Distributed Classification of Gaussian Space-Time Sources in Wireless Sensor Networks

Ashwin D'Costa, Vinod Ramachandran and Akbar M. Sayeed*
Department of Electrical and Computer Engineering
University of Wisconsin-Madison
dcosta@cae.wisc.edu, vinod@cae.wisc.edu, akbar@engr.wisc.edu
http://dune.ece.wisc.edu

Abstract

The study of sensor networks invites a fresh look at well-studied problems to account for distributed communication and computational network constraints. In this paper, we study distributed signal processing techniques for classification of objects assuming knowledge of sensor measurement statistics. Our approach is based on modeling the spatio-temporal signal field generated by an object as a bandlimited stationary ergodic Gaussian field. The model provides a simple abstraction of correlation between node measurements: it partitions the network into disjoint spatial coherence regions over which the signal remains strongly correlated, whereas the signal in distinct coherence regions is approximately uncorrelated. The size of coherence regions is determined by spatial signal bandwidths. We show that this partitioning imposes a structure on optimal distributed classification algorithms that is naturally suited to the communication constraints of the network: local high-bandwidth exchange of feature vectors within each coherence region to improve the measurement SNR, and global low-bandwidth exchange of local decisions across coherence regions to stabilize the inherent variability in the signal. We analyze classifier performance for both hard and soft decision fusion across coherence regions assuming noise-free as well as noisy communication links between nodes. Under mild conditions, the probability of error of all classification schemes (soft, hard, noisy) decays exponentially to zero with the number of independent node measurements — the error exponent depends on both the measurement and communication SNRs and decreases from soft to hard to noisy fusion. Numerical results based on real data illustrate the remarkable advantage of multiple sensor measurements in decision making.

1 Introduction

Wireless sensor networks are an emerging technology for monitoring the physical world with a densely distributed network of wireless nodes (see, e.g., [1]). Each node has limited communication and computation ability and can sense the environment in a variety of modalities, such as

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

acoustic, seismic, and infra red [1, 2, 3]. A wide variety of applications are being envisioned for sensor networks, including disaster relief, border monitoring, condition-based machine monitoring, and surveillance in battlefield scenarios. Detection and classification of objects moving through the sensor field is an important task in many applications. Exchange of sensor information between different nodes in the vicinity of the object is necessary for reliable execution of such tasks due to a variety of reasons, including limited (local) information gathered by each node, variability in operating conditions, and node failure. Indeed, development of theory and methods for collaborative signal processing (CSP) of the data collected by different nodes is a key research area for realizing the promise of sensor networks.

Some form of region-based processing is attractive in sensor networks in order to facilitate CSP between nodes and also for efficient routing of information [3]. Typically, the nodes in the network are partitioned into a number of regions and a manager node is designated within each region to facilitate CSP between the nodes in the region and for communication of information from one region to another. The CSP algorithms have to be developed under the constraints imposed by the limited communication and computational abilities of the nodes as well as their finite battery life. To this end, a key goal of CSP algorithms is to exchange the least amount of data between nodes to attain a desired level of performance. In this paper, with the above goal in mind, we investigate CSP techniques for combining the data collected by different nodes for single-target classification within a region of interest.

There are two main forms of information exchange between nodes dictated by the statistics of measured signals. If two nodes yield correlated measurements, *data fusion* is needed for optimal performance – exchange of (low-dimensional) feature vectors that yield sufficient information for the desired task. For example, estimates of signal energy at different frequencies (Fourier/spectral feature vectors) may be used for classification. On the other hand, if two nodes yield statistically independent measurements, *decision fusion* is sufficient – exchange of soft or hard decisions computed at the two nodes. In general, the measurements at different nodes would exhibit a mixture of correlated and independent components and would require a combination of data and decision fusion between nodes. In the context of sensor networks, decision fusion is clearly the more attractive choice. First, it imposes a significantly lower communication burden on the network, compared to

data fusion, since only scalars are communicated to the manager node. Second, it imposes a lower computational burden compared to data fusion since lower dimensional data has to be jointly processed at the manager node. Third, a classifier based on decision fusion requires much smaller amount of data for *training* since fewer parameters characterize the classifier.

The statistics of node measurements are determined by the signal field in space and time generated by the underlying object of interest. The signal field may be sensed by the nodes in different modalities. In the next section we present a basic model for the signal field that yields a natural approximate characterization of space-time signal statistics. In particular, the model imposes a universal structure on all CSP algorithms for decision making in which costly[1] *data fusion* is confined to local spatial coherence regions (SCR's) in which the signal is strongly correlated, and only cheaper *decision fusion* is needed across different SCR's in which the signal is nearly independent. This model forms the basis of the distributed classification schemes studied in this paper.

In any network query involving an object (such as a vehicle), the first task is typically to detect the presence of the object in a region of interest. Classification of the object follows object detection. Section 3 discusses optimal CSP algorithms for classification that assume noise-free communication links from different SCR's to the manager node. Both soft and hard decision fusion is discussed. In Section 4, we discuss the more practical hard decision fusion over noisy communication links. We show that under mild conditions, all fusion schemes (soft, hard, noisy) exhibit exponentially vanishing probability of misclassification with the number of independent measurements from different SCR's — the error exponent decreases from soft to hard to noisy fusion. In Section 5 we present numerical results based on real measurement data to illustrate the potential performance gains due to multiple node measurements. In particular, our results demonstrate a remarkable practical advantage of multiple independent node measurements: a relatively moderate number of fairly unreliable local decisions can be fused over noisy communication links to produce acceptably reliable final decisions. Section 6 presents a discussion of the results as well pointers for future work.

[1]This relates to network cost in terms of bandwidth and power expenditure.

2 A Signal Model for Sensor Measurements

2.1 Underlying Assumptions on Signal Statistics

Each signal source corresponds to a space-time signal $s(x, y, t)$ as a function of the spatial coordinates (x, y) and time t. The network nodes sample $s(x, y, t)$ in space and time. Consider a spatial region of interest, $R = D_x \times D_y = [-D_x/2, D_x/2] \times [-D_y/2, D_y/2]$ associated with a network query regarding classification of a single source. We assume that the space-time signal is a zero-mean complex circular Gaussian stationary field in the spatial and temporal dimensions.[2] While practical sources will exhibit non-stationarities, this is a reasonable assumption over the space-time region of the query. The assumption of Gaussianity is equivalent to basing CSP algorithms on second-order statistics which is a reasonable and tractable assumption for initial investigations. Specifically, $s(x, y, t)$ is represented as

$$s(x,y,t) = \int_{-B_x/2}^{B_x/2} \int_{B_y/2}^{B_y/2} \int_{-B_t/2}^{B_t/2} \phi(\nu_x, \nu_y, f) e^{j2\pi\nu_x x} e^{j2\pi\nu_y y} e^{j2\pi f t} d\nu_x d\nu_y df \tag{1}$$

where $\phi(\nu_x, \nu_y, f)$ denotes the underlying spectral representation[3] which satisfies

$$\mathrm{E}[\phi(\nu_x, \nu_y, f)\phi^*(\nu_x', \nu_y', f')] = \Phi(\nu_x, \nu_y, f)\delta(\nu_x - \nu_x')\delta(\nu_y - \nu_y')\delta(f - f') \tag{2}$$

for some $\Phi(\nu_x, \nu_y, f) \geq 0$ that represents the power spectral density (PSD) of the process. The signal correlation function is related to the PSD via a 3D Fourier transform

$$\begin{aligned} r_s(\Delta x, \Delta y, \Delta t) &= \mathrm{E}[s(x + \Delta x, y + \Delta y, t + \Delta t)s^*(x, y, t)] \\ &= \int_{-B_x/2}^{B_x/2} \int_{B_y/2}^{B_y/2} \int_{-B_t/2}^{B_t/2} \Phi(\nu_x, \nu_y, \nu_t) e^{j2\pi(\nu_x \Delta x + \nu_y \Delta y + f\Delta t)} d\nu_x d\nu_y df \end{aligned} \tag{3}$$

and both characterize the statistics of the $s(x, y, t)$.

2.2 Approximate Signal Modeling Via Spatial Coherence Regions

To enable efficient CSP, we propose an approximate signal model, based on *spatial coherence regions (SCR's)* illustrated in Fig. 1, that captures the *scales of signal variation* in the spatial dimensions. To a first approximation, the spatial scales of variation in $s(x, y, t)$ are determined by

[2] We assume a complex signal field for generality; e.g., it would be applicable for signal sources created by passband transducers that send independent information in the in-phase and quadrature components.

[3] Strictly speaking, (1) needs to be a Stieltjes integral with respect to a random measure $d\phi(\nu_x, \nu_y, f)$, where $\phi(\nu_x, \nu_y, f)$ is an orthogonal increment process, but we use the above functional definition for simplicity.

the spatial bandwidths B_x and B_y – the larger the bandwidths, the faster the signal variation in the corresponding dimension. The spatial bandwidth B_x induces a *coherence distance*, $D_{c,x} = 1/B_x$, over which the signal remains strongly correlated in the x spatial dimension. Similarly, $D_{c,y} = 1/B_y$ denotes the coherence distance in the y dimension. Thus, as illustrated in Fig. 1, we can partition the query region R into disjoint SCR's, $\{R_{c,ij}\}$, of size $D_{c,x} \times D_{c,y}$ over which the signal remains strongly correlated (approximately constant). On the other hand, it can be shown that the signal is approximately uncorrelated in distinct SCR's [4]. The uniform size of SCR's follows from the stationarity assumption.

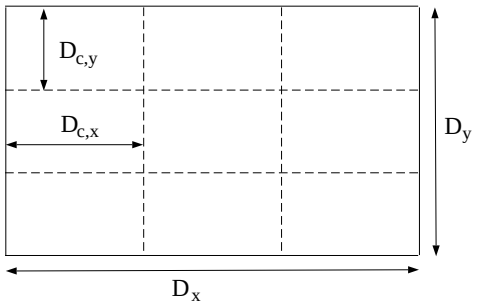

Figure 1: *A schematic illustrating the notion of* **spatial coherence regions (SCR's)** *over which $s(x, y, t)$ remains strongly correlated (approximately constant) as a function of (x, y). (a) The overall region of interest of size $D_x \times D_y$ is partitioned into SCR's of size $D_{c,x} \times D_{c,y}$ where $D_{c,x} = 1/B_x$ and $D_{c,y} = 1/B_y$ denote the coherence distances in x and y dimensions.*

Specifically, we use a *piece-wise constant (PWC) approximation* of the stationary signal commensurate with the SCR's

$$s_{pwc}(x, y, t) = \sum_{i,j} s_{ij}(t) I_{R_{c,ij}}(x, y) = \sum_{i=-\tilde{N}_s}^{\tilde{N}_s} \sum_{j=-\tilde{N}_t}^{\tilde{N}_t} s_{ij}(t) I_{D_{c,x}}(x - iD_{c,x}) I_{D_{c,y}}(y - jD_{c,y}) \quad (4)$$

where $I_X(x)$ denotes the indicator function of $(-X/2, X/2]$, $N_x = D_x/D_{c,x} = 2\tilde{N}_x + 1$, and $N_y = D_y/D_{c,y} = 2\tilde{N}_y + 1$. The PWC signal $s_{pwc}(x, y, t)$ is the projection of $s(x, y, t)$ onto the $N_s = N_x N_y$-dimensional spatial subspace spanned by the orthogonal spatial basis functions $\{u_{ij}(x, y) = I_{D_{c,x}}(x - iD_{c,x}) I_{D_{c,y}}(y - jD_{c,y})\}$. The N_s temporal processes $\{s_{ij}(t)\}$ constitute the spatial signal average in the corresponding SCR's

$$s_{ij}(t) = \frac{1}{D_{c,x} D_{c,y}} \int_{R_{c,ij}} s(x, y, t) dx dy = \frac{1}{D_{c,x} D_{c,y}} \int_{(i-1/2)D_{c,x}}^{(i+1/2)D_{c,x}} \int_{(j-1/2)D_{c,y}}^{(j+1/2)D_{c,y}} s(x, y, t) dx dy. \quad (5)$$

Example: Temporal point sources. In general, the spatial and temporal signal characteristics can be arbitrary. However, for this important class of signal sources they are intimately related. Such sources are characterized by a underlying temporal signal $s_o(t)$ with bandwidth B_t – the space-time signal is determined by $s_o(t)$ via physical signal propagation in space. For example, acoustic signals emitted by vehicles may be modeled in this fashion. For isotropic spatial propagation, $s(x,y,t) = s(r,t) = s_o(t - r/v)$ where $r = \sqrt{x^2 + y^2}$ and v is the speed of propagation. Thus, the signal is stationary along radial lines. It follows that $B_r = B_t/v$, where B_r is the bandwidth in the radial spatial dimension. The true SCR's are concentric bands around the source and the radial coherence distance D_r is given by $D_r = 1/B_r = v/B_t$. For example, for $B_t = 500$Hz, $D_r = 0.66$m whereas for $B_t = 20$Hz, $D_r = 17$m. Choosing $D_{c,x} = D_{c,y} = D_r$ for rectangular SCR's in Fig. 1 is a natural choice.

Spatial degrees of freedom. It can be shown that $s_{pwc}(x,y,t)$ preserves the most important statistical information about $s(x,y,t)$: *spatial degrees of freedom over* R. A simple intuitive way to see this is as follows. The sampling theorem states that all spatial information about $s(x,y,t)$ in contained in the samples $s[i,j;t] = s(i/B_x, j/B_y, t)$. The number of samples in the query region R equals $(D_x B_x)(D_y B_y) = (D_x/D_{c,x})(D_y/D_{c,y}) = N_s$, which is precisely equal to the number of coefficients in $s_{pwc}(t)$ in (4). In fact, at any time t, $s[i,j;t]$ corresponds to the signal sample at the center of the (i,j)-th SCR, whereas the PWC model coefficient $s_{ij}(t)$ in (5) corresponds to the signal average in the SCR. It can be shown that the temporal processes $s_{ij}(t)$'s corresponding to different SCR'ss are approximately uncorrelated [4]. Thus, there are approximately $N_s = N_x N_y$ independent spatial degrees of freedom in $s(x,y,t)$ over R which are preserved by $s_{pwc}(x,y,t)$.[4]

Assumptions on Sensor Measurements. Based on the above discussion, we make two assumptions about spatial signal variation to facilitate insight and analysis:

1. $s(x,y,t)$ is *perfectly correlated* in each SCR. That is, for any t, the signal in the (i,j)-th SCR is constant as a function of (x,y); $s(x,y,t) = s_{ij}(t)$, $(x,y) \in R_{c,ij}$.

2. the temporal processes $\{s_{ij}(t)\}$ in different SCR's are statistically independent.

[4]We note that similar approximations are widely used in the analysis of randomly time-varying communication channels in the guise of *block fading models* (see, e.g.,[5]).

In the region $R = D_x \times D_y$, there are $G = N_x N_y$ independent SCR's. From now on, we will label the SCR's by a single index: $R_{c,k}, k = 1, \cdots, G$. We assume that there are n_G nodes in each SCR, resulting in a total of $K = Gn_G$ nodes in R. We model time signal sensed by the i-th node as

$$x_i(t) = s_i(t) + n_i(t) \ , \ i = 1, \cdots, K = Gn_G \tag{6}$$

where $n_k(t)$ denotes zero-mean complex circular white Gaussian noise process. We assume that the $\{n_i(t)\}$ at different nodes are independent identically distributed (i.i.d.). The signal at each node is sampled at a sufficiently rate in disjoint blocks of N samples. Let $\{\boldsymbol{x}_i, \ i = 1, \cdots K\}$ denote the N-dimensional measurement vectors at the K nodes.

At the sensing level, there are two sources of error in decision making: i) the additive noise, and ii) the statistical variability in the source signal. The notion of SCR's illustrated in Fig. 1 imposes a structure on optimal classifiers that is naturally suited to network communication constraints and also enables mitigation of both sources of error:

- First, the n_G measurements in each SCR are *averaged* to increase the effective measurement signal-to-noise-ratio (SNR) by a factor of n_G. This high-bandwidth data fusion is limited to within SCR's.
- Second, local independent decisions from different SCR's are appropriately combined to reduce the statistical variability in the final decision.

For the remainder of the paper we assume that the measurements $\{\boldsymbol{x}_i\}$ in each SCR are averaged to yield a single N-dimensional vector $\boldsymbol{z}_k$ for each SCR

$$\boldsymbol{z}_k = \frac{1}{n_G} \sum_{i \in R_{c,k}} \boldsymbol{x}_i = \boldsymbol{s}_k + \boldsymbol{n}_k \ , \ k = 1, \cdots, G. \tag{7}$$

If the original noise variance is σ_n^2, the variance of the averaged noise becomes σ_n^2/n_G. Let $\boldsymbol{s} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a complex circular Gaussian vector with mean $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{s}]$ and covariance matrix $\boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{s}\boldsymbol{s}^H]$, where $(\cdot)^H$ denotes Hermitian transpose. Then $\boldsymbol{s}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma})$ for some $\boldsymbol{\Sigma}$ and $\boldsymbol{n}_k \sim \mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}/n_G)$ where $\boldsymbol{I}$ denotes the identity matrix.

Feature Selection. An imporant issue in classification is: what kind of measurements $\{\boldsymbol{x}_i\}$ should be collected? This is the called *feature selection* [6]. Essentially, the raw time series data

collected in each block at each node is processed to extract a relevant feature vector that best facilitates discrimination between classes. Feature selection is an important research topic in its own right but we will not discuss it here; we refer the reader to [6] for a discussion. In numerical results, we will assume a particular type of feature vectors $\{\boldsymbol{x}_i\}$ — spectral feature vectors — that can be obtained by computing a Fourier transform of the raw data. This is a natural consequence of our signal model in which the target signals are modeled as a stationary process. An important consequence of Fourier features is that the different components of $\boldsymbol{s}_k$ correspond to different frequencies and are approximately uncorrelated. The power in each component is proportional to a sample of the temporal PSD associated with the target class. Furthermore, noise statistics remain unchanged since Fourier transformation does not change the statistics of white noise. Thus, in numerical results we will explicitly substitute $\boldsymbol{\Sigma}_j = \boldsymbol{\Lambda}_j$, where $\boldsymbol{\Lambda}_j$ is a diagonal matrix whose diagonal entries are proportional to PSD samples for class j. Equivalently, if we consider $\boldsymbol{\Sigma}_j$ to be the (Toeplitz) covariance matrix of raw signal, $\boldsymbol{\Lambda}_j = \boldsymbol{U}^H \boldsymbol{\Sigma}_j \boldsymbol{U}$ denotes the matrix of eigenvalues of $\boldsymbol{\Sigma}_j$ where $\boldsymbol{U}$ is the discrete Fourier transform (DFT) matrix.[5]

3 Decision Fusion with Noise-Free Communication Links

Suppose that we are interested in classifying a single target/object in a region R. 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]). We assume that a target has already been detected. Furthermore, we assume that the target belongs to one of M possible classes. Mathematically, the classification problem can be stated as an M-ary hypothesis testing problem

$$H_j \;:\; \boldsymbol{z}_k = \boldsymbol{s}_k + \boldsymbol{n}_k \;,\; k = 1, \cdots, G, \;\;,\;\; j = 1, \cdots, M \tag{8}$$

where $\{\boldsymbol{n}_k\}$ are i.i.d. $\mathcal{CN}(\boldsymbol{0}, \sigma_n^2 \boldsymbol{I}/n_G)$ and $\{\boldsymbol{s}_k\}$ are i.i.d. $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Sigma}_j)$ under hypothesis H_j. Thus, under H_j, $\{\boldsymbol{z}_k\}$ are i.i.d. $\mathcal{CN}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_j)$ where $\widetilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_j + \sigma_n^2 \boldsymbol{I}/n_G$. All information about the targets is contained in the covariance matrices $\{\boldsymbol{\Sigma}_j\}$. In practice, $\{\boldsymbol{\Sigma}_j\}$ have to be estimated from available training data. We assume knowledge of $\mathrm{tr}(\boldsymbol{\Sigma}_j)$ and that $\mathrm{tr}(\boldsymbol{\Sigma}_j)$ (signal energy)[6] is the

[5] It is well-known that the DFT matrix diagonalizes Toeplitz matrices in the limit of large dimension [7].

[6] $\mathrm{tr}(\cdot)$ denotes the trace of a matrix (sum of the diagonal entries).

same for all j. Based on the measurement vectors $\{\boldsymbol{z}_k\}$ from the G SCR's, the manager node has to decide which one of the M classes does the target belong to. In this section, we discuss CSP algorithms for classification based on fusion of both *soft* and *hard* decisions, assuming a noise-free communication link from each SCR to the manager node. The noise-free decision fusion architecture is illustrated in Fig. 2(a).

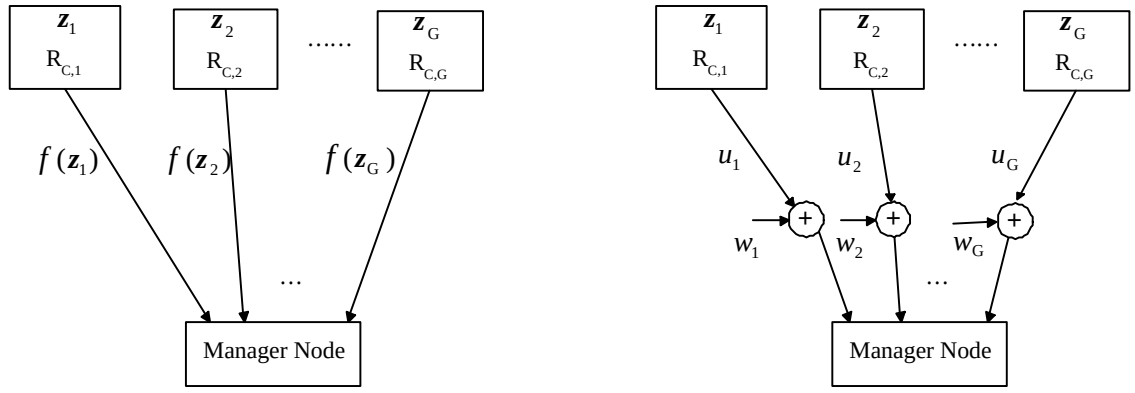

Figure 2: *(a) Noise-free decision fusion. For soft decisions,* $f(\boldsymbol{z}_k) = \{\boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k, j = 1, \cdots, M\}$*, whereas for hard decisions* $f(\boldsymbol{z}_k) = u_k = \arg\max_j p_j(\boldsymbol{z}_k)$*. (b) Noisy hard decision fusion.*

3.1 Soft Decision Fusion

For simplicity we assume that different classes are equally likely. Then, the optimal classifier chooses the class with the largest likelihood [6, 5, 8]

$$C(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) = \arg \max_{j=1,\cdots,M} p_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) \tag{9}$$

where $p_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G)$ is the probability density function (pdf) of the measurements under H_j. Since $\{\boldsymbol{z}_k\}$ are i.i.d. $\mathcal{CN}(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_j)$ under H_j

$$p_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) = \prod_{k=1}^{G} p_j(\boldsymbol{z}_k) \ , \ p_j(\boldsymbol{z}_k) = \frac{1}{\pi^N |\widetilde{\boldsymbol{\Sigma}}_j|} e^{\overline{\boldsymbol{z}}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k} \tag{10}$$

where $|\widetilde{\boldsymbol{\Sigma}}_j| = \prod_{n=1}^{N} (\lambda_j[n] + \sigma_n^2/n_G)$ denotes the determinant of $\widetilde{\boldsymbol{\Sigma}}_j$ (product of eigenvalues). It is often convenient to work with the negative log-likelihood functions

$$C(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) = \arg \min_{j=1,\cdots,M} l_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) \tag{11}$$

$$l_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) = -\frac{\log p_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G)}{G} = -\frac{1}{G} \sum_{k=1}^{G} \log p_j(\boldsymbol{z}_k). \tag{12}$$

Ignoring constants that do not depend on the class, the negative log-likelihood function for H_j takes the form

$$l_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) = \log |\widetilde{\boldsymbol{\Sigma}}_j| + \frac{1}{G} \sum_{k=1}^{G} \boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k. \tag{13}$$

Note that implementation of the optimal classifier requires that the k-th SCR communicates the local log-likehood functions for the M hypotheses, $\{\boldsymbol{z}_k^H \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z} \ , \ j = 1, \cdots, M\}$, to the manager node. The classifier at the manager node then computes l_j in (13) for $j = 1, \cdots, M$ and declares that the target belongs to the class with the smallest l_j.

3.1.1 Performance of Soft Decision Fusion

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

$$P_e(G) = \frac{1}{M} \sum_{m=1}^{M} P_{e,m}(G) \ \ , \ \ P_{e,m}(G) = P(l_j < l_m \text{ for some } j \neq m \ | H_m) \tag{14}$$

where $P_{e,m}(G)$ is the conditional error probability under H_m. Computing $P_{e,m}$ is complicated in general but we can bound it using the union bound [5]

$$P_{e,m}(G) \leq \sum_{j=1, j \neq m}^{M} P(l_j < l_m | H_m). \tag{15}$$

Note that $P_e(G) = 1 - \mathrm{PC}(G)$ where PC denotes the average probability of correct classification

$$\mathrm{PC}(G) = \frac{1}{M} \sum_{m=1}^{M} \mathrm{PC}_m(G) \ \ , \ \ \mathrm{PC}_m(G) = P(l_m \leq l_j \text{ for all } j \neq m | H_m) \tag{16}$$

and PC_m denotes the conditional probability of correct classification under H_m. Each pairwise error probability (PEP) on the right hand side of (15) depends on a decision statistic that is a weighted sum of NG χ_2^2 random variables [8]. The pdf and distribution function of the statistic can be computed in closed-form but take on tedious expressions [5].

We can obtain a reasonably tight upperbound on P_e via Chernoff bounds for the pairwise hypothesis tests. Define the average symmetric PEP, $\mathrm{PEP}_{jm}(G)$, for the binary test between H_j and H_m with G measurements, as

$$\mathrm{PEP}_{jm}(G) = \frac{1}{2}[P(l_j < l_m | H_m) + P(l_m < l_j | H_j)]. \tag{17}$$

Then (14) can be written as

$$P_e(G) \leq \frac{1}{M} \sum_{m=1}^{M} \sum_{j=1, j \neq m}^{M} P(l_j < l_m | H_m) = \frac{2}{M} \sum_{m=1}^{M} \sum_{j<m} \mathrm{PEP}_{jm}(G). \tag{18}$$

Chernoff bounding techniques can be used to obtain tight bounds for the PEP. We state some well-known results (see, e.g., [5, 9]) in the context of our set-up. Let $\mathrm{E}_m[\cdot]$ denote the expectation under H_m.

Proposition 1 (Chernoff Bounds) *For $0 \leq \theta \leq 1$, define*

$$\mu_{jm}(\theta) = \log \mathrm{E}_m \left[\left(\frac{p_j(\boldsymbol{Z})}{p_m(\boldsymbol{Z})} \right)^{\theta} \right] \leq 0. \tag{19}$$

Then, for any $0 \leq \theta \leq 1$ and for all $G \geq 1$

$$\mathrm{PEP}_{jm}(G) \leq \frac{1}{2} e^{\mu_{jm}(\theta) G} \tag{20}$$

The tightest error exponent (Chernoff information) is given by

$$D^*_{jm} = - \min_{0 \leq \theta \leq 1} \mu_{jm}(\theta) \tag{21}$$

and a simpler non-trivial exponent is provided by the Bhattacharya bound

$$D^*_{B,jm} = -\mu_{jm}(1/2) > 0. \tag{22}$$

Thus, for all $G \geq 1$ we have

$$\mathrm{PEP}_{jm}(G) \leq \frac{1}{2} e^{-D^*_{jm} G} \leq \frac{1}{2} e^{-D^*_{B,jm} G}. \tag{23}$$

In the case of noise-free soft decision fusion, it is relatively straightforward to compute $\mu_{jm}(\theta)$ (we omit the details here):

$$\mu_{jm}(\theta) = \theta \log \left| \widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1} \right| - \log \left| (1-\theta) \boldsymbol{I} + \theta \widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1} \right| \tag{24}$$

where the above expression holds for all $\theta \geq 0$ for which $(1-\theta)\boldsymbol{I} + \theta \widetilde{\boldsymbol{\Sigma}}_m \widetilde{\boldsymbol{\Sigma}}_j^{-1}$ is positive definite. The minimization in (21) can be easily performed numerically using (24).

Substituting $\mathrm{PEP}_{jm}(G)$ bounds from (23) into (18) we get two corresponding bounds for $P_e(G)$

$$P_e(G) \leq \frac{1}{M}\sum_{m=1}^{M}\sum_{j<m} e^{-D^*_{jm}G} \leq \frac{1}{M}\sum_{m=1}^{M}\sum_{j<m} e^{-D^*_{B,jm}G}. \tag{25}$$

Thus all PEP_{jm}'s decay exponentially to zero with G, and thus so does $P_e(G)$. In particular, the decay of $P_e(G)$ will be dominated by the smallest (worst) error exponent. Let $D^*_{B,min} = \min\{D^*_{B,jm}\}$ and $D^*_{min} = \min\{D^*_{jm}\}$ be the smallest pairwise Bhattacharya and Chernoff exponents. Using the dominant exponents in (25) we get

$$P_e(G) \leq \frac{C_2^M}{M} e^{-D^*_{min}G} \leq \frac{C_2^M}{M} e^{-D^*_{B,min}G} \tag{26}$$

where $C_2^M = \frac{M!}{2!(M-2)!}$ is the number of pairwise hypothesis tests.

Asymptotic Performance as $G \to \infty$. Note from (12) that by the law of large numbers, under H_m we have

$$\lim_{G\to\infty} l_j(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_G) \;\; = \;\; -\mathrm{E}_m[\log p_j(\boldsymbol{Z})] = D(p_m\|p_j) + h_m(\boldsymbol{Z}) \tag{27}$$

where $D(p_m\|p_j)$ is the Kullback-Leibler (K-L) distance between the pdf's p_j and p_m and $h_m(\boldsymbol{Z})$ is the differential entropy of $\boldsymbol{Z}$ under H_m [10]

$$D(p_m\|p_j) \;\; = \;\; \mathrm{E}_m\left[\log(p_m(\boldsymbol{Z})/p_j(\boldsymbol{Z}))\right] = \log\left(|\widetilde{\boldsymbol{\Sigma}}_j|/|\widetilde{\boldsymbol{\Sigma}}_m|\right) + \mathrm{tr}\left(\widetilde{\boldsymbol{\Sigma}}_j^{-1}\widetilde{\boldsymbol{\Sigma}}_m - \boldsymbol{I}\right) \tag{28}$$

$$h_m(\boldsymbol{Z}) \;\; = \;\; -\mathrm{E}_m[\log p_m(\boldsymbol{Z})] = \log\left((\pi e)^N|\widetilde{\boldsymbol{\Sigma}}_m|\right). \tag{29}$$

An important property of K-L distance is that $D(p_m\|p_j) \geq 0$ with equality if and only if $p_m = p_j$, in which case there is no way to distinguish between the two classes. Thus, from (27) we conclude that under H_m, l_m will always give the smallest value and thus lead to the correct decision as $G \to \infty$ as long as $D(p_j\|p_m) > 0$ for all $j \neq m$. From Proposition 1, we know that $P_e(G)$ decays to zero exponentially. Thus, we have the following result.

Proposition 2 *If all pairwise K-L distances are strictly non-negative, that is*

$$D(p_m\|p_j) > 0 \;\; \forall\, j, m, \; j \neq m \tag{30}$$

then $P_e(G)$ decays exponentially to zero with G

$$\lim_{G\to\infty} \frac{\log P_e(G)}{G} \leq -D^*_{min} \tag{31}$$

where $D^*_{min} = \min\{D^*_{jm}\}$ is the smallest pairwise Chernoff information defined in (21).

Note that (31) directly follows from (25). 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.

3.2 Hard Decision Fusion

In soft decision fusion, the k-th SCR sends M log-likelihood values $\{\boldsymbol{z}_k^T \widetilde{\boldsymbol{\Sigma}}_j^{-1} \boldsymbol{z}_k : j = 1, \cdots, M\}$, computed from its local measurement $\boldsymbol{z}_k$, to the manager node. While exchange of real-valued likelihoods puts much less communication burden on the network as compared to *data fusion* in which the feature vectors $\{\boldsymbol{z}_k\}$ are communicated to the manager node, it is attractive to reduce the communication burden even further. One way is to quantize the M likelihood values from different SCR's with sufficient number of bits. The number of bits required for accurate communication can be estimated from the differential entropy of the log-likelihoods [10, 8]. Another natural quantization strategy is to compute local *hard decisions* in each SCR based on the local measurement $\boldsymbol{z}_k$. In this section we discuss this hard decision fusion approach, assuming noise-free communication links from the SCR's to the manager node.

We assume that the locally optimal (based on the local measurement $\boldsymbol{z}_k$) hard decision is made in in k-th SCR

$$u_k = \arg \max_{j=1,\cdots,M} p_j(\boldsymbol{z}_k) \ , \ k = 1, \cdots, G. \tag{32}$$

Note that u_k is a discrete random variable with M possible values. Furthermore, since all $\{\boldsymbol{z}_k\}$ are i.i.d., so are $\{u_k\}$. Let $\{p_m[j] : j = 1, \cdots, M\}$ denote the M values of the pmf of the hard decision variable U under H_m. The pmf's of U under all hypotheses are characterized by the following probabilities

$$p_m[j] = P(U = j|H_m) = P\left(p_j(\boldsymbol{z}_k) \geq p_l(\boldsymbol{z}_k) \text{ for all } l \neq j \,|H_m\right) \ , \ j, m = 1, \cdots, M \, . \tag{33}$$

The hard decisions $\{u_k\}$ from all SCR's are communicated via noise-free links to the manager

node which makes the final optimal[7] decision

$$C_{hard}(u_1, \cdots, u_G) = \arg \max_{j=1,\cdots,M} p_j[u_1, \cdots, u_G] \tag{34}$$

$$p_j[u_1, \cdots, u_G] = \prod_{k=1}^{G} p_j[u_k] \,, \tag{35}$$

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

$$C_{hard}(u_1, \cdots, u_G) = \arg \min_{j=1,\cdots,M} l'_j[u_1, \cdots, u_G] \tag{36}$$

$$l'_j[u_1, \cdots, u_G] = -\frac{1}{G} \log p_j[u_1, \cdots, u_G] = -\frac{1}{G} \sum_{k=1}^{G} \log p_j[u_k] \,. \tag{37}$$

3.2.1 Performance of Hard Decision Fusion

While the exact calculation of P_e is difficult, it can be bounded via PEP's analogous to soft decision fusion. In particular, in this case $\mu_{jm,hard}(\theta)$ is given by

$$\mu_{jm,hard}(\theta) = \log \mathrm{E}_m[p_j^\theta[U]/p_m^\theta[U]] = \log \sum_{i=1}^{M} p_j^\theta[i] p_m^{1-\theta}[i] \tag{38}$$

which can be manipulated numerically to compute the pairwise Chernoff and Bhattacharya exponents: $D^*_{jm,hard} = -\min_{\theta \in [0,1]} \mu_{jm,hard}(\theta)$ and $D^*_{B,jm,hard} = -\mu_{jm,hard}(1/2)$. Then the PEP's and $P_{e,hard}$ can be bounded analogous to (23), (25) and (26).

For a given measurement SNR and G, we expect the PEP's to be higher compared to soft decision fusion due to local hard decisions. In fact, the pmf's of hard decision in (33) are based on decision statistics that are weighted sums of N χ^2_2 random variables, compared to NG χ^2_2 in soft decision fusion, thereby resulting in less reliable hard decisions [8].

Asymptotic Performance as $G \to \infty$. Despite the fact that the local hard decisions can be quite unreliable, the final hard decision fusion classifier can still attain perfect performance as $G \to \infty$. Note from (37) that because of law of large numbers, under H_m we have

$$\lim_{G \to \infty} l'_j[u_1, \cdots, u_G] = -\mathrm{E}_m[\log p_j[U]] = D(p_m \| p_j) + H_m(U) \tag{39}$$

[7] Optimal, given the the hard decisions $\{u_k\}$.

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 [10]

$$D(p_m \| p_j) = \sum_{i=1}^{M} p_m[i] \log(p_m[i]/p_j[i]) \quad (40)$$

$$H_m(U) = -\sum_{i=1}^{M} p_m[i] \log p_m[i]. \quad (41)$$

Thus, we see from (39) that in the limit of large G we will attain perfect classification performance as long as $D(p_m \| p_j) > 0$ for all $j \neq m$. Furthermore, as in soft decision fusion, $P_{e,hard}(G)$ decays exponentially with G.

Proposition 3 *If all pairwise K-L distances between the pmf's in (33) are strictly non-negative, then $P_{e,hard}(G)$ decays exponentially to zero with G*

$$\lim_{G \to \infty} \frac{\log P_{e,hard}(G)}{G} \leq -D^*_{min,hard} \quad (42)$$

where $D^*_{min,hard} = \min\{D^*_{jm,hard}\}$ is the smallest pairwise Chernoff information for hard decision fusion.

Note that for a given measurement SNR, the error exponent for hard decision fusion will be smaller compared to soft decision fusion, since the pairwise K-L distances between the pmf's in hard decision fusion will be smaller than those between the pdf's in soft decision fusion.

4 Decision Fusion with Noisy Communication Links

In this section, we discuss decision fusion from different SCR's using noisy communication links, as illustrated in Fig. 2(b). Given the competitive performance of noise-free hard decision fusion compared to soft decision fusion, we focus on noisy fusion of hard decisions.[8]

We assume that each SCR has a dedicated communication link to the manager node.[9] Each SCR sends an amplified version of its hard decision u_k in (32) over a noisy link

$$y_k = \alpha u_k + w_k \ , \ \ k = 1, \cdots, G \quad (43)$$

[8]Noisy *analog fusion* of soft decisions is also an attractive choice, but we do not discuss it here due to lack of space.
[9]Note that this requires large bandwidth or large latency at the manager node in the limit of large G.

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 M is odd and define $\widetilde{M} = (M-1)/2$. We assume that each SCR sends a symmetrized version of its hard decision to use minimum power: $u_k \in \{-\widetilde{M}, \cdots, \widetilde{M}\}$.[10] Given this simple communication scheme, the optimal classifier at the manager node takes the form

$$C_{noisy}(\boldsymbol{y}) = \arg\min_j l_{j,noisy}(\boldsymbol{y}) \tag{44}$$

$$l_{j,noisy}(\boldsymbol{y}) = -\frac{1}{G}\log p_{,noisy}(\boldsymbol{y}) = -\frac{1}{G}\sum_{k=1}^{G}\log p_{j,noisy}(y_k) \tag{45}$$

$$p_{j,noisy}(y) = \frac{1}{\sqrt{2\pi\sigma_w^2}}\sum_{i=-\widetilde{M}}^{\widetilde{M}} e^{-(y-\alpha i)^2/2\sigma_w^2} p_j[i]. \tag{46}$$

The exact calculation of $P_{e,noisy}$ is most complicated in this case; however, it can be bounded via Bhattacharya or Chernoff bounds for the PEP's as discussed in previous sections. In particular, in this case $\mu_{jm,noisy}(\theta)$ takes the form

$$\mu_{jm,noisy}(\theta) = \log \mathrm{E}_m[p^{\theta}_{j,noisy}(\boldsymbol{Y})/p^{\theta}_{m,noisy}(\boldsymbol{Y})] \tag{47}$$

which can be computed numerically. Most importantly, for sufficiently large measurement and communication SNR's, we again expect exponentially vanishing $P_{e,noisy}$ since from (46) we have under H_m

$$\lim_{G\to\infty} l_{j,noisy}(\boldsymbol{y}) = -\mathrm{E}_m[\log p_{j,noisy}(\boldsymbol{Y})] = D(p_{m,noisy}\|p_{j,noisy}) + h_m(\boldsymbol{Y}). \tag{48}$$

Thus, following the lines in noise-free fusion, we immediately have the following result.

Proposition 4 *If all pairwise K-L distances between noisy pdf's in (46) are strictly non-negative then $P_{e,noisy}$ decays exponentially to zero as $G \to \infty$*

$$\lim_{G\to\infty} \frac{P_{e,noisy}}{G} \le -D^*_{min,noisy} \tag{49}$$

where $D^*_{min,noisy} = \min\{D^*_{jm,noisy}\}$ is the smallest pairwise Chernoff information for noisy hard decision fusion.

We expect the K-L distances in this case to be smaller than those for noise-free hard decision fusion. In particular, they depend on both the measurement and communication SNR's.

[10] Note that this is not necessarily the optimal symbol assignment from the viewpoint of final decision.

5 Numerical Results Based on Real Data

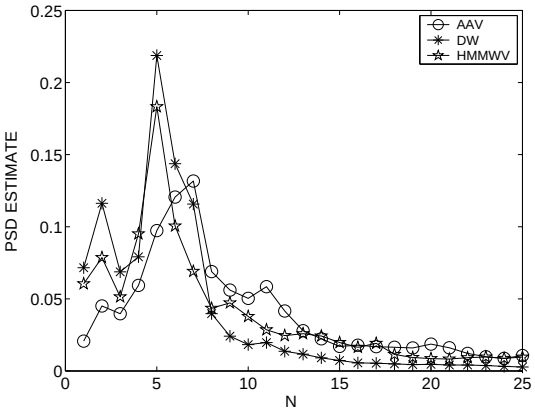

Figure 3: *Covariance matrix eigenvalues (PSD estimates) based on acoustic measurements for three vehicles: AAV, DW, and HMMWV.*

In this section, we present numerical results to illustrate the performance of the three classifiers C, C_{hard} and C_{noisy} as a function of G for different measurement and communication SNRs which are defined(in dB) as

$$\mathrm{SNR}_{meas} \;=\; 10\log_{10}(\mathrm{tr}(\boldsymbol{\Sigma}_j)/N\sigma_n^2) \;\;,\;\; \mathrm{SNR}_{comm} = 10\log_{10}\left(\sum_i (\alpha i)^2 p[i]/\sigma_w^2\right) \qquad (50)$$

where $p[i] = P(U=i) = \sum_m p_m[i]/M$ and recall that we assume constant $\mathrm{tr}(\boldsymbol{\Sigma}_j)$ (signal energy) for all classes. Our results are based on real measurements collected during the DARPA SensIT program and correspond to classifying a vehicle from $M = 3$ classes based on acoustic signals collected by microphones [3]. The three vehicle classes are: Amphibious Assault Vehicle (AAV), Dragon Wagon (DW) and Humvee (HMMWV). PSD values at $N = 25$ 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) $\boldsymbol{\Sigma}_j = \boldsymbol{\Lambda}_j$. Under H_j, the $N = 25$-dimensional averaged measurement vector $\boldsymbol{z}_k$ in the k-th SCR was simulated as

$$\boldsymbol{z}_k = \boldsymbol{\Lambda}_j^{1/2}\boldsymbol{v}_k + \boldsymbol{n}_k \;\;,\;\; k = 1, \cdots, G \qquad (51)$$

where $\{\boldsymbol{v}_k\}$ are i.i.d $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ and $\{\boldsymbol{n}_k\}$ are i.i.d. $\mathcal{CN}(\boldsymbol{0}, \sigma_n^2\boldsymbol{I})$. P_e and $P_{e,hard}$ were estimated via Monte Carlo simulation using 7000 independent sets of G measurements for each hypothesis. The pmf's $\{p_m[i]\}$ for hard decisions were also estimated via this Monte Carlo simulation.

The measurements for noisy hard decision fusion were simulated using (43) and the pmf's for hard decisions. $P_{e,noisy}$ was estimated using 10000 independent sets of measurement realizations. The simulations were done for four values for $\mathrm{SNR}_{meas} = -4, 0, 4, 10$ dB and three values for $\mathrm{SNR}_{comm} = 0, 5, 10$ dB.

Measurement SNR=-4dB

m/j	AAV	DW	HMMWV
AAV	0	0.4665	0.2584
	0	0.2773	0.1360
	0	0.0714	0.0706
	0	0.1707	0.1201
	0	0.2716	0.1583
DW	0.5159	0	0.1649
	0.2829	0	0.0827
	0.0667	0	0.0098
	0.1631	0	0.0319
	0.2737	0	0.0608
HMMWV	0.2711	0.1607	0
	0.1316	0.0794	0
	0.0739	0.0090	0
	0.1182	0.0325	0
	0.1520	0.0642	0

(a)

Measurement SNR=0dB

m/j	AAV	DW	HMMWV
AAV	0	1.3777	0.6478
	0	0.8013	0.3520
	0	0.1728	0.1548
	0	0.4281	0.2441
	0	0.7212	0.3231
DW	1.4559	0	0.4922
	0.7998	0	0.2819
	0.1442	0	0.0288
	0.3927	0	0.1129
	0.7273	0	0.2247
HMMWV	0.6880	0.5114	0
	0.3375	0.2725	0
	0.1509	0.0328	0
	0.2403	0.1161	0
	0.3073	0.2139	0

(b)

Measurement SNR=4dB

m/j	AAV	DW	HMMWV
AAV	0	3.3380	1.2145
	0	1.6520	0.6680
	0	0.3258	0.3630
	0	0.8727	0.5644
	0	1.4901	0.6556
DW	3.1455	0	1.1893
	1.6422	0	0.7002
	0.2690	0	0.0895
	0.7238	0	0.3139
	1.4654	0	0.6322
HMMWV	1.2971	1.4010	0
	0.6482	0.6921	0
	0.3707	0.0953	0
	0.5498	0.3162	0
	0.6502	0.5981	0

(c)

Measurement SNR=10dB

m/j	AAV	DW	HMMWV
AAV	0	8.6186	2.0692
	0	3.3931	1.0723
	0	0.4604	0.6317
	0	1.4600	0.9385
	0	2.8608	1.0804
DW	6.4357	0	2.9212
	3.2744	0	1.7447
	0.3537	0	0.2256
	1.1475	0	0.7527
	2.7645	0	1.4739
HMMWV	2.1531	4.2559	0
	1.0546	1.8374	0
	0.6383	0.2769	0
	0.9205	0.8524	0
	1.0371	1.6038	0

(d)

Table 1: *Pairwise K-L distances for the 3 vehicles classes for soft, hard and noisy hard decision fusion at four different* SNR_{meas}*. The first entry in each cell is the K-L distance for soft, the second for hard, and the remaining three for noisy hard decision fusion at* $\mathrm{SNR}_{comm} = 0, 5, 10$ *dB respectively. (a)* SNR_{meas} *= -4dB. (b)* SNR_{meas} *= 0dB. (c)* SNR_{meas} *= 4dB. (d)* SNR_{meas} *= 10dB.*

Table 1 shows the values of all pairwise K-L distances for soft, hard and noisy hard decison fusion for all values of SNR_{meas} and SNR_{comm}. Note that the K-L distances get larger with SNR_{meas}, as expected. For a given SNR_{meas}, the K-L distances decrease from soft to hard to noisy hard fusion depending on SNR_{comm}.

Fig. 4 shows plots of numerically estimated P_e, $P_{e,hard}$ and $P_{e,noisy}$ for the four values of SNR_{meas}. Three plots for $P_{e,noisy}$ are shown corresponding to the three values of SNR_{comm}. As expected, ideal soft decision fusion is better than ideal hard decision fusion which is in turn better than noisy hard decision fusion. The gap between ideal soft and hard decision fusion can be significant. The gap between ideal hard and noisy hard decision fusion, on the other hand, is fairly small at

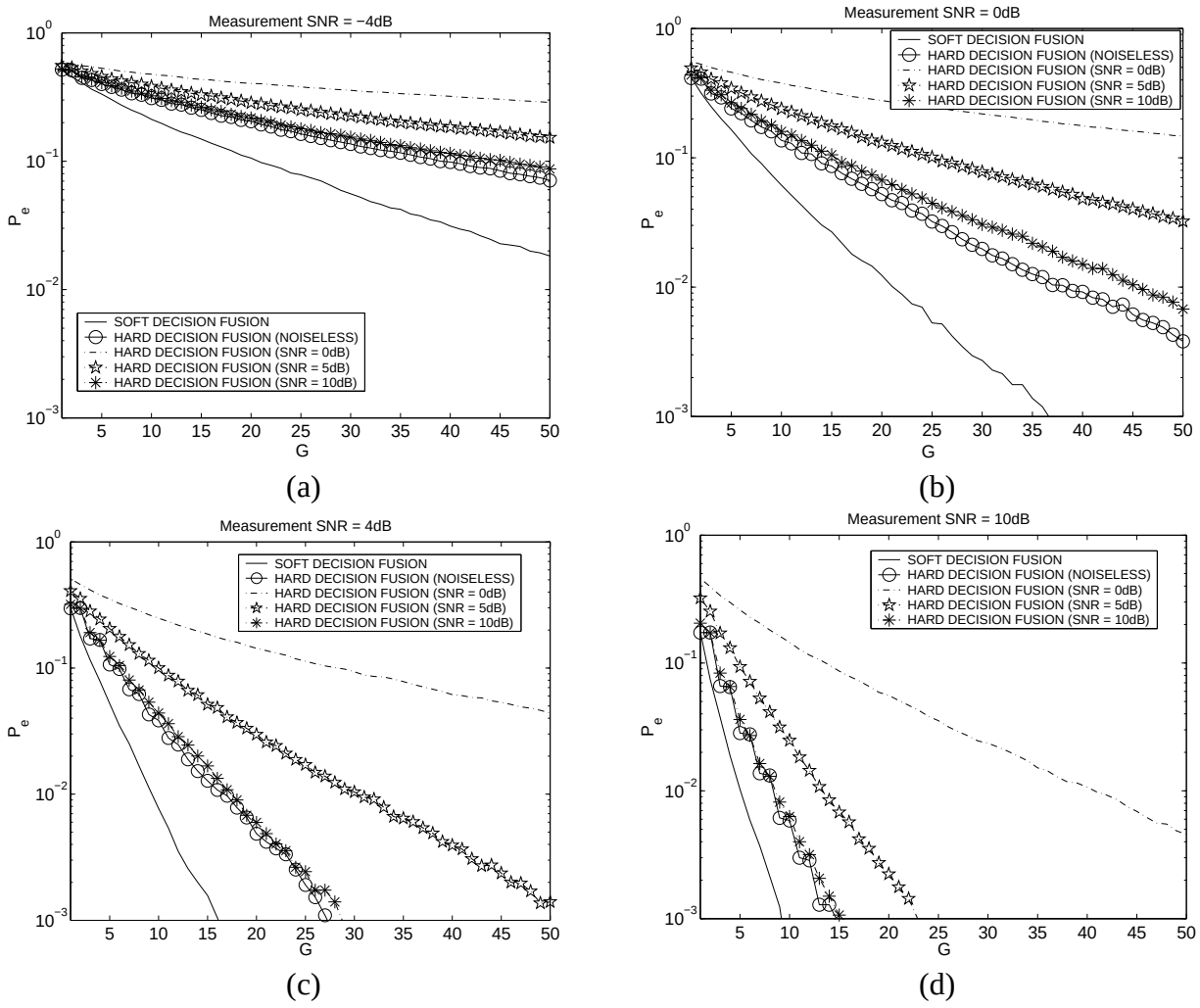

Figure 4: *P_e as a function of G for noise-free soft and hard decision fusion, and noisy hard decision fusion for different values of* SNR_{meas}*. For the noisy hard decision fusion, three plots are shown for* $\mathrm{SNR}_{comm} = 0, 5, 10$ *dB. (a)* SNR_{meas} *= -4dB. (b)* SNR_{meas} *= 0dB. (c)* SNR_{meas} *= 4dB. (d)* SNR_{meas} *= 10dB.*

$\mathrm{SNR}_{comm} = 10$dB. Note that even at $\mathrm{SNR}_{meas} = 0$dB, $P_e \approx 10^{-2}$ for soft decision fusion is attained with only $G \approx 20$ independent measurements. Furthremore, the same performance can be attained with the much simpler hard decision fusion (both ideal and noisy) around $G = 40$. At $\mathrm{SNR}_{meas} = 10$dB, which could be attained by averaging over $n_G = 10$ measurements at 0 dB within each SCR, only $G \approx 8$ independent measurements are needed to attain $P_{e,noisy} \approx 10^{-2}$ with noisy hard decision fusion (Note that $P_{e,noisy} \approx 0.2$ for $G = 1$.). 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 node decisions. And this can be achieved with simple communication schemes, as the

one in noisy hard decision fusion.

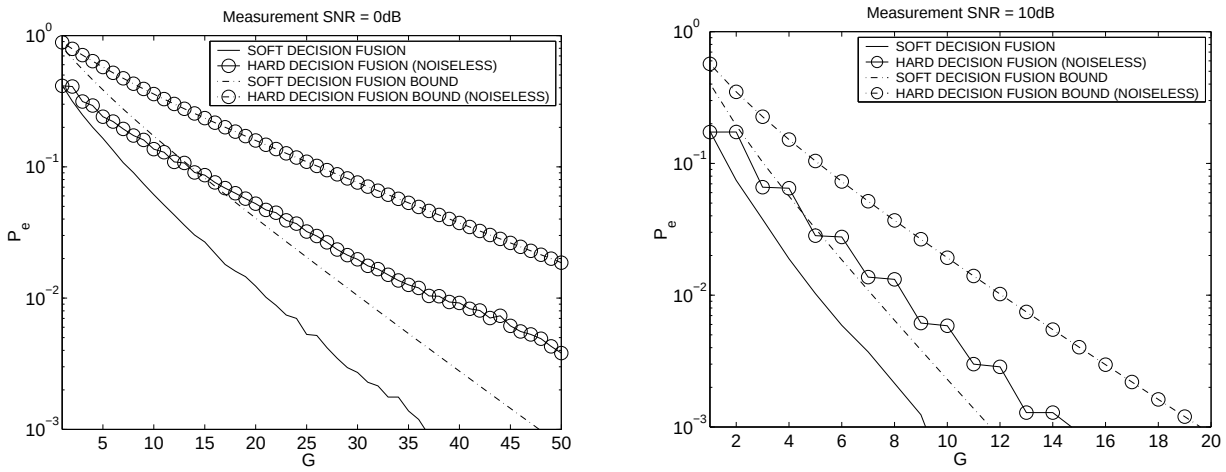

Figure 5: *Comparison of simulated P_e and P_e bounds based on the Bhattacharya PEP bounds for noiseless hard and soft decision fusion. (a)* SNR_{meas} *= 0dB. (b)* SNR_{meas} *= 10dB.*

Finally, in Fig. 5 we compare simulated P_e and $P_{e,hard}$ for noise-free soft and hard decision fusion with analytical bounds (see (25)) based on Bhattacharya PEP bounds. As evident, 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 SNR_{meas}. We note that bounds based on the Chernoff PEP bounds would be tighter.[11]

6 Discussion and Conclusions

All applications of sensor networks are built on two primary operations: i) distributed processing of data collected by the nodes, and ii) communication and routing of processed data from one part of the network to another. Furthermore, the two operation are intimately tied: information flow in a sensor network directly depends on the data collected by the nodes, and the nature of information exchange between nodes is constrained by the communication protocols. Thus, distributed signal processing techniques need to be developed in the context of communication and routing techniques and vice versa. In this paper we have discussed distributed decision making in a simple context — classification of a single object — to study basic principles that govern the interaction between information processing and information routing. Our approach is based on modeling the object signal as a bandlimited Gaussian field in space and time. This simple model partitions the

[11] Space-permitting, we plan to include them in the final revised version of the paper.

network into disjoint Spatial Coherence Regions (SCR's) whose size is inversely proportional to the spatial signal bandwidths. This partitioning of network nodes into SCR's suggests a structure on information exchange between nodes that is naturally suited to the communication constraints in the network: high-bandwidth feature-level *data-fusion* is limited to spatially local nodes within each SCR, wheras global fusion of low-bandwidth *local decisions* at each SCR is sufficient across different SCR's. The data-fusion within each SCR improves the effective measurement SNR, whereas decision-fusion across SCR's combats the statistical variability in the signal. This simple structure on the nature of information exchange between nodes applies to virtually all CSP algorithms, including distributed estimation and compression [4].

Our stochastic signal model was motivated by the fact that the signals sensed by different nodes, even when due to the same underlying source, will exhibit variability due to a multitude of factors [3]. In the face of statistical variability, we need some form of ergodicity for reliable decision making which is precisely provided by multiple independent node measurements. Our results underscore the remarkable advantage of multiple node measurements: fairly unreliable local decisions from multiple SCRs can be combined to yield arbitrary reliable final decisions. In particular, we showed that under very mild conditions, the probability of misclassification for ideal soft and hard decision fusion as well as noisy hard decision fusion decays exponentially to zero with the number of independent measurements. Furthermore, our numerical results indicate that at fairly reasonable measurement SNRs (around 0-10 dB), relatively modest number of independent measurements (10-50) can yield desirable probabilities of error ($\approx 10^{-2}$).

The performance of noisy hard decision fusion, the most attractive approach in practice, depends on both the measurement and communication SNR's as well as the number of independent measurements G. The SNR's have to be just high enough so that all pairwise K-L distances are non-vanishing; this guarantees exponentially decay in $P_e(G)$ with G. However, higher SNR's yield better error exponents. A related observation is that rapidly varying signals (with large bandwidths) require cooperation between nodes in a smaller region (size of SCR's is small) to yield a sufficient number of independent measurements. Furthermore, multiple independent measurements could be also be collected at each node over time to further improve performance at the cost of latency.

Some of the assumptions in this work warrant further comment.

Non-idealities in practical settings. Examples of non-idealities that could be incorporated in future work include non-stationary, non-Gaussianity, faulty sensors and propagation path loss.

Power-bandwidth-latency tradeoff. We assumed independent channels from different SCR's to the manager node. For large number of SCR's, this requires a large bandwidth at the manager node, or imposes a large latency in decision making. On the other hand, if the decisions from different SCR's were communicated on a single narrowband multiple-access channel, then reliable noisy decision fusion would require a higher transmission power at each node. A combination of dedicated and multiple-access channels would likely be needed based on power, bandwidth and latency constraints.

Multi-target classification. Simultaneous classification of multiple objects is a much more challenging problem. For example, the number of possible hypotheses increases exponentially with the number of objects. Several forms of sub-optimal algorithms, including tree-structured classifiers [6], subspace-based approaches [11, 12] and sub-optimal fusion schemes [13] 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 Proces sing 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. Sayeed, "A statistical signal modeling framework for optimizing sensor networks," in *in preparation for the IEEE Signal Processing Magazine*.

[5] J. G. Proakis, *Digitial Communications*. New York: McGraw Hill, 3rd ed., 1995.

[6] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley, 2nd ed., 2001.

[7] R. M. Gray, "On the asymptotic eigenvalue distribution of toeplitz matrices," *IEEE Trans. Inform. Th.*, vol. 18, pp. 725–730, Nov. 1972.

[8] 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.

[9] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[11] 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.

[12] S. Watanabe and N. Pakvasa, “Subspace method to pattern recognition,” *Proceedings of the 1st International Conference on Pattern Recognition*, pp. 25–32, Feb. 1973.

[13] 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.