How Much Information Can One Get From a Wireless *Ad Hoc* Sensor Network Over a Correlated Random Field?

Youngchul Sung, *Senior Member, IEEE*, H. Vincent Poor, *Fellow, IEEE*, and Heejung Yu, *Student Member, IEEE*

***Abstract*—New large-deviations results that characterize the asymptotic information rates for general d-dimensional (d-D) stationary Gaussian fields are obtained. By applying the general results to sensor nodes on a two-dimensional (2-D) lattice, the asymptotic behavior of *ad hoc* sensor networks deployed over correlated random fields for statistical inference is investigated. Under a 2-D hidden Gauss–Markov random field model with symmetric first-order conditional autoregression and the assumption of no in-network data fusion, the behavior of the total obtainable information [nats] and energy efficiency [nats/J] defined as the ratio of total gathered information to the required energy is obtained as the coverage area, node density, and energy vary. When the sensor node density is fixed, the energy efficiency decreases to zero with rate $\Theta(\text{area}^{-1/2})$ and the per-node information under fixed per-node energy also diminishes to zero with rate $O(N_t^{-1/3})$ as the number N_t of network nodes increases by increasing the coverage area. As the sensor spacing d_n increases, the per-node information converges to its limit D with rate $D - \sqrt{d_n}e^{-\alpha d_n}$ for a given diffusion rate α. When the coverage area is fixed and the node density increases, the per-node information is inversely proportional to the node density. As the total energy E_t consumed in the network increases, the total information obtainable from the network is given by $O(\log E_t)$ for the fixed node density and fixed coverage case and by $\Theta(E_t^{2/3})$ for the fixed per-node sensing energy and fixed density and increasing coverage case.**

***Index Terms*—*Ad hoc* sensor networks, asymptotic Kullback–Leibler information rate, asymptotic mutual information rate, conditional autoregressive model, Gauss–Markov random fields, large deviations principle, stationary Gaussian fields.**

I. Introduction

SENSOR networks have drawn much attention in recent years because of their promising applications such as scientific research, environmental monitoring, and surveillance [1]. In the design of sensor networks, there are several distinctive features. First, sensor networks are designed to sense and monitor various physical phenomena such as temperature, humidity, density of a certain gas or stress level of different locations in a structure. Many of these physical processes can be modeled as two-dimensional (2-D) random fields over a certain area, where the uncertainty of the underlying signal is captured as the randomness of samples and the proximity of samples close in location is modeled by the correlation among the samples. Second, sensors in different locations should be able to deliver the measured data to a control center (or fusion center) where the decision is made, and thus the communication capability is required as in *ad hoc* communication networks. Such communication functionality can be provided by networking sensor nodes, for example, using multihop routing. Third, energy is one of the critical issues in sensor network design since both sensing and communication require energy and it is difficult to recharge batteries in already deployed sensor nodes. Hence, it is of interest to design energy-efficient sensor networks.

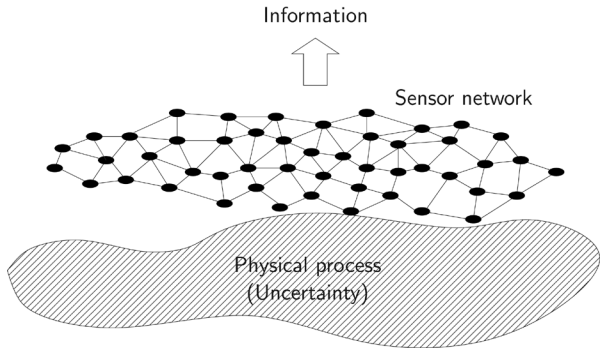

Fig. 1. *Ad hoc* sensor network over a physical process.

In this paper, we consider the design of such sensor networks, and investigate the behavior and efficiency of these networks from an information-theoretic perspective. From the information-theoretic viewpoint, the process of sensing and communication mentioned above can be viewed as extracting information (about the underlying 2-D physical process) using imperfect sensor nodes by expending energy for statistical inference such as detection or reconstruction of the sensed signal field [2], [3], as shown in Fig. 1. Relevant questions regarding the network design are as follows. How much information can one obtain from the network for given coverage and node density? How does the amount of gathered information change as we increase the coverage area or node density? How do the field correlation and

Manuscript received April 22, 2008; revised February 17, 2009. Current version published May 20, 2009. The work of Y. Sung was supported by the IT R&D program of MKE/IITA [2008-F-004-01 "5G mobile communication systems based on beam-division multiple access and relays with group cooperation".] The work of H. V. Poor was supported in part by the U. S. National Science Foundation under Grants ANI-03-38807 and CNS-06-25637.

Y. Sung and H. Yu are with the Department of Electrical Engineering, KAIST, Daejeon 305-701, Korea (e-mail: ysung@ee.kaist.ac.kr; hjyu@stein.kaist.ac.kr).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by H. Bölcskei, Associate Editor for Detection and Estimation.

Color versions of Figures 4–9 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2018333

measurement signal-to-noise (SNR) affect the amount of information obtainable from the network? What is the optimal node density? What are the information and energy tradeoffs in such a sensor network with *ad hoc* routing? Answering these questions is difficult, especially, because of the 2-D spatial correlation structure of the signal process inherent to the two dimensionality of network deployment. To circumvent this problem, several studies based on one-dimensional (1-D) spatial signal models have been conducted (see, e.g., [2], [4], [5]). However, there is an important difference between 1-D signal models and actual spatial signals. Suppose that we take observations from sensors located equidistantly along a line transect laid over an area. The observations may then be viewed as samples generated by a 1-D process along the line transect and results from time series analysis could be applied to examine their statistical properties. In the 2-D case, however, there is no natural notion of signal flow or dependence direction along the transect as there is in a more traditionally obtained time series. For samples from sensors placed over a 2-D area, it is necessary to consider the signal dependence in all direction in the plane.

A. The Approach and Summary of Results

In this paper, we consider *ad hoc* sensor networks deployed for making statistical inferences about underlying 2-D random fields, and address the above questions in a general 2-D setting. In particular, we investigate the amount of information obtainable from the network and related tradeoffs among information, coverage, density, and energy in various asymptotic settings, and reveal the fundamental behavior of large scale planar *ad hoc* sensor networks. We model the signal field as a 2-D Gauss–Markov random field (GMRF), which is suitable for many physical processes, and consider the Kullback–Leibler information (KLI) and mutual information (MI) as our information measures [6], [7]. Our approach for calculating the total obtainable information is based on the large deviations principle (LDP). Under a stationarity assumption, the amount of information from a sensor node becomes independent of sensor location as the network size grows, and the total amount of information is approximately given by the product of the number of sensor nodes and the asymptotic information rate or asymptotic per-node information. (Thus, the units of these quantities are nats/node.) To quantify the information content, we first derive closed-form expressions for the asymptotic per-node KLI and MI for stationary Gaussian fields in a general d-dimensional (d-D) lattice in the spectral domain, and then apply these results to the 2-D case. We do so by exploiting the spectral structure of d-D stationary Gaussian signals and the relationship between the eigenvalues of the block circulant approximation to a block Toeplitz matrix describing the d-D correlation structure. However, the general expressions obtained in this way render the investigation of the field correlation and SNR difficult. To address this problem, we adopt the conditional autoregression (CAR) model, which is a generalization of the autoregressive (AR) model of classical time-series analysis. We further investigate the properties of the asymptotic per-node KLI and MI as functions of the field correlation and the measurement SNR under the symmetric first-order conditional autoregression (SFCAR) model, which captures the 2-D correlation on the plane effectively. In this case, the asymptotic per-node KLI and MI are given explicitly in terms of the SNR and the field correlation. The behavior of the asymptotic per-node KLI and MI as functions of correlation strength is seen to divide into two regions depending on the value of the SNR. At high SNR, uncorrelated observations maximize the per-node information for a given SNR, whereas there is nonzero optimal correlation at low SNR. Interestingly, it is seen that there is a discontinuity in the optimal correlation strength as a function of SNR. In the perfectly correlated case, the asymptotic per-node KLI and MI are zero as expected. As a function of SNR, the asymptotic per-node information increases as $\log \mathrm{SNR}$ for a given correlation strength at high SNR. At low SNR, the two information measures show different rates of convergence to zero.

Based on the derived expressions for asymptotic per-node information and their properties under the SFCAR and the corresponding correlation function, we then investigate the fundamental behavior of large-scale *ad hoc* sensor networks deployed over correlated random fields for statistical inference. Specifically, we examine the total information [nats] (about the underlying physical process) obtainable from the network and the energy efficiency [nats/J] defined as the ratio of total gathered information to the required energy as the coverage, density and energy vary. We assume that sensors are located on a 2-D lattice and all sensor nodes in the network deliver the measured data to a fusion center in the center of the 2-D lattice via minimum hop routing without in-network data fusion. Under these assumptions, we have the following results on the tradeoffs among the information, coverage, density, and energy, and the results provide guidelines for the design of sensor networks for statistical inference about many interesting physical processes that can be modeled as 2-D correlated random fields.

(1) When the sensor node density is fixed, the amount of total information increases linearly with respect to (w.r.t.) the coverage area, and the energy efficiency decreases to zero with rate $\Theta\left(\mathrm{area}^{-1/2}\right)$ as the coverage area increases. Further, in this case the amount of information per sensor node diminishes to zero as the network size grows with fixed energy per node.

(2) As the sensor spacing d_n increases, the per-node information converges to its limit D with rate $D - \sqrt{d_n}e^{-\alpha d_n}$ for a given diffusion rate α. Hence, the per-node information saturates almost exponentially as we increase the sensor spacing.

(3) When the coverage area is fixed and the node density increases, the per-node information is inversely proportional to the node density for any nontrivial diffusion rate. Hence, the total amount of information from a given area is upper-bounded unless the random field is spatially white.

(4) As the total energy E_t consumed in the network increases, the total information obtainable from the network is given by $\Theta\left(E_t^{2/3}\right)$ for fixed node density and increasing coverage, whereas the total information increases only with rate of $O\left(\log E_t\right)$ for fixed node density and fixed coverage.

B. Related Work

Large deviations analysis of Gaussian processes in Gaussian noise has been considered previously, e.g., [8]–[13]. However, most work in this area considers only 1-D signals or time series.

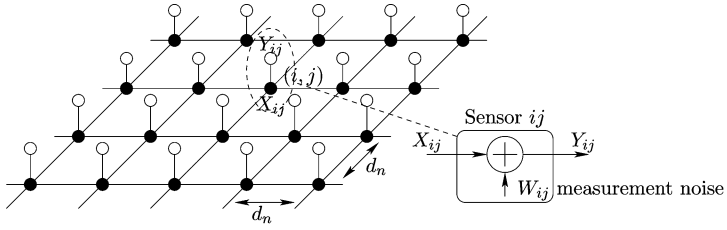

Fig. 2. Sensors on a 2-D lattice: hidden Markov structure.

A closed-form expression for the asymptotic KLI rate was obtained and its properties were investigated for 1-D hidden Gauss–Markov random processes in [12]. Large deviations analyses were used to examine the issues of optimal sensor density and optimal sampling in a 1-D signal model in [2] and [4]. For a 2-D setting, an error exponent was obtained for the detection of 2-D GMRFs in [14], where the sensors are located randomly and the Markov graph is based on the nearest neighbor dependency enabling a loop-free graph. Our work here focuses on the analysis of the fundamental behavior of 2-D sensor networks deployed for statistical inference via new large deviations results for general d-D and 2-D stationary Gaussian random fields and their application to 2-D SFCAR GMRFs, which enables us to investigate the impact of field correlation and measurement SNR on the information and the fundamental behavior of *ad hoc* sensor networks for statistical inference with preliminary presentation of the work in [15].

C. Notation and Organization

We will make use of standard notational conventions. Vectors and matrices are written in boldface with matrices in capitals. All vectors are column vectors. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^T$ indicates the transpose and $\boldsymbol{A}(i,j)$ denotes the (i,j)th element of $\boldsymbol{A}$. We reserve $\boldsymbol{I}_m$ for the identity matrix of size m (the subscript is included only when necessary). For a random vector $\boldsymbol{x}$, $\mathbb{E}_j\{\boldsymbol{x}\}$ is the expectation of $\boldsymbol{x}$ under probability density p_j, $j=0,1$. The notation $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\boldsymbol{x}$ is Gaussian distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For a set $\mathcal{A}$, $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$.

The paper is organized as follows. The background and signal model are described in Section II. In Section III, the closed-form expressions for the asymptotic KLI and MI rates are obtained in the spectral domain, and their properties are investigated as functions of the correlation and the SNR under the symmetric first-order CAR model. The tradeoffs related to *ad hoc* sensor networks deployed for statistical inference are presented in Section IV, followed by conclusions in Section V.

II. Background and Signal Model

We assume that sensors are distributed over a 2-D area and each sensor measures the underlying signal field at its location. To simplify the problem and gain insights into behavior in 2-D, we assume that sensors are located on a 2-D square lattice

$$\mathcal{I}_n \triangleq \{(i,j),\ i=0,1,\ldots,n-1, \text{ and } j=0,1,\ldots,n-1\} \tag{1}$$

where the distance between two adjacent nodes (i,j) and $(i+1,j)$ is d_n, as shown in Fig. 2. (We will use ij to denote (i,j) when there is no ambiguity of notation.) We model the 2-D signal field $\{X_{ij}, ij \in \mathcal{I}_n\}$ (or simply $\{X_{ij}\}$) sampled by sensors as a GMRF[1] w.r.t. an undirected graph in which a node corresponds to a sensor node or its signal sample. We assume that each sensor has Gaussian measurement noise. The noisy measurement Y_{ij} of Sensor ij on the 2-D lattice $\mathcal{I}_n$ is then given by

$$Y_{ij} = X_{ij} + W_{ij}, \quad ij \in \mathcal{I}_n \tag{2}$$

where $\{W_{ij}\}$ represents independent and identically distributed (i.i.d.) $\mathcal{N}(0,\sigma^2)$ noise with a known variance σ^2, and the GMRF $\{X_{ij}\}$ is assumed to be independent of the measurement noise $\{W_{ij}\}$. Thus, the observation samples form a 2-D hidden GMRF.[2] In the following, we briefly review results on GMRFs relevant to our further development.

Definition 1 (Undirected Graph): An undirected labeled graph $\mathcal{G}$ is a collection $(\mathcal{N},\mathcal{E})$ of nodes and edges, where $\mathcal{N}=\{1,2,\ldots,N\}$ is the set of nodes in the graph, and $\mathcal{E}$ is the set of edges $\{(l,m): l,m \in \mathcal{N}, \text{ and } l \neq m\}$. There exists an undirected edge between two nodes l and m if and only if $(l,m) \in \mathcal{E}$.

We will use the terms node, sample, and sensor interchangeably hereafter.

Definition 2 (GMRF): A Gaussian random vector $\boldsymbol{x} = [X_1, X_2, \ldots, X_N]^T \in \mathbb{R}^N$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} > 0$ is a GMRF w.r.t. a labeled graph $\mathcal{G}=(\mathcal{N},\mathcal{E})$ if X_l and X_m are independent given $\mathcal{X}_{-lm}$ if and only if there exists no edge between nodes l and m, where $\mathcal{X}_{-lm} \triangleq \{X_k, k \in \mathcal{N}, \text{ and } k \neq l,m\}$.

Note that a GMRF is defined using conditional independence on a graph. However, its distribution is easily characterized by the mean $\boldsymbol{\mu}$ and the precision matrix $\boldsymbol{Q}$ ($\triangleq \boldsymbol{\Sigma}^{-1}$), and is given by

$$p(\boldsymbol{x}) = (2\pi)^{-N/2}|\boldsymbol{Q}|^{1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{3}$$

and $Q_{lm} \neq 0$ if and only if $(l,m) \in \mathcal{E}$ for all $l \neq m$, i.e.,

$$Q_{lm} = 0 \Longleftrightarrow X_l \perp X_m | \mathcal{X}_{-lm}. \tag{4}$$

Note that the covariance matrix $\boldsymbol{\Sigma}$ is completely dense in general while the precision matrix $\boldsymbol{Q}$ has nonzero elements Q_{lm} only

[1]The Markov dependence structure may be restrictive. However, it is a meaningful model capturing 2-D spatial correlation structure and allowing further analysis.

[2]In this paper, we focus primarily on the spatial correlation structure of 2-D sensor fields, and the signal evolution over time is not considered.

when there is an edge between nodes l and m in the Markov random field. Hence, when the graph is not fully connected, the precision matrix is sparse [16]. The 2-D indexing scheme (i,j) in (1) and (2) can properly be converted to a 1-D scheme to apply Definitions 1 and 2. From here on, we again use the 2-D indexing scheme for convenience.

Definition 3 (Stationarity): A GMRF $\{X_{ij}\}$ on a 2-D infinite lattice $\mathcal{I}_\infty$ is said to be (second-order) stationary if the mean vector is constant and the covariance between samples X_{ij} and $X_{i'j'}$ depends only on the difference of the node index, i.e.,

$$\begin{aligned}\mathrm{Cov}(X_{ij},X_{i'j'}) &= \mathbb{E}\{(X_{ij}-\mu)(X_{i'j'}-\mu)\}\\ &= c(i-i',j-j')\end{aligned}$$

for some function $c(\cdot,\cdot)$, where μ is the mean of the stationary field.

Without loss of generality, we assume that the signal GMRF $\{X_{ij}\}$ is zero-mean.[3] For a 2-D zero-mean and stationary GMRF $\{X_{ij}\}$, the covariance $\{\gamma_{ij}\}$ is defined as

$$\gamma_{ij} \triangleq \mathbb{E}\{X_{i'j'}X_{i'+i,j'+j}\} = \mathbb{E}\{X_{00}X_{ij}\} \tag{5}$$

which does not depend on i' or j' due to the stationarity. The spectral density function of a stationary GMRF $\{X_{ij}\}$ on $\mathcal{I}_\infty$ with covariance γ_{ij} is defined as

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2}\sum_{ij\in\mathcal{I}_\infty}\gamma_{ij}e^{-\iota(i\omega_1+j\omega_2)} \tag{6}$$

where $\iota=\sqrt{-1}$ and $(\omega_1,\omega_2)\in[-\pi,\pi)^2$. Note that (6) is a 2-D extension of the conventional 1-D Fourier transform. We can express $\{\gamma_{ij}\}$ from the spectral density function via the inverse transform

$$\gamma_{ij} = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\omega_1,\omega_2)e^{\iota(i\omega_1+j\omega_2)}d\omega_1 d\omega_2. \tag{7}$$

A stationary GMRF can be implicitly specified by a CAR model, which is a natural generalization of the AR model arising in 1-D time series and which provides an efficient tool for capturing the spatial correlation structure of the sensor field considered here.

Definition 4 (The Conditional Autoregression [16]): A zero-mean CAR GMRF is defined by a set of full conditional normal distributions with mean and precision

$$\mathbb{E}\{X_{ij}|\mathcal{X}_{-ij}\} = -\frac{1}{\theta_{00}}\sum_{i'j'\in\mathcal{I}_\infty\setminus\{00\}}\theta_{i'j'}X_{i+i',j+j'} \tag{8}$$

and

$$\mathbb{E}^{-1}\{X_{ij}^2|\mathcal{X}_{-ij}\} = \theta_{00} > 0 \tag{9}$$

where $\mathcal{X}_{-ij}$ denotes the set of all variables except X_{ij}.

Note in (8) that the conditional mean of X_{ij} given all other node variables depends on nodes $(i+i',j+j')$ such that $\theta_{i'j'}\neq 0$, and the relationship between the CAR model of (8) and (9) and the precision matrix is given by

$$Q_{(i,j),(i+i',j+j')} = \theta_{i'j'}. \tag{10}$$

Hence, the Markov dependence structure on the graph is easily captured by the CAR model through (4), and $\{\theta_{i'j'}\}$ directly represent the connectivity of the Markov graph.

Theorem 1 (Spectrum of a CAR Model [16]): The GMRF defined by the CAR model of (8) and (9) is a zero-mean stationary Gaussian process on $\mathcal{I}_\infty$ with the spectral density function

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2}\frac{1}{\sum_{ij\in\mathcal{I}_\infty}\theta_{ij}\exp(-\iota(i\omega_1+j\omega_2))} \tag{11}$$

if

$$|\{\theta_{ij}\neq 0\}|<\infty,\quad \theta_{ij}=\theta_{-i,-j},\quad \theta_{00}>0 \tag{12}$$

and

$$\{\theta_{ij}\} \text{ is such that } f(\omega_1,\omega_2)>0,\quad \forall\,(\omega_1,\omega_2)\in[-\pi,\pi)^2. \tag{13}$$

Henceforth, we assume that the 2-D stochastic signal $\{X_{ij}\}$ in (2) is given by a stationary GMRF defined by the CAR model of (8) and (9) satisfying (12) and (13) as $n\to\infty$.

The SNR of the observation Y_{ij} in (2) is well defined due to the stationarity as $n\to\infty$, and is given by

$$\mathrm{SNR} = \frac{\mathbb{E}\{X_{ij}^2\}}{\mathbb{E}\{W_{ij}^2\}} = \frac{P}{\sigma^2},\quad \forall\, ij \tag{14}$$

where the signal power is constant over $(i,j)\in\mathcal{I}_\infty$ and is given, using the inverse Fourier transform of (6), by

$$P=\gamma_{00}=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\omega_1,\omega_2)d\omega_1 d\omega_2. \tag{15}$$

III. Asymptotic Information Rates: Closed-Form Expressions and Impact of Correlation and SNR

In this section, we derive closed-form expressions for the asymptotic KLI and MI rates under the 2-D CAR GMRF model discussed in the previous section. We further investigate the properties of the asymptotic information rates under a symmetric correlation assumption. For the MI, the signal model (2) is directly applicable, whereas for the KLI the probability density functions of the null (noise-only) and alternative (signal-plus-noise) distributions are given by

$$p_0(Y_{ij}): Y_{ij}=W_{ij},\quad ij\in\mathcal{I}_n \qquad \text{and} \tag{16}$$

$$p_1(Y_{ij}): Y_{ij}=X_{ij}+W_{ij},\quad ij\in\mathcal{I}_n \tag{17}$$

respectively. The asymptotic KLI rate $\mathcal{K}$ is defined as

$$\mathcal{K} = \lim_{n\to\infty}\frac{1}{|\mathcal{I}_n|}\log\frac{p_0}{p_1}(\{Y_{ij},ij\in\mathcal{I}_n\}) \quad \text{almost surely (a.s.) under } p_0 \tag{18}$$

where p_0 and p_1 are given by (16) and (17), respectively. Under a Neyman–Pearson detection formulation, the miss probability

[3] Of course, if a stationary GMRF has a known and nonzero mean, the known mean can be subtracted to yield a zero-mean field.

P_M decays exponentially in many cases, including (16) and (17), and the error exponent is defined as the exponential decay rate

$$\lim_{|\mathcal{I}_n|\to\infty} -\frac{1}{|\mathcal{I}_n|}\log P_M \tag{19}$$

where $|\mathcal{I}_n|$ is the total number of samples in $\mathcal{I}_n$. It is known that the error exponent is given by the asymptotic KLI rate $\mathcal{K}$ defined in (18) in this case [17]. Hence, a larger KLI rate (or per-node KLI) implies better detection performance with a given network size, or a smaller network size required for a given level of performance.

While the asymptotic KLI rate determines the error exponent for Neyman–Pearson detection, the asymptotic MI rate is interpreted as the amount of uncertainty reduction about the hidden signal field resulting from one observation sample, in the large sample size regime. The asymptotic MI rate $\mathcal{I}$ is given by

$$\begin{aligned}\mathcal{I} &= \lim_{n\to\infty}\frac{1}{|\mathcal{I}_n|}I(\{X_{ij}, ij\in\mathcal{I}_n\};\{Y_{ij}, ij\in\mathcal{I}_n\})\\ &= \lim_{n\to\infty}\frac{1}{|\mathcal{I}_n|}[H(\{X_{ij}, ij\in\mathcal{I}_n\})\\ &\qquad - H(\{X_{ij}, ij\in\mathcal{I}_n\}|\{Y_{ij}, ij\in\mathcal{I}_n\})].\end{aligned} \tag{20}$$

It is shown in the sequel that the asymptotic KLI rate is smaller than the asymptotic MI rate and that the two information measures converge when SNR increases. Thus, at high SNR, the two information measures are equivalent.

A. Asymptotic Information Rates in General d-Dimension

While the 2-D results are relevant to our analysis of fundamental tradeoffs in planar sensor networks, it is of theoretical interest to investigate the statistical properties of stationary Gaussian random fields in general higher dimension. In this subsection, we first derive closed-form expressions for the asymptotic KLI and MI rates for stationary Gaussian random fields in d-D, and then apply the results to the 2-D case. For a stationary d-D Gaussian random field $\{Y_{\boldsymbol{i}}, \boldsymbol{i}\in\mathbb{Z}^d\}$, where $\mathbb{Z}$ is the set of all integers, the autocovariance function under p_1 is given by

$$\gamma_{\boldsymbol{h}} = \mathbb{E}_1\{Y_{\boldsymbol{i}}Y_{\boldsymbol{i}+\boldsymbol{h}}\},\quad \boldsymbol{h}=(h_1,h_2,\ldots,h_d)\in\mathbb{Z}^d \tag{21}$$

and the corresponding Fourier transform (i.e., the power spectral density) and its inverse are given by

$$\begin{aligned}f_1(\boldsymbol{\omega}) &= \frac{1}{(2\pi)^d}\sum_{\boldsymbol{h}\in\mathbb{Z}^d}\gamma_{\boldsymbol{h}}e^{-\iota\boldsymbol{h}\cdot\boldsymbol{\omega}},\\ &\qquad \boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_d)\in[-\pi,\pi)^d\end{aligned} \tag{22}$$

and

$$\gamma_{\boldsymbol{h}} = \int e^{\iota\boldsymbol{h}\cdot\boldsymbol{\omega}}f_1(\boldsymbol{\omega})d\boldsymbol{\omega} \tag{23}$$

respectively, where the integration is over $\boldsymbol{\omega}\in[-\pi,\pi)^d$, and $\boldsymbol{h}\cdot\boldsymbol{\omega}$ denotes the inner product between $\boldsymbol{h}$ and $\boldsymbol{\omega}$. Note that (21), (22), and (23) are the extensions of (5), (6), and (7), respectively, to d-D. The null and alternative distributions arising in the KLI in d-D are given by

$$\begin{cases} p_0(Y_{\boldsymbol{i}}): Y_{\boldsymbol{i}} = W_{\boldsymbol{i}}, & \boldsymbol{i}\in\mathcal{D}_n\\ p_1(Y_{\boldsymbol{i}}): Y_{\boldsymbol{i}} = Y_{\boldsymbol{i}}^{(1)}, & \boldsymbol{i}\in\mathcal{D}_n\end{cases} \tag{24}$$

where $\{W_{\boldsymbol{i}}\}$ are i.i.d. Gaussian from $\mathcal{N}(0,\sigma^2)$, $\{Y_{\boldsymbol{i}}^{(1)}\}$ is a stationary d-D Gaussian random field with spectrum $f_1(\boldsymbol{\omega})$,[4] and

$$\mathcal{D}_n \triangleq [0,1,\ldots,n-1]^d. \tag{25}$$

Based on the previous work [18], we further exploit the relationship between the eigenvalues of block circulant and block Toeplitz matrices representing correlation structure in d-D and the i.i.d. null distribution, and obtain the KLI for (24) given by the following theorem.

Theorem 2 (Asymptotic KLI Rate in d-D): Suppose that

A.1 the alternative spectrum $f_1(\boldsymbol{\omega})$ has a positive lower bound, and

A.2 $\exists\, M<\infty$ such that $\forall\, k=1,2,\ldots,d$

$$\sum_{\boldsymbol{h}\in\mathbb{Z}^d}(1+|h_k|)|\gamma_{\boldsymbol{h}}| < M.$$

Then, the asymptotic KLI rate $\mathcal{K}$ for (24) is given by

$$\begin{aligned}\mathcal{K} &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d}\left[\frac{1}{2}\log\frac{(2\pi)^d f_1(\boldsymbol{\omega})}{\sigma^2}\right.\\ &\qquad \left. -\frac{1}{2}\left(1-\frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})}\right)\right]d\boldsymbol{\omega}\end{aligned} \tag{26}$$

$$= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} D(\mathcal{N}(0,\sigma^2)\|\mathcal{N}(0,(2\pi)^d f_1(\boldsymbol{\omega})))d\boldsymbol{\omega} \tag{27}$$

where $D(\cdot\|\cdot)$ denotes the Kullback–Leibler distance.

Proof: See Appendix I.

Theorem 2 is an extension to general d-D of the asymptotic KLI rate in 1-D obtained in [12], and shows that the frequency binning interpretation of (27) holds in the general d-D case under some regularity conditions on the alternative spectrum. Note that the integrand in (27) is the Kullback–Leibler information between two zero-mean Gaussian distributions with variances σ^2 and $(2\pi)^d f_1(\boldsymbol{\omega})$, respectively. For each d-D frequency segment $d\boldsymbol{\omega}$, the spectra can be thought of as being flat, i.e., the signals are independent, and Stein's lemma [19] can be applied for the segment. The overall KLI is the sum of contributions from each bin. The smoothness of the spectrum $f_1(\boldsymbol{\omega})$ is a sufficient condition for Assumption *A.2* for second-order stationary fields, and thus the frequency binning in Theorem 2 is valid for a wide class of spectra. Theorem 2 follows from the fact that $\mathcal{K}$ is given by the almost-sure limit of the normalized log-likelihood ratio in (18) and that we have Gaussian distributions for p_0 and p_1. That is, $\mathcal{K}$ is given by the almost sure limit

$$\begin{aligned}\mathcal{K} = \lim_{n\to\infty}\frac{1}{|\mathcal{D}_n|}&\left(\frac{1}{2}\log\frac{\det(\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|})}{\det(\boldsymbol{\Sigma}_{0,|\mathcal{D}_n|})}\right.\\ &\left.+\frac{1}{2}\boldsymbol{y}_{|\mathcal{D}_n|}^T\left(\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1}-\boldsymbol{\Sigma}_{0,|\mathcal{D}_n|}^{-1}\right)\boldsymbol{y}_{|\mathcal{D}_n|}\right) \text{ under } p_0\end{aligned} \tag{28}$$

[4]Note that $\{Y_{\boldsymbol{i}}^{(1)}\}$ need not be a hidden Markov field.

where $\boldsymbol{y}_{|\mathcal{D}_n|}$ is a vector consisting of $|\mathcal{D}_n|$ observation samples $\{Y_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{D}_n\}$ with elements arranged in lexicographic order; for example, in 2-D

$$\begin{aligned}\boldsymbol{y}_{|\mathcal{I}_n|} &= [y_1, \ldots, y_{|\mathcal{I}_n|}]^T \\ &\triangleq [Y_{00}, Y_{10}, \ldots, Y_{n-1,0}, Y_{01}, \ldots, Y_{n-1,n-1}]^T \quad (29)\end{aligned}$$

and $\boldsymbol{\Sigma}_{0,|\mathcal{D}_n|}$ and $\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}$ are the covariance matrices of $\boldsymbol{y}_{|\mathcal{D}_n|}$ under p_0 and p_1, respectively. Note that the log-likelihood ratio in (28) consists of two terms: one is a deterministic term and the other is a quadratic random term. The overall convergence follows from the convergence of each of the two terms. Note that the deterministic term in (28) is simply the mutual information between $\{X_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{D}_n\}$ and $\{Y_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{D}_n\}$ for the model

$$Y_{\boldsymbol{i}} = X_{\boldsymbol{i}} + W_{\boldsymbol{i}}, \quad \boldsymbol{i} \in \mathcal{D}_n. \quad (30)$$

Using the convergence of the first term in the right-hand side (RHS) of (28), the asymptotic MI rate $\mathcal{I}$ for d-D is given by

$$\mathcal{I} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} \frac{1}{2} \log \frac{\sigma^2 + (2\pi)^d f(\boldsymbol{\omega})}{\sigma^2} d\boldsymbol{\omega} \quad (31)$$

where $f(\boldsymbol{\omega})$ is the spectrum of the signal $\{X_{\boldsymbol{i}}\}$. This is simply a d-D extension of the 1-D MI rate in spectral form [20], and shows the validity of the $\log(1 + \mathrm{SNR})$ formula and frequency binning approach in general d-D under some regularity conditions on the spectrum; a sufficient condition is provided in Theorem 2.

Applying the d-D results to the 2-D hidden GMRF model of (16) and (17), we have the following corollary for 2-D.

Corollary 1 (Asymptotic Information Rates in 2-D): Assuming that the conditions (12) and (13) hold, the asymptotic KLI and MI rates for the hidden CAR GMRF model with (16) and (17) are given by

$$\begin{aligned}\mathcal{K} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Bigg[&\frac{1}{2} \log \frac{\sigma^2 + 4\pi^2 f(\omega_1, \omega_2)}{\sigma^2} \\ &- \frac{1}{2} \left(1 - \frac{\sigma^2}{\sigma^2 + 4\pi^2 f(\omega_1, \omega_2)} \right) \Bigg] d\omega_1 d\omega_2 \quad (32)\end{aligned}$$

and

$$\mathcal{I} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{\sigma^2 + 4\pi^2 f(\omega_1, \omega_2)}{\sigma^2} d\omega_1 d\omega_2 \quad (33)$$

where $f(\omega_1, \omega_2)$ is the 2-D spectrum of the signal GMRF $\{X_{ij}, ij \in \mathcal{I}_\infty\}$ defined in (11).

Proof: See Appendix I.

Comparing (32) and (33), we note that the asymptotic KLI rate is strictly less than the asymptotic MI rate for any positive signal spectrum, and that the two information measures converge with a fixed offset of $-1/2$ as the SNR increases without bound since $\frac{\sigma^2}{\sigma^2 + 4\pi^2 f(\omega_1, \omega_2)} \to 0$ in (32) as $\mathrm{SNR} \to \infty$. Hence, the two information measures can be equivalently used at high SNR.

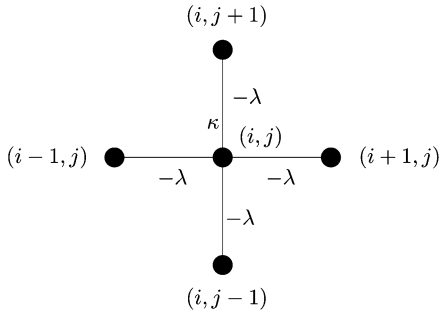

Fig. 3. Symmetric first-order conditional autoregression model.

B. Symmetric First-Order Conditional Autoregression

In the previous subsection, we have derived closed-form expressions for the asymptotic KLI and MI rates for hidden CAR GMRFs with general 2-D spectra defined in (11) in the spectral domain. However, these general spectral expressions render further analysis infeasible. To investigate the impact of the field correlation and the SNR on the information rates, we further adopt the SFCAR model, described by the conditions

$$\mathbb{E}\{X_{ij} | \mathcal{X}_{-ij}\} = \frac{\lambda}{\kappa}(X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1}), \quad (34)$$

and

$$\mathbb{E}^{-1}\{X_{ij}^2 | \mathcal{X}_{-ij}\} = \kappa > 0 \quad (35)$$

where $0 \le \lambda \le \frac{\kappa}{4}$.[5] Note that the parameters in (8) and (9) for this model are given by $\theta_{00} = \kappa$, $\theta_{1,0} = \theta_{-1,0} = \theta_{0,1} = \theta_{0,-1} = -\lambda$, and all other $\theta_{ij} = 0$. In this model, the correlation is symmetric for each set of four neighboring nodes, as seen in Fig. 3. The SFCAR model is a simple yet meaningful extension of the 1-D first-order AR model which has the conditional causal dependency only on the previous sample. Here in the 2-D SFCAR we have the conditional dependency on four neighboring nodes in the four (planar) directions. By Theorem1, the spectrum of the SFCAR is given by

$$f(\omega_1, \omega_2) = \frac{1}{4\pi^2 \kappa (1 - 2\zeta \cos \omega_1 - 2\zeta \cos \omega_2)} \quad (36)$$

where we define the edge dependence factor ζ as

$$\zeta \triangleq \frac{\lambda}{\kappa}, \quad 0 \le \zeta \le 1/4. \quad (37)$$

Note that for the range of $0 \le \zeta \le 1/4$ the 2-D spectrum (36) is always nonnegative and the conditions (12) and (13) are satisfied. Note also that $\zeta = 0$ corresponds to the i.i.d. case whereas $\zeta = 1/4$ corresponds to the perfectly correlated case, i.e., $X_{ij} = X_{i'j'}$ for all i, j, i', j'. Hence, the correlation strength can be captured in this single quantity ζ for 2-D SFCAR signals: larger ζ implies stronger correlation. The power of the SFCAR signal is obtained using the inverse Fourier transform via the relation (6), and is given by [21]

$$P = \gamma_{00} = \frac{2K(4\zeta)}{\pi\kappa} \quad \left(0 \le \zeta \le \frac{1}{4}\right) \quad (38)$$

where $K(\cdot)$ is the complete elliptic integral of the first kind. The SNR is given by

$$\mathrm{SNR} = \frac{P}{\sigma^2} = \frac{2K(4\zeta)}{\pi\kappa\sigma^2}. \quad (39)$$

[5]This is a sufficient condition to satisfy (12) and (13).

$$\mathcal{K}_s = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{1}{2}\log\left(1+\frac{\mathrm{SNR}}{(2/\pi)K(4\zeta)(1-2\zeta\cos\omega_1-2\zeta\cos\omega_2)}\right) - \frac{1}{2}\left(1-\frac{1}{1+\frac{\mathrm{SNR}}{(2/\pi)K(4\zeta)(1-2\zeta\cos\omega_1-2\zeta\cos\omega_2)}}\right)\right]d\omega_1 d\omega_2 \tag{40}$$

$$\mathcal{I}_s = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log\left(1+\frac{\mathrm{SNR}}{(2/\pi)K(4\zeta)(1-2\zeta\cos\omega_1-2\zeta\cos\omega_2)}\right)d\omega_1 d\omega_2. \tag{41}$$

Using (32), (36), and (39), we now obtain the asymptotic KLI and MI rates in the SCFAR signal case, denoted by $\mathcal{K}_s$ and $\mathcal{I}_s$ and given in the following corollary to Corollary 1.

Corollary 2: For the hidden 2-D SFCAR signal model the asymptotic per-node KLI $\mathcal{K}_s$ is given by (40) at the top of the page, and the asymptotic per-node MI $\mathcal{I}_s$ is given by (41), also at the top of the page.

Proof: The result follows upon substitution of (36) and (39) into (32) and (33), respectively. □

Note that the SNR for the hidden SFCAR model is dependent on correlation through ζ (see (39)). However, the SNR and correlation are separated in the expressions (40) and (41) for the asymptotic per-node information, which enables us to investigate the effects of each term on the per-sample information separately.

1) Properties of the Asymptotic Per-Node KLI and MI for the Hidden SFCAR Model: First, it is readily seen from Corollary 2 that the asymptotic per-node KLI $\mathcal{K}_s$ and MI $\mathcal{I}_s$ are continuously differentiable functions of the edge dependence factor $\zeta(0 \le \zeta \le 1/4)$ for a given SNR since $f: x \to K(x)$ is a continuously differentiable C^∞ function for $0 \le x < 1$ [22]. Now we examine the asymptotic behavior of $\mathcal{K}_s$ and $\mathcal{I}_s$ as functions of ζ. The values of $\mathcal{K}_s$ at the extreme correlations are given by noting that the values of the complete elliptic integral at the two extreme correlation points

$$K(0) = \frac{\pi}{2} \quad \text{and} \quad K(1) = \infty.$$

Therefore, in the i.i.d. case (i.e., $\zeta = 0$), Corollary 2 reduces to Stein's lemma [19] as expected, and $\mathcal{K}_s$ is given by

$$\mathcal{K}_s(0) = \frac{1}{2}\log(1+\mathrm{SNR}) - \frac{1}{2}\left(1-\frac{1}{1+\mathrm{SNR}}\right) \tag{42}$$

$$= D(\mathcal{N}(0,1)\|\mathcal{N}(0,1+\mathrm{SNR})). \tag{43}$$

For the perfectly correlated case ($\zeta = 1/4$), on the other hand, $\mathcal{K}_s = 0$. In fact, in this case as well as in the i.i.d. case, the two-dimensionality is irrelevant. The known result in the 1-D case [12] is applicable. With regard to $\mathcal{I}_s$, we have similar behavior at the extreme correlations. In the i.i.d. case, the mutual information is given by the well-known formula

$$\mathcal{I}_s(0) = \frac{1}{2}\log(1+\mathrm{SNR}) \tag{44}$$

whereas we have $\mathcal{I}_s = 0$ in the perfectly correlated case. Thus, both information measures are zero at perfect correlation ($\zeta = 1/4$). The limiting behavior of the asymptotic information rates near the extreme correlation values is given by Taylor's theorem. Due to the differentiability of $\mathcal{K}_s$ and $\mathcal{I}_s$ w.r.t. ζ, we have

$$\mathcal{K}_s(\zeta) = c_1 \cdot (\zeta - 1/4) + o(|\zeta - 1/4|) \tag{45}$$

and

$$\mathcal{I}_s(\zeta) = c_1' \cdot (\zeta - 1/4) + o(|\zeta - 1/4|) \tag{46}$$

in a neighborhood of $\zeta = 1/4$ for some constants c_1 and c_1' as $\zeta \to 1/4$. Similarly, we also have the linear limiting behavior for $\mathcal{K}_s$ and $\mathcal{I}_s$ in a neighborhood of $\zeta = 0$ with nonzero limiting values $D(\mathcal{N}(0,1)\|\mathcal{N}(0,1+\mathrm{SNR}))$ and $\frac{1}{2}\log(1+\mathrm{SNR})$, respectively, as $\zeta \to 0$. That is

$$\mathcal{K}_s(\zeta) = \mathcal{K}_s(0) + c_2\zeta + o(\zeta) \tag{47}$$

and

$$\mathcal{I}_s(\zeta) = \mathcal{I}_s(0) + c_2'\zeta + o(\zeta) \tag{48}$$

for some c_2 and c_2', as $\zeta \to 0$.

For intermediate values of correlation, we evaluate (40) and (41) for several different SNR values, as shown in Fig. 4. It is seen that, at high SNR, $\mathcal{K}_s$ decreases monotonically as ζ increases. Hence, i.i.d. observations yield the largest per-node information for a given value of SNR when SNR is large, as in the 1-D case [12]. As we decrease the SNR, it is seen that a second mode grows near $\zeta = 1/4$, i.e., in the strong correlation region. As we decrease the SNR further, the value of ζ of the second mode shifts toward $1/4$, and the value of the second mode exceeds that of the i.i.d. case. Hence, there is a discontinuity in the optimal correlation as a function of SNR in the 2-D case even though the maximal $\mathcal{K}_s$ itself is continuous, as seen in Fig. 5. That is, there is a phase transition for optimal correlation w.r.t. SNR: above a certain SNR value i.i.d. observations yield the best performance, whereas below that SNR point suddenly strong correlation is preferred. This is not the case for 1-D Gauss–Markov time series, where the optimal correlation maximizing the information rate is continuous w.r.t. SNR. Although it is not shown here, the per-node MI $\mathcal{I}_s$ exhibits similar behavior as a function of the edge dependence factor ζ.

With regard to $\mathcal{K}_s$ and $\mathcal{I}_s$ as functions of SNR, it is straightforward to see from (40) that they are continuously differentiable functions, and the behavior of $\mathcal{K}_s$ and $\mathcal{I}_s$ with respect to SNR is given by the following theorem.

Theorem 3 (Per-Node Information Versus SNR): The asymptotic per-node KLI $\mathcal{K}_s$ for the hidden SFCAR model is continuous and monotonically increasing as SNR increases for a given

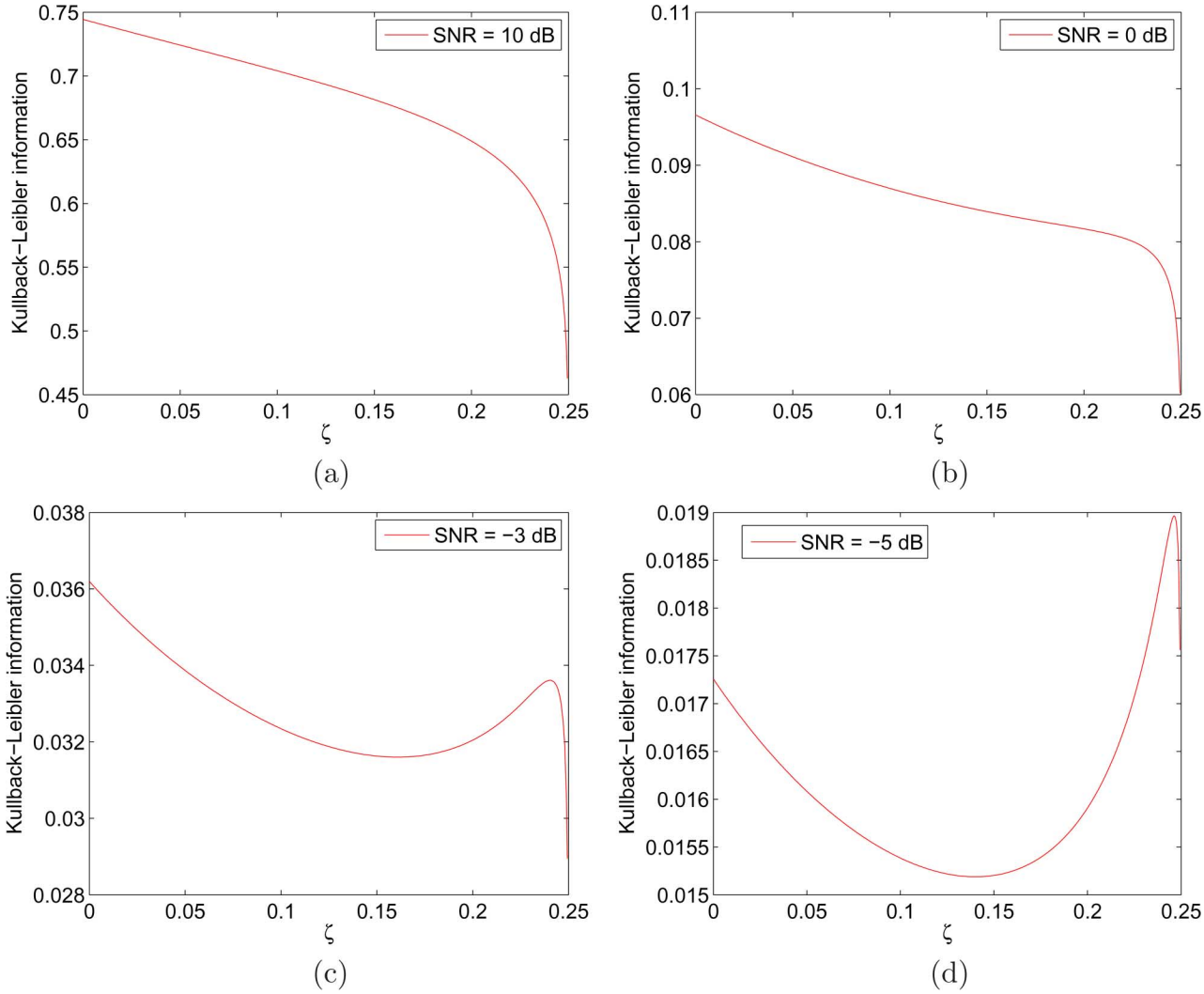

Fig. 4. $\mathcal{K}_s$ as a function of ζ: (a) SNR $= 10$ dB, (b) SNR $= 0$ dB, (c) SNR $= -3$ dB, (d) SNR $= -5$ dB.

edge dependence factor $\zeta \in [0 \quad 1/4]$. Moreover, $\mathcal{K}_s$ increases with rate $\frac{1}{2}\log \mathrm{SNR}$ as $\mathrm{SNR} \to \infty$. As SNR decreases to zero, on the other hand, $\mathcal{K}_s$ converges to zero and the rate of convergence is given by

$$\mathcal{K}_s(\mathrm{SNR}) = c_3 \cdot \mathrm{SNR}^2 + o(\mathrm{SNR}^2) \tag{49}$$

as $\mathrm{SNR} \to 0$, where c_3 is given by

$$c_3 = \frac{1}{2^6 K^2(4\zeta)} \times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{(1 - 2\zeta\cos\omega_1 - 2\zeta\cos\omega_2)^2} d\omega_1 d\omega_2. \tag{50}$$

The per-node MI $\mathcal{I}_s$ has similar properties as a function of SNR, i.e., it is a continuous and monotonically increasing function of SNR. At high SNR, it increases with rate $\frac{1}{2}\log \mathrm{SNR}$, whereas it decreases to zero with rate of convergence

$$\mathcal{I}_s(\mathrm{SNR}) = c_3' \cdot \mathrm{SNR} + o(\mathrm{SNR}) \tag{51}$$

as $\mathrm{SNR} \to 0$, where c_3' is given by

$$c_3' = \frac{1}{2^3 \pi K(4\zeta)} \times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{1 - 2\zeta\cos\omega_1 - 2\zeta\cos\omega_2} d\omega_1 d\omega_2. \tag{52}$$

Proof: See the Appendix I.

Note that the limiting behavior as $\mathrm{SNR} \to 0$ is different for $\mathcal{K}_s$ and $\mathcal{I}_s$; $\mathcal{K}_s$ decays to zero quadratically while $\mathcal{I}_s$ decreases linearly. Fig. 6 shows $\mathcal{K}_s$ and $\mathcal{I}_s$ with respect to SNR for $\zeta = 0.1$. The log SNR behavior is evident at high SNR for both information measures. Note that $\mathcal{K}_s$ and $\mathcal{I}_s$ increase with the same slope in the logarithmic scale with offset $1/2$. This is easily seen from (40) and (41) because the second term in the integrand of (40) converges to $-1/2$, and thus $\mathcal{K}_s \to \mathcal{I}_s - \frac{1}{2}$ as SNR increases. However, the offset is negligible as SNR increases. It is easy to see from (40) and (41) that for a given edge dependence factor ζ, the convergence between the two information measures is characterized by $\frac{\mathcal{K}_s}{\mathcal{I}_s} = 1 + O\left(\frac{1}{\log \mathrm{SNR}}\right)$ as $\mathrm{SNR} \to \infty$.

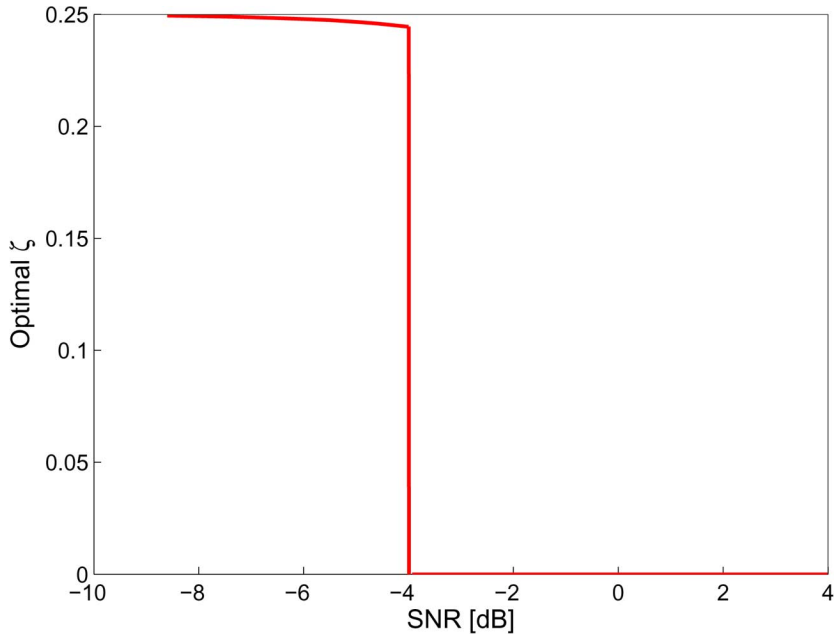

Fig. 5. Optimal ζ maximizing $\mathcal{K}_s$ versus SNR.

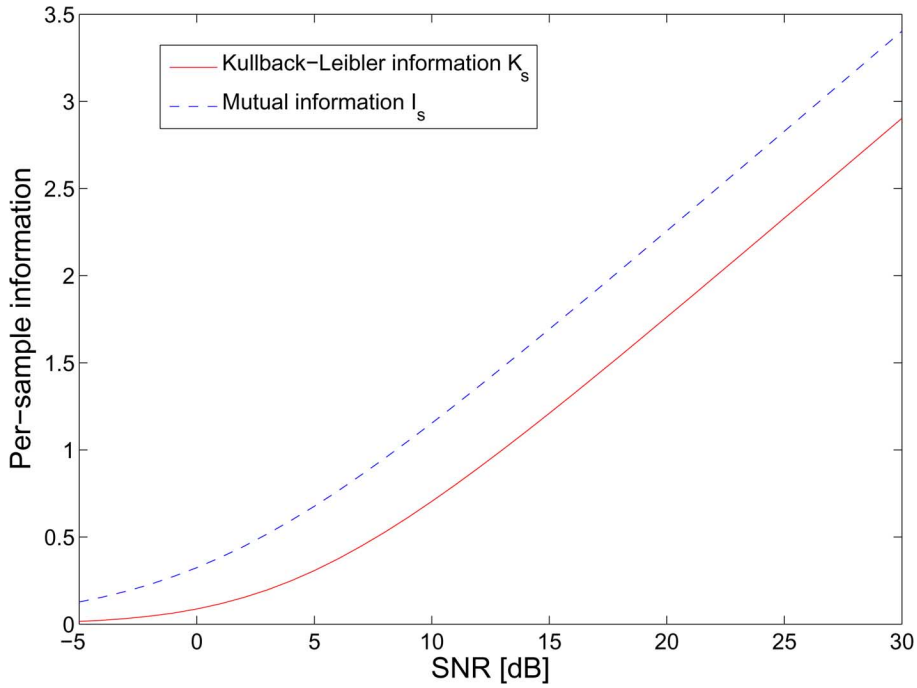

Fig. 6. $\mathcal{K}_s$ and $\mathcal{I}_s$ as functions of SNR ($\zeta = 0.1$).

IV. *Ad Hoc* Sensor Networks: Fundamental Tradeoffs Among Information, Coverage, Density, and Energy

Using the results of the previous sections, we now answer the fundamental questions, raised in Section I, concerning planar *ad hoc* sensor networks deployed over correlated random fields for statistical inference under the 2-D hidden SFCAR GMRF model. We first derive relevant physical correlation parameters for the SFCAR from the corresponding continuous-index stochastic model. Once the physical correlation parameters for the SFCAR are obtained, the analysis of information obtainable from an *ad hoc* sensor network and related trade-offs is straightforward.

A. Physical Correlation Model

We first derive how the physical correlation is related to the edge dependence factor ζ in the 2-D SFCAR model. The edge correlation coefficient ρ is defined as

$$\rho \triangleq \frac{\gamma_{01}}{\gamma_{00}} = \frac{\gamma_{10}}{\gamma_{00}} \quad (0 \le \rho \le 1) \tag{53}$$

due to the spatial symmetry, where $\gamma_{ij} = \mathbb{E}\{X_{00}X_{ij}\}$. ρ represents the correlation strength between the signal samples of two adjacent sensor nodes connected by the Markov dependence

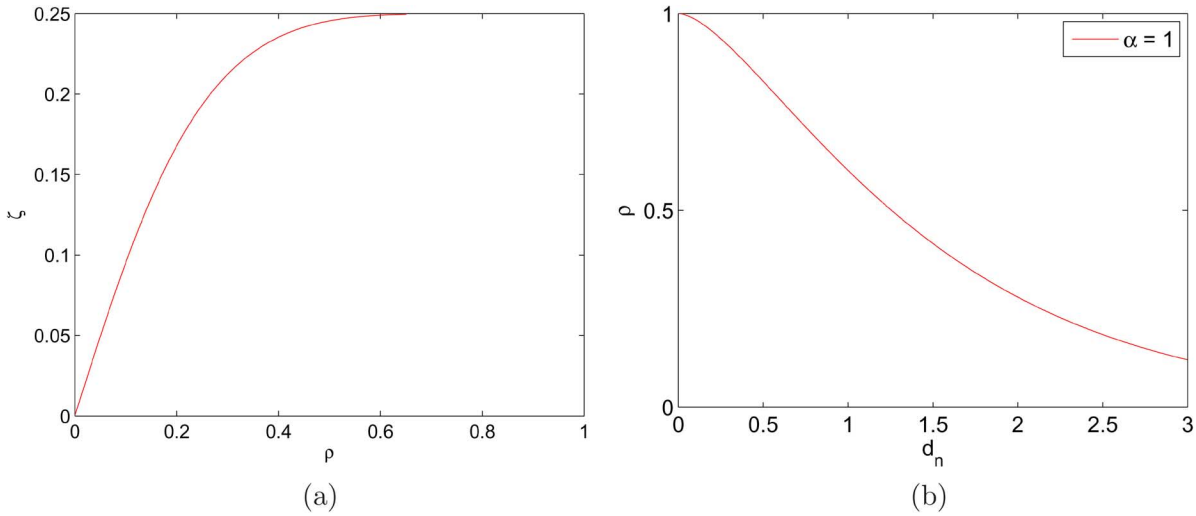

Fig. 7. (a) Edge dependence factor ζ versus edge correlation coefficient ρ and (b) ρ versus edge length d_n.

graph defined by the SFCAR model. The edge correlation coefficient ρ is obtained using the following relationship [21]:

$$\kappa\gamma_{00} = 1 + 4\zeta\kappa\gamma_{01} \Rightarrow \gamma_{01} = \frac{\kappa\gamma_{00} - 1}{4\kappa\zeta} \tag{54}$$

and by substituting (38) and (54) into (53), we have

$$\rho = \frac{(2/\pi)K(4\zeta) - 1}{(2/\pi)(4\zeta)K(4\zeta)} =: g^{-1}(\zeta). \tag{55}$$

Note that the correlation coefficient ρ is not dependent on the power factor κ in (35), as expected, even though γ_{00} and γ_{01} are. Note that function $g^{-1} : \zeta \to \rho$ is a continuous and differentiable C^1 function on the domain $0 \le \zeta \le 1/4$ due to the continuous differentiability of $K(x)$ for $0 \le x < 1$, and $g^{-1}(1) = \lim_{x\to 1} \frac{(2/\pi)K(x)-1}{(2/\pi)xK(x)} = 1$ by $K(1) = \infty$. Note also that $g^{-1}(0) = 0$ since $K(0) = \pi/2$. Thus, the inverse mapping $g : \rho \to \zeta$ from the edge correlation factor ρ to the edge dependence factor ζ, which maps zero and one to zero and $1/4$, respectively, behaves as shown in Fig. 7(a).

Now we consider the correlation coefficient ρ as a function of the sensor spacing d_n. In general, the correlation function $h : d_n \to \rho$ is a positive and monotonically decreasing function of d_n with $h(0) = 1$ and $h(\infty) = 0$. It is well known that for the 1-D first-order AR signal a corresponding underlying (continuous-index) physical model is given by the Ornstein–Uhlenbeck process

$$\frac{ds(x)}{dx} = -As(x) + Bu(x) \tag{56}$$

and its discrete-time equivalent is given by

$$\begin{cases} s_{i+1} = as_i + u_i \\ a = \mathbb{E}\{s_i s_{i-1}\}/\mathbb{E}\{s_i^2\} = e^{-Ad_n} \end{cases} \tag{57}$$

where $A \ge 0$, $B \in \mathbb{R}$, $s_i = s(id_n)$, and the input processes $u(x)$ and u_i are zero-mean white Gaussian processes. Here, d_n is the spacing between two adjacent signal samples. For the 2-D SFCAR signal, however, the same stochastic differential equation is not applicable. Note that the dependence in the signal in (56) and (57) is only on the past in 1-D space, whereas the signal (34) has symmetric dependence in all four direction in the plane. The SFCAR signal is given by the solution of a second-order difference equation

$$X_{ij} = \zeta(X_{i+1,j} + X_{i-1,j} + X_{i,j+1} + X_{i,j-1}) + \epsilon_{ij} \tag{58}$$

and the corresponding continuous-index physical model is given by the stochastic Laplace equation [23]

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 - \alpha^2\right] X(x,y) = \epsilon(x,y) \tag{59}$$

where $\alpha\ (\ge 0)$ is the physical diffusion rate, and ϵ_{ij} and $\epsilon(x,y)$ are 2-D white zero-mean Gaussian perturbations. Note that the solution of (59) is circularly symmetric, i.e., it depends only on $r = \sqrt{x^2+y^2}$, and samples of the solution $X(x,y)$ of (59) on lattice $\mathcal{I}_n$ do not form a discrete-index SFCAR GMRF. However, (59) is still the continuous-index counterpart of (58), and we use its correlation function for the SFCAR model. The correlation function corresponding to (59) is given by [23]

$$\rho = h(d_n) = \alpha d_n K_1(\alpha d_n) \tag{60}$$

where $K_1(\cdot)$ is the modified Bessel function of the second kind. Fig. 7(b) shows the correlation function w.r.t. d_n for $\alpha = 1$. The asymptotic behavior of $K_1(x)$ is given by

$$\begin{cases} K_1(x) \to \sqrt{\frac{\pi}{2x}}e^{-x}, & \text{as } x \to \infty \\ K_1(x) \to \frac{1}{x} \text{ as } x \to 0. \end{cases} \tag{61}$$

The correlation function (60) can be regarded as the representative correlation in 2-D, similar to the exponential correlation function e^{-Ad_n} in 1-D. Both functions decrease monotonically w.r.t. d_n. However, the 2-D correlation function is flat at $d_n = 0$ [23], i.e.,

$$\left(\frac{d\rho}{dd_n}\right)_{d_n=0} = 0, \tag{62}$$

and it decays with rate $\sqrt{d_n}e^{-\alpha d_n}$ as $d_n \to \infty$. Note that the 2-D correlation function has $\sqrt{d_n}$ in front of the exponential

decay as $d_n \to \infty$. However, this polynomial term is not significant and the exponential decay is dominant for large d_n. Thus, we have $\zeta = g(h(d_n))$, and for given physical parameters (with a slight abuse of notation)

$$\mathcal{K}_s(\mathrm{SNR}, \zeta) = \mathcal{K}_s(\mathrm{SNR}, g(h(d_n))) = \mathcal{K}_s(\mathrm{SNR}, d_n)$$

and

$$\mathcal{I}_s(\mathrm{SNR}, \zeta) = \mathcal{I}_s(\mathrm{SNR}, g(h(d_n))) = \mathcal{I}_s(\mathrm{SNR}, d_n).$$

We will use the arguments SNR, ζ, and d_n for $\mathcal{K}_s$ and $\mathcal{I}_s$ properly as needed for exposition.

B. Scaling Laws in Ad Hoc Sensor Networks Over Correlated Random Fields

In this subsection, we investigate the fundamental behavior of wireless flat multihop *ad hoc* sensor networks deployed for statistical inference based on the 2-D hidden SFCAR model and the corresponding correlation functions (55) and (60). We consider several criteria for determining the efficiency of the sensor network. Specifically, we consider the total amount of information [nats] obtainable from the network and the energy efficiency η of a sensor network, defined as

$$\eta = \frac{\text{total gathered information } I_t}{\text{total required energy } E_t} \text{ [nats/J]} \tag{63}$$

where the gathered information is about the underlying physical process.

In the following, we summarize the assumptions for the planar *ad hoc* sensor network that we consider.

(A.1) n^2 sensors are located on the grid $\mathcal{I}_n$ with spacing d_n, as shown in Fig. 2, and a fusion center is located at the center $(\lfloor n/2 \rfloor, \lfloor n/2 \rfloor)$. The network size is $L \times L$, where $L = nd_n$. Thus, the node density μ_n on $\mathcal{I}_n$ is given by

$$\mu_n = \frac{n^2}{L^2} = \frac{n^2}{(nd_n)^2}. \tag{64}$$

(A.2) The observations $\{Y_{ij}\}$ of sensor nodes form a 2-D hidden (discrete-index) SFCAR GMRF on the lattice for each $d_n > 0$, and the edge dependence factor is given by the correlation functions (55) and (60).

(A.3) The fusion center gathers the measurements from all nodes using minimum hop routing. Note that the links in Fig. 2 are not only the Markov dependence edges but also the routing links. The minimum hop routing requires a hop count of $|i - \lfloor n/2 \rfloor| + |j - \lfloor n/2 \rfloor|$ to deliver Y_{ij} to the fusion center.

(A.4) The communication energy per link is given by $E_c(d_n) = E_0 d_n^{\nu}$, where $\nu \geq 2$ is the propagation loss factor of the wireless channel.

(A.5) Sensing requires energy, and the sensing energy per node is denoted by E_s. Moreover, we assume that the *measurement* SNR in (14) is linearly increasing w.r.t. E_s, i.e., $\mathrm{SNR} = \beta E_s$ for some constant β.

Remark 1: Assumption *(A.2)* facilitates the analysis. Since discrete samples of a continuous-index GMRF do not form a discrete-index GMRF almost surely, we assume that for each d_n sensor samples on $\mathcal{I}_n$ form a discrete-index SFCAR GMRF, and match the correlation between two neighboring nodes with the physically meaningful correlation function (60).

Remark 2: In Assumption *(A.3)*, we assume that there is no data fusion during the information gathering, i.e., no in-network data fusion. The fusion center collects the raw measurements from all sensors.

Remark 3: We can also consider a routing graph different from the Markov dependence graph in Fig. 2. For example, sensors not directly connected to the transmitting node via the Markov dependence edge can deliver the data to the fusion center. However, this results in a reduced number of hops with a larger hop length, and the corresponding routing path consumes more energy. Thus, Assumption *(A.3)* of minimum hop routing via the Markov dependence edge ensures least energy consumption with a minimum hop routing strategy.

Remark 4: Assumption *(A.5)* does not imply that we can increase the power of the underlying signal, but it means that we can increase the SNR of effective sensor samples. Suppose that E_1 joules are required for one sensing to obtain one sample $Y_{ij}(1) = X_{ij}(1) + W_{ij}(1)$ at location ij and the measurement SNR of this sample is SNR_1. Now assume that we have M identical subsensors at location ij and obtain M samples with one sample per each subsensor, requiring $M \cdot E_1$ joules, and we take an average of M samples at location ij, yielding $Y_{ij} = (1/M)\sum_{m=1}^{M} Y_{ij}(m)$, where $Y_{ij}(m)$ denotes the sample at the mth subsensor at location ij. The measurement SNR of the effective sample Y_{ij} is given by $M \cdot \mathrm{SNR}_1$ assuming that the measurement noise is i.i.d. across the subsensors. Thus, the effective measurement SNR at each sensor can be increased linearly w.r.t. the sensing energy. However, this linear SNR model is an optimistic assumption since the observation SNR may saturate as the sensing energy is increased without bound in practical situation.

From here on, we consider various asymptotic scenarios and investigate the fundamental behavior of *ad hoc* sensor networks deployed over correlated random fields for statistical inference under Assumptions *(A.1)–(A.5)*. Our asymptotic analysis in the previous sections enables us to calculate the total information I_t for large sensor networks. The total amount of information is given approximately by the product of the number of sensor nodes in the network and the asymptotic per-node information $\mathcal{K}_s$ or $\mathcal{I}_s$, i.e.,

$$I_t = n^2 \mathcal{K}_s(\mathrm{SNR}, d_n) \quad \text{or} \quad I_t = n^2 \mathcal{I}_s(\mathrm{SNR}, d_n) \tag{65}$$

for KLI or MI, respectively. The total energy E_t required for data gathering via the minimum hop routing is given by

$$\begin{aligned} E_t &= n^2 E_s + E_c(d_n) \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (|i - \lfloor n/2 \rfloor| + |j - \lfloor n/2 \rfloor|), \\ &= \begin{cases} n^2 E_s + \frac{1}{2} n(n-1)(n+1) E_c(d_n), & \text{if } n \text{ odd} \\ n^2 E_s + \frac{1}{2} n^3 E_c(d_n), & \text{if } n \text{ even.} \end{cases} \end{aligned} \tag{66}$$

First, we consider an infinite area model with fixed density. In this case, the number of sensor nodes per unit area is fixed and the total area increases without bound as we increase n. The behavior of the information versus area and energy in this case is given in the following theorem.

Theorem 4 (Fixed Density and Infinite Area): For an *ad hoc* sensor network with a fixed and finite node density and fixed sensing energy per node, the total amount of information increases linearly w.r.t. area, but the amount of gathered information per unit energy decays to zero with rate

$$\eta = \Theta\left(\text{area}^{-1/2}\right) \tag{67}$$

for any nontrivial diffusion rate α, i.e., $0 < \alpha < \infty$, as we increase the area. Further, in this case the total amount of information obtainable from the network as a function of total consumed energy increases with rate of

$$\text{total information } I_t = \Theta\left(E_t^{2/3}\right) \tag{68}$$

for any propagation loss factor $\nu > 0$, as the total energy E_t consumed by the network increases without bound, i.e., $E_t \to \infty$.

Proof: See Appendix I.

Theorem 4 enables us to investigate the asymptotic behavior of *ad hoc* sensor networks with fixed available energy per node. From the detection perspective the error probability is given by

$$P_M \sim e^{-I_t(E_t(N_t(A)))} \tag{69}$$

for large networks, where $N_t(A)$ represents the total number of sensor nodes in the network with coverage area A. Now consider that each node has a fixed amount of energy denoted by $\bar{E}(< \infty)$. Then, the total energy in the network is given by

$$E_t = N_t(A)\bar{E}. \tag{70}$$

Note in this case that the total energy available in the network increases linearly w.r.t. the number of sensor nodes. The asymptotic behavior of *ad hoc* networks with fixed per-node energy is given by the following corollary to Theorem 4.

Corollary 3: For an *ad hoc* sensor network with a fixed and finite node density, the information amount per sensor node diminishes to zero as the network size grows, i.e.,

$$\lim_{N_t(A)\to\infty} -\frac{1}{N_t(A)} \log P_M(E_t(N_t(A))) = \lim_{N_t(A)\to\infty} O(N_t(A)^{-1/3}) = 0 \tag{71}$$

if each sensor has a finite amount of available energy.

Proof: Substitute (68), (69), and (70) into I_t, P_M, and E_t, respectively.

Corollary 3 states that a nonzero per-node information is not achievable as the coverage increases without in-network data fusion in the case that each node has only a fixed amount of energy, which is the case in most network design with fixed amount of battery. In this case, the per-node information scales with $O(N_t^{-1/3})$ as the network size grows. This result is by the communication energy required for *ad hoc* routing without in-network data fusion. Note from (66) that for the fixed density and increasing area model the sensing energy increases quadratically with n while the communication energy without in-network data fusion increases cubically with n since d_n is fixed w.r.t. n. Hence, for *ad hoc* sensor networks with large coverage areas, the communication energy dominates the sensing energy, and both the energy efficiency for information and the per-node information under fixed per-node energy constraint diminish to zero because of the slower increasing rate of the total information amount than that of the communication energy required for *ad hoc* routing without in-network data fusion.

This diminishing energy efficiency and per-node information under fixed per-node energy constraint can be fixed with in-network data fusion. Suppose that in-network data fusion is performed so that each node needs to deliver (aggregated) data only to the neighboring node along the minimum hop route to the fusion center in Fig. 2. In this case, the number of transmission associated with one node is just one and the total number of transmissions in the network is given by $\Theta(n^2)$. So, the communication energy as well as the sensing energy increases quadratically with n. Since the total amount of information also increases quadratically with n, the total amount of information as a function of total energy is given, under this aggregation scenario, by

$$I_t = \Theta(E_t), \tag{72}$$

as we increase the area, and a nonzero energy efficiency and a nonzero per-node information under fixed per-node energy constraint are achieved. Thus, in-network data fusion is essential for energy efficiency in large sensor networks.

Next, we consider the case in which the node density diminishes, i.e., $d_n \to \infty$. Especially, this case is of interest at high SNR since at high SNR less correlated samples yield larger per-node information, as seen in Section III-B1. However, the per-node information is upper-bounded as $d_n \to \infty$, and the asymptotic behavior is given by the following theorem.

Theorem 5: As $d_n \to \infty$, the per-node information $\mathcal{K}_s$ and $\mathcal{I}_s$ converge to $D(\mathcal{N}(0,1)\|\mathcal{N}(0,1+\text{SNR}))$ and $\frac{1}{2}\log\text{SNR}$, respectively, and the convergence rates are given by

$$\mathcal{K}_s(d_n) = D(\mathcal{N}(0,1)\|\mathcal{N}(0,1+\text{SNR})) - c_4\sqrt{d_n}e^{-\alpha d_n} + o\left(\sqrt{d_n}e^{-\alpha d_n}\right) \tag{73}$$

and

$$\mathcal{I}_s(d_n) = \frac{1}{2}\log(1+\text{SNR}) - c_4'\sqrt{d_n}e^{-\alpha d_n} + o\left(\sqrt{d_n}e^{-\alpha d_n}\right) \tag{74}$$

with positive constants c_4 and c_4'.

Proof: See Appendix I.

Theorem 5 explains how much gain in information is obtained from less correlated observation samples by making the sensor spacing larger. Fig. 8 shows the per-node KLI $\mathcal{K}_s$ and the communication energy E_c for each link as functions of d_n for $\alpha = 1$, $c_4 = 1$, and 10-dB SNR. The gain in information is given by $\sqrt{d_n}e^{-\alpha d_n}$ for large d_n, whereas the required per-link communication energy increases without bound, i.e., $E_c(d_n) = E_0 d_n^{\nu}$ $(\nu \geq 2)$. Since the exponential term is dominant in the gain as d_n increases, the information gained by increasing the sensor spacing d_n decreases almost exponentially fast, and no significant gain is obtained by increasing the sensor spacing further after some point. Hence, it is not effective, in

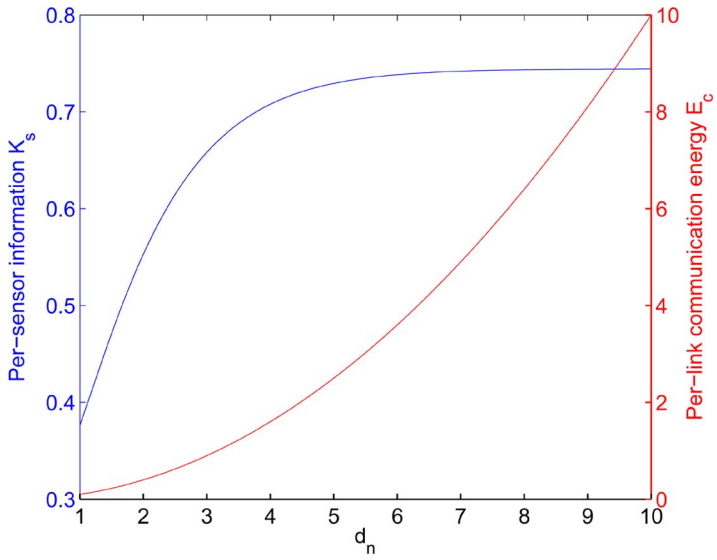

Fig. 8. Per-node information and per-link communication energy w.r.t. sensor spacing d_n (SNR = 10 dB, $\alpha = 1$, $c_4 = 1$).

terms of energy efficiency, to increase the sensor spacing too much to obtain less correlated samples at high SNR.

From Theorem 5 we have seen that increasing the sensor spacing is not so effective in terms of the information gain per unit of consumed energy since the per-link communication energy increases without bound. On the other hand, the per-link communication energy can be made arbitrarily small by decreasing the sensor spacing. To investigate the effect of diminishing communication energy E_c as $d_n \to 0$, we now consider the asymptotic case in which the node density goes to infinity for a fixed coverage area. In this case, the per-node information decays to zero as $d_n \to 0$ since $\zeta \to 1/4$ as $d_n \to 0$, and $\mathcal{K}_s(\zeta)$ and $I_s(\zeta)$ converge to zero as $\zeta \to 1/4$, as shown in Section III-B1. The asymptotic behavior in this case is given by the following theorem.

Theorem 6 (Infinite Density Model): For the infinite density model with a fixed coverage area S with nontrivial diffusion rate α, the per-node information decays to zero with convergence rate

$$\mathcal{K}_s = c_5 \mu_n^{-1} + o\left(\mu_n^{-1}\right) \tag{75}$$

for some constant c_5 as the node density $\mu_n \to \infty$. Hence, the amount of total information from the coverage area converges to the constant $c_5 S$ as $\mu_n \to \infty$. Furthermore, in the case of no sensing energy, a nonzero energy efficiency η is achievable if the propagation loss factor $\nu = 3$, and even an infinite energy efficiency[6] is achievable under Assumption *(A.4)* if $\nu > 3$ as $\mu_n \to \infty$. $\mathcal{I}_s$ has similar behavior.

Proof: See Appendix I.

Remark 5: The finite total information for the infinite density and fixed-area model follows our intuition. The maximum information provided by the samples from the continuous-index random field does not exceed the information between $X(x,y)$ and $Y(x,y)$ except in the case of spatially white fields. Here, the relevance of (62) in 2-D is evident. From (62) we have

$$\mathcal{K}_{s,2-D}(\zeta(\rho(d_n))) = c_6 \cdot d_n^2 + o(d_n^2) \tag{76}$$

as $d_n \to 0$ since $h : d_n \to \zeta$ has slope zero at $d_n = 0$ and $\mathcal{K}_s$ is a continuous and differentiable function of ζ. In the 1-D case, it is shown in [12] that $\mathcal{K}_{s,1-D}$ is also a continuous and differentiable function of $a = e^{-Ad_n}$ for $0 \le a \le 1$ with $\mathcal{K}_{s,1-D}|_{a=1} = 0$. However, the exponential correlation e^{-Ad_n} has a nonzero slope at $d_n = 0$, and thus we have

$$\mathcal{K}_{s,1-D}(a(d_n)) = c_6' \cdot d_n + o(d_n) \tag{77}$$

as $d_n \to 0$. The number of nodes in the space is given by $\Theta(n^2)$ and $\Theta(n)$ for 2-D and 1-D, respectively, and $d_n = L/n$ in both cases. Hence, the total amount of information from the coverage space (given by the product of the per-node information and the number of nodes in the space) converges to a constant both in 1-D and 2-D as the node density increases. Thus, any proper 2-D correlation function w.r.t. the sample distance should have a flat top at a distance of zero.

Remark 6: It is common that the propagation loss factor $\nu > 3$ for near-field propagation (i.e., $d_n \to 0$). Hence, infinite energy efficiency is theoretically achievable under Assumption *(A.4)* as we increase the node density for a fixed area assuming that only communication energy is required. Note that the total amount of information converges to a constant as we increase the node density. So, the infinite energy efficiency is achieved by diminishing communication energy as $d_n \to 0$.

Remark 7: Considering the sensing energy, infinite energy efficiency is not feasible even theoretically since we have in this case

$$E_t = n^2 E_s + \Theta(n^{3-\nu}) \tag{78}$$

and

$$\eta = \frac{c_5 S + o(1)}{n^2 E_s + \Theta(n^{3-\nu})}, \quad \nu \ge 2 \tag{79}$$

[6]Of course, this is under Assumption *(A.4)* for any $d_n > 0$. In reality, Assumption *(A.4)* is valid for $d_n \ge d_{min}$ for some $d_{min} > 0$.

as $n \to \infty$ for fixed coverage area. In this case, the sensing energy $n^2 E_s$ is the dominant factor for low energy efficiency, and the energy efficiency decreases to zero with rate $O\left(\mu_n^{-1}\right)$. Thus, it is critical for densely deployed sensor networks to minimize the sensing energy or processing energy for each sensor.

In the infinite density model, we have observed that energy is an important factor in efficiency. Now we investigate the change of total information w.r.t. energy. There are many possible ways to invest energy in the network. One simple way is to fix the node density and coverage area and to increase the sensing energy. We assume that the network size is sufficiently large so that our asymptotic analysis is valid. The energy-asymptotic behavior in this case is given in the following theorem under Assumptions *(A.1)–(A.5)*.

Theorem 7: As we increase the total energy E_t consumed by a sensor network (including both sensing and communication) with a fixed node density and fixed area, the total information increases with rate

$$\text{total information } I_t = O\left(\log E_t\right) \tag{80}$$

as $E_t \to \infty$.

Proof: See Appendix I.

Theorem 7 suggests a guideline for investing the excess energy. It is not efficient in terms of the total amount of gathered information to invest energy to improve the quality of sensed samples from a limited area. This only provides an increase in total information at a logarithmic rate. Note in Theorem 4 that the information gain is given by

$$I_t = \Theta\left(E_t^{2/3}\right) \tag{81}$$

as we increase the coverage area with fixed density and sensing energy even without in-network data fusion. Thus, the energy should be spent to increase the number of samples by enlarging the coverage area even if it yields less accurate samples. In this way, we can achieve the information increase with rate at least $\Theta\left(E_t^{2/3}\right)$ which is much faster than the logarithmic increase obtained by increasing the sensing energy.

C. Optimal Node Density

In the previous subsection, we investigated the asymptotic behavior of the total information obtainable from the network and the energy efficiency as the coverage, density, or energy change. We now consider another important problem in sensor network design for statistical inference about underlying random fields, namely, the optimal density problem. Here, we are given a fixed coverage area, and are interested in determining an optimal node density. The total amount of information gathered from the network increases monotonically (even if it has an upper bound) as we increase the node density, as shown in Theorem 6. Hence, the problem cannot be properly formulated without some constraint. We consider a total energy constraint in which a fixed amount of energy is available to the entire network for both sensing and communication. Thus, we consider the following problem.

Problem 1 (Optimal Density): Given a fixed coverage area with size $L \times L$ and total available energy E_t, find the density μ_n that maximizes the total information I_t obtainable from the sensor network.

The above optimization problem can be solved using our analysis based on the large deviations principle assuming the asymptotic result is still valid in the low density case, and the optimal density for the KLI measure is given by

$$\mu_n^* = \arg\max_{\mu_n} L^2 \mu_n \mathcal{K}_s(\mathrm{SNR}(E_t, \mu_n), d_n(\mu_n)) \tag{82}$$

$$\text{s.t. } n^2 E_s(\mu_n) + \frac{1}{2} n(n-1)(n+1) E_c(d_n(\mu_n)) \le E_t \tag{83}$$

where the sensing energy E_s as well as n and d_n are functions of the node density μ_n. From μ_n $(= n^2/L^2)$, we first calculate n and then $d_n = L/n$. (Here, the quantization of n to the nearest integer is not performed.) With the determined d_n, $E_c(d_n)$ is obtained from the propagation parameters E_0 and ν, and then $E_s(\mu_n)$ is obtained from the constraint (83). When $E_s(\mu_n)$ is determined, the measurement SNR is calculated using Assumption *(A.5)*, i.e., $SNR = \beta E_s$, and finally, we evaluate the per-node information $\mathcal{K}_s(\mathrm{SNR}, \zeta(\rho(d_n)))$ and $\mathcal{I}_s(\mathrm{SNR}, \zeta(\rho(d_n)))$ from Corollary 2.

Fig. 9 shows the total information obtainable from a 2-m $\times$ 2–m area as we change the node density μ_n with a fixed total energy budget of E_t joules. Other parameters that we use are given by

$$\alpha = 100, \quad \beta = 1, \quad E_0 = 0.1, \quad \text{and } \nu = 2.$$

Here, the values of E_t, E_0, and β are selected so that the minimum and maximum per-node sensing SNRs are roughly -10 to 10 dB for maximum and minimum densities, respectively. The diffusion rate $\alpha = 100$ is chosen for the edge correlation coefficient ρ to range from almost zero to 0.6 as the node density varies. It is seen in the figure that there is an optimal density for each value of E_t under either information measure. It is also seen that the total KLI is sensitive to the density change whereas the total MI is less sensitive. The existence of the optimal density is explained as follows. At low densities, we have only a few sensors in the area. So, the energy for communication is not large due to the small number of communicating nodes (see (108) below) and most of the energy is allocated to sensing. Here, the per-node sensing energy is even higher due to the small number of sensors. However, the per-node information increases only logarithmically w.r.t. the sensing energy or SNR by Theorem 7, and this logarithmic gain cannot compensate for the loss in the number of sensors. Hence, low density yields very poor performance, and large gain is obtained initially as we increase the density from very low values, as seen in Fig. 9. As we further increase the density, on the other hand, the per-node sensing energy or SNR decreases due to the increase in the overall communication and the increase in the number of sensor nodes, and the measurement SNR is in the low-SNR regime eventually, where (49) and (51) hold. From (66), we have

$$E_s(\mu_n) = \beta^{-1}\mathrm{SNR} = O(n^{-2}) \tag{84}$$

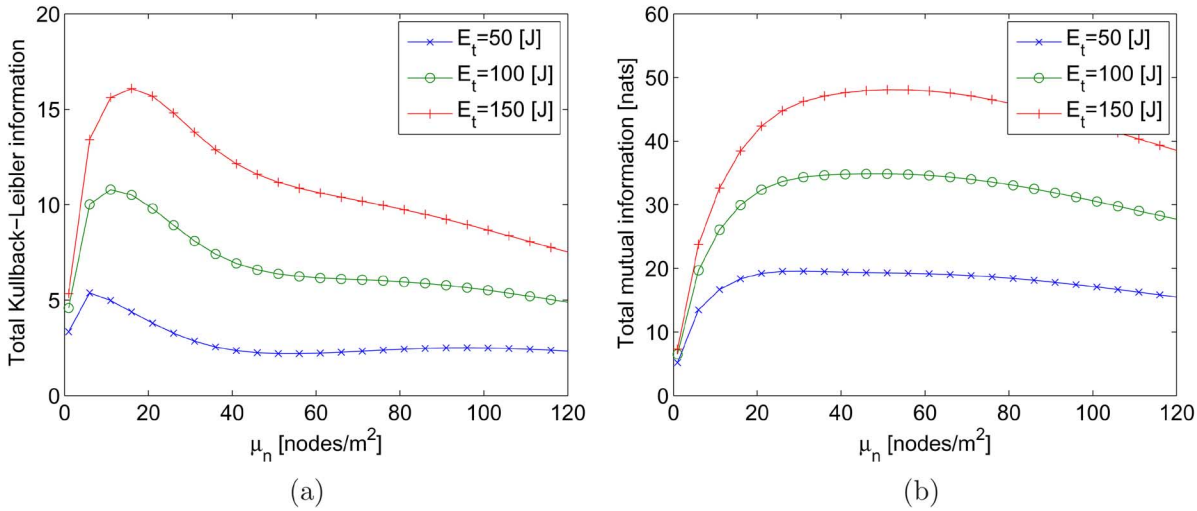

Fig. 9. (a) Total KLI versus density and (b) total MI versus density.

for fixed E_t and $E_c = E_0(L/n)^2$, as $n \to \infty$. By the quadratic decaying behavior of $\mathcal{K}_s$ at low SNR given by (49), the total KLI is given by

$$\text{total KLI} = L^2 \mu_n \mathcal{K}_s = O(n^2 n^{-4}) = O(n^{-2}) = O(\mu_n^{-1}).$$

By (51), on the other hand, the mutual information decays linearly as SNR decreases to zero, and the total MI is given by

$$\text{total MI} = L^2 \mu_n \mathcal{I}_s = O(n^2 n^{-2}) = O(1).$$

This explains the initial fast decay after the peak in Fig. 9(a) and flat curve in Fig. 9(b). In the above equations, however, the effect of ζ on $\mathcal{K}_s$ and $\mathcal{I}_s$ is not considered. As the node density increases, the sensor spacing decreases and the edge dependence factor ζ increases for a given diffusion rate α. The behavior of the per-node information as a function of ζ is shown in Fig. 4. Note in Fig. 4 that the per-node information has a second lobe at strong correlation at low SNR while at high SNR it decreases monotonically as the correlation becomes strong. The benefit of sample correlation is evident in the low energy case ($E_t = 50$ [J]) in Fig. 9(a); the second peak around $\mu_n = 95$ [nodes/m^2] is observed. Note that the second peak is not very significant. Since the per-node information decays to zero as $\zeta \to 1/4$ eventually, the total amount of information decreases eventually, as seen in the right corner of the figure, as we increases the node density further.

V. Conclusion and Discussion

In this paper, we have considered the design of sensor networks for statistical inference about correlated random fields in a 2-D setting. To quantify the information from the sensor network, we have used a spectral domain approach to derive closed-form expressions for asymptotic KLI and MI rates in general d-D and in 2-D in particular, and have adopted the 2-D hidden CAR GMRF for our signal model to capture the spatial correlation and measurement noise for samples in a 2-D sensor field. Under the first-order symmetry assumption, we have further obtained the asymptotic information rates explicitly in terms of the SNR and the edge dependence factor, and have investigated the properties of the asymptotic information rates as functions of SNR and correlation. Based on these LDP results, we have then analyzed the asymptotic behavior of *ad hoc* sensor networks deployed over 2-D correlated random fields for statistical inference. Under the SFCAR GMRF model, we have obtained fundamental scaling laws for total information and energy efficiency as the coverage, node density, and consumed energy change. The results provide guidelines for sensor network design for statistical inference about 2-D correlated random fields such as temperature, humidity, or density of a gas on a certain area.

In closing, we discuss several issues related to some of the assumptions we have used to simplify our analysis. First, of course, sensors in a real network may not be located on a 2-D grid. However, we conjecture that similar scaling behaviors w.r.t. the coverage, density and energy are valid for randomly and uniformly deployed sensors. Second, the spatial Markov assumption may be restrictive. However, it is a minimal model that captures the two-dimensionality of the signal correlation structure in all planar directions and allows analysis to be tractable. And, finally, we have not considered the temporal evolution of the spatial signal field. In case of i.i.d. temporal variation, the results here can be applied directly without modification. When the signal variation over time is correlated, the modification to spatiotemporal fields is required.

Appendix I

Proof of Theorem 2:

The asymptotic KLI rate $\mathcal{K}$ is given by the almost-sure limit

$$\mathcal{K} = \lim_{n\to\infty} \frac{1}{|\mathcal{D}_n|} \log \frac{p_0}{p_1}(\{Y_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{D}_n\}) \tag{85}$$

evaluated under p_0 [24]. We consider the following index mapping from d-D to 1-D in lexicographic order:

$$l = f_{id}(\boldsymbol{i}) \quad (\boldsymbol{i} \in [0, 1, \ldots, n-1]^d) \tag{86}$$

and the corresponding observation vector $\boldsymbol{y}_{|\mathcal{D}_n|}$ generated from $\{Y_{\boldsymbol{i}}, \boldsymbol{i} \in \mathcal{D}_n\}$. Then, $\boldsymbol{y}_{|\mathcal{D}_n|}$ is a zero-mean Gaussian vector with

the covariance matrices $\mathbf{\Sigma}_{0,|\mathcal{D}_n|}$ and $\mathbf{\Sigma}_{1,|\mathcal{D}_n|}$ under p_0 and p_1, respectively. Hence, the asymptotic KLI rate is given by

$$\mathcal{K} = \lim_{n\to\infty} \frac{1}{|\mathcal{D}_n|}\left(\frac{1}{2}\log\frac{\det(\mathbf{\Sigma}_{1,|\mathcal{D}_n|})}{\det(\mathbf{\Sigma}_{0,|\mathcal{D}_n|})} + \frac{1}{2}\boldsymbol{y}_{|\mathcal{D}_n|}^T\left(\mathbf{\Sigma}_{1,|\mathcal{D}_n|}^{-1} - \mathbf{\Sigma}_{0,|\mathcal{D}_n|}^{-1}\right)\boldsymbol{y}_{|\mathcal{D}_n|}\right) \quad (87)$$

under p_0. Now we consider the terms on the RHS of (87). First, we consider $\log\det(\mathbf{\Sigma}_{0,|\mathcal{D}_n|})$. Since $\mathbf{\Sigma}_{0,|\mathcal{D}_n|} = \sigma^2\boldsymbol{I}_{n^d}$ under the assumption of an i.i.d. null distribution, we simply have

$$\frac{1}{|\mathcal{D}_n|}\log\det\mathbf{\Sigma}_{0,|\mathcal{D}_n|} = \frac{1}{n^d}\log\det(\sigma^2\boldsymbol{I}_{n^d}) = \log\sigma^2. \quad (88)$$

Next, we consider the term $\frac{1}{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T\mathbf{\Sigma}_{0,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}$. Since $\boldsymbol{y}_{|\mathcal{D}_n|}$ is i.i.d. Gaussian, d-D is irrelevant in this case, the known result from [25, Proposition 10.8.3] is applicable, and we have

$$\frac{1}{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T\mathbf{\Sigma}_{0,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|} \to 1 \text{ almost surely} \quad (89)$$

assuming that the random vector $\boldsymbol{y}_{|\mathcal{D}_n|}$ is generated from the distribution p_0. Now we consider the term $\frac{1}{|\mathcal{D}_n|}\log\det\mathbf{\Sigma}_{1,|\mathcal{D}_n|}$. This is the entropy rate of a d-D Gaussian process, and the convergence behavior of this term is studied in [18]. It is shown in [18, p. 391] under the assumption in Theorem 2 that we have

$$\left|\log\det\mathbf{\Sigma}_{1,|\mathcal{D}_n|} - \frac{|\mathcal{D}_n|}{(2\pi)^d}\int_{[-\pi,\pi)^d}\log((2\pi)^d f_1(\boldsymbol{\omega}))d\boldsymbol{\omega}\right| = O\left(\frac{|\mathcal{D}_n|}{n}\right).$$

Applying this result, we have

$$\frac{1}{|\mathcal{D}_n|}\log\det\mathbf{\Sigma}_{1,|\mathcal{D}_n|} \to \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d}\log((2\pi)^d f_1(\boldsymbol{\omega}))d\boldsymbol{\omega}. \quad (90)$$

Finally, we consider the random term $\frac{1}{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T\mathbf{\Sigma}_{1,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}$.[7] By Lemma 2 in Appendix II, we have

$$\frac{1}{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T\mathbf{\Sigma}_{1,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|} \to \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d}\frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})}d\boldsymbol{\omega}, \quad (91)$$

almost surely as $n\to\infty$.

Combining (87)–(91), we have

$$\mathcal{K} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d}\left[\frac{1}{2}\log\frac{(2\pi)^d f_1(\boldsymbol{\omega})}{\sigma^2} - \frac{1}{2}\left(1 - \frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})}\right)\right]d\boldsymbol{\omega}. \quad (92)$$

Since

$$D(\mathcal{N}(0,\sigma_0^2)||\mathcal{N}(0,\sigma_1^2)) = \frac{1}{2}\log\frac{\sigma_1^2}{\sigma_0^2} - \frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) \quad (93)$$

equation (92) is given by

$$\mathcal{K} = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} D(\mathcal{N}(0,\sigma^2)||\mathcal{N}(0,(2\pi)^d f_1(\boldsymbol{\omega})))d\boldsymbol{\omega}. \quad (94)$$

□

[7]The proof given in [25] and [26] for the convergence of this term for the 1-D index case is not applicable for general d-D, nor is the almost-sure convergence of the term shown in [18], where the convergence of the term in probability to an integral involving the periodogram was shown. Thus, we prove the almost-sure convergence of the term in Lemma 2 separately in Appendix II.

Proof of Corollary 1:

For the 2-D hidden model we have

$$f_1(\omega_1,\omega_2) = (2\pi)^{-2}\sigma^2 + f(\omega_1,\omega_2) \quad (95)$$

where $f(\omega_1,\omega_2)$ is the CAR spectrum (11) in 2-D satisfying (12) and (13). First, $f_1(\omega_1,\omega_2)$ has a positive lower bound, and thus satisfies Assumption *A.1* in Theorem 2. It is also known in [27] that if $\boldsymbol{k} = (k_1,\ldots,k_d) \in \mathbb{N}^d$ and if $f_1(\boldsymbol{\omega})$ is of class $C^{\boldsymbol{k}}$ (i.e., differentiable up to the k_d-order w.r.t. ω_d), then

$$\limsup_{\boldsymbol{h}\to\infty} h_1^{k_1}h_2^{k_2}\ldots h_d^{k_d}|\gamma_{\boldsymbol{h}}| < \infty \quad (96)$$

where $\mathbb{N}$ is the set of all natural numbers, and $\boldsymbol{h}\to\infty$ means that at least one coordinate tends to infinity. Under the condition (12) and (13), the hidden CAR spectrum $f_1(\omega_1,\omega_2)$ in (95) is $C^{(\infty,\infty)}$, i.e., smooth both in ω_1 and ω_2. This ensures that Assumption *A.2* in Theorem 2 is satisfied, and the corollary follows by substituting (95) and $d=2$ into (26). □

Proof of Theorem 3:

The continuity is straightforward. The monotonicity is shown as follows. Let $s = 1 + \mathrm{SNR}g_\zeta(\boldsymbol{\omega})$ where

$$g_\zeta(\boldsymbol{\omega}) = ((2/\pi)K(4\zeta)(1 - 2\zeta\cos\omega_1 - 2\zeta\cos\omega_2))^{-1}.$$

Then, the partial derivative of $\mathcal{K}_s$ w.r.t. SNR is given by

$$\frac{\partial\mathcal{K}_s}{\partial\mathrm{SNR}} = \frac{1}{(2\pi)^2}\times\int_{\boldsymbol{\omega}\in[-\pi,\pi)^2}\frac{\partial}{\partial s}\left(\frac{1}{2}\log s + \frac{1}{2s} - \frac{1}{2}\right)\frac{\partial s}{\partial\mathrm{SNR}}d\boldsymbol{\omega} \quad (97)$$

where

$$\frac{\partial}{\partial s}\left(\frac{1}{2}\log s + \frac{1}{2s} - \frac{1}{2}\right) = \frac{1}{2}\frac{s-1}{s^2} = \frac{1}{2}\frac{\mathrm{SNR}g_\zeta(\boldsymbol{\omega})}{s^2} \geq 0, \quad (98)$$

and

$$\frac{\partial s}{\partial\mathrm{SNR}} = g_\zeta(\boldsymbol{\omega}) \geq 0 \quad (99)$$

for $0\leq\zeta\leq 1/4$. Hence,

$$\frac{\partial\mathcal{K}_s}{\partial\mathrm{SNR}} \geq 0, \quad (100)$$

and $\mathcal{K}_s$ increases monotonically as SNR increases for a given ζ $(0\leq a\leq 1/4)$.

As SNR $\to\infty$, we have

$$\mathcal{K}_s \approx \frac{1}{(2\pi)^2}\int_{\boldsymbol{\omega}\in[-\pi,\pi)^2}\frac{1}{2}\log(\mathrm{SNR}g_\zeta(\boldsymbol{\omega}))d\boldsymbol{\omega},$$

$$= \frac{1}{2}\log \mathrm{SNR} + \frac{1}{(2\pi)^2}\int_{\boldsymbol{\omega}\in[-\pi,\pi)^2} \frac{1}{2}\log(g_\zeta(\boldsymbol{\omega}))d\boldsymbol{\omega}.$$

Thus, we have $\frac{1}{2}\log \mathrm{SNR}$ behavior at high SNR.

For (49) and (51), take the Taylor expansion around $\mathrm{SNR} = 0$ to obtain

$$\log(1+\mathrm{SNR}g_\zeta(\boldsymbol{\omega})) = \mathrm{SNR}g_\zeta(\boldsymbol{\omega}) - \mathrm{SNR}^2 g_\zeta^2(\boldsymbol{\omega})/2 + \cdots,$$
$$\frac{1}{1+\mathrm{SNR}g_\zeta(\boldsymbol{\omega})} = 1 - \mathrm{SNR}g_\zeta(\boldsymbol{\omega}) + \mathrm{SNR}^2 g_\zeta^2(\boldsymbol{\omega}) - \cdots$$

and then integrate. □

Proof of Theorem 4:

In this case, the edge length $d_n = d$ for all n, and thus the asymptotic per-sensor information $\mathcal{K}_s(d_n)$ or $\mathcal{I}_s(d_n)$ does not change with n. Considering the KLI , we have $I_t = n^2\mathcal{K}_s(d)$, and $\mathrm{area} = \Theta(n^2)$. Hence, the total information is linear w.r.t. area. The total energy E_t required for data gathering is given by

$$E_t = n^2 E_s + E_c(d)\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(|i - \lfloor n/2\rfloor| + |j - \lfloor n/2\rfloor|)$$
$$= n^2 E_s + \Theta(n^3)E_c(d) \tag{101}$$

where the first term is the sensing energy and the second term is the energy consumed for communication. The energy efficiency is given by

$$\eta = \frac{n^2\mathcal{K}_s(d)}{n^2 E_s + \Theta(n^3)E_c(d)} = \Theta\left(\frac{1}{n}\right) \tag{102}$$

as $n \to \infty$. Since $\mathrm{area} = \Theta(n^2)$, (67) follows.

For the second statement, we have $E_t = \Theta(n^3)$. The total information is given by $n^2\mathcal{K}_s(\mathrm{SNR}, d)$. Since $\mathcal{K}_s$ is fixed, the total information is $\Theta(n^2)$ as $n \to \infty$, and we have (68). □

Proof of Theorem 5:

The proof is by the asymptotic behavior of the modified Bessel function $K_1(\cdot)$ of the second kind and Taylor expansion of $\mathcal{K}_s$ (as a function of ζ) and ζ (as a function of ρ), which is allowed because of their continuous differentiability. From (60) and (61) we have

$$\rho(d_n) = \sqrt{\frac{\pi}{2}}\alpha d_n e^{-\alpha d_n} + o\left(\alpha d_n e^{-\alpha d_n}\right) \tag{103}$$

as $d_n \to \infty$. From the continuous differentiability of $\mathcal{K}_s$ as a function of ζ in (47) and ζ as a function of ρ, we have

$$\mathcal{K}_s = D(\mathcal{N}(0,1)||\mathcal{N}(0,1+\mathrm{SNR})) - c_2\zeta + o(\zeta),$$
$$= D(\mathcal{N}(0,1)||\mathcal{N}(0,1+\mathrm{SNR})) - c_2(c_7\rho + o(\rho)) + o(c_7\rho + o(\rho)),$$
$$= D(\mathcal{N}(0,1)||\mathcal{N}(0,1+\mathrm{SNR})) - c_2c_7\rho + o(\rho),$$

for some $c_2, c_7 > 0$. Applying (103) to the above equation, we have (73). The proof for the mutual information $\mathcal{I}_s$ is similar. □

Proof of Theorem 6:

Consider a fixed area with size $L \times L$ and a lattice $\mathcal{I}_n$ on it. The sensor spacing d_n for n is given by

$$d_n = \frac{L}{n}.$$

By (62), we have

$$\rho(d_n) = 1 + c_8 \cdot d_n^2 + o(d_n^2) \tag{104}$$

for some constant c_8. By the continuous differentiability of $\mathcal{K}_s$ (as a function of ζ) and ζ (as a function of ρ), we have

$$\zeta = \frac{1}{4} + c_9 \cdot (1-\rho) + o((1-\rho)^2),$$

and

$$\mathcal{K}_s = c_1 \cdot (\zeta - 1/4) + o(\zeta - 1/4)$$

for some constant c_9. Substituting (104) into the above equations gives

$$\mathcal{K}_s = c_{10} \cdot d_n^2 + o(d_n^2) \tag{105}$$

for some constant c_{10}. The node density is given by

$$\mu_n = \frac{n^2}{L^2} = d_n^{-2}. \tag{106}$$

Substituting (106) into (105) yields (75). The total amount of information per unit area is given by

$$\mu_n\mathcal{K}_s = c_5 + o(1) \tag{107}$$

and it converges to c_5 as $n \to \infty$.

To calculate the energy efficiency, we first calculate the total communication energy consumed by the minimum hop routing, given by

$$E_t' = E_c(d_n)\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(|i - \lfloor n/2\rfloor| + |j - \lfloor n/2\rfloor|)$$
$$= \Theta(n^3)E_c(d_n) = E_0 L^\nu n^{-\nu}\Theta(n^3)$$
$$= \Theta(n^{3-\nu}) \tag{108}$$

as $n \to \infty$ (i.e., $\mu_n \to \infty$). Here, E_t' denotes the total energy considering only the communication energy. The energy efficiency in this case is given by

$$\eta' = \frac{\mu_n\mathcal{K}_s}{E_t'} \text{ [nats/J/ m}^2\text{]}. \tag{109}$$

Applying (107) and (108) to the above equation, we have the claims. □

Proof of Theorem 7:

Note that

$$E_t = n^2 E_s + \Theta(n^3)E_c(d_n).$$

In this case, n and d_n are fixed, and Theorem 3 is directly applicable. Since the number of nodes and communication energy are fixed, the sensing energy increases linearly with the total energy E_t. By Assumption *(A.5)*, the measurement SNR increases linearly with the sensing energy. Applying Theorem 3 yields (80). □

APPENDIX II

To prove Lemma 2 (this will be stated below), we briefly introduce some relevant preliminary results.

Definition 5 (Matrix Norms [18], [28]): Let $\boldsymbol{A}$ be an $n \times n$ matrix with singular value decomposition

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T = \sum_{i=1}^{n} s_i \boldsymbol{u}_i \boldsymbol{v}_i^T \tag{110}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices with columns $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$, respectively, and $\boldsymbol{S} = \operatorname{diag}(s_1, s_2, \ldots, s_n)$ with nonnegative elements $s_1 \geq s_2 \geq \cdots s_n \geq 0$. The *operator norm* of $||\boldsymbol{A}||$ is defined as

$$||\boldsymbol{A}|| = s_1 = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} ||\boldsymbol{A}\boldsymbol{x}||/||\boldsymbol{x}|| \tag{111}$$

where $||\boldsymbol{x}||$ denotes the 2-norm of $\boldsymbol{x}$. On the other hand, the trace class norm of $\boldsymbol{A}$ is defined as

$$||\boldsymbol{A}||_1 = \sum_i s_i. \tag{112}$$

Note that if $\boldsymbol{A}$ is a symmetric matrix with eigenvalues $\{\lambda_i\}$, then

$$||\boldsymbol{A}||_1 = \sum_i |\lambda_i|. \tag{113}$$

Remark 8 (The Covariance Matrix and its Circulant Approximation): Using vector notation, the covariance matrix of the vector $\boldsymbol{y}_{|\mathcal{D}_n|}$ in (29) under p_1 is given by

$$\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|} = \mathbb{E}_1\{\boldsymbol{y}_{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T\} = [\sigma_{f_{id}^{-1}(\boldsymbol{i}), f_{id}^{-1}(\boldsymbol{j})}], \quad \sigma_{f_{id}^{-1}(\boldsymbol{i}), f_{id}^{-1}(\boldsymbol{j})} = \gamma_{\boldsymbol{i}-\boldsymbol{j}},\ \boldsymbol{i}, \boldsymbol{j} \in \mathcal{D}_n \tag{114}$$

where $\boldsymbol{\gamma_h}$ is defined in (23) and f_{id} is defined in (86). With slight abuse of notation, we use $\sigma_{\boldsymbol{ij}}$ for $\sigma_{f_{id}^{-1}(\boldsymbol{i}), f_{id}^{-1}(\boldsymbol{j})}$ for the sake of exposition.

The circulant approximation $\boldsymbol{C}_{|\mathcal{D}_n|}$ to $\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}$ is obtained by treating $\mathcal{D}_n$ as a high-dimensional torus with opposite ends being neighbors, and $\boldsymbol{C}_{|\mathcal{D}_n|}$ is given by

$$\boldsymbol{C}_{|\mathcal{D}_n|} = [c_{\boldsymbol{ij}}], \quad c_{\boldsymbol{ij}} = \gamma_{\pi(\boldsymbol{i}-\boldsymbol{j})},\ \boldsymbol{i}, \boldsymbol{j} \in \mathcal{I}_n \tag{115}$$

where the mapping $\pi : \mathbb{Z}^d \to \mathbb{Z}^d$ is defined as

$$\pi(\boldsymbol{h}) = \pi(h_1, h_2, \ldots, h_d) = (h_1', h_2', \ldots, h_d') \tag{116}$$

and

$$h_k' = h_k I(|h_k| \leq n/2) + (n - |h_k|) I(|h_k| > n/2), \quad k = 1, \ldots, d. \tag{117}$$

Here,[8] $I(\cdot)$ is the indicator function. Note that $\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}$ is a block Toeplitz matrix, while $\boldsymbol{C}_{|\mathcal{D}_n|}$ is a block circulant matrix. It is known that the eigenvalues of the block circulant matrix $\boldsymbol{C}_{|\mathcal{D}_n|}$ are given by

$$\lambda_{\boldsymbol{i}} = \sum_{\boldsymbol{h} \in \mathcal{D}_n} \gamma_{\pi(\boldsymbol{h})} e^{\iota \boldsymbol{h} \cdot \boldsymbol{\omega_i}} \tag{118}$$

for $\boldsymbol{i} = (i_1, \ldots, i_d) \in \mathcal{D}_n$, where

$$\boldsymbol{\omega_i} = (\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_d}) = \left(\frac{2\pi i_1}{n}, \frac{2\pi i_2}{n}, \ldots, \frac{2\pi i_d}{n}\right). \tag{119}$$

Define the periodic approximate spectral density by

$$f_n^c(\boldsymbol{\omega}) = (2\pi)^{-d} \sum_{\boldsymbol{h} \in \mathcal{D}_n} \gamma_{\pi(\boldsymbol{h})} e^{\iota \boldsymbol{h} \cdot \boldsymbol{\omega}}. \tag{120}$$

Then, the eigenvalues of $\boldsymbol{C}_{|\mathcal{D}_n|}$ are given by

$$\lambda_{\boldsymbol{i}} = (2\pi)^d f_n^c(\boldsymbol{\omega_i}), \quad \boldsymbol{i} \in \mathcal{D}_n. \tag{121}$$

Further, it is shown in [18, Lemma 4.1.(c)] that the periodic approximate spectral density converges uniformly to the true spectral density $f_1(\boldsymbol{\omega})$, i.e.,

$$\sup_{\boldsymbol{\omega} \in [-\pi, \pi)^d} |f_n^c(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega})| \to 0 \tag{122}$$

as $n \to \infty$.

Lemma 1: Under the assumption of Theorem 2, we have the following.

(a) $f_n^c(\boldsymbol{\omega})$ is uniformly continuous for sufficiently large n.

(b)

$$\sup_{\boldsymbol{\omega} \in [-\pi, \pi)^d} \left| \frac{1}{f_n^c(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega})} \right| \to 0, \quad \text{as } n \to \infty. \tag{123}$$

(c) $1/f_n^c(\boldsymbol{\omega})$ is uniformly continuous for sufficiently large n.

Proof of Lemma 1:

(a) By assumption, $f_1(\boldsymbol{\omega})$ is continuous on the compact domain $[-\pi, \pi]^d$. By the uniform continuity theorem, $f_1(\boldsymbol{\omega})$ is uniformly continuous. For any $\epsilon > 0$, $||\boldsymbol{\omega} - \boldsymbol{\omega}'|| < \delta$ imples

$$\begin{aligned} |f_n^c(\boldsymbol{\omega}) - f_n^c(\boldsymbol{\omega}')| &\leq |f_n^c(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega}) + f_1(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega}') \\ &\quad + f_1(\boldsymbol{\omega}') - f_n^c(\boldsymbol{\omega}')| \\ &\leq |f_n^c(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega})| + |f_1(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega}')| \\ &\quad + |f_1(\boldsymbol{\omega}') - f_n^c(\boldsymbol{\omega}')| \\ &\leq \epsilon/3 + \epsilon/3 + \epsilon/3 \end{aligned}$$

for sufficiently large n. The convergence of the first and third terms is by (122) and that of the second term is by the uniform continuity of $f_1(\boldsymbol{\omega})$.

(b) Since the spectrum $f_1(\omega)$ has a positive lower bound by assumption, its inverse $1/f_1(\boldsymbol{\omega})$ is bounded from above. In addition, due to (122) there exists $M_1 > 0$ such that

$$\frac{1}{f_1(\boldsymbol{\omega})} \leq M_1 \quad \text{and} \quad \frac{1}{f_n^c(\boldsymbol{\omega})} \leq M_1 \tag{124}$$

for all $\boldsymbol{\omega} \in [-\pi, \pi)^d$ and for sufficiently large n. Then, for any $\epsilon > 0$

$$\left| \frac{1}{f_n^c(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega})} \right| = \left| \frac{1}{f_n^c(\boldsymbol{\omega})} \frac{1}{f_1(\boldsymbol{\omega})} \right| |f_n^c(\boldsymbol{\omega}) - f_1(\boldsymbol{\omega})| \tag{125}$$

$$\leq \epsilon M_1^2 \tag{126}$$

[8]The distinction of even and odd n will not be considered for simplicity, as this is merely a technical issue. In either case, the asymptotic behavior is the same.

for all $\boldsymbol{\omega} \in [-\pi, \pi)^d$ and for sufficiently large n, by (122) and (124).

(c) For any $\epsilon > 0$, $||\boldsymbol{\omega} - \boldsymbol{\omega}'|| < \delta$ implies

$$\begin{aligned}\left|\frac{1}{f_n^c(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega}')}\right| &\le \left|\frac{1}{f_n^c(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega})} + \frac{1}{f_1(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega}')} + \frac{1}{f_1(\boldsymbol{\omega}')} - \frac{1}{f_n^c(\boldsymbol{\omega}')}\right| \\ &\le \left|\frac{1}{f_n^c(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega})}\right| + \left|\frac{1}{f_1(\boldsymbol{\omega})} - \frac{1}{f_1(\boldsymbol{\omega}')}\right| + \left|\frac{1}{f_1(\boldsymbol{\omega}')} - \frac{1}{f_n^c(\boldsymbol{\omega}')}\right| \\ &\le \epsilon/3 + \epsilon/3 + \epsilon/3\end{aligned}$$

for sufficiently large n. The convergence of the first and third terms is by (123) and that of the second term is by the uniform continuity of $1/f_1(\boldsymbol{\omega})$. (The uniform continuity of $1/f_1(\boldsymbol{\omega})$ is obvious due to the uniform continuity and strict positivity of $f_1(\boldsymbol{\omega})$.) □

Lemma 2: Under the conditions of Theorem 2, we have

$$\frac{1}{|\mathcal{D}_n|}\boldsymbol{y}_{|\mathcal{D}_n|}^T \boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1} \boldsymbol{y}_{|\mathcal{D}_n|} \to \frac{1}{(2\pi)^d}\int_{[-\pi,\pi)^d} \frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})} d\boldsymbol{\omega}$$

almost surely.

Proof of Lemma 2: First, it is shown in [18, Lemma 4.1.(a)] that

$$|\mathcal{D}_n|^{-1}||\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|} - \boldsymbol{C}_{|\mathcal{D}_n|}||_1 = O\left(\frac{1}{n}\right) \tag{127}$$

as $n \to \infty$. Let $\{\lambda_{|\mathcal{D}_n|}(i),\ i = 1, 2, \ldots, |\mathcal{D}_n|\}$ be the eigenvalues of $|\mathcal{D}_n|^{-1}(\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|} - \boldsymbol{C}_{|\mathcal{D}_n|})$, where $|\mathcal{D}_n| = n^d$ for d-D. Then, by (113) and (127) we have

$$\sum_{i=1}^{n^d} |\lambda_{|\mathcal{D}_n|}(i)| = O\left(\frac{1}{n}\right). \tag{128}$$

Since the convergence of the eigenvalues of the block Toeplitz matrix $\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}$ and its block circulant approximation $\boldsymbol{C}_{|\mathcal{D}_n|}$ is uniform (the eigenvalues of these matrices are the samples of the corresponding spectra for sufficiently large n; see (121) and (122)), $\min_i |\lambda_{|\mathcal{D}_n|}(i)|$ and $\max_i |\lambda_{|\mathcal{D}_n|}(i)|$ have the same convergence rate, i.e., there exist M_2, M_3, and r_n such that

$$M_2 r_n \le \min_i |\lambda_{|\mathcal{D}_n|}(i)| \le \max_i |\lambda_{|\mathcal{D}_n|}(i)| \le M_3 r_n. \tag{129}$$

By (128) and (129) we have

$$r_n = O\left(\frac{1}{n^{d+1}}\right). \tag{130}$$

Since the spectra $f_1(\omega)$ and $f_n^c(\boldsymbol{\omega})$ have positive lower bounds by assumption, their inverses $1/f_1(\boldsymbol{\omega})$ and $1/f_n^c(\boldsymbol{\omega})$ are bounded from above. Hence, the eigenvalues of $\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1}$ and $\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}$ are bounded from above since the eigenvalues of these matrices are the samples of the corresponding inverse spectra for sufficiently large n, and thus we have

$$||\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1}|| < M_1 \quad \text{and} \quad ||\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}|| < M_1 \tag{131}$$

for all sufficiently large n.

Now consider the error between two quadratic terms

$$\begin{aligned}&\left||\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|} - |\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}\right| \\ &= \left||\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\left(\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1} - \boldsymbol{C}_{|\mathcal{D}_n|}^{-1}\right)\boldsymbol{y}_{|\mathcal{D}_n|}\right| \\ &= \left||\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}\left(\boldsymbol{C}_{|\mathcal{D}_n|} - \boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}\right)\boldsymbol{\Sigma}_{1,|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}\right| \\ &\overset{(a)}{\le} CM_1^2\sum_{i=1}^{|\mathcal{D}_n|}|\lambda_i|y_i^2 \\ &\overset{(b)}{\le} CM_1^2M_3r_n\sum_{i=1}^{|\mathcal{D}_n|}y_i^2 \\ &\overset{(c)}{\le} CM_1^2M_3O\left(\frac{1}{n}\right)\frac{1}{n^d}\sum_{i=1}^{n^d}y_i^2 \\ &\xrightarrow{(d)} 0 \text{ a.s.}\end{aligned} \tag{132}$$

for some $C > 0$. Here, step (a) is by (131) and the definition of the trace class norm (113), step (b) is by (129), and step (c) is by (130). Step (d) is by the strong law of large numbers (SLLN) on the sample mean of y_i^2. Since $\{y_i\}$ is i.i.d. $\mathcal{N}(0, \sigma^2)$ under p_0, $\frac{1}{n^2}\sum_{i=1}^{n^2} y_i^2 \to \sigma^2$ almost surely. Thus, the quadratic form using the block circulant approximation converges almost surely to that based on the true covariance matrix.

We next consider the asymptotic behavior of $|\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}$. Since $\boldsymbol{C}_{|\mathcal{D}_n|}$ is a block circulant matrix, the eigendecomposition is given by [29], [30]

$$\boldsymbol{C}_{|\mathcal{D}_n|} = \boldsymbol{W}_{|\mathcal{D}_n|}\boldsymbol{\Lambda}_{|\mathcal{D}_n|}\boldsymbol{W}_{|\mathcal{D}_n|}^H \tag{133}$$

where $\boldsymbol{W}_{|\mathcal{D}_n|}$ is the d-dimensional discrete Fourier transform (DFT) matrix which is unitary, and

$$\boldsymbol{\Lambda}_{|\mathcal{D}_n|} = \text{diag}(\lambda_{0,\ldots,0}, \ldots, \lambda_{n-1,\ldots,n-1}). \tag{134}$$

The inverse of $\boldsymbol{C}_{|\mathcal{D}_n|}$ is given by

$$\boldsymbol{C}_{|\mathcal{D}_n|}^{-1} = \boldsymbol{W}_{|\mathcal{D}_n|}\boldsymbol{\Lambda}_{|\mathcal{D}_n|}^{-1}\boldsymbol{W}_{|\mathcal{D}_n|}^H. \tag{135}$$

Define

$$\bar{\boldsymbol{y}}_{|\mathcal{D}_n|} = \boldsymbol{W}_{|\mathcal{D}_n|}^H\boldsymbol{y}_{|\mathcal{D}_n|}. \tag{136}$$

Then, $\bar{\boldsymbol{y}}_{|\mathcal{D}_n|}$ is a vector of i.i.d. Gaussian random variables since $\boldsymbol{W}_{|\mathcal{D}_n|}$ is unitary and $\boldsymbol{y}_{|\mathcal{D}_n|}$ is a vector with i.i.d. Gaussian elements under p_0. Thus, $|\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{C}_{|\mathcal{I}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}$ is given by

$$\begin{aligned}S_n &= |\mathcal{D}_n|^{-1}\boldsymbol{y}_{|\mathcal{D}_n|}^T\boldsymbol{C}_{|\mathcal{D}_n|}^{-1}\boldsymbol{y}_{|\mathcal{D}_n|} = |\mathcal{D}_n|^{-1}\bar{\boldsymbol{y}}_{|\mathcal{D}_n|}^T\boldsymbol{\Lambda}_{|\mathcal{D}_n|}^{-1}\bar{\boldsymbol{y}}_{|\mathcal{D}_n|} \\ &= \frac{1}{n^d}\sum_{\boldsymbol{i}\in\mathcal{D}_n}\frac{\bar{Y}_{\boldsymbol{i}}^2}{\lambda_{\boldsymbol{i}}} \qquad (137) \\ &= \frac{1}{n^d}\sum_{i_1=0}^{n-1}\cdots\sum_{i_d=0}^{n-1}\frac{\bar{Y}_{i_1,\ldots,i_d}^2}{\lambda_{i_1,\ldots,i_d}} \qquad (138)\end{aligned}$$

where $\{\bar{Y}_{\boldsymbol{i}},\ \boldsymbol{i} \in \mathcal{D}_n\}$ is i.i.d. zero-mean Gaussian with variance σ^2. For sufficiently large n, fix K $(0 < K < n)$ and divide the indices of each dimension such that

$$\begin{aligned}&\mathcal{I} = [0, 1, \ldots, n-1] = \mathcal{I}(0) \cup \mathcal{I}(1) \cup \cdots \mathcal{I}(K-1) \\ &\mathcal{I}(i) \cap \mathcal{I}(j) = \phi, \quad \text{if } i \ne j \text{ and}\end{aligned}$$

$$|\mathcal{I}(0)| = \ldots = |\mathcal{I}(K-2| = \lfloor n/K \rfloor,$$
$$|\mathcal{I}(K-1)| = n - (K-1)|\mathcal{I}(0)|.$$

Then, (138) is given by

$$S_n = \frac{1}{K^d} \sum_{j_1=0}^{K-1} \cdots \sum_{j_d=0}^{K-1} \left(\frac{1}{|\mathcal{I}(j_1)| \ldots |\mathcal{I}(j_d)|} \times \sum_{i_1 \in \mathcal{I}(j_1)} \cdots \sum_{i_d \in \mathcal{I}(j_d)} \frac{\bar{Y}^2_{i_1,\ldots,i_d}}{\lambda_{i_1,\ldots,i_d}} \right). \quad (139)$$

Now let $i_1, \ldots, i_d(j_1, \ldots, j_d)$ denote the index representing the center of the $(j_1, \ldots, j_d)$th hypercube. Then, by (121) we have

$$\frac{1}{\lambda_{i_1,\ldots,i_d(j_1,\ldots,j_d)}} = \frac{1}{(2\pi)^d} \frac{1}{f_n^c(\boldsymbol{\omega_j})} \quad (140)$$

$$\boldsymbol{\omega_j} = (\omega_{j_1}, \ldots, \omega_{j_d}) = \left(\frac{2\pi j_1}{K}, \ldots, \frac{2\pi j_d}{K} \right) \quad (141)$$

and

$$\frac{1}{(2\pi)^d} \frac{1}{f_n^c(\boldsymbol{\omega_j})} - \epsilon' \leq \frac{1}{\lambda_{i_1,\ldots,i_d}} \leq \frac{1}{(2\pi)^d} \frac{1}{f_n^c(\boldsymbol{\omega_j})} + \epsilon' \quad (142)$$

for all $(i_1, \ldots, i_d)$ in the $(j_1, \ldots, j_d)$th hypercube. Here, ϵ' (> 0) is independent of $(j_1, \ldots, j_d)$ since $1/f_n^c(\boldsymbol{\omega})$ is uniformly continuous over $\boldsymbol{\omega} \in [-\pi, \pi)^d$ by Lemma 1 (c). Applying (142)–(139), we have

$$V_n - \frac{\epsilon'}{n^d} \sum_{\boldsymbol{i} \in \mathcal{D}_n} \bar{Y}^2_{\boldsymbol{i}} \leq S_n \leq V_n + \frac{\epsilon'}{n^d} \sum_{\boldsymbol{i} \in \mathcal{D}_n} \bar{Y}^2_{\boldsymbol{i}}, \quad (143)$$

where

$$V_n = \frac{1}{K^d} \sum_{j_1=1}^{K} \cdots \sum_{j_d=1}^{K} \frac{1}{(2\pi)^d} \frac{1}{f_n^c(\boldsymbol{\omega_j})} \times \left(\frac{1}{|\mathcal{I}(j_1)| \ldots |\mathcal{I}(j_d)|} \sum_{i_1 \in \mathcal{I}(j_1)} \cdots \sum_{i_d \in \mathcal{I}(j_d)} \bar{Y}^2_{i_1,\ldots,i_d} \right). \quad (144)$$

By the SLLN for the sample mean of $\bar{Y}^2_{\boldsymbol{i}}$, we have

$$\sigma^2 - \epsilon'' \leq \frac{1}{|\mathcal{I}(j_1)| \ldots |\mathcal{I}(j_d)|} \sum_{i_1 \in \mathcal{I}(j_1)} \cdots \sum_{i_d \in \mathcal{I}(j_d)} \bar{Y}^2_{i_1,\ldots,i_d} \leq \sigma^2 + \epsilon'' \quad (145)$$

almost surely for sufficiently large n given K. Thus, V_n is given by

$$(\sigma^2 - \epsilon'') Z_n \leq V_n \leq (\sigma^2 + \epsilon'') Z_n \quad (146)$$

where

$$Z_n = \frac{1}{K^d} \sum_{j_1=1}^{K} \cdots \sum_{j_d=1}^{K} \frac{1}{(2\pi)^d f_n^c(\boldsymbol{\omega_j})}. \quad (147)$$

Now we take $K \to \infty$, and the Riemann sum Z_n converges to

$$Z_n \to \frac{1}{(2\pi)^d} \int_{-[\pi,\pi)^d} \frac{1}{(2\pi)^d f_1(\boldsymbol{\omega})} d\boldsymbol{\omega} \quad (148)$$

by Lemma 1 (b) and (c). Since ϵ' and ϵ'' can be made arbitrarily small by making n and K large, and $\frac{1}{(2\pi)^d} \int_{-[\pi,\pi)^d} \frac{1}{(2\pi)^d} \frac{1}{f_1(\boldsymbol{\omega})} d\boldsymbol{\omega} < M_4$ for some $M_4 > 0$ and $n^{-d} \sum_{\boldsymbol{i} \in \mathcal{D}_n} \bar{Y}_{\boldsymbol{i}} \to \sigma^2$ a.s., we have by (143), (146), and (148), that

$$|\mathcal{D}_n|^{-1} \boldsymbol{y}^T_{|\mathcal{D}_n|} \boldsymbol{C}^{-1}_{|\mathcal{D}_n|} \boldsymbol{y}_{|\mathcal{D}_n|} \to (2\pi)^{-d} \int_{\boldsymbol{\omega} \in [-\pi,\pi)^2} \frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})} d\boldsymbol{\omega} \quad (149)$$

almost surely as $n \to \infty$. By (132) and (149) we have

$$|\mathcal{D}_n|^{-1} \boldsymbol{y}^T_{|\mathcal{D}_n|} \boldsymbol{\Sigma}^{-1}_{1,|\mathcal{D}_n|} \boldsymbol{y}_{|\mathcal{D}_n|} \to (2\pi)^{-d} \int_{\boldsymbol{\omega} \in [-\pi,\pi)^2} \frac{\sigma^2}{(2\pi)^d f_1(\boldsymbol{\omega})} d\boldsymbol{\omega} \quad (150)$$

almost surely as $n \to \infty$. This concludes the proof. □

REFERENCES

[1] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Computing*, vol. 1, pp. 59–69, Jan.–Mar. 2002.

[2] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[3] M. Dong, L. Tong, and B. M. Sadler, "Impact of data retrieval pattern on homogeneous signal field reconstruction in dense sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4352–4364, Nov. 2006.

[4] Y. Sung, X. Zhang, L. Tong, and H. V. Poor, "Sensor configuration and activation for field detection in large sensor arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 1354–1365, Feb. 2008.

[5] L. S. Pillutla and V. Krishnamurthy, "Mutual information and energy tradeoff in correlated wireless sensor networks," in *Proc. 2008 IEEE Int. Conf. Communications (ICC)*, Beijing, China, May 2008, pp. 4402–4406.

[6] F. Liese and I. Vajda, "On divergence and informations in statistics and information theory," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4394–4412, Oct. 2006.

[7] S. Kullback, *Information Theory and Statistics*. Mineola, NY: Dover, 1997.

[8] M. D. Donsker and S. R. S. Varadhan, "Large deviations for stationary Gaussian process," *Commun. Math. Phys.*, vol. 97, pp. 187–210, 1985.

[9] G. R. Benitz and J. A. Bucklew, "Large deviation rate calculations for nonlinear detectors in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 358–371, Mar. 1990.

[10] W. Bryc and A. Dembo, "Large deviations for quadratic functionals of Gaussian processes," *J. Theor. Probab.*, vol. 10, no. 2, pp. 307–332, 1997.

[11] B. Bercu, F. Gamboa, and A. Rouault, "Large deviations for quadratic forms of stationary Gaussian processes," *Stochastic Processes and Their Applications*, vol. 71, pp. 75–90, 1997.

[12] Y. Sung, L. Tong, and H. V. Poor, "Neyman–Pearson detection of Gauss–Markov signals in noise: Closed-form error exponent and properties," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1354–1365, Apr. 2006.

[13] A. S. Leong, S. Dey, and J. S. Evans, "Error exponents for Neyman-Pearson detection of Markov chains in noise," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5097–5103, Oct. 2007.

[14] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss-Markov random field on nearest-neighbor graph," in *Proc. 2007 IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'07)*, Honolulu, HI, Apr. 2007, pp. III-829–III-832.

[15] Y. Sung, H. V. Poor, and H. Yu, "Information, energy and density for ad hoc sensor networks over correlated random fields: Large deviations analysis," in *Proc. 2008 Int. Symp. Information Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 1592–1596.
[16] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applicatons*. New York: Chapman & Hall/CRC, 2005.
[17] R. R. Bahadur, S. L. Zabell, and J. C. Gupta, "Large deviations, tests, and estimates," in *Asymptotic Theory of Statistical Tests and Estimation: In Honor of Wassily Hoeffding*, I. M. Chakravarti, Ed. New York: Academic, 1980.
[18] J. T. Kent and K. V. Mardia, "Spectral and circulant approximations to the likelihood for stationary Gaussian random fields," *J. Statist. Planning and Inference*, vol. 502, no. 3, pp. 379–394, 1996.
[19] R. R. Bahadur, "Some limit theorems in statistics," in *Proc. Conf. Board of the Mathematical Sciences Regional Conferences (Series in Applied Mathematics)*, Philadelphia, PA, 1971.
[20] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[21] J. Besag, "On a system of two-dimensional recurrence equations," *J. Roy. Statist. Soc. Ser. B*, vol. 43, no. 3, pp. 302–309, 1981.
[22] A. Erdélyi, *Higher Transcendental Functions, Vol. II*. New York: McGraw-Hill, 1953.
[23] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, pp. 434–449, Dec. 1954.
[24] I. Vajda, *Theory of Statistical Inference and Information*. Dordrecht, The Netherlands: Kluwer Academic, 1989.
[25] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. New York: Springer, 1991.
[26] E. J. Hannon, "The asymptotic theory of linear time-series models," *J. Appl. Probab.*, vol. 10, pp. 130–145, Mar. 1973.
[27] X. Guyon, *Random Fields on a Network: Modeling, Statistics and Applications*. New York: Springer-Verlag, 1995.
[28] I. Gohberg, S. Goldberg, and M. A. Kaashoek, *Basic Classes of Linear Operators*. Basel, Switzerland: Birkhäuser, 2004.
[29] R. J. Martin, "A note on the asymptotic eigenvalues and eigenvectors of the dispersion matrix of a second-order stationary process on a d-dimensional lattice," *J. Appl. Probab.*, vol. 23, pp. 529–535, Jun. 1986.
[30] P. J. Davis, *Circulant Matrices*. New York, NY: Chelsea, 1994.

Youngchul Sung (S'92–M'93–SM'09) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, in 2005.

From 2005 until 2007, he worked as a Senior engineer in the Corporate R & D Center at Qualcomm, Inc. , San Diego, CA. He is now an Assistant Professor in the Department of Electrical Engineering in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His research interests include statistical signal processing, asymptotic statistics, large deviations, and their applications to wireless communication systems and related areas.

H. Vincent Poor (S'72–M'77–SM'82–F'77) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1977.

From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign, Urbana, IL. Since 1990, he has been on the faculty at Princeton University, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. His research interests are in the areas of stochastic analysis, statistical signal processing and their applications in wireless networks, and related fields. Among his publications in these areas are the recent books *MIMO Wireless Communications* (Cambridge University Press, 2007), coauthored with Ezio Biglieri *et al.*, and *Quickest Detection* (Cambridge University Press, 2009), coauthored with Olympia Hadjiliadis.

Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and a former Guggenheim Fellow. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and from 2004 to 2007 as the Editor-in-Chief of these TRANSACTIONS. He is the recipient of the 2005 IEEE Education Medal. Recent recognition of his work includes the 2007 IEEE Marconi Prize Paper Award, the 2007 Technical Achievement Award of the IEEE Signal Processing Society, and the 2008 Aaron D. Wyner Distinguished Service Award of the IEEE Information Theory Society.

Heejung Yu (S'07) received the B.S. degree in radio science and engineering from the Korea University, Seoul, Korea, in 1999 and the M.S. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2001.

He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, KAIST. From 2001 to 2006, he was with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. His areas of interest include statistical signal processing and communication theory.