1}{d_k} - \frac{1}{\chi_k^2} \right)^+$$

for $d_k \le \chi_k^2$, and

$$\alpha_k^{\text{opt}} = \frac{\xi_k^2 r_k^{\text{opt}}}{g_k (1 - r_k^{\text{opt}} \sigma_k^2)}$$

Note that because of the observation noise, d_k is lower bounded by $\frac{\sigma_k^2 \chi_k^2}{\sigma_k^2 + \chi_k^2}$, $k = 1, \ldots, K$. ■

2) Average Distortion Constraint: If we wish to satisfy an average square error distortion across all sensor nodes, we can express the power optimization problem as follows:

$$\begin{array}{ll} \min & \sum_{k=1}^{K} W_k^2 \alpha_k \\ \text{s.t.} & \frac{1}{K} \text{tr}(\mathbf{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \le d_s \end{array}$$

where $\mathbf{\Gamma} = \mathbf{H}^T \mathbf{R}_{\mathbf{v}}^{-1} \mathbf{H}$, which can be rewritten as

$$\begin{array}{ll} \min & \sum_{k=1}^{K} \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{r_k}{1 - r_k \sigma_k^2} \right) \\ \text{s.t.} & \frac{1}{K} \text{tr}(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \le d_0, \; 0 \le r_k < \frac{1}{\sigma_k^2} \end{array} \tag{23}$$

where $\mathbf{R} = \text{diag}\{r_1, \ldots, r_k\}$, and $r_k = \frac{\alpha_k g_k}{\alpha_k g_k \sigma_k^2 + \xi_k^2}$, $k = 1, \ldots, K$.

Noting that $\text{tr}(\mathbf{X}^{-1})$ is convex over the set of symmetric positive definite matrices [40] and observing that $\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1} \succ 0$, (e.g., the sum of two symmetric positive definite matrices is also a symmetric positive definite matrix), we conclude that the constraint in (23) defines a convex set. Thus, we can use the Lagrangian method where the cost function is given by

$$J(\lambda_0, \mathbf{R}) = \sum_{k=1}^{K} \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{r_k}{1 - r_k \sigma_k^2} \right) + \lambda_0 \left(\frac{1}{K} \text{tr}(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} - d_s \right) + \sum_{k=1}^{K} \mu_k (r_k - \frac{1}{\sigma_k^2}) - \sum_{k=1}^{K} \gamma_k r_k.$$

Using $\frac{\partial}{\partial [\mathbf{X}]_{r,s}} \text{tr}(\mathbf{X}) = \text{tr} \frac{\partial \mathbf{X}}{\partial [\mathbf{X}]_{rs}}$, we obtain the gradient of the Lagrangian as

$$\frac{\partial J(\lambda_0, \mathbf{R})}{\partial r_k} = \frac{W_k^2 \xi_k^2}{g_k} \left(\frac{1}{1 - r_k \sigma_k^2} \right)^2 - \lambda_0 \frac{1}{K} \left[\mathbf{e}_k^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-2} \mathbf{e}_k \right] + \mu_k - \gamma_k \tag{24}$$

for $k = 1, \ldots, K$. Thus, we have the following KKT conditions:

$$W_k^2 \frac{\xi_k^2}{g_k} \left(\frac{1}{1 - r_k \sigma_k^2} \right)^2 - \lambda_0 \frac{1}{K} \left[\mathbf{e}_k^T (\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-2} \mathbf{e}_k \right] = 0, \; k = 1, \ldots, K \tag{25}$$

$$\frac{1}{K} \text{tr}(\mathbf{R} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} = d_s, \; \mu_k (r_k - 1/\sigma_k^2) = 0, \; \gamma_k (-r_k) = 0 \tag{26}$$

$$r_k - 1/\sigma_k^2 \le 0, \; r_k \ge 0 \tag{27}$$

for $0 \le r_k \le 1/\sigma_k^2$, $k = 1, \ldots, K$. One can solve for the unknowns with numerical techniques. If the local observations are independent, i.e., $\mathbf{R}_{\boldsymbol{\theta}}$ is diagonal, the optimization in (23) assumes a closed form solution as shown in the example below. For a general correlation model, we resort to numerical techniques.

Example 2: *Independent Observations:* Let $\mathbf{R}_{\boldsymbol{\theta}} = \mathrm{diag}(\chi_1^2, \ldots, \chi_K^2)$, and without loss of generality, assume that $\frac{W_1\xi_1}{\sqrt{g_1}\chi_1^2} \leq \ldots \leq \frac{W_K\xi_K}{\sqrt{g_K}\chi_K^2}$. We define

$$A(J) = \frac{Kd_s - \mathrm{tr}(\mathbf{R}_{\boldsymbol{\theta}}) + \sum_{j=1}^{J} \frac{\chi_j^4}{\chi_j^2+\sigma_j^2}}{\sum_{j=1}^{J} \frac{\chi_j^2 W_j \xi_j}{(\chi_j^2+\sigma_j^2)\sqrt{g_j}}}$$

and

$$f(J) = \frac{W_J\xi_J}{\sqrt{g_J}\chi_J^2} A(J),$$

and then determine the unique J_1 such that $f(J_1) \leq 1$ and $f(J_1+1) > 1$. Simplifying the KKT conditions in (25) and (26) and solving for r_j, we finally arrive at

$$r_j^{\text{opt}} = \begin{cases} \frac{1 - \frac{W_j\xi_j}{\sqrt{g_j}\chi_j^2} A(J_1)}{\sigma_j^2 + \frac{W_j\xi_j}{\sqrt{g_j}} A(J_1)} & j = 1, \ldots, J_1 \\ 0 & j = J_1+1, \ldots, K \end{cases} \quad (28)$$

Therefore, we have the optimal power gains

$$\alpha_j^{\text{opt}} = \frac{\xi_j^2}{g_j} \frac{r_j^{\text{opt}}}{1 - r_j^{\text{opt}}\sigma_j^2} \quad (29)$$

Note that for this case, some of the sensor nodes may be shut down depending on the noise variances and channel gains. ■

A power-distortion region, in a similar fashion to the theoretically optimum power distortion region described in Appendix A, can also be defined for the uncoded transmission system using the MMSE at the fusion center. For example, for the average distortion characterization, we have

$$\mathcal{P}^{\text{MMSE}}(d_s) = \left\{ P_1, \ldots, P_K : d_s \geq \frac{1}{K}\mathrm{tr}(\boldsymbol{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1})^{-1} \right\} \quad (30)$$

where $P_k = \alpha_k([\mathbf{R}_{\boldsymbol{\theta}}]_{kk} + \sigma_k^2)$, $k = 1, \ldots, K$. In Section IV, several comparisons between the power-distortion regions achieved by different schemes are provided. We note that if the observations at different sensor nodes are uncorrelated Gaussian random variables, it is easy to show that $\mathcal{P}^{\text{MMSE}}(d_s) = \mathcal{P}(d_s)$, i.e., using the uncoded forwarding strategy and employing the MMSE at the receiver achieves the theoretically optimum performance.

C. Asymptotic Performance for Large Networks

We next study asymptotic performance of analog transmission in large networks, e.g., $K \to \infty$. For tractable analysis, we assume independent and identically distributed observation samples e.g., $\mathbf{R}_\theta = \mathrm{diag}(\chi^2, \ldots, \chi^2)$, identical observation noise variances, e.g., $\sigma_k^2 = \sigma^2$, and identical channel noise variances, e.g., $\xi_k^2 = \xi^2$, for $k = 1, \ldots, K$. We define by $\bar{P}(d)$ the minimum transmit power per sensor node, e.g., $\bar{P}(d) = \frac{1}{K}\sum_k P_k(d)$, that is required to attain a distortion $d_k = d$, $k = 1, \ldots, K$ at each node, or an average distortion $d = \frac{1}{K}\sum_k d_k$ across all nodes. We can show that with BLUE, we have $\bar{P}(d)$ for individual and average power constraints as, respectively,

$$\bar{P}_{\text{individual}}(d) = \frac{(\chi^2+\sigma^2)\xi^2}{d-\sigma^2} E_{\mathbf{G}}\left\{\frac{1}{g}\right\} = \frac{(\chi^2+\sigma^2)\xi^2}{d-\sigma^2}\sqrt{\pi} \quad (31)$$

$$\bar{P}_{\text{average}}(d) = \frac{(\chi^2+\sigma^2)\xi^2}{d-\sigma^2}\left(E_{\mathbf{G}}\left\{\frac{1}{\sqrt{g}}\right\}\right)^2 = \frac{(\chi^2+\sigma^2)\xi^2}{d-\sigma^2} 1.5 \quad (32)$$

where $E_{\mathbf{G}}\{\cdot\}$ denotes the expectation over the distribution of the fading coefficients. If we use MMSE estimator at the receiver, we have

$$\bar{P}_{\text{individual}}(d) = \frac{(\chi^2+\sigma^2)\xi^2}{\frac{\chi^2 d}{\chi^2 - d} - \sigma^2}\sqrt{\pi} \quad (33)$$

$$\bar{P}_{\text{average}}(d) = \frac{J_1^2\chi^4\xi^2}{K^2(d-\chi^2)(\chi^2+\sigma^2) + J_1K\chi^4}\left(\frac{1}{J_1}\sum_{j=1}^{J_1}\frac{1}{\sqrt{g_j}}\right)^2 - \frac{J_1\xi^2}{K}\left(\frac{1}{J_1}\sum_{j=1}^{J_1}\frac{1}{g_j}\right) \quad (34)$$

where the $\bar{P}_{\text{average}}(d)$ in (34) can be calculated using Monte Carlo simulation.

IV. Numerical Examples

In this section, we present numerical examples for the power allocation problems and the achievable power-distortion regions studied in the paper. The power allocation problem with the MMSE does not allow for a closed form expression, so we resort to numerical techniques to solve for optimal r_k and then find the optimal α_k, $k = 1, \ldots, K$. The optimization for the distributed BLUE has a closed from solution and the optimum power gains are given by (9) or (13).

We consider the spatial correlation model defined by a correlation matrix

$$[\mathbf{R}_{\boldsymbol{\theta}}]_{i,j} = \rho^{|j-i|}, \ \rho < 1. \quad (35)$$

This matrix has a symmetric tridiagonal inverse that can be computed by

$$[\mathbf{R}_{\boldsymbol{\theta}}^{-1}]_{i,j} = \begin{cases} \frac{1}{1-\rho^2} & i = j = 1, \ K \\ \frac{1+\rho^2}{1-\rho^2} & 2 \leq i = j \leq K-1 \\ \frac{-\rho}{1-\rho^2} & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases} \quad (36)$$

Hence, the matrix $\boldsymbol{\Gamma} + \mathbf{R}_{\boldsymbol{\theta}}^{-1}$ in (23) also has a symmetric tridiagonal structure and can be analytically inverted by the method described in [41]. The analytic evaluation of the gradient in (17) and (24) improves the accuracy of the optimization.

In Figures 2 and 3, for a network with two sensors, we depict the power-distortion regions that can be achieved by the theoretically optimum system and the uncoded system with the MMSE estimator and the BLUE under various observation and channel noise variances. In Figure 2, we consider the average individual distortion criterion. For the fading scenario, we study two cases: (i) both links have equal gain (2.a and 2.c), and (ii) the link to the fusion center for one of the sensor nodes is 10 dB worse than that of the other one. From the plots, we observe that the MMSE estimator performs significantly better than the BLUE estimator for all ρ values. It is also clear that $P(d_s) \supseteq P^{\text{MMSE}}(d_s) \supseteq P^{\text{BLUE}}(d_s)$. For higher correlations, e.g., $\rho = 0.9$, or $\rho = 0.99$, we observe that it possible to satisfy an average distortion level if the transmit power for one of the sensor node is sufficiently large, and this can be achieved by an uncoded transmission with the MMSE at the receiver. This implies that by exploiting the correlation between the observations we can estimate one from the other to attain a

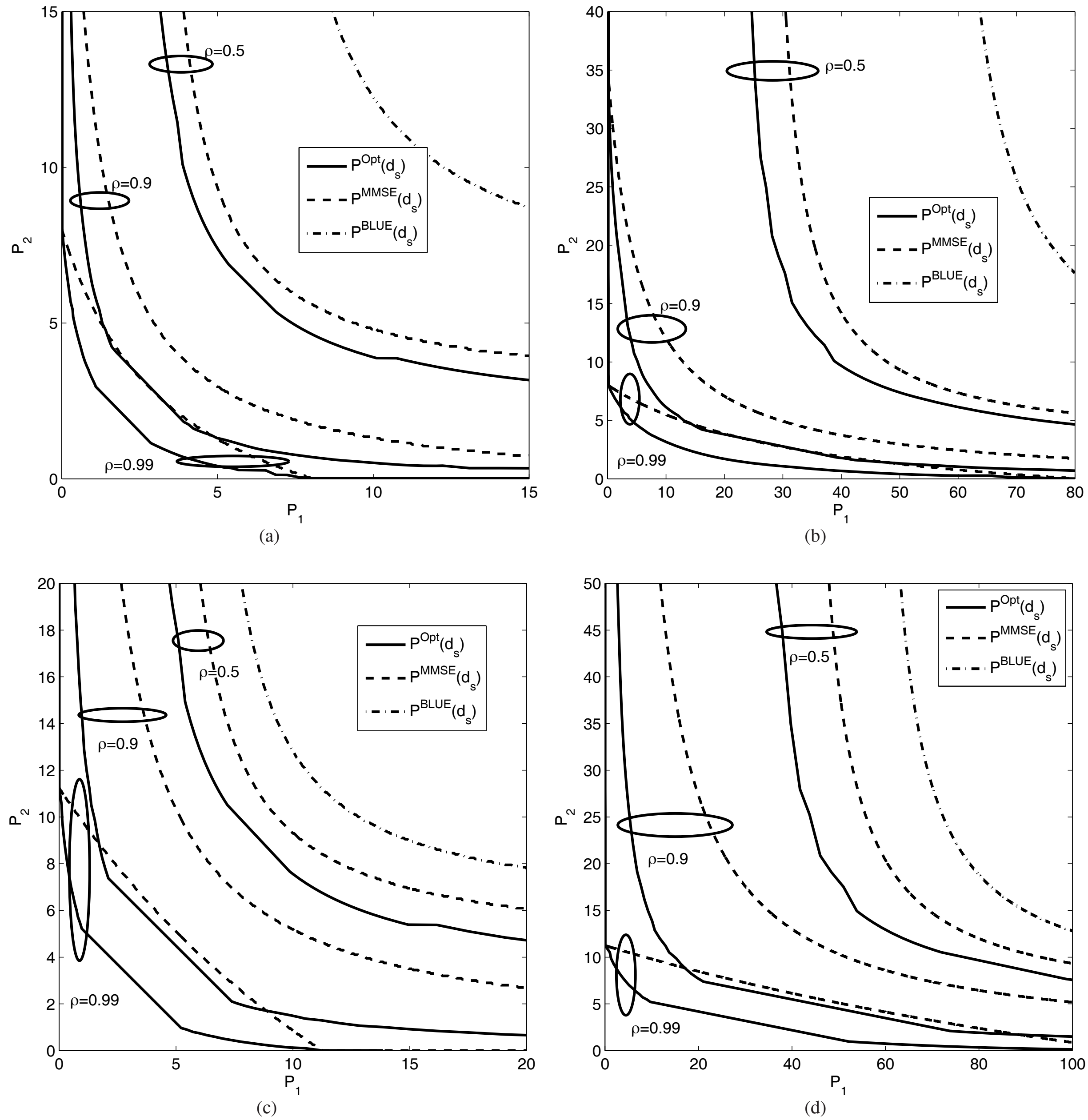


Fig. 2. Comparison of power-distortion regions attained by the theoretically optimum system and the uncoded transmission for the average distortion case. Simulation parameters are $K = 2$, $\rho = 0.5, 0.9$ and 0.99, a) $\sigma_1^2 = \sigma_2^2 = 0.1$, $d_s = 0.2$, $g_1 = g_2 = 1$, b) $\sigma_1^2 = \sigma_2^2 = 0.1$, $d_s = 0.2$, $g_1 = 0.1$, $g_2 = 1$ c) $\sigma_1^2 = \sigma_2^2 = 0.01$, $d_s = 0.1$, $g_1 = g_2 = 1$ d) $\sigma_1^2 = \sigma_2^2 = 0.01$, $d_s = 0.1$, $g_1 = 0.1$, $g_2 = 1$

certain average distortion level. Furthermore, it is also seen that for larger values of ρ, it is possible to attain a prescribed distortion level with less power.

In Figure 3, we depict power-distortion regions for the same system assuming individual distortion levels of $\mathbf{d} = [0.2\ 0.2]$. For this case, we also have $P(\mathbf{d}) \supseteq P^{\text{MMSE}}(\mathbf{d}) \supseteq P^{\text{BLUE}}(\mathbf{d})$. The power-distortion region exhibits similar properties as those in the average distortion case as the correlation levels or link qualities is varied.

Next, we study the total power (normalized by the number of sensor nodes) versus reconstruction quality performance. In Figure 4, we study $K = 2$ sensor systems and compare the MMSE estimation and BLUE against the theoretically optimum system performance. Curves in Figure 4.a and Figure 4.b are for the average distortion and individual distortion constraints, respectively. In the latter case, we assume $d_1 = d_2$. The performance is averaged over Rayleigh fading using Monte Carlo simulation. It is seen that performance in both cases are very close to each other. The MMSE estimator performance is always superior to the BLUE and the improvement is larger especially at higher source correlation. Furthermore, at lower source SNR, i.e., 10 dB, the MMSE estimation performance is about 10 dB better. The theoretically optimum system performance is about 3.4 dB better than that of the MMSE when $\rho = 0.99$ and source SNR is 20 dB. The gap is about 2.4 dB when $\rho = 0.9$.

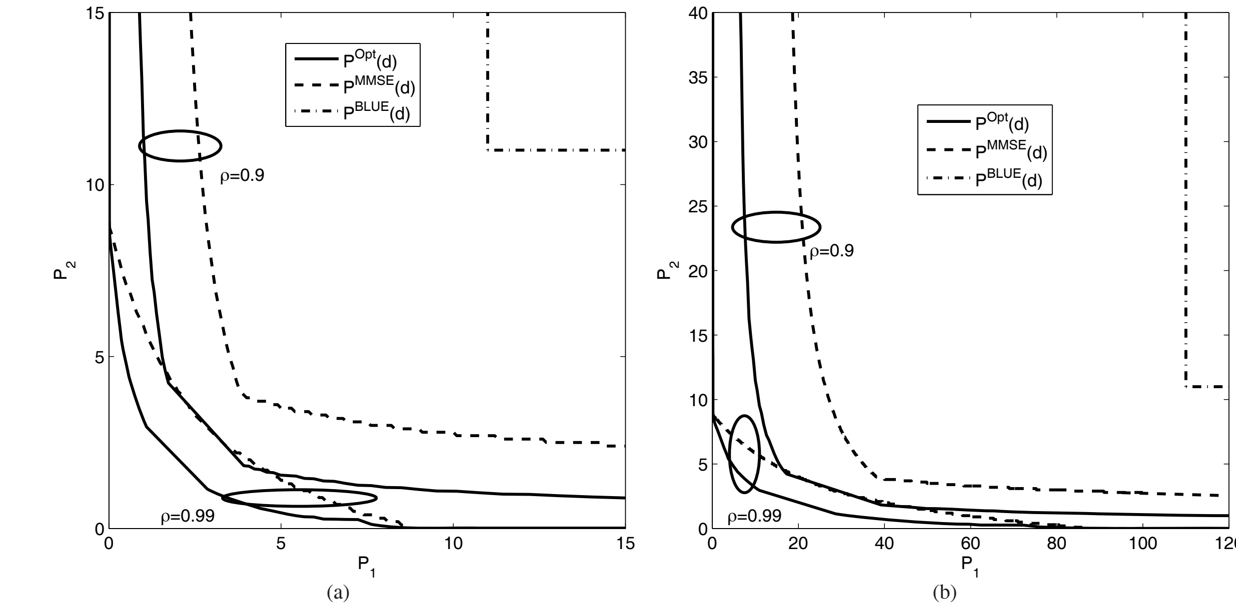


Fig. 3. Comparison of power-distortion regions attained by the theoretically optimum system and the uncoded transmission for the individual distortion case. Simulation parameters are $K = 2$, $\sigma_1^2 = \sigma_2^2 = 0.1$, $\mathbf{d} = [0.2\ 0.2]$, $\rho = 0.9$ and 0.99, a) $g_1 = g_2 = 1$, b) $g_1 = 0.1$, $g_2 = 1$

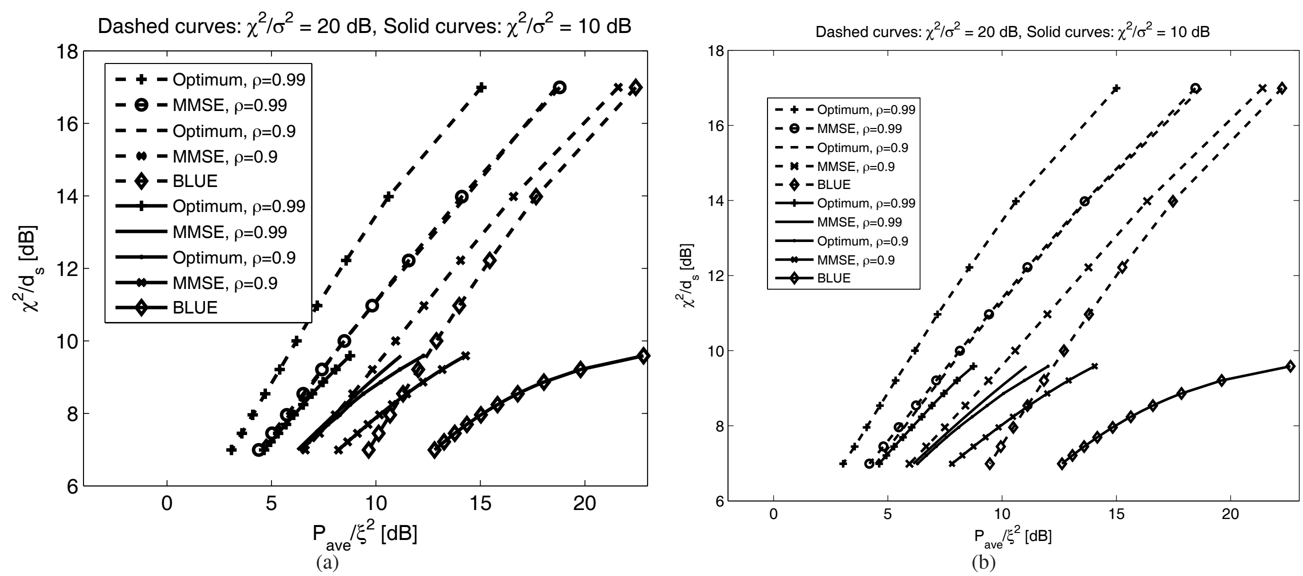
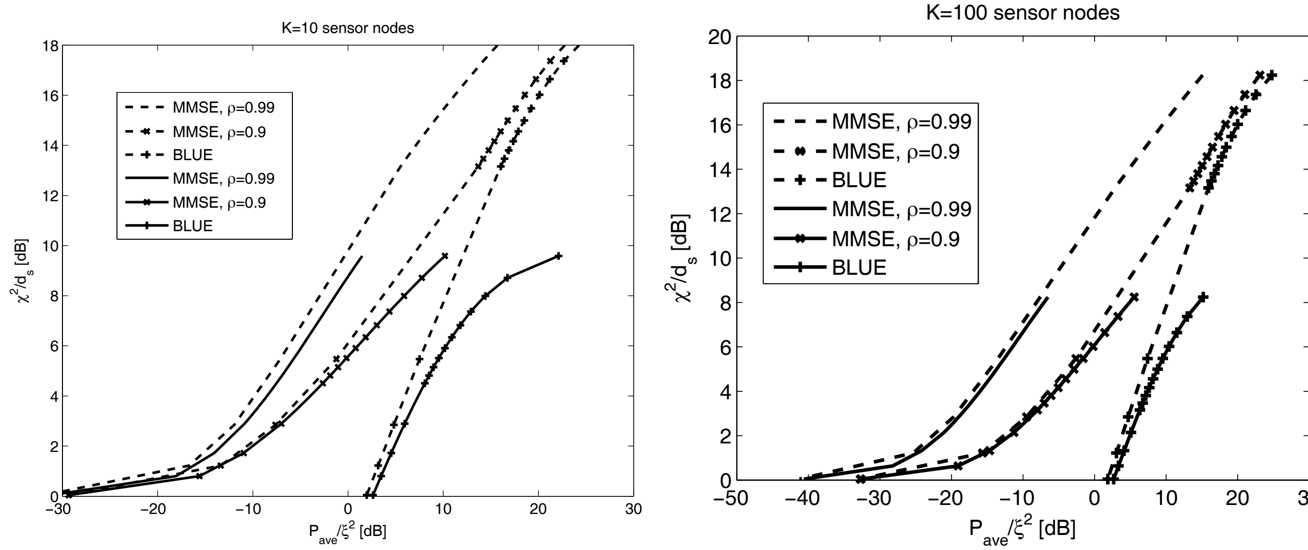


Fig. 4. Average channel SNR per sensor node vs. reconstruction SNR attained by the theoretically optimum system and the uncoded transmissions. Simulation parameters are $K = 2$, $\chi_1^2/\sigma_1^2 = \chi_2^2/\sigma_2^2 = 20$ dB for dashed curves, $\chi_1^2/\sigma_1^2 = \chi_2^2/\sigma_2^2 = 10$ dB for solid curves, $\rho = 0.9$ and 0.99, a) Average distortion constraint b) Individual distortion constraint such that $d_1 = d_2 = d$. Performance in both cases are close to each other.

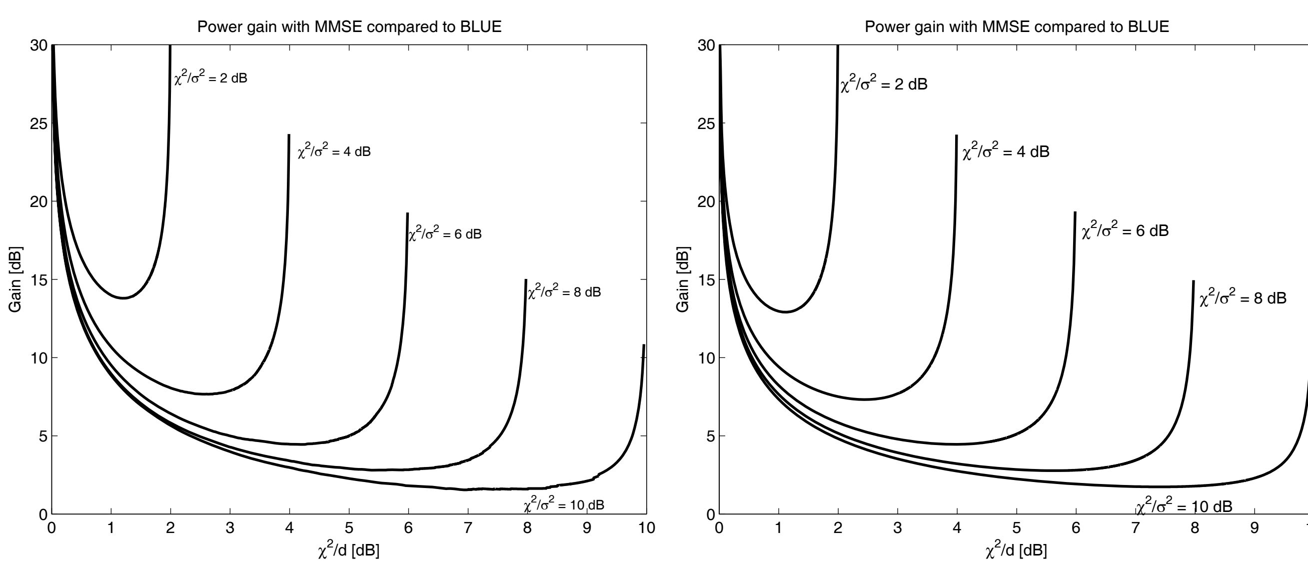
Figures 5.a and 5.b study larger networks with 10 and 100 sensor nodes, respectively. In this case, it is formidable to compute the theoretically optimum performance, therefore, we study the performance of analog transmission only. In Figure 5, we consider a network of $K = 10$ (in 5.a) and $K = 100$ (in 5.b) sensor nodes. We plot the reconstruction SNR against the average power per node required (normalized by the channel noise). It is seen that for both sizes of the network, we have very similar performance trend. In the figures, the dashed curves are for observation SNR of 20 dB at each node, while the solid curves are for 10 dB observation SNR. In particular, for the observation SNR of 10 dB, the performance of MMSE estimator is again about 10 dB better than that of the MMSE when $\rho = 0.9$ while the relative gain with MMSE compared to the BLUE increases to more than 20 dB when $\rho = 0.99$. The gains with MMSE become less for higher observation SNR, however, as the correlations become higher, we observe much better performance with MMSE. This suggests we better utilize the correlation structure of the observed field.

In the last example, we consider the asymptotic performance of analog estimation assuming that the observations are independent identical and the channels from sensor nodes to the fusion center is Rayleigh fading (See Section III-C). Figure 6 plots the relative gain between the MMSE estimation



(a) (b)

Fig. 5. Power vs. reconstruction SNR performance comparison for distributed estimation of a correlated field with optimal power allocation with the MMSE and the BLUE. Simulation parameters: $\rho = 0.9$ and 0.99, dashed curves: $\chi^2/\sigma_k^2 = 20$ dB, solid curves: $\chi^2/\sigma_k^2 = 10$ dB. Average power per node is normalized by channel noise variance. a) $K = 10$ sensor nodes, b) K=100 sensor nodes



(a) (b)

Fig. 6. Power gain with MMSE estimator compared to the BLUE for large networks when observations are interdependent. a) Average distortion criterion d, b) same individual distortion criterion d for each node

and the BLUE estimation with optimal power allocations (for both average and individual distortion criteria). Note that, the analog transmission with MMSE estimation is also the theoretically optimum system for this case. Again, the MMSE estimation is superior. The gain is more if we wish to attain very good reconstruction quality, e.g., the gains are more than 10 dB if a reconstruction quality close to the observation quality at the sensor nodes is desired.

V. Conclusions

In this paper, we addressed energy-efficient estimation of correlated data in a wireless sensor network where spatially distributed sensor nodes observe independently corrupted versions of a correlated Gaussian vector source and transmit their observations over orthogonal multiple access channels to a fusion center where a final estimate is obtained. We assumed that the communication between the sensor nodes and the fusion center is subject to fading and additive Gaussian noise. For this network model, we derived the achievable power-distortion region. We considered two different distortion characterization: (i) an individual distortion measure to assess the estimation quality at individual sensor nodes, and (ii) an average distortion measure to characterize the distortion on an average basis across sensor nodes. For both distortion measures, we studied an uncoded analog transmission strategy where the noise-corrupted sensor observations are simply amplified and forwarded. Based on the availability of the correlation matrix of the Gaussian vector source, we considered two different estimation techniques: (i) the minimum mean-square error estimation, which requires knowledge of the correlation matrix, and (ii) minimum linear unbiased estimation, which does not require any statistical knowledge. For both estimation techniques, we determined the optimal power allocation scheme and the minimum required power with which one can satisfy a certain mean-squared error distortion level. Performance comparisons of the various schemes indicate that one needs to exploit the intersensor correlations for better energy efficiency. Comparison with optimal schemes indicate that as correlations between the sensor observation becomes small, the performance gap between the uncoded scheme and the theoretically optimum scheme decreases. Furthermore, performance comparisons between the MMSE and the BLUE indicate that exploiting the correlation among sensor observations (by the MMSE) reduces the required power to attain some distortion level.

In this work, we assumed that no bandwidth expansion is allowed, e.g. $\kappa = 1$. An extension to the current work includes the design and analysis of analog error-correcting techniques for coding of noisy observations at the local sensors prior to the transmission to the fusion center. Furthermore, in this paper, we assumed that there is no temporal correlation. However, in many cases, the consecutive observations at each node are likely to have correlations. One needs to exploit these correlations for a better energy performance.

Appendix A: Achievable Power-Distortion Region

The achievable power-distortion region for our problem can be determined using the methods described in [21], [23], [38] for the case where a *common source* in noise is observed at all nodes. While separation theorem is not optimal in multiple access channel, it is shown in [23] that if the multiple access is via orthogonal channels, the separate source-channel coding attains all possible power-distortion pairs. Note that although the separation theorem generally requires infinite length source and channel codes to achieve the theoretical limits, it will be useful to compare these limits with that of the proposed uncoded scheme.

We first obtain the rate-distortion region for the network shown in Figure 1 via the auxiliary Gaussian test channels [23], [28]:

$$\mathbf{z} = \mathbf{x} + \mathbf{n} \tag{37}$$

where $\mathbf{n}$ is a zero mean Gaussian noise vector with covariance matrix $\mathbf{R_n} = \text{diag}(\eta_1^2, \ldots, \eta_K^2)$ and $\mathbf{x} = \boldsymbol{\theta} + \mathbf{w}$ is the observation vector. The set of variances η_k^2, $k = 1, \ldots, K$, for which the mean square error distortion at node k, $k = 1, \ldots, K$, is less than d_k is given by

$$\Psi_M(\mathbf{d}) \triangleq \left\{ (\eta_1^2, \ldots, \eta_K^2) : d_k \geq \left[\left((\mathbf{R_n} + \mathbf{R_w})^{-1} + \mathbf{R}_{\boldsymbol{\theta}}^{-1} \right)^{-1} \right]_{k,k}, \; k = 1, \ldots, K \right\}. \tag{38}$$

Then, the rate-distortion region follows as [21], [22], [38]

$$\mathcal{R}(\mathbf{d}) = \bigcup_{(\eta_1^2, \ldots, \eta_K^2) \in \Psi_M(\mathbf{d})} \mathcal{R}(\mathbf{d}; \eta_1^2, \ldots, \eta_K^2) \tag{39}$$

such that $\mathcal{R}(\mathbf{d}; \eta_1^2, \ldots, \eta_K^2) = \left\{ (R_1, \ldots, R_K) : \sum_{i \in \mathcal{A}} R_i \geq I(X_{\mathcal{A}}; Z_{\mathcal{A}} | Z_{\mathcal{A}^c}), \forall \mathcal{A} \subseteq \mathcal{I}_K \right\}$ where $\mathcal{I}_K = \{1, \ldots, K\}$, $\mathcal{A}$ denotes subsets of $\mathcal{I}_K$, $\mathcal{A}^c$ is the complementary set of $\mathcal{A}$, and $X_{\mathcal{A}}$ and $I(\cdot)$ denote the set $\{X_k : k \in \mathcal{A}\}$ and the mutual information, respectively. $X_{\mathcal{A}}$ and $Z_{\mathcal{A}}$ denote the random variables whose realizations are given described by (37). Note that for the special case of $K = 2$ sensors, we can show that setting $\sigma_k^2 = 0$, $k = 1, \ldots, K$, e.g., $\mathbf{R_w} = 0$, the rate-distortion region given by (39) reduces to the region given by [16]. The rate-distortion region in (39) provides the minimal required rate-tuples $(R_1, \ldots, R_k)$ to achieve distortion $\mathbf{d}$ regardless of the channels between the sensor nodes and the fusion center. As long as the capacity of the channels are greater than the required rates, one can attain $\mathbf{d}$ by using sufficiently long source codes and channel codes.

Combining the rate-distortion region with the Shannon channel capacity for each sensor node (assuming a zero-mean unit-variance additive white Gaussian channel noise), $R_k \leq \frac{1}{2} \log(1 + P_k g_k)$, we obtain the power-distortion region

$$\mathcal{P}(\mathbf{d}) = \bigcup_{(\eta_1^2, \ldots, \eta_K^2) \in \Psi_M(\mathbf{d})} \mathcal{P}(\mathbf{d}; \eta_1^2, \ldots, \eta_K^2) \tag{40}$$

where

$$\mathcal{P}(\mathbf{d}; \eta_1^2, \ldots, \eta_K^2) = \left\{ (P_1, \ldots, P_K) : \prod_{i \in \mathcal{A}} (1 + P_k g_k) \geq \left(\frac{|\mathbf{R}_{X_{\mathcal{A}} Z_{\mathcal{A}^c}}| |\mathbf{R}_Z|}{|\mathbf{R}_{X_{\mathcal{A}} Z}| |\mathbf{R}_{Z_{\mathcal{A}^c}}|} \right)^{\kappa}, \forall \mathcal{A} \subseteq \mathcal{I}_K \right\} \tag{41}$$

and κ is the source/channel code rate (e.g., for uncoded transmission, $\kappa = 1$), $X_{\mathcal{A}_1} Z_{\mathcal{A}_2}$ denotes the vector formed by stacking the elements of $X_{\mathcal{A}_1}$ and $Z_{\mathcal{A}_2}$, and $\mathbf{R}_{X_{\mathcal{A}_1} Z_{\mathcal{A}_2}}$ denotes the corresponding covariance matrix. The matrix $\mathbf{R}_{X_{\mathcal{A}_1} Z_{\mathcal{A}_2}}$ can be evaluated using (37), e.g., $E\{X_i X_j\} = [\mathbf{R}_{\boldsymbol{\theta}}]_{i,j}$, $E\{X_i Z_j\} = [\mathbf{R}_{\boldsymbol{\theta}}]_{i,j} + \sigma_i^2 \delta_{i,j}$, and $E\{Z_i Z_j\} = [\mathbf{R}_{\boldsymbol{\theta}}]_{i,j} + (\sigma_i^2 + \eta_i^2) \delta_{i,j}$ †.

The obtain the achievable region under the average distortion metric, we only need to replace the region $\Psi_M(\mathbf{d})$ in (38) by $\Psi_S(d_s) \triangleq \left\{ (\eta_1^2, \ldots, \eta_K^2) : d_s \geq \frac{1}{K} \text{tr} \left((\mathbf{R_n} + \mathbf{R_w})^{-1} + \mathbf{R}_{\boldsymbol{\theta}}^{-1} \right)^{-1} \right\}$, and $\mathbf{d}$ by d_s in all subsequent equations (39)-(41)

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E.Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, pp. 102–114, Aug. 2002.

[2] 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.

[3] D. Li, K. D. Wang, Y. H. Hu, and A. M. Sayeed, "Detection, classification and tracking of targets," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 17–29, Mar. 2002.

[4] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors–part I: fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[5] H. V. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 2nd edition, 1994.

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

† $\delta_{i,j}$ is the Kronecker delta function

[7] C. E. Shannon, “A mathematical thoery of communications,” *The Bell Syst. Techn. J.*, vol. 27, pp. 379–423, 623–656, Oct. 1948.
[8] D. Slepian anf J. Wolf, “Noisless coding of correlated sources,” *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.
[9] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information,” *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.
[10] T. Berger, *Multiterminal Source Coding*, vol. 229, pp. 171–231, Springer-Verlag, 1977, CISM Courses and Lectures.
[11] S. Y. Tung, *Multiterminal source coding*, Ph.D. thesis, Cornell University, 1978.
[12] T. Berger and R. Yeung, “Multiterminal source coding with one distortion criterion,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 228–236, Mar. 1989.
[13] Y. Oohama, “Gaussian multiterminal source coding,” *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912–1923, Nov. 1997.
[14] R. Zamir, “The rate loss in the Wyner-Ziv problem,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.
[15] R. Zamir and T. Berger, “Multiterminal source coding with high resolution,” *IEEE Trans. Inform. Theory*, vol. 45, pp. 106–117, Jan. 1999.
[16] A. B. Wagner, V. Anantharam, and P. Viswanath, “The rate region of the quadratic Gaussian two-terminal source-coding problem,” submitted for publication. Available from http://www.arxiv.org/abs/cs.IT/0510095/., 2006.
[17] A. B. Wagner and V. Anantharam, “An infeasibility result for the multiterminal source-coding problem,” submitted for publication. Available from http://arxiv.org/abs/cs.IT/0511103/., 2006.
[18] S. D. Servetto, “The region of achievable rates for multiterminal source coding–part I,” submitted for publication, 2006.
[19] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem,” *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 887–902, May 1996.
[20] H. Viswanathan and T. Berger, “The quadratic Gaussian CEO problem,” *IEEE Trans. Inform. Theory*, vol. 43, no. 9, pp. 1549–1559, Sept. 1997.
[21] Y. Oohama, “The rate-distortion function for the quadratic CEO problem,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 1057–1070, May 1998.
[22] V. Prabhakaran, D. Tse, and K. Ramchandran, “Rate region of the quadratic Gaussian CEO problem,” in *Proc. International Symposium on Information Theory*, p. 199, June 2004.
[23] J.-J. Xiao and Z.-Q. Luo, “Multiterminal source-channel communication under orthogonal multiple access channels,” submitted for publication, Mar. 2005.
[24] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks–part I: Gaussian case,” *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.
[25] J.-J. Xiao and Z.-Q. Luo, “Decentralized estimation in an inhomogeneous sensing environment,” *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.
[26] Z.-Q. Luo, “Universal decentralized estimation in a bandwidth constrained sensor network,” *EEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.
[27] T. J. Goblick, “Theoretical limitations on the transmission of data from analog sources,” *IEEE Trans. Inform. Theory*, vol. 11, no. 4, pp. 558–567, Oct. 1965.
[28] R. G. Gallager, *Information Theory and Realiable Communication*. New York: John Wiley and Sons, 1968.
[29] M. Gastpar, B. Rimoldi, and M. Vetterli, “To code, or not to code: lossy source-channel communication revisited,” *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
[30] M. Gastpar and M. Vetterli, “Source-channel communication in sensor networks,” *Lecture Notes in Computer Science*, vol. 2634, pp. 162–177, Apr. 2003.
[31] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, “Energy-efficient joint estimation in sensor networks: analog vs. digital,” in *Proc. IEEE ICASSP*, Philadelphia, PA, Mar. 2005, pp. 745–748.
[32] J. Kusuma, L. Doherty, and K. Ramchandran, “Distributed compression for sensor networks,” in *Proc. 2001 International Conference on Image Processing*, Oct. 2001, vol. 1, pp. 82–85.
[33] A. Jindal and K. Psounis, “Modeling spatially-correlated sensor network data,” in *Proc. First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks 2004*, pp. 162–171.
[34] M. Gastpar and M. Vetterli, “Power, spatio-temporal bandwidth, and distortion in large sensor networks,” *J. Select. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.
[35] W. Bajwa, A. Sayeed, and R. Nowak, “Efficient communication strategies for distributed signal field estimation,” in *Proc. Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, Nov. 2004, vol. 2, pp. 1371–1375.
[36] W. Bajwa, A. Sayeed, and R. Nowak, “Matched source-channel communication for field estimation in wireless sensor network,” in *Proc. Fourth International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, Apr. 2005, pp. 332–339.
[37] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, “A universal matched source-channel communication scheme for wireless sensor ensembles,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, vol. 5, pp. 1153–1156.
[38] J. Chen, Z. Zhang, T. Berger, and S. B. Wicker, “An upper bound on the sum-rate distortion functions and its corresponding rate allocation schemes for the CEO problem,” *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.
[39] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Englewoof Cliffs, NJ: Prentice Hall PTR, 1995.
[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[41] H. A. Yamani and M. S. Abdelmonem, “The analytic inversion of any finite symmetric tridiagonal matrix,” *J. Phys. A: Math. Gen*, vol. 30, pp. 2889–2893, 1997.

Israfil Bahceci (S’00) received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 1999, the M.S. degree in electrical engineering from Arizona State University, Tempe, in 2001, and the Ph.D. degree from Georgia Institute of Technology, Atlanta, in 2005. He worked as a Postdoctoral Fellow at the University of Waterloo from 2005 to 2006. He is currently working at the Wireless Technology Laboratories of Nortel Networks, Ottawa, Canada. His current research interests include digital communications, wireless and mobile communications, channel coding, turbo codes, coding for multi-access channels, and coding for wireless communications.

Amir K. Khandani (S’93-M’02) received the M.A.Sc. degree from the University of Tehran, Tehran, Iran, in 1985 and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 1992. He was a Research Associate with INRS Telecommunications, Quebec University, Montreal, for one year. Since 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor. Dr. Khandani is currently holding an NSERC Industrial Research Chair (funded by NSERC and Nortel Networks) on “Advanced Telecommunications Technologies” and a Canada Research Chair (Tier I) on “Wireless Systems.” He is serving as an Associate Editor for the *IEEE Transactions on Communications* in the area of coding and communication theory.