

# Distributed Multi-target Classification in Wireless Sensor Networks

Jayesh H. Kotecha, Vinod Ramachandran and Akbar M. Sayeed\*

Department of Electrical and Computer Engineering

University of Wisconsin-Madison

jkotecha@ece.wisc.edu, vinod@cae.wisc.edu, akbar@enr.wisc.edu

<http://dune.ece.wisc.edu>

## Abstract

We address the problem of distributed classification of multiple targets in a wireless sensor network. The maximum number of targets is known *a priori* but the actual number of distinct targets present in any given event is assumed unknown. Feature vectors from noise-corrupted measurements are collected at nodes in the region of interest. The target signal components of feature vectors are modeled as zero-mean Gaussian vectors. Nodes that are sufficiently separated in space provide independent measurements that can be exploited for improving classification performance. In the proposed distributed classifiers, hard decisions are made at each node which are then communicated to a manager node for the final decision via noise-free or noisy communication links. A natural way to make local decisions is to use the optimal classifier. A key problem with the optimal classifier is that the number of hypotheses increases exponentially with the number of targets. We propose two sub-optimal (mixture density and Gaussian) local classifiers that are based on a natural but coarser re-partitioning of the hypothesis space, resulting in linear complexity with the number of targets. The performance of the distributed algorithms depends on four key quantities: the target signal statistics, the measurement SNR, the communication SNR, and the number of independent sensor measurements. We analyze the asymptotic performance (in the number of measurements) of the classifiers and show that non-vanishing error exponents can be guaranteed under mild conditions. The performance of the optimal centralized classifier is analyzed and serves as a benchmark for comparing the distributed classifiers. Numerical results based on real acoustic data illustrate the performance of the proposed distributed classifiers and the advantage of independent sensor measurements for multi-target classification. In particular, the performance of the sub-optimal mixture density classifier is very close to the distributed optimal classifier, making it an attractive choice in practice.

---

\*This work was supported by the DARPA SensIT program under Grant F30602-00-2-0555.

# 1 Introduction

Wireless sensor networks promise an unprecedented opportunity to monitor the physical world via cheap wireless nodes that can sense the environment in multiple modalities, including acoustic, seismic, and infra-red [1, 2]. Distributed decision making is an important application of sensor networks; for example, the detection and classification of objects in the sensor field. Due to a variety of factors, such as measurement noise and statistical variability in target signals, collaborative processing of multiple node measurements is necessary for reliable decision making.

In practical implementation, collaborative processing of sensor measurements collected in a region of interest is typically coordinated by a manager node (see, e.g., [3]). Given the limited communication ability of sensor nodes (due to limited transmit power or energy, e.g.), a key goal in developing collaborative signal processing (CSP) algorithms is to transmit the least amount of data from the sensing nodes to the manager nodes. With this goal in mind, in this paper we extend our earlier work on CSP algorithms for single target classification [4, 5] to the more challenging multiple target setting.

Consider a network query regarding the classification of multiple targets present in a region of interest. We assume that the maximum number ( $M$ ) of *distinct* targets is known *a priori*. However, the actual number of distinct targets present in a given event is unknown. Thus, the multi-target classification problem corresponds to a  $N$ -ary hypothesis testing problem with  $N = 2^M$  hypotheses corresponding to the various possibilities for the presence or absence of each target. Thus, the complexity of the optimal classification algorithms (centralized or decentralized) increases exponentially with the number of targets  $M$ . To circumvent the exponential complexity of optimal algorithms, we propose two sub-optimal approaches to distributed multi-target classification based on a natural re-partitioning of the hypothesis space: the sub-optimal classifiers perform  $M$  (rather than  $N = 2^M$ ) tests to determine the presence or absence of each distinct target. Our results show that the sub-optimal mixture Gaussian classifier, that is optimal for the re-partitioned hypothesis space, delivers competitive performance despite its linear complexity.

The proposed algorithms are based on modeling each target as a point source whose temporal signal characteristics can be modeled as a zero-mean Gaussian process [4, 5, 6]. Each target gener-

ates a Gaussian space-time signal field whose statistical characteristics have a profound impact on classifier performance. In particular, the region of interest containing the targets can be divided into spatial coherence regions (SCR's) over which the spatial signal field remains strongly correlated. The size of the SCR's is inversely proportional to the target signal bandwidth: high-bandwidth targets induce smaller SCR's whereas low-bandwidth targets induce larger SCR's [5, 6]. A very important property of the SCR's is that the spatial signal in distinct SCR's is approximately uncorrelated (independent in the Gaussian case). Thus, the number of SCR's in the query region of interest determines the number of *independent* spatial measurements that can be collected at any given time. The sensor nodes sample the spatio-temporal signal field generated by targets.

There are two main sources of error in distributed decision making. 1) sensor measurement noise, and 2) the inherent statistical variability in the signal. The above modeling of the space-time target signal field in terms of SCR's suggests a natural structure on CSP to mitigate the two sources of error [4, 6]. First, since all nodes within each SCR sense a highly correlated target signal, the node measurements in each SCR can be aggregated to improve the effective measurement SNR. Second, the independent node measurements from distinct SCR's can be combined to reduce the impact of inherent variability in the target signal. Furthermore, since the node measurements in distinct SCR's are approximately independent, local hard decisions can be first formed in each SCR and then the lower-dimensional decisions can be communicated to the manager node to make the final decision.

The basic architecture of the proposed distributed classification algorithms is illustrated in Fig. 1. We assume that there are  $G$  SCR's in the query region of interest. In a given snapshot, an  $N_o$ -dimensional feature vector  $\mathbf{z}_k$  is collected in the  $k$ -th SCR. The optimal centralized classifier that directly operates on  $\mathbf{z}_k$  provides an upper bound on the performance of any classification (centralized or decentralized) algorithm. A local decision  $u_k$  about the  $N$  hypotheses is made in each SCR based on the local independent feature vector  $\mathbf{z}_k$ . We consider three local classification algorithms for generating  $\{u_k\}$ : the optimal local classifier, and two sup-optimal classifiers based on a natural re-partitioning of the hypothesis space (mixture Gaussian (MG) classifier and single Gaussian (SG) classifier). The local decisions are then communicated over ideal (noise-free) or noisy links to the manager nodes that optimally fuses them to make the final decision. Thus, we con-

sider three distributed classification algorithms based on the three local classification algorithms (optimal, MG, SG). The performance of the distributed algorithms is also compared with the optimal centralized classifier that serves as a benchmark. Our results indicate that the sub-optimal linear-complexity MG distributed classifier performs nearly the same as the optimal exponential complexity decentralized classifier.

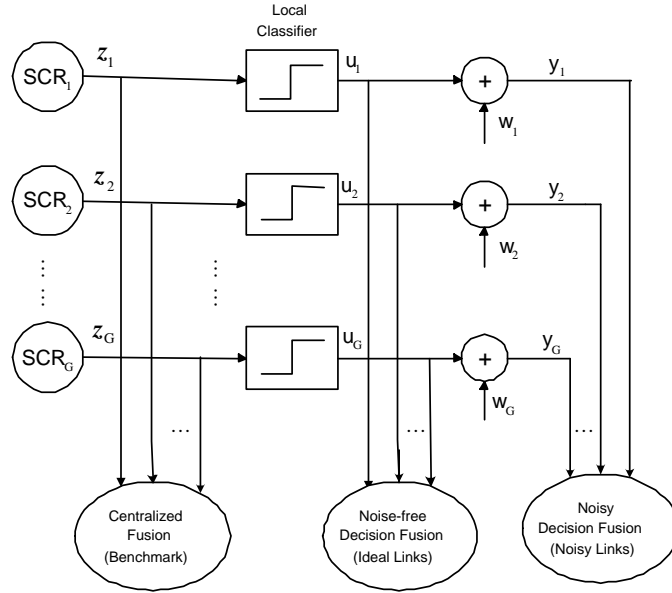


Figure 1: A generic architecture for multi-target classification algorithms. All classifiers operate on  $G$  independent, identically distributed feature vectors  $\{z_1, \dots, z_G\}$  collected in distinct spatial coherence regions (SCRs). Centralized fusion corresponds to optimal centralized classification that serves as a benchmark for all classification algorithms. In the proposed distributed classification algorithms, local hard decisions  $\{u_k\}$  about the hypotheses are generated in the different SCR's. In ideal decision fusion, the local decisions are communicated to the manager noise via noise-free communication links. In noisy decision fusion, the local decisions are communicated over noisy communication links. In both ideal and noisy decision fusion, the manager node optimally combines the received local decisions to make the final decision. Three different types of local classifiers are considered for generating the local hard decisions  $\{u_k\}$ : optimal classifier, sub-optimal mixture Gaussian classifier, and sub-optimal single Gaussian classifier. The sub-optimal local classifiers have linear complexity in the number of targets compared to the optimal classifier which has exponential complexity.

In the next section we formulate the multi-target classification problem. In Section 3 we analyze the performance of the optimal centralized classifier that serves as a benchmark and also facilitates analysis of the distributed classification algorithms. Section 4 discusses the three different local classifiers for generating the hard decisions in each SCR. In Section 5 we present the structure of the final classifier at the manager node that optimally fuses the decisions from the SCR's. Section 6 presents some numerical results and Section 7 provides some concluding remarks.

$H_j$	$b_2$	$b_1$	$\pi_j$
$H_0$	0	0	$q_2 q_1$
$H_1$	0	1	$q_2(1 - q_1)$
$H_2$	1	0	$(1 - q_2)q_1$
$H_3$	1	1	$(1 - q_2)(1 - q_1)$

Table 1: Hypothesis space for  $M = 2$  targets.

## 2 Problem Formulation

The problem of interest is the classification of multiple target/objects in a query region of the sensor field. In a practical scenario, a query for target classification will usually be preceded by a query for target detection. Target detection can be accomplished reliably with distributed energy detectors (see, e.g., [3]). However, in this paper we include the null hypothesis (no targets present) in the formulation. We assume that *at most*  $M$  *distinct* targets are present in the query region during the duration of interest. However, the actual number of targets present is unknown. Consequently, there are  $N = 2^M$  possible hypotheses, denoted by  $H_j$ ,  $j = 0, \dots, N - 1$ , corresponding to different possibilities for the presence or absence of different targets. The probability of the  $m$ -th target being present is assumed to be  $1 - q_m$  and is independent of other targets. Let  $b_m(j)$  denote the presence ( $b_m(j) = 1$ ) or absence ( $b_m(j) = 0$ ) of the  $m$ -th target in the  $j$ -th hypothesis. The prior probabilities of the different hypotheses are given by

$$\pi_j = P(H_j) = \prod_{m=1}^M [b_m(j)(1 - q_m) + (1 - b_m(j))q_m], \quad j = 0, \dots, N - 1, \quad (1)$$

where  $b_M(j) b_{M-1}(j) \dots b_1(j)$  is the binary representation of the integer  $j$ ;  $H_0$  corresponds to no target being present, whereas  $H_{N-1}$  corresponds to a all targets being present. See Table 1 for an example with  $M = 2$ .

As mentioned in the previous section, the final decision about the hypotheses is made at the manager node based on  $G$  i.i.d. effective feature vectors  $\{\mathbf{z}_k\}$  collected in  $G$  distinct SCR's. Each feature vector is of dimension  $N_o$ . The signal component of the feature vector corresponding to  $m$ -th target is modeled as a zero-mean complex Gaussian vector with covariance matrix  $\Sigma_m$ . The energy in each target is assumed to be the same; that is  $\text{tr}(\Sigma_m) = \sigma_s^2$  for all  $m$ . It follows that the signal corresponding to each  $H_j$  is also Gaussian whose covariance matrix is the sum of the

covariance matrices of the targets present. Thus, the multi-target classification problem can be stated as the following  $N$ -ary hypothesis testing problem

$$\begin{aligned}
H_j &: \mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k, \quad k = 1, \dots, G; \quad j = 0, \dots, N-1 \\
\mathbf{s}_k &\sim \mathcal{CN}(\mathbf{0}, \bar{\boldsymbol{\Sigma}}_j), \quad \bar{\boldsymbol{\Sigma}}_j = \sum_{m=1}^M b_m(j) \boldsymbol{\Sigma}_m, \quad \mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}).
\end{aligned} \tag{2}$$

Under  $H_j$ , the probability density function (pdf) of the feature vector at the  $k$ -th SCR is

$$p_j(\mathbf{z}_k) = p(\mathbf{z}_k | H_j) = \frac{1}{\pi^{N_o} |\tilde{\boldsymbol{\Sigma}}_j|} e^{-\mathbf{z}_k^H \tilde{\boldsymbol{\Sigma}}_j^{-1} \mathbf{z}_k}, \quad \tilde{\boldsymbol{\Sigma}}_j = \bar{\boldsymbol{\Sigma}}_j + \sigma^2 \mathbf{I}. \tag{3}$$

and the  $G$  feature vectors are i.i.d.

In the rest of the paper, we discuss distributed classification algorithms for the multi-target hypothesis testing problem defined in (2). The basic architecture of the classification algorithms is illustrated in Fig. 1. First, a local decision ( $u_k$ ) is made in each SCR based on its local independent feature vector  $\mathbf{z}_k$ . The local decisions from the  $G$  SCR's are communicated to the manager node (via dedicated noise-free or noisy communication links), where they are fused to obtain the final decision.

In the next section, we review the optimal *centralized* classifier whose performance serves as an ultimate upper bound for the performance of any distributed classifier.

### 3 Optimal Centralized Classifier

The optimal classifier chooses the hypothesis according to

$$C(\mathbf{z}_1, \dots, \mathbf{z}_G) = \arg \max_{j=0, \dots, N-1} p_j(\mathbf{z}_1, \dots, \mathbf{z}_G) \pi_j \tag{4}$$

$$\text{where } p_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = p(\mathbf{z}_1, \dots, \mathbf{z}_G | H_j) = \prod_{k=1}^G p_j(\mathbf{z}_k) \tag{5}$$

due to the independence of measurements  $\mathbf{z}_k$ . Using logarithms, the above problem is written as

$$C(\mathbf{z}_1, \dots, \mathbf{z}_G) = \arg \min_{j=0, \dots, N-1} l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) \tag{6}$$

$$\text{where } l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = -\frac{1}{G} \log p_j(\mathbf{z}_1, \dots, \mathbf{z}_G) \pi_j = -\frac{1}{G} \sum_{k=1}^G \log p_j(\mathbf{z}_k) - \frac{1}{G} \log \pi_j. \tag{7}$$

Ignoring constants that do not depend on the class,  $l_j$  takes the form

$$l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = \log |\tilde{\Sigma}_j| + \frac{1}{G} \sum_{k=1}^G \mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k - \frac{1}{G} \log \pi_j. \quad (8)$$

Note that implementation of the optimal centralized classifier requires that the  $k$ -th SCR communicates the local log-likelihood functions for the  $N$  hypotheses,  $\{\mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k, j = 0, \dots, N-1\}$ , to the manager node. The classifier at the manager node then computes  $l_j$  in (8) for  $j = 0, \dots, N-1$ , and makes the decision according to (6).

### 3.1 Performance of Optimal Centralized Classifier

We quantify classifier performance in terms of the average probability of error

$$P_e(G) = \sum_{m=0}^{N-1} P_{e,m}(G) \pi_m, \quad P_{e,m}(G) = P(l_j < l_m \text{ for some } j \neq m | H_m) \quad (9)$$

where  $P_{e,m}(G)$  is the conditional error probability under  $H_m$  and is explicitly written as a function of  $G$ . Computing  $P_{e,m}$  is complicated in general but we can bound it using the union bound [7]

$$P_{e,m}(G) \leq \sum_{j=0, j \neq m}^{N-1} P(l_j < l_m | H_m). \quad (10)$$

Thus,  $P_e$  can be upperbounded as

$$P_e(G) \leq \sum_{m=0}^{N-1} \sum_{j=0, j \neq m}^{N-1} P(l_j < l_m | H_m) \pi_m. \quad (11)$$

For each pair  $j, m$ , the pairwise error probability (PEP),  $\text{PEP}_{jm}(G) = P(l_j < l_m | H_m)$  depends on a decision statistic that is a weighted sum of  $NG$   $\chi_2^2$  random variables [4]. The pdf and distribution function of the statistic can be computed in closed-form but take on tedious expressions [7]. Chernoff bounding techniques can be used to obtain tight bounds for the PEP's. We state some well-known results (see, e.g., [7, 8]) in the context of our set up. Let  $E_m[\cdot]$  denote the expectation under  $H_m$ .

**Proposition 1 (Chernoff Bounds)** *For any  $0 \leq \theta \leq 1$  and for all  $G \geq 1$*

$$\text{PEP}_{jm}(G) \leq \left( \frac{\pi_j}{\pi_m} \right)^\theta e^{\mu_{jm}(\theta)G}, \quad (12)$$

$$\mu_{jm}(\theta) = \log E_m \left[ \left( \frac{p_j(\mathbf{Z})}{p_m(\mathbf{Z})} \right)^\theta \right] \leq 0. \quad (13)$$

For any  $G$ , the tightest Chernoff bound for  $\text{PEP}_{j_m}(G)$  can be obtained by minimizing (12) over  $\theta$

$$\text{PEP}_{j_m}(G) \leq \left( \frac{\pi_j}{\pi_m} \right)^{\theta^*(G)} e^{-D_{j_m}^*(G)G} \quad (14)$$

$$D_{j_m}^*(G) = -\mu_{j_m}(\theta^*(G)) \quad , \quad \theta^*(G) = \arg \min_{0 \leq \theta \leq 1} \left[ \frac{\theta}{G} \log \left( \frac{\pi_j}{\pi_m} \right) + \mu_{j_m}(\theta) \right] \quad (15)$$

The tightest error exponent (Chernoff information) in the limit of large  $G$  is given by

$$D_{j_m}^* = \lim_{G \rightarrow \infty} D_{j_m}^*(G) = - \lim_{G \rightarrow \infty} \mu_{j_m}(\theta^*(G)) = - \min_{0 \leq \theta \leq 1} \mu_{j_m}(\theta) \quad (16)$$

A simple non-trivial exponent for all  $G$  is the Bhattacharya bound which corresponds to  $\theta = 1/2$ .

For the optimal classifier,  $\mu_{j_m}(\theta)$  is given by [5]

$$\mu_{j_m}(\theta) = \theta \log \left| \tilde{\Sigma}_m \tilde{\Sigma}_j^{-1} \right| - \log \left| (1 - \theta) \mathbf{I} + \theta \tilde{\Sigma}_m \tilde{\Sigma}_j^{-1} \right| \quad (17)$$

where the above expression holds for all  $\theta \geq 0$  for which  $(1 - \theta) \mathbf{I} + \theta \tilde{\Sigma}_m \tilde{\Sigma}_j^{-1}$  is positive definite.

The minimization in (16) can be easily performed numerically using (17).

Using (11) and (14), we get the following bound for  $P_e(G)$ ,

$$P_e(G) \leq \sum_{m=0}^{N-1} \sum_{j=0, j \neq m}^{N-1} A_{j_m}(G) e^{-D_{j_m}^*(G)G}, \quad (18)$$

where  $A_{j_m}(G) = \pi_j^{\theta^*(G)} \pi_m^{1-\theta^*(G)}$ . Since all  $\text{PEP}_{j_m}$ 's decay exponentially to zero with  $G$ , so does  $P_e(G)$ . In particular, for a given SNR, the decay of  $P_e(G)$  will be dominated by the smallest (worst) error exponent.

**Asymptotic Performance as  $G \rightarrow \infty$ .** Note from (6) that by the law of large numbers, under  $H_m$  we have

$$\lim_{G \rightarrow \infty} l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = - \lim_{G \rightarrow \infty} \frac{1}{G} \sum_{k=1}^G \log p_j(\mathbf{z}_k) - \lim_{G \rightarrow \infty} \frac{1}{G} \log \pi_j \quad (19)$$

$$= -\mathbb{E}_m[\log p_j(\mathbf{Z})] \quad (20)$$

where we note that the second term becomes zero, implying that the effect of the prior probabilities on the decision vanishes in the limit and the convergence in the first term is with probability 1. But

$$-\mathbb{E}_m[\log p_j(\mathbf{Z})] = D(p_m \| p_j) + h_m(\mathbf{Z}) \quad (21)$$

where  $D(p_m \| p_j)$  is the Kullback-Leibler (K-L) distance between the pdfs  $p_j$  and  $p_m$  and  $h_m(\mathbf{Z})$  is the differential entropy of  $\mathbf{Z}$  under  $H_m$  [9]

$$D(p_m \| p_j) = E_m [\log(p_m(\mathbf{Z})/p_j(\mathbf{Z}))] = \log \left( |\tilde{\Sigma}_j|/|\tilde{\Sigma}_m| \right) + \text{tr} \left( \tilde{\Sigma}_j^{-1} \tilde{\Sigma}_m - \mathbf{I} \right) \quad (22)$$

$$h_m(\mathbf{Z}) = -E_m[\log p_m(\mathbf{Z})] = \log \left( (\pi e)^{N_o} |\tilde{\Sigma}_m| \right). \quad (23)$$

Thus, we get

$$\lim_{G \rightarrow \infty} l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) - l_m(\mathbf{z}_1, \dots, \mathbf{z}_G) = D(p_m \| p_j) \quad (24)$$

where  $D(p_m \| p_j) \geq 0$  with equality if and only if  $p_m = p_j$ . This implies that if  $D(p_m \| p_j) > 0$ ,

$$\lim_{G \rightarrow \infty} P(l_j < l_m | H_m) = 0. \quad (25)$$

Thus,  $P_e$  goes to zero as  $G \rightarrow \infty$  if and only if all the pairwise K-L distances are positive. From Proposition 1, we know that each PEP $_{jm}$  goes to zero exponentially with  $G$  if the corresponding exponent is positive. It can be easily shown that the Chernoff exponent for a particular PEP $_{jm}$  is strictly positive if and only if the corresponding K-L distance is strictly positive. Thus, we have the following result on  $P_e$  of the optimal classifier.

**Proposition 2**  $P_e(G)$  decays (exponentially) to zero with  $G$  if and only if all pairwise K-L distances are strictly positive, that is

$$D(p_m \| p_j) > 0 \quad \forall j, m, j \neq m \quad (26)$$

Specifically,

$$\limsup_{G \rightarrow \infty} \frac{\log P_e(G)}{G} \leq -D_{min}^* \quad (27)$$

where<sup>1</sup>  $D_{min}^* = \min\{D_{jm}^*\}$  is the smallest pairwise Chernoff information defined in (16).  $D_{min}^* > 0$  if and only if (26) is satisfied.

The condition in the above result would be satisfied in general except in the limit of very poor SNR. The effect of poor SNR is to reduce the values of the K-L distances.

---

<sup>1</sup>Note that  $\limsup_n x(n) = \inf_n \sup_{k \geq n} x_k$

## 4 Local Classifiers

In the optimal centralized classifier, the  $k$ -th SCR sends  $N$  log-likelihood values  $\{z_k^H \tilde{\Sigma}_j^{-1} z_k : j = 0, \dots, N-1\}$ , computed from its local measurement  $z_k$ , to the manager node (see Fig.reffig:arch). While exchange of real-valued likelihoods puts much less communication burden on the network as compared to *data fusion* in which the feature vectors  $\{z_k\}$  are communicated to the manager node, it is attractive to reduce the communication burden even further. A natural quantization strategy is to compute local *decisions* on the hypothesis in each SCR based on the local measurement  $z_k$ . The local decisions, denoted as  $\{u_k\}$ , from all SCR's are communicated to the manager node which makes the final optimal<sup>2</sup> decision. The local classifier in the  $k$ -th SCR makes a decision on the hypothesis using only its local measurement  $z_k$  according to a given decision rule  $f : \mathcal{C}^{N_o} \rightarrow \{0, 1, \dots, N-1\}$

$$u_k = f(z_k), \quad k = 1, \dots, G. \quad (28)$$

Since all  $\{z_k\}$  are i.i.d., so are  $\{u_k\}$ . The decision statistics are characterized by the probability mass function (pmf),  $\{p_m[j] : j = 0, \dots, N-1\}$ , of the decision  $U$  under different hypotheses

$$p_m[j] = P(U = j | H_m) \quad j, m = 0, \dots, N-1. \quad (29)$$

Note that since  $\{u_k\}$  are i.i.d, the pmfs  $\{p_m[j]\}$  are identical for all  $k$ .

In this section, we discuss three local classifiers which induce three different local decision rules (and corresponding  $\{p_m[j]\}$ ). This first one is the optimal classifier which has exponential complexity in the number of targets  $M$ . The other two are sub-optimal classifiers, based on a re-partitioning of the hypothesis space, which have linear complexity in  $M$ . In the next section, we study the performance of the final decision made at the manager node, considering both ideal (noise-free) and noisy communication links.

---

<sup>2</sup>Optimal, given the statistics of decisions  $\{u_k\}$  and the nature of the communication link.

## 4.1 Optimal Local Classifier

The optimal Bayesian local classifier in the  $k$ -th SCR makes the decision as

$$u_k = \arg \max_{j=0, \dots, N-1} p_j(\mathbf{z}_k) \pi_j, \quad k = 1, \dots, G. \quad (30)$$

The optimal classifier is illustrated in Fig. 2(a). The pmf's of  $U$  under different hypotheses are characterized by the following probabilities

$$p_m[j] = P(p_j(\mathbf{z}_k) \pi_j \geq p_l(\mathbf{z}_k) \pi_l \text{ for all } l \neq j | H_m), \quad j, m = 0, \dots, N-1. \quad (31)$$

The complexity of the optimal Bayes classifier increases exponentially with  $M$  since  $N = 2^M$ . We next propose two sub-optimal classifiers whose complexity grows only linearly with  $M$ .

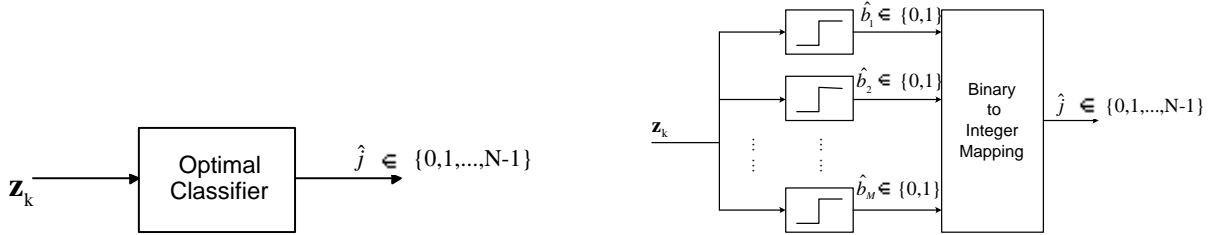


Figure 2: (a) Optimal local classifier. (b) The structure of sub-optimal local classifiers based on the re-partitioning of the hypothesis space via the presence or absence of individual targets.

## 4.2 Sub-optimal Local Classifiers

In the sub-optimal classifiers,  $M$  tests are conducted, one for each target, to determine the presence or absence of the target. Each test partitions the set of hypotheses into two sets,  $\tilde{H}_m$  and  $\tilde{H}_m^c$ , where  $\tilde{H}_m$  contains those hypotheses in which the  $m$ -th target is present, and  $\tilde{H}_m^c$  contains those hypotheses in which it is absent. Let

$$S_m = \{j \in \{0, \dots, N-1\} : b_m(j) = 1\} \quad (32)$$

$$S_m^c = \{j \in \{0, \dots, N-1\} : b_m(j) = 0\} = \{0, \dots, N-1\} - S_m, \quad (33)$$

where recall that  $b_M(j) b_{M-1}(j) \dots b_1(j)$  is the binary representation of the integer  $j$ . Then, the two hypotheses for the  $m$ -th test are

$$\tilde{H}_m = \bigcup_{j \in S_m} H_j \quad (34)$$

$$\tilde{H}_m^c = \bigcup_{j \in S_m^c} H_j = \left\{ \bigcup_{j=0}^{N-1} H_j \right\} - \tilde{H}_m. \quad (35)$$

In the example in Table 1 for  $M = 2$ ,  $\tilde{H}_1 = H_1 \cup H_3$ ,  $\tilde{H}_1^c = H_0 \cup H_2$ ,  $\tilde{H}_2 = H_2 \cup H_3$ , and  $\tilde{H}_2^c = H_0 \cup H_1$ .

Under  $\tilde{H}_m$  and  $\tilde{H}_m^c$ ,  $\mathbf{z}_k$  is distributed as a weighted sum of Gaussians (mixture Gaussian (MG)), that is,

$$p(\mathbf{z}_k | \tilde{H}_m) = \frac{1}{1 - q_m} \sum_{i \in S_m} \pi_i p_i(\mathbf{z}_k) = \frac{1}{1 - q_m} \sum_{i \in S_m} \pi_i \frac{1}{\pi^{N_o} |\tilde{\Sigma}_i|} e^{-\mathbf{z}_k^H \tilde{\Sigma}_i^{-1} \mathbf{z}_k} \quad (36)$$

$$p(\mathbf{z}_k | \tilde{H}_m^c) = \frac{1}{q_m} \sum_{i \in S_m^c} \pi_i p_i(\mathbf{z}_k) = \frac{1}{q_m} \sum_{i \in S_m^c} \pi_i \frac{1}{\pi^{N_o} |\tilde{\Sigma}_i|} e^{-\mathbf{z}_k^H \tilde{\Sigma}_i^{-1} \mathbf{z}_k}. \quad (37)$$

We now present two local classifiers based on this re-partitioning of the hypothesis space. We call the first one a mixture Gaussian classifier (MGC) which is the optimal classifier for the re-partitioned space, and the second one a single Gaussian classifier (SGC) classifier that approximates the mixture densities with a Gaussian density. Recall that each local classifier conducts  $M$  tests to determine the presence or absence of each individual target. Essentially, under  $H_j$ , the  $m$ -th test estimates the value of  $b_m(j)$ . Let  $\hat{b}_m \in \{0, 1\}$  denote the value of the  $m$ -th test. The local decision  $u_k \in \{0, \dots, N - 1\}$  in the  $k$ -th SCR is equal to the integer representation of the binary decisions  $\hat{b}_M \hat{b}_{M-1} \dots \hat{b}_1$ , as illustrated in Fig. 2(b). Let  $f_{bin-int}$  denote the binary-to-integer mapping:  $f_{bin-int}(b_M, b_{M-1}, \dots, b_1) = \sum_{m=1}^M b_m 2^{m-1}$ . Then,  $u_k(\mathbf{z}_k) = f_{bin-int}(\hat{b}_M(\mathbf{z}_k), \hat{b}_{M-1}(\mathbf{z}_k), \dots, \hat{b}_1(\mathbf{z}_k))$ .

In both MGC and SGC, the pmf's of the i.i.d. local decisions at each SCR are characterized by the following probabilities

$$p_m[j] = P(U = j | H_m) = P(\hat{b}_M = b_M(j), \dots, \hat{b}_1 = b_1(j) | H_m) \quad j, m = 0, \dots, N - 1. \quad (38)$$

### 4.2.1 Mixture Gaussian Classifier

This is the optimal classifier for the re-partitioning of the hypothesis space described above. For  $m = 1, \dots, M$ ,  $\hat{b}_m(\mathbf{z}_k)$  denotes the value of the  $m$ -th binary hypothesis test between  $\tilde{H}_m$  and  $\tilde{H}_m^c$  in the  $k$ -th SCR; that is

$$(1 - q_m)p(\mathbf{z}_k|\tilde{H}_m) \underset{\hat{b}_m=0}{\overset{\hat{b}_m=1}{\geq}} q_m p(\mathbf{z}_k|\tilde{H}_m^c) \Leftrightarrow \sum_{i \in S_m} \pi_i p(\mathbf{z}_k|H_i) \underset{\hat{b}_m=0}{\overset{\hat{b}_m=1}{\geq}} \sum_{i \in S_m^c} \pi_i p(\mathbf{z}_k|H_i). \quad (39)$$

Consider the example with  $M = 2$ . It is easy to show that for  $m = 1$ , the test in (39) can be written as be

$$q_2 [(1 - q_1)\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I} + \Sigma_1) - q_1 \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})] + (1 - q_2) [(1 - q_1)\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I} + \Sigma_1 + \Sigma_2) - q_1 \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I} + \Sigma_2)] \underset{\hat{b}_1=0}{\overset{\hat{b}_1=1}{\geq}} 0.$$

A similar expression can be written for  $m = 2$ . Observe that the above test is a weighted sum of two tests. The first expression in square brackets is the Bayes test for detecting target 1 given that target 2 is absent, which is weighted by the probability of the absence of target 2. Similarly, the second expression in square brackets is the Bayes test for detecting target 1 given that target two is present, which is weighted by the probability of the presence of target 2. It can be shown that this interpretation can be generalized for  $M > 2$ .

### 4.2.2 Single Gaussian Classifier

The SGC is obtained by approximating the distributions in (36) and (37) by single Gaussians. This is achieved by preserving the first two moments of the distributions. Thus, we have the approximations

$$\hat{p}(\mathbf{z}_k|\tilde{H}_m) = \frac{1}{\pi^N |\hat{\Sigma}_m|} e^{-\mathbf{z}_k^H \hat{\Sigma}_m^{-1} \mathbf{z}_k}, \quad \hat{\Sigma}_m = \sum_{i \in S_m} \pi_i \tilde{\Sigma}_m \quad (40)$$

$$\hat{p}(\mathbf{z}_k|\tilde{H}_m^c) = \frac{1}{\pi^N |\hat{\Sigma}_{-m}|} e^{-\mathbf{z}_k^H \hat{\Sigma}_{-m}^{-1} \mathbf{z}_k}, \quad \hat{\Sigma}_{-m} = \sum_{i \in S_m^c} \pi_i \tilde{\Sigma}_m. \quad (41)$$

For  $m = 1, \dots, M$ , the value of the  $m$ -th test in the  $k$ -th SCR is given by

$$(1 - q_m)\hat{p}(\mathbf{z}_k|\tilde{H}_m) \underset{\hat{b}_m=0}{\overset{\hat{b}_m=1}{\geq}} q_m \hat{p}(\mathbf{z}_k|\tilde{H}_m^c), \quad (42)$$

or, equivalently, in terms of log-likelihoods, it is given by

$$\hat{l}_m \underset{\hat{b}_m=0}{\overset{\hat{b}_m=1}{\gtrless}} \hat{l}_m^c \quad (43)$$

$$\hat{l}_m = \log(1 - q_m) + \log \hat{p}(\mathbf{z}_k | \tilde{H}_m), \quad \hat{l}_m^c = \log q_m + \log \hat{p}(\mathbf{z}_k | \tilde{H}_m^c). \quad (44)$$

## 5 Fusion of Local Decisions at the Manager Node

In the previous section, we discussed three different local classifiers for generating the local hard decisions  $\{u_k\}$  at the different SCR's: the optimal local classifier, the sub-optimal MG classifier and the sub-optimal SG classifier. In this section, we discuss communication of these local decisions to the manager node and the structure of the corresponding final distributed classifier at the manager node that optimally fuses the decisions. As illustrated in Fig. 1, we consider ideal (noise-free) communication links as well as communication links corrupted by additive Gaussian noise (AWGN). The performance of the final classifier at the manager node can be characterized in a unified fashion for all the three local classifiers. In the case of ideal communication links, the performance of the final classifier is governed by the pmf's of the local decisions,  $\{p_j[m] : m, k = 0, \dots, N - 1\}$ , which are different for the three local classifiers. In the case of noisy communication links, the performance is governed by noisy pdf's under different hypotheses induced by the pmf's of local decisions and the AWGN channel.

Let  $P_{e,opt,ideal}$ ,  $P_{e,mg,ideal}$  and  $P_{e,sg,ideal}$  denote the  $P_e$  of the final classifier under noise-free communication based on the optimal local classifier, sub-optimal MG classifier, and the sub-optimal SG classifier, respectively. Similarly, let  $P_{e,opt,noisy}$ ,  $P_{e,mg,noisy}$  and  $P_{e,sg,noisy}$  denote the  $P_e$  under noisy communication links. As a general trend, we expect  $P_{e,opt} \leq P_{e,mg} \leq P_{e,sg}$  and  $P_{e,ideal} \leq P_{e,noisy}$ .

### 5.1 Decision Fusion with Ideal Communication links

With ideal communication links, the final classifier at the manager node is given by

$$C_{ideal}(u_1, \dots, u_G) = \arg \max_{j=0, \dots, N-1} p_j[u_1, \dots, u_G] \pi_j = \arg \max_{j=0, \dots, N-1} \prod_{k=1}^G p_j[u_k] \pi_j \quad (45)$$

which can also be expressed in terms of log-likelihoods as

$$C_{ideal}(u_1, \dots, u_G) = \arg \min_{j=0, \dots, N-1} l_{j,ideal}[u_1, \dots, u_G] \quad (46)$$

$$l_{j,ideal}[u_1, \dots, u_G] = -\frac{1}{G} \log p_j[u_1, \dots, u_G] \pi_j = -\frac{1}{G} \sum_{k=1}^G \log p_j[u_k] - \frac{1}{G} \log \pi_j. \quad (47)$$

Note that the above expressions apply to all three local classifiers; the only difference is that different local classifiers induce different decision pmf's  $\{p_j[m] = P(U = m | H_j)\}$ .

### 5.1.1 Performance of Ideal Decision Fusion

While the exact calculation of  $P_{e,ideal}$  is difficult, it can be bounded and analyzed asymptotically analogous to the centralized classifier.

**Bound on  $P_{e,ideal}$ .** As in the centralized classifier,  $P_{e,ideal}(G)$  decays exponentially with  $G$  and is bounded analogous to (12) with  $\mu_{jm,ideal}(\theta)$  given by

$$\mu_{jm,ideal}(\theta) = \log E_m [p_j^\theta[U] / p_m^\theta[U]] = \log \sum_{i=0}^{N-1} p_j^\theta[i] p_m^{1-\theta}[i]. \quad (48)$$

Then the PEP's and  $P_{e,ideal}$  can be bounded analogous to (14) and (18).

**Asymptotic Performance as  $G \rightarrow \infty$ .** Despite the fact that the local hard decisions can be quite unreliable, the final hard decision fusion classifier can still attain perfect performance as  $G \rightarrow \infty$ . From (47) and the law of large numbers, under  $H_m$ , we have

$$\lim_{G \rightarrow \infty} l_{j,ideal}[u_1, \dots, u_G] = -E_m [\log p_j[U]] = D(p_m \| p_j) + H_m(U) \quad (49)$$

where  $D(p_m \| p_j)$  is the K-L distance between the pmf's  $p_m$  and  $p_j$  and  $H_m(U)$  is the entropy of the hard decision under  $H_m$  [9]

$$D(p_m \| p_j) = \sum_{i=0}^{N-1} p_m[i] \log(p_m[i] / p_j[i]) \quad (50)$$

$$H_m(U) = -\sum_{i=0}^{N-1} p_m[i] \log p_m[i]. \quad (51)$$

It follows that  $\lim_{G \rightarrow \infty} P_{e,ideal} = 0$  if and only if  $D(p_m \| p_j) > 0$ ,  $\forall m \neq j$ . The following result summarizes the behavior of  $P_{e,ideal}$  for all three local classifiers.

**Proposition 3**  $P_{e,ideal}(G)$  decays exponentially to zero with  $G$  if and only if all the pairwise K-L distances between the decision pmf's are strictly positive, that is,

$$D(p_j || p_m) > 0, \quad \forall j \neq m. \quad (52)$$

Specifically,

$$\limsup_{G \rightarrow \infty} \frac{\log P_{e,ideal}(G)}{G} \leq -D_{min,ideal}^* \quad (53)$$

where  $D_{min,ideal}^* = \min\{D_{jm,ideal}^*\}$  is the smallest pairwise Chernoff information corresponding to the underlying local classifiers.

Note that for a given measurement SNR, the error exponent for optimal ideal decision fusion will be smaller compared to the centralized classifier, since the pairwise K-L distances between the pmf's of the optimal local classifier will be smaller than those between the pdf's in the centralized classifier. Amongst the three local classifiers, the K-L distances would generally decrease from optimal to sub-optimal MG to sub-optimal SG classifiers.

## 5.2 Decision Fusion with Noisy Communication Links

In this section, we discuss fusion of local decisions from different SCR's using noisy communication links, as illustrated in Fig. 1. We assume that each SCR has a dedicated communication link to the manager node. (Note that this requires large bandwidth or large latency at the manager node in the limit of large  $G$ .) Each SCR sends an amplified version of its local hard decision  $u_k$  in (28) over a noisy link

$$y_k = \alpha u_k + w_k, \quad k = 1, \dots, G \quad (54)$$

where  $y_k$  denotes the received signal at the manager node from the  $k$ -th SCR and  $\{w_k\}$  are i.i.d  $\mathcal{N}(0, \sigma_w^2)$  (real Gaussian noise). Note that since  $\{u_k\}$  are i.i.d, so are  $\{y_k\}$ . Without loss of generality, assume that  $N$  is odd and define  $\tilde{N} = (N - 1)/2$ . We assume that each SCR sends a symmetrized version of its hard decision to use minimum power:  $u_k \in \{-\tilde{N}, \dots, \tilde{N}\}$ . Given this simple communication scheme, the optimal decentralized classifier at the manager node takes the

form

$$C_{noisy}(\mathbf{y}) = \arg \min_j l_{j,noisy}(\mathbf{y}) \quad (55)$$

$$l_{j,noisy}(\mathbf{y}) = -\frac{1}{G} \log p_{j,noisy}(\mathbf{y}) = -\frac{1}{G} \sum_{k=1}^G \log p_{j,noisy}(y_k) \quad (56)$$

$$p_{j,noisy}(y) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \sum_{i=-\tilde{N}}^{\tilde{N}} e^{-(y-\alpha i)^2/2\sigma_w^2} p_j[i]. \quad (57)$$

### 5.2.1 Performance of Noisy Decision Fusion

The exact calculation of  $P_{e,noisy}$  is most complicated in this case; however, it can be bounded and analyzed asymptotically as in the centralized classifier and ideal decision fusion.

**Bound on  $P_{e,noisy}$ .** As in the centralized classifier, Chernoff and Bhattacharya bounds for the PEP's can be obtained. In particular, in this case  $\mu_{jm,noisy}(\theta)$  takes the form

$$\mu_{jm,noisy}(\theta) = \log \mathbb{E}_m [p_{j,noisy}^\theta(\mathbf{Y})/p_{m,noisy}^\theta(\mathbf{Y})] \quad (58)$$

which can be estimated or computed numerically.

**Asymptotic Performance as  $G \rightarrow \infty$ .** For sufficiently large measurement and communication SNR's, we again expect  $P_{e,noisy} \rightarrow 0$  as  $G \rightarrow \infty$ . From (57), we have under  $H_m$

$$\lim_{G \rightarrow \infty} l_{j,noisy}(\mathbf{y}) = -\mathbb{E}_m [\log p_{j,noisy}(\mathbf{Y})] = D(p_{m,noisy} \| p_{j,noisy}) + h_m(\mathbf{Y}) \quad (59)$$

Hence,  $\lim_{G \rightarrow \infty} P_{e,noisy} = 0$  if and only if  $D(p_{m,noisy} \| p_{j,noisy}) > 0$ ,  $\forall m \neq j$ . The following summarizes the performance of noisy decision fusion.

**Proposition 4**  $P_{e,noisy}$  decays exponentially to zero as  $G \rightarrow \infty$  if and only if all pairwise K-L distances between noisy pdf's in (57) are strictly positive; that is,

$$D(p_{j,noisy} \| p_{m,noisy}) > 0, \quad \forall m \neq j. \quad (60)$$

Specifically,

$$\limsup_{G \rightarrow \infty} \frac{P_{e,noisy}}{G} \leq -D_{min,noisy}^* \quad (61)$$

where  $D_{min, noisy}^* = \min\{D_{jm, noisy}^*\}$  is the smallest pairwise Chernoff information for noisy decision fusion corresponding to the local classifier (optimal, MG, or SG).

For any given local classifier, we expect the K-L distances in this case to be smaller than those for ideal decision fusion. In particular, they depend on both the measurement and communication SNR's.

## 6 Numerical Results

Numerical results presented in this section are based on real data collected during the DARPA SensIT program for two ( $M=2$ ) targets: Amphibious Assault Vehicle (AAV) and Dragon Wagon (DW). PSD values at  $N_o = 25$  frequencies (within a 2kHz bandwidth) were estimated using data collected at multiple nodes. The PSD estimates are plotted in Fig. 3. The PSD values define the diagonal covariance matrices (in the Fourier domain) for the two targets:  $\Sigma_m = \Lambda_m$ ,  $m = 1, 2$ . Thus, the simulated multi-target classification problem corresponds to  $N = 4$  hypotheses. Under  $H_j$ ,  $j = 1, \dots, 3$ , the  $N_o = 25$ -dimensional measurement vector  $\mathbf{z}_k$  in the  $k$ -th SCR was simulated as

$$\mathbf{z}_k = \bar{\Lambda}_j^{1/2} \mathbf{v}_k + \mathbf{n}_k, \quad k = 1, \dots, G \quad (62)$$

where  $\{\mathbf{v}_k\}$  are i.i.d  $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ ,  $\bar{\Lambda}_j = b_1(j)\Lambda_1 + b_2(j)\Lambda_2$ , and  $\{\mathbf{n}_k\}$  are i.i.d.  $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$ .

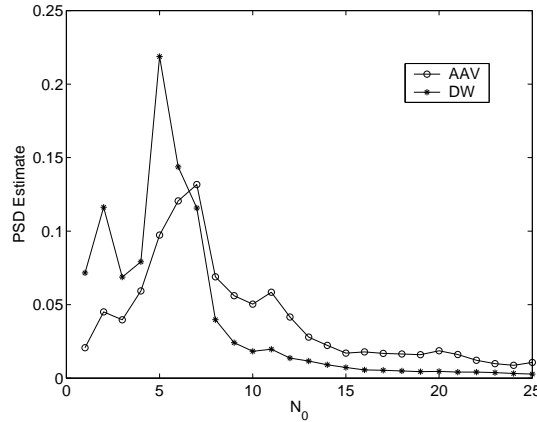


Figure 3: Covariance matrix eigenvalues (PSD estimates) based on acoustic measurements for the two target vehicles: AAV and DW.

The three proposed distributed classifiers, corresponding to the three local classifiers (optimal, MG, and SG), are compared with the optimal centralized classifier under ideal (noise-free) and noisy communication links. The optimal centralized classifier serves as a benchmark for the proposed classifiers. The  $P_e$ 's for the different classifiers are plotted as a function of  $G$  in Figures 4 and 5 for various measurement and communication SNRs which are defined as

$$\text{SNR}_{meas} = 10 \log_{10}(\sigma_s^2/N_o\sigma_n^2) , \quad \text{SNR}_{comm} = 10 \log_{10} \left( \sum_i (\alpha_i)^2 p[i]/\sigma_w^2 \right) \quad (63)$$

where  $p[i] = P(U = i) = \sum_j p_j[i]\pi_j$  and recall that we assume constant signal energy for the two targets:  $\sigma_s^2 = \text{tr}(\mathbf{\Lambda}_1) = \text{tr}(\mathbf{\Lambda}_2)$ . The  $P_e$ 's were estimated via Monte Carlo simulation using 10000 independent sets of  $G$  measurements assuming equal prior probabilities, i.e.  $q_1 = 1/2$  and  $q_2 = 1/2$ . The pmf's  $\{p_m[i]\}$  of local hard decisions for the three local classifiers were also estimated via this Monte Carlo simulation. The measurements for noisy hard decision fusion were simulated using (54) and the pmf's for hard decisions and the  $P_e$ 's were estimated using 50000 independent sets of measurement realizations. The simulations were done for four values of  $\text{SNR}_{meas} = -4, 0, 4, 10$  dB and two values for  $\text{SNR}_{comm} = 0, 10$  dB.

Fig. 4 shows plots for  $\text{SNR}_{meas} = -4\text{dB}, 0\text{dB}$ , and  $\text{SNR}_{comm} = 0\text{dB}, 10\text{dB}$ . Fig. 5 shows plots for  $\text{SNR}_{meas} = 4\text{dB}, 10\text{dB}$ , and  $\text{SNR}_{comm} = 0\text{dB}, 10\text{dB}$ . In all figures, we observe that the  $P_e$  of the centralized classifier serves as a lower bound. As expected, the  $P_e$  of all classifiers improves with  $\text{SNR}_{meas}$ . Furthermore, for any distributed classifier, the  $P_e$  under noisy links is higher than that over ideal communication links, and the  $P_e$  under noisy links approaches the  $P_e$  under ideal links with increasing  $\text{SNR}_{comm}$ . The performance of the SG classifier is worse than both the MG and optimal classifiers. Most importantly, in general, the sub-optimal MG classifier performs nearly as well as the optimal classifier for all considered values of  $\text{SNR}_{meas}$  and  $\text{SNR}_{comm}$ . This is because the MG classifier is optimal for the underlying natural re-partitioning of the hypothesis space whereas the SG classifier is an approximation to it.

As the figures show,  $P_e$  decays exponentially with  $G$  for all classifiers, albeit with different rates (error exponents). This demonstrates an important practical advantage of multiple independent measurements in sensor networks: we can attain reliable classification performance by combining a relatively moderate number of much less reliable independent local decisions. The rates for the

optimal and the MG classifier are greater than those for the SG classifier under both ideal and noisy communication links. As expected, the rate is greatest for the optimal centralized classifier. These differences in performance can be attributed to the differences in the pair-wise K-L distances shown in Table 2 for all the hypotheses for the different classifiers. We can infer the following general trends from the values of K-L distances:

- The K-L distances increase with  $\text{SNR}_{meas}$  for all classifiers.
- For a given  $\text{SNR}_{meas}$ , the K-L distances decrease from the centralized optimal classifier to the distributed optimal to MG to SG classifiers.
- For any given distributed classifier, the K-L distances are lower for noisy links than those for ideal links, and they increase with  $\text{SNR}_{comm}$ .

However, these general observations are violated in some instances as can be seen from the table. When these anomalies arise in the smallest (dominating) K-L distance under any given hypothesis, then the overall  $P_e$  trends can be anomalous. For instance, if under a given hypothesis the smallest K-L distance of any classifier was smaller at a higher  $\text{SNR}_{meas}$  than the corresponding value at a lower  $\text{SNR}_{meas}$ , it would violate the first observation, and could result in a worse  $P_e$  performance at the higher  $\text{SNR}_{meas}$ . While no such anomalies are seen in the Table, we did observe them in some isolated cases. This warrants further investigation which will be reported elsewhere.

Finally, Fig. 6 compares the union bounds for  $P_e$  with the actual simulated  $P_e$  under ideal communication links for for  $\text{SNR}_{meas}$  of -4dB and 10dB. Observe that the bounds match the error exponent fairly well but exhibit an offset due to the union bounding via PEP's. Furthermore, the bounds get tighter at higher  $\text{SNR}_{meas}$ .

We also performed simulations for M=3 targets (the third vehicle was a Humvee). A  $P_e$  plot is shown in Fig. 7 for  $\text{SNR}_{meas}$  of 10dB and  $\text{SNR}_{comm}$  of 10dB. As evident, there is a considerable loss in performance compared to M=2 targets. This is related to the fact that for M=3, the total number of hypotheses is increased from 4 to 8 whereas the dimensionality of the feature vector is the same  $N_o = 25$ . This issue also warrants further investigation and is briefly discussed in the next section.

Measurement SNR = -4 dB					Measurement SNR = 10 dB				
$m/j$	No Vehicle	AAV	DW	AAV and DW	$m/j$	No Vehicle	AAV	DW	AAV and DW
No Vehicle	0	1.4822	1.7388	3.7103	0	32.3045	25.3041	42.8035	
	0	0.9874	1.3220	2.6837	0	13.8154	13.8154	13.8154	
	0	0.9306	1.1885	2.6318	0	13.8154	13.8154	13.8154	
	0	0.7026	0.8943	2.2597	0	4.3899	1.5994	13.8154	
	0	0.4631	0.8435	1.6159	0	1.1019	3.3922	3.3751	
	0	0.4907	0.8274	1.7664	0	1.0987	3.4054	3.3784	
	0	0.4321	0.6667	1.6946	0	1.8289	1.0665	5.1495	
	0	0.0953	0.1832	0.3809	0	0.1305	0.3928	0.7419	
	0	0.1114	0.1886	0.3906	0	0.1285	0.3895	0.7365	
	0	0.0937	0.1493	0.3681	0	0.3691	0.2095	0.5956	
AAV	2.1890	0	0.4665	0.8373	196.2808	0	8.6186	3.1420	
	1.4161	0	0.2490	0.6184	13.4462	0	3.1847	1.9794	
	1.3312	0	0.2744	0.5908	13.4421	0	3.2085	1.9827	
	1.3636	0	0.1930	0.5997	13.0190	0	1.2925	1.0219	
	0.5963	0	0.0804	0.4193	1.4481	0	0.6868	1.4316	
	0.7358	0	0.0577	0.4088	1.4557	0	0.6864	1.4304	
	0.8222	0	0.0334	0.4413	4.2908	0	0.1111	0.3286	
	0.1042	0	0.0148	0.0966	0.1371	0	0.0673	0.2575	
	0.1281	0	0.0101	0.0823	0.1381	0	0.0669	0.2559	
	0.1121	0	0.0066	0.0899	0.4118	0	0.0210	0.0201	
DW	3.0301	0.5159	0	0.7118	206.5777	6.4357	0	7.7110	
	1.9142	0.2546	0	0.4413	13.6419	3.3245	0	4.6003	
	1.7656	0.2707	0	0.3973	13.6309	3.3156	0	4.6832	
	1.8378	0.1639	0	0.4067	9.9230	3.1266	0	8.4207	
	1.0868	0.0788	0	0.1697	3.9523	0.5715	0	0.5980	
	1.2157	0.0566	0	0.1631	3.9558	0.5695	0	0.5974	
	1.2476	0.0332	0	0.2291	2.6532	0.1253	0	0.9641	
	0.2012	0.0144	0	0.0356	0.3921	0.0635	0	0.0687	
	0.2124	0.0101	0	0.0340	0.3985	0.0658	0	0.0661	
	0.1762	0.0067	0	0.0477	0.2383	0.0223	0	0.0889	
AAV and DW	7.3804	1.1950	0.8890	0	434.8090	6.0817	15.2650	0	
	4.2339	0.6167	0.4538	0	13.4755	2.1402	4.3877	0	
	3.0919	0.4716	0.3256	0	13.4813	2.1261	4.0519	0	
	3.5086	0.4800	0.3191	0	13.7986	0.2804	1.1875	0	
	2.0935	0.4112	0.1741	0	8.1175	1.4928	0.7642	0	
	2.1772	0.3356	0.1414	0	8.0908	1.4796	0.7437	0	
	2.5056	0.3570	0.1820	0	5.9821	0.1510	0.4279	0	
	0.4139	0.0946	0.0358	0	0.7982	0.2589	0.0701	0	
	0.4107	0.0757	0.0310	0	0.8040	0.2612	0.0716	0	
	0.3965	0.0823	0.0423	0	0.5964	0.0186	0.0761	0	

(a)

(b)

Table 2: Pairwise K-L distances for the 4 hypotheses, corresponding to  $M=2$  targets, for  $\text{SNR}_{meas}$  of -4dB and 10dB. The first four entries in each cell correspond to the optimal centralized classifier, the optimal distributed classifier, the MG classifier and the SG classifier with ideal links. The next three entries correspond to optimum distributed classifier, the MGC and the SGC with noisy links at  $\text{SNR}_{comm}$  of 10dB. The final three values correspond to optimum distributed classifier, the MGC and the SGC with noisy links at  $\text{SNR}_{comm}$  of 0dB.

## 7 Conclusions

In this paper, we have taken a first step in attacking the challenging problem of multi-target classification in wireless sensor networks. The problem is cast as a hypothesis testing problem. A key problem with multi-target classification is that the number of hypotheses increases exponentially with the number of targets. To circumvent this exponential complexity with the number of targets, we proposed two sub-optimal algorithms based on a natural re-partitioning of the hypothesis space that result in linear complexity. Our results show that the mixture Gaussian sub-optimal classifier, that is optimal for the re-partitioned space, delivers performance comparable to the optimal distributed classifier.

There are many avenues for future investigations. An important issue is the impact of signal

path loss in sensing measurements. We assumed that all the node measurements from different SCR's were i.i.d. In practice they will be independent but will not be identically distributed: nodes farther from the target will exhibit poorer measurement SNR. Thus, we expect an optimal radius around each point source over which the fusion of node measurement should be performed (the performance will degrade if farther node measurements were fused). This will limit the number of spatial independent measurements  $G$ . However, the effective  $G$  could be increased by processing multiple independent feature vectors collected over time [6].

Another important issue is the relation between the number of targets and the dimension  $N_o$  of the feature vector. As evident from Fig. 7, the performance significantly degrades for  $M = 3$  compared to  $M = 2$  for the same  $N_o = 25$ . This is natural since the signal space is getting more crowded due to the larger number of hypotheses. This effectively reduces the pairwise K-L distances and hence performance.

Another interesting direction is the investigation of type-based classification schemes [10] that may incur no loss in asymptotic performance compared to their centralized counterparts (unlike the loss in error exponents incurred by the proposed distributed algorithms).

Finally, this paper constitutes just one approach to multi-target classification. Other forms of sub-optimal algorithms, including tree-structured classifiers [11], subspace-based approaches [12, 13] and sub-optimal fusion schemes [14] could be leveraged in this context.

## References

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2001*, vol. 4, pp. 2033–2036, 2001.
- [2] "Special issue on collaborative signal and information processing in microsensor networks," in *IEEE Signal Processing Magazine*, (S. Kumar and F. Zhao and D. Shepherd (eds.)), March 2002.
- [3] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, tracking of targets in micro-sensor networks," in *IEEE Signal Processing Magazine*, pp. 17–29, March 2002.
- [4] A. D'Costa and A. M. Sayeed, "Collaborative signal processing for distributed classification in sensor networks," in *Lecture Notes in Computer Science (Proceedings of IPSN'03)*, (Springer-Verlag, Berlin Heidelberg), pp. 193–208, (F. Zhao and L. Guibas (eds.)), April 2003.
- [5] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of gaussian space-time sources in wireless sensor networks," *submitted to the IEEE J. Sel. Areas. Commun.*, July 2003.
- [6] A. Sayeed, "A statistical signal modeling framework for integrated design of wireless sensor networks," *submitted to the IEEE Signal Processing Magazine*, October 2003.

- [7] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 3rd ed., 1995.
- [8] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [10] K. Liu and A. Sayeed, "Distributed type-based detection in wireless sensor networks," *submitted to IPSN'04*, Oct. 2003.
- [11] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley, 2nd ed., 2001.
- [12] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loeve expansion to feature selection and ordering," *IEEE Transactions on Computers*, vol. C-19, pp. 311–318, Apr. 1970.
- [13] S. Watanabe and N. Pakvasa, "Subspace method to pattern recognition," *Proceedings of the 1st International Conference on Pattern Recognition*, pp. 25–32, Feb. 1973.
- [14] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 20, pp. 226–238, Mar. 1998.

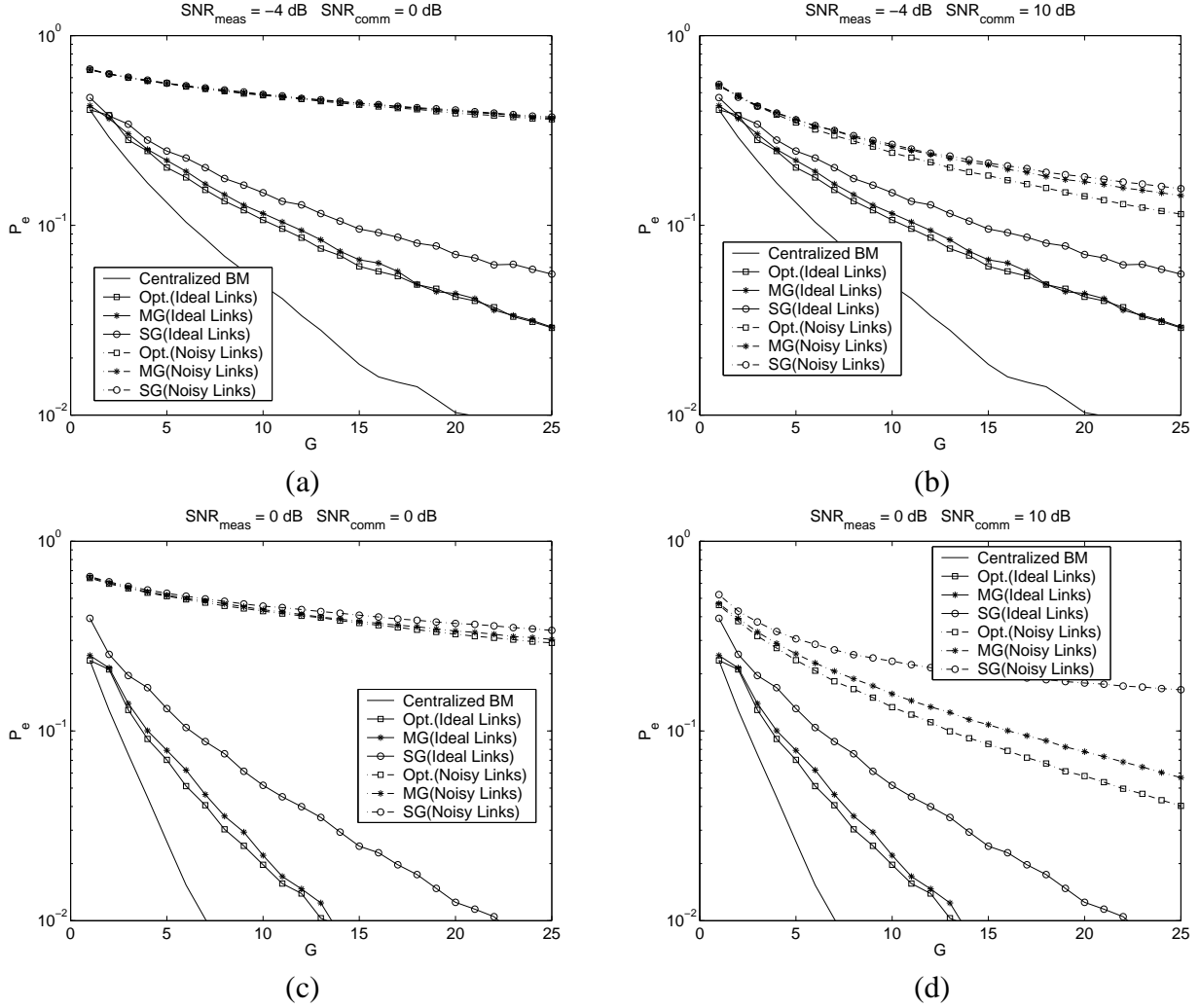


Figure 4:  $P_e$  as a function of  $G$  for  $M=2$  targets for  $\text{SNR}_{\text{meas}}$   $-4\text{dB}$  and  $0\text{dB}$  and  $\text{SNR}_{\text{comm}}$  for noisy links at  $0\text{dB}$  and  $10\text{dB}$ . The proposed (distributed) optimal, mixture Gaussian and single Gaussian classifiers, under ideal (noise-free) communication links and noisy communications links, are compared with the centralized optimal classifier which serves as a benchmark (BM). (a)  $\text{SNR}_{\text{meas}} = -4\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 0\text{dB}$  (b)  $\text{SNR}_{\text{meas}} = -4\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 10\text{dB}$  (c)  $\text{SNR}_{\text{meas}} = 0\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 0\text{dB}$  (d)  $\text{SNR}_{\text{meas}} = 0\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 10\text{dB}$ .

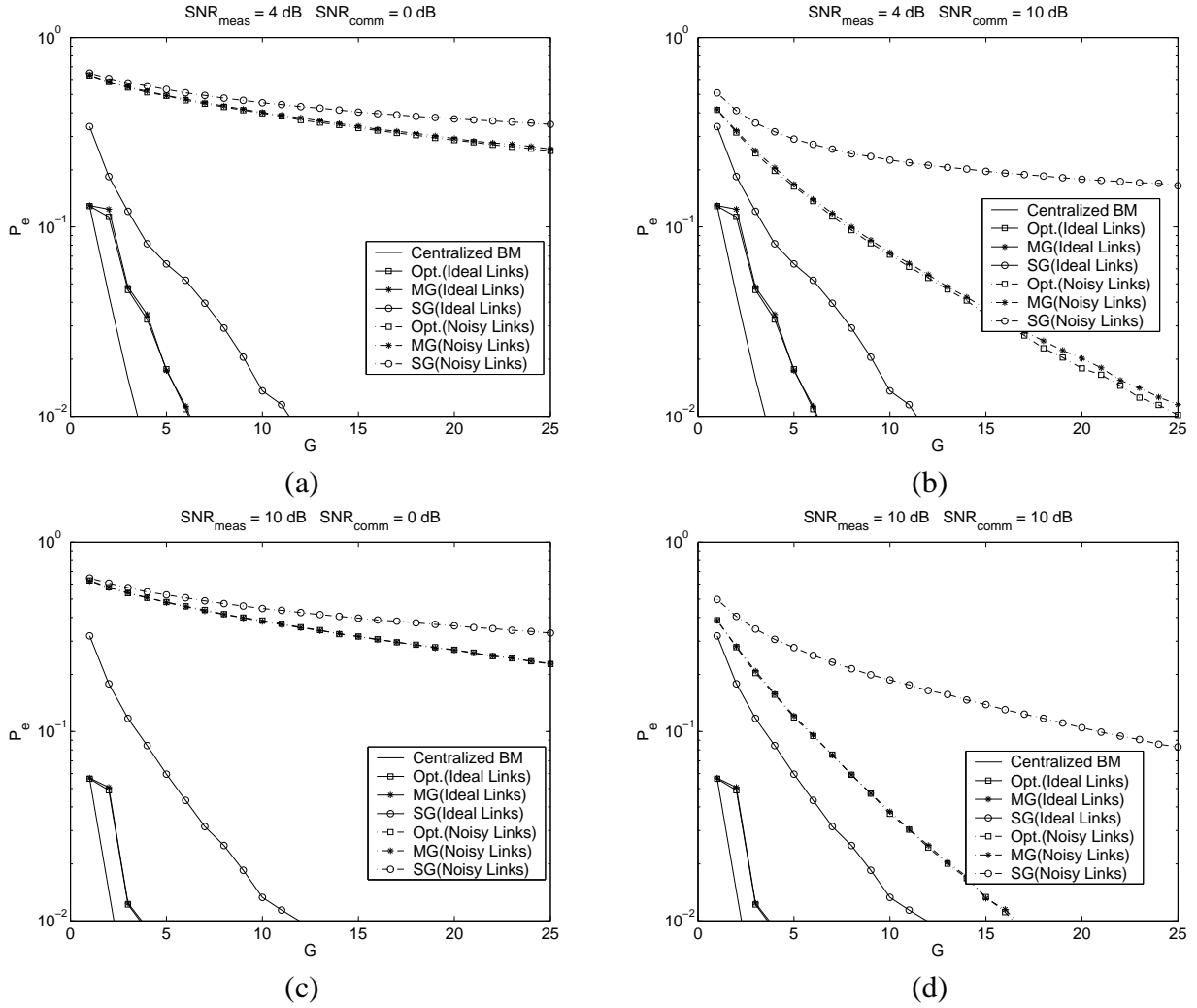


Figure 5:  $P_e$  as a function of  $G$  for  $M=2$  targets for  $\text{SNR}_{\text{meas}}$  4dB and 10dB and  $\text{SNR}_{\text{comm}}$  for noisy links at 0dB and 10dB. The proposed (distributed) optimal, mixture Gaussian and single Gaussian classifiers, under ideal (noise-free) communication links and noisy communications links, are compared with the centralized optimal classifier which serves as a benchmark (BM). (a)  $\text{SNR}_{\text{meas}} = 4\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 0\text{dB}$  (b)  $\text{SNR}_{\text{meas}} = 4\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 10\text{dB}$  (c)  $\text{SNR}_{\text{meas}} = 10\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 0\text{dB}$  (d)  $\text{SNR}_{\text{meas}} = 10\text{dB}$ ,  $\text{SNR}_{\text{comm}} = 10\text{dB}$ .

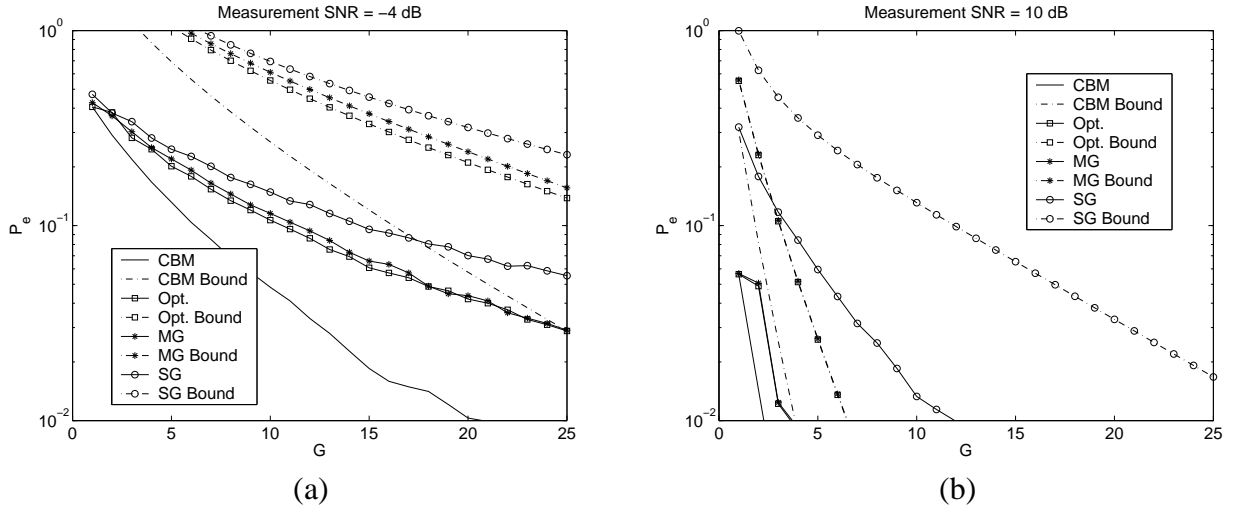


Figure 6: Comparison of  $P_e$  from simulations with  $P_e$  Chernoff bounds for the centralized optimal classifier and the proposed (distributed) optimal, mixture Gaussian and single Gaussian classifiers with ideal communication links. (a)  $\text{SNR}_{meas} = 0\text{dB}$ . (b)  $\text{SNR}_{meas} = 10\text{dB}$ .

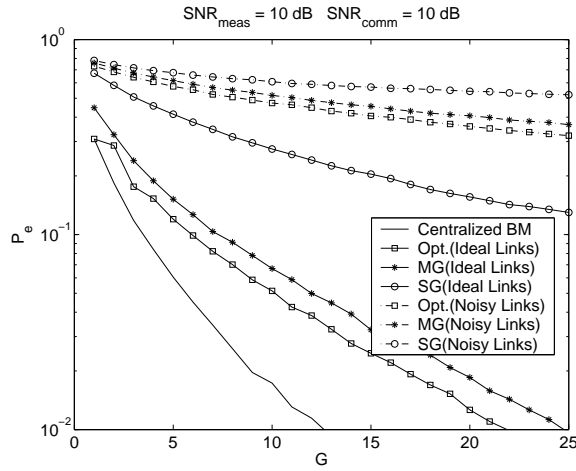


Figure 7:  $P_e$  as a function of  $G$  with  $M=3$  targets for  $\text{SNR}_{meas} 10\text{dB}$  and  $\text{SNR}_{comm}$  for noisy links at  $10\text{dB}$ . The proposed (distributed) optimal, mixture Gaussian and single Gaussian classifiers, under ideal (noise-free) communication links and noisy communications links, are compared with the centralized optimal classifier which serves as a benchmark (BM).