Type-Based Decentralized Detection in Wireless Sensor Networks

Ke Liu, *Member, IEEE*, and Akbar M. Sayeed, *Senior Member, IEEE*

***Abstract*—Distributed detection strategies for wireless sensor networks are studied under the assumption of spatially and temporally independent and identically distributed (i.i.d.) observations at the sensor nodes. Both intelligent (with knowledge of observation statistics) and dumb (oblivious of observation statistics) sensors are considered. Two types of communication channels are studied: a parallel access channel (PAC) in which each sensor has a dedicated additive white Gaussian noise (AWGN) channel to a decision center, and a multiple-access channel (MAC) in which the decision center receives a coherent superposition of the sensor transmissions. Our results show that the MAC yields significantly superior detection performance for any network power constraint. For intelligent sensors, uncoded (finite duration) communication of local log-likelihood ratios over the MAC achieves the optimal error exponent of the centralized (noise-free channel) benchmark as the number of nodes increases, even with sublinear network power scaling. Motivated by this result, we propose a distributed detection strategy for dumb sensors—histogram fusion—in which each node appropriately quantizes its temporal data and communicates its type or histogram to the decision center. It is shown that uncoded histogram fusion over the MAC is also asymptotically optimal under sublinear network power scaling with an additional advantage: knowledge of observation statistics is needed only at the decision center. Histogram fusion achieves exponential decay in error probability with the number of nodes even under a finite total network power. In principle, a vanishing error probability at a slower subexponential rate can be attained even with vanishing total network power in the limit. These remarkable power/energy savings with the number of nodes are due to the inherent beamforming gain in the MAC.**

***Index Terms*—Distributed beamforming, distributed detection, energy efficiency, error exponents, low latency, types.**

I. Introduction

DETECTION of events of interest constitutes an important application of wireless sensor networks in which spatially distributed nodes communicate to a decision center. This scenario is referred to as *decentralized or distributed detection* due to the spatially distributed nature of sensor measurements and processing involved. Noisy communication between the sensors and the decision center represents a major bottleneck limiting the performance of decentralized detection. The idealized scenario corresponding to noise-free communication is called *centralized detection* and provides a performance benchmark for any decentralized scheme. A key motivation of this work is to investigate decentralized detection strategies that approach the performance of the centralized benchmark.

To simplify discussion, we assume that sensor measurement data are independent and identically distributed (i.i.d.) across different nodes and time slots (see [1] for a physical discussion of this assumption and the associated sensor sampling requirements). The network power scaling with the number of nodes has a crucial impact on performance. We consider two important special cases: 1) individual power constraint (IPC), where each node has a constant power budget so that the total network power increases linearly with the number of nodes, and 2) total power constraint (TPC), where the total power is limited regardless of the network size, and hence the power per node diminishes with increasing number of nodes. We consider two different additive white Gaussian noise (AWGN) channels: 1) a parallel-access channel (PAC) consisting of dedicated, noninterfering AWGN links from every node to the decision center and 2) a multiple-access channel (MAC), where all the nodes share the same AWGN channel.

Many existing works assume *intelligent sensors* that have detailed knowledge of the source statistics (see, e.g., [1]–[3]). For instance, it is shown in [1] that likelihood-based fusion schemes over a PAC generally attain exponential decay in detection error probability with the number of nodes under the IPC. In this paper, we first sharpen the results of [1] and show that uncoded communication of local log-likelihood ratios (LLR) over the MAC is asymptotically optimal with *sublinear* network power scaling (between the TPC and IPC); that is, it achieves the centralized error exponent in the limit of large number of nodes. This result is a manifestation of *source-channel matching* between the MAC and the single source problem in the context of detection where uncoded transmission is the optimal communication strategy; such source-channel matching was first reported in [4] and [5] for distributed estimation of a Gaussian source.

However, sensor processing in likelihood-based fusion methods varies with the source statistics so that networkwide sensor update is required for different detection tasks. This motivates us to consider detection schemes for "dumb sensors" that are oblivious of source statistics. The main focus of this paper is a distributed detection strategy based on the method of types [6], [7]. The type of a data sequence is the histogram or empirical distribution of the sequence. In the proposed type-based decentralized detection framework, sensor nodes

Manuscript received October 11, 2005; revised June 27, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marcelo G. S. Bruno. This work was supported in part by the NSF under Grants CCR-0113385 and CCF-0431088, and by ONR under Grant N00014-01-1-0825. The results in this paper were presented in part at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, QC, Canada, May 2004.

K. Liu is with the Qualcomm, Inc., San Diego, CA 92121 USA (e-mail: kel@qualcomm.com).

A. M. Sayeed is with the Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706-1691 USA (e-mail: akbar@engr.wisc.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.2006.890821

are both "simple" and "dumb": they record the type statistics of data measurements, which can be efficiently computed via simple counters, and do not require detailed knowledge of source statistics. Knowledge of source statistics is only required at the decision center. In particular, we propose a low-complexity spatial fusion scheme, *histogram fusion*, in which sensors transmit their local type/histogram statistics in an uncoded analog fashion, which are naturally fused during transmission over the MAC to yield the global type statistic at the decision center. In particular, uncoded transmission of type statistics over the MAC is also asymptotically optimal with sublinear network power scaling. We note that for uncoded (finite duration) transmissions, power scaling with the number of nodes is equivalent to network *energy* scaling and the remarkable asymptotic optimality of LLR and type fusion with *sublinear* energy scaling is due to the coherent beamforming gain inherent to the MAC for uncoded transmissions.

The method of types has been explored in several recent works on sensor networks [8]–[10]. In particular, a type-based estimation/detection framework, type-based multiple access (TBMA), has been independently proposed by Mergen and Tong for multiple-access channels [8], [9]. The work [9] characterizes the detection error probability of TBMA over a fading MAC using tools from large deviation theory. The are two main differences between our work and [9]. First, we focus on the critical impact of power scaling and study both the temporal and spatial asymptotics of type-based detection. In particular, we show that type-based dumb sensors can achieve exponential decay in detection error probability with the number of nodes even under the stringent total power constraint. Our analysis also suggests that one can drive error probability to zero even with a diminishing amount of total power, albeit at a slower subexponential rate. Second, histogram fusion proposed in this work presents a different interpretation and analysis of type-based detectors than the ones derived from the large deviations principle with rate function in [9]. In our context, for discrete-valued (quantized) observations, histogram fusion can be viewed as an efficient realization of LLR fusion using dumb sensors and closely approximates its performance. Rather than resorting to specialized results from large deviations theory, we exploit the bounded nature of type statistics using the Hoeffding's inequality to derive informative, closed-form performance bounds, and use Chernoff bounding techniques for proving optimality. Although the technique in [9] applies in principle to non-i.i.d. data or nonidentical fading in MAC branches (which is not considered in this paper), the scaling results obtained in [9] are limited to i.i.d. data and identical fading due to mathematical tractability.

The rest of paper is organized as follows. The decentralized detection problem is formulated in the next section where we show the optimality of LLR fusion over the MAC. Section III develops type-based decentralized detection in detail. The spatial asymptotics are studied in Section IV with emphasis on analytical performance bounds and the optimality of histogram fusion over the MAC. In Section V, numerical results are presented to illustrate the theoretical results and to compare the performance of histogram fusion with other fusion methods. Concluding remarks are provided in Section VI.

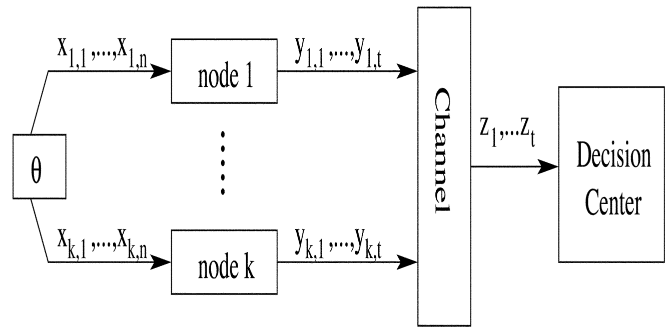

Fig. 1. Schematic illustrating decentralized detection using k distributed sensor nodes.

II. Decentralized Detection

A. System Model

Fig. 1 illustrates a "one-shot" model of the decentralized detection problem, where θ denotes the unknown source to be detected. Conditioned on a given state of the source, the underlining homogeneous sensor field generates spatio–temporal i.i.d. observation data $x_{i,j}$ according to a particular hypothesis distribution. For simplicity of exposition, we consider *binary* hypothesis testing, that is, $\theta \in \{\theta_0, \theta_1\}$, with the corresponding hypothesis distributions Q_0 and Q_1. For dumb sensors (which is the main focus of this paper), we assume that $x_{i,j}$ are from a *finite* alphabet (appropriate quantization is assumed for continuous-valued sensor data), whereas this assumption is not required for intelligent sensors. For both type of sensors, we assume that Q_0 and Q_1 have the same support.[1] For intelligent sensors, in addition to identical support, we also assume that the likelihood ratio $Q_1(x)/Q_0(x)$ has a finite second moment under both hypotheses.

The *local* observation data $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n})^T$ are encoded into the transmit sequence $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,t})^T$ by the ith sensor node $(1 \le i \le k)$. The temporal and spatial dimensions of measurement data are denoted by n and k, respectively, in Fig. 1. The parameter t in the transmit sequence specifies the total number of channel uses for communicating information about one fixed value of θ contained in n temporal measurements. A decision center makes decision $\hat{\theta}$ based on signals received from the sensors.

The encoding of observation data into transmit sequences is a form of *joint source–channel communication* adapted to the decentralized detection problem. Although the formulation (in Fig. 1) permits general source–channel codes, in this paper, we primarily focus on *identical coding* where the same encoder is applied at every sensor node. It has been shown that, under the assumption of i.i.d. observations, this scheme is asymptotically optimal in the limit of large number of sensors [2], [3].

The performance of decentralized detection is characterized by the *detection error probability* (DEP)

$$P_e = \frac{1}{2}(\alpha + \beta) = \frac{1}{2}[P(\hat{\theta} = \theta_1 \,|\, \theta_0) + P(\hat{\theta} = \theta_0 \,|\, \theta_1)]. \quad (1)$$

[1]This is to exclude unusual scenarios in which perfect detection is possible (likelihood ratio takes on an infinite value for measurements that can occur with a nonzero probability). In practice, the presence of measurement noise, e.g., Gaussian noise, would satisfy this assumption.

In particular, we are interested in characterizing the *decentralized error exponent* defined as

$$E_d = \lim_{nk\to\infty} -\frac{1}{kn}\log P_e(n,k) \tag{2}$$

where we have emphasized the dependence of DEP on both the spatial and temporal measurement dimensions. The noisy communication channel from the sensors to the decision center directly impacts the DEP and E_d and represents a key challenge in decentralized detection. We will characterize the achievable values of E_d for both temporal and spatial asymptotics (either n or k goes to infinity) as a function of the nature of the communication channel, the number of channel uses t, and the amount of power consumed by the network for communication.

An ideal scenario in which the decision center has perfect (noise-free) access to sensor measurements is commonly referred to as *centralized detection* and represents a performance benchmark for any decentralized detection scheme. The corresponding *centralized error exponent* is given by the following well-known result in classical statistical decision theory [7], [11], [12].

Lemma 1 (Chernoff):

$$E_c = \limsup_{kn\to\infty} -\frac{1}{kn}\log P_e = CI(Q_0,Q_1) \tag{3}$$

where $CI(Q_0,Q_1)\geq 0$ is called the *Chernoff information* between Q_0 and Q_1 and is defined as

$$CI(Q_0,Q_1) = -\min_{0\leq s\leq 1}\log \mathbb{E}_0\left[\frac{Q_1(X)}{Q_0(X)}\right]^s \tag{4}$$

where $\mathbb{E}_0$ is with respect to $X\sim Q_0$.

Clearly, $E_d \leq E_c$ due to nonideal (noisy) communication between the sensors and the fusion center. An overarching aim of this paper is to investigate decentralized detection schemes that can approach the performance of the centralized benchmark and the associated network power consumption requirements.

We consider two representative AWGN channel models: 1) a PAC

$$z_{i,\ell} = y_{i,\ell} + w_{i,\ell}, \quad i=1,\ldots,k; \ell=1,\ldots,t \tag{5}$$

where $w_{i,\ell}$ denotes AWGN and every sensor node has a dedicated, non-interfering parallel AWGN channel to the decision center (as in frequency division multiple access), and 2) an AWGN MAC

$$z_\ell = \sum_{i=1}^{k} y_{i,\ell} + w_\ell, \quad \ell = 1,\ldots,t \tag{6}$$

which exploits the inherent superposition in the wireless medium.

Let P_i denote the (average) power budget at the ith node (under both hypotheses)

$$\mathbb{E}_{0/1}|y_{i,j}|^2 \leq P_i. \tag{7}$$

Note that the power constraint is equivalent to an energy constraint when spatial asymptotics are considered, i.e., n and t are finite and $k\to\infty$, or when the encoder requires only finite number of channel uses. The total network power $P_{\rm tot}(k) = \sum_{i=1}^{k} P_i$ is viewed as a function of k to model network power scaling with the number of sensors. Our analysis of network power scaling is anchored around two reference power constraints: 1) individual power constraint (IPC) where each node is given a constant power budget $P_{\rm ind}$ while the total network power, $P_{\rm tot} = kP_{\rm ind}$, scales linearly with the number of nodes, and 2) total power constraint (TPC) where the total network power remains constant and the power per node, $P_i = P_{\rm tot}/k$, diminishes as k increases.

B. Distributed Detection With Intelligent Sensors

A majority of existing works (see, e.g., [1]–[3]) assume the use of "intelligent" sensors that are aware of source distributions $\{Q_0,Q_1\}$. In this case, a natural source compression strategy at each sensor is to compute the local (normalized) LLR $\gamma_i = (1/n)\log(Q_1(\mathbf{x}_i)/Q_0(\mathbf{x}_i))$ [1], [11]. This is because the global LLR (associated with all observation data) in the benchmark centralized detector is simply an average of the local LLRs

$$\gamma = \frac{1}{nk}\log\frac{Q_1(\mathbf{X})}{Q_0(\mathbf{X})} = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{n}\log\frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} = \frac{1}{k}\sum_{i=1}^{k}\gamma_i \tag{8}$$

which is sometimes referred to as (ideal) soft-decision fusion [1]. While the computation of a single LLR value γ_i at each sensor achieves substantial compression of the n-dimensional temporal observation data $\mathbf{x}_i$, it is a continuous-valued random variable for continuous-valued observation data and thus still contains an infinite amount of information. Thus, for digital communication over noisy communication links, the local LLR needs to be quantized and coded to protect against channel error [1]. In particular, one-bit quantization $\mathrm{sign}(\gamma_i)$ corresponds to *hard-decision fusion*, where the decision center fuses local hard decisions made by sensor nodes. Although the source is minimally compressed (only 1 bit to be sent in a single channel use), this approach is suboptimal even in the absence of channel noise [11], [13], [1]. Nevertheless, such an approach provides a low-complexity solution for practical sensor network applications. In particular, the study in [1] investigated several LLR-based schemes for target detection and classification under the assumption of an AWGN PAC. The results showed exponential decay in the DEP with the number of sensors when each sensor transmits with a nonvanishing power (IPC). However, the associated error exponents always incur a loss compared to E_c, which depends on the nature of LLR quantization and the transmission SNR.

The first result in this paper improves upon the work in [1] by showing that analog (uncoded and unquantized) transmission of the local LLRs is asymptotically optimal over the MAC. The received signal in this case is given by $(t=1)$

$$z = \sum_{i=1}^{k}\rho\gamma_i + w \tag{9}$$

where w is the channel AWGN and ρ is the transmission amplification constant chosen according to the power constraints: $\rho^2\mathbb{E}_{0/1}[\gamma_i^2]\leq P_{\rm ind}$ (IPC) and $\rho^2\mathbb{E}_{0/1}[\gamma_i^2]\leq P_{\rm tot}/k$ (TPC). The decision center forms the global LLR estimate as $\tilde{\gamma} = z/\rho k$ and proceeds with the LLR test ($\hat{\theta}=\theta_1$ if $\tilde{\gamma}>0$; and θ_0 otherwise).

For this result, we do not impose the assumption of finite-alphabet observation data. The assumption $\mathbb{E}|\gamma_i|^2 < \infty$ under both hypotheses is sufficient to satisfy the *average* transmission power constraints.[2]

Theorem 1: Under the individual power constraint, uncoded LLR (soft-decision) fusion over the MAC achieves the centralized error exponent in the limit of large number of sensor measurements.

Proof: See Appendix I. ■

Remark 1: The proof [(43) in Appendix I] actually shows that a sufficient condition for the asymptotic optimality $(E_d = E_c)$ of uncoded LLR fusion is that the total power grows unbounded, $P_{\text{tot}}(k) \sim k\rho^2 \to \infty$. This condition is trivially satisfied by IPC where $P_{\text{tot}}(k) \sim k$, but it can also be satisfied by a less stringent *sublinear* scaling in total power, for example, $P_{\text{tot}}(k) \sim k^{\epsilon}$ for $0 < \epsilon < 1$. An important consequence of sublinear total power scaling is that the individual power requirement at each sensor node becomes arbitrarily small $(\rho^2 \sim k^{\epsilon-1})$ as the number of nodes increases. Thus, uncoded LLR fusion is asymptotically optimal even with vanishing transmission power per sensor! This is due to the k-fold beamforming (power amplification) gain inherent in the MAC (see also [4] and [14]). We will revisit this issue later for histogram fusion.

The proof of Theorem 1 applies to continuous- or discrete-valued measurements and thus uncoded LLR (soft-decision) fusion over the MAC is asymptotically optimal in either case. However, in order to calculate the local LLRs, sensor nodes must have *a priori* knowledge of hypothesis statistics, $\{Q_0, Q_1\}$. Thus, programming a network of such intelligent sensors for different detection tasks (different $\{Q_0, Q_1\}$) would require a networkwide update which may be too costly or not feasible. Flexibility of network design thus motivates the important question: *Can optimal performance be achieved by a network of "dumb" sensors that perform local processing without knowledge of measurement statistics?* The answer is in the affirmative as we show next in our development of type-based decentralized detection.

III. Type-Based Dumb Sensors for Decentralized Detection

In this section, we develop a decentralized detection framework based on the method of types. Sensor nodes compute and transmit type/histogram statistics of the (finite alphabet) observation data and are "simple" and "dumb": computation of the type statistics amounts to a simple counting procedure, and the nodes do not require knowledge of measurement statistics. Measurement statistics are only needed at the decision center and thus different detection tasks can be performed with same type-based encoding at the sensors.

[2]Since the value of γ_i could be arbitrarily large for finite length (n) continuous-valued measurements $\mathbf{x}_i$, in practice ρ may be chosen so that $P(\rho^2\gamma_i^2 > P_{\text{ins}}) < \epsilon$ to guarantee that the instantaneous power constraint P_{ins} is exceeded with a sufficiently small probability ϵ, in which case the sensor does not transmit. With this modification, the total energy in each transmission is always bounded by $P_{\text{ins}}T(1-\epsilon)$, where T is the duration of the analog transmission in each channel use. Note that longer i.i.d. temporal observations (n) would reduce such fluctuations in γ_i.

In essence, type-based detection can be viewed as a low-cost and efficient implementation of LLR fusion via types (for finite alphabet observations) and hence enjoys its remarkable detection performance. To see this, first recall the (global) LLR computation in the benchmark centralized detector, $\gamma = \sum_{i=1}^{k} \gamma_i$, where $\gamma_i = (1/n)\log(Q_1(\mathbf{x}_i))/(Q_0(\mathbf{x}_i)) = (1/n)\sum_{j=1}^{n} f(x_{i,j})$ is the local LLR of the temporal measurements at ith node. Since $x_{i,j} \in \{1,\ldots,A\}$ is discrete, so is $f(x_{i,j}) = \log(Q_1(x_{i,j}))/(Q_0(x_{i,j})) \in \{f(1),\ldots,f(A)\}$ and we have

$$\begin{aligned}\gamma = \frac{1}{k}\sum_{i=1}^{k}\gamma_i &= \frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{n}\sum_{a=1}^{A} T_i(a)f(a)\right]\\ &= \frac{1}{k}\sum_{a=1}^{A}\left[\sum_{i=1}^{k}\frac{T_i(a)}{n}\right]f(a) \qquad (10)\end{aligned}$$

where $T_i(a)$ is the local number of "a"-measurements at the ith node and $T(a) = \sum_i T_i(a)$ is the global number of "a" measurements at all k nodes. Therefore, instead of sending its local LLR γ_i, the ith sensor can send its (normalized) local "count" values $T_i(a)/n$ $(a \in \{1,\ldots,A\})$, and the fusion center can compute the global LLR γ as in (10). Clearly this has the advantage that the measurement statistics $\{f(a)\}$ are only needed at the fusion center.

Simply put, the count vector $T_i(a)$ above is the so-called "type," which is the histogram of measurement data at node i. The method of types refers to a special set of tools dealing with type statistics (see [6], [7], and [15] for more information). We next provide a brief introduction to types that is relevant to this paper.

A. Method of Types

Let $\mathcal{A}$ be a size-A discrete alphabet with symbols written as $\{a_1,\ldots,a_A\}$. A length-n sequence drawn from $\mathcal{A}$ is denoted by $x_1,\ldots,x_n$ or $\mathbf{x}$.

Definition 1: The type $T_{\mathbf{x}}$ (or empirical probability distribution) of a sequence $x_1,\ldots,x_n$ is the relative frequency of each alphabet symbol in $\mathcal{A}$

$$T_{\mathbf{x}}(a) = N(a|\mathbf{x})/n, \quad \forall a \in \mathcal{A} \qquad (11)$$

where $N(a|\mathbf{x})$ denotes the number of occurrences of the symbol a in the sequence $\mathbf{x} \in \mathcal{A}^n$.

Different sequences may have the same type. For instance, sequences (0,1,1) and (1,1,0) have the same type (1/3,2/3) for having one "0" and two "1"s in the sequence. The type statistic measures the empirical distribution or the histogram of the observation data. It is a combinatorial property of the sequence with respect to the symbol alphabet $\mathcal{A}$ and should not be confused with the probability distribution of the sequence: for a random sequence $\mathbf{x}$ drawn from a distribution Q, its realizations can exhibit all the possible types $T_{\mathbf{x}}$. We sometimes write $T(\mathbf{x})$ to emphasize type as a statistic derived from the data.

Given $x_1,\ldots,x_n$ drawn i.i.d. according to distribution Q on $\mathcal{A}$, its probability depends only on the type

$$Q(\mathbf{x}) = \prod_{a\in\mathcal{A}} Q(a)^{nT_{\mathbf{x}}(a)} \qquad (12)$$

that is, the type is a *sufficient statistic* of the data. Then, we can write the *log-likelihood* in terms of the type

$$\frac{1}{n}\log Q(\mathbf{x}) = \sum_{a\in\mathcal{A}} T_{\mathbf{x}}(a)\log Q(a) \qquad (13)$$

or, in a compact form

$$\frac{1}{n}\log Q(\mathbf{x}) = [\log \mathbf{Q}]^T \mathbf{T}(\mathbf{x}) \qquad (14)$$

where $\log \mathbf{Q} = [\log Q(a_1), \ldots, \log Q(a_A)]^T$ and $\mathbf{T}(\mathbf{x}) = [T_{\mathbf{x}}(a_1), \ldots, T_{\mathbf{x}}(a_A)]^T$. Similarly, the LLR between Q_0 and Q_1 is given by

$$\gamma = \log\frac{Q_1(\mathbf{x})}{Q_0(\mathbf{x})} = \left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \mathbf{T}(\mathbf{x}) \qquad (15)$$

where

$$\log(\mathbf{Q}_1)/(\mathbf{Q}_0) = [\log(Q_1(a_1)/Q_0(a_1)), \ldots, \log(Q_1(a_A)/Q_0(a_A))]^T.$$

The relations (14) and (15) show that log-likelihood statistics for discrete-valued data can be computed as a *linear* function of its type. Thus, likelihood-based algorithms can be implemented via the type statistics.

From the viewpoint of compression of local temporal data at each sensor, the type statistic achieves asymptotically *zero-rate* compression as $n \to \infty$. The following result provides an upper bound on the entropy rate of type statistics (see [7, Theorem 12.1.1]).

Proposition 1: Fix sequence length n and alphabet size A. The total number of different types is no greater than $(n+1)^A$.

It follows that the entropy rate of type statistics $H(T(\mathbf{x}))/n \le A\log(n+1)/n \to 0$ as $n \to \infty$. Thus, in the absence of latency constraints, this zero-rate property of the type statistic imposes a *vanishing* communication burden as temporal dimension n increases. As long as the communication channel has a positive capacity, then the temporal type statistics, being of zero rate, can be digitally communicated to the decision center with arbitrarily high reliability through appropriate channel coding at each sensor node (see Fig. 1). Consequently, from the viewpoint of temporal asymptotics (k is fixed but n and $t \to \infty$), type-based detection can achieve the performance of the benchmark centralized detector, as shown in Theorem 2 next, which can be derived alternatively using the framework of *zero-rate, multiterminal* hypothesis testing in [13]. However, it is worth noting that temporal optimality imposes a fairly stringent requirement that the underlying value of the source θ does not change for a sufficiently long period of time.

Theorem 2 (Temporal Optimality of Type-Based Detection): Let $C > 0$ be the link capacity from each sensor node to the decision center. Assume a fixed number of sensor nodes and a constant scaling of channel uses t with sequence length, i.e., $t \sim \mathcal{O}(n)$. Then

$$E_d = \lim_{n\to\infty} -\frac{1}{nk}\log P_e(n) = E_c. \qquad (16)$$

Proof: See Appendix II.

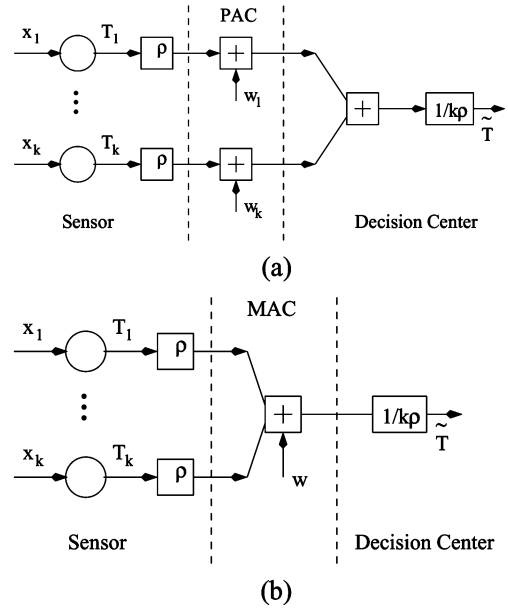

Fig. 2. Schematic illustrating histogram fusion over: (a) the PAC and (b) the MAC.

Remark 2: The above result involves digitally encoded transmission of local temporal type statistics $\{\mathbf{T}(\mathbf{x}_i)\}$ at each node in Fig. 1 and applies to the PAC or the MAC. If R denotes the transmission rate and ρ^2 the transmission power at each sensor, for the PAC we require $R < C_{\rm pac} = \log(1+\rho^2)$ where $C_{\rm pac}$ is the capacity of each parallel channel and for the MAC we require $kR < C_{\rm mac} = \log(1+k\rho^2)$, where $C_{\rm mac}$ is the sum capacity of the MAC [7]. Due to the zero-rate property of type statistics ($R \to 0$ as $n \to \infty$), the condition $C > 0$ is sufficient for both channels and is satisfied with arbitrarily small but positive transmission power ρ^2.

B. Histogram Fusion Across Sensors

We now describe a decentralized detection scheme, histogram fusion, that exploits spatial fusion of type (histogram) statistics across sensors for asymptotically ($k \to \infty$) optimal performance. Histogram fusion can be viewed as a realization of LLR (soft-decision) fusion that is tailored to the use of dumb sensors. While LLR fusion applies to continuous-valued data as well, histogram fusion requires quantization of sensor observations so that the type framework is applicable. Histogram fusion architecture is illustrated in Fig. 2 for both the PAC and the MAC. Two important characteristics of histogram fusion are 1) the local temporal type statistics at each sensor are transmitted in an analog uncoded fashion, and 2) a finite number of channel uses are required, regardless of observation length. Thus, this scheme is particularly attractive in low-latency applications or when limited temporal observation data is available at each node due to changes in the source value θ over time. Furthermore, for the MAC, the scheme achieves dramatic reductions in power (energy) consumption due to the beamforming (power amplification) gain inherent in uncoded transmission over the MAC (see also [4] and [14]).

We begin by describing the encoding and transmission at the sensor nodes. The ith sensor node computes its local type statistic $\mathbf{T}_i = [T_{\mathbf{x}_i}(a_1), \ldots, T_{\mathbf{x}_i}(a_A)]^T$ from its length-n observation sequence $\mathbf{x}_i$. In other words, each sensor counts the occurrences of alphabet symbols within its observation duration. Then, sensor nodes send these relative frequencies ($T_{\mathbf{x}_i}(a) \in [0,1]$) over the channel in an uncoded (analog) fashion via amplitude modulation. The number of channel uses needed for transmission of local type statistics equals $t = A$, the size of the (quantized) observation alphabet. In the special case of a single observation ($n = 1$), histogram fusion reduces to the pulse position modulation (PPM) scheme in [8]. As the length of the observation data increases, the higher amount of local information is encoded in the higher dynamic range (between $1/n$ and 1) of the relative frequency of each observation symbol and is preserved in the amplitude of the corresponding analog transmission.

Due to source quantization, histogram fusion requires $t = A$ channel uses compared to LLR fusion that requires a single $t = 1$ channel use. Thus, histogram fusion incurs A-times longer latency or A-times larger bandwidth consumption compared to analog LLR (soft-decision) fusion. This is the price for the flexibility afforded by dumb sensors in histogram fusion. We note the above comparison applies to analog (uncoded) transmission in both LLR and histogram fusion, which is critical for dramatic reductions in power requirements (over the MAC) in both cases due to the distributed beamforming gain. For a fair comparison, we adopt a power normalization where histogram fusion (A channel uses) and LLR fusion (one channel use) consume the same amount of power. This can be accomplished by distributing power budget equally among the A channel uses in histogram fusion. Thus, the signal amplification ρ in histogram fusion is modified as

$$\rho^2 = \begin{cases} P_{\text{ind}}/A, & \text{IPC} \\ P_{\text{tot}}/kA, & \text{TPC} \end{cases}. \tag{17}$$

We note that transmit sequences $\{y_{i,j}\}$ are bounded and also belong to a finite alphabet in histogram fusion and as a result the average power constraint in (7) also implies a stronger instantaneous constraint (hard limit) on transmit power: $|y_{i,j}|^2 \leq P_{i,\text{ins}}$.

It is straightforward to see that the global type statistic $\mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is the spatial average of local type statistics

$$\mathbf{T} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{T}_i. \tag{18}$$

Inspired by (18), histogram fusion performs a similar averaging at the decision center to obtain an estimate of the global type statistic. For the PAC, it is given by

$$\tilde{\mathbf{T}} = \frac{1}{k}\sum_{i=1}^{k} [\mathbf{T}_i + \tilde{\mathbf{w}}_i] \tag{19}$$

where $\tilde{\mathbf{w}}_i$ is a length-A noise vector associated with the ith spatial AWGN channel. The noise elements in $\tilde{\mathbf{w}}_i$ have variance $1/\rho^2$, where ρ is the amplification constant defined in (17). For the MAC, histogram fusion exploits the inherent (coherent) spatial averaging which significantly reduces the effective noise level compared to the PAC

$$\tilde{\mathbf{T}} = \sum_{i=1}^{k} \mathbf{T}_i + \frac{1}{k}\tilde{\mathbf{w}} \tag{20}$$

where each noise element in $\tilde{\mathbf{w}}$ has variance $1/\rho^2$.

Finally, the decision center computes an estimate of the global LLR from the global type estimate $\tilde{\mathbf{T}}$ via the relation (15)

$$\tilde{\gamma} = \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \tilde{\mathbf{T}}. \tag{21}$$

The decision rule is given by $\hat{\theta} = \theta_1$ if $\tilde{\gamma} \geq 0$ and θ_0 otherwise.

The relation (21) captures the essence of the proposed type-based distributed detection framework. The sensor nodes do not need knowledge of the observation statistics (Q_0 and Q_1): they simply compute the type/histogram of their local measurement data in a "dumb" fashion. The decision center forms an estimate of the global type via histogram fusion, which is then used to compute an estimate of the global LLR via (21). It is worth noting the impact of channel noise in LLR (soft-decision) fusion versus histogram fusion. In soft-decision fusion the type-to-LLR conversion is done locally at each sensor and the noise is added in the global fusion of local LLRs. In histogram fusion, noise is added during the formation of the global type estimate, which is then converted to a global LLR at the fusion center. Thus, while histogram fusion is equivalent to LLR fusion in the absence of noise [and finite alphabet observations, as in (10)], it is not equivalent in the presence of channel noise. Nevertheless, the two fusion schemes perform nearly identically for finite alphabet data as we will see in numerical results. Thus, histogram fusion can be seen as an efficient implementation of LLR fusion with a key difference: the computation of the LLR is shifted from sensor nodes to the decision center, which can thus perform different detection tasks using the same type-encoded sensor data.

IV. Spatial Asymptotics of Type-Based Decentralized Detection

In this section, we fix the temporal dimension (n) and study the asymptotic performance of histogram fusion in the limit of large number of sensor nodes. We characterize the DEP scaling behavior under the two channel models and different power constraints. The analysis shows that histogram fusion over the MAC, due to its close connection to LLR fusion, achieves optimal centralized (noise-free) error exponent even when the total network power scales sublinearly with the number of nodes (in between IPC and TPC). In principle, a diminishing total power suffices to drive the DEP asymptotically to zero, albeit at slower subexponential rate.

A. *Performance of Histogram Fusion*

We begin by presenting an informative analysis of histogram fusion using *Hoeffding's inequality* [16] that exploits the boundedness of type statistics, i.e., $T \in [0,1]$. However, we need to extend the original Hoeffding's inequality to account for the noise term in the type estimates. The proof of the following theorem is given in Appendix III.

Theorem 3 (Generalized Hoeffding's Inequality): Let $Z_i = X_i + Y_i$ where $X_i \in [a_i, b_i]$ with probability 1 and $Y_i \sim \mathcal{N}(0, \sigma_i^2)$. Assume all X_i's and Y_i's independent and $\mathbb{E}X_i = 0$. Then, for any $d > 0$

$$P\left(\sum_{i=1}^{k} Z_i \geq d\right) \leq e^{-\frac{2d^2}{\sum_{i=1}^{k}(b_i-a_i)^2+4\sigma_i^2}}. \tag{22}$$

Consider the error probability of selecting $\hat{\theta} = \theta_1$ when θ_0 is in force. Following the notation of histogram fusion in Section III, the error event can be expressed as

$$\mathcal{E} = \{\tilde{\gamma} \geq 0\} = \left\{\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \tilde{\mathbf{T}} \geq 0\right\} \tag{23}$$

which gives the error probability $P_e = Q_0(\mathcal{E})$. For the PAC, by plugging the expression (19) of $\tilde{\mathbf{T}}$ into (23), we have

$$\mathcal{E} = \left\{\sum_{i=1}^{k}\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T (\mathbf{T}_i + \tilde{\mathbf{w}}_i) \geq 0\right\}. \tag{24}$$

It follows from $\mathbb{E}_0 \mathbf{T}_i = \mathbf{Q}_0$ that

$$\begin{aligned}\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \mathbb{E}_0 \mathbf{T}_i &= -\sum_{a\in\mathcal{A}} Q_0(a) \log \frac{Q_0(a)}{Q_1(a)} \\ &= -D(Q_0 \| Q_1)\end{aligned} \tag{25}$$

where $D(Q_0\|Q_1)$ denotes the *Kullback–Leibler* (K-L) distance between Q_0 and Q_1. Centering the random vector $\mathbf{T}$, we get

$$\mathcal{E} = \left\{\sum_{i=1}^{k}\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T ([\mathbf{T}_i - \mathbf{Q}_0] + \tilde{\mathbf{w}}_i) \geq kD(Q_0\|Q_1)\right\}. \tag{26}$$

Set $X_i = [(\log \mathbf{Q}_1)/(\mathbf{Q}_0)]^T(\mathbf{T}_i - \mathbf{Q}_0)$ and $Y_i = [(\log \mathbf{Q}_1)/(\mathbf{Q}_0)]^T \tilde{\mathbf{w}}_i$. Note that the elements $\mathbf{T}_i$ are bounded between 0 and 1. It is straightforward to verify that

$$X_i \in \left[\min_a \log \frac{Q_1(a)}{Q_0(a)}, \max_a \log \frac{Q_1(a)}{Q_0(a)}\right] + D(Q_0\|Q_1) \tag{27}$$

$$\mathbb{E}_0 Y_i^2 = \frac{1}{\rho^2} \sum_a \log^2 \frac{Q_1(a)}{Q_0(a)} \tag{28}$$

where ρ satisfies the power constraints in (17). Define

$$\lambda(Q_0, Q_1) = \left(\max_a \log \frac{Q_1(a)}{Q_0(a)} - \min_a \log \frac{Q_1(a)}{Q_0(a)}\right)^2 \tag{29}$$

$$\nu(Q_0, Q_1) = \sum_a \log^2 \frac{Q_1(a)}{Q_0(a)}. \tag{30}$$

The nonnegative λ and ν measure the discrepancy between Q_0 and Q_1 analogous to a metric. It can be shown for λ (similarly for ν) that

$$\lambda(Q_0, Q_1) = \lambda(Q_1, Q_0) \tag{31}$$

$$\lambda(Q_0, Q_1) = 0 \Longleftrightarrow Q_0 = Q_1. \tag{32}$$

Now applying Theorem 3 to (26), we have

$$\begin{aligned}P_e &= Q_0\left(\sum_{i=1}^{k}(X_i + Y_i) \geq kD(Q_0\|Q_1)\right) \\ &\leq e^{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0,Q_1)+\frac{4}{\rho^2}\nu(Q_0,Q_1)}}.\end{aligned} \tag{33}$$

The bounds on DEP for the PAC naturally follow by substituting ρ^2 according to the power constraints in (17).

Theorem 4 (Spatial Scaling for the PAC): For the PAC, under the IPC

$$P_e \leq \exp\left\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0, Q_1) + \frac{4A}{P_{\text{ind}}}\nu(Q_0, Q_1)}\right\} \tag{34}$$

while under the TPC

$$P_e \leq \exp\left\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0, Q_1) + \frac{4kA}{P_{\text{tot}}}\nu(Q_0, Q_1)}\right\}. \tag{35}$$

Remark 3: Exponential DEP decay is achievable under IPC. However, the bound exhibits an error floor under TPC, that is, as $k \to \infty$

$$P_e \leq \exp\left\{-\frac{P_{\text{tot}} D^2(Q_0\|Q_1)}{2A\nu(Q_0, Q_1)}\right\}. \tag{36}$$

It suggests poor performance of histogram fusion over the PAC when there is a stringent total power constraint.

The analysis for MAC follows essentially the same steps as above, except that the expression (20) for $\tilde{\mathbf{T}}$ is used in the application of Hoeffding's inequality.

Theorem 5 (Spatial Scaling for the MAC): For the MAC, under the IPC

$$P_e \leq \exp\left\{-\frac{2k^2D^2(Q_0\|Q_1)}{k\lambda(Q_0, Q_1) + \frac{4A}{P_{\text{ind}}}\nu(Q_0, Q_1)}\right\} \tag{37}$$

whereas under the TPC

$$P_e \leq \exp\left\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0, Q_1) + \frac{4A}{P_{\text{tot}}}\nu(Q_0, Q_1)}\right\}. \tag{38}$$

Remark 4: Compared to the PAC, exponential DEP scaling is achievable under either IPC or TPC for the MAC. In fact, the asymptotic exponent in the IPC bound

$$\frac{2kD^2(Q_0\|Q_1)}{k\lambda(Q_0, Q_1) + \frac{4A}{P_{\text{ind}}}\nu(Q_0, Q_1)} \xrightarrow{k\to\infty} \frac{2D^2(Q_0\|Q_1)}{\lambda(Q_0, Q_1)} \tag{39}$$

is independent of the transmit SNR at each SNR, as if the channel effects completely disappear.

Remark 5: It is seen from (38) that as long as $kP_{\text{tot}}(k) \to \infty$ the DEP decays to zero with k. For example, even if $P_{\text{tot}}(k) = 1/\log(k) \downarrow 0$, $P_e \to 0$ but at a subexponential rate (vanishing error exponent in the limit).

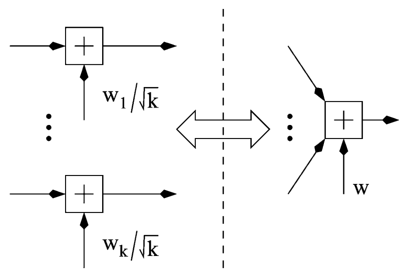

Fig. 3. Noise-level equivalence between the PAC and the MAC. All noises, $\{w_i\}$ and w have the same variance so that $\sum_{i=1}^{k} \mathrm{Var}((w_i)/(\sqrt{k})) = \mathrm{Var}(w)$.

B. Optimality of Histogram Fusion Over the MAC

While the above analysis based on Hoeffding's inequality reveals the required power scaling for exponentially decaying DEP, it is does not determine whether histogram fusion is asymptotically optimal or not. We resort to the argument in the optimality proof for LLR (soft-decision) fusion (Theorem 1) to prove this stronger result.

Theorem 6: Histogram fusion over the MAC asymptotically achieves the centralized error exponent under the individual power constraint

$$E_d^{\mathrm{hist}} = \lim_{k\to\infty} -\frac{1}{kn}\log P_e = E_c. \tag{40}$$

Proof: See Appendix IV.

Remark 6: As in the case of LLR fusion (see Remark 1), sublinear *total power scaling* is sufficient to achieve asymptotic optimality.

Remark 7: The dramatic difference in performance of histogram fusion over the PAC or the MAC is evident by comparing (20) and (19): the effective noise level in the MAC is $1/k$ times smaller than the effective noise level in the PAC, as illustrated in Fig. 3. It follows that the performance of histogram fusion over the MAC under the TPC is equivalent to its performance over the PAC under the IPC with $P_{\mathrm{tot}} = P_{\mathrm{ind}}$.

The superior performance of histogram fusion over the MAC is due to the k-fold beamforming gain inherent in coherent communication over the MAC. The beamforming gain is exploited via uncoded transmission of local statistics (type statistics in histogram fusion and the local LLRs in soft-decision fusion) which yields asymptotically optimal performance in both cases with remarkably low sensor power consumption. In a sense, the spatial summation inherent in the MAC is well matched to the additive structure of the sufficient statistics (type or LLR); that is, the global statistic is the sum of spatially distributed local statistics and is computed automatically by the physical channel itself during transmission. This source–channel matching and the optimality of uncoded communication over the MAC was first shown in [5] in the context of estimating a single Gaussian source from distributed noisy measurements. Uncoded coherent transmission by the sensor nodes was shown to attain the optimal $1/k$ distortion scaling with the number of node measurements. In a recent work [8], uncoded pulse position modulation over the MAC (a special case of histogram fusion) is shown to achieve the optimal centralized Cramér–Rao bound in the context of decentralized parameter estimation. The results in this paper, as well as the above mentioned, can be viewed as special cases of a general *source–channel matching principle: the MAC is naturally matched to inference about a single distributed source under the i.i.d. observations model, and uncoded transmission of appropriate local statistics is the optimal distributed communication strategy*. Extension of this matching principle to estimation of more general signal fields is addressed in [14] and to more general inference problems in [17].

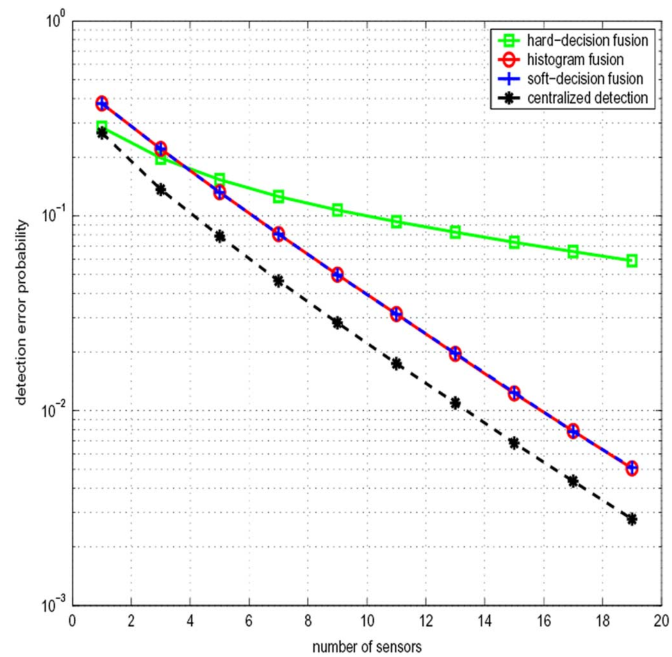

Fig. 4. DEP as a function of the number of sensors for the MAC under the IPC. $P_{\mathrm{ind}} = 5$ dB.

V. Numerical Simulations

We now present Monte Carlo simulation results to illustrate the theoretical results in this paper. Sensor observation data are assumed to be Bernoulli distributed (binary alphabet). The Bernoulli parameter (the probability of 1) is denoted by $P(1) = p_i$ for hypothesis $i \in \{0,1\}$. The two hypotheses are equally likely and in all results the data is generated for $p_0 = 0.6$ and $p_1 = 0.4$. Our focus is on spatial scaling of detection performance with the number of sensors k. The performance of histogram fusion is compared with LLR (soft-decision) fusion and hard-decision fusion, where the latter two require intelligent sensors. The performance of the centralized benchmark is also presented for reference. Each sensor transmission is based on a fixed observation sequence length of $n = 10$ in all results. The DEP values are calculated from 10^5 independent realizations of sensor observation sequences for each value of k.

Fig. 4 plots the DEP of the four schemes as a function of k for the MAC under the individual power constraint ($P_{\mathrm{ind}} = 5$ dB). As expected, the DEP of all schemes decays exponentially with k. The performance of histogram fusion is nearly identical to that of LLR fusion because the two are equivalent for discrete alphabet in the absence of noise. The figure also shows the optimality of histogram/LLR fusion over the MAC: both achieve the optimal centralized error exponent (slope of the semi-log plots).

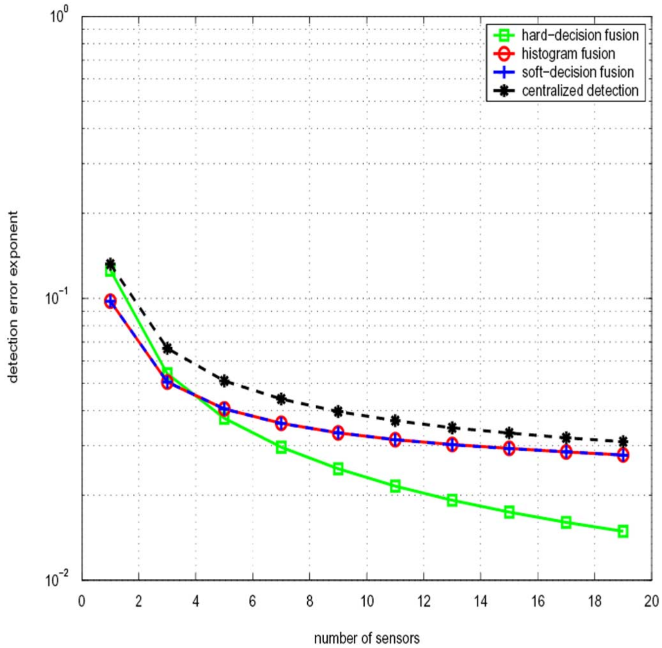

Fig. 5. Error exponent of the DEP as a function of the number of sensors for the MAC under the IPC. $P_{\rm ind} = 5$ dB.

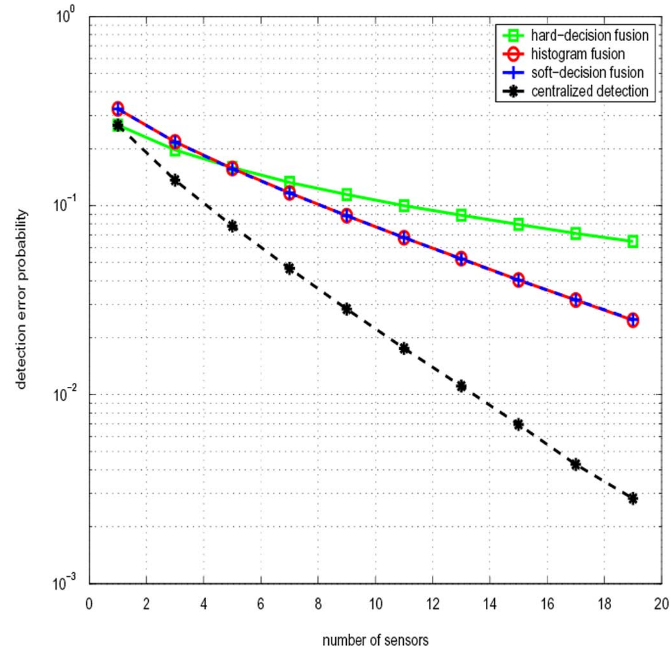

Fig. 7. DEP as a function of the number of sensors for the MAC under the TPC. $P_{\rm tot} = 10$ dB.

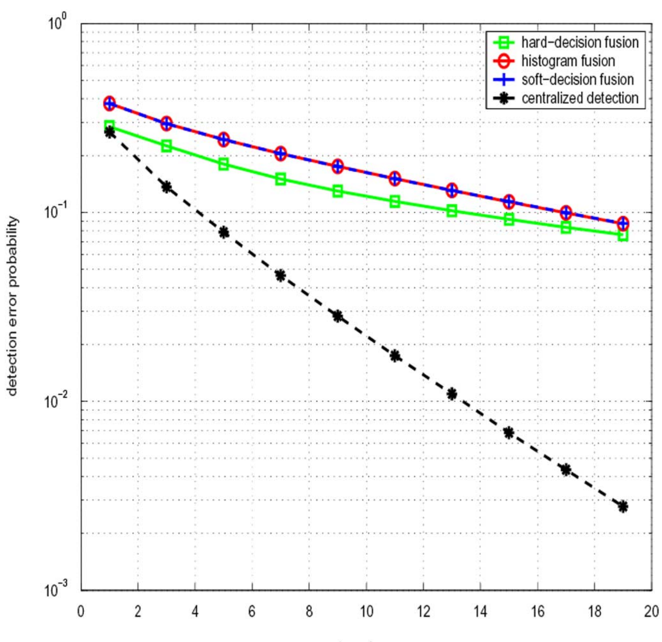

Fig. 6. DEP as a function of the number of sensors for the MAC under the TPC. $P_{\rm tot} = 5$ dB.

Barring a constant offset, the DEP curves of histogram and LLR fusion run parallel to that of centralized detection. The error exponent is calculated and plotted in Fig. 5, where the exponent of histogram/LLR fusion is seen to be converging to the centralized exponent. In contrast, there is a loss in error exponent for hard-decision fusion.

Detection performance over the MAC degrades under the stringent total power constraint, as shown in Fig. 6 ($P_{\rm tot} = 5$ dB) and Fig. 7 ($P_{\rm tot} = 10$ dB). Histogram fusion is no longer able to achieve the optimal centralized error exponent but it still achieves an exponential decay in DEP with the number of nodes. Comparing the two figures, we see that at a sufficiently low *transmit* SNR, hard-decision fusion initially outperforms histogram/LLR fusion for small k but the opposite is true eventually for sufficiently large k since the *receive* SNR increases linearly with k due to the beamforming gain in the MAC (the crossover occurs for $k > 19$ in Fig. 6). This crossover effect occurs because at sufficiently low *receive* SNR, the binary quantization in hard-decision fusion offers better immunity to channel noise. A related phenomenon was also reported in [3], where a similar cross-over between the performance of binary and analog sensor mappings was shown as a function of transmit power level. However, we note that for any given transmit SNR, the performance of histogram/LLR fusion eventually dominates hard-decision fusion for sufficiently large k in the case of the MAC.

Fusion over the PAC suffers a performance loss compared to fusion over the MAC as illustrated in Figs. 8 ($P_{\rm ind} = 10$ dB) and 9 ($P_{\rm tot} = 5$ dB). Under the individual power constraint, the DEP decays exponentially with k, albeit with a smaller exponent (compare with Fig. 4). On the other hand, the DEP exhibits an error floor under the TPC which is consistent with Remark 3. Note that Figs. 7 and 8 are identical confirming that the performance over the PAC under the IPC is equivalent to the performance over the MAC under the TPC (Remark 7).

VI. Conclusion

We have studied sensor encoding and transmission strategies for distributed detection in wireless sensor networks and analyzed the impact of the number of sensor measurements, and associated network power consumption, on detection performance. For intelligent sensors, LLR (soft-decision) fusion

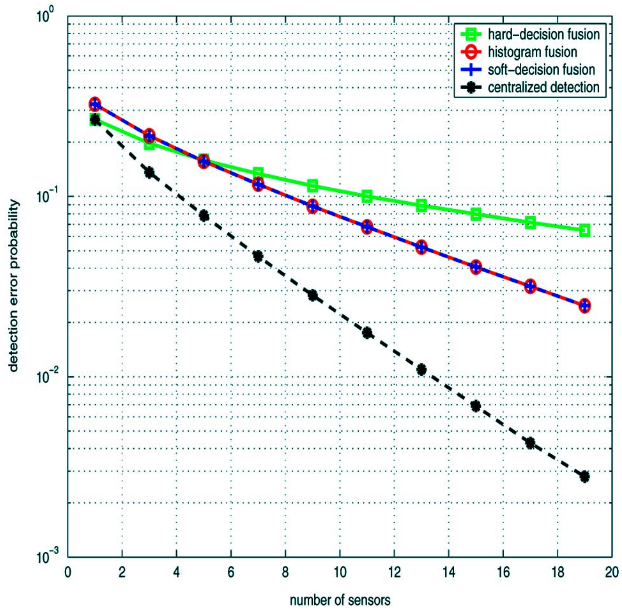

Fig. 8. DEP as a function of the number of sensors for the PAC under the IPC. $P_{\mathrm{ind}} = 10$ dB.

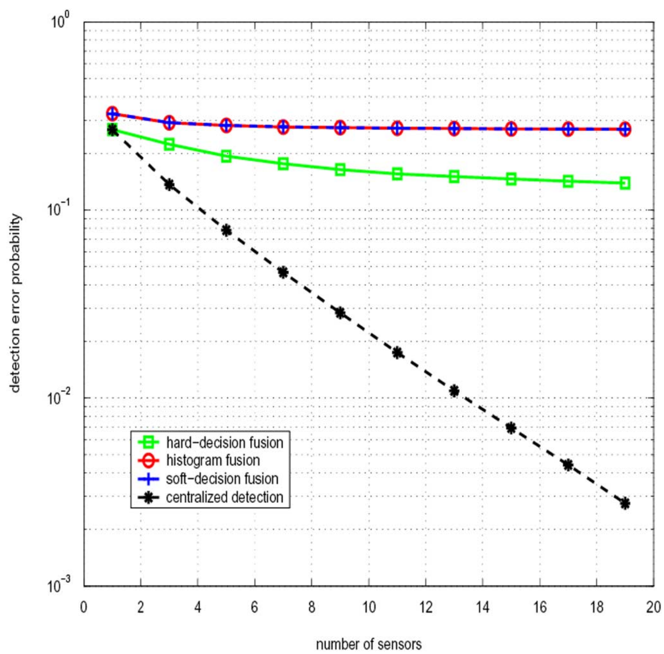

Fig. 9. DEP as a function of the number of sensors for the PAC under the TPC. $P_{\mathrm{tot}} = 10$ dB.

over the MAC is asymptotically optimal (achieves the centralized error exponent) with *sublinear* network power scaling (vanishing per-node power). The main focus of this paper is a type-based method—histogram fusion—that closely approximates the performance of LLR fusion with dumb sensors. The main source of loss in performance in histogram fusion, compared to LLR fusion, is the required quantization of the generally continuous-valued observation data. For discrete-valued data, histogram fusion over the MAC is also asymptotically optimal with sublinear network power scaling. Furthermore, both LLR and histogram fusion over the MAC exhibit exponential decay in error probability with the number of sensors even under finite total network power. Since the computation of local histograms requires dumb sensors that essentially act as counters, and the corresponding uncoded sensor transmissions require finite number of channel uses, type-based decentralized detection is a very promising low-latency strategy for large-scale, energy-, and cost-constrained wireless sensor networks. In particular, multiple detection tasks, corresponding to different hypothesis statistics, can be performed at the decision center using the same type-encoded sensor data.

The results in this paper also demonstrate the optimality of the MAC in the context of joint source-channel communication for distributed detection. The coherent spatial summation inherent to the MAC is matched to the linear structure of the global decision statistic and uncoded transmission of local statistics (type or LLR) is optimal. However, the associated synchronization requirements for coherent sensor transmissions and dynamic range considerations at the decision center need further investigation for practical implementation. Other directions for future research include studying the impact of quantization of sensor measurements and extension of type-based distributed detection to correlated measurements. The impact of fading in the MAC is studied in a more general setting in [17].

Appendix I
Proof of Theorem 1

We focus on a one-side DEP $P(\hat{\theta} = \theta_1 \,|\, \theta_0)$ thanks to the symmetry in the hypothesis testing problem. The error event in this case is given by

$$\mathcal{E} \subseteq \left\{ \sum_{i=1}^{k} \gamma_i + \frac{w}{\rho} \geq 0 \right\} = \left\{ \sum_{i=1}^{k} \log \frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} + \frac{nw}{\rho} \geq 0 \right\}. \tag{41}$$

Applying the Chernoff bound [12] and exploiting the i.i.d. data structure, for any $s > 0$, we have

$$\begin{aligned} P_e &\leq \mathbb{E}_0 \exp\left[s \sum_{i=1}^{k} \log \frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} + s\frac{nw}{\rho} \right] \\ &= \exp\left[kn \log \mathbb{E}_0 \left[\frac{Q_1(X)}{Q_0(X)} \right]^s + \frac{n^2}{2\rho^2} s^2 \right]. \end{aligned} \tag{42}$$

The optimal exponent of the centralized detection is characterized by Lemma 1. Let $s_0 \in (0, 1)$ be the optimal value in the Chernoff information $CI(Q_0, Q_1)$ as in (4). Then, at $s = s_0$

$$-\frac{1}{kn} \log P_e \geq CI(Q_0, Q_1) - \frac{n}{2k\rho^2} s_0^2 \rightarrow E_c, \quad \text{as } k \rightarrow \infty \tag{43}$$

which, together with the trivial upper bound $(E_d \leq E_c)$, proves the theorem.

Appendix II
Proof of Theorem 2

The entropy rate of $T(\mathbf{x}_i)$ at sensor i is upper bounded by

$$R_n \leq \frac{A \log(n+1)}{t} \rightarrow 0 \tag{44}$$

since $t \sim O(n)$. Let $P_{e,\text{opt}}$ denote the optimal centralized detection error probability when all the type statistics are perfectly available at the decision center. The channel error probability is denoted by P_{ch}^i for the ith channel. The probability of receiving incorrect type statistics can be union bounded by

$$P_{\text{ch}} \leq \sum_{i=1}^{k} P_{\text{ch}}^i \leq k \exp[-tE_r(R_n)] \tag{45}$$

where $E_r(R_n)$ is the channel error exponent [18] associated with rate R_n. The assumption $C > 0$ implies that

$$E_r(R_n) \to E_r(0) > 0. \tag{46}$$

Therefore, the overall detection error probability can be bounded as

$$\begin{aligned} P_e &\leq (1 - P_{\text{ch}})P_{e,\text{opt}} + P_{\text{ch}} \\ &\leq \exp[-nk(E_c + \epsilon_n)] + k\exp[-tE_r(R_n)] \\ &= \exp[-nk(E_c + \epsilon_n)](1 + \exp[-nD_n]) \end{aligned} \tag{47}$$

where $\epsilon_n \to 0$ as $n \to \infty$, and

$$D_n = \frac{t}{n}E_r(R_n) - k(E_c + \epsilon_n) + \frac{\log k}{n}. \tag{48}$$

Now, setting $t = bn$ where the constant $b = (kE_c)/(E_r(0)) + 1$, one has $D_n \to E_r(0) > 0$. Therefore

$$E_d = \lim_{n \to \infty} -\frac{1}{nk} \log P_e \geq E_c, \tag{49}$$

which, together with the trivial $E_d \leq E_c$, proves the theorem.

APPENDIX III
PROOF OF THEOREM 3

Using Chernoff method, for any $s > 0$, we have

$$\begin{aligned} P\left(\sum_{i=1}^{k} Z_i \geq d\right) &\leq \mathbb{E}\left[e^{s\left(\sum Z_i - d\right)}\right] \\ &= e^{-sd} \prod_i \mathbb{E}[e^{sX_i}]\mathbb{E}[e^{sY_i}] \end{aligned} \tag{50}$$

since X_i's and Y_i's are independent. The evaluation of the term $\mathbb{E}[e^{sY_i}]$ can be done by the standard Gaussian moment generating function; that is

$$\mathbb{E}[e^{sY_i}] = e^{\frac{1}{2}\sigma_i^2 s^2}. \tag{51}$$

We now focus on $\mathbb{E}[e^{sX}]$ where $X \in [a, b]$ and $\mathbb{E}X = 0$. The convexity of the function e^{sx} implies that

$$e^{sx} \leq \frac{e^{sb} - e^{sa}}{b - a}(x - a) + e^{sa} \tag{52}$$

where the right-hand side corresponds to the chord connecting point (a, e^{sa}) to point (b, e^{sb}) in the graph of e^{sx}. Taking expectation on both sides and using $\mathbb{E}X = 0$, we have

$$\mathbb{E}e^{sX} \leq \frac{e^{sb} - e^{sa}}{b - a}(-a) + e^{sa} = \frac{be^{sa} - ae^{sb}}{b - a}. \tag{53}$$

Setting $\theta = \frac{-a}{b-a}$ and $u = s(b - a)$, we get

$$\begin{aligned} \frac{be^{sa} - ae^{sb}}{b - a} &= \left[(1 - \theta) + \theta e^{s(b-a)}\right] e^{-\theta s(b-a)} \\ &= e^{-\theta u + \log(1 - \theta + \theta e^u)}. \end{aligned} \tag{54}$$

Note that $\theta \in (0, 1)$ and $u > 0$. We now bound $\phi(u) = -\theta u + \log(1 - \theta + \theta e^u)$ whose first and second derivatives are

$$\phi'(u) = -\theta + \frac{\theta e^u}{1 - \theta + \theta e^u} \tag{55}$$

$$\phi''(u) = \frac{\theta(1 - \theta)e^u}{(1 - \theta + \theta e^u)^2}. \tag{56}$$

Recall the Taylor expansion

$$\begin{aligned} \phi(u) &= \phi(0) + \phi'(0)u + u^2 \int_0^1 (1 - t)\phi''(tu)\, dt \\ &= u^2 \int_0^1 (1 - t)\phi''(tu)\, dt \end{aligned} \tag{57}$$

since $\phi(0) = 0$ and $\phi'(0) = 0$. But

$$\phi''(u) = \frac{\theta(1 - \theta)e^u}{(1 - \theta + \theta e^u)^2} \leq \frac{\theta(1 - \theta)e^u}{4\theta(1 - \theta)e^u} = \frac{1}{4} \tag{58}$$

where we have used the standard inequality $(x + y)^2 \geq 4xy$ for $x, y > 0$. Hence, we have

$$\mathbb{E}e^{sX} \leq e^{\phi(u)} \leq e^{\frac{u^2}{4}\int_0^1 (1-t)dt} = e^{\frac{u^2}{8}} = e^{\frac{s^2(b-a)^2}{8}}. \tag{59}$$

Therefore, the Chernoff bound (50) becomes

$$P\left(\sum_{i=1}^{k} Z_i \geq d\right) \leq e^{\frac{1}{8}s^2\left(\sum_{i=1}^{k}[(b_i - a_i)^2 + 4\sigma_i^2]\right) - sd}. \tag{60}$$

The theorem follows from optimizing the exponent $(s = (4d)/(\sqrt{\sum(b_i - a_i)^2 + 4\sigma_i^2}))$.

APPENDIX IV
PROOF OF THEOREM 6

The proof is directly adapted from that of Theorem 1. The detection error event for histogram fusion over the MAC is given by

$$\begin{aligned} \mathcal{E} &= \left\{\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \sum_{i=1}^{k} \mathbf{T}_i + \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \tilde{\mathbf{w}} \geq 0\right\} \\ &= \left\{\left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T \sum_{i=1}^{k} n\mathbf{T}_i + \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T n\tilde{\mathbf{w}} \geq 0\right\} \\ &= \left\{\sum_{i=1}^{k} \left[\log \frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)}\right] + \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T n\tilde{\mathbf{w}} \geq 0\right\} \end{aligned} \tag{61}$$

where we have used the linear relationship (15) to get the log-likelihood ratio from the type statistics. Applying the Chernoff bound and exploiting the i.i.d. data structure, for any $s > 0$, we have

$$P_e \leq \mathbb{E}_0 e^{s \sum_{i=1}^{k} \log \frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} + s \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^T n\tilde{\mathbf{w}}} = e^{kn \log \mathbb{E}_0 \left[\frac{Q_1(X)}{Q_0(X)}\right]^s + \frac{n^2}{2\rho^2}\nu(Q_0,Q_1)s^2}. \quad (62)$$

Let $s_o \in (0,1)$ optimize the Chernoff information and $\rho^2 = P_{\text{ind}}/A$ be the individual power constraint. We have

$$-\frac{1}{kn}\log P_e \geq \log \mathbb{E}_0 \left[\frac{Q_1(X)}{Q_0(X)}\right]^{s_o} - \frac{nA}{2kP_{\text{ind}}}\nu(Q_0,Q_1)s_o^2 \rightarrow E_c, \quad \text{as } k \rightarrow \infty \quad (63)$$

from which, together with the trivial upperbound $(E_d \leq E_c)$, the theorem follows.

References

[1] A. D'Costa, V. Ramachandran, and A. M. Sayeed, "Distributed classification of Gaussian space–time sources in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1026–1036, Aug. 2004.

[2] J. N. Tsitsiklis, "Decentralized detection," *Adv. Stat. Signal Process.*, vol. 2, pp. 297–344, 1993.

[3] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.

[4] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science (Proceedings of IPSN'03)*, F. Zhao and L. Guibas, Eds., Berlin, Germany, Apr. 2003, pp. 162–177.

[5] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[6] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[8] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 613–622, Feb. 2006.

[9] G. Mergen, V. Naware, and L. Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1081–1092, 2007.

[10] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[11] R. O. Duda, R. E. Hart, and D. G. Stock, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[12] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[13] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[14] W. Bajwa, A. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in sensor networks," in *Proc. Information Processing Sensor Networks (IPSN)*, Apr. 2005, pp. 332–339.

[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, pp. 13–30, 1963.

[17] K. Liu, H. E. Gamal, and A. Sayeed, "Decentralized inference over multiple access channels," *IEEE Trans. Signal Process.*, 2006, accepted for publication.

[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

Ke Liu (S'99–M'05) received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1999 and the M.S. degree in electrical engineering, the M.A. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Wisconsin–Madison in 2001, 2002, and 2004, respectively.

From 2000 to 2004, he was a Research Assistant in the Wireless Communications Laboratory at the University of Wisconsin–Madison. He was a Postdoctoral Researcher at Ohio State University, Columbus, from 2004 to 2005. In 2005, he joined Qualcomm, Inc., San Diego, CA, as a Senior Engineer working on the next-generation wireless systems. His current research interests include wireless communications, wireless networks, information theory, and signal processing.

Dr. Liu received a Graduate School Fellowship from 1999 to 2000 and was awarded the Harold A. Peterson Graduate Student Paper Award for his work on space–time coding for multiantenna systems. He served as reviewer for the IEEE Communications, Signal Processing, and Information Theory societies.

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

He was a Postdoctoral Fellow at Rice University, Houston, TX, from 1996 to 1997, and he has been with the University of Wisconsin–Madison since 1997, where he is currently Associate Professor of electrical and computer engineering. His current research interests include wireless communications, multidimensional communication theory, statistical signal processing, information theory, and applications in wireless networks.

Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1999 to 2002 and as a Guest Editor for a special issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on sensor networks in 2005.