

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

Ashwin D'Costa, *Student Member, IEEE*, Vinod Ramachandran, and Akbar M. Sayeed, *Senior Member, IEEE*

Abstract—Distributed signal processing techniques for classification of objects are studied assuming knowledge of sensor measurement statistics. The spatio-temporal signal field generated by an object is modeled as a bandlimited stationary ergodic Gaussian field. The model suggests 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. It is shown 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 signal-to-noise ratio (SNR), and global low-bandwidth exchange of local decisions across coherence regions to stabilize the inherent variability in the signal. Classifier performance is analyzed for both soft and hard 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 distributed decision making.

Index Terms—Chernoff bounds, distributed classification, error exponents, sensor networks, signal modeling.

I. 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 acoustic, seismic, and infrared [1]–[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. Fusion of sensor information from 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. This requires the development of theory and methods for collaborative signal processing (CSP) of the data collected by different nodes.

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.

There are two main forms of information fusion across 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 exhibit a mixture of correlated and independent components and require a combination of data and decision fusion between nodes. Decision fusion is clearly the more attractive choice in terms of the communication burden on the network.

The statistics of node measurements are determined by the spatio-temporal signal field generated by the underlying object of interest. 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 natural structure on CSP algorithms for decision making in which costly¹ *data fusion* is confined to local spatial coherence regions (SCRs) in which the signal is strongly correlated, and only cheaper *decision fusion* is needed across different SCRs in which the signal is nearly independent. This model forms the basis of the distributed classification schemes studied.

Manuscript received July 15, 2003; revised February 1, 2004. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) under Grant F30602-00-2-0555 and by the National Science Foundation (NSF) under ITR Grant CCR-0113385. This paper was presented in part at the 2003 International Symposium on Information Processing in Sensor Networks, Los Angeles, CA.

The authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: dcosta@cae.wisc.edu; vinod@cae.wisc.edu; akbar@engr.wisc.edu).

Digital Object Identifier 10.1109/JSAC.2004.830896

¹In terms of bandwidth and power expenditure.

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 I discusses optimal CSP algorithms for classification that assume noise-free communication links from different SCRs to the manager node. Both soft and hard decision fusion is discussed. In Section IV, 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 SCRs—the error exponent decreases from soft to hard to noisy fusion. In Section V, we present numerical results based on real measurement data to illustrate the performance of the three classification schemes. 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 VI presents a discussion of the results, as well as pointers for future work.

II. SIGNAL MODEL FOR SENSOR MEASUREMENTS

A. 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 may exhibit non-Gaussianity and nonstationarity, this is a reasonable assumption for initial investigations to gain insight. Specifically, $s(x, y, t)$ is represented as

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

where $\phi(\nu_x, \nu_y, f)$ denotes the underlying spectral representation³ which satisfies

$$\begin{aligned} \mathbb{E}[\phi(\nu_x, \nu_y, f)\phi^*(\nu'_x, \nu'_y, f')] \\ = \Phi(\nu_x, \nu_y, f)\delta(\nu_x - \nu'_x)\delta(\nu_y - \nu'_y)\delta(f - f') \end{aligned} \quad (2)$$

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 three-dimensional Fourier transform

$$\begin{aligned} r_s(\Delta x, \Delta y, \Delta t) \\ = \mathbb{E}[s(x + \Delta x, y + \Delta y, t + \Delta t)s^*(x, y, t)] \end{aligned}$$

²We assume a complex signal field for generality; e.g., it would be applicable for complex baseband modeling of bandpass signals generated by transducers.

³Strictly 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.

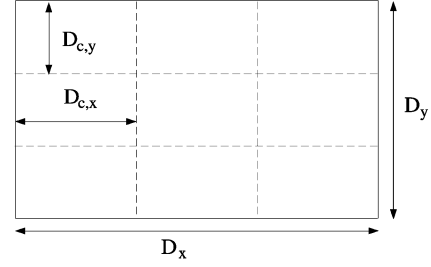


Fig. 1. Schematic illustrating the notion of SCRs. The overall region of size $D_x \times D_y$ is partitioned into SCRs of size $D_{c,x} \times D_{c,y}$, where $D_{c,x} = 1/B_x$ and $D_{c,y} = 1/B_y$ denote the coherence distances in x and y dimensions. The spatial signal field is strongly correlated within each SCR, whereas it is approximately uncorrelated across SCRs (except near the boundaries).

$$\begin{aligned} &= \int_{-B_x/2}^{B_x/2} \int_{-B_y/2}^{B_y/2} \int_{-B_t/2}^{B_t/2} \Phi(\nu_x, \nu_y, f) \\ &\times e^{j2\pi(\nu_x \Delta x + \nu_y \Delta y + f \Delta t)} d\nu_x d\nu_y df \end{aligned} \quad (3)$$

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

B. Approximate Signal Modeling Via Spatial Coherence Regions

To enable efficient CSP, we propose an approximate signal model, based on SCRs 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 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 SCRs, $\{R_{c,ij}\}$, 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 SCRs (except near the boundary of the SCRs) [4]. The uniform size of SCRs follows from the stationarity assumption.

Specifically, we use a piece-wise constant (PWC) approximation of the stationary signal commensurate with the SCRs

$$\begin{aligned} s_{\text{pwc}}(x, y, t) &= \sum_{i,j} s_{ij}(t) I_{R_{c,ij}}(x, y) \\ &= \sum_{i=-\tilde{N}_x}^{\tilde{N}_x} \sum_{j=-\tilde{N}_y}^{\tilde{N}_y} s_{ij}(t) I_{D_{c,x}}(x - iD_{c,x}) \\ &\quad \times I_{D_{c,y}}(y - jD_{c,y}) \end{aligned} \quad (4)$$

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 $s_{\text{pwc}}(x, y, t)$ 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 SCRs

$$s_{ij}(t) = \frac{1}{D_{c,x} D_{c,y}} \int_{R_{c,ij}} s(x, y, t) dx dy. \quad (5)$$

Example: Temporal Point Sources: In general, the spatial and temporal signal characteristics can be arbitrary. However, for this important class of sources they are intimately related. Such sources are characterized by an 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 and the bandwidth in the radial dimension is $B_r = B_t/v$. The true SCRs 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 an acoustic source with $B_t = 500$ 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 SCRs in Fig. 1 is a natural choice.

Spatial Degrees-of-Freedom (DoF): It can be shown that $s_{\text{pwc}}(x, y, t)$ 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 $s(x, y, t)$ is 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 $s_{\text{pwc}}(t)$ 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 SCRs are approximately uncorrelated [4]. Thus, there are approximately $N_s = N_x N_y$ independent spatial DoF in $s(x, y, t)$ over R which are preserved by $s_{\text{pwc}}(x, y, t)$. 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]).⁴

Assumptions on Sensor Measurements: Based on the above discussion, we make two assumptions about the *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 SCRs are statistically independent.

In the region $R = D_x \times D_y$, there are $G = N_s = N_x N_y$ independent SCRs. From now on, we will label the SCRs by a single index: $R_{c,k}$, $k = 1, \dots, G$. We assume that n_G node measurements are collected in each SCR, resulting in a total of $K = G n_G$ measurements. We model the temporal signal at the i th node as

$$x_i(t) = s_i(t) + n_i(t), \quad i = 1, \dots, K = G n_G \quad (6)$$

where $n_i(t)$ denotes a 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 high rate in disjoint

⁴Note that $N_x = D_x B_x$ and $N_y = D_y B_y$ are the distance-bandwidth products of the spatial signal in the x and y dimensions, analogous to the time-bandwidth product of temporal signals, and quantify the independent dimensions (degree of freedom) in the spatial signal field.

blocks of N samples. Let $\{\mathbf{x}_i, i = 1, \dots, K\}$ denote the N -dimensional measurement vectors at the K nodes.

We note that the above approximation of statistically independent SCRs is equivalent to modeling the spatial signal field as a band-limited white noise field, with the same spatial bandwidths. Thus, we are ignoring the fine spatial correlation structure in the signal (the shape of spatial PSDs) since accounting for it comes at the cost of coordination/communication across different nodes. Detailed signal characteristics are exploited via temporal processing (in-node processing) to facilitate classification, and only the independent DoF are exploited in space.

At the sensing level, there are two sources of error in decision making: 1) the measurement noise and 2) the statistical variability in the source signal. The notion of SCRs 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, 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 SCRs.
- Second, local independent decisions from different SCRs are appropriately combined to reduce the statistical variability in the final decision. Thus, low-bandwidth decision fusion is sufficient across SCRs.

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

$$\mathbf{z}_k = \frac{1}{n_G} \sum_{i \in R_{c,k}} \mathbf{x}_i = \mathbf{s}_k + \mathbf{n}_k, \quad k = 1, \dots, G. \quad (7)$$

If the original noise variance is σ_n^2 , the variance of the averaged noise becomes σ_n^2/n_G . Note that the signal component of node measurements in two adjacent SCRs will exhibit higher correlation near the common boundary. Thus, in practice, the above averaging of node measurements to improve the SNR should be done for nodes near the center of the SCR to minimize the correlation across SCRs. Let $\mathbf{s} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a complex circular Gaussian vector with mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{s}]$ and covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{s}\mathbf{s}^H]$, where $(\cdot)^H$ denotes Hermitian transpose. Then, $\{\mathbf{s}_k\}$ are i.i.d. $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ for some $\boldsymbol{\Sigma}$ and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}/n_G)$, where \mathbf{I} denotes the identity matrix.

Feature Selection: An important issue in classification is: what kind of measurements $\{\mathbf{x}_i\}$ should be collected? This is the called *feature selection* [6]. Essentially, the raw time series data collected 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 $\{\mathbf{x}_i\}$ —spectral feature vectors—that can be obtained by computing a Fourier transform of the raw data. This is a natural consequence of the stationary signal model. The covariance matrix $\boldsymbol{\Sigma}$ of the temporal signal is Toeplitz and it is well known that the discrete Fourier transform (DFT) matrix diagonalizes Toeplitz matrices in the limit of large dimension [7]; that is, $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where \mathbf{U} is the DFT matrix and $\boldsymbol{\Lambda} = \text{diag}(\lambda[1], \dots, \lambda[N])$ is the diagonal matrix of eigenvalues; the eigenvalues are proportional to the temporal PSD of

the signal at different frequencies. Thus, in numerical results we will explicitly substitute $\Sigma = \Lambda$.

III. 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

$$H_j: \mathbf{z}_k = \mathbf{s}_k + \mathbf{n}_k, \quad k = 1, \dots, G, \quad j = 1, \dots, M \quad (8)$$

where $\{\mathbf{z}_k\}$ are i.i.d. $\mathcal{CN}(\mathbf{0}, \tilde{\Sigma}_j)$ under H_j and $\tilde{\Sigma}_j = \Sigma_j + \sigma_n^2 \mathbf{I}/n_G$. All information about the targets is contained in the covariance matrices $\{\Sigma_j\}$. In practice, $\{\Sigma_j\}$ have to be estimated from available training data. We assume that $\{\Sigma_j\}$ are known *a priori* and that $\text{tr}(\Sigma_j)$ (signal energy)⁵ is the same for all j . Based on the measurement vectors $\{\mathbf{z}_k\}$ from the G SCRs, 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).

A. Soft Decision Fusion

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

$$C(\mathbf{z}_1, \dots, \mathbf{z}_G) = \arg \max_{j=1, \dots, M} p_j(\mathbf{z}_1, \dots, \mathbf{z}_G) \quad (9)$$

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

$$p_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = \prod_{k=1}^G p_j(\mathbf{z}_k) \\ p_j(\mathbf{z}_k) = \frac{e^{-\mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k}}{\pi^N |\tilde{\Sigma}_j|} \quad (10)$$

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

$$C(\mathbf{z}_1, \dots, \mathbf{z}_G) = \arg \min_{j=1, \dots, M} l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) \quad (11)$$

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

⁵ $\text{tr}(\cdot)$ denotes the trace of a matrix.

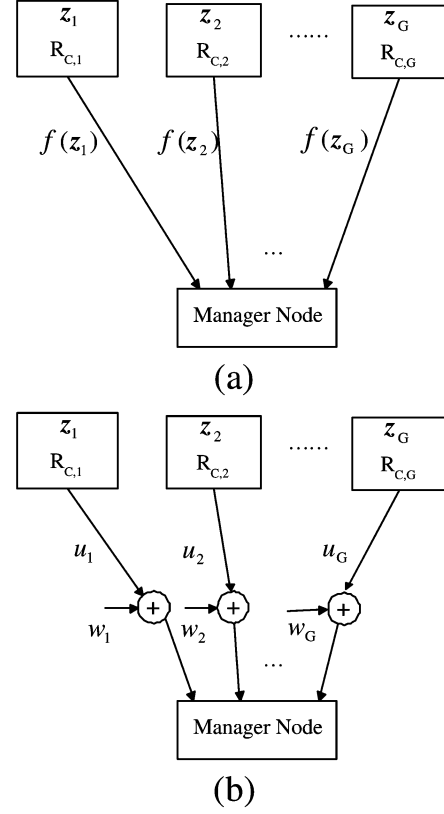


Fig. 2. (a) Noise-free decision fusion. For soft decisions $f(\mathbf{z}_k) = \{\mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k, j = 1, \dots, M\}$, whereas for hard decisions $f(\mathbf{z}_k) = u_k = \arg \max_j p_j(\mathbf{z}_k)$. (b) Noisy hard decision fusion.

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

$$l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = \log |\tilde{\Sigma}_j| + \frac{1}{G} \sum_{k=1}^G \mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k. \quad (13)$$

Note that with noise-free communication links, the optimal (centralized) classifier can be implemented in a distributed fashion by communicating the the local log-likelihood functions computed in each SCR, $\{\mathbf{z}_k^H \tilde{\Sigma}_j^{-1} \mathbf{z}_k, j = 1, \dots, M\}$, to the manager node. The manager node then computes l_j in (13) for $j = 1, \dots, M$ and computes the final decision as in (11).

1) *Performance of Soft Decision Fusion:* We quantify classifier performance in terms of the average probability of error

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

where $P_{e,m}(G)$ is the conditional error probability under H_m . It is instructive to first study the asymptotic ($G \rightarrow \infty$) behavior of l_j 's. Let $E_m[\cdot]$ denote the expectation under H_m . Note from (12) that by the law of large numbers, under H_m , we have

$$\lim_{G \rightarrow \infty} l_j(\mathbf{z}_1, \dots, \mathbf{z}_G) = -E_m[\log p_j(\mathbf{Z})] \\ = D(p_m || p_j) + h_m(\mathbf{Z}) \quad (15)$$

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

$$\begin{aligned} D(p_m || p_j) &= E_m[\log(p_m(\mathbf{Z})/p_j(\mathbf{Z}))] \\ &= \log\left(\left|\tilde{\Sigma}_j \tilde{\Sigma}_m^{-1}\right|\right) + \text{tr}(\tilde{\Sigma}_j^{-1} \tilde{\Sigma}_m - \mathbf{I}) \end{aligned} \quad (16)$$

$$h_m(\mathbf{Z}) = -E_m[\log(p_m(\mathbf{Z}))] = \log((\pi e)^N |\tilde{\Sigma}_m|). \quad (17)$$

An important property of the 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 (15), we conclude that under H_m , l_m will always give the smallest value and, thus, lead to the correct decision as $G \rightarrow \infty$ as long as $D(p_j || p_m) > 0$ for all $j \neq m$. Thus, we expect P_e to go to zero as $G \rightarrow \infty$ if all pairwise K–L distances are strictly positive.

The exact computation of P_e is complicated but we can obtain tight upper bounds. We first use the union bound for $P_{e,m}$

$$P_{e,m}(G) \leq \sum_{j=1, j \neq m}^M P(l_j < l_m | H_m). \quad (18)$$

Each pairwise error probability (PEP) on the right-hand side of (18) depends on a decision statistic that is a weighted sum of $NG\chi_2^2$ random variables [8]. The distribution function of the statistic can be computed in closed-form but takes on tedious expressions [5]. We can obtain tight upperbounds on P_e via Chernoff bounds for the pairwise hypothesis tests. Define the symmetric PEP, $\text{PEP}_{jm}(G)$, for the binary test between H_j and H_m with G measurements, as

$$\text{PEP}_{jm}(G) = \frac{1}{2}[P(l_j < l_m | H_m) + P(l_m < l_j | H_j)]. \quad (19)$$

Then, (14) can be written as

$$P_e(G) \leq \frac{2}{M} \sum_{m=1}^M \sum_{j < m} \text{PEP}_{jm}(G). \quad (20)$$

We state some well-known results (see, e.g., [10]) in the context of our setup.

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

$$\mu_{jm}(\theta) = \log E_m \left[\left(\frac{p_j(\mathbf{Z})}{p_m(\mathbf{Z})} \right)^\theta \right] \leq 0. \quad (21)$$

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

$$\text{PEP}_{jm}(G) \leq \frac{1}{2} e^{\mu_{jm}(\theta)G}. \quad (22)$$

The tightest error exponent (Chernoff exponent) is given by

$$D_{jm}^* = - \min_{0 \leq \theta \leq 1} \mu_{jm}(\theta) \quad (23)$$

and a simpler nontrivial exponent is the Bhattacharya exponent $D_{B,jm}^* = -\mu_{jm}(1/2)$. Thus, for all $G \geq 1$

$$\text{PEP}_{jm}(G) \leq \frac{1}{2} e^{-D_{jm}^* G} \leq \frac{1}{2} e^{-D_{B,jm}^* G}. \quad (24)$$

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

$$\mu_{jm}(\theta) = \theta \log \left| \tilde{\Sigma}_m \tilde{\Sigma}_j^{-1} \right| - \log \left| (1 - \theta)\mathbf{I} + \theta \tilde{\Sigma}_m \tilde{\Sigma}_j^{-1} \right| \quad (25)$$

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 minimum in (23) occurs at θ_{jm}^* , which is the solution to $\mu'_{jm}(\theta) = 0$ (see the Appendix), and can be easily computed numerically using (25). It is also shown in the Appendix that both the Chernoff and Bhattacharya exponents are positive if and only if the corresponding pairwise K–L distances are positive.

Substituting $\text{PEP}_{jm}(G)$ bounds from (24) into (20), we get two corresponding bounds for $P_e(G)$

$$P_e(G) \leq \frac{1}{M} \sum_{m=1}^M \sum_{j < m} e^{-D_{jm}^* G} \leq \frac{1}{M} \sum_{m=1}^M \sum_{j < m} e^{-D_{B,jm}^* G}. \quad (26)$$

Thus, all PEP_{jm} 's decay exponentially to zero with G , and so does $P_e(G)$, if and only if all the pairwise K–L distances are positive. The bound in (26) optimizes only the exponent. As shown in the Appendix, the following tighter asymptotic approximation (not a bound) for P_e can be obtained by estimating the slowing varying function of G that multiplies the exponential in each PEP_{jm}

$$P_e(G) \approx \frac{2}{M} \sum_{m=1}^M \sum_{j < m} \frac{e^{-GD_{jm}^*}}{\theta_{jm}^* \sqrt{2\pi G \mu''_{jm}(\theta_{jm}^*)}} \quad (\text{large } G) \quad (27)$$

where θ_{jm}^* is the minimizing value of θ in (23). The asymptotic decay of $P_e(G)$ will be dominated by the smallest (worst) error exponent. We summarize the results in the following.

Proposition 2: The probability of error, $P_e(G)$, decays exponentially to zero with G if and only if all the pairwise K–L distances are positive

$$D(p_m || p_j) > 0 \quad \forall j, m, \quad j \neq m. \quad (28)$$

A strict bound for $P_e(G)$, applicable for all $G \geq 1$, is given by (26) and a tighter approximation for large G is given by (27). The asymptotic rate of exponential decay in $P_e(G)$ is dominated by the smallest pairwise Chernoff exponent

$$\lim_{G \rightarrow \infty} \frac{\log P_e(G)}{G} = -D_{\min}^* \quad (29)$$

where $D_{\min}^* = \min\{D_{jm}^*\}$.

We note that all pairwise Chernoff exponents would be positive (and bounded) in general except in pathological cases. The size of the exponents increases with the measurement SNR.

B. Hard Decision Fusion

In soft decision fusion, the k th SCR sends M log-likelihood values $\{z_k^T \tilde{\Sigma}_j^{-1} z_k : j = 1, \dots, M\}$, computed from its local measurement z_k , to the manager node. While exchange of real-valued likelihoods puts much less communication burden on the network as compared with *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. One way is to quantize the M likelihood values from each SCR. The number of bits required for accurate communication can be estimated from the differential entropy of the log-likelihoods [9], [8]. Another natural quantization strategy is to compute local *hard decisions* in each SCR based on the local measurement z_k . In this section, we discuss this hard decision fusion approach, assuming noise-free communication links from the SCRs to the manager node.

The locally optimal (based on the local measurement z_k) hard decision in the k th SCR is given by

$$u_k = \arg \max_{j=1,\dots,M} p_j(z_k), \quad k = 1, \dots, G. \quad (30)$$

Since the $\{z_k\}$ are i.i.d., so are $\{u_k\}$. Let $\{p_m[j] : j = 1, \dots, M\}$ denote the M values of the pmf of the hard decision variable U under H_m . The pmf is characterized by the following probabilities

$$p_m[j] = P(U = j | H_m) = P(p_j(z_k) \geq p_l(z_k) \forall l \neq j | H_m) \quad (31)$$

The hard decisions $\{u_k\}$ from all SCRs are communicated via noise-free links to the manager node which makes the final optimal (given $\{u_k\}$) decision as

$$C_{\text{hard}}(u_1, \dots, u_G) = \arg \min_{j=1,\dots,M} l_{j,\text{hard}}[u_1, \dots, u_G] \quad (32)$$

$$\begin{aligned} l_{j,\text{hard}}[u_1, \dots, u_G] &= -\frac{1}{G} \log p_j[u_1, \dots, u_G] \\ &= -\frac{1}{G} \sum_{k=1}^G \log p_j[u_k]. \end{aligned} \quad (33)$$

1) Performance of Hard Decision Fusion: Despite the fact that the local hard decisions can be quite unreliable, the final fused decision becomes arbitrary reliable as $G \rightarrow \infty$. From (33), we note that under H_m

$$\begin{aligned} \lim_{G \rightarrow \infty} l_{j,\text{hard}}[u_1, \dots, u_G] &= -E_m[\log p_j[U]] \\ &= D(p_m \| p_j) + H_m(U) \end{aligned} \quad (34)$$

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]

$$D(p_m \| p_j) = \sum_{i=1}^M p_m[i] \log(p_m[i]/p_j[i]) \quad (35)$$

$$H_m(U) = -\sum_{i=1}^M p_m[i] \log p_m[i]. \quad (36)$$

Thus, we see from (34) that in the limit of large G we will attain perfect classification performance as long as $D(p_m \| p_j) > 0$ for all $j \neq m$.

We can bound the probability of error of hard decision fusion, $P_{e,\text{hard}}(G)$, via PEPs, as in soft decision fusion. In this case, $\mu_{jm,\text{hard}}(\theta)$ is given by

$$\begin{aligned} \mu_{jm,\text{hard}}(\theta) &= \log E_m [p_j^\theta[U] / p_m^\theta[U]] \\ &= \log \sum_{i=1}^M p_j^\theta[i] p_m^{1-\theta}[i] \end{aligned} \quad (37)$$

which can be used to numerically compute the pairwise Chernoff and Bhattacharya exponents: $D_{jm,\text{hard}}^* = -\min_{\theta \in [0,1]} \mu_{jm,\text{hard}}(\theta) = \mu_{jm,\text{hard}}(\theta_{jm}^*)$, where θ_{jm}^* is the solution to $\mu'_{jm,\text{hard}}(\theta) = 0$ (see the Appendix) and $D_{B,jm,\text{hard}}^* = -\mu_{jm,\text{hard}}(1/2)$. Furthermore, using the approach in the Appendix, a tighter asymptotic approximation to the PEPs can be obtained. We summarize the results in the following Proposition.

Proposition 3: The probability of error, $P_{e,\text{hard}}(G)$, decays exponentially to zero with G if and only if all the pairwise K–L distances defined in (35) are strictly positive. A strict bound for $P_{e,\text{hard}}(G)$, applicable for all $G \geq 1$, is given by (26) by using $D_{jm,\text{hard}}^*$ and $D_{B,jm,\text{hard}}^*$. A tighter approximation to $P_{e,\text{hard}}(G)$ for large G is given by (27) by using θ_{jm}^* and $\mu''_{jm,\text{hard}}(\theta_{jm}^*)$. The asymptotic rate of exponential decay in $P_{e,\text{hard}}(G)$ is dominated by the smallest pairwise Chernoff exponent

$$\lim_{G \rightarrow \infty} \frac{\log P_{e,\text{hard}}(G)}{G} = -D_{\min,\text{hard}}^* \quad (38)$$

where $D_{\min,\text{hard}}^* = \min\{D_{jm,\text{hard}}^*\}$.

Note that for a given measurement SNR, the error exponent for hard decision fusion will be smaller compared with 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. In fact, the pmf's of hard decision in (31) are based on decision statistics that are weighted sums of $N\chi_2^2$ random variables, compared with $NG\chi_2^2$ in soft decision fusion, thereby resulting in less reliable hard decisions [8].

IV. DECISION FUSION WITH NOISY COMMUNICATION LINKS

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

We assume that each SCR has a dedicated communication link to the manager node.⁶ Each SCR sends an amplified version of its hard decision u_k in (30) over a noisy link

$$y_k = \alpha u_k + w_k, \quad k = 1, \dots, G \quad (39)$$

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 $\tilde{M} = (M - 1)/2$.

⁶Note that this requires large bandwidth or large latency at the manager node in the limit of large G .

We assume that each SCR sends a symmetrized version of its hard decision to use minimum power: $u_k \in \{-\tilde{M}, \dots, \tilde{M}\}$.⁷ Given this simple communication scheme, the optimal classifier at the manager node takes the form

$$C_{\text{noisy}}(y_1, \dots, y_G) = \arg \min_j l_{j,\text{noisy}}(y_1, \dots, y_G) \quad (40)$$

$$l_{j,\text{noisy}}(y_1, \dots, y_G) = -\frac{1}{G} \sum_{k=1}^G \log p_{j,\text{noisy}}(y_k) \quad (41)$$

$$p_{j,\text{noisy}}(y) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \sum_{i=-\tilde{M}}^{\tilde{M}} e^{-\frac{(y-\alpha_i)^2}{2\sigma_w^2}} p_j[i]. \quad (42)$$

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

$$\mu_{jm,\text{noisy}}(\theta) = \log E_m [p_{j,\text{noisy}}^\theta(Y) / p_{m,\text{noisy}}^\theta(Y)] \quad (43)$$

which can be computed numerically. Most importantly, we again expect exponentially vanishing $P_{e,\text{noisy}}$ since from (42) we have under H_m

$$\begin{aligned} \lim_{G \rightarrow \infty} l_{j,\text{noisy}}(y_1, \dots, y_G) &= -E_m[\log p_{j,\text{noisy}}(Y)] \\ &= D(p_{m,\text{noisy}} \| p_{j,\text{noisy}}) + h_m(Y) \end{aligned} \quad (44)$$

where $h_m(Y) = -E_m[\log(p_m(Y))]$ denotes the differential entropy of the received signal at the manager node from each SCR. Thus, parallel to the previous two cases, we immediately have the following result.

Proposition 4: $P_{e,\text{noisy}}(G)$ decays exponentially to zero with G if and only if all the pairwise K-L distances between noisy pdf's in (42) are strictly positive. A strict bound for $P_{e,\text{noisy}}(G)$, applicable for all $G \geq 1$, is given by (26) by using $D_{jm,\text{noisy}}^*$ and $D_{B,jm,\text{noisy}}^*$. A tighter approximation to $P_{e,\text{hard}}(G)$ for large G is given by (27) by using θ_{jm}^* and $\mu_{jm,\text{noisy}}^*(\theta_{jm}^*)$. The asymptotic rate of exponential decay in $P_{e,\text{noisy}}(G)$ is dominated by the smallest pairwise Chernoff exponent

$$\lim_{G \rightarrow \infty} \frac{\log P_{e,\text{noisy}}}{G} \leq -D_{\min,\text{noisy}}^* \quad (45)$$

where $D_{\min,\text{noisy}}^* = \min\{D_{jm,\text{noisy}}^*\}$.

The K-L distances and the exponents in this case will be smaller than those for noise-free hard decision fusion. In particular, they depend on both the measurement and the communication SNRs.

V. NUMERICAL RESULTS BASED ON REAL DATA

In this section, we present numerical results to illustrate the performance of the three classifiers C, C_{hard} and C_{noisy} as a

⁷Note that this is not necessarily the optimal symbol assignment from the viewpoint of final decision.

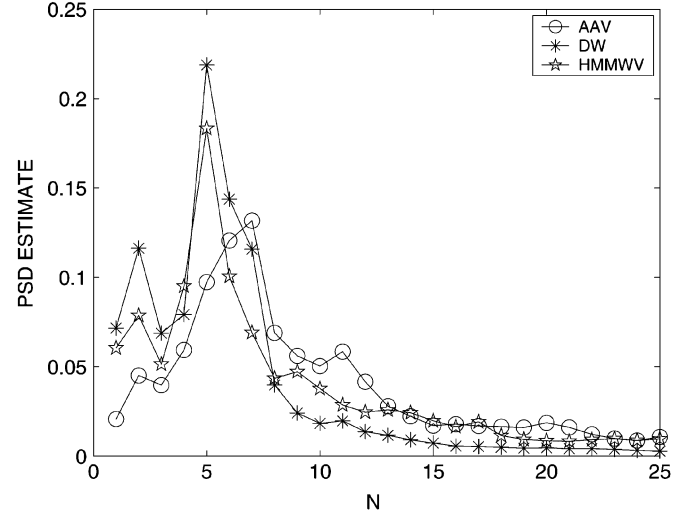


Fig. 3. Covariance matrix eigenvalues (PSD estimates) based on acoustic measurements for three vehicles: AAV, DW, and HMMWV.

function of G for different measurement and communication SNRs which are defined (in decibels) as

$$\begin{aligned} \text{SNR}_{\text{meas}} &= 10 \log_{10} (\text{tr}(\Sigma_j) / N \sigma_n^2) \\ \text{SNR}_{\text{comm}} &= 10 \log_{10} \left(\frac{\sum_i (\alpha_i)^2 p[i]}{\sigma_w^2} \right) \end{aligned} \quad (46)$$

where $p[i] = P(U = i) = \sum_m p_m[i] / M$ and recall that we assume that $\text{tr}(\Sigma_j)$ (signal energy) is the same for all classes. Our results are based on real acoustic measurements collected during the DARPA SensIT program and correspond to classifying a vehicle from $M = 3$ classes [3]. The three vehicle classes are: Amphibious Assault Vehicle (AAV), Dragon Wagon (DW) and Humvee (HMMWV). Fig. 3 plots the PSD values at $N = 25$ frequencies (within a 2-kHz bandwidth) that were estimated from data collected at multiple nodes. The PSD values define the diagonal covariance matrices (in the Fourier domain) $\Sigma_j = \Lambda_j$. Under H_j , the N -dimensional averaged measurement vector z_k in the k th SCR was simulated as

$$z_k = \Lambda_j^{1/2} \mathbf{v}_k + \mathbf{n}_k, \quad k = 1, \dots, G \quad (47)$$

where $\{\mathbf{v}_k\}$ are i.i.d $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ and $\{\mathbf{n}_k\}$ are i.i.d. $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$. P_e and $P_{e,\text{hard}}$ 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 (39) and the pmf's for hard decisions. $P_{e,\text{noisy}}$ was estimated using 10 000 independent sets of measurement realizations. The simulations were done for four values for $\text{SNR}_{\text{meas}} = -4, 0, 4, 10$ dB and three values for $\text{SNR}_{\text{comm}} = 0, 5, 10$ dB.

Table I shows the values of all pairwise K-L distances for soft, hard, and noisy hard decision 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} .

TABLE I
PAIRWISE K-L DISTANCES FOR THE THREE VEHICLES CLASSES FOR SOFT, HARD, AND NOISY HARD DECISION FUSION AT FOUR DIFFERENT SNR_{meas} . THE FIRST ENTRY IN EACH CELL IS FOR SOFT DECISIONS, THE SECOND FOR HARD DECISIONS, AND THE REMAINING THREE FOR NOISY HARD DECISIONS AT $\text{SNR}_{\text{comm}} = 0, 5, 10$ dB, RESPECTIVELY. (a) $\text{SNR}_{\text{meas}} = -4$ dB. (b) $\text{SNR}_{\text{meas}} = 0$ dB. (c) $\text{SNR}_{\text{meas}} = 4$ dB. (d) $\text{SNR}_{\text{meas}} = 10$ 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)

Fig. 5 plots the simulated P_e , $P_{e,\text{hard}}$, and $P_{e,\text{noisy}}$ for the four values of SNR_{meas} . The three plots for $P_{e,\text{noisy}}$ correspond to the different 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 decreases with SNR_{comm} and is fairly small at $\text{SNR}_{\text{comm}} = 10$ dB. Note that even at $\text{SNR}_{\text{meas}} = 0$ dB, $P_e \approx 10^{-2}$ for soft decision fusion is attained with only $G \approx 20$ independent measurements. Furthermore, the same performance can be attained with the much simpler hard decision fusion (both ideal and noisy) around $G = 40$. Note that the P_e 's are around 0.2 for all three classifiers for a single node measurement ($G = 1$). At $\text{SNR}_{\text{meas}} = 10$ dB, which could be attained by averaging over $n_G = 10$ measurements at $\text{SNR}_{\text{meas}} = 0$ dB within each SCR, only $G \approx 8$ independent measurements are needed to attain $P_{e,\text{noisy}} \approx 10^{-2}$ with noisy hard decision fusion ($P_{e,\text{noisy}} \approx 0.2$ for $G = 1$). This demonstrates an important practical advantage of multiple independent measurements in sensor networks: we can attain reliable classification performance ($P_e \approx 10^{-2}$) by combining a moderate number (10–40) of fairly unreliable ($P_e \approx 0.2$ – 0.4) independent node decisions, and this can be achieved

with simple communication schemes, as the one in noisy hard decision fusion.

Finally, in Fig. 4, we compare simulated P_e and $P_{e,\text{hard}}$ for noise-free soft and hard decision fusion with analytical bounds [see (26)] based on the Chernoff PEP bounds. As evident, the bounds perfectly match the slope (exponent) of the simulated P_e curves but exhibit an offset since they do not optimize the constants multiplying the exponentials. The tighter asymptotic approximations (27), which optimize the constants, are also plotted and are remarkably accurate even for the moderate values of G shown. We note both the bounds and approximations are tighter for soft decision fusion and at higher SNR_{meas} .

VI. DISCUSSION AND CONCLUSION

All applications of sensor networks are built on two primary operations: 1) distributed processing of data collected by the nodes and 2) communication of processed data from one part of the network to another. Furthermore, the two operations 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 limited by network communication constraints. In this paper, we have investigated

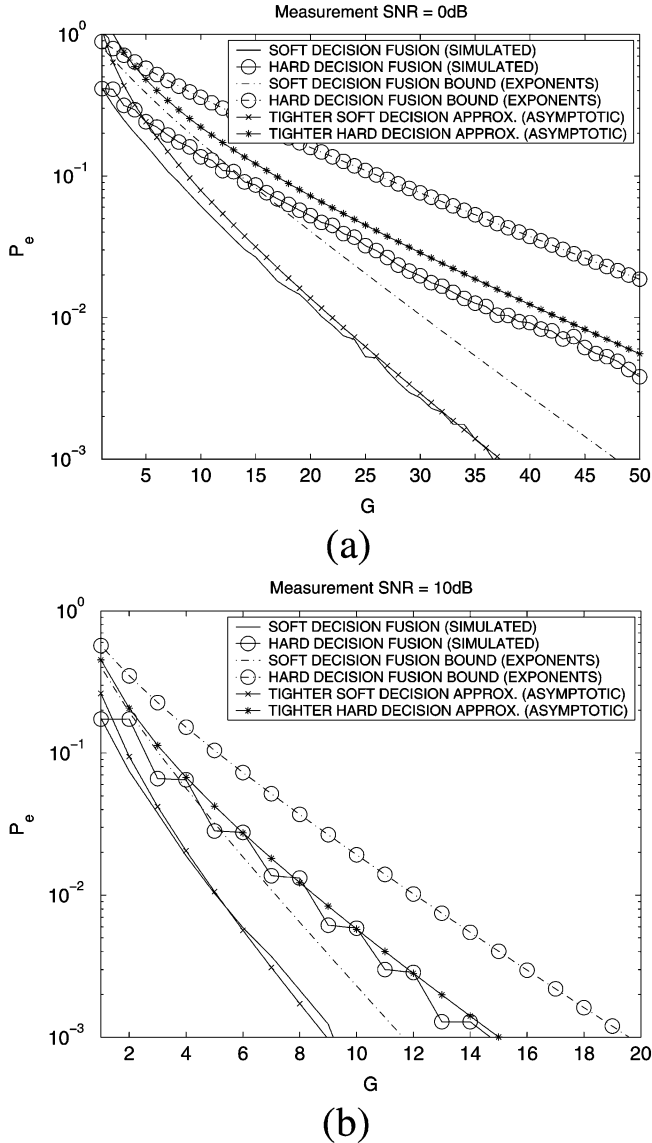


Fig. 4. Comparison of simulated P_e with analytical P_e (Chernoff) bounds (26), as well as the tighter asymptotic approximation (27) for noiseless hard and soft decision fusion. (a) $\text{SNR}_{\text{meas}} = 0 \text{ dB}$. (b) $\text{SNR}_{\text{meas}} = 10 \text{ dB}$.

distributed decision making in a simple context—classification of a single object—to study the basic principles that govern the interplay between information sensing, processing, and communication, and how it impacts the performance of distributed classification algorithms. Our approach is based on modeling the object signal as a band-limited ergodic Gaussian field in space and time. The model suggests a natural partitioning of the network (see Fig. 1) into disjoint SCRs whose size is inversely proportional to the spatial signal bandwidths—the spatial signal is strongly correlated with each SCR and approximately independent across SCRs.

From a communication perspective, this network partitioning into independent SCRs leads to a structure on information exchange that is naturally suited to network constraints: high-bandwidth feature-level *data-fusion* is limited to spatially local nodes within each SCR, whereas global fusion of

low-bandwidth *local decisions* at each SCR is sufficient across different SCRs.

From the viewpoint of classification performance, both local data fusion (of n_G measurements within each SCR) and global decision fusion (of G independent decisions from different SCRs) improve P_e . Specifically, decision fusion in all cases (ideal hard/soft, or noisy hard) leads to exponential decay in P_e with G , whereas data fusion determines the *rate* of decay in P_e (the pairwise error exponents). For given signal statistics, the error exponents increase with n_G (improvement in SNR_{meas}), as well as the dimension (N) of the feature vector N . In essence, local data fusion improves the reliability of local SCR decisions, whereas global decision fusion improves the reliability of the final decision. Decision fusion over noisy communication links incurs a loss in error exponents, which can be reduced by using coded communication. Thus, for given class signal statistics, SNR_{meas} and SNR_{comm} , the parameters N , G , and n_G can be appropriately chosen to attain a desired level of performance. In particular, our numerical results underscore the remarkable impact of decision fusion: a moderate (10–50) number of fairly unreliable ($P_e \approx 0.2$ – 0.4) local decisions can be combined to yield acceptably reliable ($P_e \approx 10^{-2}$) final decisions. A related observation is that rapidly varying spatial signals (with larger bandwidths) require cooperation between nodes in a smaller region (size of SCRs is small) to yield a sufficient number of independent measurements. Furthermore, multiple independent measurements could also be collected at each node over time to further improve performance at the cost of latency.

From the viewpoint of sensing, we believe that the simple SCR-based stationary modeling reveals fundamental insights that are applicable to more general settings. Essentially, any spatial field possesses certain number of independent DoF (G). However, these DoF would in general be nonidentical in nonstationary spatial fields: different SNRs and coherence scales (SCRs) associated with each DoF. Furthermore, the DoFs may exhibit non-Gaussian statistics (as in boundaries). Network nodes sample the spatial field and n_G reflects the over-sampling per DoF, which should in general be adapted to the coherence scale and SNR associated with each DoF. Adaptive sampling techniques (see, e.g., [11]) may be used in such contexts to determine the DoFs and the associated coherence scales (partitioning of the network into G nonuniform regions). Qualitatively, we expect G to be the dominant factor in driving the P_e down to zero, whereas the number of nodes in each region would impact the *rate* of decay of P_e . We note that this notion of coherent averaging in homogeneous regions also plays a key role in driving the mean-squared-error down in the context of signal field estimation [4], [11].

Finally, note that we assumed independent channels from different SCRs to the manager node. For large number of SCRs, 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 SCRs 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.

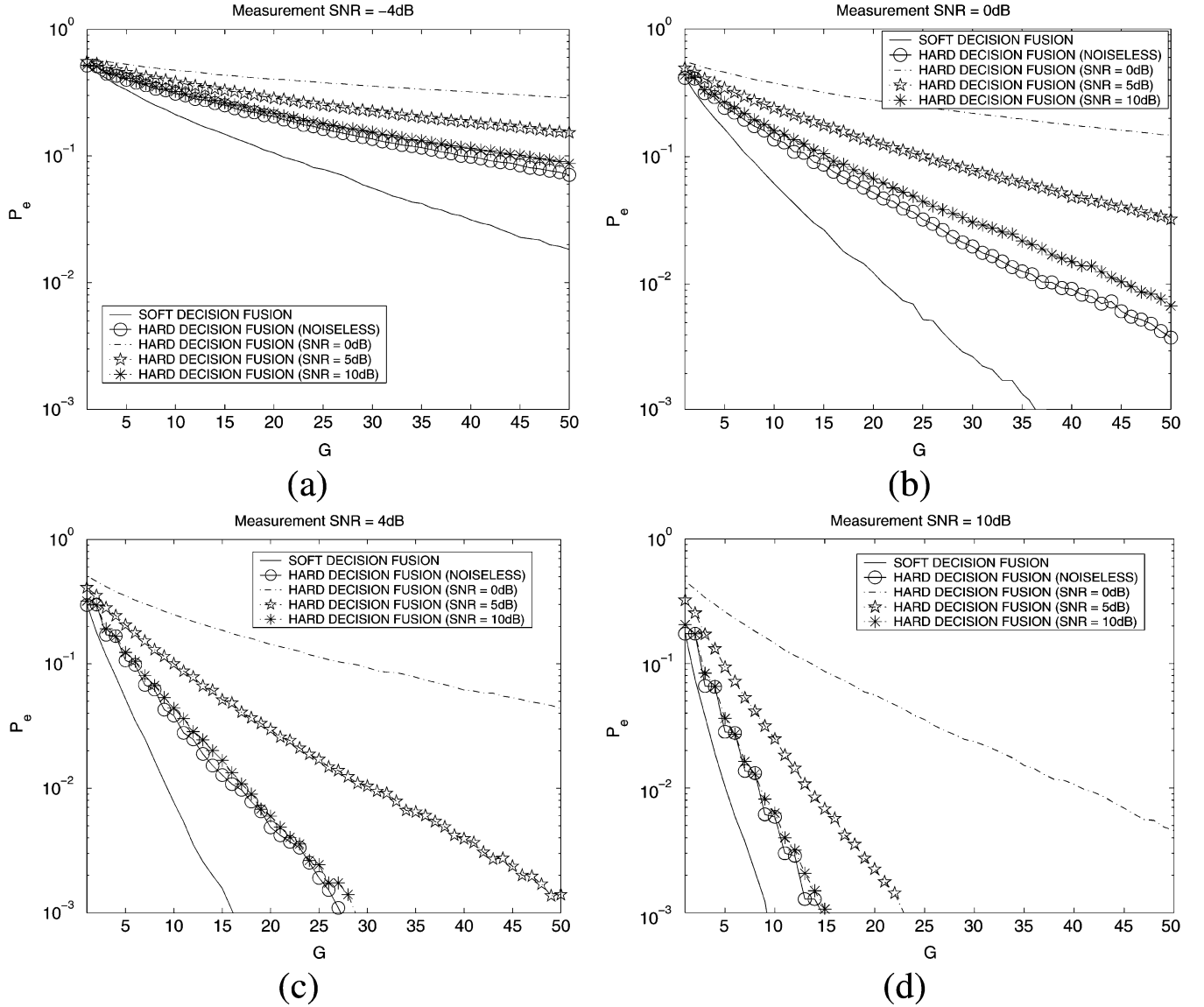


Fig. 5. P_e as a function of G for the three classifiers for different values of SNR_{meas} . Three plots are shown for noisy hard decision fusion at $\text{SNR}_{\text{comm}} = 0, 5, 10$ dB. (a) $\text{SNR}_{\text{meas}} = -4$ dB. (b) $\text{SNR}_{\text{meas}} = 0$ dB. (c) $\text{SNR}_{\text{meas}} = 4$ dB. (d) $\text{SNR}_{\text{meas}} = 10$ dB.

APPENDIX

We determine the conditions for each PEP error exponent to be positive and also provide a tighter asymptotic approximation for the PEP based on an accurate estimate of the constant multiplying the exponential in the PEP. Let H_1 and H_2 be the two hypotheses in the PEP and let X_1, \dots, X_G denote the G i.i.d. measurements. X represents the N -dimensional vector Z in soft decision fusion, U in noise-free hard decision fusion, and Y in noisy hard decision fusion. From Proposition 1, $\text{PEP}(G) \leq (1/2)e^{G\mu(\theta)}$ where

$$\mu(\theta) = \log[g(\theta)] \quad g(\theta) = \mathbb{E}_1[(p_2(X)/p_1(X))^\theta]. \quad (48)$$

Our calculations involve $\mu(\theta)$ and its three derivatives given by

$$\mu'(\theta) = g'(\theta)/g(\theta) \quad (49)$$

$$\mu''(\theta) = g''(\theta)/g(\theta) - (\mu'(\theta))^2 \quad (50)$$

$$\mu'''(\theta) = \frac{g(\theta)g'''(\theta) - g''(\theta)g'(\theta)}{g^2(\theta)} - 2\mu'(\theta)\mu''(\theta) \quad (51)$$

where the n th derivative of g is given by $g^{(n)}(\theta) = \mathbb{E}_1[(p_2(X)/p_1(X))^\theta (\log[p_2(X)/p_1(X)])^n]$. First note that $\mu(0) = \mu(1) = 0$. It is well known that $\mu(\theta)$ is convex and, thus, $\mu(\theta) \leq 0$ for $0 \leq \theta \leq 1$. The Chernoff exponent is $D^* = -\min_\theta \mu(\theta) = -\mu(\theta^*) \geq 0$, where θ^* is the solution to $\mu'(\theta) = 0$. It readily follows from (49) that $\mu'(0) = -D(p_1 \| p_2) \leq 0$ and $\mu'(1) = D(p_2 \| p_1) \geq 0$. Since the two K-L distances are positive if and only if $p_1 \neq p_2$, it follows that $\theta^* \in (0, 1)$ and $D^* = -\mu(\theta^*) > 0$ if and only if the K-L distances are positive.

The Chernoff bound $\text{PEP} \leq (1/2)e^{-GD^*}$ in Proposition 1 only optimizes the exponent. A tighter asymptotic approximation for the PEP, of the form $\text{PEP} \approx c(G)e^{-GD^*}$, can be obtained by accurately estimating $c(G)$. Using the bounding techniques in [12, eq. 5.4.23 and Appendix 5-A], it can be

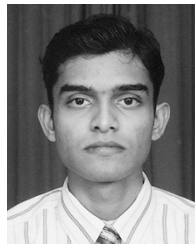
shown that $c(G) \approx O(1/\sqrt{G})$. Specifically, the PEP for large G can be approximated as

$$\text{PEP} \approx \left[\frac{1}{\theta^* \sqrt{2\pi G \mu''(\theta^*)}} + o\left(\frac{1}{\sqrt{G}}\right) \right] e^{-GD^*} \quad (52)$$

provided that $\mu(\theta) < \infty$ in a neighborhood of $\theta = 0$ and $\mu'''(\theta^*)$ is finite, which is true in our Gaussian formulation barring the pathological cases in which some pairwise K-L distances are zero or unbounded.

REFERENCES

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing 2001*, vol. 4, 2001, pp. 2033–2036.
- [2] S. Kumar, F. Zhao, and D. Shepherd, "Special issue on collaborative signal and information processing in microsensor networks," in *IEEE Signal Processing Mag.*, Mar. 2002, pp. 13–14.
- [3] D. Li, K. Wong, Y. Hu, and A. Sayeed, "Detection, classification, tracking of targets in micro-sensor networks," in *IEEE Signal Processing Mag.*, Mar. 2002, pp. 17–29.
- [4] A. Sayeed. A statistical signal modeling framework for sensor networks. UW–Madison Tech. Rep. ECE-04-1. [Online]. Available: <http://dune.ece.wisc.edu>
- [5] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [6] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [7] R. M. Gray, "On the asymptotic eigenvalue distribution of toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-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, Proc. IPSN*, F. Zhao and L. Guibas, Eds. Berlin, Germany, Apr. 2003, pp. 193–208.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [10] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [11] R. Nowak and U. Mitra, "Boundary estimation in sensor networks: Theory and methods," in *Lecture Notes in Computer Science, Proc. IPSN*, F. Zhao and L. Guibas, Eds. Berlin, Heidelberg, Apr. 2003, pp. 80–95.
- [12] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.



Ashwin D'Costa (S'03) received the B.E. degree in electronics and communication engineering from the National Institute of Technology, Trichy, India, in 1999. He is currently working toward the M.S. degree in electrical and computer engineering at the University of Wisconsin–Madison.

Since August 2001, he has been a Research Assistant in the Sensor Networks Research Group, University of Wisconsin–Madison. His research interests are in sensor networks, wireless communication systems, real-time computing systems, and

signal processing applications.



Vinod Ramachandran received the B.Tech. degree in instrumentation engineering from Anna University, Chennai, India, in 2002. He is currently working toward M.S. degree in electrical engineering at the University of Wisconsin–Madison.

His research interests include signal processing for sensor networks and wireless communications.

Mr. Ramachandran received the Donald and Esther Procknow Distinguished Graduate Fellowship in 2002.



Akbar M. Sayeed (S'89–M'97–SM'02) received the B.S. degree from the University of Wisconsin–Madison, in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, in 1993 and 1996, respectively, all in electrical and computer engineering.

While at the University of Illinois, he was a Research Assistant in the Coordinated Science Laboratory and was also the Schlumberger Fellow in signal processing from 1992 to 1995. From 1996 to 1997, he was a Postdoctoral Fellow at Rice

University, Houston, TX. Since August 1997, he has been with the University of Wisconsin–Madison, where he is currently an Associate Professor in Electrical and Computer Engineering. His research interests are in wireless communications, communication and information theory, statistical signal processing, and time-frequency analysis.

Dr. Sayeed received the National Science Foundation (NSF) CAREER Award in 1999 and the Office of Naval Research (ONR) Young Investigator Award in 2001.

He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002, and is currently co-guest editing a special issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Self-Organizing Distributed Collaborative Sensor Networks.