Optimal Distributed Detection Strategies for Wireless Sensor Networks

Ke Liu and Akbar M. Sayeed *

University of Wisconsin-Madison

kliu@cae.wisc.edu, akbar@engr.wisc.edu

Abstract

We study optimal distributed detection strategies for wireless sensor networks under the assumption of spatially and temporally i.i.d. observations at the sensor nodes. Each node computes a local statistic and communicates it to a decision center over a noisy channel. The performance of centralized detection (noise-free channel) serves as a benchmark. We address the following fundamental question: *under what network resource constraints can distributed detection achieve the same error exponent as centralized detection?* Two types of constraints are considered: 1) transmission power constraints at the nodes, and 2) the communication channel between the nodes and the decision center. Two types of channels are studied: 1) a parallel access channel (PAC) consisting of dedicated AWGN channels between the nodes and the decision center, and 2) an AWGN multiple access channel (MAC). We show that for *intelligent sensors* (with knowledge of observation statistics) analog communication of local likelihood ratios (*soft decisions*) over the MAC is asymptotically optimal (for large number of nodes) when each node can communicate with a constant power. Motivated by this result, we propose an optimal distributed detection strategy for *dumb sensors* (oblivious of observation statistics) based on the *method of types*. In this strategy, each node appropriately quantizes its temporal observation data and communicates its type or histogram to the decision center. It is shown that type-based distributed detection over the MAC is also asymptotically optimal with an additional advantage: observation statistics are needed only at the decision center. Even under the more stringet total power constraint, it is shown that both soft decision- and type-fusion result in exponentially decaying error probability.

1 Introduction

Wireless sensor networks are an emerging technology that promise an unprecedented ability to monitor the physical environment via a large number of cheap sensing devices that can communicate with a fusion center in a tetherless fashion. Detection of certain events or targets in the environment is an important application of sensor networks (see, e.g., [1, 2]). The performance of the centralized detector (noise-free communication between the nodes and the fusion center) serves as a benchmark for any distributed detection strategy. We assess the performance of distributed detection strategies in terms of the *error exponent* associated with their probability of error P_e. Specifically, we address the following fundamental question: *Under what network resources constraints can a distributed detection strategy achieve the centralized benchmark performance (error exponent) in the limit of large number of node measurements?*[1]

*This research was supported in part by NSF under grants #CCR-9875805 and #CCR-0113385, and by the ONR under grant #N00014-01-1-0825.

[1]We note that for a fixed number of nodes, optimal performance can be achieved in the limit of large number of i.i.d. temporal measurements [3]; see also Sec. 4.1.

We consider two types of resource constraints relating to the power budget for sensors and the nature of the communication link between the nodes and the decision center. Specifically, we consider two extreme power constraints: 1) individual power constraint (IPC) in which each node has constant power, and 2) total power constraint (TPC) in which the total power over all nodes is constant. We consider two types of channels: 1) a parallel access channel (PAC) consisting of dedicated AWGN channels between the nodes and the decision center, and 2) an AWGN multiple access channel (MAC). Finally, we assume that the node observations are independent and identically distributed (i.i.d.) both spatially (across nodes) and temporally (see [2] for a physical justification of this assumption).

In the next section, we formulate the distributed detection problem and first consider the case of *intelligent nodes* that have knowledge of observation statistics, an assumption that is typically assumed in existing works (see,e.g.,[1, 2]). Building on the results in [2] for the PAC under IPC, we show that analog communication of local likelihood ratios from the nodes (*soft decision-fusion*) over the MAC is asymptotically optimal; that is, it achieves the centralized error exponent in the limit of large number of nodes. Furthermore, even under the more stringent TPC, soft decision-fusion results in exponentially decaying probability of error, P_e, with the number of nodes. On the other hand, communication over the PAC incurs a significant degradation in performance compared to the MAC.

Motivated by the above key result, the second major contribution of this paper is an optimal distributed detection strategy for *dumb sensors* (oblivious of observation statistics) that is based on the *method of types* [4, 5, 6]. A type is essentially a histogram or empirical probability distribution of the observation data. In this strategy, each node appropriately quantizes its raw temporal observation data and communicates its type to the decision center (*histogram fusion*). The type-based distributed detection framework is described in Section 3 and, to the best of our knowledge, it is the first application of the method of types to sensor networks.

Section 4 analyzes the performance of type-based distributed detection. We show that histogram fusion over the MAC is also asymptotically optimal under the IPC, and results in an exponential decay in P_e even under the TPC. Furthermore, it is shown that type-based detection is essentially an equivalent implementation of soft decision-fusion with a key advantage: *the computation of likelihood ratios is shifted from the nodes to the decision center and the nodes can be completely oblivious of the observation statistics*. In Section 5, numerical results are presented to illustrate the optimal performance of soft decision-fusion and histogram-fusion under different network constraints. Concluding remarks are presented in Section 6.

2 Decentralized Detection

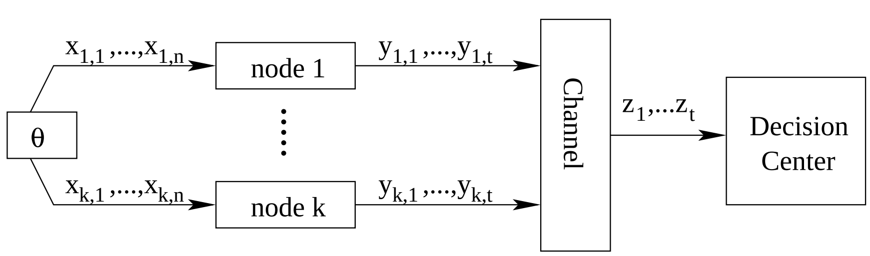


Figure 1: Architecture for decentralized detection using k spatially distributed sensor nodes. The sequence $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,n})$ denotes the observation data at node i ($1 \leq i \leq k$). The observations are generated according to a distribution Q_θ, $\theta \in \{\theta_0, \theta_1\}$.

Fig. 1 shows an achitecture for decentralized detection in wireless sensor networks that is studied in this paper. A source generates the observation data according to a distribution Q_θ corresponding to a finite number of hypotheses H_θ. In this paper, we consider *binary* hypothesis testing, that is, $\theta \in \{0,1\}$, to simplify exposition, although the essential ideas hold for general M-ary hypothesis testing. A group of k spatially distributed sensor nodes observe length-n data sequences $\mathbf{x}$ that are *independently and identically distributed* (i.i.d.) across time and space: $x_{i,j}$ ($1 \le i \le k$ and $1 \le j \le n$) denotes the observation random variable at the i-th sensor node and the j-th time instant. The observation data can be continuous or discrete, depending on the physical source and the sensing process. Each node encodes its observation sequence $\mathbf{x}_i$ into a length-t sequence $\mathbf{y}_i = f_i(\mathbf{x}_i)$ and communicates it to a (remote) decision center via a noisy communication link.[2] The decision center makes a decision $\hat{\theta}$ on the hypothesis based on its length-t received signal $\mathbf{z}$. The detection performance can be characterized by the error probabilities: $\alpha = P(\hat{\theta} = 1|H_0)$, $\beta = P(\hat{\theta} = 0|H_1)$, $P_e = (\alpha+\beta)/2$, where P_e is the average *detection error probability* (DEP) assuming equally likely hypotheses.

The design of a distributed detection strategy boils down to the design of the mappings $\{f_i, i = 1, \cdots, k\}$ at the k sensor nodes under network resource constraints. These mappings represent a *joint source-channel communication* strategy in the context of detection applications. This is a long standing open problem and only a few partial answers are available. For instance, it is shown in [7, 1] that under the i.i.d. observation model, the optimal strategy employs the same encoder at all nodes, which we adopt here as well. However, relatively little is known about the actual construction of the optimal mapping $\{f\}$, under appropriate network constraints. We consider two key network constraints in this work.

Power Constraints: Energy consumption at sensor nodes is a key consideration in the design of sensor networks. Along the lines of [1], we consider power constrained networks. Each node i is allocated a certain power budget P_i for communication: $\frac{1}{t}\sum_{j=1}^{t} \mathbb{E}\,|y_{i,j}|^2 \le P_i$. We consider two extreme power constraints: 1) *individual power constraint (IPC)* in which each node is given a constant power budget $P_i = P_{ind}$, and 2) *total power constraint (TPC)* in which the overall network power is limited to a constant P_{tot} and each sensor node has a diminishing power budget $P_i = P_{tot}/k$ as k increases.

Network Communication Channel: For simplicity, we consider two types of channels: 1) a parallel access channel (PAC) consisting of k dedicated AWGN channels, and 2) an AWGN multiple access channel (MAC). For each channel use, in the PAC, $z_i = y_i + w_i$, $i = 1, \cdots, k$, whereas in the MAC, $z = \sum_{i=1}^{k} y_i + w$. The channel noise is zero-mean with unit variance.

2.1 Centralized Detection – Performance Benchmark

Decentralized detection differs from its centralized counterpart primarily due to the noisy channel that separates data collecting nodes from the decision center. In centralized detection, the observation data at all nodes, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$, is available at the decision center, which uses the *maximum likelihood (ML) test* for optimal performance:

$$\hat{\theta} = \begin{cases} 1\,, & Q_1(\mathbf{X}) \ge Q_0(\mathbf{X}) \\ 0\,, & Q_1(\mathbf{X}) < Q_0(\mathbf{X}) \end{cases} \tag{1}$$

[2] We assume that the total channel use $t \sim O(n)$.

It is convenient to express the ML test in terms of the (normalized) log-likelihood ratio (LLR)

$$\gamma = \frac{1}{nk} \log \frac{Q_1(\mathbf{X})}{Q_0(\mathbf{X})} = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{n} \log \frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} = \frac{1}{k} \sum_{i=1}^{k} \gamma_i \ , \quad \hat{\theta} = \begin{cases} 1 \ , & \gamma \geq 0 \\ 0 \ , & \gamma < 0 \end{cases} \tag{2}$$

The performance of centralized detection is well studied (see, e.g., [4, 8, 9]). The DEP decays exponentially with the data size (kn in our setting), and the optimal error exponent (E_c) is characterized by the following well-known result.

Lemma 1 (Chernoff). *The centralized ML-test achieves the optimal error exponent*

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

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

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

where $\mathbb{E}_0$ denotes expectation with respect to Q_0.

Centralized ML detection represents the ultimate performance benchmark for any decentralized detection scheme: the error exponent of decentralized detection, E_d, is smaller than E_c in general. An overarching goal is to reduce, or even to close, such performance gap between decentralized and centralized detection by optimizing the joint source-channel communication between the sensor nodes and the decision center, subject to network resource constraints.

2.2 Distributed Detection Strategies with Intelligent Sensors

In this section, we discuss distributed detection strategies when the sensors are intelligent: they have knowledge of source statistics under different hypotheses. Evidently, this assumption is prevalent in all existing works (see, e.g., [7, 1, 2]). Although the temporal and spatial dimensions are interchangeable in centralized detection, their role in decentralized detection is fundamentally different. While each node can jointly process its temporal data, joint data processing across distributed nodes is generally not viable due to high cost and complexity associated with inter-node communication. The architecture in Fig. 1 assumes that each node can only process its *local* data. Temporal data processing in decentralized detection has been studied from the perspective of source compression (see, e.g., [3] and references therein), as n gets large (with k fixed; more on this in Section 4.1). However, the potential of *spatially distributed* processing is arguably the most important and unique aspect of wireless sensor networks. Thus, we mainly focus on the situation where the network has a large number of nodes (k), but with *limited* data (n) at each node.

In the spatial context, an attractive approach is *soft decision-fusion* [8, 2], which is inspired by the LLR test in centralized detection. Each sensor computes the value of its local LLR, γ_i, defined in (2), which is then transmitted to the decision center. If the transmission is noise-free, the decision center can implement the centralized detector simply by averaging the local LLRs as in (2). However, the decision center can only obtain noisy estimates of local LLRs, and hence, a noisy estimate of the global LLR, due to the communication channel. Therefore, soft decision-fusion generally suffers a performance loss compared to centralized detection.

By compressing the length-n observation sequence into a single LLR value, each sensor significantly reduces the amount of transmission data. The LLR value can be quantized and encoded to protect against channel error. One natural choice is a one-bit quantization strategy, *hard decision-fusion*, in which each sensor makes a local (hard) decision, $\mathrm{sign}(\gamma_i)$, and the decision center makes a final decision by fusing the local hard decisions (see, e.g., [2]). In [2], hard decision-fusion over the PAC is studied and it is shown it incurs a significant loss in error exponent (compared to E_c) even in the absence of channel noise. The next result improves upon the work in [2] and shows that, for the MAC, unquantized and uncoded soft decision-fusion is, in fact, optimal under the IPC. In this scheme, each sensor sends its real-valued (analog) LLR value directly over the MAC: $z = \sum_{i=1}^{k} \rho\gamma_i + w$, where ρ controls the signal amplification. The decision center then sets the global LLR estimate to $\tilde{\gamma} = z/\rho k$ and performs the LLR test.

Theorem 1 (Asymptotic Optimality of Soft-Decision Fusion over the MAC). *Uncoded (analog) soft decision-fusion over the MAC is asymptotically optimal under the IPC:*

$$E_d^{soft} = \lim_{k\to\infty} -\frac{1}{kn}\log P_e = E_c \tag{5}$$

Proof. The amplification factor ρ is a chosen to satisfy the IPC. Because of the symmetry, we focus on the probability of error in detecting H_1 while H_0 is true, for which the error event is

$$\mathcal{E} \subseteq \left\{\sum_{i=1}^{k}\gamma_i + \frac{w}{\rho} \geq 0\right\} = \left\{\sum_{i=1}^{k}\log\frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} + \frac{nw}{\rho} \geq 0\right\}. \tag{6}$$

Applying the Chernoff bound, for any $s > 0$ we have

$$P_e \leq \mathbb{E}_0 \exp\left[s\sum_{i=1}^{k}\log\frac{Q_1(\mathbf{x}_i)}{Q_0(\mathbf{x}_i)} + s\frac{nw}{\rho}\right] = \exp\left[kn\log\mathbb{E}_0\left[\frac{Q_1(X)}{Q_0(X)}\right]^s + \frac{n^2}{2\rho^2}s^2\right]. \tag{7}$$

Let $s_0 \in (0,1)$ optimize the Chernoff information for the centralized detector in (4). Then,

$$E_d^{soft} = -\frac{1}{kn}\log P_e \geq CI(Q_0,Q_1) - \frac{n}{2k\rho^2}s_0^2 \to E_c, \quad \text{as } k\to\infty, \tag{8}$$

which, together with the trivial upper bound ($E_d^{soft} \leq E_c$), proves the theorem. □

Theorem 1 motivates the use of uncoded soft decision-fusion for intelligent sensors: LLRs can be computed regardless of continuous or discrete random variables and simple uncoded transmission is sufficient to achieve optimal performance. However, in order to calculate the local LLR values, sensor nodes must know the source statistics in advance, which makes soft decision fusion *application-dependent* and a potential problem since network-wide updating of sensor processing algorithms may be too costly. A natural question arises: *can we attain near-optimal performance with* **dumb sensors** *which perform local processing without knowledge of source statistics?* In the following, we develop an optimal distributed detection strategy for dumb sensors using the *method of types*.

3 Type-Based Distributed Detection with Dumb Sensors

In this section, we develop a distributed detection framework using the *method of types* [4, 5, 6] in which the sensor nodes do not need knowledge of source statistics. The method of types represents a rich connection between information theory and statistics; it relates to sequences with with the same empirical distribution or "type". Since the method of types applies only to discrete random variables, in the rest of the paper we assume that the observation data comes from a discrete alphabet $\mathcal{A}$. Thus, quantization of sensor measurements is needed, as usual in practice, and the resulting performance loss can be made arbitrarily small by using sufficiently fine quantization.

The Method of Types. Let $\mathcal{A}$ be a size-A discrete alphabet with symbols $\{a_1, \ldots, a_A\}$. For probability mass functions P and Q, the entropy $H(P)$ and the *Kullback-Leibler (K-L)* distance are defined as [4]

$$H(P) = -\sum_{a \in \mathcal{A}} P(a) \log P(a) \ , \ D(P \| Q) = \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)} \ . \tag{9}$$

Definition 1 (Type). *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 $\mathcal{A}$*

$$T_{\mathbf{x}}(a) = N(a|\mathbf{x})/n, \quad \forall a \in \mathcal{A}, \tag{10}$$

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

For example, the type of the length-3 sequence $(1, 0, 1)$ from $\mathcal{A} = \{0, 1\}$ is $(1/3, 2/3)$. Different sequences may have the same type; e.g., the sequences $(1, 0, 1)$, $(0, 1, 1)$ and $(1, 1, 0)$ all have the same type. The following combinatorial result plays an important role in our context.

Proposition 1 (see Theorem 12.1.1 in [4]). *Fix sequence length n and alphabet size A. The total number of different types is upper-bounded by $(n + 1)^A$.*

If random sequence $x_1, \ldots, x_n$ is drawn i.i.d. according to Q, then the probability of $\mathbf{x}$ depends only on its type:

$$Q(\mathbf{x}) = \prod_{a \in \mathcal{A}} Q(a)^{nT_{\mathbf{x}}(a)}. \tag{11}$$

The (normalized) *log-likelihood (LLH)* of $\mathbf{x}$ can be written as

$$\frac{1}{n} \log Q(\mathbf{x}) = \sum_{a \in \mathcal{A}} T_{\mathbf{x}}(a) \log Q(a) = \frac{1}{n} \log Q(\mathbf{x}) = [\log \mathbf{Q}]^\dagger \mathbf{T} \tag{12}$$

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

$$\gamma = \log \frac{Q_1(\mathbf{x})}{Q_0(\mathbf{x})} = \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0} \right]^\dagger \mathbf{T} \tag{13}$$

where $\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0} = \left[\log \frac{Q_1(a_1)}{Q_0(a_1)}, \ldots, \log \frac{Q_1(a_A)}{Q_0(a_A)} \right]^\dagger$. Using (12) and (13), algorithms based on likelihood statistics can be equivalently implemented via the type of the observed data.

Transmission of Type Statistics from Sensor Nodes. Each node computes the type $T(\mathbf{x})$ of its length-n sequence $\mathbf{x}$: $\mathbf{T} = [T_{\mathbf{x}}(a_1), \ldots, T_{\mathbf{x}}(a_A)]^\dagger$. Although the type can be coded to cope with channel noise, we consider a simple uncoded, analog scheme inspired by Theorem 1: each relative frequency $T_{\mathbf{x}}(a) \in [0, 1]$ is simply amplitude-modulated to send over the channel. Thus, type transmission takes A channel uses, which is independent of n, and thus represents an efficient source compression. The node power budget is equally divided between the A channel uses. The signal amplification level ρ is given by

$$\rho^2 = P_{ind}/A \text{ for IPC and } \rho^2 = P_{tot}/kA \text{ for TPC.} \tag{14}$$

Type or Histogram Fusion at the Decision Center. Effectively, each sensor sends the histogram of its local data to the decision center. The decision center then fuses the local histograms from the sensor nodes to estimate the global type statistic; Fig. 2 illustrates this simple scheme, *histogram fusion*, for both the PAC and the MAC. Let $\mathbf{T}_i$ denote the type or

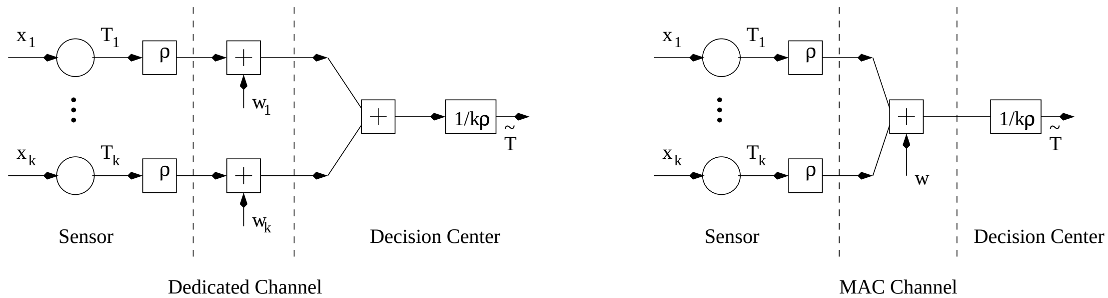


Figure 2: Histogram fusion. (a) PAC. (b) MAC.

histogram of the data $\mathbf{x}_i$ at sensor i: $\mathbf{T}_i = [T_{\mathbf{x}_i}(a_1), \ldots, T_{\mathbf{x}_i}(a_A)]^\dagger$. In the absence of channel noise, the global type statistic for the entire distributed data $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is given by $\mathbf{T} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{T}_i$. Inspired by this, the decision center averages the received (noisy) local histograms to obtain an estimate of the global type statistics

$$\tilde{\mathbf{T}} = \frac{1}{k}\sum_{i=1}^{k}[\mathbf{T}_i + \tilde{\mathbf{w}}_i] \text{ and } \tilde{\mathbf{T}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{T}_i + \frac{1}{k}\tilde{\mathbf{w}} \tag{15}$$

for the PAC and MAC, respectively. Note that $\tilde{\mathbf{w}}_i$ is a length-A noise vector associated with the i-th channel in the PAC and $\tilde{\mathbf{w}}$ is the MAC noise vector. The noise elements in $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{w}}$ have variance $1/\rho^2$, where ρ is defined in (14). Finally, motivated by the linear representation of LLH/LLR values via types in (12) and (13), the decision center makes the final decision as

$$\tilde{\gamma} = \left[\log \frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^\dagger \tilde{\mathbf{T}} \ , \ \ \hat{\theta} = \begin{cases} 1 \ , & \tilde{\gamma} > 0 \\ 0 \ , & \tilde{\gamma} < 0 \end{cases} \tag{16}$$

Remark 1. *The relation* (16) *shows that histogram fusion results in an efficient implementation of soft decision-fusion*[3] *with a key advantage: the calculation of LLR is shifted from the nodes to the decision center. Furthermore, the nodes can be both "simple" and "dumb": simple counters can implement the type computation and sensors do not need to know* Q_0 *and* Q_1.

4 Asymptotic Performance of Type-Based Detection

4.1 Temporal Asymptotics

The essence of the temporal advantage of type-based detection is best understood via the following result, which we state without proof (see [10]).

Lemma 2 (Sufficient Source Compression). *Let* $\mathbf{x} = (x_1, \ldots, x_n)$ *be a length-*n *i.i.d. random sequence from a discrete alphabet* $\mathcal{A}$. *Then, for distributions* Q_0 *and* Q_1 *on* $\mathcal{A}$

$$D(Q_{0,\mathbf{x}} \| Q_{1,\mathbf{x}}) = D(Q_{0,T(\mathbf{x})} \| Q_{1,T(\mathbf{x})}) \quad , \quad CI(Q_{0,\mathbf{x}}, Q_{1,\mathbf{x}}) = CI(Q_{0,T(\mathbf{x})}, Q_{1,T(\mathbf{x})}) \tag{17}$$

$$H(T(\mathbf{x})) \ \leq \ A\log(n+1) \tag{18}$$

[3] Here, the context is discrete random variables.

where the notation $Q_{0,T(\mathbf{x})}$ denotes the distribution of $T(\mathbf{x})$ as induced by Q_0 on $\mathbf{x}$.

Lemma 2 shows that types preserve the Chernoff information and K-L distances, and thus using types instead incurs no loss in detection performance. On the other hand, the entropy of type statistics scales *logarithmically* with sequence length (see Prop. 1) compared with linear increase in the entropy of raw data. Thus, the entropy rate of type statistics, $H(T(\mathbf{x}))/n \to 0$, as n increases – the zero-rate property of types [3]. Consequently, type statistics collected by the nodes can be reliably transmitted to the decision center, no matter how small the capacity of the sensor-decision center communication link, although at the cost of delay. This is formalized in the following result (see [10] for a proof) which can also be derived using the results in [3].

Theorem 2 (Temporal Optimality of Type-Based Detection). *Let $C > 0$ be the link capacity from each node to the decision center. Assuming a fixed number of nodes (k) and a constant scaling of channel use t with n, $t \sim O(n)$, we have: $E_d^{hist} = \lim_{n\to\infty} -\frac{1}{nk}\log P_e = E_c$.*

4.2 Spatial Asymptotics

In this section, we study the performance of type-based distributed detection in the limit of a large number of sensor nodes (k) and finite temporal data (n) at each node. It turns out that the boundedness of the type values, $T \in [0,1]$, greatly facilitates analysis via the use of *Hoeffding's inequality* [11]. However, since the noisy estimates $\tilde{T}$ of the joint type are not bounded, we use a generalization of Hoeffding's inequality (derived in [10])

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

$$P\Big(\sum_{i=1}^{k} Z_i \geq d\Big) \leq e^{-\frac{2d^2}{\sum_{i=1}^{k}(b_i-a_i)^2+4\sigma_i^2}}. \tag{19}$$

We illustrate application of Theorem 3 by considering the DEP for the PAC. The error event and DEP can be expressed as

$$\mathcal{E} = \{\tilde{\gamma} \geq 0\} = \left\{\sum_{i=1}^{k}\left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^{\dagger}(\mathbf{T}_i + \tilde{\mathbf{w}}_i) \geq 0\right\}\ ,\ \ P_e = Q_0(\mathcal{E})\ . \tag{20}$$

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

$$\left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^{\dagger}\mathbb{E}_0\,\mathbf{T}_i = -\sum_{a\in\mathcal{A}} Q_0(a)\log\frac{Q_0(a)}{Q_1(a)} = -D(Q_0\|Q_1). \tag{21}$$

Therefore, we can centralize the random vector $\mathbf{T}$ to get

$$\mathcal{E} = \left\{\sum_{i=1}^{k}\left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^{\dagger}([\mathbf{T}_i - \mathbf{Q}_0] + \tilde{\mathbf{w}}_i) \geq kD(Q_0\|Q_1)\right\}. \tag{22}$$

It is easy to verify that $X_i = \left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^{\dagger}(\mathbf{T}_i - \mathbf{Q}_0)$ and $Y_i = \left[\log\frac{\mathbf{Q}_1}{\mathbf{Q}_0}\right]^{\dagger}\tilde{\mathbf{w}}_i$ satisty the conditions of Theorem 3. Define two measures of discrepancy between Q_0 and Q_1:

$$\lambda(Q_0,Q_1) = \left(\max_a \log\frac{Q_1(a)}{Q_0(a)} - \min_a \log\frac{Q_1(a)}{Q_0(a)}\right)^2\ ,\ \ \nu(Q_0,Q_1) = \sum_a \log^2\frac{Q_1(a)}{Q_0(a)}. \tag{23}$$

Now applying Theorem 3, we have

$$P_e = Q_0\Big(\sum_{i=1}^{k}(X_i + Y_i) \geq kD(Q_0\|Q_1)\Big) \leq e^{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0,Q_1)+\frac{4}{\rho^2}\nu(Q_0,Q_1)}}. \tag{24}$$

The bounds on the DEP for the PAC follow by using ρ^2 for the IPC or TPC in (14).

Theorem 4 (Spatial Scaling for the PAC). *The DEP of histogram fusion over the PAC is bounded under the IPC and TPC, respectively, by*

$$P_e \leq \exp\Big\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0,Q_1)+\frac{4A}{P_{ind}}\nu(Q_0,Q_1)}\Big\} \;,\; P_e \leq \exp\Big\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0,Q_1)+\frac{4kA}{P_{tot}}\nu(Q_0,Q_1)}\Big\}. \tag{25}$$

Remark 2. *For the PAC, the DEP decays exponentially with k under IPC. However, the performance degrades under the more stringent TPC; the corresponding bound does not vanish.*

The analysis for the MAC follows the same steps as above by using (15) for the MAC.

Theorem 5 (Spatial Scaling for the MAC). *The DEP of histogram fusion over the MAC is bounded under the IPC and TPC, respectively, by*

$$P_e \leq \exp\Big\{-\frac{2k^2D^2(Q_0\|Q_1)}{k\lambda(Q_0,Q_1)+\frac{4A}{P_{ind}}\nu(Q_0,Q_1)}\Big\} \;,\; P_e \leq \exp\Big\{-\frac{2kD^2(Q_0\|Q_1)}{\lambda(Q_0,Q_1)+\frac{4A}{P_{tot}}\nu(Q_0,Q_1)}\Big\}. \tag{26}$$

Remark 3. *Under the IPC, not only does the DEP decay exponentially with k, but also the asymptotic exponent ($k \to \infty$) is independent of communication channel parameters, as if the channel effects completely disappear. However, such strong performance does not hold under the TPC, but even in this case DEP decays exponentially with k.*

The next result shows that histogram fusion over the MAC is asymptotically optimal. The proof is based on an argument similar to that in Theorem 1 (see [10]).

Theorem 6. *Histogram fusion over the MAC is asymptotically optimal under the IPC:*

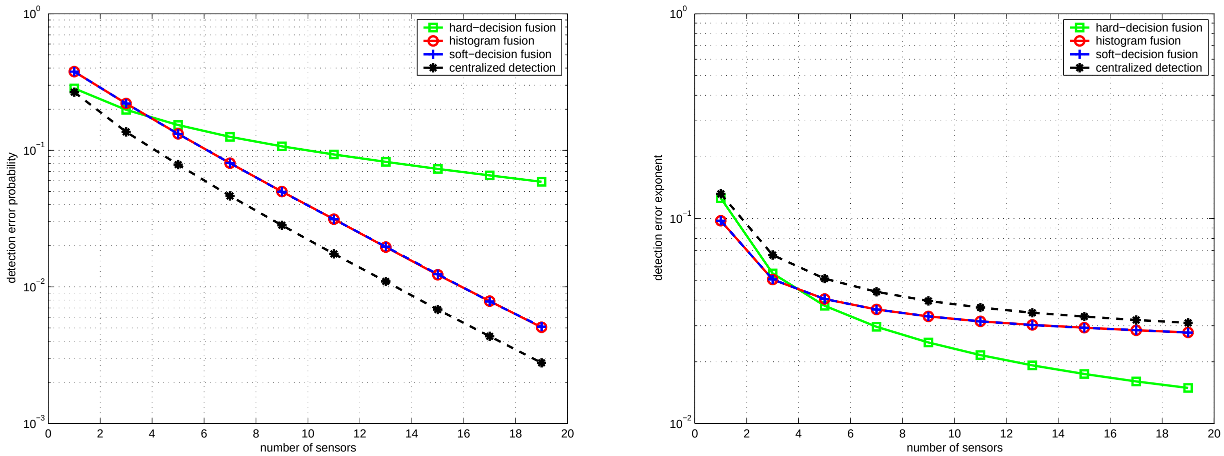
$$E_d^{hist} = \lim_{k\to\infty} -\frac{1}{kn}\log P_e = E_c. \tag{27}$$

5 Numerical Results

In this section, we illlustrate the performance of histogram fusion numerically and compare it to three other schemes: soft-decision fusion, hard-decision fusion and the centralized benchmark. Observation data are generated according to binary Bernoulli distribution with $P(1) = \theta$, whose mean θ is the parameter associated with two competing hypotheses: $\theta_0 = 0.6$ and $\theta_1 = 0.4$ are used in all simulations. The temporal data length is $n = 10$ in all examples. Fig. 3(a) illustrates the performance of the four distributed detection schemes for the MAC under the IPC. The DEP decays exponentially with the number of sensors (k) in all cases. As evident from the figure, the performance of histogram fusion is identical to that of soft-decision fusion since the observations are inherently discrete-valued, confirming our analysis that histogram fusion (using types) is essentially an efficient implementation of soft-decision fusion.

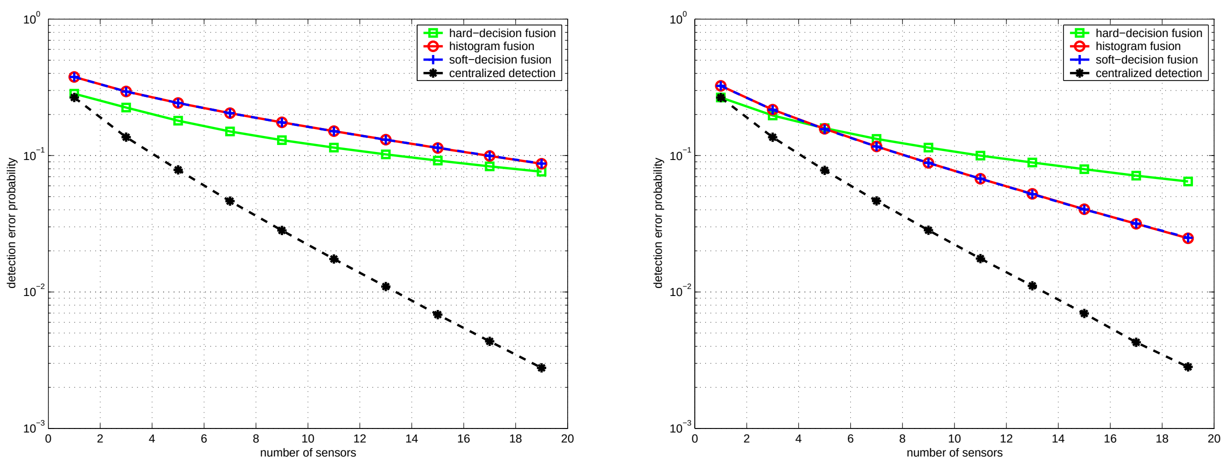
More importantly, the figure shows the optimality of histogram and soft-decision fusion over the MAC: they both achieve the same error exponent (slope of DEP curve) as the centralized benchmark. This is confirmed in Fig. 3(b), where the error exponent for the four detectors is plotted. Hard-decision fusion incurs a loss in error exponent.

Detection performance degrades under the more stringent TPC, as shown in Fig. 4. Histogram fusion and soft-decision fusion no longer achieve the optimal centralized exponent, but they still result in exponential DEP decay with the number of sensors. Comparing Figs. 4(a) and (b), we can see that hard-decision fusion outperforms histogram/soft-decision fusion at low SNR (5 dB in Fig. 4(a)), while the opposite is true at high SNR (10 dB in Fig. 4(b)). This is because in very low SNR, the binary quantization in hard decision fusion offers better resilience against channel noise. Such a phenomenon was also reported in [1] in terms of a cross-over in the performance of binary and analog sensor mappings as a function of SNR.



(a) (b)

Figure 3: The performance of various detectors for the MAC under the IPC; $P_{ind} = 5$dB. (a) DEP as a function of the number of sensors (k). (b) Error exponent of DEP as a function of k.



(a) (b)

Figure 4: DEP as a function of k for the MAC under the TPC. (a) $P_{tot} = 5$dB. (b) $P_{tot} = 10$dB

6 Conclusions

The results in this paper show the optimality of the MAC in the context of joint source-channel communication for distributed detection. The MAC is inherently matched to the distributed detection problem because the computation of the joint decision statistic is automatically accomplished by the channel itself. Furthermore, *uncoded* transmission of local statistics (type or LLR) is optimal. In [12], the same is shown to be optimal in the context of distributed estimation of a single noisy Gaussian source. Similarly, in [13], uncoded PPM transmission (a special case of type transmission) over the MAC is shown to achieve the centralized Cramer-Rao bound in parameter estimation problems. All these results are manifestations of a general *source-channel matching principle*: the MAC is naturally matched to the single source problem and uncoded transmission of appropriate statistics is optimum.

The optimality of histogram fusion, coupled with the use of dumb sensors, makes type-based decentralized detection a very promising strategy for large-scale, energy- and cost-constrained wireless sensor networks. Furthermore, in view of the zero-rate property of types, under relaxed latency requirements, type-based detection can be performed as a background process, consuming a vanishing amount of network bandwidth. However, the synchronization requirements implicit in this scheme (as in soft decision-fusion) need further investigation.

References

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

[2] A. DĆosta, V. Ramachandran, and A. M. Sayeed, "Distributed classification of gaussian space-time sources in wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1026–1036, Aug. 2004.

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

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

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

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

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

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

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

[10] K. Liu and A. M. Sayeed, "Type-based distributed detection in wireless sensor networks," *to be submitted to the IEEE Trans. Inf. Th.*, 2004.

[11] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

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

[13] G. Mergen and L. Tong, "Estimation over multiaccess channels," submitted toIEEE Trans. Signal Processing, July 2004.