A Statistical Signal Modeling Framework for Sensor Networks

Akbar M. Sayeed[*]

University of Wisconsin–Madison
akbar@engr.wisc.edu, http://dune.ece.wisc.edu

1 Introduction

Wireless sensor networks are an emerging technology that promise an unprecedented opportunity to monitor the physical world via wireless nodes that can sense the environment in various modalities, such as acoustic, seismic, infra-red [1, 2, 3, 4]. Typically each node can sense in multiple modalities but has limited computational and communication abilities due to battery powered operation. 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. While recent years have seen a surge of research activity in sensor networks, many significant challenges need to be overcome to realize the vision of sensor networks.

The key challenges are tied to two vital operations in a sensor network: 1) efficient information routing between network nodes, and 2) collaborative signal processing (CSP) between nodes to extract useful information from the data collected by the sensors. Exchange of sensor information between nodes in the region of activity is necessary due to a variety of reasons, including limited (local) information gathered by each node, variability in operating conditions, and node failure. From a communication and networking viewpoint, sensor networks are similar to ad hoc multi-hop wireless networks. However, there is a key distinction: *the information flow in a sensor network is fundamentally governed by the activity in the physical environment sensed by the nodes.* Furthermore, in view of the limited communication and computational capability of nodes, an overarching objective in the design of sensor networks is to exchange the least amount of information between nodes to enable desired information processing. On the one hand, statistical signal characteristics in a region of space directly impact the extent and nature of information exchange between nodes for CSP. On the other hand, the exchange of information between wireless nodes is constrained by the network configuration and the limited node resources. Consequently, the coupling between the networking and physical/sensing layers is much more direct in sensor networks than in general ad hoc networks. This motivates a central question at the heart of integrated design of sensor networks:

[*]This research was supported in part by NSF under grants CCR-9875805 and CCR-0113385, ONR under grant N00014-01-1-0825, and DARPA SensIT program under grant F30602-00-2-0555.

What are the key principles governing the interplay between information processing and information routing in wireless sensor networks?

In view of the wide variety of possible signal sources, judicious abstractions of signal statistics in space and time are necessary to address this question. We propose a statistical signal modeling framework based on the notion of *spatial coherence regions (SCR's)* that captures the salient second-order statistical characteristics of a wide variety of signals in space and time. The proposed signaling modeling framework enables us to address two key design facets of sensor networks:

1. Fundamental issues related to distributed signal processing and communication, including space-time sampling, signal inference and information flow requirements associated with each query. It also provides a general framework for assessing the performance and network cost of CSP algorithms.

2. Fundamental understanding of the interplay between information sensing, processing, communication and routing. It suggests a natural hierarchical structure for information exchange between nodes for enabling practical CSP algorithms and associated communication and routing strategies. In particular, it suggests a *fixed matched interface* between information processing and routing, with high bandwidth information exchange naturally restricted to local spatial regions.

A sensor network is typically partitioned into regions to facilitate distributed signal processing and information routing. In particular, a *location-centric* routing paradigm has been recently proposed [5, 6] that is inspired by the important role of signal sources in sensor networks: *network queries involve particular network regions where the relevant signal sources are located.* This fundamental characteristic of information flow is reflected in virtually all state-of-the-art routing techniques for sensor networks (see, e.g., [1, 5]). The proposed space-time signaling modeling framework naturally complements the location-centric information routing paradigm.

The next section provides a brief description of location-centric routing. The proposed signal modeling framework is presented in Section 3. Section 4 discusses the implications of the signal modeling framework in the design of sensor networks. Concluding remarks are presented in Section 5.

2 Region-Based Information Processing and Routing

Given the limited communication and computational abilities of sensor nodes, a critical consideration in the design off sensor networks is that most of information exchange and associated CSP must take place at a local level. The location-centric or region-based routing framework is precisely tailored to meet the needs of CSP

in sensor networks [5, 6]. The approach is based on the observation that CSP typically requires collaboration among nodes in a certain region and not among an arbitrarily specified set of nodes. For example, application queries regarding the concentration profile of a certain bio-chemical agent in a given area, regarding the temperature or pressure variation in a given area, or regarding unauthorized entries into a given area, all require collaboration among sensor nodes in the region of interest. Consequently, in this approach, geographic regions play the role of a node in the traditional network interface. In particular, the nodes are not individually addressable. Instead, an application first creates entities called regions, which are then addressable.

In the location-centric paradigm, the network is divided into regions for facilitating distributed information routing and processing (see Fig. 1). A region typically represents a rectangular geographic area. We assume that each node is aware of its geographic location, such as through GPS sensors, or through other location estimation schemes [7, 8, 9]. In each region, a subset of nodes is designated manager sub-region to coordinate the information exchange and processing within the region. For example, tracking of a moving object generally involves the following steps [10]:

1. **Initial region creation.** Regions are first created at potential target entry areas. These regions are then tasked to detect an incoming target. For example, Region 1 in Fig. 1.

2. **Target detection and classification.** Target detection in a region involves CSP among the nodes in the region. For example, outputs of energy detectors at each node may be communicated to the manager nodes, who in turn use a robust fusion algorithm to arrive at a consensus decision for the region. Each node may also send feature vectors to the manager nodes for target classification.

3. **Target localization and location prediction.** Target detection information (for example, the time of closest point of approach and energy detector outputs) from different nodes is used by the manager nodes to estimate the location of the target. Location estimates over a period of time are used by the manager nodes to predict target location at future time instants.

5. **Creation of new potential regions.** When the target gets close to exiting the current region, the estimates of predicted target location are used to create one or more new regions. The new regions are put on alert for target detection. When a target is detected by new region, the above steps are repeated. For example, in Fig. 1, Region 1 creates Region 2. Since Region 2 has two possible predictions for future position of the target, it creates Regions 3a and 3b.

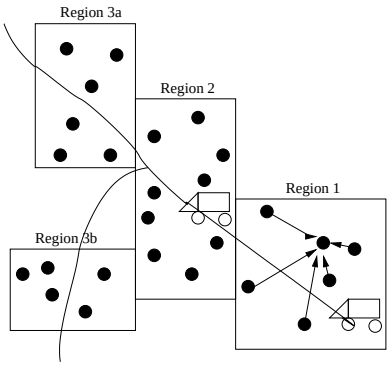

Figure 1: *Location-centric approach for tracking an object moving through the sensor field.*

The regions in location-centric routing facilitate local CSP. As we will see, the space-time signal statistics associated with a query directly impact the size of the regions and the nature of information exchange between nodes. The signal modeling framework discussed in the next section is precisely aimed at investigating this interplay between information routing and processing.

3 A Signal Model for Sensor Measurements

3.1 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, $\mathcal{R} = D_x \times D_y = [-D_x/2, D_x/2] \times [-D_y/2, D_y/2]$ associated with a network query involving 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.[1] 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

$$s(x, y, t) = \int_{-B_x/2}^{B_x/2} \int_{B_y/2}^{B_y/2} \int_{-B/2}^{B/2} \phi_s(\nu_x, \nu_y, f) e^{j2\pi f_x x} e^{j2\pi f_y y} e^{j2\pi f t} df_x df_y df \tag{1}$$

where $\phi(f_x, f_y, f)$ denotes the underlying spectral representation[2] which satisfies

$$\mathrm{E}[\phi_s(f_x, f_y, f)\phi_s^*(f_x', f_y', f')] = \Phi_s(f_x, f_y, f)\delta(f_x - f_x')\delta(f_y - f_y')\delta(f - f') \tag{2}$$

[1] 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.

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

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

$$\begin{aligned} r_s(\Delta x, \Delta y, \Delta t) &= \mathrm{E}[s(x+\Delta x, y+\Delta y, t+\Delta t)s^*(x,y,t)] \\ &= \int_{-B_x/2}^{B_x/2}\int_{B_y/2}^{B_y/2}\int_{-B/2}^{B/2} \Phi_s(f_x, f_y, f)e^{j2\pi(f_x\Delta x + f_y\Delta y + f\Delta t)}df_x df_y df \end{aligned} \quad (3)$$

and both characterize the statistics of $s(x, y, t)$. In the above equations, B_x and B_y represent the spatial signal bandwidths and B the temporal bandwidth. We assume that B_x, B_y and B are known *a priori* for all signal sources of interest. This is a reasonable assumption about very basic characteristics of $s(x, y, t)$. The power in the signal field is defined as

$$\sigma_s^2 = \mathrm{E}[|s(x,y,t)|^2] = r_s(0,0,0) = \int_{B_x}\int_{B_y}\int_{B} \Phi_s(f_x, f_y, f)df_x df_y df. \quad (4)$$

3.2 Approximate Signal Modeling Via Spatial Coherence Regions

In order to study the implications of signal statistics for sensor network operation, we propose an approximate signal model, based on *spatial coherence regions (SCR's)* illustrated in Fig. 2, that captures the *scales of signal variation* in the spatial coordinates. 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 dimension. Similarly, $D_{c,y} = 1/B_y$ denotes the coherence distance in the y dimension. Thus, as illustrated in Fig. 2, we partition the query region $\mathcal{R}$ into disjoint SCR's, $\{\mathcal{R}_{m,n}\}$, of size $D_{c,x} \times D_{c,y}$ over which the signal remains strongly correlated. On the other hand, as we will see, the signal is approximately uncorrelated in distinct SCR's. The uniform size of SCR's follows from the stationarity assumption. Furthermore, the size of each SCR decreases as the spatial bandwidths increase.

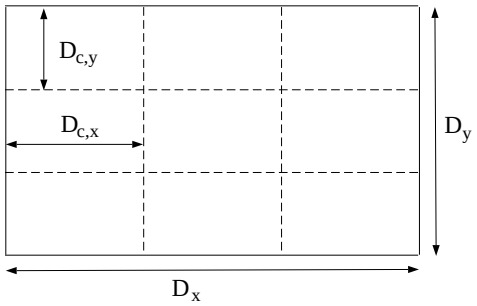

Figure 2: *A schematic illustrating the notion of* **spatial coherence regions (SCR's)** *over which $s(x, y, t)$ remains strongly correlated (approximately constant) as a function of (x, y). The region of interest, $\mathcal{R}$, of size $D_x \times D_y$ is partitioned into SCR's, $\{\mathcal{R}_{m,n}\}$, 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.*

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

$$s_{pwc}(x,y,t) = \sum_{m,n} s_{m,n}(t) I_{\mathcal{R}_{m,n}}(x,y) = \sum_{m=-\tilde{N}_x}^{\tilde{N}_x} \sum_{n=-\tilde{N}_y}^{\tilde{N}_y} s_{m,n}(t) I_{D_{c,x}}(x - mD_{c,x}) I_{D_{c,y}}(y - nD_{c,y}) \quad (5)$$

where $I_X(x)$ denotes the indicator function of the set X, $N_x = D_x/D_{c,x} = 2\tilde{N}_x + 1$, 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 basis functions $\{u_{m,n}(x,y) = I_{D_{c,x}}(x - mD_{c,x}) I_{D_{c,y}}(y - nD_{c,y})\}$. For any t, the PWC model coefficients can be computed as

$$s_{m,n}(t) = \frac{1}{D_{c,x}D_{c,y}} \int_{\mathcal{R}_{m,n}} s(x,y,t)dxdy = \frac{1}{D_{c,x}D_{c,y}} \int_{(m-1/2)D_{c,x}}^{(m+1/2)D_{c,x}} \int_{(n-1/2)D_{c,y}}^{(n+1/2)D_{c,y}} s(x,y,t)dxdy. \quad (6)$$

Note that the *distance-bandwidth (DB) products* $N_x = D_x B_x$ and $N_y = D_y B_y$ are exactly analogous to the time-bandwidth product TB of the space of signals of duration T and bandwidth B [11].

The mean-squared error (MSE) of the PWC approximation in any SCR is independent of t and the particular SCR (due to stationarity) and is given by

$$\begin{aligned} \text{mse}_{scr} &= \int_{-D_{c,x}/2}^{D_{c,x}/2} \int_{-D_{c,y}/2}^{D_{c,y/2}} \text{E}[|s(x,y,t) - s_{pwc}(x,y,t)|^2]dxdy \\ &= \frac{1}{B_x B_y} \int_{-B_x/2}^{B_x/2} \int_{-B_y/2}^{B_y/2} \Phi_s(f_x,f_y)[1 - \text{sinc}^2(f_x/B_x)\text{sinc}^2(f_y/B_y)]df_x df_y \end{aligned} \quad (7)$$

where $\Phi_s(f_x,f_y) = \int \Phi_s(f_x,f_y,f)df$ is the joint spatial PSD of the signal field. The MSE is smaller for smoother signal fields which have most of the power concentrated in the lower frequencies.

Remark: In developing the implications of the PWC approximation to sensor networks, we will assume that the component processes $s_{m,n}(t)$ are *perfectly uncorrelated* across distinct SCR's. This assumption will be justified by our subsequent analysis which shows that the impact of residual correlation between SCR's is relatively insignificant in a variety of aspects. By far the most dominant factor is the number of SCR's, $N_s = N_x N_y = (D_x B_x)(D_y B_y)$, in a given region that determines the number of independent spatial degrees of freedom in $s(x,y,t)$ and $s_{pwc}(t)$.

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 purely temporal source signal $s_o(t)$ – the space-time signal is determined by $s_o(t)$ via physical signal propagation in space. Examples of such signals include acoustic signals as well as seismic (vibrational) signals

produced by vehicles. The spatial signal bandwidth is determined by the temporal bandwidth via the speed of signal propagation. 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 constant on concentric circles around the source and is thus is stationary along radial lines. It is easy to verify that that $B_r = B/v$, where B_r is the spatial bandwidth in the radial dimension, and B is the bandwidth of $s_o(t)$. The spatial coherence regions are concentric bands around the source and the radial coherence distance $D_{c,r}$ is given by $D_{c,r} = 1/B_r = v/B = vT_c$. For example, for an acoustic source with $B = 500$Hz, $D_r = 0.66\mathrm{m}$ whereas for $B = 20$Hz, $D_r = 17\mathrm{m}$.

Distributed space-time sources. Such sources are truly distributed in space and time, such as the time-varying concentration of a chemical agent in space, or signals generated by a distributed array of temporal point sources or actuators. The temporal and spatial bandwidths of such distributed sources may be completely independent. Consequently such sources may result in space-time signals that have sharp variations in space and time. We will only consider smooth sources. Even if the space-time signals exhibit sharp variations, the limited bandwidths of the sensors as well as the spacing between them may result in a smooth effective signal that is sensed by the network.

3.3 Statistics of the PWC Approximation

We need to characterize statistics of $s_{pwc}(x,y,t)$ to study the implications of PWC modeling. That is, we need to characterize the second-order statistics of the component processes $\{s_{mn}(t)\}$. We first define an intermediate *spatially smoothed* field $z(x,y,t)$ as

$$z(x,y,t) = \frac{1}{D_{c,x}D_{c,y}} \int\int s(x',y',t) I_{D_{c,x}}(x'-x) I_{D_{c,y}}(y'-y) dx'dy'. \tag{8}$$

It follows from (6) that $\{s_{m,n}(t)\}$ are obtained via sampling: $s_{m,n}(t) = z(mD_{c,x}, nD_{c,y}, t)$. Note from (8) that $z(x,y,t)$ is also a stationary Gaussian field and its PSD is given by

$$\Phi_z(f_x,f_y,f) = \Phi_s(f_x,f_y,f)\mathrm{sinc}^2(D_{c,x}f_x)\mathrm{sinc}^2(D_{c,y}f_y) = \Phi_s(f_x,f_y,f)\mathrm{sinc}^2(f_x/B_x)\mathrm{sinc}^2(f_y/B_y) \tag{9}$$

where $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$. The correlation function of the processes $\{s_{m,n}(t)\}$ is then given by

$$\begin{aligned} r_s[\Delta m, \Delta n, \Delta t] &= \mathrm{E}[s_{m+\Delta m, n+\Delta n}(t+\Delta t) s^*_{m,n}(t)] = r_z(\Delta m/B_x, \Delta n/B_y, \Delta t) \\ &= \int_{-B_x/2}^{B_x/2}\int_{-B_y/2}^{B_y/2}\int_{-B/2}^{B/2} \Phi_z(f_x,f_x,f) e^{j2\pi(\Delta m f_x/B_x + \Delta n f_y/B_y + \Delta t f)} df_x df_y df. \end{aligned} \tag{10}$$

Consider the one dimensional process $s(x)$ and its filtered version $z(x)$ to compare the correlation structure

of $s_{pwc}(x)$ and $s(x)$. By substituting normalized frequency $\omega_x = f_x/B_x$ in (3) we have

$$r_s(\Delta x/B_x) = B_x \int_{-1/2}^{1/2} \Phi_s(\omega_x B_x) e^{j2\pi\omega_x \Delta x} d\omega_x \tag{11}$$

and similar normalized relation holds for $r_z(\Delta x/B_x)$ and $\Phi_z(\omega_x B_x)$. The correlation between PWC model coefficients is then $r_s[\Delta m] = r_z(\Delta m/B_x)$. In partcular, $\sigma^2_{s,pwc} = r_s[0] = r_z(0) < r_s(0) = \sigma^2_s$. Fig. 3 illustrates the correlation structure for a truncated Gaussian PSD: $\Phi_s(\omega_x B_x) = e^{-\omega_x^2/\alpha^2}$, $\omega_x \in [-0.5, 0.5]$, for three different values of α^2 = 0.1, 0.36, 1. The value of $\alpha^2 = 0.36$ corresponds to half-power one-sided bandwidth of Φ_s being equal to $B_x/2$. Fig. 3(a) and (b) plot the normalized (by the maximum value) PSD's $\Phi_s(\omega_x B_x)$ and $\Phi_z(\omega_x B_x)$ with respect to ω_x. Nyquist samples of $s(x)$ are uncorrelated if and only if Φ_s is constant within the bandwidth (as $\alpha^2 \to \infty$ in the Gaussian PSD). This is evident from Fig. 3(c) in which the normalized correlation $r_s(\Delta x/B_x)/r_s(0)$ is plotted with respect to the normalized lag Δx — the correlation decreases as α^2 increases. Due to smoothing, $z(x)$ exhibits higher correlation compared to $s(x)$, but not significantly so, as evident from Fig. 3(d). The values of $r_z(\Delta x/B_x)$ at integer values of Δx correspond to the correlation in PWC coefficients $\{s_m\}$: $r_s[\Delta m] = r_z(\Delta m/B_x)$.[3] We note from Fig. 3(d) that $r_s[\Delta m] \approx 0$ for $\Delta m \geq 2$.

Spatial degrees of freedom in the signal field. The above correlation analysis quantifies our earlier observation that most of the spatial correlation in $s(x, y, t)$ is limited to within each SCR, and the residual correlation across SCR's is primarily limited to adjacent SCR's for smooth signal spectra. Furthermore, this spatial correlation structure is preserved by the PWC component processes $\{s_{m,n}(t)\}$. In fact, the PWC approximation $s_{pwc}(x, y, t)$ preserves the most important statistical information about $s(x, y, t)$: *spatial degrees of freedom in* $\mathcal{R}$, which equal $N_s = (D_x B_x)(D_y B_y)$. The sampling theorem states that all *spatial* information about $s(x, y, t)$ in contained in the samples $s[m, n, t] = s(m/B_x, n/B_y, t)$, and these coefficients are approximately uncorrelated as discussed above. The number of samples in $\mathcal{R}$ equals N_s, which is precisely the number of component processes $s_{m,n}(t)$ in the PWC approximation (5). In fact, at any time t, $s[m, n, t]$ corresponds to the spatial sample at the center of the (m, n)-th SCR, whereas the PWC component $s_{m,n}(t)$ in (6) corresponds to the signal average in the SCR. The correlation analysis shows that the $s_{m,n}(t)$'s corresponding to different SCR's are also approximately uncorrelated. *Thus, there are approximately* $N_s = (D_x B_x)(D_y B_y)$ *independent spatial degrees of freedom in* $s(x, y, t)$ *over* $\mathcal{R}$ *which are preserved by* $s_{pwc}(x, y, t)$.[4]

[3] Note that the PWC coefficients $s_m = z(m/B_x)$ will be uncorrelated if Φ_z is constant within the bandwidth; that is, $\Phi_s(\omega_x B_x) \propto 1/\text{sinc}^2(\omega_x)$.

[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., [12]). Fading refers to channel variation over space or time.

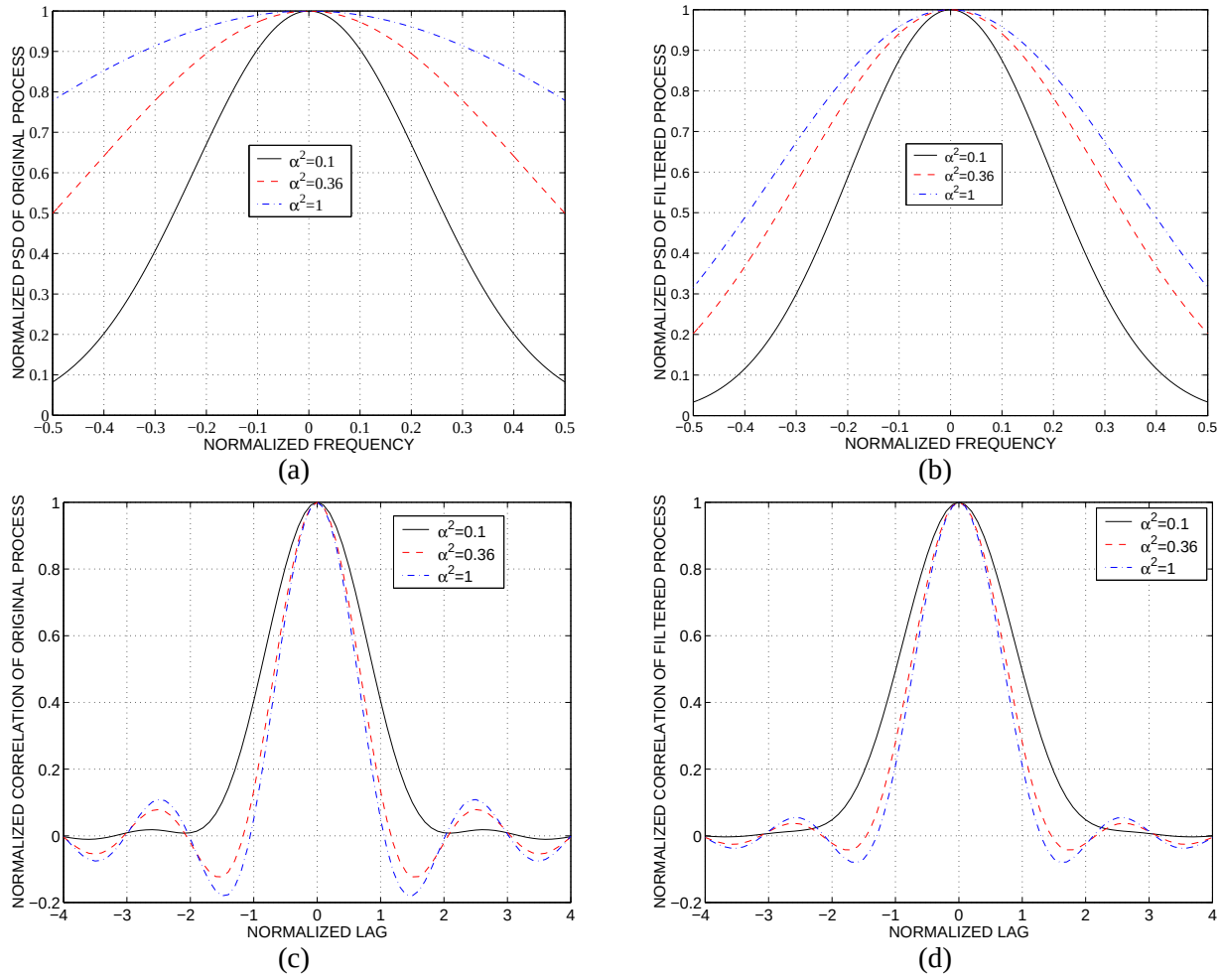

Figure 3: *Comparison of the normalized PSDs and correlation functions of the original ($s(x)$) and filtered ($z(x)$) one-dimensional processes for $\Phi_s(\omega_x B_x) = e^{-\omega_x^2/\alpha^2}$. The plots are for three values of α^2: 0.1, 0.36, and 1, and the corresponding values of σ_z^2/σ_s^2 are 0.88, 0.81, 0.79. (a) Plot of $\Phi_s(\omega_x B_x)/\max \Phi_s$ versus ω_x. (b) Plot of $\Phi_z(\omega_x B_x)/\max \Phi_z$. (c) Plot of $r_s(\Delta x/B_x)/r_s(0)$ versus Δx. (d) Plot of $r_z(\Delta x/B_x)/r_z(0)$.*

4 Implications for Design of Sensor Networks

The distributed nature of signal processing in sensor networks warrants a fresh investigation of traditional centralized approaches under network constraints. The proposed PWC signal modeling framework directly facilitates such investigations. We now discuss the implications of the PWC model in a variety of fundamental and practical issues in the design of sensor networks. Unless otherwise stated, in all cases we consider a network query relating to a spatial region $\mathcal{R} = D_x \times D_y$ of area $A = D_x D_y$ m^2. We assume that there are N uniformly[5] distributed nodes in $\mathcal{R}$. The area of each SCR is $A_c = D_{c,x} D_{c,y} = 1/B_x B_y$ m^2. Thus, there are $N_s = A/A_c = D_x B_x D_y B_y$ SCR's, $\{\mathcal{R}_{m,n}\}$, and there are $N_c = N/N_s$ nodes in each SCR.

[5]The essential ideas apply to non-uniform distribution as well.

4.1 Space-Time Sampling of the Signal Field

To extract all information about the signal field in $\mathcal{R}$, there should be at least one node in each SCR; that is, $N_c \geq 1$. $N_c = 1$ corresponds to Nyquist sampling. Equivalently, the minimum required (Nyquist) node density is $B_x B_y$ nodes per unit area.[6] Assume a higher than Nyquist node density ($N_c > 1$). An important question is: *How many* and *which* node measurements in $\mathcal{R}$ should be processed to execute the query? If the sensor measurements are noise-free, one node measurement in each SCR ($N_c = 1$) is sufficient. However, in the presence of measurement noise, additional node measurements in each SCR are extremely advantageous to improve the measurement signal-to-noise ratio. Let $q_l(t) = s_l(t) + w_l(t)$ denote the noisy measurement at the l-th node, $l = 1, \cdots, N$, where $w_l(t)$ denotes an AWGN process with PSD σ_w^2. We assume that the noise processes at different nodes are independent and identically distributed (i.i.d). Let $\mathrm{SNR}_{meas} = \sigma_s^2/\sigma_w^2$ denote the measurement SNR at each node where $\sigma_s^2 = \mathrm{E}[|s_l(t)|^2]$ denotes signal power. Estimating $s_{m,n}(t)$ in the PWC model from the $\{q_l(t)\}$ corresponds to averaging all node measurements in $\mathcal{R}_{m,n}$

$$\hat{s}_{m,n}(t) = \sum_{l \in \mathcal{R}_{m,n}} q_l(t) = s_{m,n}(t) + w'_{m,n}(t) \tag{12}$$

which is also the MMSE estimate of $s_{m,n}(t)$ (within a scaling constant). This coherent averaging in each SCR improves the *effective* SNR_{meas} by a factor of N_c: $\mathrm{SNR}_{meas} \rightarrow \mathrm{SNR}_{meas}/N_c$. Clearly the same principle applies to temporal measurements at each node as well. Oversampling above the Nyquist rate, B samples/sec, can improve the measurement SNR. Furthermore, since all the temporal samples are available at the same node, optimal MMSE estimation (rather than simple averaging over the coherence interval $T_c = 1/B$) could be used.

4.2 Nature of Information Exchange Within a Query Region

How should the sampled sensor measurements in $\mathcal{R}$ be processed to execute the query? In general, CSP of node measurements will be needed and would require some form of information exchange between nodes. Temporal measurements at each node are often processed in blocks of (Nyquist or oversampled) samples for the application at hand. Information can be exchanged between nodes at two basic levels of abstraction: *feature* or *symbol* level. A *feature* represents a lower-dimensional data representation (e.g. a lower dimensional transform of data block at each node) that contains the relevant information in the signal. A *symbol* represents a compressed version of the feature vector; for example, a quantized representation in the case of compression or a set of local decisions in the case of decision making (detection and classification). The decisions may be soft, such as a likelihood value, or hard, such as an index from a finite set.

[6] For non-uniform node spacing, a higher node density may be needed to preserve signal information.

As noted above, noisy node measurements $q_l(t)$ or features within an SCR can be coherently averaged to yield an effective feature for the SCR at the improved SNR_{meas}. This requires feature-level exchange between nodes within each SCR. Furthermore, since the node measurements in different SCR's are approximately uncorrelated, the mapping from the features to symbols may be done independently in different SCR's without signficant impact on performance. These independent symbols from different SCR's can then be communicated to a manager node for final processing. Thus, the PWC model suggests a hierarchical structure for information exchange between nodes that is naturally suited to the communication constraints of the network: *high-bandwidth feature-level exchange is confined to spatially local nodes within each SCR, whereas low-bandwidth symbol-level exchange is sufficient across spatially distant SCR's*. This structure on information exchange is illustrated in Fig. 6.

4.3 Fundamental Limits on Information in the Signal Field

What is the communication burden on the network imposed by a query? An upper bound can be obtained by characterizing the rate at which information is generated by the node measurements. This can be done by computing the (differential) *entropy rate* [13] of the space-time signal field.

Consider Nyquist samples $s[m,n,i] = s(m/B_x, n/B_y, i/B)$ of the signal field. It can be shown that the (differential) entropy[7] of each sample, or the entropy rate of the signal field in $\mathrm{bits/sample}$ is

$$\mathcal{H}_s = \mathcal{H}(\{s[n,m,i]\}) = \frac{1}{B_x B_y B} \int_{B_x}\int_{B_y}\int_{B} \log(\pi e B_x B_y B \Phi_s(f_x,f_y,f))\, df_x df_y df \ \ \mathrm{bits/sample}. \quad (13)$$

For a bandlimited white noise field (independent samples) with total power σ_s^2, $\Phi_s(f_x,f_y,f) = \sigma_s^2/(B_x B_y B)$, and the entropy rate in (13) becomes

$$\mathcal{H}_{s,wn} = \log(\pi e \sigma_s^2) \ \ \mathrm{bits/sample} \quad (14)$$

which is the differential entropy of a zero-mean complex Gaussian random variable with variance σ_s^2. Since there are $B_x B_y$ spatial samples per unit area and B temporal samples per second, the *entropy rate per unit area per second* is given by

$$\mathcal{H}_u = \mathcal{H}(\{s(x,y,t)\}) = B_x B_y B \mathcal{H}_s = \int_{B_x}\int_{B_y}\int_{B} \log(\pi e B_x B_y B \Phi_s(f_x,f_y,f))\, df_x df_y df \ \ \mathrm{bits/m^2/}s. \quad (15)$$

Thus, the entropy rate of the signal field over $\mathcal{R}$ is

$$\mathcal{H}_{\mathcal{R}} = D_x D_y \mathcal{H}_u = D_x B_x D_y B_y B \mathcal{H}_s = N_s B \mathcal{H}_s \ \ \mathrm{bits/s}. \quad (16)$$

[7]The differential entropy $\mathcal{H}(X)$ of a continuous random variable X can be misleading in terms of quantifying the number of bits required to encode it, since it can be negative. However, the entropy of an b-bit quantization of X is approximately $\mathcal{H}(X) + b$ [13].

In particular, since there is a single spatial sample in each SCR, the entropy rate of the signal field in each SCR is $\mathcal{H}_{scr} = B\mathcal{H}_s$ bits/s.

How different is the entropy rate of the PWC approximation? Recall that the processes $\{s_{m,n}(t)\}$ in the PWC model correspond to Nyquist spatial samples of the filtered process $\Phi_z(x,y,t)$. Consider Nyquist temporal samples: $s_{m,n}[i] = s_{m,n}(i/B)$. The entropy rate of the PWC approximation can be characterized as

$$\begin{aligned}
\mathcal{H}_{s,pwc} &= \mathcal{H}(\{s_{pwc}[m,n,i]\}) = \frac{1}{B_x B_y B}\int_{B_x}\int_{B_y}\int_{B}(\pi e B_x B_y B \log(\Phi_z(f_x,f_y,f))df_x df_y df \text{ bits/sample} \quad (17)\\
\mathcal{H}_{u,pwc} &= \mathcal{H}(\{s_{pwc}(x,y,t)\})\\
&= \int_{B_x}\int_{B_y}\int_{B}\log(\pi e B_x B_y B \text{sinc}^2(f_x/B_x)\text{sinc}^2(f_y/B_y)\Phi_s(f_x,f_y,f))df_x df_y f \text{ bits/m}^2\text{/s}. \quad (18)
\end{aligned}$$

Combining (13) and (17) and using (9) we note that

$$\mathcal{H}_s - \mathcal{H}_{s,pwc} = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\log\left(\frac{1}{\text{sinc}^2(\omega_x)\text{sinc}^2(\omega_y)}\right)d\omega_x d\omega_y \approx 2.77\text{bits/sample} \quad (19)$$

which is independent of Φ_s and shows that the difference in the per-sample entropy rates of $s(x,y,t)$ and $z(x,y,t)$ is about 2.77 bits/sample.[8]

The above calculations for the entropy rate of the PWC approximation include the spatial correlation between the component processes. Under the uncorrelated assumption we get

$$\begin{aligned}
\mathcal{H}_{s,pwc-uc} &= \log\left(\pi e \text{E}[|s_{pwc}(t)|^2]\right) = \log(\pi e \sigma_z^2) \text{ bits/sample} \quad (20)\\
&= \log\left(\pi e \int_{B_x}\int_{B_y}\int_{B}\text{sinc}^2(f_x/B_x)\text{sinc}^2(f_y/B_y)\Phi_s(f_x,f_y,f))df_x df_y f\right) \quad (21)\\
\mathcal{H}_{u,pwc-uc} &= B_x B_y B \mathcal{H}_{s,pwc-uc} \text{ bits/m}^2\text{/s} \quad (22)\\
\mathcal{H}_{\mathcal{R},pwc-uc} &= N_s B \mathcal{H}_{s,pwc-uc} \text{ bits/s}. \quad (23)
\end{aligned}$$

To summarize, the above calculations provide ideal estimates of the entropy rates (assuming a sufficiently high node density) and yield the following relations between per-sample rates

$$\mathcal{H}_s \leq \mathcal{H}_{s,wn} \;\;,\;\; \mathcal{H}_{s,pwc} \leq \mathcal{H}_{s,pwc-uc} \text{ and } \mathcal{H}_s - \mathcal{H}_{s,pwc} \approx 2.77 \text{ bits/sample}. \quad (24)$$

However, the entropy rate over $\mathcal{R}$ of both $s(x,y,t)$ and $s_{pwc}(x,y,t)$ is $\mathcal{O}(N_s B) = \mathcal{O}(D_x B_x D_y B_y B)$ and the difference is only in the constant. *Thus, the PWC model preserves the order of the entropy rate of the signal field as a function of the spatial degrees of freedom.*

[8]While the differential entropy can be negative, the *difference* of differential entropies of two random variables correctly quantifies the difference in the number of bits required to encode them.

Relation with network transport capacity. The above analysis shows that the rate at which information is generated in $\mathcal{R}$ is $\mathcal{O}(N_s B)$. How does this relate to the communication capacity of the network of nodes in $\mathcal{R}$? Recent results in network information theory show that the transport capacity of a network of N nodes grows as $\mathcal{O}(\sqrt{N})$ (bit-meters/s) [14]. Suppose that $D_x = D_y$. Under the assumption of uniform node distribution, there are $\mathcal{O}(\sqrt{N})$ nodes in each spatial dimension. There are two ways in which the number of nodes can increase. If the area of $\mathcal{R}$ remains fixed as N increases, then the spacing between nodes decreases. Since the information rate of the signal field remains constant at $\mathcal{O}(N_s B) = \mathcal{O}(D_x B_x D_y B_y B)$ in this case, the network should be able to transport this information for sufficiently large N. This corresponds to the case discussed in [15] to show the feasibility of sensor networks. The more interesting case is when the average node spacing remains constant. In this case, as N increases, the area of $\mathcal{R}$ and N_s increase as $\mathcal{O}(N)$. In this case, the $\mathcal{O}(\sqrt{N})$ network transport capacity is not sufficient to transport the information in the signal field. However, the results of [14] ignore cooperation between nodes. More recent results (see e.g., [16]) indicate that such cooperative routing schemes may yield $\mathcal{O}(N)$ capacity scaling, which would be sufficient to transport the signal field information even in this challenging case.

4.4 Distributed Signal Compression and Estimation

In practice, the node measurements have to be quantized for digital communication over the network.[9] What does the signal model tell us about distributed estimation and compression of the signal field? Most existing works on distributed compression in sensor networks focus on noise-free measurements (see, e.g., [18] and references therein). If the node measurements are noisy, as is usually the case, the PWC model suggests strategies for joint estimation and compression by averaging the node measurements in each SCR before quantization, thereby improving the effect SNR_{meas} and performance.

Consider a noisy signal field modeled as $q(x, y, t) = s(x, y, t) + w(x, y, t)$, where $w(x, y, t)$ is a zero-mean complex Gaussian noise field that is spatially and temporally white. Consider Nyquist samples of the noisy field: $q[m, n, i] = q(m/B_x, n/B_y, i/B) = s[m, n, i] + w[m, n, i]$ where $\mathrm{E}[|w[m, n, i]|^2] = \sigma_w^2$. First, consider *independent* (scalar) quantization of each Nyquist sample, $q = s + w$, where we suppress the indices because of stationarity. We want to generate a quantized version $\hat{q}$ of q to within a prescribed *per-sample distortion* $d_s = \mathrm{E}[|q - \hat{q}|^2]$. Rate-distortion theory [13] tells us that the minimum number of bits per sample required for

[9] Analog communication is also a possibility [17], but we do not discuss it here.

a given $0 < d_s < \sigma_q^2$ is given by the rate-distortion function

$$R_s(d_s) = \log\left(\frac{\sigma_q^2}{d_s}\right) = \log\left(\frac{\sigma_s^2 + \sigma_w^2}{d_s}\right) \ \text{bits/sample} \tag{25}$$

The distortion *per unit area per second* is $d_u = B_x B_y B d_s$ and the corresponding rate is $R_u(d_u) = B_x B_y B\ R_s(d_s)\ \text{bits/m}^2\text{/s}$. Thus, over $\mathcal{R}$, the *total distortion per second* and the required minimum bit rate are

$$d_{\mathcal{R}} = D_x D_y d_u = N_s B d_s \tag{26}$$

$$R_{\mathcal{R}}(d_{\mathcal{R}}) = D_x D_y\ R_u(d_u) = N_s B\ R_s(d_s)\ \text{bits/s}. \tag{27}$$

We note that d_s quantifies the distortion in representing each *noisy* sample $q = s + w$. The *effective* per-sample distortion in representing the *noise-free* signal sample s is thus

$$d_{s,eff} = d_s + \sigma_w^2. \tag{28}$$

Thus, d_s in (25) should be replaced with $d_{s,eff} - \sigma_w^2$ where $\sigma_w^2 < d_{s,eff} < \sigma_s^2 + 2\sigma_w^2$.

Now consider the component processes $q_{m,n}(t)$ of the PWC approximation obtained by averaging the N_c node measurements in the (m,n)-th SCR. Nyquist temporal sampling yields $q_{m,n}[i] = q_{m,n}(i/B) = z[m,n,i] + v[m,n,i]$ where we have approximated the samples of the averaged signal field with Nyquist samples of the ideal smoothed field $z(x,y,t)$ defined in (8). Note that $\mathrm{E}[|v[m,n,i]|^2] = \sigma_v^2 = \sigma_w^2/N_c$ due to averaging. Consider independent (scalar) quantization of the PWC samples $q_{m,n}[i]$. For a given per-sample distortion $0 < d_{s,pwc} < \sigma_z^2 + \sigma_v^2$, the required minimum bit rate is

$$R_{s,pwc}(d_{s,pwc}) = \log\left(\frac{\sigma_z^2 + \sigma_w^2/N_c}{d_{s,pwc}}\right) \ \text{bits/sample}. \tag{29}$$

The distortion per unit area per second is $d_{u,pwc} = B_x B_y B d_{s,pwc}$ and the required number of bits per unit area per second are given by $R_{u,pwc}(d_{u,pwc}) = B_x B_y B R_{s,pwc}$. The total distortion and corresponding bit rate per second over $\mathcal{R}$ are given for the PWC model by

$$d_{\mathcal{R},pwc} = N_s B d_{s,pwc} \tag{30}$$

$$R_{\mathcal{R},pwc}(d_{\mathcal{R}}) = N_s B\ R_{s,pwc}(d_{s,pwc})\ \text{bits/s}. \tag{31}$$

In this case, we need to account for the bias introduced by spatial smoothing, in addition to the effective noise v, in order to compute the effective distortion. This can be done by computing the per-sample MSE due to PWC

approximation which is given by $\mathrm{mse}_{pwc} = \mathrm{E}[|s[m,n,i] - s_{m,n}[i]|^2] = B_x B_y \mathrm{mse}_{scr}$, where mse_{scr} is defined in (7). Thus, for the PWC model

$$d_{s,eff} = d_{s,pwc} + \sigma_w^2/N_c + \mathrm{mse}_{pwc} \tag{32}$$

$$\mathrm{mse}_{pwc} = \int_{B_x}\int_{B_y} \Phi_s(f_x,f_y)[1 - \mathrm{sinc}^2(f_x/B_x)\mathrm{sinc}^2(f_y/B_y)]df_x df_y = \sigma_s^2 - \sigma_z^2 < \sigma_s^2. \tag{33}$$

We can substitute $d_{s,pwc} = d_{s,eff} - \sigma_w^2/N_c - \mathrm{mse}_{pwc}$ in (29) where $\sigma_w^2/N_c + \mathrm{mse}_s < d_{s,eff} < \sigma_s^2 + 2\sigma_w^2/N_c + \mathrm{mse}_s$.

Comparing (27) and (31), we note that for a given $d_{s,eff}$ the minimum bit rate over $\mathcal{R}$ is of the same order, $\mathcal{O}(N_s B)$, for both Nyquist samples and PWC model samples; the difference is only in the per-sample rate. Furthermore, for a given $d_{s,eff}$, the per-sample rate can be actually lower in the PWC model due to noise reduction. Comparing (28) and (32) we note that for a given $d_{s,eff}$

$$\frac{d_{s,pwc} - d_s}{\sigma_w^2} = 1 - \gamma \mathrm{SNR}_{meas} - 1/N_c \tag{34}$$

where $\gamma = \mathrm{mse}_{pwc}/\sigma_s^2 < 1$ and $\mathrm{SNR}_{meas} = \sigma_s^2/\sigma_w^2$. The relation (34) states that as long as $\mathrm{SNR}_{meas} < 1/\gamma$ (equivalently, $\sigma_w^2 > \mathrm{mse}_{pwc}$), a larger $d_{s,pwc}$ compared to d_s could be tolerated to maintain the same effective distortion $d_{s,eff}$. This is faciliated by increasing N_c; in particular, N_c should be chosen larger than $1/(1 - \gamma\mathrm{SNR}_{meas})$, the value at which the RHS of (34) is equal to zero. This suggests that for a given $d_{s,eff}$, PWC coding would require smaller per-sample bit rate as long as $\mathrm{SNR}_{meas} < 1/\gamma$. For $\mathrm{SNR}_{meas} > 1/\gamma$, Nyquist coding would be sufficient (no signal averaging within a SCR is needed). Fig. 4 compares Nyquist versus PWC coding in two regimes to illustrate the advantage of signal averaging in PWC coding at low measurement SNRs. More generally, it illustrates the importance of joint estimation and compression of noisy measurements in each SCR.

Independent versus Joint Compression. The above analysis considers independent quantization of Nyquist samples or PWC samples in each SCR. Since these samples have some residual correlation, better compression performance may be attained by joint (vector) coding of the samples [13]. How much does joint quantization buy us? We address this question by considering ideal joint quantization of Nyquist samples, assuming perfect collaboration across SCR's and ignoring the corresponding network communication cost. The per-sample rate-distortion function can now be expressed in terms of the PSD of $q(x,y,t)$: $\Phi_q = \Phi_s + \sigma_w^2/B_x B_y B$. We can express the per-sample distortion d_s as

$$d_s = \frac{1}{B_x B_y B}\int_{B_x}\int_{B_y}\int_{B} d_s(f_x,f_y,f)df_x df_y df \tag{35}$$

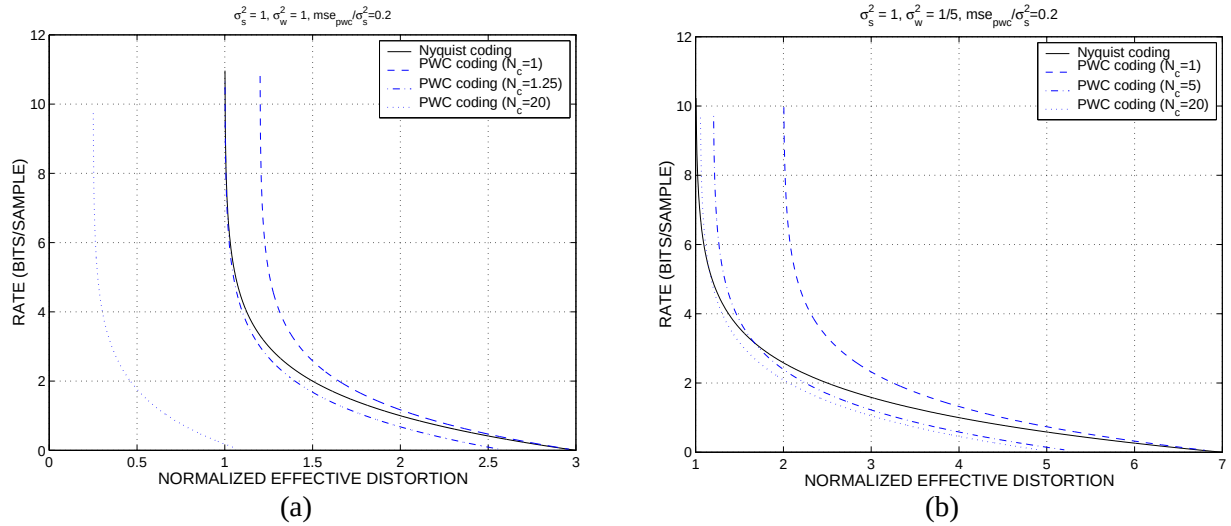

Figure 4: *Rate-distortion comparison of Nyquist coding versus PWC coding for* $\gamma = \mathrm{mse}_{pwc}/\sigma_s^2 = 1-\sigma_z^2/\sigma_s^2 = 0.2$ *(corresponding to Fig. 3). (a)* $\mathrm{SNR}_{meas} = 1 < 1/\gamma = 5$*. As evident PWC coding yields a smaller per-sample bit rate for* $N_c > 1/(1-\gamma \mathrm{SNR}_{meas}) = 1.25$*. (b)* $\mathrm{SNR}_{meas} = 5 = 1/\gamma$*. In this case (and for larger* SNR_{meas}*), Nyquist coding is sufficient; PWC coding cannot beat it no matter how large* N_c *is.*

where $d_s(f_x, f_y, f)$ denotes the distribution of d_s over the spatial and temporal bandwidths. For sufficiently large number of samples (SCR's), the required bit rate can be computed as [13][10]

$$R_{s,joint}(d_s) = \frac{1}{B_x B_y B} \int_{B_x} \int_{B_y} \int_{B} \log \left(\frac{B_x B_y B \Phi_q(f_x, f_y, f)}{d_s(f_x, f_y, f)} \right) df_x df_y df \quad \text{bits/sample} \tag{36}$$

$$d_s(f_x, f_y, f) = \min(\lambda, B_x B_y B \Phi_q(f_x, f_y, f)) \tag{37}$$

where λ is chosen to satisfy (35) for a given d_s. The solution captured by (36) and (37) corresponds to reverse waterfilling in the spectral domain. It says that for a given d_s, only those spectral components are allocated bits at which $B_x B_y B \Phi_q(f_x, f_y, f)$ is larger than λ. For sufficiently small distortion, $d_s(f_x, f_y, f) = \lambda = d_s$ and (36) reduces to

$$R_{s,joint}(d_s) = \frac{1}{B_x B_y B} \int_{B_x} \int_{B_y} \int_{B} \log \left(\frac{B_x B_y B \Phi_q(f_x, f_y, f)}{d_s} \right) df_x df_y df \quad \text{bits/sample.} \tag{38}$$

Comparing (25) and (38) for a given d_s, we note that

$$R_s(d_s) - R_{s,joint}(d_s) = \frac{1}{B_x B_y B} \int_{B_x} \int_{B_y} \int_{B} \log \left(\frac{(\sigma_s^2 + \sigma_w^2)/B_x B_y B}{\Phi_q(f_x, f_y, f)} \right) df_x df_y df \geq 0 \tag{39}$$

which we recognize as the Kullback-Liebler (K-L) distance (relative entropy) between a white PSD and Φ_q, and it is thus always non-negative. The equality in (39) holds when $q(x, y, t)$ corresponds to bandlimited white noise: $\Phi_q = \sigma_q^2/(B_x B_y B)$. In this case, for $d_s < \sigma_q^2$, (38) reduces to (25) corresponding to independent (scalar)

[10] Following the discussion on joint compression of independent random variables (different frequencies in this context).

quantization. We note that the difference in the performance of scalar versus vector quantization decreases at lower SNR_{meas}. For $\Phi_q = \Phi_s$ ($\sigma_w^2 = 0$), when Φ_s is as in Fig. 3, the difference in (39) is 0.0069, .0325, and 0.3347 $\mathrm{bits/sample}$ for $\alpha^2 = 1, 0.36$ and 1.

Thus, while vector quantization can yield a lower bit rate *per sample* than scalar quantization (for the same d_s), the difference is relatively small. More importantly, the bit rates over $\mathcal{R}$ are $\mathcal{O}(N_s B)$ for both independent and joint quantization. The same conclusion applies to scalar versus vector quantization of PWC model samples as well. Consequently, the advantage of vector quantization may not be significant when the number of SCR's (N_s) is large, or the SNR_{meas} is low, especially considering the significant communication burden imposed by vector quantization. Furthermore, part of the gain of vector quantization may be obtained by joint quantization of temporal samples (which does not require CSP). Vector quantization could be advantageous when N_s is relatively small and it would be more so at high SNR_{meas}.

4.5 Distributed Detection and Classification

Detection and classification of objects moving through the sensor field is an important application of sensor networks, e.g., detection and classification of vehicles based on acoustic measurements [10]. The approximate PWC model is more than adequate in such applications for signal sources that can be well-modeled as a stationary Gaussian field in the region of interest (e.g., acoustic measurements of vehicle engine sound [10]). As mentioned earlier, in classification applications, *feature vectors* of sufficiently high dimension are extracted from blocks of time series data at each node. For signals that are approximately stationary in time, a natural choice of feature vectors is *Fourier features* obtained by Fourier transforming the blocks of data. The advantage of Fourier feature vectors is that different feature components correspond to different frequencies and are approximately uncorrelated.[11] The required dimension of the feature vectors increases with the number of sources to be classified.

There are two main sources of error in distributed detection and classification: i) the noise in sensor measurements, and ii) the inherent variability in the source signal. In [19, 20], we have shown that the optimal classifier exploits the structure on information exchange suggested by the PWC model, illustrated in Fig. 6, to naturally combat the two sources of error. First, the feature vectors from a subset of N_c nodes in each SCR are coherently averaged to yield an effective feature vector at a higher effective SNR_{meas}. Second, the statistically

[11] It is well-known that complex sinusoids serve as eigenfunctions of stationary processes for sufficiently large values of TB, where T is the block duration and B is the signal bandwidth. The feature vectors then correspond to a subset of TB frequency components with significant power [10].

independent local node decisions (hard or soft, based on the effective feature) from the N_s SCR's are combined (at the manager node) to combat the inherent variability in the signal. In this approach, the differences in temporal PSD's of the different sources are exploited (via the feature vectors) to facilitate discrimination between classes, whereas the multiple node measurements are exploited to increase the reliability of final decisions at the manager node.

The results in [19, 20] yield two important conclusions. First, fusion of decisions from different SCR's is the most important factor influencing the reliability of the final decision — the high-bandwidth feature-level exchange within each SCR (to improve SNR_{meas}) can be avoided in many situations. Second, a moderate number (20-50) of relatively unreliable[12] local node decisions (hard or soft) from different SCR's can be combined at the manager node (even via noisy communication links) to yield remarkably reliable[13] final decisions. Fig. 5, based on real acoustic data[14] illustrates this remarkable advantage of fusion of decisions from different SCR's (see [19, 20] for a detailed discussion). These encouraging initial results also underscore the opportunity to leverage the vast literature on distributed detection, data fusion and pattern classification (see, e.g., [21, 22, 23, 24, 25, 26]) in the context of the specialized constraints of sensor networks.

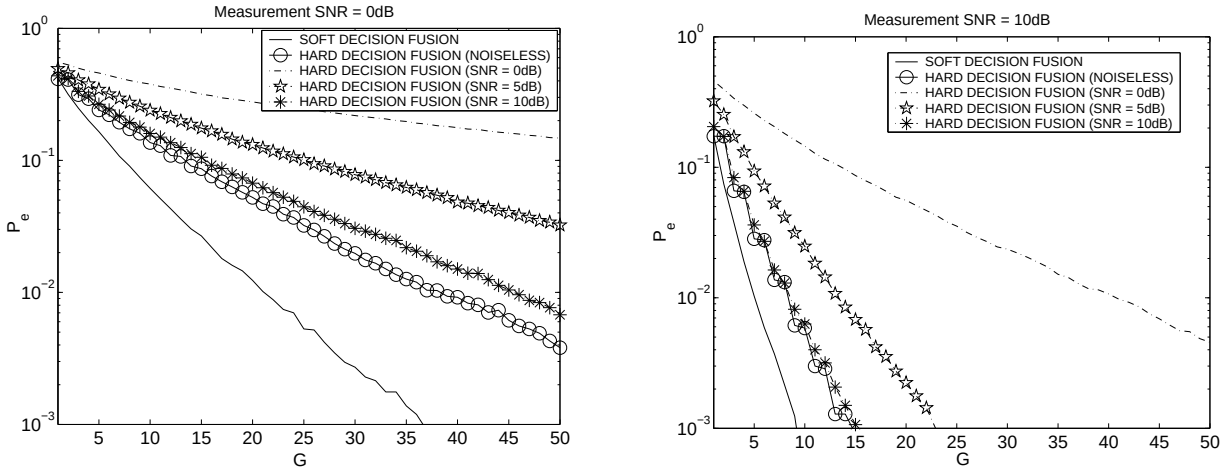

Figure 5: *Probability of error (of misclassification among 3 vehicles) as a function of $G = N_s$ (number of SCR's) for noise-free soft and hard decision fusion, and noisy hard decision fusion. In noisy hard decision fusion, the decisions are communicated over an AWGN channel and three different communication SNR's are shown: $\mathrm{SNR}_{comm} = 0, 5, 10$ dB. P_e for $G = 1$ reflects the reliability of individual node decisions. (a) SNR_{meas} = 0dB. (b) SNR_{meas} = 10dB.*

[12] With error probabilities as high as 0.2 or 0.3

[13] With an error probability of 0.01 or smaller. See Fig. 5.

[14] Collected as part of the DARPA SensIT program.

4.6 Communication and Routing Strategies

The power and bandwidth available at each node are the key resources from a communication perspective. As illustrated in Fig. 6, the nature of information exchange suggested by the PWC signal model is naturally suited to the location-centric information routing paradigm: (high rate) feature-level exchange is limited to local nodes within each SCR and (low rate) symbol-level information exchange is sufficient between (spatially distant) different SCRs. Location-centric routing can directly enable this bandwidth efficient information exchange via manager nodes in each SCR. Thus, given the knowledge of the spatial signal bandwidths associated with a network query, PWC signal modeling suggests a *fixed matched interface* between collaborative information processing and routing. For example, as suggested by [15] and our discussion in Section 4.4, joint compression and routing could be exploited in each SCR. This connection can also be exploited to assess the communication cost of CSP algorithms and to determine optimal region sizes in location-centric routing (Fig. 1) to facilitate CSP.

The PWC model also suggests natural communication strategies for communicating the symbol-level information from each SCR to a destination node or region. We illustrate the basic idea in one scenario in Fig. 6. Consider a query that requires information from the eight SCR's on the boundary to be communicated to a *single* node in the SCR at the center. If each SCR is of a sufficiently small size, the N_c nodes in each SCR could *collaborate* to send its information to the destination node. Suppose that all nodes can transmit at a power P and have a communication bandwidth W. Then, each *SCR* $\rightarrow$ *destination* link can be thought of as a coherent MISO (multiple input single output) communication link with N_c transmitting nodes. If we assume AWGN node-to-node communication links, then the capacity of this MISO link is given by (ignoring path loss)

$$C_{miso} = W \log(1 + N_c^2 \mathrm{SNR}_{comm}/W) \ \text{ bits/s} \tag{40}$$

where $\mathrm{SNR}_{comm} = P/\sigma_n^2$ is the SNR of each node-to-node link. Effectively, all the N_c nodes in the SCR collaborate to act as a coherent virtual node array that beamforms the information to the destination. Since B random variables are quantized per second in each SCR, this requires $BR_{s,pwc} < C_{miso}$ (assuming PWC coding) for reliable communication of this information, which is feasible.[15]

Since different SCR's are sending independent information to the destination node, the communication link from the N_s (=8) SCR's to the destination node is a multiple access (MAC) communication link whose *sum*

[15] We note that for this SCR $\rightarrow$ destination link, the results of [17] show that *uncoded* communication of the analog Gaussian source (rather than coded communication of compressed symbols) is optimal from a rate-distortion viewpoint.

capacity [13] is given by

$$C_{mac-miso} = W \log(1 + N_s N_c^2 \mathrm{SNR}_{comm}/W) \ \text{bits/s}. \tag{41}$$

Thus, we require $N_s B R_{s,pwc} < C_{mac-miso}$ to reliably communicate the sensed information from the N_s SCRs to the destination node. Here we see a problem: the source information rate is increasing as $\mathcal{O}(N_s)$ whereas the capacity of the MAC-MISO link is growing as $\mathcal{O}(\log(N_s))$. If all nodes have the same communication capabilities, this presents a fundamental bottleneck at the destination node, which cannot be avoided if we insist on sending information from a growing number of independent SCR's to a single node. The only way to sustain the information flow is by matching the number of independent SCRs in the source to the number of independent receiving nodes in a (distributed) destination. More sophisticated collaborative communication strategies are needed to address this key challenge in network communication.

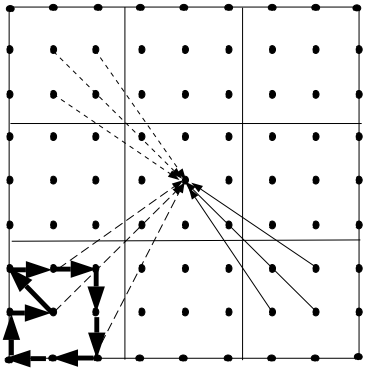

Figure 6: *Illustration of the implications of the PWC signal model for the design of sensor networks. The big square represents a region $\mathcal{R} = D_x \times D_y$ associated with a query, and the smaller squares depict the SCR's. The size of each SCR, $(1/B_x) \times (1/B_y)$, corresponds to a Nyquist spatial bin and decreases with increasing spatial signal bandwidths. The number of SCR's, $N_s = D_x B_x D_y B_y$, represents the spatial degrees of freedom in the signal field. Each SCR represents a single spatial degree of freedom – information worth a single random variable is sensed in each SCR. The nature of of information exchange suggested by the PWC model is naturally suited to the communication network constraints. Joint processing of high-bandwidth feature-level data is limited to within SCR's, and generates information at the low-bandwidth symbol level (compressed signal representation or local decisions). The independent symbols from different SCR's can then be communicated to the destination node/region for final processing. From an information processing perspective, node measurements in each SCR may be averaged to improve the measurement SNR. From a communication perspective, the nodes in each SCR may collaborate as a virtual coherent node array (and improve the communication SNR) to transport the symbol-level information to the destination node or region.*

5 Conclusion

The two vital operations in a sensor network, information processing and information routing, are intimately coupled and the nature of interplay between them is determined by the statistical characteristics of the signal field sensed by the nodes. In this paper, we have taken a first step in investigating the nature of the interplay under the assumption of a stationary Gaussian space-time signal field. We have proposed a PWC signal model

that partitions a region of interest into SCR's. The size of each SCR is inversely related to the spatial signal bandwidths and the number of SCR's quantifies the independent spatial degrees of freedom in the signal field (and hence the rate at which information is generated by it). As summarized in Fig. 6, the PWC model provides several useful insights. In particular, the primary communication burden of distributed signal processing and routing is confined to within SCR's in which high-bandwidth (feature-level) information exchange is most advantageous. Lower-bandwidth (symbol-level) information exchange is sufficient across different SCR's.

The PWC model makes two approximations about spatial signal characteristics: i) it ignores signal variation within each SCR, and ii) it ignores residual correlation across SCR's. This was motivated by the desire to minimize the communication cost associated with distributed signal processing. We have quantified the error introduced by this approximation in various contexts. Our results suggest that capturing finer spatial signal variation in each SCR, such as in signal estimation and compression, becomes important only at sufficiently high measurement SNR's and/or sufficiently low allowable distortion. Similarly, capturing the residual correlation across SCR's, such as in joint signal compression, becomes important when the number of SCR's is small and/or the measurement SNR is sufficiently high. The impact of PWC modeling is least significant in the context of distributed decision making.

The insights revealed by the PWC model suggest several directions for future work. One interesting direction is to develop a hierarchical framework for distributed signal processing that combines PWC modeling at multiple scales, as in wavelet-based signal processing [27]. Another interesting direction is to refine the signal model to account for two effects that are commonly encountered in the case of point sources: i) the effect of signal path loss, and ii) the effect of source mobility. Finally, including node-related attributes, such as remaining battery life, and directional sensing, could be fruitfully exploited in both information processing and routing.

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] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, pp. 102–114, Aug. 2002.

[3] G. Pottie, "Wireless sensor networks," *Information Theory Workshop*, Jun. 1998.

[4] "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.

[5] P. Ramanathan, "Location-centric approach for collaborative target detection, classification, and tracking in sensor networks," in *Proceedings of IEEE CAS Workshop on Wireless Communication and Networking*, Sept. 2002.

[6] P. Ramanathan, K.-C. Wang, K. K. Saluja, and T. Clouqueur, "Communication support for location-centric collaborative signal processing in sensor networks," in *Proceedings of DIMACS Workshop on Pervasive Networks*, May 2001.

[7] N. Sundaram and P. Ramanathan, "Connectivity based location estimation scheme for wireless ad hoc networks," in *in Proceedings of Globecom*, Nov. 2002.

[8] L. Doherty, K. S. J. Pister, and L. E. Ghaoui, "Convex position estimation in wireless sensor networks," in *Proceedings of INFOCOM*, pp. 1655–1663, Apr. 2001.

[9] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low cost outdoor localization for very small devices," *IEEE Personal Communications Magazine*, vol. 7, pp. 28–24, Oct. 2000.

[10] 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.

[11] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, pp. 292–300, Mar. 1976.

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

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

[14] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[15] A. Scaglione and S. Servetto, "On the interdependence of routing and data compression in multi-hop sensor networks," in *MOBICOM'02*, (Atlanta, GA), September 2002.

[16] P. Gupta and P. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Th.*, vol. 49, pp. 1877–1894, Aug. 2003.

[17] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science (Proceedings of IPSN'03)*, (Springer-Verlag, Berlin Heidelberg), pp. 162–177, (F. Zhao and L. Guibas (eds.), April 2003.

[18] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, Mar. 2002.

[19] 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.

[20] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed classification of gaussian space-time sources in wireless sensor networks," *submitted to the IEEE J. Selec. Areas. Commun.*, July 2003.

[21] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[22] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1996.

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

[24] 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.

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

[26] R. Blum, S. Kassam, and H. Poor, "Distributed detection with multiple sensors I. Advanced topics," *Proceedings of the IEEE*, vol. 85, pp. 64–79, Jan. 1997.

[27] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.