
Error Exponents for Decentralized Detection in Tree Networks *

Wee Peng Tay and John N. Tsitsiklis

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA, USA
{wptay, jnt}@mit.edu

Summary. We consider the problem of decentralized detection in a sensor network consisting of nodes arranged as a tree of bounded height. We characterize the optimal error exponent under a Neyman-Pearson formulation, and show that the Type II error probability decays exponentially with the number of nodes. Surprisingly, the optimal error exponent is often the same as that corresponding to a parallel configuration. We provide sufficient, as well as necessary, conditions for this to happen. We also consider the impact of failure-prone sensors or unreliable communications between sensors on the detection performance. Simple strategies that nearly achieve the asymptotically optimal performance in these cases are also developed.

1 Introduction

Consider a set of sensors, one of them designated as the fusion center. We are given two hypotheses H_0 and H_1 , with associated probability spaces. In this paper, we consider only simple hypothesis testing, i.e., the probability measures under both hypotheses are known to the network. The goal of the network is to make a decision on the true hypothesis based on information provided by observations made at each sensor node. This is commonly known as the decentralized detection problem. Decentralized detection in sensor networks has attracted a lot of interest in recent years, because of new technologies (especially, the availability of low-cost sensing devices) and numerous potential applications. The decentralized detection problem was first formulated and studied by [3], which considers a “parallel configuration” whereby each sensor makes an observation and sends a quantized version of that observation to a fusion center. The goal is to make a decision on the two possible hypotheses, based on the messages received at the fusion center. The main

* This paper is an abridged version of [1] and [2], and an overview of the results therein. This research was supported, in part, by the National Science Foundation under contracts ECCS-0701623 and ECS-0426453.

difference between this scenario and the classical *centralized* decision system is that the fusion center has no access to the raw observation made at each sensor. Rather, a sensor transmits a summary of its observation via a transmission function to the fusion center. (When the outputs of the transmission functions are restricted to a finite alphabet, these are known as quantizers.) The network aims to minimize the probability of error or some other cost function at the fusion center, by choosing optimal transmission functions and fusion rules. Various properties and variants of the decentralized detection problem in a parallel configuration have been extensively studied over the last twenty-five years; examples include the following: [4–8] study the properties of optimal fusion rules and quantizers at sensor nodes; [9] shows the existence of optimal strategies, and proves that likelihood ratio quantizers are optimal for a large class of problems including the decentralized detection problem; and [10–14] consider constrained decentralized detection. The reader is referred to [15, 16] for a survey of the work done in this area.

We are interested in networks operating in a regime of limited communication capabilities. Our focus on this regime reflects an emphasis on networks consisting of many, small, and inexpensive devices that have limited battery life and power, and cannot afford to communicate frequently or to transmit a lot of data. Indeed, with abundant communication capabilities, the sensors could just share all their measurements, in which case the network aspects become immaterial, and we are faced with much easier, classical, centralized information processing problems.

Suppose we have n sensors dispersed in a large geographical region. If the sensors are organized in a parallel configuration (cf. the left-hand side in Figure 1), some of the sensors may have to communicate to a far away receiver. The energy expended for communicating can be reduced significantly if the sensors are organized in an in-tree architecture, as in the right-hand side in Figure 1, with sensors sending their messages first to an intermediate aggregator. Moreover, all sensors, except the aggregators, in the right-hand side figure expend approximately the same amount of energy if each uses the same transmission function. This ensures that the lifetimes of the sensors are uniform geographically (aggregators can be special nodes that have a larger energy supply).

Special cases of the tandem configuration (sensors arranged in a serial network), and some specific tree configurations have been studied in [17], where it is shown that it is optimal for sensor nodes to employ likelihood ratio quantizers. Tree configurations are also discussed in [6, 18–23]. However, the exact form of optimal strategies in tree configurations is difficult to derive. Nevertheless, to obtain necessary conditions for the optimal transmission functions or fusion rule, one can analyze the problem using a person-by-person (PBP) optimality approach. In such an analysis, all nodes' transmission functions, except for one particular node, are fixed. Then, one can derive the form of the optimal transmission function at that particular node. Under a conditional independence assumption, and when transmission functions are quantizers,

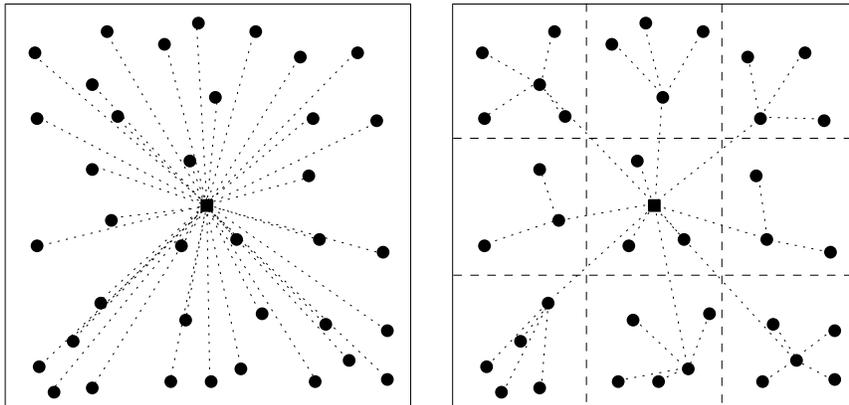


Fig. 1. Two alternative architectures of a geographically dispersed sensor network. The architecture on the left is known as a parallel configuration, while that on the right is a tree configuration.

typical results show that likelihood ratio quantizers are PBP optimal. However, finding optimal quantizer thresholds requires the solution of a nonlinear system of equations, with as many equations as there are thresholds. Closed form formulae for the optimal thresholds used in the likelihood ratio quantizers at each node are known only for trees with a small number of nodes; e.g. [6] considers configurations with at most four sensor nodes. Therefore, characterizing the overall performance is hard, even for networks of moderate size.

Because of these difficulties, to obtain useful insights into the detection performance of large scale networks, we resort to asymptotics. In the Neyman-Pearson framework, one can focus on minimizing the error exponent

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta_n, \quad (1)$$

where β_n is the Type II error probability at the fusion center and n is the number of sensors, while keeping the Type I error probability less than some given threshold. Suppose that $f_i(n)$, $i = 1, \dots, N$, are functions taking positive values. Then, it is easy to see that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^N f_i(n) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \max_{1 \leq i \leq N} f_i(n).$$

Therefore, by minimizing the error exponent (1), we also approximately minimize the probability of a “dominant” error event (one can think of a Type II error event as the disjoint union of several smaller error events, one of which has the highest probability of occurrence). In the asymptotic regime

of large n , an efficient system design strives to minimize the probability of the most likely of the “rare” error events, and this motivates a design that minimizes (1). As we will see, studying the error exponent also makes the problem tractable, and produces elegant results that provide useful insights into the original problem.

The rest of this paper is organized as follows. We describe the basic model, and introduce some concepts and notations in Section 2. In Section 3, we characterize the performance of the parallel configuration, then in Section 4, we consider more general tree networks with bounded height. We study the impact of sensor failures and unreliable communications on the detection performance in Section 5. Finally, we offer some concluding remarks in Section 6.

2 The Basic Model

We now introduce the basic model, notations and assumptions made in this paper. We consider a decentralized binary detection problem involving $n - 1$ sensors and a fusion center; we will be interested in the case when n increases to infinity. We are given two probability spaces $(\Omega, \mathcal{F}, \mathbb{P}_0)$ and $(\Omega, \mathcal{F}, \mathbb{P}_1)$, associated with two hypotheses H_0 and H_1 . We use \mathbb{E}_j to denote the expectation operator with respect to \mathbb{P}_j . Each node v observes a random variable X_v taking values in some set \mathcal{X} . Under either hypothesis H_j , $j = 0, 1$, the random variables X_v are i.i.d., with marginal distribution $\mathbb{P}_j^{\mathcal{X}}$.

Our main goal is to characterize the optimal performance (over all transmission strategies) of a tree network of sensors, under an appropriate asymptotic performance criterion. We consider tree configurations with a bounded height h . The parallel configuration is a special case of a tree, with height $h = 1$. We first define formally a tree network below.

2.1 Tree Networks

We use a directed tree $T_n = (V_n, E_n)$ to represent the sensor network. Here, V_n is the set of nodes, of cardinality n , and E_n is the set of directed arcs of the tree. One of the nodes (the “root”) represents the fusion center, and the remaining $n - 1$ nodes represent the remaining sensors. We use the special symbol f to denote the root of T_n . The arcs in E_n are oriented so that they all point towards the root or the fusion center.

A node u is a *predecessor* of node v if there exists a directed path from u to v . In this case, we also say that v is a *successor* of u . An *immediate predecessor* of node v is a node u such that $(u, v) \in E_n$. An immediate successor is similarly defined. Let the set of immediate predecessors of v be $C_n(v)$. If v is a leaf node, $C_n(v)$ is naturally defined to be empty. The *length* of a path is defined as the number of arcs in the path. The *height* of the tree T_n is the length of the longest path from a leaf to the root, and will be denoted by h_n .

Since we are interested in asymptotically large values of n , we consider a *sequence* of trees $(T_n)_{n \geq 1}$. While we could think of the sequence as representing the evolution of the network as sensors are added, we do not require the sequence E_n to be an increasing sequence of sets; thus, the addition of a new sensor to T_n may result in some arcs being deleted and some new arcs being added. We define the height of a sequence of trees to be $h = \sup_{n \geq 1} h_n$. We are interested in tree sequences of bounded height, i.e., $h < \infty$. For a tree with height h , we say that a node is at *level* k if it is connected to the fusion center via a path of length $h - k$. Hence the fusion center f is at level h .

Let $l_n(v)$ be the number of leaves of the sub-tree rooted at the node v . Thus, $l_n(f)$ is the total number of leaves.

2.2 Strategies

Consider a node v , other than the fusion center f . The node v receives messages $Y_{u,n}$ from its immediate predecessors $u \in C_n(v)$. Because of capacity or other cost constraints, node v can only transmit a summary of its received messages and its own observation X_v , if any. It uses a transmission function γ_v to form its message $Y_{v,n} = \gamma_v(X_v, \{Y_{u,n} : u \in C_n(v)\})$. Let all messages be symbols in a fixed alphabet \mathcal{T} . Thus, if the number of immediate predecessors of v is $|C_n(v)| = d$, then the transmission function γ_v maps $\mathcal{X} \times \mathcal{T}^d$ to \mathcal{T} . We also assume that for each $d \geq 0$, we are given a set of transmission functions $\Gamma(d)$ that the sensor v can choose from. For convenience, we denote $\Gamma(0)$ by Γ . This is the set of transmission functions available to leaf nodes. We assume that all transmissions are perfectly reliable, unless there is a statement to the contrary.

The role of the fusion center f is to make a decision on the true hypothesis, based on the messages it receives from its immediate predecessors. Suppose that it has d immediate predecessors. Recall that in centralized Neyman-Pearson detection, randomization can reduce the Type II error probability. Hence, we assume that the fusion center has access to a random variable, which is uniformly distributed in $[0, 1]$, and independent of everything else. The fusion center uses a randomized fusion rule $\gamma_f : \mathcal{T}^d \times [0, 1] \mapsto \{0, 1\}$ to make a decision. Let $Y_{f,n}$ be a binary-valued random variable indicating the decision of the fusion center.

A *strategy* $\gamma^{(n)}$ consists of a collection of transmission functions, one for each sensor, and a fusion rule for the fusion center. Strategies in which only the leaves make observations will be of special interest to us. In such a scenario, every other node v simply fuses the messages it has received, and forwards a message $Y_{v,n} = \gamma_v(\{Y_{u,n} : u \in C_n(v)\})$ to its immediate successor. We call a strategy of this type a *relay strategy*. A tree network in which we restrict to relay strategies will be called a *relay tree*. Finally, in a relay tree, nodes other than the root and the leaves will be called *relay nodes*.

2.3 Neyman-Pearson Hypothesis Testing

In Neyman-Pearson hypothesis testing, we require that the Type I error probability $\mathbb{P}_0(Y_{f,n} = 1)$ be no more than a given $\alpha \in (0, 1)$. A strategy $\gamma^{(n)}$ is said to be *admissible* if it meets this constraint. We define $\beta^*(T_n)$ as the infimum of $\mathbb{P}_1(Y_{f,n} = 0)$, over all admissible strategies. Similarly, we define $\beta_R^*(T_n)$ as the infimum of $\mathbb{P}_1(Y_{f,n} = 0)$, over all admissible relay strategies. Typically, $\beta^*(T_n)$ or $\beta_R^*(T_n)$ will converge to zero as $n \rightarrow \infty$. Our goal is to determine if such convergence takes place exponentially fast, and to characterize the Type II error exponent, defined by

$$g^* = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \beta^*(T_n), \quad g_R^* = \limsup_{n \rightarrow \infty} \frac{1}{l_n(f)} \log \beta_R^*(T_n).$$

For a relay tree, g_R^* is defined using $l_n(f)$ instead of n due to the fact that only the leaves make observations. Therefore, g_R^* measures the rate of error decay per observation. In the case of a parallel configuration, we use the special notation g_P^* to denote the error exponent.

For any $\gamma \in \Gamma$, let $\mathbb{P}_j^\gamma = \mathbb{P}_j^X \circ \gamma^{-1}$ be the probability law of $\gamma(X)$, and let the Kullback-Leibler divergences be

$$\bar{x}_{0,\gamma} = \mathbb{E}_0 \left[\log \frac{d\mathbb{P}_1^\gamma}{d\mathbb{P}_0^\gamma} \right], \quad \bar{x}_{1,\gamma} = \mathbb{E}_1 \left[\log \frac{d\mathbb{P}_1^\gamma}{d\mathbb{P}_0^\gamma} \right].$$

It is well known that $\bar{x}_{0,\gamma} \leq 0 \leq \bar{x}_{1,\gamma}$ [24]. Moreover, both inequalities are strict as long as the measures \mathbb{P}_j^γ are not indistinguishable. We will make the following assumptions throughout this paper. However, for our results to hold, Assumption 2 can be weakened somewhat; see [25].

Assumption 1 *The measures \mathbb{P}_0^X and \mathbb{P}_1^X are equivalent, i.e., they are absolutely continuous w.r.t. each other. Furthermore, there exists some $\gamma \in \Gamma$ such that $\bar{x}_{0,\gamma} < 0 < \bar{x}_{1,\gamma}$.*

Assumption 2 $\mathbb{E}_0 \left[\log^2 \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X} \right] < \infty$.

Suppose that the node v sends a message $Y_{v,n} = y$ to its immediate successor. Let the log-likelihood ratio of the message sent by v be

$$\mathcal{L}_{v,n}(y) = \log \frac{d\mathbb{P}_{1,n}^{(v)}}{d\mathbb{P}_{0,n}^{(v)}}(y),$$

where $d\mathbb{P}_{1,n}^{(v)}/d\mathbb{P}_{0,n}^{(v)}$ is the Radon-Nikodym derivative of the distribution of $Y_{v,n}$ under H_1 w.r.t. the distribution under H_0 . If the transmission alphabet \mathcal{T} is a discrete set, then this is just the ratio

$$\log \frac{\mathbb{P}_1(Y_{v,n} = y)}{\mathbb{P}_0(Y_{v,n} = y)}.$$

We will make extensive use of the following class of transmission functions. Recall that $l_n(v)$ is the number of leaves of the sub-tree rooted at node v .

Definition 1 A (1-bit) Log-Likelihood Ratio Quantizer (LLRQ) with threshold t for a non-leaf node v , with $|C_n(v)| = d$ immediate predecessors, is a binary-valued function on \mathcal{T}^d , defined by

$$\text{LLRQ}_{d,t}(\{y_u : u \in C_n(v)\}) = \begin{cases} 0, & \text{if } x \leq t, \\ 1, & \text{if } x > t, \end{cases}$$

where

$$x = \frac{1}{l_n(v)} \sum_{u \in C_n(v)} \mathcal{L}_{u,n}(y_u).$$

Note that if a node v uses a LLRQ, it ignores its own observation X_v and acts as a relay. If all non-leaf nodes use a LLRQ, we have a special case of a relay strategy. We assume that LLRQs are available choices of transmission functions for all non-leaf nodes. As we will see, LLRQs will play an important role in our results.

Assumption 3 For all $t \in \mathbb{R}$ and $d > 0$, $\text{LLRQ}_{d,t} \in \Gamma(d)$.

For simplicity, we define the sum of the log-likelihood ratios of the received messages at node v , as follows:

$$S_n(v) = \sum_{u \in C_n(v)} \mathcal{L}_{v,n}(Y_{v,n}).$$

3 The Parallel Configuration

We consider here the special case of a network with a parallel configuration. The following proposition shows that the Type II error probability falls exponentially fast with the number of nodes n . Moreover, an asymptotically optimal strategy consists of using identical transmission functions for each sensor. A proof can be found in [26].

Proposition 1 If Assumptions 1-2 hold, then

$$g_p^* = \inf_{\gamma \in \Gamma} \bar{x}_{0,\gamma}.$$

Moreover, the error exponent stays the same if we restrict all sensors to using the same transmission function.

The quantity $-\bar{x}_{0,\gamma}$ may be recognized as the Kullback-Leibler divergence, which measures the “discrimination” between the two probability measures \mathbb{P}_0^γ and \mathbb{P}_1^γ . Hence, the asymptotically optimal strategy is to choose a transmission function that produces the greatest discrimination between the two hypotheses.

4 Tree Architectures

In this section, we consider general tree networks with a bounded height. Our objective is to study g^* and g_R^* for different sequences of trees. Since the fusion center of a parallel configuration can simulate a relay network (by carrying out all the operations taking place at each relay node, internally in the fusion center), we have

$$g_P^* \leq g_R^*. \quad (2)$$

Recall that $l_n(f)$ is the number of leaves in the network. Let

$$z = \liminf_{n \rightarrow \infty} \frac{l_n(f)}{n}$$

be the proportion of nodes that are leaves. By comparing the performance to a centralized system where all raw observations are transmitted directly to the fusion center, and using a similar argument as above, we obtain the first inequality in the expression below,

$$\bar{x}_0 = \mathbb{E}_0 \left[\log \frac{d\mathbb{P}_1^X}{d\mathbb{P}_0^X} \right] \leq g^* \leq z g_R^*. \quad (3)$$

The second inequality follows because an optimal strategy is at least as good as an optimal relay strategy; the factor z arises because we have normalized g_R^* by $l_n(f)$ instead of n .

In the following, we provide a method to propagate error bounds along a tree network, and derive upper bounds similar to that in Cramér's Theorem for the parallel configuration [27]. We consider specifically a h -uniform tree, defined as follows.

Definition 2 (h -uniform tree) *A tree T_n is said to be h -uniform if the length of every path from a leaf to the root is exactly h . A sequence of trees $(T_n)_{n \geq 1}$ is said to be h -uniform if there exists some $n_0 < \infty$, so that for all $n \geq n_0$, T_n is h -uniform.*

It turns out that it is easier to work with h -uniform trees, and as shown on page 11, a height uniformization procedure can be performed on any given tree. Moreover, the detection performance of this height uniformized tree cannot be better than the original tree.

4.1 Error Bounds for h -Uniform Relay Trees

We consider the special case of a 1-bit h -uniform relay tree, in which all relay nodes at level k use a LLRQ with a common threshold t_k . Let $t^{(k)} = (t_1, t_2, \dots, t_k)$, for $k \geq 1$, and $t^{(0)} = \emptyset$. For $j = 0, 1$, $k \geq 1$, and $\lambda \in \mathbb{R}$, we define recursively

$$\Lambda_{j,0}(\gamma; \lambda) = \Lambda_{j,0}(\gamma, \emptyset; \lambda) = \log \mathbb{E}_j \left[\left(\frac{d\mathbb{P}_1^\gamma}{d\mathbb{P}_0^\gamma} \right)^\lambda \right],$$

$$\Lambda_{j,k}^*(\gamma, t^{(k)}) = \sup_{\lambda \in \mathbb{R}} \{ \lambda t_k - \Lambda_{j,k-1}(\gamma, t^{(k-1)}; \lambda) \}, \quad (4)$$

$$\Lambda_{j,k}(\gamma, t^{(k)}; \lambda) = \max \{ -\Lambda_{1,k}^*(\gamma, t^{(k)})(j + \lambda), \Lambda_{0,k}^*(\gamma, t^{(k)})(j - 1 + \lambda) \}. \quad (5)$$

Here, $\Lambda_{j,k}^*(\gamma, t^{(k)})$ is the Fenchel-Legendre transform of $\Lambda_{j,k-1}$ [27], and can be visualized as in Figure 2. We will be interested in the case where

$$\bar{x}_{0,\gamma} < 0 < \bar{x}_{1,\gamma}, \quad (6)$$

$$t_1 \in (\bar{x}_{0,\gamma}, \bar{x}_{1,\gamma}), \quad (7)$$

$$t_k \in (-\Lambda_{1,k-1}^*(\gamma, t^{(k-1)}), \Lambda_{0,k-1}^*(\gamma, t^{(k-1)})), \text{ for } 1 < k \leq h. \quad (8)$$

The reader is referred to [1] for an argument that shows the above requirements on the thresholds t_k to be feasible.

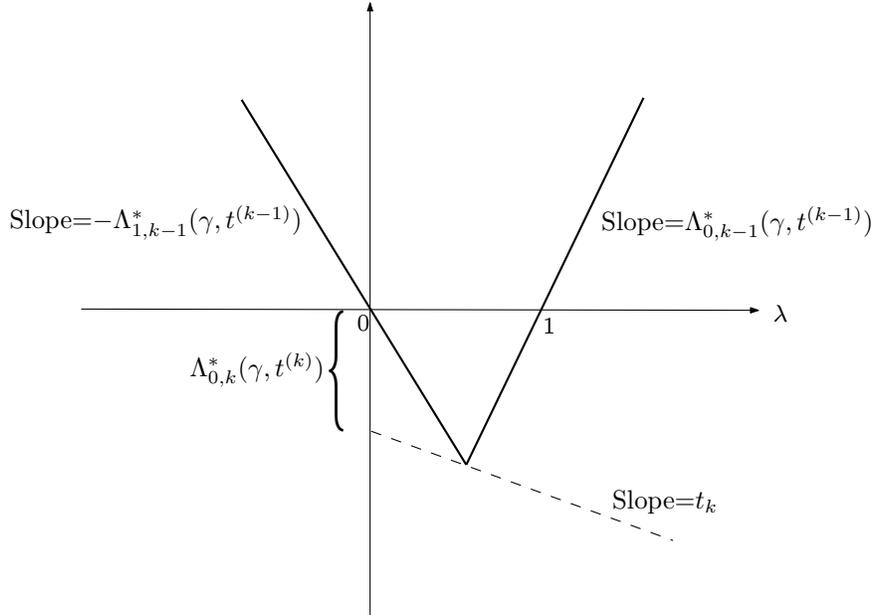


Fig. 2. Typical plot of $\Lambda_{0,k-1}(\gamma, t^{(k-1)}; \lambda)$, $k \geq 2$.

Proposition 2 below shows that the Type I and II error exponents are essentially upper bounded by $-\Lambda_{0,h}^*(\gamma, t^{(h)})$ and $-\Lambda_{1,h}^*(\gamma, t^{(h)})$ respectively. Note that we recover the classical Chernoff bound when the network has height $h = 1$, i.e., the network is a parallel configuration. Let $p_n(v)$ be the

total number of predecessors of v , i.e., the total number of nodes in the subtree rooted at v , not counting v itself. Thus, $p_n(f) = n - 1$. Recall that for a given h -uniform sequence of trees $(T_n)_{n \geq 1}$, there exists a n_0 such that for all $n \geq n_0$, T_n is h -uniform.

Proposition 2 *Fix some $h \geq 1$, and consider an h -uniform sequence of trees. Suppose that Assumptions 1-2 hold. Suppose that, for every n , every leaf node uses the same transmission function $\gamma \in \Gamma$, which satisfies (6), and that every level k node ($k \geq 1$) uses a LLRQ with threshold t_k , satisfying (7)-(8).*

(i) *For all nodes v of level $k \geq 1$ and for all $n \geq n_0$,*

$$\begin{aligned} \frac{1}{l_n(v)} \log \mathbb{P}_1 \left(\frac{S_n(v)}{l_n(v)} \leq t_k \right) &\leq -A_{1,k}^*(\gamma, t^{(k)}) + \frac{p_n(v)}{l_n(v)} - 1, \\ \frac{1}{l_n(v)} \log \mathbb{P}_0 \left(\frac{S_n(v)}{l_n(v)} > t_k \right) &\leq -A_{0,k}^*(\gamma, t^{(k)}) + \frac{p_n(v)}{l_n(v)} - 1. \end{aligned}$$

(ii) *Suppose that for all $n \geq n_0$ and for all level 1 nodes v , we have $l_n(v) \geq N$. Then, for all $n \geq n_0$, we have*

$$\begin{aligned} \frac{1}{l_n(f)} \log \mathbb{P}_1 \left(\frac{S_n(f)}{l_n(f)} \leq t_h \right) &\leq -A_{1,h}^*(\gamma, t^{(h)}) + \frac{h}{N}, \\ \frac{1}{l_n(f)} \log \mathbb{P}_0 \left(\frac{S_n(f)}{l_n(f)} > t_h \right) &\leq -A_{0,h}^*(\gamma, t^{(h)}) + \frac{h}{N}. \end{aligned}$$

4.2 Optimal Error Exponent

The following proposition shows that Type II error probabilities decay exponentially (the error exponents are negative). The bounded height assumption is crucial for this result. Indeed, for the case of a tandem configuration, the error probability seems to decay at a sub-exponential rate [28]. The following is proved in [29].

Proposition 3 *Consider a sequence of trees of height h , and suppose Assumptions 1-3 hold. Then,*

$$-\infty < g_P^* \leq g_R^* < 0 \quad \text{and} \quad -\infty < \bar{x}_0 \leq g^* < 0.$$

From (2), we have $g_P^* \leq g_R^*$, i.e., a relay network performs at best as well as a parallel configuration. Next, we want to know when a relay network has the same error exponent as a parallel configuration. A proof of the following proposition can be found in [29].

Proposition 4 *Consider a sequence of trees of height h in which $z = 1$. Suppose that Assumptions 1-3 hold. Then,*

$$g_P^* = g^* = g_R^*.$$

Furthermore, if the sequence of trees is h -uniform, the optimal error exponent does not change even if we restrict to relay strategies in which every leaf uses the same transmission function and all other nodes use a 1-bit LLRQ with the same threshold.

Proposition 4 is surprising as it establishes that the performance of every network possessing certain qualitative properties is comparable to that of a parallel configuration. This result has important ramifications: suppose that all nodes are restricted to be at most h hops away from the fusion center, then a system designer can reduce the energy consumption (e.g., by employing a h -hop spanning tree that minimizes the overall energy consumption), without losing detection efficiency, as long as the number of nodes n is large, and the proportion of leaf nodes is large. For example, consider the case of nodes uniformly distributed in a square as in Figure 1, and suppose that the cost of transmitting a message from one node to another is proportional to the Euclidean distance between the nodes. It is well known that finding a h -hop constrained Minimum Spanning Tree (MST) is NP-hard (see [30]). However, heuristics that achieve a cost of the same order of magnitude as the h -hop constrained MST can be employed to design a suitable network architecture [31]. This involves dividing the square into suitable sub-squares, and it can be verified that with high probability, as $n \rightarrow \infty$, the network we obtain has the property that $z = 1$.

If a sequence of trees is h -uniform and $z = 1$, it can be shown that the following simple relay strategy ϵ -achieves the optimal error exponent:

1. all leaf nodes transmit with the same transmission function $\gamma \in \Gamma$, such that $\bar{x}_{0,\gamma} \leq g_P^* + \epsilon/2$;
2. all other nodes use a 1-bit LLRQ with the same threshold $t = \bar{x}_{0,\gamma} + \epsilon/2$.

This is a convenient strategy since only leaf nodes need to make observations, while all the rest of the nodes act as relay nodes. Moreover, transmitting only 1 bit is sufficient, and all relay nodes use the same 1-bit LLRQ. This may be useful in situations where the nodes are simple, low-cost devices.

For a general sequence of trees with height h , we can perform a height uniformization procedure to obtain an h -uniform sequence of trees as follows. We let $A_n \subset V_n$ be the set of nodes whose immediate predecessors include leaves of the tree T_n .

Height Uniformization Procedure. Consider a tree $T_n = (V_n, E_n)$ of height h , and a node v that has at least one leaf as an immediate predecessor (i.e., $v \in A_n$). Let D_n be the set of leaves that are immediate predecessors of v , and whose paths to the fusion center f are of length $k < h$. Add $h - k$ nodes, $\{u_j : j = 1, \dots, h - k\}$, to V_n ; remove the edges (u, v) , for all $u \in D_n$; add the edges (u_1, v) , and (u_{j+1}, u_j) , for $j = 1, \dots, h - k - 1$; add the edges (u, u_{h-k}) , for all $u \in D_n$. This procedure is repeated for all $v \in A_n$. The resulting tree is h -uniform. \square

It is clear that any strategy on the height uniform tree can be simulated by a strategy on the original tree. Furthermore, it can be shown that the height uniformization procedure preserves the property that $z = 1$ [29]. As the height uniformized tree sequence cannot perform better than the original tree sequence, using the strategy as described above for h -uniform trees is an ϵ -optimal strategy for the original tree sequence.

Next, we want to consider when the sufficient condition $z = 1$ is also a necessary condition for a relay network to have the same asymptotically optimal performance as a parallel configuration. Non-trivial necessary conditions for the equality $g_R^* = g_P^*$ to hold are, in general, difficult to obtain, because they depend on the nature of the transmission functions available to the sensors. Suppose that sensors are allowed to simply forward undistorted all of the messages that they receive, then the equality $g_R^* = g_P^*$ holds trivially. Hence, we need to impose some restrictions on the set of transmission functions available, as in the assumption that follows. Let B_n be the set of nodes all of whose predecessors are leaves.

Assumption 4

- (a) *There exists a $n_0 \geq 1$ such that for all $n \geq n_0$, we have $l_n(v) > 1$ for all $v \in B_n$.*
- (b) *Let X_1, X_2, \dots be i.i.d. random variables under either hypothesis H_j , each with distribution \mathbb{P}_j^X . For $k > 1$, $\gamma_0 \in \Gamma(k)$, and $\gamma_i \in \Gamma$, $i = 1, \dots, k$, let $\xi = (\gamma_0, \dots, \gamma_k)$. Let ν_j^ξ be the distribution of $\gamma_0(\gamma_1(X_1), \dots, \gamma_k(X_k))$ under hypothesis H_j . We assume that*

$$g_P^* = \inf_{\gamma \in \Gamma} \bar{x}_{0,\gamma} < \inf_{\xi \in \Gamma(k) \times \Gamma^k} \frac{1}{k} \mathbb{E}_0 \left[\log \frac{d\nu_1^\xi}{d\nu_0^\xi} \right], \quad (9)$$

for all $k > 1$.

Assumption 4 holds in most cases of interest. There is no loss of generality in assuming part (a), because if in a relay tree we have $l_n(v) = 1$ for some $v \in B_n$, we can remove the predecessor of v , and treat v as a leaf sensor. As for part (b), it is easy to see that the L.H.S. of (9) is always less than or equal to the R.H.S., hence we have only excluded those cases where (9) holds with equality. We are essentially assuming that when the messages $\gamma_1(X_1), \dots, \gamma_k(X_k)$ are summarized (or quantized) by γ_0 , there is some loss of information, as measured by the associated Kullback-Leibler divergences.

Proposition 5 *Suppose that Assumptions 1-4 hold. Then, $g_R^* = g_P^*$ iff $z = 1$.*

5 Unreliable Networks

So far, we have assumed that all nodes are error-free, and all communications are reliable. We next study the impact of failure-prone sensors and unreliable communications on the detection performance.

5.1 Sensor Failures

We model the case of sensor failures by using a random number of nodes. Variants of the decentralized detection problem with a random number of nodes distributed in a parallel configuration have been studied in [32–34]. In [32] and [34], the authors consider the case of spatially correlated signals, and analyze the detection performance of a simple but suboptimal strategy. In [33], the objective is not to find an optimal transmission strategy. Rather, the authors assume that nodes in a parallel configuration make i.i.d. observations under either hypothesis, quantize their observations using some known quantizer that is identical for all nodes, and use a special multiple access protocol called Type-Based Random Access (TBRA) (in all problems so far, we have implicitly assumed some sort of orthogonal multiple access protocol in which messages from different nodes do not corrupt one another, whereas in TBRA, messages are combined additively over the transmission medium). In this section, our goal is to characterize the asymptotically optimal performance for tree networks with bounded height, and to develop an optimal transmission strategy, assuming i.i.d. observations and the usual orthogonal multiple access protocol. The results in this section are a summary of those in [2].

To model sensor failures, we construct a random tree as in a Galton-Watson process, but with a limited ‘time span’ of h (which corresponds to our tree having a height h). We start with the fusion center f , and let the number of immediate predecessors of f be a random variable $N_f = |C(f)|$, with distribution law μ_h . Then, we let each node v in the random set $C(f)$ have $N_v = |C(v)|$ immediate predecessors, where N_v has marginal law μ_{h-1} . We continue this process until the level 0 nodes are reached. Hence, each level k node v (with $k \geq 1$) has N_v immediate predecessors, where N_v is a random variable with law μ_k . Furthermore, we also assume that all these random variables are independent, and independent of the hypothesis. We call such a random tree a GW-tree.

We want to model a dense network, therefore we consider the case when each of the laws μ_k , $k = 1, \dots, h$, has asymptotically large mean. Let λ_k be the mean of μ_k . We let $\lambda^* = \min_{1 \leq k \leq h} \lambda_k$ increase to infinity, by allowing the laws μ_k to vary accordingly. However, we require that the distributions satisfy the following assumption.

Assumption 5 *Let \tilde{N}_k be random variables with distribution μ_k , $k = 1, \dots, h$. We have*

$$\mathbb{E}[\tilde{N}_k^2] = (1 + o(1))\lambda_k^2, \quad (10)$$

where $o(1)$ stands for a term that goes to 0 as $\lambda_k \rightarrow \infty$.

It is easy to check that both the Poisson distribution and Binomial distribution satisfy the above assumption. Under Assumption 5, Chebychev’s inequality shows that the distribution of \tilde{N}_k is clustered around its mean.

Lemma 1 For all $\eta > 0$, $\mathbb{P}(|\tilde{N}_k/\lambda_k - 1| > \eta) \rightarrow 0$ as $\lambda_k \rightarrow \infty$.

Let β_{GW}^* be the infimum of the Type II error probability, with the minimization taken over all strategies (to be more precise, we have to consider a family of strategies, see [2]), subject to the constraint that the Type I error probability is not more than $\alpha \in (0, 1)$. Our goal is to characterize the optimal error exponent

$$\limsup_{\lambda^* \rightarrow \infty} \frac{1}{\lambda(h)} \log \beta_{GW}^*,$$

where $\lambda(h) = \prod_{k=1}^h \lambda_k$ is the expected number of nodes.

Let the log-likelihood ratio of the received messages at v be S_v . Motivated by the ϵ -optimal strategies for non-random tree networks, it is natural to define the following class of transmission policies.

Definition 3 A transmission function for a level k node v is called a Mean-normalized Log-Likelihood Ratio (MLLR) quantizer at level k with threshold t if

$$Y_v = \begin{cases} 0, & \text{if } S_v/\lambda(k) \leq t, \\ 1, & \text{otherwise.} \end{cases}$$

Again, we assume that MLLR quantizers are valid quantizers for each node.

Assumption 6 Every node of level $k \geq 1$ has access to MLLR quantizers.

In the case where there are no sensor failures, i.e., $N_v = \lambda_k$ a.s. for all non-leaf nodes v , Proposition 4 shows that the Type II error probability decays exponentially fast with $\lambda(h)$, at rate g_P^* . The proposition below shows that this remains true for a GW-tree.

Proposition 6 Suppose that Assumptions 1, 2, 5 and 6 hold. Then, for all $\alpha \in (0, 1)$, the optimal error exponent of a GW-tree of height h is given by

$$\lim_{\lambda^* \rightarrow \infty} \frac{1}{\lambda(h)} \log \beta_{GW}^* = g_P^*. \quad (11)$$

Furthermore, for any $\epsilon \in (0, -g_P^*)$, and any large enough λ^* , the following strategy satisfies the Type I error probability constraint, and its error exponent is bounded above by $g_P^* + \epsilon$:

- (i) each leaf uses the same transmission function $\gamma \in \Gamma$, with $\bar{x}_{0,\gamma} \leq g_P^* + \epsilon/2 < 0$; and
- (ii) for $k \geq 1$, every level k node uses a MLLR quantizer with threshold $t_k = \bar{x}_{0,\gamma} + \epsilon/2^{h-k+1}$.

Consider the scenario where each node in the network may fail, independently, with some probability p . A network of n failure-prone sensors cannot be better (on the average) than a network of np failure-proof sensors. However, in the limit of large n , Proposition 6 shows that the asymptotically optimal performance of both networks is essentially the same.

5.2 Unreliable Communications

We now consider the case where each sensor in a tree of height h is constrained to sending one bit to its immediate successor, and the channel between any two nodes is a binary symmetric channel (BSC) with known crossover probability $\eta \in (0, 1/2)$. Suppose that $(T_n)_{n \geq 1}$ is a h -uniform tree sequence. For simplicity, we discuss only relay strategies in this section. For every non-leaf node v , we assume that $|C_n(v)| \geq c_n$, for some $c_n \rightarrow \infty$ as $n \rightarrow \infty$. This models a dense sensor network with bounded height h . We again consider the problem of minimizing the Type II error exponent, with the Type I error constrained to be no larger than $\alpha \in (0, 1)$. However, in this case, the appropriate error exponent to consider, as we will see later in Proposition 8, is

$$\limsup_{n \rightarrow \infty} \frac{1}{|C_n(f)|} \log \beta^*(T_n),$$

where we have normalized the error exponent by $|C_n(f)|$, the number of immediate predecessors of the fusion center, instead of the total number of nodes n .

Consider an immediate predecessor u of the node v . The node u transmits a 1-bit message $Y_{u,n} = y_u$ over the BSC to node v . Let $\bar{Y}_{u,n} = \bar{y}_u$ be the received message at node v . We now define LLRQs as in Definition 1, but with respect to the received messages $\bar{Y}_{u,n}$, i.e., the likelihood ratios are replaced with the Radon-Nikodym derivative of the distribution of $\bar{Y}_{u,n}$ under H_1 with respect to that under H_0 .

Let us first consider the simple case when $h = 1$, i.e., the parallel configuration. For each $\gamma \in \Gamma$ and $j = 0, 1$, we define the probability measures \mathbb{Q}_j^γ on the space $\{0, 1\}$ as follows. Let

$$\begin{aligned} \mathbb{Q}_j^\gamma(\{0\}) &= (1 - \eta)\mathbb{P}_j^\gamma(\{0\}) + \eta\mathbb{P}_j^\gamma(\{1\}), \\ \mathbb{Q}_j^\gamma(\{1\}) &= (1 - \eta)\mathbb{P}_j^\gamma(\{1\}) + \eta\mathbb{P}_j^\gamma(\{0\}). \end{aligned}$$

Then, the following proposition is a consequence of Proposition 1.

Proposition 7 *Suppose that Assumptions 1-3 hold. For $h = 1$, and all $\alpha \in (0, 1)$, the optimal error exponent is*

$$\lim_{n \rightarrow \infty} \frac{1}{|C_n(f)|} \log \beta^*(T_n) = \inf_{\gamma \in \Gamma} \mathbb{E}^{\mathbb{Q}_0^\gamma} \left[\log \frac{d\mathbb{Q}_1^\gamma}{d\mathbb{Q}_0^\gamma} \right],$$

where $\mathbb{E}^{\mathbb{Q}_0^\gamma}$ is the expectation operator under \mathbb{Q}_0^γ . Furthermore, there is no loss in optimality if we restrict all the leaf nodes in $C_n(f)$ to using the same transmission function $\gamma \in \Gamma$.

The optimal error exponent in the case $h \geq 2$ is markedly different from that in Proposition 7. Let $\text{Bern}(\eta)$ denote the Bernoulli distribution on $\{0, 1\}$ that takes the value 1 with probability η . The following proposition is proved in [2].

Proposition 8 *Suppose that Assumptions 1-3 hold. For $h \geq 2$, and for all $\alpha \in (0, 1)$, the optimal error exponent is*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{|C_n(f)|} \log \beta^*(T_n) &= -\left(\eta \log \frac{\eta}{1-\eta} + (1-\eta) \log \frac{1-\eta}{\eta}\right) \\ &= -D(\eta) < 0, \end{aligned} \tag{12}$$

where $D(\eta)$ is the Kullback-Leibler divergence function of $\text{Bern}(1-\eta)$ w.r.t. $\text{Bern}(\eta)$.

Again, it can be shown that a strategy that ϵ -achieves the optimal error exponent is the following:

- (i) All leaves use the same transmission function $\gamma \in \Gamma$, where γ is chosen so that $\mathbb{P}_0(\gamma(X) = 0) \neq \mathbb{P}_1(\gamma(X) = 0)$.
- (ii) Every level 1 node uses a LLR with threshold 0.
- (iii) All other nodes use the majority rule: send a 1 if and only if more than half of the received messages are equal to 1.
- (iv) The fusion center uses a LLR with threshold $t = -D(\eta) + \epsilon$.

Compared to the result in Proposition 3, the above proposition shows that the detection performance of a h -uniform relay tree network of height at least 2, in the presence of unreliable communications, is significantly worse than that of a similar network with reliable communications. Indeed, in the case of unreliable communications, the error probability decays exponentially fast with $|C_n(f)|$, instead of n .

6 Conclusion

We have considered Neyman-Pearson decentralized detection in sensor networks with tree architectures of bounded height. Although the problem of finding exact optimal strategies, and hence of characterizing the optimal detection performance, for a fixed number of nodes n is computationally intractable, the asymptotically optimal performance is surprisingly the same as the well-known detection performance of the parallel configuration, under most practical cases of interest. Indeed, when the leaf nodes dominate, a tree network with bounded height has the same error exponent as a parallel configuration. Under a mild condition, the property that leaf nodes dominate

the network is also shown to be a necessary condition for achieving the same optimal error exponent as the parallel configuration.

We have only considered the Neyman-Pearson criterion in this paper. A similar analysis is possible within a Bayesian framework. It is shown in [35] that error probabilities also decay exponentially fast in a Bayesian setting. However, the same performance as the parallel configuration can no longer be achieved.

We also discussed the impact of sensor failures and unreliable communications on detection performance. Our results provide a useful insight into the performance of height uniform tree sequences. Suppose that all nodes can only send 1-bit messages, and every node, except the leaf nodes, has the same number of immediate predecessors c_n . A network can operate in two modes, in a ‘sensor failure’ mode and in an ‘unreliable communications’ mode. In the ‘sensor failure’ mode, if a sensor determines that its message cannot be received reliably by its intended recipient, it remains silent instead of transmitting. In the ‘unreliable communications’ mode, it transmits its 1-bit message regardless. Our results indicate that in the ‘sensor failure’ mode, the error probability decays exponentially with c_n^h , whereas in the ‘unreliable communications’ mode, it decays exponentially with c_n . Therefore, it is better for the network to operate in the ‘sensor failure’ mode, when the height of the tree network is greater than one.

Several issues remain outstanding, and are areas for further research. Our results are valid in the large n regime; however, a significantly larger number of nodes may be needed before a relay network can approximate, in a certain sense, the performance of a parallel configuration. Another issue is that although the error exponents are the same, the ratio $\beta^*(T_n)/\beta_P^*$, where β_P^* is the optimal error probability of the parallel configuration, could be diverging to infinity as n increases. Therefore, it is of interest to study the *exact asymptotics* of this problem. Finally, the case where sensor observations are correlated remains a difficult problem. For the case of correlated sensor observations in parallel configurations, the reader is referred to [36, 37] for recent results.

References

1. W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, “Data fusion trees for detection: Does architecture matter?” in *Proc. Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sep. 2006.
2. —, “Detection in dense wireless sensor networks,” in *Proc. IEEE Wireless Commun. and Networking Conf.*, Hong Kong, Mar. 2007.
3. R. R. Tenney and N. R. Sandell, “Detection with distributed sensors,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, pp. 501–510, 1981.
4. Z. Chair and P. K. Varshney, “Optimal data fusion in multiple sensor detection systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, pp. 98–101, 1986.

5. G. Polychronopoulos and J. N. Tsitsiklis, "Explicit solutions for some simple decentralized detection problems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, pp. 282–292, 1990.
6. A. R. Reibman and L. W. Nolte, "Design and performance comparison of distributed detection networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 789–797, 1987.
7. P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Transactions on Information Theory*, vol. 38, pp. 355–361, Mar. 1992.
8. W. W. Irving and J. N. Tsitsiklis, "Some properties of optimal thresholds in decentralized detection," *IEEE Trans. Autom. Control*, vol. 39, pp. 835–838, 1994.
9. J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, pp. 550–558, 1993.
10. C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, 1996.
11. S. Appadwedula, V. V. Veeravalli, and D. Jones, "Energy-efficient detection in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 693–702, 2005.
12. J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, 2003.
13. —, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Select. Areas Commun., Special Issue on Wireless Sensor Networks*, vol. 22, no. 6, pp. 1007–1015, 2004.
14. W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Censoring sensors: Asymptotics and the value of cooperation," in *Proc. Conf. on Inform. Sci. and Sys.*, Princeton, NJ, Mar. 2006, pp. 62–67.
15. J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
16. R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: part I - fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, 1997.
17. L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. 21st IEEE Conf. Decision Control*, 1982, pp. 686–691.
18. R. Viswanathan, S. C. A. Thomopoulos, and R. Tumuluri, "Optimal serial distributed decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 366–376, 1988.
19. Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: part I- tandem structures," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 21, no. 5, pp. 1044–1059, 1991.
20. —, "Optimization of detection networks: part II- tree structures," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, no. 1, pp. 211–221, 1993.
21. J. D. Papastavrou and M. Athans, "On optimal distributed decision architectures in a hypothesis testing environment," *IEEE Trans. Autom. Control*, vol. 37, no. 8, pp. 1154–1169, 1992.
22. A. Pete, K. Pattipati, and D. Kleinman, "Optimization of detection networks with multiple event structures," *IEEE Trans. Autom. Control*, vol. 39, no. 8, pp. 1702–1707, 1994.

23. S. Alhakeem and P. K. Varshney, "A unified approach to the design of decentralized detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 9–20, 1995.
24. T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY: John Wiley & Sons, Inc., 1991.
25. P.-N. Chen and A. Papamarcou, "New asymptotic results in parallel distributed detection," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1847–1863, Nov. 1993.
26. J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.
27. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY: Springer-Verlag, 1998.
28. W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, "On the sub-exponential decay of detection probabilities in long tandems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Honolulu, HI, Apr. 2007.
29. —, "Data fusion trees for detection: Does architecture matter?" *IEEE Trans. Inf. Theory*, 2006, submitted for publication.
30. M. R. Garey and D. S. Johnson, *A guide to the theory of NP-Completeness*. New York, NY: W. H. Freeman, 1979.
31. A. E. F. Clementi, M. D. Ianni, M. A., L. M., R. G., and S. R., "Divide and conquer is almost optimal for the bounded-hop MST problem on random Euclidean instances," in *Proc. Structural Information and Communication Complexity*, Mont Saint-Michel, France, May 2005, pp. 89–98.
32. R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 4, pp. 462–472, 2005.
33. L. T. Animashree Anandkumar, "A large deviation analysis of detection over multi-access channels with random number of sensors," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 4, 2006.
34. T. Q. S. Quek, D. Dardari, and M. Z. Win, "Energy efficiency of dense wireless sensor networks: to cooperate or not to cooperate," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 459–469, 2007.
35. W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," in *Proc. of Data Compression Conf.*, Snowbird, UT, Mar. 2007.
36. J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.
37. W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," in *Proc. 43rd Allerton Annual Conference on Communication, Control, and Computing*, Sep. 2005.