A Combined Decision Fusion and Channel Coding Scheme for Distributed Fault-Tolerant Classification in Wireless Sensor Networks

Tsang-Yi Wang, *Member, IEEE,* Yunghsiang S. Han, *Member, IEEE,* Biao Chen, *Member, IEEE,* and Pramod K. Varshney, *Fellow, IEEE*

***Abstract*— In this paper, we consider the distributed classification problem in wireless sensor networks. Local decisions made by local sensors, possibly in the presence of faults, are transmitted to a fusion center through fading channels. Classification performance could be degraded due to the errors caused by both sensor faults and fading channels. Integrating channel decoding into the distributed fault-tolerant classification fusion algorithm, we obtain a new fusion rule that combines both soft-decision decoding and local decision rules without introducing any redundancy. The soft decoding scheme is utilized to combat channel fading, while the distributed classification fusion structure using error correcting codes provides good sensor fault-tolerance capability. Asymptotic performance of the proposed approach is also investigated. Performance evaluation of the proposed approach with both sensor faults and fading channel impairments is carried out. These results show that the proposed approach outperforms the system employing the MAP fusion rule designed without regard to sensor faults and the multiclass equal gain combining fusion rule.**

***Index Terms*— Distributed classification, wireless sensor networks, coding, soft-decision decoding, decision fusion, multisensor fusion, fading channels.**

I. Introduction

Wireless sensor networks (WSN) form an emerging area that has attracted enormous attention in recent years [1]. Its envisaged applications include, among others, monitoring of environments and performing tasks such as detection, classification and tracking of objects [2]–[7]. In this paper, we consider the problem of event or target classification based on observations from distributed sensors. In a decentralized multiclass classification problem, each local detector (sensor) may perform multiclass classification and transmit its decision (classification result) to a fusion center [8], [9] where a final classification decision is made. The fusion center may be a central decision unit, or, in clustered-based WSN, may simply be a cluster head. Thus the algorithm developed in this paper applies to different levels of classification fusion problems. The sensor-level decision can be represented by $\lceil \log_2 M \rceil$ information bits, where M is the number of classes to be distinguished.

However, there are many limitations for this mechanism in WSN applications. In particular, in sensor networks involving a large number of unattended sensors operating on irreplaceable power, the cost, bandwidth, and energy constraints may dictate that sensors transmit as few bits as possible. In fact, in some recent work sensors transmit binary decisions to the fusion center at which they are combined to yield multiclass decisions [10]–[12]. This approach was adopted in our previous work [13], where fault tolerant fusion rules have been developed that are robust to the presence of sensor faults. Also, it has been indicated that, in a WSN, fault-tolerance capability is critical since sensors can be damaged, blocked or run out of battery energy [1], [2]. Notice that at the heart of the distributed classification fusion approach using error correcting codes (DCFECC) is the fault-tolerant fusion rule. Unlike the conventional approach that employs the Chair-Varshney fusion rule[1] that assumes no faults [14], the fault-tolerant fusion rule provides enough distance between the decision regions corresponding to different hypotheses by a careful design and exploitation of a code matrix. The observed local decision vectors could still fall into correct decision regions even when several sensor faults are present. We provide a brief introduction to the DCFECC approach in Section II.

Although the DCFECC approach has shown excellent fault-tolerance capability in the presence of sensor faults [13], the effect of communication in WSN, and, in particular, the impact of fading channels [15], has not been considered. Fading induced transmission errors will cause degradation of the classification performance. Thus, in this paper, we explore decoding rules that are robust to fading channels at the fusion center. We name this new approach distributed classification fusion using soft-decision decoding (DCSD). Simulation results have shown that when the effect of channel fading is taken into account, the performance of the proposed

Manuscript received February 17, 2004; revised January 14, 2005; accepted July 25, 2005. The editor coordinating the review of this paper and approving it for publication was J. Evans. This work was supported in part by the W.M. Keck Foundation under grant No. CK #1329, by the SUPRIA program of the CASE Center at Syracuse University, by the National Science Foundation under grant #0501534, and by the NSC of Taiwan, R.O.C., under grants NSC 93-2218-E-260-003 and 93-2213-E-305-005.

T.-Y. Wang is with the Institute of Communications Engineering, National Sun Yat-sen University, Taiwan, R.O.C. (e-mail: tcwang@mail.nsysu.edu.tw).

Y. S. Han is with the Graduate Institute of Communication Engineering, National Taipei University, 151 University Rd. Sanhsia, Taiwan, R.O.C. (e-mail: yshan@mail.ntpu.edu.tw).

B. Chen and P. K. Varshney are with the Dept. of Electrical Engineering and Computer Science, Syracuse University, 335 Link Hall, Syracuse, NY 13244 USA (e-mail: bichen@ecs.syr.edu; varshney@syr.edu).

Digital Object Identifier 10.1109/TWC.2006.04037

[1] Chair-Varshney fusion rule is the optimal fusion rule given the local decision rules.

new decoding rule is much better than that presented in [13].

Another proposed enhancement from DCFECC is that a D-ary ($D \geq 2$) decision can be sent out from local sensors if needed (Note that in the DCFECC approach local sensors can only make binary local decisions). In this new scheme, a single fixed binary code matrix is used at the fusion center for all values of D. That is, DCSD can process multiple-bit local decision information while the code matrix is still binary. It will be shown later that when more bits of local decision information are sent, and when the channel status is good, further improvement in classification performance can be obtained when the total energy output from each sensor is fixed.

Asymptotic performance analysis for the DCSD approach is also provided in this paper. The results show that the probability of error for DCSD approaches zero under certain conditions that are not difficult to satisfy in practice.

For performance evaluation, we compare the DCSD approach with the system employing the MAP fusion rule. The results show that the DCSD approach has better performance in the presence of sensor faults. We also employ the multiclass equal gain combining (MEGC) fusion rule for the purpose of comparison. This rule is an extension of the equal gain combining (EGC) fusion rule for the distributed binary detection problem proposed in [16], [17]. Simulation results show that DCSD has the best fault-tolerance capability among these three approaches even though the MEGC rule also has robust performance. Finally, the performance is evaluated with fixed total energy sent out from each sensor node when multiple-bit local decision information is used.

This paper is organized in the following manner. A brief introduction to the DCFECC approach is given in Section II. System description of the DCSD approach is given in Section III. Section IV provides the derivation of the local decision rule by the person-by-person optimization (PBPO) procedure and the derivation of the DCSD fusion rule. Theoretical performance analysis is provided in Section V. In Section VI and Section VII, performance evaluation of the DCSD approach is provided. Finally, we conclude this paper in Section VIII.

II. A Brief Introduction to the DCFECC Approach

The DCFECC approach was used to solve the fault-tolerant multiclass classification problem [13]. The system architecture for the DCFECC system can be obtained from the system shown in Fig. 1, by replacing the fading channels with binary symmetric channels. The local sensors send binary information to the fusion center based on the observations $y_1, y_2, \ldots, y_N$. However, the fusion center makes the final decision in favor of one of the multiclass options based on the binary vector formed by the received binary local decisions. Several researchers have considered the design of fault-tolerant detection systems given *a priori* fault probability [18]–[20]. However, they only considered the binary detection problem. Moreover, the *a priori* fault probability which is needed in those approaches is difficult to estimate in many real world applications. In contrast, the DCFECC approach employs a fault-tolerant fusion rule by using error-correcting codes to provide good fault-tolerance in the multiclass classification problem.

The key to the DCFECC approach is the design and exploitation of a code matrix $\boldsymbol{T}$. Let H_ℓ, where $\ell = 0, 1, \ldots, M-1$ and $M \geq 2$, denote the M hypotheses under test at each of the N sensors. The code matrix $\boldsymbol{T}$ is an $M \times N$ matrix with binary elements $t_{\ell j} \in \{0, 1\}$, $\ell = 0, \ldots, M-1$, $j = 1, \ldots, N$. Each hypothesis $H_\ell \in \Omega = \{H_0, H_1, \ldots, H_{M-1}\}$ is associated with a row in the code matrix $\boldsymbol{T}$ in the following manner. The ℓth row vector of the matrix is the "codeword" corresponding to the ℓth hypothesis, while its jth column vector provides the decision rule for the jth sensor. That is, if the jth sensor "decides" on hypothesis H_ℓ, it sends a binary decision whose value equals $t_{\ell j}$ to the fusion center. Thus, assuming ideal transmission and no sensor faults, one needs at least $\lceil \log_2 M \rceil$ sensors for classifying M hypotheses. With more sensors, i.e., $N > \lceil \log_2 M \rceil$, the redundancy can be exploited for fault-tolerance.

In [13], the DCFECC system is designed as follows. First, a good error-correcting code matrix needs to be found. It can be obtained based on the simulated annealing or the cyclic column replacement algorithm as proposed in [13]. The obtained code matrix is then employed at the fusion center to implement the fault-tolerant fusion rule. The local decision rule at each local sensor is designed based on the minimization of the misclassification error criterion. We follow the PBPO approach given the designed code matrix. To provide fault-tolerance ability, the fusion center then performs minimum distance decoding (the fault-tolerant fusion rule) to decide on the hypotheses based on the binary inputs received (the received vector) from the local sensors. That is, the fusion center decides on the codeword that is closest in Hamming distance to the received vector, where the Hamming distance between two binary vectors is defined as the number of distinct positions between these vectors. The decision on a codeword is equivalent to making a multiclass classification decision. While the system performance in the DCFECC approach also depends on the patterns of columns in the code matrix, a larger minimum Hamming distance of the code employed usually provides the capability to tolerate more faults. The justification for this can be found in the asymptotic analysis and an example provided in [13].

It should be pointed out that large scale wireless sensor networks often have many nodes and sensor nodes are usually aggregated into several groups (or clusters) to reduce the amount of power spent on long distance data transmissions [21]. Hence, the members of each group (or cluster) are within transmission range of each other and the number of members of each group (or cluster) is 10 to 40 [1]. It is possible for each group (or cluster) to run the DCFECC scheme separately [21]. Collaborative detection processing is carried out among nodes within a group (or cluster) under the control of a manager node (or cluster head) where the fusion process is carried out. This cluster based architecture is attractive from computational point of view.

III. Problem Statement

The multiclass classification problem considered by the DCSD approach in WSN is formulated in this section and

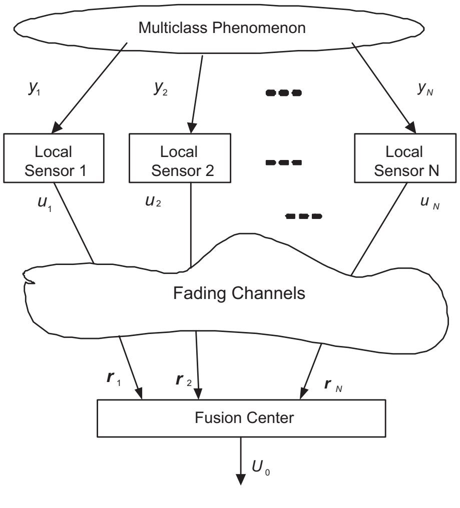


Fig. 1. Distributed classification fusion architecture for fading channels.

the system structure is depicted in Fig. 1. The *a priori* probabilities of the M hypotheses are denoted by $P(H_\ell) = P_\ell$, respectively. The observation at each local sensor or detector is represented by y_j, where $j = 1, \ldots, N$. Assume that the distribution function of y_j under each hypothesis is known with the conditional probability density functions of these observations denoted by $P(y_j|H_\ell)$.

As in the DCFECC approach, a code matrix $\boldsymbol{T}$ to perform distributed classification fusion is designed in advance by either the simulated annealing or the cyclic column replacement algorithm presented in [13]. Unlike the problem formulated in [13] where we assumed binary local sensor decisions, we consider the case where each local sensor processes its observations and makes a multilevel decision based on the corresponding column of matrix $\boldsymbol{T}$. More precisely, each sensor employs a decision rule $g_j(y_j)$ to make a multilevel D-ary decision $u_j = d$, where $d = 0, \ldots, D-1$ and $D \geq 2$. Since the multilevel local decision rule in the DCSD approach is designed according to a binary code matrix $\boldsymbol{T}$, a new distance metric, instead of the Hamming distance used in the DCFECC approach, must be employed to measure the distance between the multilevel local decision vector and the codeword in the given binary code matrix. This distance metric will be used to design the optimal local fusion rule whereby the fusion center employs soft-decision decoding. In this paper, we use the distance metric between $\boldsymbol{u} = (u_1, \ldots, u_N)$ and $\boldsymbol{T}_\ell = (t_{\ell 1}, \ldots, t_{\ell N})$ defined as

$$d(\boldsymbol{u}, \boldsymbol{T}_\ell) = \sum_{j=1}^{N} d_j(u_j, t_{\ell j}) \text{ for } \ell = 0, \ldots, M-1,$$

where

$$d_j(u_j, t_{\ell j}) = |u_j - t_{\ell j} \times (D-1)|.$$

This definition implies that different levels of a multilevel D-ary local decision have different distances from elements $t_{\ell j} \in \{0, 1\}$ in a binary code matrix. Moreover, the distances between all adjacent decision levels are equal. Consider an example of 4-ary local decisions, the distance between the local decision $u_j = 3$ and $t_{\ell j} = 0$ is $d_j(u_j = 3, 0) = 3$, and the distance between the local decision $u_j = 3$ and $t_{\ell j} = 1$ is $d_j(u_j = 3, 1) = 0$. Note that each local sensor makes its decision by itself based on its own observations and is independent of the other sensors. After processing the observations locally, possibly in the presence of faults, the local decisions u_j are mapped to a binary signal vector $\boldsymbol{b}_j = (b_{j1}, \ldots, b_{jS})$, where $S = \lceil \log_2 D \rceil$ is the number of bits to represent the local decision $u_j = d, d = 0, \ldots, D-1$. In this paper, we assume that all local decisions, u_j, take values from 0 up to $D-1$. For instance, a four-level local decision is transmitted by means of one of the 2-bit binary signal vectors, $\{11, 10, 01, 00\}$. Each bit of the binary signal vectors is transmitted to the fusion center over channels sequentially and is assumed to undergo independent fading. Due to the fact that low bit rate (long symbol duration) and short range (hence small delay spread) are used in most WSN [16], the fading channels are assumed flat. We further make the assumption of phase coherent reception. Hence, the effect of fading channels is further simplified as a real scalar multiplication given the transmitted signal.

We assume that binary antipodal signalling is employed for transmission. This results in a received vector at the fusion center consisting of real numbers, $\boldsymbol{R} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n)$, $\boldsymbol{r}_j = (r_{j1}, \ldots, r_{jS})$, where $j = 1, \ldots, N$. The r_{js}, for $s = 1, \ldots, S$, can be expressed as $r_{js} = \alpha_{js}(-1)^{b_{js}}\sqrt{E_b} + n_{js}$, where E_b is the energy per channel bit and n_{js} is a noise sample from a Gaussian process with two-sided power spectral density $N_0/2$. α_{js} is the attenuation factor that models the fading channel.

By taking advantage of the structure of the fault-tolerant fusion rule for the DCFECC approach, the received vectors at the fusion center can be treated as codewords sent from a transmitter using a channel coding scheme. That is, the local decisions $u_j, j = 1, \ldots, N$ form a D-ary codeword $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$ which is transmitted to the fusion center. Soft distance metric between received real vectors $\boldsymbol{R} = (\boldsymbol{r}_1, , \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$ and $\boldsymbol{T}_\ell, \ell = 0, \ldots, M-1$, can then be obtained based on the MAP criterion. Soft-decision decoding [22]–[25] at the fusion center can then provide a channel error protection capability. We provide a detailed derivation of soft-decision decoding in Section IV. The fault-tolerance capability for sensor faults is also achieved due to the structure of error correcting codes. Note that, in our earlier work [13], the error-correcting code matrix is only used in the decision fusion process and does not provide error correction for the local decisions transmitted over real channels. The research reported in this paper integrates channel coding/decoding to combat the effects of fading with the DCFECC classification fusion algorithm. We obtain a new soft-decision fusion rule that takes into consideration local decision rules. The soft-decision decoding scheme is utilized to combat channel fading, while the DCFECC fusion structure provides good sensor fault-tolerance ability. In the next section, we give a detailed description of the DCSD scheme.

IV. The DCSD Scheme

In this section, we derive the local decision rules and the fusion rule for the DCSD approach. For simplicity, in the design of local decision rules, the received vector at the fusion center is assumed to consist of multilevel D-ary decisions $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)$, where $u_j = d$, and $d = 0, \ldots, D-1$, and

TABLE I
THE CODE MATRIX OBTAINED BY SIMULATED ANNEALING IN [13]

H_0	0	0	0	1	0	1	0	0	1	1
H_1	1	0	0	0	0	0	0	1	0	1
H_2	1	1	1	0	1	0	1	1	0	0
H_3	0	1	0	1	1	1	1	0	0	0

that local decisions are transmitted over error-free channels. That is, the design of local decision rules is based on receiving multilevel quantized decisions instead of real numbers at the fusion center. The design of the DCSD fusion rule (decoding rule), however, assumes real-valued channel output. Note that it will have infinitely many cases (real-valued vectors) that will need to be considered if the soft decoding rule and the local decision rules are jointly optimized. Naturally, this will be computationally prohibitive.

A. The Derivation of Local Decision Rules

Define $C^{\ell}_{i_1,i_2,\ldots,i_N}$, where $i_1, i_2, \ldots, i_N \in \{0, 1, \ldots, D-1\}$, as the cost that the received word at the fusion center $\boldsymbol{u}$ equals $(i_1, i_2, \ldots, i_N)$ and the true hypothesis is H_ℓ. These costs $C^{\ell}_{i_1,i_2,\ldots,i_N}$ can be determined by the decision regions of codewords. Note that these costs are based on D-ary decisions. In order to determine these costs based on the binary codeword, the distance metric $d(\boldsymbol{u}, \boldsymbol{T}_\ell)$ defined in Section III is used. According to the designed code matrix, the decision region Ψ of a codeword $\boldsymbol{T} \in \boldsymbol{T}_w$ is given as follows:

$$\Psi(\boldsymbol{t}) = \{\boldsymbol{u} | d(\boldsymbol{u}, \boldsymbol{t}) \le d(\boldsymbol{u}, \boldsymbol{t}') \text{ for all } \boldsymbol{t}' \in \boldsymbol{T}_w\},$$

where $\boldsymbol{T}_w = \{\boldsymbol{t}_\ell | \ell = 0, \ldots, M-1\}$ is the set of all codewords, i.e., all rows of the code matrix. In order to minimize the probability of misclassification, set $C^{\ell}_{i_1,\ldots,i_N} = 0$ if $(i_1, \ldots, i_N)$ is in the decision region of $\boldsymbol{t}_\ell$ that is the row of $\boldsymbol{T}$ corresponding to the hypothesis H_ℓ; otherwise set $C^{\ell}_{i_1,\ldots,i_N} = 1$. Whenever a received vector $(i_1, \ldots, i_N)$ simultaneously belongs to decision regions of $\boldsymbol{t}_{k_0}, \boldsymbol{t}_{k_1}, \ldots, \boldsymbol{t}_{k_{q-1}}$, where $q > 1$, for all $\ell = 0, \ldots, q-1$, set $C^{\ell}_{i_1,\ldots,i_N} = (1 - 1/q)$, i.e., we assume the fusion center randomly picks one codeword among the codewords which are at the same distance from the received word $\boldsymbol{u}$. For instance, given the code matrix $\boldsymbol{T}$ illustrated in Table I and $D = 4$, we have $C^{0}_{3,2,1,2,1,2,3,0,3,3} = 0$, $C^{1}_{3,2,1,2,1,2,3,0,3,3} = 1$, $C^{2}_{3,2,1,2,1,2,3,0,3,3} = 1$, and $C^{3}_{3,2,1,2,1,2,3,0,3,3} = 1$, since the distance between $\boldsymbol{u} = (3, 2, 1, 2, 1, 2, 3, 0, 3, 3)$ and codeword $\boldsymbol{t}_0$ is the smallest.

Based on the costs assigned above, the probability of misclassification can be written as

$$\begin{aligned} P_e = \sum_{i_1,\ldots,i_N,\ell} \int_{y_1,\ldots,y_N} & P_\ell P(u_1 = i_1 | y_1) \times \cdots \times \\ & P(u_N = i_N | y_N) P(y_1, \ldots, y_N | H_\ell) C^{\ell}_{i_1,\ldots,i_N}, \end{aligned} \tag{1}$$

if the local sensors make local decisions based on their own observations independent of the hypothesis present.

Let us follow the person-by-person optimization (PBPO) procedure described in [8]. The probability of error P_e can be minimized if we set the local decision rule at sensor k as

$$P(u_k = i_k | y_k) = \begin{cases} 1, & \text{if } \boldsymbol{I}^*_k(i_k) \le \boldsymbol{I}^*_k(m) \text{ for all } i_k \text{ and} \\ & m \text{ such that } i_k \ne m, \text{ and } i_k, \\ & m = 0, \ldots, D-1; \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

where,

$$\begin{aligned} \boldsymbol{I}^*_k(i_k) = & \sum_{\ell} P(y_k | H_\ell) \sum_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_N} P_\ell \\ & P(u_1 = i_1 | H_\ell) \times \cdots \times P(u_{k-1} = i_{k-1} | H_\ell) \times \\ & P(u_{k+1} = i_{k+1} | H_\ell) \times \cdots \times P(u_N = i_N | H_\ell) \\ & \times C^{\ell}_{i_1,\ldots,i_k,\ldots,i_N} \end{aligned} \tag{3}$$

under the assumption of conditionally independent observations given any hypothesis.

It is easy to see that the local decision rule at sensor k depends on the decision rule at the other sensors. In order to obtain the PBPO local decision rules, we must solve these coupled equations. An algorithm that could be used to search for the PBPO decision rules is the iterative Gauss-Seidel cyclic coordinate descent algorithm [26], although it only converges to a local optimum and depends on the chosen initial values of $P(u_1|H_\ell), P(u_2|H_\ell), \ldots, P(u_N|H_\ell)$.

B. DCSD Fusion (Decoding) Rule

As mentioned earlier, the essence of the fusion process in the DCFECC approach is decoding. This coding structure enables us to consider the received vector at the fusion center as a codeword transmitted collectively from all local sensors. Consequently, the DCSD fusion rule is able to jointly consider the local decision rules and word-by-word decoding to achieve robust system performance via its sensor fault-tolerance and channel error correction capability.

Let codeword $\boldsymbol{t}_k$ be chosen for transmission before local decision making when the true hypothesis H_k is present. The process of local decision making can be seen as a transmission channel, and $P(u_j | t_{kj})$ can be seen as a transition probability for the channel at local sensor j. Employing the MAP criterion, the fusion rule can be stated below given the received vector $\boldsymbol{R}$:

$$\text{set } \hat{H} = H_\ell \text{ if}$$

$$P(\boldsymbol{t}_\ell | \boldsymbol{R}) \ge P(\boldsymbol{t}_k | \boldsymbol{R}), \text{ for all } \boldsymbol{t}_k \in \boldsymbol{T}_w. \tag{4}$$

If we assume equal prior probability of each hypothesis, then the MAP decoding rule is equivalent to the maximum-likelihood decoding (MLD) rule. Thus, (4) becomes

$$P(\boldsymbol{R} | \boldsymbol{t}_\ell) \ge P(\boldsymbol{R} | \boldsymbol{t}_k), \text{ for all } \boldsymbol{t}_k \in \boldsymbol{T}_w. \tag{5}$$

Assuming conditional independence of observations at the sensors and discrete memoryless channels between local decision outputs and the fusion center, (5) can be rewritten as

$$\prod_{j=1}^{N} P(\boldsymbol{r}_j | t_{\ell j}) \ge \prod_{j=1}^{N} P(\boldsymbol{r}_j | t_{kj}), \text{ for all } \boldsymbol{t}_k \in \boldsymbol{T}_w,$$

where $\boldsymbol{r}_j = (r_{j1}, r_{j2}, \ldots, r_{jS})$ is the jth component of the received vector $\boldsymbol{R}$. The above equation can be further rewritten as

$$\prod_{j=1}^{N} \sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{\ell j}) \geq \prod_{j=1}^{N} \sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{kj}), \text{ for all } \boldsymbol{t}_k \in \boldsymbol{T}_w,$$

since the received $\boldsymbol{r}_j$ does not depend on the code matrix we designed given the local decision u_j. Taking the logarithm on both sides of the above equation, we have, for all $\boldsymbol{t}_k \in \boldsymbol{T}_w$,

$$\sum_{j=1}^{N} \ln \frac{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{\ell j})}{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{kj})} \geq 0.$$

For a binary code matrix, we can define reliability of bit j of the received vector as

$$\phi_j = \ln \frac{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|0)}{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|1)}, \tag{6}$$

where

$$P(\boldsymbol{r}_j|u_j = d) = P(r_{j1}|b_{j1})P(r_{j2}|b_{j2}) \cdots P(r_{jt}|b_{jS})$$

can be obtained from the statistical model of the fading channel,[2] and

$$P(u_j = d|s) = \sum_{\ell=0}^{M-1} P(u_j = d|H_\ell)P_j(H_\ell|s), s \in \{0, 1\}.$$

The probability of $P(u_j = d|H_\ell)$ can be obtained from the local decision rule, while $P_j(H_\ell|s)$ is the probability that the hypothesis H_ℓ is true given s present at the bit j (column j of the code matrix) before local decision making, and can be expressed as

$$P_j(H_\ell|s) = \frac{P_j(s|H_\ell)P_\ell}{\sum_{\ell=0}^{M-1} P_j(s|H_\ell)P_\ell},$$

where

$$P_j(s|H_\ell) = \begin{cases} 1, & \text{if } t_{\ell j} = s; \\ 0, & \text{if } t_{\ell j} \neq s \end{cases}.$$

Then,

$$\begin{aligned} &\sum_{j=1}^{N} \ln \frac{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{\ell j})}{\sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|t_{kj})} \geq 0 \\ \Leftrightarrow\ &\sum_{j=1}^{N} ((-1)^{t_{\ell j}}\phi_j - (-1)^{t_{kj}}\phi_j) \geq 0 \\ \Leftrightarrow\ &\sum_{j=1}^{N} (\phi_j - (-1)^{t_{\ell j}})^2 \leq \sum_{j=1}^{N} (\phi_j - (-1)^{t_{kj}})^2. \end{aligned}$$

Let us define $d_F(\boldsymbol{\phi}, \boldsymbol{t}_k) = \sum_{j=1}^{N} (\phi_j - (-1)^{t_{kj}})^2$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$. The DCSD fusion rule is then

$$\text{set } \hat{H} = H_\ell, \text{ if } \ell = \arg \min_{0 \leq k \leq M-1} d_F(\boldsymbol{\phi}, \boldsymbol{t}_k). \tag{7}$$

[2] Recall that the local decision u_j is mapped into a binary signal vector $(b_{j1}, b_{j2}, \ldots, b_{jS})$ that is sent over the channel.

Note that the probabilities $P(u_j = d|0)$ and $P(u_j = d|1)$, where $j = 1, \ldots, N$, and $d = 0, \ldots, D-1$, can be computed in advance before the on-line operation if the channel and measurement statistics at the deployed locations are known. When the fusion center receives the real vector $\boldsymbol{R}$, the reliability of each sensor can be computed by (6), and the DCSD fusion rule in (7) can then be employed.

C. A Brief Description of Code Design by Simulated Annealing

As mentioned in [13], the code matrix design can not be viewed as the independent design of individual column vectors. This makes the analytical design of the code matrix quite difficult. Thus, heuristic algorithms are adopted to efficiently solve the code design problem. Since the code matrix used in the performance evaluation in this paper is designed by simulated annealing, presentation of the main ideas of the code design by simulated annealing algorithm is in order. For details, the readers please refer to [13].

1) The energy function required by the annealing process is set to the probability of misclassification as shown in (1).
2) Random changes in the code configuration are achieved by perturbation of the code matrix in the previous iteration until the pre-specified minimum Hamming distance constraint is satisfied. The minimum Hamming distance constraint is set to meet the fault-tolerance requirement.
3) Each time a new code matrix is generated, the local decision rules (2) and (3) are optimized by the Gauss-Seidel algorithm. Specifically, for each sensor j, $j = 1, \ldots, N$, compute the probability $P(u_j|H_\ell)$ by (2) and (3) with updated values of $P(u_1|H_\ell)$, …, $P(u_{j-1}|H_\ell)$, $P(u_{j+1}|H_\ell)$, …, $P(u_N|H_\ell)$.

Code design by simulated annealing could be computationally intensive if the number of sensors becomes very large. But this optimization process is actually performed off-line and the compuation issue is not that critical. Moreover, as mentioned in Section II, large scale wireless sensor networks are usually aggregated into several clusters and the number of members of each cluster is 10 to 40. For such cluster-based networks, the computational load is acceptable.

V. Asymptotic Performance Analysis

In [13], [27] the asymptotic analysis for the DCFECC approach has been provided. The major results from the analysis in [13], [27] are that the DCFECC decoding error vanishes as the minimum Hamming distance of the employed code matrix approaches infinity if local classification is such that $\beta_{\max} < \frac{1}{2}$, where $\beta_{\max} \triangleq \max_{0 \leq i \leq M-1} 1 - h_{i,i}$ and $h_{j,i}$ is the probability of classifying H_j given that H_i is the true hypothesis. Note that the probabilities $h_{j,i}, 0 \leq j, i \leq M-1$ are assumed the same for all local sensors in [13], [27].

In this section, we give the asymptotic analysis for the proposed DCSD approach. Unlike the analysis in [13], [27], we do not assume that the probabilities $h_{j,i}, 0 \leq j, i \leq M-1$ are the same for all the sensors.

For simplicity, we only consider the case of binary local sensor decisions in this asymptotic analysis. The case of D-ary ($D > 2$) local decisions can be easily obtained in a similar manner. We also assume $E_b = 1$ without losing any generality. $E[n_j^2]$ is equal to $N_0/2$ as assumed in the section III. Thus, the reliability of bit j of the received vector can be expressed as

$$\phi_j = \log \frac{q_{0,j}^0 P(r_j|u_j=0) + q_{0,j}^1 P(r_j|u_j=1)}{q_{1,j}^0 P(r_j|u_j=0) + q_{1,j}^1 P(r_j|u_j=1)},$$

where $q_{s,j}^d = P(u_j = d|t_{\ell j} = s)$. We assume $q_s^d > m_q > 0$ here, where m_q is any nonzero constant. In the following, we will omit the second subscript j in $q_{s,j}^d$ if this does not cause any confusion. We first present the following lemma without giving the proof.

Lemma 1: Let $\tilde{\phi_j} = \phi_j - E[\phi_j|H_\ell]$, then

$$E\left\{\tilde{\phi_j}^2|H_\ell\right\} \le \frac{8}{\sigma^4}\left\{E[\alpha_j^4] + \frac{1}{2}E[\alpha_j^2]N_0\right\}.$$

In the case of Rayleigh fading, which is considered in the performance evaluation in this paper, $E[\alpha_j^4]$ is bounded. Thus, $E\left\{\tilde{\phi_j}^2|H_\ell\right\}$ is also bounded for the Rayleigh fading channel.

Let $S(\boldsymbol{t_k}, \boldsymbol{t_\ell})$ be defined as the set $\{j|t_{kj} \ne t_{\ell j}, 1 \le j \le N\}$. The performance of DCSD can be seen from the following theorem.

Theorem 1: If $\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell] \to \infty$, where $Z_j^{\ell k} = \frac{1}{2}\left((-1)^{t_{\ell j}} - (-1)^{t_{kj}}\right)$, then the probability of error for DCSD approaches zero.

Proof:

$$\begin{aligned}
P_e &= \frac{1}{M}\sum_{\ell=0}^{M-1} \Pr\left(\text{decision} \ne H_\ell | H_\ell\right) \\
&\le \frac{1}{M}\sum_{\ell=0}^{M-1} \Pr\big(d_F\left(\boldsymbol{\phi}, \boldsymbol{t}_\ell\right) \ge \\
&\qquad \min_{0\le k\le M-1, k\ne\ell} d_F\left(\boldsymbol{\phi}, \boldsymbol{t}_k\right) \big| H_\ell\big) \\
&\le \frac{1}{M}\sum_{\ell=0}^{M-1} \sum_{0\le k\le M-1, k\ne\ell} \Pr\left(d_F\left(\boldsymbol{\phi}, \boldsymbol{t}_\ell\right) \ge d_F\left(\boldsymbol{\phi}, \boldsymbol{t}_k\right)|H_\ell\right) \\
&= \frac{1}{M}\sum_{\ell=0}^{M-1} \sum_{0\le k\le M-1, k\ne\ell} \Pr\left(\sum_{j=1}^{N}\left(\phi_j - (-1)^{t_{\ell j}}\right)^2 \ge \right. \\
&\qquad \left. \sum_{j=1}^{N}\left(\phi_j - (-1)^{t_{kj}}\right)^2 \right| H_\ell\Bigg) \\
&= \frac{1}{M}\sum_{\ell=0}^{M-1} \sum_{0\le k\le M-1, k\ne\ell} \Pr\left(\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k}\phi_j < 0 \right| H_\ell\Bigg) \\
&= \frac{1}{M}\sum_{\ell=0}^{M-1} \sum_{0\le k\le M-1, k\ne\ell} \Pr\left(\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k}\tilde{\phi}_j \right. \\
&< \left. - \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell] \right| H_\ell\Bigg)
\end{aligned} \tag{8}$$

Let $\sigma_{\tilde{\phi}}^2 = \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} E\left\{(Z_j^{\ell k}\tilde{\phi}_j)^2|H_\ell\right\} = \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} E\left\{\tilde{\phi_j}^2|H_\ell\right\}$, then (8) can be rewritten as

$$\begin{aligned}
P_e \le \frac{1}{M}\sum_{\ell=0}^{M-1} \sum_{0\le k\le M-1, k\ne\ell} &\Pr\left(\frac{1}{\sigma_{\tilde{\phi}}} \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k}\tilde{\phi}_j < \right. \\
&\left. -\frac{1}{\sigma_{\tilde{\phi}}} \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell] \right| H_\ell\Bigg)
\end{aligned} \tag{9}$$

Since $E\left\{\tilde{\phi_j}^2|H_\ell\right\}$ is bounded and thus the Lindeberg condition [28] holds, $\frac{1}{\sigma_{\tilde{\phi}}} \sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k}\tilde{\phi}_j$ tends to the normal distribution with zero mean and unit variance by the Lindeberg central limit theorem [28]. Since $\sigma_{\tilde{\phi}}$ will grow slower than $\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell]$ when $\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell] \to \infty$, the right hand side of (9) approaches zero if $\sum_{j\in S(\boldsymbol{t_k},\boldsymbol{t_\ell})} Z_j^{\ell k} E[\phi_j|H_\ell] \to \infty$. □

This theorem tells us that as long as the summation of positive $Z_j^{\ell k} E[\phi_j|H_\ell]$ for bit $j \in S(\boldsymbol{t_k}, \boldsymbol{t_\ell})$ asymptotically beats the summation of negative $Z_j^{\ell k} E[\phi_j|H_\ell]$ for bit $j \in S(\boldsymbol{t_k}, \boldsymbol{t_\ell})$, the probability of error for DCSD tends to zero. In the i.i.d case when all the local sensors employ an identical decision rule, the condition becomes $Z_j^{\ell k} E[\phi_j|H_\ell] > 0$ and the minimum Hamming distance goes to infinity when the number of sensors approaches infinity. By Gilbert's lower bound on achievable minimum distance [29], for a given M, such code matrix exists with the minimum Hamming distance at least satisfying

$$\lim_{N\to\infty} \frac{\log_2 M}{N} = 1 - H_2\left(\delta\right),$$

where $\delta = \lim_{N\to\infty} \frac{d_{\min}}{N}$, $d_{\min}$ is the minimum Hamming distance, and H_2 is the binary entropy function. Hence, for a fixed M, the minimum Hamming distance of the employed code matrix can easily increase to infinity when the number of sensors goes to infinity, and the condition will only fail when the code happens to use poor local classifications infinitely many times or the channel status is very bad.

VI. Performance Evaluation of the Fault-Tolerance Capability of DCSD

In this section, we investigate the performance of the DCSD approach. We compare the performance of the DCSD approach with the system employing the soft-decision MAP fusion rule (SMAPF) and the multiclass equal gain combining (MEGC) fusion rule in the presence of stuck-at faults and channel transition errors. Both SMAPF and MEGC fusion rules are obtained by extending their binary detection versions given in [16]. In the following, we give brief descriptions of the multiclass classification version of these two fusion rules. The details of derivations can be found in [30].

Soft-Decision MAP Fusion Rule (SMAPF)

The fusion rule provided here is obtained by extending the optimal likelihood ratio fusion rule for the binary detection problem given in [16]. However, unlike the rule provided in [16], we do not assume that the attenuation factor α_{js} corresponding to the fading channel is known at the fusion center. Assume that the local decision rules and the probabilities $P(\boldsymbol{r}_j|u_j = d)$, where $j = 1, \ldots, N$, and $d = 0, \ldots, D-1$, corresponding to the fading channel are known. The optimal fusion rule given the local decision rules in this case is the MAP fusion rule:

$$\text{Assign } \hat{H} \text{ to } H_\ell, \text{ if } \ell = \arg \max_{0 \le k \le M-1} P(\boldsymbol{R}|H_k)P_k. \quad (10)$$

Using the assumption of conditional independence of observations at the sensors and discrete memoryless channels, we have the following result:

$$P(\boldsymbol{R}|H_k) = \prod_{j=1}^{N} \sum_{d=0}^{D-1} P(\boldsymbol{r}_j|u_j = d)P(u_j = d|H_k). \quad (11)$$

Multiclass Equal Gain Combining Fusion Rule

The MEGC fusion rule is also obtained by extending the result for the binary classification problem considered in [16], [17], where EGC is shown to be the low signal to noise ratio approximation of the optimal likelihood ratio fusion rule when only the channel fading statistics are available. When the local sensors only make binary decisions and the fusion center does not know the attenuation factor α of fading channels (only the channel signal-to-noise ratio is known), the following MEGC fusion rule can be employed,

$$\text{Assign } \hat{H} \text{ to } H_\ell, \text{ if } \ell = \arg \max_{0 \le k \le M-1} \sum_{j=1}^{N} q_k^j r_j, \quad (12)$$

where $q_k^j = P(u_j = 0|H_k)$.

As mentioned in [16], [17], for the distributed binary detection problem in wireless sensor networks, the EGC fusion rule outperforms the maximum ratio combining (MRC) fusion rule in the sense that the detection performance is robust for a wider range of channel signal-to-noise ratio (CSNR).

Example 1

The performance of the DCSD, SMAPF, and MEGC schemes are evaluated in both fault-free (without stuck-at faults) and faulty situations (sensors in the presence of stuck-at faults), while the decisions of local sensors are transmitted over Rayleigh fading channels to the fusion center. The CSNR is defined as $\gamma = E_b/N_0 \times E[\alpha_{js}^2]$.

A system with a fusion center and ten independent local sensors is considered for the multiclass classification of four equally likely hypotheses H_0, H_1, H_2, and H_3. Note that each local sensor makes a binary decision in this evaluation. We further assume that all the sensor observations have the same characteristics, i.e. the distributions of observations at all the sensors are identical.[3] The probability density function for each hypothesis is assumed to be a Gaussian distribution with the same variance ($\sigma^2 = 1$) but with different means $0, V, 2V$, and $3V$, respectively. We define the observation signal-to-noise ratio (OSNR) at each local sensor as $20 \log_{10} V$.

[3]This assumption has been made for simplicity in this illustrative example. It may not be true in practice due to the different distances between the target and each sensor.

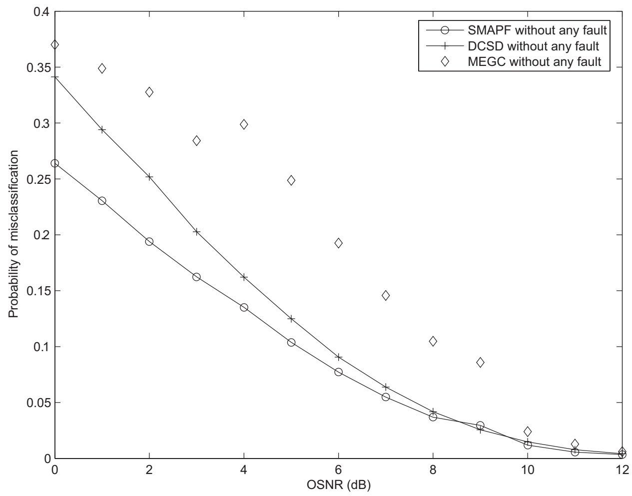


Fig. 2. Performance comparison of DCSD, MEGC, and SMAPF at 5 dB CSNR without any sensor faults.

The code matrix employed in the DCSD approach is obtained from [13] and shown in Table I. It is easy to see that the minimum Hamming distance between any two codewords in the resulting code matrix is 5. After the code matrix is chosen, the local sensor decision rules are given by equation (2) where the coefficients can be determined from (3). For the DCSD approach, the fusion rule given in (7) is employed, while the reliability is computed by using (6). As for SMAPF and MEGC, (10) and (12) are used to compute the fusion rule respectively.

In the performance comparison, we first fix the CSNR at 5 dB, and then consider the OSNR ranging from 0 dB to 12 dB with step size equal to 1 dB. For each OSNR, the computations of the local decision rules for DCSD, SMAPF, and MEGC approaches employ the Gauss-Seidel iterative algorithm. We initialize the Gauss-Seidel algorithm with the probability $P(u_k|H_\ell)$ when all the local sensors operate independently with their optimal decision rule (analytically designed) for classification corresponding to the columns of the code matrix designed for the DCSD approach. For performance evaluation with stuck-at faults, we assume that the faulty sensors always send decision 1 to the fusion center. Performance in terms of probability of misclassification is obtained by 10^5 Monte Carlo runs for each OSNR point.

Fig. 2 and Fig. 3 present simulation results for DCSD, MEGC and SMAPF approaches corresponding to no faulty sensor and two faulty sensors, respectively. The CSNR is fixed at 5 dB. From Fig. 2, when local sensors are in a fault-free state, the MEGC rule has the worst performance among all three approaches. This can be justified from the fact that the MEGC fusion rule is derived as a low CSNR approximation of the optimal likelihood ratio fusion rule. As shown in Fig. 3, DCSD has the best fault-tolerance capability even though MEGC also possesses good fault-tolerance ability.

Fig. 2 and Fig. 3 also reveal two important observations.

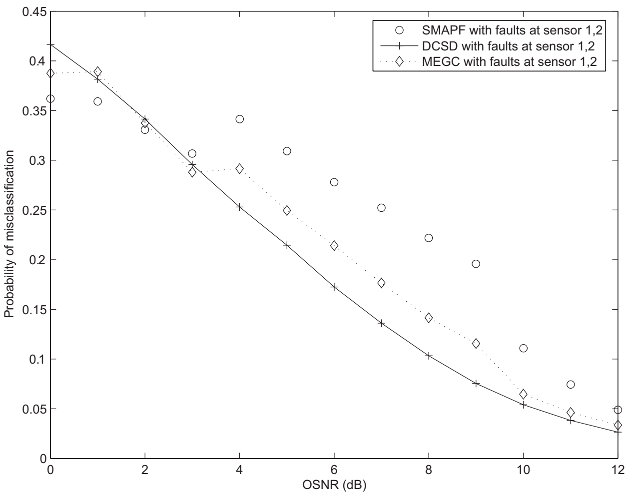


Fig. 3. Performance comparison of DCSD, MEGC, and SMAPF at 5 dB CSNR when two faults are present.

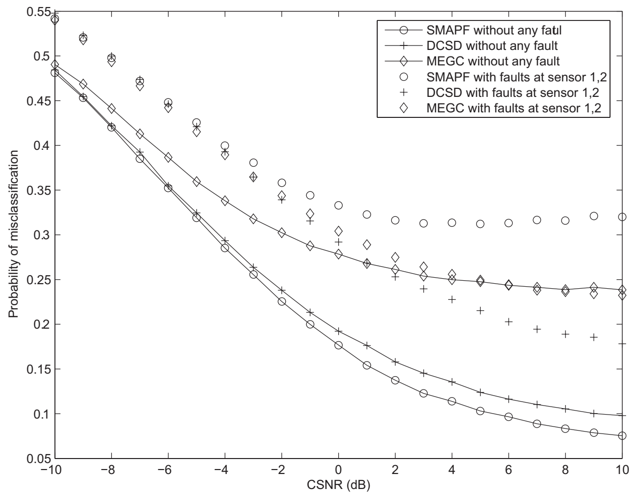


Fig. 5. Performance comparison of DCSD, MEGC, and SMAPF at 5 dB OSNR in faulty and fault-free situations.

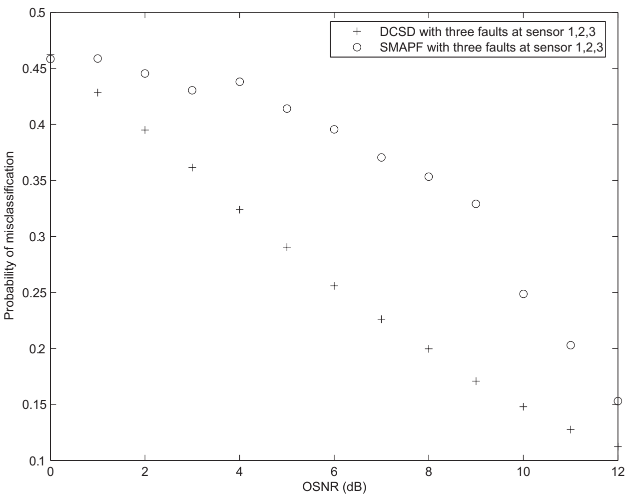


Fig. 4. Performance comparison of DCSD and SMAPF at 5 dB CSNR when three faults are present.

First, the performance of DCSD in the presence of two faulty sensors is better than that of SMAPF at most OSNR values except when OSNR is less than 1 dB. A possible reason for this situation is that the error resulting from two faulty sensors is smaller than the error resulting from the overall decision error of the network at very low OSNRs. Note that the decision error in this section means classification error resulting from the local decision rules and the fusion rule designed for the sensor networks for fault-free sensors. That is, the decision errors dominate the system performance at very low OSNRs. Second, the performance of DCSD without any fault is worse than that of the SMAPF approach at most OSNRs except at 9.0 dB. This is obvious since SMAPF employs the MAP fusion rule, which is optimal given the local decision rule. There is one unusual situation at 9.0 dB where the performance of DCSD is better. This is because the local decision rules determined by the Gauss-Seidel iterative algorithm depend on the initial conditions and do not necessarily yield the global optimum.

Fig. 4 shows the simulation results when faults occur at three sensors ($j = 1, 2, 3$). We can see, unlike the results with two faulty sensors, that the performance of the DCSD approach is better than that of the SMAPF approach at all OSNR values considered. That is, the decision errors of the network do not dominate the performance any more at low OSNRs when the number of sensor faults increases.

Fig. 5 presents the simulation results corresponding to the performance when the OSNR is fixed at 5 dB and CSNR ranges from -10 dB to 10 dB. From this figure, one can see that the MEGC approach has the worst performance while DCSD and SMAPF have similar performance in the absence of faults. However, in the presence of faulty sensors, MEGC provides a more robust performance than SMAPF, yet DCSD still is the most favorable fusion rule in this case.

Although the SMAPF and the MEGC schemes have worse fault-tolerance capability as compared with the DCSD, both schemes require less off-line computation since they do not need to search for the code matrix.

VII. Performance Evaluation of DCSD with Multibit Information

In this section, we evaluate the performance of the DCSD approach for both 1-bit and 2-bit local decision information cases, while the total energy sent out from each sensor node, E, is fixed. Since $E = S \times E_b$, where S is the number of bits that represent the local decision, this implies that the 2-bit local decision has a 3 dB degradation of per bit energy compared with the 1-bit decision. We also compare their performance with the performance of the two stage DCFECC approach. For the two stage DCFECC approach, the fusion center first estimates the local decision $u_j \in \{0, 1\}$ based on the received r_j, and then performs hard-decision decoding according to the DCFECC fault-tolerant fusion rule [13]. Since the maximum likelihood estimate (MLE) for u_j is

$$\hat{u}_j = \begin{cases} 0, & \text{if } r_j > 0; \\ 1, & \text{otherwise,} \end{cases}$$

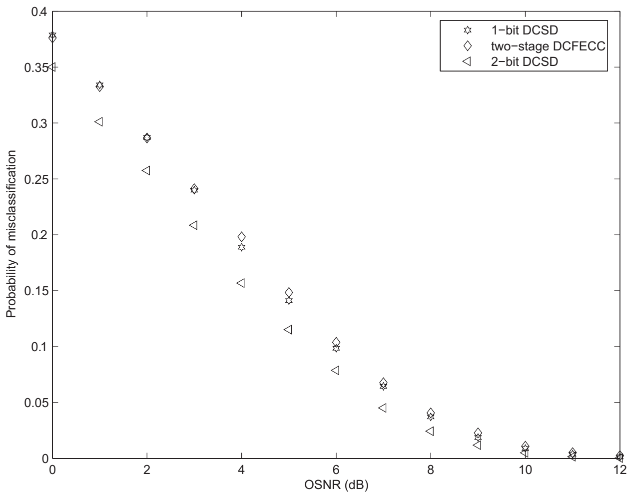


Fig. 6. Performance comparison of 1-bit DCSD, 2-bits DCSD, and two stage DCFECC (hard-decision decoding) at $E/N_0 = 10$ dB.

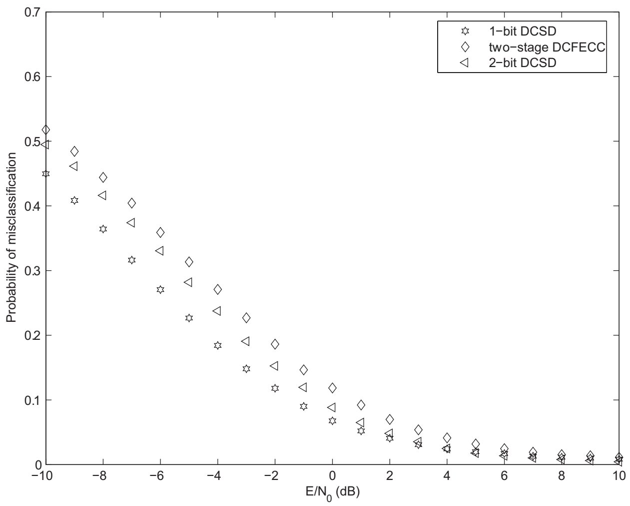


Fig. 7. Performance comparison of 1-bit DCSD, 2-bits DCSD, and two stage DCFECC (hard-decision decoding) at 10 dB OSNR.

given the statistical model of the Rayleigh fading channel [31], the fusion rule for the two stage DCFECC is

$$\text{Assign } \boldsymbol{y} \text{ to } H_\ell, \text{ if}$$

$$\ell = \arg\min_k \sum_{j=1}^{N} (\hat{u}_j - t_{kj})^2, k = 0, \ldots, M-1. \tag{13}$$

Note that DCSD employs soft-decision decoding, while the DCFECC only performs hard-decision decoding.

Example 2

In this example, ten local sensors and one fusion center are used to classify the objects coming from five equally likely hypotheses H_0, H_1, H_2, H_3, and H_4. The distribution function for each hypothesis is assumed to be a Gaussian distribution with the same variance ($\sigma^2 = 1$), but different means, $-2V$, $-V$, 0, V, and $2V$, respectively. In this example, we design a code matrix by the simulated annealing algorithm and the result is shown in Table II. For the purpose of comparison, we first compute the performance of classification as a function of OSNR while the E/N_0 is fixed at a high value. Similarly, we also investigate the performance of classification as a function of E/N_0 when OSNR is large. These two cases reduce the combined effect caused by communication channel errors and observation noise. This is then followed by the performance comparison when both OSNR and E/N_0 are moderate.

TABLE II
THE CODE MATRIX FOR 5 HYPOTHESES AND 10 SENSORS OBTAINED BY SIMULATED ANNEALING

H_0	1	0	0	1	0	1	0	0	0	1
H_1	1	1	0	0	0	0	0	1	1	1
H_2	0	1	1	0	1	1	0	0	1	0
H_3	0	1	1	1	1	0	1	1	0	0
H_4	0	0	0	0	0	0	1	0	0	0

Fig. 6 shows the performance comparison of 1-bit DCSD, 2-bit DCSD, and the two stage 1-bit DCFECC when E/N_0 is fixed at 10 dB, considered as "good" channels. Thus, the performance will be mostly affected by the decision errors with different number of information bits. From this figure, One can observe that the 2-bit DCSD scheme has the best performance, especially at low OSNRs. Therefore, the performance of DCSD can be improved by using more bits to convey local decision information when the E/N_0 is at a high value. From this figure, the performance of 1-bit DCFECC almost matches the performance of 1-bit DCSD. The reason for this result is that we simulate the performance at very high E/N_0 (10 dB) and, therefore, the performance advantage by using soft decision decoding is not significant.

Fig. 7 shows the performance comparison of 1-bit DCSD, 2-bit DCSD, and two stage 1-bit DCFECC when the OSNR is fixed at 10 dB. One can see that the system employing soft-decision decoding results in improved performance over two stage DCFECC that employs hard-decision decoding. From this figure, one can also observe that the performance of 2-bit DCSD is only better when E/N_0 is greater than 6 dB. The reason about this phenomenon will be explained later.

Figs. 8 and 9 illustrate the performance caused by the combined effect of both channel errors and observation noise. Fig. 8 shows the performance comparison of these three schemes when E/N_0 is at 0 dB. One can see that the system employing soft-decision decoding provides improved performance over the two stage 1-bit DCFECC, which employs hard-decision decoding. Unlike the result shown in Fig. 6, the performance of two stage 1-bit DCFECC does not achieve the performance of 1-bit DCSD, since the effect of channel errors is large when E/N_0 is low. Fig. 9 shows the performance comparison when the OSNR is fixed at 5 dB. One can still see that the system employing soft-decision decoding improves the performance of two stage 1-bit DCFECC, which is employing hard-decision decoding, even though the OSNR is moderate.

From Figs. 7, 8, and 9, one can see that using more bits (fixed total energy E) to convey local decision information can only improve the performance when the status of communication channel is good (i.e., E/N_0 is high). When the status of channel becomes worse (i.e., E/N_0 is low), the performance is degraded when more bits are used since the effect of channel errors is amplified with more bits to convey local decision

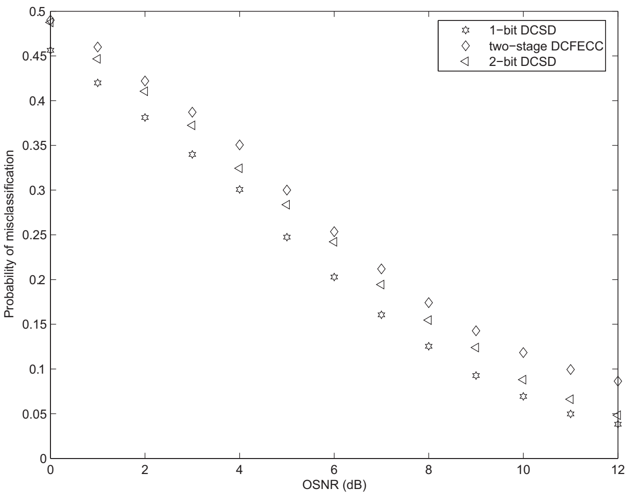


Fig. 8. Performance comparison of 1-bit DCSD, 2-bits DCSD, and two stage DCFECC (hard-decision decoding) at $E/N_0 = 0$ dB.

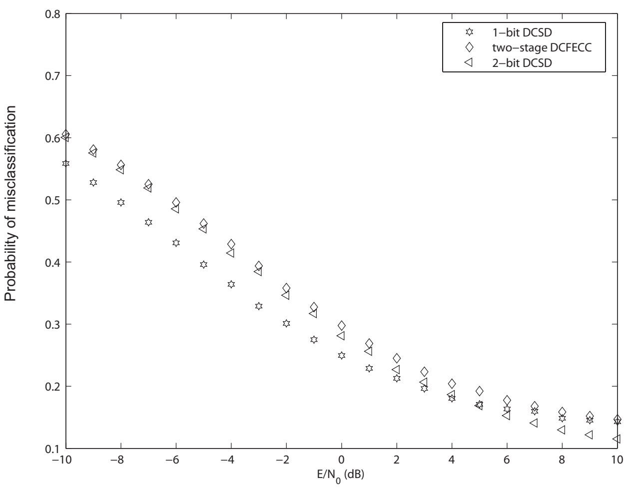


Fig. 9. Performance comparison of 1-bit DCSD, 2-bits DCSD, and two stage DCFECC (hard-decision decoding) at 5 dB OSNR.

information due to the degradation of bit SNR. In this case, the performance lost due to the channel errors is greater than the performance gained by using more bits of local decision information [4].

VIII. Conclusions

In this paper we incorporated soft-decision decoding into the error-correcting code based distributed classification fusion algorithm. By exploiting the structure of the DCFECC approach, the received vectors at the fusion center can be soft-decision decoded. Thus, the sensor fault-tolerance capability and communication channel errors can both be handled in the proposed DCSD approach without introducing any additional redundancy that may reduce channel bandwidth.

Theoretical performance analysis for the DCSD was also provided in this paper. According to this analysis, the condition that the probability of error for the DCSD approach vanishes asymptotically is not difficult to hold in real applications.

It is shown through computer simulations that even though the system that employs the optimal fusion rule has slightly better performance than that of the proposed DCSD approach in the fault-free sensor case, the DCSD approach performs much better when sensors are faulty. We also compare the performance of DCSD with that of the multiclass equal gain combining fusion rule approach. The results show that even though MEGC exhibits good fault-tolerance capability similar to DCSD, it has much worse performance in the fault-free sensors case. Finally, the performance comparison of 1-bit DCSD and 2-bit DCSD with fixed total energy sent out from the output of each sensor node is also investigated. Simulation results show that by using more bits to convey local decision information can improve the performance when the status of channel is good, but degrades the performance if the channel status is poor.

[4]This justification is only true in our performance evaluation under the assumption that no error correction coding is employed on the information sent out from local sensors and the total energy E is fixed.

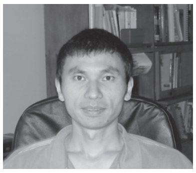
References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, pp. 102-114, Aug. 2002.
[2] L. Dan, K. D. Wong, H. H. Yu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Mag.*, vol. 19, pp. 17-29, Mar. 2002.
[3] H. Wang, J. Elson, L. Girod, D. Estrin, and K. Yao, "Target classification and localization in habitat monitoring," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2003, vol. 4, pp. 844-847.
[4] A. D'Costa and A. M. Sayeed, "Data versus decision fusion in sensor networks," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 2003, vol. 4, pp. 832-835.
[5] S. A. Aldosari and J. M. F. Moura, "Detection in decentralized sensor networks," in *Proc. IEEE International Conference on Accoustics, Speech, and Signal Processing*, May 2004, vol. 2, pp. 277-280.
[6] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1007-1015, Aug. 2004.
[7] Y. Yuan and M. Kam, "Distributed decision fusion with a random-access channel for sensor network applications," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1339-1344, Aug. 2004.
[8] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1997.
[9] F. A. Sadjadi, "Hypotheses testing in a distributed environment," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 2, pp. 134-137, Mar. 1986.
[10] B. V. Dasarathy, "Operationally efficient architectures for fusion of binary-decision sensors in multidecision environments," *Optical Engineering*, vol. 36, no. 3, pp. 632-641, Mar. 1997.
[11] Q. Zhang and P. K. Varshney, "Decentralized *M*-ary detection via hierarchical binary decision fusion," *Information Fusion*, vol. 2, pp. 3-16, Mar. 2001.
[12] C. Rorres X. Zhu, Y. Yuan, and M.Kam, "*M*-ary hypothesis testing with binary local decisions," *Information Fusion*, vol. 5, no. 3, pp. 157-167, Sep. 2004.
[13] T.-Y. Wang, Y. S. Han, P. K. Varshney, and P.-N. Chen, "Distributed fault-tolerant classification in wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 724-734, Apr. 2005.
[14] Z. Chair and P. K. Varshney, "Optimal data fusion in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 98-101, Jan. 1986.
[15] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 551-558, May 2000.
[16] B. Chen, R. Jiang, T. Kasetkasem, and P. K. Varshney, "Channel aware decision fusion in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 52, pp. 3454-3458, Dec. 2004.
[17] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1018-1027, Mar. 2006.
[18] G. G. L. Meyer and H. L. Weinert, "On the design of fault-tolerant signal detectors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 973-978, Aug. 1986.

[19] A. R. Reibman and L. W. Nolte, "Optimal fault-tolerant signal detection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 1, pp. 179-180, Jan. 1990.
[20] A. R. Reibman and L. W. Nolte, "Optimal design and performance of distributed signal detection systems with faults," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1771-1782, Oct. 1990.
[21] R. E. Van Dyck and L. E. Miller, "Distributed sensor processing over an ad hoc wireless network: Simulation framework and performance criteria," in *Proc. IEEE Military Communications Conference*, Oct. 2001, vol. 2, pp. 28-31.
[22] G. D. Forney, Jr., "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, pp. 125-131, Apr. 1966.
[23] Y. S. Han, C. R. P. Hartmann, and C.-C. Chen, "Efficient priority-first search maximum-likelihood soft-decision decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1514-1523, Sep. 1993.
[24] Y. S. Han, "A new treatment of priority-first search maximum-likelihood soft-decision decoding of linear block codes," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 3091-3096, Nov. 1998.
[25] G. C. Clark, Jr. and J. B. Cain, *Error-Correction Coding for Digital Communications*. New York: Plenum Press, 1981.
[26] Z.-B. Tang, K. R. Pattipati, and D. L. Kleinman, "An algorithm for determining the decision thresholds in a distributed detection problem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 21, no. 1, pp. 231-237, Jan./Feb. 1991.
[27] P.-N. Chen, T.-Y. Wang, Y. S. Han, P. K. Varshney, and C. Yao, "Asymptotic performance analysis for minimum-hamming-distance fusion," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar. 2005, vol. 4, pp. 865-868.
[28] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1966.
[29] R. J. McEliece, *The Theory of Information and Coding*, Advanced Book Program. Reading, MA: Addison-Wesley, 1982.
[30] T.-Y. Wang, *Fault-Tolerant Classification in Multisensor Networks Using Coding Theory*, Ph.D. dissertation, Syracuse University, 2003.
[31] J. G. Praokis, *Digital Communication*, 4th ed. New York: McGraw-Hill, 2001.

Tsang-Yi Wang (S'01-M'04) received the B.S. and M.S. degrees from the National Sun Yat-sen University, Kaohsiung, Taiwan, in 1994 and 1996, respectively, and the Ph.D. degree in electrical engineering from the Syracuse University, NY, in 2003. From 2004 to 2006, he was an Assistant Professor in the Graduate Institute of Communication Engineering at National Chi Nan University, Nantou, Taiwan. Since 2006, he has been with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, as an Assistant Professor. His research interests are in wireless sensor networks, distributed detection, and wireless communication.

Yunghsiang S. Han (S'90-M'93) was born in Taipei, Taiwan, on April 24, 1962. He received the B. S. and M. S. degrees in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 1984 and 1986, respectively, and the Ph. D. degree from the School of Computer and Information Science, Syracuse University, Syracuse, NY, in 1993. He is a winner of 1994 Syracuse University Doctoral Prize. From 1993 to 1997 he was an Associate Professor in the Department of Electronic Engineering at Hua Fan College of Humanities and Technology, Taipei Hsien, Taiwan. From 1997 to 2004 he was with the Department of Computer Science and Information Engineering at National Chi Nan University, Nantou, Taiwan. He was promoted to Full Professor in 1998. From June to October 2001 he was a visiting scholar in the Department of Electrical Engineering at University of Hawaii at Manoa, HI, and from September 2002 to January 2004 he was the SUPRIA visiting research scholar in the Department of Electrical Engineering and Computer Science and CASE center at Syracuse University, NY. He is now with the Graduate Institute of Communication Engineering at National Taipei University, Taipei, Taiwan. His research interests are in wireless networks, security, and error-control coding.



Biao Chen (S'97-M'99) received his B.E. and E.E. in Electrical Engineering from Tsinghua University, Beijing, China, in 1992 and 1994 respectively. From 1994 to 1995, he worked at AT&T (China) Inc., Beijing China. He joined the University of Connecticut, Storrs, in 1995 where he received his M.S. in Statistics and Ph.D. in Electrical Engineering, in 1998 and 1999, respectively. From 1999 to 2000 he was with Cornell University as a Post-Doc research associate. Since 2000, he has been with Syracuse University, Syracuse, NY, as an assistant professor with the Department of Electrical Engineering and Computer Science. He is on the editorial board of EURASIP Journal on Wireless Communications and Networking (JWCN) and a guest editor for a special issue on Wireless Sensor Networks of EURASIP JWCN. His area of interest mainly focuses on signal processing for wireless sensor and ad hoc networks and in multi-user MIMO systems.

Pramod K. Varshney (F'97) was born in Allahabad, India on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest honors), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively. Since 1976 he has been with Syracuse University, Syracuse, NY where he is currently a Professor of Electrical Engineering and Computer Science and the Research Director of the New York State Center for Advanced Technology in Computer Applications and Software Engineering. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, and radar signal processing. He has supervised thirty-four Ph.D. dissertations, and has published extensively.

While at the University of Illinois, Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. In 2000, he received the Third Millennium Medal from the IEEE and Chancellor's Citation for exceptional academic achievement at Syracuse University. He serves as a distinguished lecturer for the AES society of the IEEE. He was the President of International Society of Information Fusion during 2001.