Optimal Power Allocation for Distributed Detection Over MIMO Channels in Wireless Sensor Networks

Xin Zhang, *Member, IEEE*, H. Vincent Poor, *Fellow, IEEE*, and Mung Chiang, *Member, IEEE*

Abstract—In distributed detection systems with wireless sensor networks, the communication between sensors and a fusion center is not perfect due to interference and limited transmitter power at the sensors to combat noise at the fusion center's receiver. The problem of optimizing detection performance with such imperfect communication brings a new challenge to distributed detection. In this paper, sensors are assumed to have independent but nonidentically distributed observations, and a multiple-input/multiple-output (MIMO) channel model is included to account for imperfect communication between the sensors and the fusion center. The J-divergence between the distributions of the detection statistic under different hypotheses is used as a performance criterion in order to provide a tractable analysis. Optimization of performance with individual and total transmitter power constraints on the sensors is studied, and the corresponding power allocation scheme strikes a tradeoff between two factors, the communication channel quality and the local decision quality. For the case with orthogonal channels, under certain conditions, the power allocation can be solved by a weighted water-filling algorithm. Simulations show that the proposed power allocation in certain cases only consumes as little as 25% of the total power used by an equal power allocation scheme while achieving the same performance.

Index Terms—Distributed detection, multiple-input multiple-output (MIMO) channel, power allocation, wireless sensor networks (WSNs).

I. INTRODUCTION

WIRELESS sensor networks (WSNs) have received considerable attention recently. Event monitoring is a typical application of wireless sensor networks. In event monitoring, a number of sensors are deployed over a region where some phenomenon is to be monitored. Each sensor collects and possibly processes data about the phenomenon and transmits its observation or a summary of its observation to a fusion center

(FC). The FC makes a global decision about the state of the phenomenon based on the received data from the sensors, and possibly triggers an appropriate action.

The essential part of event monitoring is a detection problem, i.e., the FC needs to detect the state of the phenomenon under observation. In wireless sensor networks, due to power and communication constraints, sensors are often required to process their observations and transmit only summaries of their own findings to an FC. In this case, the detection problem associated with event monitoring becomes distributed detection (also called decentralized detection).

Distributed detection is obviously suboptimal relative to its centralized counterpart. However, energy, communication bandwidth, and reliability may favor the use of distributed detection systems. Distributed detection has been studied for several decades. Particularly, the design of optimal and suboptimal local decision and fusion rules has been extensively investigated. Tsitsiklis [31], Varshney [32], Viswanathan and Varshney [33], and Blum et al. [3] provide excellent reviews of the early work as well as extensive references.

However, most of these studies assume that a finite valued summary of a sensor is perfectly transmitted to an FC, i.e., no error occurs during the transmission. In distributed detection systems based on wireless sensor networks, this assumption may fail due to interference and limited transmitter power at sensors to combat receiver noise at the FC. The problem of optimizing detection performance with imperfect communications between the sensors and the FC over wireless channels brings a new challenge to distributed detection.

Rago et al. [28] consider a "censoring" or "send/no-send" idea. The sensors may choose to transmit data or keep silent according to a total communication rate constraint and values of their local likelihood ratios. Predd, Kulkarni, and Poor [27] examine a related protocol for the problem of distributed learning. Duman and Salehi [13] introduce a multiple access channel model to account for noise and interference in data transmission, and optimal quantization points (in the person-by-person sense) were obtained on the original observations through a numerical procedure. Chen and Willett [6] assume a general orthogonal channel model from the local sensors to the FC and investigate the optimality of the likelihood ratio test (LRT) for local sensor decisions. Chen et al. [5] formulate the parallel fusion problem with a fading channel with instantaneous channel state information (CSI) and derive the optimal likelihood ratio (LR)-based fusion rule with binary local decisions. Niu et al. [24] extend the results of [5] to the case without instantaneous CSI. Note that both [5] and [24] assume orthogonal channels between the sensors and the FC.

Chamberland and Veeravalli [7]–[9] provide asymptotic results for distributed detection in power (or equivalently, capacity) constrained wireless sensor networks. More specifically, [7] shows that, when the sensors have i.i.d. Gaussian or exponential observations and the sensors and the FC are connected with a multiple access channel with capacity R, having identical binary local decision rules at the sensors is optimal in the asymptotic regime where the observation interval goes to infinity. [8] considers a similar problem but with a total power constraint instead of a channel capacity constraint, and shows that using identical local decision rules at the sensors is optimal for i.i.d. observations. [9] considers the detection of 1-D spatial Gaussian stochastic processes. An amplify-and-relay communication strategy with power constraint is used and the channels are orthogonal with equal received signal power from each sensor. They assume sensors are scattered along 1-D space and have correlated observations of the Gaussian stochastic processes. The tradeoff between sensor density and the quality of information provided by each sensor is studied using an asymptotic analysis.

Liu and Sayeed [20] and Mergen *et al.* [21] propose the use of type based multiple access (TBMA) to transmit local information from the sensors to the FC, and present a performance analysis of detection at the FC. The results of [20] and [21] focus mainly on the case with i.i.d. observations at the sensors. Jayaweera [15] studies the fusion performance of distributed stochastic Gaussian signal detection with i.i.d. sensor observations, assuming an amplify-and-relay scheme.

Chamberland and Veeravalli [10] provide a survey of much of the recent progress in distributed detection in wireless sensor networks with resource constraints.

In this paper, we propose a distributed detection system infrastructure with a virtual multiple-input multiple-output (MIMO) channel to account for nonideal communications between a finite number of sensors and an FC. Our analysis does not consider an infinite number of sensors because in many practical cases, only a few tens of sensors are used. We assume the sensors have *independent but nonidentically distributed* observations, so they have different local decision qualities. Each sensor has an individual transmitting power constraint, and there is also a joint power constraint on the total amount of power that the sensors can expend to transmit their local decisions to the FC. The goal is to optimally distribute the joint power budget among the sensors so that the detection performance at the FC is optimized.

The J-divergence between the distributions of the detection statistic under different hypotheses is used as a performance index instead of the probability of error in order to provide a more tractable analysis. A power allocation scheme is developed with respect to the J-divergence criterion, and in-depth analysis of the special case of orthogonal channels is provided. The proposed power allocation is shown to be a tradeoff between two factors, the quality of the communication channel and the quality of the local decisions of the sensors. As will be shown in the simulations, to achieve the same performance in certain cases, the power allocation developed in this paper consumes as little as 25% of the total power used by an equal power allocation scheme.

This paper differentiates from previous work in the following aspects.

- A system with the sensors and the FC connected by a virtual MIMO channel is considered.
- The sensors have independent but nonidentically distributed observations,[1] and hence they have different local decision qualities.
- We develop the power allocation scheme for a finite number of sensors rather than asymptotically.
- To improve global detection performance within a power budget, we focus on how to efficiently and effectively transmit the local sensor decisions to the FC rather than how to design local and global decision rules.
- The proposed power allocation quantifies the tradeoff between communication channel quality and local decision quality.

The rest of the paper is organized as follows. In Section II, we introduce a distributed detection system infrastructure with a MIMO channel model. In Section III, we develop the optimal power allocation scheme with respect to the J-divergence performance index. In Section IV, we study a special case in which the sensors transmit data to the FC over orthogonal channels. In Section V, we provide numerical examples to illustrate the proposed power allocation. We conclude the paper in Section VI.

II. MODELS

Let us consider a hypothesis testing problem with two hypotheses H_0 and H_1, as shown in Fig. 1. There are K wireless sensors with observations $\mathbf{x} = [x_1, \ldots, x_K]^T$. The observations are independent of each other but are not necessarily identically distributed. The conditional probability density functions of these observations (conditioned on the underlying hypotheses) are given by $p(\mathbf{x} \mid H_i)$ for $i \in \{0, 1\}$. The sensors then make local decisions $\mathbf{u} = [u_1, \ldots, u_K]^T$ according to their local decision rules:

$$u_k = \gamma_k(x_k) = \begin{cases} 0 & \text{decide } H_0 \\ 1 & \text{decide } H_1 \end{cases} \tag{1}$$

where $k = 1, \ldots, K$. In this paper, we assume the local sensors do not communicate with each other, i.e., sensor k makes a decision independently based only on its own observation x_k. The local decision rules $\gamma_k(\cdot)$ do not have to be identical, and the false alarm probability and detection probability of sensor k are given by

$$P_F(k) = p(u_k = 1 \mid H_0) \tag{2}$$

and

$$P_D(k) = p(u_k = 1 \mid H_1). \tag{3}$$

We assume the sensors have knowledge of their observation quality in terms of P_D and P_F, which can be obtained by various standard methods from detection theory [25]. The joint conditional density functions of the local decisions are

$$p(\mathbf{u}|H_0) = \prod_{k=1}^{K} P_F(k)^{u_k} (1 - P_F(k))^{(1-u_k)} \tag{4}$$

---

[1]TBMA [20], [21] and amplify-and-relay [15] schemes usually assume that the sensors have i.i.d. observations.
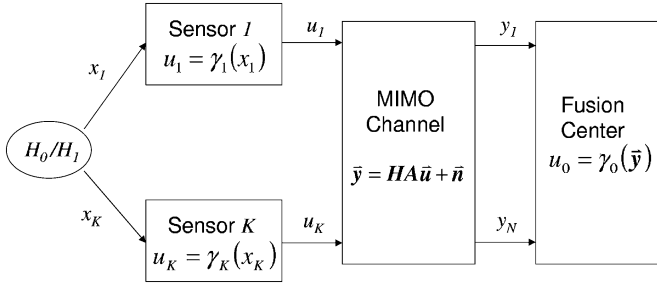
Fig. 1. Distributed detection system diagram.

and

$$p(\mathbf{u}|H_1) = \prod_{k=1}^{K} P_D(k)^{u_k}(1 - P_D(k))^{(1-u_k)}. \qquad (5)$$

The local decisions are transmitted to an FC through a MIMO channel, modelled by the following sampled baseband signal model (see, e.g., [35]):

$$\mathbf{y} = \mathbf{HAu} + \mathbf{n} \qquad (6)$$

where $\mathbf{y} = [y_1, \ldots, y_N]^T$ contains the received signals at the FC. $\mathbf{A} = \mathrm{diag}\{a_1, a_2, \ldots, a_K\}$ is a diagonal matrix, the diagonal elements of which are the amplitudes of the signals transmitted from the sensors. $\mathbf{H}$ is the channel matrix, which is assumed to be deterministic in this paper.[2] $\mathbf{n}$ is an additive noise vector which is assumed to be Gaussian with zero mean and covariance matrix $\mathbf{R}$. We assume that the channel quality (in terms of the $\mathbf{H}$ and $\mathbf{R}$ matrices) is known at the FC. This information can be obtained by channel estimation techniques. The dimension of $\mathbf{y}$, determined by the receiver design, is N, which does not have to be the same as the number of sensors K. Many different wireless channels and multiple access schemes can be expressed with the MIMO model in (6), including CDMA, TDMA, FDMA, as well as TBMA [20], [21], [34], [30].

The conditional density function of the received signals $\mathbf{y}$ at the FC given the transmitted signals $\mathbf{u}$ from the sensors is

$$p(\mathbf{y}|\mathbf{u}) = \frac{1}{|2\pi\mathbf{R}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{HAu})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{HAu})\right]. \qquad (7)$$

The conditional density functions of the received signals given the two hypotheses are

$$p(\mathbf{y} \,|\, H_i) = \sum_{\mathbf{u}} p(\mathbf{y} \,|\, \mathbf{u}) p(\mathbf{u} \,|\, H_i) \qquad (8)$$

where the summation is over all possible values of $\mathbf{u}$. The FC applies its fusion rule $\gamma_0(\,\cdot\,)$ to $\mathbf{y}$ to get a global decision

$$u_0 = \gamma_0(\mathbf{y}). \qquad (9)$$

The system is summarized in Fig. 1. We notice the Markov property of the system: $\{H_i\} \to \mathbf{x} \to \mathbf{u} \to \mathbf{y} \to u_0$ forms a Markov chain, which is used to derive (8) and will be used in the next section.

[2]We focus on the case in which the sensors and the FC have minimal movement and the environment changes slowly. In this case, the coherence time [30] of the wireless channel can be much longer than the time interval between two consecutive decisions made by the FC, and instantaneous CSI can be obtained.

In this paper, we do not focus on the design of local and global decision rules to optimize the detection performance at the FC. Instead, we focus on how to intelligently distribute a total transmitter power budget P_{tot} among the sensors, by choosing an amplitude matrix $\mathbf{A}$ within the constraint $\mathrm{Tr}[\mathbf{A}\mathbf{A}^T] \leq P_{\mathrm{tot}}$. There are also individual power constraint for each sensor, $\mathbf{A} \preceq \sqrt{\mathbf{P}_{\max}}$, to account for the maximum output power at each sensor. Here, $\sqrt{\mathbf{P}_{\max}}$ denotes the component-wise square root of $\mathbf{P}_{\max} = \mathrm{diag}\{P_{\max}(1), \ldots, P_{\max}(K)\}$, where $P_{\max}(k)$ is the transmitting power limit of sensor k. The matrix inequality $\preceq$ means $\sqrt{\mathbf{P}_{\max}} - \mathbf{A}$ is positive semidefinite.

III. OPTIMAL POWER ALLOCATION

In this section, an optimal power allocation among the sensors in the distributed detection system described in Section II is studied. We first choose a detection performance metric for our analysis.

There are three categories of commonly used detection performance metrics [25]: exact closed-form expressions of the miss probability P_M (which equals $1 - P_D$) and false alarm probability P_F (or the average error probability P_e, if prior probabilities of the hypotheses are known), distance related bounds, and asymptotic relative efficiency (ARE).

The closed-form expressions of P_M and P_F (or P_e) are hard to obtain even for centralized detection. ARE is useful for detection systems under large-sample-size (long observation duration) and weak signal conditions. Distance related bounds are upper or lower bounds on P_M and P_F (or P_e), such as the Chernoff bound, the Bhattacharyya bound, and the J-divergence [25].

In this paper, we use distance related bounds, more specifically the J-divergence, as the performance metric. The J-divergence, first proposed by Jeffreys [16], is a widely used metric for detection performance [17]–[19], [26]. It provides a lower bound on the detection error probability P_e [18] via the inequality

$$P_e > p(H_0)p(H_1)e^{-J/2}. \qquad (10)$$

We choose the J-divergence as the performance metric because it provides more tractable results in our study, it is closely related to results in information theory, such as the data processing lemma [12], and it is also closely related to other types of performance metrics. [19] shows that the ratio of the J-divergences of two test statistics is equivalent to the ARE under some circumstances. The J-divergence and the Bhattacharyya bound both belong to a more general class of distance measures, the Ali–Silvey class of distance measures [1]. The J-divergence is the symmetric version of the Kullback–Leibler (KL) distance [11], [12], and the KL distance is asymptotically the error exponent of the Chernoff bound from Stein's lemma [11].

The J-divergence between two densities, p_1 and p_0, is defined as

$$J(p_1, p_0) = D(p_1 \| p_0) + D(p_0 \| p_1) \qquad (11)$$

where $D(p_1 \| p_0)$ is the (nonsymmetric) KL distance between p_1 and p_0. $D(p_1 \| p_0)$ and $D(p_0 \| p_1)$ are defined as

$$D(p_i \| p_j) = \int \log\left(\frac{p_i}{p_j}\right) p_i. \qquad (12)$$

There is a well-known data processing lemma on the KL distance along a Markov chain [12, Lemma 3.11].

Lemma 1: The KL distance is nonincreasing along the Markov chain $\mathbf{x} \to \mathbf{u} \to \mathbf{y}$, i.e.,

$$D(p(\mathbf{x}\,|\,H_1)\|p(\mathbf{x}\,|\,H_0)) \geq D(p(\mathbf{u}\,|\,H_1)\|p(\mathbf{u}\,|\,H_0)) \quad (13)$$

and

$$D(p(\mathbf{u}\,|\,H_1)\|p(\mathbf{u}\,|\,H_0)) \geq D(p(\mathbf{y}|H_1)\|p(\mathbf{y}\,|\,H_0)). \quad (14)$$

This result can be easily generalized to the J-divergence with the following corollary.

Corollary 1: The J-divergence is nonincreasing along the Markov chain $\mathbf{x} \to \mathbf{u} \to \mathbf{y}$, i.e.,

$$J(p(\mathbf{x}\,|\,H_1), p(\mathbf{x}\,|\,H_0)) \geq J(p(\mathbf{u}\,|\,H_1), p(\mathbf{u}\,|\,H_0)) \quad (15)$$

and

$$J(p(\mathbf{u}\,|\,H_1), p(\mathbf{u}\,|\,H_0)) \geq J(p(\mathbf{y}\,|\,H_1), p(\mathbf{y}\,|\,H_0)). \quad (16)$$

Corollary 1 tells us that a performance upper bound of the detection at the FC is provided by $J(p(\mathbf{u}\,|\,H_1), p(\mathbf{u}\,|\,H_0))$. This can be achieved only when there are perfect data transmissions from the sensors to the FC, i.e., the FC receives $\mathbf{u}$ with no error.

Recall that our goal is to optimize the detection performance at the FC. This now translates into maximization of the J-divergence between the two densities of the received signals $\mathbf{y}$, with respect to the underlying hypotheses. The optimal power allocation is thus the solution to the following optimization problem:

$$\max_{\mathbf{A}} J(p(\mathbf{y}\,|\,H_1), p(\mathbf{y}\,|\,H_0))$$

$$\text{s.t. } \mathrm{Tr}\left[\mathbf{A}\mathbf{A}^T\right] \leq P_{\text{tot}}$$

$$\mathbf{0} \preceq \mathbf{A} \preceq \sqrt{\mathbf{P}_{\max}} \quad (17)$$

where the J-divergence $J(p(\mathbf{y}\,|\,H_1), p(\mathbf{y}\,|\,H_0))$ is given by

$$J(p(\mathbf{y}\,|\,H_1), p(\mathbf{y}\,|\,H_0))$$
$$= \int_{\mathbf{y}} d\mathbf{y}[p(\mathbf{y}\,|\,H_1) - p(\mathbf{y}\,|\,H_0)]\log\frac{p(\mathbf{y}\,|\,H_1)}{p(\mathbf{y}\,|\,H_0)}. \quad (18)$$

The density functions $p(\mathbf{y}\,|\,H_i)$, $i \in \{0,1\}$, are given by (7)–(8).

It can be seen that the conditional density functions $p(\mathbf{y}\,|\,H_i)$ are Gaussian mixtures. Unfortunately, the J-divergence between two Gaussian mixture densities does not have a general closed-form expression [22], [29]. In order to present the objective function in (17) in closed form, approximations must be made. An upper bound has been suggested in [29] based on the log-sum inequality [11]. However, this upper bound is not suitable for the study here, since the dependence on the power of transmitted signals is lost in the bound.

In this paper, the J-divergence of two Gaussian mixture densities $p(\mathbf{y}\,|\,H_i)$ is approximated by the J-divergence of two Gaussian densities $p_g(\mathbf{y}\,|\,H_i) = \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. The parameters of the Gaussian densities are provided by moment matching, i.e.,

$$\boldsymbol{\mu}_i = \int_{\mathbf{y}} \mathbf{y}p(\mathbf{y}\,|\,H_i)d\mathbf{y} \quad (19)$$

and

$$\boldsymbol{\Sigma}_i = \int_{\mathbf{y}} [\mathbf{y} - \boldsymbol{\mu}_i][\mathbf{y} - \boldsymbol{\mu}_i]^T p(\mathbf{y}\,|\,H_i)d\mathbf{y} \quad (20)$$

for $i \in \{0,1\}$. That is, a Gaussian mixture density is approximated by a Gaussian density with the same mean and variance as the Gaussian mixture density.

Obviously the quality of this approximation will directly affect the analysis in this paper and the difference between the optimal scheme and the proposed scheme, which is optimal for the approximated cases. It can be seen from (7) and (8) that when $\mathrm{Tr}[(\mathbf{H}\mathbf{A})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{A})] \to 0$, the Gaussian mixture density in (8) approaches a Gaussian distribution. So, we can predict the approximation will work well for the low SNR cases, and simulations in Section V show that it still works well even with received SNR as high as 10–12 dB.

We next calculate the means and covariance matrices of the Gaussian densities $p_g(\mathbf{y}\,|\,H_i)$, $i \in \{0,1\}$. From (8), (19), and the Markov property of the system

$$\boldsymbol{\mu}_i = \int_{\mathbf{y}} \mathbf{y}p(\mathbf{y}\,|\,H_i)d\mathbf{y}$$
$$= \int_{\mathbf{y}} \mathbf{y} \sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u})p(\mathbf{u}|H_i)d\mathbf{y}$$
$$= \sum_{\mathbf{u}} p(\mathbf{u}|H_i) \int_{\mathbf{y}} \mathbf{y}p(\mathbf{y}|\mathbf{u})d\mathbf{y}. \quad (21)$$

Recall that $p(\mathbf{y}\,|\,\mathbf{u})$ is a Gaussian density with mean $\mathbf{H}\mathbf{A}\mathbf{u}$, as shown in (7), so

$$\boldsymbol{\mu}_i = \sum_{\mathbf{u}} \mathbf{H}\mathbf{A}\mathbf{u}p(\mathbf{u}|H_i). \quad (22)$$

By applying (4) and (5), we have

$$\boldsymbol{\mu}_i = \mathbf{H}\mathbf{A}\boldsymbol{\beta}_i \quad (23)$$

for $i \in \{0,1\}$,

$$\boldsymbol{\beta}_1 = \sum_{\mathbf{u}} \mathbf{u}p(\mathbf{u}\,|\,H_1) = [P_D(1), \ldots, P_D(K)]^T \quad (24)$$

and

$$\boldsymbol{\beta}_0 = \sum_{\mathbf{u}} \mathbf{u}p(\mathbf{u}\,|\,H_0) = [P_F(1), \ldots, P_F(K)]^T. \quad (25)$$

Similarly, from (20) and the Markov property of the system, we have

$$\boldsymbol{\Sigma}_i = \int_{\mathbf{y}} [\mathbf{y} - \boldsymbol{\mu}_i][\mathbf{y} - \boldsymbol{\mu}_i]^T p(\mathbf{y}\,|\,H_i)d\mathbf{y}$$
$$= \int_{\mathbf{y}} [\mathbf{y} - \boldsymbol{\mu}_i][\mathbf{y} - \boldsymbol{\mu}_i]^T \sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u})p(\mathbf{u}|H_i)d\mathbf{y}$$
$$= \sum_{\mathbf{u}} p(\mathbf{u}|H_i) \int_{\mathbf{y}} [\mathbf{y} - \mathbf{H}\mathbf{A}\mathbf{u} + \mathbf{H}\mathbf{A}\mathbf{u} - \boldsymbol{\mu}_i]$$
$$\cdot [\mathbf{y} - \mathbf{H}\mathbf{A}\mathbf{u} + \mathbf{H}\mathbf{A}\mathbf{u} - \boldsymbol{\mu}_i]^T p(\mathbf{y}|\mathbf{u})d\mathbf{y}$$
$$= \mathbf{R} + \sum_{\mathbf{u}} p(\mathbf{u}|H_i)[\mathbf{H}\mathbf{A}\mathbf{u} - \boldsymbol{\mu}_i][\mathbf{H}\mathbf{A}\mathbf{u} - \boldsymbol{\mu}_i]^T. \quad (26)$$

The last step follows because $p(\mathbf{y} \mid \mathbf{u})$ is a Gaussian density with mean $\mathbf{HAu}$ and covariance matrix $\mathbf{R}$. Applying (4) and (5), we obtain

$$\Sigma_i = \mathbf{R} + \mathbf{HAB}_i\mathbf{A}^T\mathbf{H}^T \tag{27}$$

where

$$\mathbf{B}_1 = \mathrm{diag}\{P_D(1)[1-P_D(1)],\ldots,P_D(K)[1-P_D(K)]\}, \tag{28}$$

and

$$\mathbf{B}_0 = \mathrm{diag}\{P_F(1)[1-P_F(1)],\ldots,P_F(K)[1-P_F(K)]\}. \tag{29}$$

We next derive the J-divergence between the Gaussian densities, $J(p_g(\mathbf{y} \mid H_1), p_g(\mathbf{y} \mid H_0))$. From the definition of the J-divergence and the KL distance in (11) and (12), we have

$$J(p_g(\mathbf{y} \mid H_1), p_g(\mathbf{y} \mid H_0))$$
$$= \int_{\mathbf{y}} d\mathbf{y}[p_g(\mathbf{y} \mid H_1) - p_g(\mathbf{y} \mid H_0)]\log\frac{p_g(\mathbf{y} \mid H_1)}{p_g(\mathbf{y} \mid H_0)}. \tag{30}$$

Using the fact that $p_g(\mathbf{y} \mid H_i)$ are Gaussian densities $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_i, \Sigma_i)$, $i \in \{0, 1\}$, after some algebra, we obtain

$$J(p_g(\mathbf{y} \mid H_1), p_g(\mathbf{y} \mid H_0))$$
$$= \frac{1}{2}\mathrm{Tr}\left[\Sigma_0\Sigma_1^{-1} + \Sigma_1\Sigma_0^{-1}\right.$$
$$\left. + \left(\Sigma_1^{-1} + \Sigma_0^{-1}\right)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\right] - N \tag{31}$$

where N is the dimension of the received signal vector $\mathbf{y}$ at the FC. Applying the means $\boldsymbol{\mu}_i$ and covariance matrices Σ_i in (23) and (27), we have

$$J(p_g(\mathbf{y} \mid H_1), p_g(\mathbf{y} \mid H_0))$$
$$= \frac{1}{2}\mathrm{Tr}[[\mathbf{R} + \mathbf{HA}(\mathbf{B}_0 + \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{A}^T\mathbf{H}^T]$$
$$\cdot [\mathbf{R} + \mathbf{HAB}_1\mathbf{A}^T\mathbf{H}^T]^{-1}]$$
$$+ \frac{1}{2}\mathrm{Tr}[[\mathbf{R} + \mathbf{HA}(\mathbf{B}_1 + \boldsymbol{\beta}\boldsymbol{\beta}^T)\mathbf{A}^T\mathbf{H}^T]$$
$$\cdot [\mathbf{R} + \mathbf{HAB}_0\mathbf{A}^T\mathbf{H}^T]^{-1}] - N \tag{32}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0$.

The approximated optimal power allocation is the solution to the following optimization problem:

$$\max_{\mathbf{A}} J(p_g(\mathbf{y} \mid H_1), p_g(\mathbf{y} \mid H_0))$$
$$\text{s.t. } \mathrm{Tr}[\mathbf{AA}^T] \leq P_{\mathrm{tot}}$$
$$\mathbf{0} \preceq \mathbf{A} \preceq \sqrt{\mathbf{P}_{\max}}. \tag{33}$$

For the objective function given in (32), the optimization is over the amplitude matrix $\mathbf{A}$, or equivalently the power allocation among the sensors. The optimization problem can be solved by various constrained optimization techniques, and in the simulations we use the interior point method [2], [4].
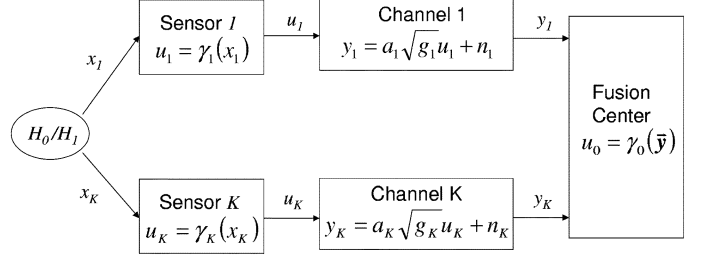

Fig. 2. Distributed detection system with orthogonal channels.

IV. SPECIAL CASE WITH ORTHOGONAL CHANNELS

A special case of the distributed detection system depicted in Fig. 1, is that in which all of the sensors have orthogonal channels for communication with the FC. A system diagram for this case is shown in Fig. 2.

Compared to the system in Fig. 1, this special case has

$$\mathbf{H} = \mathrm{diag}\left\{\sqrt{g_1},\ldots,\sqrt{g_K}\right\} \tag{34}$$
$$\mathbf{R} = \sigma^2\mathbf{I}_K \tag{35}$$

where $\mathbf{I}_K$ is a $K \times K$ identity matrix, and the noises in all the channels are independent and have the same variance σ^2. Here, g_j is the channel power gain for sensor j. By substituting the above two matrices into the optimization problem in (32) and (33), the power allocation for this special case reduces to the solution to the following optimization problem:

$$\max_{\{P_1,\ldots,P_K\}} J(P_1,\ldots,P_K)$$
$$= \sum_{j=1}^{K}\left[\frac{\sigma^2 + \alpha_F(j)g_jP_j}{\sigma^2 + \beta_F(j)g_jP_j} + \frac{\sigma^2 + \alpha_D(j)g_jP_j}{\sigma^2 + \beta_D(j)g_jP_j}\right]$$
$$\text{s.t. } \sum_{j=1}^{K}P_j \leq P_{\mathrm{tot}}$$
$$0 \leq P_j \leq P_{\max}(j), \quad j = 1,\ldots,K \tag{36}$$

where

$$\alpha_F(j) = P_F(j)(1 - P_D(j)) + P_D(j)(P_D(j) - P_F(j)) \tag{37}$$
$$\alpha_D(j) = P_D(j)(1 - P_F(j)) - P_F(j)(P_D(j) - P_F(j)) \tag{38}$$
$$\beta_F(j) = P_D(j)(1 - P_D(j)) \tag{39}$$

and

$$\beta_D(j) = P_F(j)(1 - P_F(j)). \tag{40}$$

Note that $P_j = a_j^2$ is the power allocated to sensor j for transmitting its findings to the FC. The objective function is fully decoupled, a direct result of the orthogonal channels between the sensors and the FC.

The first order partial derivative of $J(P_1,\ldots,P_K)$ with respect to P_j is given by

$$\frac{\partial}{\partial P_j}J(P_1,\ldots,P_K)$$
$$= \frac{(\alpha_F(j) - \beta_F(j))\sigma^2g_j}{(\sigma^2 + \beta_F(j)g_jP_j)^2} + \frac{(\alpha_D(j) - \beta_D(j))\sigma^2g_j}{(\sigma^2 + \beta_D(j)g_jP_j)^2}. \tag{41}$$

It has an interesting property as stated by the following lemma.

Lemma 2: The first order derivative of the objective function $J(P_1, \ldots, P_K)$ with respect to P_j is always nonnegative at any valid power allocation point $P_i \geq 0$. That is

$$\frac{\partial}{\partial P_j} J(P_1, \ldots, P_K)|_{P_i \geq 0} \geq 0. \tag{42}$$

Proof: See Appendix I. ∎

Lemma 2 tells us that the objective function in (36) is non-decreasing with increasing power budget P_{tot}. Since we are maximizing a nondecreasing function, the optimal point is always at the constraint boundary, i.e., $\sum_{j=1}^{K} P_j = P_{\text{tot}}$, or $P_j = P_{\max}(j), j = 1, \ldots, K$. This result is intuitively plausible since it makes full use of the power budget.

Practical sensors should always have $P_D > P_F$, since, if $P_D = P_F$, the sensors do not provide useful information. With this condition, we can easily prove the following corollary.

Corollary 2: If $P_D(j) > P_F(j)$, then the first order derivative of the objective function $J(P_1, \ldots, P_K)$ with respect to P_j is always strictly positive at any valid power allocation point $P_i \geq 0$.

Corollary 2 tells us that there is no stationary point inside the constraint boundary, so gradient based optimization techniques will not get stuck.

The second order partial derivative of $J(P_1, \ldots, P_K)$ with respect to P_j is given by

$$\frac{\partial^2}{\partial P_j^2} J(P_1, \ldots, P_K)$$
$$= -\sigma^2 g_j \left[\frac{[\alpha_F(j) - \beta_F(j)]\beta_F(j)}{[\sigma^2 + \beta_F(j)g_j P_j]^3} \right.$$
$$\left. + \frac{[\alpha_D(j) - \beta_D(j)]\beta_D(j)}{[\sigma^2 + \beta_D(j)g_j P_j]^3} \right]$$
$$= -\sigma^2 g_j \left[\frac{C_0\sigma^6 + C_1\sigma^4 g_j P_j + C_2\sigma^2 g_j^2 P_j^2 + C_3 g_j^3 P_j^3}{[\sigma^2 + \beta_F(j)g_j P_j]^3 [\sigma^2 + \beta_D(j)g_j P_j]^3} \right]$$
$$\tag{43}$$

where

$$C_0 = \beta_F(j)[\alpha_F(j) - \beta_F(j)] + \beta_D(j)[\alpha_D(j) - \beta_D(j)] \tag{44}$$
$$C_1 = \beta_F(j)\beta_D(j)[\alpha_F(j) - \beta_F(j) + \alpha_D(j) - \beta_D(j)] \tag{45}$$
$$C_2 = \beta_F(j)\beta_D(j)[\beta_D(j)[\alpha_F(j) - \beta_F(j)]$$
$$+ \beta_F(j)[\alpha_D(j) - \beta_D(j)]] \tag{46}$$

and

$$C_3 = \beta_F(j)\beta_D(j)[\beta_D(j)^2[\alpha_F(j) - \beta_F(j)]$$
$$+ \beta_F(j)^2[\alpha_D(j) - \beta_D(j)]]. \tag{47}$$

The second-order partial derivative in (43) is not always non-positive, which means the objective function $J(P_1, \ldots, P_K)$ is not always concave. However, the following lemma specifies the region in terms of local sensor observation quality, where the second order derivative of the objective function is indeed non-positive. We again assume that practical sensors have $P_D > P_F$.

Lemma 3: The second order partial derivative of the objective function, $((\partial^2)/(\partial P_j^2))J(P_1, \ldots, P_K) \leq 0$, for any allocated
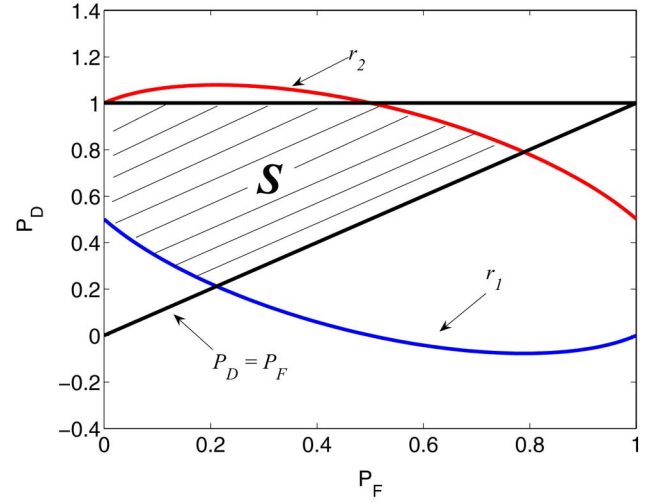

Fig. 3.   Illustration of region $\mathcal{S}$.

power $P_j \geq 0$, if and only if $(P_D(j), P_F(j)) \in \mathcal{S}$, where $\mathcal{S}$ is defined by

$$\left\{ \mathcal{S}(P_D, P_F) \middle| \frac{3}{4} - \frac{1}{2}P_F - \frac{1}{4}\sqrt{1 + 12P_F - 12P_F^2} \right.$$
$$\leq P_D \leq \frac{3}{4} - \frac{1}{2}P_F + \frac{1}{4}\sqrt{1 + 12P_F - 12P_F^2},$$
$$\left. 0 \leq P_F < P_D \leq 1 \right\}. \tag{48}$$

Proof: See Appendix II. ∎

Region $\mathcal{S}$ is depicted in Fig. 3, in which $r_1 = (3/4) - (1/2)P_F - (1/4)\sqrt{1 + 12P_F - 12P_F^2}$ and $r_2 = (3/4) - (1/2)P_F + (1/4)\sqrt{1 + 12P_F - 12P_F^2}$. We will show that, if all the sensors operate in region $\mathcal{S}$, the power allocation can be solved by a weighted waterfilling algorithm.

To derive the algorithm, we will use the technique of Lagrange multipliers [2], [4]. The Lagrangian associated with the constrained optimization problem in (36) is

$$L(P_1, \ldots, P_K, \lambda)$$
$$= \sum_{j=1}^{K} \left[\frac{\sigma_j^2 + \alpha_F(j)P_j}{\sigma_j^2 + \beta_F(j)P_j} + \frac{\sigma_j^2 + \alpha_D(j)P_j}{\sigma_j^2 + \beta_D(j)P_j} \right]$$
$$- \lambda \left(\sum_{j=1}^{K} P_j - P_{\text{tot}} \right) + \sum_{j=1}^{K} \nu_j P_j$$
$$- \sum_{j=1}^{K} \eta_j (P_j - P_{\max}(j)) \tag{49}$$

where λ, $\{\nu_j\}_{j=1}^{K}$, and $\{\eta_j\}_{j=1}^{K}$ are Lagrange multipliers. The Karush–Kuhn–Tucker necessary conditions for optimality [4] are

$$\frac{\partial}{\partial P_j} J(P_1, \ldots, P_K)|_{P_j^*} - \lambda^* + \nu_j^* - \eta_j^* = 0 \tag{50}$$

$$\lambda^* > 0, \quad \text{if } \sum_{j=1}^{K} P_j^* = P_{\text{tot}} \tag{51}$$

$$\lambda^* = 0, \quad \text{if } \sum_{j=1}^{K} P_j^* < P_{\text{tot}} \tag{52}$$

$$\nu_j^* > 0, \quad \text{if } P_j^* = 0 \tag{53}$$

$$\nu_j^* = 0, \quad \text{if } P_j^* > 0 \tag{54}$$

and

$$\eta_j^* > 0, \quad \text{if } P_j^* = P_{\max}(j) \tag{55}$$

$$\eta_j^* = 0, \quad \text{if } P_j^* < P_{\max}(j). \tag{56}$$

All variables with superscript "$*$" are at their optimal values. Since the optimal solution is always on the total power constraint boundary as indicated by Lemma 2, (52) is inapplicable except in the trivial case when $\sum_{j=1}^{K} P_{\max}(j) < P_{\text{tot}}$ and all the sensors just transmit at full power. So, in other words, we consider that the total power constraint is always active (meaning $\sum_{j=1}^{K} P_j^* = P_{\text{tot}}$) and λ^* is always positive. Similarly, ν_j^* (or η_j^*) is positive only when the constraint $P_j^* \geq 0$ (or $P_j^* \leq P_{\max}(j)$) is active.

(50) is the key equation to solve. For each fixed value of λ^*, we can solve (50) to obtain the corresponding P_j^*, ν_j^*, and η_j^*. We can then calculate the corresponding $\sum_{j=1}^{K} P_j^*$. The goal here is to find a λ^* such that $\sum_{j=1}^{K} P_j^* = P_{\text{tot}}$.

Substituting (41) into (50), we have

$$\frac{(\alpha_F(j) - \beta_F(j))\sigma^2 g_j}{(\sigma^2 + \beta_F(j)g_j P_j)^2} + \frac{(\alpha_D(j) - \beta_D(j))\sigma^2 g_j}{(\sigma^2 + \beta_D(j)g_j P_j)^2}$$
$$- \lambda^* + \nu_j^* - \eta_j^* = 0. \tag{57}$$

Let us define

$$w_{j,0} \triangleq \frac{\partial}{\partial P_j} J(P_1, \ldots, P_K)\Big|_{P_j^* = 0}$$
$$= g_j(\alpha_F(j) - \beta_F(j)) + \alpha_D(j) - \beta_D(j))/\sigma^2 \tag{58}$$

and

$$w_{j,1} \triangleq \frac{\partial}{\partial P_j} J(P_1, \ldots, P_K)\Big|_{P_j^* = P_{\max}(j)}$$
$$= \frac{(\alpha_F(j) - \beta_F(j))\sigma^2 g_j}{(\sigma^2 + \beta_F(j)g_j P_{\max}(j))^2}$$
$$+ \frac{(\alpha_D(j) - \beta_D(j))\sigma^2 g_j}{(\sigma^2 + \beta_D(j)g_j P_{\max}(j))^2}. \tag{59}$$

For sensor j operating at $(P_D(j), P_F(j)) \in \mathcal{S}$, we can see that, when $\lambda^* > w_{j,0}$, we have $P_j^* = 0$, $\eta_j^* = 0$, and $\nu_j^* = \lambda^* - w_{j,0}$. Sensor j starts to get positive power allocation $P_j^* > 0$ when $\lambda^* < w_{j,0}$, and in this case the constraint $0 \leq P_j^* \leq P_{\max}(j)$ is inactive ($\nu_j^* = \eta_j^* = 0$). λ^* is monotonically decreasing with increasing P_j^*, as long as $w_{j,1} < \lambda^* < w_{j,0}$. This can be easily verified using Lemma 3 since $((\partial^2)/(\partial P_j^2))J(P_1, \ldots, P_K)$ is always nonpositive for $(P_D(j), P_F(j)) \in \mathcal{S}$. When $\lambda^* < w_{j,1}$, P_j^* is now "clamped" at $P_{\max}(j)$, so we have $\nu_j^* = 0$ and $\eta_j^* = w_{j,1} - \lambda^* > 0$.

This case is depicted as the dashed line in Fig. 4. Regardless of the value of λ^*, we can easily verify that there is always a one-to-one mapping [through (50) or (57)] between P_j^* and λ^*, and P_j^* is nondecreasing with decreasing λ^*.

We have the following observations based on the above analysis. (1) If all the sensors operate at $(P_D(j), P_F(j)) \in \mathcal{S}$,

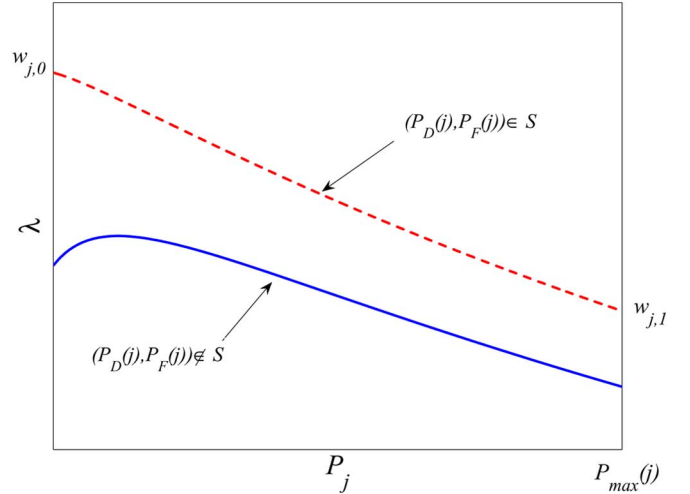Fig. 4. λ as a function of P_j for sensors operating in or not in region $\mathcal{S}$.
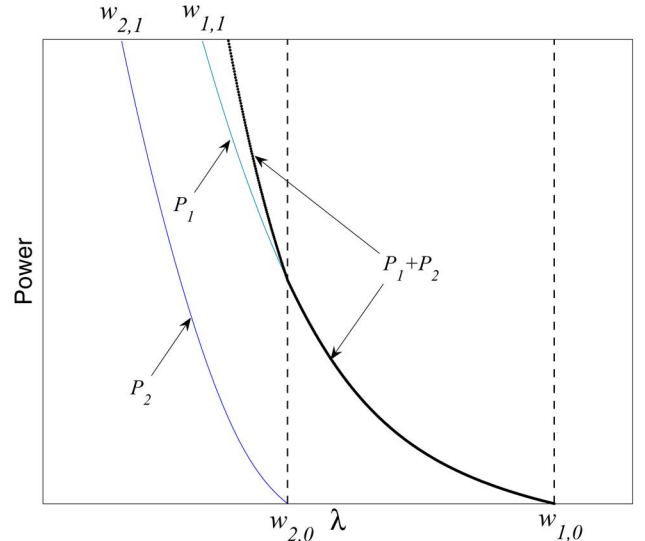

Fig. 5. Power allocation as a function of λ^* when all the sensors are operating in region $\mathcal{S}$.

$j = 1, \ldots, K$, there is a one-to-one mapping between $\sum_{j=1}^{K} P_j^*$ and λ^*, and $\sum_{j=1}^{K} P_j^*$ is nondecreasing with decreasing λ. (2) The sensors get positive power allocation with increasing power budget (hence decreasing λ^*) in a sequential fashion, and it is determined by $w_{j,0}$.

The above observations with a two-sensor case are illustrated in Fig. 5. Based on the observations, the solution can be found through a "weighted waterfilling" procedure, specified by the following algorithm. Initially the sensors send their local detection quality $\{P_D(j), P_F(j)\}_{j=1}^{K}$ to the FC, and then the algorithm is executed at the FC.

Algorithm 1:
1) The FC estimates the channel power gain $\{g_j\}_{j=1}^{K}$ of the sensors and the noise variance σ^2.
2) The FC calculates $\{w_{j,0}\}_{j=1}^{K}$ using (58) and ranks them such that $w_{j,0} \geq w_{j+1,0}$. The FC also solves for $\{w_{j,1}\}_{j=1}^{K}$ using (59). Then the FC calculates the power allocations $\{P_{k,w_{j,0}}^*\}_{k=1}^{K}$ with $\lambda^* = w_{j,0}$ for each j using (57).
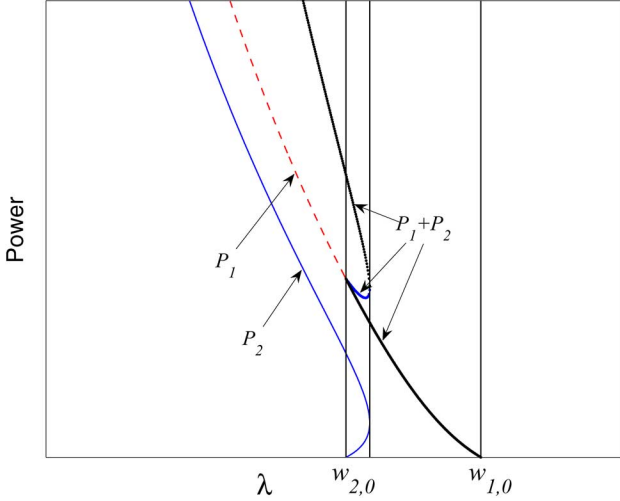
Fig. 6. Power allocation as a function of λ^* when one or more sensors are not operating in region $\mathcal{S}$.

3) The FC finds the largest j' such that $\sum_{k=1}^{K} P_{k,w_{j',0}}^* \leq P_{\text{tot}}$, and assigns $w_a = w_{j',0}$ and $w_b = w_{j'+1,0}$. If $j' = K$, the FC sets $w_b = 0$. The rest of the algorithm conducts a simple line search on λ^* between w_a and w_b such that $\sum_{k=1}^{K} P_{k,\lambda^*}^* = P_{\text{tot}}$.

4) If $\sum_{k=1}^{K} P_{k,w_b}^* - \sum_{k=1}^{K} P_{k,w_a}^* \geq \epsilon$, where ϵ is a small positive number, the algorithm goes to (5). Otherwise, the FC stops the algorithm and broadcasts the desired power allocations $\{P_{k,w_a}^*\}_{k=1}^{K}$ to the sensors.

5) The FC sets $w_c = (w_a + w_b)/2$, following bisection rule. If $\sum_{k=1}^{K} P_{k,w_c}^* \leq P_{\text{tot}}$, the FC sets $w_a = w_c$. Otherwise, the FC sets $w_b = w_c$. The algorithm goes to (4).

Algorithm 1 can be easily seen to converge because of the monotonicity between total power budget P_{tot} and λ^*. And the simple line search of λ^* between w_a and w_b should converge very quickly [2].

If sensor j operates at $(P_D(j), P_F(j)) \notin \mathcal{S}$, λ^* is monotonically increasing with P_j^* when P_j^* is small and is monotonically decreasing with P_j^* when P_j^* grows larger. This is because C_0 is negative and C_1, C_2, and C_3 are nonnegative in (43). Thus, $((\partial^2)/(\partial P_j^2))J(P_1,\ldots,P_K)|_{P_j^*=0}$ is positive, and will eventually become negative with P_j^* sufficiently large. Therefore, $(\partial)/(\partial P_j)J(P_1,\ldots,P_K)$ has a single local maximum at some $P_j^* > 0$, and it is possible that (57) has two solutions for P_j^* for a single λ^*. This case is also shown in Fig. 4.

If one or more sensors operate at $(P_D(j), P_F(j)) \notin \mathcal{S}$, the monotonicity and one-to-one mapping between λ^* and $\sum_{j=1}^{K} P_j^*$ may be invalid, as shown in Fig. 6. Therefore, the computationally efficient Algorithm 1 does not work for this case. The solution can still be obtained from general constrained optimization techniques, such as the interior point method [2], [4].

V. SIMULATIONS

In this section, numerical results are provided to illustrate the power allocation scheme developed in this paper. In the simulations, we consider the following settings. There are K sensors scattered around an FC and the distances from the sensors to the FC are $\{d_k\}_{k=1}^{K}$. The pathloss of signal power at the FC from sensor k follows the Motley–Keenan pathloss model (expressed in decibels) without the wall and floor attenuation factor [23]:

$$PL_k = PL_0 + 10n \log_{10}(d_k/d_0) \tag{60}$$

where PL_0 is a constant set to 55 dB, and d_0 is also a constant set to 1 m in the simulations. Here, n is the pathloss exponent, which is set to 2 for free space propagation. The channel power gain for sensor k is $g_k = -PL_k$ in dB. The noise variance at the FC is $\sigma^2 = -70$ dBm, and we assume the noise covariance matrix is $\mathbf{R} = \sigma^2 \mathbf{I}_k$. The maximum transmitting power of each sensor is $P_{\max} = 2$ mW (3 dBm). The total power budget in the simulations is below or equal to $2K$ mW, otherwise each sensor just uses maximum transmitting power (a trivial case). All the sensors perform Neyman-Pearson detection with false alarm probabilities set to $P_F(k) = 0.04$, $k = 1,\ldots,K$. The detection probabilities may vary according to their local observation qualities. The FC also uses a Neyman–Pearson detector targeting the same false alarm probability as the local sensors,[3] i.e., $P_{F,FC} = 0.04$.

We will investigate three scenarios: 1) two sensors with orthogonal MIMO channels; 2) two sensors with nonorthogonal MIMO channels; and 3) ten sensors with orthogonal MIMO channels.

A. Two Sensors With Orthogonal Channels

Two sensors are located $d_1 = 2$ m and $d_2 = 5$ m away from the FC, and communicate with the FC through orthogonal channels. Channel gains g_1 and g_2 are calculated by (60), and are -61 and -69 dB, respectively.

We will consider four cases with various local detection quality combinations.

Case V-A1: $P_D(1) = 0.1$, $P_D(2) = 0.9$.
Case V-A2: $P_D(1) = 0.7$, $P_D(2) = 0.9$.
Case V-A3: $P_D(1) = 0.9$, $P_D(2) = 0.9$.
Case V-A4: $P_D(1) = 0.9$, $P_D(2) = 0.7$.

In Case V-A1 one sensor does not operate in region $\mathcal{S}$, so the interior point optimization algorithm is used in this case. In Case V-A2, Case V-A3, and Case V-A4, both sensors operate in region $\mathcal{S}$, thus Algorithm 1 is used. The total power budget P_{tot} varies from -14 to 6 dBm (when each sensor transmits at full power 2 mW).

In addition to the proposed power allocation, we also include an equal power allocation and an equal received SNR power allocation for comparison. The equal power allocation simply distributes power equally among the sensors, without considering channel or local detection quality. The equal received SNR power allocation considers channel quality only and distributes power among sensors in such a way that the received signals from the sensors have the same SNR.

---

[3]The operating points in terms of targeted false alarm probabilities of the detectors at the sensors and the FC can be designed to be different from one another, and the analysis in this paper does not require the false alarm probabilities to be the same. The optimal design of the operating points of the detectors is beyond the scope of this paper, and we use the same false alarm probability for the sake of simplicity.
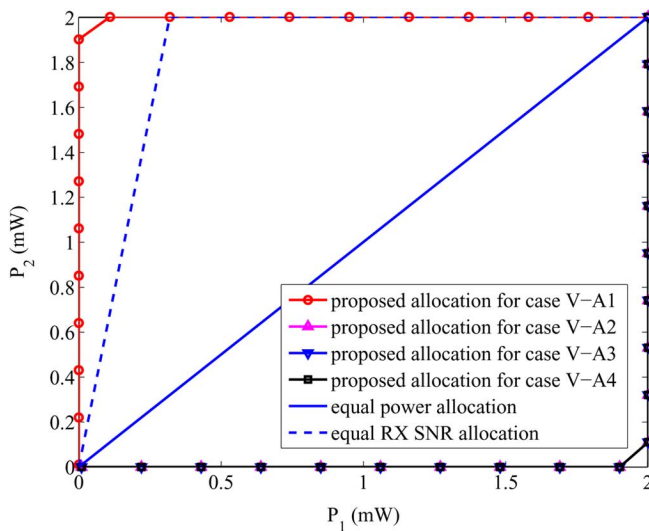
Fig. 7. Equal power allocation, equal received SNR allocation, and the proposed power allocation for the four cases in Section V-A. (Note that the curves for $\triangle$, $\triangledown$, and $\square$ overlay in this graph.)
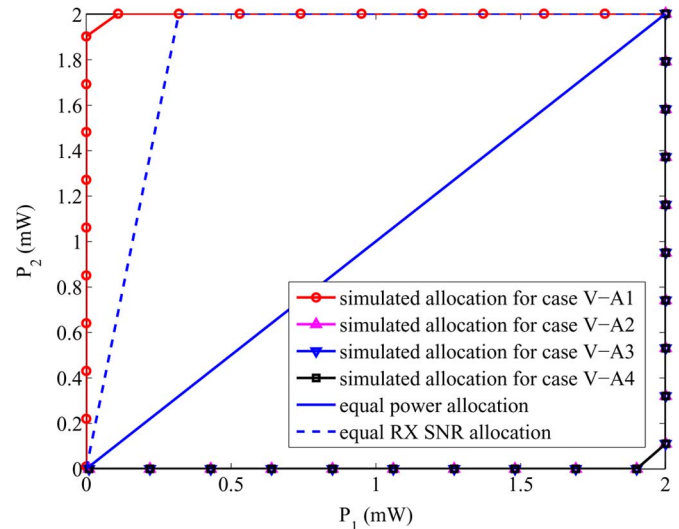

Fig. 8. Equal power allocation, equal received SNR allocation, and the simulated optimal power allocation for the four cases in Section V-A. (Note that the curves for $\triangle$, $\triangledown$, and $\square$ overlay in this graph.)

Fig. 7 shows the proposed power allocation as well as the equal power allocation and the equal received SNR power allocation, we can see that for Case V-A1 the proposed power allocation distributes all the power to sensor 2, until the maximum output power is reached for sensor 2, and then sensor 1 starts to get positive power allocation. This is because, although sensor 1 is closer to the FC (hence it has a better channel), its detection quality is much worse than that of sensor 2. For Case V-A2, the detection quality of sensor 1 is still worse than that of sensor 2, but the gap is small enough for sensor 1's better communication channel to show difference. The proposed allocation distributes all the power to sensor 1, until the maximum output power is reached for sensor 1, and then sensor 2 starts to get power allocation. For Case V-A3 and Case V-A4, sensor 1 has a better communication channel and equal or better local detection quality, so it is not surprising to see that the proposed allocation distributes power to sensor 1 as much as possible, and these two cases have the same power allocation as Case V-A2. The waterfilling effect of the proposed power allocation is obvious in this scenario. The equal power allocation and the equal received SNR allocation do not change between the four cases because they are not affected by the local detection quality.

Recall that the proposed power allocation scheme is based on the J-divergence instead of detection probability and false alarm probability. Furthermore, the J-divergence between two Gaussian mixture distributions, i.e., that of the received signals at the FC under the two hypotheses, is approximated in this optimization by the J-divergence between two Gaussian distributions, with the same means and covariance matrices as the Gaussian mixtures. We next show the quality of this approximation.

Fig. 8 shows the optimal power allocation found by simulations. The FC uses a Neyman-Pearson detector based on the likelihood ratio of the received signal $\mathbf{y}$. The optimal power allocation is the one that produces the highest $P_{D,\text{FC}}$ for a given total power budget. The optimal power allocation is

found by a brute-force grid search in a two dimensional space of all possible power allocations. For each possible power allocation point, 2×10^4 Monte Carlo runs are used to provide the corresponding $P_{D,\text{FC}}$.

We can see that the simulated optimal power allocations in Fig. 8 perfectly match the proposed power allocations in Fig. 7.

The contours of the approximated J-divergence (used as the objective function to develop the proposed power allocation) and the simulated $P_{D,\text{FC}}$ for Case V-A3 are plotted in Figs. 9 and 10. The contours for the other cases are similar and are omitted due to limited space. We can see from Figs. 9 and 10 that the two contours match each other well at any power allocation point in this scenario, including the case in which either sensor transmits at full power (2 mW or equivalently 3 dBm). When sensor 1 transmits at full power (3 dBm), the corresponding received SNR is about 12 dB. So we can see the approximation works well in this scenario even with received SNR as high as 12 dB. This is the reason for the perfect match between the proposed allocation and the simulated optimal allocation.

In Figs. 11–14, we plot the detection probability at the FC $P_{D,\text{FC}}$ as a function of the total power budget P_{tot} for the four cases. The performance of the proposed power allocation matches that of the simulated optimal power allocation very well in all four cases (the two curves overlay in the four figures), and it can save almost 3 dB in P_{tot} compared to the equal power allocation to achieve the same $P_{D,\text{FC}}$. The equal received SNR power allocation considers only the channel quality, so it performs even worse than the equal power allocation in Case V-A2 through Case V-A4, where optimally sensor 1 should get more power than sensor 2. The equal received SNR power allocation happens to be similar to the proposed power allocation only in Case V-A1.

B. Two Sensors With Nonorthogonal MIMO Channels

The setting here is similar to the setting in Section V-A, but the data transmission is over nonorthogonal channels. The
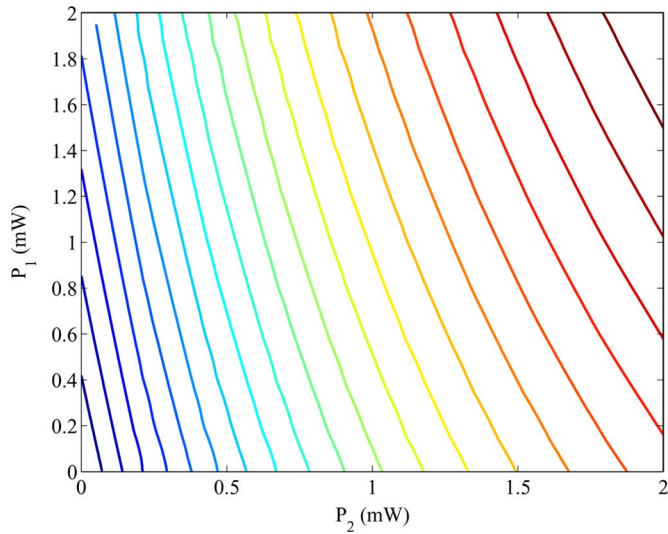
Fig. 9. Contour of the approximated J-divergence objective function for Case V-A3.
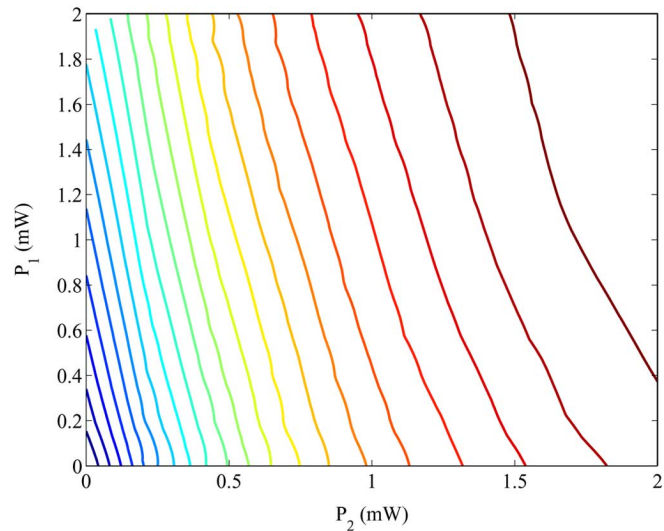

Fig. 11. FC detection probability $P_{D,\text{FC}}$ as a function of P_{tot} of Case V-A1.

Fig. 10. Contour of the simulated $P_{D,\text{FC}}$ for Case V-A3.

Fig. 12. FC detection probability $P_{D,\text{FC}}$ as a function of P_{tot} of Case V-A2.

channel matrix is given by

$$\mathbf{H} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} g_1 & 0 \\ 0 & g_2 \end{bmatrix}. \tag{61}$$

g_1 and g_2 are the same as those in Section V-A. $\rho = 0.2$ is the interference coefficient. We consider four cases, Case V-B1 through Case V-B4, with exactly the same local detection quality combinations as those of Case V-A1 through Case V-A4. The interior point optimization algorithm is used to solve the proposed power allocation for all four cases.

Fig. 15 shows the proposed power allocation as well as the equal power allocation and the equal received SNR allocation. The major difference between Figs. 15 and 7 in Section V-A is that the waterfilling effect in Case V-B2 through Case V-B4 is not as obvious as that in Case V-A2 through Case V-A4. The nonorthogonal channel makes the contribution from the sensors
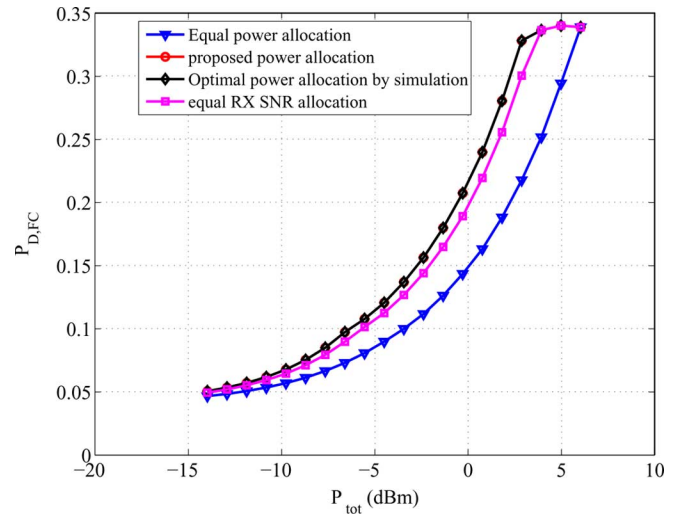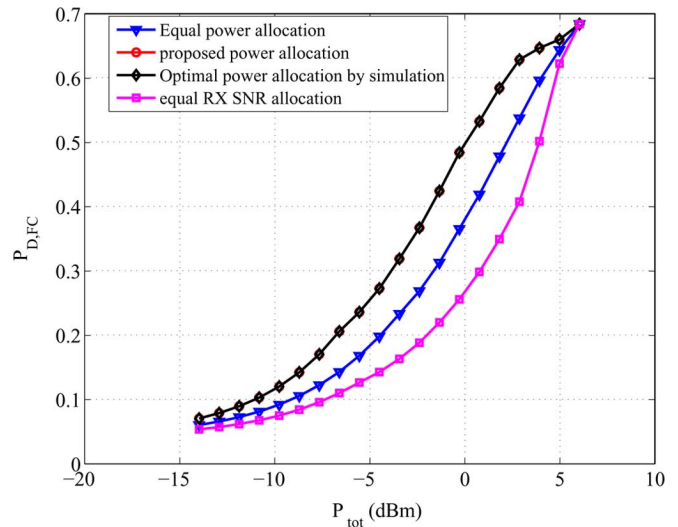
at the FC dependent. So the power allocation is less extreme in the sense that the "better" sensors take all the power.

Fig. 16 shows the optimal power allocation found by simulations for the four cases. For Case V-B1 and Case V-B4, the simulated optimal power allocation matches the proposed power allocation in Fig. 15. For Case V-B2 and Case V-B3, the simulated optimal power allocation is different from the proposed power allocation when the total power budget is high. For these two cases, $P_{D,\text{FC}}$ as a function of power allocation is quite "flat" in the high total power region. This can be seen from the wider gaps between contour lines in Fig. 18. So the difference between the proposed power allocations and the simulated optimal power allocation in the high total power region for Case V-B2 and Case V-B3 is nothing but a pronounced artifact of Monte Carlo trials. In Figs. 20 and 21, the performance of the proposed power allocation is still very close to that of the simulated power allocation.

The contours of the approximated J-divergence and the simulated $P_{D,\text{FC}}$ for Case V-B2 are plotted in Figs. 17 and 18. The
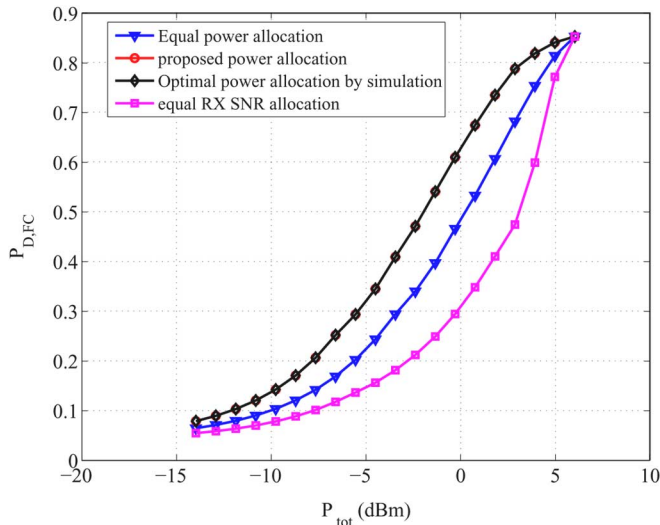
Fig. 13.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-A3.

Fig. 15.   Equal power allocation, equal received SNR allocation, and the proposed power allocation for the four cases in Section V-B.
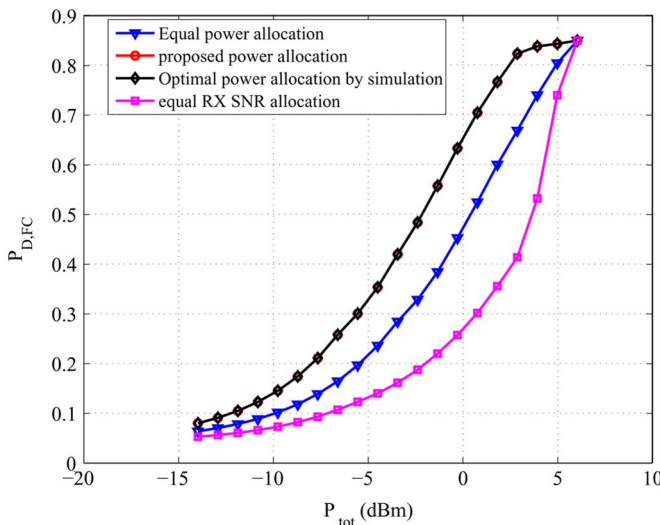
Fig. 14.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-A4.

Fig. 16.   Equal power allocation, equal received SNR allocation, and the simulated optimal power allocation for the four cases in Section V-B.

two contours match each other well, but the artifact of Monte Carlo trials in Fig. 18 is obvious, as discussed above.

In Figs. 19–22, we plot the detection probability at the FC $P_{D,\mathrm{FC}}$ as a function of the total power budget P_{tot}, for the four cases. The performance of the proposed power allocation matches that of the simulated optimal power allocation very well in all four cases (the two curves overlay in the four figures). Compared to Figs. 11–14 in Section V-A, the performance gap between the proposed power allocation and the equal power allocation (as well as the equal received SNR allocation) is slightly narrower.

C. Ten Sensors With Orthogonal MIMO Channels

Here we consider ten sensors scattered around an FC. We will investigate five cases, according to various sensor distance and detection probability combinations. In Case V-C1, Case V-C2, and Case V-C4, some of the sensors do not operate in region $\mathcal{S}$, so the interior point optimization algorithm is used in these three cases. In Case V-C3 and Case V-C5, all the sensors operate in
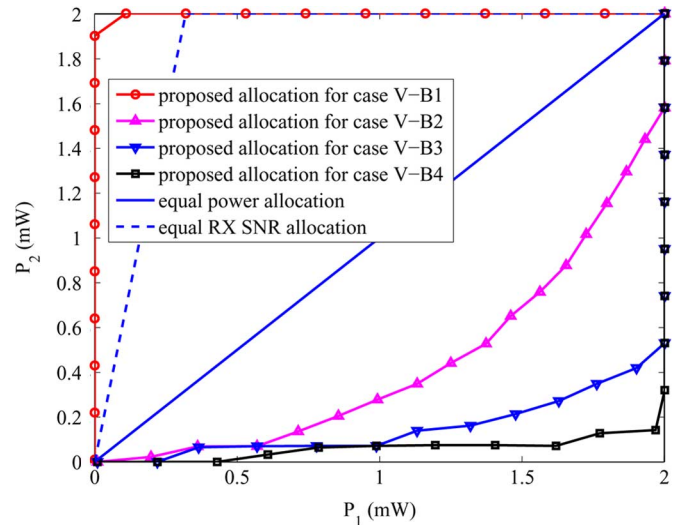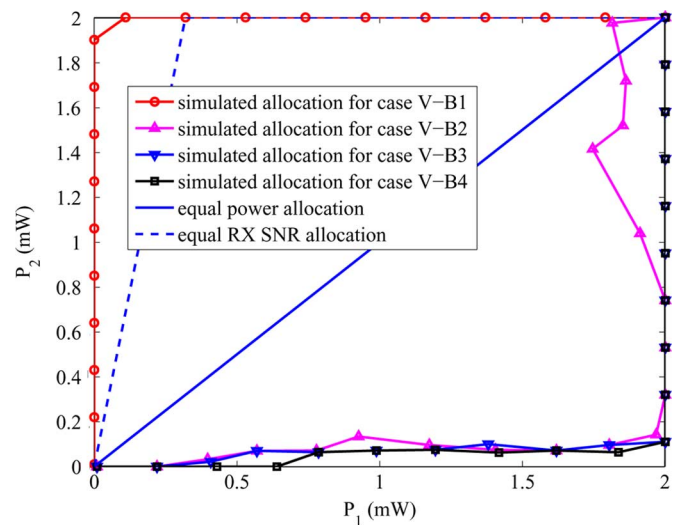
region $\mathcal{S}$, thus Algorithm 1 is used. The total power budget P_{tot} varies from -7 dBm to 13 dBm (when each sensor transmits at full power 2 mW).

1) Case V-C1: The distance between sensor j and the FC is $d_j = 2 + 0.6(j-1)$ m, e.g., $d_1 = 2$ m and $d_{10} = 7.4$ m. Sensor j has detection probability of $P_D(j) = 0.1 + 0.09(j-1)$, e.g., $P_D(1) = 0.1$ and $P_D(10) = 0.91$. In this case, sensors closer to the FC have worse detection probability.

The percentage of the total power budget P_{tot} allocated to each sensor is shown in Table I. When P_{tot} is low, more power is distributed to sensors farther away from the FC. Intuitively this is because even though sensors closer to the FC have good channel gain, their local detection qualities are much worse than those of the farther sensors. As P_{tot} increases, power is distributed more evenly among the sensors, because some sensors have already reached their maximum output power. Eventually, when the total power budget reaches 13 dBm, every sensor transmits at $P_{\mathrm{max}} = 3$ dBm.
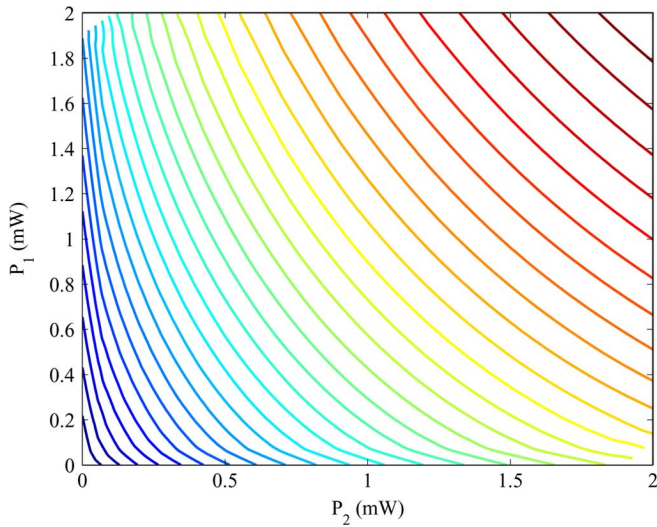
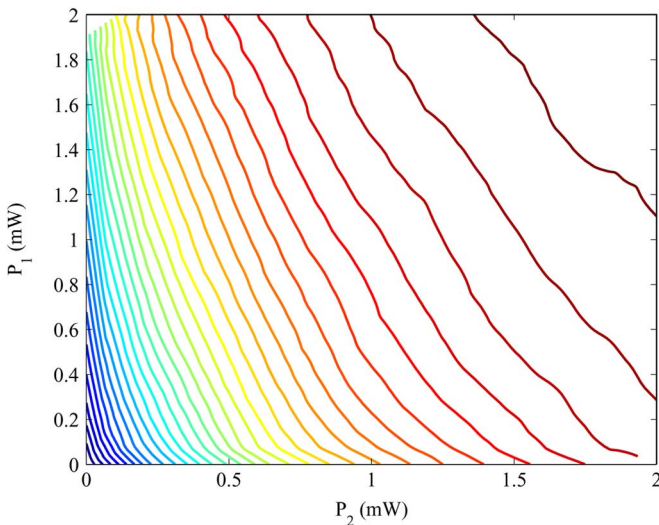Fig. 17.   Contour of the approximated J-divergence objective function for Case V-B2.

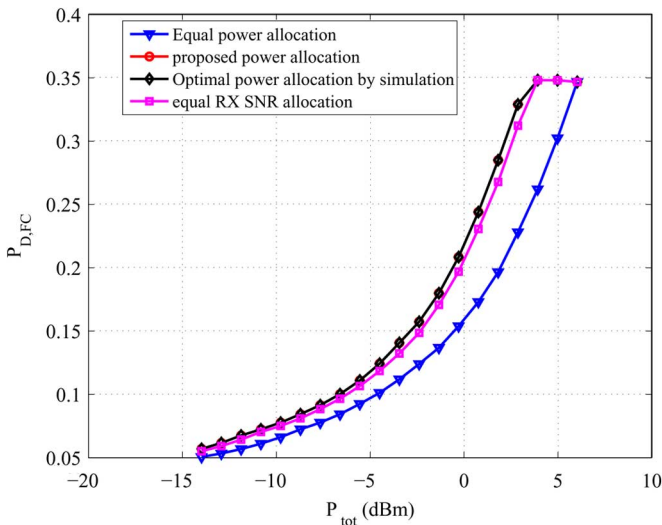Fig. 18.   Contour of the simulated $P_{D,\mathrm{FC}}$ for Case V-B2.

Fig. 19.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-B1.

Fig. 23 shows that the approximated and actual (by Monte Carlo simulation) J-divergences are very close even when every
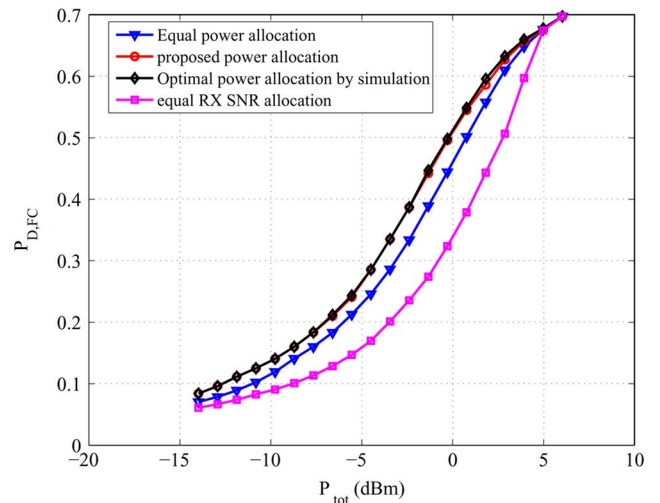

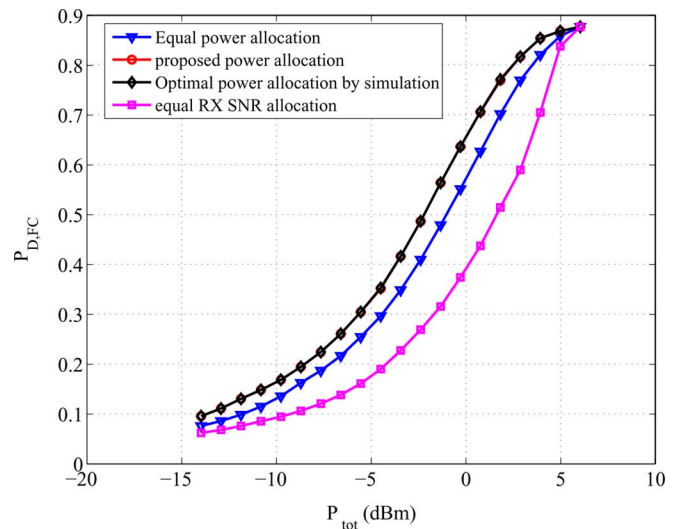Fig. 20.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-B2.

Fig. 21.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-B3.

sensor is transmitting at full power. The maximum received SNR of the closest sensor at the FC is about 12 dB. Similar to Section V-A, we can see that the approximation in J-divergence works well with SNR as high as 12 dB.

Fig. 24 shows the simulated detection probability at the FC $P_{D,\mathrm{FC}}$ of the proposed power allocation and the equal power allocation.[4] In this case, the proposed power allocation can use about 1 dB less power than equal power allocation to reach the same detection performance at the FC.

2) Case V-C2: The distance between sensor j and the FC is $d_j = 2 + 0.6(j-1)$ m. Sensor j has detection probability of $P_D(j) = 0.4 + 0.06(j-1)$, e.g., $P_D(1) = 0.4$ and $P_D(10) = 0.94$. In this case, sensors closer to the FC still have worse detection probability, but the gap in local detection qualities is not as large as that in case 1.

[4]For this scenario with ten sensors, the complexity is too high to find the optimal power allocation by brute-force search and simulations in a ten-dimensional space of all possible power allocations. So in this scenario, we will not include the optimal power allocation found by simulation that gives the highest $P_{D,\mathrm{FC}}$.
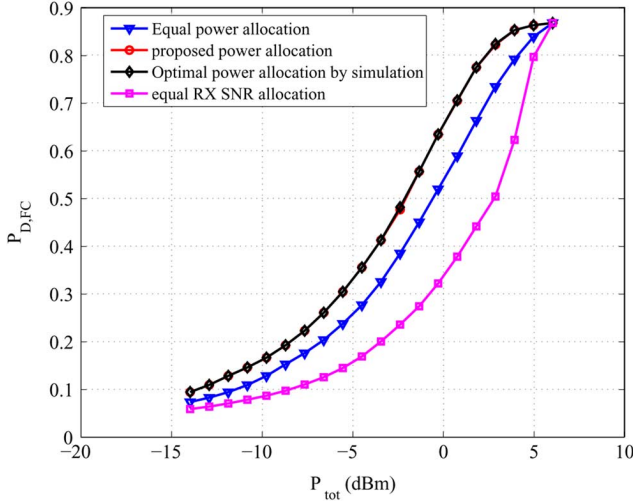
Fig. 22.   FC detection probability $P_{D,\mathrm{FC}}$ as a function of P_{tot} of Case V-B4.

TABLE I
PERCENTAGE OF THE TOTAL POWER ALLOCATED TO
EACH SENSOR FOR CASE V-C1

P_{tot} (dBm)	1	2	3	4	5	6	7	8	9	10
-7.0	0	0	0	0	0	7	15	21	26	31
-2.8	0	0	0	0	6	11	16	19	23	25
3.5	0	0	0	5	9	12	15	17	20	22
8.8	0	3	7	9	10	11	13	14	16	17
13	10	10	10	10	10	10	10	10	10	10

TABLE II
PERCENTAGE OF TOTAL POWER ALLOCATED TO EACH SENSOR FOR CASE V-C2

P_{tot} (dBm)	1	2	3	4	5	6	7	8	9	10
-7.0	100	0	0	0	0	0	0	0	0	0
-2.8	74	26	0	0	0	0	0	0	0	0
3.5	36	23	15	10	7	5	3	1	0	0
8.8	18	14	12	10	9	8	8	7	7	7
13	10	10	10	10	10	10	10	10	10	10

In this case, Table II shows that when P_{tot} is low, more power is distributed to sensors closer to the FC. The advantage in channel gain of sensors closer to the FC has offset their disadvantage in detection quality. The water filling effect is obvious here, and the sensors get positive power allocation in a sequential fashion.

In Fig. 25, the approximated and actual J-divergences are still very close. Fig. 26 shows the proposed power allocation has a maximum power saving of 4 dB (more than 50%) compared with equal power allocation.

3) Case V-C3: $d_j = 2 + 0.6(j-1)$ m, and $P_D(j) = 0.8$. In Table III power allocation is even more biased toward sensors closer to the FC compared to Case V-C2. All the sensors now have the same detection quality, but the sensors closer to the FC have the advantage in channel gain.

In Fig. 27, the approximated and actual J-divergences start to show some difference. Fig. 28 shows the proposed power allocation has a maximum power saving of more than 5 dB.
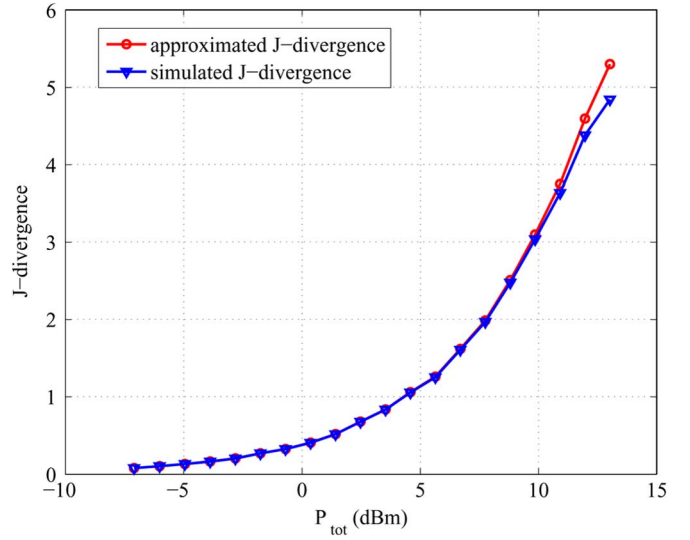
Fig. 23.   Approximated and simulated J-divergence as a function of P_{tot} for the proposed power allocation of case V-C1.
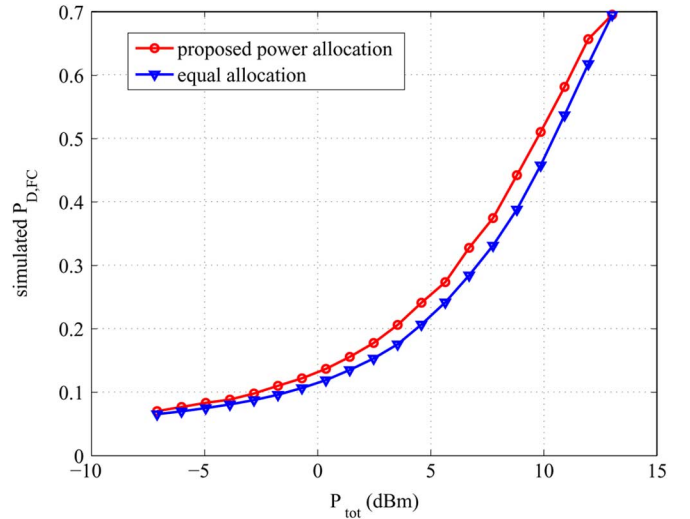
Fig. 24.   Simulated $P_{D,\mathrm{FC}}$ as a function of P_{tot} for the proposed power allocation and equal power allocation of case V-C1.

4) Case V-C4: $d_j = 2 + 0.6(j-1)$ m, and $P_D(j) = 0.94 - 0.06(j-1)$. In this case, the sensors closer to the FC have advantage in both channel gain and local detection quality. So Table IV power allocation is even more biased toward sensors closer to the FC compared with Case V-C3.

In Fig. 29, the approximated and actual J-divergences have more difference than the previous three cases, but their shapes are still quite similar. Fig. 30 shows that the proposed power allocation consumes only less than 25% (has more than 6-dB savings) of the total power used by the equal power allocation to achieve the same detection performance at the FC.

5) Case V-C5: $d_j = 4$ m, and $P_D(j) = 0.8$. In this case, all sensors have the same detection probability and distance from the FC. This case serves as a sanity check because intuitively the sensors should always have equal power allocation in this case. Table V and Fig. 31 verify this intuition.
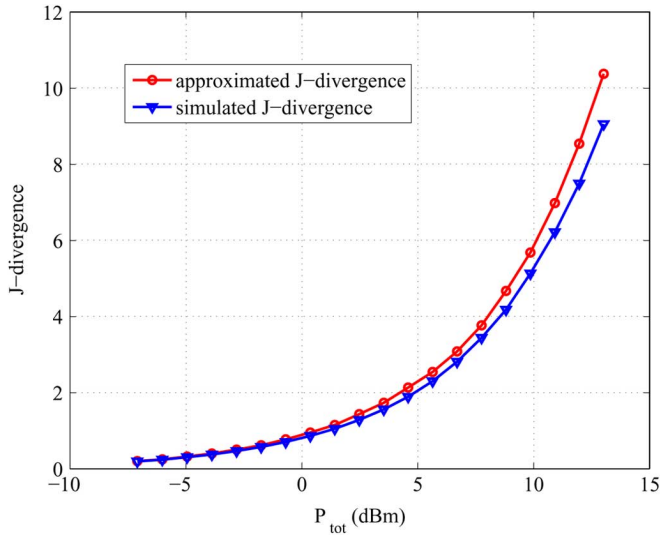
Fig. 25. Approximated and simulated J-divergence as a function of P_{tot} for the proposed power allocation of case V-C2.
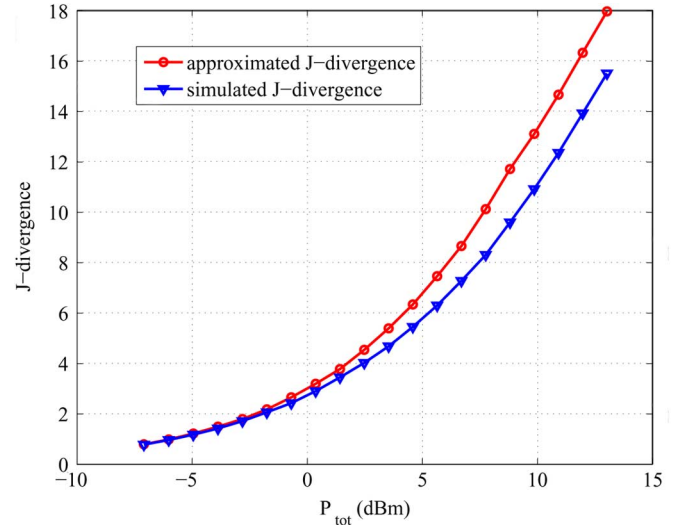

Fig. 27. Approximated and simulated J-divergence as a function of P_{tot} for the proposed power allocation of case V-C3.
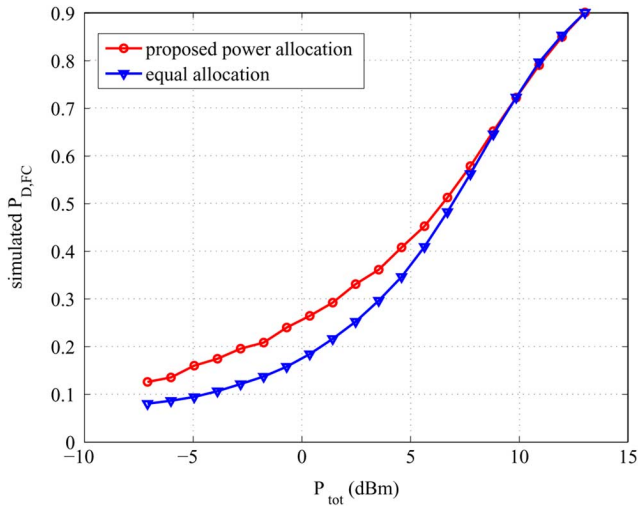

Fig. 26. Simulated $P_{D,\text{FC}}$ as a function of P_{tot} for the proposed power allocation and equal power allocation of case V-C2.
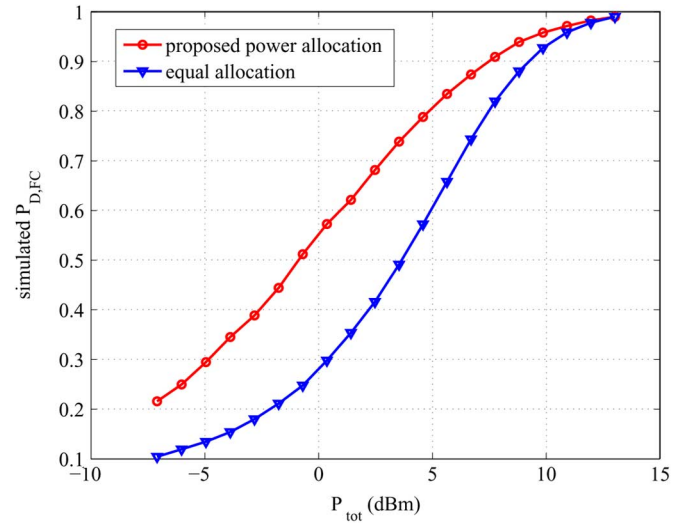

Fig. 28. Simulated $P_{D,\text{FC}}$ as a function of P_{tot} for the proposed power allocation and equal power allocation of case V-C3.

TABLE III
PERCENTAGE OF TOTAL POWER ALLOCATED TO EACH SENSOR FOR CASE V-C3

P_{tot} (dBm)	1	2	3	4	5	6	7	8	9	10
-7.0	100	0	0	0	0	0	0	0	0	0
-2.8	81	19	0	0	0	0	0	0	0	0
3.5	54	33	13	0	0	0	0	0	0	0
8.8	26	26	21	15	8	3	0	0	0	0
13	10	10	10	10	10	10	10	10	10	10

TABLE IV
PERCENTAGE OF TOTAL POWER ALLOCATED TO EACH SENSOR FOR CASE V-C4

P_{tot} (dBm)	1	2	3	4	5	6	7	8	9	10
-7.0	100	0	0	0	0	0	0	0	0	0
-2.8	100	0	0	0	0	0	0	0	0	0
3.5	73	27	0	0	0	0	0	0	0	0
8.8	26	26	26	15	7	0	0	0	0	0
13	10	10	10	10	10	10	10	10	10	10

VI. CONCLUSION

In this paper we have studied the problem of optimal power allocation for distributed detection over MIMO channels in wireless sensor networks. Our contribution is novel compared to the pervious work in the following aspects: 1) we have considered a distributed detection system with a MIMO channel to account for nonideal communications between the sensors and the FC; 2) we have assumed that there are a finite number of sensors and the sensors have independent but nonidentically distributed observations; 3) We also have assumed both individual and joint constraints on the power that the sensors can expend to transmit their local decisions to the FC; 4) we have developed a power allocation scheme to distribute the total power budget among the sensors so that the detection performance at the FC is optimized in terms of the
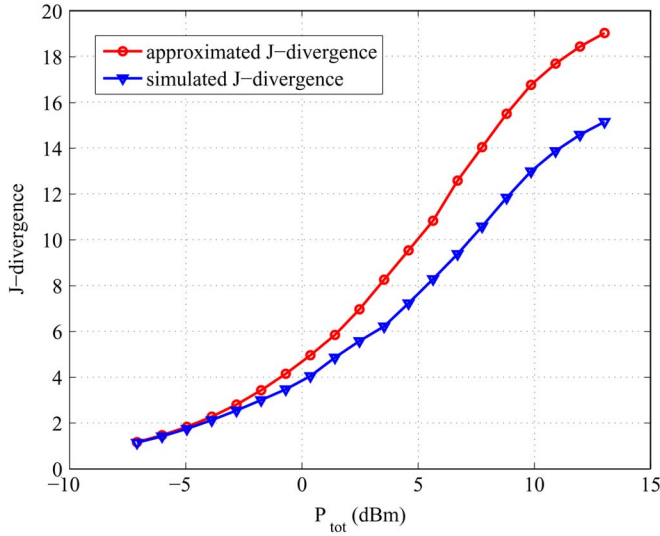
Fig. 29.   Approximated and simulated J-divergence as a function of P_{tot} for the proposed power allocation of case V-C4.
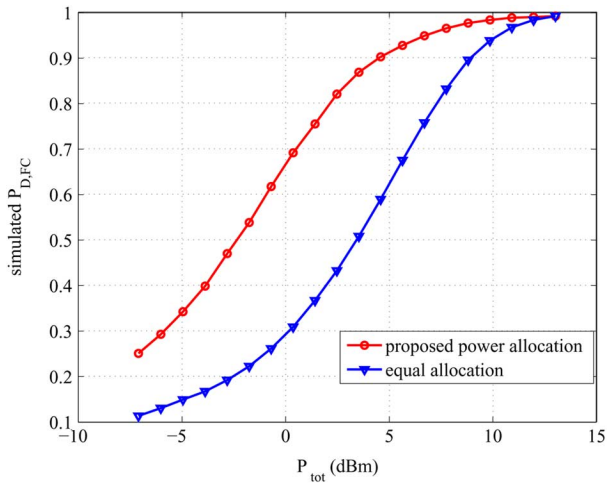

Fig. 30.   Simulated $P_{D,\text{FC}}$ as a function of P_{tot} for the proposed power allocation and equal power allocation of case V-C4.

TABLE V
PERCENTAGE OF TOTAL POWER ALLOCATED TO EACH SENSOR FOR CASE V-C5

P_{tot} (dBm)	1	2	3	4	5	6	7	8	9	10
-7.0	10	10	10	10	10	10	10	10	10	10
-2.8	10	10	10	10	10	10	10	10	10	10
3.5	10	10	10	10	10	10	10	10	10	10
8.8	10	10	10	10	10	10	10	10	10	10
13	10	10	10	10	10	10	10	10	10	10

J-divergence; and 5) the proposed power allocation quantifies the tradeoff between the quality of the local decisions of the sensors and the quality of the communication channels between the sensors and the FC. Simulations show that, to achieve the same detection performance at the FC, the proposed power allocation can use as little as 25% of the total power used by equal power allocation.
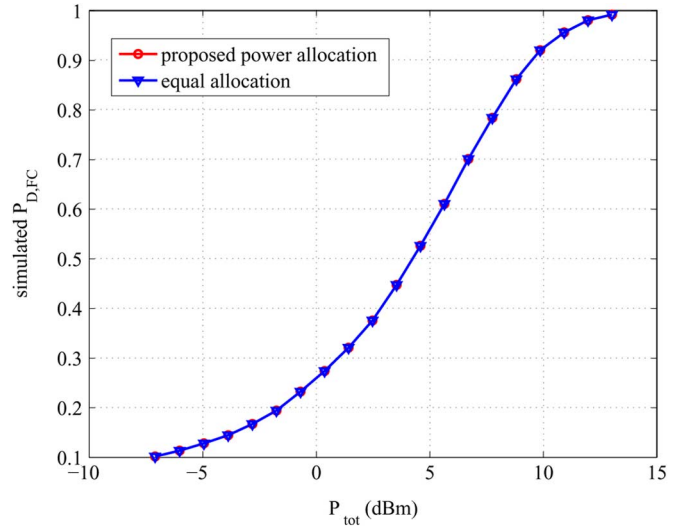

Fig. 31.   Simulated $P_{D,\text{FC}}$ as a function of P_{tot} for the proposed power allocation and equal power allocation of case V-C5. (Note that the curves for ∘ and ▽ overlay in this graph.).

APPENDIX I
PROOF OF LEMMA 2

Proof: Taking the derivative of $J(P_1,\ldots,P_K)$ in (36) with respect to P_j, we have

$$\frac{\partial}{\partial P_j} J(P_1,\ldots,P_K)\big|_{\{P_i\}_{i=1}^K, P_i \geq 0}$$
$$= \frac{(\alpha_F(j) - \beta_F(j))\sigma^2 g_j}{(\sigma^2 + \beta_F(j) g_j P_j)^2} + \frac{(\alpha_D(j) - \beta_D(j))\sigma^2 g_j}{(\sigma^2 + \beta_D(j) g_j P_j)^2}$$
$$= c[(\alpha_F(j) - \beta_F(j))(\sigma^2 + \beta_D(j) g_j P_j)^2$$
$$+ (\alpha_D(j) - \beta_D(j))(\sigma^2 + \beta_F(j) g_j P_j)^2]$$
$$= c\left[d_0\sigma^4 + 2d_1\sigma^2 g_j P_j + d_2 g_j^2 P_j^2\right] \quad (62)$$

where

$$c = \frac{\sigma^2 g_j}{(\sigma^2 + \beta_F(j) g_j P_j)^2 (\sigma^2 + \beta_D(j) g_j P_j)^2} \quad (63)$$
$$d_0 = \alpha_F(j) + \alpha_D(j) - \beta_F(j) - \beta_D(j) \quad (64)$$
$$d_1 = \alpha_F(j)\beta_D(j) + \alpha_D(j)\beta_F(j) - 2\beta_F(j)\beta_D(j) \quad (65)$$

and

$$d_2 = \alpha_F(j)\beta_D(j)^2 + \alpha_D(j)\beta_F(j)^2$$
$$- \beta_F(j)^2\beta_D(j) - \beta_F(j)\beta_D(j)^2. \quad (66)$$

Substituting $\alpha_F(j), \beta_F(j), \alpha_D(j), \beta_D(j)$ into (37)–(40), and after lengthy algebra, we have

$$d_0 = 2(P_D(j) - P_F(j))^2 \geq 0 \quad (67)$$
$$d_1 = (P_D(j) - P_F(j))^2[P_D(j)P_F(j) \quad (68)$$
$$+ (1 - P_D(j))(1 - P_F(j))] \geq 0$$

and

$$d_2 = (P_D(j) - P_F(j))^2[(P_D(j) + P_F(j))$$
$$\times (P_D(j) + P_F(j) - 1 - P_D(j)P_F(j))^2$$
$$+ P_D(j)^2 P_F(j)^2(2 - P_D(j) - P_F(j))] \geq 0. \quad (69)$$

Since c and P_j are also nonnegative, we conclude that

$$\frac{\partial}{\partial P_j} J(P_1, \ldots, P_K)\big|_{\{P_i\}_{i=1}^K, P_i \geq 0} \geq 0. \tag{70}$$

∎

APPENDIX II
PROOF OF LEMMA 3

Proof: From (43), we can easily see that $((\partial^2)/(\partial P_j^2)) J(P_1, \ldots, P_K) \leq 0$ if and only if $C_i \geq 0$, $i = 0, \ldots, 3$. From the Proof of Lemma 2, we know that $C_i \geq 0$, $i = 1, 2, 3$, so $((\partial^2)/(\partial P_j^2)) J(P_1, \ldots, P_K) \leq 0$ if and only if $C_0 \geq 0$.

We have

$$C_0 = \beta_F(j)[\alpha_F(j) - \beta_F(j)] + \beta_D(j)[\alpha_D(j) - \beta_D(j)]. \tag{71}$$

Substituting $\alpha_F(j)$, $\beta_F(j)$, $\alpha_F(j)$, and $\beta_F(j)$ into (37)–(40) and after some algebra, we have

$$C_0 = (P_D(j) - P_F(j))^2 [-2P_D(j)^2 + (3 - 2P_F(j))P_D(j) + 3P_F(j) - 2P_F(j)^2 - 1]. \tag{72}$$

Now, $C_0 \geq 0$ if and only if

$$-2P_D(j)^2 + (3 - 2P_F(j))P_D(j) + 3P_F(j) - 2P_F(j)^2 - 1 \geq 0. \tag{73}$$

Solving the above quadratic inequality leads to Lemma 3. ∎

REFERENCES

[1] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Stat. Soc.*, ser. B, vol. 28, pp. 131–142, 1966.

[2] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Boston, MA: Athena Scientific, 1999.

[3] 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.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Wiley, 1991.

[5] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3454–3458, Dec. 2004.

[6] 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.

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

[8] 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.

[9] 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.

[10] J. F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 16–25, May 2007.

[11] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kaidó, 1981.

[13] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 2, pp. 469–476, Apr. 1998.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns The Johns Hopkins Univ. Press, 1996.

[15] S. K. Jayaweera, "Bayesian fusion performance and system optimization for distributed stochastic Gaussian signal detection under communication constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1238–1250, Apr. 2007.

[16] H. Jeffreys, "An invariant form for the prior probability in estimation problems," in *Proc. Roy. Soc. A.*, 1946, vol. 186, pp. 453–461.

[17] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 2, pp. 52–60, Feb. 1967.

[18] H. Kobayashi and J. B. Thomas, "Distance measures and related criteria," in *Proc. 5th Annu. Allerton Conf. Circuit System Theory*, Oct. 1967, pp. 491–500.

[19] H. Kobayashi, "Distance measures and asymptotic relative efficiency," *IEEE Trans. Inf. Theory*, vol. 16, no. 3, pp. 288–291, May 1970.

[20] K. Liu and A. M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1899–1910, May 2007.

[21] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1081–1092, Mar. 2007.

[22] P. J. Moreno, P. P. Ho, and N. Vasconcelos, "A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications," *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 1385–1393, 2003.

[23] A. J. Motley and J. P. Keenan, "Personal communication radio coverage in buildings at 900 MHz and 1700 MHz," *Electron. Lett.*, vol. 24, pp. 763–764, 1988.

[24] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1018–1027, Mar. 2006.

[25] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 1994.

[26] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 893–900, Sep. 1977.

[27] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Consistency in models for distributed learning under communication constraints," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 52–63, Jan. 2006.

[28] 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.

[29] Y. Singer and M. K. Warmuth, "Batch and on-line parameter estimation of Gaussian mixtures based on the joint entropy," *Adv. Neural Inf. Process. Syst.*, vol. 11, pp. 578–584, 1998.

[30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[31] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Proccessing, Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993, vol. 2, pp. 297–344.

[32] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1996.

[33] 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.

[34] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[35] X. Wang and H. V. Poor, *Wireless Communication Systems: Advanced Techniques for Signal Reception*. Upper Saddle River, NJ: Prentice-Hall, 2004.

Xin Zhang (S'01–M'06) received the B.S. and M.S. degree in electrical engineering from Fudan University, Shanghai, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, in 2005.

From 2005 to 2006, he was a Postdoctoral Research Associate at Princeton University, Princeton, NJ. Currently, he is with the United Technologies Research Center, East Hartford, CT. His research interests include wireless sensor networks, statistical signal processing, detection, and target tracking.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1977.

From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. His research interests are in the areas of stochastic analysis, statistical signal processing, and their applications in wireless networks and related fields. Among his publications in these areas are the recent book *MIMO Wireless Communications* (Cambridge University Press, 2007), and the forthcoming book *Quickest Detection* (Cambridge University Press, 2009).

Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and a former Guggenheim Fellow. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and in 2004–2007 he served as the Editor-in-Chief of the IEEE Transactions on Information Theory. Recent recognition of his work includes the 2005 IEEE Education Medal, the 2007 IEEE Marconi Prize Paper Award, and the 2007 Technical Achievement Award of the IEEE Signal Processing Society.

Mung Chiang (S'00–M'03) received the B.S. (Hons.) degree in electrical engineering and mathematics and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively.

He is an Assistant Professor of Electrical Engineering and an affiliated faculty member of the Applied and Computational Mathematics and of Computer Science at Princeton University, Princeton, NJ. He conducts research in the areas of nonlinear optimization of communication systems, theoretical foundation of network architectures, algorithms for broadband access networks, and stochastic analysis of communications and networking. He is a co-editor of the new Springer book series on Optimization and Control of Communication Systems.

Dr. Chiang received the CAREER Award from the National Science Foundation, the Young Investigator Award from the Office of Naval Research, the Howard B. Wentz Junior Faculty Award, and the Engineering Teaching Commendation from Princeton University, the School of Engineering Terman Award from Stanford University, and the New Technology Introduction Award from SBC Communications, and was a Hertz Foundation Fellow and Stanford Graduate Fellow. For his work on broadband access networks and Internet traffic engineering, he was selected for the TR35 Young Technologist Award in 2007, a list of top 35 innovators in the world under age 35. His work on Geometric Programming was selected by Mathematical Programming Society as one of the top three papers by young authors in the area of continuous optimization during 2004–2007. His work on Layering As Optimization Decomposition became a Fast Breaking Paper in Computer Science by ISI citation. He also coauthored papers that were an IEEE INFOCOM best paper finalist and IEEE GLOBECOM best student paper. He has served as an Associate Editor for the IEEE Transactions on Wireless Communications, a lead Guest Editor for the IEEE Journal of Selected Areas in Communications, a Guest Editor for the IEEE/ACM Transactions on Networking and the IEEE Transactions on Information Theory, a Program Co-Chair of the Thirty-Eighth Conference on Information Sciences and Systems.