Power Control Strategy for Distributed Multiple-Hypothesis Detection

Hyoung-soo Kim, *Student Member, IEEE*, and Nathan A. Goodman, *Senior Member, IEEE*

Abstract—We introduce a local-channel power control strategy applicable to multiple-hypothesis distributed detection systems communicating over slow-fading orthogonal multiple-access channels. In earlier work, it was demonstrated that performance could be improved by adjusting transmit power to maximize the J-divergence measure of a binary detection system. The local power control strategy introduced here further improves performance by exploiting *a priori* probabilities and local sensor statistics. Moreover, the local power optimization can be combined with additional power control based on the state of the propagation channel. We extend the optimization to systems performing multiple-hypothesis detection, and evaluate outage probability for these systems. Various numerical results are shown.

Index Terms—Classification outage, detection outage, diversity, local-channel power control, optimal power allocation, outage probability, sensor network.

I. INTRODUCTION

STRATEGIES to improve the lifetime of battery-powered wireless sensor networks (WSNs) have been an intensively studied topic due to the difficulty of replacing batteries in geographically deployed sensors. In this paper, we consider distributed detection systems where a network of sensors each observe the event status of a source, make their own local decision, and forward the decision to a fusion center through a nonideal fading channel. Based on the local decisions, the fusion center makes a final decision according to a fusion rule. However, the focus here is not on the local decision or the fusion rules, but on strategies for conserving the power used by wireless sensor networks to communicate and make decisions.

Research on local decision and fusion rules for distributed sensor networks can be traced back to the early 1980s [1]–[3] and are still in process [4]–[7], but studies of optimal power control strategies are only recently being explored [8]–[11]. Thomopoulos [12] *et al.* showed that the local likelihood ratio test (LRT) is optimal in the presence of a fusion center under the Neyman–Pearson criterion where the decision scheme maximizes the probability of detection for fixed probability of false alarm. These decision rules were then evaluated and improved in the presence of network delay and channel errors in [13]. On

the other hand, Hoballah and Varshney investigated the optimal local rule under the Bayesian criterion in [3], and the rule was expanded to consider the presence of nonideal channels in [6].

Sadjadi developed local decision logic and fusion rules for generalizing the distributed detection problem of Tenney to distributed multiple-hypothesis detection [14]. He obtained optimal detectors by minimizing an average cost function and finding the optimal decision regions. Chao showed optimum partitionings of decision regions based on multibit decisions to avoid the information loss by local hardlimiting process in [15]. In [16] and [17], Wang studied local sensor decision logic and sensor fusion with a double-bound testing method for both serial distributed sensor networks and parallel decision networks. Oh introduced a multiple-target tracking algorithm with a dynamic model of multiple targets for sensor networks [18]. However, power optimization strategies unique to multiple-hypothesis scenarios have not been explored.

Recently, the emerging optimal power allocation issue has been considered in the context of estimation of an unknown parameter or detection of an unknown source. Xiao *et al.* [9] introduced optimal power scheduling for the joint estimation of a Gaussian source in an inhomogeneous Gaussian sensor network by minimizing total power consumption while satisfying a mean-squared distortion constraint. Zhang [10] introduced an optimal power allocation scheme over a multiple-access channel by maximizing J-divergence under a fixed total communication power constraint on the sensors of a distributed binary detection system.

In this paper, we introduce *local power control*, which is performed by minimizing the average power of the local $\mathcal{D}$-dimensional transmit symbol constellation for distributed multiple-hypothesis detection. Then *channel power control* is performed by maximizing the total J-divergence, which is an instantaneous performance measure based on the instantaneous channel realization. The J-divergence optimization results in optimized amplifying factors prior to transmission over the channel. We analyze the proposed *local-channel power control scheme* using detection outage probability(or classification outage), which is a long-term system performance measure averaged over many realizations of the communication channel. The outage metric is similar to communication outage probability, but applies specifically to the distributed multiple-hypothesis detection.

This paper is organized as follows. In the next section, the structure of the distributed detection system is described. In Section III, the local power optimization is explained and the total J-divergence of the overall system which is used for channel power optimization is derived. In Section IV, after briefly explaining an uniform power transmission scheme with
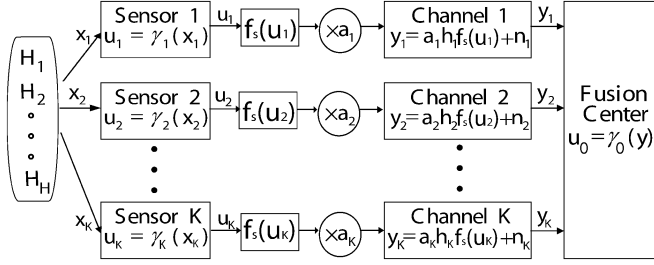
Fig. 1.　Distributed detection with a fusion center.

local power optimization, we derive the asymptotic total J-divergence with which we generalize detection outage probability for the multiple-hypothesis problem. The relationship between asymptotic total J-divergence and multiple-hypothesis detection outage is also demonstrated. Using the extended outage probability, we analyze the power gain and detection diversity achieved through local power control. In Section V, channel power optimization (symbol amplification) is combined with the local power optimization procedure of Section IV. Detection outage performance is evaluated under 1) a total power constraint and 2) total and individual power constraints. Section VI summarizes the results and gives some concluding remarks.

II. SYSTEM MODEL

We introduce a similar system setup as that used in [10], [11], and [19], but the system model is extended for multiple-hypothesis detection. At some observation time, detection of the event source (e.g., detection of multiple targets or target classification) can be abstracted as multiple hypotheses: $H_1, H_2, \ldots, H_{\mathcal{H}}$. There are $\mathcal{K}$ sensors that make local observations about the source and transmit local decisions to the fusion center. For simplicity, we assume that the local decisions are transmitted to the fusion center over orthogonal multiple-access channels as shown in Fig. 1. Specifically, sensor k collects a local observation x_k that has been corrupted by observation noise.

Each sensor then makes its own local decision u_k according to a local decision rule denoted symbolically as

$$u_k = \gamma_k(x_k) \in \{1, 2, \ldots, \mathcal{H}\}$$

Node k then maps its local decision u_k to a transmit symbol according to some modulation scheme, $f_s(u_k)$, amplifies the transmit symbol by a_k, and transmits the signal over the kth fading channel. In general, the modulation scheme $f_s(u_k)$ can be scalar (e.g., amplitude modulation) or multidimensional [e.g., phase shift keying (PSK) or frequency shift keying (FSK)]. The fading communication channel is modeled as a set of random amplitude gains h_k, and the signal received at the fusion center is corrupted by additive white Gaussian noise (AWGN). We assume that the h_k's and n_k's are both independent over k. Later in the paper we use detection outage probability [19], [20] to quantify long-term system performance over multiple independent and identically distributed (i.i.d.) realizations of the channel gain coefficients.

From now on, we use a bold capital letter for a matrix and a bold lower-case letter for a vector. $\mathcal{K}$ is the total number of channels. $\mathcal{H}$ is the total number of hypotheses, and $\mathcal{D}$ is the total number of dimensions used for the local modulation scheme.

We can characterize an individual sensor by its transition probabilities

$$P(u_k = l | H_i), \qquad \text{for } l \text{ and } i = 1, 2, \ldots, \mathcal{H}.$$

We further assume that the local observation x_k's and the local decision u_k's are independent over k when conditioned on a particular hypothesis [11], [19]. As such, the joint conditional probability mass function of the local decisions is

$$P(\mathbf{u} | H_i) = \prod_{k=1}^{\mathcal{K}} P(u_k | H_i) \tag{1}$$

where the local decision vector is $\mathbf{u} = [u_1, \ldots, u_{\mathcal{K}}]^T$.

We now allow each sensor to adaptively control its transmit symbol constellation according to its own unique decision statistics. Let $\mathbf{f_s}(\cdot)$ be the optimization of the transmit symbol constellation. The optimized modulation symbols are transmitted to the fusion center over a fading channel. At the fusion center, the received signals are

$$\mathbf{y} = \mathbf{H} \mathbf{A} \mathbf{f_s}(\mathbf{u}) + \mathbf{n} \tag{2}$$

where

$$\mathbf{H} = \text{diag}\{h_1, h_2, \ldots, h_{\mathcal{K}}\}$$
$$\mathbf{A} = \text{diag}\{a_1, a_2, \ldots, a_{\mathcal{K}}\}$$
$$\mathbf{f_s}(\mathbf{u}) = [f_s(u_1), f_s(u_2), \ldots, f_s(u_{\mathcal{K}})]^T$$
$$\mathbf{y} = [y_1, y_2, \ldots, y_{\mathcal{K}}]^T$$

and where $\mathbf{n} = \{n_1, n_2, \ldots, n_{\mathcal{K}}\}^T$ is the additive noise vector. The conditional probability density function (pdf) of the received signal given hypothesis H_i is [10]

$$p(\mathbf{y} | H_i) = \sum_{\mathbf{u}} p(\mathbf{y} | \mathbf{u}) P(\mathbf{u} | H_i). \tag{3}$$

The final decision at the fusion center, u_0, is determined by a fusion rule denoted symbolically by $u_0 = \gamma_0(\mathbf{y})$.

In general, $\{P(H_i) | i = 1, 2, \ldots, \mathcal{H}\}$ may be unknown but can be estimated with a given local transition matrix and local decision probabilities. For example, a vector of *a priori* source probabilities could be estimated according to $\widetilde{P}_k(\boldsymbol{H}) = \mathbf{L}_k^{-1} \cdot \widetilde{P}(u_k)$ where

$$\mathbf{L}_k = \begin{pmatrix} P(u_k = 1|H_1) & \cdots & P(u_k = 1|H_{\mathcal{H}}) \\ P(u_k = 2|H_1) & \cdots & P(u_k = 2|H_{\mathcal{H}}) \\ \vdots & \ddots & \vdots \\ P(u_k = \mathcal{H}|H_1) & \cdots & P(u_k = \mathcal{H}|H_{\mathcal{H}}) \end{pmatrix}$$

is given. $\widetilde{P}(u_k)$ is a locally estimated vector of decision probabilities obtained by averaging in time over many independent decisions. For example, suppose the kth sensor operates for some period of time and makes 100 decisions. If the sensor chooses the second hypothesis ($u_k = 2$) three times, then $\widetilde{P}(u_k = 2) = 3/100$. However, this approach yields estimates of the decision probabilities, not of the source prior probabilities. The two are related by $\widetilde{P}_k(\boldsymbol{H}) = \mathbf{L}_k^{-1} \cdot \widetilde{P}(u_k)$, but unfortunately this can lead to negative prior probabilities

because they are calculated from estimated $\widetilde{P}(u_k)$ rather than the true local decision probabilities. Therefore, we have instead performed a simple nonlinear optimization

$$\min \quad \left\| \widetilde{P}_k(\boldsymbol{H}) - \mathbf{L}_k^{-1} \cdot \widetilde{P}(u_k) \right\|^2$$
$$\text{s.t.} \quad \sum_{i=1}^{\mathcal{H}} \widetilde{P}_k(H_i) = 1, \qquad \widetilde{P}_k(H_i) \geq 0 \text{ for all } i.$$

The final estimate $\widetilde{P}(H_i)$ is calculated by averaging over the $\mathcal{K}$ sensor estimates $(\widetilde{P}_k(H_i))$ according to

$$\widetilde{P}(H_i) = \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \widetilde{P}_k(H_i).$$

III. POWER CONTROL STRATEGY

A. Local Power Control

In this paper, we adopt two power optimization strategies—local power control and channel power control—to take advantage of local sensor statistics and channel statistics. As for the local power control, we allow each sensor to adaptively control its transmit symbol constellation according to its own unique decision statistics. Let $\{\delta_k(l) = f_s(u_k = l) | l = 1, 2, \cdots, \mathcal{H}\}$ be the local modulation symbols of the kth sensor where the modulation function $f_s(\cdot)$ adjusts the symbols based on local decision probabilities, $\{P(u_k = h) | h = 1, 2, \cdots, \mathcal{H}\}$. The local power control minimizes the average power of the modulation constellation. Let this optimization for the kth sensor be stated as

$$\min_{\delta_k(1), \delta_k(2), \cdots, \delta_k(\mathcal{H})} PW_k = \sum_{l=1}^{\mathcal{H}} |\delta_k(l)|^2 \cdot P(u_k = l)$$
$$\text{s.t.} \qquad F(\cdot) = \text{const} \qquad (4)$$

where the constraint $F(\cdot)$ can be any communication performance measure between the kth sensor and the fusion center. The design variables are $\delta_k(1), \delta_k(2), \cdots, \delta_k(\mathcal{H})$. The basic goal of the local power control is to find an optimal local constellation configuration that minimizes the average constellation power while keeping a given communication performance. We could let the constellation configuration be completely arbitrary, but this approach would likely lead to impractical implementations and increases the optimization procedure considerably. Instead, we essentially implement $F(\cdot)$ as a *structural constraint* by assuming a symbol constellation in the form of some traditional modulation scheme (QAM, FSK, etc.) with spacing determined by desired communication performance. Once the constellation structure and size are set, we can then optimize the average power of the constellation through linear translations of the constellation and by judicious assignment of decisions to symbols in the constellation. Note that communication performance is determined by the constellation structure, not by its translational shifts. We later show an example using a translated QAM constellation including a fair comparison to the performance of a system with a nontranslated QAM modulation transmission. This example is shown in Sections IV-D and V. We also show an example using scalar modulation in Section IV-C. Optimization of the transmit symbol constellation as defined by (4) is referred to as *local power control*, while optimization of the average power transmitted by each sensor based on channel states is referred to as *channel power control*.

B. System Performance Measure, Total J-Divergence and Channel Power Control

Rather than considering optimal fusion rules, we instead argue that detection performance is generally improved if the total J-divergence J^{total} is maximized. In particular, the total J-divergence measure is defined by a weighted sum of pairwise J-divergences, $J^{ij} = J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j))$, according to [21], [22]

$$J^{\text{total}} = \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) \cdot P(H_i) \cdot P(H_j)$$
$$= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) \cdot P(H_i) \cdot P(H_j). \quad (5)$$

The second line of (5) is possible because J^{ij} is symmetric and $J^{ij} = 0$ when $i = j$. The pairwise J-divergences are $J^{ij} = (1/2)[D(p_j \| p_i) + D(p_i \| p_j)]$ where $D(p_i \| p_j)$ is the Kullback–Leibler (KL) divergence measure between two probability density functions [23]. Thus, the J-divergence is a symmetric version of the more general KL distance measure.

The KL distance is the average of the difference between two log-likelihood functions. Let $\log(p_1)$ and $\log(p_2)$, respectively, be the two conditional log-likelihood functions for hypotheses H_1 and H_2. The KL distance is defined as $E_1[\log(p_1) - \log(p_2)] = E_1[log(p_1/p_2)]$ where $E_1[\cdot]$ is the expected value with respect to p_1. From this definition, the KL distance is interpreted as the average of the log-likelihood ratio between two conditional pdfs p_1 and p_2. Because the likelihood ratio is an optimal detection method that appears in both Neyman–Pearson and Bayesian detection, we can conclude that J-divergence is closely related with detection performance. In fact, for the binary Gaussian detection problem, J-divergence becomes the signal-to-noise (SNR) at the receiver and the probability of error is $Q(\sqrt{J(\cdot)/2})$ where $Q(\cdot)$ is the Gaussian Q-function and $J(\cdot)$ is the J-divergence between the two hypotheses. Asymptotically, J-divergence determines the error exponent of the Chernoff bound from Stein's lemma [11], [23].

For distributed detection systems, the simulation results of [11] show that probability of detection can be enhanced by increasing J-divergence. Even though J-divergence is not a direct performance measure like probability of detection or probability of error, it usually leads to tractable analytical frameworks for distributed detection systems with Gaussian assumptions and has been adopted by many researchers, such as in [24]–[27]. J-divergence also provides a lower bound to the probability of error by $P_e > P(H_1)P(H_2)e^{-J/2}$ [24] in a binary detection system.

We apply J-divergence to the multiple-hypothesis detection system defined in (5) because of its relationship to detection performance and its ability to provide a tractable analysis. For example, J^{total} provides a lower bound for the multiple-hypothesis detection problem by $P_e^{bound} \geq C - J^{\text{total}}$ where P_e^{bound}

is defined as a pairwise sum of the individual probabilities of error between two hypotheses ($P_e^{bound} = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P_e^{ij}$) and C is a constant ($C = \sum_{i=1}^{\mathcal{H}} \sum_{j=1}^{\mathcal{H}} P(H_i) p(H_j)$). A detailed proof is given in Appendix VII-A. We can reduce this bound by increasing the total J-divergence. Therefore, we can design a power control strategy by maximizing the total J-divergence according to

$$
\begin{aligned}
\max_{\mathbf{A}} \quad & J^{\text{total}}(\cdot) \\
\text{s.t.} \quad & Tr[\mathbf{A}\mathbf{A}^T] \le P_{\text{tot}}, \qquad 0 \preceq \mathbf{A} \preceq \sqrt{\mathbf{P}_{\max}}
\end{aligned} \quad (6)
$$

where $\mathbf{A}$ is the $\mathcal{K} \times \mathcal{K}$ diagonal amplification matrix defined earlier, P_{tot} is a total power constraint, and $\mathbf{P}_{\max}$ is a $\mathcal{K} \times \mathcal{K}$ diagonal matrix of individual power constraints. $\sqrt{\mathbf{P}_{\max}}$ denotes the component-wise square root of $\mathbf{P}_{\max}$, and the inequality $\preceq$ means that $\sqrt{\mathbf{P}_{\max}} - \mathbf{A}$ is positive semidefinite [11]. The individual diagonal terms of $\sqrt{\mathbf{P}_{\max}}$ need not be the same.

Note that the pairwise J-divergence J^{ij} directly depends only on the conditional probability densities, $p(\mathbf{y}|H_i)$ and $p(\mathbf{y}|H_j)$, but apparently not at all on the source's prior probabilities. Hence, it is not immediately obvious how a local optimization procedure can have any effect on system performance. Closer inspection of (2) and (3), however, shows that J^{ij} does depend on the transmit symbol constellations used by the various sensors. Since the symbol constellations are the design variables in the local power optimization, J^{ij} does indeed depend on *a priori* probabilities of the source. Essentially, the local optimization reduces the average power of the symbol constellation for each sensor, which we will see later allows the amplification coefficients a_k to be increased while still meeting average power constraints on the total transmission power. As discussed in [11], with AWGN channel noise, the system conditional probabilities can be approximated by Gaussian densities and the resulting individual J-divergence J^{ij} is given as

$$
\begin{aligned}
J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) = & \frac{1}{2} Tr\left[\left[\mathbf{R} + \mathbf{H}\mathbf{A}(\mathbf{C}_i + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{H}^T \right] \right. \\
& \left. \cdot [\mathbf{R} + \mathbf{H}\mathbf{A}\mathbf{C}_j\mathbf{A}^T\mathbf{H}^T]^{-1} \right] \\
& + \frac{1}{2} Tr\left[\left[\mathbf{R} + \mathbf{H}\mathbf{A}(\mathbf{C}_j + \boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{H}^T \right] \right. \\
& \left. \cdot [\mathbf{R} + \mathbf{H}\mathbf{A}\mathbf{C}_i\mathbf{A}^T\mathbf{H}^T]^{-1} \right] - \mathcal{K}
\end{aligned} \quad (7)
$$

where $\mathbf{C}_i$ is a $\mathcal{K} \times \mathcal{K}$ diagonal matrix with elements

$$
\begin{aligned}
C_i(k, k) = & \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \{\delta_k(l) - \delta_k(m)\}^2 P(u_k = l|H_i) \\
& \cdot P(u_k = m|H_i), \qquad \text{for } di \in \{1, 2, \cdots, \mathcal{H}\}
\end{aligned}
$$

and $\boldsymbol{\xi} = \boldsymbol{\xi}_0 - \boldsymbol{\xi}_1$, and

$$
\boldsymbol{\xi}_i = \begin{bmatrix} \sum_{l=1}^{\mathcal{H}} \delta_1(l) P(u_1 = l|H_i) \\ \sum_{l=1}^{\mathcal{H}} \delta_2(l) P(u_2 = l|H_i) \\ \vdots \\ \sum_{l=1}^{\mathcal{H}} \delta_{\mathcal{K}}(l) P(u_{\mathcal{K}} = l|H_i) \end{bmatrix}.
$$

$\mathbf{R}$ is the covariance matrix of the channel noise vector $\mathbf{n}$, and $\mathcal{K}$ is the dimension of the received signal vector $\mathbf{y}$ at the fusion center. More details on the derivation of (7) are given in Appendix VII-B. Finally, the total J-divergence of the system is obtained by applying (7) to (5). In the next section, we consider the performance benefit of local optimization when average transmit power is equal across sensors.

IV. DETECTION OUTAGE OF A UNIFORM TRANSMIT STRATEGY WITH LOCAL POWER OPTIMIZATION

In this section, we discuss a scheme where local power optimization and uniform channel power control cooperate such that the average transmit power P_k of each sensor is equal, and analytically derive detection outage probability through asymptotic total J-divergence for distributed multiple-hypothesis detection. We use the detection outage to evaluate the diversity gain of uniform transmit strategy in a multiple-hypothesis system.

A. Uniform Transmit Strategy

First, define the average power transmitted by the kth sensor as $P_k = PW_k^* \cdot a_k^2$ where $PW_k^* = \sum_{l=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l)$ is the minimized constellation power obtained through local optimization. In other words, the asterisk denotes that the constellation has been selected to minimize the average power in the symbol constellation. The total transmit power constraint for all sensors is $P_{\text{tot}} = \sum_{k=1}^{\mathcal{K}} P_k$, which for uniform transmit power requires that $P_k = P_{\text{tot}}/\mathcal{K}$. Substituting for PW_k^*, the amplifying factor for the kth sensor is $a_k = \sqrt{P_{\text{tot}}/(\mathcal{K} \cdot \sum_{l=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l))}$. Note that the amplification factor is inversely related to the preamplification average power of the symbol constellation. Thus, the local optimization step has allowed larger a_k while still meeting the total power constraint. The dimension of the received signal vector $\mathbf{y}$ at the fusion center is the same as the number of sensors $\mathcal{K}$, since the system is modeled as having orthogonal communication channels. We set $\mathbf{R} = \text{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_{\mathcal{K}}^2\}$ where σ_k^2 is the noise power of the nth channel. From (5) and (7), the total J-divergence is given in (8), shown at the bottom of the page, where

$$
\begin{aligned}
\psi_{ik} = & \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \left[(\delta_k^*(l) - \delta_k^*(m))^2 \right. \\
& \left. \times P_k(u_k = l|H_i) P_k(u_k = m|H_i) \right]
\end{aligned}
$$

$$
J^{\text{total}} = \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[\sum_{k=1}^{\mathcal{K}} \frac{a_k^2 h_k^2 \left\{ \psi_{jk} - \psi_{ik} + \phi_k^{ij} \right\}}{\sigma_k^2 + a_k^2 h_k^2 \psi_{ik}} + \sum_{k=1}^{\mathcal{K}} \frac{a_k^2 h_k^2 \left\{ \psi_{ik} - \psi_{jk} + \phi_k^{ij} \right\}}{\sigma_k^2 + a_k^2 h_k^2 \psi_{jk}} \right] P(H_i) \cdot P(H_j) \quad (8)
$$

$$\phi_k^{ij} = \sum_{l=1}^{\mathcal{H}} [\delta_k^*(l) \cdot (P_k(u_k = l|H_i) - P_k(u_k = l|H_j))]^2$$

where i and j are hypothesis indices, k is the index of an individual sensor, and l, m are the indexes of transmission symbols. Applying the uniform per-sensor power constraint and defining the kth channel's SNR as $s_k = h_k^2/\sigma_k^2$, the total J-divergence becomes (9), shown at the bottom of the page. In the following, we will analytically derive the detection outage with (8), show a simulation result by the uniform transmit strategy, and then compare the analytical diversity gain of the detection outage probability with the gain obtained through the simulation result.

B. Asymptotic Total J-Divergence J_∞ and Detection Outage

Intuitively, J-divergence can be increased by increasing the number of sensors. If a total power constraint is enforced, however, there is an asymptotic limit to the increase since a finite amount of power must be distributed among more and more sensors. In this section, we derive the asymptotic total J-divergence expression and show its relationship to detection outage probability in a homogeneous sensor environment. We begin by defining the two terms that constitute each individual J-divergence, $J^{ij} = J(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)) = J_1^{ij} + J_2^{ij}$ in (5) and (8) as $J_1^{ij} = (1/2)\sum_{k=1}^{\mathcal{K}}(a_k^2 h_k^2\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\}/(\sigma_k^2 + a_k^2 h_k^2\psi_{ik}))$ and $J_2^{ij} = (1/2)\sum_{k=1}^{\mathcal{K}}(a_k^2 h_k^2\{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\}/(\sigma_k^2 + a_k^2 h_k^2\psi_{jk}))$. We use these expressions to derive the upper and lower bounds for J^{ij}. We then show that J^{ij} converges to an asymptotic value by showing that the upper and lower bounds converge to the same value as $\mathcal{K}$ goes to infinity. We first consider J_1^{ij}.

Case 1: For $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \geq 0$, the lower and upper bounds for J_1^{ij} are

$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$
$$- \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K}^2 \cdot PW_k^{*2}}$$
$$\leq 2J_1^{ij}$$
$$\leq \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*} \qquad (10)$$

where P_{tot}, s_k, and ψ_{ik} are positive values. A detailed proof is given in Appendix VII-C. Consider a homogeneous sensor network where ψ_{ik}, ψ_{jk}, and ϕ_k^{ij} are the same for each sensor and the s_k's are i.i.d. Then by the law of large numbers (LLN), as $\mathcal{K} \to \infty$ we obtain

$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$
$$\to P_{\text{tot}} E\left[\frac{s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{PW_k^*}\right]$$
$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$
$$\to P_{\text{tot}}^2 E\left[\frac{s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{PW_k^*}\right]$$

such that

$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \to 0 \qquad (11)$$

because s_k, ψ_{ik}, ϕ_k^{ij} are finite and $PW_k^* \neq 0$. Therefore, for $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \geq 0$, we have

$$\lim_{\mathcal{K} \to \infty} J_1^{ij} = \frac{1}{2} P_{\text{tot}} E\left[\frac{s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{PW_k^*}\right].$$

Case 2: For $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \leq 0$, the lower and upper bound for J_1^{ij} are

$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$
$$\leq 2J_1^{ij}$$
$$\leq \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$
$$- \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K}^2 \cdot PW_k^{*2}} \qquad (12)$$

$$J^{\text{total}} = \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \sum_{h=1}^{\mathcal{H}} |\delta_k^*(l)|^2 P(u_k = l) + P_{\text{tot}} s_k \{\psi_{ik}\}} \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot \sum_{h=1}^{\mathcal{H}} |\delta_k^*(l)|^2 \cdot P(u_k = l) + P_{\text{tot}} s_k \{\psi_{jk}\}} \right]$$
$$\times P(H_i) \cdot P(H_j)$$
$$= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^* + P_{\text{tot}} s_k \{\psi_{ik}\}} \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^* + P_{\text{tot}} s_k \{\psi_{jk}\}} \right] P(H_i) \cdot P(H_j) \qquad (9)$$

where P_{tot}, s_k, ψ_{ik}, and PW_k^* are positive. A detailed proof is given in Appendix VII-D. Similar to Case 1, for $(\psi_{jk} - \psi_{ik} + \phi_k^{ij}) \le 0$, we have

$$\lim_{\mathcal{K} \to \infty} J_1^{ij} = \frac{1}{2} P_{\text{tot}} E \left[\frac{s_k \left\{ \psi_{jk} - \psi_{ik} + \phi_k^{ij} \right\}}{PW_k^*} \right].$$

Hence, the limit is valid for any $(\psi_{jk} - \psi_{ik} + \phi_k^{ij})$, and by similar procedures it can be shown that

$$\lim_{\mathcal{K} \to \infty} J_2^{ij} = \frac{1}{2} P_{\text{tot}} E \left[\frac{s_k \left\{ \psi_{ik} - \psi_{jk} + \phi_k^{ij} \right\}}{PW_k^*} \right].$$

Therefore, by combining the asymptotic expressions for J_1^{ij} and J_2^{ij}, we obtain the following theorems.

Theorem 4.1: In a homogeneous distributed detection system with finite $E[\eta_k^{ij}]$ and nonzero PW_k^*, the large-$\mathcal{K}$ asymptotic pairwise J-divergence between $p(\mathbf{y}|H_i)$ and $p(\mathbf{y}|H_j)$ is

$$J_\infty^{ij} := \lim_{\mathcal{K} \to \infty} J^{ij}(\cdot) = P_{\text{tot}} E \left[\eta_k^{ij} \right]$$

where $\eta_k^{ij} = s_k \phi_k^{ij} / PW_k^*$.

Theorem 4.2: In a homogeneous distributed detection system with finite $E[\eta_k]$ and nonzero PW_k^*, the asymptotic total J-divergence is

$$
\begin{aligned}
J_\infty^{\text{total}} &:= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left\{ \lim_{\mathcal{K} \to \infty} J^{ij}(\cdot) \right\} \cdot P(H_i) \cdot P(H_j) \\
&= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} J_\infty^{ij} \cdot P(H_i) \cdot P(H_j) \\
&= P_{\text{tot}} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} E \left[\eta_k^{ij} \right] \cdot P(H_i) \cdot P(H_j) \\
&= P_{\text{tot}} E \left[\sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j) \right] \\
&= P_{\text{tot}} E[\eta_k]
\end{aligned}
$$

where $\eta_k = \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j)$. That is to say, η_k is linear combination of random variables η_k^{ij} where $P(H_i)$ are fixed values.

The asymptotic total J-divergence depends on the channel statistics, local statistics, and the local symbol constellation since $\eta_k^{ij} = s_k \phi_k^{ij} / PW_k^*$ where s_k is the kth channel's SNR, ϕ_k^{ij} is calculated by local statistics and symbol constellation, and PW_k^* is the minimized constellation power obtained through local optimization. In other words, The asymptotic total J-divergence increases with increased channel SNR and can be optimized by minimizing the power of the transmit constellation. From this result, we can say that the local-channel power control scheme can be used to increase instantaneous J-divergence.

C. Long-Term System Performance Measure

The total J-divergence measure in (8) is an instantaneous performance measure because it depends on an instantaneous realization of the channel gain coefficients. The measure can be used as an objective function for power control schemes. In this section, we now employ detection outage probability as a long-term

system performance measure [19], [20], [28]. This enables us to assess the average performance of our power control strategies over many realizations of the fading communication channel. Detection outage probability is defined as the probability that J^{total} falls below a specific threshold J_0. Mathematically, this is stated as

$$p_{J_0} = \text{Prob}\{J^{\text{total}} < J_0\}.$$

For the proposed distributed detection system, the following theorem holds.

Theorem 4.3: In a homogeneous distributed multiple-hypothesis detection system with a finite $E[\eta_k]$, for $J_0 < J_\infty^{\text{total}}$ and a sufficiently-large $\mathcal{K}$, the outage probability is given as [19], [28]–[30]

$$p_{J_0} \sim \exp\left(-\mathcal{K} I_\eta(c)\right)$$

or

$$-\log p_{J_0} \sim -\mathcal{K} I_\eta(c)$$

where $c = J_0 / P_{\text{tot}}$, $I_\eta(c) = \sup_{t \in R}(tc - \log M_\eta(t))$ with $M_\eta(t)$, the moment generating function of η which is an i.i.d. random variable η_k over k, and $\mathcal{K}$ determines the detection diversity order of the system.

The rate function $I_\eta(c)$ is related to the following expression. For $\{\eta_k : k \in \{1, \ldots, \mathcal{K}\}\}$ that are i.i.d. random variables, $c < E\{\eta_k\}$ when $J_0 < J_\infty^{\text{total}}$. We then have

$$\lim_{\mathcal{K} \to \infty} -\frac{1}{\mathcal{K}} \log \left[\text{Prob} \left\{ \frac{1}{\mathcal{K}} \sum_{k=1}^{\mathcal{K}} \eta_k < c \right\} \right] = I_\eta(c).$$

A detailed proof is shown in [28]–[30]. From an example of a Rayleigh fading channel system, the diversity gain can be analyzed as follows. The rate function, $I_\eta(c)$, is defined to be

$$I_\eta(c) = \sup_{t \in R}(tc - \log M_\eta(t))$$

where

$$
\begin{aligned}
\eta_k &= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \eta_k^{ij} \cdot P(H_i) \cdot P(H_j) \\
&= \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \frac{s_k \phi_k^{ij}}{PW_k^*} \cdot P(H_i) \cdot P(H_j) \\
&= \frac{s_k}{2} \cdot \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \frac{2\phi_k^{ij}}{PW_k^*} \cdot P(H_i) \cdot P(H_j) \\
&= \frac{s_k}{2} \cdot \theta_k.
\end{aligned}
$$

The variable η_k is a weighted sum of random variables η_k^{ij}. However, for given sensor statistics and a given modulation scheme, it becomes an exponential random variable due to the fixed value θ_k. Therefore, η_k is exponentially distributed with the mean $\sigma^2 \theta_k$ where $\sqrt{s_k}$ is i.i.d. Rayleigh-distributed with $p_{\sqrt{s_k}}(t) = (t/\sigma^2)\exp\{-(t^2/2\sigma^2)\}$. Since $I_\eta(c)$ is nonnegative and convex over c, we obtain

$$I_\eta(c) = \sup_{t \in R} \left[tc + \log(1 - \sigma^2 \theta \cdot t) \right] \tag{13}$$

$$= \left(\frac{1}{\sigma^2 \theta} - \frac{1}{c} \right) c + \log \left(\frac{\sigma^2 \theta}{c} \right) \tag{14}$$
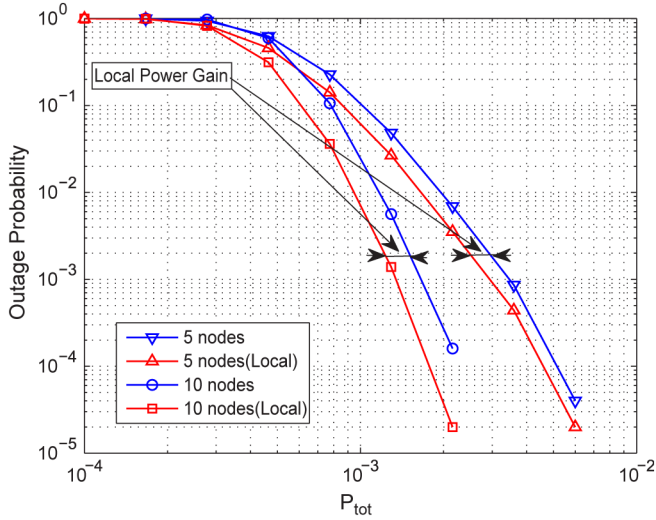
Fig. 2. Outage probability versus P_{tot} with the uniform channel power transmission strategy in heterogeneous sensor configuration. "Local" in parenthesis means that Local power control is performed.
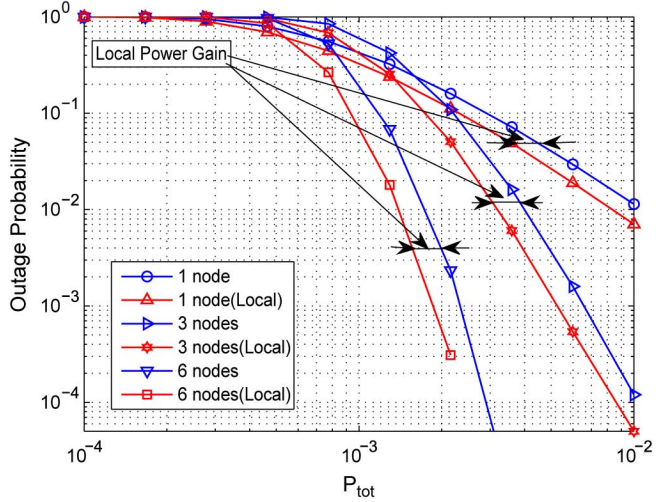

Fig. 3. Outage probability versus P_{tot} with the uniform channel power transmission, 2D-local modulation, and heterogeneous sensor configuration.

Fig. 4. Outage probability versus P_{tot}. local and channel power control strategies are applied in a six-node system.

Fig. 5. Outage probability of an optimal system with individual power constraints and a total power constraint in a nine-node system.

$$= \frac{J_0}{P_{\text{tot}}\sigma^2\theta} - \log\left(\frac{J_0}{P_{\text{tot}}\sigma^2\theta}\right) - 1. \qquad (15)$$

As $P_{\text{tot}} \longrightarrow \infty$, then $(J_0/P_{\text{tot}}\sigma^2\theta) \longrightarrow 0$. Therefore

$$\log p_{J_0} \sim -\mathcal{K} \cdot \log(P_{\text{tot}}). \qquad (16)$$

This result is verified in Figs. 2–5. The proportional gain, $\mathcal{K}$, is the detection diversity order, and if we increase the total power constraint P_{tot}, the reduction in outage probability in log scale is proportional to $\mathcal{K}$.

Fig. 2 shows detection outage probability versus the total power constraint for varying number of sensors in a four-hypothesis distributed detection system. From the figure, we can see the benefit of local power control since channel power control has not been optimized. Local power gain is shown as a left shift of the curve. The figure was generated by simulating 10^5 independent realizations of the discrete source and the fading channel coefficients (or equivalently, the channel SNR values). For each realization, total J-divergence was calculated using
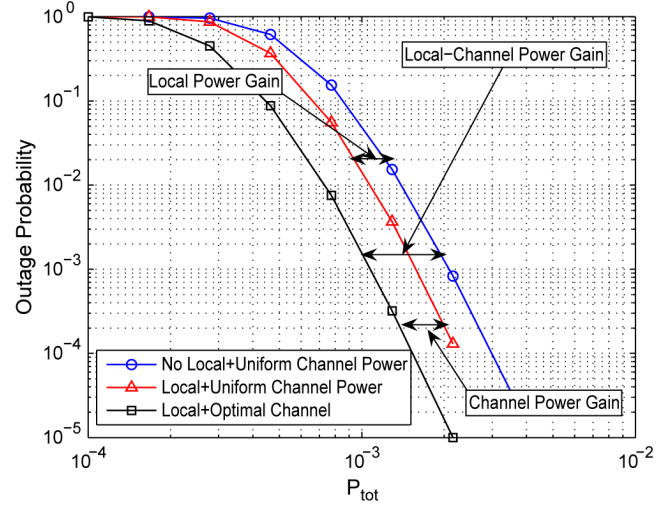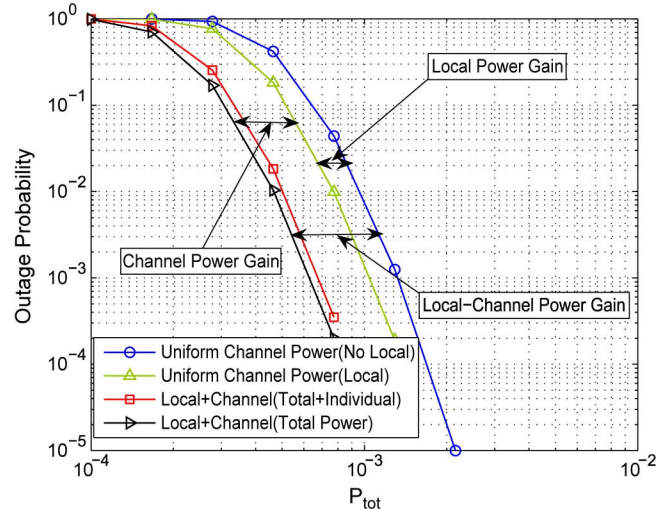
(9). Finally, we counted the number of times the total J-divergence fell below a specified threshold. The following parameters were used to generate the results shown in Fig. 2. The channel SNR was set according to $s_k = (G_0/\sigma_k^2\epsilon_k^2)|r_k|^2$ where ϵ_k is the transmission distance from user k to the fusion center ($\epsilon_k = 100$ m for all k's), $\sigma_k^2 = -70$ dBm is the channel noise power for all k's, $G_0 = -30$ dB is the nominal gain at the unit distance $\epsilon_k = 1$ m, and the $|r_k|$'s are i.i.d. Rayleigh fading random variables with unit variance. We further set ten arbitrary, but different, local transition matrices $\{\mathbf{L}_k | k = 1, 2, \cdots, 9\}$, which implies a heterogeneous sensor environment. The outage threshold was set to $J_0 = 0.05$. In this example, we considered amplitude modulation and four hypotheses. After estimating each of the local decision probabilities $\widetilde{P}(u_k)$ with 100 local decision samples, local power control is performed by minimizing the average constellation power according to

$$\min_{\delta_k(1),\delta_k(2),\cdots,\delta_k(\mathcal{H})} PW_k = \sum_{l=1}^{\mathcal{H}} |\delta_k(l)|^2 \cdot \widetilde{P}(u_k = l). \qquad (17)$$

The symbol constellation was constrained such that the minimum distance among the four symbols was unity. Therefore, we can set $[\delta_k(1), \delta_k(2), \delta_k(3), \delta_k(4)] = [\tau, \tau+1, \tau+2, \tau+3]$ where the design parameter is τ. We also ordered the symbols in descending order of estimated prior probability. This is a logical step, since minimum power should be allocated to the most frequently occurring symbols and vice versa. After substituting the optimal symbol constellation into (9), results for five and ten nodes using only local power control were generated in Fig. 2. It is seen that, at reasonably high P_{tot}, the slope of the ten-node case is two times larger than the slope of the five-node case. In the next subsection, we specifically address local power control for a system with multidimensional local modulation scheme.

D. Total J-Divergence With $\mathcal{D}$-Dimensional Local Modulation

For a more general case, we can use a multidimensional ($\mathcal{D}$-dimensional) local modulation scheme such as QAM, FSK, or any arbitrary modulation. To explicitly handle multidimensional modulation, we modify the observation vector $\mathbf{y}$ to become $\mathcal{K} \times \mathcal{D}$ matrix $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_\mathcal{D}]$ where $\mathcal{K}$ is still the number of sensors and $\mathcal{D}$ is the maximum dimension of transmitted symbols. Thus, a matrix-variate normal distribution for $P(\mathbf{Y}|H_i)$ should be considered, but each column of the observation matrix $\mathbf{Y}$ is independent of every other column under the Gaussian assumption used in (7). This implies $J(p(\mathbf{Y}|H_i), p(\mathbf{Y}|H_j)) = \sum_{d=1}^{\mathcal{D}} J(p(\mathbf{y}_d|H_i), p(\mathbf{y}_d|H_j))$, since $P(\mathbf{Y}|H_i) = P(\mathbf{y}_1|H_i)P(\mathbf{y}_2|H_i)\ldots P(\mathbf{y}_\mathcal{D}|H_i)$ where the vector $\mathbf{y}_d$ is the $\mathcal{K}$-dimensional observation.

Each optimal symbol component $\delta_{(k,d)}^*(l)$ is transmitted with uniform power $P_{\text{tot}}/\mathcal{K} \cdot \mathcal{D}$ since the dimensional components d are also independent $\{P_{(k,d)} = (P_{\text{tot}}/\mathcal{K} \cdot \mathcal{D})|k = 1, 2, \ldots, \mathcal{K}, d = 1, 2, \ldots, \mathcal{D}\}$. l is the local decision index at each sensor. The system works like a system with $\mathcal{K} \cdot \mathcal{D}$ independent components due to independent channels and independent symbol dimensional components and still satisfies the detection diversity $\mathcal{K}$ for the given local modulation scheme ($\mathcal{D}$ is a fixed value). For a given local power constraint, we can get better local power gain than that of the scalar modulation because the multidimensional modulation provides more geometrical distances between symbols. Recently, software-defined radio furnishes flexible modulation schemes that give additional source power gain. Finally, from $J(p(\mathbf{Y}|H_i), p(\mathbf{Y}|H_j)) = \sum_{d=1}^{\mathcal{D}} J(p(\mathbf{y}_d|H_i), p(\mathbf{y}_d|H_j))$ and

(19), the total J-divergence with multidimensional local modulation is given in (18), shown at the bottom of the page, where

$$\psi_{i,(k,d)} = \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \left[\left(\delta_{(k,d)}^*(l) - \delta_{(k,d)}^*(m) \right)^2 \times P_k(u_k = l|H_i)P_k(u_k = m|H_i) \right]$$

$$\phi_{(k,d)}^{ij} = \sum_{l=1}^{\mathcal{H}} \left[\delta_{(k,d)}^*(l) \cdot (P_k(u_k = l|H_i) - P_k(u_k = l|H_j)) \right]^2,$$

$PW_{(k,d)}^* = \sum_{l=1}^{\mathcal{H}} |\delta_{(k,d)}^*(l)|^2 P(u_k = l)$, d is the dimensional index of the transmission symbol, i, j are the indexes of multiple hypotheses, k is the sensor or channel node index, and l, m are the hypothesis indexes of transmission symbols. Similar system parameters from Fig. 2 were used to generate Fig. 3 for a multidimensional system. Differences include the threshold $J_0 = 0.1$ and the 2-D symbol constellation (QAM), which has less average symbol power than the 1D-symbol transmission system. To make a fair and simple comparison, we apply a simple structural constraint to the local power control by setting $\boldsymbol{\delta}_k(l = 1) = (\zeta - \tau, \zeta - \tau)$, $\boldsymbol{\delta}_k(l = 2) = (-\zeta - \tau, \zeta - \tau)$, $\boldsymbol{\delta}_k(l = 3) = (-\zeta - \tau, -\zeta - \tau)$, and $\boldsymbol{\delta}_k(l = 4) = (\zeta - \tau, -\zeta - \tau)$ where ζ is a given constant and the design parameter is τ. The square QAM structure can be translated based on the local decision probabilities. τ can be obtained based on local decision statistics $P(u_k = l)$, which then defines the optimal symbol constellation.

In Fig. 3, it is seen that, at reasonably high P_{tot}, the slope of the outage probability is proportional to the number of nodes in the system. The power gain achieved by local optimization manifests itself as a shift of the curve. An interesting point is that this local power gain is achievable even for a single-node system, and the shifts for one, three, and six nodes are approximately equal. That is to say, the gain obtained through local power control is not related to the number of nodes, but instead is strongly related to the *a priori* source statistics. However, full detection diversity order is observed even under uniform average power allocation. Although detection diversity order is related to large-$\mathcal{K}$ asymptotic divergence, it is seen in Fig. 3 that even systems with small $\mathcal{K}$ (such as one or three) achieve full diversity order. Although we have only derived diversity order using asymptotic total J-divergence for a homogeneous sensor environment, it can be seen that detection diversity also applies to a heterogeneous sensor environment. In the next section, we simultaneously apply both power control strategies—local

$$J^{\text{total}} = \frac{1}{2} \sum_{d=1}^{\mathcal{D}} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij} \right\}}{\mathcal{K}\mathcal{D} \sum_{l=1}^{\mathcal{H}} \left| \delta_{(k,d)}^*(l) \right|^2 P(u_k = l) + P_{\text{tot}} s_k \{\psi_{i,(k,d)}\}} + \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij} \right\}}{\mathcal{K}\mathcal{D} \sum_{l=1}^{\mathcal{H}} \left| \delta_{(k,d)}^*(l) \right|^2 P(u_k = l) + P_{\text{tot}} s_k \{\psi_{jk}\}} \right]$$
$$\times P(H_i) P(H_j)$$
$$= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \left[\frac{P_{\text{tot}} s_k \left\{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi_{(k,d)}^{ij} \right\}}{\mathcal{K} \cdot \mathcal{D} \cdot PW_{(k,d)}^* + P_{\text{tot}} s_k \{\psi_{ik}\}} + \frac{P_{\text{tot}} s_k \left\{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi_{(k,d)}^{ij} \right\}}{\mathcal{K} \cdot \mathcal{D} \cdot PW_{(k,d)}^* + P_{\text{tot}} s_k \{\psi_{j,(k,d)}\}} \right] P(H_i) P(H_j) \qquad (18)$$

power control and channel power control—to a distributed detection system.

V. LOCAL-CHANNEL POWER TRANSMISSION STRATEGY

A. Optimization With a Total Power Constraint

If local decision probabilities are available at each local sensor, the sensor can adjust its local transmitting symbol in an optimal way. On the other hand, if the propagation channel states are also known, then the optimized symbol constellation of each sensor can also be optimally amplified within given total and individual power constraints. Such a strategy is formulated as follows. Recall that the average transmit power transmitted by the kth sensor is $\{P_{(k,d)}|d = 1, 2, \ldots, \mathcal{D}\} = \{PW^*_{(k,d)} \cdot a^2_{(k,d)}|d = 1, 2, \ldots, \mathcal{D}\} = \{\sum_{l=1}^{\mathcal{H}} |\delta^*_{(k,d)}(l)|^2 \cdot P(u_k = l)a^2_{(k,d)}|d = 1, 2, \ldots, \mathcal{D}\}$ where $\mathcal{D}$ is the dimensionality of the symbol constellation. In the previous section, the average power P_k of each sensor was constrained to be equal, meaning the amplifying factor was decided based on the optimized symbol constellation. In this section, we use multidimensional average powers $P_{(k,d)}$ of each sensor as additional design variables. This additional freedom leads to an optimization problem defined as (19), shown at the bottom of the page, where

$$\psi_{i,(k,d)} = \sum_{l=1}^{\mathcal{H}} \sum_{m=l+1}^{\mathcal{H}} \left[\left(\delta^*_{(k,d)}(l) - \delta^*_{(k,d)}(m) \right)^2 \right.$$
$$\left. \times P_k(u_k = l|H_i)P_k(u_k = m|H_i) \right]$$
$$\phi^{ij}_{(k,d)} = \sum_{l=1}^{\mathcal{H}} \left[\delta^*_{(k,d)}(l) \cdot (P_k(u_k = l|H_i) - P_k(u_k = l|H_j)) \right]^2 .$$

The optimized constellation power of the kth sensor is $\{PW^*_{(k,d)}|d = 1, 2, \ldots, \mathcal{D}\}$. The average transmitted power in the dth dimension is $P^*_{(k,d)} = a^2_{(k,d)} \cdot PW^*_{(k,d)}$ where the asterisk means optimized value. Therefore, the amplifying factor $a_{(k,d)}$ is controlled by $\delta_{(k,d)}(l)^*$ and $P^*_{(k,d)}$, which are respectively the local power control and channel power control. Therefore, we can optimize system performance by performing both local power control and channel power control.

For a scalar modulation system, once the P^*_k's are found, the amplification factors a_k's can be found. The objective function is nondecreasing with increasing P_{tot} because

$$\left\{ \frac{\partial}{\partial P_k} J^{\text{total}}(P_1, P_2, \cdots, P_{\mathcal{K}}) \geq 0 | k = 0, 1, \cdots \mathcal{K} \right\} \quad (20)$$

where the P_k's are nonnegative. A detailed proof is given in Appendix VII-E. In the multidimensional modulation case, the total J-divergence $J^{\text{total}}(P_{(1,d)}, P_{(2,d)}, \cdots, P_{(\mathcal{K},d)})$, is still nondecreasing since it is a linear sum of nondecreasing scalar modulation systems. Therefore, we can get the optimal power allocation at the boundary of the power constraint, $\sum_{k=1}^{\mathcal{K}} \sum_{d=1}^{\mathcal{D}} P_{(k,d)} = P_{\text{tot}}$ by a convex optimization algorithm for the concave region of the object function [10], [11]. Though the objective function is not necessarily concave in general, for the parameters of interest in this paper all of the objective functions are concave and we apply the following convex optimization algorithm to find the optimal power allocation.

To get the optimal solution, we need to apply Lagrange multipliers to (19)

$$L \left(\{P_{(k,d)}|k = 1, 2, \ldots, \mathcal{L}, d = 1, 2, \ldots, \mathcal{D}\} \right) = J^{\text{total}}(\cdot)$$
$$- \lambda \left(\sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} P_{(k,d)} - P_{\text{tot}} \right) + \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \nu_{d,k} P_{(k,d)} \quad (21)$$

where λ and $\{\nu_{d,k}|k = 1, 2, \ldots, \mathcal{L}, d = 1, 2, \ldots, \mathcal{D}\}\}$ are Lagrange multipliers. After applying the derivative to (21), we can get the Karush-Kuhn-Tucker conditions and calculate the optimal power in the same manner with [11]. In cases where the objective function is not concave, another technique such as the interior point method [11], [31], [32] must be used.

For a simulation result, we use the same setup as that of Fig. 3, except we assume homogeneous local sensor statistics according to the transition matrices

$$\mathbf{L}_k = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \text{ for all } k$$

$$\max \ J^{\text{total}}(\cdot) = \frac{1}{2} \sum_{d=1}^{\mathcal{D}} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \left[\sum_{k=1}^{\mathcal{K}} \frac{P_{(k,d)} s_k \left\{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi^{ij}_{(k,d)} \right\}}{\sum_{l=1}^{\mathcal{H}} \left| \delta^*_{(k,d)}(l) \right|^2 P(u_k = l) + P_{(k,d)} s_k \{\psi_{i,(k,d)}\}} \right.$$
$$\left. + \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi^{ij}_{(k,d)} \right\}}{\sum_{l=1}^{\mathcal{H}} \left| \delta^*_{(k,d)}(l) \right|^2 P(u_k = l) + P_{(k,d)} s_k \{\psi_{j,(k,d)}\}} \right] P(H_i) P(H_j)$$
$$= \frac{1}{2} \sum_{i=1}^{\mathcal{H}} \sum_{j=i+1}^{\mathcal{H}} \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} \left[\frac{P_{(k,d)} s_k \left\{ \psi_{j,(k,d)} - \psi_{i,(k,d)} + \phi^{ij}_{(k,d)} \right\}}{PW^*_{(k,d)} + P_{(k,d)} s_k \{\psi_{i,(k,d)}\}} + \frac{P_{(k,d)} s_k \left\{ \psi_{i,(k,d)} - \psi_{j,(k,d)} + \phi^{ij}_{(k,d)} \right\}}{PW^*_{(k,d)} + P_{(k,d)} s_k \{\psi_{j,(k,d)}\}} \right] P(H_i) P(H_j)$$

$$\text{s.t.} \quad \sum_{d=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} P_{(k,d)} \leq P_{\text{tot}}, \ P_{(k,d)} \geq 0, k = 1, \ldots, \mathcal{K} \text{ and } d = 1, \ldots, \mathcal{D} \quad (19)$$

TABLE I
POWER DISTRIBUTION BY LOCAL POWER CONTROL IN PERCENTAGE OF TOTAL POWER ALLOCATED TO EACH SENSOR

P_{tot}	Scheme	Sensor									
		1	2	3	4	5	6	7	8	9	10
P1	NLP	100	0	0	0	0	0	0	0	0	0
	LP	100	0	0	0	0	0	0	0	0	0
P2	NLP	43.17	34.41	22.42	0	0	0	0	0	0	0
	LP	37.47	32.01	23.50	7.02	0	0	0	0	0	0
P3	NLP	28.04	25.76	20.98	11.21	6.79	3.82	3.39	0	0	0
	LP	23.64	22.50	19.17	12.04	8.76	6.55	6.23	1.11	0	0
P4	NLP	21.32	20.71	18.07	12.22	9.50	7.66	7.40	3.12	0	0
	LP	18.15	18.22	16.50	12.36	10.39	9.05	8.85	5.71	0.78	0
P5	NLP	17.71	17.86	16.25	12.33	10.45	9.17	8.98	5.98	1.26	0
	LP	15.04	15.60	14.62	11.90	10.56	9.63	9.50	7.30	3.83	2.03

and we optimize the local+channel power control according to (19).

Fig. 4 shows numerically generated detection outage probability versus the total power constraint for the three different optimization strategies in a six-node system. The three strategies are: 1) no power control; 2) local power control (symbol constellation optimization) with equal transmit power (from the previous section); and 3) optimization of both the symbol constellations and the amplification factors (local and channel power optimization). Note that all three curves have the same asymptotic slope but have different translations due to different power optimization schemes. Therefore, each system has the same diversity order, but for a given value of the power constraint, the curve corresponding to local-channel power control has the lowest probability of outage.

B. Optimization With Individual and Total Power Constraints

In a practical system, each sensor may have an individual power constraint imposed by its battery or certain transmission regulations. The total power constraint above does not model this scenario; therefore, we now modify the optimization problem to included individual power constraints as well. Define a new optimization problem as

$$\max \quad J^{\text{total}}(\cdot)$$
$$\text{s.t.} \quad \sum_{k=1}^{\mathcal{D}} \sum_{k=1}^{\mathcal{K}} P_{(k,d)} \leq P_{\text{tot}},$$
$$0 \leq P_{(k,d)} \leq P_k^{\max}(d), \ k = 1, \ldots, \mathcal{K} \text{ and } d = 1, \ldots, \mathcal{D}$$
$$(22)$$

where the optimization is again over the $P_{(k,d)}$'s and $P_k^{\max}(d)$ is the individual maximum allowed power. Since the additional power constraints on the individual sensors are linear constraints, there is no change on the conditions given for (19), such that the optimal solution can be similarly solved.

To find the solution to (22), the $P_{(k,d)}$'s are first optimized with the total power constraint, but without the individual power constraints. If some of the $P_{(k,d)}$'s are more than their upper limits, $P_k^{\max}(d)$, then the optimal power for that sensor lies on the boundary of the individual power constraint and they are forced to equal $P_k^{\max}(d)$. These sensors are then removed from P_{tot} and the optimization procedure continues with the

remaining sensors. This iteration continues until all power constraints are satisfied and a global optimum is obtained [28]. Fig. 5 compares detection outage probability for the case with both total and individual power constraints to the case with only the total power constraints in a nine-node system. We see that with the additional individual power constraints, full diversity is still achieved but the power gain is reduced compared with the case without individual power constraints. We consider the power allocation to each sensor in the next subsection.

C. Power Distribution Across Sensors

In Table I, homogeneous sensor statistics are adopted in order to clearly show the effect of the local power control. We define five different total power constraints in increasing order according to $P1 < P2 < P3 < P4 < P5$. We also order ten sensors in decreasing order of channel strength. In other words, the channel gain coefficient is strongest for sensor 1 and weakest for sensor 10. When we apply the lowest power constraint, all of the transmitted power is focused on sensor 1 because power is extremely limited and the first sensor's channel has the highest SNR. As more power is allowed, the power begins to be distributed to the next highest SNR channel, then the next highest, and so on. Finally, for the highest total power constraint, sensor 10 is utilized, but only for the system with local power optimization (LP). In fact, at every power level except the weakest, the system with local power control is able to exploit one additional sensor compared to the system with no local power control (NLP). Power distribution in the absence of local power control is explained in detail in [11]. The main goal of Table I is to show the effect of the local power control. By comparing LP with NLP for the same total power constraint, we can see that the system with local power control can often exploit more sensors than the system without local power control, thereby achieving higher power gain and improved outage probability. The reason is that the optimized local constellation consumes lower power for the same detection performance.

VI. CONCLUSIONS

A local-channel power control scheme applicable to distributed multiple-hypothesis detection systems in slow-fading environments is introduced. We generalized the outage probability of [19] for the multiple-hypothesis problem and showed that the detection diversity still holds. We also analyzed the

proposed power control strategy with the asymptotic total J-divergence and showed mathematically the relationship between asymptotic total J-divergence and detection outage probability. This relationship leads to the detection diversity gain under the homogenous sensor environment. We showed via simulation that detection diversity is observed even in heterogeneous sensor environment. Then, using the detection outage probability as a long-term performance measure, we showed that the distributed system is efficiently improved through both local power optimization and channel power optimization. The local power optimization is based on local decision statistics and results in optimized transmit symbol constellations while the channel power optimization is based on channel fading states and results in optimized amplifying factors. Individual power constraints were also considered.

APPENDIX

A. Relationship Between Bound Probability of Error and Total J-Divergence

The inequality $P_e > P(H_1)P(H_2)e^{-J/2}$ [24] is valid between two hypotheses with *a priori* probabilities whose sum are less than one. This can be proved by starting with

$$P_e^{12} = P(u|H_1)\acute{P}(H_1) + P(u|H_2)\acute{P}(H_2)$$
$$> \acute{P}(H_1)\acute{P}(H_2)e^{-\frac{J_{12}}{2}}.$$

Let $\acute{P}(H_1) = P(H_1)/a$ and $\acute{P}(H_2) = P(H_2)/b$ be new probabilities where a and $b \geq 1$. Then

$$P(u|H_1)\frac{P(H_1)}{a} + P(u|H_2)\frac{P(H_2)}{b}$$
$$> \frac{P(H_1)}{a}\frac{P(H_2)}{b}e^{-\frac{J_{12}}{2}}$$
$$\iff P(u|H_1)P(H_1)b + P(u|H_2)P(H_2)a$$
$$> P(H_1)P(H_2)e^{-\frac{J_{12}}{2}}.$$

Now, we extend the inequality for the multiple-hypotheses problem by defining an upper bound to the bound probability of error, $P_e^{bound} = \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P_e^{ij}$, and a probability-weighted total J-divergence, $J^{\text{total}} = \sum_{i=1}^{\mathcal{H}}\sum_{j=i+1}^{\mathcal{H}}J^{ij}\cdot P(H_i)\cdot P(H_j)$. At first, by summing the inequalities, we get

$$\sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P_e^{ij} > \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P(H_i)P(H_j)e^{-\frac{J^{ij}}{2}}.$$

By an inequality, $e^x > 1 + x$, which is derived via the Taylor series of the exponential function, we get

$$\sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P_e^{ij} > \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P(H_i)P(H_j)e^{-\frac{J^{ij}}{2}}$$
$$> \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P(H_i)P(H_j)\left(1 - \frac{J^{ij}}{2}\right).$$
$$= \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P(H_i)P(H_j) - J^{\text{total}}$$

since $J^{\text{total}} = (1/2)\sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}J^{ij}\cdot P(H_i)\cdot P(H_j)$. Finally, we get

$$P_e^{bound} > C - J^{\text{total}}$$

where $C = \sum_{i=1}^{\mathcal{H}}\sum_{j=1}^{\mathcal{H}}P(H_i)p(H_j)$.

B. Derivation of (7)

We now derive an approximate total J-divergence measure that explicitly includes the optimized transmit symbol constellation. The measure is approximate because the signals received at the fusion center are distributed according to a Gaussian mixture, which doesn't lead to closed-form expressions. Therefore, we adopt the strategy used in [10] and [19], which is to approximate the received conditional probabilities in (3) by Gaussian densities with the same mean and covariance as the mixture.

As a first step, we must first derive the mean and covariance of $\mathbf{y}$ under H_i to find the multivariate conditional pdf $p(\mathbf{y}|H_i)$. The mean vector is given as

$$\boldsymbol{\mu}_i = \int_{\mathbf{y}} \mathbf{y}p(\mathbf{y}|H_i)d\mathbf{y}$$
$$= \int_{\mathbf{y}} \mathbf{y}\sum_{\mathbf{u}} p(\mathbf{y}|\mathbf{u})P(\mathbf{u}|H_i)d\mathbf{y}$$
$$= \sum_{\mathbf{u}} P(\mathbf{u}|H_i)\int_{\mathbf{y}} \mathbf{y}p(\mathbf{y}|\mathbf{u})d\mathbf{y}$$
$$= \sum_{\mathbf{u}} P(\mathbf{u}|H_i)\mathbf{HAf_s}(\mathbf{u})$$
$$= \mathbf{HA}\begin{bmatrix} \sum_{l=1}^{\mathcal{H}}\delta_1(l)P(u_1 = l|H_i) \\ \sum_{l=1}^{\mathcal{H}}\delta_2(l)P(u_2 = h|H_i) \\ \vdots \\ \sum_{l=1}^{\mathcal{H}}\delta_{\mathcal{K}}(l)P(u_{\mathcal{K}} = l|H_i) \end{bmatrix}$$
$$= \mathbf{HA}\boldsymbol{\xi}_i \qquad (23)$$

where $i \in \{1, 2, \cdots, \mathcal{H}\}$, $\delta_k(l) = f_s(u_k = l)$ and

$$\boldsymbol{\xi}_i = \begin{bmatrix} \sum_{l=1}^{\mathcal{H}}\delta_1(l)P(u_1 = l|H_i) \\ \sum_{l=1}^{\mathcal{H}}\delta_2(l)P(u_2 = l|H_i) \\ \vdots \\ \sum_{l=1}^{\mathcal{H}}\delta_{\mathcal{K}}(l)P(u_{\mathcal{K}} = l|H_i) \end{bmatrix}.$$

The covariance matrix is given as

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

From (23), $\boldsymbol{\mu}_i\boldsymbol{\mu}_i^T = \mathbf{HA}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\mathbf{HA}$. Hence, after some algebra with $\sum_{u_k=1}^{\mathcal{H}} P(u_k|H_i) = 1$, we obtain

$$
\begin{aligned}
\boldsymbol{\Sigma}_i &= \mathbf{HA}\sum_{\mathbf{u}}\mathbf{f_s(u)f_s(u)}^T P(\mathbf{u}|H_i)\mathbf{A}^T\mathbf{H}^T \\
&\quad - \mathbf{HA}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\mathbf{A}^T\mathbf{H}^T + \mathbf{R} \\
&= \mathbf{HA}\left[\sum_{\mathbf{u}}\mathbf{f_s(u)f_s(u)}^T P(\mathbf{u}|H_i) - \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right]\mathbf{A}^T\mathbf{H}^T + \mathbf{R} \\
&= \mathbf{HAC}_i\mathbf{A}^T\mathbf{H}^T + \mathbf{R} \qquad (24)
\end{aligned}
$$

where $\mathbf{R}$ is the covariance matrix of the channel noise vector $\mathbf{n}$. In (24), $\mathbf{C}_i$ is a $\mathcal{K} \times \mathcal{K}$ diagonal matrix with elements

$$
\begin{aligned}
C_i(k,k) = \sum_{l=1}^{\mathcal{H}}\sum_{m=l+1}^{\mathcal{H}}&\{\delta_k(l) - \delta_k(m)\}^2 \\
&\times P(u_k = l|H_i)P(u_k = m|H_i)
\end{aligned}
$$

where $i \in \{1, 2, \cdots, \mathcal{H}\}$. Since the matrix $\mathbf{C}_i$ is diagonalized, J^{total} becomes a simple form even in multiple-hypotheses problem. We define multivariate normal distributions with the mean vectors, $\boldsymbol{\mu}_i$, $i \in \{1, 2, \cdots, \mathcal{H}\}$, and covariance matrices, $\boldsymbol{\Sigma}_i$, $i \in \{1, 2, \cdots, \mathcal{H}\}$. Then, by substituting the distributions into the definition of J-divergence, we get

$$
\begin{aligned}
J\left(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)\right) &= \frac{1}{2}Tr\left[\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Sigma}_i\right] \\
&+ \frac{1}{2}Tr\left[\left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\right)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T\right] - \mathcal{K} \quad (25)
\end{aligned}
$$

where $\mathcal{K}$ is the dimension of the covariance matrix [33].

Now using the mean matrices and covariance matrices from (23) and (24), we can define the following terms in (25):

$$
\begin{aligned}
\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_j &= (\mathbf{R}+\mathbf{HAC}_i\mathbf{A}^T\mathbf{H}^T)^{-1}(\mathbf{R}+\mathbf{HAC}_j\mathbf{A}^T\mathbf{H}^T) \\
\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Sigma}_i &= (\mathbf{R}+\mathbf{HAC}_j\mathbf{A}^T\mathbf{H}^T)^{-1}(\mathbf{R}+\mathbf{HAC}_i\mathbf{A}^T\mathbf{H}^T) \\
\left(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{\Sigma}_j^{-1}\right)&(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T \\
&= \left\{(\mathbf{R}+\mathbf{HAC}_i\mathbf{A}^T\mathbf{H}^T)+(\mathbf{R}+\mathbf{HAC}_j\mathbf{A}^T\mathbf{H}^T)^{-1}\right\} \\
&\quad \times \mathbf{HA}\boldsymbol{\xi}\boldsymbol{\xi}^T\mathbf{A}^T\mathbf{H}^T
\end{aligned}
$$

where $\boldsymbol{\xi} = \boldsymbol{\xi}_i - \boldsymbol{\xi}_j$.

Therefore, we have

$$
\begin{aligned}
J\left(p(\mathbf{y}|H_i), p(\mathbf{y}|H_j)\right) &= J^{ij} \\
&= \frac{1}{2}Tr\left[\left[\mathbf{R}+\mathbf{HA}(\mathbf{C}_i+\boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{H}^T\right]\right. \\
&\quad \left. \times [\mathbf{R}+\mathbf{HAC}_j\mathbf{A}^T\mathbf{H}^T]^{-1}\right] \\
&\quad + \frac{1}{2}Tr\left[\left[\mathbf{R}+\mathbf{HA}(\mathbf{C}_j+\boldsymbol{\xi}\boldsymbol{\xi}^T)\mathbf{A}^T\mathbf{H}^T\right]\right. \\
&\quad \left. \times [\mathbf{R}+\mathbf{HAC}_i\mathbf{A}^T\mathbf{H}^T]^{-1}\right] - \mathcal{K}
\end{aligned}
$$

where $Tr[\cdot]$ denotes the matrix trace operation.

C. Derivation of Case 1, (10)

In (9), let us consider J_1^{ij} for a large $\mathcal{K}$, $(\psi_{jk}-\psi_{ik}+\phi_k^{ij}) \geq 0$, and $PW_k^* > 0$. We have

$$
\begin{aligned}
&\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*} - \frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^* + P_{\text{tot}}s_k\psi_{ik}} \\
&= \frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2 PW_k^{*^2} + \mathcal{K}PW_k^* P_{\text{tot}}s_k\psi_{ik}} \\
&\leq \frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2 PW_k^{*^2}}
\end{aligned}
$$

where $PW_k^* = \sum_{h=1}^{\mathcal{H}}\|\boldsymbol{\delta}_h\|^2 P(u_k = h)$, since P_{tot}, s_k, ψ_{ik}, and PW_k^* are positive.

The inequality is arranged as follows:

$$
\begin{aligned}
&\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*} - \frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2\cdot PW_k^{*^2}} \\
&\leq \frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^* + P_{\text{tot}}s_k\psi_{ik}} \\
&\leq \frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*} \\
&\Longrightarrow \sum_{k=1}^{\mathcal{K}}\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*} \\
&\quad - \sum_{k=1}^{\mathcal{K}}\frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2\cdot PW_k^{*^2}} \\
&\leq 2J_1^{ij} \leq \sum_{k=1}^{\mathcal{K}}\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*}
\end{aligned}
$$

since

$$
\begin{aligned}
\sum_{k=1}^{\mathcal{K}}&\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^* + P_{\text{tot}}s_k\psi_{ik}} \\
&\leq \sum_{k=1}^{\mathcal{K}}\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*}.
\end{aligned}
$$

D. Derivation of Case 2, (12)

In (9), let us consider $J_1(\cdot)$ for a large $\mathcal{K}$, $(\psi_{jk}-\psi_{ik}+\phi_k^{ij}) \leq 0$, and $PW_k^* > 0$. We have

$$
\begin{aligned}
&\frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^*} - \frac{P_{\text{tot}}s_k\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}\cdot PW_k^* + P_{\text{tot}}s_k\psi_{ik}} \\
&= \frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2 PW_k^{*^2} + \mathcal{K}\cdot PW_k^* P_{\text{tot}}s_k\psi_{ik}} \\
&\geq \frac{P_{\text{tot}}^2 s_k^2\psi_{ik}\left\{\psi_{jk}-\psi_{ik}+\phi_k^{ij}\right\}}{\mathcal{K}^2\cdot PW_k^{*^2}}
\end{aligned}
$$

where $PW_k^* = \sum_{h=1}^{\mathcal{H}} \|\boldsymbol{\delta}_h\|^2 \cdot P(u_k = h)$, since P_{tot}, s_k, ψ_{ik}, and PW_k^* are positive. The inequality is arranged as follows:

$$\frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*} - \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K}^2 \cdot PW_k^{*2}}$$

$$\geq \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^* + P_{\text{tot}} s_k \psi_{ik}}$$

$$\geq \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$

$$\implies \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*} \qquad .$$

$$\leq 2 J_1^{ij}$$

$$\leq \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$

$$- \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}}^2 s_k^2 \psi_{ik} \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K}^2 \cdot PW_k^{*2}}$$

since

$$\sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^*}$$

$$\leq \sum_{k=1}^{\mathcal{K}} \frac{P_{\text{tot}} s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{\mathcal{K} \cdot PW_k^* + P_{\text{tot}} s_k \psi_{ik}}.$$

E. Derivation of Nondecreasing Characteristic of Total J-Divergence

The total J-divergence, J^{total} is a linear combination of J^{ij}. So, if J^{ij} is nondecreasing, then J^{total} is nondecreasing. The first partial derivative of $J^{ij}(\cdot)$ with respect to P_k is

$$\frac{\partial}{\partial P_k} \left[\sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \left\{\psi_{jk} - \psi_{ik} + \phi_k^{ij}\right\}}{PW_k^* + P_k s_k \{\psi_{ik}\}}\right.$$

$$\left. + \sum_{k=1}^{\mathcal{K}} \frac{P_k s_k \left\{\psi_{ik} - \psi_{jk} + \phi_k^{ij}\right\}}{PW_k^* + P_k s_k \{\psi_{jk}\}}\right]$$

$$= \frac{PW_k^*}{(PW_k^* + P_k s_k \{\psi_{ik}\})^2 (PW_k^* + P_k s_k \{\psi_{jk}\})^2}$$

$$\times \left[2 PW_k^{*2} \left(\phi_k^{ij}\right) + 2 P_k PW_k^* (\psi_{ik} + \psi_{jk})\right.$$

$$\left. + P_k^2 \left(\psi_{ik}^2 + \psi_{jk}^2\right)\right].$$

Therefore, the first partial derivative is nonnegative since PW_k^*, ϕ_k^{ij} and ψ_{ik} are nonnegative from (8). The total J-divergence is nondecreasing.

REFERENCES

[1] R. R. Tenney and N. R. Sandell, Jr., "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-17, no. 4, pp. 501–510, Jul. 1981.

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

[3] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 995–1000, Sep. 1989.

[4] 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.

[5] 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.

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

[7] R. Niu, B. Chen, and P. 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.

[8] J. 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.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power-efficient analog forwarding transmission in an inhomogeneous gaussian sensor network," in *Proc. IEEE Workshop on Signal Process. Advances in Wireless Commun.*, New York, Jan. 2005, pp. 121–125.

[10] X. Zhang, H. V. Poor, and M. Chiang, "Power allocation in distributed detection with wireless sensor networks," in *Proc. Military Commun. Conf. (MILCOM '06)*, Oct. 23–25, 2006, pp. 1–7.

[11] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4124–4140, Sep. 2008.

[12] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 5, pp. 761–765, Sep. 1989.

[13] S. C. A. Thomopoulos and L. Zhang, "Distributed decision fusion with networking delays and channel errors," *Inf. Sci.*, vol. 66, pp. 91–118, Dec. 1992.

[14] F. A. Sadjadi, "Hypotheses testing in a distributed environment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 134–137, Mar. 1986.

[15] J. J. Chao, E. Drakopoulos, and C. C. Lee, "An evidential reasoning approach to distributed multiple-hypothesis detection," in *Proc. 26th IEEE Conf. Decision and Control*, Dec. 1987, pp. 1826–1831.

[16] X. G. Wang and H. C. Shen, "Multiple hypotheses testing strategy for distributed multisensor systems," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS 2000)*, 2000, vol. 2, pp. 1440–1445.

[17] X. G. Wang, M. Moallem, and R. V. Patel, "Distributed multiple hypotheses testing with serial distributed decision fusion," in *Proc. IEEE Int. Symp. Comput. Intell. Robotics and Automation*, 2001, pp. 549–554.

[18] S. Oh, S. Sastry, and L. Schenato, "A hierarchical multiple-target tracking algorithm for sensor networks," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA 2005)*, Apr. 2005, pp. 2197–2202.

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

[20] H. Kim and S. Cui, "Detection outage and detection diversity in distributed sensor networks," in *Proc. IEEE Int. Symp. Information Theory*, Jun. 24–29, 2007, pp. 1606–1610.

[21] E. Mosca, "Probing signal design for linear channel identification," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 481–487, Jul. 1972.

[22] T. L. Grettenberg, "Signal selection in communication and radar systems," *IEEE Trans. Inf. Theory*, vol. IT-9, pp. 265–275, Oct. 1963.

[23] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.

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

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

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

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

[28] S. Cui, J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.

[29] A. Weiss, "An introduction to large deviations for communication networks," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 6, pp. 938–952, Aug. 1995.

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

[31] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, NY: Athena Scientific, Sep. 1999.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[33] J. P. Campbel, "Speaker recognition: A tutorial," *Proc. IEEE*, vol. 85, no. 9, pp. 1437–1462, Sep. 1997.

Hyoung-soo Kim (S'07) received the B.S. and M.S. degree in control and instrumentation engineering from the Hanyang University, Seoul, Korea, in 1997 and 1999. He is currently pursuing the Ph.D. degree in electrical engineering from the University of Arizona, Tucson, under the supervision of Prof. Nathan A. Goodman.

He is also currently a Graduate Research Assistant in the Laboratory for Sensor and Array Processing, Electrical and Computer Engineering Department, University of Arizona. His research interests span wireless communications, signal processing, and information theory.

Nathan A. Goodman (S'98–M'02–SM'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, in 1995, 1997, and 2002, respectively.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering, University of Arizona, Tucson. Within the department, he directs the Laboratory for Sensor and Array Processing. His research interests are in novel system concepts and signal processing techniques for radar, antenna arrays, and other sensors. From 1996–1998, he was an RF Systems Engineer for Texas Instruments, Dallas, TX. From 1998–2002, he was a Graduate Research Assistant in the Radar Systems and Remote Sensing Laboratory, University of Kansas. He serves as Deputy Editor-in-Chief of *Digital Signal Processing*. In the 2009–2010 academic year, he was also a Visiting Senior Research Engineer at the Georgia Tech Research Institute, Smyrna.

Dr. Goodman was awarded the Madison A. and Lila Self Graduate Fellowship from the University of Kansas in 1998. He was also awarded the IEEE 2001 International Geoscience and Remote Sensing Symposium Interactive Session Prize Paper Award.