ON UNIFORMLY MOST POWERFUL DECENTRALIZED DETECTION

by

Uri Rogers

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Electrical and Computer Engineering

Boise State University

May 2014

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the dissertation submitted by

Uri Rogers

Dissertation Title: On Uniformly Most Powerful Decentralized Detection

Date of Final Oral Examination: 4 April 2014

The following individuals read and discussed the dissertation submitted by student Uri Rogers, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Hao Chen, Ph.D.	Chair of the Supervisory Committee
John N. Chiasson, Ph.D.	Member, Supervisory Committee
Vishal Saxena, Ph.D.	Member, Supervisory Committee
Zachariah C. Teitler, Ph.D.	Member, Supervisory Committee
Qi Cheng, Ph.D.	External Examiner

The final reading approval of the dissertation was granted by Hao Chen, Ph.D., Chair of the Supervisory Committee. The dissertation was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

This is dedicated to my wife Kerry and daughter Nicole.

ACKNOWLEDGMENTS

First, I would like to thank my advisor Professor Hao Chen for all his guidance and support. I would also like to thank the other members of by dissertation comittee, Professor John Chiasson, Professor Vishal Saxena, and Professor Zach Teitler for the stimulating conversations, advice, and all their time. I also wish to express gratitude to the Idaho Space Grant Consortium for fellowship funding and the financial support of Boise State University.

Finally, I would like to thank my family for their sacrifice, patience, and complete support. Without them, this would not have been possible.

ABSTRACT

The theory behind Uniformly Most Powerful (UMP) composite binary hypothesis testing is mature and well defined in centralized detection where all observations are directly accessible at one central node. However, within the area of decentralized detection, UMP tests have not been researched, even though tests of this nature have properties that are highly desirable. The purpose of this research is to extend the UMP concept into decentralized detection, which we define as UMP decentralized detection (UMP-DD). First, the standard parallel decentralized detection model with conditionally independent observations will be explored. This section will introduce theorems and corollaries that define when UMP-DD exists and provide counterintuitive examples where UMP-DD tests do not exist. Second, we explore UMP-DD for directed single-rooted trees of bounded height. We will show that a binary relay tree achieves a Type II error probability exponent that is equivalent to the parallel structure even if all the observations are not identically distributed. We then show that the optimal configuration can also achieve UMP-DD performance, while the tandem configuration does not achieve UMP-DD performance. Finally, we relax the assumption of conditional independence and show under specific constraints that both the parallel and binary relay tree configurations can still be UMP-DD. Throughout, examples will be provided that tie this theoretical work together with current research in fields such as Cognitive Radio.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCDF – Complementary CDF

CDF – Cumulative Distribution Function

C-V – Chair-Varshney

FC – Fusion Center

GGD – Generalized Gaussian Distribution

h**UBRT** – h-uniform binary relay tree

h**BRT** – binary relay tree of bounded height h

HCI – hierarchical conditional independence

i.i.d. – Independent, and Identically Distributed

KKT – Karush-Kuhn-Tucker

KLD – Kullback Leibler Divergence

LRT – Likelihood-Ratio Test

LLRT – log Likelihood-Ratio Test

LMP-DD – Locally Most Powerful Decentralized Detection

NP Hard – Non-deterministic polynomial time hard

NP Complete – Non-deterministic polynomial time complete

pdf – Probability Distribution Function

pmf – Probability Mass Function

ROC – Receiver Operating Characteristics

SNR – Signal to Noise Ratio

UMP – Uniformly Most Powerful

UMP-DD – UMP Decentralized Detection

LIST OF SYMBOLS

A_n	Set containing all the directed arcs of T_n
a_k	Type I Error Probability for class k
b_k	Type II Error Probability for class k
$\mathcal{C}$	Critical region for deciding H_0 versus H_1
C_k	Set containing all V_n in class k with conditionally independent observations
C_k	Set containing all V_n in class k with conditionally dependent observations
$D\left(P\|\|Q\right)$	Kullback Leibler Divergence between distributions P and Q
f	Fusion Center / Root of the Tree
h	Height of T_n with $h = \max\left\{h_B \,:\, B \in A_n\right\}$
h'	Number of directed arcs traveled from f to a certain level of relay nodes along a given branch
h_B	Total number of directed arcs traveled from f to the leaves on branch B
l	Likelihood Ratio Test
L	log Likelihood Ratio Test
M_i	Set containing all of $v_i s$ predecessors
m_i	Total number of predecessor nodes to v_i, $m_i = \|M_i\|$
N	Total number of classes

n	Total number of nodes in the tree
P_j	Probability distribution conditioned on j
P_D	Detection Probability
P_F	False Alarm Probability
P_M	Probability of Miss
T_n	Directed single-rooted tree with n nodes.
U_0	Final inference at f
U_s	Inference at sensor s
U_v	Inference at node v_i
V_n	Set containing all the nodes of T_n
v_i	Relay or Tandem Node with arbitrary index i
$\tilde{V}_k$	Set containing all relay nodes in class k
W_s	Uncorrelated Noise at sensor s
$\mathcal{W}_s$	Correlated Noise at sensor s
α	Type I Error Probability, Size of the Test at Root Level
β	Type II Error Probability at Root Level
λ	Lagrange Multiplier
ν	Randomization value in Neyman-Pearson Testing
ς	Signal-to-Noise Ratio (SNR)
θ	Composite Signal Level
θ_k	Composite Signal Level in class k

CHAPTER 1

INTRODUCTION

Statistical inference is a field of study that attempts to *optimally* determine or estimate a state of nature based on observations regarding that state. When the number of states is finite, we refer to this as a detection problem or possibly hypothesis testing. When the number of states is infinite, we refer to this as an estimation problem. Both problems assume some knowledge of the states of nature, the uncertainty of the environment, and typically define these statistically.

The traditional detection problem occurs when all the observations are available at a single decision point and will be referred to as centralized detection. A key attribute of centralized detection is that the decision procedure is only applied once for a given set of observations at the single decision point. When the observations are quantized and the final detection decision is based on these quantized values, then this is commonly known as decentralized detection or in some cases distributed detection.

Unlike centralized detection, a key attribute of decentralized detection is that the decision making typically occurs at multiple locations and multiple layers, where these so called local decisions are fused at a central fusion node (e.g., fusion center, root of the tree) to arrive at the final decision. The multitude of decision locations and layers introduces coupling among the various decision processes and greatly increases complexity regarding the system design and optimization. As a result, many well established theories in centralized detection cannot be directly applied to decentralized detection.

The analysis and theory of centralized detection is well established, with [2–5] representing a small subset of possible references. At the same time, when the observations regarding the state of nature are statistically independent, the analysis and theory for distributed detection is well researched with an introduction available in [6]. However, there remain gaps in the decentralized detection theory, with areas that are relatively unexplored. One such gap is in the area of Uniformly Most Powerful (UMP) tests and will be our focus throughout this work.

1.1 Motivation

Within the general field of statistical inference, the largest branch of study is deciding between two states of nature and is referred to as binary hypothesis testing. An important application of binary hypothesis testing is radar, where one transmits an electromagnetic pulse and if a target is present, expects to see a reflected return signal. However, the return signal is corrupted by noise, false reflections, and other disturbances, providing an uncertain observation of the state of nature. This example provides three key points that will be important in the sequel. First, if a target is present, the reflected signal has an unknown strength that can be assumed to be greater than zero. This is referred to as a composite parameter. Second, when the target is absent (i.e., no reflected signal), there is still noise and other disturbances impeding the decision process. Finally, the target's true presence or absence is outside the control of the detection process and predominantly studied statistically.

The importance of radar detection drove early researchers to study the three conditions listed above and determine optimal methods relative to a set of constraints for detecting targets. One intuitive method is to maximize the probability of detecting the target when present, subject to a constraint on how often the target is classified as being present when the true state of nature is that it is absent. While these concepts will be formalized later, there are cases where the detection probability subject to the previous constraint is maximized for any value of the composite parameter. These tests are classified as UMP tests and are supported by a well established theory [2–5]. However, in decentralized detection, the concept of UMP tests is almost completely missing from both literature and research beyond our prior work [7–9].

The goal of this research is to introduce UMP detection into the decentralized detection arena. We endeavor to establish a theoretical framework for when a decentralized system based on a certain network topology and a set of sufficient conditions is UMP, and establish justification for why other network topologies are never UMP. We also introduce the acronym UMP-DD, which stand for UMP decentralized detection and will be used henceforth to help distinguish this effort from centralized UMP theory.

1.2 Hypothesis Testing Background and Notation

Before exploring UMP-DD in detail, we provide a brief background on the theory of hypothesis testing and the notation to be used throughout. This is by no means a thorough treatment of the subject, with the interested reader referred to [2] as a starting point. The following sections will establish notation, introduce the binary hypothesis testing model, define what composite parameters are, introduce Neyman-

Pearson statistics, and provide a rigorous definition of UMP.

1.2.1 Notation

We define $(\Omega, \mathcal{F}_\sigma)$ as a measurable space, with Ω representing an observation set, $\mathcal{F}_\sigma$ the sigma-algebra generated by Ω endowed with a $\sigma-$finite measure [3, 10, 11]. There are probability measures, P_j, associated with each hypothesis, H_j, where $j = 0$, 1 in the binary case. The measure μ is selected so it dominates all P_j (i.e., $P_j \ll \mu, \ \forall j$) and the densities are defined as $p_j \triangleq {}^{dP_j}/_{d\mu}, \ \forall j$. When $(\Omega, \mathcal{F}_\sigma)$ is a discrete observation space, the p_j is a probability mass function (pmf) and is a probability density function (pdf) on a continuous observation space. We will use the term density to imply pmf or pdf, similar to [3]. Then, for any $A \in \mathcal{F}_\sigma$, the $P_j(A) = \int_A p_j d\mu$. Suppose A is a measurable function of some random variable X, then $P_j(A) = \int_A p_j d\mu = \int_A p_j(x)\, \mu(dx)$, where we use the latter notation. Equivalence of the measures P_j and P_k occurs when both are absolutely continuous relative to one another (i.e., $P_j \ll P_k$ and $P_k \ll P_j$) and will be denoted as $P_j \equiv P_k, \ j \neq k$.

Boldface capital letters (e.g., $\mathbf{X}$, $\mathbf{U}$) are used to denote vectors of random variables and boldface lowercase letters to denote a particular realization of the random vector. The expectation operator relative to P_j is denoted at E_j and the distribution of a random variable $\boldsymbol{X}$ under H_j is $P_j^{\boldsymbol{X}}$. All vectors are column vectors and will be represented in row format as (x, y, z) or $[x, y, z]^T$, where T is the standard transpose operator. All logarithms will be base e, unless explicitly defined otherwise and $0 \log {}^0/_p := 0, \ \forall p$.

Table 1.1: Hypothesis Testing Statistics

Conditional Probability	Symbol(s)	Description
$P_1(U_0 = 0)$	P_M, β	Type II Error Probability Miss Probability
$P_1(U_0 = 1)$	P_D	Detection Probability Power of the Test
$P_0(U_0 = 1)$	P_F, α	Type I Error Probability False Alarm Rate, or Size of the Test

1.2.2 Binary Hypothesis Testing

Let H_0 and H_1 represent two states of nature or hypotheses on a measurable space $(\Omega, \mathcal{F}_\sigma, \mu)$ with $\boldsymbol{X}$ the noisy observation of this state. Then, an appropriate model is

$$
\begin{aligned}
H_0 : &\boldsymbol{X} \sim P_0 \quad \text{(Target Absent)}, \\
H_1 : &\boldsymbol{X} \sim P_1 \quad \text{(Target Present)}.
\end{aligned} \tag{1.1}
$$

Let $U_0 \in \{0, 1\}$ represent the final inference regarding H_0 versus H_1. Then, the Type I error probability is defined as $P_0(U_0 = 1)$ and the Type II error probability as $P_1(U_0 = 0)$. Table 1.1 defines all conditional probabilities, a set of notations, and the related terminology typical in hypothesis testing. Note that $P_0(U_0 = 0)$ is a correct decision implicitly defined as $1 - P_0(U_0 = 1)$.

Within this statistical framework, the goal is to minimize the probability of error, which is a function of the Type I and Type II error probabilities. When the *a priori* probabilities for H_0 and H_1 are known, then Bayesian statistics form the optimal decision rule by minimizing a weighted cost function of both error probabilities (e.g., Bayes' risk) over all decision rules and is many times referred to as the Bayes' rule for testing H_0 versus H_1 [3]. However, when the *a priori* probabilities are unknown, then another decision rule is required, such as that used under Neyman-Pearson hypothesis testing.

1.2.3 Neyman-Pearson Criterion

When the *a priori* probabilities of H_0 and H_1 are unknown, then the weighted cost function based on the Type I and Type II error probabilities is indeterminate and the optimal Bayesian statistics approach cannot be applied. An alternative approach is to minimize $P_1(U_0 = 0)$ subject to a constraint $P_0(U_0 = 1) \leq \alpha$ for some $\alpha \in (0, 1)$ and is known as Neyman-Pearson hypothesis testing or Neyman-Pearson statistics. Specifically, let γ be a decision rule for deciding between H_0 and H_1, where γ is a typically a function of some observations set, say $\boldsymbol{X}$ (i.e., $\gamma(\boldsymbol{x})$). Then, $P_0(U_0 = 1)$ and $P_1(U_0 = 0)$ are a function of γ and the Neyman-Pearson criterion can be written as

$$\min_{\gamma} P_1(U_0 = 0 : \gamma) \quad \text{subject to } P_0(U_0 = 1 : \gamma) \leq \alpha.$$

Next, let $\boldsymbol{X}$ be the observations at a fusion node and define the likelihood-ratio test (LRT) as

$$l(\boldsymbol{x}) = \frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}(\boldsymbol{x}) = \frac{dP_1^{\boldsymbol{X}}/d\mu}{dP_0^{\boldsymbol{X}}/d\mu}(\boldsymbol{x}) = \frac{p_1^{\boldsymbol{X}}(\boldsymbol{x})}{p_0^{\boldsymbol{X}}(\boldsymbol{x})}, \tag{1.2}$$

where $p_j^{\boldsymbol{X}}(\boldsymbol{x}),\ j = 0, 1$ are the associated densities, $\frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}$ is the Radon-Nikodym derivative of $P_1^{\boldsymbol{X}}$ with respect to $P_0^{\boldsymbol{X}}$, when $P_1^{\boldsymbol{X}} \ll P_0^{\boldsymbol{X}}$. Then, by the Neyman-Pearson Lemma

$$\gamma(\boldsymbol{x}) = \begin{cases} 1, & \text{if } l(\boldsymbol{x}) > \eta, \\ \nu, & \text{if } l(\boldsymbol{x}) = \eta, \\ 0, & \text{if } l(\boldsymbol{x}) < \eta \end{cases} \tag{1.3}$$

is the most powerful test of size α for a threshold η, where $\eta > 0$ and ν, a randomization factor with $\nu \in [0, 1]$, are chosen such that $P_0 (U_0 = 1 : \gamma) = \alpha$. This last condition defines a critical region, $\mathcal{C}$, where $\mathcal{C} = \{\boldsymbol{x} : \gamma(\boldsymbol{x}) = 1, \ \forall \boldsymbol{x} \in \Omega\}$, and is said to have size α. Then, the false-alarm probability as a function of $\gamma(\boldsymbol{x})$ is

$$P_F (\gamma(\boldsymbol{x})) = P_0 (U_0 = 1 : \gamma) = \int_{\mathcal{C}(\gamma)} p_0 (\boldsymbol{x}) \, \mu (d\boldsymbol{x})$$

and the detection probability is

$$P_D (\gamma(\boldsymbol{x})) = P_1 (U_0 = 1 : \gamma) = \int_{\mathcal{C}(\gamma)} p_1 (\boldsymbol{x}) \, \mu (d\boldsymbol{x}) \, .$$

Each of these concepts will be fundamental in the sequel.

1.2.4 Uniformly Most Powerful Defined

Suppose that our observations vector $\mathbf{X}$ is composed of the true signal level, $\boldsymbol{\theta}$, and noise, $\boldsymbol{W}$. There are two general classifications with respect to $\boldsymbol{\theta}$. When $\boldsymbol{\theta} = \dot{\boldsymbol{\theta}}$ for a single $\dot{\boldsymbol{\theta}} \in \Omega$ under one of the hypotheses, then this is referred to as a simple hypothesis. However, when $\boldsymbol{\theta} \in \boldsymbol{\Theta}_j$ under $H_j, \ j = 1, 2$ with $\boldsymbol{\Theta}_j \subseteq \Omega$, then this is said to be a composite hypothesis with a composite parameter. Here $\boldsymbol{\Theta}_j$ is a family of possible signal levels and may be uncountable with $\boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \Omega$. We require that $\mu (\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1) = 0$ (i.e., intersection exists only on a measure zero set), thus we can assume without loss of generality, the $\boldsymbol{\Theta}_j$ are disjoint (i.e., $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_1 = \emptyset$) and $\boldsymbol{\Theta}_0 \sqcup \boldsymbol{\Theta}_1 = \Omega$, where $\sqcup$ represents a disjoint union.

Suppose $\boldsymbol{\theta}$ is a composite parameter under both hypotheses and consider the Neyman-Pearson test of (1.3). As such, $P_0^{\boldsymbol{X}}$, $P_1^{\boldsymbol{X}}$, and γ all become functions of these unknown signal levels. Thus, the Neyman-Pearson Lemma becomes

$$\gamma(\boldsymbol{x};\boldsymbol{\theta}) = \begin{cases} 1, & \text{if } \frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}(\boldsymbol{x};\boldsymbol{\theta}) > \eta, \\ \nu, & \text{if } \frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}(\boldsymbol{x};\boldsymbol{\theta}) = \eta, \\ 0, & \text{if } \frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}(\boldsymbol{x};\boldsymbol{\theta}) < \eta. \end{cases} \tag{1.4}$$

Generally, η and ν are functions of the unknown composite parameter $\boldsymbol{\theta}$. Thus the optimal decision rule γ is also a function of the composite parameter and we are unable to select a single decision rule to meet the Type I error probability constraint α, $\forall\boldsymbol{\theta}$. In other words, the decision rule becomes a function of the unknown parameter to be detected.

There are, however, some special situations where a solution to the challenging problem of composite binary hypothesis testing exists. One such solution is when the composite test is UMP. Specifically, a UMP test is defined as follows [5].

Definition 1.1. The critical region $\mathcal{C}$ is a uniformly most powerful critical region of size α, testing the simple hypothesis H_0 against an alternative composite hypothesis H_1, if the set $\mathcal{C}$ is a best critical region of size α for testing H_0 against each simple hypothesis in H_1. A test defined by this critical region $\mathcal{C}$ is called a **Uniformly Most Powerful** (UMP) test, with significance level α, for testing the simple (possibly composite) hypothesis H_0 against the alternative composite hypothesis H_1.

This definition uses the fact that $\alpha \triangleq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta_0}} P_0^{\boldsymbol{X}}(\boldsymbol{x} : \boldsymbol{x} \in \mathcal{C})$, which indicates the following two tests are equivalent

$$H_0 : \boldsymbol{\theta} \leq \dot{\boldsymbol{\theta}} \quad \text{versus} \quad H_1 : \boldsymbol{\theta} > \dot{\boldsymbol{\theta}}$$

$$H_0 : \boldsymbol{\theta} = \dot{\boldsymbol{\theta}} \quad \text{versus} \quad H_1 : \boldsymbol{\theta} > \dot{\boldsymbol{\theta}}$$

for some $\dot{\boldsymbol{\theta}} \in \Omega$.

Because the threshold η and the randomization parameter ν can be selected independent of $\boldsymbol{\theta}$, the UMP tests are highly desirable, even though they do not always exist. Therefore, it is of both practical and theoretical importance to study hypotheses testing problems to determine whether an UMP exists or not.

1.3 Uniformly Most Powerful Decentralized Detection

Uniformly Most Powerful tests in the centralized case are essentially determined using the single LRT of (1.2) or possibly the logarithm LRT (LLRT)

$$L\left(\boldsymbol{x};\boldsymbol{\theta}\right) = \log \frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}\left(\boldsymbol{x};\boldsymbol{\theta}\right) = \log \frac{p_1^{\boldsymbol{X}}\left(\boldsymbol{x};\boldsymbol{\theta}\right)}{p_0^{\boldsymbol{X}}\left(\boldsymbol{x};\boldsymbol{\theta}\right)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{1.5}$$

where H_1 is selected with probability ν if $L\left(\boldsymbol{x};\boldsymbol{\theta}\right) = \eta$. This will be discussed in detail in Section 2.2. Contrast this with the decentralized detection problem consisting of n total nodes, where each node is capable of making a local decision, and one node performs both data fusion and the final inference. Now the number of decision rules can be as high as n, with some or potentially all of the rules statistically coupled, which will be expanded on in Section 2.2 and Section 2.3.

Clearly, the decision complexity for decentralized versus centralized detection has significantly increased. In fact, it is known that many decentralized detection problems are non-deterministic polynomial time hard (NP-hard) [12], because the decision rules are generally coupled. This complexity explains to some extent the reason why UMP-DD as a field of research has not been introduced into the literature beyond our publication [7–9] to the best of our knowledge.

1.4 Single-Rooted Tree

This work will focus on directed single-root trees of bounded height. They are directed in the sense that messaging flows from the leaves to a single-root f responsible for making the final inference. And they are bounded height in the sense that the number of arcs along a branch connecting any leaf to f is finite. The restriction to bounded height trees is made because unbounded trees are known to have a detection performance that is inferior, decaying sub exponentially [13], a point which will be expanded on in Chapter 3.

Using notation similar to [14], let $T_n = (V_n, A_n)$ denote a directed single-root tree. Here V_n is the set of vertices or nodes having cardinality n (i.e., $|V_n| = n$) and A_n are the arcs or directed edges of the tree. The fusion center responsible for the final inference is at the root of the tree and will be denoted as f. There are $n - 1$ remaining nodes that perform local sensing and/or local data fusion, where only summary information is forwarded along branches towards the root. The height of the tree is the largest number of arcs, connecting any leaf in T_n to the root. Trees where all leaves have the same height are known as $h-$uniform and will play a key role in our study.

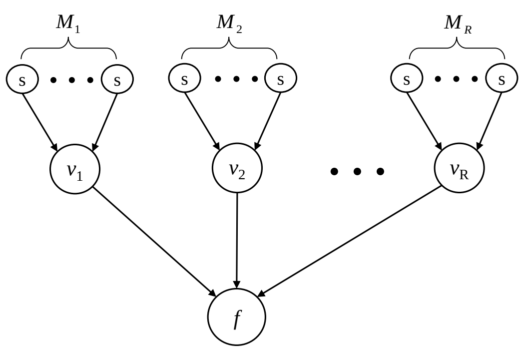
Any node performing local data fusion based on predecessor information will be denoted as v_i, where i is an index variable. A predecessor of node v_i is defined as any node having a directed path from itself to v_i. Here we denote M_i as the set containing all predecessors nodes to v_i, with M_0 as a reserved set for the predecessors of f. Nodes that only perform sensing will be called sensors or leaves and will be denoted by s. Nodes that summarize their predecessor messages and perform no sensing will be called relay nodes or relays [1]. Any network consisting of sensors and relays will be referred to as a relay network. Nodes that augment their predecessor messages with their own local observation will be called tandem or serial nodes. Any network consisting of just sensors and tandem nodes will be called a tandem configuration or network. These networks appear in Figure 1.1, and include the special case parallel network that occurs when all $n-1$ nodes are sensors (i.e., $h=1$).

We now make the observation model for the Directed Single-Rooted Trees specific. That is, each sensor s and tandem node v_i observes a random variable X in some set $\mathcal{X}$, having a marginal distribution P_j^X under hypothesis j. All decisions are quantized to 1-bit for which we elaborate all cases. Each $s \in V_n$ makes a binary decision U_s via a decision rule γ_s, such that $U_s = \gamma_s(X)$, where $\gamma_s : \mathcal{X} \to \{0,1\}$. Node v_i makes a decision U_i using a decision rule γ_i, such that $U_i = \gamma_i\left(X, \{U_s : s \in M_i\} \bigcup \{U_v : v \in M_i\}\right)$, including possibly its own observation X for a tandem node, where $\gamma_i : \mathcal{X} \times \{0,1\}^{|M_i|} \to \{0,1\}$. The decision rule γ_0 at f maps $\{0,1\}^{|M_0|} \to \{0,1\}$ to generate the final decision U_o.

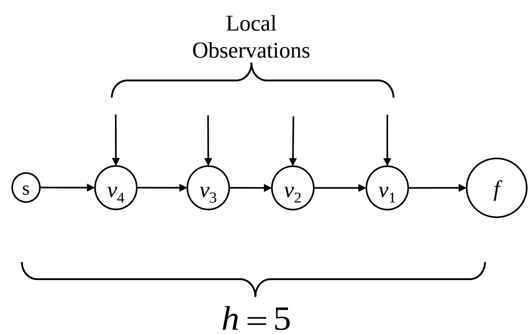
1.5 Outline

Within this work, we endeavor to meet the challenge presented at the end of Section 1.1. We study the UMP-DD problem using a directed single-rooted tree of bounded height using three different observation models. Chapter 2 investigates the parallel configuration of Figure 1.1c consisting of $n-1$ sensors and no relay or tandem

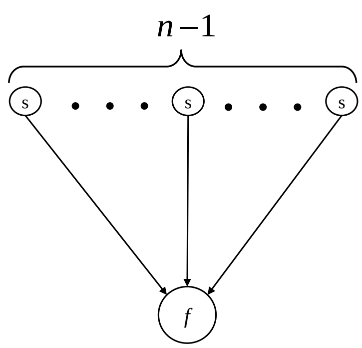
[1]Technically, a relay can perform sensing as long as the relay treats its local sensing as an independent sensor. We choose to use the historical decentralized detection definition of relay [14], which simplifies notation and has no impact on the bounded height asymptotic performance (i.e., $n \to \infty$). However, in the non-asymptotic regime, improved detection performance may be possible when treating each relays local sensing as an independent sensor.



(a) Relay Network / 2-Uniform Tree



(b) Tandem Network



(c) Parallel Configuration / 1-Uniform Tree

Figure 1.1: Directed Single-Rooted Tree Structures

nodes. This model is commonly known as the parallel or star configuration where all leaves send a summary of their local observation directly to the fusion node f. The observations at each s are conditionally independent, when conditioned on the composite parameter under each hypothesis.

Chapter 3 studies single-rooted trees where leaves pass their summary information based on conditionally independent observations via non-leaf nodes towards the fusion node f. Two topologies will be investigated, with the first being a relay tree, where v_i are relay nodes $\forall i$. The second topology is the tandem network, where v_i nodes augment the summary information from their predecessors with their own observation X.

Chapter 4 extends the results of Chapter 2 and Chapter 3 by allowing conditionally dependent observation models. The general conditionally dependent problem is a non-deterministic polynomial time complete (NP-complete) problem [12], with some special cases that offer a polynomial time solution and within this subset there exists categories of tests that are UMP-DD.

CHAPTER 2

CONDITIONALLY INDEPENDENT UMP-DD

2.1 Introduction

The majority of decentralized detection research is based on the assumption that the local sensor observations are conditionally independent. The dominant reason for this assumption is that the general problem is known to be NP-complete [12], where the majority of the problems are NP-hard. That is, when the sensors are conditionally dependent, the optimal local sensor decision and global fusion rules are coupled and the *form* of the optimal local sensor decision rules is often unknown [1, 15]. As such, this chapter will focus solely on the conditionally independent observations model with composite parameters. This is a constraint that will be relaxed in Chapter 4.

Assumption 2.1. *The sensor observations are conditionally independent given the composite parameter under each hypothesis.*

Given the conditionally independent observation assumption, the decentralized detection research is rich with numerous studies on optimal sensor rules and the corresponding fusion rule. The interested reader can refer to [6, 16, 17] and references therein for an in-depth treatment.

2.1.1 Prior Research

The UMP-DD concept that will follow is intimately connected to the selection of optimal local sensor decision rules and data fusion rules. This selection process has a long research history, starting with conditionally independent and identically distributed observations of deterministic signals. Tsitsiklis [18] showed that when each sensor's observations are conditionally independent and identically distributed (i.i.d.), identical likelihood ratio tests (i.e., equal quantizer thresholds) are asymptotically optimal when the fusion rule is also optimized. This study, however, did not address fixed fusion rules nor UMP tests where the hypotheses testing problem is composite versus simple.

The study of optimal sensor decision rules with a fixed fusion rule is more challenging relative to the asymptotic analysis with an optimized fusion rule. Under a fixed fusion rule, Irving and Tsitsiklis showed that for two sensors having conditionally i.i.d. Gaussian observations given a respective hypothesis that equal quantizers are optimal [19]. Warren and Willett extended this result by analyzing the multi-sensor model with conditionally i.i.d. observations under densities that are not sufficiently "peaky" [20]. There they showed equal quantizer thresholds to be optimal with the *And* and *Or* fusion rule for simple binary hypothesis testing (e.g., a deterministic signal). Certainly, the assumption of conditionally i.i.d. observations is highly restrictive, as it assumes each sensor receives the same noise corrupted deterministic signal. In fact, most practical problems in decentralized detection assume differing signal levels across sensors, with cooperative spectrum sensing in the cognitive radio field providing relevant examples (see [21, 22] and references therein). Additionally, Warren and Willett also showed that equal sensor thresholds were at least locally optimal for the general counting (also known as k out of N) fusion rule with i.i.d. observations. The counting fusion rule sums all local decisions, U_s, and compares the result to a threshold to decide H_1 versus H_0.

Within decentralized detection, UMP-DD tests do exist, but the terminology classifying them as such is void. As described previously, this is partially due to the complexity of decentralized detection and the mathematics required to analyze the performance of fusion rules. Hence, the UMP-DD tests explored in this chapter are fundamentally linked with the concept of LRTs and equal quantizer thresholds at every sensor in a decentralized detection system (cf. (1.2) and (1.5)).

2.1.2 Chapter Goals

This chapter will address the decentralized composite binary hypothesis testing problem given its importance to the field of decentralized detection including applications to cognitive radio. Of interest will be composite signals that are actually random across the sensors, as is typically the case for mobile devices (e.g., mobile primary and secondary users in cognitive radio). Not only is the random composite signal model no longer i.i.d., it is not even conditionally independent given a hypothesis, as each sensors observation is dependent on the random unknown signal. Given certain *sufficient conditions* regarding the random composite signal and conditionally independent observations, we will introduce the theory of UMP-DD testing. This UMP-DD theory readily supports the *fixed but unknown* composite signal model and it is straight forward to realize the results of both Irving and Tsitsiklis, and also Warren and Willett with deterministic signals as special cases. Similar to the works mentioned previously and in Section 1.4, we assume 1-bit quantized local decisions and a fixed fusion rule. The UMP-DD theory that will be introduced relies on the generalized log-concave efforts of Prékopa [23, 24] and to some extent the log-concave probability work by Bergstrom and Bagnoli [25].

In Section 2.2, we review the Karlin-Rubin Theorem, which is a fundamental method in determining centralized UMP tests. We then show that application of Karlin-Rubin to the parallel configuration is insufficient for determining UMP-DD.

2.2 Parallel Decentralized Detection and Karlin-Rubin

A standard method used to determine the existence of centralized UMP composite tests is the Karlin-Rubin Theorem as described directly in [26] and indirectly in [2, 5].

Theorem 2.2. *(Karlin-Rubin) Consider testing testing $H_0 : \boldsymbol{\theta} \leq \boldsymbol{\theta}'$ versus $H_1 : \boldsymbol{\theta} > \boldsymbol{\theta}'$. Suppose $T(\boldsymbol{x})$ is a sufficient statistic for the likelihood ratio, $\gamma_s(\boldsymbol{x})$,* with *$l(\boldsymbol{x})$* monotone in *$T(\boldsymbol{x})$. That is for any $\boldsymbol{\theta}_1 < \boldsymbol{\theta}_2$ that $\frac{dP_1^{\boldsymbol{X}}}{dP_0^{\boldsymbol{X}}}(\boldsymbol{x}; \boldsymbol{\theta})$ is a monotone function of $T(\boldsymbol{x})$. Then, for any η', the test that rejects H_0 if and only if $T(\boldsymbol{x}) > \eta'$ is a UMP level α test, where $\alpha = P_0(T(\boldsymbol{x}) > \eta')$.*

The factorization theorem of Neyman defines the sufficiency of a statistic [2, 5]. Specifically, the statistic is sufficient if the density $p_j(\boldsymbol{x}; \boldsymbol{\theta})$ can be factored as $g_{\boldsymbol{\theta}}(T(\boldsymbol{x}))\, h(\boldsymbol{x})$, where $h(\boldsymbol{x})$ does not depend upon $\boldsymbol{\theta}$. The Gaussian distribution with a shifted mean (i.e., $X_s \sim \mathcal{N}(\theta, 1)$) has $T(\boldsymbol{x}) = \sum_{s=1}^{S} x_s$ as a minimally sufficient statistic for S observations in centralized detection. Similarly, the energy detection model with $X_s \sim \mathcal{N}(0, \theta)$ has a minimally sufficient statistic $T(\boldsymbol{x}) = \sum_{s=1}^{S} x_s^2$. The Karlin-Rubin Theorem, through a sufficient statistic, can be applied to a wide array of problems in centralized detection. However, we gave a counter example in an earlier study [7] that implied Karlin-Rubin cannot be applied directly to decentralized detec-

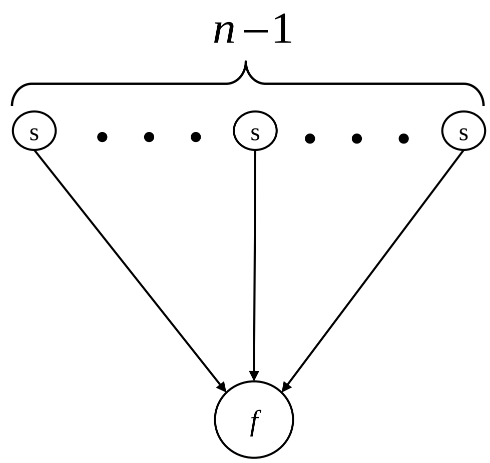


Figure 2.1: Parallel Configuration / 1-Uniform Directed Single-Rooted Tree

tion. Here, we will offer a proof to this implication after introducing the decentralized detection model to be studied.

Consider a 1-uniform directed single-rooted tree with $n-1$ leaves and a single fusion node at the root f. This configuration is depicted in Figure 2.1 and as defined previously will be referred to as a parallel configuration. Here, only the leaves are sensors. As described in Section 1.4, each s observes a conditionally independent random variable $X_s \in \mathcal{X}$ having a marginal distribution $P_j^{X_s}$ under hypothesis $j = 0, 1$ and makes a decision $U_s = \gamma_s(X_s)$, where $\gamma_s : \mathcal{X} \to \{0, 1\}$ (i.e., binary decision). The decision rule γ_0 at f maps $\{0, 1\}^{m_0} \to \{0, 1\}$ to generate the final decision U_o.

Within the parallel decentralized detection framework with conditionally independent observations given θ and H_j, one might attempt to apply the Karlin-Rubin Theorem at f and again at each local sensor to define a UMP-DD. However, such an extension is insufficient, even when testing $H_0 : \theta = \theta'$ versus $H_1 : \theta > \theta'$, when each sensors' LRT is monotone in a test statistic $T_s(X_s)$ for all $s \in V_n$ and the fusion rule is monotone. Given the importance of this claim, these terms are made specific as follows:

Condition 2.3. The fusion node f employs a *monotone fusion rule* such that the probability of deciding H_1 is a monotonic function of $\boldsymbol{U} = \{U_s\}_{s=1}^{n-1}$ and $P_j(U_0 = 1|U_s = 1, \theta) \geq P_j(U_0 = 1|U_s = 0, \theta)$ for all θ and s,

and

Condition 2.4. The local likelihood ratio, $l_s(X_s) = \frac{p_1^{X_s}(x_s;\theta)}{p_0^{X_s}(x_s;\theta')}$, is monotone in $T_s(X)$.

The earlier claim that Karlin-Rubin is insufficient for UMP-DD is unfortunate and is counterintuitive on the surface. We now provide a proof by contradiction, justifying the claim and then give a concrete counter example based on two sensors with conditionally i.i.d. observations and Laplace Noise in Section 2.5.

Proof. Clearly, each sensor has a local UMP level α_s test by Karlin-Rubin, since the local LRT is monotone in $T_s(X_s)$ $\forall s \in V_n$. As the fusion rule is monotone and the observations conditionally independent given θ and H_j, then sensor-by-sensor optimization is at least locally optimal [15]. Hence, the optimal local sensor decision rule, $\gamma_s(T_s(X_s))$, is a LRT against a threshold η_s. The selection of η_s is a function of α_s and can be arbitrarily selected in theory as long as the global α constraint is realized at f. This implies that the optimal test is achieved with an arbitrary selection of each η_s. This contradicts the optimality of equal quantizers with deterministic signals in [19, 20], a subset of the composite case. Even constraining the quantizers to be equal is insufficient, as highlighted by the counter example in Section 2.5 previously mentioned. Thus, a simple extension of the Karlin-Rubin Theorem is not sufficient for UMP-DD. $\square$

2.3 Uniformly Most Powerful Decentralized Detection

This section will define a set of sufficient conditions required to determine UMP-DD and at the same time extend the research beyond the simple i.i.d. observation case. We will require that the distributions be smooth log-concave functions, $P_0^{X_s|\theta} \equiv P_1^{X_s|\theta}$, and as such contain no point masses. Thus, the $\sigma-$finite measure μ can be taken to be the natural Lebesgue Measure.

Assumption 2.5. *The $\sigma-$finite measure μ is the Lebesgue Measure (i.e., for an interval $[a,b]$, $\mu([a,b]) = |b-a|$).*

Notice that Assumption 2.5 does not preclude singularity distributions.

Suppose the goal is to test $H_0 : \boldsymbol{\theta} \leq \boldsymbol{\theta}'$ versus $H_1 : \boldsymbol{\theta} > \boldsymbol{\theta}'$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0 \sqcup \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}$ is the true non-deterministic signal. Then, the unknown parameter $\boldsymbol{\theta}$ induces a $H_j \to \boldsymbol{\theta} \to \boldsymbol{X} \to \boldsymbol{U} \to U_0$ Markov chain for the canonical parallel decentralized testing system of Figure 2.1. Thus $\boldsymbol{X}$ is conditionally dependent given H_j, but conditionally independent given $\boldsymbol{\theta}$ assuming the θ_s $\forall s$ are i.i.d., when conditioned on H_j. That is, let θ_s follows an *a priori* distribution $P_j^{\theta_s}(\mathcal{D}_j)$ with $P_j^{\theta_s} \ll \mu$, under hypothesis j, where $\mathcal{D}_j$ is a single set of parameters defining the distribution (i.e., support, mean, variance, ...) and $\boldsymbol{\mathcal{D}}_j$ is a collection of sets. Henceforth, it will be understood that $P_j^{\theta_s}$ implies $P_j^{\theta_s}(\mathcal{D}_j)$ for notational simplicity. The exponential distribution with mean parameter τ, $\theta_s \overset{i.i.d.}{\sim} \mathrm{Exp}(\tau)$ under H_1 is an example of $P_1^{\theta_s}$. This can be used to model a large sensor network, where the target location is random within the network and the signal level falls off exponentially under a standard exponential path loss model. The zero mean Gaussian distribution with variance φ_s^2, $\theta_s \sim \mathcal{N}(0, \varphi_s^2)$, under H_1, is another example for $P_1^{\theta_s}$. This last approach is appropriate for modeling multipath fading in cognitive radio energy detection as studied in [27], which is based on the results in [28]. A more subtle example for $P_j^{\theta_s}$ is the "fixed but unknown" case where $P_j(\theta_s = \theta_j) = 1$ (i.e., singleton), where $\theta_j \in \Theta_j$ is unknown.

After each $P_j^{\theta_s}$ is defined, it is then possible to calculate an average conditional density

$$\begin{aligned} p_j^{\boldsymbol{X}}(\boldsymbol{\mathcal{D}}_j) &= \prod_{s=1}^{n-1} \int_{\Theta_j} p_j^{X_s|\theta_s} p_j^{\theta_s} \mu(d\theta_s) \\ &= \prod_{s=1}^{n-1} E_j^{\theta_s}\left[p_j^{X_s|\theta_s}\right], \end{aligned} \tag{2.1}$$

where $i = 0, 1$. In other words, each $p_j^{X_s}(\boldsymbol{\mathcal{D}}_j)$ is conditionally i.i.d in an average sense even though it is not strictly i.i.d., as the sensor observations differ via $P_j^{\theta_s}$.

Next, suppose the fusion node f applies the counting fusion rule, which is not only monotone but also an optimal choice with equal quantizer thresholds with i.i.d. observations [20]. The counting rule can be written as

$$U_o = \begin{cases} 1, & \sum_{s=1}^{n-1} U_s \geq \kappa, \\ 0, & \sum_{s=1}^{n-1} U_s < \kappa, \end{cases} \tag{2.2}$$

which is a special case of the the so-called Chair-Varshney rule [29], and is sometimes read as the k out of N fusion rule. The *Or* rule is defined as $\kappa = 1$ and the *And* rule as $\kappa = n - 1$.

Before proceeding, it is important to define the term log-concave as it will be critical to our analysis. Using the definition in [30]:

Definition 2.6. A function $g : \mathbb{R}^n \to \mathbb{R}$ is *logarithmically concave* (log-concave) if $g(x) > 0 \quad \forall x \in \mathrm{dom} g$ and $\log g$ is concave or $-\log g$ is convex. Defining $\log 0 = -\infty$, then g is log-concave if $g(x) \geq 0$ and the newly extended $\log g$ is concave.

At this stage, it is now possible to use (2.1) and define sufficient conditions for a UMP-DD test. A key point regarding the following theorem is that the conditions are not overly restrictive, as many problems in decentralized detection and cooperative spectrum sensing meet these constraints.

Theorem 2.7. *Let $T_s(X_s)$ be a sufficient statistic for the conditionally independent observations X_s having density $p_j^{X_s|\theta_s}$ where $p_j(T_s(X_s|\theta_s))$ is log-concave with respect to (w.r.t.) $T_s(X_s)$ for $j = 0, 1$. Let each $P_j^{\theta_s}$ be conditionally independent, log-concave w.r.t. θ_s, and have a convex support for all s. If $p_j(T_s(X_s); \boldsymbol{\mathcal{D}}_j) = \int_{\Theta_j} p_j(T_s(X_s|\theta_s)) p_j^{\theta_s} \mu(d\theta_s)$ is a smooth function (continuous derivatives), then equal quantizers, $\eta_1 = \eta_2 = \cdots = \eta_{n-1} = \eta$, are a UMP-DD test under the And or Or fusion rules and 1-bit quantized local decisions.*

The following Theorem by Prékopa will be used in the proof of Theorem 2.7. In [24] with Theorem VI, Prékopa showed that when $f(\boldsymbol{X}, \boldsymbol{Y})$, $\boldsymbol{X} \in \mathbb{R}^n$, $\boldsymbol{Y} \in \mathbb{R}^m$ is log-concave in $\mathbb{R}^{n+m}$ and with $\boldsymbol{A}$, a convex subset of $\mathbb{R}^m$, then

$$g(\boldsymbol{X}) = \int_{\boldsymbol{A}} f(\boldsymbol{X}, \boldsymbol{Y}) \, d\boldsymbol{y} \tag{2.3}$$

is log-concave over all of $\mathbb{R}^n$. Prékopa also presented an interesting corollary to this theorem in that log-concavity is passed from functions to their integrals as expanded upon in [25]. For example, consider the widely used Gaussian (normal) distribution, which is smooth, log-concave, and has a convex support. Prékopa's corollary implies all the associated integrals over a convex set of the Gaussian pdf are also log-concave. Thus, the Gaussian Cumulative Distribution Function (CDF), and the Gaussian Complementary CDF (CCDF) or so-called *Q-Function*, are all log-concave. This point is expanded in [25] for other log-concave densities.

The proof of Theorem 2.7 is by construction providing an approach to study UMP-DD, and is presented here instead of the appendix.

Proof. Let $p_j\left(T_s\left(X_s|\theta_s\right)\right)$, $p_j^{\theta_s}$ be log-concave w.r.t. $T_s\left(X_s\right)$ and θ_s respectively, and the support for $p_j^{\theta_s}$ be convex on a subset of Θ_j for $j = 0, 1$. Since the product of log-concave functions is log-concave, then $p_j\left(T_s\left(X_s|\theta_s\right)\right)p_j^{\theta_s}$ is log-concave. Thus, $p_j\left(T_s\left(X_s\right), \boldsymbol{\mathcal{D}}_j\right) = \int_{\Theta_j} p_j\left(T_s\left(X_s|\theta_s\right)\right)p_j^{\theta_s}\mu\left(d\theta_s\right)$ is log-concave by Prékopa Theorem VI. Note that most log-concave densities of interest have Θ_j as a connected subset of $\mathbb{R}$, which is both convex and concave.

The LRT $l\left(x\right)$ is monotone in $T_s\left(X_s\right)$ (cf. 2.4) and all densities are smooth; the critical region $\mathcal{C}_s$ for deciding H_1 is a connected interval $[\eta_s, \infty)$. Thus, the detection probability under the *And* fusion rule is

$$\begin{aligned} P_1\left(U_0 = 1\right) &= \prod_{s=1}^{n-1}\int_{\mathcal{C}_s} p_1\left(T_s\left(X_s\right), \boldsymbol{\mathcal{D}}_1\right)\,\mu\left(dx_s\right) \\ &= \prod_{s=1}^{n-1}\int_{\eta_s}^{\infty} p_1\left(T_s\left(X_s\right), \boldsymbol{\mathcal{D}}_1\right)\,\mu\left(dx_s\right) \\ &= \prod_{s=1}^{n-1}\int_{0}^{\infty} p_1\left(T_s\left(X_s\right) - \eta_s, \boldsymbol{\mathcal{D}}_1\right)\,\mu\left(dx_s\right) \end{aligned} \tag{2.4}$$

for each individual s, as $p_1\left(T_s\left(X_s\right) - \eta_s, \boldsymbol{\mathcal{D}}_1\right)$ is a log-concave function of $\left(T_s(x_s), \eta_s\right)$. Hence, $P_1\left(U_0 = 1\right)$ is a log-concave function of η_s. Similarly, the Type I error probability is log-concave in η_s with

$$P_0\left(U_0 = 1\right) = \prod_{s=1}^{n-1}\int_{0}^{\infty} p_0\left(T_s\left(X_s\right) - \eta_s, \boldsymbol{\mathcal{D}}_0\right)\,\mu\left(dx_s\right). \tag{2.5}$$

Therefore, convex minimization can be used to find the optimal $\boldsymbol{\eta} = \left\{\eta_s\right\}_{s\in V_n}$. With Neyman-Pearson criterion (cf. 1.2.3), the optimization problem is

$$\min_{\boldsymbol{\eta}} P_1\left(U_0 = 0\right) \quad \text{subject to } P_0\left(U_0 = 1\right) \leq \alpha, \tag{2.6}$$

and is a standard convex minimization problem without equality constraints [30]. Thus, the method of Lagrange multipliers can be applied to optimize (2.6). Let $\mathcal{L}\left(\boldsymbol{\eta}, \lambda\right)$ be a Lagrange multiplier function with

$$\mathcal{L}\left(\boldsymbol{\eta}, \lambda\right) = \sum_{s=1}^{n-1} -\log P_F + \lambda \left(\sum_{s=1}^{n-1} -\log\left(1 - P_F\right) + K \right), \tag{2.7}$$

where $P_D = \int_{\eta_s}^{\infty} p_1\left(T_s\left(X_s\right), \boldsymbol{\mathcal{D}}_1\right) \, \mu\left(dx_s\right)$, $P_F = \int_{\eta_s}^{\infty} p_0\left(T_s\left(X_s\right), \boldsymbol{\mathcal{D}}_0\right) \, \mu\left(dx_s\right)$, $K = \log\left(1-\alpha\right)$, and λ is a Lagrange multiplier. Setting $\frac{\partial}{\partial \eta_s}\mathcal{L}\left(\boldsymbol{\eta}, \lambda\right) = 0$ and simplifying

$$\begin{aligned} \lambda = & \frac{p_1\left(T_s\left(\eta_s\right), \boldsymbol{\mathcal{D}}_1\right) - p_1\left(T_s\left(\infty\right), \boldsymbol{\mathcal{D}}_1\right)}{1 - P_1^{X_s}\left(\eta_s\right)} \\ & \times \frac{P_0^{X_s}\left(\eta_s\right)}{p_0\left(T_s\left(\eta_s\right), \boldsymbol{\mathcal{D}}_0\right) - p_0\left(T_s\left(\infty\right), \boldsymbol{\mathcal{D}}_0\right)}, \end{aligned} \tag{2.8}$$

where the $p_j\left(T_s\left(\infty\right), \boldsymbol{\mathcal{D}}_j\right)$, $j = 0, 1$ in (2.8) are often zero for most log-concave distributions of interest.[1] With λ a fixed constant across all $n-1$ sensors, $\eta_1 = \eta_2 = \cdots = \eta_s = \eta$ is a solution to (2.6). To see this, consider the ratio of $\frac{\partial}{\partial \eta_s}\mathcal{L}\left(\boldsymbol{\eta}, \lambda\right)$ to $\frac{\partial}{\partial \eta_r}\mathcal{L}\left(\boldsymbol{\eta}, \lambda\right)$ for all $r \neq s$.

This solution can be verified as optimal using the Karush-Kuhn-Tucker (KKT) conditions, which are necessary and sufficient when the objective functions are smooth and convex [30]. With $\boldsymbol{\eta}^*$ and λ^* solutions to (2.7), the KKT conditions are

[1] As all distributions are smooth, the Lebesgue-Stieltjes and Reimann integrals are equivalent. Thus, Leibniz's rule for differentiation under the integral sign can be applied directly.

$$\text{primal: } -\log\left(1-P_F\left(\boldsymbol{\eta}^*\right)\right)+\log\left(1-\alpha\right)\leq 0$$
$$\text{dual: } \lambda^*\geq 0$$
$$\text{comp. slack.: } \lambda^*\left(-\log\left(1-P_F\left(\boldsymbol{\eta}^*\right)\right)+\log\left(1-\alpha\right)\right)=0$$
$$\text{vanishing grad. : } \nabla L\left(\boldsymbol{\eta}^*,\lambda^*\right)=0.$$

The log-concavity of the densities ensure $\lambda^* \geq 0$, $P_F \in [0,1]$ and equals $\alpha \in (0,1)$ by the smooth function assumption meeting the complementary slackness constraint, and the remaining two constraints follow by construction. The proof under *Or* fusion is similar. □

This result establishes a general framework to find and prove the existence of UMP-DD tests and is applicable to a broad set of problems, including those tests where the local sensor observations are not strictly i.i.d and deterministic. When $\theta_s = \theta \in \Theta$ is a fixed but unknown signal (i.e., singleton point mass), then the following corollary is applicable.

Corollary 2.8. *When the signal level at each sensor is fixed, but unknown, such that $\theta_s = \theta \in \Theta\ \forall\, s \in V_n$ so $p_j\left(T_s\left(X_s|\theta_s\right)\right) = p_j\left(T_s\left(X_s\right),\theta\right)$ and $p_j\left(T_s\left(X_s\right),\theta\right)$ is smooth log-concave with a convex support for $j \in \{0,1\}$, then thresholds $\eta_1 = \eta_2 = \cdots = \eta_{n-1} = \eta$ are a UMP-DD test under the And or Or fusion rules and 1-bit quantized local decisions..*

The results of Warren and Willett [20] can be viewed as a special case of Corollary 2.8, where θ is deterministic. Clearly, when one of these conditions in Corollary 2.8 is absent, then UMP-DD is not guaranteed. This claim will be justified using the Karlin-Rubin generalization counter example in Section 2.5 with i.i.d. Laplace noise.

We now address the more general k out of N fusion rule (cf. (2.2)).

Remark 2.9. Under the k out of N fusion rule and using the conditions in Theorem 2.7, then $\eta_1 = \eta_2 = \cdots = \eta_{n-1} = \eta$ are at least a local minimum or locally most powerful decentralized detection (LMP-DD) test. A full proof for this remark is similar to the proof of Theorem 2.7.

The LMP-DD designation implies that within an $\epsilon-$ball of $\eta_1 = \eta_2 = \cdots = \eta_{n-1} = \eta$ for $\{\epsilon : \epsilon > 0, \epsilon \in \mathbb{R}^{n-1}\}$ any other selection of the η_ss will result in a lower $P_1(U_0 = 1)$ for a given Neyman-Pearson size constraint α. Unfortunately, the LMP-DD label cannot be replaced with the more restrictive UMP-DD designation because both the $P_1(U_0 = 1)$ and $P_0(U_0 = 1)$ equations in (2.6) are a combinatorial summation of probabilities that are not guaranteed convex under the k out of N fusion rule, resulting in a non-convex optimization problem.

As the optimization is non-convex, then any solution to a Lagrange Optimization problem similar to Theorem 2.7 can only be declared locally optimal (LMP-DD) and not globally optimal (UMP-DD). However, we conjecture that this local optimum (LMP-DD) is in fact the global optimum (UMP-DD) under the general k out of N fusion rule, but a general proof supporting this conjecture remains an open research topic in decentralized detection.

The case for two sensors with common observations in Gaussian noise is exceptional, with the following results representing a complete UMP-DD solution that also supports randomization between the *And* and *Or* rules. Specifically, consider the k out of N fusion rule with $n = 3$ having the observation model

$$H_1 : X_s = \theta + W_s$$

$$H_0 : X_s = W_s,$$

where $W_s \overset{iid}{\sim} \mathcal{N}(0,1)$ follows a standard normal noise distribution and $\theta > 0$. Then, not only is $\eta_1 = \eta_2 = \eta$ the optimal local sensor quantizer under *And* and *Or* fusion, they are are also optimal using randomization of $k = 1$ and $k = 2$ fusion. Specifically, the randomized fusion model is

$$P_j(U_0 = 1) = \begin{cases} 1, & U_1 = U_2 = 1 \\ \upsilon, & U_1 = 1 \text{ or } U_2 = 1, \\ 0, & U_1 = U_2 = 0 \end{cases} \tag{2.9}$$

where $\upsilon \in [0,1]$ is a randomization constant randomly selecting $U_0 = 1$ with probability υ when either $U_1 = 1$ or $U_2 = 1$.

Theorem 2.10. *A two-sensor system, where both sensors observe the same fixed but unknown signal, $\theta > 0$, in zero mean i.i.d. Gaussian noise with 1-bit quantized local decisions and randomization between And and Or fusion has a UMP-DD with local quantizer thresholds $\eta_1 = \eta_2 = \eta$.*

The proof is given in Appendix A, highlighting the complexity of extensions beyond $n = 3$. However, this result does generalize the *And*, $\upsilon = 0$, and *Or*, $\upsilon = 1$, fusion results presented in [19].

2.4 UMP-DD in Decentralized Energy Detection

We now apply the results of the previous section to a collection of problems of considerable importance in decentralized detection. This includes cognitive radio models, with and without multipath fading, focused on energy detection observation models.

Consider a typical cognitive radio decentralized detection observation model.

$$\begin{aligned} H_1 &:, \quad X_s = \theta_s + W_s \\ H_0 &:, \quad X_s = W_s, \end{aligned} \tag{2.10}$$

where θ_s is the signal and W_s is the noise received by the nth cognitive radio. This model has been studied using energy detection, 1-bit quantized local decisions, and a fixed fusion rule (see [22, 28] and references therein). The choice of energy detection for cooperative spectrum sensing is common as it is simple, robust, and has relatively competitive detection performance. Optimal energy detection depends on an appropriate selection of the local decision rules and the fusion rule. Most authors assume equal quantizer thresholds and a *majority rule* form of the k out of N fusion rule. This is done without offering proof or justification as to the validity of this assumption [31–33]. Here, we offer a justification and show that energy detection with multipath fading can be UMP-DD and that the common assumption of equal quantizer thresholds is at least LMP-DD.

We consider observation distributions similar to that in [22] and [28]. Let $\theta_s \sim \mathcal{N}(0, \varphi_s^2)$ where the variance φ_s^2 is proportional to the signal energy at the sth sensor having zero mean Gaussian noise, $w_s \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. Then, a minimal sufficient statistic is

$$Y_s = \frac{1}{\sigma^2} \sum_{d=1}^{D} |X_s|^2 \quad \forall s \in V_n,$$

where D is determined from the time-bandwidth product [34] (Note that y_s is $T_s(X_s)$ in Theorem 2.7). Then, the test statistic, Y_s, follows the following distributions (i.e., $p_j(T_s(X_s|\theta_s))$) under each hypothesis (cf. [3, Sec. III.B])

$$Y_s \sim \begin{cases} \text{Gamma}\left(\frac{D}{2}, \frac{1}{2}\right) = \chi_D^2, & H_0, \\ \text{Gamma}\left(\frac{D}{2}, \frac{1}{2(1+\varsigma_s)}\right), & H_1, \end{cases} \tag{2.11}$$

where the sth sensor Signal-to-Noise Ratio (SNR) is $\varsigma_s = \varphi_s^2/\sigma^2$, χ_D^2 is the central chi-square distribution with D degrees of freedom, and $\text{Gamma}\,(a, b) = p\,(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}\, U\,(x)$ is a Gamma pdf with $\Gamma\,(a)$ the gamma function evaluated at a and $U\,(x)$ the unit step function. When the SNR at each sensor is equal, $\varsigma_s = \varsigma\ \forall s$ (i.e., $\varphi_s^2 = \varphi^2$), or the SNR follows a log-concave distribution, as can be the case for multipath fading, then these are UMP-DD problems. This follows because $\text{Gamma}\left(\frac{D}{2}, \cdot\right)$ is a smooth log-concave function for $D \geq 2$ [25]. When $\varphi_s^2 = \varphi^2\ \forall s$, the Type I and Type II error probabilities at each sensor are [28]

$$P_0\,(U_s = 1) = \frac{\Gamma\left(\frac{D}{2}, \frac{\eta}{2}\right)}{\Gamma\left(\frac{D}{2}\right)} \tag{2.12}$$

$$P_1\,(U_s = 0) = 1 - \frac{\Gamma\left(\frac{D}{2}, \frac{\eta}{2(1+\varsigma)}\right)}{\Gamma\left(\frac{D}{2}\right)} \tag{2.13}$$

for all s, where $\Gamma\,(\cdot, \cdot)$ is the upper incomplete gamma function. As described in the proof of Theorem 2.7 using Prékopas corollary, these are expected to both be log-concave [25] and for $\eta \geq 0$ is a smooth function of η. Thus, spectrum energy detection with fixed but unknown observations and no fading can be UMP-DD and is at least LMP-DD under counting fusion.

Next, we extend these results to the fading channel. Suppose that θ_s follows a Nakagami fading model with $E\,[\varsigma] = \bar{\varsigma}\ \forall s$. The density for the Nakagami channel received power follows a gamma distribution

$$p_{\text{Nak}}(\varsigma) = \frac{1}{\Gamma(m)} \left(\frac{m}{\bar{\varsigma}}\right)^m \upsilon_n^{m-1} \exp\left(-\frac{m}{\bar{\varsigma}}\varsigma_s\right), \quad \varsigma \geq 0, \tag{2.14}$$

where m is the Nakagami parameter and $\bar{\varsigma}$ is the average SNR. Note that $p_{\text{Nak}}(\varsigma)$ is $p_1^{\theta_s}$ in Theorem 2.7. With $m = 1$, then (2.14) describes Rayleigh fading, so it is a subset of the Nakagami analysis. Therefore, both fading models follow log-concave distribution[25]. Since both densities in (2.11) and the density in (2.14) are log-concave, then the average detection probability is also log-concave.

Therefore, cognitive radio spectrum energy detection with Nakagami fading can be UMP-DD and is at least LMP-DD with equal quantizer thresholds and the counting fusion rule. This result justifies using equal quantizer thresholds in decentralized detection problems, including cognitive radio cooperative spectrum sensing with hard decisions.

2.5 UMP-DD Families and Counter Examples

Before summarizing this chapter's results, we highlight a variety of decentralized detection problems that explore the breadth and depth of the theory presented, beyond spectrum sensing in cognitive radio. We typically define the noise, w_s, in (2.10) as Gaussian, but include other log-concave noise distributions meeting the constraints of Theorem 2.7 and Corollary 2.8 for generality.

Starting at the base, Corollary 2.8 provides sufficient conditions for UMP-DD when the observations are i.i.d. (i.e., $\theta_s = \theta \; \forall s$). The first application of this corollary in decentralized binary hypothesis testing with composite parameters is that of an i.i.d. shift in a fixed but unknown signal level. As discussed in the introduction, this is an extension to the results of Warren and Willett [20]. This model family is highlighted in the first three rows of Table 2.1 and includes the log-concave generalized Gaussian distribution (GGD) noise variant, where the GGD distribution is defined as $p\left(x;\mu,q,\, r\right)=\frac{r}{2q\Gamma(1/r)}e^{-(|x-\mu|/q)^{r}}$, $\Gamma\left(\cdot\right)$ is the gamma function, and r is restricted to the even integers to ensure the distribution is smooth. The second model family is energy detection with conditionally i.i.d observations. This is a simplified energy detection model with rather idealized wireless signals, as discussed in the previous section. This model family appears in the sixth row of Table 2.1. Again, the assumption of conditionally i.i.d. observations with no multipath fading is overly restrictive and leads us to the following more important result.

Theorem 2.7 provides sufficient conditions for UMP-DD when the signal levels are independently distributed, but not identical (i.e., $\theta_s \sim p_j^{\theta_s},\ j=0,1$). As with the previous paragraph, the first application of Theorem 2.7 is to those problems concerned with a shift in signal level between hypotheses. Consider the problem of detecting the presence or absence of a randomly positioned target in a set of sensors, where the target emits a weak signal level that follows an inverse exponential path loss model. Then, a model for the signal level at each sensor is the exponential distribution having mean τ with $\theta_s \sim \mathrm{Exp}\left(\tau\right)$. As the exponential distribution is log-concave, then this is a UMP-DD test. The idea is easily extended to other smooth log-concave signal level distributions and this model family appears in the fourth and fifth row of Table 2.1, where the Logistic distribution is $p\left(x;\mu,q\right)=\frac{e^{-\frac{(x-\mu)}{q}}}{q\left(1+e^{-\frac{(x-\mu)}{q}}\right)^{2}}$ and is similar to the Gaussian distribution with heavier tails. The second family of models studies an independent increase in received signal energy as described previously for energy detection with multipath fading. This model family appears in the final row of Table 2.1.

Table 2.1: UMP-DD Examples

Signal Model	Noise	Comment
$\theta_s = \theta \in \Theta$	$W_s \overset{iid}{\sim} \mathcal{N}(0,1)$	Corollary 2.8 and [20]
$\theta_s = \theta \in \Theta$	$W_s \overset{iid}{\sim} \text{Logistic}(0,1)$	Normalized Logistic Noise
$\theta_s = \theta \in \Theta$	$W_s \overset{iid}{\sim} \text{GGD}(0,1,r)$	Generalized Gaussian Noise ($r \geq 2$ Even)
$\theta_s \overset{iid}{\sim} \text{Exp}(\tau)$	$W_s \overset{iid}{\sim} \mathcal{N}(0,1)$	Exponential Signal Decay
$\theta_s \overset{iid}{\sim} \text{Exp}(\tau)$	$W_s \overset{iid}{\sim} \text{Logistic}(0,1)$	Logistic Noise
$\theta_s \overset{iid}{\sim} \mathcal{N}(0,\varphi^2)$	$W_s \overset{iid}{\sim} \mathcal{N}(0,1)$	Energy Detection
$\theta_s \sim \mathcal{N}(0,\varphi_s^2)$ log-concave Fading	$W_s \overset{iid}{\sim} \mathcal{N}(0,1)$	Energy Detection Multipath Fading

It is intuitively pleasing that equal quantizer thresholds are optimal for some cases. However, when the conditions of Theorem 2.7 or Corollary 2.8 are not met, then intuition can lead to a sub-optimal solution. The following counter example shows how intuition can fail and why the Karlin-Rubin Theorem cannot be directly extended to decentralized detection. Consider an i.i.d. observation model where the signal follows a Laplace distribution. Then, the test is no longer UMP-DD, because the Laplace pdf is not a smooth function.

Specifically, set $n = 3$ with $\theta_1 = \theta_2 = 1$, and the noise, w_s, following a zero mean i.i.d. Laplace distribution, such that

$$p^W(w|\mu,b) = \frac{1}{2b}e^{-\frac{|w-\mu|}{b}}, \tag{2.15}$$

with mean $\mu = 0$ and width parameter $b = 0.75$. We now consider the Receiver

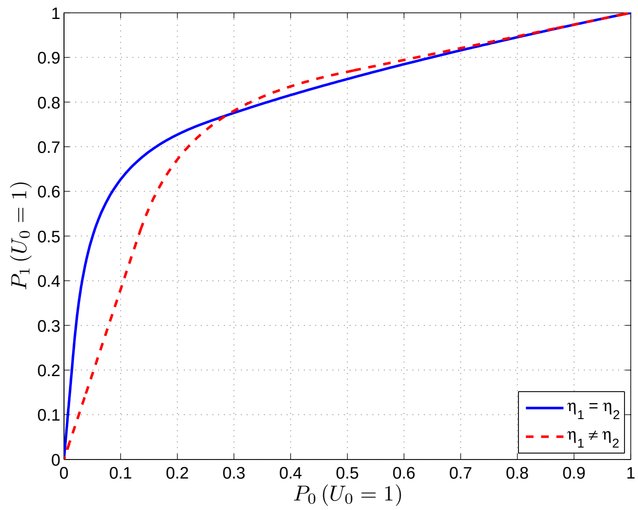


Figure 2.2: ROC Curves with Laplace Noise

Operating Characteristic (ROC) curves, which will show that asymmetric quantizer thresholds have better performance for some test sizes α. These results appear graphically in Figure 2.2 for $\eta_1 = \eta_2$, $\eta_1 \neq \eta_2$, and $\theta_1 = \theta_2 = 1$, where the latter η_2 is set such that the second sensor generates a constant detection rate of 99.9% for all η_1. These data clearly indicate that better detection performance for some false alarm rates can be achieved using asymmetric quantizer thresholds when $0.28 < P_0\,(U_0 = 1) < 0.80$. Thus, the UMP-DD does not exist for a fixed $\theta = 1$ with *And* fusion, so it clearly does not exist for all θ.

2.6 Summary

This chapter introduced UMP binary hypothesis testing with composite parameters into decentralized detection theory. This UMP-DD theory includes the introduction of a theorem that provided sufficient condition for UMP-DD when the sensor observations are randomly distributed and conditionally dependent given the hypothesis, but conditionally independent given the unknown composite parameters. A corollary provides the required sufficient conditions when the parameters are fixed but unknown. The traditional deterministic signaling model explored by previous authors can be considered a special case of this fixed but unknown UMP-DD problem. In addition, we showed that centralized UMP theory is insufficient for determining UMP-DD, including extensions of the Karlin-Rubin Theorem to decentralized detection.

The results introduced were dependent on either *And* or *Or* data fusion, but it was shown that the more generalized Counting fusion rule (k out of N) is at least LMP-DD, with a proof of UMP-DD an open research topic. Nevertheless, a complete UMP-DD solution was shown to exist for the two-sensor case, where a randomized fusion rule is allowed and the noise is zero mean i.i.d. Gaussian. For this case, randomization of the k out of N fusion rule remains UMP-DD with equal quantizers and fixed but unknown signals. These methods were then used to justify the selection of equal quantizer thresholds in a multitude of inference regimes, including spectrum sensing in cognitive radio using energy detection methods under realistic channel models such as Nakagami or Rayleigh fading.

This work established a new class of decentralized detection problems and in doing so significantly advanced decentralized detection research. It provides justification for selection of equal quantizer thresholds in complex decentralized testing systems that suffer multipath fading and non-Gaussian noise. It also showed that when there exist observation models that violate at least one of the sufficient conditions offered, then equal quantizer thresholds are not always optimal.

CHAPTER 3

RELAY NETWORKS, TANDEM CONFIGURATIONS, AND UMP-DD

3.1 Introduction

This chapter considers both UMP-DD and asymptotic detection performance for the case when the height of the single-rooted tree is greater than unity $(h > 1)$. A benefit of configurations with $h > 1$ is that local sensor decisions can be aggregated and summarized along the branches towards the root (cf. Section 1.4) with a significant reduction in energy consumption possible. This is because the average physical distance between nodes is reduced, resulting in lower transmission power requirements to achieve equivalent detection performance with non-ideal communication channels. A drawback of this approach is that summarizing the aggregated local decisions results in lost information relative to centralized detection and parallel decentralized detection systems. Intuitively, it is expected that this loss in information always results in an inferior detection performance, but the question is "Does intuition match reality?" This is a question that we study and answer for relay networks relative to the parallel network of Figure 1.1.

Similar to Chapter 2, we make the following assumption:

Assumption 3.1. *The sensor observations are conditionally independent given the composite parameter under each hypothesis.*

Using this assumption, we will show that the Type II error exponent for a relay network can be equivalent to that of a parallel configuration under the Neyman-Pearson framework and certain other conditions. These conditions include 1-bit quantization and the ability to group sensors into sub-classes where sensor observations within a class are conditionally i.i.d. in an average sense and at least conditionally independent across classes. We then progress to our ultimate goal and study whether either the relay or tandem network are UMP-DD, where the tandem configuration appears in Figure 1.1b.

A directed tree network having UMP-DD performance and a Type II error probability that decays exponentially at the same rate as the parallel configuration has many advantages. This includes a potential for lower transmit energy consumption, as the physical distance between nodes can be reduced. At the same time, a lower transmit power requirement implies a reduction in interference energy, which is beneficial, for example, to collaborative spectrum sensing in cognitive radio. There, local information is typically transmitted to a fusion node in order to make an aggregate decision regarding the primary users' activity. A challenge with this model is that local decision transmission must not impact the primary user operation, where *underlay* networks offer one solution (see [35] for a discussion on *underlay* networks). Certainly a directed tree with with low transmission power over directed (beamformed) links offers one such *underlay* solution.

3.1.1 Prior Research

Our study is limited to bounded height tree networks using asymptotic methods to compare performance. The parallel configuration discussed in Chapter 2 is one such configuration, where all the leaves are sensors and the root is the fusion center. The parallel configuration and its analysis with Bayesian and Neyman-Pearson criterion is well researched, including the study of the non-ideal channel [6, 14, 15, 18, 36–41] and references therein. Additionally, it is known that when the communication channel is ideal with conditionally independent observations at the leaves, the parallel configuration error probability decays exponentially [13, 18].

The first bounded height tree network we study is the relay network. This network consists of relay nodes that perform no local sensing when summarizing their predecessor information, as they forward that summary to the next node on the way to the root. It is known that a bounded height directed relay tree with conditionally i.i.d. observation has a Type II error probability that decays exponentially and is equivalent to that of the parallel configuration under the Neyman-Pearson criterion subject to some mild constraints [14]. The critical constraint is that the leaves (e.g., sensors) dominate the network as n increases or $\frac{|S_n|}{n} \to 1$ where $S_n = \{s : s \in V_n\}$. Again, V_n is the set of all nodes in the tree.

Under the Bayesian criterion, the bounded height relay tree is known to have an error exponent that is inferior to the parallel configuration [39]. It is also known that when the height of the tree grows without bound, the unbounded relay tree has a Type II error probability exponent that again is inferior to the parallel configuration [14], with error probability bounds presented in [42] for a balanced binary relay tree. Thus, we limit this study to the parallel configuration and bounded height relay trees with binary decisions under the Neyman-Pearson criterion because the remaining configurations listed have an error exponent that is inferior to the parallel configuration.

The tandem network, where non-leaf nodes augment the summary information from their predecessors with their own observations has been studied in [39] and [43–47] under both the Bayesian and Neyman-Pearson frameworks, respectively. In both cases, the Type II error probability was shown to decay sub-exponentially and, under some conditions, does not decay to zero, even as the $n \to \infty$. One sufficient condition for the Type II error probability to decay to zero under the Neyman-Pearson framework is that the ROC curve have infinite initial slope and zero final slope [45].

3.1.2 Chapter Goals

The initial contribution is establishing a set of sufficient conditions for when a bounded height relay tree and the parallel configuration achieve equivalent error exponents with conditionally independent, but not identical observations. These conditions include a large number of sensors, 1-bit quantization, and the ability to group the sensors into N finite classes. The class terminology we use is similar to that mentioned in [48]. Here, a class is defined as a collection of sensors that apply equivalent decision rules and the local observations are conditionally independent. Notice that this work is differentiated relative to the conditionally i.i.d. study in [14], as we only require conditionally independent sensor observations between classes. We do not study the tandem configuration error exponent under class notation, as the conditionally i.i.d. sensor observation case has already been shown to decay sub-exponentially [39, 45].

Under the Neyman-Pearson framework, we ensure the Type I error probability does not exceed a defined threshold and minimize the error exponent

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

where β^n is the fusion center Type II error probability and n represents the total number of nodes including f. Similar to the notation in [49], the error exponents are negative numbers, with the magnitude of $|g|$ referred to as the rate of decay of the Type II error probability.

After establishing sufficient conditions for a class-based relay network to achieve an equivalent error exponent to the parallel configuration in Section 3.3.1, we provide a moderately complex example in Section 3.3.4. This is done to highlight one method to study these complex relay trees since they typically do not have closed form solutions, even with 1-bit quantization. This technique can be used to study the impact on transmission power for a relay network with non-ideal communication channels operating below the asymptotic regime, for example.

The results of Section 3.3 and in [14] are then coupled with the results of Chapter 2 to show that the relay tree can be UMP-DD in Section 3.4. Similarly, we show in Section 3.5 that the tandem network is not UMP-DD. This claim is developed at the atom level for tandem configuration analysis, where an atom is a single leaf with a single tandem node, as depicted later in Figure 3.1d.

3.2 General Relay and Tandem Network Formulation

Using similar notation as Tay et al. in [14], we consider a binary decentralized detection problem with $n-1$ leaves and relays, and a fusion center denoted as f. As we introduced in Section 1.2, there are two probability measures P_0 and P_1 associated with hypotheses H_0 and H_1, respectively, on an observation space $(\Omega, \mathcal{F}_\sigma)$.

Throughout this chapter, we assume

Assumption 3.2. P_0 *and* P_1 *are absolutely continuous relative to one another, thus they are equivalent* ($P_0 \equiv P_1$).

There are N different classes of observation models, where sensor observations are conditionally i.i.d. within a class and at least conditionally independent across classes given a hypothesis H_j $j = 0, 1$. Each sensor s within the kth class observes a random variable $X_s \in \mathcal{X}_k$ for $k = 1, 2, \ldots, N$. The random variable X_s has a marginal distribution $P_j^{X_s}$ conditioned on hypothesis H_j, $j = 0, 1$.

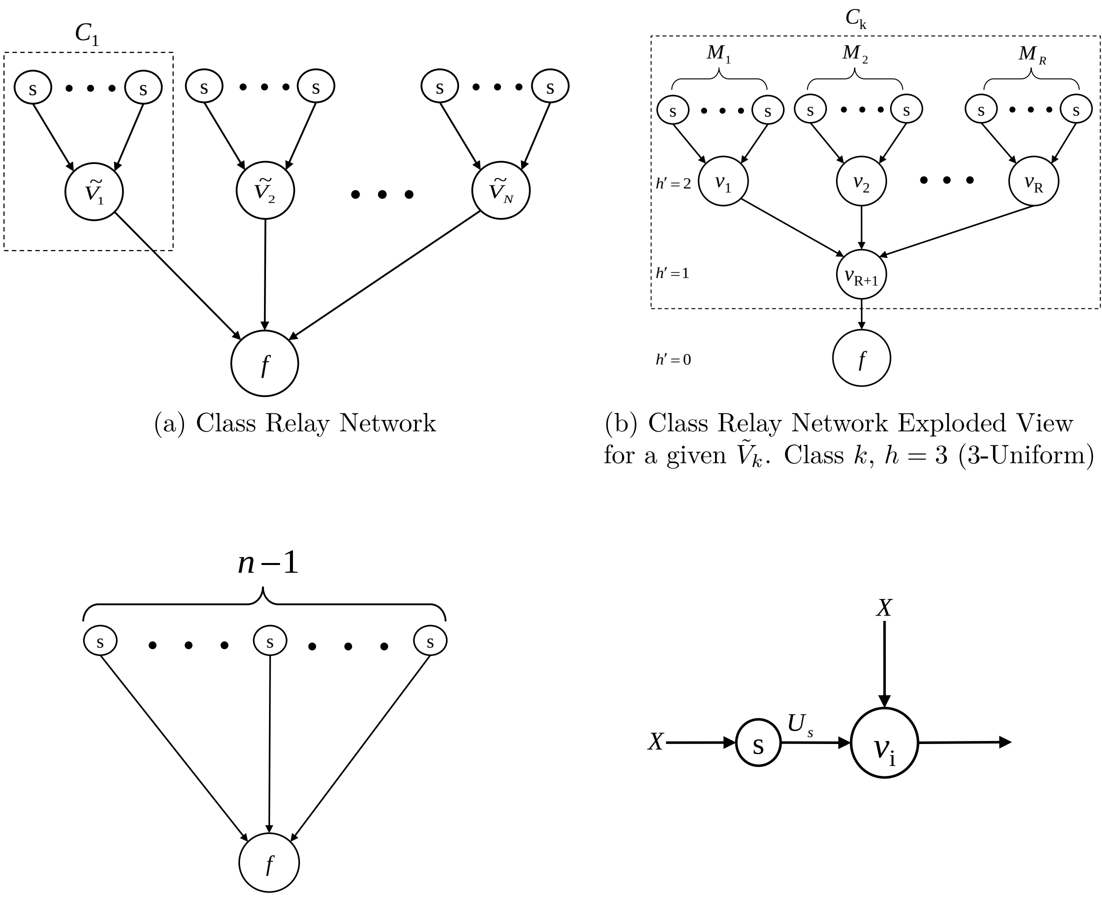
3.2.1 Tree Networks

As described in Section 1.4, we consider a directed tree network $T_n = (V_n, A_n)$ with all arcs pointing from the leaves towards the root, f, and n total nodes represented as V_n. The leaves of T_n are sensors that use 1-bit quantization and transmit their message to the next node along a single directed arc. Relay nodes only perform a relay operation, while Tandem nodes augment the received messages with their own local observation, which is denoted by the random variable $X \in \Omega$.

The height of T_n is the length of the longest path in terms of arcs traveled from a leaf to the root and will be represented by h. Trees where each leaf has the same height are known as $h-$uniform. Tay et al. in [14] showed that by using a "height uniformization procedure" that an upper bound on the error exponent for $h-$uniform relay trees readily translates to an upper bound on general relay trees of height h. Thus, we will focus our analysis on h-uniform trees, realizing that the same "uniformization procedure" can be applied at a class level, which allows conclusions made for $h-$uniform trees to be applied to non-uniform trees.

As the tandem network in our study is limited to the atom, the following notation will only apply to relay networks. Let C_k denote the set of all conditionally i.i.d. sensors in the kth class. Let $\tilde{V}_k$ denote the collection of all relay nodes, v_i, associated with the kth class. The $\tilde{V}_k$ notation eliminates the notational complexity associated with multiple sub-indexes required to define a relays location (e.g., class and height) in the tree. The Type I error probability for UMP-DD analysis in Section 3.4, however, does require a relationship to height and will be denoted by h'. Specifically, h' is the number of directed arcs traveled from f to a certain level of relay nodes in a relay tree along a given branch. For example, $h' = 0$ is f, $h' = 1$ is the relay level directly preceding f, and $h' = h_B - 1$ are the relays with only sensors as their predecessors along branch B with $h = \max\{h_B : B \in A_n\}$. This class relay network is depicted in Figure 3.1a, with an exploded view for the kth class in Figure 3.1b for an arbitrary $h = 3$ uniform relay tree.

We now augment the notation from Section 1.4 to support classes. Each $s \in C_k$ makes a binary decision U_s via a LRT γ_s^k, such that $U_s = \gamma_s^k(X_s)$, where $\gamma_s^k : \mathcal{X}_i \to \{0,1\}$. Node v_i then makes a decision U_i, again using a LRT γ_i, such that $U_i = \gamma_i(\{U_s : s \in M_i\})$ for $h' = h - 1$ or $U_i = \gamma_i(\{U_v : v_i \in M_i\})$, otherwise, where $\gamma_i : \{0,1\}^{|M_i|} \to \{0,1\}$. The relay tree decision rule γ_0 at f maps $\{0,1\}^{|M_0|} \to \{0,1\}$ to generate U_o. Any directed relay tree of height h (not necessarily uniform) meeting the preceding constraints will be referred to as a h binary relay tree (hBRT). When the same tree is $h-$uniform, the tree will be denoted as a $h-$uniform binary relay tree (hUBRT). The parallel configuration is duplicated for convenience in Figure 3.1c and has a fusion rule γ_0 that is unchanged under class notation, mapping $\{0,1\}^{n-1} \to \{0,1\}$.



(a) Class Relay Network

(b) Class Relay Network Exploded View for a given $\tilde{V}_k$. Class k, $h = 3$ (3-Uniform)

(c) Parallel Configuration

(d) Tandem Configuration Atom

Figure 3.1: Directed Single Rooted Trees

3.3 Relay Network Performance Analysis

This section studies the asymptotic detection performance between the hUBRT and the parallel configuration using a decentralized detection Neyman-Pearson framework where all sensor observation are conditionally independent, but not necessarily conditionally i.i.d.

3.3.1 Neyman-Pearson Hypothesis Testing

Under the Neyman-Pearson framework, the Type I error probability is constrained as $P_0(U_0 = 1) \leq \alpha$ for some $\alpha \in (0, 1)$. The goal is to minimize the Type II error probability $P_1(U_0 = 0)$ across all binary hypothesis testing strategies that meet the Type I constraint. Let β_R^* be the infimum of $P_1(U_0 = 0)$ over all admissible hUBRT strategies and similarly for β_P^* associated with the parallel structure. The Type II error exponents for each T_n are defined by

$$g_R^* = \limsup_{n\to\infty} \frac{1}{n} \log \beta_R^*,$$

for a bounded height tree where $\lim_{n\to\infty} \frac{N}{n} = 0$ and

$$g_P^* = \limsup_{n\to\infty} \frac{1}{n} \log \beta_P^*.$$

We desire to show that $g_R^* = g_P^*$ and will do so as follows. First, $g_P^* \leq g_R^*$ for all test strategies with an equivalent number of nodes. To see this, let Γ_R be the set of all binary relay tree test strategies and Γ_P the set for all parallel test strategies. Clearly, $\Gamma_R \subset \Gamma_P$, which implies $g_P^* \leq g_R^*$. Second, since $g_P^* \leq g_R^*$, it is sufficient to show that if there exists a strategy β_R with $g_R = g_P^*$, then the optimal strategy β_R^*

must also result in $g_P^* = g_R^*$. We will use the Kullback-Leibler divergence (KLD) to prove this second step. The KLD between two probability measures is

$$D\left(P||Q\right) = E^P\left[\log\frac{dP}{dQ}\right], \tag{3.1}$$

where E^P is the expectation relative to P, $P \equiv Q$, and ${}^{dP}/_{dQ}$ is the Radon-Nikodym derivative [10].

3.3.2 hBRT and the Parallel Configuration Error Exponent

The error exponents for both the hBRT and parallel configurations are now studied under the Neyman-Pearson framework. The key result is the following Lemma.

Lemma 3.3. *Consider a hBRT consisting of N classes of sensors with i.i.d. observations within a class and independent observations across classes. Then, $g_P^* = g_R^*$ when the Type I error probability is $\alpha \in (0,1)$, $\frac{S_n}{n} \to 1$, and the total number of nodes are equal between the hBRT and parallel configurations.*

Proof. Without loss of generality, the relay tree error probability exponent study can be limited to the special case 2UBRT (2-uniform Binary Relay Tree) configuration. To see this, consider that each class is equivalent to the conditionally i.i.d. bounded height h-uniform analysis presented in [14], when $\frac{S_n}{n} \to 1$ where $S_n = \{s : s \in V_n\}$ as defined previously. Additionally, a quick analysis of the hUBRT configuration in Figure 3.1b using $U_i = 0$ iff $\sum_{v_j \in M_i} U_j = 0$ (cf. (2.2) with $\kappa = 1$) shows that this must be the case for a bounded height tree with the $\tilde{V}_k$ constraint on decision aggregation.

Let the 2UBRT fusion rule decide H_0 if and only if $\sum_{i=1}^{N} U_i = 0$ and the v_i be ordered with the classes (i.e., $i = k$). Then, for this special case

$$\begin{aligned} g_R &= \limsup_{n\to\infty} \frac{1}{n} \log P_1 \left(\{U_i = 0 \,:\, i = 1, 2, \ldots N\}\right) \\ &= \sum_{i=1}^{N} \limsup_{n\to\infty} \frac{1}{m_i} \log P_1 \left(U_i = 0\right), \end{aligned} \tag{3.2}$$

where $m_i = |M_i|$ and each m_i increases proportionally with $\frac{n-N}{N}$.

Suppose that the Type I and Type II error probabilities for all sensors within C_k are a_k and b_k, respectively. Then, the KLD at v_k associated with $P_1\left(U_i = 0\right)$ is

$$D\left(P_0^{v_i} || P_1^{v_i}\right) = m_k \left(a_k \log\left(\frac{a_k}{b'_k}\right) + a'_k \log\left(\frac{a'_k}{b_k}\right)\right), \tag{3.3}$$

where $a'_k = 1 - a_k$, $b'_k = 1 - b_k$, and $m_i = m_k$ since $i \triangleq k$. Equations (3.2) and (3.3) imply

$$g_R = -\sum_{k=1}^{N} \left(a_k \log\left(\frac{a_k}{b'_k}\right) + a'_k \log\left(\frac{a'_k}{b_k}\right)\right). \tag{3.4}$$

We now consider the parallel configuration using a similar analysis with a_k, b_k, and m_k defined as before $\forall s \in C_k$. After straightforward but tedious calculations, the KLD at the root f is

$$D\left(P_0^f || P_1^f\right) = \sum_{k=1}^{N} \left(a_k \log\left(\frac{a_k}{b'_k}\right) + a'_k \log\left(\frac{a'_k}{b_k}\right)\right). \tag{3.5}$$

The proof of (3.5) is given in Appendix B. As the Type II error probability decays exponentially within the parallel configuration, we have $g_P^* = -D\left(P_0^f || P_1^f\right)$ where f represents the root [13]. Thus, $g_P^* = g_R$, which implies $g_P^* = g_R^*$ as desired for the hUBRT configuration.

The Tay et al. results in [14] on "height uniformization" and the upper bound on the general tree error exponent can be applied to each $\tilde{V}_i$ and its predecessor leafs. The desired $g_P^* = g_R^*$ for the hBRT configuration follows. □

The following corollary follows directly from the proof of Lemma 3.3.

Corollary 3.4. *Consider a hUBRT with N classes of sensors with i.i.d. observations within a class and independent observations across classes. Then, $g_P^* = g_R^*$ when the Type I error probability is $\alpha \in (0,1)$, $\frac{S_n}{n} \to 1$, and the total number of nodes are equal between the hUBRT and parallel configurations.*

3.3.3 Sensor Decision and Node Fusion Rules

Achieving these optimal error exponents in practice is not necessarily straightforward. Within the hBRT configuration, the Type I error must be properly distributed across classes, the class fusion rule defined, and the sensor threshold for each class determined. We now shift our focus to address these challenges.

The LLRT at each sensor $s \in V_n$, based solely on its conditionally independent observation X_s is

$$L_s(x) = \log \frac{dP_1^{X_s}}{dP_0^{X_s}}(x) = \log \frac{p_1^{X_s}(x)}{p_0^{X_s}(x)} \tag{3.6}$$

for $x \in \mathcal{X} = \bigcup_{k=1}^{N} \mathcal{X}_k$. Since X_s is i.i.d. for all $s \in C_k$, the $L_s(x)$ is compared to a common threshold η^k deciding H_1 if $L_s(x) > \eta^k$, H_0 if $L_s(x) < \eta^k$, and possibly randomizing if $L_s(x) = \eta^k$.

Similarly, let $\boldsymbol{U}^k = \{U_s\}_{s \in C_k}$ be the decisions received at node v_i $\forall s \in C_k$ under both P_0 and P_1. Then, the optimal decision fusion at node v_i is also a LLRT

$$L_{v_i}(\boldsymbol{u}) = \log \frac{dP_1^{v_i}}{dP_0^{v_i}}\left(\boldsymbol{u}^k\right) = \log \frac{p_1^{v_i}\left(\boldsymbol{u}^k\right)}{p_0^{v_i}\left(\boldsymbol{u}^k\right)}, \tag{3.7}$$

which is compared to a threshold κ_i to decide H_1 versus H_0. The optimal LLRT at node v_i can be shown to be equivalent to the so-called counting rule (i.e., $\sum_{s \in C_k} U_s \gtrless$

κ_i), where κ_i is a non-negative integer, since $\forall s \in C_k$, $U_s = \gamma_s^k(X_s)$ is an i.i.d. binary random variable. Let α_k be the required Type I error probability for C_k to obtain the global α Type I constraint. Then, κ_i is selected to obtain α_k, which may require randomization between adjacent integer thresholds.

The optimal parallel configuration fusion rule is also a LLRT similar to that in Equation (3.7), but now with $\boldsymbol{U}^n = \left\{\boldsymbol{U}^k\right\}_{k=1}^N$, the set containing the local decisions $\forall s \in V_n$. The LLRT at f is

$$L_{f,n}(\boldsymbol{u}) = \log \frac{dP_1^f}{dP_0^f}(\boldsymbol{u}^n) = \log \frac{p_1^f(\boldsymbol{u}^n)}{p_0^f(\boldsymbol{u}^n)} \tag{3.8}$$

and is compared to a threshold $\tau \in \mathbb{R}$ to decide H_1 versus H_0. The optimal LLRT at f is the Chair-Varshney (C-V) rule $\sum_{k=1}^N \sum_{s \in C_k} U_s \log\left[\frac{b'_k}{a_k}\frac{a'_k}{b_k}\right] \gtrless \tau - \sum_{k=1}^N m_k \log \frac{b_k}{a'_k} = \tau'$ [6]. Essentially, this fusion rule compares the weighted sum of all the ones received at f against a threshold τ', deciding H_1 if the sum is greater than τ', H_0 if it is less, and when equal may require randomization to meet α.

3.3.4 Numerical Example

The goal of this numerical example is to establish techniques that can be used to study (3.7) and (3.8). We will show that these techniques achieve the optimal Type II error exponent and in doing so establish one method to study the detection performance and transmission power requirements when the wireless links are not ideal. Certainly by Corollary 3.4, we know $g_P^* = g_R^*$ for all appropriately defined hBRTs. Furthermore, we know that g_R (cf. (3.4)) achieves g_R^*, so any selection of N, α, a_k, b_k will suffice and the study can be limited to a single non-trivial example. We set $h = 2$ (i.e., 2UBRT), $N = 5$, $\alpha = 0.1$, $\boldsymbol{a} = [0.20, 0.65, 0.45, 0.15, 0.50]^T$, and $\boldsymbol{b} = [0.7, 0.2, 0.4, 0.7, 0.3]^T$, where $\boldsymbol{a} = [a_1, \ldots a_N]$ and similarly for $\boldsymbol{b}$.

Notice that the decision rule for (3.8) is a weighted sum of independent Bernoulli random variables. Let this weighted sum at f be represented by $Y = \sum_{k=1}^{N} \sum_{s \in C_k} w_k U_s$, where $w_k = \log\left[\frac{b'_k}{a_k}\frac{a'_k}{b_k}\right]$. Equivalently, Y can be defined as a weighted sum of independent Binomial random variables with parameters m_i and a_k or b_k. Then, the densities $P_0^f(y)$ and $P_1^f(y)$ do not have a general closed form solution.[1] However, when n is sufficiently large (i.e., m_k large) with $0 < a_k < 1 - b_k \; \forall k$, the parallel configuration distributions at f are approximately Gaussian

$$H_0 : Y \sim \mathcal{N}\left(\sum_{k=1}^{N} w_k m_k a_k,\; \sum_{k=1}^{N} w_k^2 m_k a_k \cdot a'_k\right), \tag{3.9}$$

and similarly for $P_1^f(y)$ replacing a_k, a'_k with b'_k, b_k, respectively by the DeMoivre-Laplace Theorem [50].

Fortunately, the distributions at each fusion node v_i are well defined for the 2UBRT configuration. Specifically, $P_0^{v_i}(y_i)$ is Binomial with parameters m_k and a_k, where $Y_k = \sum_{s \in C_k} U_s$ and $P_1^{v_i}(y_i)$ is similar. However, determining the threshold κ_i requires α_k. One method of determining α_k is to set $\alpha_j = \alpha_k \; \forall j, k \in \{1, 2, \ldots, N\}$ using

$$\alpha_k = 1 - (1 - \alpha)^{1/N}. \tag{3.10}$$

Since 3.4 places no constraint on α_k beyond the global α requirement, we expect to still achieve the Optimal Type II error exponent. The following numerical example supports this claim.

Within Figure 3.2, we plot the magnitude of the normalized logarithm of the Type

[1] Henceforth we assume the $\sigma-$finite measure μ is the Lebesgue measure when the random variable is continuous and the counting measure when discrete.

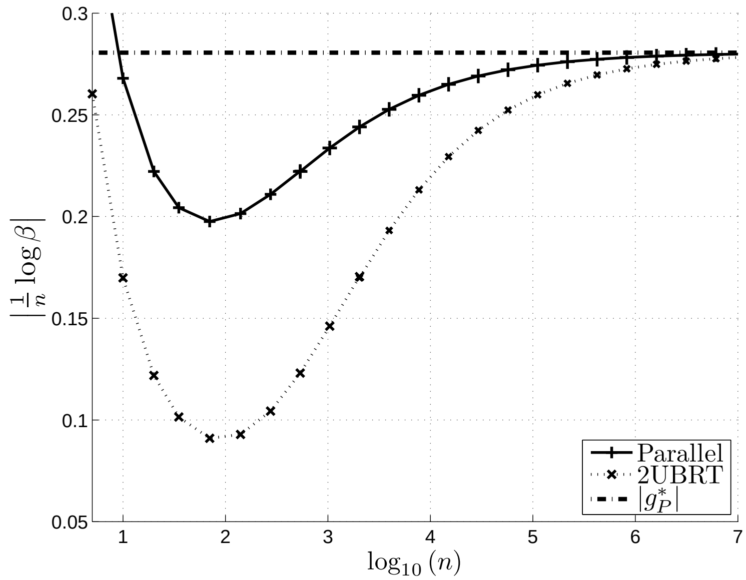


Figure 3.2: 2UBRT Asymptotic Performance Analysis: $N = 5$, with $\alpha = 0.1$, $h = 2$, $\boldsymbol{a} = [0.20,\ 0.65,\ 0.45,\ 0.15,\ 0.50]^T$, $\boldsymbol{b} = [0.7,\ 0.2,\ 0.4,\ 0.7,\ 0.3]^T$

II error probability, $\left|\frac{1}{n}\log\beta\right|$, versus n for both the parallel and 2UBRT configurations using the methods described in this subsection (cf. (3.9)-(3.10)). When n is large, the Type II error probability at relay v_i is modeled with a Gaussian approximation similar to (3.9) with Mills ratio used to bound the complementary cumulative distribution function (CCDF) for numerical simulation purposes. The specifics for calculating the Type II error probability when n is large is given in Appendix C. The asymptotic convergence point, $|g_P^*|$, is depicted as a horizontal line in Figure 3.2 for reference.

Figure 3.2 supports three main conclusions. First, the Type II error probability decays at the same exponential rate for the 2UBRT and parallel configurations as expected. Second, $\beta_P \leq \beta_h$ as expected, where β_P is the Type II error probability for the parallel configuration and β_h the Type II error probability for the hBRT configuration. Finally, when $a_j \neq a_k\ \forall j,k$ and $b_j \neq b_k\ \forall j,k$, then the inequality becomes strict in the non-asymptotic regime $\beta_P < \beta_h$.

3.4 Relay Networks and UMP-DD

Given that hBRT configurations have an equivalent error exponent to the parallel configuration when the conditions of Lemma 3.4 are met, we would not like to understand if UMP-DD can also be achieved under the same set of constraints. The following study builds upon the UMP-DD results of Chapter 2 and in particular Theorem 2.7, as related to a composite parameter $\theta_k \in \Theta_k$ associated with class C_k and $\alpha \in (0,1)$, which is a constraint on the size of the test.

3.4.1 Achieving UMP-DD in a hBRT Configuration

Suppose $T_s^k(X_s)$ is a sufficient statistic for the class C_k sensor LRT in (3.6), but with densities $p_j^{X_s|\theta_k}$, $j = 0,1$. Further, let $p_j\left(T_s^k(X_s|\theta_k)\right)$ be log-concave w.r.t. $T_s^k(X_s)$ for $j = 0,1$ and $\forall k$. Let each $P_j^{\theta_k}$ be conditionally independent, log-concave w.r.t. θ_k, and have a convex support for all s. When $p_j\left(T_s^k(X_s);\boldsymbol{\mathcal{D}}_j\right) = \int_{\Theta_k} p_j\left(T_s^k(X_s|\theta_k)\right) p_j^{\theta_k}\mu(d\theta_k)$ is a smooth function (continuous derivatives), then by Theorem 2.7 equal quantizers, $\eta_1^k = \eta_2^k = \cdots = \eta^k_{|\{s\,:\,s\in C_k\}|} = \eta^k$ are a UMP-DD test at the level $h' = h_B - 1$ under the *And* or *Or* fusion rules.

Lemma 3.5. *If all relays at level $h' = h_B - 1$ in a hBRT configuration are UMP-DD given the conditions in the preceding paragraph (i.e., per Theorem 2.7), then a hBRT configuration is UMP-DD.*

Proof. Without loss of generality, we can assume the tree is h-Uniform, which is only made to simplify the notation of the subsequent Markov Chain. Suppose all relays at level $h' = h - 1$ are UMP-DD, where $h_B = h\ \forall B \in A_n$ by assumption. Next, notice that $H_j \to \theta \to X_s \to U_s \to \boldsymbol{U}_i(h' = h-1) \to \cdots \to \boldsymbol{U}_i(h' = 1) \to \boldsymbol{U}^n(h' = 0) \to U_0$ forms a Markov Chain. Then, by the properties of a Markov Chain, any states following the state that is UMP-DD are also UMP-DD. Thus, the hUBRT configuration is UMP-DD. □

Similar to Remark 2.9, when the relays at level $h' = h_B - 1$ are LMP-DD using the more generic counting fusion rule (cf. 3.7), then the hBRT configuration is also at least LMP-DD.

3.4.2 Achieving hBRT UMP-DD and Asymptotically Optimal Performance

When the hBRT configuration is UMP-DD, it is also desirable to achieve an error exponent that is equivalent to the parallel configuration (i.e., $g_R^* = g_P^*$). This goal can be realized by applying an approach similar to that applied in Section 3.3.1 to show the hBRT configuration error exponent, g_R^* is equivalent to g_P^*.

Suppose the fusion rule at f and each relay node decides $U_i = 0$ iff $\sum_{v_j \in M_i} U_j = 0$ (cf. (2.2) with $\kappa = 1$). Then, $g_R^* = g_P^*$ occurs if an appropriate size can be realized at each h' through the tree. Let $\alpha_i(h')$ be the Type I error probability constraint for relay v_i with $\alpha_0(h' = 0) \triangleq \alpha$, where α_i is a function of h'. Then, $\alpha_i(h')$ can be defined recursively as

$$\alpha_i(h') = 1 - (1 - \alpha_j(h' - 1))^{\frac{1}{|M_i|}} \ \forall \{v_i : v_i \text{ is at height } h' \leq h_B\}, \tag{3.11}$$

where v_j is the successor relay node $\forall v_i \in M_j$, similar to (3.10).

The proceeding formulation does achieve the desired result of asymptotically optimal UMP-DD hBRT performance. This is a rather non-intuitive result, as the fusion rules and the determination of each $\alpha_i(h')$ are seemingly arbitrary. Notice that one benefit of such an approach is that the system is relatively insensitive to changes in the statistics at the leaf level (i.e., non-stationary). With that said, application of the optimal fusion rules (cf. (3.7)) for relays at $h' < h_B - 1$ and another selection method of each $\alpha_i(h')$ may result in improved detection performance outside the asymptotic regime. However, any improvements realized come at the cost of making the system more sensitive to changing statistics at the leaf level.

With the UMP-DD designation, the hBRT configuration has all the advantages discussed in Chapter 2. This includes support for composite parameters where the observation models given H_1 are no longer conditionally independent. It supports a multitude of noise models (cf. Table 2.1), multipath fading environments, and, perhaps most importantly, the ability for sensors having different locations, technology, or interference to be readily modeled via the class hierarchy. It also allows a reduction in network energy usage, as the relay network tends to reduce the average transmission power required.

3.5 Tandem Networks and UMP-DD

This section highlights that tandem networks are not UMP-DD as the optimal decision at any node v_i is a function of the composite parameter θ. Consider the tandem network atom presented in Figure 3.1d. The LRT for the atom is (cf. (1.2) and (3.6))

$$l(x, u_s) = \frac{p_1^{X,U_s|\theta}(x, u_s)}{p_0^{X,U_s|\theta}(x, u_s)}, \tag{3.12}$$

where the relay v_is observation is X and the decision of s is U_s. As U_s is a binary random variable,

$$l\left(x,u_{s}\right)=\frac{p_{1}^{X|\theta}\left(x\right)\left(p_{1}\left(U_{s}=1\,|\,\theta\right)\delta\left[u_{s}-1\right]+p_{1}\left(U_{s}=0\,|\,\theta\right)\delta\left[u_{s}\right]\right)}{p_{0}^{X|\theta}\left(x\right)\left(p_{0}\left(U_{s}=1\,|\,\theta\right)\delta\left[u_{s}-1\right]+p_{0}\left(U_{s}=0\,|\,\theta\right)\delta\left[u_{s}\right]\right)},$$
$$=\frac{p_{1}^{X|\theta}\left(x\right)\left(b'\delta\left[u_{s}-1\right]+b\delta\left[u_{s}\right]\right)}{p_{0}^{X|\theta}\left(x\right)\left(a\delta\left[u_{s}-1\right]+a'\delta\left[u_{s}\right]\right)}$$

where $\delta\left[x\right]=1$ iff $x=0$ and equals zero otherwise. Now consider the case $U_s=1$ and $U_s=0$ separately. First $U_s=1$

$$\begin{aligned} l\left(x,u_{s}\right)\restriction_{U_{s}=1}&=\frac{p_{1}^{X|\theta}\left(x\right)p_{1}\left(U_{s}=1\,|\,\theta\right)}{p_{0}^{X|\theta}\left(x\right)p_{0}\left(U_{s}=1\,|\,\theta\right)},\\ &=l\left(x\right)\frac{p_{1}\left(U_{s}=1\,|\,\theta\right)}{p_{0}\left(U_{s}=1\,|\,\theta\right)}\gtrless\tau, \end{aligned} \tag{3.13}$$

and $U_s=0$

$$l\left(x,u_{s}\right)\restriction_{U_{s}=0}=l\left(x\right)\frac{p_{1}\left(U_{s}=0\,|\,\theta\right)}{p_{0}\left(U_{s}=0\,|\,\theta\right)}\gtrless\tau, \tag{3.14}$$

where $\restriction$ is the function restriction operator. Now the relays LRT, $l\left(x\right)$ in both (3.13) and (3.14) can be expressed as

$$\begin{aligned} l\left(x\right)\restriction_{U_{s}=1}&\gtrless\tau\frac{p_{0}\left(U_{s}=1\,|\,\theta\right)}{p_{1}\left(U_{s}=1\,|\,\theta\right)}=\tau_{1}\\ l\left(x\right)\restriction_{U_{s}=0}&\gtrless\tau\frac{p_{0}\left(U_{s}=0\,|\,\theta\right)}{p_{1}\left(U_{s}=0\,|\,\theta\right)}=\tau_{0}. \end{aligned}$$

Thus, the optimal LRT $l\left(x\right)$ has two thresholds selected based on u_s and both are a function of the sensors Type I and Type II error probabilities. Then for any non-trivial distribution $P_j^{X_s},\;\; j=0,1$ and a sensor test of size $\alpha_s\in(0,1)$ the numerator and denominator of τ_1 and τ_0 remain a function of the composite parameter under both hypotheses. Hence, the tandem atom is not UMP-DD, which implies that a general tandem network composed of multiple atoms is also not UMP-DD.

3.6 Summary

This chapter studied the detection performance for binary tree structures of bounded uniform height under the Neyman-Pearson framework with different classes of sensors. We extended the work of prior authors, using sensor observations that are conditionally i.i.d. in an average sense within a given class and at least conditionally independent across classes. This model was named a hBRT configuration and was shown to achieve an equivalent error exponent to the optimal parallel configuration. That is, the Type II error probability decayed at the same exponential rate with the number of nodes between the two configurations.

We showed that the asymptotically optimal hBRT configuration can also be UMP-DD. In doing so, we extended current research beyond the conditionally i.i.d. case, providing model support for observation models that are more closely aligned to problems of interest in decentralized detection. Included are models with composite parameters, where that parameter can differ as a function of location, sensing technology, and by random effects such as multipath fading channels.

Finally, unlike the hBRT relay network, we discussed that the tandem networks Type II error probability decays sub-exponentially and showed that the tandem network is not UMP-DD. The reason for this was that the quantizers for each tandem node are a function of the prior nodes probabilities under both H_0 and H_1 and at least one of those probabilities will be a function of the composite parameter.

CHAPTER 4

CONDITIONALLY DEPENDENT DECENTRALIZED DETECTION AND UMP-DD

4.1 Introduction

The prior two chapters concluded that for a single rooted tree of bounded height, the parallel and relay networks under 1-bit quantization can achieve UMP-DD. However, those conclusions and supporting derivations were based on the assumption of conditionally independent observations (cf. 2.1 and 3.1). This is an assumption that is overly restricted for most problems in decentralized detection, particularly wireless networks [48]. Examples of such observation models include sensors that are subject to a common interference, a jamming signal, or observations with correlated noise (e.g., correlated log-normal shadowing [51]).

This chapter will focus exclusively on both the parallel and hBRT configurations discussed in Chapters 2 and 3, and will exclude the tandem configuration, since it does not achieve UMP-DD performance, even with conditionally i.i.d. observations. We will establish a set of sufficient conditions for both the parallel and hBRT configuration to establish UMP-DD, but now with conditionally dependent observation.

Assumption 4.1. *The sensor observations are conditionally dependent given the composite parameter under each hypothesis.*

Certainly a directed tree network having UMP-DD performance with conditionally dependent observations is advantageous in practice. Intuitively, one expects that as the physical separation between sensors diminishes, the probability that local wireless sensor observations become conditionally dependent and identically distributed increases. Thus, it is desirable to develop a conditionally dependent UMP-DD theory that allows sensors in close proximity to be grouped, and support the case where the observation model across groups is different (e.g., geographically dispersed groups).

4.1.1 Prior Research

It is know that while the general decentralized detection problem is NP-complete, a majority of the conditionally dependent problems are NP-hard [12]. NP-hard in the sense that the form of the optimal decision rules at the leaves is often unknown, many time coupled between leafs, and may be coupled with fusion rules throughout the tree [1, 15]. Unlike the conditionally independent model, the application of a LRT at the leaves, even for seemingly straightforward hypothesis testing problems may not be optimal (see [52, 53] and references therein).

An interesting treatment of binary hypothesis testing with two sensors ($n = 3$) observing a shift in mean, θ_i $i = 1, 2$, with noise that has a standard bivariate Gaussian distribution with correlations parameter ρ appears in [1] and is an extension of [54]. Even for this relatively simple two sensor model, determining the optimal decision rules is complex. In some regions, the optimal decision rules are single threshold quantizers, in other regions multi-level quantizers, and undefined for the remaining regions, where the regional separation is a function of ρ and the mean shift values θ_1 and θ_2. The authors in [1] denote these regions respectively as “Good,” “Bad,” and “Ugly,” with Figure 4.1 indicating the signal plane relationship to ρ, θ_1, and θ_2 for the

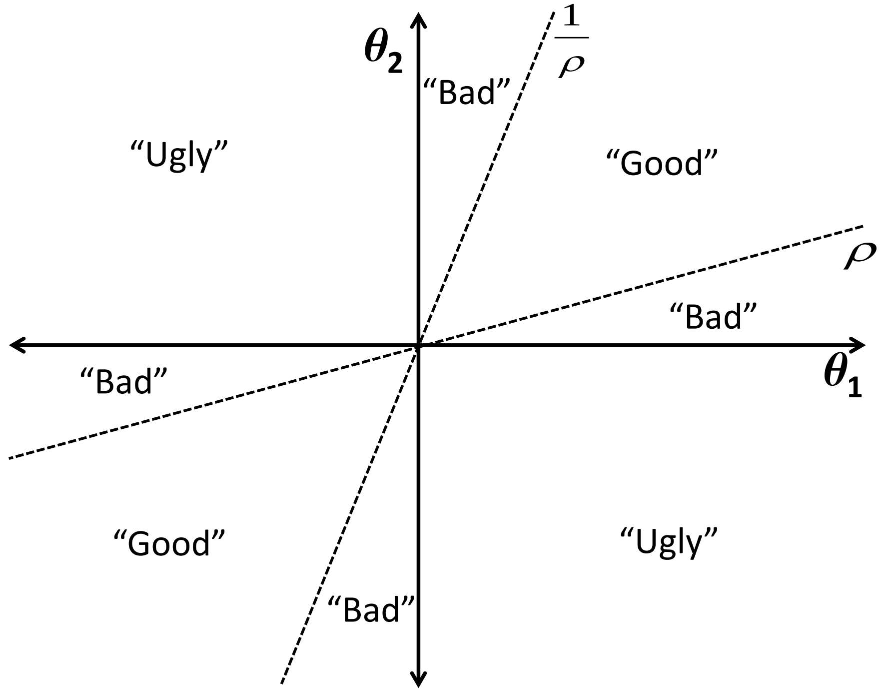


Figure 4.1: The Good, Bad, and Ugly Regions [1]

two sensors. The bounding equation in both [1] and [54] is $(\theta_1 - \rho\theta_2)(\theta_2 - \rho\theta_1) \geq 0$, where the joint constraint equations $\theta_2 \geq \rho\theta_1$ and $\theta_1 \geq \rho\theta_2$ follow directly.

A hierarchical approach for modeling conditionally dependent observations was introduced in [15]. Here, a hidden random variable was introduced in a manner that allows each sensor's observation to be conditionally independent given the hidden variable, but remain conditionally dependent given H_j, $j = 0, 1$. This approach is relatively intuitive and allows the exploration of a larger class of decentralized detection problems with conditionally dependent observations under both the Bayesian and Neyman-Pearson framework. Additionally, it is straightforward to establish the bounding equation for the "Good" region in Figure 4.1 under the hierarchical framework.

4.1.2 Chapter Goals

This chapter will show that under certain sufficient conditions the bounded height relay tree discussed in Chapters 2 and 3 with conditionally dependent observations can be UMP-DD. These sufficient conditions will be established using the hierarchical conditional independence (HCI) model under the Neyman-Pearson framework [7], but applied to N different classes of observations within a tree configuration. The observations within a single class are conditionally dependent with a known dependency and may be either conditionally dependent or independent between classes. Similar to the prior chapters, we assume 1-bit quantized local decisions throughout the tree.

Section 4.2 will review the HCI model and establish the tools that will be used to define the UMP-DD sufficient conditions. Section 4.3 then applies these tools to the parallel configuration ($h = 1$) and introduces the main contribution of this section, which defines a set of sufficient conditions to achieve UMP-DD performance with conditionally dependent observations. These results are then applied to example parallel configurations where the observation models are correlated Gaussian noise. The HCI model is then applied to relay networks in Section 4.4, leveraging the results of Section 4.3. Again, a set of sufficient conditions are required for the more general hBRT configuration to achieve UMP-DD performance.

4.2 The Hierarchical Conditional Independence Model

The HCI model for decentralized detection can be represented in terms of a Markov Chain. The parallel configuration has a conditionally dependent Markov Chain $H_j \to \boldsymbol{X} \to \boldsymbol{U}_s \to U_0$, where the joint density $p_j^{\boldsymbol{X}}$ cannot be factored as $\prod_{s \in V_n} p_j^{X_s}$. The HCI model introduces a hidden random variable, say Y, into the Markov Chain as

$H_j \rightarrow Y \rightarrow \boldsymbol{X} \rightarrow \boldsymbol{U}_s \rightarrow U_0$, such that $p_j^{\boldsymbol{X}|Y=y} = \prod_{s \in V_n} p_j^{X_s|y}$. Thus, Y induces a conditional independence on the sensor observations.

Theorem 1 in [15] indicates that when $\boldsymbol{X}$ is a continuous random variable, similar to our UMP-DD requirements in Chapter 2, and Y is a scalar random variable, then a single quantizer threshold for each sensor is optimal, subject to the following conditions:

1. The fusion center employs a *monotone fusion rule:* such that the probability of deciding 1 is a monotonic function of U_s, such that $P_j(U_0 = 1|U_s = 1, Y = y) \geq P_j(U_0 = 1|U_s = 0, Y = y)\ \forall\, y$ and $j = 0, 1$;
2. The LRT $l\,(y) = \frac{p_1^Y(y)}{p_0^Y(y)}$ is a non-decreasing function of y with $P_1^Y \ll P_0^Y$;
3. The LRT $l\,(x_s, x_s';\, y) = \frac{p_1^{X_s|y}(x_s)}{p_0^{X_s|y}(x_s')}$ is also a non-decreasing function of $y\ \forall x_s > x_s'$ with $P_1^{X_s|y} \ll P_0^{X_s|y}$.

We now use the HCI model to study the composite binary hypothesis testing problem.

4.3 Parallel Configuration, Dependent Observations, and UMP-DD

This section studies the parallel configuration UMP-DD problem of Chapter 2, depicted in Figure 1.1c, but now with dependent observations. Again we will require that the distributions be smooth log-concave functions and as such contain no point masses. Thus, the $\sigma-$finite measure μ can be taken to be the Lebesgue measure.

Assumption 4.2. *The $\sigma-$finite measure μ is the Lebesgue Measure.*

Suppose under the HCI model, the hidden variable Y with distribution P_j^Y is a function of a composite parameter $\theta \in \Theta_0 \sqcup \Theta_1$ having distribution P_j^θ with $P_j^\theta << \mu$.

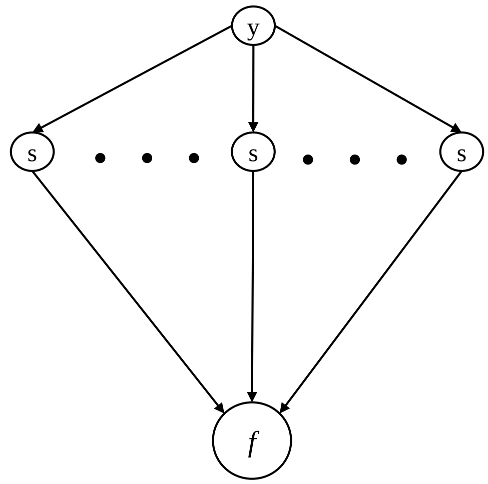


Figure 4.2: Parallel Configuration via the Hierarchical Conditional Independence (HCI) Model

The goal is to test H_0: $\theta \leq \theta'$ versus H_1: $\theta > \theta'$ under the Neyman-Pearson framework. A difference to Chapter 2 is that θ is a scalar, since Y is a scalar random variable, so θ_s is simplified to θ.

Suppose Y is selected so the Markov chain $H_j \to Y \to \boldsymbol{X} \to \boldsymbol{U} \to U_0$ meets the HCI model constraint of Section 4.2, with $\boldsymbol{X}$ conditionally dependent given H_j, but conditionally independent given Y. The parallel configuration using the HCI model is depicted in Figure 4.2. Similar to Chapter 2, the *a priori* distribution P_j^{θ} is defined by a set of parameters $\mathcal{D}_j$ (i.e., support, mean, variance, ...), where the dependence on $\mathcal{D}_j$ is implied, henceforth.

Next, suppose the FC applies the counting fusion rule

$$U_o = \begin{cases} 1, & \sum_{s=1}^{n-1} U_s \geq \kappa, \\ 0, & \sum_{s=1}^{n-1} U_s < \kappa, \end{cases} \tag{4.1}$$

which is monotone in U_s per Condition 1.

Let

$$
\begin{aligned}
P_j\left(\boldsymbol{U}_s=\boldsymbol{u}_s\right) &= \int_Y \prod_{s\in V_n} P_j^{X_s|y}\left(U_s=u_s\right) p_j^Y\left(y\right)\mu\left(dy\right) \\
&= E_j^Y\left[\prod_{s\in V_n} P_j^{X_s|y}\left(U_s=u_s\right)\right],
\end{aligned}
$$

where $u_s \in \{0,1\}$. Suppose $T_s\left(X_s|y\right)$ is a sufficient statistic for deciding $P_j^{X_s|y}\left(U_s=u_s\right)$, with $\mathcal{C}_s$ the critical region for sensor s deciding H_1. Consider the *And* fusion rule, then $P_j\left(U_0=1\right)=P_j\left(\boldsymbol{U}_s=\mathbf{1}^T\right)$, where $\mathbf{1}^T$ is a vector of all ones. Then,

$$
P_j\left(U_0=1\right)=E_j^Y\left[\prod_{s\in V_n}\int_{\mathcal{C}_s} p_j\left(T_s\left(X_s|y\right)\right)\mu\left(dx\right)\right]. \tag{4.2}
$$

Similar to Theorem 2.7 for the conditionally independent case, we have the following for the conditionally dependent case under the HCI model.

Theorem 4.3. *Let $T_s\left(X_s|y\right)$ be a sufficient statistic for the conditionally independent observations X_s having density $p_j^{X_s|Y}$ where $p_j\left(T_s\left(X_s|y\right)\right)$ is log-concave w.r.t. $T_s\left(X_s\right)$ for $j=0,1$ and $l\left(x_s,x_s';\,y\right)$ is nondecreasing in y $\forall x_s>x_s'$ (cf. 3). Let $p_j^{Y(\theta)}\left(y\right)$ be log-concave w.r.t. y, have a convex support with θ a constant, and $l\left(y\right)$ is nondecreasing in y (cf. 2). If $E_j^Y\left[\prod_{s\in V_n}\int_{\mathcal{C}_s} p_j\left(T_s\left(X_s|y\right)\right)\mu\left(dx\right)\right]$ is a smooth function (continuous derivatives), then equal quantizers, $\eta_1=\eta_2=\cdots=\eta_{n-1}=\eta$, are a UMP-DD test under the And or Or fusion rules (cf. 1) with identical distributed conditionally dependent observations, and 1-bit quantized local decisions.*

Proof. Let $l\left(x_s,x_s';\,y\right)$ be nondecreasing in y $\forall x_s>x_s'$, and $l\left(y\right)$ nondecreasing in y per HCI constraints 2 and 3. Let the fusion rule be the *And* rule per HCI constraint

1. Then there exist single threshold 1-bit quantizers at each sensor that are optimal (cf. Theorem 1 in [15]). Thus, the critical region, $\mathcal{C}_s$, in 4.2 is equivalent to $[\eta_s, \infty)$ and

$$P_j\left(U_0 = 1\right) = E_j^Y\left[\prod_{s\in V_n}\int_{\eta_s}^{\infty} p_j\left(T_s\left(X_s|y\right)\right)\mu\left(dx\right)\right],$$

and similarly for $P_j\left(U_0 = 0\right)$ with the interval $(-\infty, \eta_s)$.

Let $p_j\left(T_s\left(X_s|y\right)\right)$, $p_j^{Y(\theta)}\left(y\right)$ be log-concave w.r.t. $T_s\left(X_s\right)$ and y respectively, and the support for $p_j^{Y(\theta)}\left(y\right)$ be convex with $\theta \in \Theta_j$ a constant for $j = 0, 1$. Then, $E_j^Y\left[\prod_{s\in V_n}\int_{\eta_s}^{\infty} p_j\left(T_s\left(X_s|y\right)\right)\mu\left(dx\right)\right]$ is log-concave w.r.t. $\boldsymbol{\eta}_s$ (cf. the proof of Theorem 2.7). Then, if $E_j^Y\left[\prod_{s\in V_n}\int_{\eta_s}^{\infty} p_j\left(T_s\left(X_s|y\right)\right)\mu\left(dx\right)\right]$ is also smooth, equal quantizer thresholds are optimal where the remainder of the proof follows from that of Theorem 2.7. $\square$

Next, consider the general counting rule (e.g., k out of N), which is a monotone fusion rule per HCI constraint 1. Under the same conditions as Theorem 4.3, the conditionally dependent problem can also be LMP-DD per Remark 2.9. We now explore the results of this section via an illustrative example.

4.3.1 Detection with Composite Parameters under Correlated Gaussian Noises

The selection of local decision rules for the general decentralized binary hypothesis testing problem with composite parameters under correlated Gaussian noise does not admit a closed form solution. To see this, consider Figure 4.1 and note that for even two sensors, it is not possible to ensure operations within any particular region, "Good," "Bad," or "Ugly." However, the case where each sensors observation is based on the same composite parameter is exceptional.

That is, when $\theta_s = \theta \ \forall s$, the operating region is guaranteed to be within the "Good" region for all $\rho \in [0,1]$ with two sensors [1, 15, 52]. Suppose $\Theta_0 = \{0\}$, $\Theta_1 = \{\theta : \theta > 0\}$, which implies $\theta' = 0$. Here we set Θ_0 to be a singleton, but $\Theta_0 = \{\theta : \theta \leq 0\}$ is also possible (cf. 1.1). Notice that for the two sensor case, either formulation for Θ_0 ensures the observation space excludes the "Ugly" region.

Let the observations at each sensor be $P_j^{\boldsymbol{X}_s|\theta} \sim \mathcal{N}\left(\theta \cdot \mathbf{1}^T, \Sigma_X\right)$, $j = 0, 1$, where $\mathbf{1}^T = [1, 1, \ldots, 1]^T$ consisting of $n-1$ rows. When Σ_X has the following structure

$$\Sigma_X = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

where $\rho \in [0,1]$ and, without loss of generality, the variance down the diagonal has been normalized to one by scaling X_s appropriately. This structure admits a possible decomposition of $X_s = \theta Y + W_s \ \forall s \in V_n$ where $W_s \sim \mathcal{N}(0, 1-\rho)$ and $P_j^Y \sim \mathcal{N}(j, \rho)$, $j = 0, 1$, with a slight abuse of notation using j as the mean. This formulation with the counting fusion rule satisfies HCI constraints 1-3 and a single quantizer test for each s is optimal [15]. We will use a simple two-sensor example to indicate how the remaining conditions required by Theorem 4.3 are met, which implies the problem is UMP-DD or at least LMP-DD using (4.1).

Suppose $P_j^{\boldsymbol{X}|\theta} \sim \mathcal{N}(\theta, \theta, 1, 1, \rho)$, $j = 0, 1$. This is the standard shifted mean problem in correlated Gaussian noise, $\mathcal{W}$, with a traditional formulation of

$$H_1\left(\theta > 0\right): \quad X_s = \theta + \mathcal{W}_s, \tag{4.3}$$

$$H_0\left(\theta = 0\right): \quad X_s = \mathcal{W}_s,$$

with $n = 3$ and $\mathcal{W}_s \sim \mathcal{N}\left(0, 0, 1, 1, \rho\right)$ the standard bivariate Gaussian distribution with $\rho \in [0, 1]$.

The optimal local decision rule, $\gamma_s\ \forall s$, is a 1-bit quantizer with a graphical representation of this decomposition, depicted in Figure 4.2 by setting $n = 3$.

Suppose the *And* fusion rule is applied for a given ρ, then $P_j\left(U_0 = 1\right)$ in (4.2) is

$$P_j\left(U_0 = 1\right) = \int_{-\infty}^{\infty} Q\left(\frac{\eta_1 - y}{\sqrt{1-\rho}}\right) Q\left(\frac{\eta_2 - y}{\sqrt{1-\rho}}\right) p_j^Y\left(y\right) \mu\left(dy\right),$$

where $Q\left(x\right)$ is the standard Gaussian CCDF and $\prod_{i=1}^{n-1} Q\left(\frac{\eta_i - y}{\sqrt{1-\rho}}\right)$ follows from the conditional independence of $U_s|Y = y$. As such, $P_j\left(U_0 = 1\right)$ is smooth log-concave for $j = 0, 1$ since $Q\left(\cdot\right)$ and p_j^Y are both smooth log-concave. Notice that extensions for $n > 3$ follow similarly. Thus, by Theorem 4.3, the two sensor composite parameter detection problem with correlated identically distributed Gaussian observations is UMP-DD. The case with *Or* fusion is similar using all zeros versus an all ones approach (i.e., $P_j\left(U_0 = 0\right)$).

The results of this section are explicitly based on the requirement of identical composite parameters across all sensors (i.e., $\theta_s = \theta\ \forall s \in V_n$). While this model is applicable to some specific problems in decentralized detection, a better model is one that allows the composite parameter to vary across the sensors, even if the variation is at a class level. This will be the focus of the next section.

4.4 Relay Networks, Dependent Observations, and UMP-DD

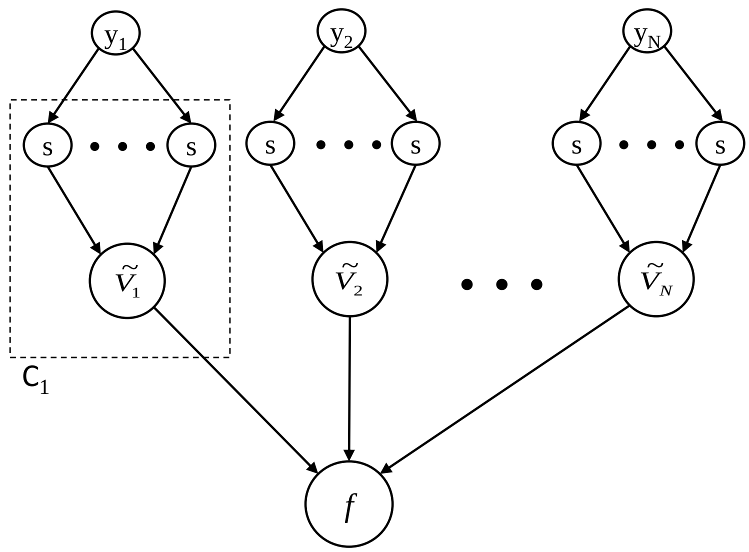
This section studies the UMP-DD problem for the relay networks of Chapter 3, depicted in Figure 3.1a with dependent observations. Similar to the prior section, we will require that the distributions be smooth log-concave functions and the $\sigma-$finite measure μ is the Lebesgue measure (c.f. 4.2).

Suppose under the HCI model, a hidden variable Y_k can be determined for each class C_k, where the observations within the kth class are conditionally dependent with a composite parameter $\theta_s = \theta_k \; \forall s \in \mathsf{C}_k$. Here, the font C_k is used to differentiate from the conditionally i.i.d. class definition, C_k, used in Chapter 3. Let the distribution $P_j^{Y_k}(\theta_k)$ be a function of the composite parameter $\theta_k \in \Theta_{0,k} \sqcup \Theta_{1,k} = \Theta_k$, where θ_k follows probability distribution $P_j^{\theta_k}$ with $P_j^{\theta_k} << \mu$. Within this section, the observations between classes are allowed to be either conditionally dependent or independent as we are not concerned with the asymptotic performance.

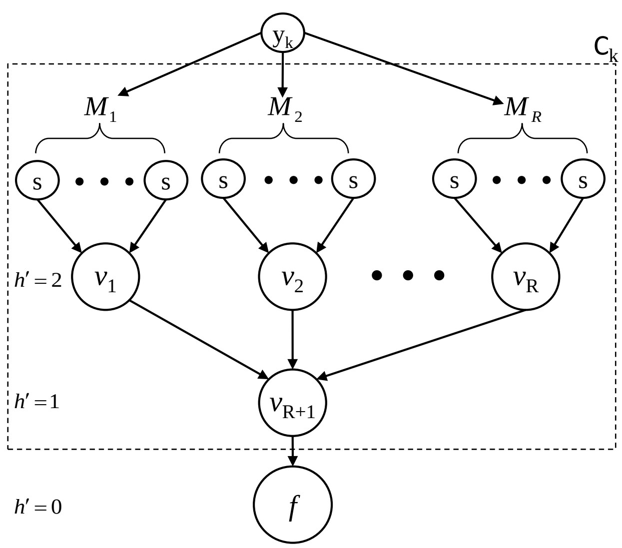
Again, the goal is to test H_0: $\theta \leq \theta'$ versus H_1: $\theta > \theta'$ under the Neyman-Pearson framework, with θ_k a scalar function of θ for a given class. By way of example, θ might represent the transmitted power and θ_k the received signal power by a collection of sensors that are physically close to one another.

Suppose Y_k is selected so the Markov chain $H_j \to \boldsymbol{Y} \to \boldsymbol{X} \to \tilde{\boldsymbol{U}} \to U_0$ meets the HCI model constraint of Section 4.2 at a class level, with $\boldsymbol{X}$ conditionally dependent given H_j, but conditionally independent $\forall s \in \mathsf{C}_k$ given Y_k. For clarity, $\boldsymbol{Y}$ is $\{Y_k\}_{k=1}^{N}$, the segregation by class is implied throughout the Markov chain, and $\tilde{\boldsymbol{U}}$ represents the sequence of local decisions from s to f along the respective branches.

The relay network under the HCI model is depicted in Figure 4.3. Again, we



(a) HCI Class Relay Network



(b) HCI Class Relay Network Exploded View for a $\tilde{V}_k$. Class k, $h = 3$ (3-Uniform)

Figure 4.3: Relay Network with Dependent Observations and HCI Model

focus on the hBRT configuration, with $\tilde{V}_k$ used to abstract the relay structure within a class, as stated in Chapter 3. The set of predecessor nodes to relay v_i is denoted by M_i, h' is the number of directed arcs traveled from f to a certain level of relay nodes along a given branch, and $h' = h_B - 1$ are the relays with only sensors as their predecessors along branch B with $h = \max\{h_B : B \in A_n\}$.

4.4.1 Achieving UMP-DD with Dependent Observations in a hBRT Configuration

We now consider a hBRT configuration with conditionally dependent observations supporting an HCI model at the class level. The same methods used to establish Theorem 4.3 for the parallel configuration can then be applied within a given class, say C_k, with a common composite parameter $\theta_k \in \Theta_k \ \forall s \in \mathsf{C}_k$. Under the Neyman-Pearson framework with test size $\alpha \in (0,1)$ at the system level, we have the following Corollary to Theorem 4.3.

Corollary 4.4. *Let $T_s^k(X_s|y_k)$ be a sufficient statistic for the conditionally independent observations X_s having density $p_j^{X_s|Y_k}$ where $p_j\left(T_s^k(X_s|y_k)\right)$ is log-concave w.r.t. $T_s^k(X_s)$ for $j = 0,1$, and $l(x_s, x_s'; y_k)$ is nondecreasing in y_k $\forall x_s > x_s'$ (cf. HCI 3). Let $p_j^{Y_k(\theta_k)}(y_k)$ be log-concave w.r.t. y_k, have a convex support with θ a constant, and $l(y_k)$ is nondecreasing in y_k (cf. HCI 2). If $E_j^{Y_k}\left[\prod_{s\in\mathsf{C}_k}\int_{\mathcal{C}_s^c} p_j\left(T_s^k(X_s|y_k)\right)\mu(dx)\right]$ is a smooth function (continuous derivatives), then equal quantizers, $\eta_1^k = \eta_2^k = \cdots = \eta_{|\{s:s\in\mathsf{C}_k\}|}^k = \eta^k$, are a UMP-DD test at the level $h' = h_B - 1$ under the* And *or* Or *fusion rules (cf. HCI 1) with identical distributed conditionally dependent observations, and 1-bit quantized local decisions.*

When all relays at level $h' = h_B - 1$ achieve UMP-DD performance, then the hBRT configuration in its entirety is also UMP-DD.

Theorem 4.5. *If all relays at level $h' = h_B - 1$ in a hBRT configuration are UMP-DD per Corollary 4.4, then a hBRT configuration is UMP-DD.*

The proof is similar to that of Lemma 3.5 and relies on the properties of the Markov Chain $H_j \to \boldsymbol{Y} \to \boldsymbol{X} \to \tilde{\boldsymbol{U}} \to U_0$. Specifically, let $\tilde{\boldsymbol{U}} = \boldsymbol{U}_{h_B-1} \to \boldsymbol{U}_{h_B-2} \to \cdots \to \boldsymbol{U}_1$. Then, when the decisions $\boldsymbol{U}_{h_B-1}$ are each UMP-DD for each class, the remaining decisions are also UMP-DD. Notice that because of the Markov Chain properties, only the relays at level $h' = h_B - 1$ are required to use the And or Or fusion rule. Each subsequent relay can use the optimal counting fusion rule (cf. 3.7) when data aggregation up the tree is constrained within a $\tilde{V}_k$. Finally, it is straightforward to show that when the relays at level $h' = h_B - 1$ are LMP-DD using the more generic counting fusion rule (cf. 3.7), then the hBRT configuration is also at least LMP-DD.

4.5 Summary

This chapter studied the detection performance for bounded height binary tree structures under the Neyman-Pearson framework with different classes of sensors. We applied the HCI model to study problems with conditionally dependent observations that resulted in UMP-DD. The HCI model required that each sensor observe the same composite parameter for the parallel configuration. This is a constraint that is overly restrictive for most decentralized detection problems of interest.

We then showed that by using the class formulation presented in Chapter 3, it was possible to have a UMP-DD system with a multitude of composite parameters allowed. The only constraint on the class definition is that the composite parameter be common across all sensors within the class. Finally, we described how this model supports the critical challenge of correlated log-normal shadowing prevalent in decentralized wireless networks.

CHAPTER 5

CONCLUSIONS

The objective of this work has been to introduce UMP into the decentralized detection theory, focusing on directed single-rooted trees of bounded height. The first chapter reviewed UMP binary hypothesis testing from the centralized detection theory perspective and introduced the notation used across the remaining chapters regarding directed single-rooted trees. This first chapter also provided the motivation regarding the results of Chapter 2 and their associated significance.

Chapter 2 introduced Theorem 2.7 that defined sufficient conditions for when the classical parallel configuration is able to achieve UMP-DD performance. These conditions include: conditionally independent sensor observations given a random composite parameter following an *a priori* known distribution, and a fixed *And* or *Or* fusion rule at the fusion center. Corollary 2.8 associated with Theorem 2.7 then established sufficient conditions for UMP-DD when the composite parameters are fixed but unknown. These methods were then applied to a multitude of inference models, including spectrum sensing in cognitive radio using energy detection methods under various channel models such as Nakagami or Rayleigh fading.

The results of Chapter 2 were then extended to the more general counting fusion rule with the UMP-DD designation replaced with the less restrictive LMP-DD identifier. However, we conjectured that the counting fusion rule remains UMP-DD, including randomization between adjacent thresholds. This conjecture was supported by the two-sensor model with a generalized fusion rule. There, we showed that all forms of the counting rule, including randomization between thresholds was in fact UMP-DD in all cases. We also showed that the Karlin-Rubin Theorem used in centralized detection theory was insufficient to determine if a single-rooted tree was UMP-DD, even with conditionally independent observations.

Chapter 3 studied the detection performance for binary tree structures of bounded uniform height under the Neyman-Pearson framework with different classes of sensors. We extended the work of prior authors who assumed i.i.d. observations by introducing observation classes, where sensor observations within a class are conditionally i.i.d. in an average sense, but only need to be independent across classes. This model was named a hBRT configuration and was shown to have a Type II error probability that decays exponentially at the same rate as the parallel configuration with the same number of nodes.

We also showed in Chapter 3 that the asymptotically optimal hBRT configuration can also be UMP-DD, extending current research based on the conditionally i.i.d. case to support more general conditionally independent models. We discussed that these results support typical decentralized detection models having composite parameters, where that parameter can differ as a function of location, sensing technology, and suffer random perturbations such as multipath or log-normal fading channels.

Chapter 3 then concluded with a study of the tandem network. There, we discussed that the tandem networks Type II error probability decays sub-exponentially and showed that the tandem network can not be UMP-DD, unlike the hBRT configuration. The reason for this was that the quantizers for each tandem node are a function of the prior nodes inference probabilities under both H_0 and H_1, and at least one of those probabilities is a function of the composite parameter itself.

Chapter 4 relaxed the assumption of conditional independence found in Chapters 2 and 3. Using the described HCI model, we showed that with conditionally dependent observations the parallel configuration resulted in UMP-DD under some very specific conditions. Specifically, each sensor was required to observe the same composite parameter, and that the conditionally dependent observation model also met the requirements defined by the described HCI model.

Within Chapter 4, we discussed that identical observations in a decentralized detection system were an overly restrictive assumption. Under this premise, we then showed that by using the class formulation presented in Chapter 3, it was possible to have a UMP-DD system under certain constraints. These constraints allowed the composite parameter to vary across classes, but required it to be equivalent for all sensors within the class. This model supported sensors that were physically close with identically distributed, but conditionally dependent observations as is the case with conditionally dependent interference or correlated noise.

5.1 Some Open Research Topics

While this effort introduced UMP into decentralized detection, there remains a number of open problems to be considered:

1. As we discussed in Chapter 2, the general counting fusion rule in the parallel configuration with conditionally independent observations in an average sense resulted in a LMP-DD designation. We conjecture that the counting rule (k out of N) is actually UMP-DD, but the proof to this conjecture remains an open research topic. It is likely that randomization between consecutive thresholds

is also UMP-DD, similar to what we showed in Theorem 2.10 for two sensors with Gaussian noise.

2. The relay network with conditionally dependent observations in Chapter 4 did not explore the asymptotic performance as was done in Chapter 3. An open question is if it is possible to achieve UMP-DD performance and an error exponent that decays exponentially fast. If it does decay exponentially, then is the rate of decay similar to that of the parallel configuration?

3. The conditionally dependent observation examples in Chapter 4 resulting in UMP-DD were shifted mean in correlated Gaussian noise. These results might be extended to conditionally dependent log-normal shadowing models if possible, with detection performance compared to the various detection methods proposed in [51, 55–58] as the numbers of sensors increases for cognitive radio applications. Notice that the asymptotic regime is explicitly covered in [55], where the Type II error probability bounded away from zero with correlated log-normal shadowing. Questions include: How does each proposed method perform asymptotically, with and without resource normalization as defined in [48]?; Does the Type II error probability for the hBRT configuration asymptotically converge to any of the proposed methods in [51, 55–58]?; If not, what is the difference in the performance?

REFERENCES

[1] P. Willett, P.F. Swaszek, and R.S. Blum. The good, bad and ugly: distributed detection of a known signal in dependent gaussian noise. *Signal Processing, IEEE Transactions on*, 48(12):3266 – 3279, Dec. 2000.

[2] Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer, 2005.

[3] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer, 1994.

[4] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 2007.

[5] RV Hogg, JW McKean, and AT Craig. *Introduction to Mathematical Statistics*. Prentice Hall, 6th edition, 2005.

[6] Pramod K. Varshney. *Distributed Detection and Data Fusion*. Springer, 1997.

[7] Hao Chen and Uri Rogers. Uniformly most powerful distributed detection and its application in cooperative spectrum sensing. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pages 1674–1676. IEEE, 2011.

[8] Uri Rogers and Hao Chen. Distributed detection and uniformly most powerful tests. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.

[9] Uri Rogers and Hao Chen. Uniformly most powerful distributed detection with cognitive radio applications. ***Submitted to*** *Aerospace and Electronic Systems (AES), IEEE Transactions on.*

[10] A.M. Bruckner, J.B. Bruckner, and B.S. Thomson. *Real Analysis*. Prentice-Hall, 1997.

[11] John Chiasson. *Introduction to Probability and Stochastic Processes*. John Wiley & Sons, 2013.

[12] John N. Tsitsiklis and Michael Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, AC-30(5):440–446, May 1985.

[13] Wee Peng Tay, J.N. Tsitsiklis, and M.Z. Win. Asymptotic performance of a censoring sensor network. *Information Theory, IEEE Transactions on*, 53(11):4191–4209, 2007.

[14] Wee Peng Tay, J.N. Tsitsiklis, and M.Z. Win. Data fusion trees for detection: Does architecture matter? *Information Theory, IEEE Transactions on*, 54(9):4155–4168, 2008.

[15] Hao Chen, Biao Chen, and P.K. Varshney. A new framework for distributed detection with conditionally dependent observations. *Signal Processing, IEEE Transactions on*, 60(3):1409 –1419, March 2012.

[16] R. Viswanathan and P.K. Varshney. Distributed detection with multiple sensors i. fundamentals. *Proceedings of the IEEE*, 85(1):54–63, 1997.

[17] Rick S. Blum, S. Kassam, and H. Poor. Distributed detection with multiple sensors - Part II: advanced topics. *Proc. IEEE*, pages 64–79, Jan. 1997.

[18] John N Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals, and Systems (MCSS)*, 1(2):167–182, 1988.

[19] W.W. Irving and J.N. Tsitsiklis. Some properties of optimal thresholds in decentralized detection. *Automatic Control, IEEE Transactions on*, 39(4):835–838, 1994.

[20] Douglas Warren and Peter Willett. Optimum quantization for detector fusion: some proofs, examples and pathology. *Journal of the Franklin Institute*, 336:323–359, 1999.

[21] S. Haykin. Cognitive radio: brain-empowered wireless communications. *Selected Areas in Communications, IEEE Journal on*, 23(2):201 – 220, 2005.

[22] Erik Axell, Geert Leus, Erik G Larsson, and H Vincent Poor. Spectrum sensing for cognitive radio: State-of-the-art and recent advances. *Signal Processing Magazine, IEEE*, 29(3):101–116, 2012.

[23] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math.(Szeged)*, 32(197):1, 1971.

[24] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math.(Szeged)*, 34:335–343, 1973.

[25] T. Bergstrom and M. Bagnoli. Log-concave probability and its applications. *Economic theory*, 26:445–469, 2005.

[26] George Casella and Roger L Berger. *Statistical Inference*. Duxbury Press, 2001.

[27] Wei Zhang, R Mallik, and K Letaief. Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 8(12):5761–5766, 2009.

[28] Fadel F Digham, M-S Alouini, and Marvin K Simon. On the energy detection of unknown signals over fading channels. *Communications, IEEE Transactions on*, 55(1):21–24, 2007.

[29] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-22(1):98 –101, Jan. 1986.

[30] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge Univ Pr, 2004.

[31] Ruixin Niu and P.K. Varshney. Performance analysis of distributed detection in a random sensor field. *Signal Processing, IEEE Transactions on*, 56(1):339–349, 2008.

[32] A.S. Rawat, P. Anand, Hao Chen, and P.K. Varshney. Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks. *Signal Processing, IEEE Transactions on*, 59(2):774–786, 2011.

[33] A. Ghasemi and E.S. Sousa. Collaborative spectrum sensing for opportunistic access in fading environments. In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pages 131–136, Nov 2005.

[34] Zhi Quan, Shuguang Cui, and Ali H Sayed. Optimal linear cooperation for spectrum sensing in cognitive radio networks. *Selected Topics in Signal Processing, IEEE Journal of*, 2(1):28–40, 2008.

[35] A. Goldsmith, S.A. Jafar, I. Maric, and S. Srinivasa. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE*, 97(5):894–914, 2009.

[36] John N. Tsitsiklis. Decentralized detection. In H. V. Poor and J. B. Thomas, editors, *Advances in Signal Processing*, volume 2, pages 297–344. JAI Press, 1993.

[37] Venugopal V. Veeravalli. Sequential decision fusion: Theory and applications. *Journal of the Franklin Institute*, 336, 1999.

[38] J.-F. Chamberland and V.V. Veeravalli. Decentralized detection in sensor networks. *Signal Processing, IEEE Transactions on*, 51(2):407 – 416, feb 2003.

[39] Wee Peng Tay, J.N. Tsitsiklis, and M.Z. Win. Bayesian detection in bounded height tree networks. *Signal Processing, IEEE Transactions on*, 57(10):4042–4051, 2009.

[40] Biao Chen and P.K. Willett. On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels. *Information Theory, IEEE Transactions on*, 51(2):693–699, 2005.

[41] A. Kashyap. Comments on "on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels". *Information Theory, IEEE Transactions on*, 52(3):1274–1275, 2006.

[42] Zhenliang Zhang, A. Pezeshki, W. Moran, S.D. Howard, and E.K.P. Chong. Error probability bounds for balanced binary relay trees. *Information Theory, IEEE Transactions on*, 58(6):3548 –3563, june 2012.

[43] R. Viswanathan, S. C A Thomopoulos, and R. Tumuluri. Optimal serial distributed decision fusion. *Aerospace and Electronic Systems, IEEE Transactions on*, 24(4):366–376, 1988.

[44] Z.-B. Tang, K.R. Pattipati, and D.L. Kleinman. Optimization of detection networks. i. tandem structures. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(5):1044–1059, 1991.

[45] J.D. Papastavrou and M. Athans. Distributed detection by a large team of sensors in tandem. *Aerospace and Electronic Systems, IEEE Transactions on*, 28(3):639–653, 1992.

[46] Venugopal Veeravalli. *Topics in decentralized detection*. PhD thesis, Univ. Illinois, Urbana-Champaign, 1992.

[47] Wee-Peng Tay, J.N. Tsitsiklis, and M.Z. Win. On the sub-exponential decay of detection error probabilities in long tandems. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pages II–837–II–840, 2007.

[48] J. Chamberland and V.V. Veeravalli. Wireless sensors in distributed detection applications. *Signal Processing Magazine, IEEE*, 24(3):16–25, May 2007.

[49] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Disquisitiones mathematicae Hungaricae. Akadémiai Kiadó, 1997.

[50] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 3rd edition, 1991.

[51] P. Agrawal and N. Patwari. Correlated link shadow fading in multi-hop wireless networks. *Wireless Communications, IEEE Transactions on*, 8(8):4024–4036, August 2009.

[52] Z.-B. Tang, K.R. Pattipati, and D.L. Kleinman. A distributed m-ary hypothesis testing problem with correlated observations. In *Decision and Control, 1989., Proceedings of the 28th IEEE Conference on*, pages 562–568 vol.1, 1989.

[53] Qing Yan and R.S. Blum. Distributed signal detection under the neyman-pearson criterion. *Information Theory, IEEE Transactions on*, 47(4):1368–1377, 2001.

[54] Po-Ning Chen and A. Papamarcou. Likelihood ratio partitions for distributed signal detection in correlated gaussian noise. In *Information Theory, 1995. Proceedings., 1995 IEEE International Symposium on*, pages 118–, 1995.

[55] A. Ghasemi and E.S. Sousa. Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing. *Communications Letters, IEEE*, 11(1):34–36, Jan 2007.

[56] J. Unnikrishnan and V.V. Veeravalli. Decentralized detection with correlated observations. In *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pages 381–385, Nov 2007.

[57] M. Renzo, L. Imbriglio, F. Graziosi, and F. Santucci. Distributed data fusion over correlated log-normal sensing and reporting channels: Application to cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 8(12):5813–5821, December 2009.

[58] V. Jamali, S. Salari, N. Reisi, J. Cances, and M. Ahmadian. Linear cooperation for spectrum sensing over correlated log-normal shadow fading channels. In *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pages 1–5, Oct 2011.

[59] ZW Birnbaum. An inequality for mill's ratio. *The Annals of Mathematical Statistics*, 13(2):245–246, 1942.

[60] M. Basseville, I.V. Nikiforov, et al. *Detection of Abrupt Changes: Theory and Application*, volume 15. Prentice Hall Englewood Cliffs, 1993.

APPENDIX A

TWO SENSOR RANDOMIZATION OF *AND* / *OR* FUSION

Proof of Theorem 2.10

The generalized detection probability is

$$\begin{aligned} P_1\left(U_0=1\right) = {} & P_1\left(U_1=U_2=1\right) \\ & + \upsilon\left(P_1\left(U_1=1, U_2=0\right) + P_1\left(U_1=0, U_2=1\right)\right). \end{aligned} \tag{A.1}$$

Under i.i.d. unit variance Gaussian noise the detection probability is

$$\begin{aligned} P_1\left(U_0=1\right) = {} & Q\left(\eta_1-\theta\right) Q\left(\eta_2-\theta\right) + \\ & \upsilon\left(Q\left(\eta_1-\theta\right) Q\left(\theta-\eta_2\right) + Q\left(\theta-\eta_1\right) Q\left(\eta_2-\theta\right)\right), \end{aligned} \tag{A.2}$$

where $Q(\cdot)$ is the standard Normal complementary distribution function. Similarly, the false alarm probability is

$$P_0\left(U_0=1\right) = Q\left(\eta_1\right) Q\left(\eta_2\right) + \upsilon\left(Q\left(\eta_1\right) Q\left(-\eta_2\right) + Q\left(-\eta_1\right) Q\left(\eta_2\right)\right). \tag{A.3}$$

With the goal to maximize $P_1\left(U_0=1\right)$ while meeting a $P_0\left(U_0=1\right) \leq \alpha$, we use a Lagrangian maximization method. Taking the partial derivative w.r.t. η_1 results in

$$\frac{\partial P_1(U_0=1)}{\partial \eta_1} = \lambda \frac{\partial P_0(U_0=1)}{\partial \eta_1}, \tag{A.4}$$

which can be written as

$$\frac{\partial P_1(U_0=1)/\partial \eta_1}{\partial P_0(U_0=1)/\partial \eta_1} = \lambda = \frac{\partial P_1(U_0=1)/\partial \eta_2}{\partial P_0(U_0=1)/\partial \eta_2}, \tag{A.5}$$

where the last equality follows trivially. Evaluating the partial derivatives results in

$$\begin{aligned}\frac{\partial P_1(U_0=1)}{\partial \eta_1} =& -\phi(\eta_1-\theta)\,Q(\eta_2-\theta)\\ &-\upsilon\phi(\eta_1-\theta)\,Q(\theta-\eta_2)+\upsilon\phi(\theta-\eta_1)\,Q(\eta_2-\theta)\end{aligned} \tag{A.6}$$

$$\frac{\partial P_0(U_0=1)}{\partial \eta_1} = -\phi(\eta_1)\,Q(\eta_2) - \upsilon\phi(\eta_1)\,Q(-\eta_2) + \upsilon\phi(-\eta_1)\,Q(\eta_2). \tag{A.7}$$

Substituting (A.6) and (A.7) into (A.5)

$$\begin{aligned}\frac{\partial P_1(U_0=1)/\partial \eta_1}{\partial P_0(U_0=1)/\partial \eta_1} =& \frac{\phi(\eta_1-\theta)}{\phi(\eta_1)}\\ &\times \frac{Q(\eta_2-\theta)+\upsilon Q(\theta-\eta_2)-\upsilon Q(\eta_2-\theta)}{Q(\eta_2)+\upsilon Q(-\eta_2)-\upsilon Q(\eta_2)}.\end{aligned} \tag{A.8}$$

Thus

$$\frac{\partial P_1(U_0=1)/\partial \eta_1}{\partial P_0(U_0=1)/\partial \eta_1} = e^{\eta_1\theta}e^{-\frac{1}{2}\theta^2}\left(\frac{(1-\upsilon)\,Q(\eta_2-\theta)+\upsilon Q(\theta-\eta_2)}{(1-\upsilon)\,Q(\eta_2)+\upsilon Q(-\eta_2)}\right), \tag{A.9}$$

and similarly for η_2 with $\phi(x) = \phi(-x)$. Setting $\frac{\partial P_1(U_0=1)/\partial \eta_1}{\partial P_0(U_0=1)/\partial \eta_1} = \frac{\partial P_1(U_0=1)/\partial \eta_2}{\partial P_0(U_0=1)/\partial \eta_2}$ and canceling common terms results in

$$\begin{aligned}&e^{-\eta_2\theta}\left(\frac{(1-2\upsilon)\,Q(\eta_2-\theta)+\upsilon}{(1-2\upsilon)\,Q(\eta_2)+\upsilon}\right) =\\ &e^{-\eta_1\theta}\left(\frac{(1-2\upsilon)\,Q(\eta_1-\theta)+\upsilon}{(1-2\upsilon)\,Q(\eta_1)+\upsilon}\right),\end{aligned} \tag{A.10}$$

where the equality holds trivially if $\eta_1 = \eta_2$.

Next, we prove that $\eta_1 = \eta_2$ is indeed the only solution of (A.10). First, this claim holds trivially for the case when $\upsilon = \frac{1}{2}$. When $\upsilon \neq \frac{1}{2}$, let $\psi = \frac{\upsilon}{1-2\upsilon}$ and consider the function

$$f(\eta) = e^{-\eta\theta}\left(\frac{Q(\eta-\theta)+\psi}{Q(\eta)+\psi}\right). \tag{A.11}$$

Then, the optimality of $\eta_1 = \eta_2$ can be established if $f(\eta)$ is monotone. This is equivalent to $\ln(f(\eta))$ monotone, where $\ln(\cdot)$ is the natural log operator.

$$\begin{aligned}\ln(f(\eta)) =& -\eta\theta + \ln(Q(\eta-\theta)+\psi) \\ & - \ln(Q(\eta)+\psi).\end{aligned} \tag{A.12}$$

Evaluating the partial derivative

$$\begin{aligned}\frac{\partial \ln(f(\eta))}{\partial \eta} =& -\theta - \frac{\phi(\eta-\theta)}{Q(\eta-\theta)+\psi} \\ & + \frac{\phi(\eta)}{Q(\eta)+\psi}.\end{aligned} \tag{A.13}$$

Let $g(\theta) = \frac{\partial \ln(f(\eta))}{\partial \eta}$. Note that $g(0) = 0$ and is the maximum value if $g'(\theta) \leq 0$, $\forall\theta \geq 0$ since $g'(\theta)$ is continuous, which implies $\ln(f(\eta))$ is decreasing. Evaluating $g'(\theta)$, we have

$$g'(\theta) = -1 - \frac{\partial\left[\frac{\phi(n-\theta)}{Q(\eta-\theta)+\psi}\right]}{\partial \theta}. \tag{A.14}$$

Setting $t = \eta - \theta$ and expanding the partial derivative

$$g'(\theta) = -1 - \frac{t \cdot \phi(t)(Q(t)+\psi) - [\phi(t)]^2}{[Q(t)+\psi]^2}, \tag{A.15}$$

or equivalently

$$g'(\theta) = \frac{-\left(Q(t)\right)^2 - Q(t)\left(t\phi(t) + 2\psi\right)}{\left[Q(t) + \psi\right]^2} + \frac{\left[\phi(t)\right]^2 - t\phi(t)\psi - \psi^2}{\left[Q(t) + \psi\right]^2}. \tag{A.16}$$

The roots of the numerator are

$$Q(t) = \frac{-t \pm \sqrt{(t^2+4)}}{2}\phi(t) - \psi. \tag{A.17}$$

Thus

$$g'(\theta) = \frac{-\left(Q(t) - \left(\frac{-t+\sqrt{(t^2+4)}}{2}\phi(t) - \psi\right)\right)}{\left[Q(t) + \psi\right]^2} \times \frac{\left(Q(t) - \left(\frac{-t-\sqrt{(t^2+4)}}{2}\phi(t) - \psi\right)\right)}{\left[Q(t) + \psi\right]^2}. \tag{A.18}$$

Since $\upsilon \in [0, 1]$ it is straightforward to show $\psi \in (-\infty, -1) \cup [0, \infty)$. First, consider the case $\psi \geq 0$. Then, $g'(\theta) \leq 0$ $\forall t$ as $Q(t) > \frac{-t+\sqrt{(t^2+4)}}{2}\phi(t)$ $\forall t \in (-\infty, \infty)$ from Birnbaum's inequality [59]. Note, the inequality still holds for $t < 0$, but is not tight.

Next, consider the case $\psi < 0$ (i.e., $\psi \in (-\infty, -1)$). Then, $Q(t) + \psi \leq Q(t) - 1 = -Q(-t)$ $\forall \psi < 0$. We desire $Q(t) + \psi \leq \frac{-t-\sqrt{(t^2+4)}}{2}\phi(t)$ to ensure both numerator terms in (A.18) are negative so $g'(\theta) \leq 0$. This is the same as $-Q(-t) \leq \frac{-t-\sqrt{(t^2+4)}}{2}\phi(t)$. Setting $t' = -t$, we have $-Q(t') \leq \frac{t'-\sqrt{((-t')^2+4)}}{2}\phi(-t')$ or equivalently $Q(t') > \frac{-t'+\sqrt{(t'^2+4)}}{2}\phi(t')$, which again holds for all t' as desired.

Therefore $g'(\theta) \leq 0$ for all θ. Hence, $f(\eta)$ is a monotone decreasing function of η since $\frac{\partial \ln(f(\eta))}{\partial \eta} = g(\theta) < 0$, $\forall \theta > 0$. Thus, $f(\eta_1) = f(\eta_2)$ if and only if $\eta_1 = \eta_2$.

□

APPENDIX B

PROOF OF $G_P^* = G_R$ BY INDUCTION ON N

Let all $s \in C_k$ have Type I error probability a_k, Type II error probability b_k, $k = i$ for the 2UBRT configuration, and m_i increases proportionally with $\frac{n-N}{N}$. Order all sensors from the N classes as $(s_1, s_2, \ldots, s_{n-1})$ for a given n, where the nth component is the root node f. Without loss of generality, let $|C_k| = 1 \ \forall k$. If this is not the case, simply increase N until the condition is met, complete the proof, then remove repeated $D\left(P_0^{v_i} || P_1^{v_i}\right)$ values because they do not change the parallel configuration error exponent. We now evaluate the induction hypothesis $g_P^* = -\sum_{i=1}^{N} D\left(P_0^{v_i} || P_1^{v_i}\right)$ using $g_P^* = E_0\left[\log \frac{dP_1^f}{dP_0^f}\left(u^N\right)\right]$, where $U^n = \{U_s\}_{s \in V_n}$.

(Base Case) Suppose $N = 2$ as $N = 1$ holds trivially. Then, $g_P^* = a_1' a_2' \log \frac{b_1 b_2}{a_1' a_2'} + a_1' a_2 \log \frac{b_1 b_2'}{a_1' a_2} + a_1 a_2' \log \frac{b_1' b_2}{a_1 a_2'} + a_1 a_2 \log \frac{b_1' b_2'}{a_1 a_2}$. Simplifying by focusing on the coefficients a_1 and a_1', we have $\left(a_1 \log \frac{b_1'}{a_1} + a_1' \log \frac{b_1}{a_1'}\right)\left(a_2' + a_2\right) = D\left(P_0^{v_1} || P_1^{v_1}\right)$. For use in the induction step, define $\left(a_2' + a_2\right)$ as the non-essential coefficients as they sum one. The coefficient pair a_2, and a_2' follow similarly with non-essential coefficients $\left(a_1' + a_1\right)$. Thus $g_P^* = -\sum_{i=1}^{2} D\left(P_0^{v_i} || P_1^{v_i}\right)$.

(Induction Step) Let $N^+ = N + 1$, and $S = \sum_{s \in V_n} U_s$. Let Q_0 be the set of cardinality one associated with $g_P^* = E_0\left[\log \frac{dP_1^f}{dP_0^f}\left(u^N\right)\right]$ when $P_0\left(S = 0\right)$, Q_1 the set terms with cardinality N^+ associated with $P_0\left(S = 1\right)$, Q_2 the set of cardinality $\binom{N^+}{2}$ when $P_0\left(S = 2\right)$, and so on until Q_{N^+} when $P_0\left(S = N^+\right)$ with cardinality $\binom{N^+}{N^+} = 1$, where $\binom{n}{k}$ is the standard Combinatorics "n choose k". Let $Q = \bigcup_{i=1}^{N^+} Q_i$ and notice that the $|Q| = 2^{N^+}$. Factor out a desired $a'_k \log \frac{b_k}{a'_k} + a_k \log \frac{b'_k}{a_k}$. Case 1: Suppose $k \neq N^+$. Then, there are 2^N non-essential coefficients. Because of symmetry, for every non-essential coefficient containing an a_{N^+}, there is an equivalent non-essential coefficient differing only by a'_{N^+}. Use $a'_{N^+} = 1 - a_{N^+}$ to cancel / eliminate the a_{N^+} component from all non-essential coefficients. After canceling, the 2^{N-1} remaining non-essential coefficients are equivalent to the factoring step for the Nth iteration and sum to one by the induction hypothesis. Case 2 with $i = N^+$ follows similarly, but using another term, say a_1, for the cancel step.

Therefore, $g_P^* = -\sum_{i=1}^{N} D\left(P_0^{v_i} || P_1^{v_i}\right)$ as desired.

While somewhat abstract, an application of the induction step to a particular N^+, say $N^+ = 4$, is straightforward but tedious, including application of the cancel step.

$\square$

APPENDIX C

ASYMPTOTIC NUMERICAL CALCULATION METHOD FOR β_P AND β_R

This section details the method used to determine the asymptotic Type II error probability for both the hBRT and parallel configuration. We first explore the hUBRT configuration and calculate the expected error exponent indirectly, then determine it directly.

Here, we desire to calculate the the KLD (cf. (3.1)) for class k with Gaussian densities $p_k \sim \mathcal{N}\left(m_k a_k, m_k^2 a_k a_k'\right)$ and $q_k \sim \mathcal{N}\left(m_k b_k', m_k^2 b_k' b_k\right)$. When the difference between the means a_k and b_k' is small, then $E^{P_k}\left[\log \frac{p_k}{q_k}\right] \approx \frac{1}{2}\boldsymbol{I}\left(p_k\right)\left(m_k b_k' - m_k a_k\right)^2$ from Lemma 4.1.3 in [60], where $\boldsymbol{I}\left(p_k\right)$ is the Fisher Information of p_k. Similar to Lemma 4.1.1 [60], $E^{Q_k}\left[\log \frac{q_k}{p_k}\right] \approx \frac{1}{2}\boldsymbol{I}\left(q_k\right)\left(p_k - q_k\right)^2$, with $E^{P_k}\left[\log \frac{p_k}{q_k}\right] \approx E^{Q_k}\left[\log \frac{q_k}{p_k}\right]$ as $\boldsymbol{I}\left(p_k\right) \approx \boldsymbol{I}\left(q_k\right)$. The $\boldsymbol{I}\left(q_k\right)$ in class k under H_1 is $\frac{1}{m_k^2 b_k' b_k}$. Thus, the KLD for class k is approximately

$$D\left(P_k || Q_k\right) \approx \frac{1}{2}\frac{\left(b_k' - a_k\right)^2}{b_k \cdot b_k'},$$

since the m_k terms cancel, as expected. Summing across each class results in

$$D\left(P || Q\right) \approx \sum_{k=1}^{N} \frac{1}{2}\frac{\left(b_k' - a_k\right)^2}{b_k \cdot b_k'} \tag{C.1}$$

as the fusion node f. As (C.1) is an indirect method of determining the KLD at f

for the hUBRT configuration, the error exponent can also be calculated directly with equivalent results.

Let η^k be the sensor threshold in class k for deciding H_0 versus H_1 from (3.6), with the size α_k selected via (3.10). Then, under the Gaussian approximation (cf. (3.9)) for a single class $Q\left(\frac{\eta^k - m_k a_k}{\sqrt{m_k a_k a'_k}}\right) = \alpha_k$. Thus,

$$\eta^k = m_k a_k + \sqrt{m_k a_k a'_k} Q^{-1}\left(\alpha_k\right), \tag{C.2}$$

with

$$\beta_R^k = Q\left(-\frac{m_k b_k - \left(m_k a_k + \sqrt{m_k a_k a'_k} Q^{-1}\left(\alpha_k\right)\right)}{\sqrt{m_k b_k \cdot b'_k}}\right), \tag{C.3}$$

where β_R^k is the 2UBRT Type II error probability in class k, and applying $1 - Q\left(x\right) = Q\left(-x\right)$. Consider $\frac{\log\left(\beta_R^k\right)}{m_k}$ with the well-known inequality $Q\left(x\right) < \frac{1}{x}\frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right), \forall x > 0$. Then,

$$\begin{aligned}\frac{\log\left(\beta_R^k\right)}{m_k} < &-\frac{1}{m_k}\log\left(-\frac{m_k b_k - \left(m_k a_k + \sqrt{m_k a_k a'_k} Q^{-1}\left(\alpha_k\right)\right)}{\sqrt{m_k b_k \cdot b'_k}}\right)\\ &-\frac{1}{2m_k}\log\left(2\pi\right)\\ &-\frac{1}{2m_k}\left(-\frac{m_k\left(b_k - a_k\right) - \sqrt{m_k a_k a'_k} Q^{-1}\left(\alpha_k\right)}{\sqrt{m_k b_k \cdot b'_k}}\right)^2.\end{aligned} \tag{C.4}$$

Evaluating the asymptotic behavior, we take the $\lim_{m_k\to\infty}\frac{\log\left(\beta_R^k\right)}{m_k}$. The first and second terms in (C.4) converge to zero and after simplification on the third term, the inequality reduces to

$$\lim_{m_k\to\infty}\frac{\log\left(\beta_R^k\right)}{m_k} < -\frac{1}{2}\frac{\left(b'_k - a_k\right)^2}{b_k \cdot b'_k}.$$

Similarly,

$$\lim_{m_k\to\infty}\frac{\log\left(\beta_R^k\right)}{m_k} > -\frac{1}{2}\frac{\left(b_k' - a_k\right)^2}{b_k \cdot b_k'}$$

under the inequality $Q(x) > \frac{x}{1+x^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\ \forall x > 0$. Thus,

$$\lim_{m_k\to\infty}\frac{\log\left(\beta_R^k\right)}{m_k} = -\frac{1}{2}\frac{\left(b_k' - a_k\right)^2}{b_k \cdot b_k'} = -g_R^k \tag{C.5}$$

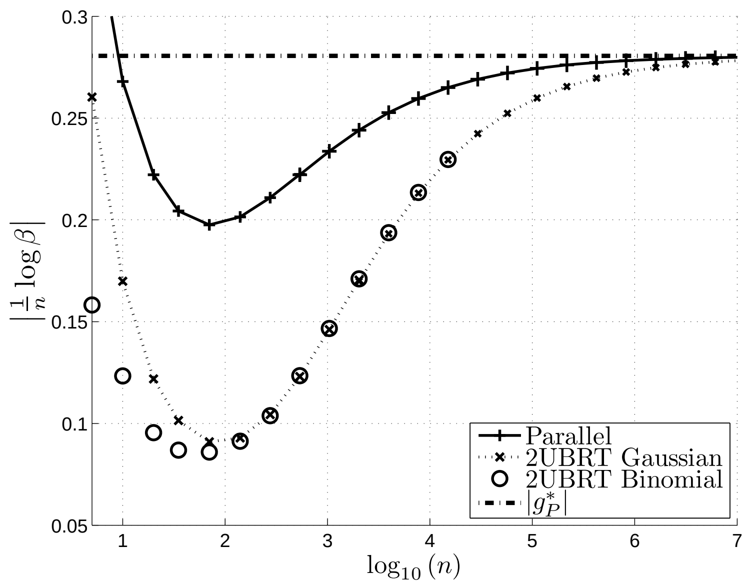
for class k, where g_R^k is the hUBRT Type II error probability exponent. Summing across classes, the hUBRT error exponent under the Gaussian approximation is

$$g_R^G = \sum_{k=1}^{N}\frac{1}{2}\frac{\left(b_k' - a_k\right)^2}{b_k \cdot b_k'}, \tag{C.6}$$

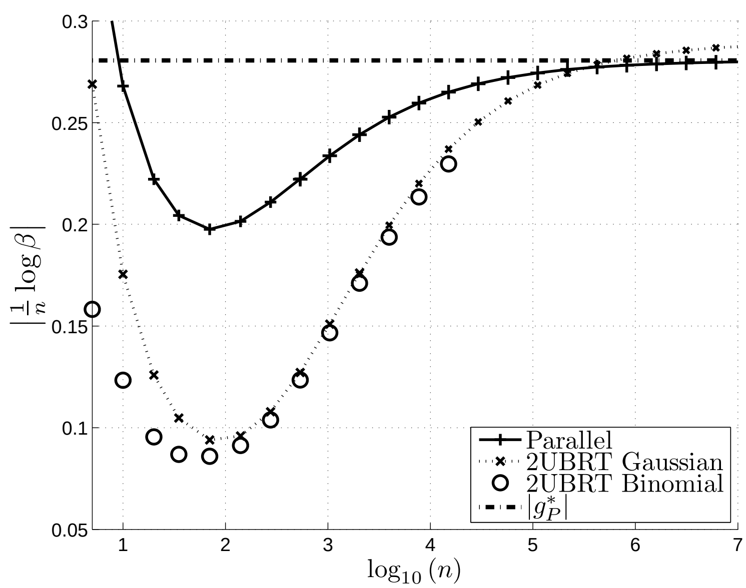
where it is assumed each m_k increases equally with n as $n \to \infty$ and the superscript G represents the Gaussian approximation model. Notice that g_R^G is equivalent to (C.1).

The hUBRT normalization factor for the relay network is then $\frac{D\left(P_0^f||P_1^f\right)}{g_R^G}$, where the numerator is the Binomial KLD (cf. (3.5)). The correction factor is then applied to the numerical simulation results using (C.4) for n small and the third term in (C.4) when n is large to generate the hUBRT data in Figure 3.2. These data are replicated in Figure C.1, but with the 2UBRT β_R calculated using a binomial random variable in class k up to numerical limits. Figure C.1a includes the Gaussian normalization factor of $\frac{D\left(P_0^f||P_1^f\right)}{g_R^G}$, with the normalization factor removed for Figure C.1b. These data indicate that the Gaussian approximation performs well with as little as 30 sensors in each class and that normalization is required for the Gaussian approximation to match the true binomial model in the non-asymptotic regime.

Determining the correction factor for the parallel configuration based on the fusion rule in (3.8) (e.g., Optimal C-V fusion rule) is straightforward for the case when both



(a) 2UBRT with Gaussian Normalization



(b) 2UBRT with No Gaussian Normalization

Figure C.1: 2UBRT Normalization Comparison: $N = 5$, with $\alpha = 0.1$, $h = 2$, $\boldsymbol{a} = [0.20,\ 0.65,\ 0.45,\ 0.15,\ 0.50]^T$, $\boldsymbol{b} = [0.7,\ 0.2,\ 0.4,\ 0.7,\ 0.3]^T$

b_k and a_k differ across all classes. Under the the Gaussian approximations for a weighted sum of binomial random variable in (3.9), the Gaussian KLD is calculated using

$$D^G\left(P_0^f||P_1^f\right) = \frac{\left(\mu_f - \mu_d\right)^2 + \sigma_f^2 - \sigma_d^2}{2\sigma_d^2} + \frac{1}{2}\log\frac{\sigma_d^2}{\sigma_f^2}, \tag{C.7}$$

where $\mu_f = \sum_{k=1}^N w_k m_k a_k$, $\sigma_f^2 = \sum_{k=1}^N w_k^2 m_k a_k a_k'$, $\mu_d = \sum_{k=1}^N w_k m_k b_k'$, $\sigma_d^2 = \sum_{k=1}^N w_k^2 m_k b_k' b_k$, and the superscript G implies the parallel configuration under the Gaussian approximation.

The correction factor for the parallel configuration is then the ratio of the two KLDs or $\frac{D\left(P_0^f||P_1^f\right)}{D^G\left(P_0^f||P_1^f\right)}$. This factor was applied to normalize the numerical calculations, resulting in the parallel configuration data presented in Figure 3.2, where an appropriate form for the third term in (C.4) is used for n large. However, when the a_k and b_k terms are common across all classes, then (C.7) must be replaced with (C.6) to determine the correction factor, as the parallel configuration fusion rule is equivalent to the class fusion rule in this particular numerical simulation case. Similarly, with a mix model where groups of class have common a_k and b_k terms, the hybrid model must be used with an appropriate mix of both (C.6) and (C.7).