Decentralized Detection in Resource-limited Sensor Network Architectures

by

Wee Peng Tay

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Author .....
Department of Electrical Engineering and Computer Science
December, 2007

Certified by .....
John N. Tsitsiklis
Clarence J Lebel Professor of Electrical Engineering
Thesis Supervisor

Certified by .....
Moe Z. Win
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by .....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Decentralized Detection in Resource-limited Sensor Network Architectures

by

Wee Peng Tay

Submitted to the Department of Electrical Engineering and Computer Science
on December, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

We consider the problem of decentralized binary detection in a network consisting of a large number of nodes arranged as a tree of bounded height. We show that the error probability decays exponentially fast with the number of nodes under both a Neyman-Pearson criterion and a Bayesian criterion, and provide bounds for the optimal error exponent. Furthermore, we show that under the Neyman-Pearson criterion, the optimal error exponent is often the same as that corresponding to a parallel configuration, implying that a large network can be designed to operate efficiently without significantly affecting the detection performance. We provide sufficient, as well as necessary, conditions for this to happen. For those networks satisfying the sufficient conditions, we propose a simple strategy that nearly achieves the optimal error exponent, and in which all non-leaf nodes need only send 1-bit messages.

We also investigate the impact of node failures and unreliable communications on the detection performance. Node failures are modeled by a Galton-Watson branching process, and binary symmetric channels are assumed for the case of unreliable communications. We characterize the asymptotically optimal detection performance, develop simple strategies that nearly achieve the optimal performance, and compare the performance of the two types of networks. Our results suggest that in a large scale sensor network, it is more important to ensure that nodes can communicate reliably with each other (e.g., by boosting the transmission power) than to ensure that nodes are robust to failures.

In the case of networks with unbounded height, we establish the validity of a long-standing conjecture regarding the sub-exponential decay of Bayesian detection error probabilities in a tandem network. We also provide bounds for the error probability, and show that under the additional assumption of bounded Kullback-Leibler divergences, the error probability is $\Omega(e^{-cn^d})$, for all $d > 1/2$, with c being a positive constant. Furthermore, the bound $\Omega(e^{-c(\log n)^d})$, for all $d > 1$, holds under an additional mild condition on the distributions. This latter bound is shown to be tight. Moreover, for the Neyman-Pearson case, we establish that if the sensors act myopically, the Type II error probabilities also decay at a sub-exponential rate.

Finally, we consider the problem of decentralized detection when sensors have access to side-information that affects the statistics of their measurements, and the network has an overall cost constraint. Nodes can decide whether or not to make a measurement and transmit a message to the fusion center ("censoring"), and also have a choice of the transmission function. We study the tradeoff in the detection performance with the cost constraint, and also the impact of sensor cooperation and global sharing of side-information. In particular, we show that if the Type I error probability is constrained to be small, then sensor cooperation is not necessary to achieve the optimal Type II error exponent.

Thesis Supervisor: John N. Tsitsiklis
Title: Clarence J Lebel Professor of Electrical Engineering

Thesis Supervisor: Moe Z. Win
Title: Associate Professor of Aeronautics and Astronautics

Acknowledgments

I want to thank my advisors, Professor John Tsitsiklis and Professor Moe Win, for all the support and encouragement that they have given me over the last few years. It has been a great honor and a privilege working with them. I have learnt a lot from them. They taught me how to combine intuition with mathematical rigor, how to find meaningful problems, and how to write scientific papers professionally. They have made a huge positive impact on my career, for which I am forever grateful.

Thanks also go to Professor Alan Willsky, for kindly agreeing to be on my thesis committee, and for providing helpful and insightful comments on my dissertation.

I also acknowledge the support of the National Science Foundation, under grants ECCS-0701623 and ANI-0335256, without which this dissertation would not have been possible.

The time spent in LIDS and MIT has been a most unforgettable one, which I will always have fond memories of. I am especially grateful to Mukul, Tony, Sheng, Xin, Ying-zong, Baris, Chung, Yuan, Thomas and Ee Hou for all the help and encouragement along the way. Thanks also go to the many other friends I have made at MIT, for their help and support, and for making life here colorful. Most of all, I want to express my deepest gratitude to my mother for being so patient, and for standing by me throughout my studies.

I dedicate this dissertation to the memory of my dearest father, who was the first to show me the beauty of Mathematics when I was still a child. My father was my greatest teacher, who helped me build a good foundation in Mathematics, and also taught me many important life lessons. I am forever grateful for his love and kindness, and his unwavering support for my academic pursuits. I will always remember him fondly in my heart.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation shows that for decentralized detection in a network configured as a tree with bounded height, the error probability falls exponentially fast with the number of nodes in the network, and the Neyman-Pearson error exponent is often the same as that of a parallel configuration. On the other hand, a tandem network exhibits sub-exponential error probability decay. More specifically, we consider a decentralized binary detection problem in a tree network, and study the detection performance when there is a large number of nodes in the network. We address some fundamental questions concerning the performance of sensor networks, and provide insights into the dependence of detection performance on network architecture and cost constraints. We also propose strategies that allow system designers to operate large networks efficiently.

In the following, we provide some background to the problem of decentralized detection, discuss some related work, and present our contributions.

1.1 Background and Related Work

The problem of optimal decentralized detection has attracted a lot of interest over the last twenty-five years. Tenney and Sandell [1] are the first to consider a decentralized detection system in which each of several sensors makes an observation and sends a summary (e.g., using a quantizer or other "transmission function") to a fusion center.

Such a system is to be contrasted to a *centralized* one, where the raw observations are transmitted directly to the fusion center. The framework introduced in [1] involves a "star topology" or "parallel configuration": the fusion center is regarded as the root of a tree, while the sensors are the leaves, directly connected to the root. Several pieces of work follow, e.g., [2–12], all of which study the parallel configuration under a Neyman-Pearson or Bayesian criterion. A common goal of these references is to characterize the optimal transmission function, where optimality usually refers to the minimization of the probability of error or some other cost function at the fusion center. A typical result is that under the assumption of (conditionally) independent sensor observations, likelihood ratio quantizers are optimal; see [6] for a summary of such results.

In this thesis, we consider resource-limited sensor networks. Sensor nodes are typically inexpensive, battery-powered devices that are deployed in large numbers over a possibly large geographical area. Such nodes commonly have limited power and may not be able to communicate over long distances. In the well studied parallel configuration described above, each node sends its information directly to the fusion center. Even though the error probabilities in a parallel configuration decrease exponentially, the energy consumption of having each sensor transmit directly to the fusion center can be too high. The energy consumption can be reduced by setting up a directed spanning in-tree, rooted at the fusion center. In a tree configuration, each non-leaf node combines its own observation (if any) with the messages it has received and forms a new message, which it transmits to another node. In this way, information from each node is propagated along a multi-hop path to the fusion center, but the information is "degraded" along the way. The study of sensor networks other than the parallel configuration is initiated in [13], which considers a tandem configuration, as well as more general tree configurations, and characterizes optimal transmission strategies under a Bayesian formulation. Tree configurations are also discussed in [14–21], under various performance objectives. In all but the simplest cases, the exact form of optimal strategies in tree configurations is difficult to derive. Most of these references focus on person-by-person (PBP) optimality and obtain nec-

essary, but not sufficient, conditions for an optimal strategy. When the transmission functions are assumed to be finite-alphabet quantizers, typical results establish that under a conditional independence assumption, likelihood ratio quantizers are PBP optimal. However, finding the optimal quantizer thresholds requires the solution of a nonlinear system of equations, with as many equations as there are thresholds. As a consequence, computing the optimal thresholds or characterizing the overall performance is hard, even for networks of moderate size.

Because of these difficulties, the analysis and comparison of large sensor networks is apparently tractable only in an asymptotic regime that focuses on the rate of decay of the error probabilities as the number of sensors increases. For example, in the Neyman-Pearson framework, one can focus on minimizing the error exponent [1]

$$g = \limsup_{n\to\infty} \frac{1}{n} \log \beta_n,$$

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. Note our convention that error exponents are negative numbers. The magnitude of the error exponent, $|g|$, is commonly referred to as the rate of decay of the Type II error probability. A larger $|g|$ would translate to a faster decay rate, hence a better detection performance. This problem has been studied in [22], for the case of a parallel configuration with a large number of sensors that receive independent, identically distributed (i.i.d.) observations.

The asymptotic performance of another special configuration, involving n sensors arranged in tandem, has been studied in [23–25], under a Bayesian formulation. Necessary and sufficient conditions for the error probability to decrease to zero as n increases have been derived. However, it has been conjectured in [6, 8, 25, 26], that even when the error probability decreases to zero, it apparently does so at a sub-exponential rate. (This is a conjecture that we will resolve in this thesis.) Accordingly, [25] argues that the tandem configuration is inefficient and suggests that

[1] Throughout this thesis, log stands for the natural logarithm.

as the number of sensors increases, the network "should expand more in a parallel than in [a] tandem" fashion.

Although a tree allows for shorter-range communications, thus making better use of communication resources, the detection performance may be worse than that of a parallel configuration. For the case where observations are obtained only at the leaves, it is not hard to see that the detection performance of a tree cannot be better than that of a parallel configuration with the same number of leaves. In this thesis, we investigate the extent to which bounded height tree networks under-perform a parallel configuration. We analyze the dependence of the optimal error exponent on the network architecture, and characterize the optimal error exponent for a large class of tree networks. We also present simple, and easily implementable strategies to nearly achieve the optimal performance. In the case of networks with unbounded height, we specifically consider the tandem network, and study the error probability decay rate.

Another way to mitigate energy consumption in a parallel configuration is to trade-off energy efficiency and detection reliability. For this purpose, "censoring networks" have been introduced in [27] and later in [28]. These references consider a binary detection problem, assume that the sensors obtain independent measurements X_i, and raise the question of deciding which sensors should transmit their measurements to a fusion center, subject to a constraint on the average number of transmitting sensors. In particular, they assume that the sensors are operating independently from each other, i.e., the censoring decisions do not involve any sensor cooperation or exchange of information. Their main results state that each sensor should base its decision on the likelihood ratio associated with its measurement, and should transmit X_i only if the likelihood ratio falls outside a "censoring interval." Subsequently, [29] and [30] consider the asymptotic performance of "constrained networks," including the case of an overall power constraint and the case of capacity-constrained communications. The question of deciding which sensors should transmit is replaced by the question of choosing the mode of sensor transmissions. There are differences between the problems considered in [27] and the problems studied in [29] and [30], but there are also

significant similarities, suggesting that a unified treatment may be possible. Such a unified treatment, at a higher level of generality, is one of the objectives of this thesis. However, we will be concerned with sensor networks with an asymptotically large number of nodes, unlike in [27, 28], where the problem of censoring is treated for the case of a fixed number of nodes.

1.2 Our Contributions

Most of the decentralized detection literature has been concerned with characterizing optimal detection strategies for particular sensor configurations; the comparison of the detection performance of different configurations is a rather unexplored area. We bridge this gap by considering the asymptotic performance of bounded height tree networks, under a Neyman-Pearson formulation. As it is not apparent that the Type II error probability decays exponentially fast with the number of nodes in the network, we first show that under the bounded height assumption, exponential decay is possible, and provide bounds for the optimal error exponent. We then obtain the rather counterintuitive result that if leaves dominate (in the sense that asymptotically almost all nodes are leaves), then bounded height trees have the same asymptotic performance as the parallel configuration, even in non-trivial cases. (Such an equality is clear in some trivial cases, e.g., the configuration shown in Figure 1-1, but is unexpected in general.) This result has important ramifications: a system designer can reduce the energy consumption in a network (e.g., by employing an h-hop spanning tree that minimizes the overall energy consumption), without losing detection efficiency, under certain conditions.

We also provide a strategy in which each non-leaf node sends only a 1-bit message, and which nearly achieves the same performance as the parallel configuration. These results are counterintuitive for the following reasons: 1) messages are compressed to only one bit at each non-leaf node so that "information" is lost along the way, whereas in the parallel configuration, no such compression occurs; 2) even though leaves dominate, there is no reason why the error exponent will be determined solely

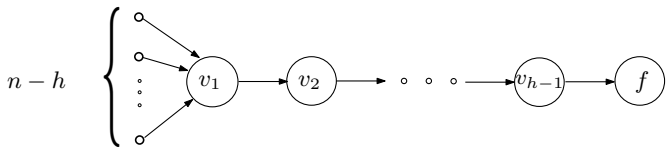

Figure 1-1: A tree network of height h, with $n-h$ leaves. Its error probability is no larger than that of a parallel configuration with $n-h$ leaves and a fusion center. If h is bounded while n increases, the optimal error exponent is the same as for a parallel configuration with n leaves.

by the leaves. For example, our discussion in Section 3.3.5 indicates that without the bounded height assumption, or if a Bayesian framework is assumed instead of the Neyman-Pearson formulation, then a generic tree network (of height greater than 1) performs strictly worse than a parallel configuration, even if leaves dominate. Under a mild additional assumption on the allowed transmission functions, we also find that the sufficient conditions for achieving the same error exponent as a parallel configuration, are necessary.

We also study the detection performance of bounded height tree networks in a Bayesian formulation, and show that although error probabilities decay exponentially fast, the decay rate is generically worse than that in the parallel configuration, unlike for the Neyman-Pearson formulation. We study specific classes of tree networks that are of practical interest, and characterize the optimal error exponents when restricted to simple, easily implementable strategies. We also compare and contrast the performance of these tree networks.

Next, we analyze the impact on the detection performance when sensors are either prone to failure or the communication links between sensors are unreliable. Specifically, we study and contrast the impact on the detection performance of either node failures (modeled by a Galton-Watson branching process) or unreliable communications (modeled by binary symmetric channels). In both cases, we focus on "dense" networks, in which we let the degree of every node (other than the leaves) become large, and we characterize the asymptotically optimal detection performance. We develop simple strategies that nearly achieve the optimal performance, and compare the performance of the two types of networks. Our results suggest that when de-

signing a large scale sensor network, it is more important to ensure that nodes can communicate reliably with each other (e.g., by boosting the transmission power) than to ensure that nodes are robust to failures. We consider the energy consumption of nodes arranged in a grid. We show that in a scheme that increases the transmission power of each node in a network so that the network performs as well as a parallel configuration with reliable communications, a tree network spanning the nodes in the grid is more energy efficient than a parallel configuration, in which all nodes transmit directly to a fusion center.

For the problem of Bayesian binary hypothesis testing in a tandem network, we show that the rate of error probability decay is always sub-exponential, thus establishing the validity of a long-standing conjecture [6,8,25,26]. Under the additional assumption of bounded Kullback-Leibler (KL) divergences, we show that for all $d > 1/2$, the error probability is $\Omega(e^{-cn^d})$, where c is a positive constant. [2] Furthermore, the bound $\Omega(e^{-c(\log n)^d})$, for all $d > 1$, holds under an additional mild condition on the distributions. This latter bound is shown to be tight. For the Neyman-Pearson formulation, we establish that if the sensors act myopically, the Type II error probabilities also decay at a sub-exponential rate.

Finally, we consider the problem of constrained decentralized detection in a parallel configuration (censoring network). We characterize the optimal error exponent, showing the tradeoff between the detection performance and the cost constraint, and derive asymptotically optimal strategies for the case where sensor decisions are only allowed to depend on locally available information. Furthermore, we show that for the Neyman-Pearson case, global sharing of side-information ("sensor cooperation") does not improve the asymptotically optimal performance, when the Type I error is constrained to be small. This implies that each sensor can make its censoring and transmission decisions only on the basis of the locally available side-information. Moreover, we show that all sensors can use the same policy, which allows for a simple distributed scheme.

[2] For two nonnegative functions f and g, we write $f(n) = \Omega(g(n))$ (resp. $f(n) = O(g(n))$) if for all n sufficiently large, there exists a positive constant c such that $f(n) \geq cg(n)$ (resp. $f(n) \leq cg(n)$). We write $f(n) = \Theta(g(n))$ if $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$.

Parts of this thesis have appeared in [31, 32], which investigated the performance of bounded height tree networks; in [33], which studied the impact of node failures and unreliable communications on detection performance; in [34], which showed that error probabilities decay sub-exponentially in tandem networks, and provided tight lower bounds for the error probability decay rate; and in [35–37], which investigated the performance of censoring sensor networks.

1.3 Thesis Outline

In Chapter 2, we introduce the basic model that underlies most of our development. We also list some assumptions that are made in most of the thesis. In Chapter 3, we study the Neyman-Pearson detection problem in bounded height tree networks, and in Chapter 4, we consider the Bayesian formulation. We analyze the impact of node failures and unreliable communications on detection performance in Chapter 5. In Chapter 6, we study the tandem network, and in Chapter 7, we consider the problem of censoring in parallel configurations. Finally, in Chapter 8, we conclude and discuss some future research directions.

Chapter 2

The Basic Model

In this chapter, we introduce the basic model that we will consider throughout this dissertation, and some common assumptions that are made in various parts of the thesis. The development of subsequent chapters builds on the model we describe here, although in some cases we will make extensions or modifications to the model. Finally, we will also sketch some mathematical results that underly most of our development.

This thesis considers decentralized binary detection problems involving $n-1$ sensors and a fusion center. We will be interested in the case where 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 sensor v observes a random variable X_v taking values in some set $\mathcal{X}$. The sensor network is configured as a directed tree, and each node's information is propagated via the arcs in the tree to the root of the tree, which is designated as the fusion center. We introduce a framework below that allows us to describe the network and its evolution as more nodes are added to the network.

2.1 Tree Networks

The configuration of the sensor network is represented by a directed tree $T_n = (V_n, E_n)$. 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 will always use the special symbol f to denote the root of T_n. We assume that the arcs are oriented so that they all point towards the fusion center. In the sequel, whenever we use the term "tree", we mean a directed, rooted tree as described above.

We will use the terminology "sensor" and "node" interchangeably. Moreover, the fusion center f will also be called a sensor, even though it plays the special role of fusing; whether the fusion center makes its own observation or not is irrelevant, since we are working in the large n regime, and we will assume it does not.

We say that 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, $C_n(v)$ 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 will 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 edges being deleted and some new edges 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., $h < \infty$.

Definition 2.1 (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.*

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, while in an h-uniform tree, all leaves are at level 0.

Let $l_n(v)$ be the number of leaves of the sub-tree rooted at node v. (These are the

leaves whose path to f goes through v.) Thus, $l_n(f)$ is the total number of leaves. Let $p_n(v)$ be the total number of predecessors of v, i.e., the total number of nodes in the sub-tree rooted at v, not counting v itself. Thus, $p_n(f) = n - 1$. We let $A_n \subset V_n$ be the set of nodes whose immediate predecessors include leaves of the tree T_n. Finally, we let $B_n \subset A_n$ be the set of nodes all of whose predecessors are leaves; see Figure 2-1.

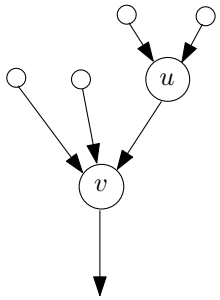

Figure 2-1: Both nodes v and u belong to the set A_n, but only node u belongs to the set B_n.

2.2 Strategies

Given a tree T_n, consider a node $v \neq f$. Node v receives messages Y_u from every $u \in C_n(v)$ (i.e., from its immediate predecessors). Node v then uses a transmission function γ_v to encode and transmit a summary $Y_v = \gamma_v(X_v, \{Y_u : u \in C_n(v)\})$ of its own observation X_v, and of the received messages $\{Y_u : u \in C_n(v)\}$, to its immediate successor.[1] We constrain all messages to be symbols in a fixed alphabet $\mathcal{T}$. Thus, if the in-degree of v is $|C_n(v)| = d$, then the transmission function γ_v maps $\mathcal{X} \times \mathcal{T}^d$ to $\mathcal{T}$. Let $\Gamma(d)$ be a given set of transmission functions that the node v can choose from. In general, $\Gamma(d)$ is a subset of the set of all possible mappings from $\mathcal{X} \times \mathcal{T}^d$ to $\mathcal{T}$. For example, $\Gamma(d)$ is often assumed to be the set of quantizers whose outputs are the result of comparing likelihood ratios to some thresholds (cf. the definition of a Log-Likelihood Ratio Quantizer in Section 3.1.2). For convenience, we denote the set

[1] To simplify the notation, we suppress the dependence of X_v, Y_v, γ_v, etc. on n.

of transmission functions for the leaves, $\Gamma(0)$, by Γ. We assume that all transmissions are perfectly reliable.

Consider now the root f, and suppose that it has d immediate predecessors. It receives messages from its immediate predecessors, and based on this information, it decides between the two hypotheses H_0 and H_1, using a fusion rule $\gamma_f : \mathcal{T}^d \mapsto \{0,1\}$.[2] Let Y_f be a binary-valued random variable indicating the decision of the fusion center.

We define a *strategy* for a tree T_n, with $n-1$ nodes and a fusion center, as a collection of transmission functions, one for each node, and a fusion rule. In some cases, we will be considering strategies in which only the leaves make observations; every other node v simply fuses the messages it has received, and forwards a message $Y_v = \gamma_v(\{Y_u : u \in C_n(v)\})$ to its immediate successor. A strategy of this type will be called a *relay strategy*. A tree network in which we restrict to relay strategies will be called a *relay tree*. If in addition, the alphabet $\mathcal{T}$ is binary, we will use the terms *1-bit relay strategy* and *1-bit relay tree*. Finally, in a relay tree, nodes other than the root and the leaves will be called *relay nodes*.

2.3 Common Assumptions

In this section, we list some of the common assumptions that we will be making in the sequel. We also introduce some notations that we will frequently use. Our first assumption involves the distribution of the random variables $\{X_v : v \in V_n\}$.

Assumption 2.1. *Under hypothesis H_j, where $j = 0, 1$, the random variables X_v are i.i.d., with marginal distribution $\mathbb{P}_j^X$.*

Assumption 2.1 will be assumed in most of the sequel except in Chapter 7. This assumption is commonly made in the literature, e.g., [6, 22, 25, 29, 30]. We note that

[2] Recall that in centralized Neyman-Pearson detection, randomization can reduce the Type II error probability, while in Bayesian detection, randomization does not improve the detection performance. Therefore, in Neyman-Pearson detection, the fusion center uses a randomized fusion rule to make its decision. Similarly, the transmission functions γ_v used by each node v, can also be randomized. We avoid any discussion of randomization here, and in most of the thesis, to simplify the exposition, and because randomization is not required asymptotically in Chapters 3-6. A detailed discussion of randomization is presented in Chapter 7.

without Assumption 2.1, the problem of decentralized detection in a parallel configuration involving only 2 nodes, and for which the space $\mathcal{X}$ (recall that this is the value space of each node's observation) is discrete, is a NP-complete problem [38]. Therefore, without the above assumption, the problems we consider would be intractable. Although this assumption is restrictive, and does not hold in many practical cases of interest, it nevertheless allows us to obtain tractable formulations and results that provide valuable insights into the more general case of correlated observations.

We denote the Kullback-Leibler (KL) divergence of two probability measures, $\mathbb{P}$ and $\mathbb{Q}$ as

$$\mathrm{D}(\mathbb{P} \,\|\, \mathbb{Q}) = \mathbb{E}^{\mathbb{P}}\Big[\log \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\Big],$$

where $\mathbb{E}^{\mathbb{P}}$ is the expectation operator with respect to (w.r.t.) $\mathbb{P}$. Suppose that X is a sensor observation. For any $\gamma \in \Gamma$, let the distribution of $\gamma(X)$ be $\mathbb{P}_j^\gamma$. Note that $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) \leq 0 \leq \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)$, with both inequalities being strict as long as the measures $\mathbb{P}_0^\gamma$ and $\mathbb{P}_1^\gamma$ are not indistinguishable.

Assumptions 2.2 and 2.3 are similar to the assumptions made in the study of the parallel configuration (see [22]).

Assumption 2.2. *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 $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < 0 < \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)$.*

Assumption 2.3. $\mathbb{E}_0\big[\log^2 \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big] < \infty$.

Assumption 2.3 implies the following lemma; see Proposition 3 of [22] for a proof.[3]

Lemma 2.1. *There exists some $a \in (0, \infty)$, such that for all $\gamma \in \Gamma$,*

$$\mathbb{E}_0\Big[\log^2 \frac{\mathrm{d}\mathbb{P}_1^\gamma}{\mathrm{d}\mathbb{P}_0^\gamma}\Big] \leq \mathbb{E}_0\Big[\log^2 \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\Big] + 1 < a,$$
$$\mathbb{E}_0\Big[\Big|\log \frac{\mathrm{d}\mathbb{P}_1^\gamma}{\mathrm{d}\mathbb{P}_0^\gamma}\Big|\Big] < a.$$

[3] In reference to that proof, the argument needs to be carried out by using the function $t \mapsto (t \log^2 t)\mathbf{1}(t \geq 1)$, which is convex for $t \geq 0$, together with the fact that $t \log^2 t < 1$ when $t \in [0, 1)$.

2.4 Mathematical Preliminaries

In this section, we state Cramèr's Theorem from Large Deviations Theory (see either [39] or [40]). This result will be useful in the proofs of subsequent chapters. We present here, without proof, a modified version of Theorem 1.3.13 of [40].

Theorem 2.1 (Cramèr)**.** *Suppose that X has distribution $\mathbb{P}$, with log-moment generating function $\varphi(\lambda) = \log \mathbb{E}[\exp(\lambda X)]$. Let the Fenchel-Legendre transform of $\varphi(\lambda)$ be $\Phi(x) = \sup\{\lambda x - \varphi(\lambda) : \lambda \in \mathbb{R}\}$. Suppose that $\mathbb{E}[|X|] < \infty$. Then,*

(i) $\mathbb{P}(X \geq x) \leq \exp(-\Phi(x))$, *for all* $x \geq \mathbb{E}[X]$.

(ii) $\mathbb{P}(X \leq x) \leq \exp(-\Phi(x))$, *for all* $x \leq \mathbb{E}[X]$.

Furthermore, suppose that there exists an open interval I such that $\varphi(\lambda) < \infty$ for all $\lambda \in I$ [4]*. Suppose that x lies inside the support of the distribution $\mathbb{P}$, and has a corresponding $\lambda \in I$, such that $\varphi'(\lambda) = x$. Then, we have for all $\epsilon > 0$,*

$$\mathbb{P}\big(\big|X - x\big| < \epsilon\big) \geq \Big(1 - \frac{\varphi''(\lambda)}{\epsilon^2}\Big) \exp\big(- \Phi(x) - |\lambda|\epsilon\big).$$

[4] Note that the function $\varphi(\cdot)$ is convex and twice differentiable over I.

Chapter 3

Network Architectures and Performance

In this chapter, we investigate the detection performance of a tree configuration under a Neyman-Pearson criterion. We restrict to trees with bounded height for two reasons. First, without a restriction on the height of the tree, performance can be poor (this is exemplified by tandem networks in which the error probability seems to decay at a sub-exponential rate; cf. Chapter 6). Second, bounded height translates to a bound on the delay until information reaches the fusion center.

We will first state the Neyman-Pearson formulation in detail, and provide some motivating examples. We then show that the detection error probability decays exponentially fast in a tree with bounded height, and provide sufficient, as well as necessary conditions, for the error exponent to be the same as that of a parallel configuration. We also propose strategies that nearly achieve the optimal error exponent. Finally, we discuss the admissibility of our proposed strategies, and provide some numerical examples.

3.1 The Neyman-Pearson Problem

In this section, we formulate the Neyman-Pearson decentralized detection problem in a tree network. We provide some motivating examples, and introduce our assump-

tions. Then, we give a summary of the main results.

Given a tree T_n, we require that the Type I error probability $\mathbb{P}_0(Y_f = 1)$ be no more than a given $\alpha \in (0, 1)$. A strategy is said to be *admissible* if it meets this constraint. We are interested in minimizing the Type II error probability $\mathbb{P}_1(Y_f = 0)$. Accordingly, we define $\beta^*(T_n)$ as the infimum of $\mathbb{P}_1(Y_f = 0)$, over all admissible strategies. Similarly, we define $\beta_R^*(T_n)$ as the infimum of $\mathbb{P}_1(Y_f = 0)$, over all admissible relay strategies. Typically, $\beta^*(T_n)$ or $\beta_R^*(T_n)$ will converge to zero as $n \to \infty$. We are interested in the question of whether such convergence takes place exponentially fast, and in the exact value of the Type II error exponent, defined by

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

Note that in the relay case, we use the total number of leaves $l_n(f)$ instead of n in the definition of g_R^*. This is because only the leaves make observations and therefore, g_R^* measures the rate of error decay per observation.

In the classical case of a parallel configuration, with $n-1$ leaves directly connected to the fusion center, the optimal error exponent, denoted as g_P^*, is given by [22]

$$g_P^* = \lim_{n\to\infty} \frac{1}{n} \log \beta^*(T_n) = -\sup_{\gamma\in\Gamma} \mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), \tag{3.1}$$

under Assumptions 2.1-2.3.

Our objective is to study g^* and g_R^* for different sequences of trees. In particular, we wish to obtain bounds on these quantities, develop conditions under which they are strictly negative (indicating exponential decay of error probabilities), and develop conditions under which they are equal to g_P^*. At this point, under Assumptions 2.1-2.3, we can record two relations that are always true:

$$g_P^* \le g_R^*, \qquad -\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X) \le g^* \le z g_R^*, \tag{3.2}$$

where $z = \liminf_{n\to\infty} l_n(f)/n$. The first inequality is true because all of the combining of messages that takes place in a relay network can be carried out internally, at the

fusion center of a parallel network with the same number of leaves. The inequality $-\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X) \leq g^*$ follows from the fact that $-\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X)$ is the classical error exponent in a centralized system where all raw observations are transmitted directly to the the fusion center. Finally, the inequality $g^* \leq z g_R^*$ follows because an optimal strategy is at least as good as an optimal relay strategy; the factor of z arises because we have normalized g_R^* by $l_n(f)$ instead of n.

For a sequence of trees of the form shown in Figure 1-1, it is easily seen that $g^* = g_R^* = g_P^*$. In order to develop some insights into the problem, we now consider some less trivial examples.

3.1.1 Motivating Examples

In the following examples, we restrict to relay strategies for simplicity, i.e., we are interested in characterizing the error exponent g_R^*. However, most of our subsequent results hold without such a restriction, and similar statements can be made about the error exponent g^* (cf. Theorem 3.1).

Example 3.1. *Consider a 2-uniform sequence of trees, as shown in Figure 3-1, where each node v_i receives messages from $m = (n-3)/2$ leaves (for simplicity, we assume that n is odd).*

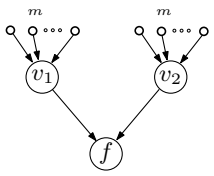

Figure 3-1: A 2-uniform tree with two relay nodes.

Let us restrict to 1-bit relay strategies. Consider the fusion rule that declares H_0 iff both v_1 and v_2 send a 0. In order to keep the Type I error probability bounded by α, we view the message by each v_i as a local decision about the hypothesis, and require that its local Type I error probability be bounded by $\alpha/2$. Furthermore, by viewing the sub-tree rooted at v_i as a parallel configuration, we can design strategies for each

sub-tree so that

$$\lim_{n\to\infty} \frac{1}{m} \log \mathbb{P}_1(Y_{v_i} = 0) = g_P^*. \tag{3.3}$$

At the fusion center, the Type II error exponent is then given by

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log \beta_n &= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1(Y_{v_1} = 0, Y_{v_2} = 0) \\
&= \frac{1}{2} \lim_{n\to\infty} \frac{1}{m} \log \mathbb{P}_1(Y_{v_1} = 0) + \frac{1}{2} \lim_{n\to\infty} \frac{1}{m} \log \mathbb{P}_1(Y_{v_2} = 0) \\
&= g_P^*,
\end{aligned}$$

where the last equality follows from (3.3). This shows that the Type II error probability falls exponentially and, more surprisingly, that $g_R^* \leq g_P^*$. In view of Eq. (3.2), we have $g_R^* = g_P^*$. It is not difficult to generalize this conclusion to all sequences of trees in which the number $n - l_n(f) - 1$ of relay nodes is bounded. For such sequences, we will also see that $g^* = g_R^*$ (cf. Theorem 3.1(iii)). □

Example 3.2. *We now consider an example in which the number of relay nodes grows with n. In Figure 3-2, we let both m and N be increasing functions of n (the total number of nodes), in a manner to be made explicit shortly.*

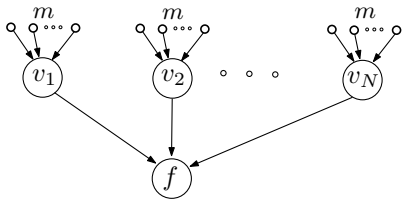

Figure 3-2: A 2-uniform tree with a large number of relay nodes.

Let us try to apply a similar argument as in Example 3.1, to see whether the optimal exponent of the parallel configuration can be achieved with a relay strategy, i.e., whether $g_R^* = g_P^*$. We let each node v_i use a local Neyman-Pearson test. We also let the fusion center declare H_0 iff it receives a 0 from all relay sensors. In order to have a hope of achieving the error exponent of the parallel configuration, we need

to choose the local Neyman-Pearson test at each relay so that its local Type II error exponent is close to $g_P^* = -\sup_{\gamma\in\Gamma} \mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma)$*. However, the associated local Type I error cannot fall faster than exponentially, so we can assume it is bounded below by* $\delta\exp(-m\epsilon)$*, for some* $\delta, \epsilon > 0$*, and for all* m *large enough. In that case, the overall Type I error probability (at the fusion center) is at least* $1-(1-\delta e^{-m\epsilon})^N$*. We then note that if* N *increases quickly with* m *(e.g.,* $N = m^m$*), the Type I error probability approaches 1, and eventually exceeds* α*. Hence, we no longer have an admissible strategy. Thus, if there is a hope of achieving the optimal exponent* g_P^* *of the parallel configuration, a more complicated fusion rule will have to be used.* □

Our subsequent results will establish that, similar to Example 3.1, the equalities $g^* = g_R^* = g_P^*$ also hold in Example 3.2. However, Example 3.2 shows that in order to achieve this optimal error exponent, we may need to employ nontrivial fusion rules at the fusion center (and for similar reasons at the relay nodes), and various thresholds will have to be properly tuned. The simplicity of the fusion rule in Example 3.1 is not representative.

In our next example, the optimal error exponent is inferior (strictly larger) than that of a parallel configuration.

Example 3.3. *Consider a sequence of 1-bit relay trees with the structure shown in Figure 3-3. Let the observations* X_v *at the leaves be i.i.d. Bernoulli random variables*

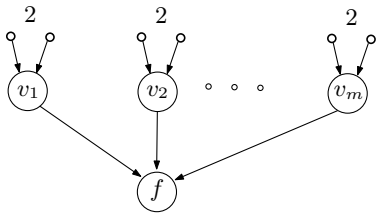

Figure 3-3: A 2-uniform tree, with two leaves attached to each level 1 node.

with parameter $1-p$ *under* H_0*, and parameter* p *under* H_1*, where* $1/2 < p < 1$*. Note that*

$$g_P^* = \mathbb{E}_0\Big[\log\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\Big] = p\log\frac{1-p}{p} + (1-p)\log\frac{p}{1-p}.$$

We can identify this relay tree with a parallel configuration involving m nodes, with each node receiving an independent observation distributed as $\gamma(X_1, X_2)$. Note that we can restrict the transmission function γ to be the same for all nodes $v_1, ..., v_m$ [22], without loss of optimality. We have

$$\lim_{n\to\infty} \frac{1}{m} \log \beta^*(T_n) = \min_{\gamma\in\Gamma(2)} \sum_{j=0}^{1} \mathbb{P}_0\big(\gamma(X_1, X_2) = j\big) \log \Big[\frac{\mathbb{P}_1\big(\gamma(X_1, X_2) = j\big)}{\mathbb{P}_0\big(\gamma(X_1, X_2) = j\big)}\Big]. \quad (3.4)$$

To minimize the right-hand side (R.H.S.) of (3.4)*, we only need to consider a small number of choices for γ. If $\gamma(X_1, X_2) = X_1$, we are effectively removing half of the original $2m$ nodes, and the resulting error exponent is $g_P^*/2$, which is inferior to g_P^*. Suppose now that γ is of the form $\gamma(X_1, X_2) = 0$ iff $X_1 = X_2 = 0$. Then, it is easy to see, after some calculations (omitted), that*

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{m} \log \beta^*(T_n) &= p^2 \log \frac{(1-p)^2}{p^2} + (1-p^2) \log \frac{1-(1-p)^2}{1-p^2} \\ &> 2\Big(p \log \frac{1-p}{p} + (1-p) \log \frac{p}{1-p}\Big),\end{aligned}$$

and

$$\lim_{n\to\infty} \frac{1}{l_n(f)} \log \beta^*(T_n) > p \log \frac{1-p}{p} + (1-p) \log \frac{p}{1-p} = g_P^*.$$

Finally, we need to consider γ of the form $\gamma(X_1, X_2) = 1$ iff $X_1 = X_2 = 1$. A similar calculation (omitted) shows that the resulting error exponent is again inferior. We conclude that the relay network is strictly inferior to the parallel configuration, i.e., $g_P^* < g_R^*$. An explanation is provided by noting that this sequence of trees violates a necessary condition, developed in Section 3.3.6 for the optimal error exponent to be the same as that of a parallel configuration; see Theorem 3.1(iv). □

A comparison of the results for the previous examples suggests that we have $g_P^* = g_R^*$ (respectively, $g_P^* < g_R^*$) whenever the degree of level 1 nodes increases (respectively, stays bounded) as n increases. That would still leave open the case of networks in which different level 1 nodes have different degrees, as in our next example.

Example 3.4. *Consider a sequence of 2-uniform trees of the form shown in Figure 3-4. Each node v_i, $i = 1, ..., m$, has $i + 1$ leaves attached to it. We will see that the optimal error exponent is again the same as for a parallel configuration, i.e., $g_R^* = g^* = g_P^*$. (cf. Theorem 3.1(ii)).* □

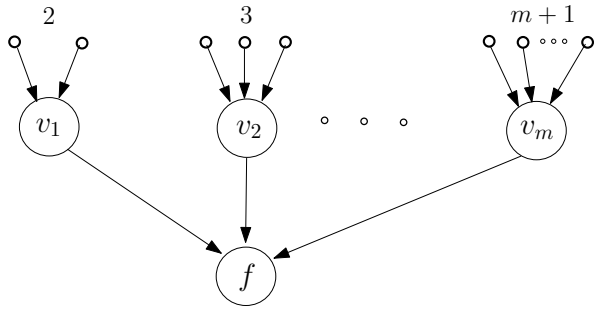

Figure 3-4: A 2-uniform tree, with $l_n(v_i) = i + 1$.

3.1.2 Assumptions and Notation

We will make use of Assumptions 2.1, 2.2 and 2.3 in most of our results. In this subsection, we list some additional assumptions and notation.

Given an admissible strategy, and for each node $v \in V_n$, we consider the log-likelihood ratio of the distribution of Y_v (the message sent by v) under H_1, w.r.t. its distribution under H_0,

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

where $\mathrm{d}\mathbb{P}_{1,n}^{(v)}/\mathrm{d}\mathbb{P}_{0,n}^{(v)}$ is the Radon-Nikodym derivative of the distribution of Y_v under H_1 w.r.t. that under H_0. If Y_v takes values in a discrete set, then this is just the log-likelihood ratio $\log\big(\mathbb{P}_1(Y_v = y)/\mathbb{P}_0(Y_v = y)\big)$. For simplicity, we let $L_{v,n} = \mathcal{L}_{v,n}(Y_v)$ and define the log-likelihood ratio of the received messages at node v to be

$$S_n(v) = \sum_{u \in C_n(v)} L_{u,n}.$$

(Recall that $C_n(v)$ is the set of immediate predecessors of v.)

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

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

where

$$x = \frac{1}{l_n(v)} \sum_{u \in C_n(v)} \mathcal{L}_{u,n}(y_u). \tag{3.5}$$

By definition, a node v that uses a LLRQ 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 will assume that LLRQs are available choices of transmission functions for all non-leaf nodes.

Assumption 3.1. *For all $t \in \mathbb{R}$ and $d > 0$, $\mathrm{LLRQ}_{d,t} \in \Gamma(d)$.*

As already discussed (cf. Eq. (3.2)), the optimal performance of a relay tree is always dominated by that of a parallel configuration with the same number of leaves, i.e., $g_P^* \le g_R^*$. In Section 3.3, we find sufficient conditions under which the equality $g_R^* = g_P^*$ holds. Then, in Section 3.3.6, we look into necessary conditions for this to be the case. It turns out that 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. For example, if the 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.

Assumption 3.2.

(a) There exists a $n_0 \ge 1$ such that for all $n \ge n_0$, we have $l_n(v) > 1$ for all v in the set B_n of nodes whose immediate predecessors are all leaves.

(b) Let $X_1, X_2, \ldots$ 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 $\gamma_i \in \Gamma$, $i = 1, \ldots, k$, let $\xi =$

$(\gamma_0, \ldots, \gamma_k)$. *We also let* ν_j^ξ *be the distribution of* $\gamma_0(\gamma_1(X_1), \ldots, \gamma_k(X_k))$ *under hypothesis* H_j. *We assume that*

$$g_P^* < \inf_{\xi \in \Gamma(k) \times \Gamma^k} \frac{1}{k} \mathbb{E}_0 \Big[\log \frac{\mathrm{d}\nu_1^\xi}{\mathrm{d}\nu_0^\xi} \Big], \tag{3.6}$$

for all $k > 1$.

Assumption 3.2 holds in most cases of interest. Part (a) results in no loss of generality: 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. Regarding part (b), it is easy to see that the left-hand side (L.H.S.) of (3.6) is always less than or equal to the R.H.S., hence we have only excluded those cases where (3.6) holds with equality. We are essentially assuming that when the messages $\gamma_1(X_1), \ldots, \gamma_k(X_k)$ are summarized (or quantized) by γ_0, there is some loss of information, as measured by the associated KL divergences.

3.1.3 Main Results

In this section, we collect and summarize the main results of this chapter. The asymptotic proportion of nodes that are leaves, defined by

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

plays a critical role.

Theorem 3.1. *Consider a sequence of trees,* $(T_n)_{n \geq 1}$, *of bounded height. Suppose that Assumptions 2.1-2.3, and Assumption 3.1 hold. Then,*

(i) $g_P^* \leq g_R^* < 0$ *and* $-\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X) \leq g^* \leq z g_R^* < 0$.

(ii) If $z = 1$, *then* $g_P^* = g^* = g_R^*$.

(iii) If the number of non-leaf nodes is bounded, or if $\min_{v \in B_n} l_n(v) \to \infty$, *then* $g_P^* = g^* = g_R^*$.

(iv) If Assumption 3.2 also holds, we have $g_R^* = g_P^*$ iff $z = 1$.

Note that part (i) follows from (3.2), except for the strict negativity of the error exponents, which is established in Proposition 3.2. Part (ii) is proved in Proposition 3.3. Part (iii) is proved in Corollary 3.1. (Recall that B_n is the set of non-leaf nodes all of whose immediate predecessors are leaves.) Part (iv) is proved in Proposition 3.5. One might also have expected a result asserting that $g_P^* \leq g^*$. However, this is not true without additional assumptions, as will be discussed in Section 3.3.6.

3.2 Error Bounds for h-Uniform Relay Trees

In this section, we consider a 1-bit h-uniform relay tree, in which all relay nodes at level k use a LLRQ with a common threshold t_k. We wish to develop upper bounds for the error probabilities at the various nodes. We do this recursively, by moving along the levels of the tree, starting from the leaves. Given bounds on the error probabilities associated with the messages received by a node, we develop a bound on the log-moment generating function at that node (cf. Eq. (3.8)), and then use the standard Chernoff bound technique to develop a bound on the error probability for the message sent by that node (cf. Eq. (3.7)).

Let $t^{(k)} = (t_1, t_2, \ldots, t_k)$, for $k \geq 1$, and $t^{(0)} = \emptyset$. For $j = 0, 1$, $k \geq 1$, and $\lambda \in \mathbb{R}$, we define recursively

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

$$\Lambda_{j,k}^*(\gamma, t^{(k)}) = \sup_{\lambda \in \mathbb{R}} \big\{ \lambda t_k - \Lambda_{j,k-1}(\gamma, t^{(k-1)}; \lambda) \big\}, \tag{3.7}$$

$$\Lambda_{j,k}(\gamma, t^{(k)}; \lambda) = \max \big\{ -\Lambda_{1,k}^*(\gamma, t^{(k)})(j + \lambda), \Lambda_{0,k}^*(\gamma, t^{(k)})(j - 1 + \lambda) \big\}. \tag{3.8}$$

The operation in (3.7) is known as the Fenchel-Legendre transform of $\Lambda_{j,k-1}(\gamma, t^{(k-1)}; \lambda)$ [39]. We will be interested in the case where

$$-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < 0 < \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma), \tag{3.9}$$

$$t_1 \in \big(-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)\big), \tag{3.10}$$

$$t_k \in \big(-\Lambda_{1,k-1}^*(\gamma, t^{(k-1)}), \Lambda_{0,k-1}^*(\gamma, t^{(k-1)})\big), \text{ for } 1 < k \leq h. \tag{3.11}$$

We now provide an inductive argument to show that the above requirements on the thresholds t_k are feasible. From Assumption 2.2, there exists a $\gamma \in \Gamma$ that satisfies (3.9), hence the constraint (3.10) is feasible. Furthermore, the $\Lambda_{j,1}^*(\gamma, t^{(1)})$ are large deviations rate functions and are therefore positive when t_1 satisfies (3.10) [39]. Suppose now that $k > 1$ and that $\Lambda_{j,k-1}^*(\gamma, t^{(k-1)}) > 0$. From (3.8), $\Lambda_{j,k-1}(\gamma, t^{(k-1)}; \lambda)$ is the maximum of two linear functions of λ (see Figure 3-5). Taking the Fenchel-Legendre transform, and since t_k satisfies (3.11), we obtain $\Lambda_{j,k}^*(\gamma, t^{(k)}) > 0$, which completes the induction.

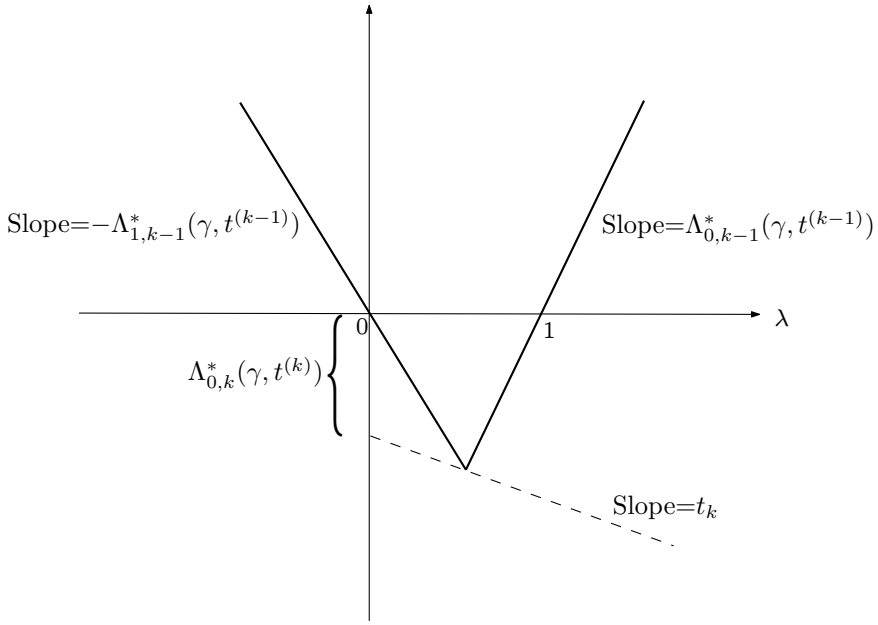

Figure 3-5: Typical plot of $\Lambda_{0,k-1}(\gamma, t^{(k-1)}; \lambda)$, $k \geq 2$.

From the definitions of $\Lambda_{j,k}$ and $\Lambda_{j,k}^*$, the following relations can be established. The proof consists of straightforward algebraic manipulations and is omitted.

Lemma 3.1. *Suppose that $\gamma \in \Gamma$ satisfies (3.9), and $t^{(h)}$ satisfies (3.10)-(3.11). For*

$k \geq 1$, we have

$$\Lambda^*_{1,k}(\gamma, t^{(k)}) = \Lambda^*_{0,k}(\gamma, t^{(k)}) - t_k,$$

and for $k \geq 2$, we have

$$\inf_{\lambda \in [0,1]} \Lambda_{0,k}(\gamma, t^{(k)}; \lambda) = -\frac{\Lambda^*_{0,k}(\gamma, t^{(k)})\Lambda^*_{1,k}(\gamma, t^{(k)})}{\Lambda^*_{0,k}(\gamma, t^{(k)}) + \Lambda^*_{1,k}(\gamma, t^{(k)})}.$$

Furthermore, the supremum in (3.7) is achieved at some $\lambda \in (-1, 0)$ for $j = 1$, and $\lambda \in (0, 1)$ for $j = 0$. For $k \geq 2$, we have

$$\Lambda^*_{1,k}(\gamma, t^{(k)}) = \frac{\Lambda^*_{1,k-1}(\gamma, t^{(k-1)})(\Lambda^*_{0,k-1}(\gamma, t^{(k-1)}) - t_k)}{\Lambda^*_{0,k-1}(\gamma, t^{(k-1)}) + \Lambda^*_{1,k-1}(\gamma, t^{(k-1)})},$$
$$\Lambda^*_{0,k}(\gamma, t^{(k)}) = \frac{\Lambda^*_{0,k-1}(\gamma, t^{(k-1)})(\Lambda^*_{1,k-1}(\gamma, t^{(k-1)}) + t_k)}{\Lambda^*_{0,k-1}(\gamma, t^{(k-1)}) + \Lambda^*_{1,k-1}(\gamma, t^{(k-1)})}.$$

Proposition 3.1 below, whose proof is provided in Section 3.5, will be our main tool in obtaining upper bounds on error probabilities. It 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. In Section 4.5, we present a class of tree networks whose error exponents are precisely $-\Lambda^*_{j,h}(\gamma, t^{(h)})$, for $j = 0, 1$ when restricted to certain strategies. Recall that $p_n(v)$ is the total number of predecessors of v, $l_n(v)$ is the number of leaves in the sub-tree rooted at v, and B_n is the set of nodes all of whose immediate predecessors are leaves.

Proposition 3.1. *Fix some $h \geq 1$, and consider a sequence of trees $(T_n)_{n \geq 1}$ such that for all $n \geq n_0$, T_n is h-uniform. Suppose that Assumptions 2.1-2.3 hold. Suppose that, for every n, every leaf uses the same transmission function $\gamma \in \Gamma$, which satisfies (3.9), and that every level k node ($k \geq 1$) uses a LLRQ with threshold t_k, satisfying (3.10)-(3.11).*

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

$$\frac{1}{l_n(v)} \log \mathbb{P}_1\Big(\frac{S_n(v)}{l_n(v)} \leq t_k\Big) \leq -\Lambda^*_{1,k}(\gamma, t^{(k)}) + \frac{p_n(v)}{l_n(v)} - 1,$$
$$\frac{1}{l_n(v)} \log \mathbb{P}_0\Big(\frac{S_n(v)}{l_n(v)} > t_k\Big) \leq -\Lambda^*_{0,k}(\gamma, t^{(k)}) + \frac{p_n(v)}{l_n(v)} - 1.$$

(ii) Suppose that for all $n \geq n_0$ and all $v \in B_n$, we have $l_n(v) \geq N$. Then, for all $n \geq n_0$, we have

$$\frac{1}{l_n(f)} \log \mathbb{P}_1\Big(\frac{S_n(f)}{l_n(f)} \leq t_h\Big) \leq -\Lambda^*_{1,h}(\gamma, t^{(h)}) + \frac{h}{N},$$
$$\frac{1}{l_n(f)} \log \mathbb{P}_0\Big(\frac{S_n(f)}{l_n(f)} > t_h\Big) \leq -\Lambda^*_{0,h}(\gamma, t^{(h)}) + \frac{h}{N}.$$

3.3 Optimal Error Exponent

In this section, we show that the Type II error probability in a sequence of bounded height trees falls exponentially fast with the number of nodes. We derive sufficient conditions for the error exponent to be the same as that of a parallel configuration. We show that if almost all of the nodes are leaves, i.e., $z = 1$, then $g^*_P = g^* = g^*_R$. The condition $z = 1$ is also equivalent to another condition that requires that the proportion of leaves attached to bounded degree nodes vanishes asymptotically. We also show that under some additional mild assumptions, this sufficient condition is necessary. We start with some graph-theoretic preliminaries.

3.3.1 Properties of Trees

In this section, we define various quantities associated with a tree, and derive a few elementary relations that will be used later.

Recall that B_n is the set of non-leaf nodes all of whose predecessors are leaves. (For an h-uniform tree, B_n is the set of all level 1 nodes.) For $N > 0$, let

$$F_{N,n} = \{v \in B_n : l_n(v) \leq N\}, \qquad F^c_{N,n} = \{v \in B_n : l_n(v) > N\}, \tag{3.12}$$

and

$$q_{N,n} = \frac{1}{l_n(f)} \sum_{v \in F_{N,n}} l_n(v), \tag{3.13}$$

where the sum is taken to be zero if the set $F_{N,n}$ is empty. Let $q_N = \limsup_{n\to\infty} q_{N,n}$. For a sequence of h-uniform trees, this is the asymptotic proportion of leaves that belong to "small" subtrees in the network.

It turns out that it is easier to work with h-uniform trees. For this reason, we show how to transform any tree of height h to an h-uniform tree.

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 ($v \in A_n$). 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, $\{u_j : j = 1, \ldots, h-k\}$, 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 $j = 1, \ldots, h-k-1$; 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. $\square$

The height uniformization procedure essentially adds more nodes to the network, and re-attaches some leaves, so that the path from every leaf has exactly h hops. Let $(T'_n = (V'_n, E'_n))_{n\geq 1}$ be the new sequence of h-uniform trees obtained from $(T_n)_{n\geq 1}$, after applying the uniformization procedure. (We are abusing notation here in that T'_n typically does not have n nodes, nor is the sequence $|V'_n|$ increasing.) Regarding notation, we adopt the convention that quantities marked with a prime are defined with respect to T'_n.

Note that $l'_n(f) = l_n(f)$. For the case of a relay network, it is seen that any function of the observations at the leaves that can be computed in T'_n can also be computed in T_n. Thus, the detection performance of T'_n is no better than that of T_n. Hence, we obtain

$$g_R^* \leq \limsup_{n\to\infty} \frac{1}{l'_n(f)} \log \beta^*(T'_n). \tag{3.14}$$

Therefore, any upper bound derived for h-uniform trees, readily translates to an upper bound for general trees. On the other hand, the coefficients q_N for the h-uniform trees T'_n (to be denoted by q'_N) are different from the coefficients q_N for the original sequence T_n. They are related as follows. The proof is given in Section 3.5.

Lemma 3.2. *For any $N, M > 0$, we have*

$$q'_N \leq h(Nq_M + N/M).$$

In particular, if $q_N = 0$ for all $N > 0$, then $q'_N = 0$ for all $N > 0$.

It turns out that the condition $z = 1$ is equivalent to the condition $q_N = 0$ for all $N > 0$. The proof is provided in Section 3.5.

Lemma 3.3. *We have $z = 1$ iff $q_N = 0$ for all $N > 0$.*

3.3.2 An Upper Bound

In this section, we develop an upper bound on the Type II error probabilities, which takes into account some qualitative properties of the sequence of trees, as captured by q_N.

Lemma 3.4. *Consider an h-uniform sequence of trees $(T_n)_{n\geq 1}$, and suppose that Assumptions 2.1-2.3, and Assumption 3.1 hold. For every $\epsilon > 0$, there exists some N such that*

$$g^*_R \leq (1 - q_N)(g^*_P + \epsilon).$$

Proof. If $g^*_P + \epsilon \geq 0$, there is nothing to prove, since $q_N \leq 1$ and $g^*_R \leq 0$. Suppose that $g^*_P + \epsilon < 0$. Choose $\gamma \in \Gamma$ such that

$$-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) \leq -\sup_{\gamma' \in \Gamma} \mathrm{D}(\mathbb{P}_0^{\gamma'} \,\|\, \mathbb{P}_1^{\gamma'}) + \frac{\epsilon}{2} = g^*_P + \frac{\epsilon}{2} < 0.$$

Let $t_k = t = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) + \epsilon/2 \leq g^*_p + \epsilon$, for $k = 1, \ldots, h$, and note that

$$-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < t < 0. \tag{3.15}$$

Because of (3.15), we have $\Lambda^*_{0,1}(\gamma, t^{(1)}) > 0$. Furthermore, using Lemma 3.1,

$$\Lambda^*_{1,1}(\gamma, t^{(1)}) = \Lambda^*_{0,1}(\gamma, t^{(1)}) - t > -t.$$

Now let $k \geq 2$, and suppose that $\Lambda^*_{1,k-1}(\gamma, t^{(k-1)}) > -t$ and $\Lambda^*_{0,k-1}(\gamma, t^{(k-1)}) > 0$. From Lemma 3.1,

$$\Lambda^*_{0,k}(\gamma, t^{(k)}) = \frac{\Lambda^*_{0,k-1}(\gamma, t^{(k-1)})(\Lambda^*_{1,k-1}(\gamma, t^{(k-1)}) + t)}{\Lambda^*_{0,k-1}(\gamma, t^{(k-1)}) + \Lambda^*_{1,k-1}(\gamma, t^{(k-1)})} > 0,$$

and

$$\Lambda^*_{1,k}(\gamma, t^{(k)}) = \Lambda^*_{0,k}(\gamma, t^{(k)}) - t_k = \Lambda^*_{0,k}(\gamma, t^{(k)}) - t > -t.$$

Hence, by induction, t_k satisfies (3.10)-(3.11), so that Proposition 3.1 can be applied.

Choose N sufficiently large so that $h/N < \Lambda^*_{0,h}(\gamma, t^{(h)})$. If $q_N = 1$, the claimed result holds trivially. Hence, we assume that $q_N \in [0, 1)$. In this case, for n sufficiently large, there exists at least one node in B_n so that $l_n(v) > N$. We remove all nodes $v \in B_n$ with $l_n(v) \leq N$, and their immediate predecessors. Then, we remove all level 2 nodes v that no longer have any predecessors, and so on. In this way, we obtain an h-uniform subtree of T_n, to be denoted by T''_n. (Quantities marked with double primes are defined w.r.t. T''_n.) We have $l''_n(v) > N$ for all $v \in B''_n$, and $l''_n(f) = \sum_{v \in F^c_{N,n}} l_n(v) = l_n(f)(1 - q_{N,n})$. Consider the following relay strategy on the tree T''_n. (Since this is a subtree of T_n, this is also a relay strategy for the tree T_n, with some nodes remaining idle.) The leaves transmit with transmission function γ, and the other nodes use a 1-bit LLRQ with threshold t. (Note that in the definition (3.5) of the normalized log-likelihood ratio, the denominator $l_n(v)$ now becomes $l''_n(v)$.)

We first show that the strategy just described is admissible. We apply part (ii) of Proposition 3.1 to T''_n, to obtain

$$\limsup_{n \to \infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_f = 1)$$

$$
\begin{aligned}
&= \limsup_{n\to\infty} \frac{l_n''(f)}{l_n(f)} \cdot \frac{1}{l_n''(f)} \log \mathbb{P}_0(Y_f = 1) \\
&\leq (1 - q_N) \limsup_{n\to\infty} \frac{1}{l_n''(f)} \log \mathbb{P}_0\Big(\frac{S_n(f)}{l_n''(f)} > t\Big) \\
&\leq (1 - q_N)\big(- \Lambda^*_{0,h}(\gamma, t^{(h)}) + \frac{h}{N}\big) < 0,
\end{aligned}
$$

hence $\mathbb{P}_0(Y_f = 1) \leq \alpha$, when n is sufficiently large.

To bound the Type II error probability, we use Proposition 3.1 and Lemma 3.1, to obtain

$$
\begin{aligned}
g_R^* &\leq \limsup_{n\to\infty} \frac{1}{l_n(f)} \log \beta^*(T_n'') \\
&\leq (1 - q_N) \limsup_{n\to\infty} \frac{1}{l_n''(f)} \log \mathbb{P}_1\Big(\frac{S_n(f)}{l_n''(f)} \leq t\Big) \\
&\leq (1 - q_N)\big(- \Lambda^*_{1,h}(\gamma, t^{(h)}) + \frac{h}{N}\big) \\
&= (1 - q_N)\big(t - \Lambda^*_{0,h}(\gamma, t^{(h)}) + \frac{h}{N}\big) \\
&\leq (1 - q_N)t \\
&\leq (1 - q_N)\big(g_P^* + \epsilon\big).
\end{aligned}
$$

This proves the lemma. □

3.3.3 Exponential decay of error probabilities

We now establish that Type II error probabilities decay exponentially. The bounded height assumption is crucial for this result. Indeed, for the case of a tandem configuration, the exponential decay property does not seem to hold.

Proposition 3.2. *Consider a sequence of trees of height h, and let Assumptions 2.1-2.3, and Assumption 3.1 hold. Then,*

$$
-\infty < g_P^* \leq g_R^* < 0 \qquad \textit{and} \qquad -\infty < -\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X) \leq g^* < 0.
$$

Proof. The lower bounds on g_R^* and g^* follow from (3.2). Note that g_P^* cannot be equal

to $-\infty$ because it cannot be better than the error exponent of a parallel configuration in which all the observations are provided uncompressed to the fusion center. The error exponent in the latter case is $-\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X)$, by Stein's Lemma, and is finite as a consequence of Assumption 2.3.

It remains to show that the optimal error exponents are negative. Every tree of height h satisfies $n \leq l_n(f)h + 1$. From (3.2), we obtain $g^* \leq g_R^*/h$. Therefore, we only need to show that $g_R^* < 0$. As discussed in connection to (3.14), we can restrict attention to a sequence of h-uniform trees.

We use induction on h. If $h = 1$, we have a parallel configuration and the result follows from [22]. Suppose that the result is true for all sequences of $(h-1)$-uniform trees. Consider now a sequence of h-uniform trees. Let $\epsilon > 0$ be such that $g_P^* + \epsilon < 0$. From Lemma 3.4, there exists some N such that $g_R^* \leq (1 - q_N)(g_P^* + \epsilon)$. If $q_N < 1$, we readily obtain the inequality $g_R^* < 0$.

Suppose now that $q_N = 1$. We only need to consider a sequence $(n_k)_{k\geq 1}$ such that $\lim_{k\to\infty} q_{N,n_k} = 1$. Using the inequality (3.26), we have

$$\frac{|F_{N,n_k}|}{l_{n_k}(f)} \geq \frac{q_{N,n_k}}{N},$$

and

$$\liminf_{k\to\infty} \frac{|F_{N,n_k}|}{l_{n_k}(f)} \geq \frac{1}{N}. \tag{3.16}$$

For each node $v \in B_n$, we remove all of its immediate predecessors (leaves) except for one, call it u. The leaf u transmits $\gamma(X_u)$ to its immediate successor v. Since node v receives only a single message, it just forwards it to its immediate successor. The resulting performance is the same as if the nodes v in B_n were making a measurement X_v and transmitting $\gamma(X_v)$ to their successor. This is equivalent to deleting all the leaves of T_n to form a new tree, T_n'', which is $(h-1)$-uniform. The above argument shows that $\beta^*(T_{n_k}) \leq \beta^*(T_{n_k}'')$.

We have $l''_{n_k}(f) = |B_{n_k}|$ and from (3.16),

$$\liminf_{k\to\infty} \frac{|B_{n_k}|}{l_{n_k}(f)} \geq \liminf_{k\to\infty} \frac{|F_{N,n_k}|}{l_{n_k}(f)} \geq \frac{1}{N}.$$

Therefore,

$$\limsup_{k\to\infty} \frac{1}{l_{n_k}(f)} \log \beta^*(T_{n_k}) \leq \frac{1}{N} \limsup_{k\to\infty} \frac{1}{l''_{n_k}(f)} \log \beta^*(T''_{n_k}).$$

By the induction hypothesis, the right-hand side in the above inequality is negative and the proof is complete. □

3.3.4 Sufficient Conditions for Matching the Performance of the Parallel Configuration

We are now ready to prove the main result of this section. It shows that when $q_N = 0$ for all $N > 0$, or equivalently when $z = 1$ (cf. Lemma 3.3), bounded height tree networks match the performance of the parallel configuration.

Proposition 3.3. *Consider a sequence of trees of height h in which $z = 1$, or equivalently $q_N = 0$ for all $N > 0$. Suppose that Assumptions 2.1-2.3, and Assumption 3.1 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.

Proof. We have shown $g_P^* \leq g_R^*$ in (3.2). We now prove that $g_R^* \leq g_P^*$. As already explained, there is no loss in generality in assuming that the sequence of trees is h-uniform (by performing the height uniformization procedure, and using Lemma 3.2).

For any $\epsilon > 0$, Lemma 3.4 yields

$$g_R^* \leq g_P^* + \epsilon.$$

Letting $\epsilon \to 0$, we obtain $g_R^* \leq g_P^*$, hence $g_R^* = g_P^*$. From (3.2) with $z = 1$, we obtain $g^* \leq g_R^* = g_P^*$.

We now show that $g^* \geq g_P^*$. Consider a tree with n nodes, $l_n(f)$ of which are leaves. We will compare it with another sensor network in which $l_n(f)$ nodes v transmit a message $\gamma_v(X_v)$ to the fusion center and $n - l_n(f) - 1$ nodes transmit their raw observations to the fusion center. The latter network can simulate the original network, and therefore its optimal error exponent is at least as good. By a standard argument (similar to the one in Proposition 3.4 below), the optimal error exponent in the latter network can be shown to be greater than or equal to

$$\limsup_{n\to\infty} \frac{l_n(f)}{n} g_P^* + \limsup_{n\to\infty} -\frac{n - l_n(f) - 1}{n} \mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X) = g_P^*,$$

hence concluding the proof. □

Fix an $\epsilon \in (0, -g_P^*)$. For any tree sequence with $z = 1$, we can perform the height uniformization procedure to obtain an h-uniform sequence of trees. In practice, this height uniformization procedure may be performed virtually at each node, so that the tree sequence simulates a h-uniform tree sequence. A simple strategy on the height uniformized tree sequence that ϵ-achieves the optimal error exponent is a relay strategy in which:

(i) all leaves transmit with the same transmission function $\gamma \in \Gamma$, where γ is chosen such that $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) \leq g_P^* + \epsilon/2$;

(ii) all other nodes use 1-bit LLRQs with the same threshold $t = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) + \epsilon/2$.

Lemmas 3.2 and 3.3, and the proof of Lemma 3.4 shows that this relay strategy ϵ-achieves the optimal error exponent $g_R^* = g^* = g_P^*$. This also shows that there is no

loss in optimality even if we restrict the relay nodes to use only 1-bit LLRQs. This may be useful in situations where the nodes are simple, low-cost devices.

Proposition 3.3 provides sufficient conditions for a sequence of trees to achieve the same error exponent as the parallel configuration. We note a few special cases in which these sufficient conditions are satisfied. The first one is the case where there is a finite bound on the number of nodes that are not leaves. In that case, z is easily seen to be 1. This is consistent with the conclusion of Example 3.1, where a simpler argument was used. The second is the more general case where nodes in B_n are attached to a growing number of leaves, which implies that $q_N = 0$ for all $N > 0$.

Corollary 3.1. *Suppose that Assumptions 2.1-2.3, and Assumption 3.1 hold. Suppose further that either of the following conditions holds:*

(i) There is a finite bound on the number of nodes that are not leaves.

(ii) We have $\min_{v \in B_n} l_n(v) \to \infty$.

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

The above corollary can be applied to Example 3.2. In that example, every level 1 node has m leaves attached to it, with m growing large as n increases. Therefore, the tree network satisfies condition (ii) in Corollary 3.1, and the optimal error exponent is $g^* = g_R^* = g_P^*$. In this case, even if the number N of level 1 nodes grows much faster than m, we still achieve the same error exponent as the parallel configuration. The above proposed strategy, in which every leaf uses the same transmission function, and every node uses the same LLRQ, will nearly achieve the optimal performance.

We are now in a position to determine the optimal error exponent in Example 3.4.

Example 3.4, revisited: Recall that in Example 3.4, every $v_i \in B_n$ has $i + 1$ of predecessors. It is easy to check that $z = 1$. From Proposition 3.3, the optimal error exponent is the same as that for the parallel configuration. □

3.3.5 Discussion of the Sufficient Conditions

Proposition 3.3 is unexpected as it establishes that the performance of a tree possessing certain qualitative properties is comparable to that of the parallel configuration. Furthermore, the optimal performance is obtained even if we restrict the non-leaf nodes to use 1-bit LLRQs. At first sight, it might appear intuitive that if the leaves dominate in a relay tree ($z = 1$), then the tree should always have the same performance as a parallel configuration. However, this intuition is misleading, as this is not the case for a Bayesian formulation, in which both the Type I and II error probabilities are required to decay at the same rate, is involved. To see this, consider the 2-uniform tree in Figure 3-1, where every node is constrained to sending 1-bit messages. Suppose we are given nonzero prior probabilities π_0 and π_1 for the hypotheses H_0 and H_1. Instead of the Neyman-Pearson criterion, suppose that we are interested in minimizing the error exponent

$$\limsup_{n\to\infty} \frac{1}{l_n(f)} \log P_e^*,$$

where P_e^* is the minimum of the error probability $\pi_0 \mathbb{P}_0(Y_f = 1) + \pi_1 \mathbb{P}_1(Y_f = 0)$, optimized over all strategies. It can be shown that to obtain the optimal error exponent, we only need to consider the following two fusion rules: (a) the fusion center declares H_0 iff both v_1 and v_2 send a 0, or (b) the fusion center declares H_1 iff both v_1 and v_2 send a 1. Then, using the results in Section 4.4 of Chapter 4, the optimal error exponent for this tree network is strictly worse than that for the parallel configuration. Similarly, if we constrain the Type I error in the Neyman-Pearson criterion to decay faster than a predetermined rate, it can be shown that the optimal Type II error exponent for a tree network can be strictly worse than that of a parallel configuration.

Note that the bounded height assumption is essential in proving $g^* = g_R^* = g_P^*$, when $z = 1$. Although our technique can be extended to include those tree sequences whose height grows very slowly compared to n (on the order of $\log|\log(n/l_n(f)-1)|$), we have not been able to find the optimal error exponent for the general case of unbounded height. We show in Chapter 6 that in a tandem network, the Bayesian error

probability decays sub-exponentially fast. The proof of Proposition 6.2 in Chapter 6 involves the construction of a tree network, with unbounded height, and in which $z = 1$. In that proof, it is also shown that such a network has a sub-exponential rate of error decay. We conjecture that this is also the case for the Neyman-Pearson formulation.

In summary, for a tree network to achieve the same Type II error exponent as a parallel configuration, we require that the tree sequence have a bounded height, satisfy the condition $z = 1$, and that the error criterion be the Neyman-Pearson criterion. Without any one of these three conditions, our results no longer hold.

3.3.6 A Necessary Condition for Matching the Performance of the Parallel Configuration

In this section, we establish necessary conditions under which a sequence of relay trees with bounded height performs as well as a parallel configuration. As noted in Section 3.1.2, any necessary conditions generally depend on the type of transmission functions available to the relay nodes. However, under an additional condition (Assumption 3.2), the sufficient condition for $g_R^* = g_P^*$ in Proposition 3.3 is also necessary.

Proposition 3.4. *Suppose that Assumptions 2.1-2.3, and Assumption 3.2 hold, and $h \geq 2$. If there exists some $N > 0$ such that $q_N > 0$ (equivalently, $z < 1$), then $g_P^* < g_R^*$.*

Proof. Fix some $N > 0$ and suppose that $q_N > 0$. Given a tree T_n, we construct a new tree T_n'', as follows. We remove all nodes other than the leaves and the nodes in $F_{N,n}$. For all the leaves u that are not immediate predecessors of some $v \in F_{N,n}$, we let u transmit its message directly to the fusion center. We add new edges (v, f), for each $v \in F_{N,n}$. This gives us a tree T_n'' of height 2, with $l_n''(f) = l_n(f)$ and $q_N'' = q_N$. The latter tree T_n'' can simulate the tree T_n, hence the optimal error exponent associated with the sequence $(T_n)_{n\geq 1}$ is bounded below by the optimal error exponent associated with the sequence $(T_n'')_{n\geq 1}$. Therefore, without loss of generality, we only need to prove the proposition for a sequence of trees of height 2, and in which $F_{N,n} = B_n$, for some

$N > 0$ such that $q_N > 0$; we henceforth assume that this is the case. The rest of the argument is similar to the proof of Stein's Lemma in Lemma 3.4.7 of [39]. Suppose that a particular admissible relay strategy has been fixed, and let β_n be the associated Type II error probability. Let $\lambda_n = \mathbb{E}_0[S_n(f)]/l_n(f)$. We show that $S_n(f)/l_n(f)$ is close to λ_n in probability. Let D_n be the set of leaves that transmit directly to the fusion center. The proof of the following lemma is in Section 3.5.

Lemma 3.5. *For all $\eta > 0$, $\mathbb{P}_0(|S_n(f)/l_n(f) - \lambda_n| > \eta) \to 0$, as $n \to \infty$.*

We return to the proof of Proposition 3.4. Given the transmission functions at all other nodes, the fusion center will optimize performance by using an appropriate likelihood ratio test, with a (possibly randomized) threshold. We can therefore assume, without loss of generality that this is the case. We let ζ_n be the threshold chosen, and note that it must satisfy

$$\mathbb{P}_0(S_n(f)/l_n(f) \leq \zeta_n) \geq 1 - \alpha. \tag{3.17}$$

From a change of measure argument (see Lemma 3.4.7 in [39]), we have for $\eta > 0$,

$$\begin{aligned}
&\frac{1}{l_n(f)} \log \beta^*(T_n) \\
&\geq \lambda_n - \eta + \frac{1}{l_n(f)} \log \mathbb{P}_0\Big(\lambda_n - \eta < \frac{S_n(f)}{l_n(f)} \leq \zeta_n\Big).
\end{aligned}$$

Using (3.17) and Lemma 3.5, we see that the last term goes to 0 as $n \to \infty$.We also have

$$\begin{aligned}
\lambda_n &= \frac{1}{l_n(f)}\Big(\sum_{v \in D_n} \mathbb{E}_0\big[\log \frac{\mathrm{d}\mathbb{P}_1^{\gamma_v}}{\mathrm{d}\mathbb{P}_0^{\gamma_v}}\big] + \sum_{v \in F_{N,n}} \mathbb{E}_0[L_{v,n}]\Big) \\
&\geq (1 - q_{N,n}) g_P^* + q_{N,n} K,
\end{aligned}$$

where, using the notation in Assumption 3.2,

$$K = \inf_{\substack{1<k\leq N \\ \xi \in \Gamma(k) \times \Gamma^k}} \frac{1}{k} \mathbb{E}_0\Big[\log \frac{\mathrm{d}\nu_1^\xi}{\mathrm{d}\nu_0^\xi}\Big] > g_P^*.$$

Then, letting $n \to \infty$, we have

$$g_R^* \geq (1 - q_N) g_P^* + q_N K - \eta,$$

for all $\eta > 0$. Taking $\eta \to 0$ completes the proof. □

The condition that there exists a finite N such that $l_n(v) \leq N$ for a non-vanishing proportion of nodes, in the statement of Proposition 3.4, can be thought of as corresponding to a situation where relay nodes are of two different types: high cost relays that can process a large number of received messages ($l_n(v) \to \infty$) and low cost relays that can only process a limited number of received messages ($l_n(v) \leq N$ for some small N). From this perspective, Proposition 3.4 states that a tree network of height greater than one, with a nontrivial proportion of low cost relays, will always have a performance worse than that of a parallel configuration.

Together with Proposition 3.3, we have shown the following.

Proposition 3.5. *Suppose that Assumptions 2.1-2.3, and Assumptions 3.1-3.2 hold. Then, $g_R^* = g_P^*$ iff $z = 1$ (or equivalently, iff $q_N = 0$ for all $N > 0$).*

Finally, we present an example in which $z < 1$ and $g^* < g_P^*$. Since there are also easy examples where $z < 1$ and $g_P^* < g^*$, this suggests that one can combine them to construct examples where $z < 1$ and $g^* = g_P^*$. Thus, unlike the case of a relay tree, $z = 1$ is not a necessary condition for $g^* = g_P$.

Example 3.5. *Consider the tree network shown in Figure 3-6, where every node makes a 3-bit observation. Each leaf then compresses its 3-bit observation to a 1-bit message, while each level 1 node is allowed to send a 4-bit message. (Recall that our framework allows for different transmission function sets $\Gamma(d)$ at the different levels.) We assume Assumptions 2.2-3.1 hold. Moreover, we assume that this network satisfies Assumption 3.2.*

Consider the following strategy: each level 1 node forwards the two 1-bit messages it receives from its two leaves to the fusion center. It then compresses its own 3-bit observation into a 2-bit message before sending it to the fusion center. Using this

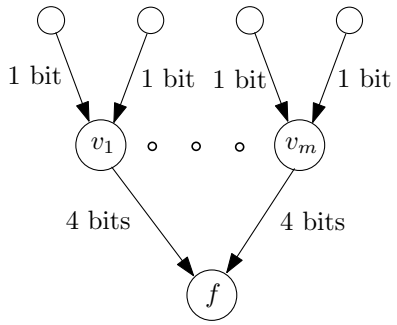

Figure 3-6: Every node makes a 3-bit observation. Leaves are constrained to sending 1-bit messages, while level 1 nodes are constrained to sending 4-bit messages.

strategy, the tree network is equivalent to a parallel configuration with $3m$ *nodes,* $2m$ *of which are constrained to sending 1-bit messages, and* m *of which are constrained to sending 2-bit messages. Clearly, this parallel configuration performs strictly better than one in which all* $3m$ *nodes are constrained to sending 1-bit messages, therefore we have* $g^* < g_P^*$. □

Example 3.5 shows that, unlike the case of relay trees, a tree can outperform a parallel configuration. On the other hand, Example 3.5 is an artifact of our assumptions. For example, if we restrict every node in this example to sending only 1 bit, the situation is reversed and we have $g_P^* < g^*$. The question of whether a parallel configuration always performs at least as well as a tree network, i.e., whether $g_P^* \leq g^*$, when every node can send the same number of bits, remains open.

3.4 Achieving the Type I Error Constraint

We have provided a strategy in Section 3.3.4 that allows a h-uniform tree sequence with $z = 1$, to achieve an error exponent to within ϵ of g_P^* (for any small $\epsilon > 0$). However, a large number of nodes may be required before the proposed strategy meets the Type I error constraint when h is large. In this section, we provide an upper bound for the number of nodes required before our proposed strategy is admissible. Clearly, this upper bound depends on the distributions under either hypothesis, so our benchmark for comparison is taken to be the number of nodes required when a similar

ϵ-optimal strategy is used in the parallel configuration. Throughout this section, we will consider only h-uniform tree sequences with $z = 1$ (since a height uniformization procedure can be performed on those tree sequences that are not height uniform).

We first record an elementary result. Let $\boldsymbol{\iota}_h$ be a vector of length h, and whose entries are all 1. Also, let $u(\gamma, t) = \Lambda^*_{0,1}(\gamma, t)/\Lambda^*_{1,1}(\gamma, t)$.

Lemma 3.6. *For $\gamma \in \Gamma$, and $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < t < \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)$, we have*

$$\frac{\Lambda^*_{0,h}(\gamma, t\boldsymbol{\iota}_h)}{\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h)} = u(\gamma, t)^{2^{h-1}}. \tag{3.18}$$

Furthermore, we have [1]

$$\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h) = \Lambda^*_{1,1}(\gamma, t) \prod_{k=0}^{h-2} \frac{1}{1 + u(\gamma, t)^{2^k}}, \tag{3.19}$$

$$\Lambda^*_{0,h}(\gamma, t\boldsymbol{\iota}_h) = \Lambda^*_{0,1}(\gamma, t) \prod_{k=0}^{h-2} \frac{u(\gamma, t)^{2^k}}{1 + u(\gamma, t)^{2^k}}. \tag{3.20}$$

Proof. From Lemma 3.1, we obtain

$$\begin{aligned}\frac{\Lambda^*_{0,h}(\gamma, t\boldsymbol{\iota}_h)}{\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h)} &= \frac{\Lambda^*_{0,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})(\Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1}) + t)}{\Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})(\Lambda^*_{0,h-1}(\gamma, t\boldsymbol{\iota}_{h-1}) - t)} \\ &= \Big(\frac{\Lambda^*_{0,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})}{\Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})}\Big)^2.\end{aligned}$$

Therefore, by induction, (3.18) holds. Lemma 3.1 also yields

$$\begin{aligned}\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h) &= \frac{\Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})}{\Lambda^*_{0,h-1}(\gamma, t\boldsymbol{\iota}_{h-1}) + \Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1})} \Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1}) \\ &= \frac{1}{1 + u(\gamma, t)^{2^{h-2}}} \Lambda^*_{1,h-1}(\gamma, t\boldsymbol{\iota}_{h-1}).\end{aligned}$$

The equation (3.19) then follows by induction. A similar argument shows (3.20), and the lemma is proved. □

We are interested in relating the Type I error exponent of a height uniform tree

[1] The products are taken to be 1 if $h = 1$.

sequence, with h greater than one, to that of a parallel configuration, when both configurations use ϵ-optimal strategies with the same Type II error exponent. Recall that in our proposed ϵ-optimal strategy, every leaf uses the same transmission function $\gamma \in \Gamma$, and every other node uses a LLRQ with the same threshold t. We shall call this the (γ, t) strategy. We consider only the case when ϵ is chosen sufficiently small so that $\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h) > \Lambda^*_{1,1}(\gamma, 0)$. There is little loss of generality in making this assumption, since we are typically interested in strategies that nearly achieve the optimal error exponent. For each n, let $\alpha_{h,n}(t)$ and $\beta_{h,n}(t)$ be the Type I and II error probabilities of a h-uniform network with n nodes, when the (γ, t) strategy is used.

Suppose that a h-uniform tree sequence uses the (γ, t) strategy. Since $\Lambda^*_{1,1}(\gamma, \cdot)$ is continuous, there exists an $s \in (-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), 0)$, such that

$$\Lambda^*_{1,1}(\gamma, s) = -\limsup_{n\to\infty} \frac{1}{n} \log \beta_{h,n}(t).$$

This implies that if a parallel configuration uses the (γ, s) strategy, then its error exponent is the same as a h-uniform tree sequence using the (γ, t) strategy. Given that both configurations have the same Type II error exponent, we can now compare their Type I error exponents. Let $a_h(t) = \limsup_{n\to\infty}(1/n) \log \alpha_{h,n}(t)$ be the Type I error exponent.

Proposition 3.6. *Suppose that Assumptions 2.1-2.3 hold. For $h > 1$, suppose a h-uniform tree sequence with $z = 1$ uses the (γ, t) strategy so that $\Lambda^*_{1,h}(\gamma, t\boldsymbol{\iota}_h) > \Lambda^*_{1,1}(\gamma, 0)$. Suppose also that a parallel configuration uses the (γ, s) strategy so that it has the same Type II error exponent as the h-uniform tree sequence. Then,*

$$a_h(t) \le K a_1(s),$$

where

$$K = \frac{\Lambda^*_{0,1}(\gamma, t)}{\Lambda^*_{0,1}(\gamma, s)} \prod_{k=0}^{h-2} \frac{u(\gamma, t)^{2^k}}{1 + u(\gamma, t)^{2^k}}.$$

Furthermore, for each $\delta \in (0,1)$, let $N_0(\delta)$ be such that for all $n \geq N_0(\delta)$,

$$0 \leq \frac{n}{l_n(f)} - 1 \leq \delta \min\{K|a_1(s)|, 1\}.$$

Then, if $n \geq \inf_{\delta\in(0,1)} \max\left\{N_0(\delta), \frac{\log\alpha}{a_1(s)}\frac{1+\delta}{K(1-\delta)}\right\}$, we have $\alpha_{h,n} \leq \alpha$.

Proof. From Proposition 3.1 and Lemma 3.6, we have

$$a_h(t) \leq -\Lambda^*_{0,h}(\gamma, t\boldsymbol{\iota}_h) = -K\Lambda^*_{0,1}(\gamma, s) = Ka_1(s),$$

where the last equality follows because $a_1(s) = -\Lambda^*_{0,1}(\gamma, s)$ (cf. Theorem 2.1).

Fix a $\delta \in (0,1)$, and consider $n \geq \max\left\{N_0(\delta), \frac{\log\alpha}{a_1(s)}\frac{1+\delta}{K(1-\delta)}\right\}$. From Proposition 3.1, we have

$$\begin{aligned}
\alpha_{h,n} &\leq \exp\Big(l_n(f)\Big(-\Lambda^*_{0,h}(\gamma, t\boldsymbol{\iota}_h) + \frac{n}{l_n(f)} - 1\Big)\Big) \\
&\leq \exp\Big(\frac{n}{1+\delta}\Big(Ka_1(s) + \frac{n}{l_n(f)} - 1\Big)\Big) \\
&\leq \exp\Big(\frac{n}{1+\delta}\Big(Ka_1(s) - \delta Ka_1(s)\Big)\Big) \\
&= \exp\Big(\frac{n}{1+\delta}(1-\delta)Ka_1(s)\Big) \leq \alpha.
\end{aligned}$$

The proposition is now proved. □

From Theorem 2.1, we have $\alpha_{1,n} \leq \exp(na_1(s))$ so that if $n \geq \log\alpha/a_1(s)$, the Type I error constraint is satisfied. Let us take $n_0 = \log\alpha/a_1(s)$ as an estimate of the required number of nodes for the (γ, s) strategy to be admissible in the parallel configuration. Then, if α is very small (e.g., on the order of 10^{-6}), Proposition 3.6 shows that approximately at most n_0/K nodes are required for the h-uniform tree sequence to meet the Type I error constraint. Moreover, since $s < 0$, we have $\Lambda^*_{0,1}(\gamma, s) \geq \Lambda^*_{0,1}(\gamma, 0)$, yielding the following lower bound for K:

$$K \geq K^0 := \frac{\Lambda^*_{0,1}(\gamma, t)}{\Lambda^*_{0,1}(\gamma, 0)} \prod_{k=0}^{h-2} \frac{u(\gamma, t)^{2^k}}{1 + u(\gamma, t)^{2^k}}. \tag{3.21}$$

Taking the reciprocal of the R.H.S. of (3.21) yields an upper bound for the number of nodes required for the (γ, t) strategy to be admissible. This upper bound applies for all tree sequences with the same height h, regardless of the network architecture.

We next consider some numerical examples below to verify our conclusions. Suppose that leaves make i.i.d. observations with Bernoulli distributions that have parameters $p_0 = 0.2$ under H_0, and $p_1 = 0.8$ under H_1. For simplicity, we assume that leaves transmit their 1-bit observations in the raw to the level 1 nodes. We consider 2-uniform tree sequences similar to that in Example 3.2. Suppose there are N level 1 nodes, each with m leaves attached. In the first case, we fix the number of relay nodes to be $N = 4$. In the second case, we let $N = m^2$ (this mirrors the case of approximating a hop constrained minimum spanning tree in the unit square; cf. Section 4.1). The thresholds used in the LLRQs for each case are given in Table 3.1 below. In all cases, the Type I error constraint is set at $\alpha = 10^{-6}$.

N	t	s	$1/K$
4	-0.4318	-0.4312	9.6547
m^2	-0.4318	-0.3714	12.5625

Table 3.1: The threshold t is used in the 2-uniform tree network, while the threshold s is chosen so that the corresponding parallel configuration has the same Type II error exponent. The ratio of the number of nodes required for the 2-uniform tree sequence to achieve the Type I error constraint, compared to that for the parallel configuration, is approximately upper bounded by $1/K$.

The Type II error probability for the case $N = 4$ is shown in Figure 3-7. We have chosen the threshold s so that the Type II error exponent of the parallel configuration is the same as that of the tree sequence, as verified by Figure 3-7.

From Figure 3-8, the parallel configuration requires approximately 198 nodes before the Type I error constraint is met. In the 2-uniform tree sequence, approximately 900 nodes are required. This is well within $1/K = 9.6547$ times of 198 nodes. In the case $N = m^2$, we require 155 nodes and 1338 nodes for the parallel configuration and the 2-uniform tree sequence respectively (cf. Figure 3-9). Again, $1/K = 12.5625$ is an upper bound for the ratio of nodes in the tree network to that in the parallel configuration. These examples verify that $1/K$ gives an approximate upper bound for

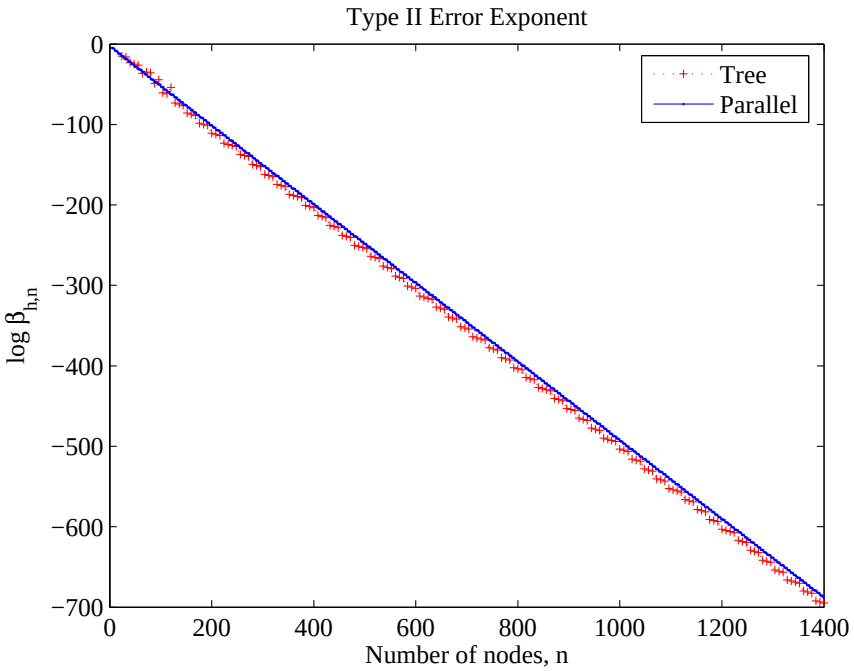

Figure 3-7: Plot of $\log \beta_{h,n}$ vs. n.

determining the number of nodes required so that our proposed strategy is admissible in a tree architecture.

The case $N = m^2$ mirrors the tree network constructed in the example in Section 4.1. As shown in that example, a tree network with a similar architecture as the case where $N = m^2$, and with a total of n nodes, has a transmission cost of $c_1 n^{2/3}$, where c_1 is a positive constant. A parallel configuration with n nodes, on the other hand, has a transmission cost of $c_2 n$, where c_2 is another positive constant. Even though the tree network requires approximately 10 times as many nodes for our strategy to be admissible (as shown numerically above), it is easy to see that if $n > 100 \cdot (c_1/c_2)^3$, the tree network with $10n$ nodes has a lower transmission cost than a parallel configuration with n nodes. As the sensor networks we are interested in have hundreds or thousands of nodes, this example shows that our proposed strategy results in an energy efficient tree network, although the number of nodes required in the tree network may be many times more than that for a parallel configuration. Moreover, as shown in Section 3.3, our proposed strategy nearly achieves the optimal

error exponent in this tree network.

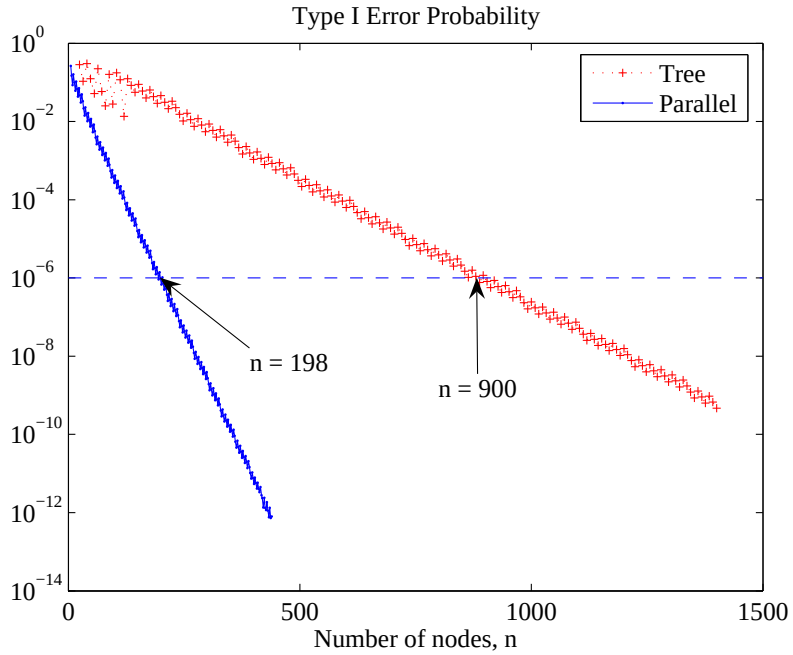

Figure 3-8: Plot of Type I error probabilities for $N = 4$.

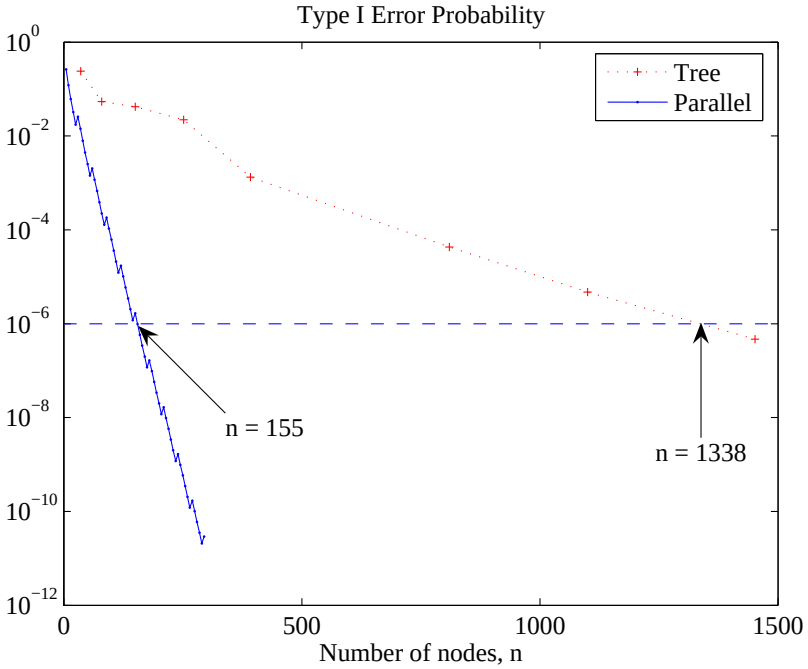

Figure 3-9: Plot of Type I error probabilities for $N = m^2$.

3.5 Proofs

In this section, we prove some of the results in this chapter.

Proof of Proposition 3.1.

We first show part (i). The proof proceeds by induction on k. Suppose that $k = 1$, which is equivalent to the well-studied case where all sensors transmit directly to a fusion center. In this case, $p_n(v) = l_n(v)$. Since $t_1 \in (-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma))$, from (2.2.13) of [39], we obtain

$$\frac{1}{l_n(v)} \log \mathbb{P}_1\Big(\frac{S_n(v)}{l_n(v)} \leq t_1\Big) \leq -\Lambda_{1,1}^*(\gamma, t_1).$$

The inequality for the Type I error probability follows from a similar argument.

Consider now the induction hypothesis that the result holds for some k. Given a k-uniform tree rooted at v, the induction hypothesis leads to bounds on the probabilities associated with the log-likelihood ratio $L_{v,n}$ of the message Y_v computed at the node v. We use these bounds to obtain bounds on the log-moment generating function of $L_{v,n}$. Recall that $L_{v,n}$ equals $\mathcal{L}_{v,n}(0)$ whenever $Y_v = 0$, which is the case if and only if $S_n(v)/l_n(v) \leq t_k$. Fix some $\lambda \in [-1, 0]$. We have

$$\begin{aligned}
&\frac{1}{l_n(v)} \log \mathbb{E}_1\big[e^{\lambda L_{v,n}}\big] \\
&= \frac{1}{l_n(v)} \log \Big[\mathbb{P}_1(Y_v = 0) e^{\lambda \mathcal{L}_{v,n}(0)} + \mathbb{P}_1(Y_v = 1) e^{\lambda \mathcal{L}_{v,n}(1)}\Big] \\
&= \frac{1}{l_n(v)} \log \Big[\mathbb{P}_1(Y_v = 0)^{1+\lambda} \mathbb{P}_0(Y_v = 0)^{-\lambda} + \mathbb{P}_1(Y_v = 1)^{1+\lambda} \mathbb{P}_0(Y_v = 1)^{-\lambda}\Big] \\
&\leq \frac{1}{l_n(v)} \log \Big[\mathbb{P}_1(Y_v = 0)^{1+\lambda} + \mathbb{P}_0(Y_v = 1)^{-\lambda}\Big].
\end{aligned}$$

Using the inequality $\log(a + b) \leq \max\{\log(2a), \log(2b)\}$, we obtain

$$\begin{aligned}
&\frac{1}{l_n(v)} \log \mathbb{E}_1\big[e^{\lambda L_{v,n}}\big] \\
&\leq \max\Big\{\frac{1+\lambda}{l_n(v)} \log \mathbb{P}_1(Y_v = 0), -\frac{\lambda}{l_n(v)} \log \mathbb{P}_0(Y_v = 1)\Big\} + \frac{\log 2}{l_n(v)}
\end{aligned}$$

$$\leq \max\big\{-(1+\lambda)\Lambda^*_{1,k}(\gamma, t^{(k)}), \lambda\Lambda^*_{0,k}(\gamma, t^{(k)})\big\} + \frac{p_n(v)}{l_n(v)} - 1 + \frac{\log 2}{l_n(v)} \tag{3.22}$$

$$\leq \Lambda_{1,k}(\gamma, t^{(k)}; \lambda) + \frac{p_n(v)}{l_n(v)} + \frac{1}{l_n(v)} - 1, \tag{3.23}$$

where (3.22) follows from the induction hypothesis.

Consider now a node u at level $k+1$. The subtree rooted at u is a $(k+1)$-uniform tree. Each level k node $v \in C_n(u)$ can be viewed as the root of a k-uniform tree and Eq. (3.23) can be applied to $L_{v,n}$. From the Markov Inequality, and since $\lambda \in [-1, 0]$, we have

$$\mathbb{P}_1\Big(\frac{S_n(u)}{l_n(u)} \leq t_{k+1}\Big) \leq e^{-\lambda l_n(u) t_{k+1}} \mathbb{E}_1\big[e^{\lambda S_n(u)}\big],$$

so that

$$\begin{aligned}
&\frac{1}{l_n(u)} \log \mathbb{P}_1\Big(\frac{S_n(u)}{l_n(u)} \leq t_{k+1}\Big) \\
&\leq -\lambda t_{k+1} + \frac{1}{l_n(u)} \sum_{v \in C_n(u)} \log \mathbb{E}_1\big[e^{\lambda L_{v,n}}\big] \\
&= -\lambda t_{k+1} + \sum_{v \in C_n(u)} \frac{l_n(v)}{l_n(u)} \cdot \frac{1}{l_n(v)} \log \mathbb{E}_1\big[e^{\lambda L_{v,n}}\big] \\
&\leq -\lambda t_{k+1} + \Lambda_{1,k}(\gamma, t^{(k)}; \lambda) + \sum_{v \in C_n(u)} \frac{p_n(v)}{l_n(u)} + \frac{|C_n(u)|}{l_n(u)} - 1 \qquad (3.24) \\
&= -\lambda t_{k+1} + \Lambda_{1,k}(\gamma, t^{(k)}; \lambda) + \frac{p_n(u)}{l_n(u)} - 1, \qquad (3.25)
\end{aligned}$$

where (3.24) follows from the induction hypothesis and (3.23). Taking the infimum over $\lambda \in [-1, 0]$ (cf. Lemma 3.1), and using (3.7), we obtain

$$\frac{1}{l_n(u)} \log \mathbb{P}_1\Big(\frac{S_n(u)}{l_n(u)} \leq t_{k+1}\Big) \leq -\Lambda^*_{1,k+1}(\gamma, t^{(k+1)}) + \frac{p_n(u)}{l_n(u)} - 1.$$

A similar argument proves the result for the Type I error probability, and the proof of part (i) is complete.

For part (ii), suppose that for all $n \geq n_0$ and all $v \in B_n$, we have $l_n(v) \geq N$. Note that $l_n(f) \geq N|B_n|$. Furthermore, the number of nodes at each level $k \geq 1$ is

bounded by $|B_n|$, which yields

$$\frac{p_n(f)}{l_n(f)} - 1 \leq \frac{n}{l_n(f)} - 1 = \frac{n - l_n(f)}{l_n(f)} \leq \frac{h|B_n|}{N|B_n|} = \frac{h}{N}.$$

Applying the results from part (i), with $k = h$, we obtain part (ii).

Proof of Lemma 3.2.

We have $l'_n(f) = l_n(f)$. Furthermore, it can be shown that $|B'_n| \leq h|B_n|$. Therefore,

$$\begin{aligned} q'_{N,n} = \frac{1}{l'_n(f)} \sum_{v \in F'_{N,n}} l'_n(v) &\leq \frac{1}{l_n(f)} N|B'_n| \\ &\leq \frac{1}{l_n(f)} Nh\big(|F_{M,n}| + |F^c_{M,n}|\big) \\ &\leq hNq_{M,n} + hN/M, \end{aligned}$$

where the last inequality follows from $|F_{M,n}| \leq \sum_{v \in F_{M,n}} l_n(v)$ and $|F^c_{M,n}| \leq l_n(f)/M$. Taking the limit superior as $n \to \infty$, we obtain

$$q'_N \leq h(Nq_M + N/M).$$

Suppose that $q_M = 0$ for all $M > 0$. Then for all $N, M > 0$, we have

$$q'_N \leq hN/M.$$

Taking $M \to \infty$, we obtain the desired result.

Proof of Lemma 3.3.

Suppose that $q_N > 0$ for some $N > 0$. Using the inequality

$$q_{N,n} = \frac{1}{l_n(f)} \sum_{v \in F_{N,n}} l_n(v) \leq \frac{N|F_{N,n}|}{l_n(f)},$$

or

$$|F_{N,n}| \geq \frac{q_{N,n}}{N} l_n(f), \tag{3.26}$$

we obtain

$$\begin{aligned}\frac{l_n(f)}{n} &\leq \frac{l_n(f)}{|F_{N,n}| + l_n(f)} \\ &\leq \frac{l_n(f)}{q_{N,n} l_n(f)/N + l_n(f)} \\ &= \frac{N}{N + q_{N,n}}.\end{aligned}$$

Letting $n \to \infty$, we obtain

$$z \leq \frac{N}{N + q_N} < 1.$$

For the converse, suppose that $q_N = 0$ for all $N > 0$. It can be seen that each non-leaf node is on a path that connects some $v \in B_n$ to the fusion center. Therefore, the number of non-leaf nodes $n - l_n(f)$ is bounded by $h|B_n|$. We have

$$\frac{n - l_n(f)}{l_n(f)} \leq \frac{h|B_n|}{l_n(f)} = h\frac{|F_{N,n}| + |F^c_{N,n}|}{l_n(f)} \leq hq_{N,n} + \frac{h}{N}.$$

Therefore,

$$\limsup_{n\to\infty} \frac{n - l_n(f)}{l_n(f)} \leq \frac{h}{N}.$$

This is true for all $N > 0$, which implies that $\lim_{n\to\infty} l_n(f)/n = 1$.

Proof of Lemma 3.5.

For each $v \in B_n$, we have $Y_v = \gamma_v(\{\gamma_u(X_u) : u \in C_n(v)\})$, for some $\gamma_v \in \Gamma(l_n(v))$. Using the first, and the second part of Lemma 2.1, there exists some $a_1 \in (0, \infty)$, such that

$$\begin{aligned}\mathbb{E}_0[L^2_{v,n}] &\leq \mathbb{E}_0\Big[\Big(\sum_{u\in C_n(v)} \log \frac{\mathrm{d}\mathbb{P}_1^{\gamma_u}}{\mathrm{d}\mathbb{P}_0^{\gamma_u}}\Big)^2\Big] + 1 \\ &\leq l_n(v)\mathbb{E}_0\Big[\sum_{u\in C_n(v)} \log^2 \frac{\mathrm{d}\mathbb{P}_1^{\gamma_u}}{\mathrm{d}\mathbb{P}_0^{\gamma_u}}\Big] + 1 \\ &\leq l_n^2(v)a_1 + 1\end{aligned}$$

$$\leq l_n^2(v)a, \tag{3.27}$$

where $a = a_1 + 1$.

To prove the lemma, we use Chebychev's inequality, and the inequalities $l_n(v) \leq N$ for $v \in F_{N,n}$, and $|D_n| \leq l_n(f)$, to obtain

$$\begin{aligned}
&\mathbb{P}_0\Big(\big|\frac{S_n(f)}{l_n(f)} - \lambda_n\big| > \eta\Big) \\
&\leq \frac{1}{\eta^2 l_n^2(f)}\Big(\sum_{v\in D_n} \mathbb{E}_0\big[\log^2 \frac{\mathrm{d}\mathbb{P}_1^{\gamma_v}}{\mathrm{d}\mathbb{P}_0^{\gamma_v}}\big] + \sum_{v\in F_{N,n}} \mathbb{E}_0[L_{v,n}^2]\Big) \\
&\leq \frac{1}{\eta^2 l_n^2(f)}\Big(\sum_{v\in D_n} a + \sum_{v\in F_{N,n}} l_n^2(v)a\Big) \qquad (3.28) \\
&\leq \frac{a}{\eta^2 l_n(f)} + \frac{a}{\eta^2 l_n(f)} \sum_{v\in F_{N,n}} \frac{l_n(v)}{l_n(f)} N \\
&\leq \frac{a(1+N)}{\eta^2 l_n(f)}, \qquad (3.29)
\end{aligned}$$

where (3.28) follows from Lemma 2.1 and (3.27). The R.H.S. of (3.29) goes to zero as $n \to \infty$, and the proof is complete.

Chapter 4

Bayesian Detection in Bounded Height Tree Networks

In Chapter 3, we considered the Neyman-Pearson detection problem for tree networks with bounded height. In this chapter, we continue this investigation by focusing on the Bayesian formulation.

We start with an example that serves to motivate some of our assumptions. We then show that for bounded height trees, error probabilities decay exponentially fast. However, the optimal error exponent is generically worse than the one associated with a parallel configuration, and is also harder to characterize exactly. In order to make further progress, we place some additional restrictions on the allowed quantization strategies, and consider tree sequences with symmetric architectures. We analyze two different classes of symmetric tree sequences, and compare their detection performances. Our results suggest that for tree sequences with the same height, having more immediate predecessors for each relay node deteriorates performance.

4.1 An Example Network

Suppose that we distribute n nodes randomly in the unit square and place a fusion center at the center of the square. We are interested in configuring the nodes so that every node is at most two hops away from the fusion center.

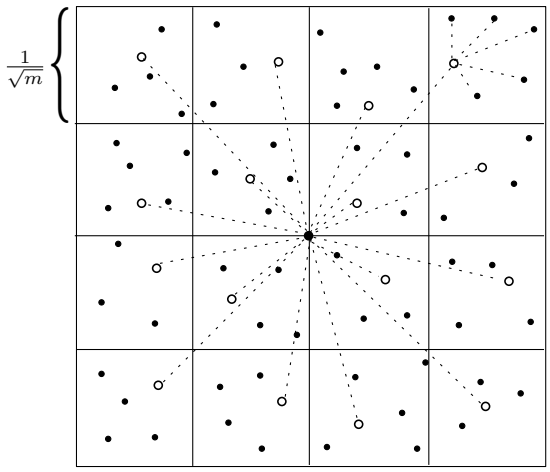

Figure 4-1: Random nodes in a unit square. The hollow circles represent the local aggregators. The dotted lines represent communication links. Only one sub-square is shown with its communication links.

One possibility (to be referred to as Design I) is to fix some m, and divide the square into m sub-squares, each with side of length $1/\sqrt{m}$ (see Figure 1). For large n, there are approximately n/m nodes in each of these sub-squares. We let all the nodes within a sub-square transmit their messages to an "aggregator" node in that sub-square. In this way, we get a "symmetric" tree network, in which every aggregator is connected to roughly the same number of nodes, with high probability. Suppose now that the communication cost is proportional to the Euclidean distance between two communicating nodes. Since the number m is fixed, the communication cost in this strategy is $\Theta(n)$.

An alternative possibility (to be referred to as Design II) is to minimize the overall communication cost by using a 2-hop Minimum Spanning Tree (MST). Because finding a hop-constrained MST is NP-hard (see [41]), let us consider a simple heuristic [42]. As before, we place an aggregator in each of the m sub-squares, and let the rest of the nodes in the sub-square send their observations to this aggregator. However, we do not fix m in this strategy. The overall expected cost is

$$F(m) = m\big(a_1 \frac{n}{m} \frac{1}{\sqrt{m}}\big) + ma_2 = \frac{a_1 n}{\sqrt{m}} + ma_2,$$

where a_1 and a_2 are nonzero constants. To minimize $F(m)$, a simple calculation shows that we should take $m = m(n) = \Theta(n^{2/3})$, which reduces the cost from $\Theta(n)$ in Design I, to $\Theta(n^{2/3})$. On the other hand, one suspects that the detection performance of Design II will be inferior to that of Design I. The results in Lemma 4.3 and Proposition 4.1 provide evidence that this is indeed the case.

Motivated by the two designs introduced above, we will consider the detection performance of two different classes of tree networks. The first class of tree networks consists of symmetric trees with a fixed number of aggregators or relay nodes, while the second class of tree networks consists of tree architectures in which the number of relay nodes increases at a certain rate (we call these the rapidly branching tree sequences; cf. Section 4.5). We characterize the detection performance of both classes over a restricted set of strategies that are easy to implement in practice in Sections 4.4 and 4.5, and compare the performance of these two classes. We show that the second class performs worse than any of the tree networks in the first class, in Proposition 4.1.

4.2 Problem Formulation

Suppose that each hypothesis H_j has positive prior probability π_j. Given a tree network T_n, a typical goal is to minimize the probability of error $P_e(T_n) = \pi_0 \mathbb{P}_0(Y_f = 1) + \pi_1 \mathbb{P}_1(Y_f = 0)$, over all strategies. Let $P_e^*(T_n)$ be the minimum probability of error (over all strategies) at the fusion center. In a sequence of trees $(T_n)_{n \geq 1}$, we seek to characterize the optimal error exponent [39],

$$\begin{aligned} \mathcal{E}^* &= \limsup_{n \to \infty} \frac{1}{n} \log P_e^*(T_n) \\ &= \max \Big\{ \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_0(Y_f = 1), \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_1(Y_f = 0) \Big\}. \end{aligned}$$

Similar to the Neyman-Pearson formulation, for a relay tree, we consider instead the optimal error exponent,

$$\mathcal{E}_R^* = \limsup_{n\to\infty} \frac{1}{l_n(f)} \log P_e^*(T_n).$$

In addition to Assumptions 2.1-2.2, we will also make the following assumption. A prime denotes differentiation w.r.t. λ, and a double prime is the second derivative w.r.t. λ.

Assumption 4.1. *Both* $\mathrm{D}(\mathbb{P}_0^X \,\|\, \mathbb{P}_1^X)$ *and* $\mathrm{D}(\mathbb{P}_1^X \,\|\, \mathbb{P}_0^X)$ *are finite, and there exists some* $b \in (0,\infty)$*, such that for all* $\gamma \in \Gamma$*,* $\Lambda''_{0,0}(\gamma;\lambda) \le b$*, for all* $\lambda \in (0,1)$*, and* $\Lambda''_{1,0}(\gamma;\lambda) \le b$*, for all* $\lambda \in (-1,0)$*.*

4.3 Exponential Decay

In this section, we show that the optimal error probability in a sequence of trees with bounded height h decays exponentially fast with the number of nodes n. (This is in contrast to general trees, where the decay can be sub-exponential, cf. Chapter 6.) When $h = 1$, we have the classical parallel configuration considered in [22], and the optimal error exponent is given by

$$\mathcal{E}_P^* = -\sup_{\gamma\in\Gamma} \Lambda_{0,1}^*(\gamma, 0) = \inf_{\gamma\in\Gamma} \min_{\lambda\in[0,1]} \Lambda_{0,0}(\gamma;\lambda) < 0. \tag{4.1}$$

Also, recall that $z = \liminf_{n\to\infty} l_n(f)/n$ is the asymptotic proportion of nodes that are leaves.

Theorem 4.1. *Suppose that Assumptions 2.1, 2.2 and 4.1 hold. Consider any sequence of trees of height* h*. Then,*

$$\mathcal{E}_P^* \le \mathcal{E}_R^* < 0, \tag{4.2}$$

and

$$\min_{\lambda\in[0,1]} \log \mathbb{E}_0\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\Big)^{\lambda}\Big] \le \mathcal{E}^* \le z\mathcal{E}_R^* < 0. \tag{4.3}$$

Furthermore, if $z = 1$, we have

$$\mathcal{E}_P^* \le \mathcal{E}^* \le \mathcal{E}_R^* \le \frac{1}{2^{h-1}}\mathcal{E}_P^*. \tag{4.4}$$

Proof. (Outline) We first note that $\mathcal{E}_R^* \ge \mathcal{E}_P^*$ holds trivially since a parallel configuration can simulate any relay tree network. Also, since a relay strategy is a possible strategy for any tree network, $z\mathcal{E}_R^* \ge \mathcal{E}^*$; the factor z is because we have normalized $\mathcal{E}_R^*$ with $l_n(f)$ instead of n. Furthermore, the first inequality in (4.3) is easily arrived at when comparing a tree network to one in which raw observations at every node is available at the fusion center. Finally, since $z \ge 1/h$, it remains to show $\mathcal{E}_R^* < 0$ and (4.4).

We first show $\mathcal{E}^* \ge \mathcal{E}_P^*$ when $z = 1$. Compare the tree network T_n to another network T_n' in which all the non-leaf nodes send their raw observations to the fusion center, and all the leaves send their messages directly to the fusion center. Clearly, T_n' can simulate T_n, and has at least as good performance as T_n. Moreover, T_n' is now a parallel configuration, and using an argument similar to that in [22], we obtain $\mathcal{E}^* \ge \mathcal{E}_P^*$.

Next, we show that $\mathcal{E}_R^* \le \mathcal{E}_P^*/2^{h-1} < 0$ for $z = 1$, and $\mathcal{E}_R^* < 0$ for $z < 1$. Any tree of height h can be transformed to an h-uniform tree using the height uniformization procedure described in Section 3.3.1. For the same reason as in Section 3.3.1, it is sufficient to show the inequalities for the height uniform counterpart of $(T_n)_{n\ge1}$, hence we can without loss of generality assume that the sequence of trees is h-uniform. We consider two cases below.

When $z = 1$, let all leaves use the same transmission function $\gamma \in \Gamma$ such that (3.9) is satisfied, and let all non-leaf nodes use a LLRQ with threshold 0 (it is easy

to verify that this satisfies (3.10)-(3.11)). Then, from Proposition 3.1, we have

$$\mathcal{E}_R^* \leq -\min\{\Lambda_{0,h}^*(\gamma, \mathbf{0}), \Lambda_{1,h}^*(\gamma, \mathbf{0})\},$$

where $\mathbf{0}$ is a vector of h 0s. From Lemma 3.1, we have

$$\Lambda_{0,h}^*(\gamma, \mathbf{0}) = \Lambda_{1,h}^*(\gamma, \mathbf{0}) = \frac{1}{2^{h-1}}\Lambda_{0,1}^*(\gamma, 0).$$

Taking the infimum over all $\gamma \in \Gamma$, we obtain the desired bound using (4.1).

We now consider a sequence of trees $(T_n)_{n\geq 1}$ with $z < 1$, and show that $\mathcal{E}_R^* < 0$. We give an outline of the proof here, since it is similar to that of Proposition 3.2. Recall that $F_{N,n}$ is the set of level 1 nodes in T_n, with less than N leaves attached. Since $z < 1$, there exists a positive N such that $\liminf_{n\to\infty} |F_{N,n}|/l_n(f) > 0$. We form a new tree T_n' by removing all leaves, and allowing each level 1 node to make its own observation. The new tree, of height $h-1$, has performance no better than the original:

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{l_n(f)} \log P_e^*(T_n) &\leq \limsup_{n\to\infty} \frac{1}{l_n(f)} \log P_e^*(T_n') \\ &\leq \liminf_{n\to\infty} \frac{|F_{N,n}|}{l_n(f)} \limsup_{n\to\infty} \frac{1}{|B_n|} \log P_e^*(T_n'),\end{aligned}$$

where B_n is the set of level 1 nodes. The desired conclusion then follows by induction on h. This concludes our proof of the theorem. □

We have shown that the rate of error probability decay in any bounded height tree network is exponential in n. However, the exact rate of decay depends on several factors, such as the probability distributions and the architecture of the network. For example, in architectures that are essentially the same as the parallel configuration or can be reduced to the parallel configuration (cf. Figure 1-1), the error exponent is $\mathcal{E}_P^*$. However, in most other cases, the error exponent is apparently strictly less than $\mathcal{E}_P^*$. To obtain some insights into the optimal error exponent, we consider specific classes of h-uniform tree networks in the next two sections. However, it turns out

that finding optimal strategies is in general difficult, so we will instead analyze simple, but suboptimal strategies.

4.4 Symmetric Tree Sequences

In this section, we consider the asymptotic performance of a special class of h-uniform tree networks, which we call r-symmetric. These are relay trees, with a bounded number of relay nodes, as in Design I in Section 4.1. Throughout this section, we assume that nodes can only send binary messages. An r-symmetric tree network is defined as follows.

Definition 4.1 (r-symmetric tree)**.** *For $h, r \geq 1$, a h-uniform tree sequence $(T_n)_{n\geq 1}$ is said to be r-symmetric if:*

(i) for all level k nodes v, where $k > 1$, $|C_n(v)| = r$, and

(ii) for all level 1 nodes v, $l_n(v)/l_n(f) \to 1/r^{h-1}$ as $n \to \infty$.

The second condition in this definition requires that when n is large, all the r^{h-1} level 1 nodes have approximately the same number of immediate predecessors.

We define a *counting quantizer (CQ)* with threshold s for a level k node v, where $k \geq 1$, as a transmission function of the form

$$Y_v = \begin{cases} 0, & \sum_{u\in C_n(v)} Y_u \leq s, \\ 1, & \text{otherwise}, \end{cases}$$

where $\sum_{u\in C_n(v)} Y_u$ is the total number of 1s that v receives from its immediate predecessors. A counting quantizer is arguably the simplest quantizer that can be implemented. A counting quantizer is equivalent to a LLRQ with an appropriate threshold, if all the messages of v's immediate predecessors are identically distributed.

It is well known that to minimize the probability of error at the fusion center, there is no loss in optimality if we restrict to LLRQs at the relay nodes [6]. Given the symmetry of our tree network, it is easy to see that if we restrict leaves to

using the same transmission function, then there is no loss in optimality if counting quantizers are used at every relay node. Without the above restriction on the leaves, it is unclear what the optimal strategy is. For tractability and to ensure that our strategies are easily implementable, we will now restrict all non-leaf nodes to using counting quantizers. We call such a strategy a *counting* strategy. Let $\mathcal{E}_S^*(r)$ denote the optimal (over all counting strategies) error exponent (in the worst-case over all r-symmetric tree sequences). We will show that with the restriction to counting strategies, using the same transmission function at the leaves results in no loss of optimality.

Consider minimizing the following objective function, [1]

$$\max\{\lambda_1 \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_0(Y_f = 1), \lambda_2 \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_1(Y_f = 0)\}, \tag{4.5}$$

where λ_1 and λ_2 are fixed positive constants. In the case of minimizing the error exponent, $\lambda_1 = \lambda_2 = 1$. We use this more general formulation because it proves to be useful below. We start with two preliminary lemmas, the first of which is proved in [22]. We provide an outline of the proof here for completeness.

Lemma 4.1. *Suppose that Assumptions 2.1, 2.2 and 4.1 hold. Consider minimizing the objective function* (4.5) *at the fusion center of a parallel configuration. Then, there is no loss in optimality if we restrict all nodes to using the same transmission function, and the fusion rule to be a counting quantizer.*

Proof. (Outline) Suppose that there are n nodes in the parallel configuration, sending messages to the fusion center using transmission functions $\gamma_1, \ldots, \gamma_n$. For any strategy, from the Neyman-Pearson Lemma, there exists a LLRQ with some threshold t such that

$$\begin{aligned}
&\max\{\frac{\lambda_1}{n} \log \mathbb{P}_0(Y_f = 1), \frac{\lambda_2}{n} \log \mathbb{P}_1(Y_f = 0)\} \\
&\geq \max\{\frac{\lambda_1}{n} \log \mathbb{P}_0\big(S_n(f)/l_n(f) > t\big), \frac{\lambda_2}{n} \log \mathbb{P}_1(S_n(f)/l_n(f) \leq t\big)\}
\end{aligned}$$

[1] We use the notation lim here, without first showing that the limit exists. The subsequent arguments can be made completely rigorous by considering a subsequence of the tree sequence, in which limits of the Type I and II error exponents exist at each non-leaf node.

$$= \max\{\frac{\lambda_1}{n}\sum_{i=1}^{n}\Lambda_{0,1}^*(\gamma_i,t), \frac{\lambda_2}{n}\sum_{i=1}^{n}\Lambda_{1,1}^*(\gamma_i,t)\} + o(1)$$
$$\geq \inf_{\gamma\in\Gamma,t\in\mathbb{R}} \max\{\lambda_1\Lambda_{0,1}^*(\gamma,t), \lambda_2\Lambda_{1,1}^*(\gamma,t)\} + o(1),$$

where $o(1)$ denotes a term that goes to 0, as n increases, and the first inequality follows from Theorem 2.1. Since the lower bound can be achieved arbitrarily closely even when nodes are restricted to using the same transmission function, and the fusion rule is restricted to the class of counting quantizers, the lemma is proved. □

Let the immediate predecessors of the fusion center f be $v_1, \ldots, v_r$, and let

$$\psi_i = -\lim_{n\to\infty}\frac{1}{l_n(v_i)}\log\mathbb{P}_0(Y_{v_i}=1),$$
$$\varphi_i = -\lim_{n\to\infty}\frac{1}{l_n(v_i)}\log\mathbb{P}_1(Y_{v_i}=0).$$

Without loss of generality, we can assume that

$$0 \leq \psi_1 \leq \psi_2 \leq \ldots \leq \psi_r < \infty, \tag{4.6}$$
$$\infty > \varphi_1 \geq \varphi_2 \geq \ldots \geq \varphi_r \geq 0. \tag{4.7}$$

Furthermore, if $\psi_i > \psi_j$, then $\varphi_i < \varphi_j$ and vice versa, for all i, j. The reason why there is no loss of generality is because if there exists $i \neq j$ with $\psi_i \leq \psi_j$ and $\varphi_i \leq \varphi_j$, then for n sufficiently large, we can use as the strategy for the sub-tree rooted at v_i, the same strategy that is used for the sub-tree rooted at v_j, and not increase the objective function at the fusion center.

Lemma 4.2. *To minimize the objective function (4.5) at the fusion center using a counting quantizer as the fusion rule, there is no loss in optimality if we restrict all immediate predecessors of f so that $\psi_i = \psi_j$, and $\varphi_i = \varphi_j$ for all i, j.*

Proof. Suppose the fusion center uses a counting quantizer with threshold s. Then,

we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_f = 1) \\
&= \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(\sum_{i=1}^{r} Y_{v_i} > s) \\
&= \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_{v_i} = 1, i = 1, 2, \ldots, s+1) \\
&= \sum_{i=1}^{s+1} \lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0(Y_{v_i} = 1) \\
&= -\frac{1}{r} \sum_{i=1}^{s+1} \psi_i, \qquad (4.8)
\end{aligned}$$

where the second equality follows because $\{Y_{v_i} = 1, i = 1, 2, \ldots, s+1\}$ is the dominating error event, and the third equality follows from independence. Similarly, we obtain

$$\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_1(Y_f = 0) = -\frac{1}{r} \sum_{i=s+1}^{r} \varphi_i. \qquad (4.9)$$

Then, the objective function (4.5) is equal to

$$\begin{aligned}
&\frac{1}{r} \max\{-\lambda_1 \sum_{j=1}^{s+1} \psi_j, -\lambda_2 \sum_{j=s+1}^{r} \varphi_j\} \\
&\geq \frac{1}{r} \max\{-\lambda_1 (s+1)\psi_{s+1}, -\lambda_2 (r-s)\varphi_{s+1}\},
\end{aligned}$$

where equality holds if we set $\psi_i = \psi_j$ and $\varphi_i = \varphi_j$ for all i, j. Hence, it is optimal to use the same strategy for each of the sub-trees rooted at the nodes $v_1, \ldots, v_r$. □

Theorem 4.2. *Consider an r-symmetric tree sequence $(T_n)_{n\geq 1}$, and suppose that Assumptions 2.1, 2.2 and 4.1 hold. Amongst all counting strategies, there is no loss in optimality if we impose the following restrictions:*

(i) all leaves use the same transmission function;

(ii) for each $k \geq 1$, all level k nodes use counting quantizers with the same threshold.

Furthermore, the optimal error exponent at the fusion center is given by[2]

$$\begin{aligned}\mathcal{E}_S^*(r) &= \lim_{n\to\infty} \frac{1}{n} \log P_e^*(T_n) \\ &= -\sup \Big\{ \Big[\Big(\prod_{k=2}^{h} \frac{s_k+1}{r} \Big) \Lambda_{0,1}^*(\gamma, t) \Big] \wedge \Big[\Big(\prod_{k=2}^{h} \frac{r-s_k}{r} \Big) \Lambda_{1,1}^*(\gamma, t) \Big] : \\ & \quad s_k = 0, \ldots, r-1, \ \textit{for } k = 2, \ldots, h; \ \gamma \in \Gamma; \ -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < t < \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma) \Big\}. \end{aligned} \tag{4.10}$$

Proof. (Outline[3]) From Lemma 4.2, an optimal counting strategy is to use the same transmission strategy at every sub-tree rooted at each $v \in C_n(f)$. Suppose that the fusion center uses, as its fusion rule, a counting quantizer with threshold s_h. Then, the objective at each $v \in C_n(f)$ is to minimize

$$\frac{1}{r} \max \Big\{ (s_h+1) \lim_{n\to\infty} \frac{1}{l_n(v)} \log \mathbb{P}_0(Y_v = 1), (r-s_h) \lim_{n\to\infty} \frac{1}{l_n(v)} \log \mathbb{P}_1(Y_v = 0) \Big\}.$$

We apply Lemma 4.2 on v, and repeat the same argument for $h-2$ steps. Therefore, we conclude that for each $k \geq 2$, there is no loss in optimality if all nodes at the same level k, use counting quantizers with the same threshold s_k. Moreover, there is no loss in optimality if each level 1 node has the same Type I and II error exponents. Applying Lemma 4.1 to each level 1 node yields that it is asymptotically optimal for all leaves to use the same transmission function γ, and all level 1 nodes to use LLRQs (which are equivalent to counting quantizers since leaves use the same transmission function) with the same threshold t . Finally, the form of the optimal error exponent is obtained by optimizing over the thresholds s_k, where $k = 2, \ldots, h$, the threshold t, and the transmission function γ. The theorem is now proved. □

Suppose that the transmission function γ in (4.10) has been fixed, and suppose

[2]The products are taken to be 1 when $h = 1$. We also use the notation $x \wedge y = \min\{x, y\}$.

[3]For any given counting strategy, a more rigorous proof will involve taking a subsequence of $(T_n)_{n\geq 1}$, so that the same performance is achieved by this subsequence using a counting strategy in which the thresholds do not vary with n.

that $h > 1$ and $r > 1$. Then, we have

$$\frac{1}{r^{h-1}} \prod_{i=2}^{h} (s_i + 1) \leq 1,$$
$$\frac{1}{r^{h-1}} \prod_{i=2}^{h} (r - s_i) \leq 1,$$

and equality cannot hold simultaneously in both expressions above. Since for each $\gamma \in \Gamma$, $\Lambda^*_{0,1}(\gamma, t)$ and $\Lambda^*_{1,1}(\gamma, t)$ are continuous in t, the error exponent in (4.10) is achieved by setting

$$\Big(\prod_{i=2}^{h} \frac{s_i + 1}{r}\Big)\Lambda^*_{0,1}(\gamma, t) = \Big(\prod_{i=2}^{h} \frac{r - s_i}{r}\Big)\Lambda^*_{1,1}(\gamma, t).$$

Hence, the error exponent is *strictly* smaller than that for the parallel configuration. This shows that using a r-symmetric tree results in a loss of efficiency as compared to the parallel configuration, if we restrict to counting strategies. In the following, we discuss some special cases.

4.4.1 On the Worst Error Exponent

When $r = 1$, the network is essentially the same, and therefore achieves the same performance, as a parallel configuration, which is the best possible. Our next result provides evidence that performance degrades as r increases. Let $(T_n(r))_{n\geq 1}$ be a r-symmetric tree sequence, where $r = 1, 2, \ldots$.

Lemma 4.3. *Suppose that Assumptions 2.1, 2.2 and 4.1 hold, and the network is restricted to counting strategies. Then, for any $r \geq 1$ and any positive integer $m > 1$, $\mathcal{E}^*_S(r) < \mathcal{E}^*_S(mr)$.*

Proof. Consider any sequence of integers k_i, where $i = 2, \ldots, h$, such that $0 \leq k_i < mr$ for all i. For each i, we can find an integer $s_i \in [0, r)$, such that $ms_i \leq k_i < m(s_i + 1)$.

We have

$$\frac{k_i+1}{mr} \le \frac{m(s_i+1)}{mr} = \frac{s_i+1}{r}, \tag{4.11}$$

$$1-\frac{k_i}{mr} \le 1-\frac{ms_i}{mr} = 1-\frac{s_i}{r}. \tag{4.12}$$

Then, it is an easy exercise to show that

$$\begin{aligned}
&\Big[\Big(\prod_{i=2}^{h}\frac{k_i+1}{mr}\Big)\Lambda^*_{0,1}(\gamma,t)\Big] \wedge \Big[\Big(\prod_{i=2}^{h}(1-\frac{k_i}{mr})\Big)\Lambda^*_{1,1}(\gamma,t)\Big] \\
&\le \Big(1-\frac{1}{mr}\Big)\Big[\Big(\prod_{i=2}^{h}\frac{s_i+1}{r}\Big)\Lambda^*_{0,1}(\gamma,t)\Big] \wedge \Big[\Big(\prod_{i=2}^{h}(1-\frac{s_i}{r})\Big)\Lambda^*_{1,1}(\gamma,t)\Big] \\
&\le -\big(1-\frac{1}{mr}\big)\mathcal{E}^*_S(r).
\end{aligned}$$

Taking the supremum over k_i, γ and t, yields $\mathcal{E}^*_S(mr) > \mathcal{E}^*_S(r)$. The proof is now complete. □

The above lemma shows that for any $m > 1$ and $r \ge 1$, $(\mathcal{E}^*_S(m^l r))_{l\ge 0}$ is an increasing sequence, which is bounded above by zero, hence it converges. We provide an upper bound for this limit (cf. Proposition 4.6) below.

Proposition 4.1. *Suppose that Assumptions 2.1, 2.2 and 4.1 hold. For any collection of symmetric tree sequences, $\{(T_n(r))_{n\ge1} : r = 1, 2, \ldots\}$, where $(T_n(r))_{n\ge1}$ is a r-symmetric tree sequence, we have*

$$\limsup_{r\to\infty} \mathcal{E}^*_S(r) \le -\sup_{\substack{\gamma\in\Gamma\\ t\in\mathbb{R}}} \Big(\frac{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}}\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}} + \Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}\Big)^{h-1}.$$

Proof. Given $\gamma \in \Gamma$, and $-\mathrm{D}(\mathbb{P}^\gamma_0 \,\|\, \mathbb{P}^\gamma_1) < t < \mathrm{D}(\mathbb{P}^\gamma_1 \,\|\, \mathbb{P}^\gamma_0)$, let

$$\delta = \frac{\Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}} + \Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}, \tag{4.13}$$

and $s = \lfloor \delta r \rfloor$. We have

$$\mathcal{E}_S^*(r) \leq -\Big[\Big(\frac{s+1}{r}\Big)^{h-1}\Lambda_{0,1}^*(\gamma,t)\Big] \wedge \Big[\Big(\frac{r-s}{r}\Big)^{h-1}\Lambda_{1,1}^*(\gamma,t)\Big].$$

Since $s/r \to \delta$ as $r \to \infty$, we obtain

$$\begin{aligned}\limsup_{r\to\infty} \mathcal{E}_S^*(r) &\leq -[\delta^{h-1}\Lambda_{0,1}^*(\gamma,t)] \wedge [(1-\delta)^{h-1}\Lambda_{1,1}^*(\gamma,t)] \\ &= -\Big(\frac{\Lambda_{0,1}^*(\gamma,t)^{\frac{1}{h-1}}\Lambda_{1,1}^*(\gamma,t)^{\frac{1}{h-1}}}{\Lambda_{0,1}^*(\gamma,t)^{\frac{1}{h-1}} + \Lambda_{1,1}^*(\gamma,t)^{\frac{1}{h-1}}}\Big)^{h-1},\end{aligned}$$

and taking the infimum over $\gamma \in \Gamma$ and $t \in \mathbb{R}$, the proposition is proved. □

Under some additional symmetry assumptions, the inequality in the above proposition is an equality. This is shown in Proposition 4.6 in Section 4.5.

4.4.2 Optimality of the Majority Decision Rule

Suppose that all the leaves use the transmission function $\gamma \in \Gamma$. Finding an optimal strategy using (4.10) requires us to search over a space with r^{h-1} elements, and also optimizing over t. The search can be daunting even for moderate values of r and h. For this reason, we now consider the case where r is odd, and the majority decision rule is used at every non-leaf node, i.e., a node transmits a 1 iff the majority of its immediate predecessors send a 1. For level 1 nodes, the majority decision rule corresponds to a LLRQ with threshold 0, while for nodes of level greater than 1, it corresponds to a counting quantizer with threshold $(r-1)/2$. In the proposition below, we develop a sufficient condition under which this strategy is optimal.

Proposition 4.2. *Consider a r-symmetric tree network with $h > 1$, and r an odd integer. Suppose that that all the leaves use the same transmission function γ. Let t_0 and t_1 be such that $\Lambda_{0,1}^*(\gamma,t_0) = r^{h-1}\Lambda_{1,1}^*(\gamma,t_0)$ and $\Lambda_{1,1}^*(\gamma,t_1) = r^{h-1}\Lambda_{0,1}^*(\gamma,t_1)$. Under Assumptions 2.1, 2.2 and 4.1, and the restriction to counting strategies, if*

$$\max\big\{\Lambda_{0,1}^*(\gamma,t_0), \Lambda_{1,1}^*(\gamma,t_1)\big\} \leq \frac{2r^{h-1}(r+1)}{r^{h-1}(r-1)+r+3}\Lambda_{0,1}^*(\gamma,0), \tag{4.14}$$

then using the majority decision rule at all relay nodes achieves the optimal error exponent, which is given by

$$\mathcal{E}_S^*(r) = -\Big(\frac{r+1}{2r}\Big)^{h-1}\Lambda_{0,1}^*(\gamma, 0).$$

Proof. If $r = 1$, the network is equivalent to the parallel configuration. It is easy to see that the sufficient condition (4.14) holds trivially, therefore our claim is consistent with the well known optimal fusion rule for the parallel configuration. Henceforth, we assume that $r > 1$.

For simplicity, let $\psi(t) = \Lambda_{0,1}^*(\gamma, t)$ and $\varphi(t) = \Lambda_{1,1}^*(\gamma, t)$. The sufficient condition (4.14) is obtained by approximating the convex functions ψ and φ with appropriate straight line segments as shown in Figure 4-2.

Suppose that each level k node uses a counting quantizer with threshold s_k, such that at least one of the thresholds $s_k \neq (r-1)/2$. Consider the case where

$$b = \prod_{k=2}^{h}(s_k + 1) < a = \big(\frac{r+1}{2}\big)^{h-1} < \prod_{k=2}^{h}(r - s_k) = c.$$

We consider the solution to the equations

$$y = \frac{b(\psi(t_0) - \psi(0))}{t_0}t + b\psi(0),$$
$$y = -\frac{c(\varphi(0) - \varphi(t_0))}{t_0}t + c\varphi(0),$$

which gives the intersection of the straight line approximations shown in Figure 4-2. Solving the linear equations, and observing that $\psi(0) = \varphi(0)$, we obtain

$$y = \frac{bc(1+d)}{c+bd}\psi(0),$$

where $d = \frac{\psi(t_0)-\psi(0)}{\varphi(0)-\varphi(t_0)}$. Since ψ and φ are convex functions,

$$\sup_t \min\{b\psi(t), c\varphi(t)\} \leq y.$$

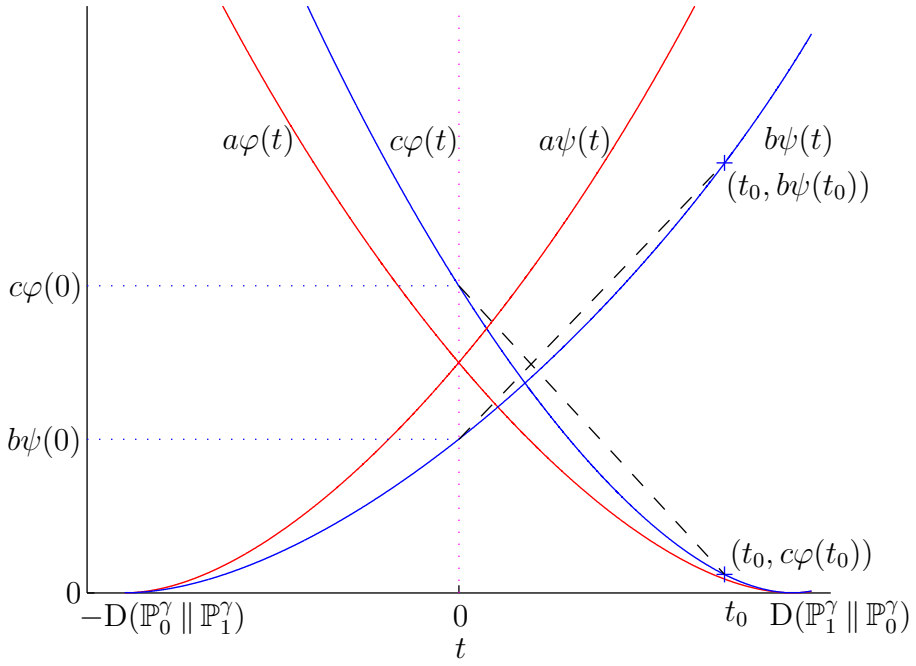

Figure 4-2: A typical plot of the rate functions.

To prove the proposition, we check that under the condition (4.14), $y \le a\psi(0)$, for all pairs (b, c) such that $b < a < c$. This is equivalent to checking that

$$d \le \frac{c(a-b)}{b(c-a)} = \frac{a}{b}\Big(1 - \frac{b-a}{c-a}\Big) - 1, \tag{4.15}$$

for all (b, c) such that $b < a < c$. The R.H.S. of (4.15) increases when b decreases (and c increases), hence the minimum value is achieved by setting $b = (r+1)^{h-2}(r-1)/2^{h-1}$, and $c = (r+1)^{h-2}(r+3)/2^{h-1}$. This yields the sufficient condition

$$\Lambda^*_{0,1}(\gamma, t_0) \le \frac{2r^{h-1}(r+1)}{r^{h-1}(r-1)+r+3}\Lambda^*_{0,1}(\gamma, 0),$$

in the case where $b < a < c$. A similar argument for the case where $c < a < b$ yields the sufficient condition

$$\Lambda^*_{1,1}(\gamma, t_1) \le \frac{2r^{h-1}(r+1)}{r^{h-1}(r-1)+r+3}\Lambda^*_{0,1}(\gamma, 0).$$

Finally, the optimal error exponent is obtained by substituting $s_k = (r-1)/2$ in Theorem 4.2, and the proposition is proved. □

To show that our sufficient condition in Proposition 4.2 is not vacuous, we provide an example in which the optimal counting strategy is not the majority decision rule.

Example 4.1. *Consider a r-symmetric network, with* $r = 45$ *and* $h = 3$. *Suppose that each leaf sends the message 1 with probability* $p_0 = 0.3$ *under hypothesis* H_0, *and with probability* $p_1 = 0.9$ *under hypothesis* H_1. *If all non-leaf nodes use the majority decision rule (the counting quantizer thresholds are* $s_2 = s_3 = 22$*), we get an error exponent of* $-129.2460/45^2$. *If counting quantizers with thresholds* $s_2 = s_3 = 23$ *are used, our error exponent is* $-129.5009/45^2$, *which is more negative than that for the majority decision rule. In fact, it can be checked numerically that these are the optimal counting quantizers that should be used. The error exponents in the cases where* s_2 *and* s_3 *are constrained to be the same, are plotted in Figure 4-3.* □

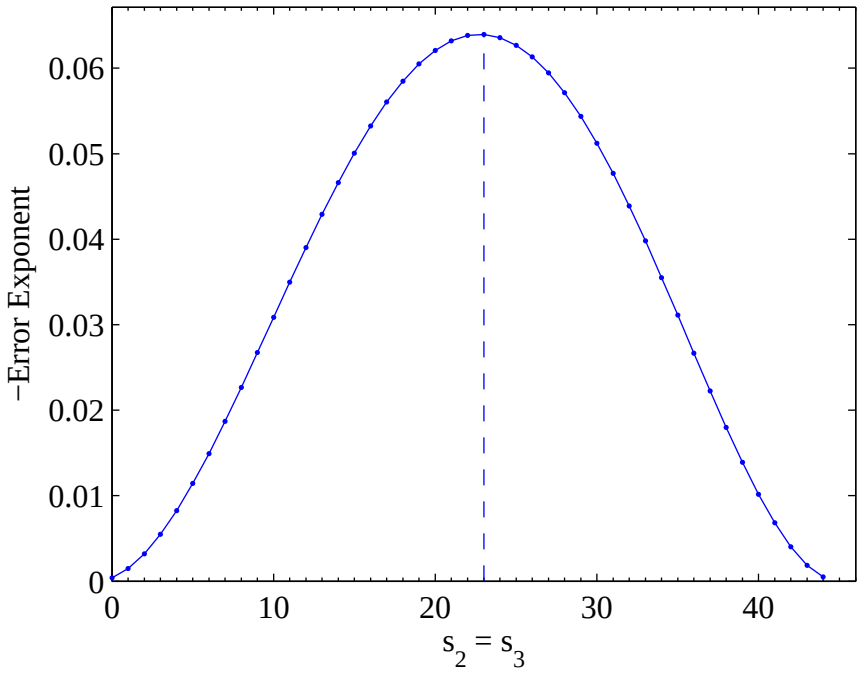

Figure 4-3: Plot of the error exponent for the cases when $s_2 = s_3$.

To illustrate the use of Proposition 4.2, we consider an example below.

Example 4.2. *We have assumed that all nodes, including the leaves, transmit 1-bit messages. However, our results still hold even if the 1-bit message restriction is removed for the leaves. In this example, we will drop the 1-bit message restriction for the leaves to simplify calculations. Consider a r-symmetric network with $r = 3$ and $h = 2$. Assume that leaves transmit their observations in the raw to their immediate successors, where each observation is i.i.d. with distribution*

$$H_0 : N(-1, \sigma^2), \quad H_1 : N(1, \sigma^2).$$

We obtain

$$\Lambda^*_{0,1}(\gamma, t) = \frac{\sigma^2}{8}\Big(t + \frac{2}{\sigma^2}\Big)^2,$$
$$\Lambda^*_{1,1}(\gamma, t) = \frac{\sigma^2}{8}\Big(t - \frac{2}{\sigma^2}\Big)^2.$$

Solving the equality $\Lambda^*_{0,1}(\gamma, t_0) = 3\Lambda^*_{1,1}(\gamma, t_0)$, yields $t_0 = (2/\sigma^2)(\sqrt{3}-1)/(\sqrt{3}+1)$. (Because of symmetry, we only need to verify (4.14) for $\Lambda^*_{0,1}(\gamma, t_0)$.) Then, we have

$$\begin{aligned}\Lambda^*_{0,1}(\gamma, t_0) &= \frac{\sigma^2}{8}\Big(\frac{\sqrt{3}-1}{\sqrt{3}+1} + 1\Big)^2 \frac{4}{\sigma^4} \\ &= \Big(\frac{\sqrt{3}}{\sqrt{3}+1}\Big)^2 \frac{2}{\sigma^2} \\ &< \frac{1}{\sigma^2} = \frac{2r^{h-1}(r+1)}{r^{h-1}(r-1)+r+3}\Lambda^*_{0,1}(\gamma, 0).\end{aligned}$$

Therefore, it is optimal (amongst all counting strategies) to use the majority rule as the fusion rule. This is verified numerically in Figure 4-4 below. □

The sufficient condition in (4.14) can be difficult to check if one does not have access to the functions $\Lambda^*_{j,1}(\gamma, t)$, $j = 0, 1$. A simpler but cruder sufficient condition is presented below; the proof is the same as in Proposition 4.2, except that we let $\mathrm{D}(\mathbb{P}^\gamma_1 \,\|\, \mathbb{P}^\gamma_0)$ play the role of t_0, and $-\mathrm{D}(\mathbb{P}^\gamma_0 \,\|\, \mathbb{P}^\gamma_1)$ the role of t_1.

Corollary 4.1. *Suppose that r is an odd integer greater than 1, and all the leaves use the same transmission function γ. Under Assumptions 2.2-4.1, and the restriction to*

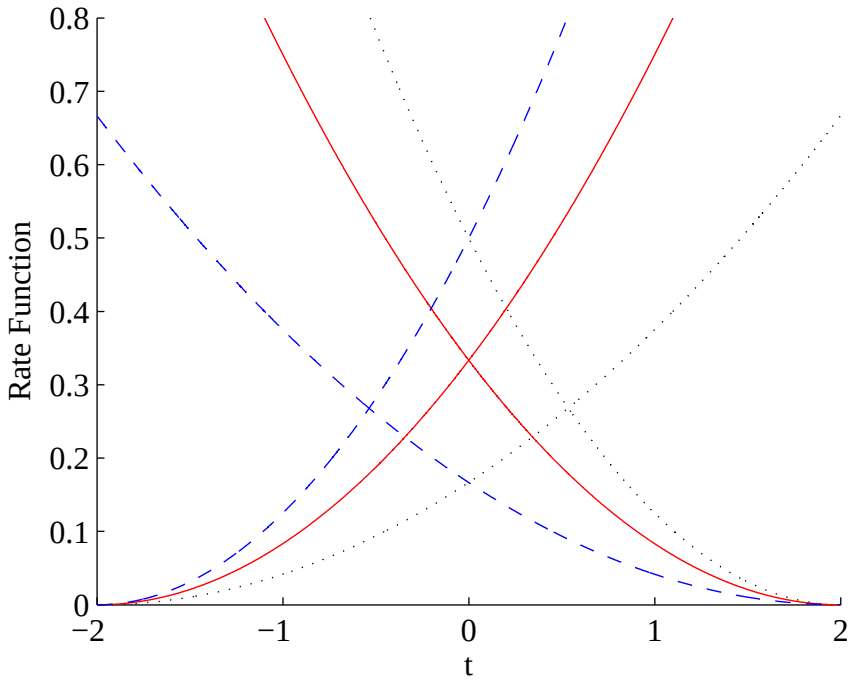

Figure 4-4: Curves of different types correspond to different strategies. Intersections of curves of the same type give the negative of the respective error exponents. Solid curves are for the majority decision strategy ($s_2 = 1$), dotted curves are for the strategy with $s_2 = 0$, and dashed curves are for the strategy with $s_2 = 2$.

counting strategies, if

$$\max\{\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)\} \leq -\frac{2(r+1)}{r-1} \inf_{\lambda \in [0,1]} \Lambda_{0,0}(\gamma; \lambda),$$

then using the majority decision rule at all non-leaf nodes achieves the optimal error exponent.

4.4.3 Binary Symmetric Networks

In the previous subsection, we considered the case where r is odd, and gave a sufficient condition under which the majority decision rule strategy is optimal. When r is even, the majority decision rule at a level k node, where $k > 1$, can be a counting quantizer with threshold either $r/2 - 1$ or $r/2$. This can lead to a large number of possibilities for the majority decision rule strategy when h is large. We now consider the special

case where $r = 2$ and h is odd. For each level k node v, where $k > 1$, there are two immediate predecessors. The two counting quantizers that v can use are either the the counting quantizer with threshold 0 (OR rule), or the counting quantizer with threshold 1 (AND rule). The proof of the following proposition is similar to that of Proposition 4.2, and is omitted.

Proposition 4.3. *Suppose that $r = 2$ and h is odd. Suppose all the leaves use the same transmission function γ. Under Assumptions 2.1, 2.2 and 4.1, and the restriction to counting strategies, if*

$$\max\{-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma), \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)\} \le 3\Lambda_{0,1}^*(\gamma, 0), \tag{4.16}$$

then it is asymptotically optimal to use LLRQs with threshold 0 for all level 1 nodes, and to use the OR rule in half of the remaining $h-1$ levels, and the AND rule in the other half of the remaining $h-1$ levels.

4.4.4 A Generalization

We generalize the concept of a r-symmetric tree network as follows.

Definition 4.2. *For a vector $\vec{r} = (r_2, r_3, \ldots, r_h)$, where r_k are positive integers, a h-uniform sequence of trees $(T_n)_{n\ge1}$ is said to be $\vec{r}$-symmetric if*

(i) for all level k vertices v, where $k > 1$, $|C_n(v)| = r_k$, and

(ii) for all level 1 nodes v, $l_n(v)/l_n(f) \to \big(\prod_{k=2}^h r_k\big)^{-1}$ as $n \to \infty$.

It is not difficult to generalize Theorem 4.2 to the following; the proof is omitted.

Proposition 4.4. *Let $\vec{r} = (r_2, r_3, \ldots, r_h)$, and suppose Assumptions 2.1, 2.2 and 4.1 hold. Given a $\vec{r}$-symmetric tree network $(T_n)_{n\ge1}$ of height $h \ge 1$, the optimal error exponent, over all counting strategies, is*

$$\mathcal{E}_S^*(\vec{r}) = \sup\Big\{\Big[\Big(\prod_{k=2}^h \frac{s_k+1}{r_k}\Big)\Lambda_{0,1}^*(\gamma,t)\Big] \wedge \Big[\Big(\prod_{k=2}^h \frac{r_k-s_k}{r_k}\Big)\Lambda_{1,1}^*(\gamma,t)\Big] : \\ s_k = 0,\ldots,r_k-1, \text{ and } \gamma\in\Gamma, -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < t < \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)\Big\}. \tag{4.17}$$

Furthermore, there is no loss in optimality if we use

1. *the same transmission function for all the leaves;*

2. *counting quantizers with the same threshold for nodes of the same level k, where $k \geq 1$.*

4.5 Rapidly Branching Tree Sequences

In the previous section, we considered a symmetric tree sequence in which the number of non-leaf nodes is bounded. In this section, we consider tree sequences in which the number of non-leaf nodes becomes large, in a certain sense, as n increases. Motivated by the example in Section 4.1, we define the following.

Definition 4.3. *A rapidly branching tree sequence is a sequence of h-uniform trees $(T_n)_{n\geq 1}$, such that*

(i) the number of immediate predecessors of each non-leaf node grows to infinity as n increases;

(ii) there exists a sequence of positive reals $(\kappa_n)_{n\geq 1}$ such that κ_n decreases to 0 as n increases, and for each level k node v, where $k \geq 2$,

$$\frac{\max_{u\in C_n(v)} l_n^2(u)}{\min_{u\in C_n(v)} l_n^2(u)} \leq \kappa_n |C_n(v)|.$$

A rapidly branching tree sequence is a sequence of trees in which the number of immediate predecessors of each node grows faster than the rate at which the network becomes "unbalanced" (in the sense that different level 1 nodes have very different numbers of leaves attached to them). The definition of a rapidly branching tree sequence implies that the number of immediate predecessors of every level 1 node grows uniformly fast.

In Design II of Section 4.1, when n is large, with high probability, we have $l_n(u) \simeq l_n(v)$ for all level 1 nodes u and v. Therefore, this tree network fits our definition of a

rapidly branching network with height $h = 2$. For a general h, a similar design can be used to approximate a h-hop MST [42]. In all of these designs, with high probability, we get a rapidly branching tree network.

Since using LLRQs for every node is known to be optimal (see e.g. [6]), we assume that every node (including leaves) has access to LLRQs. The number of nodes at each level k in a rapidly branching tree network grows with n. Similar to Section 4.4, the problem of finding optimal LLRQs for each node in a rapidly branching tree network is, in general, intractable. Therefore, we make the following simplifying assumption.

Assumption 4.2. *Every node has access to LLRQs, and every node of the same level k uses a LLRQ with the same threshold t_k.*

For notational simplicity, if each leaf uses a transmission function γ, which is a LLRQ, we identify γ with the threshold of the LLRQ, i.e., $\gamma = t_0 \in \mathbb{R}$. We will first make a slight detour and show a limit theorem for a rapidly branching tree network. The proof is provided in Section 4.6.

Proposition 4.5. *Suppose that Assumptions 2.1, 2.2 and 4.1 hold. Given a rapidly branching tree network $(T_n)_{n\geq 1}$, suppose each leaf sends its observation to its immediate successor via a transmission function $\gamma \in \Gamma$, and each level k node, where $k \geq 1$, uses a LLRQ with constant threshold t_k. Suppose that $\{\gamma, t_1, \ldots, t_h\}$ satisfy (3.9)-(3.11). Then, we have*

$$\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_1\big(Y_f = 0\big) = -\Lambda^*_{1,h}(\gamma, t^{(h)}),$$
$$\lim_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0\big(Y_f = 1\big) = -\Lambda^*_{0,h}(\gamma, t^{(h)}).$$

We now consider the Bayesian detection problem in a rapidly branching tree sequence, in which all nodes are constrained to sending only one bit reliably.

Theorem 4.3. *Consider a rapidly branching tree sequence $(T_n)_{n\geq 1}$. Suppose that*

Assumptions 2.1, 2.2, 4.1, and 4.2 hold. Then, the optimal error exponent is

$$\mathcal{E}^*_{RB} = -\sup_{\substack{\gamma\in\Gamma\\ t_1\in\mathbb{R}}} \Big(\frac{\Lambda^*_{0,1}(\gamma,t_1)^{\frac{1}{h-1}}\Lambda^*_{1,1}(\gamma,t_1)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t_1)^{\frac{1}{h-1}}+\Lambda^*_{1,1}(\gamma,t_1)^{\frac{1}{h-1}}}\Big)^{h-1}. \tag{4.18}$$

Furthermore, if the supremum is achieved by $\gamma\in\Gamma$*, and* $t_1\in(-\mathrm{D}(\mathbb{P}_0^\gamma\,\|\,\mathbb{P}_1^\gamma),\mathrm{D}(\mathbb{P}_1^\gamma\,\|\,\mathbb{P}_0^\gamma))$*, then the optimal threshold* t_k *for level* k *nodes, where* $k=2,\ldots,h$*, is*

$$t_k = \frac{\Lambda^*_{0,k-1}(\gamma,t^{(k-1)})\Lambda^*_{1,k-1}(\gamma,t^{(k-1)})^{\frac{1}{h-k+1}} - \Lambda^*_{0,k-1}(\gamma,t^{(k-1)})^{\frac{1}{h-k+1}}\Lambda^*_{1,k-1}(\gamma,t^{(k-1)})}{\Lambda^*_{0,k-1}(\gamma,t^{(k-1)})^{\frac{1}{h-k+1}}+\Lambda^*_{1,k-1}(\gamma,t^{(k-1)})^{\frac{1}{h-k+1}}}.$$

We first state a lemma whose proof is easily obtained using calculus, and is thus omitted.

Lemma 4.4. *Given* $k\geq 1$ *and* $a,b>0$*, we have*

$$\begin{aligned}&\min_{-b<x<a}\Big(\Big(\frac{a+b}{a(b+x)}\Big)^{\frac{1}{k}}+\Big(\frac{a+b}{b(a-x)}\Big)^{\frac{1}{k}}\Big)^k\\ &=\Big(\Big(\frac{1}{a}\Big)^{\frac{1}{k+1}}+\Big(\frac{1}{b}\Big)^{\frac{1}{k+1}}\Big)^{k+1},\end{aligned}$$

and the minimizer is given by

$$x^* = \frac{ab^{\frac{1}{k+1}}-a^{\frac{1}{k+1}}b}{a^{\frac{1}{k+1}}+b^{\frac{1}{k+1}}}.$$

Proof of Theorem 4.3: Suppose that under Assumptions 2.1, 2.2, 4.1, and 4.2, it is optimal for each leaf to use a LLRQ with threshold γ_n, and for each level k node, where $k\geq 1$, to use a LLRQ with threshold $t_{n,k}$. Let $(n_l)_{l\geq 1}$ be a subsequence such that

$$\lim_{l\to\infty}\frac{1}{l_{n_l}(f)}\log P_e(T_{n_l}) = \mathcal{E}^*_{RB}.$$

Since γ_n is bounded ($|\gamma_n|$ cannot diverge to infinity, otherwise every leaf reports either 1 or 0 with probability one (asymptotically), under both hypotheses), there

exists a subsequence $(u_l)_{l\geq 1}$ of $(n_l)_{l\geq 1}$ such that $\gamma_{u_l} \to \gamma \in \mathbb{R}$ as $l \to \infty$. Then, from Assumption 4.1, since $\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma)$ and $\mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)$ are bounded, the thresholds $t_{u_l,k}$ must satisfy $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) \leq t_{u_l,k} \leq \mathrm{D}(\mathbb{P}_1^\gamma \,\|\, \mathbb{P}_0^\gamma)$, for l sufficiently large, otherwise either the Type I or Type II error exponents at the fusion center is zero.

Therefore, there exists a further subsequence $(m_l)_{l\geq 1}$ of $(u_l)_{l\geq 1}$ such that for all k, $\lim_{l\to\infty} t_{m_l,k} = t_k$, for some bounded t_k. Then, for all $\epsilon > 0$, from Proposition 4.5, we obtain

$$\mathcal{E}_{RB}^* \geq -\min\{\Lambda_{0,h}^*(\gamma+\epsilon, t_1+\epsilon, \ldots, t_k+\epsilon), \Lambda_{1,h}^*(\gamma-\epsilon, t_1-\epsilon, \ldots, t_k-\epsilon)\}.$$

Taking $\epsilon \to 0$, and noting that $\Lambda_{0,h}^*$ and $\Lambda_{1,h}^*$ are continuous in all their arguments, we get

$$\mathcal{E}_{RB}^* \geq -\min\{\Lambda_{0,h}^*(\gamma, t^{(h)}), \Lambda_{1,h}^*(\gamma, t^{(h)})\}.$$

This shows that there is no loss in optimality if we restrict transmission functions to be the same for all n. Therefore, it remains to optimize over $\gamma \in \Gamma$ and over $t^{(h)}$. In this case, it is well known (using the same argument as in Corollary 3.4.6 of [39]) that the optimal fusion rule at the fusion center consists of a LLRQ with threshold $t_h = 0$. This yields

$$\begin{aligned}
\mathcal{E}_{RB}^* &= \inf_{\substack{\lambda\in[0,1]\\ \gamma, t^{(h-1)}}} \Lambda_{0,h-1}(\gamma, t^{(h-1)}; \lambda) \\
&= -\sup_{\gamma, t^{(h-1)}} \frac{\Lambda_{0,h-1}^*(\gamma, t^{(h-1)})\Lambda_{1,h-1}^*(\gamma, t^{(h-1)})}{\Lambda_{0,h-1}^*(\gamma, t^{(h-1)}) + \Lambda_{1,h-1}^*(\gamma, t^{(h-1)})} \qquad (4.19) \\
&= -\Big[\inf_{\gamma, t^{(h-2)}} \inf_{t_{h-1}} \Big\{\frac{1}{\Lambda_{0,h-1}^*(\gamma, t^{(h-1)})} + \frac{1}{\Lambda_{1,h-1}^*(\gamma, t^{(h-1)})}\Big\}\Big]^{-1} \\
&= -\Big[\inf_{\gamma, t^{(h-2)}} \inf_{t_{h-1}} \Big\{\frac{\Lambda_{0,h-2}^*(\gamma, t^{(h-2)}) + \Lambda_{1,h-2}^*(\gamma, t^{(h-2)})}{\Lambda_{0,h-2}^*(\gamma, t^{(h-2)})(\Lambda_{1,h-2}^*(\gamma, t^{(h-2)}) + t_{h-1})} \\
&\quad + \frac{\Lambda_{0,h-2}^*(\gamma, t^{(h-2)}) + \Lambda_{1,h-2}^*(\gamma, t^{(h-2)})}{\Lambda_{1,h-2}^*(\gamma, t^{(h-2)})(\Lambda_{0,h-2}^*(\gamma, t^{(h-2)}) - t_{h-1})}\Big\}\Big]^{-1}, \qquad (4.20)
\end{aligned}$$

where (4.19) and (4.20) follow from Lemma 3.1. We take $a = \Lambda_{0,h-2}^*(\gamma, t^{(h-2)})$ and

$b = \Lambda^*_{1,h-2}(\gamma, t^{(h-2)})$ in Lemma 4.4 to obtain

$$\mathcal{E}^*_{RB} = -\Big[\inf_{\gamma, t^{(h-2)}} \Big\{ \Big(\frac{1}{\Lambda^*_{0,h-2}(\gamma, t^{(h-2)})} \Big)^{1/2} + \Big(\frac{1}{\Lambda^*_{1,h-2}(\gamma, t^{(h-2)})} \Big)^{1/2} \Big\}^2 \Big]^{-1}.$$

The optimal error exponent and the optimal thresholds for the LLRQs then follow by repeating the above same argument for another $h-2$ steps. The proof is now complete. □

By taking $t_1 = 0$ in (4.18), we get the same lower bound as in (4.4). Hence one does no worse than by a factor of $1/2^{h-1}$ from the optimal error exponent of a parallel configuration.

For completeness, our next result shows that the bound in Proposition 4.1 is an equality if leaves can use LLRQs as transmission functions.

Proposition 4.6. *Suppose that the set Γ of allowable transmission functions for the leaves includes LLRQs. Then, under Assumptions 2.1, 2.2, and 4.1, we have*

$$\lim_{r\to\infty} \mathcal{E}^*_S(r) = -\sup_{\substack{\gamma\in\Gamma \\ t\in\mathbb{R}}} \Big(\frac{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}} \Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}}{\Lambda^*_{0,1}(\gamma,t)^{\frac{1}{h-1}} + \Lambda^*_{1,1}(\gamma,t)^{\frac{1}{h-1}}} \Big)^{h-1} = \mathcal{E}^*_{RB}.$$

Proof. Consider a collection of tree sequences $\{(T(n,r))_{n\geq 1} : r \geq 1\}$ such that (a) each $(T(n,r))_{n\geq 1}$ is a r-symmetric tree sequence; and (b) for each r and for each n, every level 1 node in $T(n,r)$ has the same number of leaves attached to it. Then, from Theorem 4.2, the optimal error exponent for each tree sequence $(T(n,r))_{n\geq 1}$ is $\mathcal{E}^*_S(r)$.

Suppose that there exists a subsequence $(r_m)_{m\geq 1}$ such that $g = \lim_{m\to\infty} \mathcal{E}^*_S(r_m) < \mathcal{E}^*_{RB}$. Suppose that each tree sequence $(T(n,r_m))_{n\geq 1}$ uses the asymptotically optimal counting strategy proposed in Theorem 4.2. Note that this strategy also satisfies Assumption 4.2. We shall construct a rapidly branching tree sequence from $\{(T(n,r_m))_{n\geq 1} : \ m \geq 1\}$. Fix a positive $\epsilon < \mathcal{E}^*_{RB} - g$, and let $(n_m)_{m\geq 1}$ be an increasing sequence of positive integers such that

$$\frac{1}{l_{n_m}(f)} \log P_e(T(n_m, r_m)) \leq \mathcal{E}^*_S(r_m) + \epsilon.$$

Let $\widetilde{T}_m = T(n_m, r_m)$. Then, it is an easy exercise to verify that $(\widetilde{T}_m)_{m\geq 1}$ satisfies Definition 4.3 with $\kappa_m = 1/r_m$ (which goes to 0, as $m \to \infty$). We then have

$$\frac{1}{l_{n_m}(f)} \log P_e(\widetilde{T}_m) = \frac{1}{l_{n_m}(f)} \log P_e(T(n_m, r_m)) \leq \mathcal{E}_S^*(r_m) + \epsilon.$$

Taking $m \to \infty$, we obtain $\limsup_{m\to\infty}(1/l_{n_m}(f)) \log P_e(\widetilde{T}_m) \leq g + \epsilon < \mathcal{E}_{RB}^*$, a contradiction to Theorem 4.3. Therefore, we must have $\liminf_{r\to\infty} \mathcal{E}_S^*(r) \geq \mathcal{E}_{RB}^*$. Finally, from Proposition 4.1, we obtain the desired conclusion. □

We define a strongly symmetric tree sequence $(T(n))_{n\geq 1}$ to be a tree sequence in which each tree $T(n)$ has a symmetric relay network (i.e., each node of level $k \geq 2$ has the same number of immediate predecessors), and each level 1 node has the same number of leaves. If a r-symmetric tree sequence is strongly symmetric, then the asymptotically optimal counting strategy proposed in Theorem 4.2 satisfies Assumption 4.2. In this case, Proposition 4.1 and Lemma 4.3 shows that $\mathcal{E}_S^*(r) < \mathcal{E}_{RB}^*$ for all $r < \infty$, i.e., a rapidly branching tree sequence satisfying Assumptions 2.1, 2.2, 4.1, and 4.2, has worse performance than a strongly symmetric r-symmetric tree sequence, for every r.

4.6 Proofs

In this section, we prove some of the results in this chapter.

Proof of Proposition 4.5.

Note that we cannot apply the Gärtner-Ellis Theorem directly here, since the asymptotic log moment generating function does not satisfy the regularity conditions required for the theorem to hold [39]. Therefore, our proof proceeds from first principles.

It can be shown that Definition 4.3 ensures that $n/l_n(f)$ goes to 1, as n grows large (cf. Lemma 3.3). Therefore, Proposition 3.1 yields

$$\limsup_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_1\big(Y_f = 0\big) \leq -\Lambda_{1,h}^*(\gamma, t^{(h)}), \tag{4.21}$$

$$\limsup_{n\to\infty} \frac{1}{l_n(f)} \log \mathbb{P}_0\big(Y_f = 1\big) \le -\Lambda^*_{0,h}(\gamma, t^{(h)}). \tag{4.22}$$

Next, we show the following lower bound by induction on k: for all nodes v at level k, where $k \ge 1$, and for any $\epsilon > 0$, there exists a positive N such that for all $n \ge N$, we have

$$\frac{1}{l_n(v)} \log \mathbb{P}_1\big(Y_v = 0\big) \ge -\Lambda^*_{1,k}(\gamma, t^{(k)}) - \epsilon, \tag{4.23}$$

$$\frac{1}{l_n(v)} \log \mathbb{P}_0\big(Y_v = 1\big) \ge -\Lambda^*_{0,k}(\gamma, t^{(k)}) - \epsilon. \tag{4.24}$$

When $k = 1$, it is straightforward to apply Theorem 2.1 and Assumption 4.1 to obtain (4.23) and (4.24). Details are omitted. Now assume that the claim holds for all level k nodes. We apply Theorem 2.1 to show that (4.23)-(4.24) hold for all level $k+1$ nodes v. For each $u \in C_n(v)$, let $L_{u,n}$ be the log-likelihood ratio of the message sent by u. Let

$$\varphi(\lambda) = \log \mathbb{E}_1\big[e^{\lambda S_n(v)/l_n(v)}\big] = \sum_{u\in C_n(v)} \log \mathbb{E}_1\big[e^{\lambda L_{u,n}/l_n(v)}\big],$$

and its corresponding Fenchel-Legendre transform be

$$\begin{aligned}
\Phi(x) &= \sup_{\lambda\in\mathbb{R}}\{\lambda x - \varphi(\lambda)\} \\
&= \sup_{\lambda\in\mathbb{R}} \Big\{\lambda x - \sum_{u\in C_n(v)} \log \mathbb{E}_1\big[e^{\lambda L_{u,n}/l_n(v)}\big]\Big\} \\
&\le \sum_{u\in C_n(v)} l_n(u) \sup_{\lambda\in\mathbb{R}} \Big\{\frac{\lambda}{l_n(v)} x - \frac{1}{l_n(u)} \log \mathbb{E}_1\big[e^{\lambda L_{u,n}/l_n(v)}\big]\Big\} \\
&= \sum_{u\in C_n(v)} l_n(u) \sup_{\lambda\in\mathbb{R}} \Big\{\lambda x - \frac{1}{l_n(u)} \log \mathbb{E}_1\big[e^{\lambda L_{u,n}}\big]\Big\}.
\end{aligned} \tag{4.25}$$

Note that the supremum inside the sum on the R.H.S. of (4.25) is achieved by some $\lambda \in [-1, 0]$. For such a λ, we have

$$\frac{1}{l_n(u)} \log \mathbb{E}_1\big[e^{\lambda L_{u,n}}\big]$$

$$= \frac{1}{l_n(u)} \log \Big[\mathbb{P}_1(Y_u = 0)^{1+\lambda} \mathbb{P}_0(Y_u = 0)^{-\lambda} + \mathbb{P}_1(Y_u = 1)^{1+\lambda} \mathbb{P}_0(Y_u = 1)^{-\lambda} \Big]$$
$$\geq \max \Big\{ \frac{1+\lambda}{l_n(u)} \log \mathbb{P}_1(Y_u = 0), -\frac{\lambda}{l_n(u)} \log \mathbb{P}_0(Y_u = 1) \Big\}$$
$$+ \frac{1}{l_n(u)} \log \min \big\{ \mathbb{P}_0(Y_u = 0), \mathbb{P}_1(Y_u = 1) \big\}.$$

From (4.21)-(4.22), if n is large, we have $\min \big\{ \mathbb{P}_0(Y_u = 0), \mathbb{P}_1(Y_u = 1) \big\}$ is close to 1. Together with the induction hypothesis on the node u, we then have for all n sufficiently large,

$$\frac{1}{l_n(u)} \log \mathbb{E}_1 \big[e^{\lambda L_{u,n}} \big] \geq \max \big\{ -(1+\lambda) \Lambda^*_{1,k}(\gamma, t^{(k)}), \lambda \Lambda^*_{0,k}(\gamma, t^{(k)}) \big\} - \epsilon/3$$
$$= \Lambda_{1,k}(\gamma, t^{(k)}; \lambda) - \epsilon/3.$$

Continuing from (4.25), we obtain

$$\Phi(x) \leq \sum_{u \in C_n(v)} l_n(u) \sup_{\lambda \in \mathbb{R}} \Big\{ \lambda x - \Lambda_{1,k}(\gamma, t^{(k)}; \lambda) \big] \Big\} + l_n(v) \frac{\epsilon}{3}$$
$$= l_n(v) \big(\Lambda^*_{1,k+1}(\gamma, (t^{(k)}, x)) + \epsilon/3 \big). \tag{4.26}$$

Next, we show that there exists a finite positive constant c such that for all $\lambda \in [-l_n(v), 0]$, $\varphi''(\lambda) \leq c^2 \kappa_n$. For any level k node u, we have

$$\frac{L_{u,n}}{l_n(u)} = \frac{1}{l_n(u)} \log \frac{\mathbb{P}_1(Y_u = y)}{\mathbb{P}_0(Y_u = y)},$$

where $y \in \{0, 1\}$. The induction hypothesis yields that for n large, we have $|L_{u,n}| \leq l_n(u)c$, for some constant c. Therefore, we obtain

$$\varphi''(\lambda) \leq \sum_{u \in C_n(v)} \frac{1}{l_n^2(v)} \frac{\mathbb{E}_1 \big[L^2_{u,n} \exp(\lambda L_{u,n}/l_n(v)) \big]}{\mathbb{E}_1 \big[\exp(\lambda L_{u,n}/l_n(v)) \big]}$$
$$\leq c^2 \sum_{u \in C_n(v)} \frac{l_n^2(u)}{l_n^2(v)}$$
$$\leq \frac{c^2 \max_{u \in C_n(v)} l_n^2(u)}{|C_n(v)| \min_{u \in C_n(v)} l_n^2(u)}$$

$$\leq c^2\kappa_n, \tag{4.27}$$

where the last inequality follows from Definition 4.3. Fix a $x < t_{k+1}$, and a sufficiently small positive δ that is less than $\epsilon/3$, so that $x+\delta \leq t_{k+1}$. Then, it is an easy exercise to check that if t_{k+1} satisfies (3.11), and if n is large, the conditions in Theorem 2.1 are satisfied for $X = S_n(v)/l_n(v)$, and for some $\lambda \in [-l_n(v), 0]$.

Let $l(n) = min_{v\in B_n} l_n(v)$. (Recall that B_n is the set of level 1 nodes, so we have $l_n(v) \geq l(n)$ for all non-leaf nodes v.) From Theorem 2.1, we have

$$\begin{aligned}
\frac{1}{l_n(v)} \log \mathbb{P}_1\big(Y_v = 0\big) &= \frac{1}{l_n(v)} \log \mathbb{P}_1\big(S_n(v)/l_n(v) \leq t_{k+1}\big) \\
&\geq \frac{1}{l_n(v)} \log \mathbb{P}_1\big(S_n(v)/l_n(v) \leq x + \delta\big) \\
&\geq -\frac{1}{l_n(v)}\Phi(x) - \frac{|\lambda|}{l_n(v)}\delta + \frac{1}{l_n(v)} \log\big(1 - \varphi''(\lambda)/\delta^2\big) \\
&\geq -\Lambda^*_{1,k+1}(\gamma, (t^{(k)}, x)) - \epsilon/3 - \epsilon/3 + \frac{1}{l(n)} \log\big(1 - c^2\kappa_n/\delta^2\big) \\
&\geq -\Lambda^*_{1,k+1}(\gamma, t^{(k+1)}) - \epsilon,
\end{aligned}$$

where the penultimate inequality follows from (4.26) and (4.27). The last inequality follows by taking n sufficiently large, and from the continuity of the function $\Lambda^*_{1,k+1}(\gamma, (t^{(k)}, \cdot))$. We have now shown (4.23) for all level $k+1$ nodes; the proof of (4.24) is similar. The induction is now complete, and the proposition is proved.

Chapter 5

Error Prone Sensor Networks

In Chapter 3, we studied the detection performance of bounded height tree networks, as the number of nodes increases. For a Neyman-Pearson binary hypothesis testing problem, we showed the somewhat surprising result that, under certain mild conditions, the asymptotically optimal detection performance (in terms of the error exponent) is the same as for the parallel configuration. This implies that a large network can be designed so that it is energy efficient, while matching the performance of the ideal, parallel configuration. However, in that chapter, we have not accounted for the possibility of node failures, and we have assumed that all messages are received reliably. In this chapter, we address these two issues, in the context of dense sensor networks. We also aim to obtain qualitative insights into the management of sensor networks. Throughout, we will again assume Assumption 2.1, i.e., nodes make (conditionally) i.i.d. observations under either hypothesis.

5.1 Related Work and Our Contributions

Parallel configurations with a random number of nodes have been studied by [43–45]. In [43, 44], the authors consider spatially correlated signals, and analyze the performance of a simple but suboptimal strategy. In [45], the authors assume that all nodes make i.i.d. observations under either hypothesis, quantize their observations using the same quantizer, and use a multiple access protocol that combines the sensor

messages in an additive fashion. In this thesis, we model the case of node failures by allowing the number of nodes that transmit messages to a particular node be a random variable with a known distribution. Then, we let the mean of this distribution become asymptotically large, to model a dense network. The work in this thesis differs from the previous references in several ways, including the following: (a) we are interested in evaluating the asymptotically optimal detection performance, and in designing asymptotically optimal transmission strategies; (b) we focus on trees with height greater than one. Our results show that for a dense network whose expected number of leaves is n, and under a particular assumption on the distribution of the degree of each node, the asymptotic Neyman-Pearson detection performance is the same as for a parallel configuration with n leaves, thus establishing that the randomness in the network topology does not lead to performance deterioration.

For the case of unreliable communications, we assume that all nodes are constrained to sending one-bit messages over a binary symmetric channel (BSC) with known crossover probability. To model a dense network, we let the degree of each non-leaf node grow asymptotically large. The case of the parallel configuration is covered by results in [22]. Parallel configurations with a fixed number of nodes, and with non-ideal channels between the nodes and the fusion center, have also been studied in [10–12, 21]. In this work, we study the effect of unreliable communications on the detection performance of a tree network of height greater than one, and characterize the optimal error exponent. In particular, we show that it is no longer possible to achieve the performance of a parallel configuration, in contrast to the results in Chapter 3. We also consider a scheme that allows a tree network to achieve the same performance as that of a network with reliable communications, but at the expense of increased transmission power. We compare the energy efficiency of such a scheme with that of a parallel configuration, and establish that a tree network is preferable.

Finally, we consider the Bayesian version of the problems we have described above, under some additional simplifying assumptions, and characterize the optimal error exponent.

5.2 Node Failures

We model node failures by letting the number of immediate predecessors of each node be random variables with known distributions. Although [43–45] have studied variations of this problem in a different context, they specifically assumed a Poisson distribution and considered only the parallel configuration. Our formulation involves trees with a general height $h \geq 1$, and distributions from a somewhat larger family. The main reason for introducing this larger family of distributions is to facilitate comparison with the results in Section 5.3. Since we will not be indexing our tree sequence using the number of nodes n in each tree (which is now random), we will drop the subscript n in our notation. For example, the set of immediate predecessors of a node v will be denoted as $C(v)$, instead of $C_n(v)$.

Let h be a positive integer. We form a random tree according to a Galton-Watson branching process [46] with h stages. Consider the fusion center f. Let $N_f = |C(f)|$ be a nonnegative integer random variable, with marginal law μ_h. For each node v in the random set $C(f)$, we let $N_v = |C(v)|$ be i.i.d. random variables with distribution 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 tree a *GW tree* of height h. We will sometimes use M_k to denote a generic random variable with law μ_k.

Let $\lambda_k = \mathbb{E}[M_k] < \infty$ be the mean[1] of the distribution μ_k, and let $\lambda^* = \min_{1 \leq k \leq h} \lambda_k$. We consider the case of asymptotically large λ^* to model a dense network, i.e., we let $\lambda_k \to \infty$ for all k, and allow the laws μ_k to vary accordingly. Strictly speaking, we are dealing with a sequence of random tree networks: each tree in the sequence corresponds to a different choice of the parameters λ_k, and these parameters tend to infinity along this sequence. However, we keep this underlying sequence hidden (and implicit), to prevent overburdening the notation. Let $\lambda(k) = \prod_{i=1}^{k} \lambda_i$, which is the

[1] When dealing with the distribution of the GW tree, we will use the notation $\mathbb{P}$, $\mathbb{E}$ and var, since the distribution is the same under either hypothesis.

expected number of leaves that are predecessors of a level k node.

We make the following assumption. The assumption is satisfied if M_k has Poisson distribution with mean λ_k, or if there is a constant $p \in (0,1)$ such that M_k has a Binomial distribution $\mathcal{B}(n_k, p)$ with mean $\lambda_k = n_k p$. If every N_v has a Binomial distribution, a GW tree can be interpreted as a deterministic tree network with erasure channels between nodes.

Assumption 5.1. *Let M_k be random variables with distribution μ_k and mean λ_k, $k = 1, \ldots, h$. We have*

$$\mathrm{var}[M_k]/\lambda_k^2 \to 0, \qquad \text{as } \lambda_k \to \infty. \tag{5.1}$$

Under Assumption 5.1, a straightforward application of Chebychev's inequality shows that the distribution of M_k is clustered around its mean.

Lemma 5.1. *For any $\eta > 0$, we have $\mathbb{P}(|M_k/\lambda_k - 1| > \eta) \to 0$, as $\lambda_k \to \infty$.*

We assume that there is a multiple access protocol in place, so that every node can distinguish the messages it receives from each of its immediate predecessors. This can for example, be a random access protocol that allows a large number of immediate predecessors. Lemma 5.1 shows that, with high probability, approximately λ_k orthogonal channels are required if a random access orthogonal signalling scheme is implemented.

Suppose that the distributions $\mu_1, \ldots, \mu_h$ have been fixed. A transmission *policy* for a node v specifies the transmission function of v, for each realization of the in-degree N_v. Similarly, a *GW-strategy* is defined as a mapping, which for any realization of the random tree, specifies a strategy (as defined at the end of Section 2.2) for that tree. Note that a GW-strategy requires, in general, global information on the structure of the realized tree and may be hard to implement. Given a GW-strategy π and a set of distributions $\mu = (\mu_1, \ldots, \mu_h)$, let β_π be the resulting Type II error probability, $\mathbb{P}_1(Y_f = 0)$, at the fusion center. (This is an average over all possible realizations of the tree, as well as over the distribution of the observations.) Let us

fix some $\alpha \in (0,1)$. Let β^* be the infimum of β_π, over all GW-strategies π, subject to the constraint that the Type I error probability, $\mathbb{P}_0(Y_f = 1)$, is less than or equal to α. Similar to previous chapters, our goal is to characterize the optimal error exponent

$$\limsup_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^*.$$

(Recall that $\lambda(h) = \prod_{k=1}^{h} \lambda_k$ is the expected number of leaves, as determined by μ.)

Given a GW-strategy and a level k node v, let L_v be the log-likelihood ratio (more formally, the logarithm of the Radon-Nikodym derivative) of the distribution of Y_v under H_1 with respect to that under H_0. If v is at level $k \geq 1$, we define the log-likelihood ratio of the messages it receives by

$$S_v = \sum_{u \in C(v)} L_u,$$

where the sum is taken to be 0 if $C(v)$ is empty.

Motivated by the ϵ-optimal strategies for non-random tree networks in Chapter 3, we will be interested in the case where nodes v at some level $k \geq 1$ use a transmission policy (called a Mean-normalized Log-Likelihood Ratio (MLLR) quantizer) that results in a message Y_v of the form

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

for some threshold t. We assume that all non-leaf nodes are allowed to use MLLR quantizers.

For deterministic network topologies, i.e., if $N_v = \lambda_k$, a.s., for all level k nodes v, our results in Chapter 3 show 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 5.1. *Suppose that Assumptions 2.1-2.3, and Assumption 5.1 hold, and*

that $\alpha \in (0,1)$. The optimal error exponent of a GW tree of height h is given by

$$\lim_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^* = g_P^*. \tag{5.2}$$

Furthermore, for any $\epsilon \in (0, -g_P^*)$, and any large enough λ^*, the following GW-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 $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) \leq g_P^* + \epsilon/2 < 0$; and

(ii) for $k \geq 1$, every level k node uses a MLLR quantizer with threshold $t_k = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) + \epsilon/2^{h-k+1}$.

To prove the proposition, we will first lower bound the optimal error exponent. We will then derive a matching upper bound by showing that the proposed GW-strategy comes within ϵ of the lower bound.

5.2.1 The Lower Bound

In this section, we show that in the limit, as $\lambda^* \to \infty$, and for any GW-strategy, the error exponent is lower bounded by g_P^*. We first show an elementary fact.

Lemma 5.2. *Suppose that X and Y are non-negative random variables with $\mathbb{E}[X] \leq a$ and $\mathbb{E}[Y] \leq b$, and that the event A has probability $\mathbb{P}(A) > c_1 + c_2$, where $c_1, c_2 > 0$. Then, there exists some $\omega \in A$ such that $X(\omega) \leq a/c_1$ and $Y(\omega) \leq b/c_2$.*

Proof. From Markov's Inequality,

$$\mathbb{P}(X > a/c_1) \leq c_1, \quad \mathbb{P}(Y > b/c_2) \leq c_2.$$

Therefore, by the union bound, we have

$$\mathbb{P}\big((\{X > a/c_1\} \cup \{Y > b/c_2\}) \cap A\big) \leq c_1 + c_2.$$

This implies that

$$\mathbb{P}(\{X \leq a/c_1\} \cap \{Y \leq b/c_2\} \cap A) \geq \mathbb{P}(A) - c_1 - c_2 > 0.$$

Hence, there exists some $\omega \in A$ such that $X(\omega) \leq a/c_1$ and $Y(\omega) \leq b/c_2$. □

In the following lemma, we show that $l(f)$ (the actual number of leaves) and $\lambda(h)$ (the expected number of leaves) are close (in probability), in the limit of large λ^*.

Lemma 5.3.

(a) $\mathbb{E}[l(f)] = \lambda(h)$ *and* $\text{var}[l(f)]/\lambda^2(h) \to 0$, *as* $\lambda^* \to \infty$.

(b) For all $\delta > 0$, $\mathbb{P}(|l(f)/\lambda(h) - 1| > \delta) \to 0$, *as* $\lambda^* \to \infty$.

Proof.

(a) We use induction on h. For $h = 1$, (a) follows from Assumption 5.1. Suppose that the claim holds for GW trees of height $h-1$, and consider a GW tree of height h. Recall that N_f is the cardinality of the set $C(f)$ of immediate predecessors of the fusion center f. For $u \in C(f)$, we observe that $l(u)$ is the number nodes in a GW tree of height $h-1$, rooted at u. The induction hypothesis yields $\mathbb{E}[l(u)] = \lambda(h-1)$ and $\text{var}[l(u)]/\lambda^2(h-1) \to 0$, as $\lambda^* \to \infty$. Furthermore, the random variables $l(u)$ are i.i.d. Let w be a typical element of $C(f)$. We have

$$\mathbb{E}[l(f)] = \mathbb{E}\Big[\sum_{u \in C(f)} l(u)\Big] = \mathbb{E}[N_f]\,\mathbb{E}[l(w)] = \lambda(h).$$

Using a well known formula for the variance of the sum of a random number of i.i.d. random variables, we also have

$$\begin{aligned}\frac{\text{var}[l(f)]}{\lambda^2(h)} &= \frac{\text{var}[N_f]\big(\mathbb{E}[l(w)]\big)^2 + \mathbb{E}[N_f]\text{var}[l(w)]}{\lambda^2(h)} \\ &= \frac{\text{var}[N_f]}{\lambda_h^2} \cdot \frac{\lambda^2(h-1)}{\lambda^2(h-1)} + \frac{\lambda_h}{\lambda_h^2} \cdot \frac{\text{var}[l(w)]}{\lambda^2(h-1)},\end{aligned}$$

which, using the induction hypothesis and Assumption 5.1, converges to 0.

(b) This is an immediate consequence of Chebychev's Inequality.

□

We are now ready to prove the lower bound for the optimal error exponent.

Lemma 5.4. *Suppose that Assumptions 2.1-2.3, and Assumption 5.1 hold, and that* $\alpha \in (0,1)$. *Then,*

$$\liminf_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^* \geq g_P^*.$$

Proof. Suppose that $g := \liminf_{\lambda^* \to \infty} \frac{1}{\lambda(h)} \log \beta^* < g_P^*$. Fix $\epsilon > 0$ and $\delta \in (0,1)$ such that $(g+\epsilon)/(1+\delta) < g_P^*$. Then, there exists a sequence of distributions $(\mu_1, \ldots, \mu_h)$ along which $\lambda^* \to \infty$, such that for the kth element of that sequence we have $\lambda(h) = \zeta_k$, where $\zeta_1 \geq 1$, $\zeta_{k+1} \geq \frac{1+\delta}{1-\delta}\zeta_k$, $k = 1, 2, \ldots$, and

$$\lim_{k \to \infty} \frac{1}{\zeta_k} \log \beta^* = g.$$

Let $\mathcal{G}$ be the set of all trees with height less than or equal to h, and let R_k be a random tree, generated according to the GW process. It follows that there exists some $K_1 > 0$ such that for all $k \geq K_1$, we have

$$\begin{aligned} \mathbb{E}[\mathbb{P}_1(Y_f = 0 \mid R_k)] =& \mathbb{P}_1(Y_f = 0) \leq e^{\zeta_k(g+\epsilon)}, \\ \mathbb{E}[\mathbb{P}_0(Y_f = 1 \mid R_k)] =& \mathbb{P}_0(Y_f = 1) \leq \alpha. \end{aligned}$$

Fix a $c \in (\alpha, 1)$. From Lemma 5.3(b), $\mathbb{P}(l(f) \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]) \to 1$, as $k \to \infty$. Since $(1+c)/2 < 1$, we can choose a $K \geq K_1$, such that for all $k \geq K$,

$$\mathbb{P}\big(l(f) \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]\big) > \frac{1+c}{2} = \frac{1-c}{2} + c.$$

Using Lemma 5.2, for each $k \geq K$, there exists some tree $r_k \in \mathcal{G}$ with n_k leaves, where $n_k \in [(1-\delta)\zeta_k, (1+\delta)\zeta_k]$, so that

$$\mathbb{P}_1\big(Y_f = 0 \mid R_k = r_k\big) \quad \leq \quad \frac{2}{1-c} e^{\zeta_k(g+\epsilon)}, \tag{5.3}$$

$$\mathbb{P}_0\big(Y_f = 1 \mid R_k = r_k\big) \;\leq\; \frac{\alpha}{c} \;<\; 1. \tag{5.4}$$

From (5.3),

$$\begin{aligned}\frac{1}{n_k}\log \mathbb{P}_1\big(Y_f = 0 \mid R_k = r_k\big) &\leq \frac{\zeta_k}{n_k}(g+\epsilon) + \frac{1}{n_k}\log\frac{2}{1-c} \\ &\leq \frac{g+\epsilon}{1+\delta} + \frac{1}{n_k}\log\frac{2}{1-c}.\end{aligned}$$

Letting $k \to \infty$, we obtain

$$\limsup_{k\to\infty} \frac{1}{n_k}\log \mathbb{P}_1\big(Y_f = 0 \mid R_k = r_k\big) \leq \frac{g+\epsilon}{1+\delta} < g_P^*.$$

Recall that g_P^* is the optimal Type II error exponent (as $k \to \infty$) of a parallel configuration with n_k nodes sending messages directly to the fusion center, subject to the constraint that the Type I error probability is less than or equal to α/c (cf. (5.4) and [22]). Since such a parallel configuration can simulate the tree r_k, we obtain a contradiction, which proves the desired result. □

5.2.2 Achievability

In this subsection, we fix some $\epsilon \in (0, -g_P^*)$, consider a GW-strategy of the form described in Proposition 5.1, and show that it performs as claimed. In particular, for $k \geq 1$, every level k node v sends a 0 (or, for the fusion center, it declares H_0) iff $S_v \leq \lambda(k)t_k$.

We first show that this strategy results in a Type II error exponent within ϵ of g_P^*. Consider a node v at level $k \geq 1$. Since $\exp(-S_v)$ is the ratio of the likelihood under H_0 to that under H_1, of the received messages at node v, we have $\mathbb{E}_1[\exp(-S_v)] = 1$. Hence, from the Chernoff bound, we obtain

$$\frac{1}{\lambda(k)}\log \mathbb{P}_1\Big(\frac{S_v}{\lambda(k)} \leq t_k\Big) \leq \frac{1}{\lambda(k)}\log\Big(e^{\lambda(k)t_k}\mathbb{E}_1[e^{-S_v}]\Big) = t_k. \tag{5.5}$$

In particular, for $v = f$, we have $k = h$ and

$$\frac{1}{\lambda(h)} \log \mathbb{P}_1\big(S_f \leq \lambda(h) t_h\big) \leq t_h = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) + \frac{\epsilon}{2}$$
$$\leq g_P^* + \epsilon. \tag{5.6}$$

By taking $\epsilon \to 0$ in (5.6), we obtain the claimed upper bound on the Type II error exponent.

It only remains to verify that this strategy meets the Type I error constraint, when λ^* is sufficiently large. This is accomplished by the following lemma.

Lemma 5.5. *Suppose that Assumptions 2.1-2.3, and Assumption 5.1 hold. Let v be a level k node, with $k \geq 1$. For the particular GW-strategy proposed in Proposition 5.1, we have $\mathbb{P}_0(Y_v = 0) \to 1$, as $\lambda^* \to \infty$.*

Proof. We proceed by induction on k. We start by considering the case $k = 1$. Let u be a typical immediate predecessor of v. We have

$$\frac{\mathbb{E}_0[S_v]}{\lambda(1)} = \frac{\mathbb{E}[N_v]}{\lambda(1)} \cdot \mathbb{E}_0[L_u] = \mathbb{E}_0[L_u] = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma).$$

Furthermore, using a well known formula for the variance of the sum of a random number of i.i.d. random variables,

$$\frac{\mathrm{var}_0[S_v]}{\lambda^2(1)} = \frac{\mathrm{var}[N_v](\mathbb{E}_0[L_u])^2 + \mathbb{E}[N_v]\mathrm{var}_0[L_u]}{\lambda^2(1)},$$

which converges to zero as $\lambda(1) \to \infty$, because $\mathrm{var}[N_v]/\lambda^2(1)$ converges to zero (Assumption 5.1), $\mathbb{E}_0[L_u] = -\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < \infty$, $\mathbb{E}[N_u] = \lambda(1)$, and $\mathrm{var}_0[L_u] \leq \mathbb{E}_0[L_u^2] < \infty$ (from Assumption 2.3 and Proposition 3 of [22]). Since the threshold t_1 used by v satisfies $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < t_1$, Chebychev's inequality yields $\mathbb{P}_0(S_v > \lambda(1)t_1) \to 0$, and, therefore, $\mathbb{P}_0(Y_v = 0) \to 1$.

Suppose now that the induction hypothesis holds for $k - 1$, where $k \geq 2$. Let v be a level k node, and let u be a typical immediate predecessor of v. Using the facts

$\mathbb{P}_0(Y_u = 0) \to 1$ and $\mathbb{P}_0(Y_u = 1) \to 0$ in the second equality below, we have

$$\begin{aligned}
&\limsup_{\lambda^* \to \infty} \frac{\mathbb{E}_0[L_u]}{\lambda(k-1)} \\
&= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k-1)} \left(\mathbb{P}_0(Y_u = 0) \log \frac{\mathbb{P}_1(Y_u = 0)}{\mathbb{P}_0(Y_u = 0)} + \mathbb{P}_0(Y_u = 1) \log \frac{\mathbb{P}_1(Y_u = 1)}{\mathbb{P}_0(Y_u = 1)} \right) \\
&= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k-1)} \log \mathbb{P}_1(Y_u = 0) \\
&= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k-1)} \log \mathbb{P}_1\big(S_u \le \lambda(k-1)t_{k-1}\big) \\
&\le t_{k-1},
\end{aligned}$$

where the last inequality follows from (5.5), applied to u.

Using a similar argument, we have

$$\begin{aligned}
\limsup_{\lambda^* \to \infty} \frac{\mathrm{var}_0[L_u]}{\lambda^2(k-1)} &\le \limsup_{\lambda^* \to \infty} \frac{\mathbb{E}_0[L_u^2]}{\lambda^2(k-1)} \\
&= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda^2(k-1)} \log^2 \mathbb{P}_1(S_u \le \lambda(k-1)t_{k-1}) \\
&\le t_{k-1}^2. \qquad (5.7)
\end{aligned}$$

We then obtain

$$\begin{aligned}
\limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k)} \mathbb{E}_0[S_v] &= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k)} \mathbb{E}[N_v] \mathbb{E}_0[L_u] \\
&= \limsup_{\lambda^* \to \infty} \frac{1}{\lambda(k-1)} \mathbb{E}_0[L_u] \\
&\le t_{k-1}.
\end{aligned}$$

Furthermore,

$$\frac{\mathrm{var}_0[S_v]}{\lambda^2(k)} = \frac{\mathrm{var}[N_v](\mathbb{E}_0[L_u])^2 + \mathbb{E}[N_v]\mathrm{var}_0[L_u]}{\lambda_k^2 \cdot \lambda^2(k-1)},$$

which converges to zero as $\lambda^* \to \infty$, because $\mathrm{var}[N_v]/\lambda_k^2 \to 0$ (Assumption 5.1), $\mathbb{E}[N_v]/\lambda_k^2 = 1/\lambda_k \to 0$, and both $\mathbb{E}_0[L_u]/\lambda(k-1)$ and $\mathrm{var}_0[L_u]/\lambda^2(k-1)$ are bounded. Since $t_{k-1} < t_k$, Chebychev's inequality shows that $\mathbb{P}_0(S_v > \lambda(k)t_k) \to 0$, and therefore, $\mathbb{P}_0(Y_v = 0) \to 1$. □

5.2.3 Discussion

We have shown that the optimal error exponent for a tree network with node failures is g_P^*, the same as for a parallel configuration with a large but deterministic number of nodes, and developed a strategy that achieves the optimal performance, as close as desired. In our ϵ-optimal strategy, every non-leaf node uses a MLLR quantizer. Hence, there is no loss in optimality if we restrict each of the non-leaf nodes to sending only one bit.

Another advantage of this strategy is that every non-leaf node only needs to know the received messages from its immediate predecessors and the distributions $\mu_1, \ldots, \mu_h$; no additional information on the topology of the realized tree is required. While it might be possible, in a static network, as part of the setup process, to inform each node of the topology of the network, this would be too difficult or costly in a mobile or time-varying network. The model that we have adopted, i.e., modeling the immediate predecessors of each sensor as a random set, can be applied to a mobile network, in which a node does not know a priori how many nodes will be within transmission range. See [43] for a related model, employed in a similar spirit.

5.3 Unreliable Communications

In this section, we consider the case where messages are restricted to be binary, and the channel between any two nodes is a binary symmetric channel (BSC) with known crossover probability $\eta \in (0, 1/2)$. Recall that the sequence of trees $(T_n)_{n\geq 1}$ models the evolution of the network as more nodes are added. We assume that for some n_0, and for all $n \geq n_0$, T_n is a h-uniform tree, i.e., all leaves are exactly h hops away from the fusion center (this is done for simplicity, to reduce the number of cases that we need to consider; an extension to more general types of trees is possible). For every non-leaf node v, we assume that $|C_n(v)| \geq c_n$, for some sequence c_n of positive integers that diverges to infinity as n increases. Similar to the previous section, we are interested in characterizing the Type II error exponent at the fusion center, when the Type I error probability is constrained to be less than or equal to a given $\alpha \in (0, 1)$.

However, in this case, it turns out that the relevant error exponent is

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

where β^* is the minimum Type II error probability at the fusion center, for the tree T_n, optimized over all strategies that satisfy the Type I error constraint. Note that we have normalized the error exponent using $|C_n(f)|$ (instead of $l_n(f)$, the total number of leaves), even though every leaf makes an observation. The reason for this will become apparent in Proposition 5.3 below.

Consider a non-leaf node v. It receives a message from each node $u \in C_n(v)$, and forms a message Y_v, which it sends to its immediate successor, w. Because of the noisy channel, the message received by w, denoted by Z_v, may be different from Y_v. Let $\bar{L}_v$ be the log-likelihood ratio of the distribution of Z_v under H_1 with respect to that under H_0. Since Z_v is binary, the random variable $\bar{L}_v$ takes one of the two values $\log(\mathbb{P}_1(Z_v = z)/\mathbb{P}_0(Z_v = z))$, $z = 0, 1$, depending on whether Z_v is 0 or 1. Let

$$S_v = \sum_{u\in C_n(v)} \bar{L}_u,$$

which is the sum of the log-likelihood ratios of the received messages at node v.

We will be interested in the case where nodes v at some level $k \geq 1$ use LLRQs as their transmission functions, i.e.,

$$Y_v = \begin{cases} 0, & \text{if } S_v/|C_n(v)| \leq t, \\ 1, & \text{otherwise,} \end{cases}$$

for some threshold t.

5.3.1 The Case $h = 1$

Let us first consider the simple case where $h = 1$, i.e., the parallel configuration. For every $\gamma \in \Gamma$, let

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

and

$$\begin{aligned} e_{0,\gamma} &= q_0^\gamma(0)\cdot\log\frac{q_1^\gamma(0)}{q_0^\gamma(0)} + q_0^\gamma(1)\cdot\log\frac{q_1^\gamma(1)}{q_0^\gamma(1)}, \\ e_{1,\gamma} &= q_1^\gamma(0)\cdot\log\frac{q_1^\gamma(0)}{q_0^\gamma(0)} + q_1^\gamma(1)\cdot\log\frac{q_1^\gamma(1)}{q_0^\gamma(1)}, \end{aligned}$$

For an interpretation, note that if u is a leaf that employs the transmission function γ, then $e_{j,\gamma} = \mathbb{E}_j[\bar{L}_u]$. Let $e_0 = \inf_{\gamma\in\Gamma} e_{0,\gamma}$. The following proposition follows immediately from [22].

Proposition 5.2. *Suppose that Assumptions 2.1-2.3 hold. Then, for $h = 1$, and for any $\alpha \in (0,1)$, we have*

$$\lim_{n\to\infty}\frac{1}{|C_n(f)|}\log\beta^* = e_0$$

Furthermore, the optimal error exponent does not change if we restrict all the leaves to use the same transmission function $\gamma \in \Gamma$.

As shown in [22], the optimal error exponent e_0 can be achieved to within some ϵ, by letting all leaves use a transmission function γ that satisfies $e_{0,\gamma} \le e_0 + \epsilon/2$, and letting the fusion center use a LLRQ with threshold $t = e_{0,\gamma} + \epsilon/2$.

5.3.2 The General Case

We henceforth assume that $h \geq 2$. We have the following proposition, which shows that the optimal error exponent is the same as that of a parallel configuration in which the nodes in $C_n(f)$ have perfect knowledge of the true hypothesis. Intuitively, as n becomes large, each node $v \in C_n(f)$ discriminates between the two hypotheses with vanishing probabilities of error. Let $\text{Bern}(\eta)$ denote the Bernoulli distribution on $\{0, 1\}$ that takes value 1 with probability η. Let

$$D(\eta) = \eta \log \frac{\eta}{1-\eta} + (1-\eta) \log \frac{1-\eta}{\eta},$$

which is the KL divergence function of $\text{Bern}(1-\eta)$ w.r.t. $\text{Bern}(\eta)$.

Proposition 5.3. *Suppose that Assumptions 2.1-2.3 hold, $h \geq 2$, and $\alpha \in (0, 1)$. Then, the optimal error exponent is*

$$\lim_{n\to\infty} \frac{1}{|C_n(f)|} \log \beta^* = -D(\eta) < 0.$$

Furthermore, for any $\epsilon > 0$, as $n \to \infty$, the following strategy satisfies the Type I error probability constraint, and also satisfies $\limsup_{n\to\infty}(1/|C_n(f)|) \log \mathbb{P}_1(Y_f = 0) \leq -D(\eta) + \epsilon$:

(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 node at level 1 uses a LLRQ, with a threshold t that satisfies $e_{0,\gamma} < t < e_{1,\gamma}$.

(iii) All other nodes other than the fusion center, 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 LLRQ with threshold $t = -D(\eta) + \epsilon$.

Proof. (Outline) Similar to the proof of Proposition 5.1, we first lower bound the optimal error exponent. Consider the fusion center f. Suppose a genie tells each

$v \in C_n(f)$ the true hypothesis, and each node v sends this information to the fusion center. Because of the BSC from each sensor v to f, the received message at f has distribution Bern(η) under H_0, and Bern($1-\eta$) under H_1. From Stein's Lemma [39], the optimal error exponent is $-D(\eta)$. The performance in the absence of the genie cannot be better. Therefore,

$$\liminf_{n\to\infty} \frac{1}{|C_n(f)|} \log \beta^* \geq -D(\eta). \tag{5.8}$$

We now turn to the proof of the upper bound. Consider the strategy described in the proposition. Let v be a node at level 1. This node v receives a message Z_u from each leaf $u \in C_n(v)$. These messages are binary, conditionally i.i.d., but with a different distribution under each hypothesis. Moreover, v receives at least c_n such messages. In such a case, it is well known [39] (and also easy to show from laws of large numbers) that if the node v uses a LLRQ with a threshold t that satisfies $e_{0,\gamma} < t < e_{1,\gamma}$, then the error probabilities at node v decay exponentially fast with c_n; that is, there exist some Δ and $\delta > 0$ such that

$$\mathbb{P}_0(Y_v = 1) \leq \Delta e^{-c_n\delta}, \quad \mathbb{P}_1(Y_v = 0) \leq \Delta e^{-c_n\delta}, \quad \forall n. \tag{5.9}$$

Taking into account the statistics of the BSC, we have

$$\mathbb{P}_0(Z_v = 1) \leq \eta + \Delta e^{-c_n\delta}, \quad \mathbb{P}_1(Z_v = 0) \leq \eta + \Delta e^{-c_n\delta}, \ \forall n. \tag{5.10}$$

In particular, for n sufficiently large, and for all level 1 nodes v, we have $\mathbb{P}_0(Z_v = 1) < 1/2$ and $\mathbb{P}_1(Z_v = 0) < 1/2$. Consider now a node w at level 2. This node receives at least c_n independent messages Z_v from each $v \in C_n(w)$, where these messages have error probabilities $\mathbb{P}_0(Z_v = 1) < 1/2$ and $\mathbb{P}_1(Z_v = 0) < 1/2$. The node w then uses a majority rule to form its message Y_w. It is easy to show, using laws of large numbers, that (5.9) holds for Y_w, with possibly different constants Δ and δ. Then, (5.10) also holds for Z_w. Continuing inductively, we conclude that there exist constants $\Delta > 0$ and $\delta > 0$, such that for all nodes v, (5.10) holds.

Consider now the fusion center, and a typical node $v \in C_n(f)$. From (5.10), if n is sufficiently large, the message Z_v received by the fusion center has KL divergence at least $D(\eta) - \epsilon/2$ (note that $D(\cdot)$ is continuous and decreasing over $(0, 1/2)$). It then follows, from Cramér's Theorem [39], that the Type II error exponent at the fusion center is less than or equal to $-D(\eta) + \epsilon$, if a LLRQ with threshold $t = -D(\eta) + \epsilon$ is used at the fusion center. Moreover, the Type I error exponent is strictly negative in this case, so that the Type I error probability can be brought to below α when n is sufficiently large. The proof is now complete. □

5.3.3 Discussion

We have established that the detection performance of a tree network in which the communication channel between two nodes is a BSC, and which has a height $h \geq 2$, is the same as if every immediate predecessor of the fusion center had perfect knowledge of the true hypothesis. On the other hand, when compared to the case of reliable communications (where the error probability falls exponentially fast with the number of nodes cf. Chapter 3), the performance is significantly degraded. Thus, channel noise can be detrimental.

Consider a tree network in which all non-leaf nodes have the same number of immediate predecessors c_n. Suppose that each node estimates its channel to its immediate successor, and sends its message only if that message will be received reliably. In this case, the number of immediate predecessors of a node of level $k \geq 1$ has a Binomial distribution $\mathcal{B}(c_n, 1-\eta)$ with mean $\lambda_k = c_n(1-\eta)$. In Section 5.2, we showed that the Type II error probability, when the network is operating in this manner, falls exponentially with $\lambda(h) = c_n^h(1-\eta)^h$. On the other hand, Proposition 5.3 shows that the minimum error probability achievable when messages are sent regardless of channel conditions, falls exponentially with c_n. Hence, our results suggest that in a dense sensor network of height $h \geq 2$, if a node determines that it cannot reliably transmit its message to its immediate successor, it is better for the node to remain silent. Our results also suggest that when designing a large scale sensor network, it is more important to ensure that there is reliable communication between nodes (e.g.,

by using sufficient transmission power), than to guard against node failures.

It has been argued in [47, 48] that, contrary to popular belief, in most practical scenarios, multi-hop wireless sensor networks are less energy efficient than the simple parallel configuration, if the network is used for the purpose of forwarding data to the fusion center (with no fusion involved at intermediate nodes). The model we have considered, however, requires intermediate data to be fused on the way to the fusion center into 1-bit messages. In this case, it is easy to see that a tree network is more energy efficient than a parallel configuration (for a more detailed comparison, see the example in Chapter 4). On the other hand, with unreliable communications, Proposition 5.3 indicates that the detection performance of a tree network is much worse than that of a parallel configuration, leading to an interesting tradeoff. In Section 5.3.5, we will consider an example where nodes are arranged on a grid. We will show that it is possible to construct a tree that is more energy-efficient than the parallel configuration, while guaranteeing that the Type II error probability decays exponentially fast with the number of nodes. However, we first need to develop a result on the optimal error exponent, when the channel crossover probability η decreases with the number of nodes.

5.3.4 Error Exponent with Small Channel Crossover Probabilities

In Proposition 5.3, we showed that the Type II error probability decays exponentially fast with $|C_n(f)|$, when the channel error probability η is fixed. In this section, we let η go to zero as n increases, which corresponds to increasing the transmit power of each node.[2] Under an assumption on the rate at which η goes to zero, we show that the Type II error probability can be made to decay exponentially fast with n, at rate g_P^*.

Proposition 5.4. *Suppose that Assumptions 2.1-2.3 hold. Suppose also the following:*

1. *If $h = 1$, then $\lim_{n\to\infty} \eta = 0$.*

[2]We suppress the dependence of η on n in the notation.

2. Let $l_M = \max_{v \in C_n(f)} l_n(v)$. *If* $h \geq 2$, *then* $\limsup_{n\to\infty} \frac{1}{l_M} \log \eta \leq g_P^*$.

Fix an $\epsilon \in (0, -g_p^*/h)$. *Suppose that all leaves use the same transmission function* $\gamma \in \Gamma$, *chosen so that* $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma) < g_P^* + \epsilon$. *Suppose also that each level* k $(k \geq 1)$ *node* v *sends a message 0 iff* $S_v/l_n(v) \leq t_k := g_P^* + k\epsilon$. *Then, for* n *sufficiently large, we have for every level* k *node* v,

$$\frac{1}{l_n(v)} \log \mathbb{P}_1(Y_v = 0) \leq t_k, \tag{5.11}$$

$$\frac{1}{l_n(v)} \log \mathbb{P}_0(Y_v = 1) \leq -\epsilon_k < 0, \tag{5.12}$$

where $\epsilon_1, \ldots, \epsilon_h$ *are positive reals less than or equal to* ϵ. *In particular, for any* $h \geq 1$ *and* $\alpha \in (0, 1)$, *the optimal error exponent is*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^* = g_P^*.$$

Proof. If $h = 1$, the situation is similar to the case considered in Section 5.3.1. As $\eta \to 0$, $e_{0,\gamma}$ approaches $-\mathrm{D}(\mathbb{P}_0^\gamma \,\|\, \mathbb{P}_1^\gamma)$, and e_0 approaches g_P^*, which leads to the desired result. The details are omitted.

We now consider the case where $h \geq 2$. From the Chernoff bound, we have

$$\frac{1}{l_n(v)} \log \mathbb{P}_1\Big(\frac{S_v}{l_n(v)} \leq t_k\Big) \leq \frac{1}{l_n(v)} \log \Big(e^{l_n(v)t_k} \mathbb{E}_1[e^{-S_v}]\Big) = t_k,$$

hence (5.11) follows. To show the inequality (5.12), we proceed by induction on k. When $k = 1$, the inequality follows from Cramér's Theorem [39]. Suppose that (5.12) holds for all level k nodes. Consider a level $k+1$ node v. For any $s \in [0, 1]$, we have from the Chernoff bound,

$$\begin{aligned}
&\frac{1}{l_n(v)} \log \mathbb{P}_0\Big(\frac{S_v}{l_n(v)} > t_{k+1}\Big) \\
&\leq -st_{k+1} + \frac{1}{l_n(v)} \log \mathbb{E}_0[\exp(sS_v)] \\
&= -st_{k+1} + \frac{1}{l_n(v)} \sum_{u \in C_n(v)} \log \mathbb{E}_0[\exp(s\bar{L}_u)]
\end{aligned}$$

$$
\begin{aligned}
&\leq -st_{k+1} + \\
&\quad \frac{1}{l_n(v)} \sum_{u \in C_n(v)} \log \left\{ \mathbb{P}_1(Z_u = 0)^s + \mathbb{P}_0(Z_u = 1)^{1-s} \right\} \\
&\leq -st_{k+1} + \frac{1}{l_n(v)} \sum_{u \in C_n(v)} \max \left\{ s \log \mathbb{P}_1(Z_u = 0), \right. \\
&\qquad \left. (1-s) \log \mathbb{P}_0(Z_u = 1) \right\} + \frac{|C_n(v)|}{l_n(v)} \log 2 \\
&\leq -st_{k+1} + \max\{st_k, -(1-s)\epsilon_k\} + \frac{2|C_n(v)|}{l_n(v)} \log 2. \qquad (5.13)
\end{aligned}
$$

The last inequality follows because

$$
\begin{aligned}
s \log \mathbb{P}_1(Z_u = 0) &\leq s \log(\mathbb{P}_1(Y_u = 0) + \eta) \\
&\leq s \log \left(e^{l_n(u) t_k} + e^{l_n(u)(g_p^* + \epsilon)} \right) \\
&\leq l_n(u) s t_k + \log 2.
\end{aligned}
$$

(In the penultimate inequality, we used (5.11), and the assumption on the decay rate of η; in the last inequality, we used the fact $g_P^* + \epsilon \leq t_k$.) Similarly, using the induction hypothesis instead of (5.11), we have

$$
(1-s) \log \mathbb{P}_0(Z_u = 1) \leq -l_n(u)(1-s)\epsilon_k + \log 2,
$$

hence inequality (5.13) holds. We choose s in the R.H.S. of (5.13) so that $st_k = -(1-s)\epsilon_k$. Note that $t_k < 0$ and $\epsilon_k > 0$, which together guarantee that $0 < s < 1$. Recall that every non-leaf node is assumed to have degree at least c_n, which grows to infinity. Thus, for n sufficiently large, (5.13) implies that

$$
\begin{aligned}
\frac{1}{l_n(v)} \log \mathbb{P}_0(Y_v = 1) &\leq -s\epsilon + \frac{2|C_n(v)|}{l_n(v)} \log 2 \\
&\leq -s\epsilon + \frac{2}{c_n} \log 2 \\
&\leq -s\epsilon/2 := -\epsilon_{k+1},
\end{aligned}
$$

hence (5.12) holds for level $k+1$ nodes. The induction is now complete.

To complete the proof of the proposition, since $l_n(f)/n \to 1$ as $n \to \infty$ (cf. Lemma 3.3), (5.11) yields

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta^* = \limsup_{n\to\infty} \frac{1}{l_n(f)} \log \beta^* \leq g_p^* + h\epsilon,$$

while (5.12) ensures that the Type I error probability is less than α for n sufficiently large. Finally, the optimal error exponent is obtained by letting ϵ go to 0, and the proposition is proved. □

Using a similar argument as in the proof of the above proposition, it can be shown that the condition

$$\limsup_{n\to\infty} \frac{1}{l_M} \log \eta < 0, \tag{5.14}$$

is sufficient for a tree network of height $h \geq 2$ to achieve a Type II error probability that decays exponentially fast with n, although the error exponent can be worse (less negative) than g_P^*.

5.3.5 Energy Efficiency Comparison

In this subsection, we consider $l_n(f)$ nodes arranged on a grid, with neighboring nodes unit distance apart. The $l_n(f)$ nodes are the leaves of our network, but we are otherwise free to configure the network, and to possibly introduce additional nodes that will serve as message relays. We will compare the energy consumption of a parallel configuration with that of a tree network of height $h \geq 2$, under the assumptions of Proposition 5.4. In both cases, the fusion center is placed at the center of the entire grid.

To construct a tree of height h, we add new nodes at levels $1, \ldots, h-1$, as follows (see Figure 5-1). Let r be a positive integer which is a perfect square. Partition the grid of nodes into r equal sub-squares, each of which is called a level $h-1$ sub-square. At the center of each sub-square, we place a new node, which serves as a level $h-1$ node. Next, partition each level $h-1$ sub-square into r further sub-squares, and

place a new node at the center of each of the latter sub-squares. These are the level $h-2$ nodes, which send their messages to the level $h-1$ node of that sub-square. We repeat this procedure $h-1$ times. Finally, all the leaves in a level 1 sub-square send their messages to the level 1 node in that sub-square.

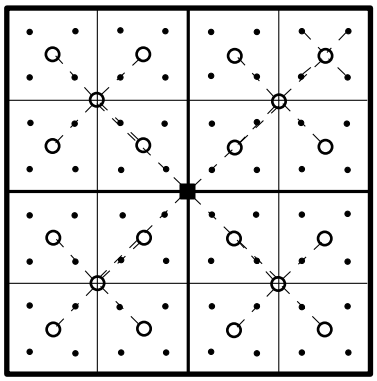

Figure 5-1: A tree network of height 3, with $r = 4$. The circles represent the new nodes that we have added. The dotted lines indicate communication links. Only one level 1 sub-square (the top right one) is shown with all its communication links.

The total number of nodes is $n = l_n(f) + (r^h - 1)/(r-1)$. As we consider progressively larger values of $l_n(f)$, we adjust the value of r used in the above construction, so that $l_n(f)/r^{h-1} \to \infty$, and $l_n(f)/r^h \to 0$, as $n \to \infty$. We compare the performance and energy consumption of this tree network with that of a parallel configuration in which all $l_n(f)$ nodes send their messages directly to the fusion center. (Since $l_n(f)/n \to 1$ as $n \to \infty$, the results would also be the same for a parallel configuration with n, instead of $l_n(f)$, nodes.)

In the tree network ($h \geq 2$) that we have constructed, the condition

$$\limsup_{n\to\infty} \frac{1}{l_M} \log \eta \leq g_P^*$$

is not only sufficient, but also necessary for the Type II error exponent to be g_P^*. To see this, suppose that $Z_1, \ldots, Z_r$ are messages received at the fusion center. For the Type I error constraint to be satisfied, there exists $(z_1, \ldots, z_r)$ such that $\gamma_f(z_1, \ldots, z_r) = 0$. Moreover, for any $Y_1, \ldots, Y_r \in \{0, 1\}$, we have $\mathbb{P}_1(Z_1 = z_1, \ldots, Z_r = z_r \mid Y_1, \ldots, Y_r) \geq$

η^r. Therefore, we obtain

$$\begin{aligned}\beta^* &\geq \mathbb{P}_1(Z_1 = z_1, \ldots, Z_r = z_r) \\ &= \mathbb{E}_1[\mathbb{P}_1(Z_1 = z_1, \ldots, Z_r = z_r \mid Y_1, \ldots, Y_r)] \geq \eta^r.\end{aligned}$$

Hence, if $\limsup_{n\to\infty}(1/l_M)\log\eta > g_P^*$, we would have $\lim_{n\to\infty}(1/l_n(f))\log\beta^* > g_P^*$, since $l_M = l_n(f)/r$.

We assume that each node employs antipodal signalling, and the received signal is corrupted by additive white Gaussian noise with variance $N_0/2$: a node receives a $N(\sqrt{E_b}, N_0/2)$ random variable if a 1 is sent by its immediate predecessor, and a $N(-\sqrt{E_b}, N_0/2)$ random variable if a 0 is sent. The recipient node performs a maximum a posterior probability test to determine if a 1 or 0 was sent. The resulting channel error probability is

$$\eta = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \approx \frac{1}{2}\sqrt{\frac{N_0}{E_b\pi}}e^{-E_b/N_0},$$

where $Q(\cdot)$ is the Gaussian complementary error function. To satisfy the conditions in Proposition 5.4, we choose E_b as follows:

1. if $h = 1$, let $E_b = E(n)$, where $E(n) \to \infty$ as $n \to \infty$;

2. if $h \geq 2$, let $E_b = c \cdot l_n(f)/r$, where $c \geq -g_P^*/N_0$ is a constant.

We also assume a path-loss model, so that the received bit energy at each receiver node is $E_b = E_0/D^a$, where D is the transmission distance, a is the path-loss exponent, and E_0 is the transmission energy expended by the transmitting node. Therefore, the transmission energy of a node is $E_0 = E_b D^a$. In line with standard wireless communications models [49], we take $2 \leq a \leq 4$.

Let E_{CT} be the circuit processing energy required by each node, and E_{CR} be the receiver circuit energy incurred by a receiver node per message received [48]. The total energy E_P expended by a parallel configuration is given below. The first term is the receiver circuit energy of the fusion center, the second term is the processing

energy expended by all the nodes, the third term is the total transmission energy, and $c(n)$ is the average path-loss D^a suffered by the nodes. We have

$$E_P = l_n(f)E_{CR} + (l_n(f)+1)E_{CT} + l_n(f)E(n)c(n).$$

Since more than half of the nodes are at distance at least $\sqrt{l_n(f)}/4$ from the fusion center, we obtain

$$\begin{aligned} E_P &\geq l_n(f)(E_{CR}+E_{CT}) + l_n(f)E(n)\cdot\frac{1}{2}\Big(\frac{\sqrt{l_n(f)}}{4}\Big)^a \\ &= \Omega\big(n^{1+a/2}E(n)\big). \end{aligned}$$

For the tree network with height $h \geq 2$, we have the following upper bound on the total energy consumption E_T. The first term is the total processing energy of all the nodes, the second term is the receiver circuit energy expended by nodes from level 1 to level h, the third term is an upper bound on the transmission energy expended by nodes from level 1 to level $h-1$, and the last term is an upper bound on the transmission energy expended by the leaves. So, we have

$$\begin{aligned} E_T &\leq nE_{CT} + (n-1)E_{CR} + E_b\sum_{k=1}^{h-1} r^k\left(\frac{\sqrt{l_n(f)}}{r^{\frac{k-1}{2}}}\right)^a + l_n(f)E_b\left(\frac{\sqrt{l_n(f)}}{r^{\frac{h-1}{2}}}\right)^a \\ &\leq n(E_{CT}+E_{CR}) + c\frac{l_n(f)}{r}\sum_{k=1}^{h-1} r^k\cdot\frac{n^{a/2}}{r^{k-1}} + l_n(f)\cdot c\frac{l_n(f)}{r}\cdot\frac{n^{a/2}}{r^{h-1}} \\ &\leq n(E_{CT}+E_{CR}) + c(h-1)n^{1+a/2} + cn^{1+a/2}\frac{l_n(f)}{r^h} \\ &= O(n^{1+a/2}). \end{aligned}$$

The above analysis shows that for large n, $E_T < E_P$. Hence, the tree network consumes less energy than the parallel configuration, if both networks are designed to have the same error exponent g_P^*.

5.4 The Bayesian Problem

In this section, we consider the Bayesian formulation of the problems analyzed in Sections 5.2 and 5.3, under some additional simplifying assumptions.

Suppose that we are given positive prior probabilities π_0 and π_1 for each hypothesis. Let $P_e = \pi_0 \mathbb{P}_0(Y_f = 1) + \pi_1 \mathbb{P}_1(Y_f = 0)$ be the probability of error at the fusion center, and let P_e^* be the minimum probability of error, where the minimization is over all strategies. We assume that the fusion center always uses the optimal fusion rule, namely the maximum a posteriori probability rule. In this section, we assume that all nodes are constrained to sending 1-bit messages. We also make the following assumption on the observations at the leaves.

Assumption 5.2. *Either one of the following holds:*

(i) The observations X_i at the leaves take values in a finite set.

(ii) Assumption 2.3 and the condition $\mathbb{E}_1\big[\log^2\big(\mathrm{d}\mathbb{P}_1^X/\mathrm{d}\mathbb{P}_0^X\big)\big] < \infty$ hold.

For each $\gamma \in \Gamma$, let

$$\Lambda(\gamma) = \min_{s\in[0,1]} \log \mathbb{E}_0\big[\big(\frac{\mathrm{d}\mathbb{P}_1^\gamma}{\mathrm{d}\mathbb{P}_0^\gamma}\big)^s\big].$$

Under Assumptions 2.2 and 5.2, it is known that the optimal error exponent for a parallel configuration with a deterministic number of nodes is given by

$$\Lambda^* = \inf_{\gamma\in\Gamma} \Lambda(\gamma).$$

According to Propositions 2 and 3 of [22], Assumptions 2.1-2.2 and 5.2 imply the following lemma.

Lemma 5.6. *Suppose that Assumptions 2.1-2.2, and 5.2 hold. Then, for any choice of transmission functions $\gamma_1, \ldots, \gamma_n$ used by the n leaves in a parallel configuration, the resulting probability of error, $P_e(n)$, assuming that all transmissions are reliable,*

satisfies

$$P_e(n) = \exp\Big\{ \sum_{i=1}^{n} \Lambda(\gamma_i) + f(n) \Big\},$$

where $f(n)$ *is a function such that* $\lim_{n\to\infty} f(n)/n = 0$.

In the next two subsections, we consider separately the cases of node failures and unreliable communications, in the Bayesian framework.

5.4.1 Node Failures

For tractability, we consider only the case where for all $k \geq 1$, μ_k is the Poisson distribution with mean λ_k. We have the following proposition, which yields the optimal error exponent in the presence of node failures. Unlike the Neyman-Pearson case, where the Type II error probability decays exponentially fast with the expected number, $\lambda(h)$ of leaves, the Bayesian error probability decays exponentially with λ_h.

Proposition 5.5. *Suppose that Assumptions 2.1-2.2, and 5.2 hold.*

(a) If $h = 1$*, the optimal error exponent is given by*

$$\lim_{\lambda_1\to\infty} \frac{1}{\lambda_1} \log P_e^* = -1 + e^{-\Lambda^*}.$$

(b) If $h \geq 2$*, the optimal error exponent is given by*

$$\lim_{\lambda^*\to\infty} \frac{1}{\lambda_h} \log P_e^* = -1.$$

Furthermore, the optimal error exponent remains unchanged if we restrict all leaves to use the same transmission function $\gamma \in \Gamma$*, and all other nodes to use a majority rule.*

Proof. (a) Suppose that $h = 1$. For every n, we have from Lemma 5.6, $P_e(n) \geq \exp\{n\Lambda^* + f(n)\}$. Furthermore, $P_e = \mathbb{E}[P_e(N)]$, where N has a Poisson distribution with mean λ_1. Fix some $\epsilon > 0$. Let n_0 be such that $|f(n)| \leq n\epsilon$, for every $n > n_0$.

Let $m = \sup_{1 \le n \le n_0}\{|f(n)|\}$, and notice that $|f(n)| \le m + n\epsilon$. We have

$$\begin{aligned} P_e &\ge \sum_{n=0}^{\infty} e^{-\lambda_1} \frac{\lambda_1^n}{n!} e^{n(\Lambda^* - \epsilon) - m} \\ &= e^{-m-\lambda_1} \sum_{n=0}^{\infty} \frac{1}{n!} \Big(\lambda_1 e^{\Lambda^* - \epsilon}\Big)^n \\ &= \exp\{\lambda_1 (e^{\Lambda^* - \epsilon} - 1) - m\} \end{aligned}$$

Therefore,

$$\liminf_{\lambda_1 \to \infty} \frac{1}{\lambda_1} \log P_e^* \ge -1 + e^{\Lambda^* - \epsilon}.$$

Since ϵ was arbitrary, it follows that

$$\liminf_{\lambda_1 \to \infty} \frac{1}{\lambda_1} \log P_e^* \ge -1 + e^{\Lambda^*}.$$

For a corresponding upper bound, let all leaves use a transmission function γ^* such that $\Lambda(\gamma^*) \le \Lambda^* + \epsilon$. We then have

$$\begin{aligned} P_e^* &\le \sum_{n=0}^{\infty} e^{-\lambda_1} \frac{\lambda_1^n}{n!} e^{n(\Lambda^* + \epsilon) + f(n)} \\ &\le \sum_{n=0}^{\infty} e^{-\lambda_1} \frac{\lambda_1^n}{n!} e^{n(\Lambda^* + 2\epsilon) + m} \\ &= \exp\{\lambda_1 (e^{\Lambda^* + 2\epsilon} - 1) + m\}. \end{aligned}$$

We take logarithms, divide by λ_1, and take the limit as $\lambda_1 \to \infty$. Using also the fact that ϵ was arbitrary, we obtain that

$$\limsup_{\lambda_1 \to \infty} \frac{1}{\lambda_1} \log P_e^* \le -1 + e^{\Lambda^*}.$$

(b) (Outline) Suppose now that $h \ge 2$. Notice that there is a probability $e^{-\lambda_h}$ that the fusion center has no predecessors, and a further probability of $\min\{\pi_0, \pi_1\}$ of making an error, so that $P_e^* \ge \min\{\pi_0, \pi_1\} e^{-\lambda_h}$. It follows that $\limsup_{\lambda^* \to \infty} (1/\lambda_h) \log P_e^* \ge -1$.

For a corresponding upper bound, consider the case where all leaves use the same transmission function, and all other nodes use a majority rule. An easy induction argument shows that for every immediate predecessor of the fusion center, $\mathbb{P}_0(Y_u = 1)$ and $\mathbb{P}_1(Y_u = 0)$ can be brought arbitrarily close to zero, as $\lambda^* \to \infty$. This brings us to a situation similar to the one considered in part (a), except that now Λ^* can be replaced by an arbitrarily negative constant. A calculation similar to the one in part (a) yields $\limsup_{\lambda_h \to \infty} (1/\lambda_1) \log P_e^* \leq -1$. □

5.4.2 Unreliable Communications

In the case of unreliable communications, the corresponding results are obtained easily.

Proposition 5.6. *Suppose that Assumptions 2.1-2.2, and 5.2 hold.*

(i) If $h = 1$, it is optimal to have all leaves use the same transmission function, and the optimal error exponent is given by

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{|C_n(f)|} \log P_e^* \\ &= \inf_{\gamma \in \Gamma} \min_{s \in [0,1]} \log \Big(\sum_{z=0}^{1} (q_0^\gamma(z))^{1-s} (q_1^\gamma(z))^s \Big). \end{aligned}$$

(ii) For $h \geq 2$, it is optimal to have all leaves use the same transmission function γ, where γ is chosen so that $\Lambda(\gamma) < 0$, and to have all intermediate nodes use a majority rule. Furthermore, the optimal error exponent is given by

$$\lim_{n\to\infty} \frac{1}{|C_n(f)|} \log P_e^* = \frac{1}{2} \log \big(4\eta(1-\eta) \big).$$

Proof. (Outline) Part (i) follows from Theorem 1 of [22]. As for part (ii), an argument similar to the proof of Proposition 5.3 shows that the probability of error at each intermediate node converges to zero, so that the messages received by the fusion center have asymptotic distributions Bern(η) or Bern($1-\eta$), under H_0 or H_1, respectively. The final result then follows immediately from Chernoff's bound [39]. □

Chapter 6

Performance of Tandem Networks

In the previous chapters, we have focused our attention on networks that have bounded height. We now shift our attention to networks whose heights grow with the number of nodes. Specifically, we consider the special case of a tandem or serial network, in which the height of the network grows linearly with the number of nodes. We study the rate of error probability decay and show that it is always sub-exponential, establishing the validity of a long-standing conjecture.

6.1 Background and Related Work

Consider a tandem network, as shown in Figure 6-1, with n nodes. We assume Assumption 2.1, i.e., each node i makes an i.i.d. observation X_i under either hypothesis. We also assume that each node i is constrained to sending a 1-bit message Y_i to sensor $i+1$. Let $\pi_j > 0$ be the prior probability of hypothesis H_j, and let $P_e(n) = \pi_0 \mathbb{P}_0(Y_n = 1) + \pi_1 \mathbb{P}_1(Y_n = 0)$ be the probability of error at node n, under some particular strategy. The goal of a system designer is to design a strategy so that the probability of error $P_e(n)$ is minimized. Let $P_e^*(n) = \inf P_e(n)$, where the infimum is taken over all possible strategies.

The problem of finding optimal strategies has been studied in [14, 16, 26], while the asymptotic performance of a long tandem network (i.e., $n \to \infty$) is considered in [23–26, 50, 51] (some of these works do not restrict the message sent by each node

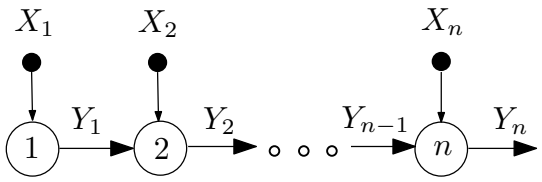

Figure 6-1: A tandem network.

to be binary). In the case of binary communications, [24, 25] find necessary and sufficient conditions under which the error probability goes to zero in the limit of large n. To be specific, recall that under the basic model, $\mathbb{P}_j^X$ is the distribution of a sensor observation under hypothesis H_j. Then, the error probability stays bounded away from zero iff there exists a $B < \infty$ such that $|\log(\mathrm{d}\mathbb{P}_1^X/\mathrm{d}\mathbb{P}_0^X)| \leq B$ almost surely. When the log-likelihood ratio is unbounded, numerical examples have indicated that the error probability goes to zero much slower than exponentially. This is to be contrasted with the case of a parallel configuration where the error probability decays exponentially fast with the number of nodes n [22]. This suggests that a tandem configuration performs worse than a parallel configuration, when n is large. It has been conjectured in [6, 8, 25, 26] that indeed, the rate of decay of the error probability is sub-exponential. However, a proof is not available. The goal of this chapter is to prove this conjecture.

We first note that there is a caveat to the sub-exponential decay conjecture: the probability measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ need to be equivalent, i.e., absolutely continuous w.r.t. each other (cf. first part of Assumption 2.2). Indeed, if there exists a measurable set A such that $\mathbb{P}_0^X(A) > 0$ and $\mathbb{P}_1^X(A) = 0$, then an exponential decay rate can be achieved as follows: each node always declares 1 until some node m observes a $X_m \in A$, whereupon all nodes $i \geq m$ declare 0. For this reason, we assume throughout that the measures $\mathbb{P}_0^X$ and $\mathbb{P}_1^X$ are equivalent. Under this assumption, we show that

$$\lim_{n\to\infty} \frac{1}{n} \log P_e^*(n) = 0.$$

When the error probability goes to zero, we would also like to quantify the best

possible (sub-exponential) decay rate. In this spirit, we find lower bounds on the probability of error, under the further assumption of bounded KL divergences. In particular, we show that for any $d > 1/2$, and some positive constant c, the error probability is $\Omega(e^{-cn^d})$. Under some further mild assumptions, which are valid in most practical cases of interest, we establish the bound $\Omega(e^{-c(\log n)^d})$ for all $d > 1$, and show that it is tight.

6.2 Sub-exponential Decay

In this section we show that the rate of decay of the error probability is always sub-exponential. Although the proof is simple, we have not been able to find it in the literature. Instead, all works on this topic, to our best knowledge, have only conjectured that the decay is sub-exponential, with numerical examples as supporting evidence [8, 25, 26].

We first state an elementary fact that we will make use of throughout this chapter.

Lemma 6.1. *Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are two equivalent probability measures. If $A_1, A_2, \ldots$ is a sequence of measurable events such that $\mathbb{P}(A_n) \to 0$, as $n \to \infty$, then $\mathbb{Q}(A_n) \to 0$, as $n \to \infty$.*

Proof. For $m > 0$, let $R = \mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$, and $B_m = \{R \leq m\}$. We have

$$\mathbb{P}(B_m^c) = \int_{\{R>m\}} \frac{1}{R} \, \mathrm{d}\mathbb{Q} \leq \frac{1}{m},$$

which implies that

$$\mathbb{P}(R = \infty) = \lim_{m\to\infty} \mathbb{P}(B_m^c) = 0.$$

Since $\mathbb{P}$ and $\mathbb{Q}$ are equivalent measures, we have $\mathbb{Q}(R = \infty) = 0$. For all $m > 0$, and $n \geq 1$, we have

$$\mathbb{Q}(A_n) \leq \mathbb{Q}(A_n \cap B_m) + \mathbb{Q}(B_m^c)$$

$$\leq m\mathbb{P}(A_n \cap B_m) + \mathbb{Q}(B_m^c)$$
$$\leq m\mathbb{P}(A_n) + \mathbb{Q}(B_m^c).$$

Taking $n \to \infty$, and then $m \to \infty$, we obtain the desired conclusion by noting that $\mathbb{Q}(B_m^c) \to \mathbb{Q}(R = \infty) = 0$, as $m \to \infty$. □

Let $L_i = \log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}(X_i)$ be the log-likelihood ratio associated with the observation made by node i. From [5, 6, 14, 25], there is no loss in optimality if we require each sensor to form its messages by using a LLRQ, i.e., a rule of the form

$$Y_i = \begin{cases} 0, & \text{if } L_i \leq t_{i,n}(y), \\ 1, & \text{otherwise,} \end{cases} \tag{6.1}$$

where $t_{i,n}(y)$ is a threshold whose value depends on the message $Y_{i-1} = y$ received by node i. In the sequel, we will assume, without loss of optimality, that all nodes use a LLRQ. The next lemma follows easily from the existence results in [5], and Proposition 4.2 in [6].

Lemma 6.2. *There exists an optimal strategy under which each node uses a LLRQ, with thresholds that satisfy* $t_{i,n}(1) \leq t_{i,n}(0)$ *for all* $i = 1, \ldots, n$.

Proof. Fix the number of nodes n. As already noted, there is no loss in optimality if we require each node to form its messages by using a LLRQ. From this, it is easily shown that for all $i = 1, \ldots, n$, $\mathbb{P}_1(Y_i = y)/\mathbb{P}_0(Y_i = y)$ is nondecreasing in $y \in \{0, 1\}$.

Consider node i, where $i \geq 2$, and suppose that $Y_{i-1} = y \in \{0, 1\}$. Since node i uses a LLRQ, it chooses its message by comparing

$$L_i + \log \frac{\mathbb{P}_1(Y_{i-1} = y)}{\mathbb{P}_0(Y_{i-1} = y)}$$

to a threshold t. Comparing with (6.1), we have

$$t_{i,n}(y) = t - \log \frac{\mathbb{P}_1(Y_{i-1} = y)}{\mathbb{P}_0(Y_{i-1} = y)}.$$

Since $\mathbb{P}_1(Y_{i-1}=y)/\mathbb{P}_0(Y_{i-1}=y)$ is nondecreasing in y, we have $t_{i,n}(1) \le t_{i,n}(0)$. $\square$

In view of Lemma 6.2, we can restrict to strategies of the form

$$\gamma_i(X_i, Y_{i-1}) = \begin{cases} 0, & \text{if } L_i \le t_{i,n}(1), \\ 1, & \text{if } L_i > t_{i,n}(0), \\ Y_{i-1}, & \text{otherwise,} \end{cases}$$

where $t_{i,n}(1) \le t_{i,n}(0)$. Note that this is the type of strategies used in [24] to show that the error probability converges to zero.

Proposition 6.1. *Suppose that Assumption 2.1, and the first part of Assumption 2.2 hold. Then, the rate of decay of the error probability in a tandem network is sub-exponential, i.e.,*

$$\lim_{n\to\infty} \frac{1}{n} \log P_e^*(n) = 0.$$

Proof. Suppose that $P_e^*(n) \to 0$ as $n \to \infty$, else the proposition is trivially true. Fix some n and consider an optimal strategy for the tandem network of length n. We have, for all i,

$$\mathbb{P}_0(Y_i=1) = \mathbb{P}_0\big(L_i > t_{i,n}(0)\big) \cdot \mathbb{P}_0(Y_{i-1}=0) + \mathbb{P}_0\big(L_i > t_{i,n}(1)\big) \cdot \mathbb{P}_0(Y_{i-1}=1), \quad (6.2)$$

$$\mathbb{P}_1(Y_i=0) = \mathbb{P}_1\big(L_i \le t_{i,n}(0)\big) \cdot \mathbb{P}_1(Y_{i-1}=0) + \mathbb{P}_1\big(L_i \le t_{i,n}(1)\big) \cdot \mathbb{P}_1(Y_{i-1}=1). \quad (6.3)$$

From (6.2) and (6.3), with $i=n$, and applying Lemma 6.2, we have

$$\begin{aligned} P_e^*(n) &= \pi_0 \mathbb{P}_0(Y_n=1) + \pi_1 \mathbb{P}_1(Y_n=0) \\ &= \pi_0 \Big(\mathbb{P}_0\big(L_n > t_{n,n}(0)\big) + \mathbb{P}_0\big(t_{n,n}(1) < L_n \le t_{n,n}(0)\big) \cdot \mathbb{P}_0(Y_{n-1}=1) \Big) \\ &\quad + \pi_1 \Big(\mathbb{P}_1\big(L_n \le t_{n,n}(1)\big) + \mathbb{P}_1\big(t_{n,n}(1) < L_n \le t_{n,n}(0)\big) \cdot \mathbb{P}_1(Y_{n-1}=0) \Big) \quad (6.4) \\ &\ge \min_{j=0,1} \mathbb{P}_j\big(t_{n,n}(1) < L_n \le t_{n,n}(0)\big) \cdot P_e^*(n-1). \quad (6.5) \end{aligned}$$

From (6.4), in order to have $P_e^*(n) \to 0$ as $n \to \infty$, we must have $\mathbb{P}_0(L_n > t_{n,n}(0)) \to 0$

and $\mathbb{P}_1(L_n \leq t_{n,n}(1)) \to 0$, as $n \to \infty$. Because $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent measures, from Lemma 6.1, we have $\mathbb{P}_1(L_n > t_{n,n}(0)) \to 0$ and $\mathbb{P}_0(L_n \leq t_{n,n}(1)) \to 0$, as $n \to \infty$. Hence, $\mathbb{P}_j(t_{n,n}(1) < L_n \leq t_{n,n}(0)) \to 1$ for $j = 0, 1$. Therefore, from (6.5), the error probability cannot decay exponentially fast. □

We have established that the decay of the error probability is sub-exponential. This confirms that the parallel configuration performs much better than the tandem configuration when n is large. It now remains to investigate the best performance that a tandem configuration can possibly achieve. In the next section, we use a more elaborate technique to derive a lower bound for the error probability.

6.3 Rate of Decay

In this section, we show that under the assumption of bounded KL divergences, the error probability is $\Omega(e^{-cn^d})$, for some positive constant c and for all $d > 1/2$. Under some additional assumptions, the lower bound is improved to $\Omega(e^{-c(\log n)^d})$, for any $d > 1$. The ideas in this section are inspired by the methods in [14] and Chapter 4. In particular, we rely on a sequence of comparisons of the tandem configuration with other tree configurations, whose performance can be quantified using the methods of Chapters 3 and 4.

Our results involve the KL divergences, which for convenience, we denote as

$$D_0 = \mathbb{E}_0\big[\log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big],$$
$$D_1 = \mathbb{E}_1\big[\log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big].$$

We assume that $-\infty < D_0 < 0 < D_1 < \infty$, throughout this section.

Let k and m be positive integers, and let $n = km$. Let us compare the following two networks: (i) a tandem network, as in Figure 6-1, with n nodes, where each node i obtains a single observation X_i; (ii) a modified tandem network $T(k, m)$, as in Figure 6-2, with k nodes, where each node v_i obtains m (conditionally) independent observations $X_{(i-1)m+1}, \ldots, X_{im}$, given either hypothesis. In both networks a node

sends a binary message to its successor. It should be clear that when we keep the total number of observations $n = km$ the same in both networks, the network $T(k, m)$ can perform at least as well as the original one. Indeed, each node v_i in the modified network can emulate the behavior of m nodes in tandem in the original network.

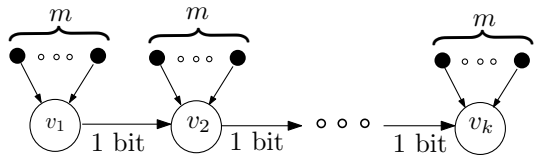

Figure 6-2: A modified tandem network $T(k, m)$ that outperforms a tandem network with $n = km$ nodes.

Therefore, it suffices to establish a lower bound for the error probability in the network $T(k, m)$. Towards this goal, we will use some standard results in Large Deviations Theory, notably Cramér's Theorem [39], in the lemma below.

Lemma 6.3. *Suppose that Assumption 2.1, and the first part of Assumption 2.2 hold. Suppose also that* $-\infty < D_0 < 0 < D_1 < \infty$. *Let* $S_m = \sum_{i=1}^m L_i$, *and for* $j = 0, 1$, *let* $\Lambda_j^*(t) = \sup_{\xi \in \mathbb{R}} \{\xi t - \log \mathbb{E}_j \big[\big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X} \big)^\xi \big] \}$.

(i) For every $\epsilon > 0$, *there exist* $a \in (0, 1)$, $c > 0$, *and* $M \geq 1$, *such that for all* $m \geq M$,

$$\mathbb{P}_0(S_m/m > D_1 + \epsilon) \geq a e^{-mc},$$
$$\mathbb{P}_1(S_m/m \leq D_0 - \epsilon) \geq a e^{-mc}.$$

(ii) Suppose that $\mathbb{E}_1 \big[\big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X} \big)^s \big] < \infty$ *for some* $s > 0$. *Then, there exists some* $\epsilon > 0$, *such that* $\Lambda_1^*(D_1 + \epsilon) > 0$, *and*

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1 - e^{-m\Lambda_1^*(D_1+\epsilon)}, \qquad \forall\ m \geq 1.$$

(iii) Suppose that $\mathbb{E}_0 \big[\big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X} \big)^s \big] < \infty$ *for some* $s < 0$. *Then, there exists some* $\epsilon > 0$,

such that $\Lambda_0^*(D_0 - \epsilon) > 0$, *and*

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1 - e^{-m\Lambda_0^*(D_0-\epsilon)}, \qquad \forall\ m \geq 1.$$

(iv) For every $\epsilon > 0$, *there exists some* $M \geq 1$ *such that*

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1/2, \qquad \forall\ m \geq M.$$

Moreover, if for some integer $r \geq 2$, $\mathbb{E}_1\big[\big|\log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big|^r\big] < \infty$, *then there exists some* $c_r > 0$ *such that*

$$\mathbb{P}_1(S_m/m \leq D_1 + \epsilon) \geq 1 - \frac{c_r}{m^{r/2}\epsilon^r}, \qquad \forall\ m \geq 1.$$

(v) For every $\epsilon > 0$, *there exists some* $M \geq 1$ *such that*

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1/2, \qquad \forall\ m \geq M.$$

Moreover, if for some integer $r \geq 2$, $\mathbb{E}_0\big[\big|\log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big|^r\big] < \infty$, *then there exists some* $c_r > 0$ *such that*

$$\mathbb{P}_0(S_m/m > D_0 - \epsilon) \geq 1 - \frac{c_r}{m^{r/2}\epsilon^r}, \qquad \forall\ m \geq 1.$$

Proof. Note that part (iii) is essentially a restatement of part (ii), with a different measure. A similar remark applies for (iv) and (v).

Part (i) follows directly from Cramér's Theorem (see Theorem 2.1). To show part (ii), we note that $\varphi(\xi) = \log \mathbb{E}_1\big[\big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big)^\xi\big]$ is a convex function of ξ, with $\varphi(-1) = \varphi(0) = 0$, and $\varphi'(0) = D_1$. Therefore, $\varphi(\xi)$ is a nondecreasing function for $\xi \in [0, s]$, and $\varphi(s)/s \geq D_1$. Choose $t > \varphi(s)/s$, then we have $\Lambda_1^*(t) \geq st - \varphi(s) > 0$, i.e., $\Lambda_1^*(D_i + \epsilon) > 0$, where $\epsilon = t - D_1 > 0$. The probability bound in part (ii) follows from Cramér's Theorem. A similar argument holds for part (iii).

Next, we prove part (iv). The first claim follows from the Weak Law of Large

Numbers (WLLN), applied to the random variables $L_1, L_2, \ldots$. Now, suppose that $\mathbb{E}_1\big[\big|\log\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big|^r\big] < \infty$, for some integer $r \geq 2$. We make use of the following estimate of the moment of S_m/m (see e.g. Lemma 5.3.1 of [52]): there exists a constant $c_r > 0$ such that

$$\mathbb{E}_1\big[\big|S_m/m - D_1\big|^r\big] \leq \frac{c_r}{m^{r/2}}, \qquad \forall\ m \geq 1.$$

From Markov's Inequality, we obtain

$$\begin{aligned}\mathbb{P}_1(S_m/m > D_1 + \epsilon) &\leq \frac{1}{\epsilon^r}\mathbb{E}_1\big[\big|S_m/m - D_1\big|^r\big] \\ &\leq \frac{c_r}{m^{r/2}\epsilon^r}.\end{aligned}$$

A similar argument holds for part (v), and the lemma is proved. □

We now state our main result. Part (ii) of the following proposition is a general lower bound that always holds; part (i) is a stronger lower bound, under an additional assumption. Note that the condition in part (i) implies that $\mathbb{E}_j\big[\big|\log\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big|^r\big] < \infty$ for all r, but the reverse implication is not always true.

Proposition 6.2. *Suppose that Assumption 2.1, and the first part of Assumption 2.2 holds. Suppose also that $-\infty < D_0 < 0 < D_1 < \infty$.*

(i) Suppose that there exists some $\epsilon' > 0$ such that for all $s \in [-\epsilon', 1+\epsilon']$, $\mathbb{E}_0\big[\big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big)^s\big] < \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{(\log n)^d} \log P_e^*(n) = 0,$$

for all $d > 1$.

(ii) For all $d > 1/2$, we have

$$\lim_{n\to\infty} \frac{1}{n^d} \log P_e^*(n) = 0.$$

Furthermore, if for some integer $r \geq 2$, $\mathbb{E}_j\big[\big|\log \frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\big|^r\big] < \infty$ for both $j = 0, 1$, then the above is true for all $d > 1/(2 + r/2)$.

Proof. Let us fix m and k, and an optimal strategy for the modified network $T(k, m)$. Let Y_{v_i} be the 1-bit message sent by node v_i, under that strategy. Let

$$S_{i,m} = \sum_{l=1}^{m} L_{(i-1)m+l}, \tag{6.6}$$

which is the log-likelihood ratio of the observations obtained at node v_i. For the same reasons as in Lemma 6.2, an optimal strategy exists and can be taken to be a LLRQ, of the form

$$Y_{v_i} = \begin{cases} 0, & \text{if } S_{i,m}/m \leq t_{i,m}(y), \\ 1, & \text{otherwise,} \end{cases}$$

where $t_{i,m}(y)$ is a threshold whose value depends on the message y received by node v_i from node v_{i-1}. For the same reasons as in Lemma 6.2, we can assume that the optimal strategy is chosen such that $t_{i,m}(1) \leq t_{i,m}(0)$, for all $m \geq 1$, and for all $i \geq 1$.

Let $q_{0,i} = \mathbb{P}_0(Y_{v_i} = 1)$ and $q_{1,i} = \mathbb{P}_1(Y_{v_i} = 0)$ be the Type I and II error probabilities at node v_i. Suppose that the conditions in part (i) of the proposition hold. Let $\delta = \min\{\Lambda_0^*(D_0 - \epsilon), \Lambda_1^*(D_1 + \epsilon)\}$. From parts (ii)-(iii) of Lemma 6.3, there exists $\epsilon > 0$, such that $\delta > 0$. Let us fix such an ϵ, and let $a \in (0, 1)$, $c > 0$, and $M \geq 1$ be as in Lemma 6.3(i). We first show a lower bound on the Type I and II error probabilities $q_{j,i}$.

Lemma 6.4. *There exists some $\bar{M}$ such that for every $i \geq 1$, and every $m \geq \bar{M}$, either*

$$q_{0,i} \geq \frac{a}{2} e^{-mc} (1 - e^{-m\delta})^i, \tag{6.7}$$

or

$$q_{1,i} \geq \frac{a}{2} e^{-mc} (1 - e^{-m\delta})^i. \tag{6.8}$$

Proof. The proof proceeds by induction on i. When $i = 1$, the result is an immediate consequence of Lemma 6.3(i). Indeed, if the threshold t used by node v_1 satisfies $t \leq D_1$, then $q_{0,1} \geq ae^{-mc}$, and if $t \geq D_0$, then $q_{1,1} \geq ae^{-mc}$.

Assume now that $i > 1$ and that the result holds for $i - 1$. We will show that it also holds for i. Let $S_{i,m}$ be as defined in (6.6). We have for $i > 1$,

$$q_{0,i} = (1 - q_{0,i-1})\mathbb{P}_0(S_{i,m}/m > t_{i,m}(0)) + q_{0,i-1}\mathbb{P}_0(S_{i,m}/m > t_{i,m}(1)), \tag{6.9}$$

$$q_{1,i} = (1 - q_{1,i-1})\mathbb{P}_1(S_{i,m}/m \leq t_{i,m}(1)) + q_{1,i-1}\mathbb{P}_1(S_{i,m}/m \leq t_{i,m}(0)). \tag{6.10}$$

We start by considering the case where $q_{0,i-1} < 1/2$ and $q_{1,i-1} < 1/2$. Suppose that $t_{i,m}(0) \leq D_1 + \epsilon$. From (6.9) and Lemma 6.3(i), we have for all $m \geq M$,

$$\begin{aligned} q_{0,i} &\geq \frac{1}{2}\mathbb{P}_0(S_{i,m}/m > D_1 + \epsilon) \\ &\geq \frac{a}{2}e^{-mc} \\ &\geq \frac{a}{2}e^{-mc}(1 - e^{-m\delta})^i. \end{aligned}$$

Similarly, if $t_{i,m}(1) \geq D_0 - \epsilon$, we have $q_{1,i} \geq ae^{-mc}(1 - e^{-m\delta})^i/2$.

It remains to consider the case where $t_{i,m}(0) > D_1 + \epsilon$ and $t_{i,m}(1) < D_0 - \epsilon$. From (6.9) and Lemma 6.3(iii), we obtain

$$\begin{aligned} q_{0,i} &\geq q_{0,i-1}\mathbb{P}_0(S_{i,m}/m > D_0 - \epsilon) \\ &\geq q_{0,i-1}(1 - e^{-m\delta}). \end{aligned}$$

Similarly, from (6.10) and Lemma 6.3(ii), we have

$$\begin{aligned} q_{1,i} &\geq q_{1,i-1}\mathbb{P}_1(S_{i,m}/m \leq D_1 + \epsilon) \\ &\geq q_{1,i-1}(1 - e^{-m\delta}). \end{aligned}$$

Using the induction hypothesis, either (6.7) or (6.8) holds.

We next consider the case where $q_{0,i-1} \geq 1/2$ and $q_{1,i-1} < 1/2$. If either

a) $t_{i,m}(1) \geq D_0 - \epsilon$, or

b) $t_{i,m}(0) > D_1 + \epsilon$ and $t_{i,m}(1) < D_0 - \epsilon$,

we obtain, via the same argument as above, the desired conclusion. Suppose then that $t_{i,m}(0) \leq D_1 + \epsilon$ and $t_{i,m}(1) < D_0 - \epsilon$. From (6.9) and the WLLN, we have for some $\bar{M}$ sufficiently large, and for all $m \geq \bar{M}$,

$$\begin{aligned} q_{0,i} &\geq \frac{1}{2}\mathbb{P}_0(S_{i,m}/m > t_{i,m}(1)) \\ &\geq \frac{1}{2}\mathbb{P}_0(S_{i,m}/m > D_0 - \epsilon) \geq \frac{1}{4}, \end{aligned}$$

so that the claim holds trivially. The case $q_{0,i-1} < 1/2$ and $q_{1,i-1} \geq 1/2$ is similar.

We finally consider the case where $q_{0,i-1} \geq 1/2$ and $q_{1,i-1} \geq 1/2$. If $t_{i,m}(1) \leq D_1$, then

$$q_{0,i} \geq \frac{1}{2}\mathbb{P}_0(S_{i,m}/m > D_1) \geq \frac{a}{2}e^{-mc}.$$

If on the other hand, $t_{i,m}(1) > D_1$, then $t_{i,m}(0) \geq t_{i,m}(1) > D_1 > D_0$, and

$$q_{1,i} \geq \frac{1}{2}\mathbb{P}_1(S_{i,m}/m \leq D_0) \geq \frac{a}{2}e^{-mc}.$$

This concludes the proof of the lemma. □

We return to the proof of part (i) of Proposition 6.2. Fix some $d > 1$ and some $l \in (1/d, 1)$. Let $k = k(m) = \exp(m^l)$. For a tandem network with n nodes, since $k(m)m = \exp(m^l)m$ is increasing in m, we have $\exp((m-1)^l)(m-1) < n \leq \exp(m^l)m$, for some m. Since the tree network $T(k(m), m)$ outperforms a tandem network with n nodes, we have

$$\begin{aligned} P_e^*(n) &\geq \pi_0 q_{0,k(m)} + \pi_1 q_{1,k(m)} \\ &\geq \min\{\pi_0, \pi_1\}\frac{a}{2}e^{-mc}(1 - e^{-m\delta})^{k(m)}, \end{aligned} \tag{6.11}$$

where the last inequality follows from Lemma 6.4. Note that

$$\begin{aligned}
&\frac{1}{(\log(k(m)m))^d}\log\left(e^{-mc}(1-e^{-m\delta})^{k(m)}\right)\\
&=-\frac{mc}{(m^l+\log m)^d}+\frac{e^{m^l}}{(m^l+\log m)^d}\log\left(1-e^{-m\delta}\right)\\
&=-\frac{mc}{(m^l+\log m)^d}+\frac{e^{m^l-m\delta}}{(m^l+\log m)^d}\log\left(1-e^{-m\delta}\right)^{e^{m\delta}}. \qquad (6.12)
\end{aligned}$$

Since $dl > 1$ and $l < 1$, the R.H.S. of (6.12) converges to 0 as $m \to \infty$. Moreover, since

$$1 \le \frac{\log(k(m)m)}{\log n} \le \frac{m^l+\log m}{(m-1)^l+\log(m-1)} \to 1,$$

as $m \to \infty$, we have from (6.11),

$$\lim_{n\to\infty}\frac{1}{(\log n)^d}\log P_e^*(n) = 0,$$

which proves part (i) of the proposition.

For part (ii), the argument is the same, except that we use parts (iv) and (v) of Lemma 6.3 (instead of parts (ii) and (iii)), and the inequalities (6.7) and (6.8) are replaced by

$$q_{0,i} \ge \frac{a}{2}e^{-mc}\frac{1}{2^i},$$

and

$$q_{1,i} \ge \frac{a}{2}e^{-mc}\frac{1}{2^i},$$

respectively, and we let $k = m^l$ where $l \in (1/d-1, 1)$, for $1/2 < d < 1$. The conclusion when $\mathbb{E}_j\left[\left|\log\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\right|^r\right] < \infty$ for some integer $r \ge 2$ can be derived similarly. □

6.4 Tightness

Part (i) of Proposition 6.2 translates to a bound of the form $\Omega(e^{-c(\log n)^d})$, for every $d > 1$. In this section, we show that this family of bounds is tight, in the sense that it cannot be extended to values of d less than one. This is accomplished by constructing an example in which the error probability is $O(e^{-c(\log n)^d})$, with $d = 1$, i.e., the error probability is of the order $O(n^{-c})$ for some $c > 0$.

Our example involves a Gaussian hypothesis testing problem. We assume that under H_j, X_1 is distributed according to a normal distribution with mean 0 and variance σ_j^2, where $0 < \sigma_0^2 < 1/2 < \sigma_1^2$. We first check that the condition in part (i) of Proposition 6.2 is satisfied. We have

$$\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}(x) = \frac{\sigma_0}{\sigma_1} e^{-\frac{x^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)},$$

and (using the formula for the moment generating function of a χ^2 distribution),

$$\begin{aligned}\mathbb{E}_0\Big[\Big(\frac{\mathrm{d}\mathbb{P}_1^X}{\mathrm{d}\mathbb{P}_0^X}\Big)^s\Big] &= \Big(\frac{\sigma_0}{\sigma_1}\Big)^s \mathbb{E}_0\big[e^{\frac{s}{2}\left(1-\sigma_0^2/\sigma_1^2\right)(X_1/\sigma_0)^2}\big] \\ &= \Big(\frac{\sigma_0}{\sigma_1}\Big)^s \Big(\frac{1}{1 - s\left(1 - \sigma_0^2/\sigma_1^2\right)}\Big)^{1/2} < \infty,\end{aligned}$$

if $s < 1/(1 - \sigma_0^2/\sigma_1^2)$. Hence, the condition in part (i) of Proposition 6.2 is satisfied.

Fix some n and let $a_n = \sqrt{\log n}$. We analyze the rate of decay of error probability of a particular sub-optimal strategy considered in [25], which is the following:

$$\gamma_1(X_1) = \begin{cases} 0, & \text{if } X_1^2 \le a_n^2, \\ 1, & \text{otherwise,} \end{cases}$$

and for $i \ge 2$,

$$\gamma_i(X_i, Y_{i-1}) = \begin{cases} 0, & \text{if } X_i^2 \le a_n^2 \text{ and } Y_{i-1} = 0, \\ 1, & \text{otherwise.} \end{cases}$$

Thus, the decision at node n is $Y_n = 1$ iff we have $X_i^2 > a_n^2$ for some $i \le n$.

Proposition 6.3. *With the above described strategy, the probability of error is $O(n^{-c})$, for some $c > 0$.*

Proof. Let $Q(\cdot)$ be the Gaussian complementary error function, i.e., $Q(x) = \mathbb{P}(Z \geq x)$, where Z is a standard normal random variable. We use the well-known bound $Q(x) \leq \exp(-x^2/2)$ (see, e.g., [46]). The Type I error probability is given by

$$\begin{aligned}
\mathbb{P}_0(Y_n = 1) &= \mathbb{P}_0(X_i^2 > a_n^2 \text{ for some } i) \\
&\leq n\mathbb{P}_0(X_1^2 > a_n^2) \\
&= 2nQ\big(a_n/\sigma_0\big) \\
&\leq 2ne^{-a_n^2/2\sigma_0^2} \\
&= 2n^{1-\frac{1}{2\sigma_0^2}},
\end{aligned}$$

which is of the form $O(n^{-c})$, with $c > 0$.

The Type II error probability is

$$\begin{aligned}
\mathbb{P}_1(Y_n = 0) &= \Big(\mathbb{P}_1(X_1^2 \leq a_n^2)\Big)^n \\
&= \Big(1 - \mathbb{P}_1(X_1^2 > a_n^2)\Big)^n \\
&\leq e^{-n\mathbb{P}_1(X_1^2 > a_n^2)}. \qquad (6.13)
\end{aligned}$$

From the lower bound $Q(x) \geq \frac{1}{x\sqrt{2\pi}}(1 - \frac{1}{x^2})\exp(-x^2/2)$ (see [46]), we have

$$\begin{aligned}
n\mathbb{P}_1(X_1^2 > a_n^2) &= 2nQ\big(a_n/\sigma_1\big) \\
&\geq \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma_1}{a_n}\Big(1 - \frac{\sigma_1^2}{a_n^2}\Big)e^{-a_n^2/2\sigma_1^2}n \\
&= \sqrt{\frac{2}{\pi}} \cdot \frac{\sigma_1}{\sqrt{\log n}}\Big(1 - \frac{\sigma_1^2}{\log n}\Big)n^{1-\frac{1}{2\sigma_1^2}} \\
&= \Omega(n^{d_1}),
\end{aligned}$$

where $d_1 > 0$. From (6.13), we obtain that $\mathbb{P}_1(Y_n = 0) = O(\exp(-n^{d_1}))$. Hence, the overall error probability is dominated by the Type I error probability, and this

strategy achieves a decay rate of n^{-c} for some positive constant c. □

We note that in most cases, the rate n^{-c} is not achievable. For example, consider the more common case of detecting the presence of a known signal in Gaussian noise: under H_0, the distribution of X_1 is normal with mean $-\mu$ and variance 1, while under H_1, the distribution is normal with mean μ and variance 1. A numerical computation indicates that the optimal error probability decay is of the order $\exp(-c\sqrt{\log n})$ (see [26] and Figure 6-3). Finding the exact decay rate analytically for particular pairs of distributions seems to be difficult because there is no closed form solution for the optimal thresholds used in the LLRQ decision rule at each node [25], except for distributions with certain symmetric properties [26].

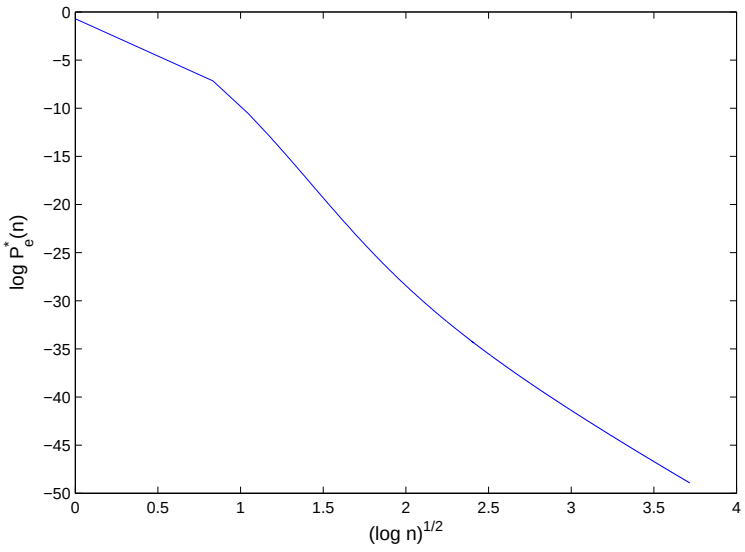

Figure 6-3: A plot of the optimal error probability as a function of the number of nodes, for the problem of detecting the presence of a known signal in Gaussian noise. The optimal thresholds for the LLRQs at each node are given in [26]. For large n, the plot is almost linear.

6.5 The Neyman-Pearson Problem

In this section, we consider a simplified version of the detection problem in a long tandem, under a Neyman-Pearson framework. We will establish that the probability

of Type II error decays sub-exponentially, if we restrict the message sent by each node to be a Neyman-Pearson optimal decision *at that node*.

It is well known that in centralized Neyman-Pearson detection, randomization can reduce the Type II error probability. Accordingly, we assume that each node i also has access to a random variable V_i, independent of the hypothesis or the observations, which acts as the randomization variable. We assume *independent randomization* [6], i.e., that the random variables V_i are independent. Given the received message Y_{i-1}, the randomization variable V_i, and its own observation X_i, each node i chooses Y_i so as to minimize $\mathbb{P}_1(Y_i = 0)$, subject to the constraint $\mathbb{P}_0(Y_i = 1) \leq \alpha$, where $\alpha \in (0, 1)$ is a given threshold. We call such a strategy a *myopic* one.

Let $\beta_n^*(\alpha)$ be the Type II error probability, $\mathbb{P}_1(Y_n = 0)$, for node n, when a myopic strategy is used. It is well known that there is again no loss in optimality if we restrict the nodes to using randomized LLRQs, i.e., each node i uses a rule of the form

$$Y_i = \begin{cases} 0, & \text{if } L_i \leq t_i(Y_{i-1}, V_i), \\ 1, & \text{otherwise,} \end{cases}$$

where the randomized threshold $t_i(Y_{i-1}, V_i)$ depends on both the message Y_{i-1} and the randomization variable V_i. It is also easy to see that it suffices for V_i to take values in a space $\mathcal{V}$ of cardinality two. We finally have $t_i(1, v) \leq t_i(0, v)$ for all i and all $v \in \mathcal{V}$. The proof of this fact is similar to that of Lemma 6.2, and is omitted.

Proposition 6.4. *Suppose that independent randomization is used, and the probability measures $\mathbb{P}_0$ and $\mathbb{P}_1$ are equivalent. Then, for all $\alpha \in (0, 1)$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^*(\alpha) = 0.$$

Proof. It is easily seen that $0 \leq \beta_{n+1}^*(\alpha) \leq \beta_n^*(\alpha)$, and therefore $\beta_n^*(\alpha)$ converges as $n \to \infty$. (To see this, note that node $n + 1$ could just set $Y_{n+1} = Y_n$, thus achieving a probability of error equal to $\beta_n^*(\alpha)$.) If $\lim_{n\to\infty} \beta_n^*(\alpha) > 0$, the result of the proposition is immediate. Therefore, without loss of generality, we assume that $\lim_{n\to\infty} \beta_n^*(\alpha) = 0$.

Suppose that the tandem network uses a myopic strategy. Then, we have $\mathbb{P}_0(Y_n = 1) = \alpha$, for all $n \geq 1$. The recursive relations (6.2)-(6.3) still hold, so we have

$$\mathbb{P}_1(Y_n = 0) = \mathbb{P}_1\big(L_n \leq t_n(0, V_n)\big) \cdot \mathbb{P}_1(Y_{n-1} = 0) + \mathbb{P}_1\big(L_n \leq t_n(1, V_n)\big) \cdot \mathbb{P}_1(Y_{n-1} = 1), \tag{6.14}$$

which implies that

$$\mathbb{P}_1(Y_n = 0) = \mathbb{P}_1\big(L_n \leq t_n(1, V_n)\big) + \mathbb{P}_1\big(t_n(1, V_n) < L_n \leq t_n(0, V_n)\big) \cdot \mathbb{P}_1(Y_{n-1} = 0). \tag{6.15}$$

Since $\mathbb{P}_1(Y_n = 0) \to 0$, as $n \to \infty$, we must have $\mathbb{P}_1(L_n \leq t_n(1, V_n)) \to 0$, as $n \to \infty$. By Lemma 6.1, we obtain $\mathbb{P}_0(L_n \leq t_n(1, V_n)) \to 0$, and $\mathbb{P}_0(L_n > t_n(1, V_n)) \to 1$, as $n \to \infty$.

Using the recursive relation (6.2) for the Type I error, we obtain

$$\begin{aligned} \alpha &= \mathbb{P}_0(Y_n = 1) \\ &= \mathbb{P}_0(L_n > t_n(0, V_n)) \cdot \mathbb{P}_0(Y_{n-1} = 0) + \mathbb{P}_0(L_n > t_n(1, V_n)) \cdot \mathbb{P}_0(Y_{n-1} = 1) \\ &= \mathbb{P}_0(L_n > t_n(0, V_n))(1 - \alpha) + \mathbb{P}_0(L_n > t_n(1, V_n))\alpha. \end{aligned} \tag{6.16}$$

We take the limit of both sides. Since $\mathbb{P}_0(L_n > t_n(1, V_n)) \to 1$, we obtain $\mathbb{P}_0(L_n > t_n(0, V_n))(1 - \alpha) \to 0$. By Lemma 6.1, it follows that $\mathbb{P}_1(L_n > t_n(0, V_n)) \to 0$. Since we also have $\mathbb{P}_1(L_n \leq t_n(1, V_n)) \to 0$, we obtain $\mathbb{P}_1(t_n(1, V_n) < L_n \leq t_n(0, V_n)) \to 1$. From (6.15), it follows that $\mathbb{P}_1(Y_n = 0)$ decays sub-exponentially fast, and the proof is complete. □

Myopic strategies are, in general, suboptimal. If we allow general strategies, the Type II error probability decay rate, can come arbitrarily close to exponential, as illustrated by the example in Section 6.4. Indeed, in that example, we exhibit a (suboptimal) strategy whose Type I error probability converges to zero, and which achieves a Type II error probability $\mathbb{P}_1(Y_n = 0)$ of order $O(\exp(-n^{d_1}))$, where $d_1 \in$

$(0, 1)$. We can choose d_1 to be arbitrarily close to 1 (by choosing a large σ_1 in the example), so that the error probability decay is almost exponential. However, whether the optimal Type II error probability decay rate is guaranteed to be sub-exponential (as is the case for the Bayesian problem) remains an open problem.

Chapter 7

Censoring Sensor Networks

In the previous chapters, we studied the relationship between the architecture of a tree network, and its detection performance. In this chapter, we focus on the parallel configuration, and study the tradeoff between energy efficiency and detection performance.

7.1 Motivation and Overview

In censoring networks, nodes can decide whether or not to make a measurement and transmit that measurement to the fusion center. When sensors are operating independently from each other, [27,28] show that each sensor should base its censoring decision on the likelihood ratio associated with its measurement. In this chapter, we consider the censoring problem in a more general context, and for a large number of nodes. We allow the possibility of sensors having access to some side-information R_i, which can be used to choose between transmission modes. We use the term "side-information" in a very general way to refer to some observable that affects the operation of each sensor. In general, R_i could provide information on the quality of the channel from sensor i to the fusion center, or on the quality of the measurement available at sensor i. The choice of what side-information is available depends on the specific problem and its constraints. We illustrate our framework by presenting two motivating examples, which will be revisited in Section 7.8.

Example 7.1 (Fading Channels)**.** *Consider a large number n of sensors deployed for the purposes of detection, that is, testing between two hypotheses, H_0 and H_1. Each sensor i obtains an independent measurement $\tilde{X}_i$ (with a different distribution under each hypothesis), which it can encode and transmit to a fusion center through a noisy channel. The message received at the fusion center is of the form*

$$Y_i = Q_i \tilde{\gamma}_i(\tilde{X}_i) + W_i,$$

where Q_i is a stochastic fading coefficient, and W_i is zero-mean Gaussian noise with known variance σ_i^2, independent of everything else. In order to conserve power, or to avoid divulging the presence of the sensors, we introduce a constraint that under "normal conditions" (that is, under the null hypothesis H_0), the expected number of transmitting sensors is bounded by nc, where $c \in (0, 1]$ is a given constant. Then, the sensor network is faced with the problem of choosing which sensors should transmit their measurements to the fusion center. Suppose that the network has knowledge of the channel state information Q_i and σ_i, $i = 1, \ldots, n$. Obviously, we would like to choose only those sensors that have a favorable channel to the fusion center, so the choice should be based on $R_i = (Q_i, \sigma_i)$. (In some cases, σ_i is a known constant, then the choice is made based only on $R_i = Q_i$.) Furthermore, we would like to examine and compare a cooperative scheme (the decision to transmit or not by each sensor depends on the channel parameters of all *sensors) and a distributed scheme (the decision of each sensor depends only on the local channel parameters). Finally, we may want to optimize the choice of the "transmission function" $\tilde{\gamma}_i$ from within a class of possible such functions.* □

Example 7.2 (Spatial Signal)**.** *Consider the problem of detecting a spatial signal on the domain $[-1, 1]$ (or more generally on a bounded subset of $\mathbb{R}^d$). The sensors are placed randomly and uniformly in the set $[-1, 1]$, with the fusion center at the origin. Let R_i be the location of sensor i. This serves as the side-information that is available. There are two possible spatial signals $s_0(\cdot)$ and $s_1(\cdot)$, and we wish to detect which of the two is present. Each sensor i makes a noisy measurement of the local*

signal. We assume that the power required for a sensor to transmit its measurement depends on the distance from the sensor to the fusion center. Given a constraint on the total power used by the sensors, which ones should be chosen to transmit? It is not necessarily the case that the sensors closest to the fusion center should be the ones transmitting; for instance, if the spatial signals $s_0(r)$ and $s_1(r)$ are equal when r is in the vicinity of the fusion center, the sensors close to the fusion center do not have any information worth transmitting. In Section 7.8.2, we will give an example where the transmitting sensors should be the ones furthest away from the fusion center. □

In our formulation, we allow the sensors to cooperate, in the sense that the sensors' censoring decisions can be made by the fusion center, on the basis of the entire vector $(R_1, \ldots, R_n)$ of side-information values at each sensor. This can arise, for example, when the local pieces of side-information are some low-resolution data that can be transmitted to the fusion center inexpensively, or when the fusion center is able to monitor the state of the channels from the sensors to itself. Nevertheless, we will establish that when the Type I error probability is asymptotically small, optimal performance can be achieved even in the absence of such cooperation, by having each sensor make its censoring and transmission decisions only on the basis of the locally available side-information. Furthermore, all the sensors can use the same policy, which shows that a simple distributed scheme is asymptotically optimal. The case where there is no cooperation is the asymptotic counterpart of the censoring problem considered in [27] (cf. Section 7.5.3).

We then proceed to consider the Bayesian counterpart of the above formulation, except that for reasons described in Section 7.9, the cooperation among sensors is explicitly ruled out. We characterize the asymptotically optimal performance and the strategies that achieve it. We show that an optimal scheme is to divide the sensors into two groups, each group using the same policy. We also show how some of the results in [29] and [30] can be derived by converting the problems studied therein to our framework. Finally, we provide a generalization of some of the results in [30].

7.2 Problem Formulation

In this section, we introduce our model. We will use the notation $z^{(n)}$ to denote a vector $(z_1, \ldots, z_n)$, where the components of the vector may be numbers, random variables, or functions.

7.2.1 The Basic Elements of the Model

We consider a hypothesis testing problem involving two hypotheses, H_0 and H_1. There are n sensors and a fusion center. Each sensor i observes some side-information R_i, which is a random variable taking values in a set $\mathcal{R}$, and a measurement X_i taking values in a set $\mathcal{X}$. In addition, there is an auxiliary random variable V, taking values in a set $\mathcal{V}$ of our own choosing, which will be used as the "seed" whenever a randomized decision is to be made. These are all the basic random variables in our model. We assume a suitably large measurable space $(\Omega, \mathscr{F})$ so that all random variables can be defined on that space, for any number n of sensors. To avoid technical distractions, we will not delve into measurability issues.

Under each hypothesis H_j, $j = 0, 1$, we assume that we have a measure $\mathbb{P}_j$, and a corresponding expectation operator $\mathbb{E}_j$, with the following properties.

1. The random variable R_i is distributed according to a given marginal probability law μ_j, for every i.

2. Conditioned on $R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n$, the measurements X_i are (conditionally) independent, and each X_i is distributed according to a given regular conditional distribution $\nu_j(\cdot \mid r_i)$.

3. The random variable V is independent of the random variables R_i and X_i, with a distribution that is the same under both hypotheses, and which will be of our choosing.

Note that we have only specified the marginal distributions of the variables R_i. Regarding their joint distribution, we will be making in the sequel one of the following alternative assumptions:

1. Under either hypothesis, the random variables R_i are independent (and therefore i.i.d.).

2. Under either hypothesis, the sequence $(R_1, R_2, \ldots)$ is stationary and ergodic. In this case, we also assume $\mu_0 = \mu_1$, so that the variables R_i provide no information about the true hypothesis.

7.2.2 Sensor Policies and Strategies

There are two types of decisions to be made at each sensor: deciding whether to make a measurement (not censoring), and if a measurement is made, deciding what to transmit to the fusion center. These decisions are to be made based on available information, according to a set of rules (policies). We describe here the types of policies to be considered.

We assume that $R^{(n)}$ is known at the fusion center (in a mathematically equivalent scenario, we could have each sensor communicate its side-information to every other sensor) and that the same is true for the auxiliary random variable V. Based on $R^{(n)}$ and V, we let the fusion center decide which of the sensors should make a measurement X_i. (This is what we term as *cooperation*: the decision depends on the side-information of all sensors.) Subsequently, each uncensored sensor is to generate a message to the fusion center.

Formally, we define a *pure censoring policy* for sensor i as a function $\xi_i : \mathcal{R}^n \mapsto \{0, 1\}$. Let the set of pure censoring policies be Ξ. A *pure transmission policy* for sensor i is a function $\gamma_i : \mathcal{X} \times \mathcal{R} \mapsto \mathcal{Y}$, where $\mathcal{Y}$ is a (possibly infinite) transmission alphabet. These policies are called *pure* because they do not make use of the randomization variable V. We restrict pure transmission policies to belong to a given set Γ. The pair (ξ_i, γ_i) is called a *pure policy* for sensor i.

We allow censoring and transmission policies to be randomized, by considering $\pi_i = (\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}$, which is a collection of pure policies indexed by $\mathcal{V}$. We call π_i a *policy* for sensor i. We envisage the following sequence of events. A realization v of the randomization variable V is generated (this can be done at the fusion center,

with the result communicated to all sensors, or at each sensor using a common seed). Sensor i then uses the pure policy $(\xi_{i,v}, \gamma_{i,v})$. It is censored (no measurement is made) if and only if $\xi_{i,v}(R^{(n)}) = 0$. If on the other hand $\xi_{i,v}(R^{(n)}) = 1$, a message $Y_i = \gamma_{i,v}(X_i, R_i)$ is transmitted to the fusion center. Although we say that the message Y_i is transmitted to the fusion center, our formulation allows for the inclusion of channel noise in the transmission function $\gamma_{i,v}$. More specifically, suppose that the message $\tilde{Y}_i = \tilde{\gamma}_{i,v}(\tilde{X}_i, R_i)$ is transmitted over a noisy channel so that $Y_i = f(\tilde{Y}_i, R_i, W_i)$ is received at the fusion center. Here, f is the channel transfer function and W_i is a random variable conditionally independent of $\tilde{X}_i$, given R_i. Then, we can define $X_i = (\tilde{X}_i, W_i)$, and the transmission function as $\gamma_{i,v}(X_i, R_i) = f(\tilde{\gamma}_{i,v}(\tilde{X}_i, R_i), R_i, W_i)$. As an example, consider Example 7.1 of Section 7.1. In our present notation, we have $X_i = (\tilde{X}_i, W_i)$, $R_i = (Q_i, \sigma_i)$, and $\gamma_{i,v}(X_i, R_i) = Q_i\tilde{\gamma}_{i,v}(\tilde{X}_i) + W_i$. Therefore, in the sequel, we will assume that the message received at the fusion center is the same as Y_i. For convenience, we also assume that X_i and $\gamma_{i,v}(X_i, R_i)$ are always defined, even if sensor i is censored and nothing gets transmitted.

A collection $\pi^{(n)}$ of policies, one for each sensor, all of which involve the same set $\mathcal{V}$ and the same randomization variable V, together with the distribution of V, will be called a *strategy*. We will often abuse terminology, however, and will be referring to $\pi^{(n)}$ as a strategy.

7.2.3 Resource Constraints

We assume that when sensor i makes a measurement X_i and transmits Y_i to the fusion center, certain resources are consumed, and therefore a cost is incurred, possibly depending on the side-information at that sensor. To model such costs, we introduce a function $\rho : \mathcal{R} \times \Gamma \mapsto [0, \infty)$, and interpret $\rho(r, \gamma)$ as the cost incurred by a sensor i that uses a pure policy $\pi = (\xi, \gamma)$, if the side-information at that sensor takes on the value r, and the sensor is not censored, i.e., $\xi(R^{(n)}) = 1$. When the sensor is censored, we assume that no cost is incurred, so that the resulting expected cost at sensor i (under H_0) equals $\rho(\pi) = \mathbb{E}_0[\xi(R^{(n)})\rho(R_i, \gamma)]$. For a more general (randomized) policy $\pi = (\xi_v, \gamma_v)_{v\in\mathcal{V}}$, $\rho(\pi)$ is defined to be equal to $\mathbb{E}_0[\xi_V(R^{(n)})\rho(R_i, \gamma_V)]$, where the

expectation is taken with respect to both $R^{(n)}$ and V. We will say that a strategy $\pi^{(n)} = (\pi_1, \ldots, \pi_n)$ is *admissible* if

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\pi_i) \leq c, \tag{7.1}$$

where c is a given constant.

Note that the resource constraint is in place only under H_0. The presumption here is that H_0 (the "null hypothesis") corresponds to a "normal" situation. Thus, we are constraining the resource utilization to be low under normal circumstances, but allow higher resource utilization under exceptional circumstances. However, in a Bayesian formulation, we will define $\mathbb{E} = q_0\mathbb{E}_0 + q_1\mathbb{E}_1$, where q_j is the prior probability of hypothesis H_j, and will replace $\mathbb{E}_0$ with $\mathbb{E}$ in the definition of $\rho(\pi_i)$.

The following are two examples of resource constraints. Many other choices are possible, to reflect particular constraints of interest to a system designer.

Example 7.3 (Proportional Censoring)**.** *If $\rho(r, \gamma) = 1$ for all $r \in \mathcal{R}$ and all $\gamma \in \Gamma$, then (7.1) becomes a constraint on the average proportion of sensors that make a measurement.* □

Example 7.4 (Power constraints)**.** *Suppose that $\rho(r, \gamma) = \mathbb{E}_0[|\gamma(X_1, r)|^2 \mid R_1 = r]$. In this case, (7.1) becomes a constraint on the average transmission power.* □

7.2.4 The Fusion Center

The fusion center receives the messages Y_i from each sensor. Based on this information, together with the side-information $R^{(n)}$ and the random variable V, it decides between the two hypotheses. Recall that in classical (centralized) Neyman-Pearson hypothesis testing, randomization can reduce the Type II error probability. Accordingly, we assume that the fusion center has access to another random variable V' which is uniformly distributed in [0,1], and independent of everything else. We then let the fusion center use a *randomized fusion rule* $\phi : \mathcal{Y}^n \times \mathcal{R}^n \times \mathcal{V} \times [0, 1] \to \{0, 1\}$ to select one of the two hypotheses. Let $\hat{H}_n = \phi(Y^{(n)}, R^{(n)}, V, V')$, which is a binary

random variable indicating the selected hypothesis. In the above expression, and in order to keep notation simple, we assume that whenever sensor i is censored, Y_i is set to a special symbol y^*.

We summarize the elements of our model in the following definition.

Definition 7.1. *An* overall strategy *consists of the following.*

1. *A set $\mathcal{V}$, and the distribution of a $\mathcal{V}$-valued random variable V;*

2. *an admissible strategy $\pi^{(n)}$ (i.e., one that satisfies the resource constraints);*

3. *a fusion rule ϕ.*

For given n and c, and a given overall strategy, the Type I error and the Type II error probabilities $\mathbb{P}_0(\hat{H}_n = 1)$ and $\mathbb{P}_1(\hat{H}_n = 0)$ are well defined. In a Neyman-Pearson formulation (Section 7.3), we will aim at minimizing the probability of the Type II error (more precisely, its error exponent), subject to a constraint on the Type I error probability. In a Bayesian formulation (Section 7.9), we will aim at minimizing a weighted average of these two error probabilities.

7.2.5 Independent Randomization

Our model allows for randomization based on a globally known randomization variable V, whose distribution is subject to our choice. Such a V can be generated at each sensor using a common seed, or it can be generated at the fusion center and communicated to the sensors. As discussed in [6], the above model of *dependent randomization* includes the special case of *independent randomization,* where the sensors rely on locally generated independent random variables.

Formally, we will say that we have *independent randomization* if the set $\mathcal{V}$ is a Cartesian product of n copies of another set $\mathcal{V}_0$, i.e., $\mathcal{V} = \mathcal{V}_0^n$, and is endowed with a product measure, so that V is of the form $V = (V_1, \ldots, V_n) = V^{(n)}$, where the V_i are independent.

7.2.6 Local and Homogeneous Strategies

Loosely speaking, in a local strategy every sensor i has access only to an independent, locally generated random variable V_i and its own side-information R_i, thus allowing for distributed implementation. Furthermore, in a homogeneous local strategy, every sensor responds in the same way to its local variables. In the definition below, $v = (v_1, \ldots, v_n)$.

Definition 7.2.

1. *A* policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}}$ *is said to be* local *(for sensor i), if (i) independent randomization is used; (ii) $\xi_v(r^{(n)})$ can be expressed as a function of only r_i and v_i; and (iii) $\gamma_v(x, r)$ can be expressed as a function of only x, r, and v_i.*

2. *A* strategy $\pi = (\pi_1, \ldots, \pi_n)$ *is said to be* local *if each π_i is a local policy for sensor i.*

A local policy for sensor i is denoted as $\pi_i = (\xi_{i,v_i}, \gamma_{i,v_i})_{v_i \in \mathcal{V}_0}$, where the functions ξ_{i,v_i} and γ_{i,v_i} are now functions whose arguments are the local random variables X_i and R_i.

Definition 7.3. *A local strategy is said to be* homogeneous *if the independent random variables V_i are identically distributed, and if the policy of every sensor is identified with the same local policy.*

Let us remark that for a homogeneous local strategy associated with a common local policy π, the resource constraint (7.1) simplifies to $\rho(\pi) \leq c$. We let $\Pi(c)$ be the set of local policies that satisfy this constraint.

For the reader's convenience, we summarize the notation introduced so far:

X_i	Measurement of sensor i.
R_i	Side-information of sensor i.
V	Randomization variable.
ξ_i	Pure censoring policy for sensor i.
γ_i	Pure transmission policy for sensor i.
$\pi_i = (\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}$	A policy for sensor i. Given V, $\xi_{i,V}$ is a pure censoring policy, and $\gamma_{i,V}$ is a pure transmission policy.
$\pi^{(n)}$	A strategy $(\pi_1, \ldots, \pi_n)$.
$\rho(\pi)$	Expected cost of policy π.
$\Pi(c)$	Set of local policies π satisfying the resource constraint $\rho(\pi) \leq c$.

7.3 The Neyman-Pearson Problem

Given an overall strategy for the n-sensor problem, we will use β_n to denote the resulting Type II error probability, $\mathbb{P}_1(\hat{H}_n = 0)$. For any given n, c, and α, we define

$$\beta_n^*(c, \alpha) = \inf \beta_n,$$

where the infimum is taken over all overall strategies that satisfy the resource constraint (7.1), as well as the constraint $\mathbb{P}_0(\hat{H}_n = 1) \leq \alpha$.

The above optimization problem is intractable, even in the absence of censoring. Even if it were tractable, implementing an optimal cooperative censoring strategy would involve complicated feedback from the fusion center to the sensors. We will see however, that the problem becomes tractable if n is large and α is small, and under an asymptotic optimality criterion. We will focus on the case of an asymptotically small Type I error probability and the associated optimal error exponent

$$\lim_{\alpha \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \beta_n^*(c, \alpha).$$

7.3.1 Assumptions and Notation

Our main assumptions for this chapter are similar in nature to those in Section 2.3. Recall that under H_j, the measure μ_j describes the distribution of R_i, and $\nu_j(\cdot \mid \cdot)$ describes the conditional distribution of X_i given R_i. We use the notation $\mu \ll \mu'$ to indicate that a measure μ is absolutely continuous with respect to another measure μ'.

Assumption 7.1. *We have $\mu_0 \ll \mu_1$, and for every $r \in \mathcal{R}$, $\nu_0(\cdot \mid r) \ll \nu_1(\cdot \mid r)$.*

Let $\mathrm{d}\mu_0/\mathrm{d}\mu_1$ be the Radon-Nikodym derivative (likelihood ratio) of the measures μ_0 and μ_1. Similarly, we define a function $\ell_{01} : \mathcal{X} \times \mathcal{R} \mapsto [0, \infty)$, so that for every $r \in \mathcal{R}$, $\ell_{01}(X_i \mid r)$ is the likelihood ratio between the two hypotheses, when X_i is observed, given that $R_i = r$. Formally, this is the Radon-Nikodym derivative of the measures $\nu_0(\cdot \mid r)$ and $\nu_1(\cdot \mid r)$ on the set $\mathcal{X}$.

In the same vein, for any pure transmission policy $\gamma \in \Gamma$, we define a function $\ell_{01}^{\gamma} : \mathcal{Y} \times \mathcal{R} \mapsto [0, \infty)$, so that for every $r \in \mathcal{R}$, $\ell_{01}^{\gamma}(Y_i \mid r)$ is the likelihood ratio between the two hypotheses, when $Y_i = \gamma(X_i, R_i)$ is received at the fusion center, given that $R_i = r$. Formally, this is the Radon-Nikodym derivative of the measures $\nu_0^{\gamma}(\cdot \mid r)$ and $\nu_1^{\gamma}(\cdot \mid r)$ on the set $\mathcal{Y}$, where $\nu_j^{\gamma}(\cdot \mid r)$ is the measure $\nu_j(\cdot \mid r)$ restricted to the σ-algebra generated by $Y_i = \gamma(X_i, r)$.

Let us fix a strategy $\big((\xi_{i,v}, \gamma_{i,v})_{v \in \mathcal{V}}\big)_{1 \le i \le n}$, and recall that $(\xi_{i,V}, \gamma_{i,V})$ is the resulting pure policy of sensor i, as determined by V. With the above introduced notation, the likelihood ratio calculated at the fusion center, on the basis of the available information $(R^{(n)}, Y^{(n)}, V)$, is

$$\prod_{i : \xi_{i,V}(R^{(n)}) = 1} \ell_{01}^{\gamma_{i,V}}(Y_i \mid R_i) \prod_{i=1}^{n} \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i). \tag{7.2}$$

For convenience, we define the random variables Z_i and S_n by

$$Z_i = -\xi_{i,V}(R^{(n)}) \log \ell_{01}^{\gamma_{i,V}}(Y_i \mid R_i) - \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i), \tag{7.3}$$

and

$$S_n = \sum_{i=1}^{n} Z_i, \tag{7.4}$$

so that S_n is the negative of the log-likelihood ratio at the fusion center.

The amount of relevant information contained in Y_i, given that $R_i = r$ and that sensor i employs a pure transmission policy $\gamma \in \Gamma$, is quantified by the KL divergence, defined by

$$I(r, \gamma) = \mathbb{E}_0\Big[\log \ell_{01}^{\gamma}(Y_i \mid r) \;\Big|\; R_i = r \Big].$$

Assumption 7.2. *We have*

$$\mathbb{E}_0\big[\log^2 \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1) \big] < \infty,$$
$$\mathbb{E}_0\big[\log^2 \ell_{01}(X_1 \mid R_1) \big] < \infty.$$

We record a consequence of Assumption 7.2, whose proof is similar to that of Lemma 2.1.

Lemma 7.1. *We have* $\mathbb{E}_0\big[\log(\mathrm{d}\mu_0/\mathrm{d}\mu_1) \big] < \infty$. *Furthermore, for every* $\gamma \in \Gamma$, *we have* $0 \leq I(R_1, \gamma) < \infty$, $\mathbb{P}_0$*-a.s., and*

$$\mathbb{E}_0\Big[\log^2 \ell_{01}^{\gamma}(Y_1 \mid R_1) \;\Big|\; R_1 \Big] \leq a(R_1), \quad \mathbb{P}_0\text{-a.s.},$$

for some function $a(\cdot)$ *that satisfies* $\mathbb{E}_0[a(R_1)] < \infty$.

7.4 The I.I.D. Case

In this section, we characterize the optimal exponent for the Neyman-Pearson problem. Furthermore, we show that the optimal exponent does not change when we restrict to homogeneous local strategies. Throughout this section, we assume that under either hypothesis, the random variables R_i are i.i.d.

According to Stein's Lemma [39], in the absence of censoring or side-information, and if all sensors use the same pure transmission policy, the error exponent is the

negative of the associated KL divergence. By a similar argument, if the sensors use a common local policy $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0} \in \Pi(c)$, we expect to obtain an error exponent equal to

$$\lambda(\pi) = -\mathbb{E}_0\Big[\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big] - \mathbb{E}_0\big[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})\big].$$

It is then natural to optimize over all admissible local policies $\pi \in \Pi(c)$, and define

$$\lambda^*(c) = -\mathbb{E}_0\Big[\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big] - \sup_{\pi \in \Pi(c)} \mathbb{E}_0\big[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})\big],$$

where the optimization includes the choice of the local randomization variable V_1 and its distribution.

We will show that $\lambda^*(c)$ is the optimal error exponent, even if we remove the restriction to homogeneous local strategies, in the limit as α goes to zero. In deriving the required lower bound, we will not be able to invoke standard results from Large Deviations Theory, because the summands in the log-likelihood ratio are all affected by the overall side-information $R^{(n)}$, and are not independent. For this reason, the proof of the lower bound will proceed from first principles.

Our main result is as follows.

Theorem 7.1. *Suppose that Assumptions 7.1 and 7.2 hold, and that the random variables R_i are i.i.d., under either hypothesis.*

(i) For every $\alpha \in (0,1)$ and $c > 0$, the optimal error exponent $\liminf_{n\to\infty} (1/n) \log \beta_n^*(c, \alpha)$ is bounded below by $\lambda^*(c)/(1-\alpha)$.

(ii) For every $\alpha \in (0,1)$ and $c > 0$, there exists a sequence of admissible homogeneous local strategies (one for each n) that satisfy the Type I error constraint, and such that the corresponding Type II error probabilities β_n satisfy

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n = \lambda^*(c).$$

(iii) For every $c > 0$,

$$\lim_{\alpha \to 0} \liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(c, \alpha) = \lambda^*(c).$$

Furthermore, if the random variables R_i are stationary and ergodic, and $\mu_0 = \mu_1$, then (iii) still holds.

We record an elementary fact that will be used later.

Lemma 7.2. *The function $\lambda^* : [0, \infty) \mapsto (-\infty, 0]$ is convex, and in particular, continuous on $(0, \infty)$.*

Proof. Suppose that $c = \delta c_1 + (1 - \delta)c_2$ for some $\delta \in [0, 1]$. Fix some $\epsilon > 0$ and consider two local policies $\pi^k \in \Pi(c_k)$, $k = 1, 2$, (so that $\rho(\pi^k) \leq c_k$), which satisfy

$$\lambda(\pi^k) \leq \lambda^*(c_k) + \epsilon.$$

Consider a new local policy π^0 that uses π^1 with probability δ, and π^2 with probability $1 - \delta$. We then have

$$\rho(\pi^0) = \delta\rho(\pi^1) + (1 - \delta)\rho(\pi^2) \leq \delta c_1 + (1 - \delta)c_2 = c,$$

so that $\pi^0 \in \Pi(c)$. Furthermore,

$$\begin{aligned}\lambda^*(c) \leq \lambda(\pi^0) &= \delta\lambda(\pi^1) + (1 - \delta)\lambda(\pi^2) \\ &\leq \delta\lambda^*(c_1) + (1 - \delta)\lambda^*(c_2) + \epsilon.\end{aligned}$$

The result follows by letting ϵ decrease to zero. □

7.4.1 Proof of the Lower Bound

In this subsection, we prove the first part of Theorem 7.1. Throughout this subsection, α is held at a fixed value.

Suppose that a strategy $\pi^{(n)}$ has been fixed. Since we are interested in a lower bound on the resulting error exponent, we assume that the fusion center uses the best possible fusion rule. As the fusion center is faced with a classical Neyman-Pearson problem, where the information available is $(Y^{(n)}, R^{(n)}, V, V')$, a corresponding opti-

mal fusion rule is a likelihood ratio test of the form:

$$\hat{H}_n = 0 \qquad \text{if and only if} \qquad \frac{S_n}{n} \leq T_n,$$

where T_n is a possibly randomized threshold (determined by V'). It is convenient to consider the expected value of the log-likelihood ratio, given that the side-information has been revealed and the randomization variable has been realized. We thus define

$$\begin{aligned}\Lambda_n &= \frac{1}{n}\mathbb{E}_0[S_n \mid R^{(n)}, V] \\ &= -\frac{1}{n}\sum_{i=1}^{n} \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i) - \frac{1}{n}\sum_{i=1}^{n} \xi_{i,V}(R^{(n)}) I(R_i, \gamma_{i,V}),\end{aligned}$$

where the second equality follows from (7.3)-(7.4) and the definition of $I(r,\gamma)$. We start by showing that S_n/n is asymptotically close (in probability) to Λ_n.

Lemma 7.3. *For every $\eta > 0$, $\lim_{n\to\infty} \mathbb{P}_0\Big(\big|\frac{S_n}{n} - \Lambda_n\big| < \eta\Big) = 1$.*

Proof. We condition on $R^{(n)}$ and V. Then $S_n - n\Lambda_n$ becomes a sum of (conditionally) independent random variables, each having (conditional) variance bounded by $a(R_i)$ [cf. Lemma 7.1]. Chebychev's inequality yields

$$\begin{aligned}&\mathbb{P}_0\Big(\big|\frac{S_n}{n} - \Lambda_n\big| \geq \eta \;\Big|\; R^{(n)}, V\Big) \\ &\leq \frac{1}{n^2\eta^2}\mathbb{E}_0\Big[(S_n - n\Lambda_n)^2 \;\Big|\; R^{(n)}, V\Big] \\ &\leq \frac{1}{n^2\eta^2}\mathbb{E}_0\Big[\sum_{i=1}^{n} \log^2 \ell_{01}^{\gamma_{i,V}}(Y_i \mid R_i) \;\Big|\; R^{(n)}, V\Big] \\ &\leq \frac{1}{n^2\eta^2}\sum_{i=1}^{n} a(R_i). \qquad (7.5)\end{aligned}$$

Taking unconditional expectations of both sides, we obtain

$$\mathbb{P}_0\Big(\big|\frac{S_n}{n} - \Lambda_n\big| \geq \eta\Big) \leq \frac{1}{n\eta^2}\mathbb{E}_0[a(R_1)],$$

which converges to zero because $\mathbb{E}_0[a(R_1)] < \infty$. □

The next lemma is crucial in that it relates the amount of information provided by an admissible strategy to the best possible exponent $\lambda^*(c)$ under local admissible strategies. The key idea is that as far as sensor i is concerned, the side-information at the other sensors has the same effect as using additional local randomization variables.

Lemma 7.4. *For any sequence of admissible strategies (one for each n), and for every sequence of measurable subsets A_n of Ω, we have*

$$\liminf_{n\to\infty} \mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] \geq \lambda^*(c).$$

Proof. Suppose that a sequence of admissible strategies $\pi^{(n)}$ has been fixed, and let $(\xi_{i,v}, \gamma_{i,v})_{v\in\mathcal{V}}$ be the policy of sensor i. We have

$$\mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] \geq -\mathbb{E}_0\Big[\frac{1}{n}\sum_{i=1}^n \log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i)\mathbf{1}_{A_n}\Big] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_0\Big[\xi_{i,V}(R^{(n)})I(R_i,\gamma_{i,V})\Big]. \quad (7.6)$$

To bound the first term,

$$\begin{aligned}
&\mathbb{E}_0\Big[\frac{1}{n}\sum_{i=1}^n \log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i)\mathbf{1}_{A_n}\Big] \\
&= \mathbb{E}_0\Big[\Big(\frac{1}{n}\sum_{i=1}^n \log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i) - \mathbb{E}_0\Big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big]\Big)\mathbf{1}_{A_n}\Big] + \mathbb{E}_0\Big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big]\mathbb{P}_0(A_n) \\
&\leq \mathbb{E}_0\Big[\ \Big|\frac{1}{n}\sum_{i=1}^n \log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i) - \mathbb{E}_0\Big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big]\Big|\ \Big] + \mathbb{E}_0\Big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big]. \qquad (7.7)
\end{aligned}$$

In the limit as $n\to\infty$, the first term in the R.H.S. of (7.7) converges to 0 (because the R_i are i.i.d. and the L^1 ergodic theorem applies), which leaves the term $\mathbb{E}_0\big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\big]$.

For any $v\in\mathcal{V}$, let $\tilde{\xi}_{i,v}(R_i) = \mathbb{E}_0[\xi_{i,v}(R^{(n)}) \mid R_i]$. Then,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \mathbb{E}_0\Big[\xi_{i,V}(R^{(n)})I(R_i,\gamma_{i,V})\Big] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_0\Big[\tilde{\xi}_{i,V}(R_i)I(R_i,\gamma_{i,V})\Big]
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_0\Big[\tilde{\xi}_{i,V}(R_1) I(R_1, \gamma_{i,V})\Big] \tag{7.8}$$

$$\leq \sup_{\pi \in \Pi(c)} \mathbb{E}_0\Big[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})\Big]. \tag{7.9}$$

The equality in (7.8) follows from the stationarity of $R^{(n)}$. The inequality in (7.9) is obtained by considering a local policy $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v}\in\bar{\mathcal{V}}}$, where $\bar{\mathcal{V}} = \{1, \ldots, n\} \times [0,1] \times \mathcal{V}$, defined as follows. Let $\bar{V} = (J, U, V)$, where J is uniform on $\{1, 2, \ldots, n\}$, U is uniform on $[0,1]$, and J, U and V are independent. Let $\delta_{\bar{V}} = \gamma_{J,V}$, and for every $r \in \mathcal{R}$, let

$$\tau_{\bar{V}}(r) = \begin{cases} 1, & \text{if } U \leq \tilde{\xi}_{J,V}(r), \\ 0, & \text{otherwise.} \end{cases}$$

In particular, if $J = i$, $U = u$ and $V = v$, the new local policy $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v}\in\bar{\mathcal{V}}}$ applies the pure transmission policy $\gamma_{i,v}$, and censors if $u > \tilde{\xi}_{i,v}(r)$, when the local side-information is r. Then, $(\tau_{\bar{v}}, \delta_{\bar{v}})_{\bar{v}\in\bar{\mathcal{V}}} \in \Pi(c)$ and the R.H.S. of (7.8) is equal to $\mathbb{E}_0[\tau_{\bar{V}}(R_1) I(R_1, \delta_{\bar{V}})]$. Hence, (7.9) follows.

Combining the above with (7.7), we obtain

$$\begin{aligned} &\liminf_{n\to\infty} \mathbb{E}_0[\Lambda_n \mathbf{1}_{A_n}] \\ &\geq -\mathbb{E}_0\Big[\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big] - \sup_{\pi\in\Pi(c)} \mathbb{E}_0[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})] \\ &= \lambda^*(c), \end{aligned}$$

and the lemma is proved. □

Lemma 7.5. *For all $\alpha \in (0,1)$ and $c \geq 0$, we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(c, \alpha) \geq \frac{\lambda^*(c)}{1-\alpha}.$$

Proof. Fix some $\eta > 0$. For every n, consider an admissible strategy $\pi^{(n)}$ and a fusion rule with $\mathbb{P}_0(\hat{H}_n = 0) \geq 1 - \alpha$. We use a change of measure argument, similar to that

in the proof of Stein's Lemma in [39], to get

$$\begin{aligned}
\beta_n &= \mathbb{P}_1\Big(\frac{S_n}{n} \le T_n\Big) \\
&= \mathbb{E}_0\Big[e^{S_n}\mathbf{1}\Big(\frac{S_n}{n} \le T_n\Big)\Big] \\
&\ge \mathbb{E}_0\Big[e^{S_n}\mathbf{1}\Big(\Lambda_n - \eta < \frac{S_n}{n} \le T_n\Big)\Big] \\
&\ge \mathbb{E}_0\Big[e^{n(\Lambda_n-\eta)}\mathbf{1}\Big(\Lambda_n - \eta < \frac{S_n}{n} \le T_n\Big)\Big]. \qquad (7.10)
\end{aligned}$$

Let $A_n = \{\omega : \Lambda_n - \eta < \frac{S_n}{n} \le T_n\}$. Then,

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}_0(A_n) \\
&\ge \liminf_{n\to\infty} \Big(\mathbb{P}_0\Big(\frac{S_n}{n} \le T_n\Big) + \mathbb{P}_0\Big(\frac{S_n}{n} > \Lambda_n - \eta\Big) - 1\Big) \\
&\ge 1 - \alpha > 0,
\end{aligned}$$

where we have made use of Lemma 7.3. Hence, for sufficiently large n, we can condition on A_n and obtain

$$\begin{aligned}
\frac{1}{n}\log\beta_n &\ge \frac{1}{n}\log\mathbb{E}_0\Big[e^{n(\Lambda_n-\eta)}\mathbf{1}_{A_n}\Big] \\
&= \frac{1}{n}\log\mathbb{P}_0(A_n) + \frac{1}{n}\log\mathbb{E}_0\Big[e^{n(\Lambda_n-\eta)} \;\Big|\; A_n\Big] \\
&\ge \frac{1}{n}\log\mathbb{P}_0(A_n) + \mathbb{E}_0[\Lambda_n|\; A_n] - \eta, \qquad (7.11)
\end{aligned}$$

where the last step follows from Jensen's inequality. Applying Lemma 7.4, we have

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n}\log\beta_n^* &\ge \liminf_{n\to\infty} \frac{\mathbb{E}_0[\Lambda_n\mathbf{1}_{A_n}]}{\mathbb{P}_0(A_n)} - \eta \\
&\ge \frac{\lambda^*(c)}{1-\alpha} - \eta.
\end{aligned}$$

The result follows by letting $\eta \to 0$. □

Lemma 7.5 concludes the proof of the lower bound (part (i) of Theorem 7.1).

7.4.2 Proof of the Upper Bound.

In this section we construct a sequence of admissible homogeneous local strategies (one for each n), under which the lower bound is asymptotically attained.

For each n, consider a strategy involving a common local policy $\pi^n = (\xi_v^n, \gamma_v^n)_{v\in\mathcal{V}_0} \in \Pi(c)$, such that

$$\lim_{n\to\infty} \lambda(\pi^n) = \lambda^*(c). \tag{7.12}$$

Consider the fusion rule $\hat{H}_n = \phi_n(Y^{(n)}, R^{(n)}, V)$ that selects H_0 if and only if

$$\frac{S_n}{n} \leq \lambda(\pi^n) + n^{-\frac{1}{4}}.$$

Let

$$\mathcal{S}(R^{(n)}, V) = \big\{Y^{(n)} \mid \phi_n(Y^{(n)}, R^{(n)}, V) = 0\big\}. \tag{7.13}$$

Since the random variables Z_i are i.i.d., with mean $\lambda(\pi^n)$ and variance bounded by some constant a (cf. Assumption 7.2), we have

$$\frac{1}{n^2}\mathbb{E}_0\big[\big(S_n - n\lambda(\pi^n)\big)^2\big] \leq \frac{a}{n},$$

and hence

$$\mathbb{P}_0\Big(\Big|\frac{S_n}{n} - \lambda(\pi^n)\Big| \leq n^{-\frac{1}{4}}\Big) \to 1. \tag{7.14}$$

This implies that $\mathbb{P}_0(\hat{H}_n = 1) \to 0$ as $n \to \infty$. It follows that for any given $\alpha > 0$, the constraint $\mathbb{P}_0(\hat{H}_n = 1) \leq \alpha$ will be satisfied for large n. We let $\tilde{\beta}_n$ be the minimum possible Type II error (over all fusion rules), when we use the common local policy π^n at all sensors. In particular, $\tilde{\beta}_n \leq \mathbb{P}_1(\hat{H}_n = 0)$.

The next lemma is a modification of Stein's Lemma. The proof is almost the same as the standard proof (see [53]) but we include it for completeness, and because we want to check that it remains valid in the case where $\mu_0 = \mu_1$ and the variables R_i

are stationary and ergodic (as opposed to i.i.d.), as will be discussed in Section 7.6.

Lemma 7.6. *For the above defined common local policy π^n, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \tilde{\beta}_n = \lambda^*(c).$$

Proof. We have

$$\begin{aligned}
&\mathbb{P}_1\big(\hat{H}_n = 0 \mid R^{(n)}, V\big) \\
&= \int_{\mathcal{S}(R^{(n)},V)} \prod_{i:\xi^n_{V_i}(R_i)=1} \nu_1^{\gamma^n_{V_i}}(\mathrm{d}y_i \mid R_i) \\
&= \int_{\mathcal{S}(R^{(n)},V)} \Big\{ e^{-\sum_{i=1}^n \xi^n_{V_i}(R_i) \log \ell_{01}^{\gamma^n_{V_i}}(y_i|R_i)} \cdot \prod_{i:\xi^n_{V_i}(R_i)=1} \nu_0^{\gamma^n_{V_i}}(\mathrm{d}y_i \mid R_i) \Big\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\tilde{\beta}_n &\le \mathbb{P}_1(\hat{H}_n = 0) \\
&= \mathbb{E}_1\big[\mathbb{P}_1(\hat{H}_n = 0 \mid R^{(n)}, V)\big] \\
&= \mathbb{E}_0\Big[\mathbb{P}_1(\hat{H}_n = 0 \mid R^{(n)}, V) \exp\Big(-\sum_{i=1}^n \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_i)\Big)\Big] \\
&= \mathbb{E}_0\Big[\int_{\mathcal{S}(R^{(n)},V)} e^{S_n} \prod_{i:\xi^n_{V_i}(R_i)=1} \nu_0^{\gamma^n_{V_i}}(\mathrm{d}y_i \mid R_i)\Big].
\end{aligned}$$

Recall that on the set $\mathcal{S}(R^{(n)}, V)$, we have $S_n \le n(\lambda(\pi^n) + n^{-1/4})$, which yields

$$\begin{aligned}
\tilde{\beta}_n &\le \exp\big(n(\lambda(\pi^n) + n^{-\frac{1}{4}})\big)\mathbb{E}_0[\mathbb{P}_0(\hat{H}_n = 0 \mid R^{(n)}, V)] \\
&\le \exp\big(n(\lambda(\pi^n) + n^{-\frac{1}{4}})\big).
\end{aligned}$$

Thus,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \tilde{\beta}_n &\le \lim_{n\to\infty} \big(\lambda(\pi^n) + n^{-\frac{1}{4}}\big) \\
&= \lambda^*(c). \qquad (7.15)
\end{aligned}$$

To show the lower bound, we mimic the proof of Lemma 7.5, with $\lambda(\pi^n)$ replacing Λ_n in that proof. Then from (7.10), with T_n as the threshold for the optimal Neyman-Pearson test,

$$\begin{aligned}\tilde{\beta}_n &\geq e^{n(\lambda(\pi^n)-n^{-1/4})}\mathbb{P}_0\Big(\lambda(\pi^n) - n^{-\frac{1}{4}} < \frac{S_n}{n} \leq T_n\Big) \\ &\geq e^{n(\lambda(\pi^n)-n^{-1/4})}\Big(1-\alpha+\mathbb{P}_0\big(\frac{S_n}{n} > \lambda(\pi^n) - n^{-\frac{1}{4}}\big) - 1\Big).\end{aligned}$$

Taking logarithms, dividing by n, and then letting $n \to \infty$, and using (7.14),

$$\liminf_{n\to\infty} \frac{1}{n} \log \tilde{\beta}_n \geq \lambda^*(c). \tag{7.16}$$

The lemma is proved. □

This concludes the proof of the upper bound (part (ii) of Theorem 7.1). Part (iii) is an immediate consequence of parts (i) and (ii). The last part, involving a stationary and ergodic sequence of random variables R_i will be discussed in Section 7.6.

7.4.3 The Role of Randomization

Suppose that the cost function is independent of the transmission function, i.e., $\rho(r,\gamma) = \tilde{\rho}(r)$ for some nonnegative function $\tilde{\rho}$. Suppose also that the set of pure transmission policies Γ is of the form $\Gamma = \prod_{r\in\mathcal{R}} \Gamma(r)$, where $\Gamma(r)$ is a set of allowed pure transmission policies $\gamma(\cdot, r)$, when the side-information takes the value r. We interpret this as each sensor i being able to choose its own transmission policy separately for each possible value of the side-information R_i. Then, finding an asymptotically optimal strategy is simplified because

$$\lambda^*(c) = -\mathbb{E}_0\big[\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\big] - \sup \mathbb{E}_0\big[\xi_{V_1}(R_1) \sup_{\gamma\in\Gamma(R_1)} I(R_1,\gamma)\big],$$

where the first supremum is taken over all local censoring policies $(\xi_v)_{v\in\mathcal{V}_0}$ that satisfy

$$\mathbb{E}_0[\xi_{V_1}(R_1)\tilde{\rho}(R_1)] \leq c.$$

In particular, a pure transmission policy γ can be used. In achieving $\lambda^*(c)$, we choose transmission policies that maximize the Kullback-Leibler (KL) divergence between the distributions of the messages Y_i, separately for each possible value of R_i. It is intuitively clear that this has to be the case, since the KL divergence quantifies the discrimination between two distributions. Randomization goes a long way to simplify the form of an optimal policy, as can be seen in the proof of Lemma 7.4. If we are restricted to pure strategies, then homogeneous local strategies need not be asymptotically optimal. Indeed, suppose that $\mathcal{R}$ and Γ are finite, so that we have to choose from a finite set of pure local policies, $\{(\xi^k, \gamma^k) : k = 1, \ldots, K\}$. We are then faced with the optimization problem

$$\begin{aligned} \max_{x_1,\ldots,x_K} \quad & \sum_{k=1}^{K} x_k \mathbb{E}_0\big[\xi^k(R_1) I(R_1, \gamma^k)\big], \\ \text{s.t.} \quad & \sum_{k=1}^{K} x_k \mathbb{E}_0\big[\xi^k(R_1) \rho(R_1, \gamma^k)\big] \le c, \\ & \sum_{k=1}^{K} x_k = 1, \\ & x_k \ge 0, \quad \forall i, \end{aligned}$$

where x_k is the proportion of sensors that use the pure policy (ξ^k, γ^k). This is a linear program with two constraints, and generically, an optimal solution will have two nonzero variables. Let x^* and $1 - x^*$ be the optimal values of these two variables, and assume (to simplify the discussion) that $x^* n$ is integer. As long as we are restricted to pure local strategies, then we have to divide the sensors into two groups, one group consisting of $x^* n$ sensors that use one local policy and another group consisting of $(1-x^*)n$ sensors that use another local policy. Thus, in the Neyman-Pearson problem, randomization and non-homogeneity are alternatives to each other.

7.5 Discussion of the Neyman-Pearson Results

In this section, we discuss some variations, extensions and generalizations of the results in Section 7.4.

7.5.1 Cooperation can Improve Performance

We observe that if $\alpha \in (0,1)$ is held fixed, there is a gap between the upper and lower bounds in Theorem 7.1. We discuss here the extent to which cooperation among the sensors improves detection performance.

If we do not allow cooperation among the sensors, i.e., if $\pi^{(n)}$ is local, we can use an argument similar to the proof of Lemma 7.6 to show that $\lambda^*(c)$ is a lower bound, for any fixed $\alpha \in (0,1)$. In particular, homogeneous local strategies are asymptotically optimal.

If cooperation is allowed, and α is fixed, the optimal exponent can be less than $\lambda^*(c)$, as shown in Proposition 7.1 below. In that case, asymptotically optimal strategies are difficult to find, and we do not believe that a simple closed form expression for the optimal error exponent is possible.

Proposition 7.1. *For a fixed $\alpha \in (0,1)$, we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(c,\alpha) \leq \lambda^*\big(\frac{c}{1-\alpha}\big).$$

Proof. Let us fix n and α, and some $\epsilon \in (0,\alpha)$. Consider the following strategy. With probability $p = \alpha - \epsilon$, we use the censoring policy $\xi_i \equiv 0$ for all i, and always declare H_1. In this case, we satisfy the resource constraint (7.1) with c replaced by $c_1 = 0$, and have a Type I error probability of $\alpha_1 = 1$. Let

$$\alpha_2 = \frac{\epsilon}{1-\alpha+\epsilon}, \qquad c_2 = \frac{c}{1-\alpha+\epsilon}.$$

With probability $1-p$, we use a homogeneous local strategy involving a common

local policy $\pi \in \Pi(c_2)$ that satisfies

$$\lambda(\pi) \leq \lambda^*(c_2) + \frac{1}{n},$$

and a fusion rule that achieves a Type I error probability of α_2. Note that $\alpha = p\alpha_1 + (1-p)\alpha_2$ and $c = pc_1 + (1-p)c_2$, so that the composite strategy we have constructed is admissible. Using Lemma 7.6, its Type II error probability β_n satisfies

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(c, \alpha) \leq \lim_{n\to\infty} \frac{1}{n} \log \beta_n = \lambda^*(c_2) = \lambda^*\Big(\frac{c}{1-\alpha+\epsilon}\Big).$$

Taking $\epsilon \to 0$ and using the continuity of $\lambda^*(\cdot)$, we obtain the desired result. □

From Proposition 7.1, the improvement in the error exponent when using cooperation instead of using homogeneous, local strategies has a magnitude of at least

$$\lambda^*(c) - \lambda^*\Big(\frac{c}{1-\alpha}\Big),$$

which is strictly positive when c is small enough, and is upper bounded by (cf. Lemma 7.5)

$$\Big(1 - \frac{1}{1-\alpha}\Big)\lambda^*(c) = \frac{\alpha}{1-\alpha}|\lambda^*(c)|.$$

We see that in a severely constrained network (small c), the price paid for not cooperating is positive but not very large. Thus, the communications overhead and resulting complexity may not justify the use of cooperative censoring.

7.5.2 Generalized Sensor Policies

In this section, we provide a generalization of our framework, by allowing a more general class of policies. In the preceding, each sensor could choose separately a censoring policy and a transmission policy. Here, these two choices will be subsumed under a single choice of a "generalized policy" $\bar{\zeta}_i$. We will see that when specialized to our earlier setting, the generalized formulation will also allow the choice of transmission functions to be made cooperatively, on the basis of the global side-information vector

$R^{(n)}$.

Formally, we define a *(generalized) policy* as a function $\bar{\zeta} : \mathcal{X} \times \mathcal{R}^{(n)} \times \mathcal{V} \mapsto \mathcal{Y}$. A sensor that uses policy $\bar{\zeta}$, transmits $Y_i = \bar{\zeta}(X_i, R^{(n)}, V)$ to the fusion center. Assuming independent randomization, the notion of a local policy for sensor i is defined as before, namely, the dependence on $(R^{(n)}, V^{(n)})$ is only through (R_i, V_i). Once more, a local strategy is called homogeneous if every sensor uses the same mapping from (X_i, R_i, V_i) to Y_i.

Let $r^{(n,i)}$ be the vector $r^{(n)}$ after removing the ith component. As before, for every $r^{(n,i)} \in \mathcal{R}^{n-1}$ and for every $v \in \mathcal{V}$, we require that the function $\zeta_{r^{(n,i)},v}$, defined by $\zeta_{r^{(n,i)},v}(x, r_i) = \bar{\zeta}(x, (r^{(n,i)}, r_i), v)$ be in a given set Γ of functions from $\mathcal{X} \times \mathcal{R}$ to $\mathcal{Y}$. The function $\zeta_{r^{(n,i)},v}$ is called a *pure* local policy. A generalized policy $\bar{\zeta}$ can be viewed as a random choice of a *pure* local policy $\zeta_{R^{(n,i)},V}$, based on the value of $R^{(n,i)}$ and V. We assume that every pure policy ζ consumes an amount $\tilde{\rho}(\zeta)$ of a certain resource. Given a generalized policy $\bar{\zeta}$ for sensor i, the policy of that sensor is chosen to be $\zeta_{R^{(n,i)},V}$. The cost is defined to be $\mathbb{E}_0\big[\tilde{\rho}(\zeta_{R^{(n,i)},V})\big]$, where the expectation is taken over $R^{(n,i)}$ and V. Similar to (7.1), we are interested in *admissible* strategies $(\bar{\zeta}^1, \ldots, \bar{\zeta}^n)$, that satisfy the constraint $\sum_{i=1}^n \mathbb{E}_0\big[\tilde{\rho}(\zeta^i_{R^{(n,i)},V})\big] \leq nc$.

Example 7.5 (Censoring)**.** *Consider the setting of Section 7.2, and assume without loss of generality, that that there is a special element y^* of $\mathcal{Y}$ that is never used, i.e., $\gamma(x, r) \neq y^*$, for every $x \in \mathcal{X}$, $r \in \mathcal{R}$, and $\gamma \in \Gamma$. Given a pure local policy (ξ, γ), we will represent a sensor i that decides to censor ($\xi_i(R_i) = 0$) as one that transmits a "dummy" message equal to y^*. Such a dummy message carries the same information to the fusion center as censoring (the absence of a message). We let $\zeta(x, r) = y^*$ whenever $\xi(r) = 0$, and $\zeta(x, r) = \gamma(x, r)$ otherwise. If $\rho(\cdot, \cdot)$ is the resource function used in our earlier formulation, it is natural to define $\tilde{\rho}(\zeta) = \mathbb{E}_0[\xi(R_1)\rho(R_1, \gamma)]$.* □

Example 7.6 (Power Constraints)**.** *Suppose that $\mathcal{Y} = \mathbb{R}$, and that the cost of a pure local policy ζ at sensor i is $\tilde{\rho}(\zeta) = \mathbb{E}_0[|\zeta(X_i, R_i)|^2]$. Then, $\tilde{\rho}(\zeta)$ corresponds to the expected power consumed by ζ. In this setting, a message with value equal to zero can also be viewed as a censoring decision.* □

When specialized to the censoring problem of earlier sections, the main difference introduced by the current framework is the following: the transmission function γ_i used by sensor i can now be chosen on the basis of not only the randomization variable V, but also the side-information $R^{(n,i)}$ at the other sensors.

For any pure local policy ζ for sensor 1, let $I(r,\zeta)$ be the KL divergence associated with the measurement $Y_1 = \zeta(X_1, R_1)$. For a randomized local policy $\bar{\zeta} : \mathcal{X}\times\mathcal{R}\times\mathcal{V}_0 \mapsto \mathcal{Y}$ for sensor 1, let ζ_{V_1} be a random variable whose realized value is the pure local policy $\zeta_v(\cdot,\cdot) = \bar{\zeta}(\cdot,\cdot,v)$ whenever $V_1 = v$. We have the following generalization of Theorem 7.1, where $\beta_n^*(c,\alpha)$ is, as before, the optimal error probability.

Theorem 7.2. *Suppose that Assumptions 7.1 and 7.2 hold, and that the random variables R_i are i.i.d., under either hypothesis. Then, for every $c > 0$,*

$$\begin{aligned}&\lim_{\alpha\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\beta_n^*(c,\alpha)\\&= -\mathbb{E}_0\Big[\log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(R_1)\Big] - \sup\mathbb{E}_0[I(R_1,\zeta_{V_1})],\end{aligned}$$

where the supremum is taken over all local policies $\bar{\zeta}$ for sensor 1 that satisfy $\mathbb{E}_0[\tilde{\rho}(\bar{\zeta})] \le c$. Furthermore, there exists a sequence of homogeneous local strategies that asymptotically achieves the optimal error exponent.

The proof of Theorem 7.2 is similar to the proof of Theorem 7.1. The main difference is that we need to replace the transmission policies $\gamma_{i,V}$ with generalized policies $\bar{\zeta}_i$, and eliminate the censoring policies. However, with generalized policies, an extension to the case where the R_i are stationary and ergodic is not apparent (in contrast to the results of Section 7.6).

7.5.3 Unknown Side-Information

So far, we have been assuming that even in the case of no cooperation (local strategies), the fusion center has access to the entire vector $R^{(n)}$. We will now consider the case of no cooperation when the fusion center does not have access to $R^{(n)}$. Thus, the only information available at the fusion center is $V^{(n)} = (V_1,\ldots,V_n)$, the identity

of the sensors that are censoring, and the messages of the sensors that do not censor. (Note that just the act of censoring provides some information to the fusion center.) Reference [27] considers a setting in which (when translated to our framework) we have R_i equal to the local likelihood ratio $l_i(X_i)$ of the measurement, $\rho(r, \gamma) \equiv 1$, and the transmission policy is $\gamma_i(X_i, R_i) = X_i$. Reference [27] shows that for any fixed n, it is optimal to choose the censoring regions to be intervals, i.e., if $l_i(X_i)$ falls within some interval $(t_{i,1}, t_{i,2})$, then the sensor does not send its measurement to the fusion center. Note that [27] assumes only that the measurements X_i are independent, but even when they have identical distributions, each sensor uses a different censoring interval. Optimizing over $(t_{i,1}, t_{i,2})$, for all i, can be a daunting task even if the number of sensors n is moderate [27]. Hence it is of interest to examine whether the problem simplifies when the variables R_i are i.i.d. and n is large.

From our discussion in Section 7.5.1, we expect that homogeneous local censoring strategies are asymptotically optimal. This is indeed the case if we assume that the fusion center knows each sensor's policy. For example, $V^{(n)}$ can be determined beforehand, and made known at every sensor and the fusion center, while the fusion center has a table of all the censoring policies employed by the sensors. For $j = 0, 1$, let υ_j^γ be the distribution of $Y_1 = \gamma(X_1, R_1)$, under hypothesis H_j. Let

$$\tilde{I}(\gamma) = \mathbb{E}_0\Big[\log \frac{\mathrm{d}\upsilon_0^\gamma}{\mathrm{d}\upsilon_1^\gamma}\Big].$$

We have the following result.

Proposition 7.2. *Suppose Assumptions 7.1 and 7.2 hold and that the random variables R_i are i.i.d., under either hypothesis. Suppose that the fusion center knows V but not $R^{(n)}$, and that we are restricted to local strategies. Then, homogeneous strategies are asymptotically optimal as $n \to \infty$, for every $\alpha \in (0, 1)$. Furthermore, the optimal exponent is equal to $\hat{\lambda}(c)$, defined by*

$$\hat{\lambda}(c) = -\sup_{\pi \in \Pi(c)} \mathbb{E}_0\Big[\xi_{V_1}(R_1)\tilde{I}(\gamma_{V_1}) + (1 - \xi_{V_1}(R_1)) \log \frac{\mathbb{P}_0(\xi_{V_1}(R_1) = 0 \mid V_1)}{\mathbb{P}_1(\xi_{V_1}(R_1) = 0 \mid V_1)}\Big].$$

Proof. (Outline) We first note that using Assumption 7.2, and an argument as in Lemma 7.1, there exists $b \in (0, \infty)$, such that for all $\gamma \in \Gamma$ we have $\mathbb{E}_0[\log^2 \frac{\mathrm{d}\upsilon_0^\gamma}{\mathrm{d}\upsilon_1^\gamma}] \leq b$.

In the current setting, a censoring decision can be viewed as a transmission of a special symbol to the fusion center. We redefine Z_i so that

$$Z_i = -\xi_{i,V_i}(R_i) \log \frac{\mathrm{d}\upsilon_0^\gamma}{\mathrm{d}\upsilon_1^\gamma}(Y_i) - (1 - \xi_{i,V_i}(R_i)) \log \frac{\mathbb{P}_0(\xi_{i,V_i}(R_i) = 0 \mid V_i)}{\mathbb{P}_1(\xi_{i,V_i}(R_i) = 0 \mid V_i)}.$$

We first check the inequality $\liminf_{n\to\infty} \mathbb{E}_0[\Lambda_n] \geq \hat{\lambda}(c)$, which is obtained as in Lemma 7.4. The rest of the proof proceeds as in Section 7.4. □

7.6 The Ergodic Case

We now consider the case where $(R_i)_{i=1}^\infty$ is a stationary and ergodic sequence, and each R_i has the same distribution under either hypothesis. This case is of interest, because in many situations, the side-information at the sensors is correlated. For example, in the sensor network described in Example 7.1 of Section 7.1, if the sensors are geographically densely co-located, then we would expect the fading channels from the sensors to the fusion center to have correlated characteristics. Note also that in that example, the side-information does not provide any information on the true hypothesis.

We now assume that $\mu_0 = \mu_1$. [1] We have the following result, which shows that cooperation is unnecessary in the asymptotic regime of large n and small α.

Theorem 7.3. *Suppose $(R_i)_{i=1}^\infty$ is a stationary and ergodic sequence, $\mu_0 = \mu_1$, and Assumptions 7.1 and 7.2 hold. Then,*

$$\lim_{\alpha\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \beta_n^*(c, \alpha) = - \sup_{\pi \in \Pi(c)} \mathbb{E}_0\big[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})\big].$$

The proof of the above theorem is similar to that in Section 7.4. The proof of the lower bound in Section 7.4.1 still holds. For the upper bound, we require the

[1] One of the reasons for this assumption is that the asymptotic KL rate of the stochastic process $(R_i)_{i\geq 1}$ may not exist [54].

following result.

Lemma 7.7. *Suppose that $(R_i)_{i=1}^{\infty}$ is a stationary ergodic sequence, and that Assumptions 7.1 and 7.2 hold. Then under hypothesis H_0, for any homogeneous local strategy involving a common local policy π, we have $S_n/n \to \lambda(\pi)$ in probability.*

Proof. We have

$$\begin{aligned}\frac{S_n}{n} &= \frac{1}{n}\sum_{i=1}^{n} Z_i \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_0[Z_i \mid R_i] + \frac{1}{n}\sum_{i=1}^{n} \big(Z_i - \mathbb{E}_0[Z_i \mid R_i]\big). \end{aligned} \tag{7.17}$$

Since the sequence $(R_i)_{i\geq 1}$ is stationary and ergodic, the first term $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_0[Z_i \mid R_i]$ on the R.H.S. of (7.17) converges in probability to $\lambda(\pi)$ (cf. Birkhoff's Ergodic Theorem [46]). For the second term, we have for each $\epsilon > 0$,

$$\begin{aligned}&\mathbb{P}_0\Big(\Big|\sum_{i=1}^{n} \big(Z_i - \mathbb{E}_0[Z_i \mid R_i]\big)\Big| > n\epsilon\Big) \\ &\leq \frac{1}{n^2\epsilon^2}\mathrm{var}\Big(\sum_{i=1}^{n} \big(Z_i - \mathbb{E}_0[Z_i \mid R_i]\big)\Big) \\ &= \frac{1}{n^2\epsilon^2}\mathbb{E}_0\Big[\mathrm{var}\Big(\sum_{i=1}^{n} \big(Z_i - \mathbb{E}_0[Z_i \mid R_i]\big)\Big| R^{(n)}\Big)\Big] &(7.18)\\ &= \frac{1}{n^2\epsilon^2}\mathbb{E}_0\Big[\sum_{i=1}^{n} \mathrm{var}\big(Z_i - \mathbb{E}_0[Z_i \mid R_i] \mid R^{(n)}\big)\Big] &(7.19)\\ &\leq \frac{1}{n^2\epsilon^2}\mathbb{E}_0\Big[\sum_{i=1}^{n} a(R_i)\Big] &(7.20)\\ &= \frac{1}{n\epsilon^2}\mathbb{E}_0[a(R_1)], \end{aligned}$$

where (7.18) follows because $\mathbb{E}_0[Z_i \mid R_i] = \mathbb{E}_0[Z_i \mid R^{(n)}]$, and (7.19) follows because given $R^{(n)}$, the Z_i are independent. The last inequality (7.20) follows from Lemma 7.1. Therefore, as $n \to \infty$, the second term on the R.H.S. of (7.17) converges in probability to 0, and the lemma is proved. □

To complete the proof of Theorem 7.3, we proceed as in Section 7.4.2, except that

we fix an $\epsilon > 0$ and consider a homogeneous local strategy involving a common local policy $\pi^\epsilon \in \Pi(c)$ that satisfies $\lambda(\pi^\epsilon) \leq \lambda^*(c) + \epsilon$. With this strategy, from Lemma 7.7, $S_n/n \to \lambda(\pi^\epsilon)$ in probability under hypothesis H_0. Hence, we have the same result as (7.14) with π^ϵ replacing π^n, and η replacing $n^{-1/4}$, for some fixed $\eta > 0$. Corresponding changes are made in Lemma 7.6. The proof of Theorem 7.3 is now complete.

7.7 Optimal Censoring

To find an optimal common local policy π, we need to maximize

$$f(\pi) = \mathbb{E}_0[\xi_{V_1}(R_1) I(R_1, \gamma_{V_1})]$$

over all $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0} \in \Pi(c)$, i.e., over all π that satisfy

$$\rho(\pi) = \mathbb{E}_0[\xi_{V_1}(R_1) \rho(R_1, \gamma_{V_1})] \leq c.$$

We now show that it is sufficient to consider local policies that randomize between only two pure local policies. In particular, each sensor need only use an extra bit to communicate to the fusion center which policy it has chosen.

Suppose that a common local policy π has been fixed, including the range $\mathcal{V}_0$ of the randomization variable V_1, except that the distribution μ of V_1 is left unspecified. Let π_v be the pure local policy obtained when $V_1 = v$. To optimize the distribution of V_1, we have to maximize $\int f(\pi_v)\, \mu(\mathrm{d}v)$, subject to $\int \rho(\pi_v)\, \mu(\mathrm{d}v) \leq c$, over all measures μ on $\mathcal{V}_0$. If $\mathcal{V}_0$ were finite, this would be a linear programming problem over the unit simplex, together with one additional constraint. As is well known, the optimum would be attained on an edge of the feasible set, that is, there would exist an optimal μ whose support consists of at most two points [55]. The lemma that follows states that the optimality of two-point distributions remains valid even when $\mathcal{V}_0$ is infinite (except that the optimum need not be attained), and establishes our claim that we only need to consider local policies that randomize between two pure local policies.

(As in the rest of the chapter, we omit the standard measurability conditions that are needed in this lemma.) The proof is provided in Section 7.11 for completeness.

Lemma 7.8. *Let $\mathcal{M}$ be the set of probability measures on a set $\mathcal{V}_0$, and let f, g be given nonnegative functions from $\mathcal{V}_0$ into $[0,\infty)$. Then,*

$$\sup\Big\{\int f(v)\,\mu(\mathrm{d}v) : \mu\in\mathcal{M},\ \int g(v)\,\mu(\mathrm{d}v)\le c\Big\} \tag{7.21}$$

$$= \sup\{uf(v_1)+(1-u)f(v_2)) : u\in[0,1], v_1,v_2\in\mathcal{V}_0, ug(v_1)+(1-u)g(v_2)\le c\}, \tag{7.22}$$

Furthermore, if the supremum in (7.21) is finite and is attained, then the supremum in (7.22) is also attained.

We close with a characterization of an optimal local censoring policy $(\xi_v)_{v\in\mathcal{V}_0}$, given that a local transmission policy $(\gamma_v)_{v\in\mathcal{V}_0}$ and the distribution of V_1 have been fixed. Let $\bar{\xi}(r,v)=\xi_v(r)$, $\tilde{I}(r,v)=I(r,\gamma_v)$ and $\tilde{\rho}(r,v)=\rho(r,\gamma_v)$. We then need to optimize $\mathbb{E}_0\big[\bar{\xi}(R_1,V_1)\tilde{I}(R_1,V_1)\big]$ over all $\bar{\xi}:\mathcal{R}\times\mathcal{V}_0\mapsto\{0,1\}$ that satisfy $\mathbb{E}_0\big[\bar{\xi}(R_1,V_1)\tilde{\rho}(R_1,V_1)\big]\le c$. It is an easy exercise (whose proof is omitted) to show that there exists an optimal censoring policy of the following form. There is a threshold t such that $\bar{\xi}(r,v)=1$ if $\tilde{I}(r,v)/\tilde{\rho}(r,v)>t$ and $\bar{\xi}(r,v)=0$ if $\tilde{I}(r,v)/\tilde{\rho}(r,v)<t$. Randomization is only used to make a censoring decision when $\tilde{I}(r,v)/\tilde{\rho}(r,v)=t$, and a binary randomization variable at each sensor suffices. This is a solution of the "water-filling" type, whereby the uncensored "states" (r,v) are chosen starting with those with a higher value of $\tilde{I}(r,v)/\tilde{\rho}(r,v)$, and continuing until the resource constraint is met. Note also that for a pure transmission policy γ, the relevant ratio is $I(r,\gamma)/\rho(r,\gamma)$, which has the intuitive interpretation of information content per unit resource consumed.

7.8 Applications

In this section, we revisit Examples 7.1 and 7.2 from Section 7.1, and illustrate the form of an optimal censoring policy. Given our focus on the censoring policy, we will assume that all sensors send their observations "in the clear" to the fusion center,

i.e., the pure local transmission policy $\gamma(x, r) = x$ is employed at all the sensors. Accordingly, γ will be suppressed in our notation below.

7.8.1 Fading Channels

We will focus on a special case of the problem posed in Example 7.1. We consider a wireless sensor network transmitting measurements to a fusion center over slowly fading Gaussian channels. We assume that $\rho(r) = 1$ for all r, so that we are only concerned with restricting the number of sensors transmitting. Depending on the condition of the channel, we will naturally want to allow sensors to transmit over good channels and allow sensors that have bad channels to censor. This raises the issue of identifying the key parameters of the channel on the basis of which censoring decisions should be made.

Suppose that

$$H_0 : X_i \sim N(-m, \sigma^2),$$
$$H_1 : X_i \sim N(m, \sigma^2),$$

and that the fusion center receives

$$Y_i = Q_i X_i + W_i,$$

where Q_i is the fading coefficient, with a known density $g(\cdot)$, and $W_i \sim N(0, \sigma_i^2)$. Assume that the channel characteristics $R_i = (Q_i, \sigma_i)$ are stationary and ergodic, with the same stationary distribution under either hypotheses. This can be used to model the case where sensors are placed in a line so that an ergodic assumption on the distribution of the variables R_i is reasonable. (Random (i.i.d.) placement of sensors is another example.) Since this is a slow fading channel, each sensor can measure R_i. From Theorem 7.3, the important design parameter is

$$I(Q_i, \sigma_i) = \mathbb{E}_0\Big[-\frac{(Y_i + Q_i m)^2}{2(Q_i^2\sigma^2 + \sigma_i^2)} + \frac{(Y_i - Q_i m)^2}{2(Q_i^2\sigma^2 + \sigma_i^2)}\Big]$$

$$= \frac{2Q_i^2 m^2}{Q_i^2\sigma^2 + \sigma_i^2}.$$

According to Theorem 7.3 and the discussion in the previous section, we want to censor when $|Q_i/\sigma_i|$ is small. Thus, an asymptotically optimal censoring policy (where censoring is based on the channel characteristics) is of the form

$$\xi(R_i) = \begin{cases} 1, & \text{if } \left|\frac{Q_i}{\sigma_i}\right| > \eta, \\ 0, & \text{otherwise,} \end{cases}$$

where η depends on the value of c and the density $g(\cdot)$. Note that randomization when $|Q_i/\sigma_i| = \eta$ is unnecessary, because this event happens with zero probability.

7.8.2 Detection of Spatial Signals

Consider n sensor nodes, placed uniformly and independently in $[-1, 1]$, with the fusion center at the origin, for the purpose of detecting a spatial signal. Consider the hypotheses

$$H_0 : X_i = s_0(R_i) + W_i \quad \forall i,$$
$$H_1 : X_i = s_1(R_i) + W_i \quad \forall i,$$

where each $s_j(\cdot)$ is a known spatial signal, and $W_i \sim N(0, \sigma^2)$ is Gaussian noise. When sensor i sends its measurement X_i to the fusion center, it consumes power $\rho(R_i)$ (assumed positive), which depends on its relative position to the fusion center. We constrain the overall average power to be less than a given positive constant c. From Theorem 7.1, each sensor should use a common local censoring policy $(\xi_v)_{v\in\mathcal{V}_0}$, obtained by maximizing $\mathbb{E}_0[\xi_{V_1}(R_1)I(R_1)]$ subject to $\mathbb{E}_0[\xi_{V_1}(R_1)\rho(R_1)] \le c$. According to the discussion in Section 7.7, a sensor i should be censored when $I(R_i)/\rho(R_i)$ is below a threshold. As a specific illustration, let $s_0(r) = 1 - r$ and $s_1(r) = 1 + r$. Then

$$I(r) = \mathbb{E}_0\Big[-\frac{1}{2\sigma^2}\big((X_1 - 1 + r)^2 - (X_1 - 1 - r)^2\big)\Big]$$

$$= \frac{2r^2}{\sigma^2}.$$

Suppose $\rho(r) = 1 + a|r|^d$, where $2 \leq d \leq 4$. (This is in line with standard models used for power decay in a wireless network, see [49]. The unit cost is due to the power used to make the measurement X_i.) Then, we have

$$\frac{I(r)}{\rho(r)} = \frac{2r^2}{\sigma^2(1 + a|r|^d)}.$$

A specific case is shown in Figure 7-1. We have taken $a = 1$, $d = 2$, $\sigma = 1$, and a constraint of $c = 19/12$. As shown, only sensors at a large enough distance from the fusion center should transmit their measurements X_i.

7.9 The Bayesian Problem with Local Censoring

We now consider the decentralized detection problem with censoring in the Bayesian context. Let the prior probability of H_j be $q_j > 0$, for $j = 0, 1$. We define $\mathbb{P} = q_0\mathbb{P}_0 + q_1\mathbb{P}_1$ and let $\mathbb{E}$ be the expectation operator w.r.t. $\mathbb{P}$. As in the Neyman-Pearson case, we allow sensors to use randomized sensor policies. In contrast to unconstrained Bayesian problems, simple examples show that randomization results in improved performance when the number of sensors n is finite. However, we will show that for the asymptotic problem considered here (no cooperation), randomization is unnecessary. In the process, we will also characterize the optimal error exponent and associated local policies.

A strategy $\pi^{(n)}$ is admissible if (7.1) is satisfied, with $\mathbb{E}$ replacing $\mathbb{E}_0$. For any admissible strategy $\pi^{(n)}$, let $P_{e,n}(\pi^{(n)})$ denote the resulting probability of error at the fusion center. We will always assume that the fusion center uses an optimal fusion rule, namely the MAP rule. Let $P^*_{e,n}$ be the infimum of $P_{e,n}(\pi^{(n)})$ over all admissible strategies. We are interested in finding asymptotically optimal local strategies that achieve

$$\liminf_{n\to\infty} \frac{1}{n} \log P^*_{e,n}.$$

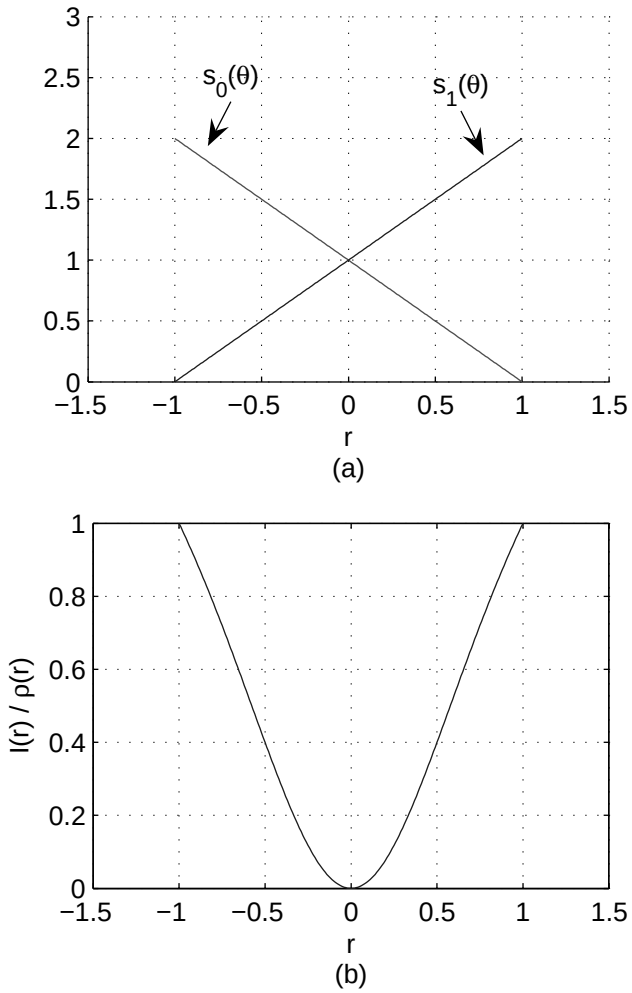

Figure 7-1: The first graph shows the spatial signals plotted as a function of sensor location. Let $c = 19/12$. The second graph shows a plot of $I(r)/\rho(r)$. A sensor is censored unless its location is in $[-1, -0.5]$ or $[0.5, 1]$.

Before we launch into the analysis, let us consider a simple example that shows that cooperation among the sensors is *strictly* better than using local strategies.

Example 7.7. *Suppose that the random variables R_i belong to $\{0, 1\}$, are i.i.d. under either hypothesis, and that $\mathbb{P}_j(R_1 = 0) = \mathbb{P}_j(R_1 = 1) = 1/2$, for $j = 0, 1$. We assume that all sensors are restricted to using the transmission function $\gamma(x, r) = x$.*

We assume that the distribution of X_1 under the two hypotheses is the same when $R_1 = 0$, but different when $R_1 = 1$. Thus, it is only those sensors with $R_1 = 1$ that have useful information to transmit. Under mild conditions (including the special case

where X_1 has a finite range), it is a well-known consequence of the Chernoff bound that if exactly m sensors have $R_i = 1$ and transmit to the fusion center, the probability of error is of the form $e^{m\Lambda^*}g(m)$, where Λ^* is a negative constant determined by the distributions of X_1 and where $g(m)$ satisfies $\lim_{m\to\infty}(\log g(m))/m = 0$. In particular, for every $\epsilon > 0$, we can find some positive a, b, such that $ae^{-\epsilon m} \le g(m) \le be^{\epsilon m}$.

Let $\rho(r,\gamma) = 1$ and $c = 1/4$. Thus, the resource constraint (7.1) becomes $\mathbb{E}[N] \le n/4$, where N is the (random) number of sensors that are not censored.

Assume for simplicity that $n/4$ is integer. Consider the following cooperative censoring strategy.

1. *If $\sum_{i=1}^n R_i \le n/4$, sensor i transmits if only if $R_i = 1$.*

2. *If $\sum_{i=1}^n R_i > n/4$, among those sensors with $R_i = 1$, arbitrarily choose $n/4$ of them to transmit.*

Using the Chernoff bound, we have $\mathbb{P}(N < n/4) \le e^{-dn}$ for some positive constant d. Let $P_{e,n}$ be the probability of error at the fusion center. We have

$$\begin{aligned} P_{e,n} &\le \mathbb{P}(N < n/4) + \mathbb{P}(N = n/4)e^{\Lambda^* n/4}g(n/4) \\ &\le e^{-dn} + be^{(\Lambda^*+\epsilon)n/4}. \end{aligned}$$

Suppose that when $R_1 = 1$, the distribution of X_1 is such that $-d < \Lambda^*$, which is certainly possible. Since $\epsilon > 0$ is arbitrary, we obtain $\liminf_{n\to\infty}(1/n)\log P_{e,n} \le \Lambda^*/4$.

Consider now a local and pure censoring strategy. In a best strategy of this kind, every sensor with $R_1 = 0$ is censored, and $\mathbb{E}[N] = n/4$. The only way to achieve this is as follows: $n/2$ sensors are always censored; the remaining sensors are censored if and only if $R_1 = 0$. Thus, N is binomial with parameters $n/2$ and $1/2$. After averaging over all possible values of N, the probability of error satisfies

$$\begin{aligned} P_{e,n} &= \mathbb{E}[e^{\Lambda^* N}g(N)] \ge a\mathbb{E}[e^{(\Lambda^*-\epsilon)N}] \\ &= a\Big(\frac{1}{2} + \frac{1}{2}e^{\Lambda^*-\epsilon}\Big)^{n/2}. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we obtain $\liminf_{n\to\infty}(1/n)\log P_{e,n} \geq (1/2)\log((1+e^{\Lambda^*})/2)$. This is strictly greater than $\Lambda^*/4$, which shows that the cooperative strategy constructed earlier has a better error exponent. Later on, we show that randomization cannot improve performance, bringing us to conclude that cooperative strategies can be strictly better than local ones.

The essence of this example is that in the local case, we have much less control over the tails of the distribution of N; the possibility of N having a large deviation results in a deterioration in the error exponent. □

In general, optimal cooperative strategies are difficult to find. As the cooperative strategy may also be practically infeasible, we will focus our attention on finding an optimal *local* strategy. For the remainder of this chapter, the words "policy" and "strategy" will always mean "local policy" and "local strategy," respectively.

7.9.1 Notation and Assumptions

Let $\ell_{10}(\cdot \mid r)$ be the Radon-Nikodym derivative of the measure $\nu_1(\cdot \mid r)$ w.r.t. $\nu_0(\cdot \mid r)$. For $\gamma \in \Gamma$, let $\ell_{10}^{\gamma}(\cdot \mid r)$ be the Radon-Nikodym derivative of the measure $\nu_1^{\gamma}(\cdot \mid r)$ w.r.t. $\nu_0^{\gamma}(\cdot \mid r)$. Let $\ell_\mu(R_1) = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(R_1)$, and for any $s \in [0,1]$, and pure local transmission policy $\gamma \in \Gamma$, let

$$\Lambda(s, r, \gamma) = \log \mathbb{E}_0\big[e^{s \log \ell_{10}^{\gamma}(Y_1|r)} \mid R_1 = r\big].$$

Finally, for a randomized local policy $\pi = (\xi_v, \gamma_v)_{v\in\mathcal{V}_0}$, let

$$\begin{aligned}\Phi(s,\pi) &= \log \mathbb{E}_0\big[e^{s(\xi_V(R_1)\log \ell_{10}^{\gamma_V}(Y_1|R_1) + \log \ell_\mu(R_1))}\big] \\ &= \log \mathbb{E}_0\Big[\Big(\xi_V(R_1)e^{\Lambda(s,R_1,\gamma_V)} + 1 - \xi_V(R_1)\Big)\ell_\mu^s(R_1)\Big]. \end{aligned} \tag{7.23}$$

For a policy $\pi = (\xi_v, \gamma_v)_{v\in\mathcal{V}_0}$, if $\xi_v(r) = 1$ for all $r \in \mathcal{R}$ and $v \in \mathcal{V}_0$, we will write $\pi = (1, \gamma_v)_{v\in\mathcal{V}_0}$. We will make the following assumptions.

Assumption 7.3.

(i) Conditioned on either hypothesis, the random variables R_i are i.i.d. Furthermore, μ_0 and μ_1 are equivalent measures.

(ii) The (regular) conditional distributions $\nu_0(\cdot \mid r)$ and $\nu_1(\cdot \mid r)$ are equivalent for every $r \in \mathcal{R}$.

(iii) We are restricted to local strategies.

(iv) We have $\rho(r, \gamma) > 0$, for every $r \in \mathcal{R}$ and pure policy $\gamma \in \Gamma$.

(v) There exists an open interval $I = (\tau_1, \tau_2) \subset [0, 1]$, such that for all pure policies $\gamma \in \Gamma$, we have $\arg\min_{s\in[0,1]} \Phi(s, (1, \gamma)) \in I$. Furthermore, for $k = 1$ and 2, the following holds:

$$\left|\mathbb{E}_0\left[\left(\ell_{10}(X_1|R_1)\ell_\mu(R_1)\right)^{\tau_k} \log\left(\ell_{10}(X_1|R_1)\ell_\mu(R_1)\right)\right]\right| < \infty.$$

(vi) For the same open interval I as in (v) above, there exists a $b \in (0, \infty)$ such that

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathbb{E}_0[\ell_\mu^s(R_1)] \leq b,$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathbb{E}_0\left[(\ell_{10}^\gamma(Y_1 \mid R_1)\ell_\mu(R_1))^s\right] \leq b,$$

for all $s \in I$ and all $\gamma \in \Gamma$.

Note that there is little loss of generality in imposing Assumption 7.3(iv). Indeed, if $\rho(r, \gamma) = 0$ for some r and some pure policy γ, then we can always transmit Y_i when $R_i = r$, without incurring any cost. So instead of censoring in the state $R_i = r$, the sensor can always choose to transmit using this particular γ.

Assumptions 7.3(v)-(vi) are required for the same technical reasons as in [22], which also gives rather general conditions under which they are satisfied.[2] In general, the open interval I can be taken to be $(0, 1)$. Indeed, it can be shown that, under Assumptions 7.3(i)-(ii), and for any pure transmission policy $\gamma \in \Gamma$, the minimizer s^* of $\min_{s\in[0,1]} \Phi(s, (1, \gamma))$ is in the interior of $[0, 1]$. If we take $I = (0, 1)$, Assumption

[2] Although [22] deals with the case of a finite transmission alphabet $\mathcal{Y}$, the results therein can be easily generalized to the case of infinite alphabets.

7.3(v) reduces to the condition that the KL divergences $-\mathbb{E}_0[\log\big(\ell_{10}(X_1 \mid R_1)\ell_\mu(R_1)\big)]$ and $\mathbb{E}_1[\log\big(\ell_{10}(X_1 \mid R_1)\ell_\mu(R_1)\big)]$ are bounded. But we only need the weaker version of Assumptions 7.3(v)-(vi), as stated. This allows us to include cases where Assumptions 7.3(v)-(vi) hold automatically. For example, if Γ is a finite set of transmission policies, the interval I only needs to include certain, finitely many, values of s, and we can choose $I = (a, b)$, where $0 < a < b < 1$. Then, it is easy to show that under Assumptions 7.3(i)-(ii), Assumptions 7.3(v)-(vi) hold automatically. We will make use of this fact in Sections 7.10.3 and 7.10.4. Another sufficient condition for Assumptions 7.3(v)-(vi) is Assumption 7.3(i)-(ii) together with an assumption similar to Assumption 7.2 (see Proposition 3 of [22]).

The main reason for introducing Assumption 7.3 is the following lemma, which is proved in Section 7.11.

Lemma 7.9. *Suppose that Assumption 7.3 holds. Then, there exists some* $b_1 \in (0, \infty)$ *such that for all* $s \in I$ *and for all* π, $|\Phi(s, \pi)| \leq b_1$, $\left|\frac{\mathrm{d}}{\mathrm{d}s}\Phi(s, \pi)\right| \leq b_1$ *and* $\frac{\mathrm{d}^2}{\mathrm{d}s^2}\Phi(s, \pi) \leq b_1$.

We record a result from [22], based on the results in [56], which will underlie the rest of our development. This result can also be obtained from Theorem 1.3.13 of [40]. The result states that, if the conclusion of Lemma 7.9 holds, then

$$\begin{aligned}\frac{1}{n}\log P_{e,n}(\pi^{(n)}) &= \min_{s\in I}\frac{1}{n}\sum_{i=1}^{n}\Phi(s, \pi_i) + o(1)\\ &= \min_{s\in[0,1]}\frac{1}{n}\sum_{i=1}^{n}\Phi(s, \pi_i) + o(1),\end{aligned} \tag{7.24}$$

where $o(1)$ stands for a term that vanishes as $n \to \infty$, uniformly over all sequences $\pi^{(n)}$. Given this result, we can just focus on the problem of optimizing the R.H.S. of (7.24), while ignoring the $o(1)$ term.

7.9.2 Optimal Strategy

In this subsection, we prove that asymptotic optimality can be obtained by dividing the sensors into two groups with sensors in each group using a common pure policy.

Theorem 7.4. *Under Assumption 7.3,*

$$\lim_{n\to\infty} \frac{1}{n} \log P^*_{e,n} = \inf \min_{s\in[0,1]} \{u\Phi(s,\pi_1) + (1-u)\Phi(s,\pi_2)\} \tag{7.25}$$

where the infimum is taken over all $u \in [0,1]$, and all pure policies π_1 and π_2 that satisfy $u\rho(\pi_1) + (1-u)\rho(\pi_2) \le c$.

Proof. Fix some $s \in [0,1]$. Let $\bar{\psi}_1, \ldots, \bar{\psi}_n$ be some (possibly randomized) policies. Let $\psi_{i,v}$ be the pure policy obtained when $V_i = v$. Using the definition (7.23) of $\Phi(s, \bar{\psi}_i)$ and Jensen's inequality, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \Phi(s,\bar{\psi}_i) &= \frac{1}{n}\sum_{i=1}^{n} \log\big(\exp(\Phi(s,\bar{\psi}_i))\big) \\
&= \frac{1}{n}\sum_{i=1}^{n} \log \mathbb{E}\big[\exp(\Phi(s,\psi_{i,V_i}))\big] \\
&\ge \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\big[\Phi(s,\psi_{i,V_i})\big]. \qquad (7.26)
\end{aligned}$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n} \rho(\bar{\psi}_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\big[\rho(\psi_{i,V_i})\big].$$

Note that taking the average over all i in the above expressions is equivalent to taking an expectation over a uniformly chosen random i. Let $U = (J, V_J)$, where J is chosen uniformly over $\{1,\ldots,n\}$. We minimize the R.H.S. of (7.26), $\mathbb{E}[\Phi(s,\psi_U)]$, subject to the constraint $\mathbb{E}[\rho(\psi_U)] \le c$. Applying Lemma 7.8, with $f(u) = \Phi(s,\psi_u)$ and $g(u) = \rho(\psi_u)$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} \Phi(s,\bar{\psi}_i) \ge \inf\{u\Phi(s,\pi_1) + (1-u)\Phi(s,\pi_2)\},$$

where the infimum is taken over all $u \in [0,1]$, and pure policies π_1, π_2, satisfying the

resource constraint $u\rho(\pi_1) + (1-u)\rho(\pi_2) \leq c$. Hence from (7.24),

$$\liminf_{n\to\infty} \frac{1}{n} \log P_{e,n}^* \geq \inf \min_{s\in[0,1]} \{u\Phi(s,\pi_1) + (1-u)\Phi(s,\pi_2)\}. \tag{7.27}$$

To achieve the lower bound, suppose that u and π_i, $i = 1, 2$, attain the infimum in (7.27) to within $\epsilon > 0$ and that s^* is a minimizing value of s in (7.27). We assign $\lfloor un \rfloor$ sensors to use policy π_1, and $\lfloor (1-u)n \rfloor$ sensors to use policy π_2. We censor any remaining sensor. Then, from (7.24),

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log P_{e,n}^* &\leq u\Phi(s^*,\pi_1) + (1-u)\Phi(s^*,\pi_2) \\ &\leq \inf \min_{s\in[0,1]} \{u\Phi(s,\pi_1) + (1-u)\Phi(s,\pi_2)\} + \epsilon,\end{aligned}$$

and taking $\epsilon \to 0$ completes the proof. □

Let us remark that similar results are easily obtained for the case where the side-information is not transmitted to the fusion center (cf. Section 7.5.3).

7.9.3 Characterization of the Optimal Exponent

In this section and the next, we will consider the case where $\mathcal{R}$ is finite, for two reasons. First, in many practical cases, because of the limited channel between each sensor and the fusion center, the side-information can be assumed to take values from a finite alphabet. Second, when $\mathcal{R}$ is finite, the analysis is simplified and results in a simple form for the censoring policies. So without loss of generality, we will take $\mathcal{R} = \{1, 2, \ldots, |\mathcal{R}|\}$. Let $p_{jr} = \mathbb{P}_j(R_1 = r)$ and $p_r = q_0 p_{0r} + q_1 p_{1r}$.

Let us fix two pure local transmission policies γ_1 and γ_2. Let

$$K_s(c, \{\gamma_1, \gamma_2\}) = \inf_{\xi_1,\xi_2,u} \{u\Phi(s,(\xi_1,\gamma_1)) + (1-u)\Phi(s,(\xi_2,\gamma_2)))\}, \tag{7.28}$$

where the infimum is taken over all $u \in [0,1]$ and pure censoring policies ξ_1, ξ_2 that satisfy

$$u\mathbb{E}[\xi_1(R_1)\rho(R_1,\gamma_1)] + (1-u)\mathbb{E}[\xi_2(R_2,\gamma_2)\rho(R_2,\gamma_2)] \leq c.$$

From Theorem 7.4 and under Assumption 7.3, we have

$$\lim_{n\to\infty} \frac{1}{n} \log P^*_{e,n} = \inf_{\gamma_1,\gamma_2\in\Gamma} \min_{s\in[0,1]} K_s(c, \{\gamma_1, \gamma_2\}). \tag{7.29}$$

Note that given γ_1 and γ_2, the minimization in (7.28) has an optimal solution. (This is because $\mathcal{R}$ is finite, and therefore there are only finitely many possible pure censoring policies.) Let z_1 be the value of $\mathbb{E}[\xi_1(R_1)\rho(R_1,\gamma_1)]$ in such an optimal solution. It follows that ξ_1 must minimize $\Phi(s,(\xi_1,\gamma_1))$ (and therefore $\exp\big(\Phi(s,(\xi_1,\gamma_1))\big)$ as well), subject to the constraint $\mathbb{E}[\xi_1(R_1)\rho(R_1,\gamma_1)] \le z_1$. Note that $\exp\big(\Phi(s,(\xi_1,\gamma_1))\big)$ is equal to

$$\mathbb{E}_0[\xi(R_1)h_s(R_1,\gamma)] + \mathbb{E}_0[\ell^s_\mu(R_1)],$$

where $h_s(r,\gamma) = (e^{\Lambda(s,r,\gamma)} - 1)\ell^s_\mu(r) \le 0$.

We can now give a characterization of the optimal ξ_1, similar to the one at the end of Section 7.7. For any $r \in \mathcal{R}$, let

$$m_{s,\gamma}(r) = \frac{p_{0r}h_s(r,\gamma)}{p_r\rho(r,\gamma)}.$$

Proposition 7.3. *Suppose Assumption 7.3 holds. Suppose that $\mathcal{R}$ is finite and that optimal choices of u, s, γ_1, γ_2 have been fixed. Then, there exist thresholds t_1, t_2 such that the corresponding optimal censoring functions ξ_1, ξ_2 satisfy the following: for each $r \in \mathcal{R}$, if $m_{s,\gamma_i}(r) < t_i$, then $\xi_i(r) = 1$, otherwise $\xi_i(r) = 0$.*

7.10 Special Cases and Examples for the Bayesian Problem

We now examine some special cases that will lead to simplified versions of Theorem 7.4.

7.10.1 No Side-Information

In this subsection, we consider the case of no side-information, which is equivalent to having $\mathcal{R} = \{1\}$. Accordingly, r will be suppressed from our notation below. For example, the cost incurred by a sensor making a measurement and transmitting it via a transmission function $\gamma \in \Gamma$ is denoted by $\rho(\gamma)$. We will show that when c in the resource constraint (7.1) is sufficiently small, we can restrict all uncensored sensors to use the same policy.

Note that there are only two possible pure censoring policies, $\xi = 0$ and $\xi = 1$. In the absence of side-information, the likelihood ratio $\ell_\mu(R_1)$ is identically equal to 1. Using the definition of $\Phi(s, (\xi, \gamma))$, for $\xi = 1$ and $\xi = 0$, respectively, we obtain $\Phi(s, (1, \gamma)) = \Lambda(s, \gamma) \leq 0$, and $\Phi(s, (0, \gamma)) = 0$. Let $\Lambda^*(\gamma) = \min_{s \in [0,1]} \Lambda(s, \gamma)$.

Corollary 7.1. *Suppose that Assumption 7.3 holds, and* $0 < c \leq \inf_{\gamma \in \Gamma} \rho(\gamma)$. *Then, in the absence of side-information, the optimal exponent is equal to* $\inf_{\gamma \in \Gamma} c\Lambda^*(\gamma)/\rho(\gamma)$. *This remains the optimal error exponent even under the additional restriction that censored sensors are chosen in advance, and all uncensored sensors use the same policy.*

Proof. From Theorem 7.4, we know that at most two different pure policies $\pi_i = (\xi_i, \gamma_i)$, $i = 1, 2$, need be considered. Suppose that one of those policies, say π_2, involves censoring, namely $\xi_2 = 0$. Then, $\Phi(s, \pi_2) = 0$ and $\rho(\pi_2) = 0$. Clearly, the other policy should not censor, so that $\pi_1 = (1, \gamma_1)$. For any choice of γ_1, the optimal choice of u in (7.25) is to let $u = c/\rho(\gamma_1) \in [0, 1]$, leading to an exponent of $c\Lambda^*(\gamma_1)/\rho(\gamma_1)$. Optimizing over all $\gamma_1 \in \Gamma$, we obtain the claimed error exponent.

Suppose now that neither of the policies π_1 and π_2 involves censoring. Since $c \leq \rho(\gamma)$ for all $\gamma \in \Gamma$, in order to satisfy the resource constraint $u\rho(\gamma_1) + (1-u)\rho(\gamma_2) \leq c$, we must have $\rho(\gamma_1) = \rho(\gamma_2) = c$. In this case, for all $u \in [0, 1]$ and all $s \in [0, 1]$, $u\Phi(s, \pi_1) + (1 - u)\Phi(s, \pi_2) \geq \min_{i=1,2} \Lambda^*(\gamma_i) \geq \inf_{\gamma \in \Gamma} c\Lambda^*(\gamma)/\rho(\gamma)$. The corollary is now proven. □

7.10.2 Finite Transmission Policy Sets

In the next two subsections, we study problem formulations in which there is no side-information, and in which we temporarily restrict transmission policies to belong to a finite subset G of Γ. According to the discussion in Section 7.9.1, this restriction implies that under Assumptions 7.3(i)-(ii), Assumptions 7.3(v)-(vi) hold automatically. This will allow us to apply Corollary 7.1 to two problems that have been considered in [29] and [30]. Let $P^*_{e,n}(G)$ be the minimum error probability, when we are restricted to transmission policies in G.

7.10.3 Total Power Constraint

In [29], the authors consider the Bayesian problem with no side-information and a power constraint of the form

$$\sum_{i=1}^{L_A} \rho(\gamma_i) \leq A, \tag{7.30}$$

where L_A, the number of sensors is not fixed in advance. The cost $\rho(\gamma)$ is assumed to be positive for all γ.

Let G be a finite subset of Γ, and let $\mathcal{G}$ be the collection of all such subsets G. Let $P_e(G, A)$ be the minimum probability of error when using transmission policies from G that satisfy (7.30). Recall that we define $\Lambda^*(\gamma) = \min_{s\in[0,1]} \Lambda(s, \gamma)$. Reference [29] shows that

$$\inf_{G\in\mathcal{G}} \liminf_{A\to\infty} \frac{1}{A} \log P_e(G, A) = \inf_{\gamma\in\Gamma} \frac{\Lambda^*(\gamma)}{\rho(\gamma)}, \tag{7.31}$$

so that it is asymptotically optimal to have all sensors use the same transmission policy. We will re-derive (7.31) from Corollary 7.1.

To see the connection with our framework, fix a $G \in \mathcal{G}$. Note that for all $\gamma \in G$, there exists a $\delta > 0$, such that $\rho(\gamma) \geq \delta$. Under the constraint (7.30), the number of sensors that can be used is bounded by $n = \lfloor A/\delta \rfloor$. With n defined in this manner,

the constraint (7.30) is equivalent to the constraint

$$\sum_{i=1}^{n} \xi_i \rho(\gamma_i) \leq n\delta,$$

where $\xi_i \in \{0, 1\}$. Note that the limit $A \to \infty$, considered in [29], is equivalent to the limit $n \to \infty$ in our framework.

Therefore, under Assumptions 7.3(i), (ii), and (iv), Corollary 7.1 shows that the optimal error exponent is

$$\lim_{n\to\infty} \frac{1}{n} \log P^*_{e,n}(G) = \delta \inf_{\gamma \in G} \frac{\Lambda^*(\gamma)}{\rho(\gamma)}.$$

By taking the infimum over all $G \in \mathcal{G}$, we recover (7.31). This argument shows that it is asymptotically optimal to use $\lfloor A/\rho(\gamma) \rfloor$ sensors, all of which employ the same pure transmission policy γ, chosen by carrying out the minimization in the R.H.S. of (7.31).

This discussion elucidates the relationship of a power constraint (in which the number of transmitting sensors is not fixed) to our constrained censoring problem. The decentralized detection problem considered in [22] can be viewed as one where c is so large that censoring is never needed. The problem in this subsection can be viewed as one involving a very small c. In this case, one group of sensors sets $\xi \equiv 0$, and another uses the transmission policy that asymptotically achieves (7.31). In comparison, the general formulation in this chapter also gives the solution for all c, in between these two extremes.

7.10.4 Constrained Capacity

Yet another connection can be made to the problem considered in [30], which is summarized as follows. Consider a network of L sensors, z_l, $l = 1, 2, \ldots, L$. Each sensor z_l observes a sequence of measurements $\{X_{l,t} : t = 1, 2, \ldots, T\}$, and there is no side-information. All the measurements are assumed to be conditionally i.i.d. given the hypothesis, over time and across sensors. At each time t, sensor z_l sends $Y_{l,t} =$

$\gamma_l(X_{l,t})$ to the fusion center, where the transmission function is a b_l-bit quantizer, where b_l is a positive integer. Let the set of allowed b-bit transmission functions be Γ_b. We are interested in minimizing the error exponent

$$\liminf_{T\to\infty} \frac{1}{T} \log P_{e,T}(\gamma^{(L)}),$$

where $P_{e,T}(\gamma^{(L)})$ is the probability of error at the fusion center, assuming that a MAP fusion rule is used. The minimization is to be carried out over the number of sensors L and transmission strategies satisfying the overall capacity constraint,

$$\sum_{l=1}^{L} b_l \leq B, \tag{7.32}$$

where B is a given positive integer. Let us call the above problem Q1. This problem, in general, does not have a closed form solution. Reference [30] finds sufficient conditions under which using B identical sensors (sensors using the same transmission policy), each sending one bit of information, is optimal. We will apply our results to arrive at the same conditions as in [30], and also characterize the solution for special values of B.

As a first step, we will relax the constraints in problem Q1. We view each sensor z_l over the time periods $1, 2, \ldots, T$, as T different sensors $z_{l,t}$, and hence remove the constraint that all $z_{l,t}$ must use the same transmission policy γ_l. Because of (7.32), $L \leq B$. Hence, we can imagine that we are starting with $n = TB$ sensors, some of which will be censored, and rewrite (7.32) as

$$\sum_{t=1}^{T} \sum_{l=1}^{B} \xi_{l,t} b_{l,t} \leq TB, \tag{7.33}$$

where $\xi_{l,t} \in \{0, 1\}$. For each $b \in \{1, \ldots, B\}$ for which Γ_b is nonempty, consider a nonempty finite subset of Γ_b, and use G to denote the union of these subsets over b.

Let $\mathcal{G}$ be the collection of all such G. With $n = TB$, we wish to minimize

$$B \inf_{G \in \mathcal{G}} \lim_{n \to \infty} \frac{1}{n} \log P^*_{e,n}(G).$$

We will obtain an optimal solution to the latter problem, which we call problem Q2. If from that optimal solution we can derive a strategy that does not change the error exponent and yet meets the constraints that $\gamma_{l,t} = \gamma_l$ and $\xi_{l,t} = \xi_l$ for all t, then we will have found an optimal solution to problem Q1. In particular, sufficient conditions for problem Q2 to have all sensors using the same one-bit transmission policy are also sufficient for problem Q1 to have B identical one-bit sensors.

To put problem Q2 into our constrained censoring context, let $\rho(\gamma) = b$ for every $\gamma \in \Gamma_b$, and note that $c = 1$. Let $K^*_b = \inf_{\gamma \in \Gamma_b} \Lambda^*(\gamma)$. (If Γ_b is empty, we set $K^*_b = \infty$.)

Proposition 7.4. *Suppose Assumptions 7.3(i)-(ii) hold.*

(i) For problem Q2,

$$\inf_{G \in \mathcal{G}} \lim_{n \to \infty} \frac{1}{n} \log P^*_{e,n}(G) = \min_{1 \le b \le B} \frac{1}{b} K^*_b.$$

In particular, using the same one-bit transmission function at all uncensored sensors is asymptotically optimal iff $K^*_1 = \min_{1 \le b \le B}(K^*_b/b)$.

(ii) Let $b^* = \arg\min_{1 \le b \le B}(K^*_b/b)$. For problem Q1, if B/b^* is an integer, we can restrict to using $L = B/b^*$ sensors, all of them using the same b^*-bit transmission policy, without affecting the optimal exponent.

Proof. Part (i) follows from Corollary 7.1. For part (ii), let γ_{b^*} achieve $K^*_{b^*}$ to within $\epsilon > 0$. Let each of the L sensors in problem Q1 use γ_{b^*}. This comes within ϵ of the optimal exponent for problem Q2, and therefore for problem Q1 as well. □

Let $H = -\min_{s \in [0,1]} \log \mathbb{E}_0\left[\left(\frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_0}\right)^s\right]$, and note that $K^*_b \ge -H$. Hence, for any $b \in \{2, \ldots, B\}$,

$$\frac{1}{b} K^*_b \ge -\frac{H}{2}.$$

So if $K_1^* \leq -H/2$, we meet the sufficient conditions for problem Q2 to achieve the optimal error exponent with identical one-bit sensors. In that case, it is also optimal for problem Q1 to have B identical one-bit sensors. This recovers Proposition 2 of [30].

On the other hand, suppose that B is an even integer and that $K_2^*/2 < K_1^*$. Then, it is strictly suboptimal to use B identical one-bit sensors for problem Q1. This is the content of Proposition 3 in [30]. For a general B that is not an integer multiple of b^*, the solution to Q1 involves an integer program, which can be difficult to solve for large B. However, as B increases to infinity, we can approach the optimal performance by using $\lfloor B/b^* \rfloor$ b^*-bit sensors.

7.10.5 ρ Independent of the Transmission Function

Suppose that for every value r of the side-information, all the transmission functions in Γ have the same cost, e.g., that the process of transmission under state r requires the same energy for all γ. Then, we can assume $\rho(r, \gamma) = \tilde{\rho}(r)$ for some nonnegative function $\tilde{\rho}$. Suppose also that the set of transmission policies Γ is of the form $\Gamma = \prod_{r \in \mathcal{R}} \Gamma(r)$, where $\Gamma(r)$ is the set of allowed transmission policies $\gamma(\cdot, r)$, when the side-information takes the value r. Let

$$\Lambda^*(s, r) = \inf_{\gamma \in \Gamma(r)} \Lambda(s, r, \gamma),$$

and

$$\Phi^*(s, \xi) = \log \mathbb{E}_0 \big[\big(\xi(R_1) e^{\Lambda^*(s, R_1)} + 1 - \xi(R_1) \big) \ell_\mu^s(R_1) \big].$$

Corollary 7.2. *Assume that $\rho(r, \gamma) = \tilde{\rho}(r)$ for all $\gamma \in \Gamma(r)$, and that Assumption 7.3 holds. Then,*

$$\lim_{n \to \infty} \frac{1}{n} \log P_{e,n}^* = \inf \min_{s \in [0,1]} (u \Phi^*(s, \xi_1) + (1 - u) \Phi^*(s, \xi_2)), \tag{7.34}$$

where the infimum is taken over all $u \in [0,1]$ *and censoring policies* ξ_i *that satisfy*

$$u\mathbb{E}[\xi_1(R_1)\tilde{\rho}(R_1)] + (1-u)\mathbb{E}[\xi_2(R_1)\tilde{\rho}(R_1)] \leq c.$$

Furthermore, it is optimal to use the same transmission policy for all sensors.

Proof. The result is obtained from Theorem 7.4, by observing that the constraints do not affect the optimization with respect to γ_1 and γ_2, and that

$$\begin{aligned}\inf_{\gamma_1,\gamma_2} \big(u\Phi(s,(\xi_1,\gamma_2)) + (1-u)\Phi(s,(\xi_2,\gamma_2))\big) \\ = u\Phi^*(s,\xi_1) + (1-u)\Phi^*(s,\xi_2).\end{aligned}$$

□

In this case, we use the same transmission policy at all sensors, and at most two different censoring policies. Suppose that s^* is a minimizing value of s in (7.34). Then, for any $r \in \mathcal{R}$, we can use a transmission function $\gamma(\cdot, r) \in \Gamma(r)$ that minimizes $\Lambda(s^*, r, \gamma)$, if the minimum is attained.

7.11 Proofs

In this section, we prove some of the results in this chapter.

Proof of Lemma 7.8.

Let $a = \sup\Big\{ \int f(v)\,\mu(\mathrm{d}v) : \mu \in \mathcal{M},\ \int g(v)\,\mu(\mathrm{d}v) \leq c\Big\}$ and $b = \sup\{uf(v_1) + (1-u)f(v_2) : \quad u \in [0,1], v_1, v_2 \in \mathcal{V}_0, ug(v_1) + (1-u)g(v_2) \leq c\}$. Clearly, $a \geq b$. It remains to show that $a \leq b$. Assume that $a < \infty$, and fix an $\epsilon > 0$. Choose a $\mu \in \mathcal{M}$ such that $\int g(v)\,\mu(\mathrm{d}v) \leq c$, and

$$a \leq \int f(v)\,\mu(\mathrm{d}v) + \epsilon. \tag{7.35}$$

Let C be the convex hull of the set $\{(f(v), g(v)) : v \in \mathcal{V}_0\}$. It can be shown that the point $\big(\int f(v)\,\mu(\mathrm{d}v), \int g(v)\,\mu(\mathrm{d}v)\big) \in C$ (see for example pg. 25 of [57]). There-

fore, there exists a finite set of points $\{v_1, \ldots, v_k\} \subset \mathcal{V}_0$, and nonnegative scalars $\alpha_1, \ldots, \alpha_k$ summing to 1, such that $\int f(v)\,\mu(\mathrm{d}v) = \sum_{i=1}^k \alpha_i f(v_i)$ and $\int g(v)\,\mu(\mathrm{d}v) = \sum_{i=1}^k \alpha_i g(v_i)$. Consider the linear program in which we maximize $\sum_{i=1}^k \alpha_i f(v_i)$ over the nonnegative scalars $\alpha_1, \ldots, \alpha_k$ so that $\sum_{i=1}^k \alpha_i g(v_i) \le c$ and $\sum_{i=1}^k \alpha_i = 1$. From a well known result in linear programming [55], there exists an optimal solution to this linear program with at most two of the α_i being non-zero. Hence, $\int f(v)\,\mu(\mathrm{d}v) \le b$. From (7.35), we have $a \le b + \epsilon$, and since ϵ is arbitrary, we obtain $a \le b$. The case where $a = \infty$ has a similar proof. The proof is now complete.

Proof of Lemma 7.9.

For any fixed pure policy π and a given $s \in I = (\tau_1, \tau_2)$, it is well known that the first and second derivatives of $\Phi(s, \pi)$, w.r.t. s, are finite. What needs to be proved, is that these derivatives are *uniformly* bounded for all policies π.

For a given $\pi = (\xi_v, \gamma_v)_{v \in \mathcal{V}_0}$, let

$$\begin{aligned}\varphi(s) &= \mathbb{E}_0\big[(\xi_V(R_1)e^{\Lambda(s,R_1,\gamma_V)} + 1 - \xi_V(R_1))\ell_\mu^s(R_1)\big] \\ &= \mathbb{E}_0\big[\xi_V(R_1)(\ell_{10}^{\gamma_V}(Y_1 \mid R_1)\ell_\mu(R_1))^s + \big(1 - \xi_V(R_1)\big)\ell_\mu^s(R_1)\big], \end{aligned} \tag{7.36}$$

so that $\Phi(s, \pi) = \log \varphi(s)$. Then, for each $s \in I$,

$$\frac{\mathrm{d}^2}{\mathrm{d}s^2}\Phi(s,\pi) = \frac{1}{\varphi(s)}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\varphi(s) - \Big(\frac{\mathrm{d}}{\mathrm{d}s}\Phi(s,\pi)\Big)^2. \tag{7.37}$$

To prove the lemma, it suffices to show that for all policies π and all $s \in I$, $\varphi(s)$ is uniformly bounded away from 0, and $\frac{\mathrm{d}}{\mathrm{d}s}\varphi(s)$ and $\frac{\mathrm{d}^2}{\mathrm{d}s^s}\varphi(s)$ are uniformly bounded. We do this in several steps below. To keep the notation simple, we will abbreviate $\ell_{10}(X_1 \mid R_1)$ to ℓ_{10}, $\ell_{10}^{\gamma_V}(Y_1 \mid R_1)$ to $\ell_{10}^{\gamma_V}$, and $\ell_\mu(R_1)$ to ℓ_μ.

(a) For every pure transmission policy $\gamma \in \Gamma$, and every $s \in [0,1]$ and $r \in \mathcal{R}$, we have $\Lambda(s, r, \gamma) \le 0$, because $\Lambda(0, r, \gamma) = \Lambda(1, r, \gamma) = 0$, and $\Lambda(s, r, \gamma)$ is a convex function of s for each r (see Lemma 2.2.5 of [39]). Therefore, using Jensen's

inequality,

$$\begin{aligned}\varphi(s) &\geq \mathbb{E}_0\big[e^{\Lambda(s,R_1,\gamma_V)}\ell_\mu^s\big] = \mathbb{E}_0\big[(\ell_{10}^{\gamma_V}\ell_\mu)^s\big] \\ &\geq \mathbb{E}_0\big[(\ell_{10}\ell_\mu)^s\big] = \psi(s). \end{aligned} \tag{7.38}$$

Using the same technique as in the proof of Proposition 3 of [22], we can show that $\psi(s) \geq \epsilon \in (0,1)$, for all $s \in I$, hence $\varphi(s) \geq \epsilon$. This implies that $|\Phi(s,\pi)| \leq |\log\epsilon|$, for all $s \in I$.

(b) Let $\Psi(s) = \log\psi(s)$. Then, from (7.38), for all $s \in [0,1]$,

$$\Phi(s,\pi) = \log\varphi(s) \geq \Psi(s).$$

It is easily shown that $\Phi(0,\pi) = \Phi(1,\pi) = \Psi(0) = \Psi(1) = 0$, and that both functions are convex in $s \in [0,1]$. Then, for $s \in I$,

$$\begin{aligned}&\Big|\frac{\mathrm{d}}{\mathrm{d}s}\Phi(s,\pi)\Big| \leq \sup_{s\in I}\Big|\frac{\mathrm{d}}{\mathrm{d}s}\Psi(s)\Big| \\ &= \sup_{s\in I}\frac{1}{\psi(s)}\big|\mathbb{E}_0\big[(\ell_{10}\ell_\mu)^s\log(\ell_{10}\ell_\mu)\big]\big| \\ &\leq \frac{1}{\epsilon}\max\{\big|\mathbb{E}_j\big[(\ell_{10}\ell_\mu)^{\tau_k}\log(\ell_{10}\ell_\mu)\big]\big| : k = 1,2\} < \infty. \end{aligned} \tag{7.39}$$

(We used here the convexity of Ψ, which implies that the magnitude of its derivative is maximized at one of the end points τ_k.) The finiteness of the R.H.S. of (7.39) follows from Assumption 7.3(v).

(c) From (b) above, $|\frac{\mathrm{d}}{\mathrm{d}s}\varphi(s)| < \infty$ for all $s \in I$. So, for the same reason as in Lemma 2.2.5(c) of [39], we can differentiate $\varphi(s)$ twice under the expectation operator. Hence, from (7.36), we have

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\varphi(s) &= \mathbb{E}_0\Big[\xi_V(R_1)(\ell_{10}^{\gamma_V}\ell_\mu)^s\log^2\big(\ell_{10}^{\gamma_V}\ell_\mu\big)\Big] \\ &\quad + \mathbb{E}_0\Big[\big(1-\xi_V(R_1)\big)\ell_\mu^s\log^2\ell_\mu\Big] \\ &\leq \mathbb{E}_0\Big[(\ell_{10}^{\gamma_V}\ell_\mu)^s\log^2\big(\ell_{10}^{\gamma_V}\ell_\mu\big)\Big] + \mathbb{E}_0\Big[\ell_\mu^s\log^2\ell_\mu\Big] \end{aligned}$$

$$= \mathbb{E}_0\Big[\frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathbb{E}_0\big[(\ell_{10}^{\gamma_V}\ell_\mu)^s \mid V\big]\Big] + \frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathbb{E}_0[\ell_\mu^s] \leq 2b,$$

from Assumption 7.3(vi).

The steps above show that $\frac{\mathrm{d}^2}{\mathrm{d}s^2}\Phi(s,\pi)$ is uniformly bounded and completes the proof of the lemma.

Chapter 8

Summary and Future Work

We have studied the asymptotic detection performance of tree networks with bounded height, under both a Neyman-Pearson criterion and a Bayesian criterion. Similar to the parallel configuration, we have shown that the optimal error probability decays exponentially fast with the number of nodes in the network. In addition, we have shown, rather surprisingly, that under the Neyman-Pearson formulation, if the number of leaves dominates, the network can achieve the same performance as if all nodes were transmitting directly to the fusion center. Moreover, this can be achieved (after performing a height uniformization procedure) by a simple strategy in which all leaves use the same transmission function, while all other nodes act as 1-bit relays and use a LLRQ with a common threshold. Of course, in practice, it would be wasteful to have only the leaf sensors make observations, if n is not large enough.

Several other issues remain outstanding, and are areas for further research. An intriguing question, which has been left unanswered, is whether the inequality $g_P^* \leq g^*$ is always true under the bounded height assumption, when every node is constrained to sending the same number of bits. Another issue is that although the error exponents are the same for relay networks in which leaves dominate, the performance of a relay network could be significantly worse than that of a parallel configuration, in the sense that 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. Additionally, we have not addressed

the case when there is feedback from a node to its immediate predecessors [20, 58]. We expect that similar techniques to the ones developed in Chapter 3 may be useful for characterizing the error exponent when there is feedback.

Under the Bayesian formulation, the error exponent is generically worse than that of the parallel configuration. To provide insights into the Bayesian detection performance, we study specific classes of tree networks that are restricted to simple strategies. In particular, we consider simple counting strategies in symmetric tree networks, and characterize the optimal detection performance over this class of strategies. Although we have not been able to show that restricting to counting strategies results in no loss of optimality, we conjecture this to be the case. We also compare the detection performance of symmetric tree networks (with a fixed number of relay nodes) to that of rapidly branching tree networks. It is shown that for these classes of tree networks and transmission strategies, the Bayesian detection performance deteriorates with the height of the tree architecture.

We also studied the effects of node failures and unreliable communications in a dense sensor network, arranged as a tree of bounded height. In the case of node failures, we showed that the optimal error probability falls exponentially with the expected number of leaves in the network. In the case of unreliable communications, the optimal error probability falls exponentially with the number of immediate predecessors of the fusion center, regardless of the height. This suggests that, in practice, it is preferable to have a node faced with an unreliable channel remain silent (as if it had failed). It also suggests that, when designing a large scale sensor network, it is more important to ensure that nodes can communicate reliably with each other (e.g., by boosting the transmission power) than to ensure that nodes are robust to failures.

We have assumed that the leaves make (conditionally) i.i.d. observations, even though our sensor network is dense. While this assumption may sometimes hold (e.g., in the context of detecting a known signal in white noise), our assumption is restrictive and will often be violated. On the other hand, without the i.i.d. assumption, finding exact optimal strategies is a NP-complete problem, even in the case of a parallel configuration [38]. It is therefore important to investigate the possibility of

approximately optimal approaches to the case of dependent observations, and simple strategies with appealing approximately optimal properties. For some recent work in the case of correlated observations in a parallel configuration, we refer the reader to [59–64].

Other than node failures and unreliable communications, another threat to a sensor network is malicious tampering of some nodes so that they report false information to the fusion center [65,66]. It would be of interest to characterize the impact of such Byzantine sensors on the detection performance.

We have shown that, in Bayesian decentralized detection, using a long tandem of sensors, the rate of decay of the error probability is sub-exponential. In order to obtain more precise bounds, we introduced a modified tandem network, which outperforms the original one, and used tools from Large Deviations Theory. Under the assumption of bounded KL divergences, we have shown that the error probability is $\Omega(e^{-cn^d})$, for all $d > 1/2$. Under the further assumption that the moments (under H_0) of order s of the likelihood ratio are finite for all s in an interval that contains $[0, 1]$ in its interior, we have shown that the lower bound can be improved to $\Omega(e^{-c(\log n)^d})$, for all $d > 1$, and that this latter bound is tight.

In our model, we have assumed binary communication between sensors, and we have been concerned with a binary hypothesis testing problem. The question of whether k-valued messages (with $k > 2$) will result in a faster decay rate, or even an exponential decay rate, remains open. In the case of m-ary hypothesis testing using a tandem network where each sensor observation is a Bernoulli random variable, [50] shows that using $(m + 1)$-valued messages is necessary and sufficient for the error probability to decrease to 0 as n increases. However, it is unknown what the decay rate is. Nevertheless, we conjecture that the error decay rate is always sub-exponential.

We finally note that under a Neyman-Pearson formulation, the picture is less complete. We have shown the sub-exponential decay of the Type II error probability, but only for a particular (myopic) sensor strategy. The case of general strategies is an interesting open problem.

We have studied the tandem network, thus as a next step, it is worthwhile to

understand the rate at which the error probability decays for general tree networks that do not have the bounded height constraint, and the dependence of the error exponent on the rate at which the height of the tree increases. Although the technique developed in Chapter 3 can be extended to determine bounds on the error exponent of tree sequences whose height grows very slowly compared to n (on the order of $\log|\log(n/l_n(f)-1)|$), we have not been able to find the optimal error exponent for the general case of unbounded height. A more detailed analysis of how the architecture of the network affects the error exponent will be required to achieve this.

Finally, we have formulated a general framework involving censoring in a sensor network. We allow the sensors to censor based on some side-information, while taking into account a general cost function that depends only on the side-information and the transmission policy used by the sensor. We allow the sensors to cooperate with each other and show that for a Neyman-Pearson formulation, such cooperation is not necessary in the asymptotic regime of large number of sensors and small Type I error. Every sensor can independently use the same (generally, randomized) local policy. An optimal policy is found by maximizing an informational quantity subject to a cost constraint. This maximization captures the *tradeoff* between the error exponent and the resource constraint.

In the Bayesian context, we have shown that, in the absence of sensor cooperation, asymptotic optimality is obtained by dividing the sensors into two groups, with every sensor in each group using the same pure policy. We have also shown how to find optimal strategies in some special cases, and the relationship of our results to other works.

Most of our results on censoring networks can be extended in various directions. For example, we may have multiple resource constraints. With k constraints, we will generally need the local randomization variable V_1 to have a range of cardinality $k+1$ (Neyman-Pearson case), or to divide the sensors into $k+1$ groups, with every sensor in each group using the same policy (Bayesian case). Extensions to the case of more than two hypotheses are also possible, along the lines of [22].

Bibliography

[1] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, pp. 501–510, 1981.

[2] 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.

[3] 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.

[4] P. Willett and D. Warren, "The suboptimality of randomized tests in distributed and quantized detection systems," *IEEE Trans. Inf. Theory*, vol. 38, pp. 355–361, Mar. 1992.

[5] J. N. Tsitsiklis, "Extremal properties of likelihood-ratio quantizers," *IEEE Trans. Commun.*, vol. 41, pp. 550–558, 1993.

[6] ——, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.

[7] 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.

[8] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: part I - fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, 1997.

[9] B. Chen and P. K. Varshney, "A Bayesian sampling approach to decision fusion using hierarchical models," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[10] B. Chen and P. K. Willett, "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 693–699, Feb. 2005.

[11] A. Kashyap, "Comments on "On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels"," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1274–1275, Mar. 2006.

[12] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3349–3358, Jul. 2006.

[13] L. K. Ekchian and R. R. Tenney, "Detection networks," in *Proc. IEEE Conference on Decision and Control*, 1982, pp. 686–691.

[14] 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.

[15] 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.

[16] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman, "Optimization of detection networks: part I - tandem structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 5, pp. 1044–1059, 1991.

[17] ——, "Optimization of detection networks: part II - tree structures," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 211–221, 1993.

[18] 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.

[19] 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.

[20] 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.

[21] Y. Lin, B. Chen, and P. K. Varshney, "Decision fusion rules in multi-hop wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 2, pp. 475–488, Apr. 2005.

[22] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Math. Control, Signals, Syst.*, vol. 1, pp. 167–182, 1988.

[23] M. E. Hellman and T. M. Cover, "Learning with finite memory," *Ann. of Math. Statist.*, vol. 41, no. 3, pp. 765–782, 1970.

[24] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. of Math. Statist.*, vol. 40, no. 3, pp. 828–835, 1969.

[25] J. D. Papastavrou and M. Athans, "Distributed detection by a large team of sensors in tandem," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 639–653, 1992.

[26] P. Swaszek, "On the performance of serial networks in distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 254–260, 1993.

[27] 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.

[28] 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.

[29] J.-F. Chamberland and V. V. Veeravalli, "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.

[30] ——, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, 2003.

[31] 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.

[32] ——, "Bayesian detection in bounded height tree networks," in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2007.

[33] ——, "Detection in dense wireless sensor networks," in *Proc. IEEE Wireless Commun. and Networking Conf.*, Hong Kong, Mar. 2007.

[34] ——, "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.

[35] ——, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007.

[36] ——, "Censoring sensors: Asymptotics and the value of cooperation," in *Proc. Conf. on Inform. Sci. and Sys.*, Princeton, NJ, Mar. 2006, pp. 62–67.

[37] ——, "Asymptotically optimal distributed censoring," in *Proc. IEEE Int. Symp. on Inform. Theory*, Seattle, WA, Jul. 2006, pp. 625–629.

[38] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, pp. 440–446, 1985.

[39] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY: Springer-Verlag, 1998.

[40] D. Stroock, *Probability Theory: An Analytic View*. Cambridge, UK: Cambridge University Press, 1993.

[41] M. R. Garey and D. S. Johnson, *A guide to the theory of NP-Completeness*. New York, NY: W. H. Freeman, 1979.

[42] 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.

[43] 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.

[44] 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.

[45] A. Anandkumar and L. Tong, "A large deviation analysis of detection over multiaccess channels with random number of sensors," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, 2006.

[46] R. Durrett, *Probability: Theory and Examples*, 2nd ed. New York, NY: Duxbury Press, 1995.

[47] R. Min and A. P. Chandrakasan, "Top five myths about the energy consumption of wireless communication," *ACM Mobile Comp. and Comm. Rev.*, vol. 6, no. 4, pp. 65–67, 2002.

[48] E. Björnemo, M. Johansson, and A. Ahlén, "Two hops is one too many in an energy-limited wireless sensor network," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 2007, pp. 181–184.

[49] D. N. C. Tse and P. Viswanath, *Fundatmentals of Wireless Communications*. New York, NY: Cambridge University Press, 2005.

[50] J. Koplowitz, "Necessary and sufficient memory size for m-hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 44–46, 1975.

[51] B. Chandrasekaran and C. C. Lam, "A finite-memory deterministic algorithm for the symmetric hypothesis testing problem," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 40–44, 1975.

[52] P. J. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2001, vol. 1.

[53] R. E. Blahut, *Principles and Practice of Information Theory*. New York, NY: Addison-Wesley, 1987.

[54] R. R. Bahadur, S. L. Zabell, and J. C. Gupta, "Large deviations, tests, and estimates," in *Asymptotic Theory of Statistical Tests and Estimation: In Honor of Wassily Hoeffding*, I. M. Chakravarti, Ed. New York, NY: Academic Press, 1980.

[55] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, 1st ed. Belmont, MA: Athena Scientific, 1997.

[56] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels, I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[57] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[58] S. Alhakeem and P. Varshney, "Decentralized bayesian detection with feedback," *IEEE Trans. Syst., Man, Cybern.*, vol. 26, no. 4, pp. 503–513, Jul. 1996.

[59] E. Drakopoulos and C. C. Lee, "Optimum multisensor fusion of correlated local decisions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 27, no. 4, pp. 593–606, Jul. 1991.

[60] M. Kam, Q. Zhu, and W. S. Gray, "Optimal data fusion of correlated local decisions in multiple sensor detection systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 916–920, 1992.

[61] R. S. Blum and S. A. Kassam, "Optimum distributed detection of weak signals in dependent sensors," *IEEE Trans. Inf. Theory*, vol. 38, no. 3, pp. 1066–1079, May 1992.

[62] R. S. Blum, S. A. Kassam, and H. Poor, "Distributed detection with multiple sensors: part II - advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[63] J.-F. Chamberland and V. V. Veeravalli, "How dense should a sensor network be for detection with correlated observations?" *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5099–5106, Nov. 2006.

[64] W. Li and H. Dai, "Distributed detection in large-scale sensor networks with correlated sensor observations," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Sep. 2005.

[65] K. Kursawe, "Optimistic Byzantine agreement," in *Proc. IEEE Symp. on Reliable Distributed Systems*, Oct. 2002, pp. 262–267.

[66] S. Marano, V. Matta, and L. Tong, "Distributed inference in the presence of Byzantine sensors," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2006.