# Innovations Diffusion: A Spatial Sampling Scheme for Distributed Estimation and Detection

Zhi Quan, William J. Kaiser, and Ali H. Sayed, *Fellow, IEEE*

***Abstract*—We consider a wireless network with distributed processing capabilities for estimation or detection applications. Due to limited communication resources, the network selects only a subset of sensor measurements for estimation or detection as long as the resulting fidelity is tolerable. We present a distributed sampling scheme based on the concept of *innovations diffusion* to select the sensor nodes. In the proposed scheme, sensor selection is accomplished through local communication and signal processing. In order to conserve energy and prolong system lifetime, the proposed algorithm selects a nearly minimum number of active sensors to ensure a desired fidelity for each working period. Extensive simulations illustrate the effectiveness of the proposed sampling scheme.**

***Index Terms*—Diffusion, distributed processing, estimation and detection, innovations, sampling, wireless networks.**

## I. Introduction

A large class of wireless sensor networks (WSNs) is concerned with estimating or detecting an underlying physical phenomenon over time and space in a noisy environment. The network consolidates data collected by sensor nodes in order to reconstruct the state of nature such as estimating or detecting a field variable given the sensor observations. Such sensor networks are typically designed for applications including environmental monitoring, military surveillance, and space exploration [2]. Two key issues in these settings are the fidelity at which the field variable can be estimated or detected and the cost of operating the sensor network.

Because wireless sensor devices are usually battery-powered and battery replacement is difficult, energy efficiency becomes important for sensor networks and has a direct influence on the system lifetime. For sensor networks with dense deployment, it is necessary to select a group of sensors that are more informative for data fusion purposes and to set other nodes inactive (or sleeping) to save energy. While there might not always exist a direct relation between energy efficiency and the number of active sensors, reducing the number of active sensors generally leads to less energy consumption [3]. In addition, energy efficiency is related to distributed processing through local communication and computation. The goal of this paper is to propose a methodology to select a nearly minimum number of sensor nodes for distributed inference applications with some desired fidelity.

Manuscript received June 07, 2007; revised August 28, 2008. First published October 31, 2008; current version published January 30, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniel P. Palomar. This work was supported in part by the NSF by grants ECS-0601266 and ECS-0725441. A shorter version of this work appeared in the proceedings of the Information Processing Sensors Networking (IPSN) Conference, Cambridge, MA, April 2007.

The authors are with the Department of Electrical Engineering, University of California at Los Angeles, CA 90095 USA (e-mail: quan@ee.ucla.edu; kaiser@ee.ucla.edu; sayed@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2008264

### A. Parameter Estimation

Consider a network of $N$ sensor nodes estimating an unknown deterministic parameter $\boldsymbol{\theta} \in \mathbb{R}^m$. For sensor $i$, which has a vector of $n_i$ measurements, its observation is assumed to be distorted by a matrix $\mathbf{H}_i \in \mathbb{R}^{n_i \times m}$ and corrupted by additive noise, i.e.

$$\mathbf{y}_i = \mathbf{H}_i \boldsymbol{\theta} + \mathbf{v}_i, \quad i = 1, 2, \ldots, N. \tag{1}$$

Equation (1) can be written compactly as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v} \tag{2}$$

where $\mathbf{y} = \operatorname{col}\{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$, $\mathbf{H} = \operatorname{col}\{\mathbf{H}_1, \ldots, \mathbf{H}_N\}$, and $\mathbf{v} = \operatorname{col}\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$. The measurement noise $\mathbf{v}$ is zero mean and has covariance matrix $\mathbf{C} = \mathrm{E}\left(\mathbf{v}\mathbf{v}^T\right)$. Linear models of the form (1) are adopted in many sensing applications due to their mathematical tractability.

The objective is to estimate the parameter $\boldsymbol{\theta}$ in (2) from a subset of the measurements $\{\mathbf{y}_i\}_{i=1}^N$. At the beginning of each estimation period, the network selects a subset of the sensor measurements, indexed by $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$, for estimation as long as some desired estimation fidelity can be guaranteed. We use the notation $\mathbf{H}_{\mathcal{A}_n} = \operatorname{col}\{\mathbf{H}_{i_1}, \mathbf{H}_{i_2}, \ldots, \mathbf{H}_{i_n}\}$ and $\mathbf{y}_{\mathcal{A}_n} = \operatorname{col}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$ to denote the corresponding data matrices and observation vectors, respectively. Let $\mathbf{C}_{\mathcal{A}_n}$ be the partial matrix selected from $\mathbf{C}$ with rows and columns corresponding to $\{i_1, i_2, \ldots, i_n\}$. With the data model in (1), the linear minimum-variance-unbiased-estimator (m.v.u.e.) [4] of $\boldsymbol{\theta}$ using the observations in $\mathbf{y}_{\mathcal{A}_n}$ is given by

$$\hat{\boldsymbol{\theta}}\left(\mathbf{y}_{\mathcal{A}_n}\right) = \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{y}_{\mathcal{A}_n} \tag{3}$$

and the resulting minimum mean-squared error (MMSE) is

$$\begin{aligned} \mathbf{D}\left(\mathbf{y}_{\mathcal{A}_n}\right) &= \mathrm{E}\left(\left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n})\right]\left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n})\right]^T\right) \\ &= \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \end{aligned} \tag{4}$$

where it is assumed that all inverses exist, whenever necessary. It is desired that the MMSE of the estimator be less than or equal

to some desired distortion $D_0$. In other words, we would like to select the sensors $\{i_1, \ldots, i_n\}$ to meet the following accuracy requirement

$$D_{\mathcal{A}_n} = \mathrm{Tr}\left[\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_n}\right)\right] \leq D_0. \tag{5}$$

We shall proceed in Section III to develop a method to select the subset $\mathbf{y}_{\mathcal{A}_n}$ of the sensors to estimate the unknown parameter $\boldsymbol{\theta}$ according to (5) in a *distributed* way. The method will require each sensor to locally compute the m.v.u.e. of $\boldsymbol{\theta}$ under a certain distortion constraint by communicating only with its nearby nodes. The spatial sampling problem is also relevant in the context of signal detection, as we now explain.

### B. Signal Detection

To detect a signal of interest, we consider a simple binary hypothesis test of the form

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{v}\ (\boldsymbol{\theta} = 0) \\ H_1 &: \mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}\ (\boldsymbol{\theta} \neq 0) \end{aligned} \tag{6}$$

where $\mathbf{y} = \mathrm{col}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ are the sensor observations, $\mathbf{v} = \mathrm{col}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$ are the observation noises, and $\mathbf{H} = \mathrm{col}\{\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_N\}$ are the distortion matrices. It is assumed that the noise is Gaussian distributed with zero mean and covariance matrix $\mathbf{C} = \mathrm{E}\left(\mathbf{v}\mathbf{v}^T\right)$. Then, the probability density functions of $\mathbf{y}$ under different hypotheses are expressed as

$$f\left(\mathbf{y}|H_0\right) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y}\right) \tag{7}$$

and [see (8) at the bottom of the page], where $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$. The optimal detector derived from the Neyman-Pearson formulation checks the log-likelihood ratio test [5]:

$$L(\mathbf{y}) \triangleq \log \frac{f(\mathbf{y}|H_1)}{f(\mathbf{y}|H_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{9}$$

where $\gamma$ is the test threshold. It can be verified that the log-likelihood ratio can be expressed as

$$L(\mathbf{y}) = \mathbf{s}^T\mathbf{C}^{-1}\mathbf{y} - \frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \tag{10}$$

and that $L(\mathbf{y})$ is also Gaussian distributed under different hypotheses, namely

$$H_0 : L(\mathbf{y}) \sim \mathcal{N}\left(-\frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}, \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}\right) \tag{11}$$

$$H_1 : L(\mathbf{y}) \sim \mathcal{N}\left(\frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}, \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}\right). \tag{12}$$

Consequently, the probabilities of false alarm and detection have the following forms:

$$P_{FA} = \Pr\left\{L(\mathbf{y}) \geq \gamma | H_0\right\} = Q\left(\frac{\gamma + \frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}\right) \tag{13}$$

and

$$P_D = \Pr\left\{L(\mathbf{y}) \geq \gamma | H_1\right\} = Q\left(\frac{\gamma - \frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}{\sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}}\right). \tag{14}$$

To evaluate the detection performance, we adopt the deflection coefficient [5], which is defined as

$$\begin{aligned} d^2(\mathbf{y}) &= \frac{\left[\mathrm{E}\left(L\left(\mathbf{y}|H_1\right)\right) - \mathrm{E}\left(L\left(\mathbf{y}|H_0\right)\right)\right]^2}{\mathrm{Var}\left[L(\mathbf{y})|H_0\right]} \\ &= \mathbf{s}^T\mathbf{C}^{-1}\mathbf{s} \\ &= \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\boldsymbol{\theta} \end{aligned} \tag{15}$$

where $\mathrm{Var}(\cdot)$ means the variance. The deflection coefficient provides a good measure for the detection performance when the observations are Gaussian distributed; the detection performance improves as the value of the deflection coefficient increases. This is illustrated by representing the probability of detection in terms of any given probability of false alarm as

$$\begin{aligned} P_D &= Q\left[Q^{-1}\left(P_{FA}\right) - \sqrt{\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s}}\right] \\ &= Q\left[Q^{-1}\left(P_{FA}\right) - d\right]. \end{aligned} \tag{16}$$

Given a fixed probability of false alarm, if the deflection coefficient is greater than a certain constraint, i.e., $d^2 > d_0^2$, then we have

$$P_D(d^2) > P_D(d_0^2). \tag{17}$$

Our second objective in this paper is to select a subset of the observations $\mathcal{A}_n$ such that the resulting deflection coefficient is larger than or equal to $d_0^2$. In other words, the detector should guarantee at least a certain probability of detection $P_D(d_0^2)$ for a given probability of false alarm. The target deflection coefficient $d_0^2$ is determined by the application requirement.

### C. Overview and Main Results

Distributed algorithms are attractive in large-scale networks where a centralized solution is infeasible, nonscalable, or too costly. A number of distributed solutions have been proposed for detection, estimation, and inference purposes [6]–[12]. In these solutions, each node does not require access to global information and can deliver performance by communicating solely with nearby nodes. The major advantage of distributed algorithms is

$$f\left(\mathbf{y}|H_1\right) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{s})^T\mathbf{C}^{-1}(\mathbf{y} - \mathbf{s})\right] \tag{8}$$

that they save the significant cost of transmitting all data to a fusion center and each sensor can operate as a data sink to achieve robustness.

Motivated by these observations, we develop in this paper a distributed sensor selection strategy. We consider a wireless sensor network with many redundant nodes, each of which can observe a physical phenomenon in the field. For each work period, the network selects a subset of sensor measurements, $\mathcal{A}_n$, to achieve a desired fidelity. We refer to the procedure of selecting these sensor measurements as *sensor sampling*. Intuitively, if the noise level is low, a small number of sensors is sufficient to achieve the desired fidelity; however, if the noise condition is severe, more sensors should be activated for accurate inference. We develop a sampling scheme that exploits what we call *innovations diffusion*. Innovation refers to the new information that a sensor measurement contributes to the reduction of the inference error relative to prior measurements, and diffusion refers to the process by which the innovation is communicated across the network. The proposed sampling procedure will be accomplished in a distributed manner whereby each active sensor will locally compute a m.v.u.e. $\hat{\boldsymbol{\theta}}$ or detector based on local measurements, and the algorithm will activate a nearly minimum number of sensors to ensure a desired fidelity (e.g., $D_0$ or $d_0^2$) for each work period.

### D. Related Work and Comparisons

The problem of sensor selection has been investigated for various purposes in the literature. Reference [13] proposed an information driven sensor query (IDSQ) algorithm for tracking applications, where only a *single* sensor (leader) is active at a given time. After obtaining a measurement, the leader passes its measurement to the most informative sensor in the network, which will become the next leader. This work was extended in [14] by selecting the sensor measurement with the maximum mutual information. Reference [15] used a local greedy strategy to select the next most informative sensor node to reduce information entropy for target location with the assumption that the probability distribution of target location estimation is known *a priori*. The work in [16] addressed the sensor selection problem for the bounded uncertainty sensing model in order to minimize the error in estimating the target position. While these works focus on sensor selection for target tracking applications, this paper distinguishes itself from these works by using innovation, a new information measure, to activate a subset of sensors for estimation or detection applications.

There are several works addressing energy efficiency for sensor selection in sensor networks. In [3], a sensor selection algorithm was proposed to reconstruct the data image of a spatially bandlimited physical phenomenon based on blue noise masking. In such an application, spatial resolution can be traded for energy efficiency by reducing the number of sensors used to observe the area. Reference [17] proposed an adaptive sensor control scheme to minimize the number of active sensors assuming that all sensors are identical and the quality-of-service (QoS) can be expressed as a function of the sensor number. In [18], an innovations-based scheme was proposed to select sensor measurements to estimate an unknown parameter under a desired distortion constraint. In [19], a power scheduling scheme was proposed to minimize the total transmit power while satisfying a given estimation requirement. The result implied that sensors with bad channels and bad observation qualities should be turned off in order to conserve energy. While these prior works considered a centralized network in which there is a fusion center responsible for the network management and *global* computation, a new class of distributed processing structures and protocols would be desirable to decentralize these sensor selection schemes. These works suggest that reducing the number of active sensors would lead to energy efficiency. Sharing this idea with previous works, our proposed algorithm presents a completely distributed sampling scheme, through which innovations can propagate effectively across the network for estimation or detection purposes.

The paper is organized as follows. Section II describes the system model for innovations diffusion sampling in sensor networks. The proposed sampling scheme for estimation and detection is presented in Section III. Section IV provides performance analysis for the innovations diffusion sampling algorithm and Section V discusses some important issues of its practical implementation. Simulation results illustrating the effectiveness of the proposed algorithm are given in Section VI. Section VII concludes the paper.

## II. System Model

Consider a wireless network with $N$ sensor nodes spatially distributed in the field. The network wishes to select a subset of sensor measurements to estimate an unknown parameter or detect a known deterministic signal in a *distributed* manner by relying on local computations and inter-sensor communications.

### A. Network Graph

The sensor network is represented as an undirected graph $G(\mathcal{V},\mathcal{E})$, where $\mathcal{V}=\{1,2,\ldots,N\}$ denotes the set of vertices (sensor nodes) and $\mathcal{E}\subset\{(i,j)\mid i,j\in\mathcal{V}\}$ is the edge set. A graph is connected if there exists a path in $\mathcal{E}$ for any two vertices $i$ and $j$. In this paper, we assume that each sensor transmits at a constant power level $P$ and the receiver has an ambient noise power level $N_0$. Let $X(i)$ represent the position of sensor $i$. Then, the transmission from sensor $i$ is successfully received by sensor $j$ if

$$\frac{P}{N_0 d_{ij}^{\alpha}}\geq\beta \tag{18}$$

where $d_{ij}=|X(i)-X(j)|$ is the distance between sensors $i$ and $j$, and $\alpha(2<\alpha<6)$ is the associated pathloss coefficient. Equation (18) models a situation where a minimum signal-to-noise ratio (SNR) is necessary for successful receptions and the signal power decays with distance $r$ according to the rule $1/r^{\alpha}$. Thus, the receiver should be within a distance of

$$r_\beta=\sqrt[\alpha]{\frac{P}{N_0\beta}} \tag{19}$$

from the transmitter in order for a reliable wireless link to be ensured. In this way, the edge set is defined by

$$\mathcal{E}\triangleq\{(i,j)\mid d_{ij}\leq r_\beta\} \tag{20}$$

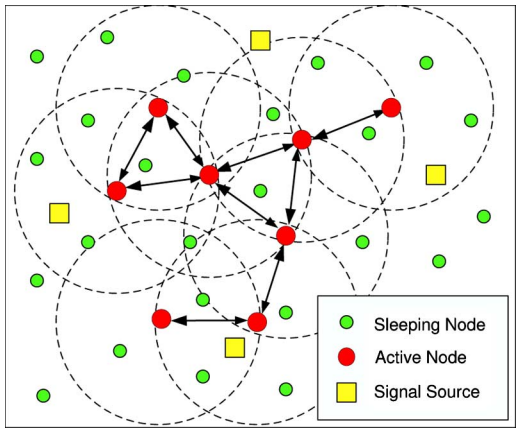


Fig. 1. A schematic representation of distributed sampling in wireless sensor networks.

and the neighbor set of sensor $i$ is defined as

$$\mathcal{N}(i) \triangleq \{j \mid (i,j) \in \mathcal{E}\} \tag{21}$$

so that sensors $i$ and $j$ are termed *neighbors* if their distance is less than or equal to $r_\beta$. As illustrated in Fig. 1, sensors within the connectivity radius $r_\beta$ are directly connected and are neighbors of each other.

For any sampling decision $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$, the active and inactive neighbors of sensor $i_k (i_k \in \mathcal{A}_n)$ are given by

$$\mathcal{N}_A(i_k) = \mathcal{N}(i_k) \cap \mathcal{A}_n \tag{22}$$

and

$$\mathcal{N}_I(i_k) = \mathcal{N}(i_k) \setminus \mathcal{A}_n. \tag{23}$$

The local knowledge of each sensor $i$ contains not only its own distortion matrix $\mathbf{H}_i$ and noise covariance $\mathrm{E}\mathbf{v}_i\mathbf{v}_i^T$, but also its neighbors' distortion matrices, noise covariances, and correlations. Let $\mathcal{N}'(i) = \{\mathcal{N}(i), i\}$. Then, each sensor $i$ is assumed to have access to the information $\{\mathbf{H}_{\mathcal{N}'(i)}, \mathbf{C}_{\mathcal{N}'(i)}\}$ at the stage of deployment.

### B. Spatial Correlation Models

The computation of the linear m.v.u.e. of $\boldsymbol{\theta}$ as in (3) and the optimal detector from the noisy measurements as in (9), can be facilitated if the network has *a priori* knowledge of the covariance structure $\mathbf{C}$. In practice, the matrix $\mathbf{C}$ can be estimated from repeated measurements at all sensors. Nevertheless, this approach is not efficient if the number of sensors $N$ is large. Therefore, a model that characterizes the correlation between sensors will be useful. In this paper, we consider three different correlation models as follows.

*1) Uncorrelated Noise (UN):* The measurement noise at each sensor is uncorrelated with other sensors. If the noises are Gaussian, then they are independent with each other as well.

*2) Near Correlation (NC):* In this model, each sensor is only correlated with its neighboring nodes and is independent of other sensors beyond the distance $r_\beta$, i.e., $\mathbf{C}_{ij} = \mathbf{0}$ if $j \notin \mathcal{N}(i)$. This is motivated by the fact that the correlation decays with the distance between sensors and approaches zero if the corresponding sensors are far apart [20].

*3) Far Correlation (FC):* The measurement noise at each sensor is not only correlated with its own neighbors, but also with other sensors far away. The correlation is assumed to be a decreasing function of the Euclidean distance between the corresponding nodes [20], [21], i.e.

$$\mathbf{C}_{ij} = \mathbf{F}(\mathbf{C}_{ii}, \mathbf{C}_{jj}, d_{ij}) \tag{24}$$

where $\mathbf{F} : \mathbb{R}^{n_i \times n_i} \times \mathbb{R}^{n_j \times n_j} \times \mathbb{R} \mapsto \mathbb{R}^{n_i \times n_j}$.

These correlation models will be used to test our distributed sampling algorithms.

## III. Innovations Diffusion Sampling

In this section, we present a distributed sampling scheme for estimation and detection under a certain distortion constraint. Starting with an initial sensor, the set of active sensors collaboratively activates one sleeping sensor at each time. The procedure continues until the set of active sensors achieves the desired estimation or detection fidelity. Although there is no direct relation between the system lifetime and the number of active sensors, reducing the number of active sensors generally leads to energy efficiency. In general, finding the exact minimum number of measurements to achieve a desired fidelity belongs to a class of combinatorial optimization problems [22] and is typically NP-hard. Thus, the proposed approach will pursue a greedy heuristic that provides useful suboptimal approximations with polynomial complexity. Assuming that each inactive sensor can listen to the control channel for the awakening message, at each time the sampling algorithm should select the most informative sensor with respect to the previous selected ones in order for the fidelity to be satisfied with a nearly minimum number of active sensors. The resulting procedure will exhibit good performance at reasonable cost and will be tractable for both analysis and implementation.

### A. Uncorrelated Noise

Consider a selected group of sensor observations $\mathbf{y}_{\mathcal{A}_n}$, where $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$. If the noises are spatially uncorrelated, i.e., $\mathbf{C}_{kl} = \mathbf{0}$ for $k \neq l$, then (4) becomes

$$\mathbf{D}^{-1}(\mathbf{y}_{\mathcal{A}_n}) = \sum_{k=1}^{n} \mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k} \tag{25}$$

and, from (15), the deflection coefficient is

$$d^2(\mathbf{y}_{\mathcal{A}_n}) = \boldsymbol{\theta}^T \left( \sum_{k=1}^{n} \mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k} \right) \boldsymbol{\theta}. \tag{26}$$

These expressions decouple the contribution of each sensor to the total inference fidelity (MMSE or the deflection coefficient). Each term $\mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k}$ has the essential properties of an information measure in that it is

1) nonnegative definite,
2) and additive for independent observations.

Intuitively, the more information sensor $i_k$ has (i.e., the larger $\mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k}$ is), the smaller the MMSE and the larger the

deflection coefficient. This suggests that at each time the network should choose the most informative sensor in order to maximally reduce the MMSE or increase the deflection coefficient. Specifically, assume that the nodes $\mathcal{A}_{n-1}$ have been selected. Then, at time $n$ we would choose from the *coverage* of $\mathcal{A}_{n-1}$ a sensor $i_n$ whose information measure $\mathbf{H}_{i_n}^T \mathbf{C}_{i_n}^{-1} \mathbf{H}_{i_n}$ is the largest. In this way, the resulting inference fidelity will be the best compared with other possible choices.

### B. Correlated Noise

However, the noises are generally spatially correlated in practice. In this case, the contributions of the individual sensors are coupled with each other in the MMSE expression $\mathbf{D}(\mathcal{A}_n)$ in (4) and in the deflection coefficient (15). In other words, it is not sufficient to examine the term $\mathbf{H}_{i_n}^T \mathbf{C}_{i_n}^{-1} \mathbf{H}_{i_n}$ and seek the one with the largest value as in the case of uncorrelated noise. We thus need to develop a procedure to find the most informative sensor with respect to the previous selected ones. To achieve this goal, we start by whitening the observation data subject to the order dictated by the choice of sensors, and then obtain a set of transformed measurements with uncorrelated noises.

Suppose that we have already selected $n-1$ sensors, i.e., $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$. For every possible $i_n \notin \mathcal{A}_{n-1}$, we define its innovation [4], [23] as

$$\mathbf{e}_{i_n} \triangleq \mathbf{y}_{i_n} - \hat{\mathbf{y}}_{i_n|\mathcal{A}_{n-1}} \tag{27}$$

where $\hat{\mathbf{y}}_{i_n|\mathcal{A}_{n-1}}$ denotes the linear least-mean-squares estimator (l.l.m.s.e. or projection) of $\mathbf{y}_{i_n}$ given the previously selected measurements $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$. It is straightforward to see that

$$\hat{\mathbf{y}}_{i_n|\mathcal{A}_{n-1}} = \mathbf{H}_{i_n}\boldsymbol{\theta} + \hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} \tag{28}$$

where $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$ represents the projection of $\mathbf{v}_i$ onto the same linear space $\mathcal{L}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$. The quantity $\mathbf{e}_{i_n}$ in (27) possesses the new information contained in sensor $i_n$ and not in any of the previously selected measurements. Now note that $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$ is given by [4]:

$$\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} = \mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} \mathbf{v}_{\mathcal{A}_{n-1}} \tag{29}$$

where

$$\mathbf{B}_{i_n|\mathcal{A}_{n-1}} = \begin{pmatrix} \mathbf{C}_{i_n i_1} & \mathbf{C}_{i_n i_2} & \ldots & \mathbf{C}_{i_n i_{n-1}} \end{pmatrix}. \tag{30}$$

Combining (1), (27), (28), and (29) gives

$$\begin{aligned} \mathbf{e}_{i_n} &= \mathbf{v}_{i_n} - \hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} \\ &= \begin{pmatrix} -\mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{\mathcal{A}_{n-1}} \\ \mathbf{v}_{i_n} \end{pmatrix}. \end{aligned} \tag{31}$$

It can be verified that $\mathbf{e}_{i_n} \perp \mathbf{v}_j$ for any $j \in \mathcal{A}_{n-1}$, that is, $\mathrm{E}\left(\mathbf{e}_{i_n} \mathbf{v}_j^T\right) = 0$. Fig. 2 illustrates a geometric interpretation of the projection relationship between $\mathbf{e}_i$, $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$, and $\mathbf{v}_i$. Let

$$\mathbf{P}_{i_n|\mathcal{A}_{n-1}} \triangleq \begin{pmatrix} -\mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix}. \tag{32}$$

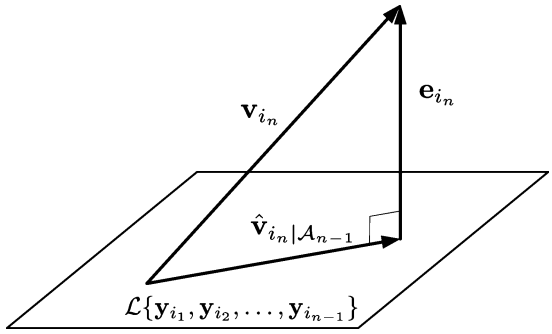


Fig. 2. A geometric interpretation of the relation between vectors $\mathbf{v}_{i_n}$, $\mathbf{e}_{i_n}$, and $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$ with respect to the affine space $\mathcal{L}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$.

*Lemma 1 (Innovations):* Given a sampling decision $\mathcal{A}_n$, the corresponding innovation process $\{\mathbf{e}_{i_k}\}_{k=1}^n$ has the important property

$$\mathrm{E}\left(\mathbf{e}_{i_k} \mathbf{e}_{i_l}^T\right) = \begin{cases} \mathbf{0} & i_k \neq i_l \\ \mathbf{Q}_{i_k} & \text{otherwise} \end{cases} \tag{33}$$

where $\mathbf{Q}_{i_k}$ is the covariance matrix of $\mathbf{e}_{i_k}$ and is given by

$$\mathbf{Q}_{i_k} = \mathbf{C}_{i_k i_k} - \mathbf{B}_{i_k|\mathcal{A}_{k-1}} \mathbf{C}_{\mathcal{A}_{k-1}}^{-1} \mathbf{B}_{i_k|\mathcal{A}_{k-1}}^T. \tag{34}$$

*Proof:* Refer to Appendix I. ■

We now introduce a transformed form of $\mathbf{y}_{i_n}$ with respect to $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$ as follows:

$$\mathbf{z}_{i_n} \triangleq \mathbf{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \mathbf{y}_{\mathcal{A}_{n-1}} \\ \mathbf{y}_{i_n} \end{pmatrix} \tag{35}$$

which can be written as

$$\mathbf{z}_{i_n} = \mathbf{G}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{\theta} + \mathbf{e}_{i_n} \tag{36}$$

where

$$\mathbf{G}_{i_n|\mathcal{A}_{n-1}} \triangleq \mathbf{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \mathbf{H}_{\mathcal{A}_{n-1}} \\ \mathbf{H}_{i_n} \end{pmatrix}. \tag{37}$$

Note in particular that for all $i_k \in \mathcal{A}_{n-1}$, it holds that $\mathrm{E}\left[\left(\mathbf{z}_{i_n} - \mathrm{E}\mathbf{z}_{i_n}\right)^T \left(\mathbf{y}_{i_k} - \mathrm{E}\mathbf{y}_{i_k}\right)\right] = 0$. That is,

$$\mathbf{z}_{i_n} \perp \mathcal{L}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}. \tag{38}$$

Moreover, the main fact to note is that now the linear m.v. u.e. of $\boldsymbol{\theta}$ given $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_n}\}$ coincides with the m.v.u.e. of $\boldsymbol{\theta}$ using $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$. This result is stated as follows.

*Theorem 1 (Equivalent Estimation):* Consider the linear model in (1). For a given set of sensor measurements $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$, the minimum-variance-unbiased linear estimator of $\boldsymbol{\theta}$ is equivalent to the estimator obtained by using the transformed measurements $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_n}\}$, i.e.

$$\begin{aligned} \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n}) &= \hat{\boldsymbol{\theta}}(\mathbf{z}_{\mathcal{A}_n}) \\ &= \left( \sum_{k=1}^n \mathbf{G}_{i_k|A_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{G}_{i_k|A_{k-1}} \right)^{-1} \\ &\quad \times \sum_{k=1}^n \mathbf{G}_{i_k|A_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{z}_{i_k}. \end{aligned} \tag{39}$$

Moreover, the resulting MMSEs are identical, i.e.,

$$\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_n}\right) = \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} = \left(\sum_{k=1}^{n} \mathbf{G}_{i_k|\mathcal{A}_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{G}_{i_k|\mathcal{A}_{k-1}}\right)^{-1}. \quad (40)$$

*Proof:* See Appendix II. ∎

The key advantage of working with the transformed quantities $\{\mathbf{z}_{i_k}\}_{k=1}^n$ over the original measurements $\{\mathbf{y}_{i_k}\}_{k=1}^n$ is that the noises $\{\mathbf{e}_{i_k}\}_{k=1}^n$ in the model (36) are now uncorrelated and the contributions of the individual sensors to the MMSE can be decoupled, as was the case with uncorrelated noise in Section III-A. We can also obtain a similar result for the optimal detector, as stated in the following theorem.

*Theorem 2 (Equivalent Detection):* Consider the binary hypothesis test in (6). For a given set of sensor measurements $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$, the optimal detector is equivalent to the detector obtained by using the transformed measurements $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_n}\}$, i.e.

$$L(\mathbf{y}_{\mathcal{A}_n}) = L(\mathbf{z}_{\mathcal{A}_n}) = 2\boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \mathbf{Q}_{\mathcal{A}_n}^{-1} \mathbf{z}_{\mathcal{A}_n} - \boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \mathbf{Q}_{\mathcal{A}_n}^{-1} \mathbf{G}_{\mathcal{A}_n} \boldsymbol{\theta}. \quad (41)$$

Moreover, the resulting deflection coefficients are identical, i.e.

$$d^2(\mathbf{y}_{\mathcal{A}_n}) = d^2(\mathbf{z}_{\mathcal{A}_n}). \quad (42)$$

*Proof:* See Appendix III. ∎

### C. Iterative Diffusion Sampling

Diffusion is the means by which innovations become available by communications throughout the current active sensors. The sampling algorithm assumes an active node at the beginning, i.e., $\mathcal{A}_1 = \{i_1\}$, which serves as the seed to activate other sensors. Sensor $i_1$ chooses within its inactive neighbors a sensor and then activates it. At the $n$th step, the current set of active sensors $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$ makes a connected network and collaboratively activates one sleeping sensor that is within the coverage of $\mathcal{A}_{n-1}$. To meet the fidelity requirement with a nearly minimum number of measurements, the sampling algorithm should choose the sensor that is most informative with respect to the previous sensors $\mathcal{A}_{n-1}$. Before we proceed to describe the algorithm, we need to define a utility function that measures the innovation of the sensor measurement.

*1) Estimation:* Motivated by (40), we define a *utility matrix* as

$$\mathbf{U}_{i_n|\mathcal{A}_{n-1}} \triangleq \mathbf{G}_{i_n|\mathcal{A}_{n-1}}^T \mathbf{Q}_{i_n}^{-1} \mathbf{G}_{i_n|\mathcal{A}_{n-1}} \quad (43)$$

which has the essential properties of an information measure. Recall that the set of sensor observations $\mathcal{A}_{n-1}$ has an MMSE $D_{\mathcal{A}_{n-1}} = \mathrm{Tr}\left[\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right)\right]$. From (40) and for any $i_n$, the MMSE with measurements $\{\mathcal{A}_{n-1}, i_n\}$ can be written as

$$D_{\mathcal{A}_n} = \mathrm{Tr}\left[\left(\mathbf{D}^{-1}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right) + \mathbf{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right]. \quad (44)$$

Then, we can define a utility function as the difference between $D_{\mathcal{A}_{n-1}}$ and $D_{\mathcal{A}_n}$, i.e., [see (45) at the bottom of the page], where the last equality follows from the matrix inversion lemma.[1] Although the definition of the utility function is not unique, it can be seen that $u(i_n|\mathcal{A}_{n-1})$ in (45) indicates a good measure of the new information provided by sensor $i_n$.

*2) Detection:* For detection of a deterministic parameter $\boldsymbol{\theta}$, we define the utility function as

$$u(i_n|\mathcal{A}_{n-1}) \triangleq \boldsymbol{\theta}^T \mathbf{G}_{i_n|\mathcal{A}_{n-1}}^T \mathbf{Q}_{i_n}^{-1} \mathbf{G}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{\theta} \quad (46)$$

according to (26).

The basic strategy of diffusion sampling is to successively choose one sensor with maximum utility from among the sleeping nodes within the coverage of $\mathcal{A}_{n-1}$, i.e.

$$i_n = \arg \max_{j \in \bigcup_{k=1}^{n-1} \mathcal{N}_I(i_k)} u\left(j|\mathcal{A}_{n-1}\right) \quad (47)$$

by local computation and message exchange between neighboring sensors. The procedure continues until the desired fidelity ($D_0$ or $d_0^2$) is achieved. The details of the sampling algorithm are presented in Algorithm 1.

**Algorithm 1** Innovations Diffusion Sampling

0: Start with $n = 1$ and $\mathcal{A}_1 = \{i_1\}$.

1: **while** $D_{\mathcal{A}n-1} > D_0$ [or $d^2(\mathbf{y}_{\mathcal{A}_{n-1}}) < d_0^2$] **do**

2: For each $i_k \in \mathcal{A}_{n-1}$, find $i_k^{\max} = \arg\max_{j \in \mathcal{N}_I(i_k)} u(j|\mathcal{A}_{n-1})$

3:**repeat**

4: Each active node $i_k$ sends a message $M_{i_k}$ including the index $i_k^{\max}$ and the associated utility $u(i_k^{\max}|\mathcal{A}_{n-1})$ to its active neighbors $\mathcal{N}_A(i_k)$.

5: Upon receiving the messages from neighboring nodes, each sensor $i_k$ in $\mathcal{A}_{n-1}$ updates $i_k^{\max} : i_k^{\max} = \arg\max_{j \in \mathcal{N}_A(i_k)} u(i_j^{\max}|\mathcal{A}_{n-1})$ and stores the associated utility.

[1] $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$ [4].

$$\begin{aligned} u(i_n|\mathcal{A}_{n-1}) &\triangleq D_{\mathcal{A}_{n-1}} - D_{\mathcal{A}_n} \\ &= \mathrm{Tr}\left[\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right) - \left(\mathbf{D}^{-1}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right) + \mathbf{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right] \\ &= \mathrm{Tr}\left[\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right)\left(\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right) + \mathbf{U}_{i_n|\mathcal{A}_{n-1}}^{-1}\right)^{-1}\mathbf{D}\left(\mathbf{y}_{\mathcal{A}_{n-1}}\right)\right] \end{aligned} \quad (45)$$

6:**until** all the active sensors reach consensus.

7: Let $i_n = i_k^{\max}$.

8: Sensor $i_n$ is activated by its closest active neighbor and inherits the necessary information related to $\mathcal{A}_{n-1}$.

9: Each active sensor adds $i_n$ into the set of active sensors, i.e., $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$ and $n = n + 1$.

10: **end while**

### D. Recursive Algorithm

When the number of active nodes $n$ increases, the computational complexity of inverting the correlation matrix $\mathbf{C}_{\mathcal{A}_n}$ in (32) can become prohibitively high, e.g., $\mathcal{O}(m^3)$, where $m = \sum_{i=1}^{n} n_i$. Nevertheless, the computational complexity can be reduced by relating $\mathbf{C}_{\mathcal{A}_n}$ to $\mathbf{C}_{\mathcal{A}_{n-1}}$.

The correlation of the nodes $\mathcal{A}_n$ can be recursively represented as follows:

$$\mathbf{C}_{\mathcal{A}_n} = \begin{pmatrix} \mathbf{C}_{\mathcal{A}_{n-1}} & \mathbf{B}_{i_n|\mathcal{A}_{n-1}}^T \\ \mathbf{B}_{i_n|\mathcal{A}_{n-1}} & \mathbf{C}_{i_n i_n} \end{pmatrix}. \tag{48}$$

Then

$$\mathbf{C}_{\mathcal{A}_n}^{-1} = \begin{pmatrix} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \begin{pmatrix} -\mathbf{C}_{\mathcal{A}_{n-1}}^{-1}\mathbf{B}_{i_n|\mathcal{A}_{n-1}}^T \\ \mathbf{I} \end{pmatrix} \times \mathbf{Q}_{i_n}^{-1} \begin{pmatrix} -\mathbf{B}_{i_n|\mathcal{A}_{n-1}}\mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix} \tag{49}$$

where $\mathbf{Q}_{i_n} = \mathbf{C}_{i_n i_n} - \mathbf{B}_{i_n|\mathcal{A}_{n-1}}\mathbf{C}_{\mathcal{A}_{n-1}}^{-1}\mathbf{B}_{i_n|\mathcal{A}_{n-1}}^T$ is a small $n_{i_n} \times n_{i_n}$ matrix and its inversion is generally straightforward. With this recursion, the computational complexity of matrix inversion can be reduced by exploiting the previous result, and each sensor is able to efficiently compute the estimator or detector and the associated inference error.

### E. Diffusion Protocol

The communication protocols for innovation diffusion can be implemented in a couple of ways.

*1) Global Knowledge:* Assume that each node has full knowledge of the distortion matrices $\{\mathbf{H}_k\}$ and the noise covariance matrices $\{\mathbf{C}_{kl}\}$ of its neighbors. Then each active sensor can locally compute (1) without communicating with other sensors. The messages exchanged between sensors only need to contain the index and the associated utility function, i.e.

$$M_{i_k} = \{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1})\}. \tag{50}$$

*2) Local Knowledge:* In the case where each sensor has only local knowledge about its neighbors, the diffusion sampling algorithm can be implemented by adding necessary information into the messages that are exchanged among active nodes. For example, the message sent by sensor $i_k$ in line 4 could contain the distortion matrix and noise covariance of sensor $i_k^{\max}$ in addition to $\{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1})\}$, i.e.

$$M_{i_k} = \{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1}), \mathbf{H}_{i_k^{\max}}, \mathbf{C}_{i_k^{\max} i_k^{\max}}\}. \tag{51}$$

When sensor $i_n$ is activated in line 8, it inherits the information of $\{\mathbf{H}_{\mathcal{A}_{n-1}}, \mathbf{C}_{\mathcal{A}_{n-1}}\}$ from its closest active neighbor and locally constructs its own $\{\mathbf{H}_{\mathcal{A}_n}, \mathbf{C}_{\mathcal{A}_n}\}$.

To avoid transmitting redundant information, in line 4 each active sensor sends the message $M_{i_k}$ to its neighbors only if the message content has been updated; otherwise, it turns off radio to save energy. That is, the innovations computation of a candidate node is performed locally by its neighbors and there is no need to transmit its distortion and correlation matrices. Only when a new node is activated, its distortion and correlation information will be shared by the set of active nodes.

*3) Location-Aware Diffusion:* As the noise correlation between sensors can be modeled as a function of their locations (or distances) as in (24), the correlation information is not transmitted in the message exchanged among nodes, i.e.

$$M_{i_k} = \{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1}), \mathbf{H}_{i_k^{\max}}\}. \tag{52}$$

For the UN and NC models, constructing $\mathbf{C}_{\mathcal{A}_n}$ is straightforward. If sensor noises are characterized by the FC model, then each active sensor can locally construct $\mathbf{C}_{\mathcal{A}_n}$ using (24).

Within the set of active nodes, each examines its neighbors and finds the best local candidate. The computation of innovations is only performed locally without invoking any message exchange. Each node then exchanges the information of its candidate such that the most informative node can be identified in a distributed way. Meanwhile, each sensor does not have to transmit the information of its candidate every time unless its candidate has been updated by its active neighboring nodes. This simple mechanism can avoid unnecessary transmission and save a significant amount of energy. Since only the information of the newly selected node is shared among the current set of active nodes, the diffusion procedure does not cause much overhead.

Since the number of active nodes is finite, consensus can be reached in line 6 within a finite number of iterations. Then, each sensor can locally construct the matrices $\mathbf{H}_{\mathcal{A}_n}$ and $\mathbf{C}_{\mathcal{A}_n}$, which will be used for activating the next sensor if $D_{\mathcal{A}_n} > D_0$ or $d^2(\mathbf{y}_{\mathcal{A}_n}) < d_0^2$. After the sampling stage is completed, $\mathbf{H}_{\mathcal{A}_n}$ and $\mathbf{C}_{\mathcal{A}_n}$ are used for data fusion. In brief, the algorithm successively adds the sensor that has the maximum utility into the set of selected sensors until the desired inference fidelity is satisfied. This scheme ensures that the set of selected nodes make a connected graph at any time and the number of active sensors is no more than necessary.

## IV. Energy Performance Analysis

It is useful to analyze the proposed innovations diffusion algorithm from a probabilistic perspective. Recall that the innovations diffusion algorithm, starting with an initial node, repeatedly activates one node at a time until a certain inference fidelity is satisfied. As such, the number of active sensors is a random number that determines the amount of energy consumed for computation and communication. That is, the less active sensors we have, the more energy will be saved. We estimate the expected number of active sensors in the sequel.

For simplicity, we consider the scalar observation $y_i = \mathbf{H}_i\boldsymbol{\theta} + v_i$, $i = 1, 2, \ldots, N$, where $y_i$, $\mathbf{H}_i\boldsymbol{\theta}$, and $v_i$ are scalars. The innovation utility derived from (43) is a scalar with the form

$$U_{i_n|\mathcal{A}_{n-1}} \triangleq G_{i_n|\mathcal{A}_{n-1}}^T Q_{i_n}^{-1} G_{i_n|\mathcal{A}_{n-1}} \tag{53}$$

where $G_{i_n|\mathcal{A}_{n-1}}$ and $Q_{i_n}$ are the corresponding scalar forms derived from (37) and (34). Thus, the inverse of the MMSE can be expressed as

$$D_{\mathcal{A}_n}^{-1} = \sum_{k=1}^{n} U_{i_k|\mathcal{A}_{k-1}}. \tag{54}$$

The algorithm will stop if, and only if, $D_{\mathcal{A}_{N_a}}^{-1} \geq D_0^{-1}$, where $N_a$ is the stopping rule for a set of random variables $\{U_{i_n|\mathcal{A}_{n-1}}; n \geq 1\}$. Thus, we can define an indicator function of the event $\{N_a \geq k\}$ as

$$I_k = \begin{cases} 1, & N_a \geq k \\ 0, & N_a < k. \end{cases} \tag{55}$$

Considering that the sensors are randomly deployed in a network with a large number of sensors, we can model innovation diffusion as a renewal process for which the $\{U_{i_n|\mathcal{A}_{n-1}}; n \geq 1\}$ are independent identically distributed (i.i.d.) random variables. Let $\mu_u = \mathbf{E}\left[U_{i_n|\mathcal{A}_{n-1}}\right]$. From Wald's equality [24]

$$\begin{aligned} \mathbf{E}\left[D_{\mathcal{A}_{N_a}}^{-1}\right] &= \mathbf{E}\left[\sum_{k=1}^{\infty} U_{i_k|\mathcal{A}_{k-1}} I_k\right] \\ &= \mu_u \sum_{k=1}^{\infty} \mathbf{E}(I_k) \\ &= \mu_u \mathbf{E}\left(N_a\right) \\ &= \mu_u \bar{N}_a \end{aligned}$$

where $\bar{N}_a = \mathbf{E}\left(N_a\right)$. Consequently, the expected number of active sensors can be estimated as

$$\bar{N}_a = \left\lceil \frac{1}{\mu_u D_0} \right\rceil \tag{56}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x$.

We can now approximate the energy consumption of the distributed algorithm. Suppose that each node spends an amount of energy $E^{\text{comm}}$ on message passing and $E^{\text{comp}}$ on computation at each time. Specifically, $E^{\text{comm}}$ is proportional to the power function of the radio range, i.e., $E^{\text{comm}}(r_\beta) \sim r_\beta^\alpha$, and $E^{\text{comp}}$ is proportional to the expected number of neighbors $|\mathcal{N}|$ of a single node, i.e., $E^{\text{comp}}(|\mathcal{N}|) \sim |\mathcal{N}|$. In the worst case where the topology of the active nodes is a straight line, the amount of energy needed to accomplish the sampling requirement can be estimated as[2]

$$\begin{aligned} &E^{\text{total}}\left(N_a\right) \\ &\quad = E^{\text{comm}}(r_\beta)\left(\sum_{k=1}^{N_a-1} k(k-1)\right) \\ &\qquad + E^{\text{comp}}(|\mathcal{N}|)\left(\sum_{k=1}^{N_a-1} k\right) \\ &\quad = E^{\text{comm}}(r_\beta)\frac{N_a(N_a-1)(N_a-2)}{3} \\ &\qquad + E^{\text{comp}}(|\mathcal{N}|)\frac{N_a^2 - N_a}{2} \\ &\quad = \mathcal{O}(N_a^3). \end{aligned}$$

[2]The notation $y = \mathcal{O}(g(N))$ denotes that there exits a constant $\kappa$ such that $\lim_{N \to \infty} y/g(N) \leq \kappa$.

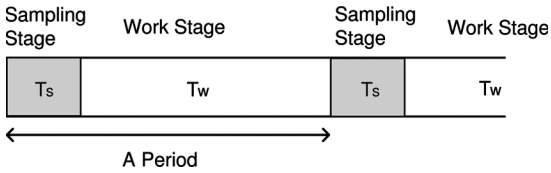


Fig. 3. The sampling operation is shown along the time scale. Once a subset of sensors has been activated at the end of the sampling stage, the sensors are used to compute the linear m.v.u.e. or the detector in the work stage. In general, the duration of the work stage is much longer than that of the sampling stage, i.e., $T_s \gg T_w$, so that the cost caused by sampling does not result in much overhead.

Namely, the expected energy consumption is of the order of $N_a^3$. The analysis suggests that by choosing the most informative sensors and reducing the number of active sensors, the innovations diffusion algorithm can save a significant amount of energy that might be unnecessarily used otherwise.

## V. Implementation Issues

In this section, we address some practical implementation issues of the proposed sampling algorithm and compare it with some related work in the literature.

### A. Continuous Sampling

The network operation is divided into periods. As shown in Fig. 3, each work period begins with a sampling stage when a group of sensors is activated using the sampling algorithm. Following the sampling stage is the work stage, during which the active sensors are used to jointly estimate the unknown parameter $\boldsymbol{\theta}$ or detect the signal of interest and will be powered off at the end of the work period. Without loss of generality, the duration of the work stage is assumed to be much longer than that of the sampling stage such that the sampling cost accounts for only a small fraction of sensor energy. For example, the sampling stage can be completed in a few seconds or minutes while the work stage will last up to hours or days.

The sampling algorithm supports the continued operation mode. In this mode, given a seed node at the beginning of each work period, the network selects a subset of sensors for estimation or detection. At the end of the work period, these active sensors $\mathcal{A}_n$ collaboratively find a sensor within their neighborhood $\bigcup_{i \in \mathcal{A}_n} \mathcal{N}_I(i)$, as the seed node for the next sampling stage according to a certain criterion. The active sensors activate the new seed node and turn themselves off afterwards. Then, a new work period starts and a new group of sensors will be activated using the same mechanism. The procedure repeats as the sensor network operates.

There are many options to choose the seed node according to different application requirements. For example, the network would like to avoid the hot spot problem (i.e., some sensors are more active than others and thus would die at an early stage) and maintain its connectivity as long as possible. A simple strategy would be to choose the sensor with the maximum leftover energy so that the energy load can be evenly distributed over all the sensors in the network. For a system that wishes to use a smaller number of sensors for estimation or detection during some period, it may want to start with a sensor with better observation

quality. These options would provide system designers with a deal of flexibility for practical network deployment.

### B. Asynchronous Implementation

The proposed sampling algorithm can be implemented in an asynchronous manner such that the synchronization penalty can be substantially reduced [25]. During the sampling stage, each active node can compute the innovations of its neighbors independently, and the computations can vary widely in time with an attendant effect on the final computed results. In the work stage, the active sensors can also compute the estimator or detector independently and asynchronously with whatever data they have and they do not have to wait for all the measurements to become available if the requirement on the resulting fidelity is not stringent. With asynchronous implementation, the networked sensors neither need any global clock to time the operations, nor have to wait for predetermined data to become available.

### C. Scalability and Robustness

Network expansion, shrinkage or sensor replacement cause no disturbance in the proposed distributed scheme. In practice, it is quite possible that some sensors will fail or be blocked due to lack of energy, physical damage, or environmental interference. Thus, the network may shrink due to sensor failure, or may expand with newly deployed sensors. By updating the data of the neighbors, the newly arriving sensors can be transparently incorporated into the network and the failed sensors can be easily replaced. The network scalability is related to the convergence rate of consensus. For example, if the sensor network adapts CSMA/CD in the medium access control (MAC) layer, then the collision probability will increase as the number of active sensors grows, thus taking more time for the algorithm to reach consensus.

## VI. Numerical Studies

This section presents numerical results that illustrate the effectiveness of the proposed sampling scheme. We are particularly interested in the inference error and the operational system lifetime of the network.

### A. Simulation Setup

We randomly generate $N$ sensor nodes in the unit square $[0,1]$ by $[0,1]$. Consider a simple linear model $y_i = \theta + v_i$, $i = 1,2,\ldots,N$, where $\theta$ as a scalar. The noise covariance matrix $\mathbf{C}$ is randomly generated according to the spatial correlation model

$$C_{ij} = \begin{cases} \sigma_i^2 & i = j \\ \sigma_i \sigma_j \exp\left(-\eta d_{ij}^2\right) & i \neq j \end{cases} \tag{57}$$

where $\sigma_i^2$, $i = 1,2,\ldots,N$, are randomly generated with a uniform distribution in $(0,1]$. The correlation is characterized as a decreasing function of the distance between the corresponding nodes. $\eta > 0$ is a parameter that measures the degree of correlation between nodes. If $\eta$ is small, the correlation is high; otherwise, the correlation is weak.

### B. Inference Error Analysis

*1) Estimation:* The fidelity of computing the estimator $\hat{\boldsymbol{\theta}}$ is influenced by the choice of the correlation model. Suppose that

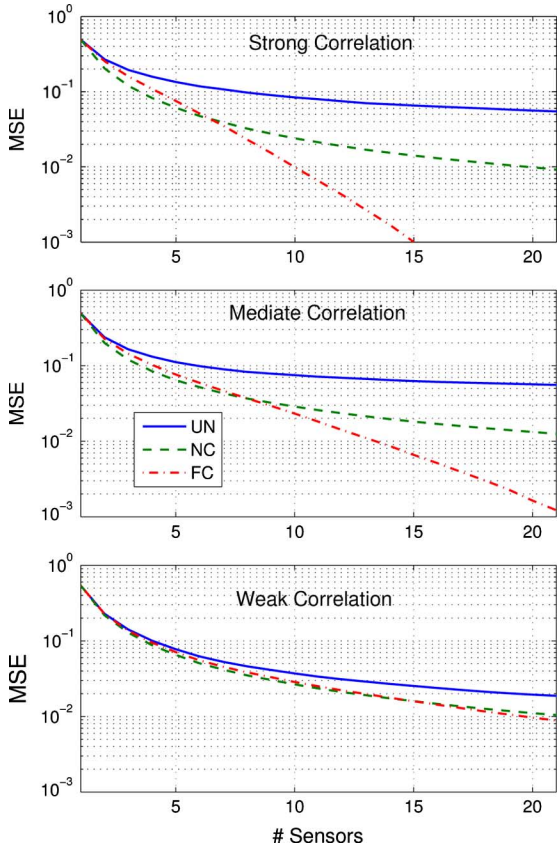


Fig. 4. The MSE's of the UN, NC, and FC models ($r_\beta = 0.15$). For the scenarios of strong, medium, and weak correlation, we let $\eta = 5, 10$, and 50, respectively. The plots are results averaged over 2000 simulations.

$\mathbf{C}_{\mathcal{A}_n}$ is the actual noise covariance matrix and $\hat{\mathbf{C}}_{\mathcal{A}_n}$ is the estimated noise covariance matrix used to compute the estimator. Then, the estimator based on measurements $\mathbf{y}_{\mathcal{A}_n}$ using $\hat{\mathbf{C}}_{\mathcal{A}_n}$ is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1 &= \left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{y}_{\mathcal{A}_n} \\ &= \boldsymbol{\theta} + \left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{v}_{\mathcal{A}_n} \end{aligned} \tag{58}$$

and the corresponding MSE is

$$\begin{aligned} \mathbf{D}_1 &= \mathrm{E}\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_1\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_1\right)^T\right] \\ &= \mathrm{E}\left[\left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{v}_{\mathcal{A}_n} \mathbf{v}_{\mathcal{A}_n}^T \right. \\ &\qquad \left. \times\, \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n} \left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1}\right] \\ &= \left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{C}_{\mathcal{A}_n} \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n} \\ &\qquad \times \left(\mathbf{H}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1}. \end{aligned} \tag{59}$$

The MSE of the FC model is actually the MMSE of the m.v.u.e. Fig. 4 shows how the MSEs of the UN and NC models differ from the MMSE of the m.v.u.e. with respect to the number of sensors under various degrees of correlation.

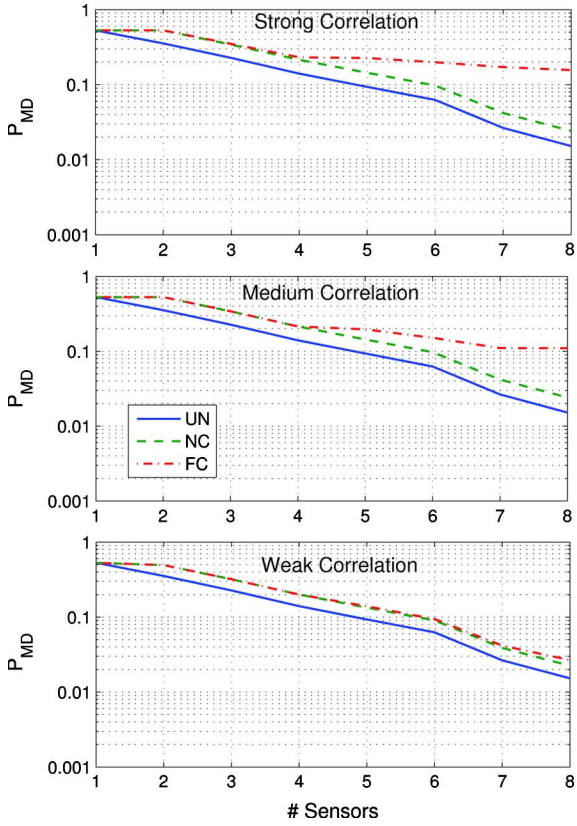


Fig. 5. The probabilities of missed detection ($P_{MD} = 1 - P_D$) of UN, NC, and FC models for a given probability of false alarm $P_{FA} = 0.1$ ($r_\beta = 0.15$). For the scenarios of strong, medium, and weak correlation, we let $\eta = 5, 10$, and 50, respectively. The plots are results averaged over 2000 simulations.

It can be shown that the performance of the UN and NC models approximates that of the FC model when the correlation between the nodes is not strong. Therefore, the condition of weak correlation may relax the requirement of sensor location for accurately estimating the covariance matrix $\mathbf{C}_{\mathcal{A}_n}$.

*2) Detection:* The modeling error on the correlation between sensor observations also influences the detection performance. According to (16), the probability of detection from an estimated noise covariance matrix $\hat{\mathbf{C}}_{\mathcal{A}_n}$ is given by

$$P_{D_1}(\mathbf{y}_{\mathcal{A}_n}) = Q\left[Q^{-1}(P_{FA}) - \sqrt{\mathbf{s}_{\mathcal{A}_n}^T \hat{\mathbf{C}}_{\mathcal{A}_n} \mathbf{s}_{\mathcal{A}_n}}\right]. \quad (60)$$

The probabilities of missed detection for a given probability of false alarm under different covariance models are plotted in Fig. 5. If the UN and NC models are used to approximate the FC case, then the results would be overoptimistic. It can be observed that the difference between the UN, NC, and FC models is negligible when correlation between nodes becomes small.

### C. Operational Lifetime

In this paper, we consider a network of $N = 200$ sensors with connectivity depending on the radio range $r_\beta$ as shown in Fig. 6. Let the degree of correlation be $\eta = 5.0$. The sensors are assumed to have a unit of initial energy after deployment. Each work period starts with a sampling stage and ends with a work stage, where the duration of the work stage is assumed to be much longer than that of the sampling stage such

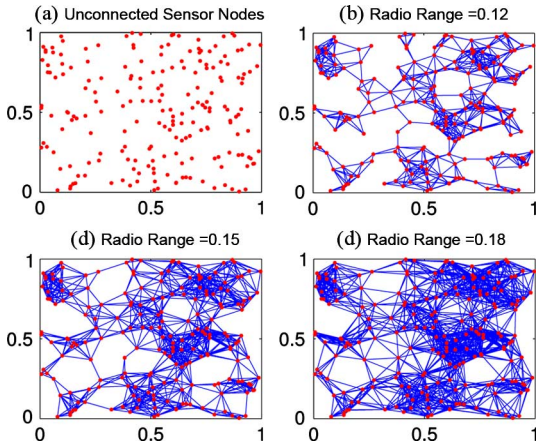


Fig. 6. The connected network of $N = 200$ randomly distributed sensors with different radio ranges. (a) $r_\beta = 0$. (b) $r_\beta = 0.12$. (c) $r_\beta = 0.15$. (d) $r_\beta = 0.18$.

that the cost caused by sampling does not result in significant overhead. During the sampling stage starting with a seed node that has maximum leftover energy, a group of sensors are activated for inference (estimation or detection) later in the work stage, and then are powered off at the end of the period. When the next period starts, a new group of sensors will be activated and the procedure repeats until the sensors deplete their energy. We neglect the energy consumed by sensors during the inactive duration, assuming the amount of energy required for listening to the awakening signal is much less than the amount required for computation and communication. It is assumed that each sensor has an omnidirectional radio antenna at its transmitter, with which the sensor can transmit a common message to its neighbors simultaneously with a single broadcast. In the sequel, we will compare the energy efficiency performance of the proposed innovations diffusion (ID) algorithm with that of a randomized selection (RS) scheme, which randomly adds a new sensor from the neighboring nodes without using innovations.

*1) Constant Transmit Power:* In this scenario, the transmitter at each sensor has a constant transmit power level $P(r_\beta)$ for a given radio range $r_\beta$. We assume that during the sampling stage the computational energy of (1) expended by sensor $i$ is $E_i^{\text{comp}} = K_1$ and the internode communication energy expended by sensor $i$ in each iteration is proportional to

$$E_i^{\text{comm}}(r_\beta) = \frac{K_2 P(r_\beta)}{\log_2(1+\beta)} \quad (61)$$

where $P(r_\beta) = r_\beta^\alpha N_0 \beta$, $K_1$ and $K_2$ are certain constants. During the work stage, the total energy expended by sensor $i$ is assumed to be

$$E_i^{\text{work}}(r_\beta) = \frac{K_3 P(r_\beta)}{\log_2(1+\beta)}. \quad (62)$$

In our simulations, we set $E_i^{\text{comp}} = 0.00001$, $E_i^{\text{comm}}(0.12) = 0.0001$, $E_i^{\text{work}}(0.12) = 0.01$, and $\alpha = 3.5$.

Fig. 7 illustrates the operational lifetime of an estimation network with different radio ranges against the MSE constraint $D_0$. As seen from the plots, the network lifetime increases when the fidelity constraint is more relaxed. The reason behind this is that

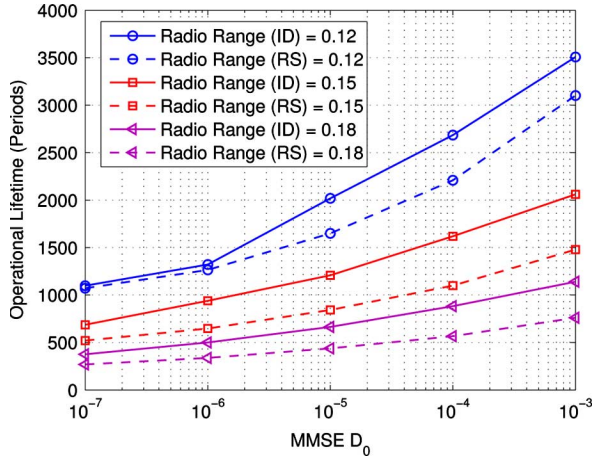


Fig. 7. The operational lifetime versus the estimation fidelity constraint in a wireless sensor network with constant transmit power. For each radio range, the results are averaged over 100 simulations.

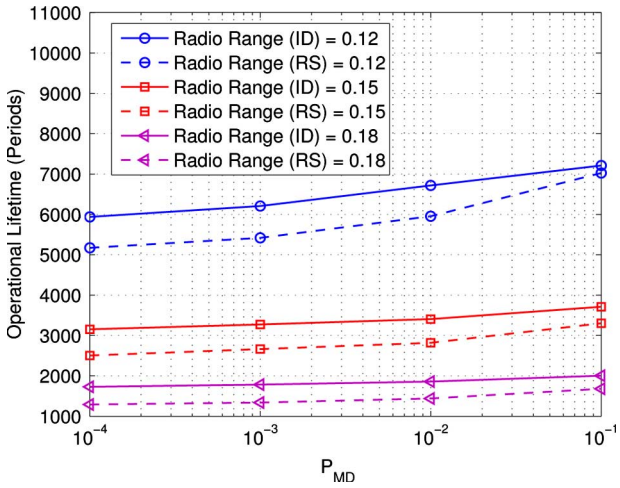


Fig. 8. The operational lifetime versus the probability of missed detection ($P_{MD} = 1 - P_D$) constraint in a wireless sensor network with constant transmit power. The probability of false alarm is fixed at $P_{FA} = 0.1$. For each radio range, the results are averaged over 100 simulations.

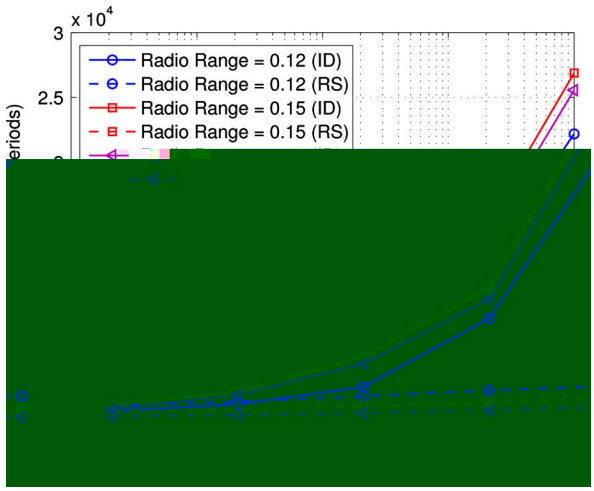


Fig. 9. The operational lifetime versus the estimation fidelity constraint in a wireless sensor network with adaptive transmit power. For each constraint on the radio range, i.e., $r_{\max} = 0.12, 0.15$, and 0.18, the results are averaged over 100 simulations.

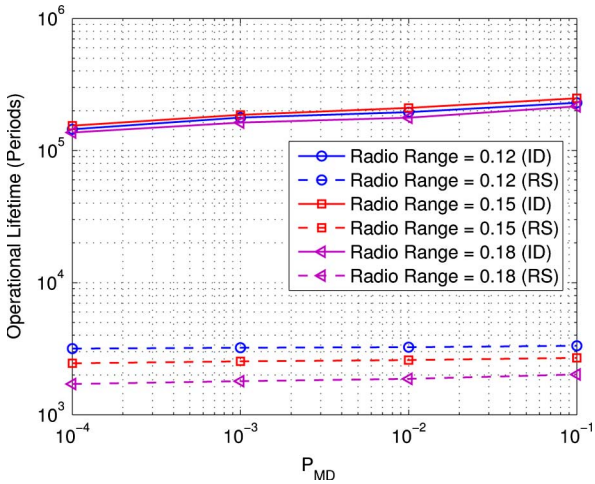


Fig. 10. The operational lifetime versus the probability of missed detection $(1 - P_d)$ constraint in a wireless sensor network with adaptive transmit power. For each constraint on the radio range, i.e., $r_{\max} = 0.12, 0.15$, and 0.18, the results are averaged over 100 simulations.

during each work period, a slack fidelity requirement requires only a small number of sensors while a stringent fidelity constraint needs a larger number of active sensors. The number of sensors activated during each work period has a direct influence on the network lifetime. It can be observed that the innovations diffusion algorithm has the network last longer than does the randomized selection method.

The radio ranges of the sensors are also of importance for the operational lifetime. Provided that the sensor positions are fixed, the number of neighboring nodes of each sensor is determined by the radio range $r_\beta$. If a sensor has a larger radio range, i.e., the transmitter has a larger transmit power, then it can communicate with more neighboring sensors. Consequently, a large radio range will be able to keep good connectivity when some sensor nodes run out of energy. However, a larger radio range requires a larger transmit power level at the transmitter, which may drain out the energy or sensors more quickly. Given a constant transmit power level $P(r_\beta)$ at each sensor, Fig. 7 depicts the system operational lifetimes with respect to different radio ranges. It can be observed that the sensors run out of energy more quickly with a larger radio range. Fig. 8 shows the operational lifetime of a detection network with different radio ranges against the constraint on the missed detection probabilities. It has observations similar to Fig. 7.

*2) Adaptive Transmit Power:* We consider that each sensor can adjust its transmit power such that the desired SNR $\beta$ at the receiver can be achieved with a minimum transmit power level at the transmitter. Thus, the transmit power of active sensor $i$ depends on the distance from its current active neighbors, i.e.

$$P_i = \max_{j \in \mathcal{N}_A(i)} d_{ij}^{\alpha} N_0 \beta. \tag{63}$$

In the simulation setup, the amounts of energy expended during the sampling and work stages are respectively given by (61) and (62) as well, except that the transmit power now uses (63). Specifically, we set $E_i^{\text{comp}} = 0.00001$, $E_i^{\text{comm}} = 0.0001 \times P_i/P(0.12)$, $E_i^{\text{work}} = 0.01 \times P_i/P(0.12)$, and $\alpha = 3.5$.

When choosing the radio range, there is a tradeoff between energy expenditure and maintaining the network connectivity. Using large radio ranges will draw off the energy quickly. On the other hand, larger radio ranges may keep the network connected and may activate a more informative sensor with larger probability at each iteration, and the number of active sensors required to satisfy the desired fidelity may be reduced resulting in a longer operational lifetime. The estimation and detection simulation results are presented in Figs. 9 and 10, respectively, with different radio constraints (i.e., $r_{\max} = 0.12, 0.15$, and $0.18$). For the innovations diffusion algorithm with adaptive transmit power, a medium radio range (e.g., $r_{\max} = 0.15$) has the best performance in terms of the system lifetime.

Compared to the constant transmit power, the innovation diffusion algorithm equipped with adaptive transmit power can achieve a significant improvement in the system lifetime because the desired SNR at the receiver can be obtained with a minimum transmit power level and the sensor energy can be more efficiently used. On the other hand, the randomized selection scheme does not show significant performance gain in the transmit power adaptation when compared with the fixed transmit power scheme.

## VII. Conclusion

This paper has proposed a distributed sampling scheme based on innovations diffusion for estimation or detection in sensor networks. It has been shown that the proposed distributed sampling algorithm requires at most a total amount of energy $\mathcal{O}(\bar{N}^3)$ to accomplish the sampling task. Consequently, it suggests that selecting the most informative sensor measurements for inference will reduce the number of active sensors and lead to energy efficiency. One main advantage of the proposed scheme is that it can be implemented efficiently in an asynchronous and scalable way. The performance of the spatial sampling scheme has been evaluated and illustrated through numerical simulations.

## Appendix I
## Proof of Lemma 1

*Proof:* For $i_k \neq i_l$, we assume $i_k < i_l$. According to (31), we have $\mathbf{e}_{i_k} \in \mathcal{L}\{\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_k}\}$ and $\mathbf{e}_{i_l} \perp \mathcal{L}\{\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_{l-1}}\}$. Since $\mathcal{L}\{\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_k}\} \subseteq \mathcal{L}\{\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_{l-1}}\}$, it follows that $\mathbf{e}_{i_l} \perp \mathbf{e}_{i_k}$. For $i_k = i_l$, it can be shown that $E\left(\mathbf{e}_{i_k}\mathbf{e}_{i_k}^T\right) = \mathbf{Q}_{i_k}$ by substituting (31) and (32) into (33), i.e.,

$$\begin{aligned}\mathbf{Q}_{i_k} &= \left(-\mathbf{B}_{i_k|\mathcal{A}_{k-1}}\mathbf{C}_{\mathcal{A}_{k-1}}^{-1} \quad \mathbf{I}\right) \\ &\quad \times \begin{pmatrix} \mathbf{C}_{\mathcal{A}_{k-1}} & \mathbf{B}_{i_k|\mathcal{A}_{k-1}}^T \\ \mathbf{B}_{i_k|\mathcal{A}_{k-1}} & \mathbf{C}_{i_k i_k} \end{pmatrix} \\ &\quad \times \begin{pmatrix} -\mathbf{C}_{\mathcal{A}_{k-1}}^{-1}\mathbf{B}_{i_k|\mathcal{A}_{k-1}}^T \\ \mathbf{I} \end{pmatrix} \\ &= \mathbf{C}_{i_k i_k} - \mathbf{B}_{i_k|\mathcal{A}_{k-1}}\mathbf{C}_{\mathcal{A}_{k-1}}^{-1}\mathbf{B}_{i_k|\mathcal{A}_{k-1}}^T. \end{aligned} \tag{64}$$

This completes the proof. ■

## Appendix II
## Proof of Theorem 1

*Proof:* Assume that we have a set of selected sensors $\mathcal{A}_n$. For all $i_k \in \mathcal{A}_n$, the projection matrices $\mathbf{P}_{i_k|\mathcal{A}_{k-1}}$ defined in (32) can be combined into a block lower triangular matrix as shown in

$$\mathbf{P}_{\mathcal{A}_n} = \begin{pmatrix} \mathbf{I} & & & \\ -\mathbf{B}_{i_2|\mathcal{A}_1}\mathbf{C}_{\mathcal{A}_1}^{-1} & \mathbf{I} & & \\ & \ddots & \ddots & \\ & & -\mathbf{B}_{i_n|\mathcal{A}_{n-1}}\mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix}. \tag{65}$$

Similarly, we can combine the transformed measurements $\mathbf{z}_{i_k}$ in (36) and the matrices $\mathbf{G}_{i_k|\mathcal{A}_{k-1}}$ in (37), respectively, in compact forms as

$$\mathbf{z}_{\mathcal{A}_n} = \operatorname{col}\left(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_n}\right) = \mathbf{P}_{\mathcal{A}_n}\mathbf{y}_{\mathcal{A}_n} \tag{66}$$

and

$$\begin{aligned}\mathbf{G}_{\mathcal{A}_n} &= \operatorname{col}\left(\mathbf{G}_{i_1|\mathcal{A}_0}, \mathbf{G}_{i_2|\mathcal{A}_1}, \ldots, \mathbf{G}_{i_n|\mathcal{A}_{n-1}}\right) \\ &= \mathbf{P}_{\mathcal{A}_n}\mathbf{H}_{\mathcal{A}_n}. \end{aligned} \tag{67}$$

According to Lemma 1, the covariance matrix of the vector $\operatorname{col}\left(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_n}\right)$ is the block diagonal matrix

$$\begin{aligned}\mathbf{Q}_{\mathcal{A}_n} &= \begin{pmatrix} \mathbf{Q}_{i_1} & & & \\ & \mathbf{Q}_{i_2} & & \\ & & \ddots & \\ & & & \mathbf{Q}_{i_n} \end{pmatrix} \\ &= \mathbf{P}_{\mathcal{A}_n}\mathbf{C}_{\mathcal{A}_n}\mathbf{P}_{\mathcal{A}_n}^T. \end{aligned} \tag{68}$$

Now the MMSE using $\mathbf{z}_{\mathcal{A}_n}$ is given by

$$\begin{aligned}\mathbf{D}_{\mathcal{A}_n} &= \left(\sum_{k=1}^{n}\mathbf{G}_{i_k|\mathcal{A}_{k-1}}^T\mathbf{Q}_{i_k}^{-1}\mathbf{G}_{i_k|\mathcal{A}_{k-1}}\right)^{-1} \\ &= \left(\mathbf{G}_{\mathcal{A}_n}^T\mathbf{Q}_{\mathcal{A}_n}^{-1}\mathbf{G}_{\mathcal{A}_n}\right)^{-1} \\ &= \left[\mathbf{H}_{\mathcal{A}_n}^T\underbrace{\mathbf{P}_{\mathcal{A}_n}^T\left(\mathbf{P}_{\mathcal{A}_n}^T\right)^{-1}}_{\mathbf{I}}\mathbf{C}_{\mathcal{A}_n}^{-1}\underbrace{\mathbf{P}_{\mathcal{A}_n}^{-1}\mathbf{P}_{\mathcal{A}_n}}_{\mathbf{I}}\mathbf{H}_{\mathcal{A}_n}\right]^{-1} \\ &= \left(\mathbf{H}_{\mathcal{A}_n}^T\mathbf{C}_{\mathcal{A}_n}^{-1}\mathbf{H}_{\mathcal{A}_n}\right)^{-1}. \end{aligned} \tag{69}$$

It follows that both $\mathbf{z}_{\mathcal{A}_n}$ and $\mathbf{y}_{\mathcal{A}_n}$ result in the same MMSE. Moreover, the linear m.v.u.e. using $\mathbf{z}_{\mathcal{A}_n}$ can be written as

$$\begin{aligned}&\hat{\boldsymbol{\theta}}\left(\mathbf{z}_{\mathcal{A}_n}\right) \\ &= \left(\sum_{k=1}^{n}\mathbf{G}_{i_k|\mathcal{A}_{k-1}}^T\mathbf{Q}_{i_k}^{-1}\mathbf{G}_{i_k|\mathcal{A}_{k-1}}\right)^{-1}\sum_{k=1}^{n}\mathbf{G}_{i_k|\mathcal{A}_{k-1}}^T\mathbf{Q}_{i_k}^{-1}\mathbf{z}_{i_k} \\ &= \left(\mathbf{H}_{\mathcal{A}_n}^T\mathbf{C}_{\mathcal{A}_n}^{-1}\mathbf{H}_{\mathcal{A}_n}\right)^{-1}\mathbf{G}_{\mathcal{A}_n}\mathbf{Q}_{\mathcal{A}_n}^{-1}\mathbf{z}_{\mathcal{A}_n} \\ &= \left(\mathbf{H}_{\mathcal{A}_n}^T\mathbf{C}_{\mathcal{A}_n}^{-1}\mathbf{H}_{\mathcal{A}_n}\right)^{-1}\mathbf{H}_{\mathcal{A}_n}^T\mathbf{C}_{\mathcal{A}_n}^{-1}\underbrace{\mathbf{P}_{\mathcal{A}_n}^{-1}\mathbf{P}_{\mathcal{A}_n}}_{\mathbf{I}}\mathbf{y}_{\mathcal{A}_n} \\ &= \hat{\boldsymbol{\theta}}\left(\mathbf{y}_{\mathcal{A}_n}\right). \end{aligned}$$

Thus, the result is established. ■

## APPENDIX III
## PROOF OF THEOREM 2

*Proof:* According to (10), the log-likelihood function from the original observations $\mathbf{y}_{\mathcal{A}_n}$ is given by

$$\begin{aligned} L(\mathbf{y}_{\mathcal{A}_n}) &= 2\mathbf{s}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{y}_{\mathcal{A}_n} - \mathbf{s}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{s}_{\mathcal{A}_n} \\ &= 2\boldsymbol{\theta}^T \mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{y}_{\mathcal{A}_n} - \boldsymbol{\theta}^T \mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n} \boldsymbol{\theta}. \end{aligned} \quad (70)$$

From (66) and (67), $L(\mathbf{y}_{\mathcal{A}_n})$ can be written as

$$\begin{aligned} L(\mathbf{y}_{\mathcal{A}_n}) &= 2\boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \underbrace{\mathbf{P}_{\mathcal{A}_n}^{-T} \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{P}_{\mathcal{A}_n}^{-1}}_{\mathbf{Q}_{\mathcal{A}_n}^{-1}} \mathbf{z}_{\mathcal{A}_n} \\ &\quad - \boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \underbrace{\mathbf{P}_{\mathcal{A}_n}^{-T} \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{P}_{\mathcal{A}_n}^{-1}}_{\mathbf{Q}_{\mathcal{A}_n}^{-1}} \mathbf{G}_{\mathcal{A}_n} \boldsymbol{\theta} \\ &= 2\boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \mathbf{Q}_{\mathcal{A}_n}^{-1} \mathbf{z}_{\mathcal{A}_n} - \boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \mathbf{Q}_{\mathcal{A}_n}^{-1} \mathbf{G}_{\mathcal{A}_n} \boldsymbol{\theta} \\ &= L(\mathbf{z}_{\mathcal{A}_n}). \end{aligned} \quad (71)$$

Consequently, $\mathbf{y}_{\mathcal{A}_n}$ and $\mathbf{z}_{\mathcal{A}_n}$ have the same deflection coefficients, i.e.

$$d^2(\mathbf{y}_{\mathcal{A}_n}) = d^2(\mathbf{z}_{\mathcal{A}_n}) = \boldsymbol{\theta}^T \mathbf{G}_{\mathcal{A}_n}^T \mathbf{Q}_{\mathcal{A}_n}^{-1} \mathbf{G}_{\mathcal{A}_n} \boldsymbol{\theta} \quad (72)$$

■

## REFERENCES


[1] Z. Quan, W. J. Kaiser, and A. H. Sayed, "A spatial sampling scheme based on innovations diffusion in sensor networks," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Cambridge, MA, Apr. 2007, pp. 323–330.

[2] I. F. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, pp. 393–422, Mar. 2002.

[3] M. Perillo, Z. Ignjatovic, and W. Heinzelman, "An energy conservation method for wireless sensor networks employing a blue noise spatial sampling technique," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Berkeley, CA, Apr. 2004, pp. 116–123.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York: Wiley, 2003.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[6] V. Delouille, R. Neelamani, and R. Baraniuk, "Robust distributed estimation in sensor networks using the embedded polygons algorithm," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Berkeley, CA, Apr. 2004, pp. 405–413.

[7] R. Nowak, U. Mitra, and R. Willett, "Estimating inhomogeneous fields using wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, pp. 999–1006, Aug. 2004.

[8] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Berkeley, CA, Apr. 2004, pp. 20–27.

[9] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Los Angeles, CA, Apr. 2005, pp. 63–70.

[10] Y. Sung, S. Misra, L. Tong, and A. Ephremides, "Signal processing for application-specific ad hoc networks," *IEEE Signal Process. Mag.*, vol. 23, no. 5, pp. 74–83, Sep. 20006.

[11] V. Saligrama, M. Alanyali, and O. Savas, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Trans. Signal Process.*, vol. 54, pp. 4118–4132, Nov. 2006.

[12] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55, pp. 4064–4077, Aug. 2007.

[13] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 61–72, 2002.

[14] E. Ertin, J. Fisher, and L. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Palo Alto, CA, Apr. 2003, pp. 405–416.

[15] H. Wang, G. Pottie, K. Yao, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Berkeley, CA, Apr. 2004, pp. 36–45.

[16] V. Isler and R. Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," in *Proc. Inf. Process. Sens. Netw. (IPSN)*, Los Angeles, CA, Apr. 2005, pp. 151–158.

[17] Z. Hu, J. Zhang, and L. Tong, "Adaptive sensor activity control in many-to-one sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1525–1534, Aug. 2006.

[18] Z. Quan and A. H. Sayed, "Innovations-based sampling over spatially-correlated sensors," in *Proc. IEEE ICASSP*, Honolulu, HI, Apr. 2007, pp. 509–512.

[19] J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, pp. 413–422, Feb. 2006.

[20] N. A. C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.

[21] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-temporal correlation: theory and applications for wireless sensor networks," *Comput. Netw.*, vol. 45, pp. 245–259, 2004.

[22] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[23] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[24] R. G. Gallager, *Discrete Stochastic Processes*. New York: Kluwer, 1996.

[25] D. P. Bertsimas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.



**Zhi Quan** received the B.E. degree in communication engineering from Beijing University of Posts and Telecommunications, China, and the M.S. degree in electrical engineering from Oklahoma State University, Stillwater.

He is working toward the Ph.D. degree in electrical engineering at the University of California, Los Angeles (UCLA). He was a visiting researcher with Princeton University, Princeton, NJ, during summer 2007, and was an engineering intern with Qualcomm during summer 2008. His current research interests include statistical signal processing, wireless communication and networking, cognitive radios, and multimedia.

Mr. Quan was the recipient of the UCLA Chancellor's Dissertation Fellowship (2008–2009).



**William J. Kaiser** received the Ph.D. degree in solid-state physics from Wayne State University, Detroit, MI, in 1984.

From 1977 to 1986, he was a member of the Ford Motor Co. Research Staff, where his development of automotive sensor and embedded system technology resulted in large volume commercial sensor production. At Ford, he also developed the first spectroscopies directed to microelectronics systems based on scanning tunneling microscopy. From 1986 to 1994, he initiated the NASA Microinstrument program for distributed sensing at the Jet Propulsion Laboratory, Pasadena, CA. In 1994, he joined the Electrical Engineering Department, University of California, Los Angeles (UCLA). Along with Professor Pottie, he initiated the first wireless networked microsensor programs with a vision of linking the Internet to the physical world through distributed monitoring. This continued research includes the topics of low-power embedded computing for wireless networked sensing, biomedical embedded computing, robotic sensor systems for environmental monitoring, and distributed sensing for energy and water resource management. He served as Electrical Engineering Department Chairman from 1996 to 2000. He has more than 170 publications and 23 patents.

Professor Kaiser received the Peter Mark Award of the American Vacuum Society, the NASA Medal for Exceptional Scientific Achievement, the Arch Colwell Best Paper Award of the Society of Automotive Engineers, the Best Paper Award at BodyNets 2008, two R&D 100 Awards, the Allied Signal Faculty Research Award, the Brian P. Copenhaver Award for Innovation in Teaching with Technology, and the UCLA Gold Shield Faculty Award.

**Ali H. Sayed** (F'01) received the Ph.D. degree in 1992 from Stanford University, Stanford, CA.

He is Professor and Chairman of Electrical Engineering with the University of California, Los Angeles (UCLA) and Principal Investigator of the Adaptive Systems Laboratory. He has published widely, with over 300 articles and five books, in the areas of statistical signal processing, estimation theory, adaptive filtering, signal processing for communications and wireless networking, and fast algorithms for large structured problems. He is coauthor of the textbook *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000), of the research monograph *Indefinite Quadratic Estimation and Control* (Philadelphia, PA: SIAM, 1999), and coeditor of *Fast Algorithms for Matrices with Structure*(Philadelphia, PA: SIAM, 1999). He is also the author of the textbooks *Fundamentals of Adaptive Filtering* (Hoboken, NJ: Wiley, 2003), and *Adaptive Filters* (Hoboken, NJ: Wiley, 2008). He has contributed several encyclopedia and handbook articles.

Dr. Sayed is a Fellow of IEEE for his contributions to adaptive filtering and estimation algorithms. He has served on the editorial boards of the IEEE *Signal Processing Magazine*, the *European Signal Processing Journal*, the *International Journal on Adaptive Control and Signal Processing*, and the SIAM *Journal on Matrix Analysis and Applications*. He has also served as the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2003–2005), and the EURASIP *Journal on Advances in Signal Processing* (2006–2007). He is a member of the Signal Processing for Communications and the Signal Processing Theory and Methods Technical Committees of the IEEE Signal Processing Society. He has served on the Publications (2003–2005), Awards (2005), and Conference (2007-present) Boards of the IEEE Signal Processing Society. He served on the Board of Governors (2007–2008) of the same Society and is now serving as Vice-President of Publications (2009–present). His work has received several recognitions including the 1996 IEEE Donald G. Fink Award, 2002 Best Paper Award from the IEEE Signal Processing Society, 2003 Kuwait Prize in Basic Sciences, 2005 Terman Award, 2005 Young Author Best Paper Award from the IEEE Signal Processing Society, and two Best Student Paper Awards at international meetings (1999 and 2001). He has served as a 2005 Distinguished Lecturer of the IEEE Signal Processing Society and as General Chairman of ICASSP 2008.