# Channel-aware distributed classification in wireless sensor networks using binary local decisions

Mohammad Fanaei, Matthew C. Valenti, Natalia A. Schmid, and Vinodkrishnan Kulathumani

Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, WV, USA

## ABSTRACT

This paper considers the problem of distributed multi-hypothesis classification in the context of wireless sensor networks. The goal is to reliably classify an underlying hypothesis at a fusion center using simple localized decisions at individual sensors. The fusion-center classification must be performed despite the presence of faults in both local sensor decisions and transmission channels between the sensors and fusion center. Local sensor nodes make binary classifications based on their noisy observations and send their decisions to the fusion center through parallel additive white Gaussian noise channels. The fusion center then uses these noisy versions of local decisions to perform a global classification. In contrast with other similar approaches for multi-hypothesis classification based on combined binary decisions, our approach exploits the relationship between the influence fields of different hypotheses and the accumulated noisy versions of local binary decisions as received by the fusion center, where the influence field of a hypothesis is defined to be the spatial region in its surrounding in which it can be sensed using some specific modality. The main contribution of this paper is the formulation of local and fusion decision rules that maximize the probability of correct global classification at the fusion center, along with an algorithm for channel-aware global optimization of the local and fusion center decision thresholds. The performance of the proposed classification system is investigated through practical scenarios. Performance analysis results show that the proposed approach could simplify decision making at local sensors while achieving acceptable performance in terms of the global probability of correct classification at the fusion center.

**Keywords:** Distributed detection and classification, $M$-ary hypothesis testing, channel-aware classification, binary local decisions, influence field, fusion center, wireless sensor networks.

## 1. INTRODUCTION

Wireless sensor networks (WSNs) are generally formed by a large number of densely-deployed sensors with limited capabilities that cooperate with each other to achieve a common goal. One of the most important applications of such networks is distributed detection and classification of an object, event, or some phenomenon, also called here an underlying hypothesis, which is the first step in a wider range of applications such as tracking, identification, and parameter estimation[1]. In a WSN performing distributed detection and classification, distributed local sensors observe the conditions of their surrounding environment, process their local observations, and send their processed data to a fusion center, which then makes the ultimate global decision. Different aspects of this problem have attracted a lot of interest in the research community throughout the last three decades. In this paper, we investigate the problem of classifying an underlying hypothesis at the fusion center of a WSN using local binary decisions received from geographically distributed sensors through impaired channels. This problem is formulated and solved based on the differences in the *influence fields* of different hypotheses, where the influence field of a hypothesis is defined to be the spatial region in its surrounding in which it can be sensed using some specific modality[2].

In the realm of distributed detection and classification in WSNs, most of the attention has been given to the binary hypothesis testing problem in which the fusion center is designed to detect the presence or absence of an underlying hypothesis based on local binary decisions received from distributed sensors. In recent years, this problem has been considered for a practical case of non-ideal channels between local sensors and the fusion

E-mail: mfanaei@mix.wvu.edu, {Matthew.Valenti, Natalia.Schmid, and Vinod.Kulathumani}@mail.wvu.edu.

center in which the decisions of local sensors are not reliably received at the fusion center. See Chamberland and Veeravalli[3], Chen et al.[4], and references therein for a survey on recent developments in this field.

The problem of testing $M$ hypotheses using sensory data in WSNs has been investigated in some capacity (see, for instance[5]). In general, decisions made by the local sensors in this case are sent to the fusion center using at least $\lceil \log_2 M \rceil$ information bits where $M$ is the number of hypotheses to be classified. However, two main constraints of WSNs make this approach undesirable: First, the processing power of the local sensors is limited. Therefore, they may not be able to distinguish between different hypotheses. Second, the bandwidth and energy resources of WSNs are limited. Therefore, it is desired to send the local sensor decisions to the fusion center with as few bits as possible. These observations and requirements motivated us to design a distributed $M$-ary classification WSN in which the local sensors make binary (rather than $M$-ary) local decisions and send them to the fusion center. The fusion center then uses the local decisions collectively and makes a global inference about the underlying hypothesis based on the known influence fields of different hypotheses.

There has been some recent work in the literature that has investigated distributed $M$-ary hypothesis testing in WSNs using local binary decisions. In Wang et al.[6], a fault-tolerant distributed multi-hypothesis classification fusion approach is proposed based on binary error correcting codes. In this approach, an error-correcting code matrix is designed in which each row forms a codeword that corresponds to one of the $M$ hypotheses to be classified. Moreover, each column of the code matrix corresponds to the binary decision rule of the corresponding local sensor. Each local sensor makes its binary decision based on the corresponding column of the designed code matrix and sends it to the fusion center through parallel channels. More precisely, when sensor $i$ detects hypothesis $H_j$, it sends the binary element in the $j$th row and $i$th column of the code matrix to the fusion center. The fusion center then makes a final $M$-ary decision on the underlying hypothesis based on the binary received local decisions using the minimum Hamming distance decoding criterion, where the Hamming distance between two binary vectors is defined as the number of distinct positions between the vectors. The performance of this multi-hypothesis classification WSN depends on the minimum Hamming distance of the designed code matrix. Note that in the classification algorithm proposed by Wang et al.[6], local sensors still need to make an $M$-ary classification. Having made that classification, each sensor sends a binary decision to the fusion center. Therefore, this approach addresses the constraints of WSNs related to limited bandwidth and energy resources. However, it does not alleviate the requirement of high processing capability at local sensors.

The approach proposed in Wang et al.[6] does not consider the impact of fading channels between distributed local sensors and the fusion center. In fact, in this approach the channels are assumed to be binary symmetric channels. This weakness has been addressed in Wang et al.[7], which has a similar problem statement, but a different decoding rule is devised that is robust to flat fading channels with phase coherent reception at the fusion center. Another proposed enhancement of[7] compared to[6] is that it allows the local sensors to send multi-level $D$-ary (rather than binary) decisions to the fusion center, if needed, while the fusion center still uses a fixed binary code matrix for all values of $D$. The fusion center in this architecture uses a soft-decision decoding rule to measure the distance between the received multi-level local decision vector and the codeword in the given binary code matrix. It is shown in[7] that when more bits of local decision information are sent to the fusion center, the classification performance can be improved while the total energy output from each sensor is fixed. In Wang et al.[8], the ideas presented in[7] for binary code matrix are extended by using a $D$-ary code matrix with $D > 2$ when $\log_2 D$ bits of local decision information are used at the fusion center. In Pai et al.[9], the approach presented in[7] is further refined in a multiple-observation scenario while keeping the sensor complexity low. In this two-dimensional $M$-ary coded classification scheme, each sensor makes $D$ (rather than one) independent observations and then sends $D$ bits as the result of its local decisions for $D$ observations (one bit for each observation rather than $D$ bits for one observation as in[7]). Each hypothesis is then represented by a two-dimensional codeword and the binary code matrix becomes three dimensional.

In Zhu et al.[10], the problem of $M$-ary hypothesis classification in WSNs using binary local decisions is solved through modelling each local sensor by a set of $M$ transition probabilities that specifies the probability that the sensor sends 1 to the fusion center for different underlying hypotheses. Moreover, the fusion center is modelled by a set of $M$ conditional misclassification probabilities, given any hypothesis. The authors of[10] have developed conditions for which the average probability of misclassification at the fusion center asymptotically goes to zero as the number of local sensors goes to infinity. Moreover, they have used a genetic algorithm-based approach to find the optimum local decision thresholds.

Zhang and Varshney[11] have considered the fusion of binary decision tree classifiers in a multi-hypothesis classification WSN. Binary decision trees make a sequence of binary decisions in a hierarchical manner and are easier to design. They are also more efficient. In[11], this hierarchical tree structure is used to break the complex $M$-ary hypothesis testing problem into a set of much simpler binary decision fusion problems. Each sensor uses a binary decision tree to make its decision and sends it to the fusion center through an ideal communication channel. The fusion center then combines the local decisions to make the global inference about the underlying hypothesis. Since each set of received local decisions corresponds to a unique path from the root node to a terminal node of the binary decision tree, it can be encoded as a sequence of binary decisions made by all the sensors in the corresponding path. Detailed analysis of designing the binary decision tree for local sensors and fusion center, designing the decision rules at the internal nodes of sensor binary decision trees, designing the optimum fusion rule, and designing the system topology including communication structure of the WSN has been presented in[11].

In all of the aforementioned references, the conditional observations at different local sensors, given any underlying hypothesis, are assumed to be independent. In Nguyen et al.[12], the problem of decentralized Bayesian detection in WSNs with $M$ hypotheses is considered when local sensors make conditionally dependent observations. Each sensor is modelled as a quantizer that makes a $D$-ary decision based on its observation and sends it to a fusion center through an ideal communication channel. The fusion center then makes a decision on the actual hypothesis based on the local decisions it receives from distributed sensors so that the average probability of misclassification is minimized. It is shown that, due to the conditional dependence between sensor observations, the threshold based decision rules (or likelihood ratios) at local sensors are no longer optimal. The same problem has been considered in a more general form in Tang et al.[13] with conditionally correlated observations, given any underlying hypothesis, perfect communication channels between local sensors and fusion center, and $D$-ary local decisions. The person-by-person optimization (PBPO) algorithm has been used to optimize the local sensor and fusion center decision rules iteratively.

In some applications involving distributed $M$-ary hypothesis testing in WSNs, local knowledge of sensors may not be sufficient for making an $M$-ary decision or it may be very costly to have sensors capable of doing such a classification. As an example, consider a surveillance system consisting of a densely-deployed sensor network whose ultimate goal is to detect and classify an intruder, which can be an armed soldier, a car, or a tank. Suppose that the sensors are simple magnetometers that can only measure the strength of a magnetic field in their limited surrounding region. Since all three hypotheses can have the same magnetic field at a local sensor location, the sensors may not be able to distinguish between these hypotheses based on only their local observations. In other words, they can only detect the existence of a magnetic field in their surrounding, i.e. a local binary hypothesis problem. On the other hand, if the fusion center has access to local binary decisions made at all of the distributed sensors, it can make a global inference about the underlying hypothesis based on, for example, the number of sensors that have detected a magnetic field of one of the three hypotheses. More precisely, the number (and possibly location) of sensors that detect the presence of magnetic field of an object determines the area of coverage of the object's influence field, and can be considered as a measure of the area associated with the influence field of the underlying hypothesis. This intuition motivated our work to design a distributed multi-hypothesis classification strategy for WSNs that uses binary local inferences to make its final optimal decision based on the knowledge of the influence fields of different underlying hypotheses.

Bapat et al.[14] have previously used the idea of influence fields for classification of objects in a large scale sensor network. The objective of[14] was to obtain requirements on the underlying sensor network density to ensure accurate classification in the presence of false decisions at local sensors, channel fading and channel contention. On the other hand, the objective of this paper is to obtain conditions on local and fusion center decision thresholds in the presence of noisy observations that maximize classification accuracy at the fusion center for a given density of sensor deployment.

The rest of this paper is organized as follows: Section 2 describes the model of the distributed parallel fusion WSN that we will consider in our analysis. In Section 3, the system is analyzed, and the optimum parameters of a specifically-defined fusion center decision rule are derived. Moreover, different methods of local versus global

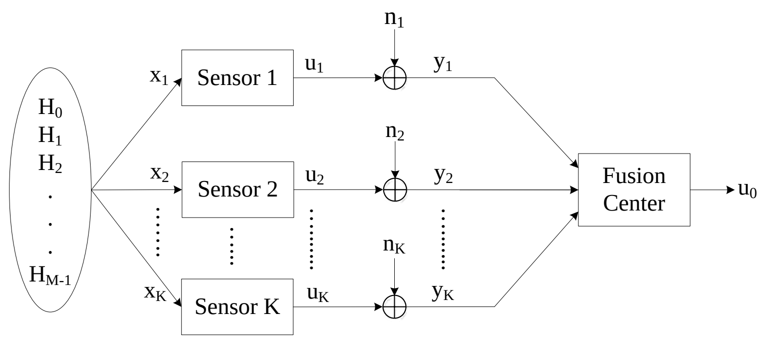


Figure 1. System model of the proposed multi-hypothesis classification system.

optimization of local decision rules are discussed. Section 4 presents the numerical results of the analytical performance evaluation of the proposed classification system and studies the effects of different parameters of the classification network on its performance. Finally, in Section 5 we conclude our discussions and summarize the main achievements of this work.

## 2. SYSTEM MODEL

Consider a WSN deployed as a parallel distributed classification system as shown in Fig. 1. The system is formed by $K$ sensors distributed in an environment covering an area $S$, and a fusion center. There are $M$ independent and mutually exclusive hypotheses $H_0$, $H_1$, ..., and $H_{M-1}$, $M \geq 2$, with the following known prior probabilities:

$$p_j \quad = \quad P[H = H_j], \;\; j = 0, 1, \ldots, M-1, \tag{1}$$

where $H$ is a random variable representing the underlying hypothesis. Note that $H_0$ is the null or rejection hypothesis and its existence means that none of the other $M-1$ hypotheses has occurred. Each non-null hypothesis is associated with a known *influence field* defined as the spatial region in its surrounding in which it can be sensed using some specific modality[2]. As an example, suppose that the sensors are simple magnetometers and the non-null hypotheses define the presence of a car or a tank. The regions in which the car or tank can be sensed by the magnetometers are called their influence fields. The influence field of hypothesis $H_j$ is denoted by $A_j$, $j = 1, 2, \ldots, M-1$. It is assumed that the entire influence field of the underlying hypothesis is inside the sensing area $S$. If the sensors are distributed uniformly within the sensing area, the average number of sensors that can be placed in the influence field of hypothesis $H_j$ will be $K_j = \lfloor \frac{A_j}{S} K \rfloor$. Throughout this paper, we assume that the center of the influence field of the underlying hypothesis is known or has been reliably estimated*. Assuming that the center of the underlying influence field is known, for each sensor $i$ the set of hypotheses is divided into two disjoint subsets: the set of hypotheses that sensor $i$ can be inside their influence fields, denoted by $\mathcal{C}_i^1$, and the set of hypotheses that sensor $i$ cannot be inside their influence fields, denoted by $\mathcal{C}_i^0$. On the other hand, assuming uniform sensor distribution within the sensing environment, for each underlying hypothesis $H_j$, on average there are $K_j$ sensors that can be inside its influence field and $K - K_j$ sensors that cannot.

Let $\mathbf{x} = [x_1, x_2, \ldots, x_K]$ be the vector of local sensor observations. It is assumed that the conditional observations of different sensors, given any specific underlying hypothesis, are independent. In other words,

$$p(\mathbf{x}|H_j) \quad = \quad \prod_{i=1}^{K} p(x_i|H_j), \;\; j = 0, 1, \ldots, M-1. \tag{2}$$

Throughout this paper, we assume that the conditional observation of each sensor $i$, given any hypothesis $H_j$, is modelled as

$$H_j : \;\; x_i \quad = \quad \begin{cases} v_i, & \text{if } H_j \in \mathcal{C}_i^0 \\ s + v_i, & \text{if } H_j \in \mathcal{C}_i^1 \end{cases}, \;\; j = 0, 1, \ldots, M-1, \; i = 1, 2, \ldots, K, \tag{3}$$

*For more information on distributed estimation in WSNs, an interested reader is referred to[15] and references therein.

where $v_i$'s are samples of zero-mean white Gaussian noise with variance $\sigma_o^2$, i.e. $v_i \sim \mathcal{N}(0, \sigma_o^2)$, $i = 1, 2, \ldots, K$. Noise samples are independent and identically distributed (i.i.d.). More specifically, if the sensor can be inside the influence field characterizing hypothesis $H_j$, it observes a noisy version of the *constant* strength of the influence field, $s$. Otherwise, it observes only noise. Note that this model implies that the strength of the influence field of all non-null hypotheses is assumed to be the same and constant over the entire influence field. This assumption makes analysis tractable. Furthermore, it is valid in a lot of applications such as the one mentioned at the beginning of this section.

Each local sensor makes a binary decision based on its sensory data. To be specific, assume that the local decision of any sensor, $u_i$, $i = 1, 2, \ldots, K$, is made based on a local binary decision rule as

$$u_i = \gamma_i(x_i) = \begin{cases} 0, & \text{if } x_i < \beta_i \\ 1, & \text{if } x_i > \beta_i \end{cases}, \tag{4}$$

where $\beta_i$ is the optimal local decision threshold for sensor $i$. Note that this decision rule might not be the optimal local decision rule for our distributed classification system. However, it is very simple and allows us to achieve an acceptable performance in terms of probability of correct classification at the fusion center without requiring the sensors to be able to distinguish between different hypotheses. In other words, the sensors are able to distinguish only the occurrence or not occurrence of the $M-1$ non-null hypotheses and it is the fusion center that makes the final $M$-ary decision based on the accumulated local binary decisions, from the sensors. This local decision rule has two main advantages addressing the stringent processing capability and bandwidth limitations of WSNs. The first advantage is that the sensors do not need to differentiate between $M-1$ non-null hypotheses, and hence their required local processing is very limited. The second advantage is that the transmission of the local decisions to the fusion center can be done using a binary scheme, and hence the bandwidth required for this communication will be limited.

Let $\mathbf{u} = [u_1, u_2, \ldots, u_K]$ be the vector of binary decisions made by the local sensors. Each $u_i$ is communicated to the fusion center over an additive white Gaussian noise (AWGN) channel resulting in a noisy output $y_i$. Channels between different sensors and the fusion center are parallel. The input to the fusion center is a vector of noisy local decisions, $\mathbf{y} = [y_1, y_2, \ldots, y_K]$, where each one of its entries,

$$y_i = u_i + n_i, \quad i = 1, 2, \ldots, K, \tag{5}$$

is a decision made by sensor $i$ and observed by the fusion center. Additive noise, $n_i$, in each of the $K$ parallel channels is assumed to be zero-mean independent and identically distributed Gaussian with variance $\sigma_n^2$, i.e. $n_i \sim \mathcal{N}(0, \sigma_n^2)$, $i = 1, 2, \ldots, K$.

The fusion center has to make the final $M$-ary decision, $u_0$, about the underlying hypothesis by using noisy versions of the local binary decisions from distributed sensors. In other words,

$$u_0 = \gamma_0(\mathbf{y}) \in \{0, 1, \ldots, M-1\}, \tag{6}$$

where $\gamma_0(.)$ is a multi-variate function. In the next section, we propose a simple yet powerful fusion center decision rule and analyze the performance of the proposed $M$-ary classification system.

## 3. FUSION RULE DERIVATION

Suppose that the receiver at the fusion center is designed to add all received $y_i$'s, $i = 1, 2, \ldots, K$, and form a decision metric (or test statistic) $S_0 = \sum_{i=1}^{K} y_i$ based on which the final $M$-ary decision is made. Note that the final decision metric is sought in the form of linear combination of noisy local decisions. Other more complicated decision metrics can be considered but this linear rule has the advantages of being simple and computationally efficient. It can be observed that $S_0$ is an appropriate yet simple decision metric, which captures differences in the influence fields of different hypotheses. In other words, it is intuitive that $S_0$ tends to have larger values if the influence field of the underlying hypothesis is larger, since more sensors are in the underlying influence field, and therefore have made $u_i = 1$ as their decisions. If the influence fields of different hypotheses have enough separation and appropriate thresholds are found for the values of $S_0$ associated with different hypotheses, the system can achieve an acceptable performance in terms of probability of correct classification at the fusion center.

Assume that $\Gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_{M-1}\}$ is the set of decision thresholds based on which the fusion center classifies the underlying hypothesis using $S_0$ as its decision metric. In other words, assume that the fusion center's decision rule is

$$u_0 \;=\; j \text{ if and only if } \alpha_j \leq S_0 < \alpha_{j+1}, \;\; j = 0, 1, \ldots, M-1, \tag{7}$$

where $\alpha_0 = -\infty$ and $\alpha_M = \infty$. The optimum values for this set of decision thresholds at the fusion center are derived in this section so that the maximum probability of correct classification at the fusion center can be achieved. Moreover, the effect of channel-aware global optimization of local sensor decision thresholds $\beta_i$, $i = 1, 2, \ldots, K$, on the probability of correct classification at the fusion center is examined.

Based on equation (3) as a model for the conditional local observation of sensor $i$, $x_i$, and equation (4) as the sensor's local binary decision rule, it can be shown that the conditional probability density function (pdf) of the $i$th local sensor's decision, $u_i$, given hypothesis $H_j$, is

$$f_{U_i|H_j}(u_i|H_j) \;=\; \begin{cases} \left\{1 - Q\left(\frac{\beta_i}{\sigma_O}\right)\right\}\delta[u_i] + Q\left(\frac{\beta_i}{\sigma_O}\right)\delta[u_i - 1], & \text{if } H_j \in \mathcal{C}_i^0 \quad i = 1, 2, \ldots, K \\ \left\{1 - Q\left(\frac{\beta_i - s}{\sigma_O}\right)\right\}\delta[u_i] + Q\left(\frac{\beta_i - s}{\sigma_O}\right)\delta[u_i - 1], & \text{if } H_j \in \mathcal{C}_i^1 \quad j = 0, 1, \ldots, M-1 \end{cases} \tag{8}$$

where $\delta[.]$ is the discrete Dirac delta function and $Q(.)$ is the complementary distribution function of the standard Gaussian random variable defined as

$$Q(x) \;=\; \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt. \tag{9}$$

From equation (5), it can be easily observed that, given any decision at sensor $i$, $u_i$, the corresponding received signal at the fusion center, $y_i$, has a Gaussian distribution with mean $u_i$ and variance $\sigma_n^2$, i.e. $y_i|u_i \sim \mathcal{N}(u_i, \sigma_n^2)$, $i = 1, 2, \ldots, K$. Therefore, the conditional pdf of the received signal at the fusion center, given the corresponding decision at the local sensor, is given by

$$f_{Y_i|U_i}(y_i|u_i) \;=\; \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_i - u_i)^2}{2\sigma_n^2}}, \;\; i = 1, 2, \ldots, K. \tag{10}$$

Since $K$ parallel channels are independent, $y_i$'s received from different sensors at the fusion center are also independent.

Then, the conditional pdf of $y_i$, $i = 1, 2, \ldots, K$, given hypothesis $H_j$, $j = 0, 1, \ldots, M-1$, can be written as

$$f_{Y_i|H_j}(y_i|H_j) \;=\; f_{Y_i|U_i}(y_i|u_i)\; f_{U_i|H_j}(u_i|H_j). \tag{11}$$

Substituting $f_{U_i|H_j}(u_i|H_j)$ and $f_{Y_i|U_i}(y_i|u_i)$ from equations (8) and (10) in equation (11) results in

$$f_{Y_i|H_j}(y_i|H_j) \;=\; \begin{cases} \frac{1}{\sqrt{2\pi\sigma_n^2}}\left\{e^{-\frac{y_i^2}{2\sigma_n^2}}\left[1 - Q\left(\frac{\beta_i}{\sigma_O}\right)\right] + e^{-\frac{(y_i-1)^2}{2\sigma_n^2}} Q\left(\frac{\beta_i}{\sigma_O}\right)\right\}, & \text{if } H_j \in \mathcal{C}_i^0 \\ \frac{1}{\sqrt{2\pi\sigma_n^2}}\left\{e^{-\frac{y_i^2}{2\sigma_n^2}}\left[1 - Q\left(\frac{\beta_i - s}{\sigma_O}\right)\right] + e^{-\frac{(y_i-1)^2}{2\sigma_n^2}} Q\left(\frac{\beta_i - s}{\sigma_O}\right)\right\}, & \text{if } H_j \in \mathcal{C}_i^1 \end{cases}. \tag{12}$$

Using the conditional pdf of the received local decision from each sensor $i$ at the fusion center, $y_i$, under any underlying hypothesis, $H_j$, its conditional moment generating function (MGF) can be evaluated as

$$\Phi_{Y_i|H_j}(\nu) \;=\; \mathbb{E}\left[e^{\nu Y_i}\middle|H_j\right] \;=\; \int_{-\infty}^{\infty} e^{\nu y_i} f_{Y_i|H_j}(y_i|H_j)\; dy_i \;=\; \mathcal{L}\left\{f_{Y_i|H_j}(y_i|H_j)\right\}\Big|_{-\nu}, \tag{13}$$

where $\mathbb{E}$ and $\mathcal{L}$ denote the expected value of a random variable and Laplace transform of a function, respectively. Note that at the last step, the variable of the Laplace transform is changed to $-\nu$. Substituting $f_{Y_i|H_j}(y_i|H_j)$ from equation (12) into (13) results in

$$\Phi_{Y_i|H_j}(\nu) = \begin{cases} e^{\frac{\sigma_n^2\nu^2}{2}} \left\{ \left[ 1 - Q\left(\frac{\beta_i}{\sigma_O}\right)\right] + Q\left(\frac{\beta_i}{\sigma_O}\right) e^{\nu} \right\}, & \text{if } H_j \in \mathcal{C}_i^0 \\ e^{\frac{\sigma_n^2\nu^2}{2}} \left\{ \left[ 1 - Q\left(\frac{\beta_i - s}{\sigma_O}\right)\right] + Q\left(\frac{\beta_i - s}{\sigma_O}\right) e^{\nu} \right\}, & \text{if } H_j \in \mathcal{C}_i^1 \end{cases}. \tag{14}$$

This is the conditional MGF of $y_i$, given the underlying hypothesis $H_j$.

Using the MGF of $y_i$, $i = 1, 2, \ldots, K$, conditioned on hypothesis $H_j$, $j = 0, 1, \ldots, M-1$, we can calculate the conditional MGF of the fusion center's decision metric, $S_0 = \sum_{i=1}^{K} y_i$, given hypothesis $H_j$, as

$$\begin{aligned} \Phi_{S_0|H_j}(\nu) &= \mathbb{E}\left[e^{\nu S_0}\middle|H_j\right] = \mathbb{E}\left[e^{\nu \sum_{i=1}^{K} Y_i}\middle|H_j\right] = \mathbb{E}\left[\prod_{i=1}^{K} e^{\nu Y_i}\middle|H_j\right] \\ &\overset{(a)}{=} \prod_{i=1}^{K} \mathbb{E}\left[e^{\nu Y_i}|H_j\right] = \prod_{i=1}^{K} \Phi_{Y_i|H_j}(\nu) \end{aligned} \tag{15}$$

where (a) is due to the independence of $y_i$'s under a given hypothesis $H_j$.

As mentioned in Section 2, assuming uniform sensor distribution within the sensing environment, for any hypothesis $H_j$, $j = 0, 1, \ldots, M-1$, there are on average $K_j$ sensors for which $H_j \in \mathcal{C}_i^1$ and $K - K_j$ sensors for which $H_j \in \mathcal{C}_i^0$. Therefore, substituting $\Phi_{Y_i|H_j}(\nu)$ from equation (14) into (15), the conditional MGF of the fusion center's decision metric, given hypothesis $H_j$, can be written as

$$\begin{aligned} \Phi_{S_0|H_j}(\nu) &= \exp\left(\frac{K\sigma_n^2\nu^2}{2}\right) \left[\prod_{i=1}^{K_j} \left\{1 - Q\left(\frac{\beta_i - s}{\sigma_O}\right)\right\} + Q\left(\frac{\beta_i - s}{\sigma_O}\right) e^{\nu}\right] \\ &\quad \times \left[\prod_{i=K_j+1}^{K} \left\{1 - Q\left(\frac{\beta_i}{\sigma_O}\right)\right\} + Q\left(\frac{\beta_i}{\sigma_O}\right) e^{\nu}\right]. \end{aligned} \tag{16}$$

Note that $\Phi_{S_0|H_j}(\nu)$ can be simplified using algebraic manipulations to the final form of

$$\Phi_{S_0|H_j}(\nu) = \exp\left(\frac{K\sigma_n^2\nu^2}{2}\right) \sum_{\ell=0}^{K} a_\ell \exp(\ell\nu), \tag{17}$$

where $a_\ell$, $\ell = 0, 1, \ldots, K$, is a function of appropriate subsets of $Q\left(\frac{\beta_i - s}{\sigma_O}\right)$, $i = 1, 2, \ldots, K_j$, and $Q\left(\frac{\beta_i}{\sigma_O}\right)$, $i = K_j + 1, K_j + 2, \ldots, K$.

Based on the result of equation (13), the conditional pdf of the fusion center's decision metric, $S_0$, given hypothesis $H_j$, can be calculated from its MGF as

$$f_{S_0|H_j}(s_0|H_j) = \mathcal{L}^{-1}\left\{\Phi_{S_0|H_j}(-\nu)\right\}, \quad j = 0, 1, \ldots, M-1, \tag{18}$$

where $\mathcal{L}^{-1}$ is the inverse Laplace transform of a function. Substituting $\Phi_{S_0|H_j}(\nu)$ from equation (17) into (18) results in the conditional pdf of the fusion center's decision metric, given hypothesis $H_j$, as follows:

$$f_{S_0|H_j}(s_0|H_j) = \frac{1}{\sqrt{2\pi K\sigma_n^2}} \sum_{\ell=0}^{K} a_\ell \exp\left[-\frac{(s_0 - \ell)^2}{2K\sigma_n^2}\right]. \tag{19}$$

Based on Bayesian decision theory, the minimum error probability decision rule for the $M$-ary classification of the underlying hypothesis using the fusion center's decision metric $S_0$ is

$$\begin{aligned}\hat{H}_j &= \underset{j\in\{0,1,\ldots,M-1\}}{\arg\max}\; f_{H_j|S_0}\left(H_j|s_0\right) \\ &= \underset{j\in\{0,1,\ldots,M-1\}}{\arg\max}\; p_j f_{S_0|H_j}\left(s_0|H_j\right).\end{aligned} \tag{20}$$

Therefore, substituting $f_{S_0|H_j}\left(s_0|H_j\right)$ from equation (19) into (20) results in the minimum error probability decision rule at the fusion center, which achieves the maximum probability of correct classification. Moreover, the optimum decision thresholds at the fusion center, $\Gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_{M-1}\}$, can be found as the intersection of different a posteriori pdfs, $f_{H_j|S_0}\left(H_j|s_0\right)$, $j = 0, 1, \ldots, M-1$. The fusion center will then classify the underlying hypothesis based on its decision rule summarized in equation (7) by using its decision metric, $S_0 = \sum_{i=1}^{K} y_i$.

It can be observed from equation (16) that the coefficients $a_\ell$, $\ell = 0, 1, \ldots, K$, in $f_{S_0|H_j}\left(s_0|H_j\right)$ are functions of the local decision thresholds, $\beta_i$, $i = 1, 2, \ldots, K$. Based on the Bayesian decision theory, the *locally* optimal decision rule of sensor $i$ for a *binary* decision making on whether a non-null hypothesis has occurred or not is in the form of

$$P\left[H_j \in \mathcal{C}_i^1 | x_i\right] \underset{u_i=0}{\overset{u_i=1}{\gtrless}} P\left[H_j \in \mathcal{C}_i^0 | x_i\right], \tag{21}$$

which can be rewritten as

$$\frac{f_{X_i|\left\{H_j\in\mathcal{C}_i^1\right\}}\left(x_i|H_j \in \mathcal{C}_i^1\right)}{f_{X_i|\left\{H_j\in\mathcal{C}_i^0\right\}}\left(x_i|H_j \in \mathcal{C}_i^0\right)} \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \frac{P\left[H_j \in \mathcal{C}_i^0\right]}{P\left[H_j \in \mathcal{C}_i^1\right]}. \tag{22}$$

Considering the conditional observation model of sensor $i$, given hypothesis $H_j$, defined in equation (3), locally optimal decision rule of sensor $i$ derived in (22) can be simplified as

$$x_i \underset{u_i=0}{\overset{u_i=1}{\gtrless}} \frac{s}{2} + \frac{\sigma_O^2}{s} \ln\left(\frac{P\left[H_j \in \mathcal{C}_i^0\right]}{P\left[H_j \in \mathcal{C}_i^1\right]}\right). \tag{23}$$

If we compare the above local decision rule with the one defined in equation (4), the locally optimal decision threshold of sensor $i$ can be defined as

$$\beta_{i,\text{Local}} = \frac{s}{2} + \frac{\sigma_O^2}{s} \ln\left(\frac{P\left[H_j \in \mathcal{C}_i^0\right]}{P\left[H_j \in \mathcal{C}_i^1\right]}\right). \tag{24}$$

It can be observed from (24) that $\beta_{i,\text{Local}}$ depends only on the variance of additive observation noise, $\sigma_O^2$. However, this decision threshold might not result in the globally optimized probability of correct classification at the fusion center. In this paper, our goal is to find the *globally* optimal local decision thresholds that result in the maximum probability of correct classification at the fusion center. It should be noted that these globally optimal local decision thresholds depend on variances of both observation noise and channel noise, $\sigma_O^2$ and $\sigma_n^2$. In the next section, we present the results of our analysis using a numerical scenario and discuss the effects of such a global optimization compared to a local optimization of the local decision thresholds in the performance of the $M$-ary classification system.

## 4. NUMERICAL ANALYSIS

In this section, the performance of the proposed channel-aware multi-hypothesis classification WSN architecture is evaluated for a typical numerical scenario. First, the parameters of the WSN under analysis are specified. Then, the effects of observation signal-to-noise ratio (SNR) and channel SNR on the performance of the classification system are investigated. Moreover, the performance enhancement that can be achieved by optimizing local sensors' decision thresholds globally rather than locally is discussed. Finally, the effects of the number of distributed local sensors on the performance of the proposed classification system are evaluated.

## 4.1 Parameter Specification of Analyzed WSN

The performance of the proposed multi-hypothesis classification system is analyzed for a typical WSN. Suppose that a WSN is formed by $K = 15$ sensors distributed over an area with size $S = 15$. The goal is to classify the distributed observed data as being generated by $M = 3$ hypotheses, $H_0$, $H_1$, and $H_2$, with known prior probabilities $p_0 = P[H = H_0] = 0.6$, $p_1 = P[H = H_1] = 0.3$, and $p_2 = P[H = H_2] = 0.1$. The influence fields of the non-null hypotheses are of size $A_1 = 5$ and $A_2 = 15$. Therefore, assuming uniform distribution of local sensors within the observation environment, $K_1 = 5$ and $K_2 = 15$ are the average number of sensors that are in the influence field of $H_1$ and $H_2$, respectively. Assume that $s = 1$ is the normalized strength of the observable influence field of non-null hypotheses.

Since there are two non-null hypotheses, $H_1$ and $H_2$, in this example, local sensors are divided into two disjoint groups. The first group is composed of $K_1 = 5$ sensors that can be inside the influence field of either of the non-null hypotheses. The second group is formed by the other $K - K_1 = 10$ sensors that can only be inside the influence field of hypothesis $H_2$. It is intuitive to assume that the decision thresholds of the sensors in each one of these two groups are the same. Therefore, the set of local decision thresholds is composed of 15 elements each one of them is associated with one sensor. The first five elements of this set are all equal. Similarly, the last ten elements of this set are all equal. In the rest of this section, we refer to these two local decision thresholds as $\beta_1$ and $\beta_2$, respectively.

## 4.2 Effects of Observation and Channel SNR on Classification Performance

The average optimized probability of correct classification at the fusion center, $P_c$, versus observation SNR (SNRO) is shown in Fig. 2 for different values of channel SNR (SNRC). SNRO and SNRC are defined as

$$\mathsf{SNRO} = \frac{1}{2\sigma_O^2} \text{ and } \mathsf{SNRN} = \frac{1}{2\sigma_n^2},$$

where $\sigma_O^2$ and $\sigma_n^2$ are variances of observation noise and channel noise, respectively. Fig. 3 shows the optimized average probability of correct classification at the fusion center versus channel SNR for different values of observation SNR. SNR values are given in dB. Solid lines show $P_c$ when the local decision thresholds are optimized globally through an exhaustive search over all possible local threshold values between zero and two with step 0.05. Dotted lines show $P_c$ when the local decision thresholds are optimized locally and derived based on equation (24). For the WSN under consideration, $\beta_{1,\text{Local}}$ and $\beta_{2,\text{Local}}$ can be written as

$$\begin{aligned} \beta_{1,\text{Local}} &= \frac{1}{2} + \sigma_O^2 \ln\left(\frac{p_0}{p_1 + p_2}\right) \\ \beta_{2,\text{Local}} &= \frac{1}{2} + \sigma_O^2 \ln\left(\frac{p_0 + p_1}{p_2}\right). \end{aligned} \tag{25}$$

As it can be observed from Figs. 2 and 3, the average probability of correct classification approaches one as SNR increases. Moreover, under all SNR regimes, channel-aware classification system, which is based on the globally optimal local decision thresholds, outperforms conventional classification system, which is based on locally optimal local decision thresholds. This crucial point is further demonstrated in detail in Table 1. In this table, the values of local decision thresholds derived from both local optimization and global optimization are shown in different columns. Moreover, the corresponding optimized average probability of correct classification at the fusion center is shown for each case. When the observation SNR is fixed, the locally optimum decision thresholds are also fixed based on equation (25). However, globally optimum decision thresholds change with both observation SNR and channel SNR. In the last column of the table, the percentage improvement in average probability of correct classification at the fusion center due to the global optimization of decision thresholds is shown. Note that as SNR increases, the achievable percentage improvement decreases. In other words, global optimization of decision thresholds does not improve the average probability of correct classification at high SNRs. A justification for this observation is that for high SNRs, the probability of correct classification at the fusion center is very high (near one) by itself and it cannot be increased further.

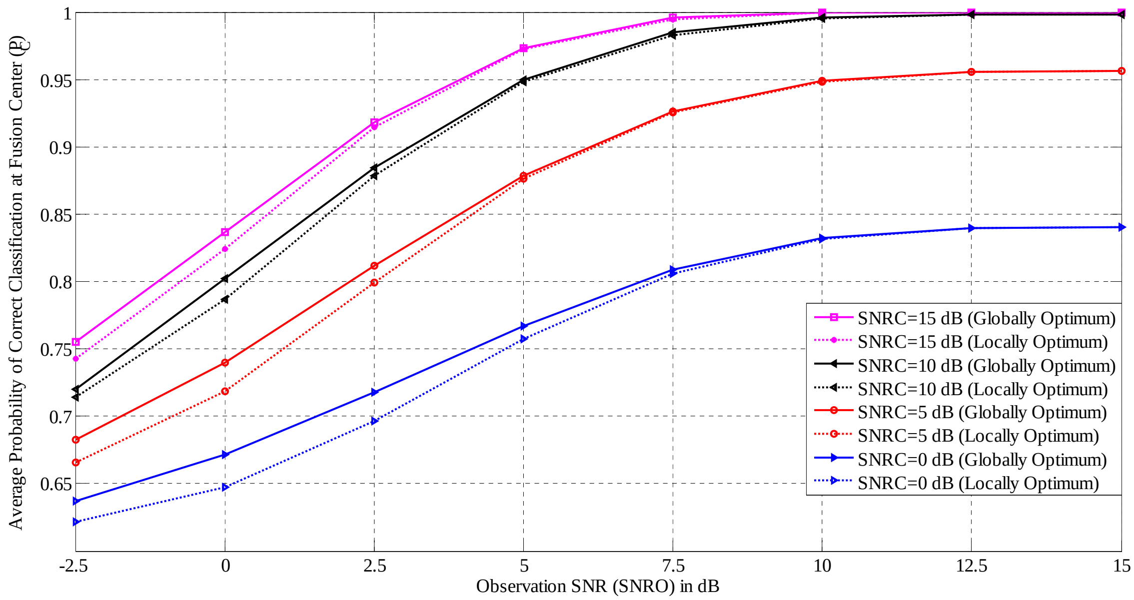


Figure 2. Optimized average probability of correct classification at the fusion center versus observation SNR (SNRO) for different values of channel SNR (SNRC). The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

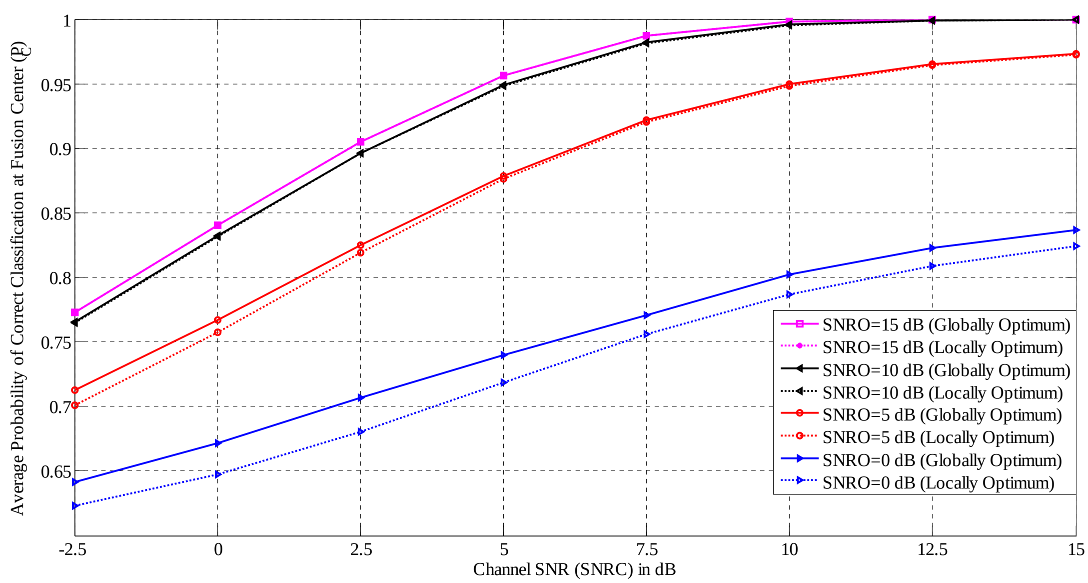


Figure 3. Optimized average probability of correct classification at the fusion center versus channel SNR (SNRC) for different values of observation SNR (SNRO). The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

### 4.3 Effect of the Number of Sensors $K$ on Classification Performance

The performance of the proposed multi-hypothesis classification system is a function of the number of distributed sensors in the observation environment. In Fig. 4, the optimized average probability of correct classification at the fusion center versus the number of distributed sensors in the observation environment, $K$, is shown for different values of observation and channel SNR. Solid lines are used to indicate that the local decision thresholds are optimized globally through an exhaustive search over all possible threshold values between zero and two with step 0.05. Dotted lines show $P_c$ when the local decision thresholds are optimized locally and derived based

Table 1. Performance improvement due to globally optimizing local decision thresholds.

| SNRO | SNRC | $\boldsymbol{\beta} = \{\beta_1, \beta_2\}$ | | $P_c$ | | Percentage Improvement |
|---|---|---|---|---|---|---|
| | | Local | Global | Local | Global | |
| 0 | 0 | $\{0.70, 1.6\}$ | $\{0.55, 0.65\}$ | 0.647 | 0.672 | 3.82% |
| | 5 | | $\{0.55, 0.75\}$ | 0.719 | 0.740 | 2.92% |
| | 10 | | $\{0.5, 1.15\}$ | 0.787 | 0.803 | 1.97% |
| | 15 | | $\{0.4, 1.3\}$ | 0.824 | 0.837 | 1.59% |
| 5 | 0 | $\{0.56, 0.85\}$ | $\{0.5, 0.6\}$ | 0.757 | 0.767 | 1.33% |
| | 5 | | $\{0.5, 0.8\}$ | 0.876 | 0.878 | 0.27% |
| | 10 | | $\{0.5, 0.9\}$ | 0.949 | 0.951 | 0.16% |
| | 15 | | $\{0.5, 0.85\}$ | 0.972 | 0.973 | 0.11% |
| 10 | 0 | $\{0.52, 0.61\}$ | $\{0.5, 0.55\}$ | 0.831 | 0.832 | 0.06% |
| | 5 | | $\{0.5, 0.65\}$ | 0.948 | 0.949 | 0.03% |
| | 10 | | $\{0.5, 0.75\}$ | 0.995 | 0.996 | 0.06% |
| | 15 | | $\{0.5, 0.8\}$ | 0.9996 | 0.9997 | 0.02% |
| 15 | 0 | $\{0.51, 0.54\}$ | $\{0.5, 0.5\}$ | 0.840 | 0.841 | $\simeq 0\%$ |
| | 5 | | $\{0.5, 0.55\}$ | 0.955 | 0.956 | $\simeq 0\%$ |
| | 10 | | $\{0.5, 0.65\}$ | 0.997 | 0.998 | $\simeq 0\%$ |
| | 15 | | $\{0.5, 0.7\}$ | 0.999 | 0.999 | $\simeq 0\%$ |

on equation (25). As it can be seen in Fig. 4, as the number of local sensors increases, the optimized average probability of correct classification at the fusion center also increases. Furthermore, our classification system shows an acceptable performance in terms of average probability of correct classification for moderate number of local sensors. Notice that since the proposed classification system works based on the number of sensors that are in the influence field of each hypothesis, if the number of sensors is very small, the number of sensors that can be in the influence field of different hypotheses is almost the same. Therefore, the value of conditional decision metric under different hypotheses is not distinct enough for the fusion center to be able to distinguish between them. This problem is more important for low SNR regimes.

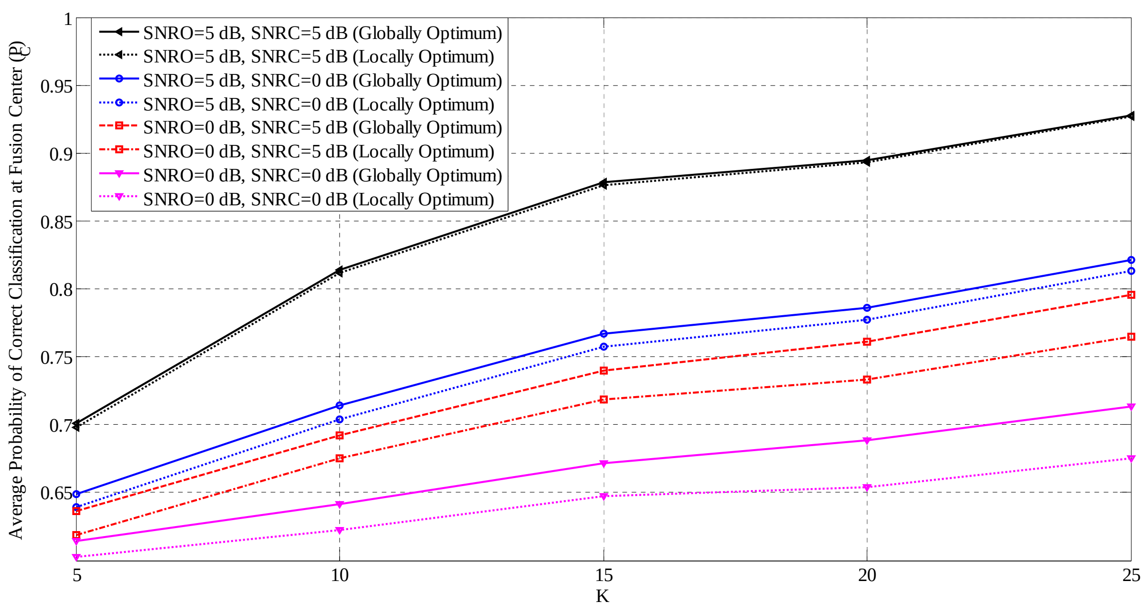


Figure 4. Optimized average probability of correct classification at the fusion center versus the number of distributed sensors in the observation environment ($K$) for different values of observation and channel SNR. The local decision thresholds are optimized either globally (solid lines) or locally (dotted lines).

## 5. CONCLUSIONS

In this paper, we designed a method to optimize the performance of a decentralized WSN deployed as a multi-hypothesis classification system. Local sensors employ a simple binary decision rule and make decisions based on their noisy observations. These binary decisions are sent to the fusion center through parallel AWGN channels. The fusion center then forms a decision metric as the linear combination of these noisy local decisions, which will be used to perform a global multi-hypothesis classification based on the known influence fields of different hypotheses. Fusion decision rule was formulated and numerical performance analysis of an example WSN was presented to investigate the effects of the observation and channel SNR, and the number of local distributed sensors, on the classification performance. Numerical analysis results showed that the proposed approach could simplify decision making at the local sensors while achieving an acceptable performance in terms of the global average probability of correct classification at the fusion center. Furthermore, it was shown that a global optimization of the local decision thresholds can improve the probability of correct classification at the fusion center compared to the case in which local thresholds are only locally optimized.

## REFERENCES


1. D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets in distributed sensor networks," *IEEE Signal Processing Magazine* **19**, pp. 17–29, March 2002.
2. A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita, "A line in the sand: a wireless sensor network for target detection, classification, and tracking," *Computer Networks* **46**, pp. 605–634, December 2004.
3. J.-F. Chamberland and V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing Magazine* **24**, pp. 16–25, May 2007.
4. B. Chen, L. Tong, and P. Varshney, "Channel-aware distributed detection in wireless sensor networks," *IEEE Signal Processing Magazine* **23**, pp. 16–26, July 2006.
5. F. A. Sadjadi, "Hypotheses testing in a distributed environment," *IEEE Transactions on Aerospace and Electronic Systems* **22**, pp. 134–137, March 1986.
6. T.-Y. Wang, Y. Han, P. Varshney, and P.-N. Chen, "Distributed fault-tolerant classification in wireless sensor networks," *IEEE Journal on Selected Areas in Communications* **23**, pp. 724–734, April 2005.
7. T.-Y. Wang, Y. Han, B. Chen, and P. Varshney, "A combined decision fusion and channel coding scheme for distributed fault-tolerant classification in wireless sensor networks," *IEEE Transactions on Wireless Communications* **5**, pp. 1695–1705, July 2006.
8. T.-Y. Wang, Y. Han, and P. Varshney, "Fault-tolerant distributed classification based on non-binary codes in wireless sensor networks," *IEEE Communications Letters* **9**, pp. 808–810, September 2005.
9. H.-T. Pai, Y. S. Han, and J.-T. Sung, "Two-dimensional coded classification schemes in wireless sensor networks," *IEEE Transactions on Wireless Communications* **7**, pp. 1450–1455, May 2008.
10. X. Zhu, Y. Yuan, C. Rorres, and M. Kam, "Distributed $M$-ary hypothesis testing with binary local decisions," *Information Fusion* **5**, pp. 157–167, September 2004.
11. Q. Zhang and P. K. Varshney, "Towards the fusion of distributed binary decision tree classifiers," in *2nd International Conference on Information Fusion (Fusion'99)*, (Sunnyvale, CA, USA), July 1999.
12. K. Nguyen, T. Alpcan, and T. Basar, "Distributed hypothesis testing with a fusion center: The conditionally dependent case," in *47th IEEE Conference on Decision and Control (CDC'08)*, pp. 4164–4169, (Cancun, Mexico), December 2008.
13. Z.-B. Tang, K. Pattipati, and D. Kleinman, "A distributed $M$-ary hypothesis testing problem with correlated observations," *IEEE Transactions on Automatic Control* **37**, pp. 1042–1046, July 1992.
14. S. Bapat, V. Kulathumani, and A. Arora, "Reliable estimation of influence fields for classification and tracking in an unreliable sensor network," in *24th IEEE Symposium on Reliable Distributed Systems (SRDS'05)*, pp. 60–69, (Orlando, FL, USA), October 2005.
15. J. Li and G. AlRegib, "Distributed estimation in energy-constrained wireless sensor networks," *IEEE Transactions on Signal Processing* **57**, pp. 3746–3758, October 2009.