Two-Dimensional Coded Classification Schemes in Wireless Sensor Networks

Hung-Ta Pai, *Member, IEEE*, Yunghsiang S. Han, *Senior Member, IEEE*, and Jing-Tian Sung

***Abstract*— This work proposes a novel fault-tolerant classification system based on distributed detection and two-dimensional channel coding. A rule is then derived to reduce the search space such that the optimal code matrix can be found. Simulation results reveal that the proposed scheme has higher classification reliability and better capability of fault tolerance than previous methods. Moreover, a code matrix using repetition codes is presented. The proposed scheme with the repetition code has a lower memory requirement at each sensor and higher detection flexibility than that with the optimal code matrix while only having a slightly lower performance. Finally, an asymptotic performance analysis is provided for the proposed scheme.**

***Index Terms*— Sensor networks, networks and systems, detection and estimation, source/channel coding, transmission technology.**

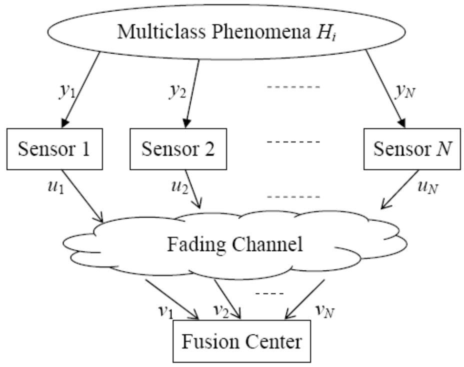

Fig. 1. Structure of a wireless sensor network for distributed detection using N sensors.

I. Introduction

SENSORS in wireless sensor networks (WSNs) detect environmental variations and then transmit the detection results to a fusion center [1]. The fusion center collects all detection results and determines the phenomenon that has occurred. To lower the transmission burden, the detection result is typically denoted by a local decision. The local decision is made by the sensor.

Some sensors may have unrecognized faults under severe conditions. Wang *et al.* [2] proposed Distributed Classification Fusion using Error-Correcting Codes (DCFECC) for fault-tolerance by combining the distributed detection theory [3] with the concept of error-correcting codes in communication systems [4]. One sample is detected in each of N sensors for a given phenomenon. A codeword consisting of N symbols is designed for each phenomenon. In other words, a one-dimensional code ($1 \times N$) corresponds to a phenomenon. Thus, M phenomena form an $M \times N$ code matrix. Each symbol with one bit is assigned to each sensor. A local decision is made from the detection result and is represented with the assigned symbol. Binary decisions from local sensors, possibly in the presence of faults, are forwarded to the fusion center that determines the final decision. Since each codeword in the code matrix is chosen apart from each other, it can tolerate faults made on local decisions when making the final decision. This approach not only provides an improved fault-tolerance capability but also reduces computation time and memory requirements at the fusion center. DCSD (distributed classification fusion using soft-decision decoding) [5] was later developed by improving DCFECC. DCSD adopts a symbol with L bits, instead of one bit [5]. However, the misclassification probability remains high in the extreme case, i.e., with many faulty sensors and very low Signal-to-Noise Ratios (SNRs). Moreover, the multi-bit symbol increases the sensor complexity (cost).

This work proposes an approach to reduce the misclassification probability in multiple-observation scenario while keeping the sensor complexity low. $L-1$ more observations are employed at each sensor and then L bits are required to represent the L observations (one bit for one observation) rather than one observation in [5]. Each phenomenon is represented by a two-dimensional (2-D) codeword ($L \times N$) and then a code matrix ($M \times L \times N$) is needed.

The key contributions of this work are summarized as follows. (1) Develop a two-dimensional coded classification scheme. The scheme has lower complexity and misclassification probability than the previous works. (2) Derive a rule such that the optimal code matrix can be found through full search. (3) Demonstrate that the repetition code is a practical choice as the code matrix. (4) Prove that the misclassification probability of the scheme approaches to zero asymptotically.

II. Fault-Tolerant Distributed Detection

Figure 1 depicts a wireless sensor network for distributed detection with N sensors deployed for collecting environment variation data and a fusion center for making a final decision of detections. When one of phenomena H_i, where $i = 1, 2, \ldots, M$, occurs, all sensors observe the same phenomenon. One observation y_j is undertaken at the jth sensor. The

Manuscript received May 14, 2006; revised July 24, 2007 and November 3, 2007; accepted February 7, 2008. The associate editor coordinating the review of this paper and approving it for publication was A. Stefanov. This work was supported by the National Science Council, Taiwan, R.O.C., under the project No. 94-2213-E-305-002 and NSC 93-2213-E-305-005.

H.-T. Pai and Y. S. Han are with the Graduate Institute of Communication Engineering, National Taipei University (e-mail: htpai@mail.ntpu.edu.tw).

J. T. Sung was with the Department of Electrical Engineering, National Taiwan University of Science and Technology. He is now with the Institute for Information Industry, Networks & Multimedia Institute, Taiwan.

Digital Object Identifier 10.1109/TWC.2008.060259.

TABLE I
THE 4×10 OPTIMAL CODE MATRIX

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

observation is normally a real number represented by infinite number of bits. Transmitting the real number to the fusion center would consume too much power, so a local decision, u_j, is made instead. For a phenomenon, if only L bits are allowed to send the local decision from the sensor to the fusion center, then the L bits are used to represent the decision. The fusion center collects all local decisions and makes a global decision according to them.

In the DCFECC approach, $L = 1$ is set, and an $M \times N$ code matrix $\mathbf{T}$ is designed using the simulated annealing algorithm [2], [6]. The application of the code matrix is derived from error-correcting codes. The code matrix is adopted herein not only to correct transmission errors but also to resist faulty sensors since the incorrect local decision of the faulty sensor can be regarded as a transmission error too. Table I shows an example of $\mathbf{T}$. Row i of the matrix is a codeword $\mathbf{c}_i = (c_{i,1}, c_{i,2}, \ldots, c_{i,N})$ corresponding to hypothesis H_i and $c_{i,j}$ is a 1-bit symbol corresponding to the decision of sensor j. Notably, the Hamming distance, denoted as $d(\mathbf{c}_i, \mathbf{c}_k)$, between two codewords $\mathbf{c}_i$ and $\mathbf{c}_k$ is defined to be the number of positions in which the symbol differ and the minimum Hamming distance of a code matrix is the smallest Hamming distance between any two distinct codewords of the matrix [4].

Let v_j be the received local decision at the fusion center, where $v_j \in \{0, 1\}$ and $\mathbf{v} = (v_1, v_2, ..., v_N)$. A cost function is then defined as $C_{\mathbf{v},\mathbf{c}_i} = 1 - \frac{1}{q}$ if $\mathbf{c}_i$ is one of q solutions of $\arg\min_{\mathbf{c}_k} d(\mathbf{v}, \mathbf{c}_k)$; otherwise, $C_{\mathbf{v},\mathbf{c}_i} = 1$, which indicates a classification error. Hence, the Bayes risk function [3] represents the probability of misclassification,

$$P_e = \sum_{i,\mathbf{v}} \int_{\mathbf{y}} p(\mathbf{v}, \mathbf{y}, H_i) C_{\mathbf{v},\mathbf{c}_i}, \qquad (1)$$

where $\mathbf{y} = (y_1, y_2, ..., y_N)$ represents all values observed at sensors. Let $\mathbf{u} = (u_1, u_2, \ldots, u_N)$ be all local decisions at sensors, and make the following assumptions:

Assumption 1: Observations at all sensors are conditionally independent, i.e., $p(\mathbf{y}|H_i) = p(y_1, y_2, \ldots, y_N|H_i) = \prod_{j=1}^{N} p(y_j|H_i)$.
This assumption is due to the fact that all sensors observe the same phenomenon and they are assumed to locate apart enough.

Assumption 2: The jth local decision, u_j, only depends on the jth observation, y_j.
This assumption is due to the fact that there are no communications between sensors when they make their own local decisions.

Assumption 3: The jth received local decision, v_j, at the fusion center only depends on the jth local decision, u_j.
This assumption is due to the fact that the fusion center receives the local decisions one by one from sensors. This can be easily satisfied in practice when TDMA transmission protocol is employed.

Equation (1) can then be recast as

$$P_e = \sum_{i,\mathbf{u},\mathbf{v}-v_j} \int_{\mathbf{y}} p(H_i) \left[p(\mathbf{v}_{j=1}|\mathbf{u}) p(\mathbf{u}|\mathbf{y}) p(\mathbf{y}|H_i) C_{\mathbf{v}_{j=1},\mathbf{c}_i} + p(\mathbf{v}_{j=0}|\mathbf{u}) p(\mathbf{u}|\mathbf{y}) p(\mathbf{y}|H_i) C_{\mathbf{v}_{j=0},\mathbf{c}_i} \right], \qquad (2)$$

where $\mathbf{v}_{j=b_v} = (v_1, \ldots, v_{j-1}, b_v, v_{j+1}, \ldots, v_N)$ is a vector with jth elements equaling to b_v, $b_v \in \{0, 1\}$, and $\mathbf{v} - v_j$ represents the elements of $\mathbf{v}$ except v_j, i.e., $\mathbf{v} - v_j = (v_1, ..., v_{j-1}, v_{j+1}, ..., v_N)$. The local decision rule of sensor j is

$$\sum_i p(y_j|H_i) D_j(\mathbf{T}) \underset{u_j=0}{\overset{u_j=1}{\gtrless}} 0, \qquad (3)$$

where

$$D_j(\mathbf{T}) = \sum_{i,\mathbf{u}-u_j,\mathbf{v}-v_j} \int_{\mathbf{y}-y_j} p(H_i) p(\mathbf{v} - v_j|\mathbf{u} - u_j) \times p(\mathbf{u} - u_j|\mathbf{y} - y_j) p(\mathbf{y} - y_j|H_i) \times \{[C_{\mathbf{v}_{j=0},\mathbf{c}_i} \left(1 - p_{10}^j\right) + C_{\mathbf{v}_{j=1},\mathbf{c}_i} p_{10}^j] - [C_{\mathbf{v}_{j=0},\mathbf{c}_i} p_{01}^j + C_{\mathbf{v}_{j=1},\mathbf{c}_i} \left(1 - p_{01}^j\right)]\}. \qquad (4)$$

Notably, $\mathbf{u} - u_j$ and $\mathbf{y} - y_j$, similar to $\mathbf{v} - v_j$, represent the elements of $\mathbf{u}$ and $\mathbf{y}$ except u_j and y_j, respectively, and $p_{b_v b_u}^j = p(v_j = b_v|u_j = b_u), b_v, b_u \in \{0, 1\}$. Equation (4) shows that the jth local decision rule depends on not only the code matrix $\mathbf{T}$ but also on the value of $p(\mathbf{u} - u_j|\mathbf{y} - y_j)$, which is derived from the local decision rules in the other sensors. Wang *et al.* adopted a person-by-person optimization to determine all of the local decision rules. The decision region at sensor j can be represented by a set of thresholds such that a local decision rule associated with this threshold set can be performed to determine u_j when y_j is observed. DCSD approach utilizes multiple bits ($L > 1$) and soft decoding, respectively, to improve the reliability of the local and final decisions. However, it still makes one observation at each sensor and one-dimensional code matrix is employed.

III. TWO-DIMENSIONAL CODED DETECTION SCHEME

A. System architecture

After performing one observation at each sensor, $L-1$ more observations are asked one by one from the fusion center to increase the reliability of the fusion result. The vector $\mathbf{y}_j = (y_{1,j}, y_{2,j}, \ldots, y_{L,j})^T$, where $y_{l,j}$ is the observation l and a real number, denotes the observation data at the jth sensor. The vector $\mathbf{u}_j = (u_{1,j}, u_{2,j}, \ldots, u_{L,j})^T$, where $u_{l,j} \in \{0, 1\}$, represents the local decision at sensor j. The matrix of all local decisions is given by $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N)$. Furthermore, this work adopts binary modulation because it is simple [5].[1]

[1]Other modulation schemes can also be employed. However, the calculation of the reliability of the received local decision becomes more complicated.

The matrix received at the fusion center is given by $\tilde{\mathbf{V}} = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_N)$, where $\tilde{\mathbf{v}}_j = (\tilde{v}_{1,j}, \tilde{v}_{2,j}, \ldots, \tilde{v}_{L,j})^T$ and [2]

$$\tilde{v}_{l,j} = \alpha_{l,j} (-1)^{u_{l,j}} \sqrt{\frac{E_s}{L}} + n_{l,j}. \quad (5)$$

Notice that $\alpha_{l,j}$ is the attenuation factor, E_s is the total transmission energy per sensor, and $n_{l,j}$ is the additive white Gaussian noise (AWGN) with the two-sided power spectral density $N_0/2$. Since the total energy E_s is shared by the L transmission bits, the energy of each transmitted bit is only E_s/L. This work later demonstrates that the 2-D coding scheme improves the overall performance despite the small transmission bit energy. We first make the following assumption:

Assumption 4: Observations at a sensor are independent, i.e., $p(y_{1,j}, y_{2,j}, \ldots, y_{L,j}|H_i) = \prod_{l=1}^{L} p(y_{l,j}|H_i)$, for all i, j. This assumption is due to the fact that the fusion center gets on detection result from each sensor in each round such that the interval between two successive detections at the same sensor can be assumed long enough to satisfy the above assumption. Finally, from Assumptions 1 and 4,

$$p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N|H_i) = \prod_{j=1}^{N} p(\mathbf{y}_j|H_i) = \prod_{l=1}^{L}\prod_{j=1}^{N} p(y_{l,j}|H_i). \quad (6)$$

The $M \times N$ code matrix must be replaced with an $M \times L \times N$ code matrix. An $L \times N$ codeword is then assigned to each of the M hypotheses. Restated, the 2-D $L \times N$ codeword $\mathbf{C}_i = [c_{i,l,j}]_{1\le l\le L, 1\le j\le N}$, corresponds to the ith hypothesis, H_i, where $c_{i,l,j} \in \{0,1\}$ is a one-bit symbol. By only considering the hard decision, $v_{l,j}$, of $\tilde{v}_{l,j}$, the probability of misclassification is given by

$$P_e = \sum_{i,\mathbf{V}} \int_{\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_N} p(\mathbf{V}, \mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N, H_i) C_{\mathbf{V},\mathbf{C}_i} \, d\mathbf{y}_1 d\mathbf{y}_2 \ldots d\mathbf{y}_N, \quad (7)$$

where the definition of $C_{\mathbf{V},\mathbf{C}_i}$ is the same as that of $C_{\mathbf{v},\mathbf{c}_i}$ in Section II and $\mathbf{V} = [v_{l,j}]_{1\le l\le L, 1\le j\le N}$. From (6), (7) can be regarded as (1) with a $M \times LN$ code matrix. Therefore, the local decision rule of sensor j can be derived as (3) except that the code matrix $\mathbf{T}$ is of size $M \times LN$.

The maximum-likelihood (ML) soft-decision decoding rule is applied at the fusion center, as in DCSD. That is, the received vector at the fusion center is decoded as the ith hypothesis if $p\left(\tilde{\mathbf{V}}|\mathbf{C}_i\right) \ge p\left(\tilde{\mathbf{V}}|\mathbf{C}_k\right)$ for $\mathbf{C}_k, k = 1, 2, ..., M$, where $\tilde{\mathbf{V}} = [\tilde{v}_{l,j}]_{1\le l\le L, 1\le j\le N}$. From Assumptions 2 and 3, the inequality can be rewritten as $\prod_{l=1}^{L}\prod_{j=1}^{N} p(\tilde{v}_{l,j}|c_{i,l,j}) \ge \prod_{l=1}^{L}\prod_{j=1}^{N} p(\tilde{v}_{l,j}|c_{k,l,j})$. Since $\tilde{v}_{l,j}$ does not depend on $c_{i,l,j}$ given $u_{l,j}$, the above equation can be expanded to

$$\prod_{l=1}^{L}\prod_{j=1}^{N}\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)\, p(u_{l,j}=b_u|c_{i,l,j}) \ge \prod_{l=1}^{L}\prod_{j=1}^{N}\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)\, p(u_{l,j}=b_u|c_{k,l,j})$$

$$\Rightarrow \sum_{l=1}^{L}\sum_{j=1}^{N} \ln \frac{\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)\,(u_{l,j}=b_u|c_{i,l,j})}{\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)\,(u_{l,j}=b_u|c_{k,l,j})} \ge 0. \quad (8)$$

Because $c_{i,l,j}$ and $c_{k,l,j}$ are binary, the bit reliability of the received matrix at the fusion center can be defined as $\lambda_{l,j} = \ln \frac{\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)(u_{l,j}=b_u|c_{i,l,j}=0)}{\sum_{b_u=0}^{1} p(\tilde{v}_{l,j}|u_{l,j}=b_u)(u_{l,j}=b_u|c_{k,l,j}=1)}$. Equation (8) is then equivalent to $\sum_{l=1}^{L}\sum_{j=1}^{N}\left[(-1)^{c_{i,l,j}}\lambda_{l,j} - (-1)^{c_{k,l,j}}\lambda_{l,j}\right] \ge 0$, i.e., $\sum_{l=1}^{L}\sum_{j=1}^{N}\left[\lambda_{l,j} - (-1)^{c_{i,l,j}}\right]^2 \le \sum_{l=1}^{L}\sum_{j=1}^{N}\left[\lambda_{l,j} - (-1)^{c_{k,l,j}}\right]^2$.

To obtain $\lambda_{l,j}$, $p(\tilde{v}_{l,j}|u_{l,j}=b_u)$, which is related to the characteristics of $\alpha_{l,j}$, must be calculated first. For instance, if $\alpha_{l,j}$ in (5) is Rayleigh-distributed [7], then $p(\alpha_{l,j}) = \frac{2\alpha_{l,j}}{E[\alpha_{l,j}^2]} \exp\left(\frac{-\alpha_{l,j}^2}{E[\alpha_{l,j}^2]}\right)$, $\alpha_{l,j} \ge 0$, where $E\left[\alpha_{l,j}^2\right]$ is the expected value of $\alpha_{l,j}^2$. Hence, from (5), $p(\tilde{v}_{l,j}|u_{l,j}=b_u) = \int_0^\infty \frac{p(\alpha_{l,j})}{\sqrt{\pi N_0}} \exp\left(-\frac{\tilde{v}_{l,j} - \alpha_{l,j}(-1)^{b_u}\sqrt{\frac{E_s}{L}}}{N_0}\right) d\alpha_{l,j}$. Furthermore, as described in [5], $p(u_{l,j} = b_u|b_c) = \sum_{i=1}^{M} p(u_{l,j} = b_u|H_i) p(H_i|b_c)$, $b_c \in \{0,1\}$, where $p(u_{l,j} = b_u|H_i)$ can be determined from the local decision rule at sensor j and $p(H_i|b_c) = \frac{p(b_c|H_i)}{\sum_{k=1}^{M} p(b_c|H_k)}$. Similar to $p(u_{l,j} = b_u|c_{i,l,j})$, $p(u_{l,j} = b_u|c_{k,l,j})$ can also be obtained.

The above derivation shows that the 2-D coded detection scheme is equivalent to DCSD with LN sensors and 1-bit symbols. This finding can be easily verified by replacing the subscripts (l,j) with j and $\sum_{l=1}^{L}\sum_{j=1}^{N}$ with $\sum_{j=1}^{LN}$. However, since the size of the code matrix becomes $M \times LN$ and then the number of candidate code matrices is 2^{MLN}, the simulated annealing approach [2] takes much longer to derive a near-optimal code matrix.

B. Optimal code matrix design

If several candidate code matrices correspond to the same probability of misclassification, only one of them had to be checked, while the other could be excluded from the optimal code matrix search. The number of the candidate code matrices can be significantly reduced as revealed in the following theorem.

Theorem 1: Based on the assumptions in Section II and assuming the transmission channel is symmetric, if two columns of a code matrix are exchanged or any column of a code matrix is taken one's complement to form a new code matrix, then the

[2] Soft-decision decoding is adopted such that the elements of $\tilde{\mathbf{V}}$ are real numbers.

new matrix and the original matrix have the same probability of misclassification.

Proof: From the definition of the cost function and (1), if two columns of the code matrix are exchanged, neither P_e nor the result of the cost function change. Next, the one's complement of column k of a code matrix $\mathbf{T} = \left(\mathbf{c}_1^T, \mathbf{c}_2^T, \ldots, \ldots, \mathbf{c}_M^T\right)^T$ is computed, and the new code matrix is obtained as $\bar{\mathbf{T}} = \left(\bar{\mathbf{c}}_1^T, \bar{\mathbf{c}}_2^T, \ldots, \bar{\mathbf{c}}_M^T\right)^T$, where $\mathbf{c}_i = (c_{i,1}, c_{i,2}, \ldots, c_{i,N})$, $\bar{\mathbf{c}}_i = (c_{i,1}, c_{i,2}, \ldots, \bar{c}_{i,k}, \ldots, c_{i,N})$, and $\bar{c}_{i,k}$ represents the one's complement of $c_{i,k}$. From the definition of the cost function, we get (a) $C_{\mathbf{v}_{k=0},\bar{\mathbf{c}}_i} = C_{\mathbf{v}_{k=1},\mathbf{c}_i}$ and $C_{\mathbf{v}_{k=1},\bar{\mathbf{c}}_i} = C_{\mathbf{v}_{k=0},\mathbf{c}_i}$. The minimum probability of misclassification using $\mathbf{T}$ is denoted as $P_{e,min}(\mathbf{T}) = P_e\left(\mathbf{T}, D_1^{opt}(\mathbf{T}), D_2^{opt}(\mathbf{T}), \ldots, D_k^{opt}(\mathbf{T}), \ldots, D_N^{opt}(\mathbf{T})\right)$, where $D_j^{opt}(\mathbf{T})$ is the optimum local decision at sensor j. The same local decision rules as above are now utilized, but with local decision rule $D_k(\bar{\mathbf{T}})$ and with $\bar{\mathbf{T}}$ as the code matrix. Because the transmission channel is symmetric, (b) $p(v_k = 0|u_k = 0) = p(v_k = 1|u_k = 1)$ and $p(v_k = 0|u_k = 1) = p(v_k = 1|u_k = 0)$ results. That is, $p_{00}^k = p_{11}^k$ and $p_{01}^k = p_{10}^k$. From (4), $D_k(\bar{\mathbf{T}}) = -D_k^{opt}(\mathbf{T})$. Consequently, (3) indicates that (c) the value of $p(u_k = 1|y)$ ($p(u_k = 0|y)$) using $D_k(\bar{\mathbf{T}})$ ($D_k(\bar{\mathbf{T}})$) equals the value of $p(u_k = 0|y)$ ($p(u_k = 1|y)$) using $D_k^{opt}(\mathbf{T})$ ($D_k^{opt}(\mathbf{T})$). From (a), (b), (c), and (2), $P_e\left(\bar{\mathbf{T}}, D_1^{opt}(\mathbf{T}), D_2^{opt}(\mathbf{T}), \ldots, D_k(\bar{\mathbf{T}}), \ldots, D_N^{opt}(\mathbf{T})\right) = P_{e,min}(\mathbf{T})$. Since the other local decision rules are optimal for $\mathbf{T}$, not $\bar{\mathbf{T}}$, $P_{e,min}(\mathbf{T}) = P_e\left(\bar{\mathbf{T}}, D_1^{opt}(\mathbf{T}), D_2^{opt}(\mathbf{T}), \ldots, D_k(\bar{\mathbf{T}}), \ldots, D_N^{opt}(\mathbf{T})\right) \geq P_{e,min}(\bar{\mathbf{T}})$. Similarly, $P_{e,min}(\bar{\mathbf{T}}) \geq P_{e,min}(\mathbf{T})$ can be derived. Therefore, $P_{e,min}(\bar{\mathbf{T}}) = P_{e,min}(\mathbf{T})$. □

Every column of a code matrix has 2^M combinations. According the above theorem, any combination has the same characteristics as its one's complement does. Moreover, neither all 0's nor all 1's can differentiate one hypothesis from the others. Thus, there are only $(2^M - 2)/2 = 2^{M-1} - 1$ useful combinations with different characteristics. Next, since the column interchange of a code matrix does not modify P_e, the number of candidate code matrices to be checked is given by $B = \frac{(LN+2^{M-1}-2)!}{(LN)!(2^{M-1}-2)!}$. Thus, $B \ll 2^{MLN}$. For example, when $M = 4$ and $LN = 10$, the number of matrices to be searched is reduced from 1.1×10^{12} to only 8008. Hence, this rule has made the exhaustive search possible to find an optimal code matrix.

C. Code matrix using repetition codes and asymptotic performance analysis

Since L local decisions need L sets of thresholds in each sensor, the complexity of the sensor is increased in the proposed 2-D scheme. This work further proposes a repetition code as a component code of the proposed 2-D coded scheme instead of the optimal code. Restated, code bits $c_{i,j,k}$ are the same for all $1 \leq j \leq L$ in the 2-D codeword $\mathbf{C}_i$ corresponding to hypothesis H_i, where $1 \leq i \leq M$ and $1 \leq k \leq N$. Thus, each sensor needs only one set of thresholds. The sensor applies the same threshold set to make a local decision for each detection at each sensor. Therefore, the code matrix and the local decision rule can be designed to be the same as those in DCSD with $L = 1$. That is, one dimension of a codeword is the original code and the other is a repetition code with length L. Consequently, the overall codeword $\mathbf{C}_i$ can be derived from $\mathbf{c}_i$ by repeating each component of $\mathbf{c}_i$ L times.

The proposed scheme with the repetition code is more flexible, as well as cheaper, than that with the optimal code matrix proposed in Section III-A. In the proposed scheme with the optimal code matrix, the maximum number of detections, and the threshold set for each detection must be determined before deploying the sensor. Since the environment may vary very widely, the detection result with the pre-defined parameters may not be able to reach the required probability of misclassification. Conversely, the fusion center can ask the local sensors to make required number of detections to achieve the required probability of misclassification in the proposed scheme with the repetition code. No extra threshold sets are required in this case.

In DEFECC, the probability of misclassification approaches zero as $N \to \infty$ if $d_{min}^{(N)}$ of the $M \times N$ code matrix meets

$$\delta_N \equiv \frac{1}{N}\left\lfloor \frac{(d_{min}^{(N)} - 1)}{2} \right\rfloor > \max_{1 \leq i \leq M} \epsilon_i, \tag{9}$$

where $d_{min}^{(N)}$ is the minimum Hamming distance of the $M \times N$ code matrix and ϵ_i is the probability of making a wrong decision given H_i for each sensor. By following the similar argument, the probability of misclassification in the proposed scheme approaches zero as $L \to \infty$ or $N \to \infty$ if $d_{min}^{(LN)}$ of the $M \times LN$ code matrix satisfies $\delta_{LN} \equiv \frac{1}{LN}\left\lfloor \frac{(d_{min}^{(LN)}-1)}{2} \right\rfloor > \max_{1 \leq i \leq M} \epsilon_i$.

By (9), δ_N can reasonably be used as a performance metric in any DEFECC-like scheme. The following proof demonstrates that if $L \geq 1$, then $\delta_{LN} \geq \delta_N$ such that the proposed scheme is expected perform better than DEFECC when channel SNR (CSNR) is high. The largest minimum Hamming distance of any code matrix with length LN, is clearly at least L times that of any code matrix with length N. This rule can be shown using a repetition code in the proposed 2-D scheme. Since $\delta_{LN} = \frac{1}{LN}\left\lfloor \frac{(d_{min}^{(LN)}-1)}{2} \right\rfloor \geq \frac{1}{LN}\left\lfloor \frac{L d_{min}^{(N)}-1}{2} \right\rfloor > \frac{1}{N}\left\lfloor \frac{d_{min}^{(N)}-1}{2} \right\rfloor$ when $L \geq 1$, $\delta_{LN} \geq \delta_N$. This rule also indicates that using a repetition code as a component code of the proposed scheme yields a better performance than DEFECC. A similar conclusion can be made when faulty sensors occur in the system.

IV. Performance Evaluation

The proposed scheme was evaluated using several simulations, each comprising 10^6 Monte Carlo tests. Four hypotheses H_1, H_2, H_3, and H_4, were detected and classified with $N = 10$ sensors, $L = 2$, and a fusion center. These hypotheses were assumed to have Gaussian-distributed probability density functions with the same standard deviation σ^2 and means 0, 1, 2, and 3, respectively. At each sensor, observation SNR (OSNR) was defined as $-10 \times \log_{10} \sigma^2$. The attenuation factors $\alpha_{l,j}$ had identical and independent Rayleigh distributions with $E\left[\alpha_{l,j}^2\right] = 1$. Furthermore, $E[\sigma_{l,j}^2] = 1$ was assumed, and

TABLE II
THE $4 \times 2 \times 10$ OPTIMAL CODE MATRIX

Hypothesis	First Detection										Second Detection									
H_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H_2	0	0	0	0	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1
H_3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
H_4	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0

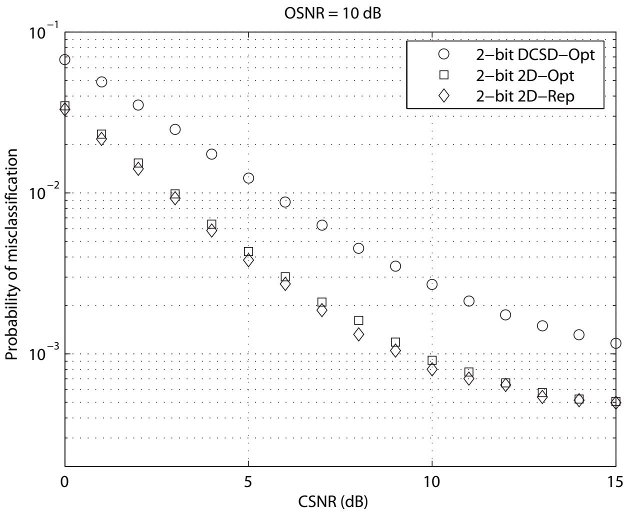

Fig. 2. Performance comparison of 1-bit DCSD-Opt, 2-bit DCSD-Opt, 2-bit 2D-Opt, and 2-bit 2D-Rep at OSNR = 10 dB.

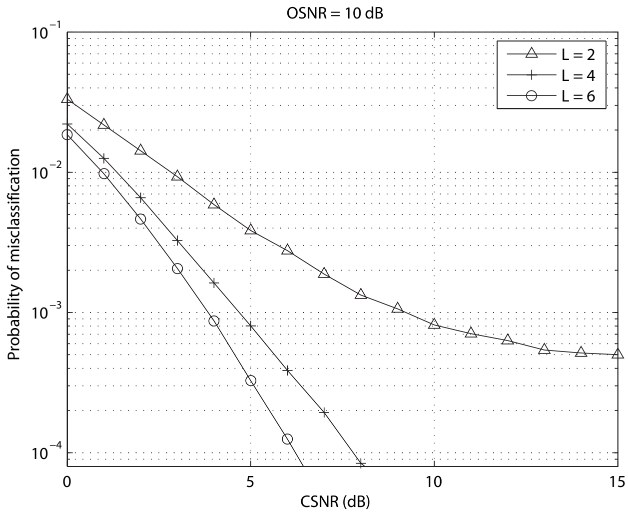

Fig. 3. Performance comparison of 2-bit, 4-bit, and 6-bit 2D-Rep at OSNR = 10 dB,

then CSNR is $10 \times \log_{10}(E_s/N_0)$. The performance was measured using the probability of misclassification, P_e.[3] For $L = 2$, the number of all candidate code matrices is $2^{80} \approx 1.2 \times 10^{24}$. If the minimum Hamming distance must be greater than or equal to 10 in order to enhance the fault-tolerance capability of the system, then the full search only check 6668 candidate matrices. Table II presents the optimal code matrix.

The performance of the following detection schemes is compared: DCSD using the code matrix in Table I (DCSD-Opt), the proposed scheme using the code matrix in Table II (2D-Opt), and the proposed scheme using the code matrix in Table I and the repetition code (2D-Rep). In the first set of simulations, the OSNR was set to be 10 dB, and the CSNRs varied from 0 dB to 15 dB. Figure 2 lists the simulation results. The proposed scheme (2D-Opt or 2D-Rep) outperforms DCSD. For example, the proposed scheme has a gain of more than 5 dB over the DCSD with the optimal code matrix on CSNR when OSNR is 10 dB and the probability of misclassification is 0.001. Moreover, the proposed scheme using the repetition code performs almost identical to that of the proposed scheme using the optimal code matrix.

The second set of simulations demonstrate that the proposed scheme with 2-D repetition codes could reduce the misclassification probability to a very small value as the number of observations increases. The results are shown in Figure 3.

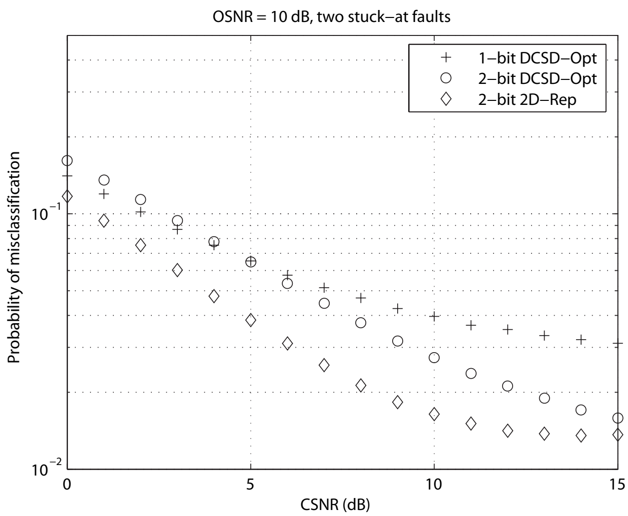

Fig. 4. Performance comparison of 1-bit DCSD-Opt, 2-bit DCSD-Opt, and 2-bit 2D-Rep at OSNR = 10 dB when two stuck-at faults are present.

For example, the misclassification probability can be reduced to 10^{-4} at CSNR= 6 dB when OSNR is 10 dB and 6 decisions of observations for each local sensor are transmitted to the fusion center; however, the misclassification probability cannot be reduced to such small value when only 2 decisions of observations for each local sensor are transmitted to the fusion center. Finally, simulations were conducted with faulty sensors. Figure 4 illustrates the performance when sensors 1 and 6 with stuck-at faults always sent out 1. The proposed scheme using the repetition code still outperformed other

[3]Even though we do not present the simulation results here it can be shown that OSNR has larger impact on the misclassification probability than CSNR. As it will be shown later in this section, to perform more decisions of observations at each sensor effectively increases OSNR at each sensor such that smaller misclassification probability can be obtained.

methods. Even though not demonstrated here, when sensors 1 and 6 with random faults randomly sent out 0 or 1 to the fusion center with the same probability, the proposed scheme using the repetition code also outperformed other methods.

V. CONCLUSIONS

This work presents a 2-D coding scheme to reduce the probability of misclassification in wireless sensor networks. A method for determining the optimal code matrix for the proposed scheme is also described. By using the optimal code, the proposed scheme can classify phenomena more reliably than DCSD in [2] and also has a lower memory requirement at each sensor. The code matrix uses repetition codes rather than the optimal code to further decrease the sensor cost. The performance penalty for using repetition codes to design the code matrix was found to be small. Additionally, the proposed scheme performed better than DCSD in the presence of faulty sensors since its code matrix has a larger minimum Hamming distance.

REFERENCES

[1] S. A. Aldosari and J. M. F. Moura, "Detection in decentralized sensor networks," in *Proc. ICASSP 2004*, Montreal, Canada, May 2004, pp. ii-277–280.

[2] 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.

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

[4] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. New York, NY: Elsevier, 1977.

[5] T.-Y. Wang, Y. S. Han, B. Chen, and P. K. Varshney, "A combined decision fusion and channel coding scheme for distributed fault-tolerant classification in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1695–1705, July 2006.

[6] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Norwell, MA: Reidel, 1982.

[7] J. G. Proakis, *Digital Communication*, 4th ed., New York, NY: McGraw-Hill, 2001.