

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

Jun Fang, Hongbin Li, *Senior Member, IEEE*, and Shaoqian Li

Abstract— We consider a decentralized detection problem in a power-constrained wireless sensor networks (WSNs), 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*. Given a constraint on the total amount of transmit power, we investigate the optimal linear precoding design for each sensor node. More specifically, in order to achieve the best detection performance, shall sensor nodes transmit their raw data to the fusion center (FC), or transmit compressed versions of their original data? The optimal power allocation among sensors is studied as well. Also, assuming a fixed total transmit power, we examine how the detection performance behaves with the number of sensors in the network. A new concept “detection outage” is proposed to quantify the reliability of the overall detection system. Finally, decentralized detection with unknown signals is studied. Numerical results are conducted to corroborate our theoretical analysis and to illustrate the performance of the proposed algorithm.

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

I. INTRODUCTION

Decentralized detection is an important problem that has attracted much attention over the past decade [1]–[17]. 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]–[14] 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], [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. Also, in many works, it is assumed that the local binary decision can be reliably reported to the FC. This assumption may fail in wireless sensor networks as the information is transmitted over wireless links.

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 [16], 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 [16], 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.

Instead, we are interested in examining the following fundamental question: shall each node transmit its raw data to the FC, or shall each node send a compressed version of the original data to the FC? Since the total transmit power is fixed, sending more messages means that a single message is transmitted with less power, which results in a poor link quality. The FC, however, can collect more information from sensors. On the other hand, sending less messages renders a better channel quality, but with less information provided to the FC. The choice between these two strategies seems difficult before conducting a thorough mathematical analysis. This optimal precoding design problem will be investigated in this paper. Note that although linear precoding design for decentralized detection remains new, its counterpart for distributed estimation has been extensively investigated, e.g. [18], [19]. In addition, the asymptotic behavior of the overall detection performance with an increasing number of sensors is examined, and a generalized likelihood ratio test (GLRT) is proposed for the scenario of unknown signals. We noticed that the problem of decentralized detection in a power-constrained sensor network was also studied in [15], in which the optimal transmission mapping strategy was investigated in the asymptotic regime where the total transmit power tends to infinity.

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

Jun Fang and Shaoqian Li are with the National Key Laboratory on Communications, University of Electronic Science and Technology of China, Chengdu 610054, China, Emails: Jun.Fang@ieee.org; lsq@uestc.edu.cn

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

This work was supported in part by the National Science Foundation under Grant ECCS-0901066.

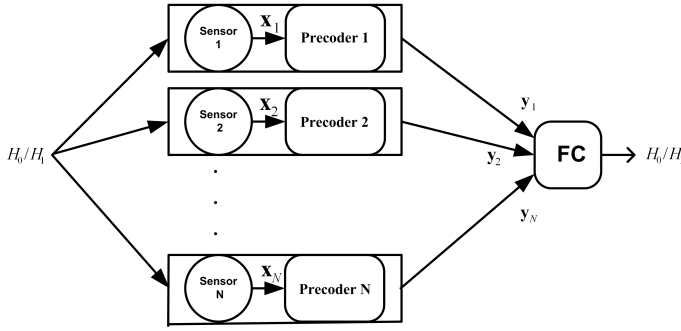


Fig. 1. Decentralized detection in a power-constrained network. Each node processes its vector observations through a linear precoder. Messages are then sent to the FC via wireless channels.

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 $\theta \in \mathbb{R}^p$. The binary hypothesis testing problem is formulated as follows:

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

where $\mathbf{H}_n \in \mathbb{R}^{q_n \times p}$ is the known observation matrix defining the input/output relation, $\mathbf{x}_n \in \mathbb{R}^{q_n}$ 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. 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 n th sensor is given by

$$\mathbf{y}_n = \mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n \quad n = 1, \dots, N \quad (2)$$

where \mathbf{v}_n represents the additive channel noise, and is assumed Gaussian with zero-mean and covariance matrix $\sigma_{v_n}^2 \mathbf{I}$. The channel matrix is implicitly set equal to an identity matrix as the multiplicative effect can be removed, given the knowledge of the channel state information.

The FC, based upon the received data $\{\mathbf{y}_n\}$, forms a final decision concerning the presence or absence of θ . 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 θ 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

$$\begin{aligned} & P_0 E[\|\mathbf{C}_n \mathbf{x}_n\|_2^2 | H_0] + P_1 E[\|\mathbf{C}_n \mathbf{x}_n\|_2^2 | H_1] \\ &= P_0 \text{tr} \{ \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \} \\ & \quad + P_1 \text{tr} \{ \mathbf{C}_n \mathbf{H}_n \theta \theta^T \mathbf{H}_n^T \mathbf{C}_n^T + \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T \} \\ &= \text{tr} \{ \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T + P_1 \mathbf{C}_n \mathbf{H}_n \theta \theta^T \mathbf{H}_n^T \mathbf{C}_n^T \} \end{aligned} \quad (3)$$

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 [15], 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 pre-specified, 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 $\mathbf{y} \triangleq [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \dots \ \mathbf{y}_N^T]^T$ denote the vector received at the FC, \mathbf{y}_n is a Gaussian random vector with its mean and covariance matrix given by

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

in which

$$\Sigma_n \triangleq \mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T + \sigma_{v_n}^2 \mathbf{I} \quad (5)$$

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

$$P_e = P(H_0 | H_1) P_1 + P(H_1 | H_0) P_0 \quad (6)$$

where $P(H_i | H_j)$ is the probability of deciding H_i when H_j is true. According to [20], in order to achieve a minimum P_e , the decision rule is a likelihood ratio test (LRT) given as follows:

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

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

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

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

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

where

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

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

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

where

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

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_e . Recalling results in the previous section, we know that the FC makes a global decision based on

$$u \underset{H_0}{\overset{H_1}{\geq}} \log \eta \quad (12)$$

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 [20]:

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

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

$$\chi = \sigma_u^2 \quad (14)$$

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_e is equivalent to

$$\max_{\{\mathbf{C}_n\}} \sigma_u^2 = \sum_{n=1}^N \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{C}_n^T (\mathbf{C}_n \mathbf{R}_{w_n} \mathbf{C}_n^T + \sigma_{v_n}^2 \mathbf{I})^{-1} \mathbf{C}_n \mathbf{H}_n \boldsymbol{\theta} \quad (15)$$

As we mentioned before, 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

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

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.

A. Optimum Precoding Design

Let us suppose, for the time being, that a power allocation is pre-specified and given as $\{T_1, T_2, \dots, T_N\}$. Then the optimum precoding matrix for each sensor can be obtained by solving

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

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. The optimization (17) is complicated in its current form. To make

the problem simplified, we, in the following, perform a series of matrix transformations. Define

$$\begin{aligned}\tilde{\mathbf{C}}_n &\triangleq \mathbf{C}_n \mathbf{R}_{w_n}^{\frac{1}{2}} \\ \mathbf{G}_n &\triangleq \mathbf{R}_{w_n}^{-\frac{1}{2}} \mathbf{H}_n \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{R}_{w_n}^{-\frac{1}{2}}\end{aligned}\quad (18)$$

and substitute them into (17), the optimization becomes

$$\begin{aligned}\max_{\tilde{\mathbf{C}}_n} \quad & \text{tr} \left\{ (\tilde{\mathbf{C}}_n \tilde{\mathbf{C}}_n^T + \sigma_{v_n}^2 \mathbf{I})^{-1} \tilde{\mathbf{C}}_n \mathbf{G}_n \tilde{\mathbf{C}}_n^T \right\} \\ \text{s.t.} \quad & \text{tr} \left\{ \tilde{\mathbf{C}}_n \tilde{\mathbf{C}}_n^T + P_1 \tilde{\mathbf{C}}_n \mathbf{G}_n \tilde{\mathbf{C}}_n^T \right\} = T_n\end{aligned}\quad (19)$$

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 non-negative diagonal elements, i.e. $d_{ii} \geq 0$. Substituting the SVD into (19), we arrive at a new optimization that searches for an optimal orthonormal matrix \mathbf{V} and an optimal diagonal matrix \mathbf{D} (\mathbf{U} can be any orthonormal matrix as it turns out that \mathbf{U} is independent of the optimization problem)

$$\begin{aligned}\max_{\{\mathbf{V}, \mathbf{D}\}} \quad & \text{tr} \left\{ \mathbf{D} (\mathbf{D}^2 + \sigma_{v_n}^2 \mathbf{I})^{-1} \mathbf{D} \mathbf{V}^T \mathbf{G}_n \mathbf{V} \right\} \\ \text{s.t.} \quad & \text{tr} \left\{ \mathbf{D}^2 + P_1 \mathbf{D}^2 \mathbf{V}^T \mathbf{G}_n \mathbf{V} \right\} = T_n\end{aligned}\quad (20)$$

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

$$\begin{aligned}\text{(i)} \quad & f_{ii} \geq 0 \\ \text{(ii)} \quad & \sum_{i=1}^{q_n} f_{ii} = \lambda_{\max}(\mathbf{G}_n)\end{aligned}\quad (21)$$

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 $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ and noting that \mathbf{G}_n is a rank-one matrix (c.f. (18)), where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of \mathbf{A} .

Treating \mathbf{F} as a new optimization variable, the optimization (20) can be re-expressed as

$$\begin{aligned}\max_{\{d_{ii}, f_{ii}\}} \quad & \sum_{i=1}^{q_n} \frac{d_{ii}^2 f_{ii}}{d_{ii}^2 + \sigma_{v_n}^2} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} d_{ii}^2 (1 + P_1 f_{ii}) = T_n \\ & f_{ii} \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} f_{ii} = \lambda_{\max}(\mathbf{G}_n)\end{aligned}\quad (22)$$

which, as we can see, involves only the diagonal elements of \mathbf{F} , while irrespective of its off-diagonal entries. The solution to (22) is given in the following lemma.

Lemma 1: The optimal solution to (22) is given by

$$f_{ii}^* = \begin{cases} \lambda_{\max}(\mathbf{G}_n) & i = 1 \\ 0 & \text{otherwise} \end{cases}\quad (23)$$

$$d_{ii}^* = \begin{cases} \sqrt{\frac{T_n}{1 + P_1 \lambda_{\max}(\mathbf{G}_n)}} & i = 1 \\ 0 & \text{otherwise} \end{cases}\quad (24)$$

Proof: See Appendix A. \blacksquare

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 (17), is a matrix with its first row a nonzero vector, whereas all other rows equal to zeros, i.e.

$$\mathbf{C}_n^*[i, :] = \begin{cases} \varphi \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{R}_{w_n}^{-1} & i = 1 \\ \mathbf{0} & i \in \{2, \dots, q_n\} \end{cases}\quad (25)$$

where

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

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 (24). From $\mathbf{F} = \mathbf{V}^T \mathbf{G}_n \mathbf{V}$, it is easy to deduce that the orthonormal matrix \mathbf{V} that yields (23) must be

$$\mathbf{V}^* = \mathbf{U}_{g_n}\quad (26)$$

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

$$\begin{aligned}\mathbf{C}_n^* &= \mathbf{D}^* \mathbf{U}_{g_n}^T \mathbf{R}_{w_n}^{-\frac{1}{2}} \\ &= \begin{cases} d_{11}^* (\mathbf{U}_{g_n}[:, 1])^T \mathbf{R}_{w_n}^{-\frac{1}{2}} & i = 1 \\ \mathbf{0} & i \in \{2, \dots, q_n\} \end{cases} \\ &= \begin{cases} \varphi \boldsymbol{\theta}^T \mathbf{H}_n^T \mathbf{R}_{w_n}^{-1} & i = 1 \\ \mathbf{0} & i \in \{2, \dots, q_n\} \end{cases}\end{aligned}\quad (27)$$

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. \blacksquare

Note that the optimal solution (25) has only one nonzero row. This suggests that in order to achieve best detection performance, a compression-transmission strategy, other than a non-compression transmission strategy, should be adopted, and 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.

B. Optimum Power Allocation

In previous subsection, we studied the optimum precoding design when a power assignment among sensors is specified.

It turns out that the optimal precoding is a compression vector which converts each sensor's observations into a single message. Substituting the optimum precoding vector back into (16), we obtain the following power allocation problem

$$\begin{aligned} \max_{\{T_n\}} \quad & \sum_{n=1}^N \frac{T_n \lambda_{\max}(\mathbf{G}_n)}{T_n + \sigma_{v_n}^2 (1 + P_1 \lambda_{\max}(\mathbf{G}_n))} \\ \text{s.t.} \quad & \sum_{n=1}^N T_n \leq T_{\text{total}} \\ & T_n \geq 0 \end{aligned} \quad (28)$$

It is easy to verify that the optimization problem (28) is convex because its Hessian matrix, which is a diagonal matrix in this case, is positive semidefinite on the convex set defined by the linear constraints. Although (28) 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 D. Briefly speaking, for a threshold ϕ that is uniquely determined by a procedure described in Appendix D, we have

$$T_n = \begin{cases} \frac{1}{\beta_n} \left(\sqrt{\frac{\alpha_n}{\phi}} - 1 \right) & \alpha_n \geq \phi \\ 0 & \text{otherwise} \end{cases} \quad (29)$$

where

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

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

C. Summary and Numerical Results

For clarity, we now summarize the proposed optimal solution.

- 1) Given the prior knowledge of the noise statistics $\{\mathbf{R}_{w,n}\}$, $\{\sigma_{v_n}^2\}$, the signal $\boldsymbol{\theta}$, and the observation matrices $\{\mathbf{H}_n\}$, compute $\{\alpha_n\}$ and $\{\beta_n\}$.
- 2) Given the total power constraint T_{total} , find the optimal power allocation among sensors via (28). The solution of (28) is elaborated in Appendix D.
- 3) With the optimal power assignment, determine the optimal precoding matrices $\{\mathbf{C}_n\}$ via (17), whose solution is given by (25).

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. The observation matrix, observation noise covariance matrix, and the channel noise variance are set equal to $\mathbf{H} = \mathbf{I}$, $\mathbf{R}_w = 0.5\mathbf{I}$, and $\sigma_v^2 = 0.5$, respectively. 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

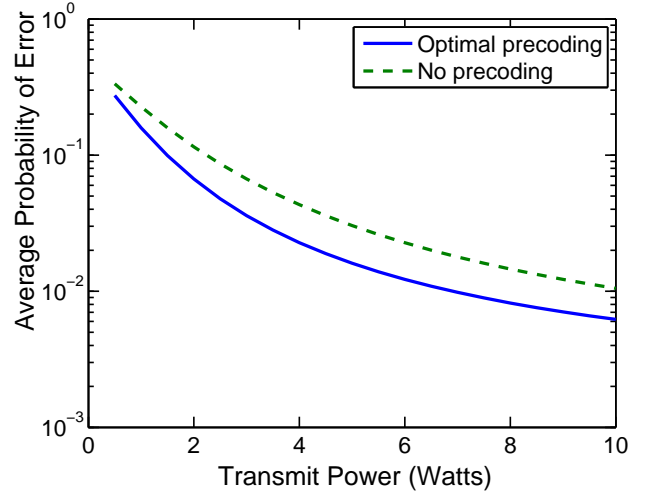


Fig. 2. Average probability of error vs. transmit power for optimal precoding and no precoding strategies.

be seen that the optimal compression strategy outperforms the non-compression strategy, which corroborates our theoretical analysis.

The detection performance under different power allocation schemes is also investigated. We set $N = 20$, and $\mathbf{H}_n = \mathbf{I}$, $\mathbf{R}_{w_n} = 0.5\mathbf{I}$ for all n . The channel signal-to-noise ratio (SNR) is assumed to be $|r_n|^2$, where $|r_n|$'s are independent and identically distributed (i.i.d.) Rayleigh-fading random variables with unit variance. Since the channel gain is normalized to unity in our problem formulation (c.f. (2)), we set $1/\sigma_{v_n}^2 = |r_n|^2$. 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 precoding vectors (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.

D. Extension to Neyman-Pearson Detection

The extension of our theoretical results to the Neyman-Pearson variant of the detection problem is straightforward. This is because the decision rule for the Neyman-Pearson detector is still an LRT, except that the threshold is determined by the prescribed 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 $P_1 = 0$) [15].

The Neyman-Pearson detection aims at maximizing the detection probability subject to a given false alarm probability. The decision rule is also a LRT which is given as

$$L(\mathbf{y}) = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} \underset{H_0}{\overset{H_1}{\geq}} \eta \quad (30)$$

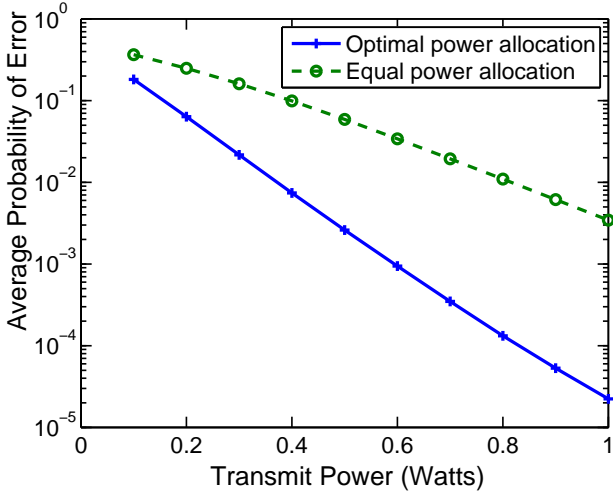


Fig. 3. Average probability of error vs. transmit power for optimal power allocation and equal power allocation schemes.

where η is the threshold determined by the specified false alarm probability. Following a similar derivation, we know 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 (25) and the optimal power allocation (28) 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, we analyze the impact of number of sensors on the overall detection performance, given the total amount of transmit power fixed. The channels between sensors and the FC are assumed i.i.d. channels. Note that as we indicated earlier, since the channel gain is normalized to unity in our problem formulation (see (2)), we, alternatively, set $\sigma_{v_n}^2$ a random variable to simulate i.i.d. channels.

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 $\mathbf{H}_n = \mathbf{I}$, and $\mathbf{R}_{w_n} = \sigma_w^2 \mathbf{I}, \forall n$. When optimal precoding vectors (conditional on the equal-power allocation) are used, according to (28), the deflection coefficient χ is given by

$$\begin{aligned} \chi &= \sum_{n=1}^N \frac{T_{\text{total}} \lambda_{\max}(\mathbf{G}_n)}{T_{\text{total}} + N \sigma_{v_n}^2 (1 + P_1 \lambda_{\max}(\mathbf{G}_n))} \\ &\stackrel{(a)}{=} \sum_{n=1}^N \frac{T_{\text{total}} \|\boldsymbol{\theta}\|_2^2}{\sigma_w^2 T_{\text{total}} + N \sigma_{v_n}^2 (\sigma_w^2 + P_1 \|\boldsymbol{\theta}\|_2^2)} \end{aligned} \quad (31)$$

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

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

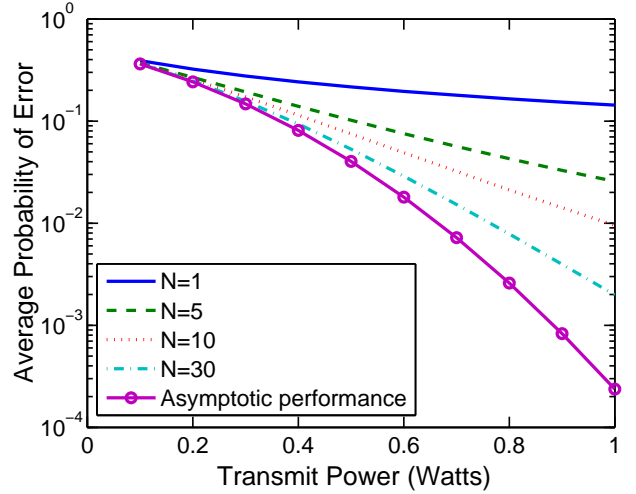


Fig. 4. Average probability of error vs. transmit power for different number of sensors.

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

$$\begin{aligned} \chi &= \sum_{n=1}^N \frac{\rho_1}{\rho_2 + N \sigma_{v_n}^2} \\ &= \sum_{n=1}^N \frac{1}{N} \left(\frac{\rho_1}{\sigma_{v_n}^2} - \frac{\rho_1 \rho_2}{N \sigma_{v_n}^4 + \rho_2 \sigma_{v_n}^2} \right) \\ &\stackrel{N \rightarrow \infty}{=} \rho_1 E[1/\sigma_{v_n}^2] \triangleq \chi_{\infty} \end{aligned} \quad (32)$$

where the last equality follows from the strong Law of Large Numbers (LLN). 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 $\mathbf{H}_n = \mathbf{I}$, $\mathbf{R}_{w_n} = 0.5\mathbf{I}$ for all n . We set $\sigma_{v_n}^2 = 1/|r_n|^2$ to simulate i.i.d. channels, where $|r_n|$'s are 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.

The detection diversity gain can be explored from a different perspective. Inspired by the notion of ‘‘estimation outage probability’’ proposed in [21], we introduce a 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.

$$P_{\text{outage}} \triangleq \Pr\{P_D < \tau_D | P_{\text{FA}}\} \quad (33)$$

Recall that the test statistic u is a Gaussian random variable with its mean and variance under null and alternative hypotheses given by (10). Therefore for a prescribed false alarm

probability, the detection probability P_D is given as

$$\begin{aligned} P_D &= \Pr(u > \eta | H_1) = Q\left(\frac{\eta - \Delta - \sigma_u^2}{\sigma_u}\right) \\ &= Q(Q^{-1}(P_{FA}) - \sigma_u) \end{aligned} \quad (34)$$

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

$$\begin{aligned} P_{\text{outage}} &= \Pr(Q(Q^{-1}(P_{FA}) - \sigma_u) < \tau_D) \\ &= \Pr(Q^{-1}(P_{FA}) - \sigma_u > Q^{-1}(\tau_D)) \\ &= \Pr(\sigma_u < Q^{-1}(P_{FA}) - Q^{-1}(\tau_D)) \\ &= \Pr(\chi < \zeta) \end{aligned} \quad (35)$$

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

From (32), 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 $\{\rho_1/\sigma_{v_n}^2\}$. According to the large deviation theory [22], for any $\zeta < \chi_\infty$, we have the outage probability decreasing exponentially with N as follows

$$P_{\text{outage}} \sim \exp(-NI_\varpi(\zeta/\rho_1)) \quad (36)$$

where \sim means asymptotic convergence as N becomes large, ϖ is the common distribution of $\varpi_n \triangleq 1/\sigma_{v_n}^2$, and $I_\varpi(x)$ is the rate function of ϖ :

$$I_\varpi(x) = \sup_{t \in \mathbb{R}} (tx - \log M_\varpi(t)) \quad (37)$$

with $M_\varpi(t)$ the moment-generating function of ϖ . From (36), we see that if the specified P_{FA} and τ_D satisfy the following condition:

$$(Q^{-1}(P_{FA}) - Q^{-1}(\tau_D))^2 < \chi_\infty \quad (38)$$

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, the condition (38) 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 $P_1 = 0$, $P_{FA} = 0.1$ and $\tau_D = 0.9$, 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 the condition (38) is satisfied as long as $T_{\text{total}} \geq 1$. From Fig. 5, we see that the outage probability decreases dramatically even 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 θ to be detected. A fundamental assumption made in previous sections is that the signal θ 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 precoding vector for each sensor.

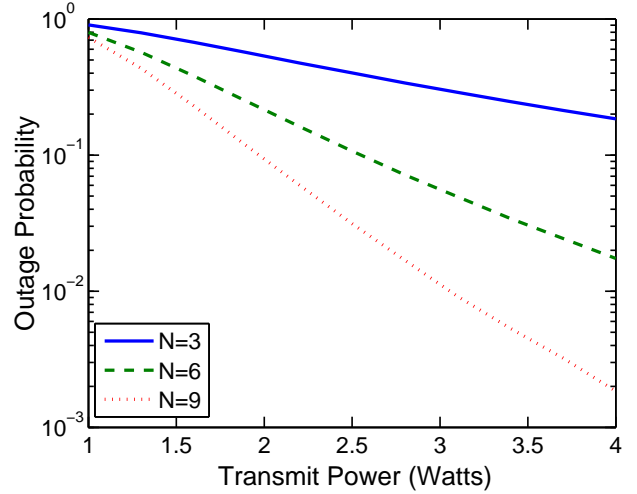


Fig. 5. Outage probability vs. total transmit power for different number of sensors.

A. GLRT Detector

Suppose that the precoding vectors $\{c_n\}$ are pre-determined, 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

$$L_G(\mathbf{y}) = \frac{p(\mathbf{y}|\hat{\theta}; H_1)}{p(\mathbf{y}|H_0)} > \eta \quad (39)$$

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

$$\begin{aligned} p(\mathbf{y}|\theta; H_1) &= \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \\ &\times \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{P}\theta)^T \Sigma^{-1}(\mathbf{y} - \mathbf{P}\theta)\right\} \end{aligned} \quad (40)$$

in which Σ is a diagonal matrix with its n th diagonal element given by $c_n \mathbf{R}_{w_n} c_n^T + \sigma_{v_n}^2$, and

$$\mathbf{P} \triangleq \begin{bmatrix} c_1 \mathbf{H}_1 \\ c_2 \mathbf{H}_2 \\ \vdots \\ c_N \mathbf{H}_N \end{bmatrix} \quad (41)$$

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

$$\hat{\theta} = (\mathbf{P}^T \Sigma^{-1} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1} \mathbf{y} \quad (42)$$

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 $\hat{\theta}$ back into (39), thus we have

$$\ln L_G(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \Sigma^{-1} \mathbf{P} (\mathbf{P}^T \Sigma^{-1} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1} \mathbf{y} \quad (43)$$

or we decide H_1 if

$$\mathbf{y}^T \Sigma^{-1} \mathbf{P} (\mathbf{P}^T \Sigma^{-1} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1} \mathbf{y} > \eta' \quad (44)$$

It is shown in [20, Section 6.5] that when $N \rightarrow \infty$, 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 plus or minus signs 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 $\text{sgn}(\mathbf{x})$ be a sign column vector with its elements given by $\text{sgn}(x_i)$, where $\text{sgn}(x_i) = 1$ if $x_i > 0$, and $\text{sgn}(x_i) = -1$ otherwise. We design the precoding vector for each sensor as follows

$$\mathbf{c}_n = \psi_n (|\mathbf{r}_n| \odot \text{sgn}(\mathbf{g}_n))^T \mathbf{R}_{w_n}^{-\frac{1}{2}} \quad \forall n \quad (45)$$

where \mathbf{r}_n is a column vector whose entries are randomly generated according to a Gaussian distribution with zero mean and unit variance, $|\mathbf{r}_n|$ is a vector whose entries are the absolute values 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 $P_1 = 0$). Recalling that the matrix \mathbf{P} defined in (41) has to be full column rank, generating the precoding vectors in a random manner guarantees that \mathbf{P} is full column rank with a high probability.

It is interesting to examine how well this heuristic precoding design performs. We consider a homogeneous scenario where sensors have identical observation matrices and observation noise covariance matrices, i.e. $\mathbf{H}_n = \mathbf{H}$, $\mathbf{R}_{w_n} = \mathbf{R}_w$ for all n . Also, we assume an equal power allocation throughout our following discussion. The deflection coefficient is then given by (c.f. (16))

$$\chi = \sum_{n=1}^N (\mathbf{c}_n \mathbf{R}_w \mathbf{c}_n^T + \sigma_{v_n}^2)^{-1} \mathbf{c}_n \mathbf{H} \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{c}_n^T \triangleq \sum_{n=1}^N \chi_n \quad (46)$$

in which

$$\chi_n \triangleq (\mathbf{c}_n \mathbf{R}_w \mathbf{c}_n^T + \sigma_{v_n}^2)^{-1} \mathbf{c}_n \mathbf{H} \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{c}_n^T \quad (47)$$

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

$$\mathbf{c}_n \mathbf{R}_w \mathbf{c}_n^T + P_1 \mathbf{c}_n \mathbf{H} \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{c}_n^T = T \quad (48)$$

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

$$\begin{aligned} v_n &\triangleq \mathbf{c}_n \mathbf{R}_w \mathbf{c}_n^T \\ \mu_n &\triangleq \frac{\mathbf{c}_n \mathbf{H} \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{c}_n^T}{\mathbf{c}_n \mathbf{R}_w \mathbf{c}_n^T} \end{aligned} \quad (49)$$

The individual deflection coefficient can be re-expressed in terms of v_n and μ_n

$$\chi_n = \frac{v_n \mu_n}{v_n + \sigma_{v_n}^2} \quad (50)$$

and the power constraint (48) can be rewritten as

$$v_n(1 + P_1 \mu_n) = T \quad (51)$$

Solving v_n from the power constraint (51), and substituting it back into (50), we arrive at

$$\chi_n = \frac{T \mu_n}{T + \sigma_{v_n}^2 (1 + P_1 \mu_n)} \quad (52)$$

Clearly, the individual deflection coefficient χ_n is a monotonically increasing function of μ_n . If $\boldsymbol{\theta}$ is known, the optimal precoding vector which maximizes μ_n can be determined, and it can be easily derived that the maximum μ_n is equal to

$$\mu_{\max} = \lambda_{\max}(\mathbf{R}_w^{-\frac{1}{2}} \mathbf{H} \boldsymbol{\theta} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{R}_w^{-\frac{1}{2}}) = \mathbf{g}^T \mathbf{g} = \sum_{i=1}^q g_i^2 \quad (53)$$

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 (45), μ_n is given by

$$\begin{aligned} \mu_n &= \frac{(|\mathbf{r}_n| \odot \text{sgn}(\mathbf{g}))^T \mathbf{g} \mathbf{g}^T (|\mathbf{r}_n| \odot \text{sgn}(\mathbf{g}))}{(|\mathbf{r}_n| \odot \text{sgn}(\mathbf{g}))^T (|\mathbf{r}_n| \odot \text{sgn}(\mathbf{g}))} \\ &= \frac{(\sum_{i=1}^q |r_{n_i} g_i|)^2}{\sum_{i=1}^q r_{n_i}^2} \end{aligned} \quad (54)$$

For notational convenience, let χ_{sub} and χ_{opt} respectively denote the overall deflection coefficients attained by the precoding design (45) and the optimal precoding design. The ratio of these two deflection coefficients is then given as

$$\begin{aligned} \frac{\chi_{\text{sub}}}{\chi_{\text{opt}}} &= \frac{\sum_n \frac{T \mu_n}{T + \sigma_{v_n}^2 (1 + P_1 \mu_n)}}{\sum_n \frac{T \mu_{\max}}{T + \sigma_{v_n}^2 (1 + P_1 \mu_{\max})}} > \frac{\sum_n \frac{T \mu_n}{T + \sigma_{v_n}^2 (1 + P_1 \mu_{\max})}}{\sum_n \frac{T \mu_{\max}}{T + \sigma_{v_n}^2 (1 + P_1 \mu_{\max})}} \\ &\approx \frac{1}{N} \sum_{n=1}^N \frac{\mu_n}{\mu_{\max}} \stackrel{N \rightarrow \infty}{=} E[\mu_n / \mu_{\max}] \end{aligned} \quad (55)$$

which converges to $E[\mu_n / \mu_{\max}]$ as the number of sensors increases. Utilizing (53)–(54), we have

$$\begin{aligned} E\left[\frac{\mu_n}{\mu_{\max}}\right] &= E\left[\frac{(\sum_{i=1}^q |r_{n_i} g_i|)^2}{(\sum_{i=1}^q r_{n_i}^2)(\sum_{i=1}^q g_i^2)}\right] \\ &\geq E\left[\frac{\sum_{i=1}^q r_{n_i}^2 g_i^2}{(\sum_{i=1}^q r_{n_i}^2)(\sum_{i=1}^q g_i^2)}\right] \\ &= \frac{1}{\sum_{i=1}^q g_i^2} \sum_{i=1}^q \left\{ g_i^2 E\left[\frac{r_{n_i}^2}{\sum_{i=1}^q r_{n_i}^2}\right] \right\} \\ &\approx \frac{1}{\sum_{i=1}^q g_i^2} \sum_{i=1}^q \left\{ g_i^2 E[r_{n_i}^2] E\left[\frac{1}{\sum_{i=1}^q r_{n_i}^2}\right] \right\} \\ &= \frac{\sum_{i=1}^q g_i^2}{(q-2)(\sum_{i=1}^q g_i^2)} > \frac{1}{q} \end{aligned} \quad (56)$$

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 (55)–(56), we conclude that the ratio of the deflection coefficient achieved by the precoding design (45) to that attained by the optimal precoding design is within $1/q$.

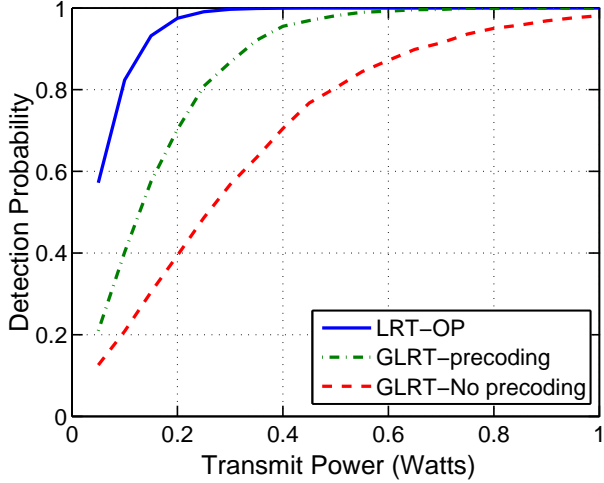


Fig. 6. Detection probability vs. total transmit power for GLRT with precoding (45) and no precoding, and NP test with optimal precoding (OP).

Simulations are conducted to illustrate the performance of the GLRT with precoding design (45) (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 θ and employs optimal precoding design (denoted as NP-OP). In our simulations, we set $P_1 = 0$, $\mathbf{H}_n = \mathbf{I}$, $\mathbf{R}_{w_n} = 0.5\mathbf{I}$ for all n , and $\theta = [\cos(1) \cos(2) \cos(3)]^T$. There are 100 sensors. The channels between sensors and the FC are generated in a same way as we did in previous examples. The false alarm probability is set to $P_{FA} = 0.05$. The detection probabilities of the GLRT and NP-OP are shown in Fig. 6. We see that GLRT with precoding (45) 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.

VII. CONCLUSIONS

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 theoretical analysis indicates that the optimal precoder is a compression vector which converts each sensor's original measurements into a single message, and the optimal precoder is exactly a matched filter in a vector form. Although matched filter detection is a well-studied topic, its optimality in a distributed power-constrained network has never been established before. The optimal power allocation among sensors was examined as well. It is found that the optimal power allocation is a water-filling type scheme.

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, a new 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.

APPENDIX A PROOF OF THEOREM 1

Let $a_i \triangleq d_{ii}^2/\sigma_{v_n}^2$, and $b_i \triangleq P_1 f_{ii}$. The optimization (22) can be rewritten as

$$\begin{aligned} \max_{\{a_i, b_i\}} \quad & \sum_{i=1}^{q_n} \frac{a_i b_i}{a_i + 1} \iff \min_{\{a_i, b_i\}} \sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} a_i(1 + b_i) = \frac{T_n}{\sigma_{v_n}^2} \triangleq \tilde{T}_n \\ & a_i \geq 0 \quad \forall i \\ & b_i \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} b_i = P_1 \lambda_{\max}(\mathbf{G}_n) \triangleq \lambda \end{aligned} \quad (57)$$

The above optimization involves optimizing two sets of variables $\{a_i\}$ and $\{b_i\}$. To solve (57), we first optimize one set of variables, given that the other set of variables are fixed. Suppose that $\{b_i\}$ are pre-determined, and are arranged in a descending order, i.e. $b_1 \geq b_2 \geq \dots \geq b_{q_n}$. Then optimizing $\{a_i\}$ conditional on fixed $\{b_i\}$ can be formulated as

$$\begin{aligned} \min_{\{a_i\}} \quad & \sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} \\ \text{s.t.} \quad & \sum_{i=1}^{q_n} a_i(1 + b_i) = \tilde{T}_n \\ & a_i \geq 0 \quad \forall i \end{aligned} \quad (58)$$

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

$$a_i = \left[\sqrt{\frac{b_i}{\phi(1 + b_i)}} - 1 \right]^+ \quad \forall i \quad (59)$$

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 $\mathbf{b} \triangleq [b_1 \ b_2 \ \dots \ b_{q_n}]$. Substituting $a_i^*(\mathbf{b})$ back into (57),

we come to an optimization involving only $\{b_i\}$:

$$\begin{aligned} \min_{\{b_i\}} \quad & \sum_{i=1}^{q_n} \frac{b_i}{a_i^*(\mathbf{b}) + 1} \\ \text{s.t.} \quad & b_i \geq 0 \quad \forall i \\ & \sum_{i=1}^{q_n} b_i = \lambda \end{aligned} \quad (60)$$

In the following, we show that the optimal solution to (60) is given by

$$b_i^* = \begin{cases} \lambda & i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (61)$$

Notice that the parameter ϕ in (59) needs to be determined through an iterative search. Therefore we cannot directly substitute the solution of $a_i^*(\mathbf{b})$ into (60). 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

$$\pi(\mathbf{b}) \triangleq \sum_{i=1}^2 \frac{b_i}{a_i^*(\mathbf{b}) + 1} \quad (62)$$

$$\mathbf{b}^{(0)} \triangleq [\lambda \ 0]$$

In Appendix C, we proved that $\mathbf{b}^{(0)}$ is the optimal solution to (60) for $q_n = 2$, that is,

$$\pi(\mathbf{b}^{(0)}) > \pi(\mathbf{b}) \quad (63)$$

for any $\mathbf{b} \neq \mathbf{b}^{(0)}$ satisfying the constraints defined in (60). Therefore for $q_n = 2$, the optimal solution to (57) is given by

$$\begin{aligned} \{b_1^*, b_2^*\} &= \{\lambda, 0\} \\ \{a_1^*, a_2^*\} &= \left\{ \frac{\tilde{T}_n}{1+\lambda}, 0 \right\} \end{aligned} \quad (64)$$

In other words, we have

$$\frac{\lambda}{\frac{\tilde{T}_n}{1+\lambda} + 1} \geq \sum_{i=1}^2 \frac{b_i}{a_i + 1} \quad (65)$$

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 (57) is lower bounded by

$$\begin{aligned} \sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} &= \sum_{i=1}^2 \frac{b_i}{a_i + 1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i + 1} \\ &\stackrel{(a)}{\geq} \frac{\check{b}_1}{\frac{\tilde{T}_{n,1}}{1+b_1} + 1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i + 1} \end{aligned} \quad (66)$$

in which $\check{b}_1 \triangleq b_1 + b_2, \check{\tilde{T}}_{n,1} \triangleq \tilde{T}_{n,1} + \tilde{T}_{n,2}$, and the inequality (a) comes by utilizing (65). The above objective function can

be further lower bounded as

$$\begin{aligned} \sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} &\geq \frac{\check{b}_1}{\frac{\tilde{T}_{n,1}}{1+b_1} + 1} + \sum_{i=3}^{q_n} \frac{b_i}{a_i + 1} \\ &= \frac{\check{b}_1}{\frac{\tilde{T}_{n,1}}{1+b_1} + 1} + \frac{b_3}{a_3 + 1} + \sum_{i=4}^{q_n} \frac{b_i}{a_i + 1} \\ &\geq \frac{\check{b}_2}{\frac{\tilde{T}_{n,2}}{1+b_2} + 1} + \sum_{i=4}^{q_n} \frac{b_i}{a_i + 1} \end{aligned} \quad (67)$$

in which $\check{b}_2 \triangleq b_1 + b_2 + b_3, \check{\tilde{T}}_{n,2} \triangleq \tilde{T}_{n,1} + \tilde{T}_{n,2} + \tilde{T}_{n,3}$, and the inequality, again, comes by using (65). So on and so forth, we can reach that the objective function is eventually lower bounded by

$$\sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} \geq \frac{\lambda}{\frac{\tilde{T}_n}{1+\lambda} + 1} \quad (68)$$

and this lower bound is attained only when

$$b_i^* = \begin{cases} \lambda & i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (69)$$

$$a_i^* = \begin{cases} \frac{\tilde{T}_n}{1+\lambda} & i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (70)$$

Therefore (69)–(70) are the optimal solution to (57). The proof is completed here.

APPENDIX B

AN ANALYTICAL SOLUTION TO (58)

The Lagrangian function associated with (58) is given by

$$\begin{aligned} L(a_i; \phi; \nu_i) &= \sum_{i=1}^{q_n} \frac{b_i}{a_i + 1} - \phi \left(\tilde{T}_n - \sum_{i=1}^{q_n} a_i(1 + b_i) \right) - \sum_{i=1}^{q_n} \nu_i a_i \end{aligned} \quad (71)$$

which gives the following KKT conditions [23]:

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

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

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

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

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

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

$$a_i = \sqrt{\frac{b_i}{\phi(1 + b_i)} - \nu_i} - 1 \quad \forall i \quad (72)$$

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 (72) becomes

$$a_i = \left[\sqrt{\frac{b_i}{\phi(1 + b_i)} - 1} \right]^+ \quad \forall i \quad (73)$$

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

$$\phi = \left[\frac{\sum_{i=1}^K \sqrt{b_i(1+b_i)}}{\tilde{T}_n + \sum_{i=1}^K (1+b_i)} \right]^2 \quad (74)$$

Now substituting ϕ back to (73), we get a new solution $\{a'_1, a'_2, \dots, a'_K, a'_{K+1}, \dots, a'_{q_n}\}$. 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 (63)

Note that for the two-dimensional case, the feasible region $\{\mathbf{b} = [b_1 \ b_2]\}$ of the optimization problem (60) 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

$$\mathcal{R}_1 \triangleq \{\mathbf{b} = [\lambda - \delta \ \delta] \mid \delta \in (0, \min(\lambda/2, \tau))\} \quad (75)$$

where $\tau > 0$ is a threshold such that if $\delta < \tau$, then the optimal solution $\{a_i\}$ to (58) conditional on $\mathbf{b} \in \mathcal{R}_1$ has the following form:

$$\mathbf{a}^*(\mathbf{b}) = [a_1^*(\mathbf{b}) \ 0] \quad (76)$$

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

$$\mathcal{R}_2 \triangleq \{\mathbf{b} = [\lambda - \delta \ \delta] \mid \delta \in [\tau, \lambda/2]\} \quad (77)$$

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 (60), 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

$$\begin{aligned} \{a_i^*(\mathbf{b}^{(0)})\} &= \left\{ \frac{\tilde{T}_n}{1+\lambda}, 0 \right\} \\ \{a_i^*(\mathbf{b})\} &= \left\{ \frac{\tilde{T}_n}{1+\lambda-\delta}, 0 \right\} \end{aligned} \quad (78)$$

Substituting the optimal solution $\{a_i^*(\mathbf{b})\}$ into (62), we have

$$\begin{aligned} \pi(\mathbf{b}^{(0)}) &= \frac{\lambda(1+\lambda)}{\tilde{T}_n + 1 + \lambda} \\ \pi(\mathbf{b}) &= \frac{(\lambda - \delta)(1 + \lambda - \delta)}{\tilde{T}_n + 1 + \lambda - \delta} + \delta \end{aligned} \quad (79)$$

and

$$\pi(\mathbf{b}) - \pi(\mathbf{b}^{(0)}) = \frac{\tilde{T}_n^2 \delta + \tilde{T}_n \delta}{(\tilde{T}_n + 1 + \lambda - \delta)(\tilde{T}_n + 1 + \lambda)} \quad (80)$$

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

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

holds. Also, from (80), 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

$$\mathcal{R}_2 \triangleq \{\mathbf{b} = [\lambda/2 + \delta' \ \lambda/2 - \delta'] \mid \delta' \in [0, \lambda/2 - \tau]\} \quad (81)$$

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 (58) conditional on \mathbf{b} has the following form:

$$\mathbf{a}^*(\mathbf{b}) = [a_1^*(\mathbf{b}) \ a_2^*(\mathbf{b})] \quad (82)$$

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

$$\begin{aligned} \pi(\mathbf{b}) &= \sum_{i=1}^2 \frac{b_i}{a_i^*(\mathbf{b}) + 1} \stackrel{(a)}{=} \sqrt{\phi} \sum_{i=1}^2 \sqrt{b_i(1+b_i)} \\ &\stackrel{(b)}{=} \frac{\left(\sum_{i=1}^2 \sqrt{b_i(1+b_i)} \right)^2}{\tilde{T}_n + \sum_{i=1}^2 (1+b_i)} \\ &= \frac{\lambda + \lambda^2}{\tilde{T}_n + 2 + \lambda} + \frac{2}{\tilde{T}_n + 2 + \lambda} \kappa(\delta') \end{aligned} \quad (83)$$

where (a) comes by utilizing (59), (b) follows from (74), and

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

$$\begin{aligned} \alpha &\triangleq 1 + \lambda + \frac{\lambda^2}{2} \\ \beta &\triangleq \frac{\lambda^4}{16} + \frac{\lambda^3}{4} + \frac{\lambda^2}{4} \end{aligned}$$

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

$$\tilde{\kappa}(t) \triangleq \sqrt{t^2 - \alpha t + \beta} - \frac{\lambda^2}{4} + t \quad (84)$$

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

$$\frac{\partial \tilde{\kappa}(t)}{\partial t} = \frac{2t - \alpha}{\sqrt{t^2 - \alpha t + \beta}} + 1 \quad (85)$$

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

$$\begin{aligned} \alpha - 2t &> \sqrt{t^2 - \alpha t + \beta} \\ \Rightarrow \frac{\partial \tilde{\kappa}(t)}{\partial t} &< 0 \end{aligned} \quad (86)$$

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

$$\pi(\mathbf{b}^{(0)}) < \pi(\mathbf{b}^{(1)}) < \pi(\mathbf{b}^{(2)}) \quad (87)$$

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

APPENDIX D

AN ANALYTICAL SOLUTION TO (28)

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

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

The Lagrangian function L associated with (28) is given by

$$\begin{aligned} L(T_n; \phi; \nu_n) \\ = - \sum_{n=1}^N \frac{\alpha_n T_n}{\beta_n T_n + 1} - \phi \left(T_{\text{total}} - \sum_{n=1}^N T_n \right) - \sum_{n=1}^N \nu_n T_n \end{aligned} \quad (88)$$

which gives the following KKT conditions [23]:

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

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

$$T_n = \frac{1}{\beta_n} \left[\sqrt{\frac{\alpha_n}{\phi - \nu_n}} - 1 \right] \quad \forall n \quad (89)$$

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 (89) becomes

$$T_n = \frac{1}{\beta_n} \left[\sqrt{\frac{\alpha_n}{\phi}} - 1 \right]^+ \quad \forall n \quad (90)$$

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 $K \in \{1, \dots, N\}$ active nodes, according to (90), these K nodes must be $\{k_1, k_2, \dots, k_K\}$, where $\{k_i\}$ is a set of indices such that $\alpha_{k_1} \geq \alpha_{k_2} \geq \dots \geq \alpha_{k_K}$. Therefore ϕ can be solved by substituting $\{T_{k_1}, T_{k_2}, \dots, T_{k_K}\}$ into the second KKT condition, where T_{k_i} is given by

$$T_{k_i} = \frac{1}{\beta_{k_i}} \left[\sqrt{\frac{\alpha_{k_i}}{\phi}} - 1 \right] \quad (91)$$

Now we substitute ϕ back to (90). We will get a new solution $\{T'_{k_1}, T'_{k_2}, \dots, T'_{k_K}, T'_{k_{K+1}}, \dots, T'_{k_N}\}$. If this new solution is exactly identical to the one we assumed before, i.e. $\{T_{k_1}, T_{k_2}, \dots, T_{k_K}, 0, \dots, 0\}$. 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 [24].

REFERENCES

- [1] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, pp. 501–510, July 1981.
- [2] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-23, no. 1, pp. 24–30, Jan. 1987.
- [3] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, pp. 995–1000, Sept. 1989.
- [4] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulas, "Optimal distributed decision fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, pp. 761–765, Sept. 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 Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 554–568, April 1996.
- [7] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I - fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, January 1997.
- [8] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II - advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, January 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 Transactions on Signal Processing*, 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 Transactions on Information Theory*, vol. 47, no. 4, pp. 1368–1377, May 2001.
- [11] J. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 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 Transactions on Information 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 Transactions on Signal Processing*, vol. 53, no. 8, pp. 2617–2624, August 2005.
- [14] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.
- [15] —, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

- [16] P. Bianchi and J. Jakubowicz, "Linear precoders for the detection of a Gaussian process in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 882–894, Mar. 2011.
- [17] 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 Journal on Selected Areas in Communications*, vol. 28, no. 7, pp. 1149–1158, Sept. 2010.
- [18] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4284–4299, August 2007.
- [19] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3833–3832, July 2009.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Upper Saddle River, NJ: Prentice Hall, 1998.
- [21] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, September 2007.
- [22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones & Bartlett, 1993.
- [23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2003.
- [24] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, February 2006.