

Information Retrieval and Processing in Sensor Networks: Deterministic Scheduling Versus Random Access

Min Dong, *Member, IEEE*, Lang Tong, *Fellow, IEEE*, and Brian M. Sadler, *Fellow, IEEE*

Abstract—We investigate the effect of medium-access control (MAC) used in information retrieval by a mobile access point (AP) on information processing in large-scale sensor network, where sensors are unreliable and subject to outage. We focus on a 1-D sensor network and assume that the location information is available locally at each sensor and unavailable to the AP. For a fixed collection interval, two types of MAC schemes are considered: the deterministic scheduling, which collects data from predetermined sensors locations, and random access, which collects data from random locations. We compare the signal estimation performance of the two MACs, using the expected maximum distortion as the performance measure. For large sensor networks with fixed density, we show that there is a critical threshold on the sensor outage probability P_{out} : For $P_{\text{out}} < e^{-\lambda(1+o(1))}$, where λ is the throughput of the random access protocol, the deterministic scheduling provides better reconstruction performance. However, for $P_{\text{out}} > e^{-\lambda(1+o(1))}$, the performance degradation from missing data samples due to sensor outage does not justify the effort of scheduling; simple random access outperforms the optimal scheduling.

Index Terms—Estimation, information retrieval, random access, random field, sampling, scheduling, sensor network.

I. INTRODUCTION

FOR many applications, a sensor network operates in three phases. In the first phase, sensors take measurements that form a snapshot of the signal field at a particular time. The measurements are stored locally. The second phase is information retrieval in which data are collected from individual sensors. The last phase is information processing in which data from sensors are processed centrally with a specific performance metric.

An appropriate network architecture for such applications is Sensor Networks with Mobile Access (SENMA) [1]. Shown in Fig. 1, SENMA has two types of nodes: low-power low-

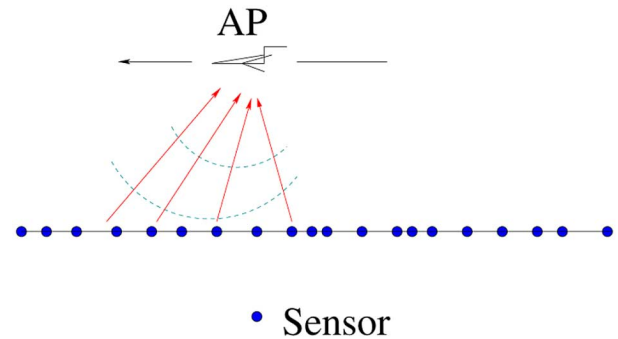


Fig. 1. A 1-D reachback sensor network with a mobile AP.

complexity sensors randomly deployed in large number and a few powerful mobile access points that communicate with sensors. Sensors communicate to the mobile access points (APs) directly. The use of mobile APs enable data collections from specific areas of the network, either by scheduling or by random access.

We focus on the latter two phases in SENMA: information retrieval and processing. Specifically, we examine the effect of medium-access control (MAC) for information retrieval on information processing in a network in which sensors are unreliable, possibly with low duty cycle, and subject to outage. In such cases, it is not possible for mobile access points to collect data from all sensors. We consider two types of MAC: deterministic scheduling that schedules transmissions from specific sensor locations, and random access that collects packets randomly from the sensor field.

For information processing, packets collected by mobile APs form a sampled signal field, either randomly or deterministically with a specific pattern. Two factors affect the performance of information processing: the sampling pattern and the MAC throughput. The former tells us how the signal field is sampled, and the latter provides the amount of samples we can obtain during a collection time using a specific MAC. While it may appear obvious that collecting data from optimally chosen locations using (centralized) deterministic scheduling gives better performance, there are several nontrivial practical complications. For networks with finite sensor density, the scheduling scheme must have a finite scheduling resolution. By this we mean that, because the probability that a randomly deployed sensor exists at a particular location is zero, the deterministic scheduling protocol must be modified to schedule, not a sensor at a particular location, but sensor(s) in the neighborhood of a

Manuscript received June 15, 2006; revised February 12, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mounir Ghogho. This work was supported in part by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011, the National Science Foundation under Contract CNS-0435190, and the Army Research Office under Grant ARO-W911NF-06-1-0346.

M. Dong is currently with Corporate Research & Development, Qualcomm Inc., San Diego, CA 92121 USA (e-mail: mdong@qualcomm.com).

L. Tong is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ece.cornell.edu).

B. M. Sadler is with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: bsadler@arl.army.mil).

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.2007.897911

location. Even with such a modification, there is still non-zero probability that the scheduled transmission location is void of sensors, or the batteries of those scheduled sensors have run out at the time of data collection.

The possibility of sensor outage brings the question of whether a deterministic MAC is sufficiently robust to practical imperfections. In this paper, we compare deterministic scheduling and random access for the application of reconstructing a signal field with Poisson distributed sensors of finite density. To make the problem more tractable, we study the performance in a 1-D signal field, which provides insight into the two-dimensional problem. A minimum mean-square error (MMSE) estimator is used for signal reconstruction, and the performance metric is the expected maximum mean-square error (MSE) of the entire signal field. A key parameter is the sensor outage probability P_{out} —a function of sensor density, sensor duty cycle, and scheduling resolution—which gives the probability that there is no active sensor within the scheduled resolution interval. We show that, for large networks, there is an outage probability threshold beyond which the deterministic central scheduling is inferior to distributed random access.

While the literature on sensor networks grows rapidly, little attention has been paid to the interaction between MAC and signal reconstruction. We approach this problem by making the connection of MAC to sampling and parametrize a MAC scheme into a particular sampling scheme. Specifically, we treat the deterministic scheduling to random access as periodic sampling with missing samples to random sampling. However, because the sampling dictated by the two MAC schemes comes from the specific sensor network application, the periodic and random sampling in our problem possess their own unique characteristics, and thus are different from the traditional sampling problem set up [2]–[8]. Therefore, the literature on the studies of sampling does not provide an answer to our problem. The problem of current setting is considered in [9] for the case when the sensor density is infinite without sensor outage. Using an asymptotic analysis, we show that the optimal deterministic scheduling always gives better performance than random access, although the performance gain depends on the level of measurement signal-to-noise ratio (SNR); the gain in the high-SNR regime is substantial and diminishes at low SNR. The intuition for the latter is that, in the low-SNR regime, the reconstruction performance is dominated by noise rather than by scheduling. In this paper, we generalize that analysis to the case when the sensor density is finite. Moreover, we also bring the effect of MAC throughput (the probability of successful transmission) in random access on the performance into the comparison. Besides the practical significance of such a generalization, our results require the notion of sensor outage, drawing the connection of sensor outage in disjoint regions with the classical problem of head runs in coin tossing [10] and the evaluation of asymptotic extreme statistics of head runs [11].

A. Related Work

In terms of MAC design problem for sensor networks, it has attracted a growing interest. Many MAC protocols have been proposed aiming to the special needs and requirements, such

as system throughput and energy efficiency, for both ad hoc sensor networks [12]–[14] and reachback sensor networks [15], [16]. In terms of sampling, there have been extensive studies in the literature. Many studies and performance analysis on the interpolation methods based on a periodic or random sample scheme have been conducted. See [3] and [4] and references therein. In terms of periodic sampling, the effect of jittering, or missing samples have been studied [6], [17]. Interpolation, sequential estimation, or prediction using random sampling has also been studied [3], [6], [18], [19]. In terms of the comparison of different type of sampling, the comparisons of periodic sampling with Poisson sampling, and time-jittering effect on periodic sampling have been considered for some special cases [3], [17], [20]. Despite these results, the discussion on the comparison of different sampling schemes is still very limited. Infinite data samples on a time (or spatial) line are assumed in most cases. For finite-window samples, the explicit expression is difficult to find and, often, either bounds on the reconstruction error are derived or numerical comparisons are sorted [3]. For the application of sampling in sensor networks, in [21], the authors discuss the tradeoff between frequent (equal-spaced) sampling with 1-bit information per sample and less frequent sampling with k -bit information per sample in a one dimensional signal field.

B. Organization

The rest of this paper is organized as follows. In Section II, we introduce the source and sensor network models. In Section III, we explain the deterministic and random access MAC we consider for data retrieval, followed by the description of the estimator and distortion measure for signal reconstruction. In Section IV, we obtain the expression for the reconstruction distortion under both MAC schemes for a fixed collection time. We then provide the asymptotic performance comparison under the two MAC schemes in Section V, where we obtain the asymptotic behavior of distortion ratio and the threshold which determines the relative performance of the deterministic scheduling and random access. Simulation results are presented in Section VI. Finally, we conclude in Section VII.

II. SYSTEM MODEL

A. Source Model

We consider a 1-D field \mathcal{A} of length D . The signal of interest in \mathcal{A} at time t is denoted by $S_t(x)$, where $x \in \mathcal{A}$. We assume that the spatial dynamic of $S_t(x)$ is a 1-D homogeneous *Ornstein Uhlenbeck field* [23] governed by the following linear stochastic differential equation:

$$dS_t(x) = -fS_t(x)dx + \sigma dW_t(x) \quad (1)$$

where $f > 0$ and σ are known, and the process $\{W_t(x)\}$ is a standard Brownian motion. The signal $S_t(x) \sim \mathcal{N}(0, \sigma_s^2)$, where $\sigma_s^2 = (\sigma^2/2f)$, is the stationary solution of (1). This process is often used to model physical phenomenon such as

temperature, vibration, etc. It is both Gaussian and Markovian [23, Ch. 2]. Given $S_t(x_0)$, we have

$$\begin{aligned} E\{S_t(x)|S_t(x_0)\} &= e^{\left[\int_{x_0}^x -f dx\right]} S_t(x_0) \\ &= e^{-f(x-x_0)} S_t(x_0) \quad (x_0 < x). \end{aligned} \quad (2)$$

Being homogeneous in \mathcal{A} , $S_t(x)$ has the autocorrelation

$$\begin{aligned} E\{S_t(x_0)S_t(x_1)\} &= E\{S_t(x_0)E\{S_t(x_1)|S_t(x_0)\}\} \\ &= \sigma_s^2 e^{-f(x_1-x_0)} \quad (x_0 < x_1) \end{aligned} \quad (3)$$

which is only a function of distance between the two points x_1 and x_0 .

B. Sensor Network Model

Consider a network where sensors are randomly deployed in \mathcal{A} . We assume that the distribution of sensors in \mathcal{A} forms a 1-D homogeneous spatial Poisson field with local density ρ sensors/unit area. In particular, in an interval of size l , the number of sensors $N(l)$ is a Poisson random variable with probability

$$P[N(l) = k] = e^{-\rho l} \frac{(\rho l)^k}{k!} \quad (4)$$

and mean $E[N(l)] = \rho l$. Given $N(l) = k$, the k sensors are independent and identically distributed (i.i.d.) in this interval with uniform distribution. Finally, the number of sensors in any two disjoint intervals are independent.

After deployment, the sensors obtain their location information through some positioning method [24]–[26]. At time t , all sensors measure their local signals and form a snapshot of the signal field. The measurement of a sensor at location x is given by

$$Y_t(x) = S_t(x) + Z_t(x) \quad (5)$$

where $Z_t(x)$ is spatially i.i.d. zero mean white Gaussian measurement noise with variance σ_Z^2 and is independent of $S_t(x)$. Note that we consider information retrieval and processing only based on the measurement data at time t . Thus, we drop the time index for brevity in the following presentation.

Each sensor then stores its local measurement, along with their sensor location information, in the form of a packet for collection.

III. INFORMATION RETRIEVAL AND PROCESSING

A. MAC for Information Retrieval

When the mobile AP is ready for data collection, it retrieves packets from sensors within a fixed collection time using a given MAC scheme. Assuming slotted transmissions in a collision channel,¹ we consider two types of MAC schemes for data retrieval: deterministic scheduling and random access.

We assume that the AP has no knowledge of the sensor locations before data collection, and only each sensor knows its

own location information. If the AP had sensor location information, it could exploit it. However, this implies a great deal of overhead and additional networking burden. Sensor deployment is often random, and additional sensors may be added at a later time; each sensor's operational condition may also be uncertain (sensors may be out of operation, or in different duty cycles). All these hinder *a priori* knowledge of sensor location information at the AP, and a large overhead burden occurs if the AP were to acquire this before data collection, especially in a large sensor field. Therefore, we focus on the case where the sensors know their individual locations, but location information is not required at the AP.

1) *Deterministic Scheduler π_D* : A deterministic scheduling collects data from predetermined locations. Given a collection time of M slots, one particular scheme is to preselect M sensors to transmit. If the AP has the knowledge of sensor locations, there is an optimal selection of M sensors to transmit. Without the knowledge of sensor location, however, it is nontrivial to select a good set of M sensors at the initial stage before collection. Somehow, the AP will need to travel across the sensor field and learn where the sensors are, and the overhead of this initial learning process can be significant. Furthermore, battery expiration, addition of new sensors, and different sensor duty cycles can all make the eligible sensor set time varying. In addition, some sensors may not have new data to transmit. That also makes the deterministic selection of M sensors for data collection inappropriate.

Instead, we consider a deterministic scheme that schedules sensors from the intervals centered at M equally spaced locations to access the channel. Ideally, if sensors exist at these equally spaced locations, such a scheduling is optimal for the distortion measure we consider in Section III-C [9]. However, when there is no sensor at a scheduled location, deterministic schedulings result in missed samples. For randomly distributed sensors, the probability that a sensor exists at a particular location is zero. Thus, we need to consider scheduling with finite resolution.

Specifically, the AP obtains the equally spaced locations based on the field length D and the collection time M slots. It then travels through the field for data collection, where the scheduler π_D enables a resolution interval of length ϵ centered at each desired location, and the AP collects one packet from each interval.² If there are no active sensors in a resolution interval, we say a sensor outage occurs. Let P_{out} be the probability of sensor outage. For the target P_{out} , the smallest interval length ϵ that π_D enables should satisfy

$$P_{\text{out}} = P[N(\epsilon) = 0] = e^{-\rho\epsilon}. \quad (6)$$

In other words, the resolution interval is determined by P_{out} and ρ as

$$\epsilon(P_{\text{out}}, \rho) = \frac{\ln 1/P_{\text{out}}}{\rho}. \quad (7)$$

²This can be implemented through carrier sensing by letting the sensor closest to the center of the interval to transmit [22], or using a control channel.

¹A packet is correctly received only if no other users attempt transmission.

Notice that the maximum number M_{\max} of disjoint intervals that π_D can enable is

$$M_{\max} = \frac{D}{\epsilon(P_{\text{out}}, \rho)} - 1 = \frac{D\rho}{\ln 1/P_{\text{out}}} - 1. \quad (8)$$

In summary, for given target sensor outage probability P_{out} and network density ρ , the deterministic scheduler π_D first sets the resolution interval length $\epsilon(P_{\text{out}}, \rho)$ according to (7). For given M slots of collection time ($M < M_{\max}$), it next enables M intervals of length $\epsilon(P_{\text{out}}, \rho)$ centered at the equally spaced locations. Then, the AP schedules one packet transmission (data sample) from each interval. Note that the AP can use target P_{out} and choose resolution interval based on P_{out} , or it may choose to only use D/M as the resolution interval, and $P_{\text{out}} = e^{-\rho D/M}$ is the resulting outage probability. We choose the former way in the following presentation, which separates the notion of collection interval ϵ from the collection time M to allow more flexible design choices under practical constraints.³ We note that both finite and asymptotic analytical results derived below are the same under either approach.

2) *Random Access π_R* : In the random access, in contrast, sensors contend to access the channel with equal priority. We assume that packets from sensors have equal reception probability,⁴ and therefore the origins of received packets in \mathcal{A} are random. In terms of the sample pattern, these received data samples can be viewed as if they are results of the AP randomly sampling the sensor measurement data in \mathcal{A} . Specifically, the mobile AP activates sensors in an area for transmissions with a given random access protocol. Depending on the application, the area enabled can be either the whole field \mathcal{A} or randomly chosen subregions of \mathcal{A} . For an enabled collection area in the network, let λ_n be the MAC throughput (probability of successful transmission per slot) for the area with n activated sensors. We assume that $\lim_{n \rightarrow \infty} \lambda_n$ exists.

B. Viewing MAC as Sampling Scheme

As we have mentioned above, a MAC scheme determines a specific sampling scheme: It dictates the data sample amount and underlying sampling pattern. Therefore, in terms of the effect on signal reconstruction, we can view a MAC scheme as a sampling scheme. In particular, for the two types of MAC schemes we consider, the deterministic scheduler can be viewed as periodic sampling. With the existence of sensor outage in a resolution interval, we can view the effect of scheduling as periodic sampling with missing samples. For random access, on the other hand, the resulting sampling pattern is random with uniform dis-

³Using interval D/M may not be always possible or desirable; We consider the resolution interval to be a design choice. Specifically, the access point usually has to be confined to an angle of vision. If the antenna of the access point has the fixed beamwidth, then it is appropriate to fix the retrieval interval size, instead of a function of M . At the sensor field, if the segment D/M is large, sensors may not be able to hear each other. This increases the chance of collision and resulting loss of data in that interval. If the sensor density is high, it is not necessary to enable the whole D/M segment, but rather a small interval may be sufficient.

⁴In this work, we assumed all nodes have equal reception probabilities (desired for reconstruction purpose). For large spatial network, this can be realized by dividing the field into smaller regions and the mobile AP enables one region at a time.

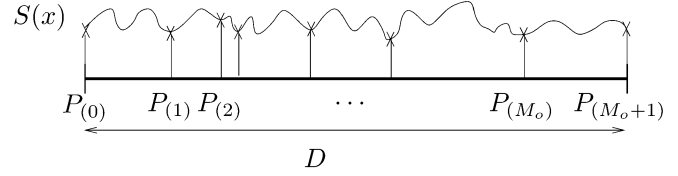


Fig. 2. Example of the locations of received packets.

tribution. Therefore, we can view the retrieval result as random sampling (in both pattern and amount) with uniform distribution. Note that the sampling induced by the MAC scheme is different from ones in traditional sampling problems: The size of the sensor field is finite; for random access, besides the sampling location is random, the number of samples in a given collection time is limited and random with the mean equal to $E(\lambda_n)$ where the expectation is taken over n , which is Poisson distributed. In this paper, we are interested in the comparison of the performance of the two sampling schemes dictated by the MAC schemes within the same collection time, and their asymptotic behavior of the MAC performance in a large network as the retrieval time becomes large.

C. Information Processing and Performance Measure

Assume that the AP retrieves packets from \mathcal{A} with a MAC scheme for M slots, and M_o packets in \mathcal{A} are received. Note that M_o is random for both π_R and π_D (due to sensor outage). We reconstruct the original signal based on these received data samples, and measure the signal reconstruction performance.

To avoid the boundary effect for signal reconstruction, we assume that there is a sensor deployed at each of the two boundaries of \mathcal{A} , and we are able to obtain their measurements. The locations of all the received packets⁵ are denoted by $\mathbf{p}_{M_o} = \{P_1, \dots, P_{M_o}, 0, D\}$. We denote the corresponding order statistics of $\{P_1, \dots, P_{M_o}\}$ by $P_{(1)} < \dots < P_{(M_o)}$. Also, we denote $P_{(0)} = 0$ and $P_{(M_o+1)} = D$. Fig. 2 shows a realization of sampling locations \mathbf{p}_{M_o} of the signal field in \mathcal{A} .

We estimate $S(x)$ at location x using its two immediate neighbor samples by the MMSE smoothing. That is, for $P_{(i)} < x < P_{(i+1)}$,

$$\hat{S}(x) = E[S(x) | Y(P_{(i)}), Y(P_{(i+1)})]. \quad (9)$$

In the noiseless measurement case where $Y(x) = S(x)$, the above estimator is in fact optimal in terms of MSE. In other words

$$\hat{S}(x) = E[S(x) | \{Y(p), p \in \mathbf{p}_{M_o}\}], \quad x \in \mathcal{A}. \quad (10)$$

The reason is that, due to the Markov property of $S(x)$ in (1), the signal $S(x)$, where $P_{(i)} < x < P_{(i+1)}$, is only a function of $S(P_{(i)})$ and $S(P_{(i+1)})$. When measurement noise is present, the estimator in (9) is suboptimal. Nonetheless, its simple structure and easy computation lends itself as an attractive practical estimator.

⁵For convenience, M_o only denotes the number of packets not from the two boundary sensors of \mathcal{A} .

Let the measurement SNR be $\text{SNR} \triangleq \sigma_s^2/\sigma_z^2$. Given \mathbf{p}_{M_o} , we use the maximum mean square estimation error in \mathcal{A} as the performance measure for the reconstruction distortion in a given realization of received samples:

$$\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR}) \triangleq \max_{x \in \mathcal{A}} E \left\{ \left| \hat{S}(x) - S(x) \right|^2 | \mathbf{p}_{M_o} \right\}. \quad (11)$$

Note that maximum MSE is often used to capture the worst case performance. Specific to the application of environmental monitoring by sensor network, information from each local area is needed in order to observe any abnormal behavior or activity. For information accuracy, it requires that the error of the reconstructed field at each location should be less than a threshold. The maximum MSE precisely provides this information, and therefore is used here as the performance measure.

A MAC scheme specifies where and how data packets should be transmitted, in other words, how the signal field $\{S(x) : x \in \mathcal{A}\}$ is sampled; it not only determines the number of samples received, but also specifies the distribution of sample locations \mathbf{p}_{M_o} . The performance of a MAC scheme π is then measured by the expected value of signal distortion metric in (11) under a specific sampling scheme dictated by π . In other words, the expected maximum distortion of signal reconstruction with $M_o = m$ distinct received packets is then given by

$$\bar{\mathcal{E}}(m, \text{SNR}) \triangleq E \{ \mathcal{E}(\mathbf{p}_m, \text{SNR}) | M_o = m \} \quad (12)$$

where the expectation is taken over \mathbf{p}_{M_o} for $M_o = m$. The expected maximum distortion of a MAC π during the collection time M slots is then given by

$$\bar{\bar{\mathcal{E}}}(M, \text{SNR}) \triangleq E \{ \bar{\mathcal{E}}(M_o, \text{SNR}) \} \quad (13)$$

where the expectation is taken over M_o .

Our objective is to compare the signal reconstruction distortion $\bar{\bar{\mathcal{E}}}(M, \text{SNR})$ of the two MAC schemes at different sensor outage condition P_{out} , when each scheme is used to retrieve data for a fixed collection time.

IV. SIGNAL RECONSTRUCTION DISTORTION AND MAC

In this section, we obtain the average maximum distortion $\bar{\bar{\mathcal{E}}}(M, \text{SNR})$ of the two schemes π_D and π_R .

Given M_o received data samples, the maximum distortion $\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR})$ in (11) can be found by comparing the maximum distortion of $S(x)$ in each interval $P_{(i)} < x < P_{(i+1)}$, for $i = 0, \dots, M_o$. Because the estimate $\hat{S}(x)$ in (9) is only a function of the two immediate neighbor samples of $S(x)$, we have

$$\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR}) = \max_{0 \leq i \leq M_o} \max_{P_{(i)} < x < P_{(i+1)}} E \left\{ \left| \hat{S}(x) - S(x) \right|^2 | P_{(i)}, P_{(i+1)} \right\}. \quad (14)$$

Note that the signal $S(x)$ and the measurement noise $Z(x)$ are independent and Gaussian. Given P_{M_o} , $S(x)$ and $\{Y(p), p \in P_{M_o}\}$ are jointly Gaussian. Therefore, the MMSE estimator

$\hat{S}(x)$ in (9) is linear. The expressions for $\hat{S}(x)$ and its MSE are obtained by

$$\begin{aligned} \hat{S}(x) &= E \left\{ S(x) \mathbf{S}_p^{(i)H} \right\} \left(E \left\{ \mathbf{S}_p^{(i)} \mathbf{S}_p^{(i)H} \right\} + \sigma_n^2 \mathbf{I} \right)^{-1} \mathbf{S}_p^{(i)} \\ &\quad E \left\{ \left| \hat{S}(x) - S(x) \right|^2 | P_{(i)}, P_{(i+1)} \right\} \\ &= E |S(x)|^2 - E \left\{ S(x) \mathbf{S}_p^{(i)T} \right\} \\ &\quad \times \left(E \left\{ \mathbf{S}_p^{(i)} \mathbf{S}_p^{(i)T} \right\} + \sigma_n^2 \mathbf{I} \right)^{-1} E \left\{ \mathbf{S}_p^{(i)} S(x) \right\} \end{aligned} \quad (15)$$

where $\mathbf{Y}_p^{(i)} = [Y(P_{(i)}), Y(P_{(i+1)})]^T$, $\mathbf{S}_p^{(i)} = [S(P_{(i)}), S(P_{(i+1)})]^T$, and $P_{(i)} < x < P_{(i+1)}$. With (3), we can show that the maximum MSE of $\hat{S}(x)$, for $P_{(i)} < x < P_{(i+1)}$, is only a function of the distance between $P_{(i)}$ and $P_{(i+1)}$

$$\begin{aligned} \mathcal{E}(d_i^{(M_o)}, \text{SNR}) &\triangleq \max_{P_{(i)} < x < P_{(i+1)}} E \left\{ \left| \hat{S}(x) - S(x) \right|^2 | P_{(i)}, P_{(i+1)} \right\} \\ &= \frac{\frac{1}{\text{SNR}} + 1 - e^{-f d_i^{(M_o)}}}{\frac{1}{\text{SNR}} + 1 + e^{-f d_i^{(M_o)}}} \sigma_s^2 \end{aligned} \quad (16)$$

where $d_i^{(M_o)} \triangleq P_{(i+1)} - P_{(i)}$, for $i = 0, \dots, M_o$. The superscript of d_i denotes the number of received distinct data samples. The maximum distortion $\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR})$ in (14) is therefore determined only by the maximum of distances $\{d_i^{(M_o)}\}$ between any two adjacent data samples

$$\mathcal{E}(\mathbf{p}_{M_o}, \text{SNR}) = \frac{\frac{1}{\text{SNR}} + 1 - e^{-f d_{\max}^{(M_o)}}}{\frac{1}{\text{SNR}} + 1 + e^{-f d_{\max}^{(M_o)}}} \sigma_s^2 \triangleq \mathcal{E}(d_{\max}^{(M_o)}, \text{SNR}) \quad (17)$$

where $d_{\max}^{(M_o)} = \max_{0 \leq i \leq M_o} d_i^{(M_o)}$, which is a monotone increasing function of $d_{\max}^{(M_o)}$. The average maximum distortion with $M_o = m$ received samples in (12) is now rewritten as

$$\bar{\mathcal{E}}(m, \text{SNR}) = E \left\{ \mathcal{E}(d_{\max}^{(M_o)}, \text{SNR}) | M_o = m \right\}. \quad (18)$$

A. Random Access π_R

As we have mentioned in Section III, for the random-access MAC π_R , given M_o received samples, their locations $\{P_1, \dots, P_{M_o}\}$ are i.i.d. random variables with uniform distribution $U(0, D)$. Consequently, the maximum sample distance $d_{\max}^{(M_o)}$ is random. The cumulative distribution $F_{d_{\max}}(x; m)$ of $d_{\max}^{(M_o)}$ for $M_o = m$ can be calculated from the distribution of the order statistics $\{P_{(1)}, \dots, P_{(M_o)}\}$. The joint density of $\{P_{(1)}, \dots, P_{(m)}\}$ is given by [27]

$$f_{P_{(1)}, \dots, P_{(m)}}(x_1, \dots, x_m) = \begin{cases} \frac{m!}{D^m}, & \text{if } 0 < x_1 < \dots < x_m < D \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Using the transformation of random variables, we have the joint distribution of sample distances $\{d_i^{(m)}\}_{i=0}^{m-1}$ given by

$$\begin{aligned} f_{d_0, \dots, d_{m-1}}(x_0, \dots, x_{m-1}) &= \begin{cases} \frac{m!}{D^m}, & \text{if } x_i \geq 0, \text{ and } \sum_{i=0}^{m-1} x_i \leq D \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \quad (20)$$

Since $d_m^{(m)} = D - \sum_{i=0}^{m-1} d_i^{(m)}$, we have

$$F_{d_{\max}}(x; m) = \Pr \left(d_0 \leq x, \dots, d_{m-1} \leq x, D - \sum_{i=0}^{m-1} d_i \leq x \right) \\ = \int_0^x \cdots \int_0^x \frac{m!}{D^m} \mathbf{1}_{[0 \leq D - \sum_{i=0}^{m-1} x_i \leq x]} \\ \times (x_0, \dots, x_{m-1}) dx_0 \cdots dx_{m-1} \quad (21)$$

where $\mathbf{1}_B(x)$ is the indicator function: $\mathbf{1}_B(x) = 1$ for $x \in B$, and $\mathbf{1}_B(x) = 0$ otherwise. By calculating the above, the expression of $F_{d_{\max}}(x; m)$ is given by (22), shown at the bottom of the page. The distribution of $d_{\max}^{(M_o)}$ is plotted in Fig. 3 for several values of M_o as examples.

Combining (18) and (22), and applying $E(X) = \int P(X > t)dt$ for $X \geq 0$, the expected maximum distortion for π_R with $M_o = m$ packets is obtained by

$$\bar{\mathcal{E}}_R(m, \text{SNR}) = \int_0^\infty P \left(\mathcal{E} \left(d_{\max}^{(m)}, \text{SNR} \right) > x \right) dx \\ = \mathcal{E} \left(\frac{D}{m+1}, \text{SNR} \right) \\ + \int_{\frac{D}{m+1}}^D (1 - F_{d_{\max}}(t; m)) \mathcal{E}'(t, \text{SNR}) dt \quad (23)$$

where t is such that $\mathcal{E}(t, \text{SNR}) = x$, and $\mathcal{E}'(t, \text{SNR})$ is the derivative of $\mathcal{E}(t, \text{SNR})$ given by

$$\mathcal{E}'(t, \text{SNR}) = \frac{2 \left(\frac{1}{\text{SNR}} + 1 \right) f e^{-ft}}{\left(\frac{1}{\text{SNR}} + 1 + e^{-ft} \right)^2} \sigma_s^2.$$

Finally, the expected maximum distortion of π_R in (13) is obtained by

$$\bar{\mathcal{E}}_R(M, \text{SNR}) = E \{ \bar{\mathcal{E}}_R(M_o, \text{SNR}) \} \\ = \sum_{m=0}^M \bar{\mathcal{E}}_R(m, \text{SNR}) P(M_o = m). \quad (24)$$

B. Deterministic Scheduler π_D

For the deterministic scheduler π_D with target sensor outage probability P_{out} , it enables M intervals of length $\epsilon(P_{\text{out}}, \rho)$ centered at $Di/(M+1)$, for $i = 1, \dots, M$ and collects one sample from each interval. The actual number of received samples may be less than M due to sensor outage. In the later presentation,

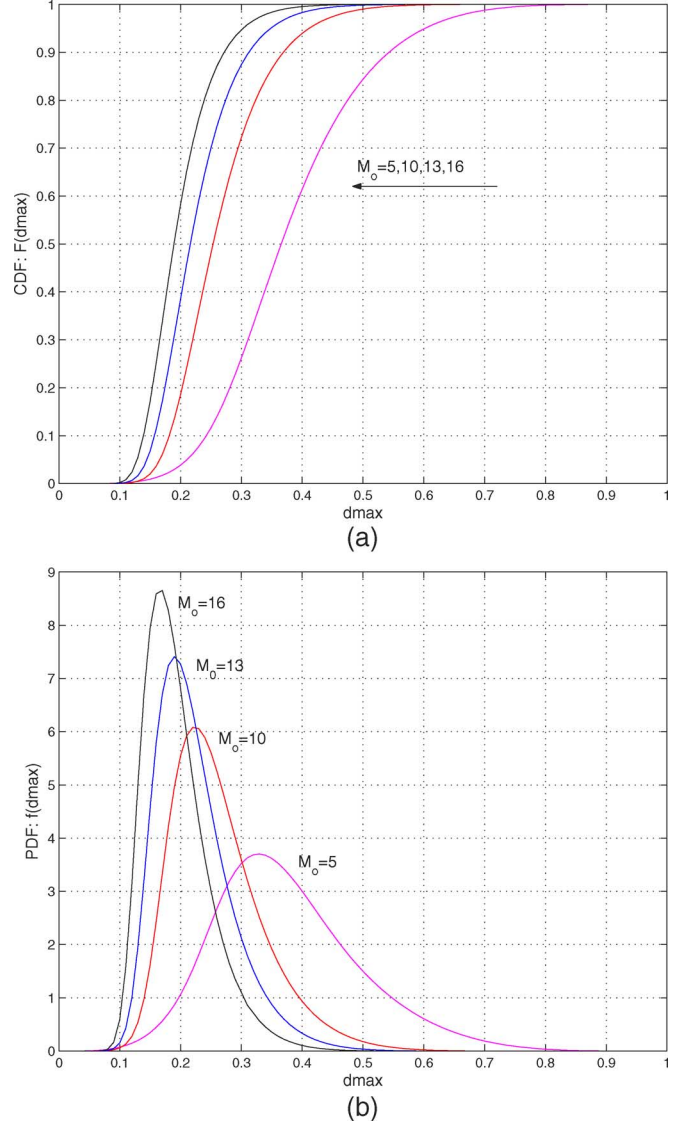


Fig. 3. Distribution of $d_{\max}^{(M_o)}$ under π_R ($D = 1$): (a) CDF of $d_{\max}^{(M_o)}$; (b) PDF of $d_{\max}^{(M_o)}$.

we use ϵ instead of $\epsilon(P_{\text{out}}, \rho)$ for simplicity. The readers should keep in mind of the relationship of ϵ with ρ and P_{out} . From (13) and (18), we have

$$\bar{\mathcal{E}}_D(M, \text{SNR}) = E \{ \mathcal{E}(d_{\max}, \text{SNR}) \} \quad (25)$$

where d_{\max} now is the maximum sample distance in a realization of the received data samples, which is determined by the

$$F_{d_{\max}}(x; m) = \begin{cases} 0 & \text{if } 0 \leq x < \frac{D}{m+1} \\ \sum_{i=0}^k (-1)^i \binom{m+1}{i} \left[(m-i+1) \frac{x}{D} - 1 \right]^m & \text{if } \frac{D}{m-k+1} \leq x < \frac{D}{m-k}, k = 0, \dots, m-2 \\ 1 - (m+1) \left(1 - \frac{x}{D} \right)^m & \text{if } \frac{D}{2} \leq x \leq D \\ 1 & \text{if } x > D. \end{cases} \quad (22)$$

largest number L_M of consecutive intervals that sensor outage occurs. Then, we have

$$\frac{D(L_M + 1)}{M + 1} - \epsilon(P_{\text{out}}, \rho) \leq d_{\text{max}} \leq \frac{D(L_M + 1)}{M + 1} + \epsilon(P_{\text{out}}, \rho). \quad (26)$$

The expected maximum distortion $\bar{\mathcal{E}}_D(M, \text{SNR})$ with π_D is therefore bounded by

$$\begin{aligned} E \left[\mathcal{E} \left(\frac{D(L_M + 1)}{M + 1} - \epsilon(P_{\text{out}}, \rho), \text{SNR} \right) \right] \\ \leq \bar{\mathcal{E}}_D(M, \text{SNR}) \\ \leq E \left[\mathcal{E} \left(\frac{D(L_M + 1)}{M + 1} + \epsilon(P_{\text{out}}, \rho), \text{SNR} \right) \right] \end{aligned} \quad (27)$$

where

$$\begin{aligned} E \left[\mathcal{E} \left(\frac{D(L_M + 1)}{M + 1} \pm \epsilon(P_{\text{out}}, \rho), \text{SNR} \right) \right] \\ = \sigma_s^2 \sum_{k=0}^M \frac{\frac{1}{\text{SNR}} + 1 - e^{-f\left(\frac{D(k+1)}{M+1} \pm \epsilon(P_{\text{out}}, \rho)\right)}}{\frac{1}{\text{SNR}} + 1 + e^{-f\left(\frac{D(k+1)}{M+1} \pm \epsilon(P_{\text{out}}, \rho)\right)}} P[L_M = k]. \end{aligned} \quad (28)$$

To calculate the bounds, we need to find the cumulative distribution of L_M denoted by $F_{L_M}(l)$. It is given by

$$F_{L_M}(l) = \sum_{k=0}^M C_M^{(k)}(l) P_{\text{out}}^k (1 - P_{\text{out}})^{M-k} \quad (29)$$

where $C_M^{(k)}(l)$ is the number of realizations in which exact k intervals have sensor outage, but no more than l of these are consecutive.

Since sensors are Poisson distributed, for disjoint resolution intervals, the occurrence of sensor outage in each of them are independent with probability P_{out} . This can be viewed as a sequence of a biased coin tossing, where heads appear with P_{out} . Then, the distribution of L_M is equivalent to that of the longest head run (consecutive heads) in a sequence of M biased coin tosses. The coefficient $C_M^{(k)}(l)$ therefore can be viewed as the number of sequences of length M in which exactly k heads occur, but no more than l of these occur consecutively. The longest run in a sequence of coin tossing has been studied extensively [10], [11]. It is shown in [10] that $C_M^{(k)}(l)$ can be calculated by the following recursion:

$$C_M^{(k)}(l) = \begin{cases} 0, & \text{if } k = M, l < M \\ \binom{M}{k}, & \text{if } k \leq l \\ \sum_{i=0}^l C_{M-1-i}^{(k-i)}(l), & \text{if } l < k < M. \end{cases} \quad (30)$$

Therefore, the upper and lower bounds on $\bar{\mathcal{E}}_D(M, \text{SNR})$ can be calculated using (27), (29), and (30).

V. SCHEDULING VERSUS RANDOM ACCESS

In this section, we perform an analysis for large sensor networks. The average number of sensors in the network, denoted as \bar{N} , is given by $\bar{N} = \rho D$. We are interested in the comparison

of the asymptotic performance of the two schemes when \bar{N} is large.

Under the mechanisms of deterministic scheduler π_D and random access π_R described in Section III-A, we compare the the signal reconstruction distortion of the two schemes. The following theorem describes the asymptotic behavior of reconstruction distortion $\bar{\mathcal{E}}(M, \text{SNR})$ when M becomes large.

Theorem 1: Let P_{out} be given. Suppose that (1) $\rho D \rightarrow \infty$ as $M \rightarrow \infty$; (2) $D \ln M / M \rightarrow 0$ as $M \rightarrow \infty$, the reconstruction distortions under π_D and π_R are given by⁶

$$\begin{aligned} \bar{\mathcal{E}}_D(M, \text{SNR}) = \frac{\sigma_s^2}{1 + 2\text{SNR}} + 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{\left(\frac{1}{\text{SNR}} + 2\right)^2} \\ \times \left(D \frac{\log M}{M} + O\left(\frac{\log \log M}{M}\right) \right) \end{aligned} \quad (31)$$

where $\log \triangleq \log_{1/P_{\text{out}}}$.

$$\begin{aligned} \bar{\mathcal{E}}_R(M, \text{SNR}) = \frac{\sigma_s^2}{1 + 2\text{SNR}} + 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{\left(\frac{1}{\text{SNR}} + 2\right)^2} \\ \times \left(D \frac{\ln(M\lambda(1 + o(1)))}{M\lambda(1 + o(1))} + O\left(\frac{\ln \ln M}{M}\right) \right) \end{aligned} \quad (32)$$

where $\lambda = \lim_{n \rightarrow \infty} \lambda_n$.

Proof: See Appendix I. ■

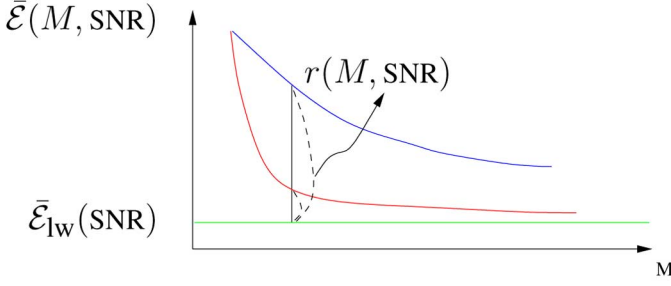
Theorem 1 shows us how reconstruction distortion scales with M as the network grows large (either density, or both size and density). Note that the condition on the growth of network population \bar{N} in Theorem 1 allows the network size D grow but below rate $M / \ln M$ with increased density ρ . One special case of this condition is the case where the size of the network D is fixed and the density ρ grows. Note from the first term of (31) and (32) that, there exists a lower bound on the distortion, we denote it by $\bar{\mathcal{E}}_{\text{lw}}(\text{SNR})$:

$$\bar{\mathcal{E}}_{\text{lw}}(\text{SNR}) = \frac{\sigma_s^2}{1 + 2\text{SNR}}. \quad (33)$$

From (17), we see that this bound corresponds to the case when we let the density of the sensor network go to infinity (thus there exists a sensor at any location), and collect all the measurement of the field. The distortion in the second term of (31) and (32) corresponds to the distortion due to the particular sampling scheme dictated by deterministic scheduler π_D and random access π_R . It is this term that represents the difference of signal reconstruction under different MAC schemes.

The distortion expressions in (31) and (32) show that in large sensor network, as we use more and more time slots to collect the data samples, the reconstruction distortion decreases to the lower bound $\bar{\mathcal{E}}_{\text{lw}}(\text{SNR})$. The decreasing rate depends on the underlying sampling scheme dictated by the MAC scheme, as well as the signal spatial correlation. From the second terms in

⁶A function $f(x)$ is $O(g(x))$ if $\lim_{x \rightarrow \infty} f(x)/g(x) = c$ where $c < \infty$ is constant; $f(x)$ is $o(g(x))$ if $\lim_{x \rightarrow \infty} f(x)/g(x) = 0$, in particular, when $g(x) = 1$, $f(x)$ is $o(1)$ if $\lim_{x \rightarrow \infty} f(x) = 0$.

Fig. 4. Distortion decay ratio $r(M, \text{SNR})$.

(31) and (32), we see that, for π_D , the decreasing rate is at least $(\ln 1/P_{\text{out}})/f$, and for π_R , the rate is at least λ/f .

Given the asymptotic expression in Theorem 1, we now compare the performance of the two schemes π_D and π_R . Note that we could directly compare the distortion $\bar{\mathcal{E}}(M, \text{SNR})$ of the two schemes. But as we mentioned above, the term representing the distortion due to the specific MAC scheme used is the second term in (31) and (32). Therefore, we will only compare the second term, i.e., the excess distortion

$$\bar{\mathcal{E}}(M, \text{SNR}) - \bar{\mathcal{E}}_{\text{lw}}(\text{SNR}).$$

We will see in Theorem 2 that comparing the excess distortion provides a clear expression for the threshold on P_{out} according to which the relative performance of π_R and π_D changes. For convenience, we define the ratio of the excess distortion of π_R to that of π_D as

$$r \triangleq \frac{\bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}_{\text{lw}}(\text{SNR})}{\bar{\mathcal{E}}_D(M, \text{SNR}) - \bar{\mathcal{E}}_{\text{lw}}(\text{SNR})}. \quad (34)$$

Fig. 4 illustrates the ratio r of the two MAC schemes. The relative performance of π_R and π_D , therefore, depends on the value of r comparing with 1: If $r > 1$, π_D is better than π_R ; otherwise, π_R outperforms π_D . The following theorem shows the condition for $r > 1$ (or < 1) in a large sensor network.

Theorem 2: Let P_{out} be given. Suppose that 1) $\rho D \rightarrow \infty$ as $M \rightarrow \infty$ and 2) $D \ln M/M \rightarrow 0$ as $M \rightarrow \infty$. Then, $r > 1$ for $P_{\text{out}} < e^{-\lambda(1+o(1))}$ and $r < 1$ for $P_{\text{out}} > e^{-\lambda(1+o(1))}$.

Proof: By (31) and (32) in Theorem 1, we have the ratio r as

$$r = \ln P_{\text{out}}^{-\frac{1}{\lambda(1+o(1))}} \cdot \frac{1 + O\left(\frac{\ln \ln M}{\ln M}\right)}{1 + O\left(\frac{\ln \ln M}{\ln M}\right) + \ln \frac{1}{P_{\text{out}}} O\left(\frac{1}{\ln M}\right)}. \quad (35)$$

We see that the second term goes to 1 as M grows. Therefore, the ratio r is determined by the first leading term. Comparing it to 1, we have the threshold on P_{out} . ■

From (35), it is clear that the relative performance r of π_R and π_D depends critically on P_{out} . As we see, although the deterministic scheduling is designed for an optimal retrieval pattern, its performance is less robust in the sense that it suffers from missing data samples due to sensor outage. Theorem 2 gives us the transition of the relative performance of the two strategies.

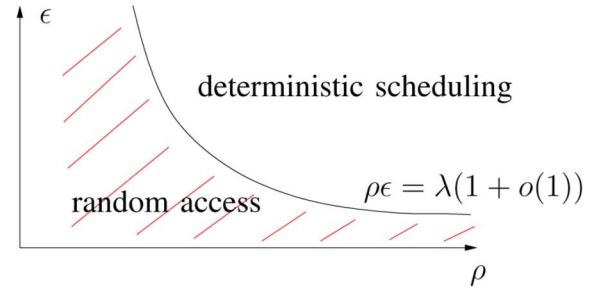


Fig. 5. Favored region of the two MAC schemes.

For $P_{\text{out}} < e^{-\lambda(1+o(1))}$, the gain of scheduling transmission with desired pattern is evident, although there exists a mild performance loss because of sensor outage. As we expected, the deterministic scheduler provides a better reconstruction performance. In [9], we have considered the case when the sensor density ρ goes to infinity, and therefore there is no sensor outage. We have shown that, in fact, the excess distortion ratio r grows as $\ln M + O(\ln \ln M)$. When $P_{\text{out}} > e^{-\lambda(1+o(1))}$, however, Theorem 1 shows that the performance loss of π_D due to missing data samples does not justify the effort of scheduling the desired retrieval pattern, and random access outperforms the optimal scheduling.

Note that there are two factors of a MAC scheme affecting the signal reconstruction: the sampling pattern and the throughput of MAC. The former determines where the samples are from, and the latter determines how many samples the AP can obtain. Because of sensor outage, the resulting sampling pattern by using π_D degrades the performance due to missing data samples. For the random access π_R , on the other hand, the number of samples is limited by the throughput. The threshold $e^{-\lambda(1+o(1))}$ on P_{out} for the relative performance clearly indicates the above two factors: the condition $P_{\text{out}} < e^{-\lambda(1+o(1))}$ is equivalent to $\rho\epsilon > \lambda(1+o(1))$, i.e., we should set the length of the resolution interval so that the expected number of sensors in the resolution interval is greater than $\lambda(1+o(1))$. Fig. 5 depicts the favored region of each MAC scheme. The two regions are divided by the curve $\rho\epsilon = \lambda(1+o(1))$. For fixed ρ , the resolution interval ϵ needs to be larger than $\lambda(1+o(1))/\rho$ to let π_D perform better. On the other hand, if we keep the ϵ of π_D fixed and vary ρ , then for ρ lower than a certain value, the expected number of sensors in ϵ is too small, and it is better to perform random access. Our result implies that if the average number of received data sampling in random access is more than that in deterministic scheduler, then we may just let sensors perform random access instead of careful planning for desired retrieval pattern. This shows that although the deterministic scheduler can be optimally designed, it relies on the network parameters and is therefore less robust to the imperfect knowledge of the network conditions.

VI. SIMULATIONS

We now numerically compare the reconstruction performance of π_D and π_R for various scenarios. The expected maximum distortion of π_R and π_D is calculated using (24) and (27), respectively.

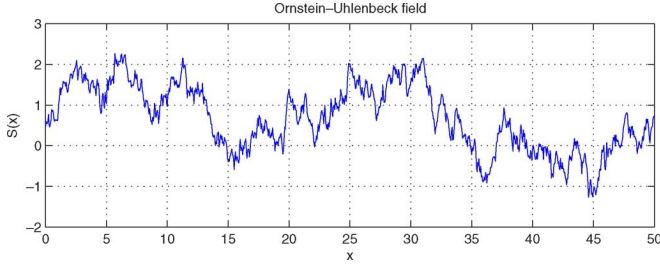


Fig. 6. Signal field $S(x)$ ($D = 500$ m, $f = 0.02$ m, $\sigma_s^2 = 1$).

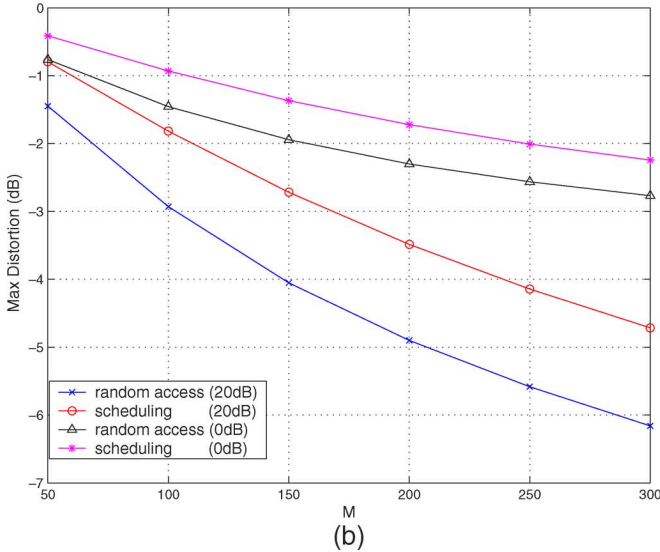
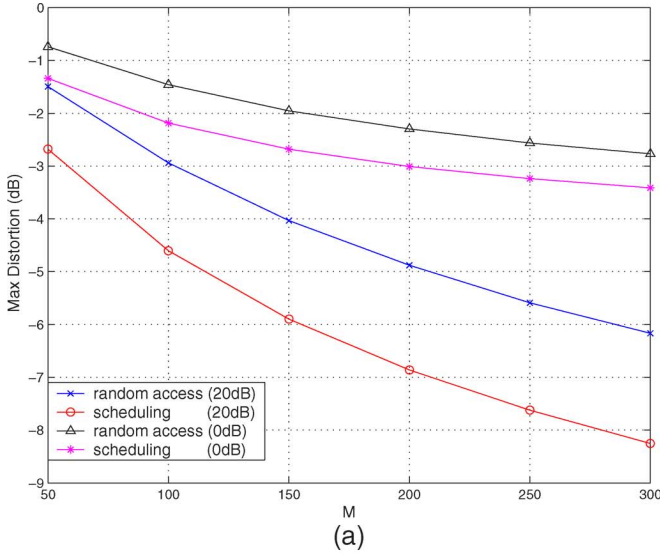


Fig. 7. Expected maximum distortion $\bar{\mathcal{E}}(M, \text{SNR})$ versus M ($\rho = 5$ sensors/m). (a) $P_{\text{out}} = 0.5$; (b) $P_{\text{out}} = 0.75$.

A. Signal Field and System Setup

We consider a field \mathcal{A} with $D = 500$ m. For the signal of interest $S(x)$, an example of the *Ornstein-Uhlenbeck field* is shown in Fig. 6. We set $\sigma_s^2 = 1$ and $f = 0.05$ m (i.e., for $|x_1 - x_0| = 1$ m, $E(S(x_1)S(x_0)) = 0.95$). We set the sensor density $\rho = 5$ sensors/m and randomly deploy Poisson distributed sensors in \mathcal{A} . For the random access, we choose a simple ALOHA with delay-first transmission without feedback, where each of

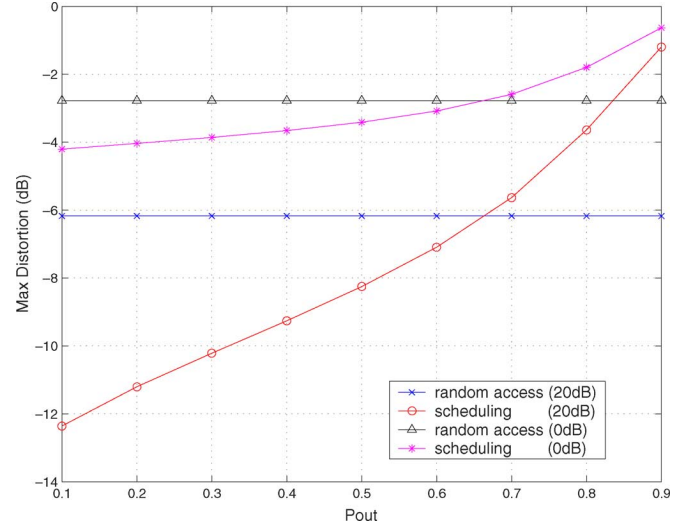


Fig. 8. $\bar{\mathcal{E}}(M, \text{SNR})$ versus P_{out} ($\rho = 5$ sensors/m, $M = 300$).

the n activated sensors transmits its packet with probability $1/n$. In this case, the throughput, i.e., the probability of a successful transmission, is $\lambda_n = (1 - 1/n)^{n-1}$ and $\lambda = \lim_{n \rightarrow \infty} \lambda_n = 1/e$.

B. Distortion versus Collection Time M

We plot in Fig. 7 the expected maximum distortion $\bar{\mathcal{E}}(M, \text{SNR})$ of π_R and π_D with various M for $\text{SNR} = 0, 20$ dB, respectively. Fig. 7(a) and (b) are the plots for probability of sensor outage $P_{\text{out}} = 0.5$ and $P_{\text{out}} = 0.75$, respectively. We observe that as M increases, on average the received data samples at the AP increases, and $\bar{\mathcal{E}}(M, \text{SNR})$ of both π_R and π_D decreases. However, we see that the relative reconstruction performance of π_D and π_R depends on the different value of P_{out} we set. For the relatively high P_{out} we set, the reconstruction performance of the random access π_R is uniformly better than that of the deterministic scheduler π_D .

C. Distortion Versus Sensor Outage Probability P_{out}

To see the transition of the relative performance of π_D and π_R , we plot $\bar{\mathcal{E}}(M, \text{SNR})$ of π_D and π_R with $M = 300$ for various P_{out} at $\text{SNR} = 0, 20$ dB in Fig. 8. Note that since P_{out} is the design parameter for π_D , the performance of π_R does not change for different P_{out} . We observe that clearly there exist a threshold on P_{out} approximately at $0.67 (\approx e^{-e^{-1}})$, such that the performance of π_D is worse than π_R for all P_{out} greater than this value, and better for P_{out} less than this value.

We also compare the ratio r of excess distortion obtained in our simulations with the theoretical expression of r in (35) of Theorem 2, where we use the coefficients in (36) of Lemma 1 and (57) of Lemma 3 to replace the O 's in (35). Fig. 9 shows the curves of r versus P_{out} for $\text{SNR} = 20$ dB and $M = 300$. We see that the simulation matches the result in Theorem 2.

D. Distortion Versus SNR

In Fig. 10, we plot the expected maximum distortion $\bar{\mathcal{E}}(M, \text{SNR})$ of π_R and π_D with various SNR for $M = 300$. We plot $\bar{\mathcal{E}}(M, \text{SNR})$ of π_D at $P_{\text{out}} = 0.5, 0.75$, where the

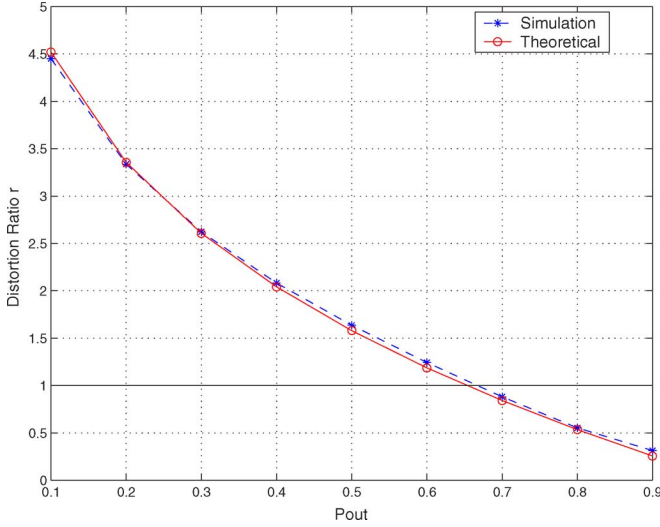


Fig. 9. Distortion ratio r versus P_{out} ($\rho = 5$ sensors/m, $M = 300$, SNR = 20 dB).

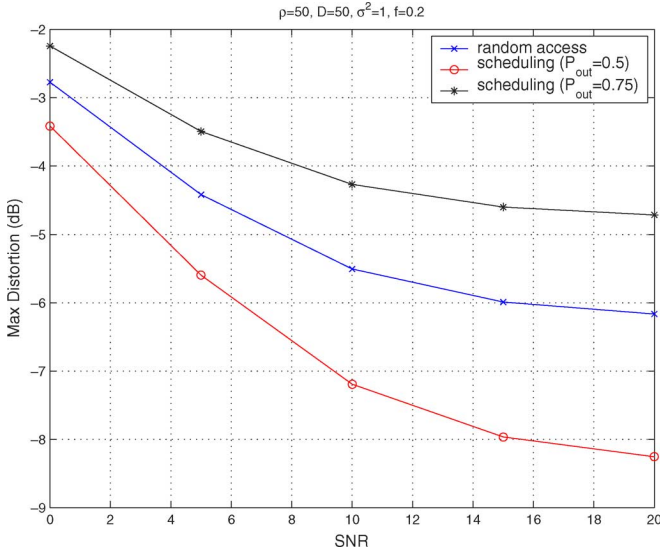


Fig. 10. $\bar{E}(M, \text{SNR})$ versus SNR ($\rho = 5$ sensors/m, $M = 300$).

performance of π_D is better than that of π_R at $P_{\text{out}} = 0.5$ and worse at $P_{\text{out}} = 0.75$. For each case, we see that as SNR becomes large, the performance gap between π_R and π_D increases. As expected, when noise level is high, the distortion is mainly dominated by the measurement noise, therefore the performance gap is small. When SNR becomes higher, the effect on the reconstruction distortion due to using different MAC scheme becomes more significant, and the performance gap becomes larger.

E. Distortion versus Signal Spatial Correlation

Finally, we compare the performance under various signal spatial correlation levels. From (3), the signal correlation coefficient is given by $a \triangleq e^{-f}$, where $a \in (0, 1)$. In Fig. 11(a) and (b), we plot $\bar{E}(M, \text{SNR})$ of π_R and π_D with various a for SNR = 20 and 0 dB, respectively. We set $M = 300$ and $P_{\text{out}} = 0.75$. For f ranges from 0.02 to 0.8, the correlation coefficient a ranges from 0.98 (high correlation) to 0.45

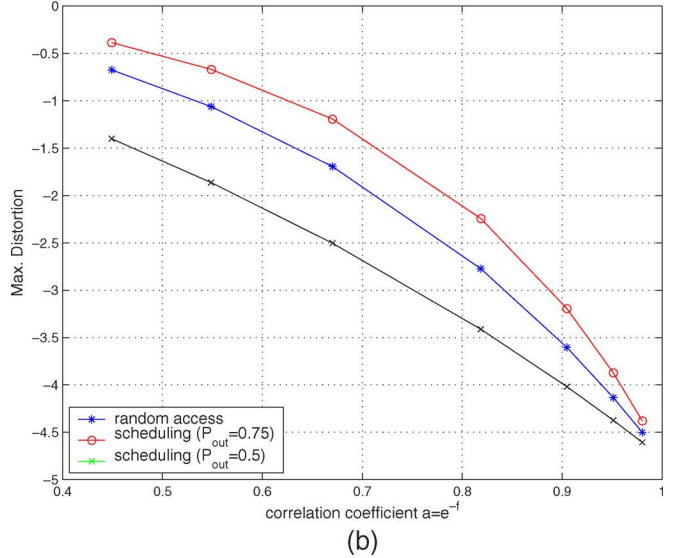
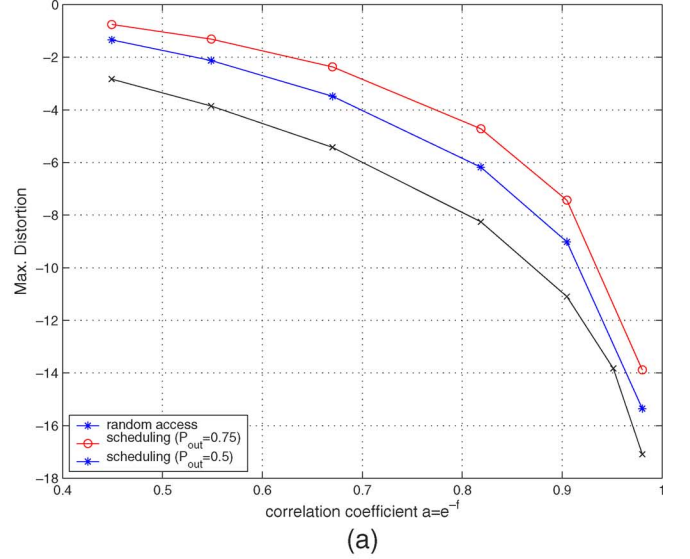


Fig. 11. $\bar{E}(M, \text{SNR})$ versus correlation coefficient a ($\rho = 5$ sensors/m, $M = 300$). (a) SNR = 20 dB; (b) SNR = 0 dB.

(low correlation). As expected, the lower spatial correlation a , the higher distortion. For the performance gap at different correlation level, from Theorem 2, we see that f is not in the leading term of the expression for r . Therefore, for M relatively large, the spatial correlation does not affect the ratio r . For high SNR, where the limiting distortion $\bar{E}(\infty, \text{SNR})$ is very small, the excess distortion dominates $\bar{E}(M, \text{SNR})$. Thus, the performance gap (in decibels) is relatively unchanged. For low SNR, however, $\bar{E}(\infty, \text{SNR})$ is relatively large. For high spatial correlation, $\bar{E}(M, \text{SNR})$ is dominated by the limiting distortion $\bar{E}(\infty, \text{SNR})$. Thus, the performance gap (in decibels) when a is high is less significant than that when a is low.

VII. CONCLUSION

In this paper, we investigate how MAC for information retrieval affects information processing. Specifically, we consider the performance of signal field reconstruction based on received data. Data retrieval with different MAC schemes results in distinct sampling pattern of the field, and throughput of MAC

determines the number of samples received by the AP. Our analysis demonstrates that, for MAC design in sensor networks, a cross-layer approach that integrates application and MAC should be considered. In particular, affecting the choices of MAC are sensor density and reconstruction performance, both dictated by the source characteristics and network parameters.

Using the maximum reconstruction distortion as our performance measure, we examined two MAC strategies: the deterministic scheduling designed to collect data from equally spaced locations and the random access MAC. We have shown that, although the scheduling is designed according to the optimal retrieval pattern, its performance suffers from the possibility of sensor outage, which results in missing data samples and, therefore, is less robust to the uncertainty of the network parameters. Specifically, the relative reconstruction performance is shown to depend critically on the outage probability P_{out} of sensors, which determines the average number of active sensors in a resolution interval. Our analysis shows that, for networks with large sensor population, whether a deterministic scheduler is better than random access depends on whether the critical threshold is $e^{-\lambda(1+o(1))}$. The former is better only if $P_{\text{out}} < e^{-\lambda(1+o(1))}$. Based on this, the favored region of each MAC is shown. The two regions are divided by $\rho\epsilon = \lambda(1 + o(1))$, where $\rho\epsilon$ is the expected number of sensors in the resolution interval ϵ , and $P_{\text{out}} < e^{-\lambda(1+o(1))}$ corresponds to $\rho\epsilon > \lambda(1 + o(1))$. This indicates that if the average number of samples obtained using the deterministic scheduler is less than that of random access in a given collection time, then one may just let sensors perform random access.

Although our results provide insight of the relation of MAC and application performance in sensor network, many problems remain for further study. In this paper, the analysis and comparison of different MAC schemes for signal reconstruction is limited to the 1-D sensor field. Even though the results provide insight into the 2-D network, detailed studies of the interaction of MAC and signal reconstruction performance in the 2-D field are still of interest. The extension to the 2-D case is nontrivial as the signal modeling and the reconstruction metric expression become much more complex. In order to make the problem more tractable for analysis, some level of simplified approximation for the 2-D models and signal estimation method may be assumed. For example, the field can be quantized into discrete grids, and a Gauss–Markov field on the lattice, often used for 2-D modeling, can be applied to study our problem. Also, in this study, we focus on the MMSE signal reconstruction for the random signal model considered in (1). A more generalized random signal model can be considered for further detailed examination. The performance in many other applications, such as signal prediction, spectral estimation, is also of great interest for investigation.

APPENDIX I PROOF OF THEOREM 1

A. Asymptotic Distortion of Deterministic Scheduler π_D

To calculate the expected maximum distortion $\bar{\mathcal{E}}_D(M, \text{SNR})$ of π_D , we first need to find the statistics of L_M , the largest

number of consecutive intervals that have sensor outage. For Poisson-distributed sensors, the sensor outage in each disjoint interval of length ϵ is i.i.d. with probability P_{out} . The problem of finding L_M essentially is equivalent to that of finding the longest run of consecutive heads for a sequence of M biased coin tossing, where the probability of head is P_{out} . Using the extreme value theory of head runs [11], we show that L_M possesses an almost sure behavior. For convenience, let

$$\log = \log_{1/P_{\text{out}}}.$$

The result is shown in the following lemma.

Lemma 1: Let L_M be defined as before. Then

$$P\left(\liminf_{M \rightarrow \infty} \left[1 - \frac{\log \log M}{\log M} \leq \frac{L_M}{\log M} < 1 + \frac{2 \log \log M}{\log M}\right]\right) = 1. \quad (36)$$

Therefore

$$\frac{L_M}{\log M} \rightarrow 1 \quad \text{a.s.} \quad (37)$$

Proof: Theorem 3 and Theorem 4 of [11] provide almost sure behavior of L_M . We state these theorems below.

Theorem 3: Let $\{d(M)\}$ be a nondecreasing sequence of integers. Then, $P\{L_M \geq d(M) \text{ i.o.}\}$ is 0 or 1 as $\sum_M P_{\text{out}}^{d(M)}$ finite or infinite.

Proof: See [11]. ■

Theorem 4: Let $\{d(M)\}$ be a nondecreasing sequence of integers with $\liminf_{M \rightarrow \infty} MP_{\text{out}}^{d(M)} / \log \log M > 0$ and $\limsup_{M \rightarrow \infty} MP_{\text{out}}^{d(M)} / d(M) < \infty$. Then, $P\{L_M < d(M) \text{ i.o.}\}$ is 0 or 1, as determined by whether $\sum_M P_{\text{out}}^{d(M)} \exp(-M(1 - P_{\text{out}})P_{\text{out}}^{d(M)})$ is finite or infinite.

Proof: See [11]. ■

Let $d(M) = \log M + 2 \log \log M$. By Theorem 3

$$\sum_{M=2}^{\infty} P_{\text{out}}^{\log M + 2 \log \log M} = \sum_{M=2}^{\infty} \frac{1}{M(\log M)^2} < \infty. \quad (38)$$

This implies

$$P\left(\liminf_{M \rightarrow \infty} [L_M < \log M + 2 \log \log M]\right) = 1. \quad (39)$$

On the other hand, for $d(M) = \log M - \log \log M$, it satisfies the two conditions in Theorem 4:

$$\liminf_{M \rightarrow \infty} \frac{MP_{\text{out}}^{d(M)}}{\log \log M} = \liminf_{M \rightarrow \infty} \frac{\log M}{\log \log M} > 0 \quad (40)$$

$$\limsup_{M \rightarrow \infty} \frac{MP_{\text{out}}^{d(M)}}{d(M)} = \limsup_{M \rightarrow \infty} \frac{\log M}{\log M - \log \log M} < \infty. \quad (41)$$

By Theorem 4

$$\begin{aligned} \sum P_{\text{out}}^{d(M)} \exp\left(-M(1 - P_{\text{out}})P_{\text{out}}^{d(M)}\right) \\ = \sum_{M=2}^{\infty} \frac{\log M}{M^{1 + \frac{1 - P_{\text{out}}}{\ln 1/P_{\text{out}}}}} < \infty. \end{aligned} \quad (42)$$

This implies

$$P\left(\liminf_{M \rightarrow \infty} [L_M \geq \log M - \log \log M]\right) = 1. \quad (43)$$

Combining (39) and (43), (36) and (37) follows. \blacksquare

Using Lemma 1, $\exists \bar{M}$, such that $\forall M > \bar{M}$, the lower bound of the expected maximum distortion of π_D in (27) can be rewritten as

$$\begin{aligned} & E\left[\mathcal{E}\left(\frac{D(L_M+1)}{M+1} - \epsilon, \text{SNR}\right)\right] \\ &= \sigma_s^2 E\left[\frac{\frac{1}{\text{SNR}} + 1 - e^{-fD\left(\frac{\log M}{M+1} \frac{L_M}{\log M} + \frac{1}{M+1} - \frac{\epsilon}{D}\right)}}{\frac{1}{\text{SNR}} + 1 + e^{-fD\left(\frac{\log M}{M+1} \frac{L_M}{\log M} + \frac{1}{M+1} - \frac{\epsilon}{D}\right)}}\right] \\ &= \sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1 - e^{-fD\left[\frac{\log M}{M+1}(1+O(\frac{\log \log M}{\log M})) + \frac{1}{M+1} - \frac{\epsilon}{D}\right]}}{\frac{1}{\text{SNR}} + 1 + e^{-fD\left[\frac{\log M}{M+1}(1+O(\frac{\log \log M}{\log M})) + \frac{1}{M+1} - \frac{\epsilon}{D}\right]}} \\ &= \sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1 - e^{-fD\left(\frac{\log M}{M+1} + O(\frac{\log \log M}{M}) + O(\frac{1}{M})\right)}}{\frac{1}{\text{SNR}} + 1 + e^{-fD\left[\frac{\log M}{M+1} + O(\frac{\log \log M}{M}) + O(\frac{1}{M})\right]}} \quad (44) \end{aligned}$$

where since $\epsilon \leq D/(M+1)$, we have $1/(M+1) - \epsilon/D = O(1/M)$. For function $g(x) = (a - e^{-x})/(a + e^{-x})$, as $x \rightarrow 0$, using Taylor's Theorem, we have

$$g(x) = \frac{a-1}{a+1} + \frac{2ax}{(a+1)^2} + O(x^2). \quad (45)$$

If the growth rate for D is such that $D \ln M/M \rightarrow 0$ as $M \rightarrow \infty$, then we can apply (45) to (44), and we obtain the expected maximum distortion as

$$\begin{aligned} & E\left[\mathcal{E}\left(\frac{D(L_M+1)}{M+1} - \epsilon, \text{SNR}\right)\right] \\ &= \frac{\sigma_s^2}{1+2\text{SNR}} + 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{\left(\frac{1}{\text{SNR}} + 2\right)^2} \\ &\quad \times \left(D \frac{\log M}{M} + O\left(\frac{\log \log M}{M}\right) + O\left(\frac{1}{M}\right)\right) + O\left(\frac{\log^2 M}{M^2}\right) \\ &= \frac{\sigma_s^2}{1+2\text{SNR}} + 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{\left(\frac{1}{\text{SNR}} + 2\right)^2} \\ &\quad \times \left(D \frac{\log M}{M} + O\left(\frac{\log \log M}{M}\right) + O\left(\frac{1}{M}\right)\right). \quad (46) \end{aligned}$$

The upper bound in (27) can be similarly derived and has the same order as in (44). Therefore, we have the asymptotic expression in (31).

B. Asymptotic Distortion of Random Access π_R

1) *Asymptotic Expression of $\bar{\mathcal{E}}_R(M_o, \text{SNR})$ for Large M_o :* Now, we obtain the distortion of the random access π_R , given M_o received packets from distinct sensors. From (17), the maximum distortion of π_R is a function of $d_{\max}^{(M_o)}$

$$d_{\max}^{(M)} = \max_{0 \leq i \leq M+1} P_{(i)} - P_{(i-1)} \quad (47)$$

where $P_{(i)}, i = 1, \dots, M$ is the order statistics of M i.i.d. uniformly distributed random variables. The property of $\{P_{(i)}\}$ is given in the following lemma.

Lemma 2: Let E_n be i.i.d. unit exponential random variables, and $\Gamma_n = E_1 + \dots + E_n$, for $n \geq 1$. Then, the joint distribution of $(\Gamma_1/\Gamma_{n+1}, \dots, \Gamma_n/\Gamma_{n+1})$ is the same as that of the order statistics from n i.i.d. random variables with uniform distribution $U(0,1)$, i.e.,

$$f_{\frac{\Gamma_1}{\Gamma_{n+1}}, \dots, \frac{\Gamma_n}{\Gamma_{n+1}}}(u_1, \dots, u_n) = \begin{cases} n!, & \text{if } 0 < u_1 < \dots < u_n < 1 \\ 0, & \text{otherwise.} \end{cases} \quad (48)$$

Proof: The joint distribution of E_1, \dots, E_n is given by

$$f_{E_1, \dots, E_{n+1}}(x_1, \dots, x_{n+1}) = e^{-\sum_{i=1}^{n+1} x_i} \quad (49)$$

where $x_i > 0, 1 \leq i \leq n+1$. Note that $\Gamma_i = \sum_{j=1}^i E_j$. For any $t > 0$, let $u_i = (1/t) \sum_{j=1}^i x_j$. By transformation of random variables, we then have the joint density of $(\Gamma_1/t, \dots, \Gamma_{n+1}/t)$ given by

$$f_{\frac{\Gamma_1}{t}, \dots, \frac{\Gamma_{n+1}}{t}}(u_1, \dots, u_{n+1}) = \frac{e^{-tu_{n+1}}}{t^{n+1}}, \quad 0 < u_1, \dots, u_{n+1}. \quad (50)$$

Notice that Γ_{n+1} has a gamma density

$$f_{\Gamma_{n+1}}(t) = \frac{e^{-t} t^n}{n!}, \quad t \geq 0.$$

It follows that

$$f_{\frac{\Gamma_{n+1}}{t}}(1) = \frac{e^{-t} t^{n+1}}{n!}, \quad t > 0. \quad (51)$$

Therefore, from (50) and (51), we have the joint density

$$\begin{aligned} & f_{\frac{\Gamma_1}{\Gamma_{n+1}}, \dots, \frac{\Gamma_n}{\Gamma_{n+1}}}(u_1, \dots, u_n) \\ &= f_{\frac{\Gamma_1}{t}, \dots, \frac{\Gamma_n}{t} | \Gamma_{n+1}=t}(u_1, \dots, u_n) \\ &= n!, \quad 0 < u_1 < \dots < u_n < 1. \end{aligned} \quad (52)$$

By Lemma 2, then the joint distribution of the order statistics $(P_{(1)}, \dots, P_{(M)})$ is the same as that of $D((\Gamma_1/\Gamma_{M+1}), \dots, (\Gamma_M/\Gamma_{M+1}))$. Therefore, the joint distribution of $(d_0^{(M)}, \dots, d_M^{(M)})$ is the same as that of $D((\Gamma_1 - \Gamma_0/\Gamma_{M+1}), \dots, (\Gamma_{M+1} - \Gamma_M/\Gamma_{M+1}))$, where $\Gamma_0 = 0$. Thus, the maximum distance $d_{\max}^{(M)}$ can be represented by

$$\begin{aligned} d_{\max}^{(M)} &= \frac{D}{\Gamma_{M+1}} \max_{1 \leq i \leq M+1} \{\Gamma_i - \Gamma_{i-1}\} \\ &= \frac{D}{\Gamma_{M+1}} \max_{1 \leq i \leq M+1} E_i. \end{aligned} \quad (53)$$

Therefore, we have

$$(M+1)d_{\max}^{(M)} = D \frac{M+1}{\Gamma_{M+1}} \max_{1 \leq i \leq M+1} E_i. \quad (54)$$

By the Strong Law of Large Numbers, we have

$$\frac{\Gamma_{M+1}}{M+1} = \frac{\sum_{i=1}^{M+1} E_i}{M+1} \rightarrow 1 \quad \text{a.s.} \quad (55)$$

Let $A_M = \max_{1 \leq i \leq M} E_i$. From the asymptotic theory of extreme statistics, we have the almost sure asymptotic property of A_M as follows.

Lemma 3 (Logarithmic Growth Rate of A_M [28, p. 262], [29, p. 215]): Let A_M be the maximum of M i.i.d. unit exponential random variables. Then

$$P\left(\liminf_{M \rightarrow \infty} \left[1 - \frac{\ln \ln M}{\ln M} \leq \frac{A_M}{\ln M} < 1 + \frac{2 \ln \ln M}{\ln M}\right]\right) = 1. \quad (56)$$

Combining (54)–(56), we have

$$P\left(\liminf_{M \rightarrow \infty} \left[\left|\frac{(M+1)d_{\max}^{(M)}}{\ln M} - D\right| < \frac{2 \ln \ln M}{\ln M}\right]\right) = 1. \quad (57)$$

Thus, similarly as in (44)–(46), if $D \ln M_o / M_o \rightarrow 0$ as $M_o \rightarrow \infty$, for large enough M_o of received samples, the expected maximum distortion of π_R can be obtained as

$$\begin{aligned} \bar{\mathcal{E}}_R(M_o, \text{SNR}) &= \sigma_s^2 \frac{1}{1 + 2\text{SNR}} + 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{\left(\frac{1}{\text{SNR}} + 2\right)^2} \\ &\quad \times \left(D + O\left(\frac{\ln \ln M_o}{\ln M_o}\right)\right) \frac{\ln M_o}{M_o}. \end{aligned} \quad (58)$$

We now obtain the limiting distribution of $\bar{\mathcal{E}}_R(M, \text{SNR})$. Without loss of generality, we consider that the AP activates the sensors in the whole field for transmissions. Note that, given $N = n$, the throughput λ_n is just the probability of a successful transmission. Thus, M_o is binomial distributed with

$$P(M_o = k | N = n) = \binom{M}{k} \lambda_n^k (1 - \lambda_n)^{M-k}. \quad (59)$$

Because $\lambda_n \rightarrow \lambda$, for a given $c > 0$, $\exists n_1$, s.t. for all $n > n_1$, $|\lambda_n - \lambda| < c$. By the Chebyshev inequality, for fixed c , we have

$$P(M_o \leq M\lambda(1-2c) | N = n) < \frac{\lambda_n(1-\lambda_n)}{M\lambda^2c^2} < \frac{1}{M\lambda^2c^2} \quad (60)$$

$$P(M_o > M\lambda(1+2c) | N = n) < \frac{\lambda_n(1-\lambda_n)}{M\lambda^2c^2} < \frac{1}{M\lambda^2c^2}. \quad (61)$$

Now we bound $\bar{\mathcal{E}}_R(M, \text{SNR})$. Note that $\bar{\mathcal{E}}_R(M_o, \text{SNR})$ decreases with M_o . We have

$$\begin{aligned} &\bar{\mathcal{E}}_R(M\lambda(1+2c), \text{SNR}) P(M_o \leq M\lambda(1+2c); \rho) \\ &\leq E[\bar{\mathcal{E}}_R(M_o, \text{SNR})] \\ &\leq P(M_o \leq M\lambda(1-2c); \rho) \bar{\mathcal{E}}_R(0, \text{SNR}) \\ &\quad + \bar{\mathcal{E}}_R(M\lambda(1-2c), \text{SNR}). \end{aligned} \quad (62)$$

Therefore

$$\begin{aligned} &\bar{\mathcal{E}}_R(M\lambda(1+2c), \text{SNR}) \lim_{\rho \rightarrow \infty} P(M_o \leq M\lambda(1+2c); \rho) \\ &\leq \lim_{\rho \rightarrow \infty} E[\bar{\mathcal{E}}_R(M_o, \text{SNR})] \\ &\leq \lim_{\rho \rightarrow \infty} P(M_o \leq M\lambda(1-2c); \rho) \bar{\mathcal{E}}_R(0, \text{SNR}) \\ &\quad + \bar{\mathcal{E}}_R(M\lambda(1-2c), \text{SNR}). \end{aligned}$$

Note that

$$\begin{aligned} &\lim_{\rho \rightarrow \infty} P(M_o \leq M\lambda(1-2c); \rho) \\ &= \sum_{n=n_1}^{\infty} P(M_o \leq M\lambda(1-2c) | N = n) P(N = n) \\ &< \frac{1}{M\lambda^2c^2}. \end{aligned} \quad (63)$$

Similarly

$$\lim_{\rho \rightarrow \infty} P(M_o > M\lambda(1+2c); \rho) < \frac{1}{M\lambda^2c^2}. \quad (64)$$

Following $\bar{\mathcal{E}}_R(M_o, \text{SNR})$ in (58) for large M_o , we have the limiting distortion as

$$\begin{aligned} \bar{\bar{\mathcal{E}}}_R(\infty, \text{SNR}) &= \lim_{M \rightarrow \infty} \lim_{\rho \rightarrow \infty} E[\bar{\mathcal{E}}_R(M_o, \text{SNR})] \\ &= \sigma_s^2 \frac{1}{1 + 2\text{SNR}}. \end{aligned} \quad (65)$$

2) *Lower Bound on $\bar{\bar{\mathcal{E}}}_R(M, \text{SNR}) - \bar{\bar{\mathcal{E}}}(\infty, \text{SNR})$:* We now lower bound $\bar{\bar{\mathcal{E}}}_R(M, \text{SNR}) - \bar{\bar{\mathcal{E}}}(\infty, \text{SNR})$ for large M and \bar{N} . We have

$$\begin{aligned} Pr(M_o = k; \rho) &= E[Pr(M_o = k | N)] \\ &\geq \sum_{n=0}^{\infty} P(M_o = k | N = n) P(N = n). \end{aligned} \quad (66)$$

The lower bound on the excess distortion is given by

$$\begin{aligned} &\bar{\bar{\mathcal{E}}}(M, \text{SNR}) - \bar{\bar{\mathcal{E}}}(\infty, \text{SNR}) \\ &= E\left[\underbrace{\bar{\mathcal{E}}(k, \text{SNR}) - \bar{\bar{\mathcal{E}}}(\infty, \text{SNR})}_{\bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR})}\right] \\ &= \sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) \\ &\geq \sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) \sum_{n=n_0}^{\infty} P(M_o = k | N = n) P(N = n) \\ &= \sum_{n=n_0}^{\infty} P(N = n) \\ &\quad \times \left(\sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k | N = n)\right). \end{aligned} \quad (67)$$

Since $\lambda = \lim_{n \rightarrow \infty} \lambda_n$, for any $\varepsilon_1 > 0$, $\exists n_2$, s.t. for all $n > n_2$

$$M\lambda(1-\varepsilon_1) < M\lambda_n < M\lambda(1+\varepsilon_1). \quad (68)$$

Also we have

$$\begin{aligned} &\sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k | N = n) \\ &> \sum_{k=M\lambda(1-\varepsilon_1)}^{M\lambda(1+\varepsilon_1)} \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k | N = n). \end{aligned} \quad (69)$$

Since $\bar{\mathcal{E}}(k, \text{SNR})$ is a decreasing function of k , we have

$$\sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k | N = n) > \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1+\varepsilon_2), \text{SNR}) \times P[M\lambda(1-\varepsilon_2) \leq M_o \leq M\lambda(1+\varepsilon_2) | N = n]. \quad (70)$$

Combining (67) and (70), we have

$$\bar{\mathcal{E}}(M, \text{SNR}) - \bar{\mathcal{E}}(\infty, \text{SNR}) > \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1+\varepsilon_1), \text{SNR}) \times \sum_{n=n_2}^{\infty} P(N=n) P[M\lambda(1-\varepsilon_1) \leq M_o \leq M\lambda(1+\varepsilon_1) | N=n]. \quad (71)$$

The tail probability of M_o can be again bounded by the Chebyshev inequality. Then, $\forall \varepsilon_3 > 0$, $\exists \bar{M}_1$, s.t. for all $M > \bar{M}_1$

$$P[M\lambda(1-\varepsilon_1) \leq M_o \leq M\lambda(1+\varepsilon_1) | N=n] \geq 1 - \varepsilon_2. \quad (71)$$

Since the number of sensors in \mathcal{A} is Poisson distributed, for any $n_o < \infty$

$$\lim_{\bar{N} \rightarrow \infty} P(N \leq n_o) = \lim_{\bar{N} \rightarrow \infty} e^{-\bar{N}} \sum_{k=0}^{n_o} (\bar{N})^k / k! = 0. \quad (72)$$

Thus, for any $\varepsilon_3 > 0$

$$\exists \bar{N}_o(M) > 0, \text{ s.t., for all } \bar{N} > \bar{N}_o(M), \\ P(N \leq \max\{n_o, n_2\}) < \varepsilon_3. \quad (73)$$

Thus, for $M > \bar{M}_1$ and $\bar{N} > \bar{N}_o(M)$

$$\begin{aligned} \bar{\mathcal{E}}(M, \text{SNR}) - \bar{\mathcal{E}}(\infty, \text{SNR}) &> (1-\varepsilon_2)(1-\varepsilon_3) \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1+\varepsilon_1), \text{SNR}) \\ &\triangleq (1-\varepsilon) \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1+\varepsilon_1), \text{SNR}). \end{aligned} \quad (74)$$

From (58), for large M_o , we have

$$\begin{aligned} \bar{\mathcal{E}}_{\text{ex}}(M_o, \text{SNR}) &= 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{(\frac{1}{\text{SNR}} + 2)^2} \\ &\times \left(D + O\left(\frac{\ln \ln M_o}{\ln M_o}\right) \right) \frac{\ln M_o}{M_o}. \end{aligned} \quad (75)$$

Therefore, by (74), for large M and $\bar{N} > \bar{N}_o(M)$, we have the lower bound

$$\begin{aligned} \bar{\mathcal{E}}(M, \text{SNR}) - \bar{\mathcal{E}}(\infty, \text{SNR}) &> 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{(\frac{1}{\text{SNR}} + 2)^2} \\ &\times \left(D \frac{\ln M\lambda(1+\varepsilon_2)}{M\lambda(1+\delta)} + O\left(\frac{\ln \ln M}{M}\right) \right) \end{aligned} \quad (76)$$

where $\delta = (\varepsilon + \varepsilon_2)/(1-\varepsilon)$.

3) *Upper Bound on $\bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}(\infty, \text{SNR})$:* For the upper bound, we have

$$\begin{aligned} \bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}_R(\infty, \text{SNR}) &= \sum_{k=0}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) \\ &= \sum_{k=M\lambda(1-\varepsilon_1)}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) \end{aligned}$$

$$+ \sum_{k=0}^{M\lambda(1-\varepsilon_2)-1} \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho). \quad (77)$$

Since $\bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR})$ is a decreasing function of k , for the first term, we have

$$\begin{aligned} \sum_{k=M\lambda(1-\varepsilon_1)}^M \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) &< \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1-\varepsilon_1), \text{SNR}) \cdot \end{aligned} \quad (78)$$

For the second term, we have

$$\begin{aligned} \sum_{k=0}^{M\lambda(1-\varepsilon_1)-1} \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) &< \bar{\mathcal{E}}_{\text{ex}}(0, \text{SNR}) P(M_o < M\lambda(1-\varepsilon_2); \rho). \end{aligned} \quad (79)$$

As in the derivation of the lower bound, $\exists \bar{N}_o(M)$, s.t. for all $\bar{N} > \bar{N}_o(M)$

$$P(N \leq \max\{n_o, n_2\}) < \frac{1}{M}. \quad (80)$$

Therefore

$$\begin{aligned} P(M_o < M\lambda(1-\varepsilon_1); \rho) &< P(N \leq \max\{n_o, n_2\}) \\ &+ \sum_{n=\max\{n_o, n_2\}}^{\infty} P(M_o < M\lambda(1-\varepsilon_1) | N=n) P(N=n) \end{aligned} \quad (81)$$

$$< \frac{c_1}{M} \quad (82)$$

for some $c_1 > 0$, where the bound on the second term in (81) is similar to that in (60). Therefore

$$\sum_{k=0}^{M\lambda(1-\varepsilon_1)-1} \bar{\mathcal{E}}_{\text{ex}}(k, \text{SNR}) P(M_o = k; \rho) < \frac{c_1 \bar{\mathcal{E}}_{\text{ex}}(0, \text{SNR})}{M}. \quad (83)$$

Therefore, combining (78) and (83), $\exists \bar{N}_o(M)$, for $\bar{N} > \bar{N}_o(M)$, the excess distortion in (77) is upper bounded by

$$\bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}_R(\infty, \text{SNR}) \leq \bar{\mathcal{E}}_{\text{ex}}(M\lambda(1-\varepsilon_1), \text{SNR}) + \frac{c_3}{M} \quad (84)$$

for some $c_3 > 0$. From (75), for large M and $\bar{N} > \bar{N}_o(M)$, we have

$$\begin{aligned} \bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}_R(\infty, \text{SNR}) &\leq 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{(\frac{1}{\text{SNR}} + 2)^2} \\ &\times \left(D \frac{\ln M\lambda(1-\varepsilon_2)}{M\lambda(1-\varepsilon_2)} + O\left(\frac{\ln \ln M}{M}\right) \right). \end{aligned} \quad (85)$$

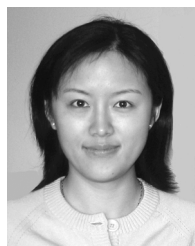
Combining the lower and upper bounds in (76) and (84), we have, for large M and $\bar{N} > \bar{N}_o(M)$

$$\begin{aligned} \bar{\mathcal{E}}_R(M, \text{SNR}) - \bar{\mathcal{E}}_R(\infty, \text{SNR}) &= 2f\sigma_s^2 \frac{\frac{1}{\text{SNR}} + 1}{(\frac{1}{\text{SNR}} + 2)^2} \\ &\times \left(D \frac{\ln M\lambda(1+o(1))}{M\lambda(1+o(1))} + O\left(\frac{\ln \ln M}{M}\right) \right). \end{aligned} \quad (86)$$

With (65), the asymptotic expression in (32) follows. \blacksquare

REFERENCES

- [1] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," presented at the 2003 Military Communications Int. Symp., Boston, MA, Oct. 2003.
- [2] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, pp. 10–21, Jan. 1949.
- [3] E. Masry, "Polynomial interpolation and prediction of continuous-time processes from random samples," *IEEE Trans. Inf. Theory*, vol. 43, pp. 776–783, Mar. 1997.
- [4] I. Bilinskis and A. Mikelsons, *Randomized Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [5] J. Jacod, "Random sampling in estimation problems for continuous Gaussian processes with independent increments," *Stochastic Process. Appl.*, pp. 181–204, 1993.
- [6] O. A. Z. Leneman and J. Lewis, "Random sampling of random processes: Mean-square comparison of various interpolators," *IEEE Trans. Autom. Control*, vol. 11, pp. 396–403, 1966.
- [7] W. M. Brown, "Sampling with random jitter," *J. SIAM*, vol. 11, pp. 460–473, 1961.
- [8] D. Klamer and E. Masry, "Polynomial interpolation of randomly sampled bandlimited functions and processes," *SIAM J. Appl. Math.*, vol. 42, pp. 1004–1019, 1982.
- [9] M. Dong, L. Tong, and B. 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.
- [10] M. Schilling, "The longest run of heads," *College Math. J.*, vol. 21, no. 3, pp. 196–207, 1990.
- [11] L. Gordon, M. Schilling, and M. Waterman, "An extreme value theory for long head runs," *Prob. Theory Related Fields*, vol. 72, pp. 279–287, 1986.
- [12] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocols for wireless sensor networks," in *Proc. IEEE INFOCOM*, New York, Jun. 2002, pp. 1567–1576.
- [13] K. Sohrabi, J. Gao, V. Ailawadhi, and G. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 16–27, Oct. 2000.
- [14] R. Iyer and L. Kleinrock, "QoS control for sensor networks," in *Proc. Int. Conf. Communications (ICC)*, May 2003, vol. 1, pp. 517–521.
- [15] P. Venkatasubramanian, S. Adireddy, and L. Tong, "Sensor networks with mobile agents: Optimal random access and coding," *IEEE J. Sel. Areas Commun. (Special Issue on Sensor Networks)*, vol. 22, no. 6, pp. 1058–1068, Aug. 2004.
- [16] Q. Zhao and L. Tong, "Quality-of-service specific information retrieval for densely deployed sensor network," presented at the 2003 Military Communications Int. Symp., Boston, MA, Oct. 2003.
- [17] A. V. Balakrishnan, "A note on the sampling principle for continuous signals," *IEEE Trans. Inf. Theory*, vol. 3, pp. 143–146, 1957.
- [18] O. A. Z. Leneman, "Random sampling of random processes: Optimum linear interpolation," *J. Franklin Inst.*, vol. 281, pp. 302–314, 1966.
- [19] M. Micheli and M. I. Jordan, "Random sampling of a continuous-time stochastic dynamical system," in *Proc. 15th Int. Symp. Mathematical Theory Networks Systems (MTNS)*, Univ. of Notre Dame, South Bend, IN, Aug. 2002.
- [20] B. Liu and T. Stanley, "Error bounds for jittered sampling," *IEEE Trans. Autom. Control*, vol. 10, no. 4, pp. 449–454, Oct. 1965.
- [21] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A 'bit-conservation principle'," in *Information Processing in Sensor Networks, Proc. 2nd Int. Workshop*, Palo Alto, CA, Apr. 22–23, 2004, pp. 17–31.
- [22] Z. Yang, M. Dong, L. Tong, and B. M. Sadler, "MAC protocols for optimal information retrieval pattern in sensor networks with mobile access," *EURASIP J. Wireless Commun. Netw.*, no. 4, pp. 493–504, 2005.
- [23] D. Cox and H. Miller, *The Theory of Stochastic Processes*. New York: Wiley, 1965.
- [24] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low cost outdoor localization for very small devices," *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 28–34, Oct. 2000.
- [25] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proc. 9th Annu. Int. Conf. Mobile Computing Networking*, San Diego, CA, 2003, pp. 81–95.
- [26] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003.
- [27] S. Resnick, *Adventures in Stochastic Processes*. Boston, MA: Birkhäuser, 1994.
- [28] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Melbourne, FL: Krieger, 1987.
- [29] S. Resnick, *A Probability Path*. Boston, MA: Birkhäuser, 2001.



Min Dong (S'00–M'05) received the B.Eng. degree from Tsinghua University, Beijing, China, in 1998 and the Ph.D. degree in electrical engineering with a minor in applied mathematics from Cornell University, Ithaca, NY, in 2004.

She is currently with Corporate Research and Development, Qualcomm, Inc., San Diego, CA, where she has been actively involved in the orthogonal frequency-division multiple-access (OFDMA) system design and performance analysis. Her research interests are in the general areas of communications,

signal processing, and mobile networks.

Dr. Dong received the 2004 IEEE Signal Processing Society Best Paper Award.



Lang Tong (S'87–M'91–SM'01–F'05) received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1987 and 1991, respectively.

He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, Stanford, CA, in 1991. Prior to joining Cornell University, he was on faculty at West Virginia University and the University of Connecticut. He was also the

2001 Cor Wit Visiting Professor at the Delft University of Technology, Delft, The Netherlands. He is currently the Irwin and Joan Jacobs Professor in Engineering at Cornell University, Ithaca, NY. His research is in the general area of statistical signal processing, wireless communications and networking, and information theory.

Dr. Tong received the 1993 Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 Best Paper Award (with M. Dong) from the IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society (with P. Venkatasubramanian and S. Adireddy). He is also a coauthor of five student paper awards. He received the Young Investigator Award from the Office of Naval Research. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION THEORY, and IEEE SIGNAL PROCESSING LETTERS.



Brian M. Sadler (M'90–SM'02–F'06) received the B.S. and M.S. degrees from the University of Maryland, College Park, and the Ph.D. degree from the University of Virginia, Charlottesville, all in electrical engineering.

He is a Senior Research Scientist at the Army Research Laboratory (ARL), Adelphi, MD. He was a Lecturer at the University of Maryland and has been lecturing at The Johns Hopkins University, Baltimore, MD, since 1994 on statistical signal processing and communications. His research interests

include signal processing for mobile wireless and ultra-wideband systems, sensor signal processing and networking, and associated security issues.

Dr. Sadler is currently an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON SIGNAL PROCESSING and has been a Guest Editor for several journals, including IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and the IEEE Signal Processing Magazine. He is a member of the IEEE Signal Processing Society Sensor Array and Multi-Channel Technical Committee, and received a Best Paper Award (with R. Kozick) from the Signal Processing Society in 2006.