Research Article

Decentralized Detection in Wireless Sensor Networks with Channel Fading Statistics

Bin Liu and Biao Chen

Department of Electrical Engineering and Computer Science (EECS), Syracuse University, 223 Link Hall,
Syracuse, NY 13244-1240, USA

Existing channel aware signal processing design for decentralized detection in wireless sensor networks typically assumes the clairvoyant case, that is, global channel state information (CSI) is known at the design stage. In this paper, we consider the distributed detection problem where only the channel fading statistics, instead of the instantaneous CSI, are available to the designer. We investigate the design of local decision rules for the following two cases: (1) fusion center has access to the instantaneous CSI; (2) fusion center does not have access to the instantaneous CSI. As expected, in both cases, the optimal local decision rules that minimize the error probability at the fusion center amount to a likelihood ratio test (LRT). Numerical analysis reveals that the detection performance appears to be more sensitive to the knowledge of CSI at the fusion center. The proposed design framework that utilizes only partial channel knowledge will enable distributed design of a decentralized detection wireless sensor system.

1. INTRODUCTION

While study of decentralized decision making can be traced back to the early 1960s in the context of team decision problems (see, e.g., [1]), the effort significantly intensified since the seminal work of [2]. Classical distributed detection [3–7], however, typically assumes error-free transmission between the local sensors and the fusion center. This is overly idealistic in the emerging systems with stringent resource and delay constraints, such as the wireless sensor network (WSN) with geographically dispersed lower-power low-cost sensor nodes. Accounting for nonideal transmission channels, channel aware signal processing for distributed detection problem has been developed in [8–10]. The optimal local decision rule was still shown to be a monotone likelihood ratio partition of its observation space, provided the observations were conditionally independent across the sensors. It was noted recently that such optimality is preserved for a more general setting [11].

The work in [8–10] assumed a clairvoyant case, that is, global information regarding the transmission channels between the local sensors and the fusion center is available at the design stage. This approach is theoretically significant as it provides the best achievable detection performance

to which any suboptimal approach needs to be compared. However, its implementation requires the exact knowledge of global channel state information (CSI) which may be costly to acquire. In the case of fast fading channel, the sensor decision rules need to be synchronously updated for different channel states; this adds considerable overhead which may not be affordable in resource constrained systems.

To make the channel aware design more practical, the requirement of the global CSI in the distributed signaling design needs to be relaxed. In the present work, only partial channel knowledge instead of the global CSI is assumed to be available. In the context of WSN, a reasonable assumption is the availability of channel fading statistics, which may remain stationary for a sufficiently long period of time. Therefore, the updating rate of the decision rules is more realistic. In this paper, we consider the distributed detection problem where the designer only has the channel fading statistics instead of the instantaneous CSI. In this case, a sensible performance measure is to use the average error probability at the fusion center where the averaging is performed with respect to the channel state. We restrict ourselves to binary local sensor outputs and derive the necessary conditions for optimal local decision rules that minimize the average error probability at the fusion center for the following two
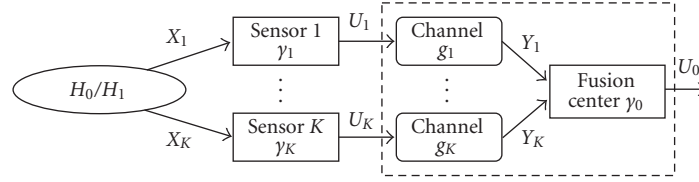
FIGURE 1: A block diagram for a wireless sensor network tasked for binary hypothesis testing with channel fading statistics.

cases: (1) CSIF: the fusion center has access to the instantaneous CSI. (2) NOCSIF: the fusion center does *not* have access to the instantaneous CSI. We note here that for the CSIF case, the design of sensor decision rules does not require CSI; the CSI is only used in the fusion rule design, which is reasonable due to the typical generous resource constraint at the fusion center. Its computation, however, is very involved and has to resort to exhaustive search. On the other hand, as to be elaborated, the NOCSIF case can be reduced to the channel aware design where one averages the channel transition probability with respect to the fading channel using the known fading statistics.

We show that the sensor decision rules amount to local LRTs for both cases. Compared with the existing channel aware design based on CSI, the proposed approaches have an important practical advantage: the sensor decision rules remain the same for different CSI, as long as the fading statistics remain unchanged. This enables distributed design as no global CSI is used in determining the local decision rules. We also demonstrate through numerical examples that the proposed schemes suffer small performance loss compared with the CSI-based approach, as long as the CSI is available at the fusion center, that is, used in the fusion rule design.

The paper is organized as follows. Section 2 describes the system model and problem formulation. In Section 3, we establish, for both the CSIF and NOCSIF cases, the optimality of LRTs at local sensors for minimum average error probability at the fusion center. Numerical examples are presented in Section 4 to evaluate the performance of these two cases. Finally, we conclude in Section 5.

2. STATEMENT OF THE PROBLEM

Consider the problem of testing two hypotheses, denoted by H_0 and H_1, with respective prior probabilities π_0 and π_1. A total number of K sensors are used to collect observations X_k, for $k = 1, \ldots, K$. We assume throughout this paper that the observations are conditionally independent, that is,

$$p(X_1, \ldots, X_K \mid H_i) = \prod_{k=1}^{K} p(X_k \mid H_i), \quad i = 0, 1. \quad (1)$$

Without this assumption, the distributed detection design becomes much more complicated from an algorithmic viewpoint [12]. The optimal design is not completely understood even in the simplest case [13]. Upon observing X_k, each local sensor makes a binary decision[1]

$$U_k = \gamma_k(X_k), \quad k = 1, \ldots, K. \quad (2)$$

The decisions U_k are sent to a fusion center through parallel transmission channels characterized by

$$p(Y_1, \ldots, Y_K \mid U_1, \ldots, U_K, g_1, \ldots, g_K) = \prod_{k=1}^{K} p(Y_k \mid U_k, g_k), \quad (3)$$

where $\mathbf{g} = \{g_1, \ldots, g_K\}$ represents the CSI. Thus, from (3), the channels are orthogonal to each other, which can be achieved through, for example, partitioning time, frequency, or combinations thereof. For the CSIF case, the fusion center takes both the channel output $\mathbf{y} = \{Y_1, \ldots, Y_K\}$ and the CSI, $\mathbf{g}$, and makes a final decision using the optimal fusion rule to obtain $U_0 \in \{H_0, H_1\}$,

$$U_0 = \gamma_0(\mathbf{y}, \mathbf{g}). \quad (4)$$

For the case of NOCSIF, the fusion output depends on the channel output and the channel fading statistics,

$$U_0 = \gamma_0(\mathbf{y}), \quad (5)$$

where the dependence of fading channel statistics is implicit in the above expression. An error happens if U_0 differs from the true hypothesis. Thus, the error probability at the fusion center, conditioned on a given $\mathbf{g}$, is

$$P_{e0}(\gamma_0, \ldots, \gamma_K \mid \mathbf{g}) \triangleq P_r(U_0 \neq H \mid \gamma_0, \ldots, \gamma_K, \mathbf{g}), \quad (6)$$

where H is the true hypothesis. Our goal is, therefore, to design the optimal mapping $\gamma_k(\cdot)$ for each sensor and the fusion center that minimizes the average error probability, defined as

$$\min_{\gamma_0(\cdot), \ldots, \gamma_K(\cdot)} \int_{\mathbf{g}} P_{e0}(\gamma_0, \ldots, \gamma_K \mid \mathbf{g}) p(\mathbf{g}) d\mathbf{g}, \quad (7)$$

where $p(\mathbf{g})$ is the distribution of CSI. A simple diagram illustrating the model is given in Figure 1.

---

[1] The extension to the case with multibit local decision is straightforward by following the same spirit as in [10].

We point out here that integrating the transmission channels into the fusion rule design has been investigated before [14, 15]. The optimal fusion rule in the Bayesian sense amounts to the maximum a posteriori probability (MAP) decision, that is,

$$\frac{f(Y_1, \ldots, Y_K \mid H_1)}{f(Y_1, \ldots, Y_K \mid H_0)} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \frac{\pi_0}{\pi_1}. \quad (8)$$

Given specified local sensor signaling schemes and the channel characterization, this MAP decision rule can be obtained in a straightforward manner. As such, we will focus on the local decision rule design without further elaborating on the optimal fusion rule design.

3. DESIGN OF OPTIMAL LOCAL DECISION RULES

As in [9, 10], we adopt in the following a person-by-person optimization (PBPO) approach, that is, we optimize the local decision rule for the kth sensor given fixed decision rules at all other sensors and a fixed fusion rule. As such, the conditions obtained are necessary, but not sufficient, for optimality. Denote

$$\mathbf{u} = [U_1, U_2, \ldots, U_K], \qquad \mathbf{x} = [X_1, X_2, \ldots, X_K], \quad (9)$$

the average error probability at the fusion center is

$$
\begin{aligned}
P_{e0} &= \int_{\mathbf{g}} \sum_{i=0}^{1} \pi_i P(U_0 = 1 - i \mid H_i, \mathbf{g}) \, p(\mathbf{g}) d\mathbf{g} \\
&= \int_{\mathbf{g}} \sum_{i=0}^{1} \pi_i \int_{\mathbf{y}} P(U_0 = 1 - i \mid \mathbf{y}, \mathbf{g}) \sum_{\mathbf{u}} p(\mathbf{y} \mid \mathbf{u}, \mathbf{g}) \\
&\quad \times \int_{\mathbf{x}} P(\mathbf{u} \mid \mathbf{x}) p(\mathbf{x} \mid H_i) p(\mathbf{g}) \, d\mathbf{x} \, d\mathbf{y} \, d\mathbf{g},
\end{aligned} \quad (10)
$$

where, different from the CSI-based channel aware design, the local decision rules do *not* depend on the instantaneous CSI. Next, we will derive the optimal decision rules by further expanding the error probability with respect to the kth decision rule $\gamma_k(\cdot)$ for the two different cases.

3.1. The CSIF case

We first consider the case where the fusion center knows the instantaneous CSI. Define, for $k = 1, \ldots, K$ and $i = 0, 1$,

$$\mathbf{u}^k = [U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_K],$$

$$\mathbf{u}^{ki} = [U_1, \ldots, U_{k-1}, U_k = i, U_{k+1}, \ldots, U_K], \quad (11)$$

$$\mathbf{y}^k = [Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_K].$$

We can expand the average error probability in (10) with respect to the kth decision rule $\gamma_k(\cdot)$, and we get

$$
\begin{aligned}
P_{e0} &= \int_{X_k} P(U_k = 1 \mid X_k) \\
&\quad \times [\pi_0 p(X_k \mid H_0) A_k - \pi_1 p(X_k \mid H_1) B_k] dX_k + C,
\end{aligned} \quad (12)
$$

where

$$
\begin{aligned}
C &= \int_{X_k} \Bigg[\sum_{i=0}^{1} \pi_i p(X_k \mid H_i) \\
&\quad \times \int_{\mathbf{y}} \int_{\mathbf{g}} P(U_0 = 1 - i \mid \mathbf{y}, \mathbf{g}) \sum_{\mathbf{u}^k} p(\mathbf{y} \mid \mathbf{u}^{k0}, \mathbf{g}) \\
&\quad \times p(\mathbf{g}) p(\mathbf{u}^k \mid H_i) d\mathbf{g} \, d\mathbf{y} \Bigg] dX_k
\end{aligned} \quad (13)
$$

is a constant with regard to U_k, and

$$
\begin{aligned}
A_k &= \int_{\mathbf{y}} \int_{\mathbf{g}} P(U_0 = 1 \mid \mathbf{y}, \mathbf{g}) \Bigg[\sum_{\mathbf{u}^k} (p(\mathbf{y} \mid \mathbf{u}^{k1}, \mathbf{g}) - p(\mathbf{y} \mid \mathbf{u}^{k0}, \mathbf{g})) \\
&\quad \times p(\mathbf{g}) P(\mathbf{u}^k \mid H_0) \Bigg] d\mathbf{g} \, d\mathbf{y},
\end{aligned} \quad (14)
$$

$$
\begin{aligned}
B_k &= \int_{\mathbf{y}} \int_{\mathbf{g}} P(U_0 = 0 \mid \mathbf{y}, \mathbf{g}) \Bigg[\sum_{\mathbf{u}^k} (p(\mathbf{y} \mid \mathbf{u}^{k0}, \mathbf{g}) - p(\mathbf{y} \mid \mathbf{u}^{k1}, \mathbf{g})) \\
&\quad \times p(\mathbf{g}) P(\mathbf{u}^k \mid H_1) \Bigg] d\mathbf{g} \, d\mathbf{y}.
\end{aligned} \quad (15)
$$

To minimize P_{e0}, one can see from (12) that the optimal decision rule for the kth sensor is

$$
P(U_k = 1 \mid X_k) = \begin{cases} 0, & \pi_0 p(X_k \mid H_0) A_k > \pi_1 p(X_k \mid H_1) B_k, \\ 1, & \text{otherwise.} \end{cases} \quad (16)
$$

Let us further take a look at A_k in (14). We can rewrite it as

$$
\begin{aligned}
A_k &= \int_{\mathbf{y}^k} [P(U_0 = 1 \mid \mathbf{y}^k, U_k = 1) \\
&\quad - P(U_0 = 1 \mid \mathbf{y}^k, U_k = 0)] p(\mathbf{y}^k \mid H_0) d\mathbf{y}^k.
\end{aligned} \quad (17)
$$

Then, as shown in Appendix A, $A_k > 0$ as long as

$$L(U_k) \triangleq \frac{P(U_k \mid H_1)}{P(U_k \mid H_0)} \quad (18)$$

is a monotone increasing function of U_k (monotone LR index assignment), that is,

$$L(U_k = 1) > L(U_k = 0). \quad (19)$$

Similarly, $B_k > 0$ if condition (19) is satisfied. This immediately leads to the following result.

Theorem 1. *For the distributed detection problem with unknown CSI only at local sensors, the optimal local decision rule for the kth sensor amounts to the following LRT assuming condition* (19) *is satisfied,*

$$
P(U_k = 1 \mid X_k) = \begin{cases} 1, & \dfrac{p(X_k \mid H_1)}{p(X_k \mid H_0)} \geq \dfrac{\pi_0 A_k}{\pi_1 B_k}, \\ 0, & \dfrac{p(X_k \mid H_1)}{p(X_k \mid H_0)} < \dfrac{\pi_0 A_k}{\pi_1 B_k}, \end{cases} \quad (20)
$$

where A_k and B_k are defined in (14) *and* (15), *respectively.*

An alternative derivation, along the line of [11], is given in Appendix B.

Although the optimal local decision rule for each local sensor is explicitly formulated in (20), it is not amenable to direct numerical evaluation: in (14) and (15), the integrand involves the fusion rule that is a highly nonlinear function of the CSI $\mathbf{g}$, making the integration formidable. The only possible way of finding the optimal local decision rules appears to be an exhaustive search, whose complexity becomes prohibitive when K is large.

This numerical challenge motivates an alternative approach: instead of minimizing the average error probability, one can first marginalize the channel transition probability followed by the application of standard channel aware design [8–10]. That is, we first compute $p(\mathbf{y} \mid \mathbf{u})$ by marginalizing out the channel $\mathbf{g}$ using the channel fading statistics:

$$p(\mathbf{y} \mid \mathbf{u}) = \int_{\mathbf{g}} p(\mathbf{y} \mid \mathbf{u}, \mathbf{g}) p(\mathbf{g}) d\mathbf{g}. \tag{21}$$

With this marginalization, we can use the channel aware design approach [9] that tends to the "averaged" transmission channel. The difference between this alternative approach and that of minimizing (7) is in the way that the channel fading statistics, $p(\mathbf{g})$, is utilized. In using (7), the variable to be averaged is $P_{e0}(\gamma_0, \ldots, \gamma_K \mid \mathbf{g})$, which is a highly nonlinear function of the decision rules, whereas the alternative approach uses $p(\mathbf{g})$ to obtain the marginalized channel transition probability, thus enabling the direct application of the channel aware approach. This difference can also be explained using Figure 1. The alternative approach averages each transmission channel over respective channel statistics g_i to obtain $p(Y_i \mid U_i)$, while the CSIF case averages all the transmission channels and the fusion center (the part in the dashed box in Figure 1) over channel statistics $\mathbf{g}$ to obtain

$$P(U_0 \mid \mathbf{u}) = \int_{\mathbf{g}} \int_{\mathbf{Y}} P(U_0 \mid \mathbf{y}, \mathbf{g}) p(\mathbf{y} \mid \mathbf{u}, \mathbf{g}) p(\mathbf{g}) d\mathbf{y} \, d\mathbf{g}. \tag{22}$$

It turns out that this alternative approach is a direct consequence of the NOCSIF, as elaborated below.

3.2. The NOCSIF case

Here we consider the case where the fusion center does not know the instantaneous CSI. Therefore, we have

$$P(U_0 = 1 - i \mid \mathbf{y}, \mathbf{g}) = P(U_0 = 1 - i \mid \mathbf{y}). \tag{23}$$

As the fusion rule no longer depends on $\mathbf{g}$, the average error probability in (10) can be rewritten as

$$P_{e0} = \sum_{i=0}^{1} \pi_i \int_{\mathbf{Y}} \sum_{\mathbf{u}} P(U_0 = 1 - i \mid \mathbf{y}) \left(\int_{\mathbf{g}} p(\mathbf{y} \mid \mathbf{u}, \mathbf{g}) p(\mathbf{g}) d\mathbf{g} \right)$$
$$\times \int_{\mathbf{x}} P(\mathbf{u} \mid \mathbf{x}) p(\mathbf{x} \mid H_i) d\mathbf{x} \, d\mathbf{y}, \tag{24}$$

where integration with respect to $\mathbf{g}$ is carried out first. Notice that the term $\left(\int_{\mathbf{g}} p(\mathbf{y} \mid \mathbf{u}, \mathbf{g}) p(\mathbf{g}) d\mathbf{g} \right)$ precisely describes the marginalized transmission channels (cf. (21)). This leads to the standard channel aware design where the transmission channels are characterized by $p(\mathbf{y} \mid \mathbf{u})$. From [9], we have a result resembling that of Theorem 1 except that A_k, B_k, and C are replaced by A_k', B_k', and C',

$$A_k' = \int_{\mathbf{y}} P(U_0 = 1 \mid \mathbf{y}) \left[\sum_{\mathbf{u}^k} \left(p(\mathbf{y} \mid \mathbf{u}^{k1}) - p(\mathbf{y} \mid \mathbf{u}^{k0}) \right) \right.$$
$$\left. \times P(\mathbf{u}^k \mid H_0) \right] d\mathbf{y},$$

$$B_k' = \int_{\mathbf{y}} P(U_0 = 0 \mid \mathbf{y}) \left[\sum_{\mathbf{u}^k} \left(p(\mathbf{y} \mid \mathbf{u}^{k0}) - p(\mathbf{y} \mid \mathbf{u}^{k1}) \right) \right.$$
$$\left. \times P(\mathbf{u}^k \mid H_1) \right] d\mathbf{y},$$

$$C' = \int_{X_k} \left[\sum_{i=0}^{1} \pi_i p(X_k \mid H_i) \int_{\mathbf{Y}} P(U_0 = 1 - i \mid \mathbf{y}) \sum_{\mathbf{u}^k} p(\mathbf{y} \mid \mathbf{u}^{k0}) \right.$$
$$\left. \times p(\mathbf{u}^k \mid H_i) d\mathbf{y} \right] dX_k. \tag{25}$$

Contrary to the CSIF case, A_k', B_k' for the NOCSIF case are much easier to evaluate.

4. PERFORMANCE EVALUATION

In this section, we first use a two-sensor example to evaluate the performance of both the CSIF and NOCSIF cases and compare them with the clairvoyant case where the global channel information is known to the designer. For convenience, we call the clairvoyant case the CSI case. Consider the detection of a known signal S in zero-mean complex Gaussian noises that are independent and identically distributed (i.i.d.) for the two sensors, that is, for $k = 1, 2$,

$$H_0 : X_k = N_k, \qquad H_1 : X_k = S + N_k, \tag{26}$$

with N_1 and N_2 being i.i.d. $\mathcal{CN}(0, \sigma_1^2)$. Without loss of generality, we assume $S = 1$ and $\sigma_1^2 = 2$.

Each sensor makes a binary decision based on its observation X_k,

$$U_k = \gamma_k(X_k), \qquad U_k \in \{0, 1\}, \tag{27}$$

and then transmits it through a Rayleigh fading channel to the fusion center. Notice that $U_k \in \{0, 1\}$ implies an on-off signaling, thus enabling the detection at the fusion center in the absence of CSI (i.e., the NOCSIF case). The channel output is

$$Y_k = g_k U_k + W_k, \tag{28}$$

where g_1, g_2 are independently distributed with zero-mean complex Gaussian distributions $\mathcal{CN}(0, \sigma_{g_1}^2)$ and $\mathcal{CN}(0, \sigma_{g_2}^2)$,
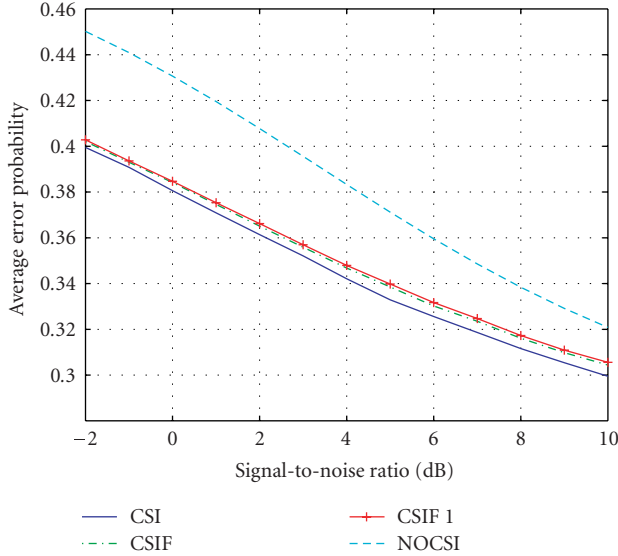
FIGURE 2: Average error probability versus channel SNR for identical channel distribution.

respectively. W_1, W_2 are i.i.d. zero-mean complex Gaussian noises with distribution $\mathcal{CN}(0, \sigma_2^2)$.

We first consider a symmetric case where the CSI is identically distributed, that is, $\sigma_{g_1^2} = \sigma_{g_2^2}$. Without loss of generality, we assume $\sigma_{g_1^2} = \sigma_{g_2^2} = 1$. In Figure 2, with the assumption of equal prior probability, the average error probabilities as a function of the average signal-to-noise ratio (SNR) of the received signal at the fusion center are plotted for the CSIF and NOCSIF cases, along with the CSI case. We also plot a curve, legended with "CSIF1" in Figure 2, where the local sensors use thresholds obtained via the NOCSIF approach but the fusion center implements a fusion rule that utilizes the CSI $\mathbf{g}$. The motivation is twofold. First, estimating $\mathbf{g}$ at the fusion center is typically feasible. Second and more importantly, the threshold design for NOCSIF is much simpler compared with CSIF, as explained in Section 3.1.

In evaluating the performance of CSIF case, we consider the following two methods to get the optimal thresholds for local sensors.

Exhaustive search method

We first partition the space of likelihood ratio of observations into many small disjoint cells. For each cell, we compute the average error probability at the fusion center by using the center point of the cell (for small enough cell size, the center point can be considered "representative" of the whole cell) as the local thresholds. After evaluating the performance for all the cells, we choose the one with smallest average error probability as our optimal thresholds. Intuitively, one can get arbitrarily close to the optimal thresholds by decreasing the cell size, which proportionally increases the computational complexity.

Greedy search

(1) choose initial thresholds $\mathbf{t}^{(0)}$, for example, the thresholds obtained via the NOCSIF case and set $r = 0$ (iteration index);
(2) compute the average error probability using $\mathbf{t}^{(0)}$ as local thresholds;
(3) choose several directions. For example, for two-sensor case, we can choose $(1, 0)$, $(-1, 0)$, $(0, 1)$, and $(0, -1)$ as our directions;
(4) choose a small stepsize ϵ;
(5) move the current thresholds $\mathbf{t}^{(r)}$ one stepsize along each direction and compute the average error probability using the new thresholds;
(6) compare the error probability of using current thresholds $\mathbf{t}^{(r)}$ and new thresholds and assign the thresholds corresponding to the smallest one as $\mathbf{t}^{(r+1)}$;
(7) if $\mathbf{t}^{(r+1)} = \mathbf{t}^{(r)}$, stop; otherwise, set $r = r + 1$ and go to step (5).

Greedy search method is much faster than exhaustive search method since we do not need to compute the average error probability corresponding to every point. But it can only guarantee convergence to a local minimum point.

As seen from Figure 2, the CSI case has the best performance and NOCSIF case has the worst performance. This is true because the designer has the most information in the clairvoyant case and the least information in NOCSIF case. The CSIF case is only slightly worse than the CSI case but is much better than the NOCSIF case. The difference of CSIF1 and CSIF is almost indistinguishable. The explanation is that the performance is much more sensitive to the fusion rule than to the local sensor thresholds. This phenomenon has been observed before: the error probability versus threshold plot is rather flat near the optimum point, hence is robust to small changes in thresholds.

We also consider an asymmetric case where $\sigma_{g_1^2} \neq \sigma_{g_2^2}$. As plotted in Figure 3 where $\sigma_{g_1^2} = 1$ and $\sigma_{g_2^2} = 3$, all three cases with known CSI at the fusion center have similar performance and are much better than the NOCSIF case. This is consistent with the symmetric case.

As we stated above, for the system with only two local sensors, a mixed approach (CSIF1) achieves almost same performance to that of the CSIF case. In the following, we show that the same holds true even in the large system regime, that is, as K increasing. We first consider the Bayesian framework. As K goes to infinity, all the local sensors use the same local decision rule [16] and the optimal local thresholds are determined by maximizing the Chernoff information to achieve the best error exponent,

$$C(p(Y \mid H_0), p(Y \mid H_1))$$
$$= -\min_{0 \le s \le 1} \log \left[\int p(Y \mid H_0)^s p(Y \mid H_1)^{1-s} dY \right]. \tag{29}$$

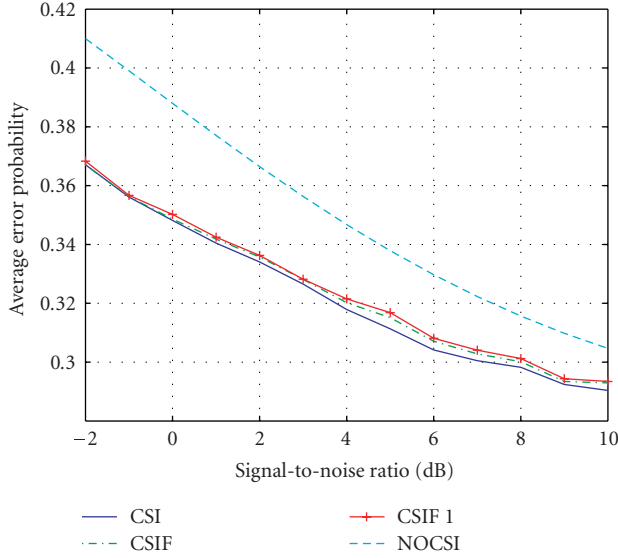For simplicity, we consider another performance measure, Bhattacharyya's distance, which is an approximation of

FIGURE 3: Average error probability versus channel SNR for non-identical channel distribution.
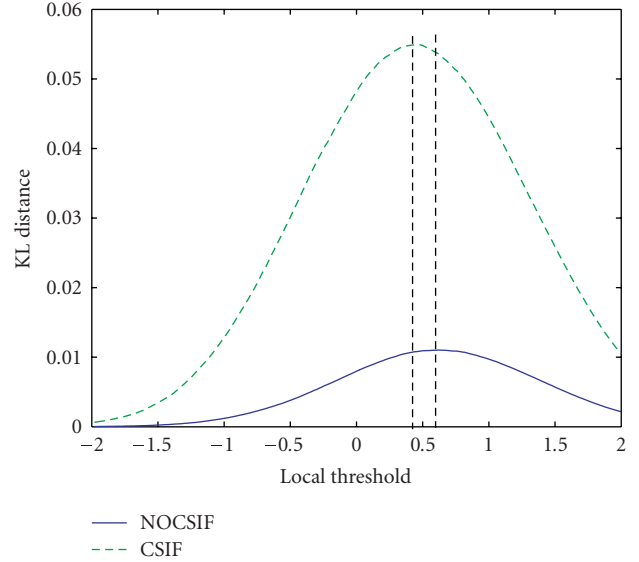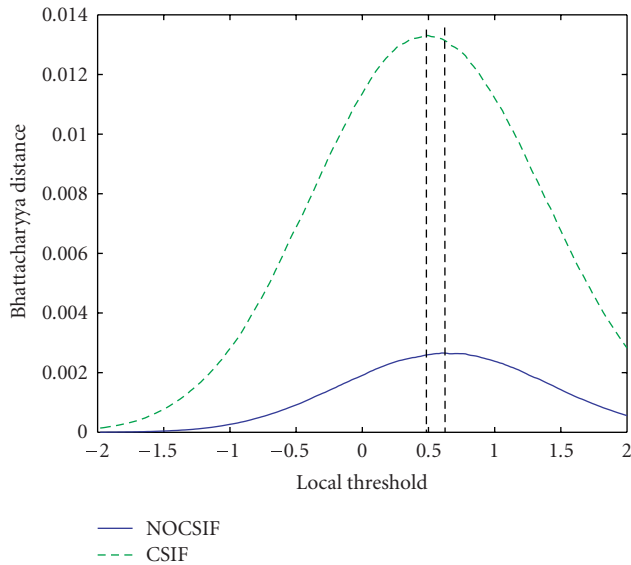

FIGURE 4: Bhattacharyya's distance versus the local threshold at the local sensors.

Chernoff information by setting $s = 1/2$,

$$
\begin{aligned}
&B(p(Y \mid H_0), p(Y \mid H_1)) \\
&= -\log\left[\int p(Y \mid H_0)^{1/2} p(Y \mid H_1)^{1/2} dY \right] \\
&= -\log\left[\int \left(\frac{p(Y \mid H_0)}{p(Y \mid H_1)} \right)^{1/2} p(Y \mid H_1) dY \right].
\end{aligned}
\tag{30}
$$

From Figure 4, which gives Bhattacharyya's distance as a function of local threshold for both CSIF and NOCSIF cases

FIGURE 5: KL distance versus the local threshold at the local sensors.

with $\sigma_2^2 = 2$ and $\sigma_g^2 = 1$, we can observe that the optimal threshold obtained in NOCSIF case is close to that of the CSIF case.

Alternatively, under the Neyman-Pearson framework, the Kullack-Leibler (KL) distance gives the asymptotic error exponent,

$$
\begin{aligned}
&KL(p(Y \mid H_0), p(Y \mid H_1)) \\
&= E_{H_0}\left[\log\left(\frac{p(Y \mid H_0)}{p(Y \mid H_1)} \right) \right] \\
&= \int p(Y \mid H_0) \log\left(\frac{p(Y \mid H_0)}{p(Y \mid H_1)} \right) dY.
\end{aligned}
\tag{31}
$$

In Figure 5, with the same setting as in Figure 4, the KL distance as a function of local threshold for both CSIF and NOCSIF cases is plotted and we can also observe that the optimal local thresholds obtained in NOCSIF and CSIF cases are close to each other.

5. CONCLUSIONS

In this paper we investigated the design of the distributed detection problem with only channel fading statistics available to the designer. Restricted to conditional independent observations and binary local sensor decisions, we derive the necessary conditions for optimal local sensor decision rules that minimize the average error probability for the CSIF and NOCSIF cases. Numerical results indicate that a mixed approach where the sensors use the decision rules from the NOCSIF approach while the fusion center implements a fusion rule using the CSI achieves almost identical performance to that of the CSIF case.

APPENDICES

A. PROOF OF $A_K > 0$ AND $B_K > 0$

$P(H_1 \mid \mathbf{y}^k, U_k)$

$$= \frac{p(H_1, \mathbf{y}^k, U_k)}{p(\mathbf{y}^k, U_k)}$$

$$= \frac{p(\mathbf{y}^k, U_k \mid H_1)P(H_1)}{p(\mathbf{y}^k, U_k \mid H_0)P(H_0) + p(\mathbf{y}^k, U_k \mid H_1)P(H_1)}$$

$$= \frac{\pi_1 p(\mathbf{y}^k \mid U_k, H_1)P(U_k \mid H_1)}{\pi_0 p(\mathbf{y}^k \mid U_k, H_0)P(U_k \mid H_0) + \pi_1 p(\mathbf{y}^k \mid U_k, H_1)P(U_k \mid H_1)}.$$
(A.1)

Since the observations, $X_1, \ldots, X_K$, are conditionally independent and the local decision in each sensor depends only on its own observation, the local decisions are also conditionally independent for a given hypothesis. In addition, the local decisions are transmitted through orthogonal channels, thus the channel output for one sensor is conditionally independent to the channel input from another sensor. Therefore,

$$p(\mathbf{y}^k \mid U_k, H_1) = p(\mathbf{y}^k \mid H_1).$$
(A.2)

Defining the likelihood ratio function for the local decision at the kth sensor,

$$L(U_k) \triangleq \frac{P(U_k \mid H_1)}{P(U_k \mid H_0)}.$$
(A.3)

Then

$P(H_1 \mid \mathbf{y}^k, U_k)$

$$= \frac{\pi_1 p(\mathbf{y}^k \mid H_1)P(U_k \mid H_1)}{\pi_0 p(\mathbf{y}^k \mid H_0)P(U_k \mid H_0) + \pi_1 p(\mathbf{y}^k \mid H_1)P(U_k \mid H_1)}$$

$$= \frac{\pi_1 p(\mathbf{y}^k \mid H_1)L(U_k)}{\pi_0 p(\mathbf{y}^k \mid H_0) + \pi_1 p(\mathbf{y}^k \mid H_1)L(U_k)}$$
(A.4)

which is a monotone increasing function of $L(U_k)$. Then, given a monotone LR index assignment for the local output, that is, $L(U_k = 1) > L(U_k = 0)$, $P(H_1 \mid \mathbf{y}^k, U_k)$ is a monotone increasing function of U_k. Thus,

$$P(H_1 \mid \mathbf{y}^k, U_k = 0) < P(H_1 \mid \mathbf{y}^k, U_k = 1).$$
(A.5)

Similarly, we can get

$$P(H_0 \mid \mathbf{y}^k, U_k = 0) > P(H_0 \mid \mathbf{y}^k, U_k = 1).$$
(A.6)

The optimum fusion rule is a maximum a posteriori rule (i.e., to minimize error probability). Thus deciding $U_0 = 1$ for a given $\mathbf{y}^k$ and $U_k = 0$ requires

$$P(H_1 \mid \mathbf{y}^k, U_k = 0) > P(H_0 \mid \mathbf{y}^k, U_k = 0).$$
(A.7)

From (A.5) and (A.6), (A.7) implies

$$P(H_1 \mid \mathbf{y}^k, U_k = 1) > P(H_0 \mid \mathbf{y}^k, U_k = 1).$$
(A.8)

Therefore, we must have $U_0 = 1$ for the same $\mathbf{y}^k$ with $U_k = 1$. Thus

$$P(U_0 = 1 \mid \mathbf{y}^k, U_k = 0) < P(U_0 = 1 \mid \mathbf{y}^k, U_k = 1)$$
(A.9)

and further,

$$A_k > 0.$$
(A.10)

Similarly, one can show

$$P(U_0 = 0 \mid \mathbf{y}^k, U_k = 0) > P(U_0 = 0 \mid \mathbf{y}^k, U_k = 1)$$
(A.11)

and further,

$$B_k > 0.$$
(A.12)

B. ALTERNATIVE DERIVATION FOR OPTIMAL LOCAL DECISION RULES

From [7, Proposition 2.2], with a cost function $F : \{0, 1\} \times \mathbb{Z} \times \{H_0, H_1\} \to \mathcal{R}$ and a random variable $\mathbf{z}$ which takes values in the set $\mathbb{Z}$ and is independent of X conditioned on any hypothesis, the optimal decision rule that minimizes the cost $E[F(\gamma(X), \mathbf{z}, H)]$ is

$$\gamma(X) = \arg \min_{U=0,1} \sum_{j=0}^{1} P(H_j \mid X)\alpha(H_j, U),$$
(B.1)

where

$$\alpha(H_j, d) = E[F(U, \mathbf{z}, H_j) \mid H_j].$$
(B.2)

Assuming the cost that we want to minimize at the fusion center can be written as

$$J = E[C(\gamma_0(\mathbf{y}, \mathbf{g}), \mathbf{u}, H)]$$
(B.3)

and we further have

$$J = E[C(\gamma_0(\mathbf{y}, \mathbf{g}), \mathbf{u}, H)]$$

$$= E[C(\gamma_0(Y_k, \mathbf{y}^k, g_k, \mathbf{g}^k), \gamma_k(X_k), \mathbf{u}^k, H)]$$

$$= E[E[C(\gamma_0(Y_k, \mathbf{y}^k, g_k, \mathbf{g}^k), \gamma_k(X_k), \mathbf{u}^k, H) \mid \mathbf{u}^k, \mathbf{y}^k, g_k, \mathbf{g}^k, X_k, H]]$$

$$= E[E[C(\gamma_0(Y_k, \mathbf{y}^k, g_k, \mathbf{g}^k), \gamma_k(X_k), \mathbf{u}^k, H) \mid X_k, g_k]]$$

$$= E\left[\int_{Y_k} C(\gamma_0(Y_k, \mathbf{y}^k, g_k, \mathbf{g}^k), \gamma_k(X_k), \mathbf{u}^k, H) \times p(Y_k \mid \gamma_k(X_k), g_k)dY_k\right],$$
(B.4)

where $\mathbf{g}^k = [g_1, \ldots, g_{k-1}, g_{k+1}, g_K]$ and (B.4) follows that conditioned on $\gamma_k(X_k)$ and g_k, the channel output Y_k is independent of everything else.

Defining $X = X_k$, $\mathbf{z} = (\mathbf{y}^k, \mathbf{u}^k, \mathbf{g})$,

$$F(U_k, \mathbf{z}, H)$$
$$= \int_{Y_k} C(\gamma_0(Y_k, \mathbf{y}^k, \mathbf{g}), U_k, \mathbf{u}^k, H) p(Y_k \mid U_k, g_k) dY_k \tag{B.5}$$

and substituting them into (B.1), we obtain the optimal local decision rule for kth sensor:

$$\gamma_k(X_k) = \arg \min_{U_k=0,1} \sum_{j=0}^{1} P(H_j \mid X_k) \alpha_k(H_j, U_k), \tag{B.6}$$

where

$$\alpha_k(H_j, U_k)$$
$$= E\left[\int_{Y_k} C(\gamma_0(Y_k, \mathbf{y}^k, \mathbf{g}), U_k, \mathbf{u}^k, H_j) p(Y_k \mid U_k, g_k) dY_k \mid H_j \right]. \tag{B.7}$$

Since

$$P(H_j \mid X_k) = \frac{p(X_k \mid H_j)\pi_j}{\sum_{i=0}^{1} p(X_k \mid H_i)\pi_i}, \tag{B.8}$$

(B.6) is equivalent to

$$\gamma_k(X_k) = \arg \min_{U_k=0,1} \sum_{j=0}^{1} \pi_j p(X_k \mid H_j) \alpha_k(H_j, U_k)$$
$$= \arg \min_{U_k=0,1} \sum_{j=0}^{1} p(X_k \mid H_j) b_k(H_j, U_k), \tag{B.9}$$

where

$$b_k(H_j, U_k) = \pi_j \alpha_k(H_j, U_k). \tag{B.10}$$

Thus, the optimal local decision rule for the kth sensor is

$$\gamma_k(X_k) = \begin{cases} 0, & P(X_k \mid H_0)(b_k(H_0,1) - b_k(H_0,0)) \\ & \quad \geq P(X_k \mid H_1)(b_k(H_1,0) - b_k(H_1,1)), \\ 1, & \text{otherwise.} \end{cases} \tag{B.11}$$

In the approach proposed in this paper, the objective is to minimize the average error probability at the fusion center,

$$J = \int_{\mathbf{g}} P(\gamma_0(\mathbf{y}, \mathbf{g}) \neq H) p(\mathbf{g}) d\mathbf{g}. \tag{B.12}$$

Thus we have

$$\alpha_k(H_j, U_k) = \int_{\mathbf{y}} \int_{\mathbf{g}} P(U_0 = 1 - j \mid \mathbf{y}, \mathbf{g})$$
$$\times \left[\sum_{\mathbf{u}^k} p(\mathbf{y} \mid \mathbf{u}^k, U_k, \mathbf{g}) p(\mathbf{g}) P(\mathbf{u}^k \mid H_j) \right] d\mathbf{g}\, d\mathbf{y}. \tag{B.13}$$

It is straightforward to see that $(\alpha_k(H_0,1) - \alpha_k(H_0,0))$ is the same as A_k in (14) and $(\alpha_k(H_1,0) - \alpha_k(H_1,1))$ is the same as B_k in (15). Therefore, (B.11) is equivalent to (20) in Theorem 1.

REFERENCES

[1] R. Radner, "Team decision problems," *Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[2] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, no. 4, pp. 501–510, 1981.

[3] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, NY, USA, 1997.

[4] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 23, no. 1, pp. 24–30, 1987.

[5] I. Y. Hoballah and P. K. Varshney, "Distributed Bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 995–1000, 1989.

[6] S. C. A. Thomopoulos, R. Viswanathan, and D. K. Bougoulias, "Optimal distributed decision fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 5, pp. 761–765, 1989.

[7] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds., pp. 297–344, JAI Press, Greenwich, Conn, USA, 1993.

[8] T. M. Duman and M. Salehi, "Decentralized detection over multiple-access channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 469–476, 1998.

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

[10] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3349–3358, 2006.

[11] A. Kashyap, "Comments on "on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels"," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1274–1275, 2006.

[12] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, 1985.

[13] P. K. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad, and ugly: distributed detection of a known signal in dependent gaussian noise," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3266–3279, 2000.

[14] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Fusion of decisions transmitted over fading channels in wireless sensor networks," in *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1184–1188, Pacific Grove, Calif, USA, November 2002.

[15] R. Niu, B. Chen, and P. K. Varshney, "Decision fusion rules in wireless sensor networks using fading channel statistics," in *Proceedings of the 37th Annual Conference on Information Sciences and Systems (CISS '03)*, Baltimore, Md, USA, March 2003.

[16] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 550–558, 1993.