Decentralized Detection in Sensor Network Architectures with Feedback

by

Spyridon Ilias Zoumpoulis

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

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2009

Author .....
Department of Electrical Engineering and Computer Science
May 22, 2009

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

Certified by .....
O. Patrick Kreidl
Research Affiliate, Laboratory for Information and Decision Systems
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Decentralized Detection in Sensor Network Architectures with Feedback

by

Spyridon Ilias Zoumpoulis

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 2009, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

We investigate a decentralized detection problem in which a set of sensors transmit a summary of their observations to a fusion center, which then decides which one of two hypotheses is true. The focus is on determining the value of feedback in improving performance in the regime of asymptotically many sensors. We formulate the decentralized detection problem for different network configurations of interest under both the Neyman-Pearson and the Bayesian criteria. In a configuration with feedback, the fusion center would make a preliminary decision which it would pass on back to the local sensors; a related configuration, the daisy chain, is introduced: the first fusion center passes the information from a first set of sensors on to a second set of sensors and a second fusion center. Under the Neyman-Pearson criterion, we provide both an empirical study and theoretical results. The empirical study assumes scalar linear Gaussian binary sensors and analyzes asymptotic performance as the signal-to-noise ratio of the measurements grows higher, to show that the value of feeding the preliminary decision back to decision makers is asymptotically negligible. This motivates two theoretical results: first, in the asymptotic regime (as the number of sensors tends to infinity), the performance of the "daisy chain" matches the performance of a parallel configuration with twice as many sensors as the classical scheme; second, it is optimal (in terms of the exponent of the error probability) to constrain all decision rules at the first and second stage of the "daisy chain" to be equal. Under the Bayesian criterion, three analytical results are shown. First, it is asymptotically optimal to have all sensors of a parallel configuration use the same decision rule under exponentially skewed priors. Second, again in the asymptotic regime, the decision rules at the second stage of the "daisy chain" can be equal without loss of optimality. Finally, the same result is proven for the first stage.

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

Thesis Supervisor: O. Patrick Kreidl
Title: Research Affiliate, Laboratory for Information and Decision Systems

Acknowledgments

I want to thank my advisors, Professor John Tsitsiklis and Dr. O. Patrick Kreidl, for their time and guidance; I consider myself privileged for having the chance to work with this advising scheme. John Tsitsiklis's deep insight and intuition motivated most of the interesting questions and results in this thesis; working with such a great teacher and researcher was a meaningful experience per se, and I am grateful for the lessons I have learnt, as I look forward to working with him in my Ph.D. Pat Kreidl's deep understanding and ability to combine intuition with mathematical rigor, along with his energy and enthusiasm, were often what pushed me towards progress; many of the ideas developed in this thesis were inspired through discussions with him.

I also want to thank my parents for their invaluable support during my studies so far. Their dedication to the values of education and constant personal growth has been a source of inspiration and courage; my natural attraction to the "quest" is owed to them.

This research was carried out at the Laboratory for Information and Decision Systems, partly with the support of the National Science Foundation, under grant ECCS-0701623 (Data Fusion Architectures).

Any findings expressed in this thesis do not necessarily reflect the views of the National Science Foundation.

Contents

List of Figures

Chapter 1

Introduction

1.1 Scope, background, and related work

Interest in sensor networks, data fusion, and distributed signal processing has virtually exploded in recent years, because of new technologies (especially the availability of low-cost sensing devices) and numerous potential applications. Research in the field involves problems that are both practically relevant and intellectually deep. On the application side, data fusion and sensor networks play a prominent role in a vast range of contexts; on the intellectual side, the development of new mathematical methods for new types of problems has obvious intrinsic intellectual merit.

The following two paragraphs, taken verbatim from the NSF proposal [14] which resulted in the funds that partly supported our research work, convey accurately the scope of our project.

"In the general context of the models that we will be considering, a sensor network consists of sensors (nodes) each of which makes a (generally noisy) observation of one or more random variables, related to a phenomenon of interest. The sensors use their observation, as well as messages received from other sensors, to form and transmit their own messages. Messages propagate through the network until eventually one (e.g. a fusion center) or multiple sensors form a final decision.

We are particularly interested in sensor networks operating in a regime of limited communication capabilities. Our focus on this regime reflects an emphasis on net-

works consisting of many, small, and inexpensive devices that have limited battery life and power, and cannot afford to communicate frequently or to transmit a lot of data. With abundant communication capabilities, the sensors could just share all their measurements, in which case the network aspects become immaterial, and we are faced with much easier, centralized information processing problems. In addition, we focus on the important issue of scalability, as the number of sensors increases. This is because there is abundant interest in sensor networks involving very large numbers of sensors, and also because some very difficult problems sometimes become tractable in the asymptotic regime. Our approach here is philosophically similar to the successful study of scalability issues in wireless networks, although the technical details are very different. Finally, our end goal is not only to optimize a given sensor network, but to derive important lessons on the merits of different network architectures. This is to be contrasted with the majority of the literature on decentralized detection, which assume a star or parallel configuration, with every sensor transmitting a message directly to a fusion center."

In a centralized scheme, each sensor communicates all of its observations to the fusion center; we discuss instead decentralized detection (introduced in [11]), where each sensor sends only a summary of its observations to the fusion center with a message that takes values in a finite alphabet. The fusion center then decides on one of the alternative hypotheses. The problem is to identify how each peripheral sensor should decide what message(s) to send, and how the fusion center should interpret these messages to make a final decision, in a manner that minimizes the probability of error. In our work performance is analyzed in the asymptotic regime, through the use of error exponents. The key point of our research is to assess the value (in terms of performance improvement) of feeding a preliminary decision back to sensors in tree networks.

A lot of research followed the seminal paper of Tenney and Sandell ([11]); for a review, see [13]. For conditionally dependent observations, the decision version of the problem becomes NP-complete (and the problem itself NP-hard). Under the assumption of conditional independence, the optimal decision rule for each sensor

takes the form of a likelihood ratio test, with a suitably chosen threshold. In turn, an optimization over the set of all thresholds can yield the desired solution.

We briefly note several pieces of work that study decentralized detection under a Bayesian or Neyman-Pearson criterion. By no means are the works mentioned here exhaustive; rather, we list those that most closely relate to the analyses we will perform in this work. [1] presents a unified approach to the study of decentralized detection systems with any configuration, and introduces and evaluates the performance of a new toplogical structure, namely a decentralized detection system with peer communication; this structure features feedback and assumes that sensors have memory. [2] investigates a decentralized detection system with feedback and memory using the Bayesian formulation and finds the system probability of error to be no larger than that for a parallel configuration without feedback, provided the mechanism for feedback is properly designed. In addition, analysis and experiments show that the probability of error decreases to zero as the number of fed back observations goes to infinity. [7] considers a Neyman-Pearson formulation to compare two different feedback architectures: the first permits feedback only from the fusion center back to the local decision makers, while the second permits feedback among all peripheral sensors. The superiority of the latter structure over the former is empirically established and it is shown that the contribution of feedback decreases exponentially with the signal-to-noise ratio and the number of local sensors. [9] explores the use of feedback (of all sensor decisions to all sensors) and successive retesting and rebroadcasting of the updated decisions until a consensus is reached, an operation identified as "parley". Two modes of operation of "parley" are considered. In the first, all sensors are as correct as possible at all times; convergence to a consensus is demonstrated to be quite fast, at the expense of performance. In the second, the decision reached by the consensus is constrained to be optimum in the sense that it would match that of a centralized processor, at the expense of haste.

The underlying non-convexity of the problem of optimizing the decision rules as the number of sensors increases has promoted research into more tractable asymptotic formulations (i.e., when the number of sensors approaches infinity). [12] focuses on

optimizing the asymptotic error exponent, defined by the normalized logarithm of the probability of error as the number of nodes goes to infinity, for the case of a parallel configuration with a large number of sensors that receive conditionally independent, identically distributed observations. [10] considers the broader problem of decentralized binary detection in a network consisting of a large number of nodes arranged as a tree of bounded height; the error probability is shown to decay exponentially fast with the number of nodes under both the Bayesian and the Neyman-Pearson formulation. Necessary and sufficient conditions are provided for the optimal error exponent to be the same as that corresponding to a parallel configuration, under the Neyman-Pearson criterion. [4] studies decentralized binary detection in wireless sensor networks where each sensor transmits its data over a multiple access channel. Under constraints for the capacity of the wireless channel, it is proven that for the problem of detecting deterministic signals in additive Gaussian noise, having a set of identical binary sensors is asymptotically optimal, as the number of observations per sensor goes to infinity.

Finally, [6] escapes the asymptotic regime and the limitations of the parallel configuration to discuss message-passing algorithms for online measurement processing and offline strategy optimization that exploit sparse graph structure of a sensor network. While the parallel configuration is not sparse from the fusion center's perspective, the associated algorithm can be applied for problems of up to 10 nodes.

1.2 Contributions and thesis outline

In Chapter 2 we formally introduce the basic model that overarches the analysis of this thesis. We first define the detection problem for all network configurations discussed in this thesis: after defining the classical decentralized detection problem, we formulate problems associated with other configurations of interest. We then provide a framework in which the asymptotics of decentralized detection shall be discussed.

Chapter 3 provides some motivating examples, as well as easy-to-draw compar-

isons between the performance of different configurations. In the first part of this chapter, the "simulation" argument (i.e., simulating the communication capabilities of a network configuration with the communication capabilities of another) is used to compare the performance of configurations proposed in Chapter 2 in the optimal error exponent sense. In the second part, we consider a classical decentralized detection model (as defined in Subsection 2.1.1) when all sensors are scalar linear Gaussian binary detectors, and analyze asymptotic performance as the measurements under the two hypotheses become more and more informative.

In Chapter 4 we prove that under the Neyman-Pearson formulation, the performance of the "daisy chain" configuration (Subsection 2.1.4) is asymptotically equal to the performance of the parallel configuration with twice as many sensors as the classical scheme (Subsection 2.1.2), in the optimal error exponent sense. We first argue that the "daisy chain" cannot be worse than the parallel configuration with twice as many sensors as the classical scheme; we then prove that it cannot perform better. We conclude that the value of feeding the preliminary decision to a second set of sensors is asymptotically negligible. We also prove that there is no loss of optimality asymptotically in the "daisy chain" if all decision rules at both stages are constrained to be equal.

Chapter 5 solves the Bayesian detection problem for the "daisy chain" configuration. [12] proves that for the parallel topology, it is asymptotically optimal to let all sensors use the same decision rule in deciding what to transmit. In particular, it is asymptotically optimal to have all sensors perform identical likelihood ratio tests, using the same threshold. First we extend this result to the case of exponentially skewed priors; we then make the connection between the result for exponentially skewed priors and Bayesian detection in the "daisy chain". Specifically, we prove that under the mild assumption that the Type I and Type II error probabilities of the first stage decay exponentially with the number of sensors, it is asymptotically optimal to have all sensors in the second stage perform identical likelihood ratio tests. Finally, we prove that it is asymptotically optimal to have all sensors in the first stage perform identical likelihood ratio tests as well. In all cases, optimality is meant in terms of

the overall system's optimal error exponent.

Finally, in Chapter 6 we summarize and suggest directions for future research.

Chapter 2

The Basic Model

In this chapter we formally introduce the basic model that overarches the analysis of this thesis. We first define the detection problem for all network topologies discussed in our work: after defining the classical decentralized detection problem, which assumes a parallel configuration, we formulate problems associated with other configurations of interest. When defining the problem for configurations other than the parallel, we focus on aspects in which the problem deviates from the parallel configuration problem, without repeating common aspects. We then provide a framework in which the asymptotics of decentralized detection shall be discussed.

2.1 Network configurations of interest for decentralized detection

2.1.1 The classical decentralized detection problem

In the Bayesian formulation, the state of the environment can be captured by one of M alternative hypotheses H_0, H_1, ..., H_{M-1}, with known positive prior probabilities $\Pr(H_i)$. Let set $\mathcal{X}$ be endowed with a σ-field $\mathcal{F}_{\mathcal{X}}$ of measurable sets. There are N sensors indexed $1, \ldots, N$ and each sensor i observes a realization of a random variable X_i, which takes values in $\mathcal{X}$. We assume that conditioned on hypothesis H_j, the random variables X_i are independent and identically distributed (i.i.d.) with

a known conditional distribution $\mathcal{P}_j$. Let D be a positive integer; let $\mathcal{T}$ be a predetermined symbol alphabet with $|\mathcal{T}| = D$. Each sensor i evaluates a D-valued message $Y_i \in \mathcal{T}$ as a function of its observation: $Y_i = \gamma_i(X_i)$, where the function $\gamma_i : \mathcal{X} \mapsto \mathcal{T}$ (assumed measurable) is the decision rule (or transmission function) of sensor i. Messages $Y_1, \ldots, Y_N$ are communicated to a fusion center which uses a fusion rule $\gamma_0 : \mathcal{T}^N \mapsto \{0, \ldots, M-1\}$ and declares hypothesis H_j to be true if and only if $Y_0 = \gamma_0(Y_1, \ldots, Y_N) = j$. The objective is to choose the rules $\gamma_0, \gamma_1, \ldots, \gamma_N$ so as to minimize the probability of an erroneous decision at the fusion center.

In the Neyman-Pearson formulation, the problem is set up in the same way but priors $\Pr(H_i)$ need not be defined. For the case $M = 2$ the objective is to choose the rules $\gamma_0, \gamma_1, \ldots, \gamma_N$ so as to minimize the missed detection probability $\mathcal{P}_1(Y_0 = 0)$ subject to a constraint $\mathcal{P}_0(Y_0 = 1) \leq \alpha$ on the false alarm probability, where α is a constant such that $0 < \alpha < 1$.

The parallel configuration we just described is shown in Figure 2-1. We are mostly interested in the case where $M = 2$ and $D = 2$ (each peripheral sensor transmits one bit to the fusion center) or $D = 4$ (each peripheral sensor transmits two bits to the fusion center).

2.1.2 Decentralized detection with double the number of sensors

We will also consider the configuration in which there are $2N$ sensors, indexed $1, \ldots, N, N+1, \ldots, 2N$. The setup is analogous to the one described in Section 2.1.1. Messages $Y_1, \ldots, Y_N, Y_{N+1}, \ldots, Y_{2N}$ are communicated to a fusion center which uses a fusion rule $\gamma_0 : \mathcal{T}^{2N} \mapsto \{0, \ldots, M-1\}$ and declares hypothesis H_j to be true if and only if $Y_0 = \gamma_0(Y_1, \ldots, Y_N, Y_{N+1}, \ldots, Y_{2N}) = j$. In both the Bayesian and the Neyman-Pearson formulation, the objective is to choose the rules $\gamma_0, \gamma_1, \ldots, \gamma_N, \gamma_{N+1}, \ldots, \gamma_{2N}$ according to the respective criterion.

The parallel configuration just described is shown in Figure 2-2. We are mostly interested in the case where $M = 2$ and $D = 2$ (each peripheral sensor transmits one

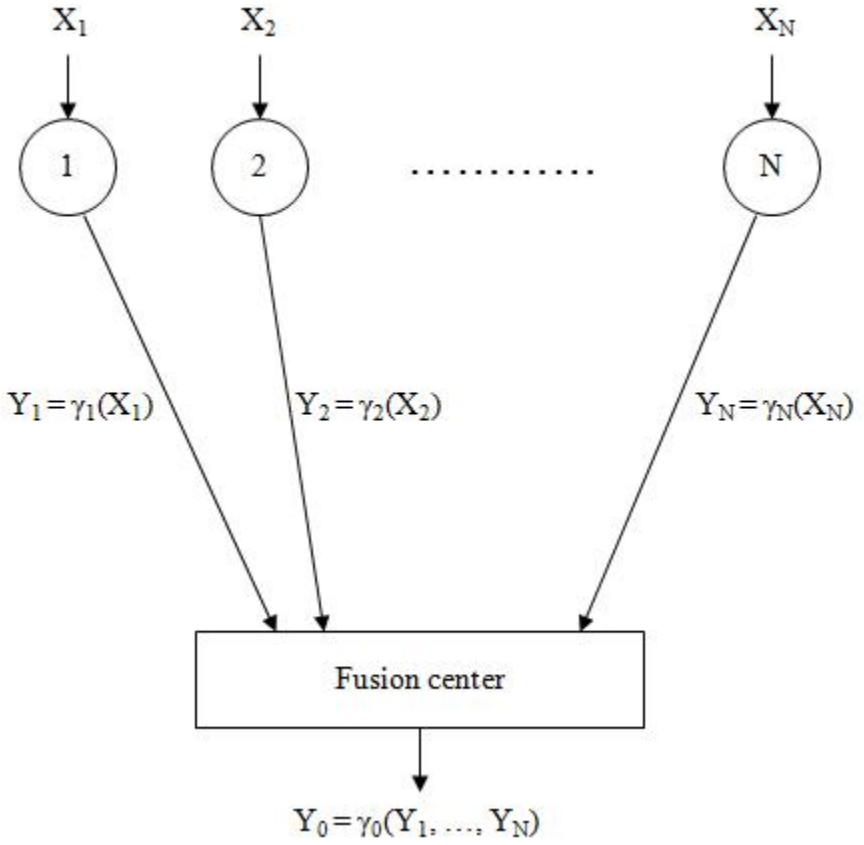

Figure 2-1: The parallel configuration.

bit to the fusion center).

2.1.3 Decentralized detection with simple feedback

Let D_1, D_2 be positive integers; let $\mathcal{T}_1$, $\mathcal{T}_2$ be predetermined symbol alphabets with $|\mathcal{T}_1| = D_1$ and $|\mathcal{T}_2| = D_2$. We consider a two-stage process. At the first stage, each sensor i evaluates a D_1-valued message $Y_i \in \mathcal{T}_1$ as a function of its observation: $Y_i = \gamma_i(X_i)$, where the function $\gamma_i : \mathcal{X} \mapsto \mathcal{T}_1$ (assumed measurable) is the decision rule (or transmission function) of sensor i in this first stage. Messages $Y_1, \ldots, Y_N$ are communicated to a fusion center which uses a fusion rule $\gamma_0 : \mathcal{T}_1^N \mapsto \{0, \ldots, M-1\}$. Intuitively, the fusion center "believes" hypothesis H_j to be true if and only if $Y_0 = \gamma_0(Y_1, \ldots, Y_N) = j$. The fusion center sends its current belief Y_0 back to each of the N sensors. At the second stage, each sensor i now evaluates a D_2-valued message

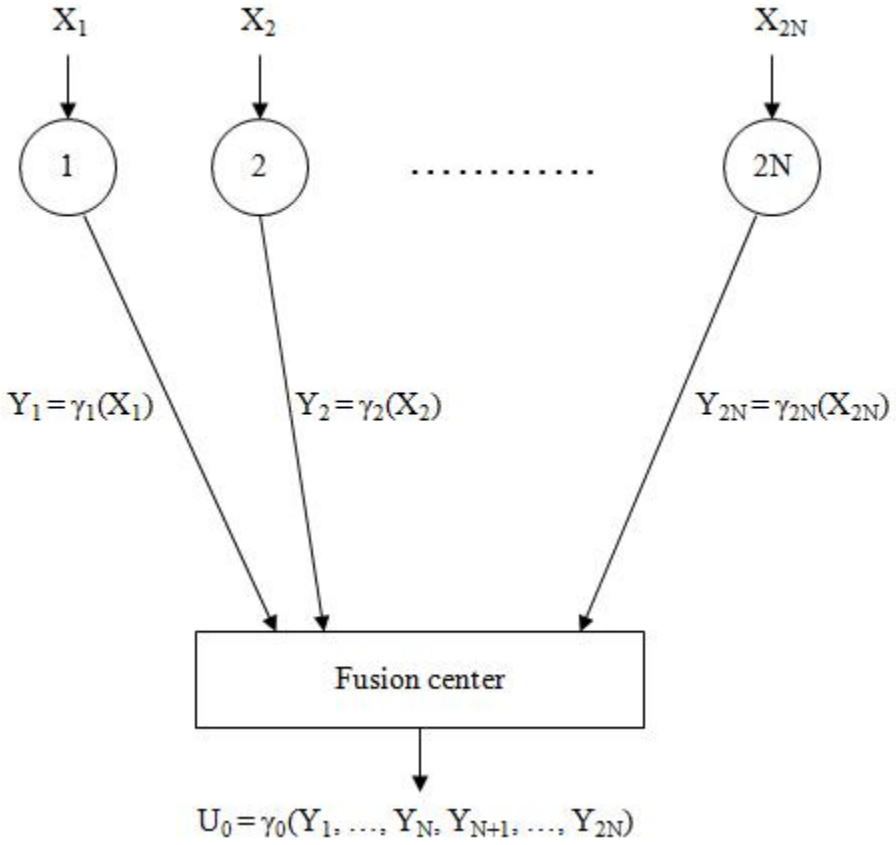

Figure 2-2: The parallel configuration with double the number of sensors.

$U_i \in \mathcal{T}_2$ as a function of its observation and the feedback: $U_i = \delta_i(X_i, Y_0)$, where the function $\delta_i : \mathcal{X} \times \{0, \ldots, M-1\} \mapsto \mathcal{T}_2$ (assumed measurable) is the decision rule (or transmission function) of sensor i in this second stage. Messages $U_1, \ldots, U_N$ are communicated to the fusion center which uses a fusion rule $\delta_0 : \mathcal{T}_2^N \mapsto \{0, \ldots, M-1\}$ and declares hypothesis H_j to be true if and only if $U_0 = \delta_0(U_1, \ldots, U_N) = j$. In both the Bayesian and the Neyman-Pearson formulation, the objective is to choose the rules $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ according to the respective criterion.

The above described parallel configuration with feedback is shown in Figure 2-3. We are mostly interested in the case when $M = 2$ and $D_1 = D_2 = 2$ (each sensor transmits one bit to the fusion center in each stage of detection/communication).

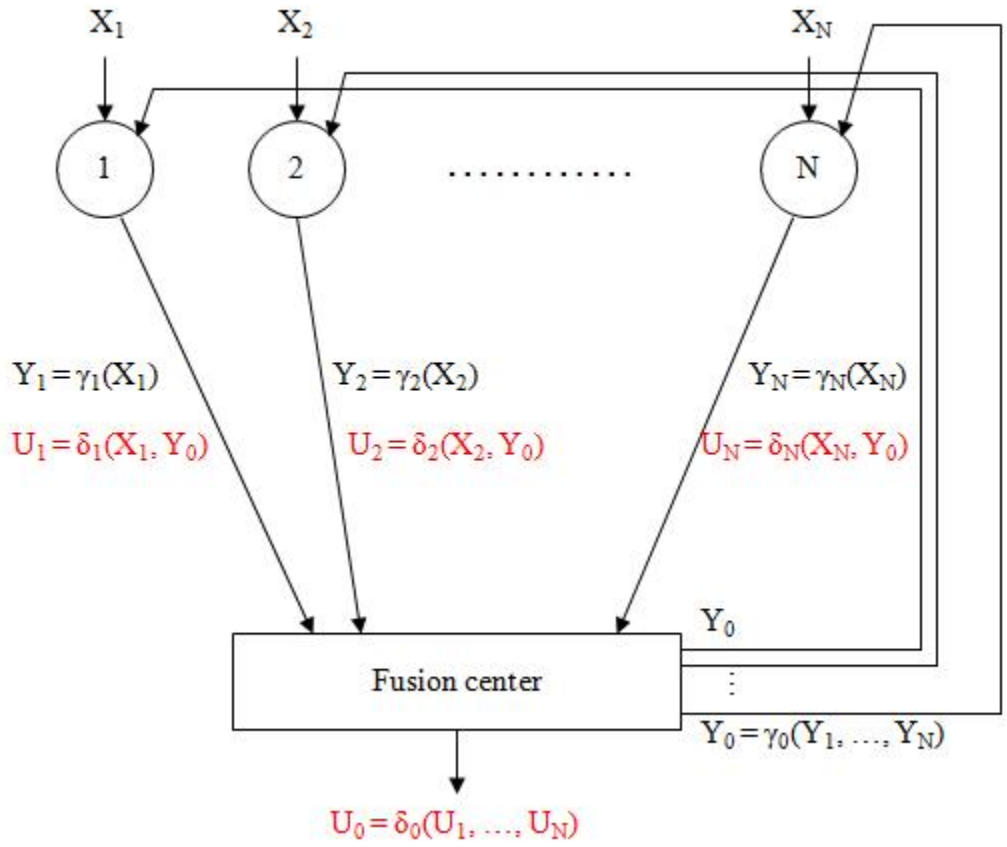

Figure 2-3: The parallel configuration with feedback.

2.1.4 Decentralized detection with two nested groups of sensors - The "daisy chain" configuration

In this configuration there are $2N$ sensors indexed $1, \ldots, N, N+1, \ldots, 2N$ and each sensor i observes a realization of a random variable X_i, which takes values in $\mathcal{X}$. Let D_1, D_2 be positive integers; let $\mathcal{T}_1$, $\mathcal{T}_2$ be predetermined symbol alphabets with $|\mathcal{T}_1| = D_1$ and $|\mathcal{T}_2| = D_2$. Each sensor i in the set of sensors indexed $1, \ldots, N$ evaluates a D_1-valued message $Y_i \in \mathcal{T}_1$ as a function of its observation: $Y_i = \gamma_i(X_i)$, where the function $\gamma_i : \mathcal{X} \mapsto \mathcal{T}_1$ (assumed measurable) is the decision rule (or transmission function) of sensor i. Messages $Y_1, \ldots, Y_N$ are communicated to a fusion center which uses a fusion rule $\gamma_0 : \mathcal{T}_1^N \mapsto \{0, \ldots, M-1\}$; intuitively, it "believes" hypothesis H_j to be true if and only if $Y_0 = \gamma_0(Y_1, \ldots, Y_N) = j$. The fusion center sends its belief Y_0 to each of the N sensors indexed $N+1, \ldots, 2N$. Each sensor i in the set of

sensors indexed $N+1, \ldots, 2N$ now evaluates a D_2-valued message $U_{i-N} \in \mathcal{T}_2$ as a function of its observation and the fusion center's belief: $U_{i-N} = \delta_{i-N}(X_i, Y_0)$, where the function $\delta_{i-N} : \mathcal{X} \times \{0, \ldots, M-1\} \mapsto \mathcal{T}_2$ (assumed measurable) is the decision rule (or transmission function) of sensor i. Messages $U_1, \ldots, U_N$, along with the fusion center's belief Y_0, are communicated to a second (global) fusion center which uses a fusion rule $\delta_0 : \mathcal{T}_2^N \times \{0, \ldots, M-1\} \mapsto \{0, \ldots, M-1\}$ and declares hypothesis H_j to be true if and only if $U_0 = \delta_0(U_1, \ldots, U_N, Y_0) = j$. In both the Bayesian and the Neyman-Pearson formulation, the objective is to choose the rules $\gamma_0, \gamma_1, \ldots, \gamma_N$, $\delta_0, \delta_1, \ldots, \delta_N$ according to the respective criterion.

The above described configuration is shown in Figure 2-4. We are mostly interested in the case when $M = 2$ and $D_1 = D_2 = 2$ (each sensor transmits a total of one bit).

2.2 The asymptotic regime

It is widely known that, as the number of sensors grows, the probability of error goes to zero exponentially fast for any "reasonable" set of decision rules. This calls for a more refined measure of performance as $N \to \infty$. We describe the large N regime for the classical decentralized detection problem of Subsection 2.1.1, in both the Bayesian and the Neyman-Pearson formulations. The formulation of the asymptotic regime of other network configurations of interest is similar.

2.2.1 Bayesian asymptotics

As [12] notices, having fixed the decision rules $\gamma_1, \ldots, \gamma_N$ of the sensors, the optimal decision for the fusion center γ_0 is the maximum a posteriori (MAP) probability rule. Thus, we will be concerned only with optimization with respect to $\gamma_1, \ldots, \gamma_N$. We follow the notation used in [12] to denote any such set of decision rules by γ^N. Let Γ be the set of all measurable functions from $\mathcal{X}$ into $\mathcal{T}$, and Γ^N be the Cartesian product of Γ with itself N times. For any $\gamma^N \in \Gamma^N$, we define $J_N(\gamma^N)$ to be the probability of an erroneous decision by the fusion center. For any given N and choice

of decision rules γ^N, we define the exponent of the error probability to be

$$r_N(\gamma^N) = \frac{\log J_N(\gamma^N)}{N}. \tag{2.1}$$

Let

$$R_N = \inf_{\gamma^N \in \Gamma^N} r_N(\gamma^N) \tag{2.2}$$

be the optimal error exponent. We focus on the asymptotic optimal error exponent, $\lim_{N\to\infty} R_N$, which we denote g_P^*. In the relevant literature, g_P^* is often termed "the optimal error exponent", as opposed to "the asymptotic optimal error exponent"; we respect this convention. In this thesis, optimal error exponent refers to g_P^*, unless otherwise indicated.

2.2.2 Neyman-Pearson asymptotics

Let $M = 2$, and let Γ be again the set of all measurable functions from $\mathcal{X}$ into $\mathcal{T}$. We allow the decision rule of the fusion center γ_0 to be randomized; as [12] notes, the final decision of the fusion center may depend on the decisions of the local decision-makers as well as an internally generated random variable. Let Γ_0 be the set of all candidate decision rules γ_0. For any given choice of decision rules $(\gamma_0, \gamma_1, \ldots, \gamma_N) \in \Gamma_0 \times \Gamma^N$, we define the Type I and Type II error probabilities to be respectively:

$$J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N) = \mathcal{P}_0(\gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N)) = 1), \tag{2.3}$$

$$J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N) = \mathcal{P}_1(\gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N)) = 0). \tag{2.4}$$

We require that $J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N)$ be no more than a given $\alpha \in (0, 1)$ and we are interested in minimizing $J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N)$ over all $\gamma_0, \gamma_1, \ldots, \gamma_N$ satisfying

$$J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N) \leq \alpha. \tag{2.5}$$

We define

$$r_N(\gamma_0, \gamma_1, \ldots, \gamma_N) = \frac{\log J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N)}{N}. \tag{2.6}$$

Let

$$R_N = \inf_{(\gamma_0,\gamma_1,\ldots,\gamma_N)\in\Gamma_0\times\Gamma^N \text{ s.t. } J_N^I(\gamma_0,\gamma_1,\ldots,\gamma_N)\leq\alpha} r_N(\gamma_0,\gamma_1,\ldots,\gamma_N). \tag{2.7}$$

Suppose that X_i is a sensor observation. For any $\gamma \in \Gamma$, let the distribution of $\gamma(X_i)$ under hypothesis H_j be $\mathcal{P}_j^\gamma$. We list some of the common assumptions we will be making in the sequel. For the proofs of all results in this thesis pertaining to Neyman-Pearson detection, we make these assumptions, unless otherwise indicated.

Assumption 2.1. *The measures $\mathcal{P}_0$ and $\mathcal{P}_1$ are equivalent, i.e. they are absolutely continuous with respect to each other. Furthermore, there exists some $\gamma \in \Gamma$ such that $-D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) < 0 < D(\mathcal{P}_1^\gamma \| \mathcal{P}_0^\gamma)$.*

Assumption 2.2. $\mathbf{E}\left[\log^2 \frac{d\mathcal{P}_1}{d\mathcal{P}_0}\right] < \infty$*, where $\frac{d\mathcal{P}_1}{d\mathcal{P}_0}$ is the Radon-Nikodym derivative of the two measures, and the expectation is taken with respect to $\mathcal{P}_0$.*

It was shown in [12] that under Assumptions 2.1 and 2.2, the asymptotic optimal error exponent, which we denote g_P^*, is given by

$$g_P^* = \lim_{N\to\infty} R_N = -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma), \tag{2.8}$$

where D denotes the Kullback-Leibler divergence of two probability measures. In the relevant literature, g_P^* is often termed "the optimal error exponent", as opposed to "the asymptotic optimal error exponent"; we respect this convention. In this thesis, optimal error exponent refers to g_P^*, unless otherwise indicated.

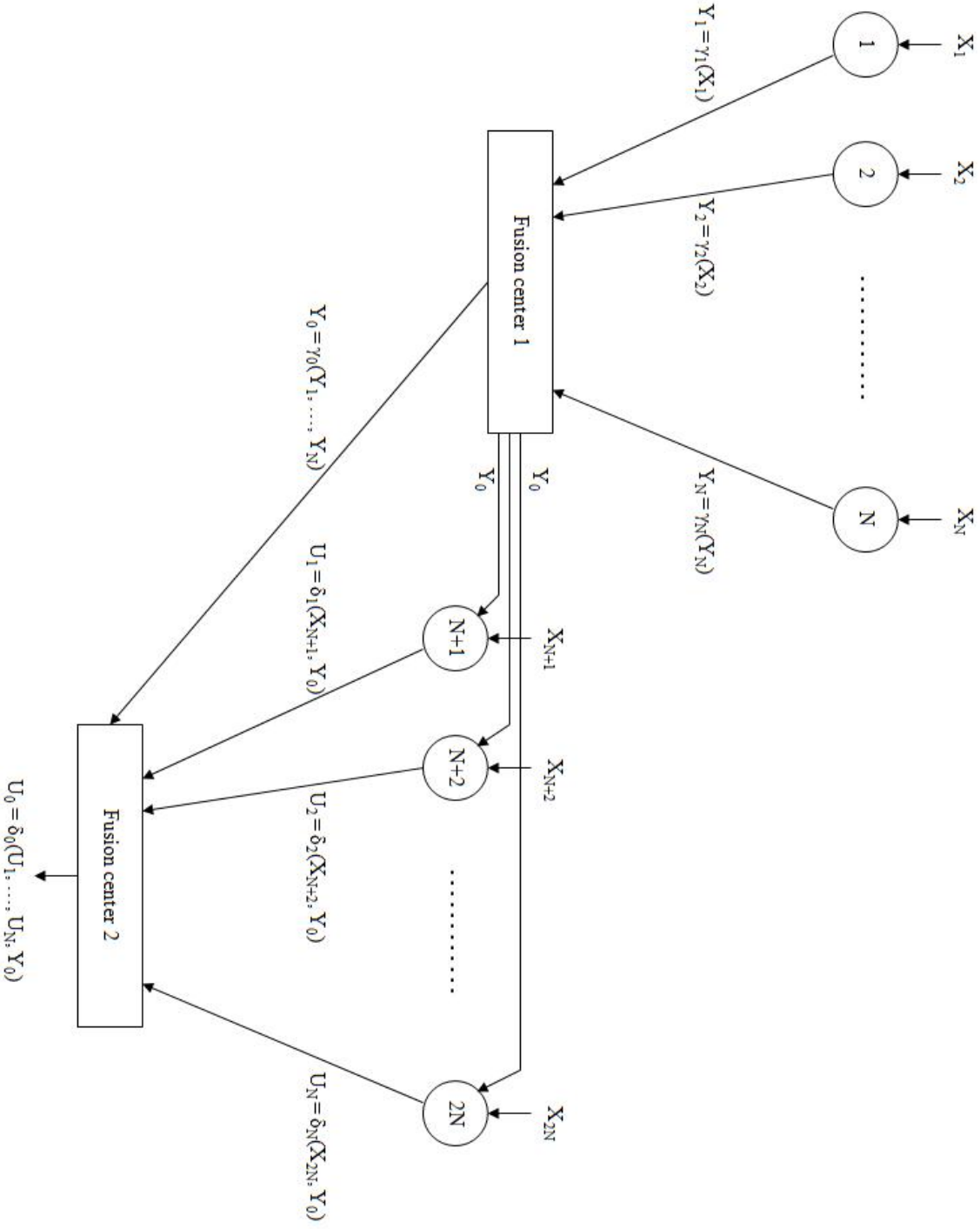

Figure 2-4: The "daisy chain" configuration.

Chapter 3

Motivation and Simple Examples

In this chapter we provide some motivating examples, as well as easy-to-draw comparisons between the performance of different configurations. In the first part of this chapter, the "simulation" argument (i.e., simulating the communication capabilities of a network configuration with the communication capabilities of another) is used to compare the performance of configurations proposed in Chapter 2, in the optimal error exponent sense. In the second part, we study the classical decentralized detection problem (as defined in Subsection 2.1.1) when all sensors are scalar linear Gaussian binary detectors, and analyze asymptotic performance as the measurements under the two hypotheses become more and more informative, i.e. as the signal-to-noise ratio grows higher. The results motivate the analysis in Chapter 4: the value of feeding the preliminary decision to decision makers is asymptotically negligible.

3.1 Comparison between configurations using the "simulation" argument

For finitely many sensors, the metric of the performance of a configuration would be the probability of an erroneous decision, with smaller probability of error implying better performance. In the asymptotic regime, we are interested in the exponent of the error probability. Assuming exponential decay of error probabilities, the error

exponents of interest are negative, and smaller values correspond to better performance.

We restrict our attention to the Neyman-Pearson paradigm. Assuming all exponents are well defined, let g_P^*, g_{2b}^*, g_{2N}^*, g_F^*, g_{DC}^* denote the optimal error exponent of the parallel configuration when all sensors send one bit to the fusion center ($D = 2$), the parallel configuration when all sensors send 2 bits to the fusion center ($D = 4$), the parallel configuration with twice as many sensors, the simple feedback network and the "daisy chain", respectively. g_P^* and g_{DC}^* are formally defined in 2.2.2 and 4.1, while the definition of g_{2b}^*, g_{2N}^*, and g_F^* is similar.

Proposition 3.1. *Assuming g_P^*, g_{2b}^*, g_{2N}^*, g_F^*, g_{DC}^* are all well defined,*

$$-\infty < -D(\mathcal{P}_0||\mathcal{P}_1) \leq g_F^* \leq g_{2b}^* \leq g_P^* < 0$$

$$-\infty < -D(\mathcal{P}_0||\mathcal{P}_1) \leq g_{DC}^* \leq g_{2N}^* \leq g_P^* < 0$$

Proof. No configuration of interest can do better than a network in which all the observations are provided uncompressed to the fusion center, in which case the error exponent is $-D(\mathcal{P}_0||\mathcal{P}_1)$ by the Stein lemma (see [5]), and is finite as a consequence of Assumption 2.2.

The simple feedback configuration can simulate communication in the parallel configuration when all sensors send 2 bits to the fusion center; the latter can simulate communication in the parallel configuration when all sensors can only send 1 bit to the fusion center.

Similarly, $g_{DC}^* \leq g_{2N}^*$ is Lemma 4.1. The parallel configuration with $2N$ sensors can simulate the communication capabilities of the parallel configuration with N sensors.

It remains to show that $g_P^* < 0$, which follows from [12]. □

Figures 3-1 and 3-2 summarize the above results.

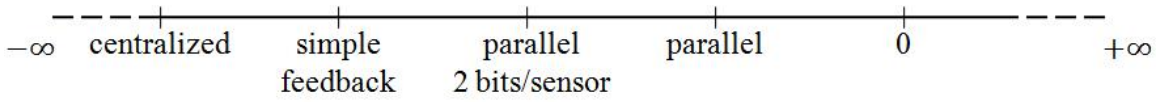

Figure 3-1: The real axis of error exponents: comparisons with the simple feedback configuration.

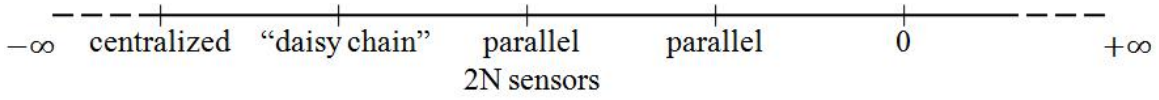

Figure 3-2: The real axis of error exponents: comparisons with the "daisy chain".

3.2 Neyman-Pearson detection with scalar linear Gaussian binary detectors

Let us consider a classic decentralized detection model with $M = 2$, $D = 2$ under the Neyman-Pearson formulation. We assume that conditioned on hypothesis H_0, the random variables X_i are independent and identically distributed (i.i.d.) with a known conditional distribution $\mathcal{P}_0$ that is normal with mean $-\mu < 0$ and variance $\sigma^2 > 0$; conditioned on hypothesis H_1, the random variables X_i are independent and identically distributed (i.i.d.) with a known conditional distribution $\mathcal{P}_1$ that is normal with mean μ and variance σ^2. The densities $f_{X|H}(x|H_0)$ and $f_{X|H}(x|H_1)$, correspond to the measurement distributions $\mathcal{P}_0$ and $\mathcal{P}_1$ respectively.

The error exponent of a parallel configuration in which all the observations are provided uncompressed to the fusion center is $-D(\mathcal{P}_0||\mathcal{P}_1)$ by the Stein lemma (see [5]); as discussed in Section 3.1, summarizing the raw data leads to worse error exponents. Nevertheless we show that for the Neyman-Pearson decentralized detection problem with Gaussian measurements as described above, the performance converges to the centralized optimal performance in the error exponent sense as measurements become more informative (i.e., as the means of distributions $\mathcal{P}_0$, $\mathcal{P}_1$ are pulled further apart).

Calculating the error exponent for a centralized system, we have

$$\begin{aligned}
-D(\mathcal{P}_0||\mathcal{P}_1) &= -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}} \log \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}\, dx \\
&= -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}} \log e^{-\frac{2x\mu}{\sigma^2}}\, dx \\
&= \frac{2\mu}{\sigma^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}} x\, dx \\
&= \frac{2\mu}{\sigma^2}(-\mu) \\
&= -\frac{2\mu^2}{\sigma^2}. \qquad (3.1)
\end{aligned}$$

Let $\mathcal{L}(x)$ denote the likelihood ratio at x for a decision rule $\gamma \in \Gamma$. Let "H_0'', "H_1'' stand for deciding to send a bit in favor of H_0 and H_1 respectively. A likelihood ratio test will be of the form

$$\begin{aligned}
\mathcal{L}(x) = \frac{f_{X|H}(x|H_1)}{f_{X|H}(x|H_0)} &\gtrless_{\text{“}H_0''}^{\text{“}H_1''} \epsilon &\iff \\
\frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+\mu)^2}{2\sigma^2}}} &\gtrless_{\text{“}H_0''}^{\text{“}H_1''} \epsilon &\iff \\
e^{\frac{2x\mu}{\sigma^2}} &\gtrless_{\text{“}H_0''}^{\text{“}H_1''} \epsilon &\iff \\
x &\gtrless_{\text{“}H_0''}^{\text{“}H_1''} \frac{\sigma^2 \log \epsilon}{2\mu}\ ,
\end{aligned}$$

where ϵ is a parameter. Let us define $x^t_{\epsilon,\mu,\sigma} = \frac{\sigma^2 \log \epsilon}{2\mu}$. The choice of ϵ and therefore of threshold $x^t_{\epsilon,\mu,\sigma}$ determines decision rule γ completely. Suppose that X is a sensor observation and, as in Subsection 2.2.2, let the distribution of $\gamma(X)$ under hypothesis H_j be $\mathcal{P}_j^\gamma$. It is clear that the distributions $\mathcal{P}_0^\gamma$, $\mathcal{P}_1^\gamma$ will be Bernoulli with parameter $1-\Phi(\frac{x^t_{\epsilon,\mu,\sigma}+\mu}{\sigma})$ and $1-\Phi(\frac{x^t_{\epsilon,\mu,\sigma}-\mu}{\sigma})$ respectively, where $\Phi(\,\cdot,\,)$ is the cumulative distribution function of a standard normal random variable. The Kullback-Leibler divergence of

$\mathcal{P}_1^\gamma$ from $\mathcal{P}_0^\gamma$ is

$$D(\mathcal{P}_0^\gamma||\mathcal{P}_1^\gamma)=\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}+\mu}{\sigma}\right)\log\frac{\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}+\mu}{\sigma}\right)}{\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}-\mu}{\sigma}\right)} + \left(1-\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}+\mu}{\sigma}\right)\right)\log\frac{1-\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}+\mu}{\sigma}\right)}{1-\Phi\left(\frac{x^t_{\epsilon,\mu,\sigma}-\mu}{\sigma}\right)}. \tag{3.2}$$

We gain some insight by running some numerical simulations involving the KL-divergence of interest. We fix $\sigma = 1$. Figure 3-3 plots the negative K-L divergence of $\mathcal{P}_1^\gamma$ from $\mathcal{P}_0^\gamma$ as a function of the threshold $x^t_{\epsilon,1,1}$ for $\mu = 1$, while Figure 3-4 plots the same (negative) K-L divergence for various values of the mean μ. Notice that the function has a unique local minimum, and that the K-L divergence takes greater values as μ increases.

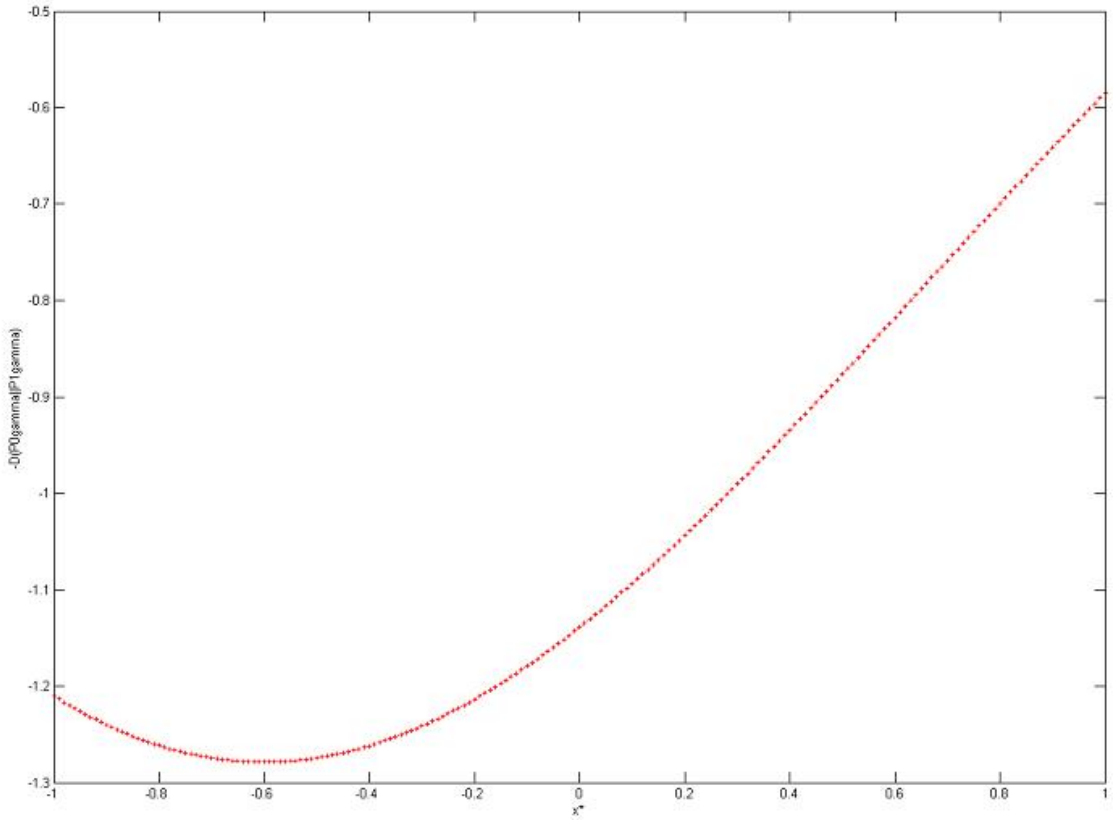

Figure 3-3: Plot of the negative K-L divergence $-D(\mathcal{P}_0^\gamma||\mathcal{P}_1^\gamma)$ as a function of the threshold $x^t_{\epsilon,1,1}$ (for $\mu = 1$).

It is of interest to investigate how the decentralized detection model defined above

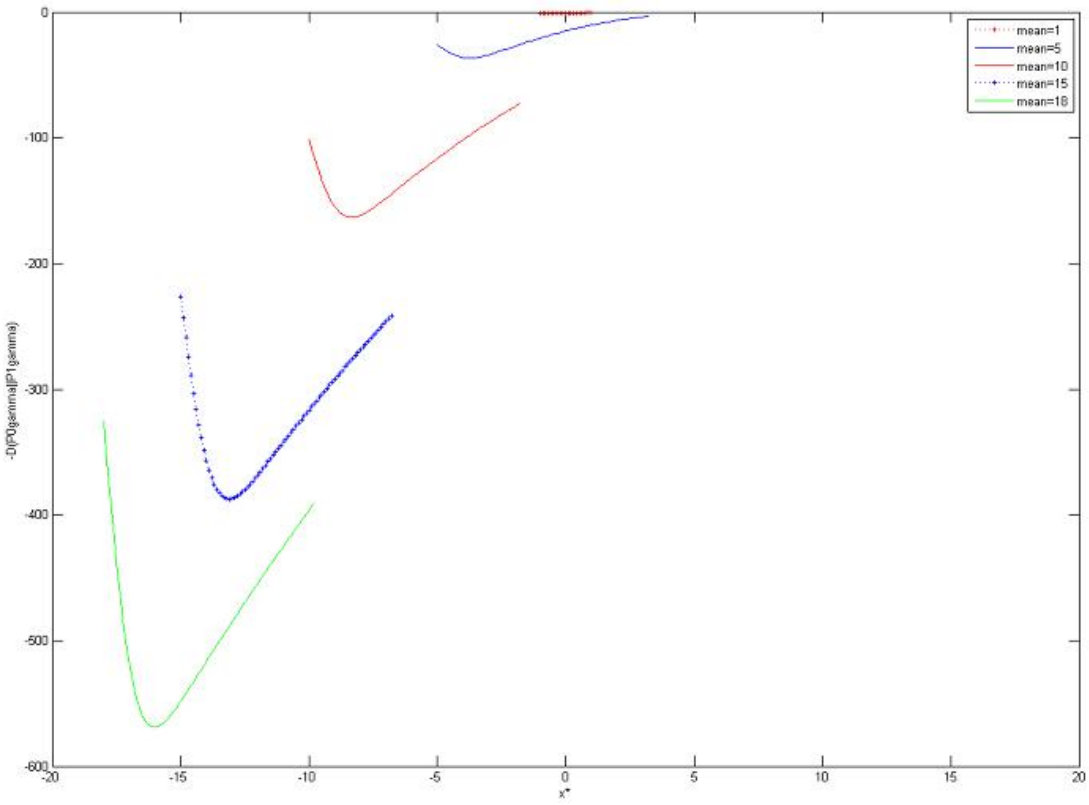

Figure 3-4: Plot of the negative K-L divergence $-D(\mathcal{P}_0^\gamma||\mathcal{P}_1^\gamma)$ as a function of the threshold $x^t_{\epsilon,\mu,1}$ for various values of the mean.

performs in comparison to centralized detection. Figure 3-5 plots the ratio of the error exponent for the centralized system over the optimal error exponent for the decentralized parallel detection model as a function of the mean μ. Notice that the ratio appears to converge as μ grows. The same behavior is observed when we compare the decentralized detection model where each sensor can send two bits (instead of one) to the fusion center to the centralized detection system (Figure 3-6). As expected, by the simulation argument of Section 3.1, the ratio of the error exponent in centralized detection over the optimal error exponent in decentralized detection where each sensor can send two bits to the fusion center takes lower values than the ratio of the error exponent in centralized detection over the optimal error exponent in decentralized detection where each sensor can only send one bit.

The observed converging behavior motivates the following result:

Proposition 3.2. *The ratio of the error exponent of the centralized detection over*

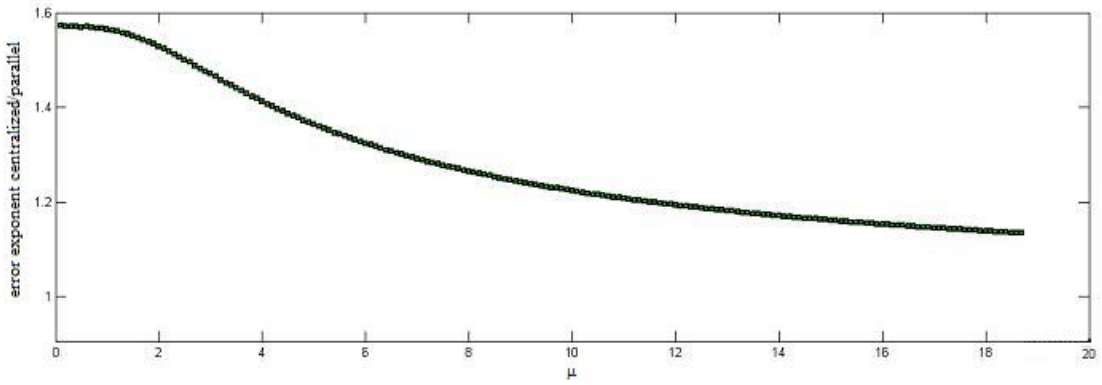

Figure 3-5: Plot of the ratio of K-L divergences as a function of the mean: error exponent in centralized detection over optimal error exponent in decentralized detection.

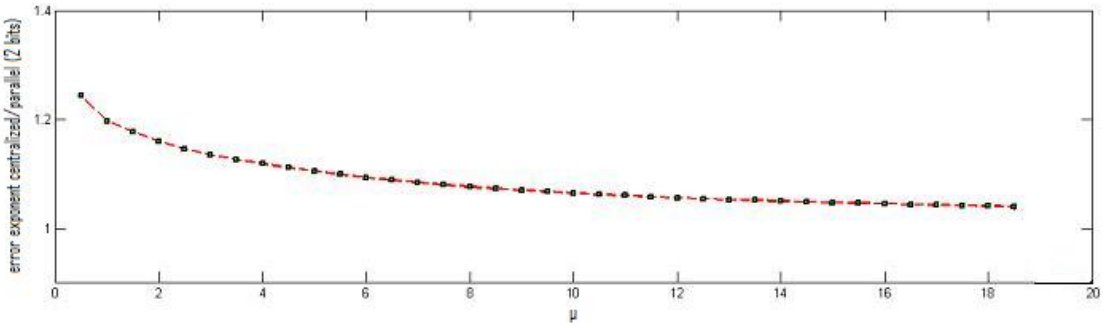

Figure 3-6: Plot of the ratio of K-L divergences as a function of the mean: error exponent in centralized detection over optimal error exponent in decentralized detection where each sensor can send two bits.

the optimal error exponent of a decentralized detection model with Gaussian detectors as defined above converges to 1 as $\mu \to \infty$*. That is,*

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{g_P^*} = 1,$$

where g_P^* denotes the optimal error exponent of the parallel configuration.

Proof. Fix $c > 0$ and let $\epsilon = e^{\frac{(-\mu+c)2\mu}{\sigma^2}}$ so that $x^t_{\epsilon,\mu,\sigma} = -\mu + c$. Let γ_c be the associated decision rule. We denote the probability density function of a standard

normal random variable by $\phi(\,\cdot\,)$. Then

$$\lim_{\mu\to\infty}\frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-D(\mathcal{P}_0^{\gamma_c}||\mathcal{P}_1^{\gamma_c})} = \lim_{\mu\to\infty}\frac{-\frac{2\mu^2}{\sigma^2}}{-\left(\Phi(\frac{c}{\sigma})\log\frac{\Phi(\frac{c}{\sigma})}{\Phi(\frac{-2\mu+c}{\sigma})}+(1-\Phi(\frac{c}{\sigma}))\log\frac{1-\Phi(\frac{c}{\sigma})}{1-\Phi(\frac{-2\mu+c}{\sigma})}\right)}$$

$$\begin{aligned}
&= \lim_{\mu\to\infty}\frac{\frac{2\mu^2}{\sigma^2}}{\Phi(\frac{c}{\sigma})(\log\Phi(\frac{c}{\sigma})-\log\Phi(\frac{-2\mu+c}{\sigma}))+(1-\Phi(\frac{c}{\sigma}))(\log(1-\Phi(\frac{c}{\sigma}))-\log(1-\Phi(\frac{-2\mu+c}{\sigma})))}\\
&= \lim_{\mu\to\infty}\frac{\frac{4\mu}{\sigma^2}}{\Phi(\frac{c}{\sigma})(-1)\frac{1}{\Phi(\frac{-2\mu+c}{\sigma})}\phi(\frac{-2\mu+c}{\sigma})\frac{-2}{\sigma}+(1-\Phi(\frac{c}{\sigma}))(-1)\frac{1}{1-\Phi(\frac{-2\mu+c}{\sigma})}(-1)\phi(\frac{-2\mu+c}{\sigma})\frac{-2}{\sigma}} \qquad (3.3)\\
&= \lim_{\mu\to\infty}\frac{\frac{4\mu}{\sigma^2}}{\frac{2\Phi(\frac{c}{\sigma})\phi(\frac{-2\mu+c}{\sigma})}{\sigma\Phi(\frac{-2\mu+c}{\sigma})}-\frac{2(1-\Phi(\frac{c}{\sigma}))\phi(\frac{-2\mu+c}{\sigma})}{\sigma(1-\Phi(\frac{-2\mu+c}{\sigma}))}}\\
&= \lim_{\mu\to\infty}\frac{\frac{4\mu}{\sigma^2}}{\frac{2}{\sigma}\phi(\frac{-2\mu+c}{\sigma})(\frac{\Phi(\frac{c}{\sigma})}{\Phi(\frac{-2\mu+c}{\sigma})}-\frac{1-\Phi(\frac{c}{\sigma})}{1-\Phi(\frac{-2\mu+c}{\sigma})})}\\
&= \lim_{\mu\to\infty}\frac{\frac{4\mu}{\sigma^2}}{\frac{2}{\sigma}\phi(\frac{-2\mu+c}{\sigma})\frac{\Phi(\frac{c}{\sigma})-\Phi(\frac{-2\mu+c}{\sigma})}{\Phi(\frac{-2\mu+c}{\sigma})(1-\Phi(\frac{-2\mu+c}{\sigma}))}}\\
&= \lim_{\mu\to\infty}\left(\frac{\mu\Phi(\frac{-2\mu+c}{\sigma})}{\phi(\frac{-2\mu+c}{\sigma})}\frac{2}{\sigma\frac{\Phi(\frac{c}{\sigma})-\Phi(\frac{-2\mu+c}{\sigma})}{1-\Phi(\frac{-2\mu+c}{\sigma})}}\right), \qquad (3.4)
\end{aligned}$$

with Equation (3.3) following from l'Hôpital's rule. It is known that

$$x^{-1}-x^{-3}e^{-\frac{x^2}{2}} < \sqrt{2\pi}(1-\Phi(x)) < x^{-1}e^{-\frac{x^2}{2}}, \qquad x>0. \qquad (3.5)$$

Note that $\Phi(\frac{-2\mu+c}{\sigma}) = 1-\Phi(\frac{2\mu-c}{\sigma})$. For $\mu > \frac{c}{2}$, by Equation (3.5), we have:

$$\lim_{\mu\to\infty}\frac{\mu\Phi(\frac{-2\mu+c}{\sigma})}{\phi(\frac{-2\mu+c}{\sigma})} \le \lim_{\mu\to\infty}\frac{\mu\frac{1}{\sqrt{2\pi}}\frac{\sigma}{2\mu-c}e^{-\frac{(\frac{2\mu-c}{\sigma})^2}{2}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{-2\mu+c}{\sigma})^2}{2}}} = \frac{\sigma}{2} \qquad (3.6)$$

and

$$\lim_{\mu\to\infty}\frac{\mu\Phi(\frac{-2\mu+c}{\sigma})}{\phi(\frac{-2\mu+c}{\sigma})} \ge \lim_{\mu\to\infty}\frac{\mu\frac{1}{\sqrt{2\pi}}(\frac{\sigma}{2\mu-c}-\frac{\sigma}{(2\mu-c)^3})e^{-\frac{(\frac{2\mu-c}{\sigma})^2}{2}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{-2\mu+c}{\sigma})^2}{2}}} = \lim_{\mu\to\infty}\left(\frac{\mu\sigma}{2\mu-c}-\frac{\mu\sigma}{(2\mu-c)^3}\right) = \frac{\sigma}{2}. \qquad (3.7)$$

By Equations (3.6) and (3.7), Equation (3.4) becomes

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-D(\mathcal{P}_0^{\gamma_c}||\mathcal{P}_1^{\gamma_c})} = \frac{\sigma}{2}\frac{2}{\sigma\Phi(\frac{c}{\sigma})} = \frac{1}{\Phi(\frac{c}{\sigma})}.$$

Nevertheless, $\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^{\gamma}\|\mathcal{P}_1^{\gamma})} \leq \lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-D(\mathcal{P}_0^{\gamma_c}||\mathcal{P}_1^{\gamma_c})} = \frac{1}{\Phi(\frac{c}{\sigma})}$, and since c was is arbitrary, it follows that

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^{\gamma}\|\mathcal{P}_1^{\gamma})} \leq \lim_{c\to\infty} \frac{1}{\Phi(\frac{c}{\sigma})} = 1. \tag{3.8}$$

Because $-D(\mathcal{P}_0||\mathcal{P}_1) \leq -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^{\gamma}\|\mathcal{P}_1^{\gamma}) < 0$,

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^{\gamma}\|\mathcal{P}_1^{\gamma})} \geq 1. \tag{3.9}$$

By Equations (3.8) and (3.9), $\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^{\gamma}\|\mathcal{P}_1^{\gamma})} = 1$ and the proof of the proposition is complete. □

A consequence of Proposition 3.2 is that the ratio of the error exponent of the centralized scheme over the optimal error exponents to the left of the error exponent of the parallel configuration on the real axes of Section 3.1 also converges to 1 as the measurement model becomes more informative, i.e. as the means of measurement distributions $\mathcal{P}_0$ and $\mathcal{P}_1$ grow apart, making the SNR higher. Specifically,

Proposition 3.3. *Consider the measurement model introduced above. Then*

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{g_F^*} = 1$$

and

$$\lim_{\mu\to\infty} \frac{-D(\mathcal{P}_0||\mathcal{P}_1)}{g_{DC}^*} = 1,$$

where g_F^* and g_{DC}^* denote the optimal error exponent of the simple feedback configuration and the "daisy chain", respectively.

The corollary that follows from Propositions 3.2 and 3.3 gives us the motivation for the analysis of Chapter 4 as well as future work proposed in Chapter 6. Feeding

the preliminary decision to sensors before making the final decision bears no value for performance improvement in the regime of infinitely many sensors and infinite SNR; in fact, having N sensors, each communicating 1 bit to the fusion center, suffices asymptotically as SNR grows higher.

Corollary 3.1. *Consider the measurement model introduced above. Then*

$$\lim_{\mu\to\infty} \frac{g_F^*}{g_P^*} = 1$$

and

$$\lim_{\mu\to\infty} \frac{g_{DC}^*}{g_P^*} = 1,$$

where g_P^*, g_F^*, and g_{DC}^* denote the optimal error exponent of the parallel configuration, the simple feedback configuration, and the "daisy chain", respectively.

The question that arises naturally is whether there is some converging behavior in the position of the optimizing threshold (the threshold $x_{\epsilon,\mu,\sigma}^t$ that minimizes the negated K-L divergence between $\mathcal{P}_0^\gamma$ and $\mathcal{P}_1^\gamma$), which we denote $x_{\mu,\sigma}^{t*}$. Fixing again $\sigma = 1$ for the purpose of running numerical simulations, Figures 3-7 and 3-8 suggest that $x_{\mu,\sigma}^{t*}$ is to first order equal to $-\mu$. In particular, we have the the following conjecture, which rests unproven:

Conjecture 3.1. $\lim_{\mu\to\infty} \frac{x_{\mu,\sigma}^{t*} - (-\mu)}{\mu - (-\mu)} = 0.$

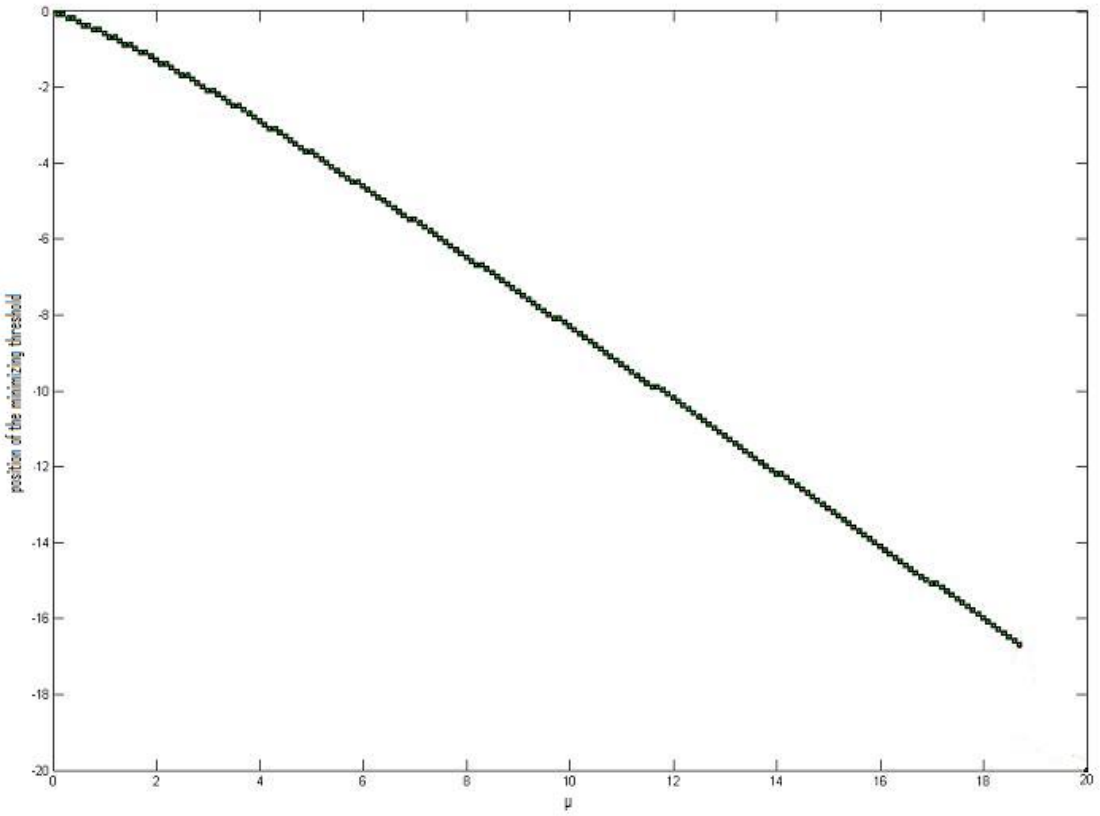

Figure 3-7: Plot of the position of the minimizing threshold $x_{\mu,1}^{t*}$ as a function of μ.

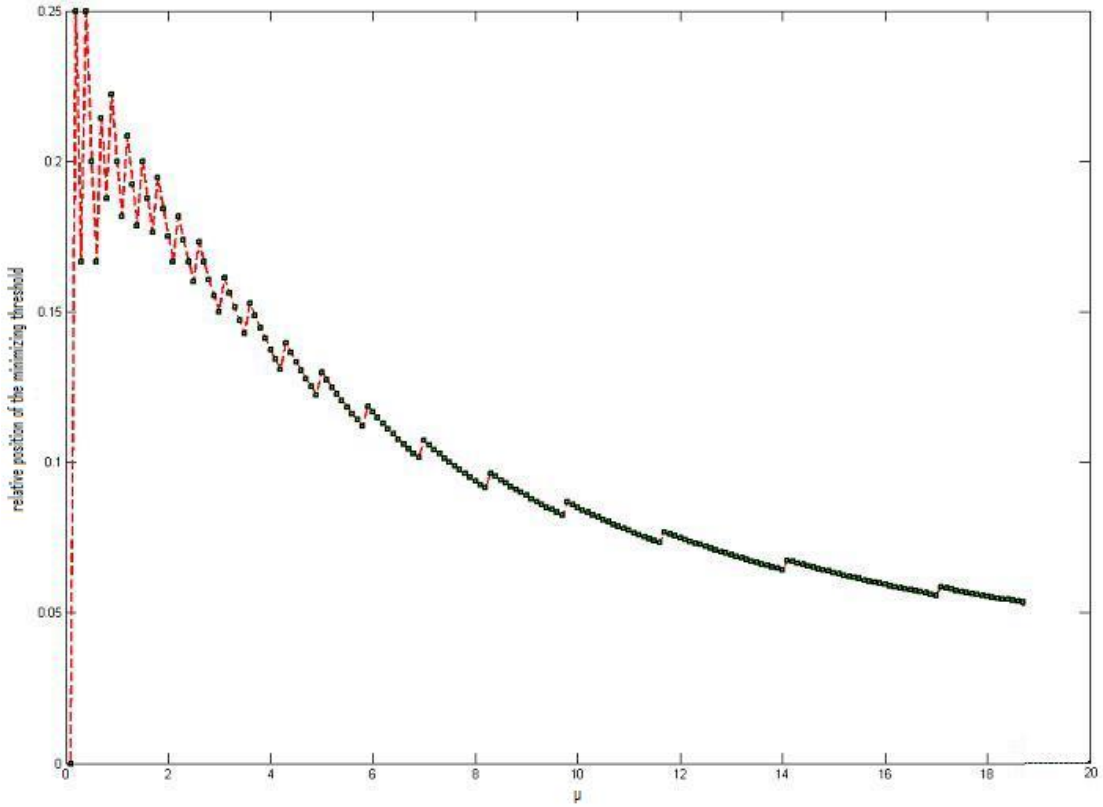

Figure 3-8: Plot of the relative position of the minimizing threshold $x_{\mu,1}^{t*}$ with respect to $-\mu$ and $+\mu$ as a function of μ.

Chapter 4

Neyman-Pearson Detection in the "Daisy Chain"

In this chapter we prove the result that under the Neyman-Pearson formulation, the performance of the "daisy chain" configuration (Subsection 2.1.4) is asymptotically equal to the performance of the parallel configuration with twice as many sensors as the classical scheme (Subsection 2.1.2), in the optimal error exponent sense. We first show that the "daisy chain" cannot be worse than the parallel configuration with twice as many sensors as the classical scheme; we then prove that it cannot perform better. We conclude that the value of feeding the preliminary decision to a second set of sensors is asymptotically negligible. In the last section, we prove that there is no loss of optimality asymptotically in the "daisy chain" if all decision rules at each stage are constrained to be equal. Throughout the chapter, we use g_{DC}^* to denote the optimal error exponent of the "daisy chain", and g_{2N}^* to denote the optimal error exponent of the parallel configuration with $2N$ sensors.

4.1 Neyman-Pearson asymptotics for the "daisy chain"

In Subsection 2.2.2 the asymptotics of the parallel configuration with N sensors were defined in a way that makes the extension to the other configurations of interest clear. Nevertheless, we define here the asymptotics specifically for the "daisy chain", for the sake of completeness.

Let $M = 2$, let Γ be the set of all measurable functions from $\mathcal{X}$ into $\mathcal{T}_1$, and let Δ be the set of all measurable functions from $\mathcal{X} \times \{0,1\}$ into $\mathcal{T}_2$. Given that the decision of the first stage is $Y_0 = y_0$, let $\Delta_{Y_0=y_0}$ be the set of all measurable functions from $\mathcal{X} \times \{y_0\}$ into $\mathcal{T}_2$. Let Γ_0 be the set of all candidate decision rules γ_0, Δ_0 be the set of all candidate decision rules δ_0. For any given choice of decision rules $(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) \in \Gamma_0 \times \Gamma^N \times \Delta_0 \times \Delta^N$, we define the Type I and Type II error probabilities to be, respectively:

$$J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) = \mathcal{P}_0(U_0 = 1) = \\ \mathcal{P}_0(\delta_0(\delta_1(X_{N+1}, \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))), \ldots, \delta_N(X_{2N}, \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))), \\ \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))) = 1), \tag{4.1}$$

$$J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) = \mathcal{P}_1(U_0 = 0) = \\ \mathcal{P}_1(\delta_0(\delta_1(X_{N+1}, \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))), \ldots, \delta_N(X_{2N}, \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))), \\ \gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N))) = 0). \tag{4.2}$$

(Remember that U_0 is the decision of the second fusion center). We require that $J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N)$ be no more than a given $\alpha \in (0,1)$ and we are interested in minimizing $J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N)$ over all $\gamma_0, \gamma_1,$ $\ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying

$$J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) \leq \alpha. \tag{4.3}$$

We define

$$r_N(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) = \frac{\log J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N)}{N}. \tag{4.4}$$

Let

$$R_N = \inf_{(\gamma_0,\ldots,\gamma_N,\delta_0,\ldots,\delta_N)\in\Gamma_0\times\Gamma^N\times\Delta_0\times\Delta^N s.t. J_N^I(\gamma_0,\ldots,\gamma_N,\delta_0,\ldots,\delta_N)\leq\alpha} r_N(\gamma_0, \ldots, \gamma_N, \delta_0, \ldots, \delta_N). \tag{4.5}$$

We focus on the optimal error exponent, $\lim_{N\to\infty} R_N$, which we denote g_{DC}^*.

Suppose that X_i is a sensor observation, $1 \leq i \leq N$; for any $\gamma \in \Gamma$, let the distribution of $\gamma(X_i)$ under hypothesis H_j be $\mathcal{P}_j^\gamma$. Suppose that X_i is a sensor observation, $N+1 \leq i \leq 2N$, and Y_0 is the decision of the first stage; for any $\delta \in \Delta$, let the distribution of $\delta(X_i, Y_0)$ under hypothesis H_j be $\mathcal{P}_j^\delta$. If it is given that the decision of the first stage is $Y_0 = y_0$, then for any $\delta_{Y_0=y_0} \in \Delta_{Y_0=y_0}$, let the distribution of $\delta(X_i, y_0)$ under hypothesis H_j be $\mathcal{P}_j^{\delta_{Y_0=y_0}}$. Since $\Gamma \subset \Delta$, it follows that $-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) \geq -\sup_{\delta\in\Delta} D(\mathcal{P}_0^\delta \| \mathcal{P}_1^\delta)$. Nevertheless, it is the case that

$$-\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) = -\sup_{\delta_{Y_0=y_0}\in\Delta_{Y_0=y_0}} D(\mathcal{P}_0^{\delta_{Y_0=y_0}} \| \mathcal{P}_1^{\delta_{Y_0=y_0}}). \tag{4.6}$$

4.2 The "daisy chain" is not worse than $2N$ sensors in parallel

Lemma 4.1. $g_{DC}^* \leq g_{2N}^*$

Proof. The "daisy chain" can simulate the communication of sensors in the tree of Figure 4-1, which by Theorem 3.1(ii) of [10] (notice that the parameter z in [10] is equal to 1) is asymptotically as good as the parallel configuration with $2N$ sensors. It follows that

$$g_{DC}^* \leq g_{2N}^* = 2g_P^* = -2\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma). \tag{4.7}$$

□

4.3 The "daisy chain" is not better than $2N$ sensors in parallel

Lemma 4.2. $g^*_{DC} \geq g^*_{2N}$

Proof. The tree in Figure 4-2 showcases the following probabilities:

$$\begin{aligned}
\beta_2^N &= \mathcal{P}_0(\gamma_0(\gamma_1(X_1), \ldots, \gamma_N(X_N)) = 0) = \mathcal{P}_0(Y_0 = 0) \\
\beta_1^N &= \mathcal{P}_0(Y_0 = 1) \\
\beta_4^N &= \mathcal{P}_1(Y_0 = 0) \\
\beta_3^N &= \mathcal{P}_1(Y_0 = 1) \\
\alpha_2'^N &= \mathcal{P}_0(\delta_0(\delta_1(X_{N+1}, Y_0), \ldots, \delta_N(X_{2N}, Y_0), Y_0) = 0 | Y_0 = 0) = \mathcal{P}_0(U_0 = 0 | Y_0 = 0) \\
\alpha_2^N &= \mathcal{P}_0(U_0 = 1 | Y_0 = 0) \\
\alpha_1'^N &= \mathcal{P}_0(U_0 = 0 | Y_0 = 1) \\
\alpha_1^N &= \mathcal{P}_0(U_0 = 1 | Y_0 = 1) \\
\alpha_4^N &= \mathcal{P}_1(U_0 = 0 | Y_0 = 0) \\
\alpha_4'^N &= \mathcal{P}_1(U_0 = 1 | Y_0 = 0) \\
\alpha_3^N &= \mathcal{P}_1(U_0 = 0 | Y_0 = 1) \\
\alpha_3'^N &= \mathcal{P}_1(U_0 = 1 | Y_0 = 1).
\end{aligned}$$

It becomes clear that the Type I and Type II error probabilities are given by

$$J_N^I(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) = \mathcal{P}_0(U_0 = 1) = \beta_2^N \alpha_2^N + \beta_1^N \alpha_1^N \tag{4.8}$$

and

$$J_N^{II}(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N) = \mathcal{P}_1(U_0 = 0) = \beta_4^N \alpha_4^N + \beta_3^N \alpha_3^N \tag{4.9}$$

respectively. To show that $g^*_{DC} \geq g^*_{2N}$, it suffices to show the following:

Proposition 4.1. *For any* $\epsilon \in (0, \frac{1}{2})$,

$$\liminf_{N\to\infty} \inf_{(\gamma_0,\gamma_1,\ldots,\gamma_N,\delta_0,\delta_1,\ldots,\delta_N)\in\Gamma_0\times\Gamma^N\times\Delta_0\times\Delta^N s.t. \beta_2^N\alpha_2^N+\beta_1^N\alpha_1^N\leq\epsilon} \frac{\log(\beta_4^N\alpha_4^N+\beta_3^N\alpha_3^N)}{N} \geq -2\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma\|\mathcal{P}_1^\gamma) = g_{2N}^*. \tag{4.10}$$

Proof. By the Stein lemma (see [5]) and the "Conditional" Stein lemma (presented in the Appendix), and using Equation (4.6), the following are true:

1. For any $\epsilon_1 \in (0,1)$ and for any $\gamma_0, \gamma_1, \ldots, \gamma_N$ satisfying $\beta_1^N \leq \epsilon_1$, we have

$$\frac{\log \beta_4^N}{N} \geq -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma\|\mathcal{P}_1^\gamma) + f_1(N,\epsilon_1),$$

   where f_1 is a function with the property $\lim_{N\to\infty} f_1(N,\epsilon_1) = 0$, for all $\epsilon_1 \in (0,1)$, and which does not depend on $\gamma_0, \gamma_1, \ldots, \gamma_N$. While this result does not follow directly from the usual formulation of the Stein lemma, it may be proven by changing the proof of the Stein lemma in [3] according to the small variation described in the proof of Theorem 2 in [12], and using Assumptions 2.1 and 2.2.

2. For any $\epsilon_2 \in (0,1)$ and for any $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying $\alpha_2^N \leq \epsilon_2$, and if $Y_0 = 0$, we have

$$\frac{\log \alpha_4^N}{N} \geq -\sup_{\delta_{Y_0=0}\in\Delta_{Y_0=0}} D(\mathcal{P}_0^{\delta_{Y_0=0}}\|\mathcal{P}_1^{\delta_{Y_0=0}}) + f_2(N,\epsilon_2) = -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma\|\mathcal{P}_1^\gamma) + f_2(N,\epsilon_2),$$

   where f_2 is a function with the property $\lim_{N\to\infty} f_2(N,\epsilon_2) = 0$, for all $\epsilon_2 \in (0,1)$, and which does not depend on $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$. This result follows by the small variation of the proof of the ("Conditional") Stein lemma discussed in item 1, and as a consequence of Assumptions 2.1 and 2.2.

3. For any $\epsilon_3 \in (0,1)$ and for any $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying $\alpha_1^N \leq \epsilon_3$, and if $Y_0 = 1$, we have

$$\frac{\log \alpha_3^N}{N} \geq -\sup_{\delta_{Y_0=1}\in\Delta_{Y_0=1}} D(\mathcal{P}_0^{\delta_{Y_0=1}}\|\mathcal{P}_1^{\delta_{Y_0=1}}) + f_3(N,\epsilon_3) = -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma\|\mathcal{P}_1^\gamma) + f_3(N,\epsilon_3),$$

where f_3 is a function with the property $\lim_{N\to\infty} f_3(N,\epsilon_3) = 0$, for all $\epsilon_3 \in (0,1)$, and which does not depend on $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$.

Additionally, by reversing the semantics of the decision of the first stage:

4. For any $\epsilon_4 \in (0,1)$ and for any $\gamma_0, \gamma_1, \ldots, \gamma_N$ satisfying $\beta_2^N \leq \epsilon_4$, we have

$$\frac{\log \beta_3^N}{N} \geq -\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) + f_4(N,\epsilon_4),$$

where f_4 is a function with the property $\lim_{N\to\infty} f_4(N,\epsilon_4) = 0$, for all $\epsilon_4 \in (0,1)$, and which does not depend on $\gamma_0, \gamma_1, \ldots, \gamma_N$.

Fix N. Since $\beta_2^N + \beta_1^N = 1$, $\beta_2^N > \frac{1}{2}$ and $\beta_1^N > \frac{1}{2}$ cannot hold simultaneously. We therefore have the following cases:

1. $\beta_2^N \leq \frac{1}{2}$. In this case, the inequality $\beta_2^N\alpha_2^N + \beta_1^N\alpha_1^N \leq \epsilon$ implies that there exists a constant $\epsilon_5 \in (0,1)$ such that $\alpha_1^N \leq \epsilon_5$. By items 3 and 4 above, it follows that for any $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying $\beta_2^N\alpha_2^N + \beta_1^N\alpha_1^N \leq \epsilon$, we have

$$\frac{\log(\beta_3^N\alpha_3^N)}{N} \geq -2\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) + f_3(N,\epsilon_5) + f_4(N,\frac{1}{2}).$$

2. $\beta_1^N \leq \frac{1}{2}$. In this case, the inequality $\beta_2^N\alpha_2^N + \beta_1^N\alpha_1^N \leq \epsilon$ implies that there exists a constant $\epsilon_6 \in (0,1)$ such that $\alpha_2^N \leq \epsilon_6$. By items 1 and 2 above, it follows that for any $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying $\beta_2^N\alpha_2^N + \beta_1^N\alpha_1^N \leq \epsilon$, we have

$$\frac{\log(\beta_4^N\alpha_4^N)}{N} \geq -2\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) + f_1(N,\frac{1}{2}) + f_2(N,\epsilon_6).$$

It follows that for any $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ satisfying $\beta_2^N\alpha_2^N + \beta_1^N\alpha_1^N \leq \epsilon$, we have

$$\frac{\log(\beta_4^N\alpha_4^N + \beta_3^N\alpha_3^N)}{N} \geq -2\sup_{\gamma\in\Gamma} D(\mathcal{P}_0^\gamma \| \mathcal{P}_1^\gamma) + h(N,\epsilon_5,\epsilon_6), \tag{4.11}$$

where h is a function with the property $\lim_{N\to\infty} h(N,\epsilon_5,\epsilon_6) = 0$, for all ϵ_5, ϵ_6, and which does not depend on $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$.

Equation (4.10) then follows by taking the infimum of both sides of Equation (4.11) over all $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ and letting N tend to infinity. □

Theorem 4.1. $g^*_{DC} = g^*_{2N}$

Proof. The theorem follows directly from Lemmata 4.1 and 4.2. □

4.4 No loss of optimality with equal decision rules in each stage

Theorem 4.2. $g^*_{DC} = g^*_{2N}$ *if in definition* (4.5) *we impose the additional constraint* $\gamma_1 = \ldots = \gamma_N = \delta_1 = \ldots = \delta_N$.

Proof. Since $g^*_{DC} \geq g^*_{2N}$ (by Lemma 4.2), and the optimal error exponent for the "daisy chain" under the constraint that $\gamma_1 = \ldots = \gamma_N = \delta_1 = \ldots = \delta_N$ cannot be better than the optimal error exponent for the "daisy chain" without the constraint, it follows that $g^*_{DC} \geq g^*_{2N}$ if in definition (4.5) we impose the additional constraint $\gamma_1 = \ldots = \gamma_N = \delta_1 = \ldots = \delta_N$.

Furthermore, the "simulation" argument can be used to claim that the optimal error exponent for the "daisy chain" under the constraint that $\gamma_1 = \ldots = \gamma_N = \delta_1 = \ldots = \delta_N$ is at least as good as the optimal error exponent for the tree configuration of Figure 4-1 where there is no communication of the decision of the first fusion center to sensors $N+1, \ldots, 2N$ under the constraint that every leaf (i.e., every peripheral sensor) uses the same decision rule. By Proposition 3.3 of [10], and specifically Lemmas 3.2, 3.3, and the proof of Lemma 3.4 therein, the latter optimal error exponent under the constraint that every sensor uses the same decision rule matches the optimal error exponent for the parallel configuration with $2N$ sensors. It follows that $g^*_{DC} \leq g^*_{2N}$ if in definition (4.5) we impose the additional constraint $\gamma_1 = \ldots = \gamma_N = \delta_1 = \ldots = \delta_N$.

The theorem follows. □

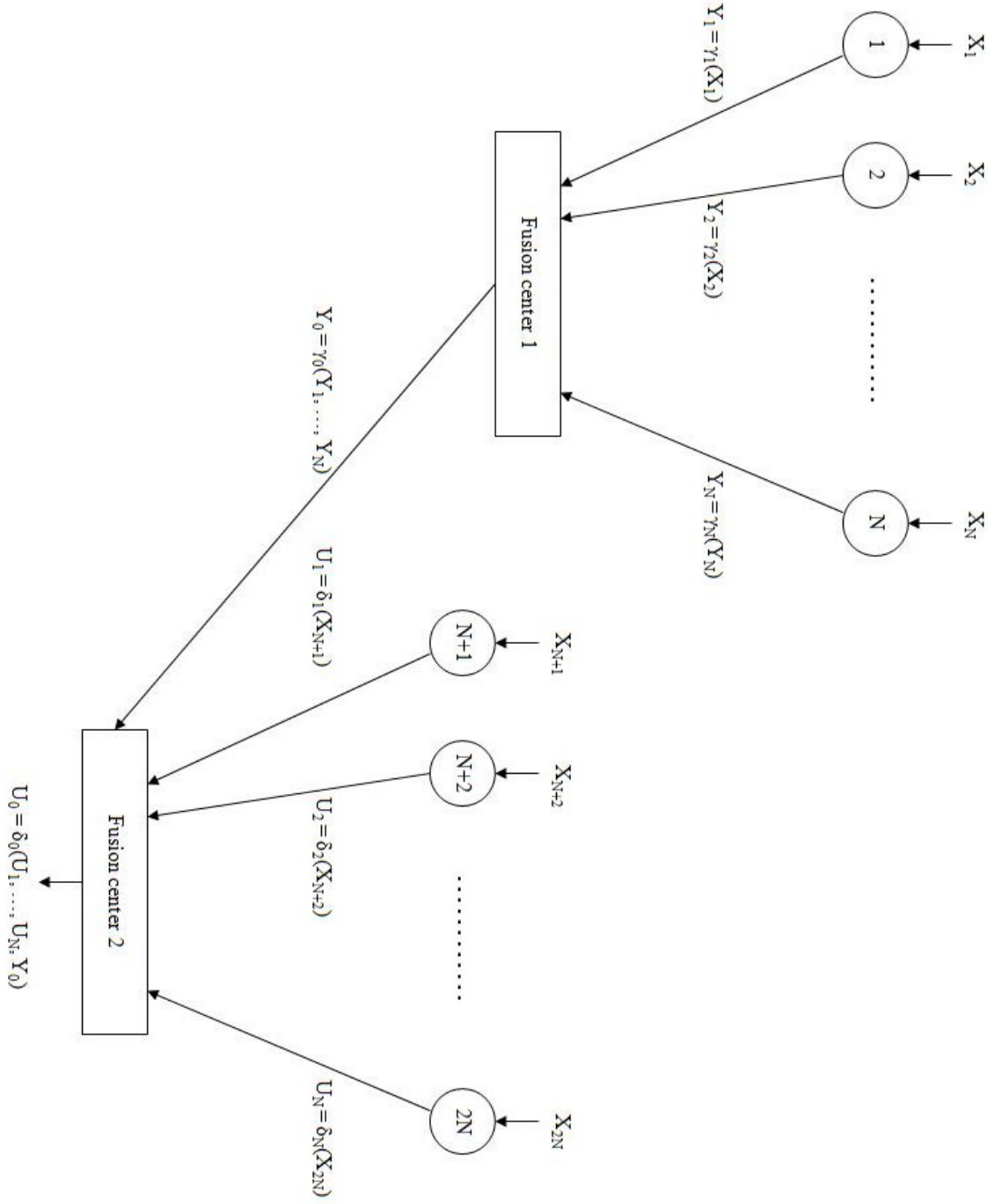

Figure 4-1: The "daisy chain" configuration without a feedback link from the first fusion center to the second set of sensors.

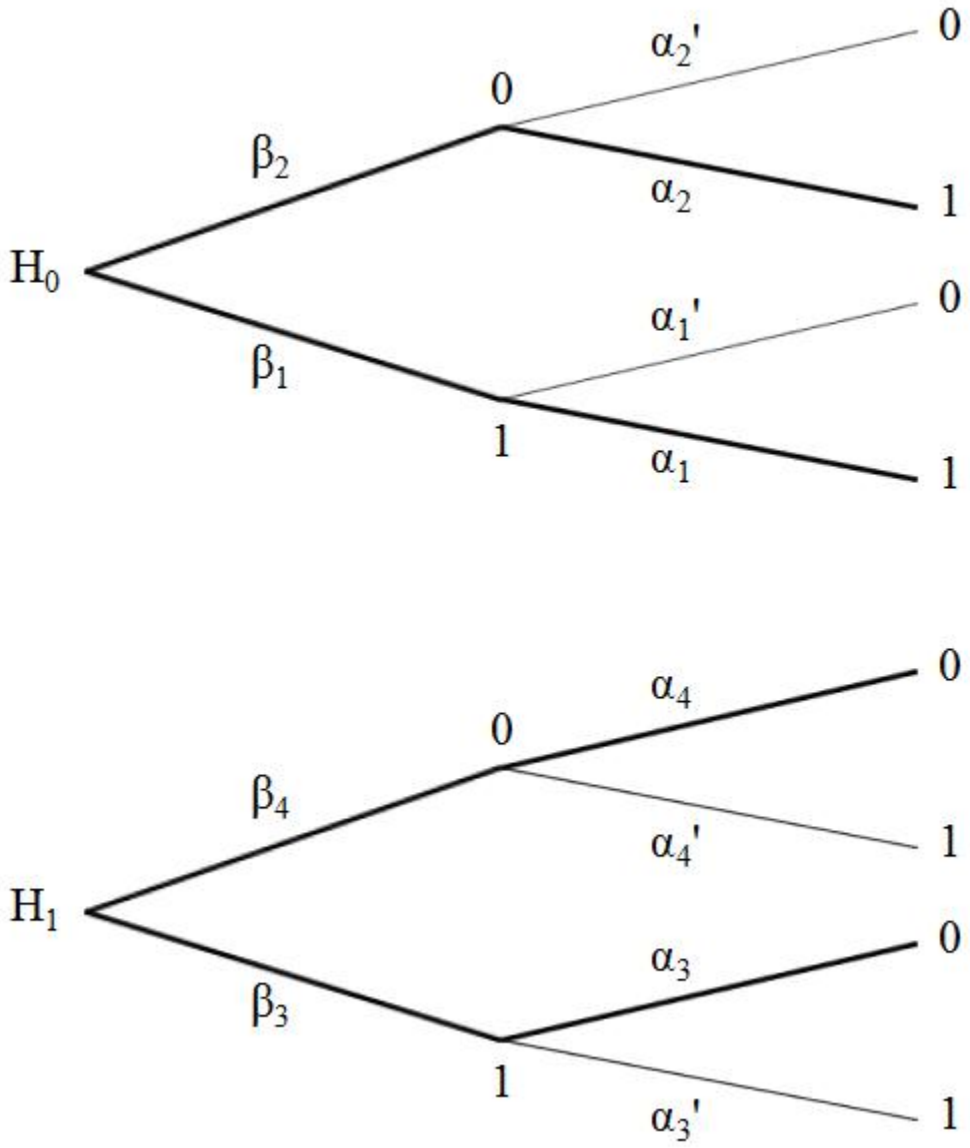

Figure 4-2: The two-stage decision tree for the "daisy chain".

Chapter 5

Bayesian Detection in the "Daisy Chain"

In this chapter, we study the Bayesian detection problem for the "daisy chain" configuration. [12] proves that for the parallel configuration, it is asymptotically optimal to let all sensors use the same decision rule in deciding what to transmit. In particular, it is asymptotically optimal to have all sensors perform identical likelihood ratio tests, using the same threshold. First we extend this result to the case of exponentially skewed priors; we then make the connection between the result for exponentially skewed priors and Bayesian detection in the "daisy chain". Specifically, we prove that under the mild assumption that the Type I and Type II error probabilities of the first stage decay exponentially with the number of sensors, it is asymptotically optimal to have all sensors in the second stage perform identical likelihood ratio tests. Finally, we prove that it is asymptotically optimal to have all sensors in the first stage perform identical likelihood ratio tests as well. In all cases, optimality is meant in terms of the overall system's optimal error exponent.

5.1 Bayesian asymptotics for the "daisy chain"

In Subsection 2.2.1, the asymptotics of the parallel configuration with N sensors were defined in a way that makes the extension to the other configurations of interest clear.

Nevertheless, we hereby formulate the asymptotics specifically for the "daisy chain", for the sake of completeness.

The optimal decision for the second fusion center δ_0, having fixed the decision rules $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$ of the sensors and the first fusion center, is determined by the maximum a posteriori (MAP) probability rule. Thus we will be concerned only with optimization with respect to $\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_0, \delta_1, \ldots, \delta_N$. Any set of decision rules $\gamma_1, \ldots, \gamma_N$ will be denoted by γ^N, and any set of rules $\delta_1, \ldots, \delta_N$ will be denoted by δ^N. Let Γ be the set of all measurable functions from $\mathcal{X}$ into $\mathcal{T}_1$, and Γ^N be the Cartesian product of Γ with itself N times. Let Δ be the set of all measurable functions from $\mathcal{X} \times \{0, \ldots, M-1\}$ into $\mathcal{T}_2$, and Δ^N be the Cartesian product of Δ with itself N times. Let Γ_0 be the set of all candidate decision rules γ_0. For any given choice of decision rules $(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_1, \ldots, \delta_N) \in \Gamma_0 \times \Gamma^N \times \Delta^N$, we define $J_N(\gamma_0, \gamma^N, \delta^N)$ to be the probability of an erroneous decision by the fusion center. For any given N and choice of decision rules $\gamma_0, \gamma^N, \delta^N$, we define the exponent of the error probability to be

$$r_N(\gamma_0, \gamma^N, \delta^N) = \frac{\log J_N(\gamma_0, \gamma^N, \delta^N)}{N}. \tag{5.1}$$

Let

$$R_N = \inf_{(\gamma_0, \gamma^N, \delta^N) \in \Gamma_0 \times \Gamma^N \times \Delta^N} r_N(\gamma_0, \gamma^N, \delta^N). \tag{5.2}$$

We focus on the optimal error exponent, $\lim_{N\to\infty} R_N$, which we denote g^*_{DC}.

5.2 Exponentially skewed priors in the parallel configuration with N sensors

We return to our definitions in the formulation of the Bayesian asymptotic regime for the parallel configuration given in 2.2.1 (as opposed to the definitions introduced in 5.1 for the "daisy chain"). We restrict to $M = 2$ and we let the priors $\pi_0^N = \Pr(H_0)$ and $\pi_1^N = \Pr(H_1)$ change with the number of sensors N. We assume that $\lim_{N\to\infty} \frac{\log \pi_0^N}{N} =$

$-\beta$ for some $\beta > 0$. As in [12], we let Γ_0^N be the set of all $\gamma^N \in \Gamma^N$ with the property that all elements in the set $\{\gamma_1, \ldots, \gamma_N\}$ are identical. Let $Q_N = \inf_{\gamma^N \in \Gamma_0^N} r_N(\gamma^N)$ be the optimal exponent when we restrict to sets of decision rules in Γ_0^N.

Theorem 5.1. *Subject to Assumption 5.1 below,* $\lim_{N\to\infty}(Q_N - R_N) = 0$.

Proof. The structure of the proof, and sometimes the wording, is intentionally left the same as in the proof of Theorem 1 of [12] so that the reader who is familiar with [12] can follow readily and pinpoint the modifications. Although we could only present the parts of the proof that are modified, we repeat the entire proof for the sake of completeness, so that the reader who is not familiar with [12] can follow as well.

Having fixed some $\gamma \in \Gamma$, we can consider the mapping from the true hypothesis H_i to the decision of a sensor using decision rule γ as a noisy communication channel defined by the probabilities $p_i^\gamma(d) = \mathcal{P}_i(\gamma(X) = d)$ where X is the observation of the sensor. As in [8], we quantify the ability of such a channel to discriminate between hypotheses H_i and H_j $(i \neq j)$ by defining the function $\mu_{ij}(\gamma, s)$, $s \in [0, 1]$:

$$\mu_{ij}(\gamma, s) = \log \sum_{d=1}^{D} (p_i^\gamma(d))^{1-s} (p_j^\gamma)^s. \tag{5.3}$$

The convention $0^0 = 0$ is used. If $\mu_{ij}(\gamma, s)$ is not infinite, then it is infinitely differentiable as a function of s, and its derivatives are continuous on $[0, 1]$, provided we define the derivative at an endpoint as the limit of the derivative when we approach the endpoint from the interior.

For any fixed γ, the function $\mu_{ij}(\gamma, s)$ is equal to $\log \mathbf{E}\left[e^{sX}\right]$, where X is the log-likelihood ratio of the distributions associated with $p_j^\gamma(\,\cdot\,)$ and $p_i^\gamma(\,\cdot\,)$. A suitable minimization involving the moment generating function of a random variable X yields tight bounds on the probability of large deviations of X from its mean. Since here X is the log-likelihood ratio, the minimization leads to tight bounds on the probability of error.

We have the following lemma, whose wording is intentionally left the same as in Lemma 1 of [12]. Notice that the measurements can be assumed to take values in

$\{1, \ldots, D\}$, as opposed to taking values generally in an alphabet of cardinality D, without loss of generality.

Lemma 5.1. *Let there be two hypotheses H' and H''. Let $Z_1, \ldots, Z_N$ be measurements taking values in a finite set $\{1, \ldots, D\}$, which are conditionally independent given the true hypothesis, and assume that the conditional distribution of X_i, when H is true, is given by $p_H^i(d) = \Pr(X_i = d|H)$. Let $\mu(i,s) = \log \sum_{d=1}^{D} (p_{H'}^i(d))^{1-s} (p_{H''}^i(d))^s$ and $\mu(s) = \sum_{i=1}^{N} \mu(i,s)$. Assume that $\mu(i,s), \mu'(i,s), \mu''(i,s)$ exist and are finite for $s \in [0,1]$, where primes on μ stand for differentiation with respect to s. Let $\hat{s}$ minimize $\mu(s) - \beta N s$ over $s \in [0,1]$. Then,*

(a) There exists a decision rule for deciding between H' and H'', on the basis of the measurements $Z_1, \ldots, Z_N$, and a constant $c_a > 0$, such that for every $\epsilon > 0$, there exists N_ϵ^a such that for every $N \geq N_\epsilon^a$,

$$\pi_1^N \Pr(\textit{decide } H' | H'' \textit{ is true}) + \pi_0^N \Pr(\textit{decide } H'' | H' \textit{ is true}) \leq e^{c_a \epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s}}.$$

(b) There exists a constant $c_b > 0$ such that for every ϵ with $0 < \epsilon < \beta$, there exists N_ϵ^b such that for every $N \geq N_\epsilon^b$, for any decision rule for deciding between H' and H'', on the basis of the measurements $Z_1, \ldots, Z_N$,

$$\begin{aligned} &\pi_1^N \Pr(\textit{decide } H' | H'' \textit{ is true}) + \pi_0^N \Pr(\textit{decide } H'' | H' \textit{ is true}) \\ &\qquad \geq e^{-c_b \epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s} - \sqrt{2\mu''(\hat{s})}}. \end{aligned}$$

Proof. Because $\lim_{N\to\infty} \frac{\log \pi_0^N}{N} = -\beta$ and $\beta > 0$, it follows that for every $\epsilon > 0$ there exists $N_\epsilon^{'}$ such that if $N \geq N_\epsilon^{'}$, then

$$e^{-\beta N - \epsilon N} \leq \pi_0^N \leq e^{-\beta N + \epsilon N} \tag{5.4}$$

and

$$1 - e^{-\beta N + \epsilon N} \leq \pi_1^N \leq 1 - e^{-\beta N - \epsilon N}. \tag{5.5}$$

We will prove each part of the lemma separately.

Part (a) follows from [8], Theorem 5 (Equations (3.13) and (3.14)). Indeed, fixing $\epsilon > 0$:

$$\begin{aligned}
&\pi_1^N \Pr(\text{decide } H' | H'' \text{ is true}) + \pi_0^N \Pr(\text{decide } H'' | H' \text{ is true}) \\
&\qquad \leq 1 \cdot \Pr(\text{decide } H' | H'' \text{ is true}) + \pi_0^N \Pr(\text{decide } H'' | H' \text{ is true}) \\
&\qquad \leq \mathrm{e}^{\mu(s) - s\mu'(s)} + e^{-\beta N + \epsilon N} e^{\mu(s) + (1-s)\mu'(s)} \\
&\qquad = e^{\mu(s) - s\mu'(s)} + e^{\epsilon N} e^{\mu(s) - s\mu'(s) + \mu'(s) - \beta N}
\end{aligned} \tag{5.6}$$

for all $s \in (0,1)$ and $N \geq N_\epsilon^{'}$, with the second step following from [8], Equations (3.13) and (3.14). At $\hat{s}$ the bound becomes

$$e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s})} + e^{\epsilon N} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) + \mu'(\hat{s}) - \beta N}. \tag{5.7}$$

If $\hat{s} \in (0,1)$, then $\mu'(\hat{s}) = \beta N$ by definition of $\hat{s}$, and therefore

$$e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s})} + e^{\epsilon N} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) + \mu'(\hat{s}) - \beta N} = (1 + e^{\epsilon N}) e^{\mu(\hat{s}) - \beta N \hat{s}}.$$

Because there exists $N_\epsilon^{''}$ such that $(1 + e^{\epsilon N}) \leq e^{2\epsilon N}$ for $N \geq N_\epsilon^{''}$, the result follows by letting $c_a = 2$ and $N_\epsilon^a = \max(N_\epsilon^{'}, N_\epsilon^{''})$. If $\hat{s} = 0$, we may take the limit of (5.6), as $s \downarrow 0$. It can be easily checked that $\mu'(0) = \sum_{i=1}^{N} -D(p_{H'}^i || p_{H''}^i)$, where D denotes the Kullback-Leibler divergence of two distributions, a nonnegative quantity. The bound in (5.7) becomes $e^{\mu(0)} + e^{\epsilon N} e^{\mu(0) + \mu'(0) - \beta N}$, and we have

$$e^{\mu(0)} + e^{\epsilon N} e^{\mu(0) + \mu'(0) - \beta N} \leq e^{\mu(0)} + e^{\epsilon N} e^{\mu(0)} = e^{\mu(0)} (1 + e^{\epsilon N})$$

for $N \geq N_\epsilon^{'}$. Because there exists $N_\epsilon^{''}$ such that $(1 + e^{\epsilon N}) \leq e^{2\epsilon N}$ for $N \geq N_\epsilon^{''}$, the result follows by letting $c_a = 2$ and $N_\epsilon^a = \max(N_\epsilon^{'}, N_\epsilon^{''})$. The argument for $\hat{s} = 1$ flows analogously.

Part (b) follows from [8], Equations (3.40), (3.41), and (3.42), as we proceed to

show. For fixed $0 < \epsilon < \beta$, there exists $\hat{N}_\epsilon$ such that

$$\pi_1^N \geq 1 - e^{-\beta N + \epsilon N} \geq \frac{1}{2} \tag{5.8}$$

for $N \geq \hat{N}_\epsilon$. Defining $Q_s(\,\cdot\,)$ as in Equation (3.27) of [8], and $Region_1$, $Region_2$, and $Region_s$ as in Equations (3.30) and (3.33) of [8], we have

$$\begin{aligned}
&\pi_1^N \Pr(\text{decide } H' | H'' \text{ is true}) + \pi_0^N \Pr(\text{decide } H'' | H' \text{ is true}) \\
&\qquad \geq \tfrac{1}{2} \Pr(\text{decide } H' | H'' \text{ is true}) + \pi_0^N \Pr(\text{decide } H'' | H' \text{ is true}) \\
&\qquad \geq \tfrac{1}{2} e^{\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}} \textstyle\sum_{{Region_1}^c \cap Region_s} Q_s(\,\cdot\,) \\
&\qquad\qquad + e^{-\beta N - \epsilon N} e^{\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}} \textstyle\sum_{{Region_2}^c \cap Region_s} Q_s(\,\cdot\,) \\
&\qquad \geq \tfrac{1}{2} e^{\mu(s) - s\mu'(s) - \sqrt{2\mu''(s)}} \textstyle\sum_{{Region_1}^c \cap Region_s} Q_s(\,\cdot\,) \\
&\qquad\qquad + e^{-\beta N - \epsilon N} e^{\mu(s) + (1-s)\mu'(s) - \sqrt{2\mu''(s)}} \textstyle\sum_{{Region_2}^c \cap Region_s} Q_s(\,\cdot\,) \\
&\qquad = \tfrac{1}{2} e^{\mu(s) - s\mu'(s) - \sqrt{2\mu''(s)}} \textstyle\sum_{{Region_1}^c \cap Region_s} Q_s(\,\cdot\,) \\
&\qquad\qquad + e^{-\epsilon N} e^{\mu(s) - s\mu'(s) - \sqrt{2\mu''(s)} + \mu'(s) - \beta N} \textstyle\sum_{{Region_2}^c \cap Region_s} Q_s(\,\cdot\,)
\end{aligned}$$

for all $s \in (0,1)$ and $N \geq \hat{N}_\epsilon$, with the second step following from Equations (3.40) and (3.41) of [8]. At $\hat{s}$ the bound becomes

$$\begin{aligned}
&\tfrac{1}{2} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) - \sqrt{2\mu''(\hat{s})}} \textstyle\sum_{{Region_1}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,) \\
&\qquad + e^{-\epsilon N} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) - \sqrt{2\mu''(\hat{s})} + \mu'(\hat{s}) - \beta N} \textstyle\sum_{{Region_2}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,).
\end{aligned}$$

If $\hat{s} \in (0,1)$, then $\mu'(\hat{s}) = \beta N$ by definition of $\hat{s}$. Because there exists $\hat{\hat{N}}_\epsilon$ such that $\frac{1}{2} \geq e^{-\epsilon N}$ for all $N \geq \hat{\hat{N}}_\epsilon$, it follows that

$$\begin{aligned}
&\frac{1}{2} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) - \sqrt{2\mu''(\hat{s})}} \sum_{{Region_1}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,) \\
&\qquad + e^{-\epsilon N} e^{\mu(\hat{s}) - \hat{s}\mu'(\hat{s}) - \sqrt{2\mu''(\hat{s})} + \mu'(\hat{s}) - \beta N} \textstyle\sum_{{Region_2}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,) \\
&\qquad \geq e^{-\epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s} - \sqrt{2\mu''(\hat{s})}} \\
&\qquad\qquad \cdot \Big(\textstyle\sum_{{Region_1}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,) + \sum_{{Region_2}^c \cap Region_{\hat{s}}} Q_{\hat{s}}(\,\cdot\,) \Big) \\
&\qquad \geq \tfrac{1}{2} e^{-\epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s} - \sqrt{2\mu''(\hat{s})}}
\end{aligned}$$

for all $N \geq \max(\hat{N}_\epsilon, \hat{\hat{N}}_\epsilon)$, with the last step following from Equation (3.42) of [SGB]. Because there exists $\hat{\hat{\hat{N}}}_\epsilon$ such that $\frac{1}{2}e^{-\epsilon N} \geq e^{-2\epsilon N}$ for $N \geq \hat{\hat{\hat{N}}}_\epsilon$, the result follows by letting $c_b = 2$ and $N_\epsilon^b = \max(\hat{N}_\epsilon, \hat{\hat{N}}_\epsilon, \hat{\hat{\hat{N}}}_\epsilon)$. If $\hat{s} = 0$ or $\hat{s} = 1$, an argument similar to the one used in the proof of part (a) of this lemma applies. □

Assumption 5.1. *For $i \neq j$,*

(a) $|\mu_{ij}(\gamma, s)| < \infty$, for all $\gamma \in \Gamma$ and $s \in [0, 1]$.

(b) There exists a finite constant A such that $\left|\mu''_{ij}(\gamma, s)\right| \leq A$, for all $\gamma \in \Gamma$ and $s \in [0, 1]$.

As explained in [12], the restrictions imposed by Assumption 5.1 are of minor practical significance.

The proof of Theorem 1 in [12] can be employed almost as is to complete our proof. The proof in [12] assumes M distinct hypotheses, whereas we are only interested in the $M = 2$ case; we can thus give a simplified proof, without employing linear programming theory. We argue that the decision rules $\gamma_1, \ldots, \gamma_N$ should be chosen so as to minimize

$$\min_{s\in[0,1]} \left(\sum_{k=1}^{N} \mu_{01}(\gamma_k, s) - \beta N s \right). \tag{5.9}$$

We show that all γ_k's can be taken equal.

Let

$$\hat{\Lambda} = \min_{\gamma\in\Gamma} \min_{s\in[0,1]} \left(\mu_{01}(\gamma, s) - \beta s\right). \tag{5.10}$$

(To keep the proof simple, we assume that the minima in (5.10) are attained.)

Let us fix some ϵ with $0 < \epsilon < \beta$, some $N \geq \max(N_\epsilon^a, N_\epsilon^b)$, and some collection $\gamma^N \in \Gamma^N$ of decision rules. Part (b) of Lemma 5.1 yields

$$\begin{aligned} J_N(\gamma^N) &= \pi_1^N \Pr(\text{decide } H_0 | H_1 \text{ is true}) + \pi_0^N \Pr(\text{decide } H_1 | H_0 \text{ is true}) \\ &\geq e^{-c_b\epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s} - \sqrt{2\mu''(\hat{s})}} \\ &= e^{-c_b\epsilon N} e^{\sum_{k=1}^{N} \mu_{01}(\gamma_k, \hat{s}_{01}) - \beta N \hat{s}_{01} - \sqrt{2\sum_{k=1}^{N} \mu''_{01}(\gamma_k, \hat{s}_{01})}}, \end{aligned} \tag{5.11}$$

where $\hat{s}_{01}$ minimizes $\sum_{k=1}^{N} \mu_{01}(\gamma_k, s) - \beta N s$ over $s \in [0,1]$. By Assumption 5.1(b), and the definitions of $\hat{s}_{01}$ and $\hat{\Lambda}$,

$$\begin{aligned} J_N(\gamma^N) &\geq e^{-c_b \epsilon N} e^{\min_{s\in[0,1]} \left(\sum_{k=1}^{N} \mu_{01}(\gamma,s) - \beta N s \right) - \sqrt{2NA}} \\ &\geq e^{-c_b \epsilon N} e^{N\hat{\Lambda} - \sqrt{2NA}}. \end{aligned} \tag{5.12}$$

This shows that $R_N \geq -c_b \epsilon + \hat{\Lambda} - \sqrt{\frac{2A}{N}}$. We take the limit as $N \to \infty$ and use the fact that ϵ was arbitrary to obtain

$$\liminf_{N\to\infty} R_N \geq \hat{\Lambda}. \tag{5.13}$$

Let us fix some $\epsilon > 0$ and some $N \geq \max(N_\epsilon^a, N_\epsilon^b)$. Let $\hat{\gamma}$ be the solution to (5.10). We now define a collection γ^N of decision rules to be used by the N sensors: for each k, let $\gamma_k = \hat{\gamma}$.

We estimate the probability of error under this particular γ_N. Using Lemma 5.1(a), we have

$$\begin{aligned} J_N(\gamma^N) &= \pi_1^N \Pr(\text{decide } H_0 | H_1 \text{ is true}) + \pi_0^N \Pr(\text{decide } H_1 | H_0 \text{ is true}) \\ &\leq e^{c_a \epsilon N} e^{\mu(\hat{s}) - \beta N \hat{s}} \\ &= e^{c_a \epsilon N} e^{\sum_{k=1}^{N} \mu_{01}(\gamma_k, \hat{s}_{01}) - \beta N \hat{s}_{01}} \\ &= e^{c_a \epsilon N} e^{\min_{s\in[0,1]} \left(\sum_{k=1}^{N} \mu_{01}(\hat{\gamma},s) - \beta N s \right)} \\ &= e^{c_a \epsilon N} e^{\min_{s\in[0,1]} (N \mu_{01}(\hat{\gamma},s) - \beta N s)} \\ &= e^{c_a \epsilon N} e^{N\hat{\Lambda}}. \end{aligned} \tag{5.14}$$

Taking logarithms and dividing by N, we obtain

$$Q_N \leq \frac{\log J_N(\gamma^N)}{N} \leq c_a \epsilon + \hat{\Lambda}.$$

Taking the limit as $N \to \infty$ and using the fact that ϵ was arbitrary, we conclude that

$$\limsup_{N\to\infty} Q_N \leq \hat{\Lambda} \tag{5.15}$$

By (5.13), (5.15), and the inequality $R_N \leq Q_N$, the theorem follows. □

5.3 No loss of optimality with equal decision rules at the second stage of the "daisy chain"

We are now ready to consider the problem of Bayesian detection in the "daisy chain". We restrict to the case with $M = 2$, and constant priors $\Pr(H_0)$ and $\Pr(H_1)$. Remember that Y_0 denotes the decision of the first stage (the first fusion center), and U_0 denotes the decision of the second stage (the second fusion center). We prove the following result:

Theorem 5.2. *Let us restrict to sequences $\{\gamma^N\}$ for the decision rules of the first stage such that the probabilities of error at the first stage satisfy*

$$\lim_{N\to\infty} \frac{\log \Pr(Y_0 = 0 \,|\, H_1)}{N} = -\beta_2 \quad \textit{and}$$
$$\lim_{N\to\infty} \frac{\log \Pr(Y_0 = 1 \,|\, H_0)}{N} = -\beta_1$$

for some $\beta_1, \beta_2 > 0$. It is then asymptotically optimal to have all sensors in the second stage of the "daisy chain" use the same decision rule.

Remark 5.3. An intuitive way of understanding the result is that once the fusion center of the first stage has formed its belief, we are left with a single set of N sensors (corresponding to the sensors of the second stage of the "daisy chain") which observe the exponentially skewed (varying with N) priors $\pi_0^N = \Pr(H_0 \,|\, Y_0)$ and $\pi_1^N = \Pr(H_1 \,|\, Y_0)$. As we proved in Section 5.2, there is no loss of optimality if all sensors (i.e., sensors $N+1, \ldots, 2N$ of the "daisy chain") use the same decision rule.

Proof. By Bayes' rule,

$$\Pr(H_0 \,|\, Y_0 = 1) = \frac{\Pr(Y_0 = 1 \,|\, H_0) \Pr(H_0)}{\Pr(Y_0 = 1 \,|\, H_0) \Pr(H_0) + \Pr(Y_0 = 1 \,|\, H_1) \Pr(H_1)}.$$

It follows that

$$\begin{aligned}
&\lim_{N\to\infty} \frac{\log \Pr(H_0 \,|\, Y_0 = 1)}{N} \\
&\qquad = \lim_{N\to\infty} \left(\frac{\log \left(\Pr(Y_0 = 1 \,|\, H_0) \Pr(H_0) \right)}{N} \right. \\
&\qquad\qquad \left. - \frac{\log \left(\Pr(Y_0 = 1 \,|\, H_0) \Pr(H_0) + \Pr(Y_0 = 1 \,|\, H_1) \Pr(H_1) \right)}{N} \right) \\
&\qquad = -\beta_1 - 0 \\
&\qquad = -\beta_1,
\end{aligned}$$

and similarly

$$\lim_{N\to\infty} \frac{\log \Pr(H_1 \,|\, Y_0 = 0)}{N} = -\beta_2.$$

Fix N and decision rules $(\gamma_0, \gamma_1, \ldots, \gamma_N, \delta_1, \ldots, \delta_N) \in \Gamma_0 \times \Gamma^N \times \Delta^N$. Then the probability of an erroneous decision by the second fusion center is

$$\begin{aligned}
J_N(\gamma_0, \gamma^N, \delta^N) = {} & \Pr(\text{erroneous decision} \,|\, Y_0 = 0) \Pr(Y_0 = 0) \\
& + \Pr(\text{erroneous decision} \,|\, Y_0 = 1) \Pr(Y_0 = 1).
\end{aligned} \tag{5.16}$$

The first conditional probability is

$$\begin{aligned}
&\Pr(\text{erroneous decision} \,|\, Y_0 = 0) \\
&\quad = \Pr(U_0 = 0 \,|\, H_1, Y_0 = 0) \Pr(H_1 \,|\, Y_0 = 0) + \Pr(U_0 = 1 \,|\, H_0, Y_0 = 0) \Pr(H_0 \,|\, Y_0 = 0).
\end{aligned} \tag{5.17}$$

The second conditional probability is

$$\begin{aligned}
&\Pr(\text{erroneous decision} \,|\, Y_0 = 1) \\
&\quad = \Pr(U_0 = 0 \,|\, H_1, Y_0 = 1) \Pr(H_1 \,|\, Y_0 = 1) + \Pr(U_0 = 1 \,|\, H_0, Y_0 = 1) \Pr(H_0 \,|\, Y_0 = 1).
\end{aligned} \tag{5.18}$$

By redefining the $\mu(\,\cdot\,)$'s of Section 5.2 using conditional probabilities, so as to account

for the decision of the first stage, and noticing that

$$\lim_{N\to\infty} \frac{\log \pi^N_{1,Y_0=0}}{N} = \lim_{N\to\infty} \frac{\log \Pr(H_1 \,|\, Y_0=0)}{N} = -\beta_2,$$
$$\pi^N_{0,Y_0=0} = \Pr(H_0 \,|\, Y_0 = 0) = 1 - \pi^N_{1,Y_0=0},$$

and

$$\lim_{N\to\infty} \frac{\log \pi^N_{0,Y_0=1}}{N} = \lim_{N\to\infty} \frac{\log \Pr(H_0 \,|\, Y_0=1)}{N} = -\beta_1,$$
$$\pi^N_{1,Y_0=1} = \Pr(H_1 \,|\, Y_0 = 1) = 1 - \pi^N_{0,Y_0=1},$$

we can bound the right-hand side of each of Equations (5.17) and (5.18) above and below using Lemma 5.1. In particular, we define the exponents

$$\hat{\Lambda}_{\beta_2} = \lim_{N\to\infty} \inf_{(\gamma_0,\gamma^N,\delta^N)\in\Gamma_0\times\Gamma^N\times\Delta^N \text{ s.t } Y_0=0} \frac{\log \Pr(\text{erroneous decision} \,|\, Y_0 = 0)}{N} \tag{5.19}$$

and

$$\hat{\Lambda}_{\beta_1} = \lim_{N\to\infty} \inf_{(\gamma_0,\gamma^N,\delta^N)\in\Gamma_0\times\Gamma^N\times\Delta^N \text{ s.t } Y_0=1} \frac{\log \Pr(\text{erroneous decision} \,|\, Y_0 = 1)}{N}. \tag{5.20}$$

Furthermore, we can use Bayes' rule to write

$$\Pr(Y_0 = 0) = \Pr(Y_0 = 0 \,|\, H_0)\Pr(H_0) + \Pr(Y_0 = 0 \,|\, H_1)\Pr(H_1) \qquad \text{and}$$
$$\Pr(Y_0 = 1) = \Pr(Y_0 = 1 \,|\, H_0)\Pr(H_0) + \Pr(Y_0 = 1 \,|\, H_1)\Pr(H_1).$$

Therefore $\lim_{N\to\infty} \Pr(Y_0 = 0) = \Pr(H_0)$, $\lim_{N\to\infty} \Pr(Y_0 = 1) = \Pr(H_1)$, and

$$\lim_{N\to\infty} \frac{\log \Pr(Y_0 = 0)}{N} = \lim_{N\to\infty} \frac{\log \Pr(Y_0 = 1)}{N} = 0.$$

It follows that $\hat{\Lambda}$ in the proof of Theorem 5.1 can be replaced with $\hat{\Lambda}_D(\beta_1, \beta_2)$, where we define

$$\hat{\Lambda}_D(\beta_1, \beta_2) = \max(\hat{\Lambda}_{\beta_1}, \hat{\Lambda}_{\beta_2}). \tag{5.21}$$

Notice that the two exponents $\hat{\Lambda}_{\beta_2}$, $\hat{\Lambda}_{\beta_1}$ arise from the two summands respectively on the right-hand side of Equation (5.16); $\hat{\Lambda}_D(\beta_1, \beta_2)$ is the "slowest" (dominating) exponent out of the two, which determines the rate of decay of the probability of

error in (5.16).

The proof of Theorem 5.1 can be then followed verbatim to show that all sensors in the second stage can use the same decision rule without loss of optimality (in the sense of the overall network's error exponent). □

5.4 No loss of optimality with equal decision rules at the first stage of the "daisy chain"

Having proven that there is no loss of optimality if all decision rules at the second stage of the "daisy chain" are constrained to be equal, we are now ready to prove that there is no loss of optimality if all sensors in the first stage use the same decision rule as well. We prove the following result:

Theorem 5.4. *Let us restrict to sequences* $\{\gamma^N\}$ *for the decision rules of the first stage such that the probabilities of error at the first stage satisfy*

$$\lim_{N\to\infty} \frac{\log \Pr(Y_0 = 0 \,|\, H_1)}{N} < 0 \qquad \textit{and}$$
$$\lim_{N\to\infty} \frac{\log \Pr(Y_0 = 1 \,|\, H_0)}{N} < 0.$$

It is then asymptotically optimal to have all sensors in the first stage of the "daisy chain" use the same decision rule.

Sketch of Proof. Some technical details are omitted, and therefore the reasoning presented here is not a complete proof.

In Section 5.3 we argued that given exponents $-\beta_1 < 0, -\beta_2 < 0$ for the probabilities of error at the first stage, the optimal error exponent for the overall network is given by $\hat{\Lambda}_D(\beta_1, \beta_2) = \max(\hat{\Lambda}_{\beta_1}, \hat{\Lambda}_{\beta_2})$. This means that to optimize the network's performance, the decision rules need to give rise to appropriate exponents $-\beta_1, -\beta_2$ so as to minimize $\hat{\Lambda}_D(\beta_1, \beta_2)$. Using the notation of Section 5.1, the optimal error

exponent of the "daisy chain" is

$$\inf_{\substack{\beta_1,\beta_2>0:\ \exists\gamma_0\in\Gamma_0,\{\gamma^N\} \text{ with } \{\gamma^N\}_i\in\Gamma\ \forall i \text{ s.t.} \\ \lim_{N\to\infty}\frac{\log \Pr(Y_0=0\,|\,H_1)}{N}=-\beta_1 \text{ and } \lim_{N\to\infty}\frac{\log \Pr(Y_0=1\,|\,H_0)}{N}=-\beta_2}} \max(\hat{\Lambda}_{\beta_1},\hat{\Lambda}_{\beta_2}). \tag{5.22}$$

For now we assume that the infimum is attained, although in a complete proof this assumption will not be needed. This optimization problem will yield optimizing pairs $(\hat{\beta}_1,\hat{\beta}_2)$, and therefore pairs of exponents $(-\hat{\beta}_1,-\hat{\beta}_2)$. Proving the theorem reduces to proving that any such pair of exponents $(-\hat{\beta}_1,-\hat{\beta}_2)$ is attainable with equal rules in a parallel single-stage configuration.

In a parallel single-stage configuration, fix a sequence of decision rules $\{\gamma^N\}$ for the sensors. As the fusion rule γ_0 varies, different pairs of exponents $(-\beta_1,-\beta_2)$ $(\beta_1,\beta_2\geq 0)$ are attained asymptotically. Let us refer to the set of all such possible pairs of exponents as the "exponent set" for decision rules $\{\gamma^N\}$. The curve in Figure 5-1 traces all points on the exponent set for the sequence of decision rules $\{\gamma^N\}$. Notice that given the sequence of decision rules $\{\gamma^N\}$, the optimal decision rule for the fusion center is the MAP rule, at which $-\beta_1=-\beta_2$.

Crossings between curves corresponding to different sequences of decision rules are possible. This motivates the definition of the "envelope" of many exponent sets. If $\mathcal{F}$ is a non-empty collection of exponent sets, we define the "envelope" of all exponent sets in $\mathcal{F}$ as follows: a pair of exponents $(-\beta_1,-\beta_2)$ $(\beta_1,\beta_2\geq 0)$ belongs to the "envelope" if and only if

(i) $(\text{ -}\beta_1,-\beta_2)\in\overline{\bigcup_{\mathcal{F}}F}$, where an overline denotes the closure of a set, and

(ii) there does not exist a pair of exponents $(-\beta_1',-\beta_2')$ such that $(-\beta_1',-\beta_2')\in F'$ for some exponent set $F'\in\mathcal{F}$ and $-\beta_1'<-\beta_1,-\beta_2'<-\beta_2$.

Let $\mathcal{F}_{all}$ be the collection of all exponent sets. We refer to the envelope of all exponent sets (in $\mathcal{F}_{all}$) as the "mixed efficient frontier", since any pair of exponents on it results from possibly unequal decision rules. We use E_M to denote the "mixed efficient frontier", shown in the curve of Figure 5-2. It is easy to argue that any pair $(-\hat{\beta}_1,-\hat{\beta}_2)$ lies on the mixed efficient frontier.

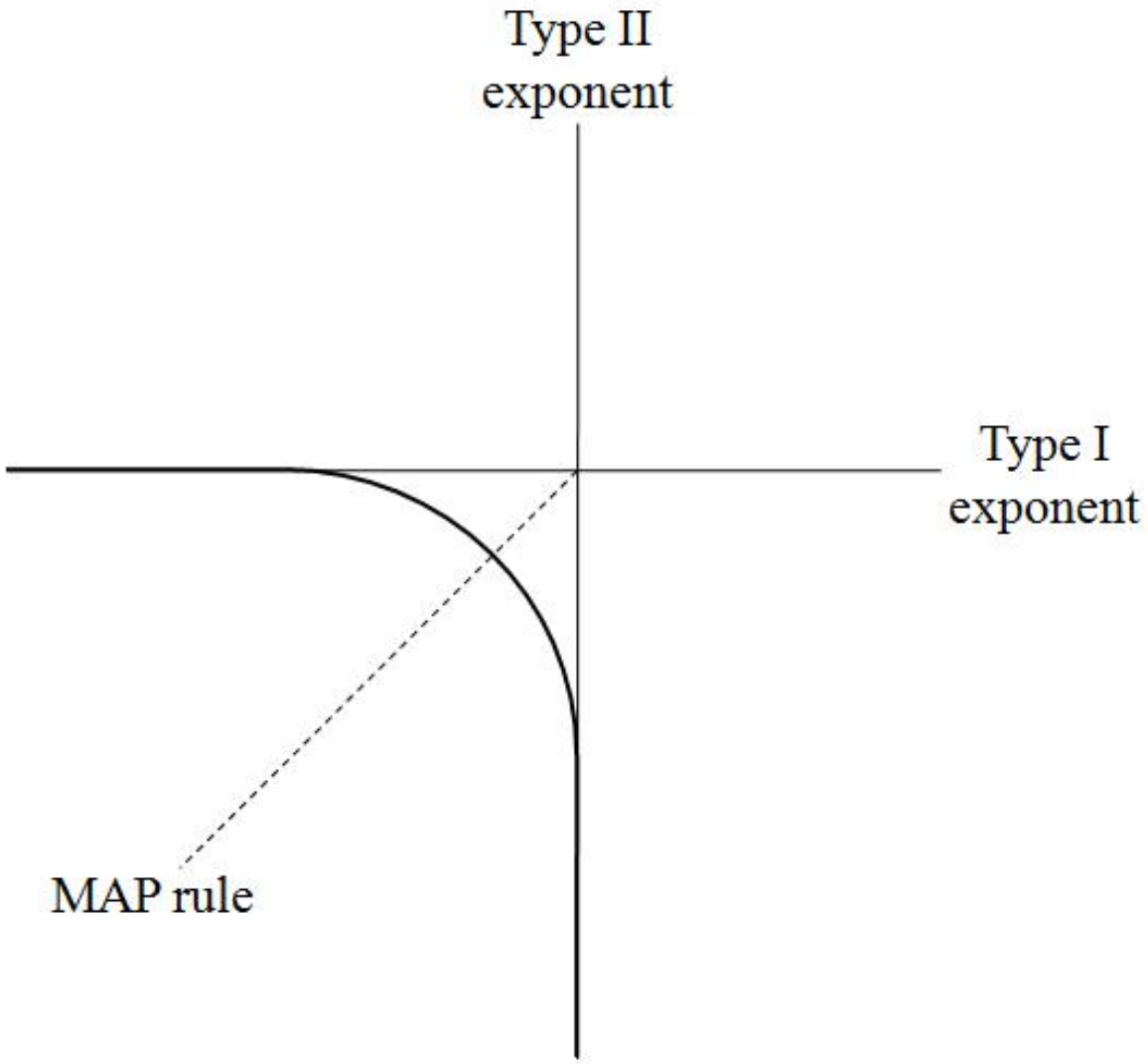

Figure 5-1: The "exponent set" for a sequence of decision rules.

Fixing γ and a sequence of decision rules $\{\gamma^N\}$ such that $\gamma_1 = \ldots = \gamma_N = \ldots = \gamma$, different pairs of exponents $(-\beta_1, -\beta_2)$ $(\beta_1, \beta_2 \geq 0)$ are attained asymptotically as the fusion rule γ_0 varies. Let $\mathcal{F}_{equal}$ be the collection of all exponent sets for sequences of decision rules $\{\gamma^N\}$ such that $\gamma_1 = \ldots = \gamma_N = \ldots$. We can consider the envelope of all exponent sets in $\mathcal{F}_{equal}$, which we refer to as the "pure efficient frontier", and denote by E_P.

Assume for the purpose of contradiction that there exists a point A on E_M which lies outside E_P (see Figure 5-3). There then exists a pair of exponentially skewed priors $\Pr(H_0), \Pr(H_1)$ for which the translation of the mixed efficient frontier is such

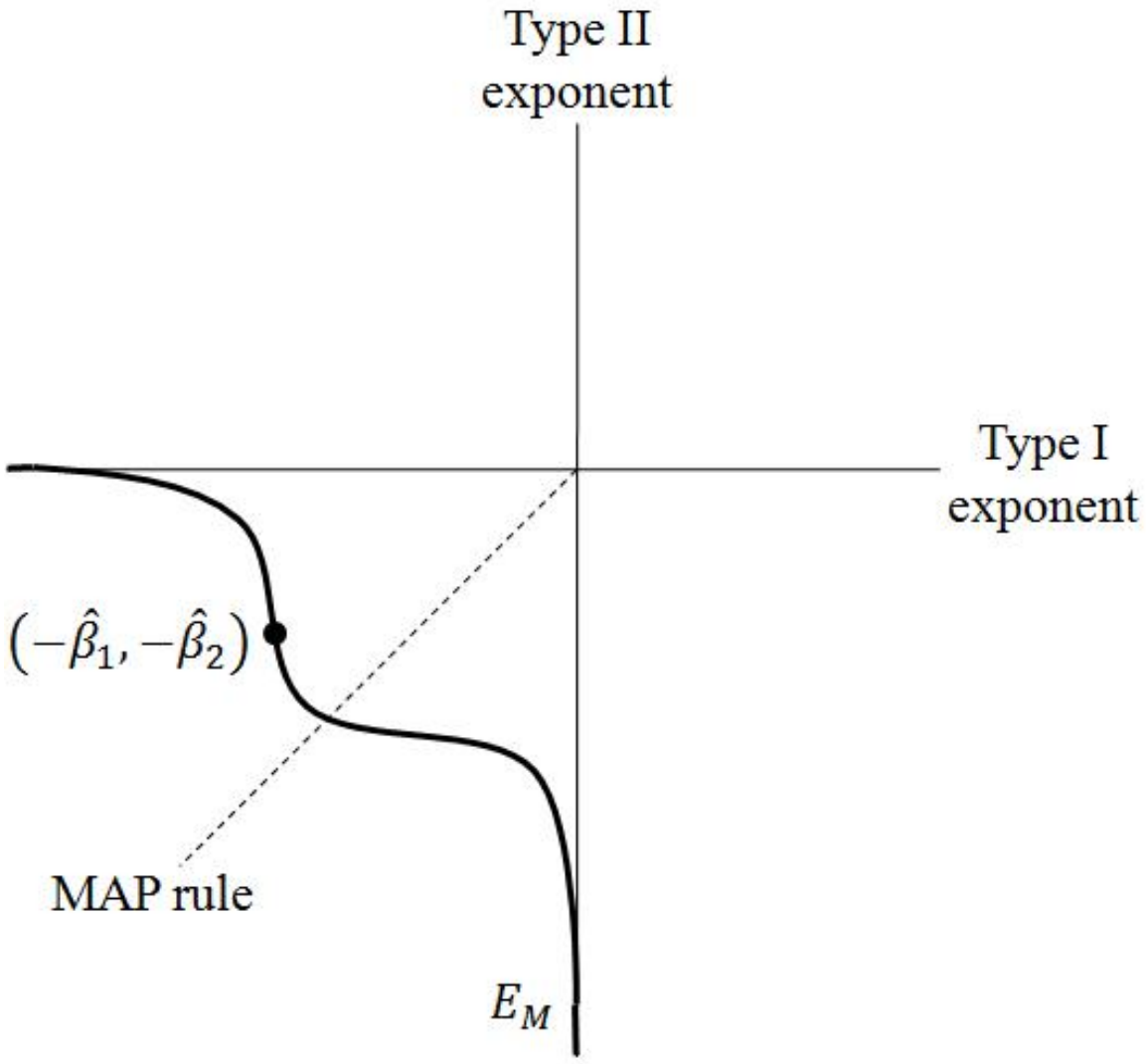

Figure 5-2: The "mixed efficient frontier".

that the translation of point A (call it A^T) is now on the slope-1 line. Notice that skewing the priors results in translation of the pure efficient frontier, too, from E_P to E_P^T. This shows that if the fusion rule is constrained to use the MAP rule, the strategy $\{\gamma^N\}$ on the pure efficient frontier is strictly worse (leads to a greater error exponent asymptotically) than the strategy on the mixed efficient frontier which corresponds to point A^T. This contradicts our result in Section 5.2, that there is no loss of optimality if all sensors are constrained to use the same decision rule in a parallel configuration under exponentially skewed priors.

It follows that any pair $(-\hat{\beta}_1, -\hat{\beta}_2)$ lies on the pure efficient frontier, which com-

pletes the proof.

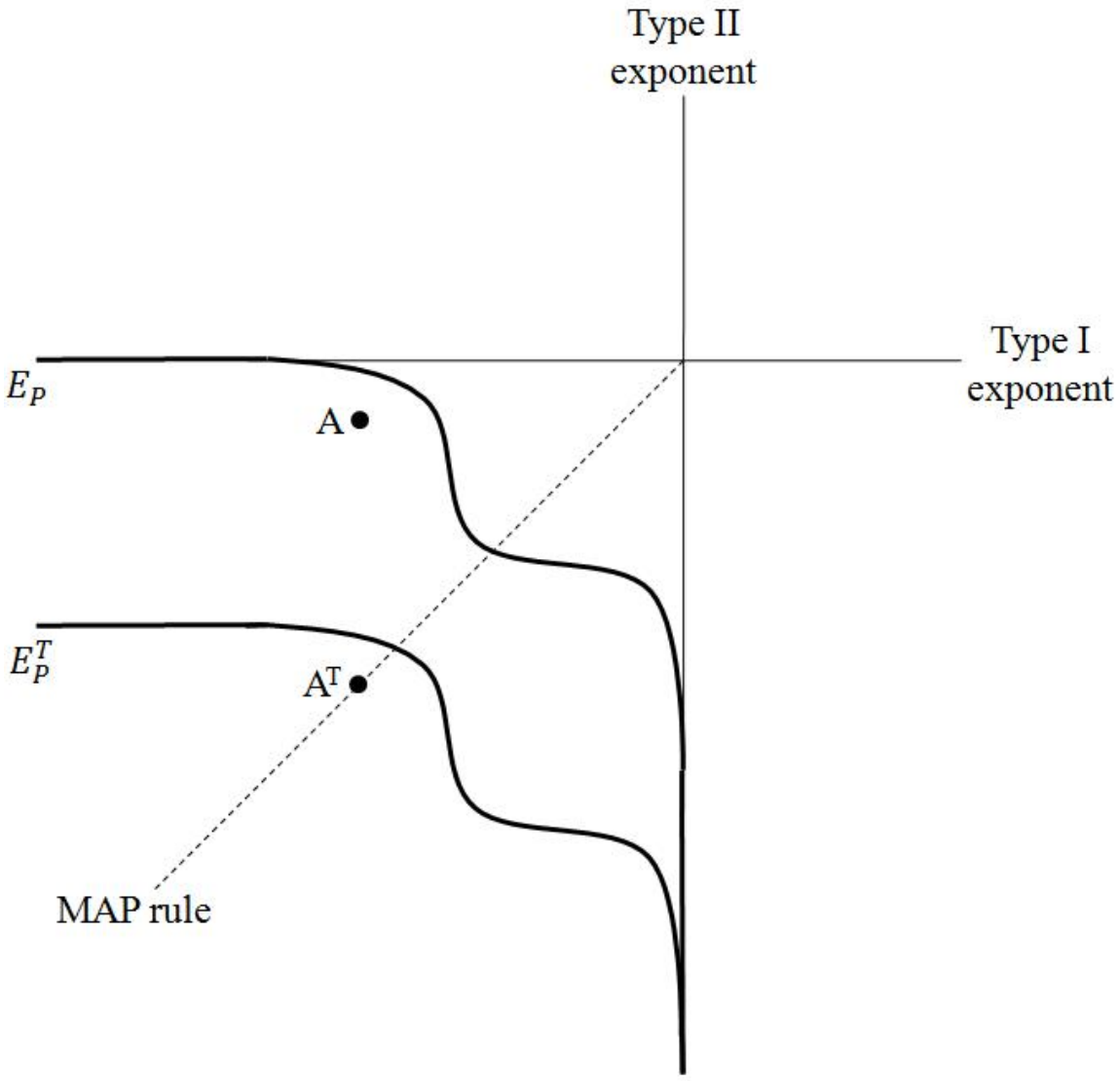

Figure 5-3: Introducing skewed priors translates both the mixed efficient frontier and the pure efficient frontier.

Chapter 6

Conclusion

6.1 Summary

In this thesis we have studied a decentralized detection problem and focused on determining the value of feeding a preliminary decision to sensors, which take it into consideration along with their observation when making their local decision, in improving the detection performance of a sensor network architecture. Performance is studied in terms of exponents of the error probability; that is, the analysis is done in the asymptotic regime, as the number of sensors in the network becomes very large.

We define the detection problem for all network configurations of interest: the parallel configuration with N sensors, the parallel configuration with $2N$ sensors, the simple feedback configuration (with N sensors), and a newly introduced configuration, the "daisy chain", in which the preliminary decision of a first set of N sensors is passed on through a first fusion center to a second set of N sensors and a second fusion center, which is responsible for the global decision. Our mathematical framework encompassed all of these configurations, as well as a definition of the asymptotic regime.

We use the "simulation" argument (i.e., simulating the communication capabilities of a network configuration with the communication capabilities of another) to compare the performance of configurations of interest in the optimal error exponent sense. We consider a classical decentralized detection model where all sensors

are scalar linear Gaussian binary detectors, and analyze asymptotic performance as the measurements under the two hypotheses become more and more informative, to conclude that the value of feeding the preliminary decision to decision makers is asymptotically negligible.

We prove that under the Neyman-Pearson formulation, the performance of the "daisy chain" configuration is asymptotically equal to the performance of the parallel configuration with twice as many sensors as the classical scheme, in the optimal error exponent sense. We first show that the "daisy chain" cannot be worse than the parallel configuration with twice as many sensors as the classical scheme; we then prove that it cannot perform better. We conclude that the value of feeding the preliminary decision to a second set of sensors is asymptotically negligible. We also prove that there is no loss of optimality asymptotically in the "daisy chain" if all decision rules at both stages are constrained to be equal.

We solve the Bayesian detection problem for the "daisy chain" configuration. [12] proves that for the parallel topology, it is asymptotically optimal to let all sensors use the same decision rule in deciding what to transmit. In particular, it is asymptotically optimal to have all sensors perform identical likelihood ratio tests, using the same threshold. First we extend this result to the case of exponentially skewed priors; we then make the connection between the result for exponentially skewed priors and Bayesian detection in the "daisy chain". Specifically, we prove that under the mild assumption that the Type I and Type II error probabilities of the first stage decay exponentially with the number of sensors, it is asymptotically optimal to have all sensors in the second stage perform identical likelihood ratio tests. Finally, we prove that it is asymptotically optimal to have all sensors in the first stage perform identical likelihood ratio tests as well. In all cases, optimality is meant in terms of the overall system's optimal error exponent.

6.2 Future work

Several issues remain outstanding and are areas of future research. The simple feedback configuration (which is defined in Subsection 2.1.3) is what motivated the study of the "daisy chain" (which is defined in Subsection 2.1.4); the "daisy chain" became the most salient configuration in our research work, leaving an intriguing question unanswered: what is the value (in performance improvement) of feeding the preliminary decision back to the same set of sensors that generated it, like the simple feedback configuration suggests? We conjecture that for the Neyman-Pearson criterion, the performance of the simple feedback configuration is asymptotically equal to the performance of the parallel configuration when each sensor transmits two bits, in the optimal error exponent sense. This result would be in the spirit of our theorem stating that the performance of the "daisy chain" is asymptotically equal to the performance of the parallel configuration with $2N$ sensors. The two results can be thought of in the light of Figures 3-1 and 3-2, respectively.

Another extension to our study of the "daisy chain" is to let the first fusion center send more than one bit to the second set of sensors and the second fusion center. We conjecture that, for binary hypothesis testing and the Neyman-Pearson formulation, sending more that one bit has no value asymptotically over only sending one bit. Supposing we are allowed to send N bits from the first fusion center to the second set of sensors and the second fusion center, there is no better rule than for the N-bit bus to the second stage of the "daisy chain" to be a copy of the N local decisions, coming from the first set of N sensors (each sensor transmits each decision, which is one bit, to the first fusion center). Because of the assumption of identical sensors in both stages, nevertheless, it suffices for the sensors in the second stage and the second fusion center to know what the count is of the N decisions coming from the first set of N sensors, i.e. how many of the N bits are "1". This can be done if the first stage communicates with the second stage through a $\log N$-bit bus, instead of a N-bit bus. Therefore the problem is reduced from whether there is value in sending N bits from the first stage to the second, to whether there is value in sending $\log N$ bits from the

first stage to the second, over only sending one bit. A further reduction of the $\log N$ bits to $\log \log N$ bits using the same argument would result in loss of information, and a reduction from $\log N$ bits to only one bit remains unproven.

Finally, a natural further step to our study of the "daisy chain" is to prove our belief that under the Bayesian formulation feedback results in strict improvement of the performance.

Appendix A

Stein Lemmas

We hereby present the Stein lemma (see [5]) and explain what we call the "Conditional" Stein lemma.

A.1 Stein lemma

Theorem A.1. *Let $X_1, X_2, \ldots, X_n$ be independent and identical distributed with distribution $Q()$. Consider the hypothesis test between two alternatives, $Q = P_1$ and $Q = P_2$, where $D(P_1||P_2) < \infty$. Let $A_n \subseteq \mathcal{X}^n$ be an acceptance region for hypothesis H_1. Let the probabilities of error be $\alpha_n = P_1^n(A_n^c)$, $\beta_n = P_2^n(A_n)$, and for $0 < \epsilon < \frac{1}{2}$, define $\beta_n^\epsilon = min_{A_n \subseteq \mathcal{X}^n, \alpha_n < \epsilon} \beta_n$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta_n^\epsilon = -D(P_1||P_2),$$

where D is the Kullback-Leibler divergence of two distributions.

A.2 "Conditional" Stein lemma

We introduce this corollary of the Stein lemma to employ it specifically in the analysis of Neyman-Pearson detection in the "daisy chain" with $M = D_1 = D_2 = 2$. Because the final decision U_0 is a function of the first fusion center's belief Y_0, the Stein

lemma can characterize the probabilities of error of the final decision with respect to the original hypotheses H_0, H_1 conditioned on the decision of the first stage. We use this observation for facts 2 and 3 in the proof of Proposition 4.1.

Bibliography

[1] S. Alhakeem and P. K. Varshney. A unified approach to the design of decentralized detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):9–20, 1995.

[2] S. Alhakeem and P. K. Varshney. Decentralized bayesian detection with feedback. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(4):503–513, 1996.

[3] R. R. Bahadur. *Some Limit Theorems in Statistics*. SIAM, Philadelphia, 1971.

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

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New Jersey, second edition, 2006.

[6] O. P. Kreidl. *Graphical models and message-passing algorithms for network-constrained decision problems*. PhD dissertation, Massachusetts Institute of Technology.

[7] D. A. Pantos, K. W. Halford, D. Kazakos, and P. Papantoni-Kazakos. Distributed binary hypothesis testing with feedback. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(1):21–42, 1995.

[8] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels, I. *Information and Control*, 10:65–103, 1967.

[9] P. F. Swaszek and P. Willett. Parley as an approach to distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):447–457, 1995.

[10] W. P. Tay. *Decentralized detection in resource-limited sensor network architectures*. PhD dissertation, Massachusetts Institute of Technology.

[11] R. R. Tenney and Jr. N.R. Sandell. Detection with distributed sensors. *IEEE Transactions on Aerospace and Electronin Systems*, 17:501–510, 1981.

[12] J. N. Tsitsiklis. Decentralized detection by a large number of sensors. *Mathematics of Control, Signals, and Systems*, 1:167–182, 1988.

[13] J. N. Tsitsiklis. Decentralized detection. *Advances in Statistical Signal Processing*, 2:297–344, 1993.

[14] J. N. Tsitsiklis. Data fusion architectures. Proposal, NSF, 2006.