Optimal Precoding Design and Power Allocation for Decentralized Detection of Deterministic Signals

Jun Fang, *Member, IEEE*, Hongbin Li, *Member, IEEE*, Zhi Chen, and Shaoqian Li

***Abstract*—We consider a decentralized detection problem in a power-constrained wireless sensor network (WSN), in which a number of sensor nodes collaborate to detect the presence of a deterministic vector signal. The signal to be detected is assumed known *a priori*. Each sensor conducts a local linear processing to convert its observations into one or multiple messages. The messages are conveyed to the fusion center (FC) by an uncoded amplify-and-forward scheme, where a global decision is made. Given a total network transmit power constraint, we investigate the optimal linear processing strategy for each sensor. Our study finds that the optimal linear precoder has the form of a matched filter. Depending on the channel characteristics, one or multiple versions of the filtered/compressed message should be reported to the FC. In addition, assuming a fixed total transmit power, we examine how the detection performance behaves with the number of sensors in the network. Analysis shows that increasing the number of sensors can substantially improve the system detection reliability. Finally, decentralized detection with unknown signals is studied and a heuristic precoding design is proposed. Numerical results are conducted to corroborate our theoretical analysis and to illustrate the performance of the proposed algorithm.**

***Index Terms*—Decentralized detection, detection outage, precoding design, wireless sensor networks.**

I. Introduction

DECENTRALIZED detection is an important problem that has attracted much attention over the past decade [1]–[20]. In a wireless sensor network (WSN), a large number of sensors are deployed in an area to monitor the environment. Each sensor makes noisy observations of a binary hypothesis on the state of the environment and transmits its data to the fusion center (FC), where a final decision regarding the state of nature is made. Due to stringent power/bandwidth constraints, each sensor needs to compress its original data before the transmission. A typical processing is to conduct a local detection at each node. The local binary decision is then sent to the FC for reaching a global decision. A large number of studies [1]–[15] were carried out in this context. A key problem that appeared in the above setting is the optimization of local decision rules such that the probability of detection error is minimized. It was shown in [2], [3], and [5] that for both Bayesian and Neyman-Pearson criteria, the optimal local sensor decision for a binary hypotheses testing problem is a likelihood ratio test (LRT). This property drastically reduces the search space for an optimal collection of local detectors [14]. Nevertheless, the search of optimal local detectors is still exponentially complex because the optimal local thresholds are generally different and need to be jointly determined along with the global fusion rule. In some other studies [16]–[18], the observations of each sensor are encoded into a real-valued summary message. The message are sent to the FC via noisy channels to form a global decision. The transmission/decision strategy, namely, which sensor should report (termed as "censoring" in [17] and [18]) and what should be transmitted, was studied by explicitly taking into account the power/rate constraints. In particular, the work [18] considered a problem formulation that admits a general class of network constraints and transmission modes.

In this paper, the problem of decentralized detection is studied under an explicit total transmit power constraint. Battery-powered wireless sensor networks are plagued with stringent energy constraints. It is therefore of utmost importance to incorporate energy awareness into the decentralized detection algorithm design. We suppose that each sensor uses a simple analog amplify-and-forward transmission scheme to transmit their data. As in [19], the local processing at each sensor node is confined to be a linear operator, which is referred to as *linear precoding*. This linear precoding allows for a simple implementation and is suitable for low-cost sensors with limited computational resources. However, unlike [19], in our study, we do not restrict the linear precoder to be a compression vector. In fact, since we already imposed a power constraint, there is no need to explicitly specify the number of messages sent by each sensor. This is also a major difference between our work and the works [16]–[18] aforementioned.

We are interested in examining the following fundamental questions: under a transmit power constraint, what is the optimal linear processing strategy at sensor nodes? shall we transmit a single compressed message, or multiple compressed messages, or just send the raw data to the FC? Sending more messages and sending one message have their own advantages: the former provides a diversity whereas the latter renders a better channel quality. The choice between these different strategies seems difficult before conducting a thorough mathematical analysis. Note

Manuscript received September 02, 2011; revised January 09, 2012; accepted February 28, 2012. Date of publication March 12, 2012; date of current version May 11, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ta-Sung Lee. This work was supported in part by the National Science Foundation by Grant ECCS-0901066, the National Science Foundation of China by Grant 61172114, the Program for New Century Excellent Talents in University (China) by Grant NCET-09-0261, and the Fundamental Research Funds for the Central Universities (China) by Grant ZYGX2010J011.

J. Fang, Z. Chen, and S. Li are with the National Key Laboratory on Communications, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: JunFang@uestc.edu.cn; chenzhi@uestc.edu.cn; lsq@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.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.2012.2190598

that although linear precoding design for decentralized detection remains new, its counterpart for distributed estimation has been extensively investigated, e.g. [21], [22]. In addition, the asymptotic behavior of the overall detection performance with an increasing number of sensors is examined in this paper, and a generalized likelihood ratio test (GLRT) is proposed for the scenario of unknown signals.

We briefly discuss the relationship between our work and [23], [24]. Although [23], [24] studied a decentralized estimation problem, some formulations and concepts (e.g., outage probability) in [23], [24] are closely connected to our work. Under similar energy constraint formulations, [23], [24] studied an optimal power allocation with the objective of minimizing the estimation error. Interestingly, it turns out that, although with different performance criteria, both our work and [23], [24] eventually arrive at a similar power allocation problem that has a common water-filling solution.

The rest of the paper is organized as follows. In Section II, we introduce the data model, basic assumptions, and the decentralized detection problem. Section III first develops an optimal Bayesian decision rule at the FC. The optimal precoding design and optimal power allocation (among sensors) are studied in Section IV. The impact of number of sensors on the overall detection performance is analyzed in V. Decentralized detection with unknown parameters is discussed in VI, followed by concluding remarks in Section VII.

II. Problem Formulation

We consider a binary hypothesis testing problem in which a number of sensors collaborate to detect the presence of a known deterministic vector signal $\boldsymbol{\theta} \in \mathbb{R}^p$. The binary hypothesis testing problem is formulated as follows:

$$\begin{aligned} H_0: \quad & \mathbf{x}_n = \mathbf{w}_n, \quad \forall n = 1, \ldots, N \\ H_1: \quad & \mathbf{x}_n = \mathbf{H}_n\boldsymbol{\theta} + \mathbf{w}_n, \quad \forall n = 1, \ldots, N \end{aligned} \tag{1}$$

where $\mathbf{H}_n \in \mathbb{R}^{q_n \times p}$ is the known observation matrix defining the input/output relation, and generally we have $\mathbf{H}_n\boldsymbol{\theta} \neq \mathbf{0}$ for some n in order to distinguish between these two hypotheses, $\mathbf{x}_n \in \mathbb{R}^{q_n}$ denotes the sensor's vector observation, $\mathbf{w}_n \in \mathbb{R}^{q_n}$ denotes the additive multivariate Gaussian noise with zero mean and covariance matrix $\mathbf{R}_{w_n}$, and the noise is assumed independent across the sensors. Unlike many existing works, the signal to be detected here is assumed to be a vector instead of a scalar. Vector models arise from a variety of scenarios. For example, if the underlying phenomena to be detected is a dynamic process, we can obtain vector signals by sampling the dynamic process at different time instances. Sensing of a target using multiple modalities (e.g., optical, chemical, thermal, magnetic, ultrasonic, etc.) also leads to multidimensional signals.

Let $\mathbf{C}_n$ denote the precoding matrix for sensor n. Without loss of generality, we assume that $\mathbf{C}_n$ is a $q_n \times q_n$ matrix that could be full rank or rank deficient, which accommodates different linear processing strategies. Each sensor uses an uncoded analog amplify-and-forward scheme to transmit its data to the FC. The signal at the FC received from the nth sensor is then given by

$$\mathbf{y}_n = \boldsymbol{\Gamma}_n\mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n \quad n = 1, \ldots, N \tag{2}$$

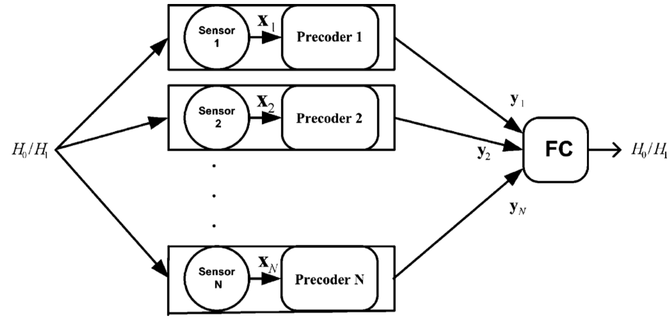


Fig. 1. Decentralized detection in a power-constrained network. Each node processes its vector observations through a linear precoder. Messages are then sent to the FC via wireless channels.

where $\boldsymbol{\Gamma}_n$ denotes the fading multiplicative channel matrix, and $\mathbf{v}_n$ represents the additive Gaussian channel noise with zero-mean and covariance matrix $\sigma_{v_n}^2\mathbf{I}$. The knowledge of the channel state information is assumed available at the FC.

The FC, based upon the received data $\{\mathbf{y}_n\}$, forms a final decision concerning the presence or absence of $\boldsymbol{\theta}$. Fig. 1 provides an illustration of the decentralized detection. The problem of interest is to determine the precoding matrix for each sensor, and to develop an optimal detector to detect $\boldsymbol{\theta}$ for the FC. Note that a transmit power constraint has to be imposed on the sensor nodes, otherwise we can always ensure ideal links between sensors and the FC by scaling the precoding matrices with an arbitrarily large factor. Let P_0 and P_1 denote the prior probabilities of the hypotheses H_0 and H_1, respectively. The average power radiated from sensor n is given by

$$\begin{aligned} & P_0 E\left[\|\mathbf{C}_n\mathbf{x}_n\|_2^2 \,|\, H_0\right] + P_1 E\left[\|\mathbf{C}_n\mathbf{x}_n\|_2^2 \,|\, H_1\right] \\ &\quad = P_0 \mathrm{tr}\left\{\mathbf{C}_n\mathbf{R}_{w_n}\mathbf{C}_n^T\right\} \\ &\qquad + P_1 \mathrm{tr}\left\{\mathbf{C}_n\mathbf{H}_n\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{C}_n^T + \mathbf{C}_n\mathbf{R}_{w_n}\mathbf{C}_n^T\right\} \\ &\quad = \mathrm{tr}\left\{\mathbf{C}_n\mathbf{R}_{w_n}\mathbf{C}_n^T + P_1\mathbf{C}_n\mathbf{H}_n\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{C}_n^T\right\}. \end{aligned} \tag{3}$$

However, in some detection applications, determining the prior probabilities of the respective hypotheses may not be possible. In this case, Neyman-Pearson detection without requiring the prior probabilities can be used. If the target/event to be detected occurs with a very small but unknown probability (this is exactly the case for many disaster detection applications), it is reasonable to consider a power constraint under hypothesis H_0 only [16], i.e., (3) with $P_1 = 0$. More discussions of the Neyman-Pearson detection will be provided later in this paper.

In the following, assuming that the precoding matrices are prespecified, we will first develop a Bayesian detector at the FC. The precoding matrix design is then investigated based on the detection performance analysis.

III. Bayesian Detector

Suppose that the precoding matrices $\{\mathbf{C}_n\}$ are prescribed. Let $\mathbf{y} \triangleq [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \ldots \ \mathbf{y}_N^T]^T$ denote the vector received at the FC, $\mathbf{y}_n$ is a Gaussian random vector with its mean and covariance matrix given by

$$\mathbf{y}_n \sim \begin{cases} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_n) & H_0 \\ \mathcal{N}(\boldsymbol{\Gamma}_n\mathbf{C}_n\mathbf{H}_n\boldsymbol{\theta}, \boldsymbol{\Sigma}_n) & H_1 \end{cases} \tag{4}$$

in which

$$\boldsymbol{\Sigma}_n \triangleq \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T + \sigma_{v_n}^2 \mathbf{I}. \tag{5}$$

Our objective is to design a decision rule that minimizes the average probability of error, i.e.,

$$P_{\rm e} = P(H_0 \,|\, H_1) P_1 + P(H_1 \,|\, H_0) P_0 \tag{6}$$

where $P(H_i \,|\, H_j)$ is the probability of deciding H_i when H_j is true. According to [25], in order to achieve a minimum $P_{\rm e}$, the decision rule is a likelihood ratio test (LRT) given as follows:

$$L(\mathbf{y}) = \frac{p(\mathbf{y} \,|\, H_1)}{p(\mathbf{y} \,|\, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} \triangleq \eta. \tag{7}$$

Noting that $\{\mathbf{y}_n\}$ are mutually independent for a given hypothesis, the LRT can be further expressed as

$$\begin{aligned} L(\mathbf{y}) &= \frac{\prod_{n=1}^N p(\mathbf{y}_n \,|\, H_1)}{\prod_{n=1}^N p(\mathbf{y}_n \,|\, H_0)} \\ &= \exp\left\{ \sum_{n=1}^N \left(\mathbf{y}_n^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \right.\right. \\ &\quad \left.\left. - \frac{1}{2} \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \right) \right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \end{aligned} \tag{8}$$

Taking logarithm on both sides of (8), the Bayesian decision rule can finally be put in the following form:

$$\sum_{n=1}^N \boldsymbol{\omega}_n^T \mathbf{y}_n + \Delta \underset{H_0}{\overset{H_1}{\gtrless}} \log \eta \tag{9}$$

where

$$\begin{aligned} \boldsymbol{\omega}_n &\triangleq \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \\ \Delta &\triangleq - \sum_{n=1}^N \frac{1}{2} \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta}. \end{aligned}$$

Δ is a constant independent of the observed data. Hence the LRT-based fusion rule is in fact a weighted linear combination of the data $\{\mathbf{y}_n\}$.

Define

$$u \triangleq \sum_{n=1}^N \boldsymbol{\omega}_n^T \mathbf{y}_n + \Delta.$$

Since u is a summation of a set of Gaussian random variables, u also follows a Gaussian distribution. It can be readily derived that its mean and variance under hypotheses H_0 and H_1 are given, respectively, as

$$u \sim \begin{cases} \mathcal{N}\left(\Delta, \sigma_u^2\right) & H_0 \\ \mathcal{N}\left(\Delta + \sigma_u^2, \sigma_u^2\right) & H_1 \end{cases} \tag{10}$$

where

$$\begin{aligned} \sigma_u^2 &\triangleq \sum_{n=1}^N \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \\ &= \sum_{n=1}^N \left\{ \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \right. \\ &\quad \left. \times \left(\boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T + \sigma_{v_n}^2 \mathbf{I} \right)^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \right\} \end{aligned} \tag{11}$$

are dependent on the precoding matrices $\{\mathbf{C}_n\}$. Clearly, the detection performance of the Bayesian detector fundamentally relies on the choice of these precoding matrices.

IV. Precoding Design & Power Allocation

In this section, we examine the problem of the precoding design, aiming at minimizing the probability of error $P_{\rm e}$. Recalling results in the previous section, we know that the FC makes a global decision based on

$$u \underset{H_0}{\overset{H_1}{\gtrless}} \log \eta \tag{12}$$

where u is a Gaussian random variable with mean $\mu_{u,0} = \Delta$ if H_0 is true, otherwise $\mu_{u,1} = \Delta + \sigma_u^2$; the variance under the null and alternative hypotheses remains the same. This hypothesis testing problem is called the *mean-shifted Gauss-Gauss* problem. For this type of detection problem, the detection performance is monotonic with the deflection coefficient χ [25]

$$\chi \triangleq \frac{(\mu_{u,1} - \mu_{u,0})^2}{\sigma_u^2} \tag{13}$$

that is, $P_{\rm e}$ decreases monotonically with χ. With $\mu_{u,0} = \Delta$ and $\mu_{u,1} = \Delta + \sigma_u^2$, it is easy to derive that

$$\chi = \sigma_u^2 \tag{14}$$

which indicates that the larger the variance σ_u^2, the better the detection performance. As shown in (11), σ_u^2 is a function of $\{\mathbf{C}_n\}$. Therefore the problem of minimizing $P_{\rm e}$ is equivalent to

$$\begin{aligned} \max_{\{\mathbf{C}_n\}} \ \sigma_u^2 = \sum_{n=1}^N &\left\{ \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \right. \\ &\left. \times \left(\boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T + \sigma_{v_n}^2 \mathbf{I} \right)^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \right\}. \end{aligned} \tag{15}$$

As aforementioned, we have to impose a transmit power constraint on the sensor nodes, otherwise the optimization is ill-posed since we can always ensure ideal links between sensors and the FC by scaling the precoding matrices with an arbitrarily large factor. To make the problem meaningful, we hereby impose an average total transmit power constraint. The precoding design can therefore be formulated as follows:

$$\begin{aligned} \max_{\{\mathbf{C}_n\}} \ &\sum_{n=1}^N \left\{ \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \right. \\ &\quad \left. \times \left(\boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T + \sigma_{v_n}^2 \mathbf{I} \right)^{-1} \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \right\} \\ \text{s.t.} \ &\sum_{n=1}^N \text{tr}\left\{ \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T + P_1 \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta}\boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \right\} \leq T_{\text{total}}. \end{aligned} \tag{16}$$

The above optimization can be decoupled into two sequential subtasks, namely, a power allocation (among sensors) problem and a set of independent precoding design problems. Let us suppose, for the time being, that a power allocation is prespecified and given as $\{T_1, T_2, \ldots, T_N\}$. Then the optimal precoding matrix for each sensor can be obtained by solving

$$\begin{aligned} \max_{\mathbf{C}_n} \ &\text{tr}\left\{ \left(\boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T + \sigma_{v_n}^2 \mathbf{I} \right)^{-1} \right. \\ &\quad \left. \times \boldsymbol{\Gamma}_n \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta}\boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \boldsymbol{\Gamma}_n^T \right\} \\ \text{s.t.} \ &\text{tr}\left\{ \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T + P_1 \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta}\boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T \right\} = T_n \end{aligned} \tag{17}$$

where the power constraint is represented as an equality instead of an inequality because the objective function is a monotonically increasing function of the transmit power. In the following, we first study the optimal precoding design by considering a simple but important channel scenario. Its extension to a general channel case is then followed.

A. Optimum Precoding Design: A Simple and Important Case

Due to size and cost limitations, each sensor node is very likely to be equipped with only one transmit antenna. If multiple messages need to be sent to the FC, a time-division multiplexing technique can be used, in which case the channel matrix $\mathbf{\Gamma}_n$ is diagonal. Also, we assume that its diagonal elements are identical, i.e., $\mathbf{\Gamma}_n = \gamma_n \mathbf{I}, \forall n$. This could be the case for slowly varying channels. The optimization therefore can be reduced to

$$\begin{aligned} \max_{\mathbf{C}_n} \quad & \mathrm{tr}\left\{\left(\mathbf{C}_n\mathbf{R}_{w_n}\mathbf{C}_n^T + \bar{\sigma}_{v_n}^2\mathbf{I}\right)^{-1}\mathbf{C}_n\mathbf{H}_n\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{C}_n^T\right\} \\ \text{s.t.} \quad & \mathrm{tr}\left\{\mathbf{C}_n\mathbf{R}_{w_n}\mathbf{C}_n^T + P_1\mathbf{C}_n\mathbf{H}_n\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{C}_n^T\right\} = T_n \end{aligned} \tag{18}$$

where $\bar{\sigma}_{v_n}^2 \triangleq \sigma_{v_n}^2/\gamma_n^2$. The optimization (18) is complicated in its current form. To simplify the problem, we perform a series of matrix transformations in the following. Define

$$\begin{aligned} \tilde{\mathbf{C}}_n &\triangleq \mathbf{C}_n\mathbf{R}_{w_n}^{\frac{1}{2}} \\ \mathbf{G}_n &\triangleq \mathbf{R}_{w_n}^{-\frac{1}{2}}\mathbf{H}_n\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-\frac{1}{2}} \end{aligned} \tag{19}$$

and substitute them into (18), the optimization becomes

$$\begin{aligned} \max_{\tilde{\mathbf{C}}_n} \quad & \mathrm{tr}\left\{\left(\tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T + \bar{\sigma}_{v_n}^2\mathbf{I}\right)^{-1}\tilde{\mathbf{C}}_n\mathbf{G}_n\tilde{\mathbf{C}}_n^T\right\} \\ \text{s.t.} \quad & \mathrm{tr}\left\{\tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T + P_1\tilde{\mathbf{C}}_n\mathbf{G}_n\tilde{\mathbf{C}}_n^T\right\} = T_n \end{aligned} \tag{20}$$

Furthermore, let $\tilde{\mathbf{C}}_n = \mathbf{U}\mathbf{D}\mathbf{V}^T$ denote the singular value decomposition (SVD) of $\tilde{\mathbf{C}}_n$, in which we drop the subscript n for those matrices $\{\mathbf{U}, \mathbf{D}, \mathbf{V}\}$ for simplicity. Without loss of generality, we assume that the diagonal matrix $\mathbf{D}$ has nonnegative diagonal elements, i.e., $d_{ii} \geq 0$. Substituting the SVD into (20), we arrive at a new optimization that searches for an optimal orthonormal matrix $\mathbf{V}$ and an optimal diagonal matrix $\mathbf{D}$ ($\mathbf{U}$ is canceled and therefore can be any orthonormal matrix)

$$\begin{aligned} \max_{\{\mathbf{V},\mathbf{D}\}} \quad & \mathrm{tr}\left\{\mathbf{D}\left(\mathbf{D}^2 + \bar{\sigma}_{v_n}^2\mathbf{I}\right)^{-1}\mathbf{D}\mathbf{V}^T\mathbf{G}_n\mathbf{V}\right\} \\ \text{s.t.} \quad & \mathrm{tr}\left\{\mathbf{D}^2 + P_1\mathbf{D}^2\mathbf{V}^T\mathbf{G}_n\mathbf{V}\right\} = T_n. \end{aligned} \tag{21}$$

Let $\mathbf{F} \triangleq \mathbf{V}^T\mathbf{G}_n\mathbf{V}$, and f_{ii} denote the ith diagonal element of $\mathbf{F}$. We have the following properties regarding the diagonal elements $\{f_{ii}\}$:

$$\begin{aligned} &\text{(i)} \quad f_{ii} \geq 0 \\ &\text{(ii)} \quad \sum_{i=1}^{q_n} f_{ii} = \lambda_{\max}(\mathbf{G}_n). \end{aligned} \tag{22}$$

In above properties, the first follows from the fact that $\mathbf{F}$ is a positive-semidefinite matrix. The second can be easily derived by resorting to the trace identity $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ and noting that $\mathbf{G}_n$ is a rank-one matrix (cf. (19)), where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$.

Treating $\mathbf{F}$ as a new optimization variable, the optimization (21) can be reexpressed as

$$\begin{aligned} \max_{\{d_{ii}, f_{ii}\}} \quad & \sum_{i=1}^{q_n} \frac{d_{ii}^2 f_{ii}}{d_{ii}^2 + \bar{\sigma}_{v_n}^2} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} d_{ii}^2(1 + P_1 f_{ii}) = T_n \\ & f_{ii} \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} f_{ii} = \lambda_{\max}(\mathbf{G}_n) \end{aligned} \tag{23}$$

which, as we can see, involves only the diagonal elements of $\mathbf{F}$, while irrespective of its off-diagonal entries. The solution to (23) is given in the following lemma.

Lemma 1: The optimal solution to (23) is given by

$$f_{ii}^* = \begin{cases} \lambda_{\max}(\mathbf{G}_n) & i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

$$d_{ii}^* = \begin{cases} \sqrt{\frac{T_n}{1+P_1\lambda_{\max}(\mathbf{G}_n)}} & i = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Proof: See Appendix A. ■

Utilizing Lemma 1, we can determine the optimal precoding matrix. The results are summarized as follows.

Theorem 1: The optimal precoding matrix, that is, the optimal solution to (18), is a matrix with its first row a nonzero vector, whereas all other rows equal to zeros, i.e.

$$\mathbf{C}_n^*[i,:] = \begin{cases} \varphi_n\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-1} & i = 1 \\ \mathbf{0} & i \in \{2, \ldots, q_n\} \end{cases} \tag{26}$$

where

$$\varphi_n \triangleq \frac{1}{\left\|\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-\frac{1}{2}}\right\|_2}\sqrt{\frac{T_n}{1 + P_1\lambda_{\max}(\mathbf{G}_n)}}$$

is a scaling factor to satisfy the power constraint.

Proof: Clearly, we have

$$\mathbf{C}_n^* = \tilde{\mathbf{C}}_n^*\mathbf{R}_{w_n}^{-\frac{1}{2}} = \mathbf{U}^*\mathbf{D}^*(\mathbf{V}^*)^T\mathbf{R}_{w_n}^{-\frac{1}{2}}.$$

The optimal $\mathbf{D}^*$ is a diagonal matrix with its diagonal elements given by (25). From $\mathbf{F} = \mathbf{V}^T\mathbf{G}_n\mathbf{V}$, it is easy to deduce that the orthonormal matrix $\mathbf{V}$ that yields (24) must be

$$\mathbf{V}^* = \mathbf{U}_{g_n} \tag{27}$$

where $\mathbf{U}_{g_n}$ is an orthonormal matrix obtained from the eigenvalue decomposition (EVD): $\mathbf{G}_n = \mathbf{U}_{g_n}\mathbf{D}_{g_n}\mathbf{U}_{g_n}^T$, in which the diagonal elements of $\mathbf{D}_{g_n}$ are arranged in a descending order. Also, we assume $\mathbf{U}^* = \mathbf{I}$ since $\mathbf{U}^*$ can be any orthonormal matrix. Therefore we have

$$\begin{aligned} \mathbf{C}_n^* &= \mathbf{D}^*\mathbf{U}_{g_n}^T\mathbf{R}_{w_n}^{-\frac{1}{2}} \\ &= \begin{cases} d_{11}^*\left(\mathbf{U}_{g_n}[:,1]\right)^T\mathbf{R}_{w_n}^{-\frac{1}{2}} & i = 1 \\ \mathbf{0} & i \in \{2, \ldots, q_n\} \end{cases} \\ &= \begin{cases} \varphi_n\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-1} & i = 1 \\ \mathbf{0} & i \in \{2, \ldots, q_n\} \end{cases} \end{aligned} \tag{28}$$

where the last equality comes from the fact that $\mathbf{G}_n$ is a rank-one matrix and the eigenvector of $\mathbf{G}_n$ corresponding to the nonzero/largest eigenvalue is equal to $\mathbf{R}_{w_n}^{-\frac{1}{2}}\mathbf{H}_n\boldsymbol{\theta}/\|\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-\frac{1}{2}}\|_2$. The proof is completed here. ■

Remark 1: Note that the optimal solution (26) has only one nonzero row. This suggests that in order to achieve best detection performance, each sensor's local measurements should be compressed into only one message. Also, it can be readily observed that the compression/precoding vector is exactly a matched filter in a vector form. Matched filter detection in a conventional context (i.e., centralized and no power constraint) is a well-studied topic. Nevertheless, to our best knowledge, the optimality of the matched filter in distributed power-constrained networks has never been established before.

Remark 2: Although the optimality of transmitting one single message is established for this simple channel scenario, as we will show in the next subsection, its optimality is no longer valid for the general channel case.

B. Optimum Precoding Design: A General Channel Case

A general channel matrix $\boldsymbol{\Gamma}_n$ may arise as a result of coherent transmissions with the aid of multiple transmit antennas[1]. The optimization (17), although has a more complex formulation than (18), can still be solved by following the same procedure and utilizing Lemma 1. Suppose that $\boldsymbol{\Gamma}_n \in \mathbb{R}^{q_n\times q_n}$ (the extension to a non-square matrix is considered in Appendix D) is full rank and $\boldsymbol{\Gamma}_n = \mathbf{U}_{\gamma_n}\mathbf{D}_{\gamma_n}\mathbf{V}_{\gamma_n}^T$ denotes the corresponding SVD. Without loss of generality, the singular values are arranged in a descending order. Re-defining $\tilde{\mathbf{C}}_n \triangleq \mathbf{V}_{\gamma_n}^T\mathbf{C}_n\mathbf{R}_{w_n}^{\frac{1}{2}}$, the optimization (17) can be re-expressed as

$$\begin{aligned}\max_{\tilde{\mathbf{C}}_n}\quad & \mathrm{tr}\left\{\left(\tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T+\sigma_{v_n}^2\mathbf{D}_{\gamma_n}^{-2}\right)^{-1}\tilde{\mathbf{C}}_n\mathbf{G}_n\tilde{\mathbf{C}}_n^T\right\}\\ \text{s.t.}\quad & \mathrm{tr}\left\{\tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T+P_1\tilde{\mathbf{C}}_n\mathbf{G}_n\tilde{\mathbf{C}}_n^T\right\}=T_n. \end{aligned}\tag{29}$$

Substituting the SVD of $\tilde{\mathbf{C}}_n = \mathbf{U}\mathbf{D}\mathbf{V}^T$ into (29), similarly we arrive at an optimization that searches for an optimal orthonormal matrix $\mathbf{V}$ and an optimal diagonal matrix $\mathbf{D}$

$$\begin{aligned}\max_{\{\mathbf{V},\mathbf{D}\}}\quad & \mathrm{tr}\left\{\mathbf{D}\left(\mathbf{D}^2+\sigma_{v_n}^2\mathbf{D}_{\gamma_n}^{-2}\right)^{-1}\mathbf{D}\mathbf{V}^T\mathbf{G}_n\mathbf{V}\right\}\\ \text{s.t.}\quad & \mathrm{tr}\left\{\mathbf{D}^2+P_1\mathbf{D}^2\mathbf{V}^T\mathbf{G}_n\mathbf{V}\right\}=T_n\end{aligned}\tag{30}$$

which can be further written as follows:

$$\begin{aligned}\max_{\{d_{ii},f_{ii}\}}\quad & \sum_{i=1}^{q_n}\frac{d_{ii}^2f_{ii}}{d_{ii}^2+\left(\sigma_{v_n}^2/d_{\gamma_n,i}^2\right)}\\ \text{s.t.}\quad & \sum_{i=1}^{q_n}d_{ii}^2(1+P_1f_{ii})=T_n\\ & f_{ii}\geq 0\quad\forall i\\ & \sum_{i=1}^{q_n}f_{ii}=\lambda_{\max}(\mathbf{G}_n)\end{aligned}\tag{31}$$

where $d_{\gamma_n,i}$ denotes the ith singular value of $\boldsymbol{\Gamma}_n$. To gain an insight into solving (31), we construct a new optimization which uses a surrogate function upper-bounding the objective function (31)

$$\begin{aligned}\max_{\{d_{ii},f_{ii}\}}\quad & \sum_{i=1}^{q_n}\frac{d_{ii}^2f_{ii}}{d_{ii}^2+\left(\sigma_{v_n}^2/d_{\gamma_n,1}^2\right)}\\ \text{s.t.}\quad & \sum_{i=1}^{q_n}d_{ii}^2(1+P_1f_{ii})=T_n\\ & f_{ii}\geq 0\quad\forall i\\ & \sum_{i=1}^{q_n}f_{ii}=\lambda_{\max}(\mathbf{G}_n).\end{aligned}\tag{32}$$

We can readily see that the optimal solution to (32) is given by (24)–(25). Besides, (32) and (31) achieve the same objective function value for the given solution (24)–(25). Therefore we can quickly infer that the optimal solution to (31) is given by (24)–(25) as well.

The optimal precoding matrix for the general channel case can be obtained by tracing back from the optimal solution of (31). The results are summarized as follows.

Theorem 2: Suppose that $\boldsymbol{\Gamma}_n \in \mathbb{R}^{q_n\times q_n}$ is full rank. The optimal precoding matrix, that is, the optimal solution to (17), is given by

$$\mathbf{C}_n^* = \varphi_n\mathbf{V}_{\gamma_n}[:,1]\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-1}\tag{33}$$

where $\mathbf{V}_{\gamma_n}[:,1]$ is the right singular vector of $\boldsymbol{\Gamma}_n$ associated with the largest singular value.

Remark 1: We observe that the optimal solution (33) has a structure with its rows all identical except with different scaling factors. This suggests that multiple messages should be transmitted to the FC for the general channel case. Nevertheless, these multiple messages are obtained by one single compressed message multiplied by different amplification gains that are proportional to the components of $\mathbf{V}_{\gamma_n}[:,1]$. Again, the compression/precoding vector, $\boldsymbol{\theta}^T\mathbf{H}_n^T\mathbf{R}_{w_n}^{-1}$, used to produce the single compressed message has a classic matched filter form. The objective of transmitting multiple identical messages with different amplification gains is to improve the signal quality by exploiting the channel diversity. Note that if $\boldsymbol{\Gamma}_n$ is diagonal, its right singular vector is a unit column vector with only one non-zero entry. Hence in this case, only one message needs to be sent to the FC. This is exactly the case we discussed in previous subsection.

Remark 2: Although $\boldsymbol{\Gamma}_n$ is assumed a square matrix, the above results hold valid for a nonsquare channel matrix $\boldsymbol{\Gamma}_n \in \mathbb{R}^{r\times q_n}$. This can be easily derived and the details are provided in Appendix D.

C. Optimum Power Allocation

In previous subsections, we studied the optimum precoding design when a power assignment among sensors is specified. Substituting the optimum precoder back into (16), we obtain the following power allocation problem

$$\begin{aligned}\max_{\{T_n\}}\quad & \sum_{n=1}^{N}\frac{T_n\lambda_{\max}(\mathbf{G}_n)}{T_n+\bar{\sigma}_{v_n}^2(1+P_1\lambda_{\max}(\mathbf{G}_n))}\\ \text{s.t.}\quad & \sum_{n=1}^{N}T_n\leq T_{\text{total}}\\ & T_n\geq 0.\end{aligned}\tag{34}$$

[1]Although it is unlikely that a low-cost sensor node is equipped with multiple transmit antennas, we still consider the general channel case to provide a complete treatment of the problem.

For the general channel case, the power allocation problem remains the same except that $\bar{\sigma}_{v_n}^2$ is replaced by $\sigma_{v_n}^2/d_{\gamma_n,1}^2$. It is easy to verify that the optimization problem (34) is convex. Although (34) is efficiently solvable by numerical methods, it can also be solved analytically by resorting to the Lagrangian function and Karush-Kuhn-Tucker (KKT) conditions, which leads to a water-filling type power allocation scheme. The details are elaborated in Appendix E. Briefly speaking, for a threshold ϕ that is uniquely determined by a procedure described in Appendix E, we have

$$T_n = \begin{cases} \frac{1}{\beta_n}\left(\sqrt{\frac{\alpha_n}{\phi}} - 1\right) & \alpha_n \geq \phi \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

where

$$\alpha_n \triangleq \frac{\lambda_{\max,n}}{\bar{\sigma}_{v_n}^2(1 + P_1\lambda_{\max,n})}$$
$$\beta_n \triangleq \frac{1}{\bar{\sigma}_{v_n}^2(1 + P_1\lambda_{\max,n})}$$

and $\lambda_{\max,n}$ stands for $\lambda_{\max}(\mathbf{G}_n)$ for notational convenience.

Remark: We see that the optimal power allocation requires the knowledge of the observation noise statistics $\{\mathbf{R}_{w_n}\}$, the observation matrices $\{\mathbf{H}_n\}$, as well as the channel state information $\{\mathbf{\Gamma}_n\}$ and the channel noise variance $\{\sigma_{v_n}^2\}$. Although it provides the best performance, optimal power allocation usually involves considerable communication overheads from sensors to the FC.

D. Summary and Numerical Results

For clarity, we now summarize the proposed optimal solution.

1) Given the prior knowledge of the noise statistics, the signal $\boldsymbol{\theta}$, and the observation and channel matrices, compute $\{\alpha_n\}$ and $\{\beta_n\}$.
2) Given the total power constraint T_{total}, find the optimal power allocation among sensors via (34). The solution of (34) is elaborated in Appendix E.
3) With the optimal power assignment, determine the optimal precoding matrices $\{\mathbf{C}_n\}$ via (17), whose solution is given by (26) for the case $\mathbf{\Gamma}_n = \gamma_n\mathbf{I}, \forall n$ and (33) for the general channel case.

We now provide numerical examples to verify the analytical results. In the simulations, the prior probabilities of the null and alternative hypotheses are assumed identical. The vector parameter is a three-dimensional vector with its entries equal to one, i.e., $\boldsymbol{\theta} = [1\ 1\ 1]^T$. We first consider a single-sensor system which has only one sensor node. We set $\mathbf{H} = \mathbf{I}$, $\mathbf{\Gamma} = \mathbf{I}$, $\mathbf{R}_w = 0.5\mathbf{I}$, and $\sigma_v^2 = 0.5$. Fig. 2 shows the average probability of error P_e as a function of the transmit power for both optimal precoding and no precoding, in which no precoding corresponds to sending the original data, i.e., $\mathbf{C} = \mathbf{I}$. It can be seen that the optimal compression strategy outperforms the noncompression strategy, which corroborates our theoretical analysis.

The detection performance under different power allocation schemes is also investigated. We set $N = 20$, $\mathbf{\Gamma}_n = \gamma_n\mathbf{I}$, $\mathbf{H}_n = \mathbf{I}$, $\sigma_{v_n}^2 = 1$ and $\mathbf{R}_{w_n} = 0.5\mathbf{I}$ for all n. The absolute channel gains, $\{|\gamma_n|\}$, are assumed independent and identically distributed (i.i.d.) Rayleigh-fading random variables with unit variance. Fig. 3 plots the detection performance of two different power allocation schemes, namely, an optimal power allocation and an equal power allocation. Results are averaged over one million independent runs. For both schemes, optimal precoders (conditioned on optimal and equal power allocation) are used. From Fig. 3, we see that for i.i.d. Rayleigh-fading channels, optimal power allocation presents a clear performance advantage over the equal power allocation scheme.

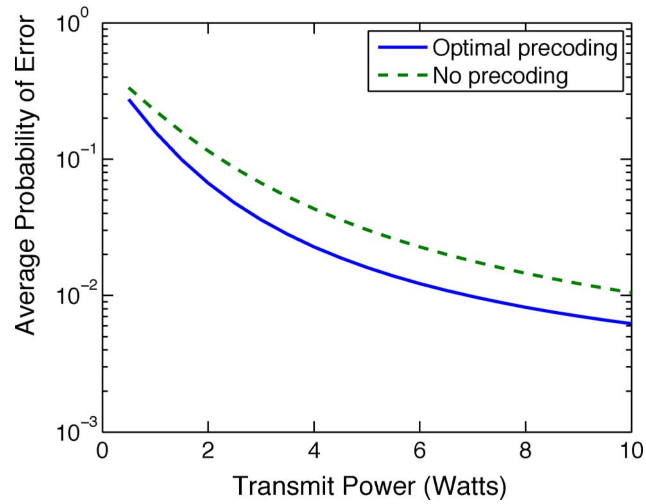


Fig. 2. Average probability of error versus transmit power for optimal precoding and no precoding strategies.

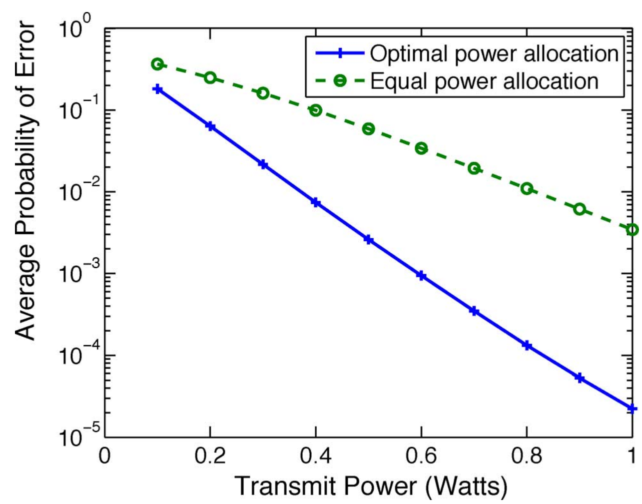


Fig. 3. Average probability of error vs. transmit power for optimal power allocation and equal power allocation schemes.

E. Extension to Neyman-Pearson Detection

The extension of our theoretical results to the Neyman-Pearson variant of the detection problem is straightforward. The Neyman-Pearson detection aims at maximizing the detection probability subject to a given false alarm probability. The decision rule is still a LRT, except that its threshold is determined by the specified false alarm probability. As indicated earlier, in the Neyman-Pearson formulation, the prior probabilities of the null and alternative hypotheses are unknown. Nevertheless, when the event/target to be detected has a rare occurrence, the power constraint could be a constraint on the behavior of the system under hypothesis H_0 (corresponding to $P_1 = 0$) [16]. Following a similar derivation, it is easy to show that the precoding design under the Neyman-Pearson framework is still given by the optimization (16), but with $P_1 = 0$. Therefore the optimal precoding design and the optimal power allocation hold valid for the Neyman-Pearson detector, simply with P_1 replaced by zero. It can be readily observed that the optimal precoding design for Neyman-Pearson detector still has a matched filter structure, but with a different scaling factor to satisfy the power constraint.

V. Equal Power Allocation: Detection Diversity

In this section, given a fixed total transmit power, we analyze the impact of the number of sensors on the overall detection performance, assuming the channels between sensors and the FC experience i.i.d. fading. Throughout this section, we assume $\mathbf{\Gamma}_n = \gamma_n \mathbf{I}, \forall n$, in which the channel gains $\{\gamma_n\}$ are i.i.d. random variables following a certain distribution.

To facilitate our analysis, we consider an equal-power allocation scheme in which all sensors transmit the same amount of power. Also, we assume a homogeneous scenario where $\mathbf{H}_n = \mathbf{I}, \mathbf{R}_{w_n} = \sigma_w^2 \mathbf{I}$, and $\sigma_{v_n}^2 = \sigma_v^2$ for all n. When optimal precoders (conditional on the equal-power allocation) are used, according to (34), the deflection coefficient χ is given by

$$\chi = \sum_{n=1}^{N} \frac{T_{\text{total}} \lambda_{\max}(\mathbf{G}_n)}{T_{\text{total}} + \bar{\sigma}_{v_n}^2 N (1 + P_1 \lambda_{\max}(\mathbf{G}_n))} \stackrel{(a)}{=} \sum_{n=1}^{N} \frac{T_{\text{total}} \|\boldsymbol{\theta}\|_2^2}{\sigma_w^2 T_{\text{total}} + \bar{\sigma}_{v_n}^2 N (\sigma_w^2 + P_1 \|\boldsymbol{\theta}\|_2^2)} \tag{36}$$

where (a) comes from the fact that $\lambda_{\max}(\mathbf{G}_n) = \|\boldsymbol{\theta}\|_2^2 / \sigma_w^2$. For notational convenience, define

$$\rho_1 \triangleq \frac{T_{\text{total}} \|\boldsymbol{\theta}\|_2^2}{\sigma_w^2 + P_1 \|\boldsymbol{\theta}\|_2^2} \quad \rho_2 \triangleq \frac{\sigma_w^2 T_{\text{total}}}{\sigma_w^2 + P_1 \|\boldsymbol{\theta}\|_2^2}$$

When the total number of sensors, N, increases without bound, χ asymptotically approaches

$$\chi = \sum_{n=1}^{N} \frac{\rho_1}{\rho_2 + N \bar{\sigma}_{v_n}^2} = \sum_{n=1}^{N} \frac{1}{N} \left(\frac{\rho_1}{\bar{\sigma}_{v_n}^2} - \frac{\rho_1 \rho_2}{N \bar{\sigma}_{v_n}^4 + \rho_2 \bar{\sigma}_{v_n}^2} \right) \stackrel{N \to \infty}{=} \frac{\rho_1}{\sigma_v^2} E\left[\gamma_n^2\right] \triangleq \chi_\infty \tag{37}$$

where the last equality follows from $\bar{\sigma}_{v_n}^2 = \sigma_v^2 / \gamma_n^2$ and the strong Law of Large Numbers (LLN) under the assumption of i.i.d. $\{\gamma_n\}$. The detection performance under different number of sensors is illustrated in Fig. 4. In this example, we assume that $P_0 = P_1 = 0.5$, $\boldsymbol{\theta} = [1\ 1\ 1]^T$, and $\mathbf{H}_n = \mathbf{I}$, $\mathbf{R}_{w_n} = 0.5\mathbf{I}$, $\sigma_{v_n}^2 = 1$ for all n. $|\gamma_n|$'s are assumed i.i.d. Rayleigh-fading random variables with unit variance. Results are averaged over one million independent random realizations. The asymptotic performance when the number of sensors increases without bound is also included for comparison. We see from Fig. 4 that, for a fixed amount of transmit power, the detection performance improves notably as we increase the number of sensor nodes, which suggests that exploiting channel diversity can achieve a substantial performance improvement.

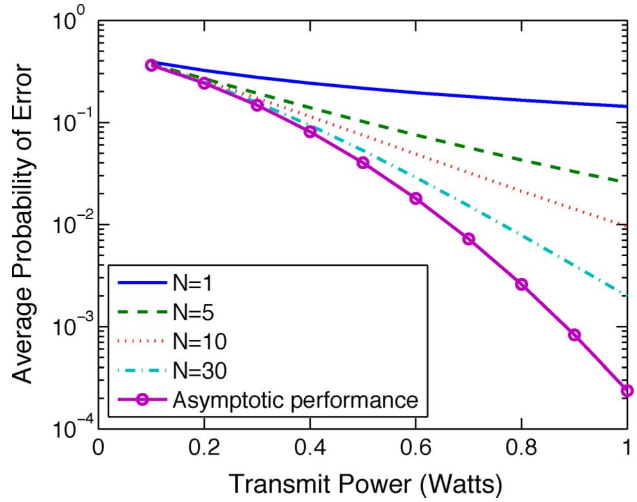


Fig. 4. Average probability of error versus transmit power for different number of sensors.

The detection diversity gain can be explored from a different perspective. In [24], the notion of "estimation outage probability" was proposed to quantify the reliability of the overall estimation system. Following [24], two new concepts called "detection diversity" and "detection outage probability" were introduced in [15]. Inspired by these two works, we hereby adopt the concept "detection outage probability" to quantify the reliability of the detection system. The detection outage probability is defined as the probability of the detection probability being less than a specified requirement given a certain false alarm probability, i.e.

$$P_{\text{outage}} \triangleq \Pr\{P_{\text{D}} < \tau_{\text{D}} \,|\, P_{\text{FA}}\}. \tag{38}$$

Note that the detection probability P_{D} is for a given channel realization, while the outage probability that P_{D} is less than a specified requirement is calculated by taking into account all possible channel realizations. The definition here is slightly different from that defined in [15], in which the outage probability is defined as the probability that a different performance metric called "J-divergence" is smaller than a certain threshold.

Recall that the test statistic u is a Gaussian random variable with its mean and variance under the null and alternative hypotheses given by (10). Therefore for a prescribed false alarm probability, the detection probability P_{D} is given as

$$P_{\text{D}} = \Pr(u > \eta \,|\, H_1) = Q\left(\frac{\eta - \Delta - \sigma_u^2}{\sigma_u}\right) = Q(Q^{-1}(P_{\text{FA}}) - \sigma_u) \tag{39}$$

where $Q(x)$ denotes the Q-function. Utilizing the above result, the detection outage probability can be rewritten as

$$\begin{aligned} P_{\text{outage}} &= \Pr(Q(Q^{-1}(P_{\text{FA}}) - \sigma_u) < \tau_{\text{D}}) \\ &= \Pr(Q^{-1}(P_{\text{FA}}) - \sigma_u > Q^{-1}(\tau_{\text{D}})) \\ &= \Pr(\sigma_u < Q^{-1}(P_{\text{FA}}) - Q^{-1}(\tau_{\text{D}})) \\ &= \Pr(\chi < \zeta) \end{aligned} \tag{40}$$

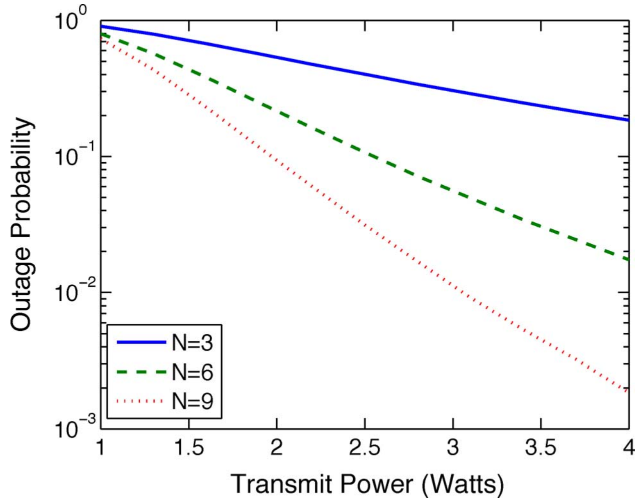


Fig. 5. Outage probability versus total transmit power for different number of sensors.

in which $\zeta \triangleq (Q^{-1}(P_{\text{FA}}) - Q^{-1}(\tau_{\text{D}}))^2$. We see that the detection outage probability is in fact the probability of the deflection coefficient being less than a certain threshold.

From (37), it can be observed that when N is sufficiently large, the deflection coefficient χ is approximately equal to the sample mean of i.i.d. random variables $\{\gamma_n^2 \rho_1/\sigma_v^2\}$. According to the large deviation theory [26], for any $\zeta < \chi_\infty$, we have the outage probability decreasing exponentially with N as follows:

$$P_{\text{outage}} \sim \exp\left(-N I_{\varpi}\left(\zeta\sigma_v^2/\rho_1\right)\right) \tag{41}$$

where $\sim$ means asymptotic convergence as N becomes large, ϖ is the common distribution of $\varpi_n \triangleq \gamma_n^2$, and $I_\varpi(x)$ is the rate function of ϖ

$$I_\varpi(x) = \sup_{t\in\mathbb{R}}(tx - \log M_\varpi(t)) \tag{42}$$

with $M_\varpi(t)$ the moment-generating function of ϖ. From (41), we see that if the specified P_{FA} and τ_{D} satisfy the following condition:

$$(Q^{-1}(P_{\text{FA}}) - Q^{-1}(\tau_{\text{D}}))^2 < \chi_\infty \tag{43}$$

then the detection outage probability can be made arbitrarily small by increasing the number of sensors N, even with the total transmit power fixed. Note that since χ_∞ is proportional to the total transmit power, (43) can always be met for a sufficiently large transmit power. The behavior of the outage probability with different number of sensors is illustrated in Fig. 5. We set $P_1 = 0$, $P_{\text{FA}} = 0.1$, and $\tau_{\text{D}} = 0.9$, and assume other simulation parameters the same as in previous example. Results are averaged over one million independent random realizations. It can be verified that (43) is satisfied as long as $T_{\text{total}} \geq 1$. From Fig. 5, we see that the outage probability decreases considerably even though we slightly increase the number of sensors.

VI. Decentralized Detection With Unknown Signals

From preceding analyses, we see that the decision rule at the FC, the precoding design, and the power allocation all require the knowledge of the signal $\boldsymbol{\theta}$ to be detected. A fundamental assumption made in previous sections is that the signal $\boldsymbol{\theta}$ is known *a priori* or the signal can be estimated from the training data before the detection task is performed. In the following, we discuss, if the knowledge of the signal to be detected is not available, how to form a final decision at the FC and design the precoder for each sensor. Since the optimality of compression-transmission strategy is already established in previous sections, we are only concerned about the precoding vector design. The channels are assumed equal to $\boldsymbol{\Gamma}_n = \gamma_n \mathbf{I}, \forall n$ throughout this section.

A. GLRT Detector

Suppose that the precoding vectors $\{\mathbf{c}_n\}$ are predetermined, we can use a generalized likelihood ratio test (GLRT) which replaces the unknown signal with their maximum likelihood estimates (MLEs). In the case there are no unknown parameters under H_0, the GLRT decides H_1 if

$$L_G(\mathbf{y}) = \frac{p(\mathbf{y}\,|\,\hat{\boldsymbol{\theta}}; H_1)}{p(\mathbf{y}\,|\,H_0)} > \eta \tag{44}$$

where $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ found by maximizing

$$p(\mathbf{y}\,|\,\boldsymbol{\theta}; H_1) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{P}\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{P}\boldsymbol{\theta})\right\} \tag{45}$$

in which $\boldsymbol{\Sigma}$ is a diagonal matrix with its nth diagonal element given by $\gamma_n^2\mathbf{c}_n\mathbf{R}_{w_n}\mathbf{c}_n^T + \sigma_{v_n}^2$, and

$$\mathbf{P} \triangleq \begin{bmatrix} \gamma_1\mathbf{c}_1\mathbf{H}_1 \\ \gamma_2\mathbf{c}_2\mathbf{H}_2 \\ \vdots \\ \gamma_N\mathbf{c}_N\mathbf{H}_N \end{bmatrix}. \tag{46}$$

The MLE of $\boldsymbol{\theta}$ can be solved by taking the logarithm of $p(\mathbf{y}\,|\,\boldsymbol{\theta}; H_1)$ and setting the first derivative equal to zero, which gives

$$\hat{\boldsymbol{\theta}} = (\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{P})^{-1}\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}. \tag{47}$$

Note that $\mathbf{P}$ has to be full column rank, otherwise the MLE requires solving an ill-posed inverse problem (more details regarding the choice of the precoding vectors such that $\mathbf{P}$ is full column rank will be provided later). Substituting $\hat{\boldsymbol{\theta}}$ back into (44), thus we have

$$\ln L_G(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{P}(\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{P})^{-1}\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} \tag{48}$$

or we decide H_1 if

$$\mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{P}(\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{P})^{-1}\mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} > \eta'. \tag{49}$$

It is shown in ([25], Section 6.5) that when $N \to \infty$, the GLRT statistic $2\ln L_G(\mathbf{y})$ under hypothesis H_0 follows a chi-squared distribution with p degrees of freedom, which does not depend on any unknown parameters. Therefore the threshold required to maintain a constant P_{FA} can be found.

B. Precoding Design With Unknown Signals

When $\boldsymbol{\theta}$ is unknown or the estimate of $\boldsymbol{\theta}$ is not available, determining the optimal precoding vectors is not possible. In this case, we propose a heuristic method for precoding design.

In practice, the sign of each component of the vector $\mathbf{g}_n \triangleq \mathbf{R}_{w_n}^{-\frac{1}{2}}\mathbf{H}_n\boldsymbol{\theta}$ may be obtained from the signal dynamic range or estimated from the observations. This knowledge can be exploited for precoding vector design. Let $\mathrm{sgn}(\mathbf{x})$ be a sign column vector with its elements given by $\mathrm{sgn}(x_i)$, where $\mathrm{sgn}(x_i)=1$ if $x_i>0$, and $\mathrm{sgn}(x_i)=-1$ otherwise. We design the precoding vector for each sensor as follows:

$$\mathbf{c}_n = \psi_n(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}_n))^T\mathbf{R}_{w_n}^{-\frac{1}{2}} \quad \forall n \tag{50}$$

where $\mathbf{r}_n$ is a column vector whose entries are randomly generated according to a Gaussian distribution with zero mean and unit variance, $|\mathbf{r}_n|$ takes the absolute value of each entry of $\mathbf{r}_n$, $\odot$ denotes the entry-wise multiplication, and ψ_n is a scaling factor which ensures that the precoding vector satisfies the specified power constraint (note that ψ_n can be determined without the knowledge of $\boldsymbol{\theta}$ if we set $P_1=0$). The rationale behind this heuristic design is to preserve the signal energy as much as possible by using the sign information. To see this, note that under the alternative hypothesis, the compressed message can be written as

$$\begin{aligned}\mathbf{c}_n\mathbf{x}_n &= \psi_n(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}_n))^T\mathbf{R}_{w_n}^{-\frac{1}{2}}(\mathbf{H}_n\boldsymbol{\theta}+\mathbf{w}_n)\\ &= \psi_n(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}_n))^T(\mathbf{g}_n+\tilde{\mathbf{w}}_n)\end{aligned} \tag{51}$$

where $\tilde{\mathbf{w}}_n$ is the whitened observation noise. Utilizing the sign knowledge of $\mathbf{g}_n$, the precoding vector design (50) preserves the signal energy by aligning the signs of the signal components of $\mathbf{g}_n$. This explains the use of the term $\mathrm{sgn}(\mathbf{g}_n)$ in (50). On the other hand, as mentioned earlier, the matrix $\mathbf{P}$ defined in (46) has to be full column rank, otherwise the GLRT detector involves an ill-posed inverse problem. Therefore the term $\mathrm{sgn}(\mathbf{g}_n)$ is entry-wise multiplied by a randomly generated vector $|\mathbf{r}_n|$ which guarantees that $\mathbf{P}$ is full column rank with a high probability. This heuristic design shares the same rationale as the optimal precoding design since the matched filter solution (26) and (33) has the effect of maximizing the signal energy while suppressing the noise energy.

It is interesting to examine how well this heuristic precoding design performs. We consider a homogeneous scenario where $\mathbf{H}_n=\mathbf{H}$, and $\mathbf{R}_{w_n}=\mathbf{R}_w$ for all n. Also, we assume an equal power allocation throughout our following discussion. The deflection coefficient is then given by

$$\begin{aligned}\chi &= \sum_{n=1}^{N}\left(\mathbf{c}_n\mathbf{R}_w\mathbf{c}_n^T+\bar{\sigma}_{v_n}^2\right)^{-1}\mathbf{c}_n\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{c}_n^T\\ &\triangleq \sum_{n=1}^{N}\chi_n\end{aligned} \tag{52}$$

in which

$$\chi_n \triangleq \left(\mathbf{c}_n\mathbf{R}_w\mathbf{c}_n^T+\bar{\sigma}_{v_n}^2\right)^{-1}\mathbf{c}_n\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{c}_n^T \tag{53}$$

denotes the individual deflection coefficient for each sensor. The precoding vector, at the same time, has to satisfy the transmit power constraint

$$\mathbf{c}_n\mathbf{R}_w\mathbf{c}_n^T + P_1\mathbf{c}_n\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{c}_n^T = T \tag{54}$$

where $T=T_{\mathrm{total}}/N$ since we assume an equal power allocation. Define

$$\upsilon_n \triangleq \mathbf{c}_n\mathbf{R}_w\mathbf{c}_n^T \quad \mu_n \triangleq \frac{\mathbf{c}_n\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{c}_n^T}{\mathbf{c}_n\mathbf{R}_w\mathbf{c}_n^T}. \tag{55}$$

The individual deflection coefficient can be reexpressed in terms of υ_n and μ_n

$$\chi_n = \frac{\upsilon_n\mu_n}{\upsilon_n+\bar{\sigma}_{v_n}^2} \tag{56}$$

and the power constraint (54) can be rewritten as

$$\upsilon_n(1+P_1\mu_n)=T \tag{57}$$

Solving υ_n from the power constraint (57), and substituting it back into (56), we arrive at

$$\chi_n = \frac{T\mu_n}{T+\bar{\sigma}_{v_n}^2(1+P_1\mu_n)} \tag{58}$$

Clearly, the individual deflection coefficient χ_n is a monotonically increasing function of μ_n. If $\boldsymbol{\theta}$ is known, the maximum μ_n attained by the optimal precoding vector is equal to

$$\begin{aligned}\mu_{\max} &= \lambda_{\max}\left(\mathbf{R}_w^{-\frac{1}{2}}\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T\mathbf{R}_w^{-\frac{1}{2}}\right)=\mathbf{g}^T\mathbf{g}\\ &= \sum_{i=1}^{q} g_i^2\end{aligned} \tag{59}$$

where $\mathbf{g}\triangleq \mathbf{R}_w^{-\frac{1}{2}}\mathbf{H}\boldsymbol{\theta}$, q is the dimension of $\mathbf{g}$. On the other hand, for the heuristic precoding design (50), μ_n is given by

$$\begin{aligned}\mu_n &= \frac{(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}))^T\mathbf{g}\mathbf{g}^T(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}))}{(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}))^T(|\mathbf{r}_n|\odot \mathrm{sgn}(\mathbf{g}))}\\ &= \frac{\left(\sum_{i=1}^{q}|r_{n_i}g_i|\right)^2}{\sum_{i=1}^{q}r_{n_i}^2}.\end{aligned} \tag{60}$$

For notational convenience, let χ_{sub} and χ_{opt}, respectively, denote the overall deflection coefficients attained by the heuristic precoding design and the optimal precoding design. The ratio of these two deflection coefficients is then given as

$$\begin{aligned}\frac{\chi_{\mathrm{sub}}}{\chi_{\mathrm{opt}}} &= \frac{\sum_n \frac{T\mu_n}{T+\bar{\sigma}_{v_n}^2(1+P_1\mu_n)}}{\sum_n \frac{T\mu_{\max}}{T+\bar{\sigma}_{v_n}^2(1+P_1\mu_{\max})}} > \frac{\sum_n \frac{T\mu_n}{T+\bar{\sigma}_{v_n}^2(1+P_1\mu_{\max})}}{\sum_n \frac{T\mu_{\max}}{T+\bar{\sigma}_{v_n}^2(1+P_1\mu_{\max})}}\\ &\overset{(a)}{=} \frac{1}{\mu_{\max}}\frac{\sum_n\xi_n\mu_n}{\sum_n\xi_n} \overset{a.s}{\to} \frac{1}{\mu_{\max}}\frac{1}{\sum_n\xi_n}E\left[\sum_n\xi_n\mu_n\right]\\ &= \frac{1}{\mu_{\max}}E[\mu_n]=E[\mu_n/\mu_{\max}]\end{aligned} \tag{61}$$

where in (a), we define $\xi_n \triangleq \frac{T}{T+\bar{\sigma}^2_{v_n}(1+P_1\mu_{\max})}$, the approximation in the second line comes from the Law of Large Numbers for independent but non identically distributed random variables: the sample average converges almost surely to the expected value, i.e., $\bar{X} \triangleq \frac{1}{n}(X_1 + \cdots + X_n) \overset{a.s}{\rightarrow} E[\bar{X}]$. Note that $\{\mu_n\}$ are i.i.d., hence $\{\xi_n \mu_n\}$ are independent but nonidentically distributed. We see that the ratio converges to $E[\mu_n/\mu_{\max}]$ as the number of sensors increases. Utilizing (59)–(60), we have

$$
\begin{aligned}
E\left[\frac{\mu_n}{\mu_{\max}}\right] &= E\left[\frac{\left(\sum_{i=1}^{q} |r_{n_i} g_i|\right)^2}{\left(\sum_{i=1}^{q} r_{n_i}^2\right)\left(\sum_{i=1}^{q} g_i^2\right)}\right] \\
&\geq \frac{1}{\sum_{i=1}^{q} g_i^2} \sum_{i=1}^{q} \left\{ g_i^2 E\left[\frac{r_{n_i}^2}{\sum_{i=1}^{q} r_{n_i}^2}\right]\right\} \\
&\approx \frac{1}{\sum_{i=1}^{q} g_i^2} \sum_{i=1}^{q} \left\{ g_i^2 E\left[r_{n_i}^2\right] E\left[\frac{1}{\sum_{i=1}^{q} r_{n_i}^2}\right]\right\} \\
&= \frac{\sum_{i=1}^{q} g_i^2}{(q-2)\left(\sum_{i=1}^{q} g_i^2\right)} > \frac{1}{q}
\end{aligned} \tag{62}
$$

where the last equality comes from the fact that $\{r_{n_i}^2\}$ are i.i.d. chi-square random variables with one degree-of-freedom. Combining (61)–(62), we conclude that the ratio of the deflection coefficient achieved by the precoding design (50) to that attained by the optimal precoding design is within $1/q$.

Simulations are conducted to illustrate the performance of the GLRT with precoding design (50) (denoted as GLRT-precoding), and its comparison with the GLRT with no precoding (that is, $\mathbf{C}_n = \mathbf{I}, \forall n$), and the Neyman-Pearson test which assumes the knowledge of $\boldsymbol{\theta}$ and employs optimal precoding design (denoted as NP-OP). In our simulations, we set $P_1 = 0$, $\mathbf{H}_n = \mathbf{I}$, $\mathbf{R}_{w_n} = 0.5\mathbf{I}$, and $\sigma^2_{v_n} = 1$ for all n, and $\boldsymbol{\theta} = [\cos(1) \;\; \cos(2) \;\; \cos(3)]^T$. The absolute coefficients $\{|\gamma_n|\}$ are assumed i.i.d. Rayleigh-fading random variables with unit variance. There are 100 sensors. The false alarm probability is set to $P_{\text{FA}} = 0.05$. The detection probabilities of the GLRT and NP-OP are shown in Fig. 6. We see that GLRT with precoding (50) presents a clear performance advantage over GLRT with no precoding. This suggests that a properly designed precoding, even not optimal, is more energy-efficient than no precoding. Also, it can be observed that to achieve a same detection performance, the GLRT with precoding requires about twice of the transmit power needed by NP-OP.

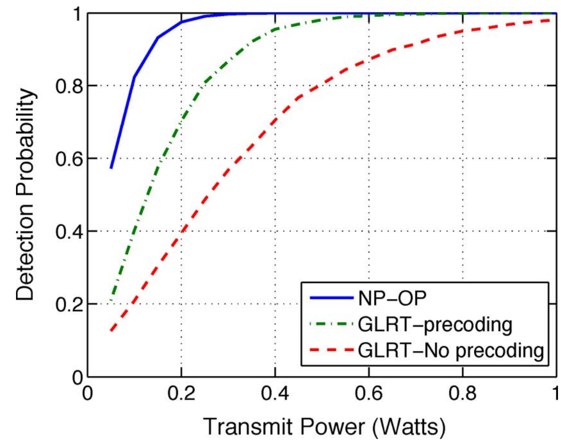


Fig. 6. Detection probability versus total transmit power for GLRT with precoding (50) and no precoding, and NP test with optimal precoding (OP).

VII. Conclusion

We considered a decentralized detection problem in which a number of sensors collaborate to detect the presence of a deterministic vector signal. The sensor network is subject to a total power constraint, and each sensor uses an analog amplify-and-forward transmission scheme to send their data to the FC. In this context, we studied the optimal precoding design for each sensor, aiming at minimizing the probability of detection error at the FC. Our study found that the optimal precoder has a form of a matched filter which converts each sensor's original measurements into a single message, and depending on the channel characteristics, one or multiple copies of this compressed message should be transmitted to the FC. More specifically, if the channel matrix is diagonal, then only one message needs to be sent, otherwise multiple versions of the compressed message which are multiplied by different amplification factors should be transmitted to the FC. Note that although matched filter detection is a well-studied topic, its optimality in a distributed power-constrained network has never been established before.

Given a fixed power constraint, the impact of the number of sensors on the overall detection performance was analyzed. Numerical results showed that a substantial performance improvement can be achieved by exploiting channel diversity. Besides, the concept "outage probability" was introduced to quantify the system detection reliability. Our analysis suggests that if a certain condition is satisfied, then the outage probability can be made arbitrarily small by increasing the number of sensors. Finally, a GLRT detector and a heuristic precoding design were proposed when the exact knowledge of the signal to be detected is not available. Numerical results were provided to illustrate its performance and its comparison with the Neyman-Pearson detector which assumes the knowledge of the signal.

The analog amplify-and-forward scheme considered in this paper, albeit simple, is not the best scheme in terms of energy-efficiency. Other digital communication schemes (e.g., QAM and QPSK) could provide a better receiver quality at a lower energy cost. Nevertheless, the optimal precoding and transmission strategies for digital communications schemes are still unclear. This is a topic worthy of future investigation. We believe that the formulations and the approach adopted in [23] are helpful to tackle the power-constrained decentralized detection problem with digital communication schemes. In addition, precoding design for the unknown signal case and its corresponding performance of the GLRT detector is lightly touched and deserves future study.

APPENDIX A
PROOF OF LEMMA 1

Let $a_i \triangleq d_{ii}^2/\bar{\sigma}_{v_n}^2$, and $b_i \triangleq P_1 f_{ii}$. The optimization (23) can be rewritten as

$$\begin{aligned} \max_{\{a_i,b_i\}} \quad & \sum_{i=1}^{q_n} \frac{a_i b_i}{a_i+1} \Longleftrightarrow \min_{\{a_i,b_i\}} \quad \sum_{i=1}^{q_n} \frac{b_i}{a_i+1} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} a_i(1+b_i) = \frac{T_n}{\bar{\sigma}_{v_n}^2} \triangleq \tilde{T}_n \\ & a_i \geq 0 \quad \forall i \\ & b_i \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} b_i = P_1 \lambda_{\max}(\mathbf{G}_n) \triangleq \lambda. \end{aligned} \tag{63}$$

The above optimization involves optimizing two sets of variables $\{a_i\}$ and $\{b_i\}$. To solve (63), we first optimize one set of variables, given that the other set of variables are fixed. Suppose that $\{b_i\}$ are predetermined, and are arranged in a descending order, i.e., $b_1 \geq b_2 \geq \cdots \geq b_{q_n}$. Then optimizing $\{a_i\}$ conditional on fixed $\{b_i\}$ can be formulated as

$$\begin{aligned} \min_{\{a_i\}} \quad & \sum_{i=1}^{q_n} \frac{b_i}{a_i+1} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} a_i(1+b_i) = \tilde{T}_n \\ & a_i \geq 0 \quad \forall i \end{aligned} \tag{64}$$

which can be analytically solved by resorting to the Lagrangian function and Karush-Kuhn-Tucker (KKT) conditions (details are elaborated in Appendix B). The optimal solution is given by

$$a_i = \left[\sqrt{\frac{b_i}{\phi(1+b_i)}} - 1\right]^+ \quad \forall i \tag{65}$$

where $[x]^+$ is equal to x if $x > 0$, otherwise it is zero; ϕ is a parameter that is uniquely determined from the procedure described in Appendix B.

Let $\{a_i^*(\mathbf{b})\}$ denote the optimal solution conditional on given $\mathbf{b} \triangleq [b_1 \ b_2 \ \ldots \ b_{q_n}]$. Substituting $a_i^*(\mathbf{b})$ back into (63), we come to an optimization involving only $\{b_i\}$:

$$\begin{aligned} \min_{\{b_i\}} \quad & \sum_{i=1}^{q_n} \frac{b_i}{a_i^*(\mathbf{b})+1} \\ \text{s.t.} \quad & b_i \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} b_i = \lambda. \end{aligned} \tag{66}$$

In the following, we show that the optimal solution to (66) is given by

$$b_i^* = \begin{cases} \lambda & i=1 \\ 0 & \text{otherwise.} \end{cases} \tag{67}$$

Notice that the parameter ϕ in (65) needs to be determined through an iterative search. Therefore we cannot directly substitute the solution of $a_i^*(\mathbf{b})$ into (66). To make the problem tractable, we start from a two-dimensional case $q_n = 2$. The extension to arbitrary dimension q_n can be accomplished based on the two-dimensional results, which will be shown later. Define

$$\begin{aligned} \pi(\mathbf{b}) &\triangleq \sum_{i=1}^{2} \frac{b_i}{a_i^*(\mathbf{b})+1} \\ \mathbf{b}^{(0)} &\triangleq [\lambda \ \ 0]. \end{aligned} \tag{68}$$

In Appendix C, we proved that $\mathbf{b}^{(0)}$ is the optimal solution to (66) for $q_n = 2$, that is

$$\pi(\mathbf{b}^{(0)}) < \pi(\mathbf{b}) \tag{69}$$

for any $\mathbf{b} \neq \mathbf{b}^{(0)}$ satisfying the constraints defined in (66). Therefore for $q_n = 2$, the optimal solution to (63) is given by

$$\begin{aligned} \{b_1^*, b_2^*\} &= \{\lambda, 0\} \\ \{a_1^*, a_2^*\} &= \left\{\frac{\tilde{T}_n}{1+\lambda}, 0\right\}. \end{aligned} \tag{70}$$

In other words, we have

$$\frac{\lambda}{\frac{\tilde{T}_n}{1+\lambda}+1} \leq \sum_{i=1}^{2} \frac{b_i}{a_i+1} \tag{71}$$

for any $\{a_1, a_2, b_1, b_2\}$ satisfying $a_i > 0$, $b_i > 0$, $\forall i$, and $b_1 + b_2 = \lambda$, $a_1(1+b_1) + a_2(1+b_2) = \tilde{T}_n$.

We now discuss the generalization of our results to arbitrary dimensional case. Again, suppose that $\{b_i\}$ are arranged in descending order, and let $\tilde{T}_{n,i} \triangleq a_i(1+b_i)$. Then the objective function of (63) is lower bounded by

$$\begin{aligned} \sum_{i=1}^{q_n} \frac{b_i}{a_i+1} &= \sum_{i=1}^{2} \frac{b_i}{a_i+1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i+1} \\ &\overset{(a)}{\geq} \frac{\ddot{b}_1}{\frac{\ddot{T}_{n,1}}{1+\ddot{b}_1}+1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i+1} \end{aligned} \tag{72}$$

in which $\ddot{b}_1 \triangleq b_1 + b_2$, $\ddot{T}_{n,1} \triangleq \tilde{T}_{n,1} + \tilde{T}_{n,2}$, and the inequality (a) comes by utilizing (71). The above objective function can be further lower bounded as

$$\begin{aligned} \sum_{i=1}^{q_n} \frac{b_i}{a_i+1} &\geq \frac{\ddot{b}_1}{\frac{\ddot{T}_{n,1}}{1+\ddot{b}_1}+1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i+1} \\ &= \frac{\ddot{b}_1}{\frac{\ddot{T}_{n,1}}{1+\ddot{b}_1}+1} + \frac{b_3}{a_3+1} + \sum_{i=4}^{q_n} \frac{b_i}{a_i+1} \\ &\geq \frac{\ddot{b}_2}{\frac{\ddot{T}_{n,2}}{1+\ddot{b}_2}+1} + \sum_{i=4}^{q_n} \frac{b_i}{a_i+1} \end{aligned} \tag{73}$$

in which $\ddot{b}_2 \triangleq b_1 + b_2 + b_3$, $\ddot{T}_{n,2} \triangleq \tilde{T}_{n,1} + \tilde{T}_{n,2} + \tilde{T}_{n,3}$, and the inequality, again, comes by using (71). So on and so forth, we can reach that the objective function is eventually lower bounded by

$$\sum_{i=1}^{q_n} \frac{b_i}{a_i+1} \geq \frac{\lambda}{\frac{\tilde{T}_n}{1+\lambda}+1} \tag{74}$$

and this lower bound is attained only when

$$b_i^* = \begin{cases} \lambda & i=1 \\ 0 & \text{otherwise} \end{cases} \quad (75)$$

$$a_i^* = \begin{cases} \frac{\tilde{T}_n}{1+\lambda} & i=1 \\ 0 & \text{otherwise.} \end{cases} \quad (76)$$

Therefore (75)–(76) are the optimal solution to (63). The proof is completed here.

Appendix B
An Analytical Solution to (64)

The Lagrangian function associated with (64) is given by

$$L(a_i;\phi;\nu_i) = \sum_{i=1}^{q_n} \frac{b_i}{a_i+1} - \phi\left(\tilde{T}_n - \sum_{i=1}^{q_n} a_i(1+b_i)\right) - \sum_{i=1}^{q_n} \nu_i a_i \quad (77)$$

which gives the following KKT conditions [27]:

$$-\frac{b_i}{(a_i+1)^2} + \phi(1+b_i) - \nu_i = 0 \quad \forall i$$

$$\tilde{T}_n - \sum_{i=1}^{q_n} a_i(1+b_i) = 0$$

$$\nu_i a_i = 0 \quad \forall i$$

$$\nu_i \geq 0 \quad \forall i$$

$$a_i \geq 0 \quad \forall i.$$

By solving the first equation of the above KKT conditions, we obtain

$$a_i = \sqrt{\frac{b_i}{\phi(1+b_i) - \nu_i}} - 1 \quad \forall i \quad (78)$$

Also, the KKT conditions: $\nu_i a_i = 0$, $\nu_i \geq 0$, and $a_i \geq 0$ imply that we have either $\{\nu_i = 0, a_i > 0\}$ or $\{\nu_i > 0, a_i = 0\}$. Therefore (78) becomes

$$a_i = \left[\sqrt{\frac{b_i}{\phi(1+b_i)}} - 1\right]^+ \quad \forall i \quad (79)$$

where $[x]^+$ is equal to x if $x > 0$, otherwise it is zero. The Lagrangian multiplier ϕ and the number of nonzero elements $(a_i > 0)$ can be uniquely determined from the second equation of the KKT conditions. The procedure is described as follows.

Suppose we have $K \in \{1,\ldots,q_n\}$ nonzero elements, i.e., $a_i > 0, \forall i = 1,\ldots,K$ (note that $\{a_i\}$ are in descending order since we assume $b_1 \geq b_2 \geq \cdots \geq b_{q_n}$). Therefore ϕ can be solved by substituting $\{a_1, a_2, \ldots, a_K\}$ into the second KKT condition:

$$\phi = \left[\frac{\sum_{i=1}^K \sqrt{b_i(1+b_i)}}{\tilde{T}_n + \sum_{i=1}^K (1+b_i)}\right]^2. \quad (80)$$

Now substituting ϕ back to (79), we get a new solution $\{a'_1, a'_2, \ldots, a'_K, a'_{K+1}, \ldots, a'_{q_n}\}$. If for this new solution, we have $a_i = 0$ for $i > K$, then it is the true solution we are looking for; otherwise we have to choose another K to repeat the above procedure.

Appendix C
Proof of Inequality (69)

Note that for the two-dimensional case, the feasible region $\{\mathbf{b} = [b_1 \;\; b_2]\}$ of the optimization problem (66) is in fact a line segment between the two points $[\lambda \;\; 0]$ and $[\lambda/2 \;\; \lambda/2]$ (note that we assume $b_1 \geq b_2$ without loss of generality). Let $\mathcal{R}$ denote the set which consists of all feasible solutions except $\mathbf{b}^{(0)}$. We divide the region $\mathcal{R}$ into two disjoint regions. One of the two disjoint regions is defined as

$$\mathcal{R}_1 \triangleq \{\mathbf{b} = [\lambda - \delta \;\; \delta] \quad | \quad \delta \in (0, \min(\lambda/2, \tau))\} \quad (81)$$

where $\tau > 0$ is a threshold such that if $\delta < \tau$, then the optimal solution $\{a_i\}$ to (64) conditional on $\mathbf{b} \in \mathcal{R}_1$ has the following form:

$$\mathbf{a}^*(\mathbf{b}) = [a_1^*(\mathbf{b}) \;\; 0]. \quad (82)$$

Note that δ has to be smaller than $\lambda/2$ to ensure that $\{b_i\}$ are arranged in descending order. If $\tau \geq \lambda/2$, then $\mathcal{R}_1 = \mathcal{R}$. For the case $\tau < \lambda/2$, the complementary region is given by

$$\mathcal{R}_2 \triangleq \{\mathbf{b} = [\lambda - \delta \;\; \delta] \quad | \quad \delta \in [\tau, \lambda/2]\}. \quad (83)$$

It can be easily verified that $\mathcal{R}_1 \cup \mathcal{R}_2 = \mathcal{R}$. Clearly, the two disjoint regions are obtained by breaking the line segment into two pieces, with $\mathcal{R}_1$ corresponding to the line segment between the points $[\lambda \;\; 0]$ and $[\lambda - \tau \;\; \tau]$ (end points are not included), and $\mathcal{R}_2$ corresponding to the line segment between $[\lambda - \tau \;\; \tau]$ and $[\lambda/2 \;\; \lambda/2]$.

To prove that $\mathbf{b}^{(0)}$ is the optimal solution to (66), we first show that $\pi(\mathbf{b}^{(0)}) < \pi(\mathbf{b})$ for any $\mathbf{b} \in \mathcal{R}_1$. It is easy to derive that the optimal solutions $\{a_i^*(\mathbf{b})\}$ conditional on $\mathbf{b}^{(0)}$ and $\mathbf{b} \in \mathcal{R}_1$ are, respectively, given as

$$\left\{a_i^*\left(\mathbf{b}^{(0)}\right)\right\} = \left\{\frac{\tilde{T}_n}{1+\lambda}, 0\right\}$$

$$\{a_i^*(\mathbf{b})\} = \left\{\frac{\tilde{T}_n}{1+\lambda-\delta}, 0\right\}. \quad (84)$$

Substituting the optimal solution $\{a_i^*(\mathbf{b})\}$ into (68), we have

$$\pi\left(\mathbf{b}^{(0)}\right) = \frac{\lambda(1+\lambda)}{\tilde{T}_n + 1 + \lambda}$$

$$\pi(\mathbf{b}) = \frac{(\lambda-\delta)(1+\lambda-\delta)}{\tilde{T}_n + 1 + \lambda - \delta} + \delta \quad (85)$$

and

$$\pi(\mathbf{b}) - \pi(\mathbf{b}^{(0)}) = \frac{\tilde{T}_n^2\delta + \tilde{T}_n\delta}{(\tilde{T}_n + 1 + \lambda - \delta)(\tilde{T}_n + 1 + \lambda)}. \quad (86)$$

Therefore for any $\mathbf{b} \in \mathcal{R}_1$ $(0 < \delta < \min(\lambda/2, \tau))$

$$\pi\left(\mathbf{b}^{(0)}\right) < \pi(\mathbf{b})$$

holds. Also, from (86), we know that $\pi(\mathbf{b})$ increases with an increasing δ. It means that from the starting point $[\lambda \;\; 0]$, when the point $\mathbf{b}$ comes closer to the end point $[\lambda - \tau \;\; \tau]$, the function value $\pi(\mathbf{b})$ increases.

We now prove $\pi(\mathbf{b}^{(0)}) < \pi(\mathbf{b})$ for any $\mathbf{b} \in \mathcal{R}_2$. We first show that for $\mathbf{b} \in \mathcal{R}_2$, $\pi(\mathbf{b})$ increases with an increasing δ. Note that the region $\mathcal{R}_2$ can be rewritten as

$$\mathcal{R}_2 \triangleq \{\mathbf{b} = [\lambda/2 + \delta' \ \ \lambda/2 - \delta'] \quad | \quad \delta' \in [0, \lambda/2 - \tau]\}. \quad (87)$$

Therefore proving that $\pi(\mathbf{b})$ increases with an increasing δ is equivalent to showing that $\pi(\mathbf{b})$ decreases with an increasing δ'. For any $\mathbf{b} \in \mathcal{R}_2$, the optimal solution $\{a_i\}$ of (64) conditional on (b) has the following form:

$$\mathbf{a}^*(\mathbf{b}) = [a_1^*(\mathbf{b}) \ \ a_2^*(\mathbf{b})] \quad (88)$$

where $a_i^*(\mathbf{b}) > 0$ for $i = 1, 2$. Substituting the optimal solution $\mathbf{a}^*(\mathbf{b})$ into $\pi(\mathbf{b})$, we have

$$\begin{aligned}\pi(\mathbf{b}) &= \sum_{i=1}^{2} \frac{b_i}{a_i^*(\mathbf{b}) + 1} \overset{(a)}{=} \sqrt{\phi} \sum_{i=1}^{2} \sqrt{b_i(1+b_i)} \\ &\overset{(b)}{=} \frac{\left(\sum_{i=1}^{2} \sqrt{b_i(1+b_i)}\right)^2}{\tilde{T}_n + \sum_{i=1}^{2}(1+b_i)} \\ &= \frac{\lambda + \lambda^2}{\tilde{T}_n + 2 + \lambda} + \frac{2}{\tilde{T}_n + 2 + \lambda}\kappa(\delta')\end{aligned} \quad (89)$$

where (a) comes by utilizing (65), (b) follows from (80), and

$$\begin{aligned}\kappa(\delta') &\triangleq \sqrt{b_1 b_2 (1 + \lambda + b_1 b_2)} - b_1 b_2 \\ &= \sqrt{\delta'^4 - \alpha\delta'^2 + \beta} - \frac{\lambda^2}{4} + \delta'^2 \\ \alpha &\triangleq 1 + \lambda + \frac{\lambda^2}{2} \\ \beta &\triangleq \frac{\lambda^4}{16} + \frac{\lambda^3}{4} + \frac{\lambda^2}{4}\end{aligned}$$

Let $t \triangleq \delta'^2$, and define

$$\tilde{\kappa}(t) \triangleq \sqrt{t^2 - \alpha t + \beta} - \frac{\lambda^2}{4} + t \quad (90)$$

We compute the first derivative of $\tilde{\kappa}(t)$:

$$\frac{\partial \tilde{\kappa}(t)}{\partial t} = \frac{2t - \alpha}{\sqrt{t^2 - \alpha t + \beta}} + 1 \quad (91)$$

It is easy to verify that for any $\lambda > 0$, and $\lambda^2/4 \geq t \geq 0$, we have

$$\begin{aligned}&\alpha - 2t > \sqrt{t^2 - \alpha t + \beta} \\ \Rightarrow &\frac{\partial \tilde{\kappa}(t)}{\partial t} < 0\end{aligned} \quad (92)$$

Therefore $\tilde{\kappa}(t)$ is a monotonically decreasing function of t for $\lambda^2/4 \geq t \geq 0$. Consequently, $\kappa(\delta')$ decreases with an increasing δ' for $\lambda/2 \geq \delta' \geq 0$, so does the function $\pi(\mathbf{b})$. In other words, for $\mathbf{b} \in \mathcal{R}_2$, $\pi(\mathbf{b})$ increases with an increasing δ. It means that from the starting point $[\lambda - \tau \ \ \tau]$, when the point $\mathbf{b}$ approaches the end point $[\lambda/2 \ \ \lambda/2]$, the function value $\pi(\mathbf{b})$ increases. Due to the continuity of the function $\pi(\mathbf{b})$, hence we have

$$\pi\left(\mathbf{b}^{(0)}\right) < \pi\left(\mathbf{b}^{(1)}\right) < \pi\left(\mathbf{b}^{(2)}\right) \quad (93)$$

for any $\mathbf{b}^{(1)} \in \mathcal{R}_1$ and $\mathbf{b}^{(2)} \in \mathcal{R}_2$. The proof is completed here.

APPENDIX D
EXTENSION TO NON-SQUARE CHANNEL MATRIX

To see that the results in Theorem 2 hold valid for non-square channel matrix, we consider two different cases.

- If $r > q_n$, the SVD of $\mathbf{\Gamma}_n$ can be written as

$$\mathbf{\Gamma}_n = \mathbf{U}_{\gamma_n} \begin{bmatrix} \mathbf{D}_{\gamma_n} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_{\gamma_n}^T$$

  where $\mathbf{U}_{\gamma_n} \in \mathbb{R}^{r \times r}$, $\mathbf{D}_{\gamma_n} \in \mathbb{R}^{q_n \times q_n}$, and $\mathbf{V}_{\gamma_n} \in \mathbb{R}^{q_n \times q_n}$. Substituting the SVD of $\mathbf{\Gamma}_n$ into (17), we reach the same optimization (29).
- If $r < q_n$, the SVD of $\mathbf{\Gamma}_n$ can be written as

$$\mathbf{\Gamma}_n = \mathbf{U}_{\gamma_n} [\mathbf{D}_{\gamma_n} \quad \mathbf{0}] \mathbf{V}_{\gamma_n}^T$$

  where $\mathbf{U}_{\gamma_n} \in \mathbb{R}^{r \times r}$, $\mathbf{D}_{\gamma_n} \in \mathbb{R}^{r \times r}$, and $\mathbf{V}_{\gamma_n} \in \mathbb{R}^{q_n \times q_n}$. Substitute the SVD into (17), and partition $\tilde{\mathbf{C}}_n$ into two parts:

$$\tilde{\mathbf{C}}_n = \begin{bmatrix} \tilde{\mathbf{C}}_{n,1} \\ \tilde{\mathbf{C}}_{n,2} \end{bmatrix}$$

  where $\tilde{\mathbf{C}}_{n,1} \in \mathbb{R}^{r \times q_n}$, and $\tilde{\mathbf{C}}_{n,2} \in \mathbb{R}^{(q_n - r) \times q_n}$. The optimization (17) can be written as

$$\begin{aligned}\max_{\tilde{\mathbf{C}}_n} \quad & \mathrm{tr}\left\{\left(\tilde{\mathbf{C}}_{n,1}\tilde{\mathbf{C}}_{n,1}^T + \sigma_{v_n}^2 \mathbf{D}_{\gamma_n}^{-2}\right)^{-1} \tilde{\mathbf{C}}_{n,1}\mathbf{G}_n\tilde{\mathbf{C}}_{n,1}^T\right\} \\ \text{s.t.} \quad & \mathrm{tr}\left\{\tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T + P_1\tilde{\mathbf{C}}_n\mathbf{G}_n\tilde{\mathbf{C}}_n^T\right\} = T_n\end{aligned} \quad (94)$$

  Since $\tilde{\mathbf{C}}_{n,2}$ has nothing to do with the objective function, it should be set to a null matrix to save the energy. Consequently we arrive at an optimization that is equivalent to (29).

APPENDIX E
AN ANALYTICAL SOLUTION TO (34)

For notational convenience, let $\lambda_{\max,n}$ stand for $\lambda_{\max}(\mathbf{G}_n)$. Define

$$\begin{aligned}\alpha_n &\triangleq \frac{\lambda_{\max,n}}{\bar{\sigma}_{v_n}^2 (1 + P_1\lambda_{\max,n})} \\ \beta_n &\triangleq \frac{1}{\bar{\sigma}_{v_n}^2 (1 + P_1\lambda_{\max,n})}\end{aligned}$$

The Lagrangian function L associated with (34) is given by

$$\begin{aligned}&L(T_n; \phi; \nu_n) \\ &= -\sum_{n=1}^{N} \frac{\alpha_n T_n}{\beta_n T_n + 1} - \phi\left(T_{\text{total}} - \sum_{n=1}^{N} T_n\right) - \sum_{n=1}^{N} \nu_n T_n\end{aligned} \quad (95)$$

which gives the following KKT conditions [27]:

$$\begin{aligned}-\frac{\alpha_n}{(\beta_n T_n + 1)^2} + \phi - \nu_n = 0 \quad &\forall n \\ T_{\text{total}} - \sum_{n=1}^{N} T_n = 0 & \\ \nu_n T_n = 0 \quad &\forall n \\ \nu_n \geq 0 \quad &\forall n \\ T_n \geq 0 \quad &\forall n\end{aligned}$$

By solving the first equation of the above KKT conditions, we obtain

$$T_n = \frac{1}{\beta_n}\left[\sqrt{\frac{\alpha_n}{\phi - \nu_n}} - 1\right] \quad \forall n \tag{96}$$

Also, the KKT conditions: $\nu_n T_n = 0$, $\nu_n \geq 0$, and $T_n \geq 0$ imply that we have either $\{\nu_n = 0, T_n > 0\}$ or $\{\nu_n > 0, T_n = 0\}$. Therefore (96) becomes

$$T_n = \frac{1}{\beta_n}\left[\sqrt{\frac{\alpha_n}{\phi}} - 1\right]^+ \quad \forall n \tag{97}$$

where $[x]^+$ is equal to x if $x > 0$, otherwise it is zero. The Lagrangian multiplier ϕ and the number of active sensors (those are assigned nonzero power) can be uniquely determined from the power constraint.

Suppose we have $K \in \{1, \ldots, N\}$ active nodes, according to (97), these K nodes must be $\{k_1, k_2, \ldots, k_K\}$, where $\{k_i\}$ is a set of indices such that $\alpha_{k_1} \geq \alpha_{k_2} \geq \cdots \geq \alpha_{k_N}$. Therefore ϕ can be solved by substituting $\{T_{k_1}, T_{k_2}, \ldots, T_{k_K}\}$ into the second KKT condition, where T_{k_i} is given by

$$T_{k_i} = \frac{1}{\beta_{k_i}}\left[\sqrt{\frac{\alpha_{k_i}}{\phi}} - 1\right]. \tag{98}$$

Now we substitute ϕ back to (97). We will get a new solution $\{T'_{k_1}, T'_{k_2}, \ldots, T'_{k_K}, T'_{k_{K+1}}, \ldots, T'_{k_N}\}$. If this new solution is exactly identical to the one we assumed before, i.e., $\{T_{k_1}, T_{k_2}, \ldots, T_{k_K}, 0, \ldots, 0\}$, then it is the true solution we are looking for; otherwise we have to choose another K to repeat the above procedure. Also, it has been proved that such a solution is unique and always exists [23].

REFERENCES

[1] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, pp. 501–510, Jul. 1981.

[2] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 24–30, Jan. 1987.

[3] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, pp. 995–1000, Sep. 1989.

[4] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, pp. 761–765, Sep 1989.

[5] J. N. Tsistsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing, Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993, vol. 2.

[6] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[7] 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.

[8] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[9] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad, and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3266–3279, Dec. 2000.

[10] Q. Yan and R. S. Blum, "Distributed signal detection under the Neyman-Pearson criterion," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1368–1377, May 2001.

[11] J. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.

[12] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[13] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2617–2624, Aug. 2005.

[14] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[15] H.-S. Kim, J. Wang, P. Cai, and S. Cui, "Detection outage and detection diversity in a homogeneous distributed sensor network," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2875–2881, Jul. 2009.

[16] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[17] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Energy-efficient detection in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 693–702, Apr. 2005.

[18] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.

[19] P. Bianchi and J. Jakubowicz, "Linear precoders for the detection of a Gaussian process in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 882–894, Mar. 2011.

[20] Y. Yang, R. S. Blum, and B. M. Sadler, "A distributed energy-efficient framework for Neyman-Pearson detection of fluctuating signals in large-scale sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1149–1158, Sep. 2010.

[21] I. D. 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.

[22] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3833–3832, Jul. 2009.

[23] 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.

[24] S. Cui, J.-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.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice-Hall, 1998.

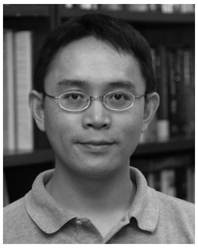
[26] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones and Bartlett, 1993.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

Jun Fang (M'08) received the B.Sc. degree and M.Sc. degrees in electrical engineering from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree in electrical engineering from National University of Singapore, Singapore, in 2006.

During 2006, he was with the Department of Electrical and Computer Engineering, Duke University, as a Postdoctoral Research Associate. From January 2007 to December 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ. Since January 2011, he has been with the National Key Laboratory on Communications, University of Electronic of Science and Technology of China. His research interests include statistical signal processing, distributed signal processing for wireless sensor networks, and wireless communications. He has authored/coauthored more than 20 papers in leading IEEE journals.

Dr. Fang received the Outstanding Paper Award at the 2011 IEEE Africon Conference (coauthored with P. Wang, H. Li, and N. Han). He served as a Guest Editor for IEEE *Communications Magazine*, and was a TPC member for the 11th and 13th IEEE International Conference on Communication Systems.



Hongbin Li (M'99) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering, University of Florida. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory, Rome, NY, in summers 2003 and 2004. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is a Full Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for Excellence In Research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He is a member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He is/has been an Editor or Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING, and served as a Guest Editor for EURASIP *Journal on Applied Signal Processing*, Special Issue on Distributed Signal Processing Techniques for Wireless Sensor Networks.

Zhi Chen received the B.Eng., M.Eng., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC), in 1997, 2000, and 2006, respectively.

In April 2006, he joined the National Key Laboratory on Communications (NCL), UESTC, where he is an Associate Professor. He was a visiting scholar at the University of California, Riverside, during 2010–2011. His current research interests include relay and cooperative communications, multiuser beamforming in cellular networks, interference coordination and cancellation, THz communication.

Dr. Chen has served as a reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Shaoqian Li received the B.S.E. degree in communication technology from Northwest Institute of Telecommunication (renamed Xidian University) in 1982 and the M.S.E. degree from the University of Electronic Science and Technology of China (UESTC) in 1984.

He is now a Professor and the Director of National Key Lab of Communication in UESTC. He is also a member of National High Technology R&D Program, which is also known as 863 Program Communications Group. His research interests include wireless information and communication theory, mobile and personal communications, anti-interference technology in wireless communications, spread-spectrum, and frequency-hopping techniques.