



**HAL**  
open science

# On cooperative and concurrent detection in distributed hypothesis testing

Pierre Escamilla

► **To cite this version:**

Pierre Escamilla. On cooperative and concurrent detection in distributed hypothesis testing. Information Theory [cs.IT]. Institut Polytechnique de Paris, 2019. English. NNT : 2019IPPAT007 . tel-02525159

**HAL Id: tel-02525159**

**<https://theses.hal.science/tel-02525159>**

Submitted on 30 Mar 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT  
POLYTECHNIQUE  
DE PARIS

NNT : 2019IPPAT007

Thèse de doctorat



# On cooperative and concurrent detection in distributed hypothesis testing

Thèse de doctorat de l'Institut Polytechnique de Paris  
préparée à Télécom ParisTech

École doctorale n°626 de l'Institut Polytechnique de Paris (ED IP Paris)  
Spécialité de doctorat : Réseaux, Information et Communications

Thèse présentée et soutenue à Boulogne-Billancourt, le 3 décembre 2019, par

**PIERRE ESCAMILLA**

Composition du Jury :

M. Sheng Yang Professeur, CentraleSupélec	Président
M. Michael Gastpar Professeur, EPFL	Rapporteur
M. Albert Guillén i Fàbregas Professeur, Universitat Pompeu Fabra	Rapporteur
M. Ligong Wang Chargé de recherche, CNRS	Examineur
M. Malcolm Egan Chargé de recherche, INRIA	Examineur
Mme Michèle Wigger Professeur, Télécom ParisTech	Directeur de thèse
M. Abdellatif Zaidi Professeur, Université Paris-Est Marne la Vallée	Co-directeur de thèse





**Titre :** Tests d'hypothèses distribués: coopération et détection concurrente

**Mots clés :** statistiques, théorie de l'information, test d'hypothèses

**Résumé :** L'inférence statistique prend une place prépondérante dans le développement des nouvelles technologies et inspire un grand nombre d'algorithmes dédiés à des tâches de détection, d'identification et d'estimation. Cependant il n'existe pas de garantie théorique pour les performances de ces algorithmes. Dans cette thèse, nous considérons un réseau simplifié de capteurs communicant sous contraintes pour tenter de comprendre comment des détecteurs peuvent se partager au mieux les informations à leur disposition pour détecter un même événement ou des événements distincts. Nous investiguons différents aspects de la coopération entre détecteurs et comment des besoins contradictoires peuvent être satisfaits au mieux dans le cas de tâches de détection. Plus spécifiquement nous étudions un problème de test d'hypothèse où chaque détecteur doit maximiser l'exposant de décroissance de l'erreur de Type II sous une contrainte d'erreur de Type I donnée. Comme il y a plusieurs détecteurs intéressés par des informations distinctes, un compromis entre les vitesses de décroissance atteignables va apparaître. Notre but est de caractériser la région des

compromis possibles entre exposants d'erreurs de Type II. Dans le cadre des réseaux de capteurs massifs, la quantité d'information est souvent soumise à des limitations pour des raisons de consommation d'énergie et de risques de saturation du réseau. Nous étudions donc, en particulier, le cas du régime de communication à taux de compression nul (i.e. le nombre de bits des messages croît de façon sous-linéaire avec le nombre d'observations). Dans ce cas, nous caractérisons complètement la région des exposants d'erreurs de Type II dans les configurations où les détecteurs peuvent avoir des buts différents. Nous étudions aussi le cas d'un réseau avec des taux de compressions positifs (i.e. le nombre de bits des messages augmente de façon linéaire avec le nombre d'observations). Dans ce cas, nous présentons des sous-parties de la région des exposants d'erreur de Type II. Enfin, nous proposons dans le cas d'un problème point à point avec un taux de compression positif une caractérisation complète de l'exposant de l'erreur de Type II optimal pour une famille de tests gaussiens.

**Title :** On cooperative and concurrent detection for distributed hypothesis testing

**Keywords :** statistics, information theory, hypothesis testing

**Abstract :** Statistical inference plays a major role in the development of new technologies and inspires a large number of algorithms dedicated to detection, identification and estimation tasks. However, there is no theoretical guarantee for the performance of these algorithms. In this thesis we try to understand how sensors can best share their information in a network with communication constraints to detect the same or distinct events. We investigate different aspects of detector cooperation and how conflicting needs can best be met in the case of detection tasks. More specifically we study a hypothesis testing problem where each detector must maximize the decay exponent of the Type II error under a given Type I error constraint. As the detectors are interested in different information, a compromise between the achievable decay exponents of the Type II error appears. Our goal is to characterize the region of possible trade-offs bet-

ween Type II error decay exponents. In massive sensor networks, the amount of information is often limited due to energy consumption and network saturation risks. We are therefore studying the case of the zero rate compression communication regime (i.e. the messages size increases sub-linearly with the number of observations). In this case we fully characterize the region of Type II error decay exponent. In configurations where the detectors have or do not have the same purposes. We also study the case of a network with positive compression rates (i.e. the messages size increases linearly with the number of observations). In this case we present subparts of the region of Type II error decay exponent. Finally, in the case of a single sensor single detector scenario with a positive compression rate, we propose a complete characterization of the optimal Type II error decay exponent for a family of Gaussian hypothesis testing problems.

---

## Contents

---

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Background and motivations . . . . .	11
1.2	Summary of the main contributions . . . . .	11
1.3	Organization of the thesis . . . . .	13
<b>2</b>	<b>Prerequisites</b>	<b>15</b>
2.1	Notations . . . . .	15
2.2	Types and typical sets . . . . .	16
2.3	Sanov's theorem . . . . .	18
2.4	Typicality . . . . .	18
2.5	Blowing up lemma . . . . .	19
<b>3</b>	<b>Prior-art</b>	<b>21</b>
3.1	Centralized hypothesis testing . . . . .	21
3.1.1	Receiver operating characteristic (ROC) . . . . .	21
3.1.2	Error exponent . . . . .	23
3.2	Hypothesis testing with communication constraints . . . . .	24
3.2.1	The work of Ahlswede and Csiszár . . . . .	25
3.2.2	No binning: Han's Scheme . . . . .	26
3.2.3	Role of binning: Shimokawa, Han and Amari scheme . . . . .	26
3.2.4	Exponent-rate function when testing against conditional independence. . . . .	27
3.2.5	Some examples . . . . .	28
3.2.6	Zero-rate hypothesis testing . . . . .	29

3.2.7	Model with cooperation . . . . .	31
3.2.8	Model with multiple detectors . . . . .	33
3.3	Other extensions to networks and other lines of works . . . . .	35
<b>4</b>	<b>Zero-rate distributed hypothesis testing</b>	<b>37</b>
4.1	<b>Formal Problem Statement</b> . . . . .	38
4.2	Coherent detection and concurrent detection with $P_X = \bar{P}_X$ . . . . .	39
4.3	Concurrent detection with $P_X \neq \bar{P}_X$ . . . . .	40
4.3.1	Concurrent detection with $P_X \neq \bar{P}_X$ and $W_1 \geq 3$ . . . . .	40
4.3.2	Concurrent detection with $P_X \neq \bar{P}_X$ and $W_1 = 2$ . . . . .	41
4.4	Numerical Example . . . . .	43
4.5	Benefits of Cooperation . . . . .	44
<b>5</b>	<b>Positive-rate distributed hypothesis testing</b>	<b>47</b>
5.1	Special cases . . . . .	48
5.1.1	High rate regime . . . . .	48
5.1.2	Degraded side information at first detector . . . . .	50
5.1.3	Degraded side information at second detector . . . . .	50
5.1.4	Testing against independence under coherent detection . . . . .	51
5.2	General achievability results . . . . .	52
5.2.1	A simple scheme with cooperation . . . . .	52
5.2.2	Achievable regions . . . . .	56
5.3	Examples for coherent detection . . . . .	58
<b>6</b>	<b>On vector Gaussian hypothesis testing</b>	<b>61</b>
6.1	Formal problem statement . . . . .	61
6.2	Optimal exponent for a class of vector Gaussian hypothesis testing problems . . . . .	63
6.3	Proof of Theorem 6.1 . . . . .	64
6.4	Discussion . . . . .	69
<b>7</b>	<b>Summary</b>	<b>71</b>
7.1	Coherent detection and non-coherent detection with cooperation . . . . .	71
7.2	Distributed hypothesis testing in the Gaussian case . . . . .	72
<b>A</b>	<b>Proof of Theorem 2.7</b>	<b>73</b>
<b>B</b>	<b>Proof of Proposition 4.1</b>	<b>77</b>

<b>C</b>	<b>Proof of Proposition 4.2</b>	<b>81</b>
<b>D</b>	<b>Proof of Theorem 4.4</b>	<b>85</b>
<b>E</b>	<b>Proof of converse to Theorem 5.3</b>	<b>91</b>
<b>F</b>	<b>Proof of Theorem 5.5</b>	<b>95</b>
<b>G</b>	<b>Proof of Theorem 5.7</b>	<b>99</b>
<b>H</b>	<b>Proof of Equation (6.36)</b>	<b>101</b>
<b>I</b>	<b>Proof of Equivalence between (6.41) and constraint C</b>	<b>103</b>





---

## List of Figures

---

1.1	A Heegard-Berger type source coding model with unidirectional conferencing for multi-terminal hypothesis testing. . . . .	12
3.1	Approximation of ROCs in logarithm scales (also known as Detection Error Tradeoff (DET characteristics) for various algorithms presented at FRVT [24]. The approximation is done using a base of labeled face images. The False non-match rate corresponds to the ratio of matching faces pairs (two face images of the same person) in the base that where classified as being those of distinct persons. The False match rate corresponds to the ratio of non-matching faces pairs (two face images of distinct persons) in the base that where classified as being those of the same person. . . . .	22
3.2	Binary hypothesis testing with one sided data compression. . . . .	24
3.3	Binary hypothesis testing with one-sided data compression and two stream of side information at the Detector. . . . .	27
3.4	Optimal exponent-rate function given by Theorem 3.3 for Example 3.1 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.1. . . . .	28
3.5	Lower bounds for the exponent-rate function given by Theorems 3.6 and 3.7 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.2. . . . .	29
3.6	Lower bounds for the exponent-rate function given by Theorems 3.6 and 3.8 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.3. . . . .	30
3.7	Binary hypothesis testing in a cooperative MAC studied in [47]. . . . .	31
3.8	Binary hypothesis testing in a Heegard-Berger network studied in [28]. . . . .	33

---

4.1	A Heegard-Berger type source coding model with unidirectional conferencing for multi-terminal hypothesis testing. . . . .	37
4.2	Exponents region of Example 4.1, see [12] for implementation details. On the left: exponent regions $\mathcal{E}_0(2, 2, \epsilon_1, \epsilon_2)$ and $\mathcal{E}_0(2, 0, \epsilon_1, \epsilon_2)$ for coherent detection. On the right: exponent regions $\mathcal{E}_0(2, 2, \epsilon_1, \epsilon_2)$ , and $\mathcal{E}_0(2, 0, \epsilon_1, \epsilon_2)$ for concurrent detection. . . . .	44
5.1	A Heegard-Berger type source coding model with unidirectional conferencing for multi-terminal hypothesis testing. . . . .	47
5.2	Equivalent system without cooperation when $X \text{---} Y_2 \text{---} Y_1$ under both hypotheses. . .	50
5.3	Equivalent point to point system when $X \text{---} Y_1 \text{---} Y_2$ under both hypotheses. . . . .	51
5.4	Error-exponent region of Example 5.1. . . . .	59
6.1	Vector Gaussian hypothesis testing problem . . . . .	61
6.2	Problem $P_2$ . . . . .	64
6.3	Rate-exponent region for Example 6.1. . . . .	70

---

## Acknowledgements

---

I would like to take this opportunity to thank all those who supported me during my thesis. First and foremost, I would like to express my sincere gratitude to my advisors Prof. Michèle WIGGER and Prof. Abdellatif ZAIDI for the continuous support of my Ph.D study and research, for their patience, motivation and enthusiasm. I have learned so much and it has been a wonderful experience to work with them. Next, I would like to mention my colleagues at Télécom ParisTech in our group: Dr. Sarah KAMEL, Dr. Longguang LI, Homa NIKBAKHT, Dr. K. G. NAGANANDA, Samet GELINCIK, Mehrasa AHMADIPOUR, Dr. Qifa YAN and Dr. Sadaf SALEHKALAIBAR and at HUAWEI FRC: Yiğit UĞUR and Dr. Iñaki ESTELLA. I value very much the discussions I had with them. I am also deeply thankful to my colleague and friends at Télécom ParisTech and Huawei FRC for all the fun we had in the last three years: Vincent CORLAY, Mathieu LECONTE, Jeremie LEGAY, Alexis DECURNINGE, Paul FERRAND, Lorenzo MAGGI, Pierre-Louis POIRION, Paolo MEDIGLIANI and Zaid ALIBOKUS. I'd like also to thanks all my friends for their priceless help and friendship. I would like to thank all my family and in particular my wife Juliette, my parents and my sisters for their support throughout all my life. Last but not least I would like to thanks Mr DESHAYES for his priceless work.



## List of abbreviations

**DET:** Detection Error Trade-off

**FRVT:** Face Recognition Vendor Test

**IoT:** Internet of Things

**NIST:** National Institute of Standards and Technology

**p.d.f.:** Probability density function

**p.m.f.:** Probability mass function

**ROC:** Receiver Operating Characteristic



### 1.1 Background and motivations

New technologies based on statistical inference allow to perform a number of estimation, identification and detection tasks and the optimal performances for such tasks are desirable. For instance the National Institute of Standards and Technology (NIST) organizes empirical evaluations of the solutions developed by laboratories and companies around the world for facial recognition [24]. But these evaluations strongly depend on the types of photos taken. Facial illumination, quality of the photos, the angle of view, the expression of the subjects and their degrees of cooperation in the shooting are all parameters that make vary the behavior of an algorithm and the optimal performances for this task are not known. A first step towards a theoretical estimation of the optimal performances for face recognition algorithms is to understand the effect of compression constraints on statistical inference. This is a critical parameter also in the case of massive sensors networks found in the Internet of Things (IoT). These sensors networks are multi-purpose and can be used to detect a fire as well as the amount of light in a house. They thus have a number of simultaneous statistical testing tasks. Our work tackles with these two problems from a theoretic perspective.

### 1.2 Summary of the main contributions

In this thesis we consider first the single-sensor two-detectors system in Figure 1.1 where Detector 1, after receiving a message from the Sensor, can send a message to Detector 2. This additional message allows the detectors to collaborate in their decision and one of the the goals of our work is to quantify



the increase in the type II error exponents enabled by this cooperation. We show that even a single bit of communication between the detectors (the guess about the hypothesis at the transmitting detector) can provide an unbounded gain in the type II error exponent of the detector receiving the bit.

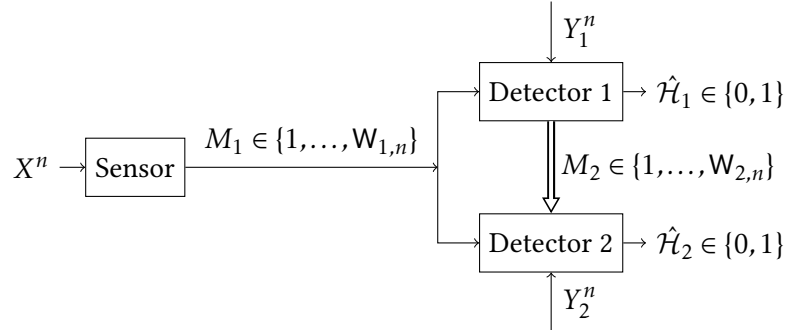


Figure 1.1: A Heegard-Berger type source coding model with unidirectional conferencing for multiterminal hypothesis testing.

Decentralized detection systems are of major importance for various applications such as autonomous vehicles or other complex monitoring systems. These systems use multiple detection pipelines that base their decisions on common or individual observations and often these decisions are fused at one or several central detectors. Our scenario can model both multiple detection pipelines and, through the cooperation link, fusion of various decisions. We consider the case with fixed-length communication. In our two-detectors setup where each detector aims at maximizing the error exponent under one of the two hypotheses, two cases can be distinguished: both detectors aim at maximizing their exponents under the same hypothesis (we refer to this setup as *coherent detection*) or the two detectors aim at maximizing their exponents under different hypotheses (we refer to this setup as *concurrent detection*). In this thesis we consider both scenarios. Under coherent detection or under concurrent detection when the sensor can send more than a single bit or cannot distinguish the two hypotheses, the exponents region is a rectangle. In these cases, each detector achieves the same exponent as if it were the only detector in the system. In contrast, a tradeoff arises under concurrent detection if the sensor can distinguish the two hypotheses but can only send a single bit to the detectors. A comparison with the optimal exponents regions without cooperation [9], allows us to exactly quantify the benefits of detector cooperation in this setup with fixed communication alphabets. All results explained in this paragraph remain valid when the alphabets size are not fixed but grow sublinearly in the length of the observed sequences. They also generalize to an arbitrary number of hypotheses. Whereas for two detectors a tradeoff between the exponents arises only when the sensor sends a single bit to the detectors, in a multi-hypothesis testing scenario with the number of distinct hypotheses,  $H$ , is superior or equal to 3 such a tradeoff can arise whenever the number of communicated bits does not exceed  $\log_2 H$ . The case of positive communication rates is also investigated. On a more technical level, the presence of a cooperation link between the detectors seems to make the

problem of identifying the optimal type II exponents significantly more difficult. For example, without cooperation, the set of achievable exponents for testing against independence has been solved in [41], and it is achievable with a simple scheme that does not rely on binning. With cooperation, we managed to identify the optimal exponents only under the additional assumption that the observations at the two detectors are independent under both hypotheses and the cooperation rate is zero. In the general case, binning is necessary, which makes generally it hard to prove optimality of the achieved exponent. Notable exceptions are the results included in [25, 30, 26, 33, 44]. For the sake of simplicity, in Chapter 5 we therefore only present and analyze a simple coding scheme without binning and also without Heegard-Berger [16] coding. We prove that this simple scheme is optimal in a special case of testing against independence where it achieves an exponent at Detector 2 equal to the sum of the exponents at both detectors in a non-cooperative setup. Cooperation between detectors thus allows to accumulate the error exponents at the detectors. The testing against independence problem considered in this paper differs from the one in [47], where the first detector cannot achieve a positive error exponent. The exponents region can significantly differ under the cooperative and the non-cooperative setup, in particular when based on its own observation the sensor can guess the hypothesis, communicate this guess to the detectors, and adapt the communication to this guess. With this strategy, the exponents region achieved by our simple scheme is a rectangle under concurrent detection, which means that each detector's exponent is the same as in a setup where the other detector is not present. Under coherent detection or concurrent detection when the sensor cannot distinguish the two hypotheses, the exponents region achieved by our scheme shows a tradeoff between the two exponents.

The major goal of the second part of this thesis is to characterize the rate-exponent function of a single-sensor single-detector model. We extend the known single-letter characterizations of the rate-exponent function in the Gaussian vector case (testing against independence and testing against conditional independence) to a more general class and provide explicit forms for them. We start by giving a single-letter characterization of the rate-exponent function generalizing the one of [25, Theorem 3] to a broader class of problems. Then we show that the rate-exponent functions of two systems whose observations are in a one to one relationship are equal. Finally, using previous results we provide a single-letter characterization of the rate-exponent function for a class of Gaussian hypothesis testing problem by showing that the rate-exponent function of each problem of this class is equal to the rate-exponent function of a problem previously characterized and we provide for it an explicit form based on the result of [39, 44].

### 1.3 Organization of the thesis

The remainder of this thesis is organized as follows. In Chapter 2 we introduce the distributed hypothesis testing problem and the information theoretic tools used in this thesis. In Chapter 3 we review known

results on hypothesis testing (i.e., collocated sensors). In Chapter 4 we present our result for the zero rate cases for the single-sensor two-detectors model: we present the error-exponent region for both concurrent and coherent detection for fixed length alphabet and sublinear growing number of bits. In Chapter 5 we present our results on positive rates for the same model: we give optimal results for special cases and an achievability result for the general problem. In Chapter 6 we present our results for the single-sensor single-detector model: we give optimal results for a family of Gaussian hypothesis tests.

In this section we present some basic information theory tools that will be used all along this report. First we present the notations, and then the method of types. Then we present the notion of typicality (for more details see [6] and [21]), and finally we present a last tool called the Blowing-up lemma (for more details see [7]).

## 2.1 Notations

Throughout, we use the following notation. Caligraphic letters  $\mathcal{X}$ ,  $\mathcal{Y}$  and  $\mathcal{Z}$  denote finite sets. Random variables are denoted by capital letters and their realizations by lower case, e.g.,  $X$  and  $x$ . The cardinality of a set  $\mathcal{X}$  is denoted by  $|\mathcal{X}|$  and the set of probability distributions on  $\mathcal{X}$  is denoted by  $\mathcal{P}(\mathcal{X})$ . A random or deterministic  $n$ -tuple  $X_1, \dots, X_n$  or  $x_1, \dots, x_n$  is abbreviated as  $X^n$  or as  $x^n$ . The  $n$ -fold Cartesian product of the set  $\mathcal{X}$  is noted  $\mathcal{X}^n$  and the probability that  $n$  independent drawings with p.m.f.  $P_X \in \mathcal{P}(\mathcal{X})$  results in a sequence  $x^n$  denoted by  $P_X^{\otimes n}(x^n)$ . For  $P_X \in \mathcal{P}(\mathcal{X})$  the set of all  $x \in \mathcal{X}$  such that  $P(x) \neq 0$  is called the support of  $P_X$  and noted  $\text{supp}(P_X) \subseteq \mathcal{X}$ . When two random variables  $(X, Y)$  are independent given a third random variable  $Z$  (i.e.  $P_{XYZ} = P_Z P_{X|Z} P_{Y|Z}$ ),  $(X, Z, Y)$  form a Markov chain and we note  $X \text{---} Z \text{---} Y$ . For random variables  $X, Y$  and  $\bar{X}, \bar{Y}$  over the same alphabet  $\mathcal{X} \times \mathcal{Y}$  with p.m.f.s  $P_{XY}$  and  $\bar{P}_{\bar{X}\bar{Y}}$  satisfying  $P_X \ll \bar{P}_{\bar{X}}$  (i.e., for every  $x_0 \in \mathcal{X}$ , if  $\bar{P}_{\bar{X}}(x_0) = 0$  then also  $P_X(x_0) = 0$ ),  $D(P_X \parallel \bar{P}_{\bar{X}})$  denotes the Kullback-Leiber divergence between  $P_X$  and  $\bar{P}_{\bar{X}}$ , and if for all  $y \in \mathcal{Y}$ ,  $P_{X|Y=y} \ll \bar{P}_{\bar{X}|Y=y}$ ,  $D(P_X \parallel \bar{P}_{\bar{X}}|Y)$  denotes the Kullback-Leiber divergence between  $X$  and  $\bar{X}$  given  $Y$ , i.e.:

$$D(P_X \parallel \bar{P}_{\bar{X}}) = \sum_{x \in \text{supp}(P_X)} P_X(x) \log \frac{P_X(x)}{\bar{P}_{\bar{X}}(x)}$$

$$D(P_X || \bar{P}_X | Y) = \sum_{y \in \mathcal{Y}} P_Y(y) D(P_{X|Y=y} || \bar{P}_{X|Y=y})$$

$H(\cdot)$  denotes entropy, i.e.:

$$H(P_X) = - \sum_{x \in \text{supp}(P_X)} P_X(x) \log P_X(x).$$

All along this report we will use the convention  $0 \log(0) = 0 \log(\frac{0}{0}) = 0$ , and if  $P_X \ll \bar{P}_X$  do not hold,  $D(P_X || \bar{P}_X) = \infty$ . For a sequence  $x_n$  the limit superior is

$$\overline{\lim}_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left( \sup_{m > n} x_m \right),$$

and the limit inferior is

$$\underline{\lim}_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left( \inf_{m > n} x_m \right).$$

Boldface upper case letters denote random vectors or matrices, e.g.,  $\mathbf{X}$ , where context should make the distinction clear. We denote the covariance of a zero mean, real-valued, vector  $\mathbf{X}$  with p.d.f.  $P_X$  by  $\mathbf{K}_X = \mathbb{E}_{P_X}[\mathbf{X}\mathbf{X}^\dagger]$  and with p.d.f.  $\bar{P}_X$  by  $\bar{\mathbf{K}}_X = \mathbb{E}_{\bar{P}_X}[\mathbf{X}\mathbf{X}^\dagger]$ , where  $(\cdot)^\dagger$  indicates transpose. Similarly, we denote the cross-correlation of two zero-mean vectors  $\mathbf{X}$  and  $\mathbf{Y}$  with p.d.f.  $P_{XY}$  by  $\mathbf{K}_{XY} = \mathbb{E}_{P_{XY}}[\mathbf{X}\mathbf{Y}^\dagger]$  with p.d.f.  $P_X$  by  $\bar{\mathbf{K}}_{XY} = \mathbb{E}_{\bar{P}_{XY}}[\mathbf{X}\mathbf{Y}^\dagger]$ . We denote the conditional covariance matrix of  $\mathbf{X}$  given  $\mathbf{Y}$  with p.d.f.  $P_{X|Y}$  by  $\mathbf{K}_{X|Y} = \mathbb{E}_{P_{X|Y}}[\mathbf{X}\mathbf{X}^\dagger | \mathbf{Y}]$  and the conditional covariance matrix of  $\mathbf{X}$  given  $\mathbf{Y}$  with p.d.f.  $\bar{P}_{X|Y}$  by  $\bar{\mathbf{K}}_{X|Y} = \mathbb{E}_{\bar{P}_{X|Y}}[\mathbf{X}\mathbf{X}^\dagger | \mathbf{Y}]$ . For a matrix  $\mathbf{M}$ ,  $\mathbf{M}^+$  denote the Moore-Penrose pseudo-inverse of  $\mathbf{M}$  and  $|\mathbf{M}|_+$  denotes the pseudo-determinant of  $\mathbf{M}$ .

## 2.2 Types and typical sets

The type of a sequence  $x^n \in \mathcal{X}^n$  and the joint type of the sequences  $x^n$  and  $y^n \in \mathcal{Y}^n$  are the p.m.f.s  $P_{x^n} \in \mathcal{P}(\mathcal{X})$  and  $P_{x^n, y^n} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  defined by letting  $P_{x^n}(x)$  and  $P_{x^n, y^n}(x, y)$  be the relative frequency of  $x$  among  $x_1, \dots, x_n$  and of  $(x, y)$  among  $(x_1, y_1), \dots, (x_n, y_n)$ :

$$P_{x^n}(x) = \frac{|\{i : x_i = x\}|}{n} \quad (2.1)$$

$$P_{x^n, y^n}(x, y) = \frac{|\{i : (x_i, y_i) = (x, y)\}|}{n} \quad (2.2)$$

Also, let

$$P_{x^n|y^n}(x|y) = \frac{P_{x^n, y^n}(x, y)}{P_{y^n}(y)} \quad (2.3)$$

be the conditional law induced by  $P_{x^n, y^n}(x, y)$  and  $P_{y^n}(y)$ . The set of all possible types of sequences  $x^n \in \mathcal{X}^n$  is denoted  $\mathcal{P}_n(\mathcal{X})$ :

$$\mathcal{P}_n(\mathcal{X}) = \{P_X \in \mathcal{P}(\mathcal{X}) : \exists x^n \in \mathcal{X}^n \text{ satisfying } P_{x^n} = P_X\}. \quad (2.4)$$

For any  $P_X \in \mathcal{P}_n(\mathcal{X})$ , the set of all possible sequences  $x^n$  with type  $P_X$  is the type class of  $P_X$ ,

$$\mathcal{T}^n(P_X) = \{x^n \in \mathcal{X}^n : P_{x^n} = P_X\}. \quad (2.5)$$

The following lemma states that  $\mathcal{P}_n(\mathcal{X})$  grows at polynomial speed with  $n$ :

**Lemma 2.1**

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}. \quad (2.6)$$

*Proof:* see [5]. ■

**Proposition 2.2** For all  $x^n \in \mathcal{T}^n(P_X)$ :

$$Q^{\otimes n}(x^n) = \exp \left[ -n(H(P_X) + D(P_X \| Q_X)) \right]. \quad (2.7)$$

*Proof:* see [5]. ■

We have the following bounds on the cardinality and the probability of the type class:

**Proposition 2.3** For any type  $P_X \in \mathcal{P}_n(\mathcal{X})$ :

$$(n+1)^{-|\mathcal{X}|} \exp[nH(P_X)] \leq |\mathcal{T}^n(P_X)| \leq \exp[nH(P_X)] \quad (2.8)$$

$$(n+1)^{-|\mathcal{X}|} \exp[-nD(P_X \| Q_X)] \leq Q_X^{\otimes n}(\mathcal{T}^n(P_X)) \leq \exp[-nD(P_X \| Q_X)] \quad (2.9)$$

*Proof:* Plugging  $Q_X = P_X$  into equation (2.7) yields  $P^{\otimes n}(x^n) = \exp[-nH(P_X)]$ , and thus

$$P_X^{\otimes n}(\mathcal{T}^n(P_X)) = \sum_{x^n \in \mathcal{T}^n(P_X)} P^{\otimes n}(x^n) = |\mathcal{T}^n(P_X)| \exp[-nH(P_X)]. \quad (2.10)$$

The second inequality in Equation (2.8) now holds simply because

$$1 \geq P^{\otimes n}(\mathcal{T}^n(P_X)). \quad (2.11)$$

The first inequality in Equation (2.9) holds by (2.11) and the following inequalities

$$\begin{aligned} P_X^{\otimes n}(\mathcal{T}^n(P_X)) &\stackrel{(a)}{=} \max_{\tilde{P}_X \in \mathcal{P}_n(\mathcal{X})} P_X^{\otimes n}(\mathcal{T}^n(\tilde{P}_X)) \\ &\stackrel{(b)}{\geq} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \sum_{\tilde{P}_X \in \mathcal{P}_n(\mathcal{X})} P_X^{\otimes n}(\mathcal{T}^n(\tilde{P}_X)) \\ &\stackrel{(c)}{=} \frac{1}{|\mathcal{P}_n(\mathcal{X})|} \\ &\stackrel{(d)}{\geq} (n+1)^{-|\mathcal{X}|}. \end{aligned} \quad (2.12)$$

Here, (a) holds of (2.7); (b) because the maximum is superior to the average; (c) because the type classes form a partition of  $\mathcal{X}^n$ ; and (d) because of Equation (2.6). Finally Equation (2.9) follows directly from equations (2.8) and (2.10). This concludes the proof. ■

The method of types has various application in information theory. In the following we present one of the, Sanov's theorem.

## 2.3 Sanov's theorem

**Theorem 2.4 (Sanov)** Let  $X^n = (X_1, \dots, X_n)$  be i.i.d.  $\sim Q_X$ . For any subset of types  $\mathcal{L} \subseteq \mathcal{P}(\mathcal{X})$ . The corresponding subset of  $\mathcal{X}^n$ ,

$$\mathcal{A}_n = \bigcup_{P_X \in (\mathcal{L} \cap \mathcal{P}_n(\mathcal{X}))} \mathcal{T}^n(P_X),$$

satisfies

$$\Pr(X^n \in \mathcal{A}_n) \leq (n+1)^{|\mathcal{X}|} e^{-n \min_{P_X \in \mathcal{L}} D(P_X \| Q_X)}. \quad (2.13)$$

If in addition, the set  $\mathcal{L}$  is the closure of its interior,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log(\Pr(X^n \in \mathcal{A}_n)) = \min_{P_X \in \mathcal{L}} D(P_X \| Q_X). \quad (2.14)$$

*Proof:* See [5, Theorem 11.4.1]. ■

## 2.4 Typicality

We now introduce the notion of typical sequences and typical sets which is central in information theory. Different notions of typicality can be encountered in the literature. In particular the weak typicality [5, Chapter 3] is opposed to the more restrictive strong typicality that we present below.

**Definition 2.1** Fix  $\mu > 0$ . A  $\mu$ -typical sequence with respect to (w.r.t.) a p.m.f.  $P_X \in \mathcal{P}(\mathcal{X})$  is a sequence  $x^n$  that satisfies:

$$|P_{x^n}(x) - P_X(x)| < \mu P_X(x) \quad \text{for all } x \in \mathcal{X}. \quad (2.15)$$

The subset

$$\mathcal{T}_\mu^n(P_X) = \{x^n : |P_{x^n}(x) - P_X(x)| < \mu P_X(x) \quad \text{for all } x \in \mathcal{X}\} \subset \mathcal{X}^n, \quad (2.16)$$

is called the  $\mu$ -typical set w.r.t.  $P_X$ . We define in a similar manner the set of the jointly  $\mu$ -typical sequences,  $\mathcal{T}_\mu^n(P_{XY})$ .

**Proposition 2.5** Fix  $\mu$  such that  $\mu_X \geq \mu > 0$  and  $x^n \in \mathcal{T}_\mu^n(P_X)$ . If  $X$  is generated i.i.d. given  $P_X$ , then:

$$e^{-n(1+\mu)H(P_X)} \leq P_X^{\otimes n}(x^n) \leq e^{-n(1-\mu)H(P_X)}, \quad (2.17)$$

$$(1 - \delta_\mu(n)) e^{n(1-\mu)H(P_X)} \leq |\mathcal{T}_\mu^n(P_X)| \leq e^{n(1+\mu)H(P_X)}, \quad (2.18)$$

$$1 - \delta_\mu(n) \leq \Pr(X^n \in \mathcal{T}_\mu^n(P_X)) \leq 1, \quad (2.19)$$

where  $\delta_\mu(n) = 2^{|\mathcal{X}|} \exp[-n\mu^2 \min_{x \in \text{supp}(P_X)} P_X(x)]$ .

*Proof:* see [21, Theorem 1.1]. ■

As mentioned before there are multiple definitions of typical sets that are adapted to different contexts. Strong typicality is well adapted to distributed hypothesis testing problem, but we can find some demonstration using other notions of typicality. A noticeable example is the demonstration of Chernoff-Stein Lemma given in [5, Section 11.8] which uses a coarser version than strong-typicality (i.e. relative entropy based typicality) yet sufficient for this purpose.

## 2.5 Blowing up lemma

We recall the “Blowing-Up” lemma [7, Theorem 5.4].

**Lemma 2.6 (“Blowing-Up” lemma)** *Let  $Y^n = (Y_1, \dots, Y_n)$  be independent random variables in  $\mathcal{Y}^n$ , distributed according to  $W^n(Y^n|X^n = x^n)$  for some fixed vector  $x^n \in \mathcal{X}^n$  and a stochastic mapping  $W : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ . Let  $\delta_n \rightarrow 0$  be a given sequence. Then, there exist sequences  $k_n$  and  $\gamma_n$  satisfying:*

$$\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{\gamma_n}{n} = 0$$

such that for every subset  $\mathcal{A}_n \subset \mathcal{Y}^n$  for which

$$W^n(\mathcal{A}_n|X^n = x^n) \geq \exp(-n\delta_n) \tag{2.20}$$

we have

$$W^n(\Gamma^{k_n} \mathcal{A}_n|X^n = x^n) \geq 1 - \gamma_n \tag{2.21}$$

where  $\Gamma^{k_n} \mathcal{A}_n$  denotes the  $\Gamma^{k_n}$ -neighborhood of  $\mathcal{A}_n$ , defined as

$$\Gamma^{k_n} \mathcal{A}_n := \left\{ \hat{y}^n \in \mathcal{Y}^n : \exists y^n \in \mathcal{A}_n \text{ satisfying } d_H(y^n, \hat{y}^n) \leq k_n \right\} \tag{2.22}$$

and  $d_H(\cdot, \cdot)$  denotes the Hamming distance.

The Blowing-up lemma is the key element of the proof of the following theorem due to Shalaby and Papamarcou (see [31, Theorem 3]) which is at the origin of numerous strong converses in the zero-rate case.

**Theorem 2.7 (Shalaby and Papamarcou 94’)** *Fix  $\rho > 0$ ,  $\epsilon \in (0, 1)$ , and let  $M_n$  be a sequence of integers such that:*

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log M_n = 0. \tag{2.23}$$

Then there exists a sequence:

$$\nu_n = \nu_n(\rho, \epsilon, M_n, |\mathcal{X}|, |\mathcal{Y}|) \rightarrow 0, \tag{2.24}$$



such that for every  $Q_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  that satisfies for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ ,  $Q_{XY}(x, y) \geq \rho$ , and every  $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ ,  $C \subseteq \mathcal{X}^n$ ,  $F \subseteq \mathcal{Y}^n$  that satisfy

$$P_X^{\otimes n}(C) \geq \frac{1-\epsilon}{M_n}, \quad P_Y^{\otimes n}(F) \geq \frac{1-\epsilon}{M_n}, \quad (2.25)$$

the following is true:

$$Q_{XY}^{\otimes n}(C \times F) \geq \exp\left[-n\left(D(P_{XY} \| Q_{XY}) + \nu_n\right)\right]. \quad (2.26)$$

*Proof:* See Appendix A. ■

### 3.1 Centralized hypothesis testing

#### 3.1.1 Receiver operating characteristic (ROC)

Consider a Detector observing a sequence  $X^n \in \mathcal{X}^n$  with p.m.f. which depends on one of the two hypotheses. Under hypothesis

$$\mathcal{H} = 0: \quad \{X_t\}_{t=1}^n \text{ i.i.d. } P_X \quad (3.1)$$

under hypothesis

$$\mathcal{H} = 1: \quad \{X_t\}_{t=1}^n \text{ i.i.d. } \bar{P}_X. \quad (3.2)$$

The Detector then chooses an *acceptance region*  $\mathcal{A}_n \subseteq \mathcal{X}^n$ . It produces  $\hat{\mathcal{H}} = 0$  if  $X^n \in \mathcal{A}_n$  and  $\hat{\mathcal{H}} = 1$  otherwise. We define the following error probabilities:

$$\alpha_n := \Pr\{\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\} = P^{\otimes n}(\mathcal{A}_n^c), \quad (3.3)$$

$$\beta_n := \Pr\{\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\} = \bar{P}^{\otimes n}(\mathcal{A}_n) \quad (3.4)$$

The Neyman-Pearson lemma shows that an optimal acceptance region  $\mathcal{A}_{T,n}$  can be found via the likelihood ratio test:

$$\mathcal{A}_{T,n} = \left\{ x^n \in \mathcal{X}^n : \frac{P_X^{\otimes n}(x^n)}{\bar{P}_X^{\otimes n}(x^n)} > T \right\}, \quad (3.5)$$

where  $T$  is a positive scalar to be determined based on the desired regime of operation. Denote the probabilities of Type I and Type II error of this test by  $\alpha_{n,T}^*$  and  $\beta_{n,T}^*$ .

**Lemma 3.1 (Neymann-Pearson Lemma)** For any acceptance region  $\mathcal{A}_n \subseteq \mathcal{X}^n$  with associated probabilities of Type I and Type II error,  $\alpha_n$  and  $\beta_n$ , for all values of  $T$  the following holds:

$$\alpha_n \leq \alpha_{n,T}^* \Rightarrow \beta_n \geq \beta_{n,T}^* \quad (3.6)$$

*Proof:* see [5, Theorem 11.7.1]. ■

**Definition 3.1** For any  $\alpha_n^* \in (0;1)$ :

$$\beta_n^*(\alpha_n^*) = \min_{T: \alpha_n^* \geq \alpha_{n,T}^*} \beta_{n,T}^* \quad (3.7)$$

is the optimal receiver operating characteristic (ROC).

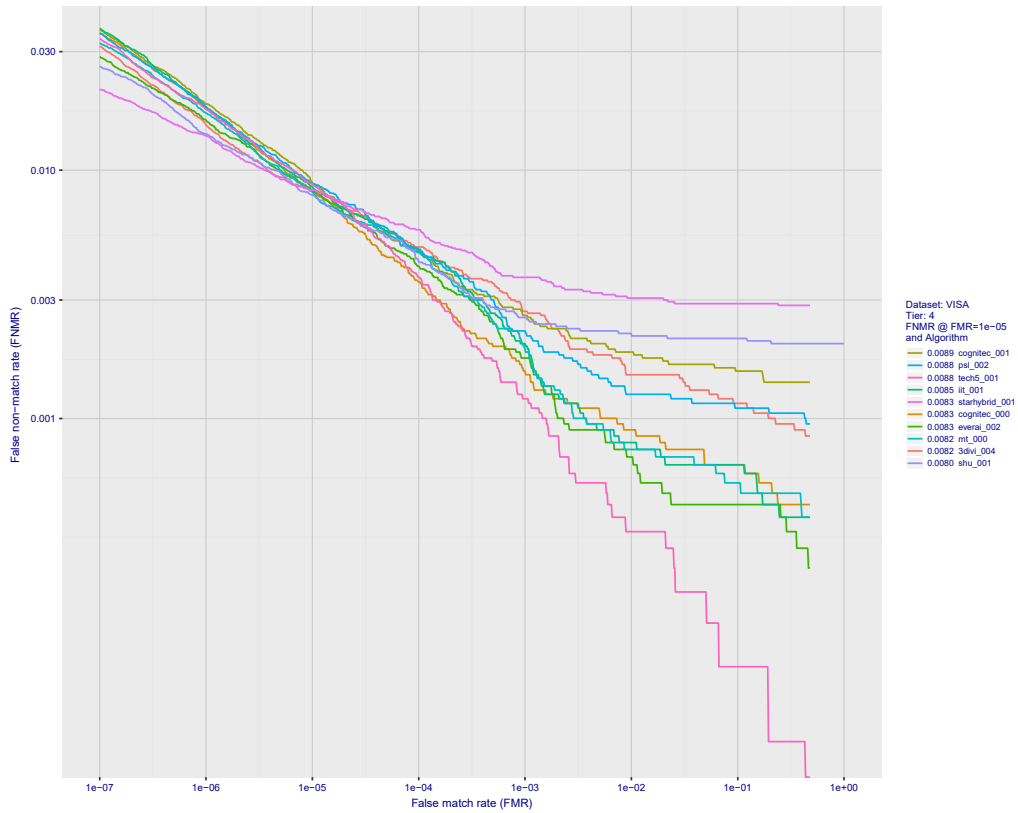


Figure 3.1: Approximation of ROCs in logarithm scales (also known as Detection Error Tradeoff (DET characteristics) for various algorithms presented at FRVT [24]. The approximation is done using a base of labeled face images. The False non-match rate corresponds to the ratio of matching faces pairs (two face images of the same person) in the base that where classified as being those of distinct persons. The False match rate corresponds to the ratio of non-matching faces pairs (two face images of distinct persons) in the base that where classified as being those of the same person.

**Remark 3.1** Lemma 3.1 states that the likelihood ratio tests ensures the best Type II error probability for a given Type I error probability. It does not give a closed form of the optimal ROC. Actually such a result

is hard to obtain in general. Almost all detection algorithms enable to make vary the tradeoff between the Type I error and the Type II error and so they are also characterized by a theoretical ROC. As for the optimal ROC, it seems difficult to obtain a closed form for them. Researchers compute an approximation of this ROCs. Here we represent some face recognition algorithms ROCs extracted from the NIST report on the ongoing Face Recognition Vendor Test [24] in Figure 3.1.

### 3.1.2 Error exponent

As stated in Remark 3.1, the closed form of the ROC curve is not known in general for a finite  $n$ . Noticeable work in that direction is [35] which provides approximation to this ROC. In a regime where the number of observations increases infinitely, the optimal probabilities of the Type I error and Type II error decay to 0 and the compromise between these two quantities vanishes. However, there is always a compromise between their decay rate. This tradeoff has been studied in [17] and [4]. In what follows we limit ourselves to the study of the case where only the Type II error is restricted to decay exponentially fast because the problem, when it is transposed to the distributed case, is of formidable complexity and somehow contains a major difficulty to be solved before solving the general case. In the asymptotic regime as  $n \rightarrow \infty$  for any fixed bound  $\epsilon > 0$  on the Type I error probability  $\alpha_n$ , it is possible to let the Type II error probability  $\beta_n$  decrease exponentially fast to 0 as  $n \rightarrow \infty$ . We have a closer look and analyze the exponential speed of this convergence.

**Definition 3.2 (Achievability for the centralized setup)** Fix  $\epsilon \in (0, \frac{1}{2})$ , an error-exponent  $\theta$  is achievable if for all blocklengths  $n$  there exists an acceptance region  $\mathcal{A}_n \in \mathcal{X}^n$  so that the corresponding Type I and Type II error probabilities  $\alpha_n$  and  $\beta_n$  satisfy:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \leq \epsilon, \quad (3.8)$$

and

$$\theta \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n. \quad (3.9)$$

We introduce the Chernoff-Stein lemma, which characterizes it.

**Lemma 3.2 (Chernoff-Stein lemma)** Fix  $\epsilon \in (0, 1)$  and define  $E(\epsilon)$  the supremum of the set of all achievable error-exponents. Then:

$$E(\epsilon) = D(P_X || \bar{P}_X). \quad (3.10)$$

*Proof:* see [5, Theorem 11.8.3] ■

**Remark 3.2** Chernoff-Stein lemma is an asymptotic first-order approximation of the optimal ROC curve. It provides an approximation of the decay rate of the Type II error with the number of observations. Note that

the error exponent does not depend on the Type I error. It is a strong converse result in the sense that there is a sharp threshold between the achievable exponents and the others.

### 3.2 Hypothesis testing with communication constraints

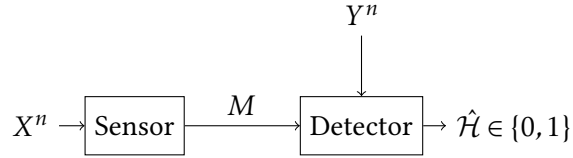


Figure 3.2: Binary hypothesis testing with one sided data compression.

Consider a two-terminal problem with a Sensor observing the sequence  $X^n$  and a detector observing  $Y^n$  (see figure 3.2). Here  $(X_i, Y_i)$  take values in the alphabet  $\mathcal{X} \times \mathcal{Y}$ . The p.m.f. of the tuple  $(X^n, Y^n)$  depends on one of two hypotheses. Under hypothesis

$$\mathcal{H} = 0: \quad \{(X_t, Y_t)\}_{t=1}^n \text{ i.i.d. } P_{XY} \quad (3.11)$$

and under hypothesis

$$\mathcal{H} = 1: \quad \{(X_t, Y_t)\}_{t=1}^n \text{ i.i.d. } \bar{P}_{XY} \quad (3.12)$$

The Sensor applies an encoding function

$$\phi_n: \mathcal{X}^n \rightarrow \mathcal{M} \triangleq \{0, 1, \dots, W_n - 1\} \quad (3.13)$$

to its observed source sequence  $X^n$  and sends the resulting index

$$M = \phi_n(X^n) \quad (3.14)$$

to the Detector. The Detector then applies a decision function to the pair  $(M, Y^n)$ :

$$\psi_n: \mathcal{M} \times \mathcal{Y}^n \rightarrow \{0, 1\}. \quad (3.15)$$

to decide on the hypothesis

$$\hat{\mathcal{H}} \triangleq \psi_n(M, Y^n). \quad (3.16)$$

We define the following error probabilities:

$$\alpha_n := \Pr\{\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\}, \quad (3.17)$$

$$\beta_n := \Pr\{\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\}. \quad (3.18)$$

**Definition 3.3 (Achievability under Rate-Constraints)** Given rate  $R \geq 0$ , an error-exponent  $\theta$  is said achievable if for all blocklengths  $n$  there exist functions  $\phi_n$  and  $\psi_n$  as in Equation (3.13), and Equation (3.15) so that the following limits hold:

$$\lim_{n \rightarrow \infty} \alpha_n = 0, \quad (3.19)$$

$$\theta \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n, \quad (3.20)$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_n \leq R. \quad (3.21)$$

**Definition 3.4 (Exponent-rate function)** For any rate  $R \geq 0$ , define the exponent-rate function  $E(R)$  as the supremum of all achievable error-exponents.

### 3.2.1 The work of Ahlswede and Csiszár

The first record on the subject of distributed hypothesis testing known by the author is [1]. In this work Ahlswede and Csiszár considered the scenario depicted in Figure 3.2. When under hypothesis

$$\mathcal{H} = 0 : \{(X_t, Y_t)\}_{t=1}^n \text{ i.i.d. } P_{XY} \quad (3.22)$$

and under hypothesis

$$\mathcal{H} = 1 : \{(X_t, Y_t)\}_{t=1}^n \text{ i.i.d. } \bar{P}_{XY} = P_X \times P_Y, \quad (3.23)$$

they derived a single-letter expression for the rate-exponent function denoted in this case  $E_I(R)$ :

**Theorem 3.3 (Exponent-rate function when testing against independence)** For every  $R > 0$ , when  $\bar{P}_{XY} = P_X \times P_Y$ :

$$E_I(R) = \max_{\substack{U: I(U; X) \leq R \\ U \text{---} X \text{---} Y \\ |\mathcal{U}| \leq |\mathcal{X}| + 1}} I(U; Y). \quad (3.24)$$

This case is known as testing against independence. The authors also gave a general lower bound for any binary hypothesis testing in Section 3.2:

**Theorem 3.4 (Lower bound for the exponent-rate function in the general case)** For every  $R > 0$ :

$$E(R) \geq D(P_X \| \bar{P}_X) + \max_{\substack{U: I(U; X) \leq R \\ U \text{---} X \text{---} Y}} D(P_{UY} \| \bar{P}_{UY}). \quad (3.25)$$

with  $\bar{P}_{UY} = P_{U|X} \bar{P}_{XY}$ .

Some basic properties of the exponent-rate function were also derived in [1].

**Proposition 3.5** The exponent-rate function  $E(R)$  is monotonically increasing and concave for  $R \geq 0$  and continuous for positive  $R$ .

### 3.2.2 No binning: Han's Scheme

Han presented a similar scheme which is optimal in the case of testing against independence and improves the result of Ahlswede and Csiszár in the general case (see [14]).

**Theorem 3.6 (Lower bound for the exponent-rate function in the general case)** *For every  $R > 0$ :*

$$E(R) \geq E_H(R) \triangleq \max_{\substack{U: I(U; X) \leq R \\ U \text{---} X \text{---} Y \\ |\mathcal{U}| \leq |\mathcal{X}| + 1}} \min_{\substack{\bar{P}_{UXY}: \bar{P}_{UX} = P_{UX} \\ \bar{P}_{UY} = P_{UY}}} D(P_{UXY} \| \bar{P}_{UXY}). \quad (3.26)$$

with  $\bar{P}_{UXY} = P_{U|X} \bar{P}_{XY}$ .

**Remark 3.3** *It appears that the inner bound of Theorem 5.4 coincides with the exponent-rate function in the case of testing against independence and improve the inner bound presented in Theorem 3.4 in the general case.*

**Remark 3.4** *We consider the (trivial) extreme case where the link is of high rate so that under hypothesis  $\mathcal{H} = 0$ , the Detector can obtain the sequence  $X^n$  with high probability. In this case the Detector attains the error-exponent of a centralized setup where it observes  $(X^n, Y^n)$ . And so as long as  $R \geq H(X)$ , Lemma 3.2 applied and the exponent-rate function verifies*

$$E(R) = E_C \triangleq D(P_{XY} \| \bar{P}_{XY}). \quad (3.27)$$

*In addition Lemma 3.2 enables us to obtain the following (trivial) upper bound: for every  $R \geq 0$ :*

$$E(R) \leq E_C. \quad (3.28)$$

### 3.2.3 Role of binning: Shimokawa, Han and Amari scheme

The scheme used in [1, 14] is based on classic source coding arguments. In [32], these arguments are completed using binning. In the former contributions the Sensor sends the index of the compression codeword. In the later the Sensor sends the index of a chosen codeword's bin and the decoder uses its side-information to retrieve the chosen codeword in the bin. This is similar to the Wyner-Ziv source coding scheme [42]. Nonetheless in standard source coding a typicality test is used to retrieve the chosen codeword. This requires knowledge of the joint distribution between the codeword sent by the sensor and the side information. But in case of distributed hypothesis testing this joint law is not known as it can be either  $P_{XY}$  or  $\bar{P}_{XY}$  depending on whether  $\mathcal{H} = 0$  or  $\mathcal{H} = 1$ . In [32], a minimum entropy decoder strategy is used to overcome this uncertainty. The sensor searches for a unique codeword that minimizes the empirical joint entropy with the side-information. This scheme improves the exponent of [14].

**Theorem 3.7 (Shimokawa Han and Amari 87')** *For every  $R > 0$ :*

$$E(R) \geq E_{\text{SHA}}(R) \triangleq \max_{\substack{U: I(U; X|Y) \leq R \\ |\mathcal{U}| \leq |\mathcal{X}| + 2 \\ U \text{---} X \text{---} Y}} \min \{E_1(P_{U|X}), E_2(R, P_{U|X})\} \quad (3.29)$$

where:

$$\begin{aligned}
 E_1(P_{U|X}) &\triangleq \min_{\substack{\tilde{P}_{UXY}: \tilde{P}_{UX}=P_{UX} \\ \tilde{P}_{UY}=P_{UY}}} D(\tilde{P}_{UXY} \| \tilde{P}_{UXY}), & (3.30) \\
 E_2(R, P_{U|X}) &\triangleq \begin{cases} \infty & \text{if } I(U; X) \geq R \\ \min_{\substack{\tilde{P}_{UXY}: \tilde{P}_{UX}=P_{UX} \\ \tilde{P}_Y=P_Y, H(\tilde{P}_{UY}) \leq H(P_{UY})}} D(P_{UXY} \| \tilde{P}_{UXY}) + [R - I(U; X|Y)]_+ & \text{otherwise.} \end{cases}
 \end{aligned}$$

Here  $\tilde{P}_{UXY} = P_{U|X} \tilde{P}_{XY}$

*Proof:* see [32]. ■

### 3.2.4 Exponent-rate function when testing against conditional independence.

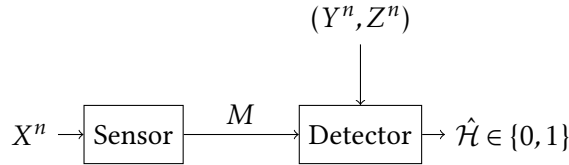


Figure 3.3: Binary hypothesis testing with one-sided data compression and two stream of side information at the Detector.

In [25], a scheme using binning is shown to be optimal for testing against conditional independence defined in the following. Consider the same two-terminal problem as in the last section. The Sensor observes the sequence  $X^n$  and the Detector the sequences  $Y^n$  and  $Z^n$  ( see Figure 3.2.4). Here  $(X_i, Y_i, Z_i)$  take values in the alphabet  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ . The joint p.m.f. of the tuple  $(X^n, Y^n, Z^n)$  depends on the hypothesis. Under hypothesis

$$\mathcal{H} = 0: \quad \{(X_t, Y_t, Z_t)\}_{t=1}^n \text{ i.i.d. } P_{XYZ} \quad (3.31)$$

and under hypothesis

$$\mathcal{H} = 1: \quad \{(X_t, Y_t, Z_t)\}_{t=1}^n \text{ i.i.d. } \tilde{P}_{XYZ} = P_Z P_{X|Z} P_{Y|Z}. \quad (3.32)$$

The sensor behaves as in the previous section (see 3.13 and (3.14)). The Detector applies the decision function

$$\psi_n: \mathcal{M} \times \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \{0, 1\}. \quad (3.33)$$

to decide on the hypothesis

$$\hat{\mathcal{H}} \triangleq \psi_n(M, Y^n, Z^n). \quad (3.34)$$

The exponent-rate function is defined similarly to before and denoted  $E_{CI}(R)$ . We have the following theorem:



**Theorem 3.8 (Exponent-rate function when testing against conditional independence)** For every  $R > 0$ :

$$E_{\text{CI}}(R) = \max_{\substack{U: I(U; X|Z) \leq R \\ U \leftrightarrow X \leftrightarrow Z \\ |\mathcal{U}| \leq |\mathcal{X}| + 1}} I(U; Y|Z). \quad (3.35)$$

*Proof:* See [25, Theorem 3]. ■

**Remark 3.5** In [25] this result was obtained using a distinct scheme from the one of Shimokawa, Han and Amari in [32]. As the joint p.m.f  $P_{XZ}$  and  $\bar{P}_{XZ}$  are equal, it is possible for this problem to use a “classic” binning strategy where the Detector uses its side information  $Z^n$  to retrieve the codeword chosen by the Sensor in the bin which is indexed by the message it receives. Then it tests if the codeword decoded is typical with its side information given the p.m.f. of  $\mathcal{H} = 0$ . This scheme is shown to achieve the optimal exponent-rate function. Nonetheless in the same contribution the scheme used in [32] is also shown to be optimal for this very same scenario.

### 3.2.5 Some examples

**Example 3.1 (Testing against independence)** Consider a setup where  $X, Y$  are binary with pmfs

$$\begin{cases} P_{XY}(0,0) = 0.4 & P_{XY}(0,1) = 0.1 \\ P_{XY}(1,0) = 0.1 & P_{XY}(1,1) = 0.4 \end{cases},$$

and  $\bar{P}_{XY} = P_X P_Y$ . In that case the optimal exponent-rate function is given in Theorem 3.3 and plotted in Figure 3.4. On the same figure we plotted the error-exponent obtained for the centralized setup.

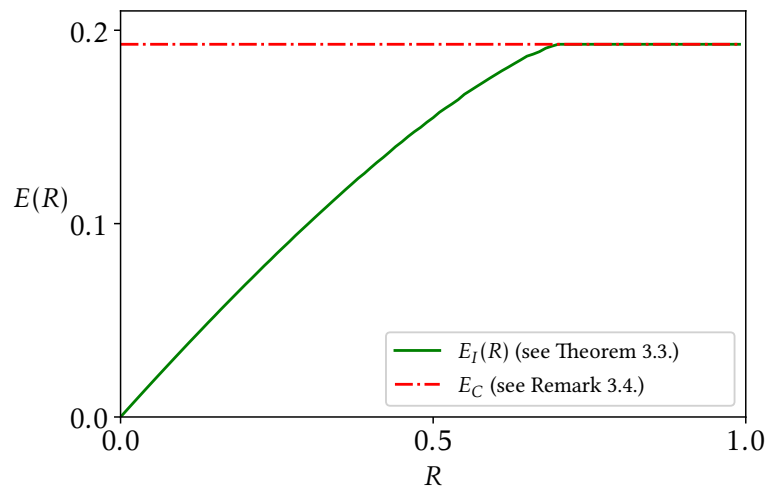


Figure 3.4: Optimal exponent-rate function given by Theorem 3.3 for Example 3.1 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.1.

**Example 3.2 (Case where binning is beneficial)** Consider a setup where  $X, Y$  are binary with pmfs

$$\begin{cases} P_{XY}(0,0) = 0.45 & P_{XY}(0,1) = 0.05 \\ P_{XY}(1,0) = 0.05 & P_{XY}(1,1) = 0.45 \end{cases}, \quad \begin{cases} \bar{P}_{XY}(0,0) = 0.4 & \bar{P}_{XY}(0,1) = 0.1 \\ \bar{P}_{XY}(1,0) = 0.1 & \bar{P}_{XY}(1,1) = 0.4 \end{cases}. \quad (3.36)$$

On Figure 3.5 we represent achievable error-exponent given by Theorem 3.6 and Theorem 3.7. In that case binning is beneficial and for a given rate  $R \geq 0$ , the error-exponent obtained with Theorem 3.7 is always larger than the one obtained with Theorem 3.6. On the same figure we plotted in red the exponent obtained for the centralized setup given in Equation (3.10).

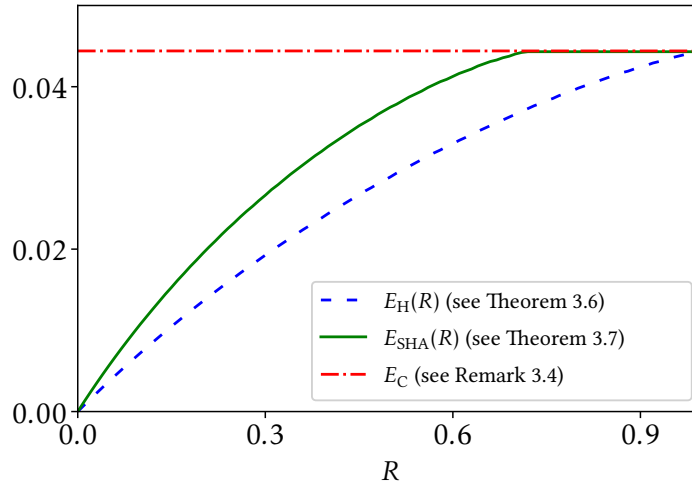


Figure 3.5: Lower bounds for the exponent-rate function given by Theorems 3.6 and 3.7 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.2.

**Example 3.3 (Testing against conditional independence)** Consider a setup as in Figure 3.2.4 where  $X, Y, Z$  are binary with p.m.f.s

$$\begin{cases} P_{XYZ}(0,0,0) = 1.e-4 & P_{XYZ}(0,0,1) = 5.e-2 & P_{XYZ}(0,1,0) = 4.99e-2 & P_{XYZ}(0,1,1) = 3.e-1 \\ P_{XYZ}(1,0,0) = 3.e-1 & P_{XYZ}(1,0,1) = 2.5e-1 & P_{XYZ}(1,1,0) = 2.5e-2 & P_{XYZ}(1,1,1) = 2.5e-2 \end{cases}$$

and  $\bar{P}_{XYZ} = P_Z P_{X|Z} P_{Y|Z}$ . On Figure 3.6 we represent the achievable error-exponent given by Theorem 3.6 and Theorem 3.7. As shown in Theorem 3.8 the error-exponent of Theorem 3.7 is optimal. On the same figure we plotted in red the error-exponent obtained for the centralized setup given in Equation (3.10).

### 3.2.6 Zero-rate hypothesis testing

Zero rate hypothesis testing was introduced in [14], which studied the general distributed binary hypothesis testing problem and determined the maximal error-exponent when the Sensor can send only one bit, i.e. the function  $\phi_n$  defined in 3.13 is restricted to

$$W_n = 2. \quad (3.37)$$

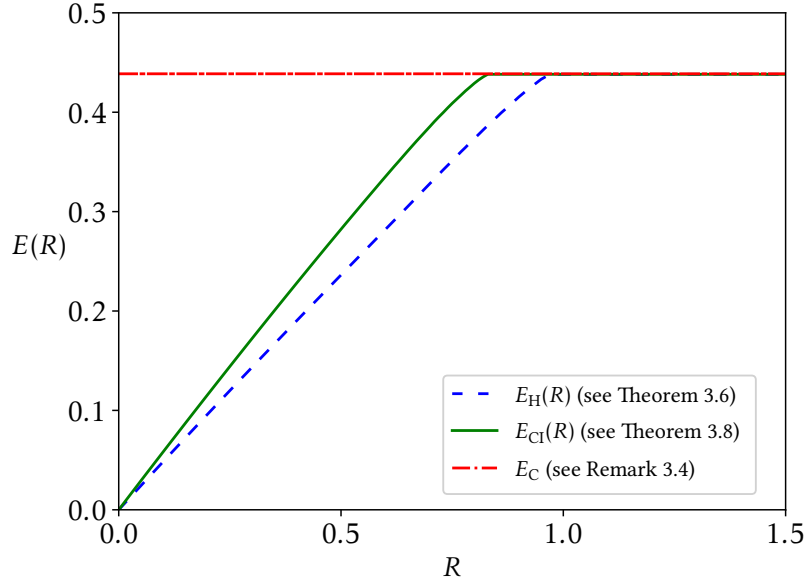


Figure 3.6: Lower bounds for the exponent-rate function given by Theorems 3.6 and 3.8 and the trivial upper bound for the exponent-rate function of Remark 3.4 for Example 3.3.

In [31] this result is extended to any function  $\phi_n$  satisfying

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log W_n = 0, \quad (3.38)$$

which corresponds to the case where only a sublinear number of bits are sent.

**Definition 3.5 (Achievability for zero rate compression)** For  $\epsilon \in (0, 1)$ , an error-exponent  $\theta$  is said achievable under zero rate constraint, if for all blocklengths  $n$  there exist functions  $\phi_n$  and  $\psi_n$  as in (3.13) and (3.15) so that (3.38) holds and

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \leq \epsilon \quad (3.39a)$$

$$\underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n \geq \theta. \quad (3.39b)$$

If  $W_n = 2$ , see (3.37), we say that  $\theta$  is achievable under complete compression.

**Definition 3.6** The supremum of all achievable error-exponent for sublinear number of bits is called the maximum error-exponent for zero rate compression and denoted  $E^*(\epsilon)$ . For  $W_n = 2$  it is called the maximum error-exponent under complete compression and it is denoted  $E_2(\epsilon)$ .

For this scenario we have this result:

**Theorem 3.9 (Maximal error-exponent for zero rate compression)** Let  $\bar{P}_{XY}(x, y) > 0$ , for all  $(x, y) \in \mathcal{X} \times \mathcal{Y}$ . Then for  $\epsilon \in (0, 1)$ :

$$E^*(\epsilon) = E_2(\epsilon) = \min_{\substack{\bar{P}_{XY}: \bar{P}_X = P_X, \\ \bar{P}_Y = P_Y}} D(\bar{P}_{XY} \| \bar{P}_{XY}). \quad (3.40)$$

*Proof:* For the achievability see [14, Theorem 5] and for the converse see [31, Theorem 1]. ■

### 3.2.7 Model with cooperation

In [47] a three-terminal hypothesis testing problem is considered with a Sensor 1 observing the sequence  $X^n$ , a Sensor 2 observing  $Y_1^n$ , and a Detector observing  $Y_2^n$  (see Figure 3.7). The joint probability mass function (p.m.f.) of the tuple  $(X^n, Y_1^n, Y_2^n)$  depends on one of two hypotheses

$$\mathcal{H} = 0: \quad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } P_{XY_1Y_2} \quad (3.41)$$

and under hypothesis

$$\mathcal{H} = 1: \quad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } \bar{P}_{XY_1Y_2}. \quad (3.42)$$

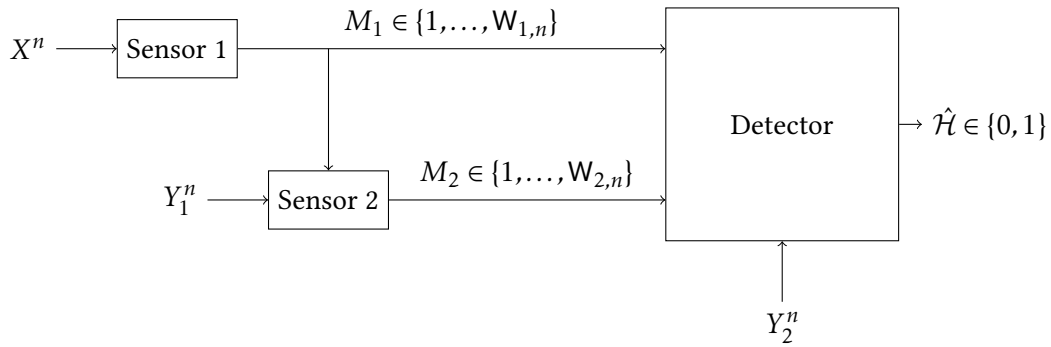


Figure 3.7: Binary hypothesis testing in a cooperative MAC studied in [47].

Sensor 1 applies an encoding function

$$\phi_{1,n}: \mathcal{X}^n \rightarrow \mathcal{M}_1 \triangleq \{0, 1, \dots, W_{1,n} - 1\} \quad (3.43)$$

to its observed source sequence  $X^n$  and sends the resulting index

$$M_1 = \phi_{1,n}(X^n) \quad (3.44)$$

to Sensor 2 and the Detector. Sensor 2 applies an encoding function

$$\phi_{2,n}: \mathcal{Y}_1^n \times \mathcal{M}_1 \rightarrow \mathcal{M}_2 \triangleq \{0, 1, \dots, W_{2,n} - 1\} \quad (3.45)$$

to its observed source sequence  $Y_1^n$  and sends the resulting index

$$M_2 = \phi_{2,n}(Y_1^n, M_1) \quad (3.46)$$

to the Detector. The Detector then applies a decision function to the triple  $(M_1, M_2, Y_2^n)$ :

$$\psi_n: \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Y}_2^n \rightarrow \{0, 1\}. \quad (3.47)$$

to decide on the hypothesis

$$\hat{\mathcal{H}} = \psi_n(M_1, M_2, Y_2^n). \quad (3.48)$$

We define the following error probabilities:

$$\alpha_n := \Pr\{\hat{\mathcal{H}} = 1 | \mathcal{H} = 0\}, \quad (3.49)$$

$$\beta_n := \Pr\{\hat{\mathcal{H}} = 0 | \mathcal{H} = 1\}. \quad (3.50)$$

**Definition 3.7 (Achievability under rate-constraints for model in Figure 3.7)** *Given rates  $R_1 \geq 0$ ,  $R_2 \geq 0$  and small positive number  $\epsilon \in (0, 1)$  an error-exponent  $\theta$  is said achievable for this model if for all blocklengths  $n$  there exist functions  $\phi_{1,n}$ ,  $\phi_{2,n}$  and  $\psi_n$  as in Equation (3.43), Equation (3.45), and Equation (3.47) so that the following limits hold:*

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n \leq \epsilon, \quad (3.51)$$

$$\theta \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n, \quad (3.52)$$

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_{1,n} \leq R_1, \quad (3.53)$$

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_{2,n} \leq R_2. \quad (3.54)$$

**Definition 3.8 (Exponent-rate function)** *For any rate  $R_1 \geq 0$ ,  $R_2 \geq 0$ , let  $E_{\text{ZL}}(R_1, R_2)$  be the supremum all achievable error-exponents.*

For given rates  $R_1 \geq 0$  and  $R_2 \geq 0$ , define the following set of auxiliary random variables:

$$\mathcal{S}_{\text{ZL}} = \left\{ (U_1, U_2); \begin{array}{l} U_1 \text{ --- } X \text{ --- } (Y_1, Y_2) \\ U_2 \text{ --- } Y_1 \text{ --- } (X, Y_2) \\ I(U_1; X) \leq R_1 \\ I(U_2; Y_1 | U_1) \leq R_2 \\ |\mathcal{U}_1| \leq |\mathcal{X}| + 1 \\ |\mathcal{U}_2| \leq |\mathcal{Y}_1| |\mathcal{U}_1| + 1 \end{array} \right\} \quad (3.55)$$

and for any pair of auxiliary random variable  $(U_1, U_2) \in \mathcal{S}_{\text{ZL}}$ , let

$$\mathcal{L} = \left\{ \begin{array}{l} \tilde{P}_{U_1 X} = P_{U_1 X} \\ \tilde{P}_{U_1 U_2 X Y_1 Y_2} : \tilde{P}_{U_1 U_2 Y_1} = P_{U_1 U_2 Y_1} \\ \tilde{P}_{U_1 U_2 Y_2} = P_{U_1 U_2 Y_2} \end{array} \right\}. \quad (3.56)$$

A lower bound to the exponent-rate function is derived for this setup. It is optimal when in addition:

$$\tilde{P}_{X Y_1 Y_2} = P_{X Y_1} P_{Y_2}. \quad (3.57)$$

**Theorem 3.10 (Zhao and Lai)** Then for  $R_1 \geq 0$  and  $R_2 \geq 0$ , the exponent-rate function for this problem,  $E_{\text{ZL}}(R_1, R_2)$ , satisfies

$$E_{\text{ZL}}(R_1, R_2) \geq \max_{(U_1, U_2) \in \mathcal{S}_{\text{ZL}}} \min_{\tilde{P}_{U_1 U_2 X Y_1 Y_2} \in \mathcal{L}} D(\tilde{P}_{U_1 U_2 X Y_1 Y_2} \| P_{U_1|X} P_{U_2|U_1 Y_1} Q_{X Y_1 Y_2}). \quad (3.58)$$

If in addition (3.57) holds, then

$$E_{\text{ZL}}(R_1, R_2) = \max_{(U_1, U_2) \in \mathcal{S}_{\text{ZL}}} I(U_2 U_1; Y_2). \quad (3.59)$$

*Proof:* see [47]. ■

### 3.2.8 Model with multiple detectors

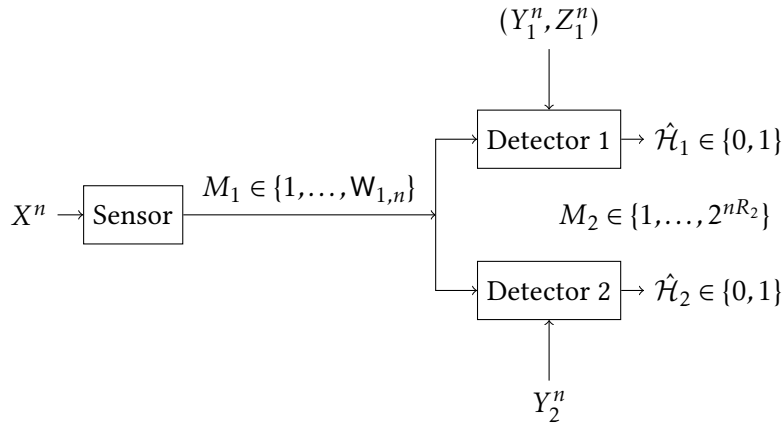


Figure 3.8: Binary hypothesis testing in a Heegard-Berger network studied in [28].

In [28] a three-terminal problem with a Sensor observing the sequence  $X^n$ , a Detector 1 observing  $(Y_1^n, Z_1^n)$ , and a Detector 2 observing  $(Y_2^n, Z_2^n)$  as in Figure 3.8 is considered. The joint probability mass function (p.m.f.) of the tuple  $(X^n, Y_1^n, Z_1^n, Y_2^n, Z_2^n)$  depends on one of two hypotheses. Under hypothesis

$$\mathcal{H} = 0: \quad \{(X_t, Y_{1,t}, Z_{1,t}, Y_{2,t}, Z_{2,t})\}_{t=1}^n \text{ i.i.d. } P_{X Y_1 Z_1 Y_2} \quad (3.60)$$

and under hypothesis

$$\mathcal{H} = 1: \quad \{(X_t, Y_{1,t}, Z_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } P_{X Z_1 Z_2} P_{Y_1|Z_1} P_{Y_2}. \quad (3.61)$$

The Sensor applies an encoding function as in (3.43) to its observed source sequence  $X^n$  and sends the resulting index

$$M_1 = \phi_{1,n}(X^n) \quad (3.62)$$

to Detector 1 and Detector 2. Detector 1 applies a decision function to the triple  $(M_1, Y_1^n, Z_1^n)$ :

$$\psi_{1,n}: \mathcal{M}_1 \times \mathcal{Y}_1^n \times \mathcal{Z}_1^n \rightarrow \{0, 1\}. \quad (3.63)$$

to decide on the hypothesis

$$\hat{\mathcal{H}}_1 = \psi_{1,n}(M_1, Y_1^n, Z_1^n). \quad (3.64)$$

and The Detector 2 applies a decision function to the pair  $(M_1, Y_2^n)$ :

$$\psi_{2,n}: \mathcal{M}_1 \times \mathcal{Y}_2^n \times \rightarrow \{0, 1\}. \quad (3.65)$$

to decide on the hypothesis

$$\hat{\mathcal{H}}_2 = \psi_{2,n}(M_1, Y_2^n). \quad (3.66)$$

We define the following error probabilities:

$$\alpha_{1,n} := \Pr\{\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0\}, \quad \alpha_{2,n} := \Pr\{\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0\}, \quad (3.67)$$

$$\beta_{1,n} := \Pr\{\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1\}, \quad \beta_{2,n} := \Pr\{\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1\}. \quad (3.68)$$

**Definition 3.9 (Achievability under rate-constraints)** *Given rate  $R_1 \geq 0$ , a pair of error-exponents  $(\theta_1, \theta_2)$  is said achievable for this model if for all blocklengths  $n$  there exist functions  $\phi_{1,n}$ ,  $\psi_{1,n}$  and  $\psi_{2,n}$  as in (3.43), (3.63), and (3.65) so that the following limits hold:*

$$\lim_{n \rightarrow \infty} \alpha_{1,n} = 0, \quad \lim_{n \rightarrow \infty} \alpha_{2,n} = 0, \quad (3.69)$$

$$\theta_1 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{1,n}, \quad \theta_2 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{2,n}, \quad (3.70)$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_{1,n} \leq R_1. \quad (3.71)$$

**Definition 3.10 (Error-Exponents Region under Rate-Constraint)** *For rate  $R_1 \geq 0$  the closure of the set of all achievable exponent pairs  $(\theta_1, \theta_2)$  is called the error-exponents region  $\mathcal{E}_{\text{SWT}}(R_1)$ .*

**Definition 3.11** *We say that  $Z_1$  is less noisy than  $Y_2$ , if for all auxiliary random variables  $U$  satisfying the Markov Chain  $U \text{---} X \text{---} (Y_1, Z_1, Y_2)$  the following inequality holds:*

$$I(U; Z_1) \geq I(U; Y_2). \quad (3.72)$$

Let

$$\mathcal{E}_{\text{SWT}}^{\text{in}} \triangleq \bigcup_{\substack{R_1 \geq I(U_0; X) + I(U_1; X | Z_1), \\ (U_0, U_1) \text{---} X \text{---} (Y_1, Z_1, Y_2)}} \left\{ (\theta_1, \theta_2) : \begin{array}{l} \theta_1 \leq I(U_1, U_0; Y_1 | Z_1), \\ \theta_2 \leq I(U_2, U_0; Y_2 | Z_2), \theta_1 \geq 0, \quad \theta_2 \geq 0 \end{array} \right\} \quad (3.73)$$

**Theorem 3.11 (Salehkalaibar, Wigger and Timo)** *Assume (3.61), then for  $R_1 \geq 0$ , then*

$$\mathcal{E}_{\text{SWT}}^{\text{in}} \subseteq \mathcal{E}_{\text{SWT}}(R_1). \quad (3.74)$$

*If in addition  $Z_1$  is less noisy than  $Y_2$ , then*

$$\mathcal{E}_{\text{SWT}}^{\text{in}} = \mathcal{E}_{\text{SWT}}(R_1). \quad (3.75)$$

*Proof:* see [28]. ■

### 3.3 Other extensions to networks and other lines of works

As mentioned before the first records on the subject are [1] and [14]. A variant of the problem presented in [15] where Type I errors are constrained to decrease exponentially fast with the number of observations has been studied in [15, 36]. In these studies, the Type II error exponents are characterized in a scenario where two sensors communicate with one detector, one at zero rate and the other with an arbitrary communication rate. In the same articles achievability results are obtained also for positive communication rates. A general review on statistical inference is also provided in [36]. In particular, it presents an approach to the problem based on the geometry of information. This approach is introduced more specifically in [2] and used in [40] to study the finite length scenario.

The distributed hypothesis testing problem has connections to the lossy source coding. In particular, the tools used to characterize the reliability function in source coding can be used also to characterize achievable type II exponent in distributed hypothesis testing (see e.g. [20]). In source coding many types of networks have been studied: several sensors [3, 38], several decoders [16], interactive coding [18]. Similarly, recent works explored these networks for hypothesis testing: with multiple sensors [14, 25, 45, 44, 39] with multiple detectors [28], with interactive terminals [37, 43, 19] and with multi-hop networks [46, 48, 41, 29, 9, 44]. In any case, it appears that many strategies used in source coding make it possible to transmit compress data efficiently and that, when properly completed, these strategies make it possible to obtain a certain number of general achievability results. The optimal coding scheme for the distributed source coding problem is also optimal for the distributed testing against independence problem (see [25, 44]). Other problems such as distributed hypothesis over noisy channels [27, 33] and scenarios with privacy constraints [23, 22, 34, 13] have been considered.

Our work mainly focus on a single sensor multiple detectors network where detectors may cooperate [9, 10, 11]:

- P. Escamilla, M. Wigger, and A. Zaidi. “Distributed hypothesis testing with concurrent detections”. *2018 IEEE International Symposium on Information Theory (ISIT’18)*. June 2018, pp. 166–170. DOI: 10.1109/ISIT.2018.8437906
- P. Escamilla, A. Zaidi, and M. Wigger. “Distributed hypothesis testing with collaborative detection”. *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Oct. 2018, pp. 512–518. DOI: 10.1109/ALLERTON.2018.8635828
- P. Escamilla, M. Wigger, and A. Zaidi. “Distributed hypothesis testing: cooperation and concurrent detection”. *revision for publication in the IEEE Transactions of Information Theory* (2019)

As we will see in more detail, the setups in [41, 28] and [47] are special cases of our model, and our new results recover these previous results as special cases.





---

Zero-rate distributed hypothesis testing

---

Based on our work published in [9, 10, 11], we focus on the case of zero-rate communication in the multiterminal binary hypothesis testing scenario shown in Figure 4.1. In this chapter communications are restricted to a sublinear number of bits:

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_{1,n} \leq W_1, \quad \text{and} \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_{2,n} \leq W_2.$$

We investigate the setups where both detectors aim at maximizing the error exponent under the same hypothesis (we refer to this setup as *coherent detection*) or the setup where the two detectors aim at maximizing their exponent under different hypotheses (we refer to this setup as *concurrent detection*).

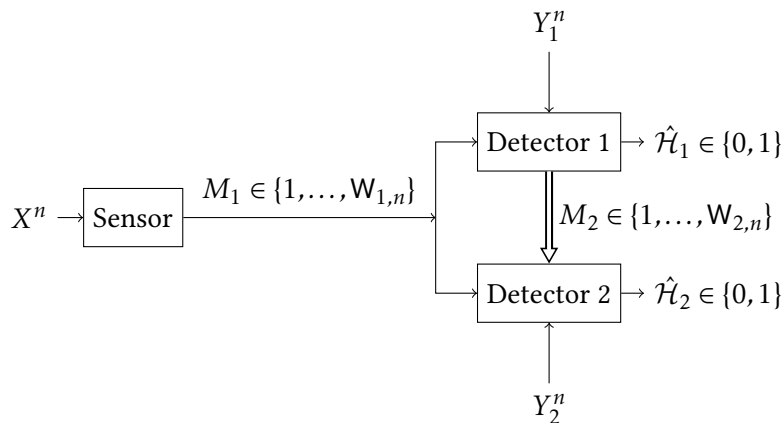


Figure 4.1: A Heegard-Berger type source coding model with unidirectional conferencing for multiterminal hypothesis testing.

For simplicity, we assume that  $P_{XY_1}(x, y_1) > 0$  and  $\bar{P}_{XY_1Y_2}(x, y_1, y_2) > 0$  for all  $(x, y_1, y_2) \in \mathcal{X}_1 \times \mathcal{Y}_2 \times \mathcal{Y}_2$ . The

main result in this chapter is the exact characterization of the error-exponents region for both setups of coherent and concurrent detection. We first introduce the problem formally in section 4.1, then we consider the case of coherent and concurrent detection when  $P_X = \bar{P}_X$  in section 4.2. Under this assumption we show that the two detectors behave as if they were alone in the network to perform the test. The case of coherent and concurrent detection when  $P_X \neq \bar{P}_X$  is discussed in section 4.3. In contrast of the previous case, a trade-off between the two optimal exponents appears. We then illustrate the two last sections with numerical examples in section 4.4. Finally we discuss the benefit of cooperation with zero rate communication in section 4.5.

## 4.1 Formal Problem Statement

Consider a three-terminal problem with a Sensor observing the sequence  $X^n$ , a Detector 1 observing  $Y_1^n$ , and a Detector 2 observing  $Y_2^n$ . Under hypothesis

$$\mathcal{H} = 0: \quad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } P_{XY_1Y_2} \quad (4.1)$$

and under hypothesis

$$\mathcal{H} = 1: \quad \{(X_t, Y_{1,t}, Y_{2,t})\}_{t=1}^n \text{ i.i.d. } \bar{P}_{XY_1Y_2} \quad (4.2)$$

The Sensor applies an encoding function  $\phi_1 : \mathcal{X} \rightarrow \mathcal{M}_1$  as in (3.43) to its observed source sequence  $X^n$  and sends the resulting index

$$M_1 = \phi_{1,n}(X^n) \quad (4.3)$$

to both decoders. Detector 1 then applies two functions to the pair  $(M_1, Y_1^n)$ , an encoding function:

$$\phi_{2,n}: \mathcal{M}_1 \times \mathcal{Y}_1^n \rightarrow \mathcal{M}_2 \triangleq \{0, 1, \dots, W_{2,n} - 1\}, \quad (4.4)$$

and a decision function

$$\psi_{1,n}: \mathcal{M}_1 \times \mathcal{Y}_1^n \rightarrow \{0, 1\}. \quad (4.5)$$

It sends the index

$$M_2 = \phi_{2,n}(M_1, Y_1^n) \quad (4.6)$$

to Detector 2, and decides on the hypothesis

$$\hat{\mathcal{H}}_1 \triangleq \psi_{1,n}(M_1, Y_1^n). \quad (4.7)$$

Detector 2 applies a decision function

$$\psi_{2,n}: \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Y}_2^n \rightarrow \{0, 1\} \quad (4.8)$$

to the triple  $(M_1, M_2, Y_2^n)$  to produce the decision

$$\hat{\mathcal{H}}_2 \triangleq \psi_{2,n}(M_1, M_2, Y_2^n). \quad (4.9)$$

Both detectors are required to have vanishing probabilities of error under both hypotheses. Moreover, for Detector 2, we require that the probability of error under  $\mathcal{H} = 1$  decays exponentially fast with the largest possible exponent. For Detector 1, we consider two scenarios: *coherent detection* and *concurrent detection*. Under coherent detection, Detector 1 wishes to maximize the exponential decay of the probability of error under  $\mathcal{H} = 1$ . Under concurrent detection, Detector 1 wishes to maximize the exponential decay of the probability of error under  $\mathcal{H} = 0$ . In a unifying manner, we define, for  $h_1 \in \{0, 1\}$  and  $\bar{h}_1 = (h_1 + 1) \bmod 2$ , the following error probabilities:

$$\alpha_{1,n} := \Pr\{\hat{\mathcal{H}}_1 = \bar{h}_1 | \mathcal{H} = h_1\}, \quad (4.10)$$

$$\beta_{1,n} := \Pr\{\hat{\mathcal{H}}_1 = h_1 | \mathcal{H} = \bar{h}_1\}, \quad (4.11)$$

$$\alpha_{2,n} := \Pr\{\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0\}, \quad (4.12)$$

$$\beta_{2,n} := \Pr\{\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1\}. \quad (4.13)$$

**Definition 4.1 (Achievability with fixed length communication alphabets.)** Given  $\bar{h}_1 \in \{0, 1\}$ ,  $(\epsilon_1, \epsilon_2) \in (0, 1)^2$  and communication alphabet sizes  $W_1, W_2 \geq 0$ , an error-exponents pair  $(\theta_1, \theta_2)$  is said achievable if for all blocklengths  $n$  there exist functions  $\phi_{1,n}$ ,  $\phi_{2,n}$ ,  $\psi_{1,n}$  and  $\psi_{2,n}$  as in (3.43), (3.45), (4.5), and (4.8) so that the following limits hold:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_{1,n} \leq \epsilon_1, \quad \overline{\lim}_{n \rightarrow \infty} \alpha_{2,n} \leq \epsilon_2, \quad (4.14)$$

$$\theta_1 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{1,n}, \quad \theta_2 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{2,n}, \quad (4.15)$$

and

$$\overline{\lim}_{n \rightarrow \infty} W_{1,n} \leq W_1, \quad \overline{\lim}_{n \rightarrow \infty} W_{2,n} \leq W_2. \quad (4.16)$$

**Definition 4.2 (Error-exponents region for fixed communication alphabets)** For fixed  $\bar{h}_1 \in \{0, 1\}$  and communication alphabet sizes  $W_1, W_2 \geq 0$ , the closure of the set of all achievable exponent pairs  $(\theta_1, \theta_2)$  is called the error-exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$ .

## 4.2 Coherent detection and concurrent detection with $P_X = \bar{P}_X$

**Proposition 4.1 (Coherent Detection)** For coherent detection,  $\bar{h}_1 = 1$ ,  $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ , and for all values  $W_1 \geq 2$  and  $W_2 \geq 2$ , the error-exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$  is the set of all non-negative rate pairs

$(\theta_1, \theta_2)$  satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1} : \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1}}} D\left(\tilde{P}_{XY_1} \parallel \bar{P}_{XY_1}\right) \quad (4.17)$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2} : \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\left(\tilde{P}_{XY_1Y_2} \parallel \bar{P}_{XY_1Y_2}\right). \quad (4.18)$$

*Proof:* For completeness the proof is given in Appendix B. The achievability and converse parts are similar to [14] and [31].  $\blacksquare$

**Proposition 4.2 (Concurrent Detection with  $P_X = \bar{P}_X$ )** Under concurrent detection, i.e. when  $\bar{h}_1 = 0$ ,  $(\epsilon_1, \epsilon_2) \in (0, 1)^2$  and when  $P_X = \bar{P}_X$ , then for all values  $W_1 \geq 2$  and  $W_2 \geq 2$ , the error-exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$  is the set of all non-negative rate pairs  $(\theta_1, \theta_2)$  satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1} : \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D\left(\tilde{P}_{XY_1} \parallel P_{XY_1}\right) \quad (4.19)$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2} : \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\left(\tilde{P}_{XY_1Y_2} \parallel \bar{P}_{XY_1Y_2}\right). \quad (4.20)$$

*Proof:* For completeness the proof is given in Appendix C. The achievability and converse parts are similar to [14] and [31].  $\blacksquare$

### 4.3 Concurrent detection with $P_X \neq \bar{P}_X$

We now consider concurrent detection,  $\bar{h}_1 = 0$ , and  $P_X \neq \bar{P}_X$ . Here the optimal exponents region depends on whether the alphabet size  $W_1$  equals 2 or is larger. This is due to the fact that a type signaling strategy similar to the coding used previously (see Appendices A and B) for this particular case requires at least a ternary message for communication between Encoder and detectors, and so, in the case of communication with total compression, another strategy must be considered.

#### 4.3.1 Concurrent detection with $P_X \neq \bar{P}_X$ and $W_1 \geq 3$

We first assume

$$W_1 \geq 3 \quad \text{and} \quad W_2 \geq 2, \quad (4.21)$$

and present a coding scheme for this scenario.

Pick a small positive number  $\mu > 0$  such that the typical sets  $\mathcal{T}_\mu^n(P_X)$  and  $\mathcal{T}_\mu^n(\bar{P}_X)$  do not intersect:

$$\mathcal{T}_\mu^n(P_X) \cap \mathcal{T}_\mu^n(\bar{P}_X) = \emptyset. \quad (4.22)$$

Sensor: Given that it observes  $X^n = x^n$ , it sends

$$M_1 = \begin{cases} 0 & \text{if } x^n \in \mathcal{T}_\mu^n(P_X) \\ 1 & \text{if } x^n \in \mathcal{T}_\mu^n(\bar{P}_X) \\ 2 & \text{otherwise.} \end{cases} \quad (4.23)$$

Detector 1: Given that it observes  $Y_1^n = y_1^n$  and  $M_1 = m_1$ , it decides

$$\hat{\mathcal{H}}_1 = \begin{cases} 1 & \text{if } m_1 = 1 \text{ and } y_1^n \in \mathcal{T}_\mu^n(\bar{P}_{Y_1}) \\ 0 & \text{otherwise.} \end{cases} \quad (4.24)$$

It sends

$$M_2 = \begin{cases} 0 & \text{if } m_1 = 0 \text{ and } y_1^n \in \mathcal{T}_\mu^n(P_{Y_1}) \\ 1 & \text{otherwise} \end{cases} \quad (4.25)$$

to Detector 2.

Detector 2: Given that it observes  $Y_2^n = y_2^n$  and messages  $M_1 = m_1$  and  $M_2 = m_2$ , it decides

$$\hat{\mathcal{H}}_2 = \begin{cases} 0 & \text{if } m_1 = m_2 = 0 \text{ and } y_2^n \in \mathcal{T}_\mu^n(P_{Y_2}) \\ 1 & \text{otherwise.} \end{cases} \quad (4.26)$$

**Proposition 4.3 (Concurrent Detection when  $P_X \neq \bar{P}_X$  and  $W_1 \geq 3$ )** Under concurrent detection (when  $\bar{h}_1 = 1$ ), and for all  $(\epsilon_1, \epsilon_2) \in (0, 1)^2$  and all values  $W_1 \geq 3$  and  $W_2 \geq 2$ , the error-exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$  is the set of all non-negative rate pairs  $(\theta_1, \theta_2)$  satisfying

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \tilde{P}_X = \bar{P}_X \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| P_{XY_1}) \quad (4.27)$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \tilde{P}_X = P_X \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}). \quad (4.28)$$

*Proof*: The achievability follows by the above coding scheme; and the converse is similar to that of Proposition 4.2. ■

The exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$  in these first three Propositions 4.1–4.3 is rectangular, and each of the detectors can simultaneously achieve the optimal exponent as if it were the only detector in the system. As we see in the following, this is not always the case.

### 4.3.2 Concurrent detection with $P_X \neq \bar{P}_X$ and $W_1 = 2$

In the rest of this section, we assume

$$W_1 = 2 \quad \text{and} \quad W_2 \geq 2, \quad (4.29)$$

and present the optimal error-exponents region for this case. It is achieved by the following coding scheme.

Pick a real number  $r$ , a small positive number  $\mu > 0$  satisfying (4.22), and the function  $b: \{0, 1\} \rightarrow \{0, 1\}$  either as

$$b(0) = b(1) = 0 \quad (4.30)$$

or as

$$b(0) = 0 \quad \text{and} \quad b(1) = 1. \quad (4.31)$$

We then assign each type  $\pi \in \mathcal{P}_n(\mathcal{X}^n)$  that satisfies

$$|\pi - P_X| > \mu \quad \text{and} \quad |\pi - \bar{P}_X| > \mu \quad (4.32)$$

to one of two sets  $\Gamma_0$  or  $\Gamma_1$ . If  $b(0) = b(1) = 0$ , then we assign all these types to the set  $\Gamma_1$ . Otherwise, we assign them between the two sets according to the following rule:

$$\pi \in \Gamma_{b(1)} \iff \min_{\substack{\tilde{P}_{XY_1}: \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| P_{XY_1}) + r \geq \min_{\substack{\tilde{P}_{XY_1 Y_2}: \tilde{P}_X = \pi \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1 Y_2} \| \bar{P}_{XY_1 Y_2}), \quad (4.33)$$

and  $\pi \in \Gamma_{b(0)}$  otherwise. Given that it observes  $X^n = x^n$ , the Sensor sends

$$M_1 = \begin{cases} b(0) & \text{if } x^n \in \mathcal{T}_\mu^n(P_X) \\ b(1) & \text{if } x^n \in \mathcal{T}_\mu^n(\bar{P}_X) \\ 0 & \text{if } P_{x^n} \in \Gamma_0 \\ 1 & \text{if } P_{x^n} \in \Gamma_1. \end{cases} \quad (4.34)$$

Detector 1: Given that it observes  $Y_1^n = y_1^n$  and received message  $M_1 = m_1$ , Detector 1 decides

$$\hat{\mathcal{H}}_1 = \begin{cases} 1 & \text{if } m_1 = b(1) \quad \text{and} \quad y_1^n \in \mathcal{T}_\mu^n(\bar{P}_{Y_1}) \\ 0 & \text{otherwise.} \end{cases} \quad (4.35)$$

It sends

$$M_2 = \begin{cases} 0 & \text{if } m_1 = b(0) \quad \text{and} \quad y_1^n \in \mathcal{T}_\mu^n(P_{Y_1}) \\ 1 & \text{otherwise} \end{cases} \quad (4.36)$$

to Detector 2.

Detector 2: Given that it observes  $Y_2^n = y_2^n$  and received messages  $M_1 = m_1$  and  $M_2 = m_2$ , it decides

$$\hat{\mathcal{H}}_2 = \begin{cases} 0 & \text{if } m_1 = b(0) \quad \text{and} \quad m_2 = 0 \quad \text{and} \quad y_2^n \in \mathcal{T}_\mu^n(P_{Y_2}) \\ 1 & \text{otherwise.} \end{cases} \quad (4.37)$$

The described scheme achieves the following optimal error-exponents region.

**Theorem 4.4 (Concurrent Detection when  $P_X \neq \bar{P}_X$  and  $W_1 = 2$ )** Under concurrent detection and for all  $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ , for all values  $W_1 = 2$  and  $W_2 \geq 2$ , the error-exponents region  $\mathcal{E}_0(W_1, W_2, \epsilon_1, \epsilon_2)$  is the set of all nonnegative rate pairs  $(\theta_1, \theta_2)$  that satisfy

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \tilde{P}_X \in \Gamma_{b(1)} \\ \tilde{P}_{Y_1} = \bar{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| P_{XY_1}), \quad (4.38)$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2} : \tilde{P}_X \in \Gamma_b(0), \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D\left(\tilde{P}_{XY_1Y_2} \parallel \bar{P}_{XY_1Y_2}\right). \quad (4.39)$$

for some real  $r$  and one of the mappings in (4.30) and (4.31), and the corresponding sets  $\Gamma_0$  and  $\Gamma_1$ .

*Proof:* See Appendix D. ■

**Remark 4.1 (Sending a sublinear number of bits)** *A close inspection of the converse proofs for Propositions 4.1–4.3 and Theorem 4.4 shows that they remain valid when the alphabet sizes are not fixed but grow sublinearly in the blocklength  $n$ , i.e., when*

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_{1,n} = 0, \quad \text{and} \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_{2,n} = 0.$$

*Propositions 4.1–4.3 and Theorem 4.4 remain thus valid when communication is limited to a sublinear number of bits. More precisely this is due to the fact that the key argument of the proof to the converse part is generalized to any alphabet whose size grows sub-exponentially fast with the number of observations. Indeed the key argument of this converse is Theorem 2.7 still valid in this case.*

**Remark 4.2 (Extension to many hypotheses)** *Most of the results in this section can be extended to a scenario with more than two hypotheses. For  $H = 2$  the error-exponents region showed a tradeoff in the exponents under concurrent detection only when  $W_1 = W_2 = 2$ . In contrast, for  $H \geq 3$ , a tradeoff arises for a variety of pairs  $W_1, W_2$ . The minimum required values for  $W_1$  and  $W_2$  leading to a rectangular error-exponents region is always less than the number of hypotheses which have distinct  $X$ -marginals and the number of hypotheses which have distinct  $Y_1$ -marginals.*

## 4.4 Numerical Example

We now present an example for concurrent detection with  $P_X \neq \bar{P}_X$ .

**Example 4.1** *Consider a setup where  $X, Y_1, Y_2$  are binary with pmfs*

$$\left\{ \begin{array}{ll} P_{XY_1Y_2}(0, 0, 0) = 0.1 & P_{XY_1Y_2}(0, 0, 1) = 0.15 \\ P_{XY_1Y_2}(0, 1, 0) = 0.1125 & P_{XY_1Y_2}(0, 1, 1) = 0.1375 \\ P_{XY_1Y_2}(1, 0, 0) = 0.0875 & P_{XY_1Y_2}(1, 0, 1) = 0.1625 \\ P_{XY_1Y_2}(1, 1, 0) = 0.0825 & P_{XY_1Y_2}(1, 1, 1) = 0.1675 \\ \bar{P}_{XY_1Y_2}(0, 0, 0) = 0.1675 & \bar{P}_{XY_1Y_2}(0, 0, 1) = 0.0825 \\ \bar{P}_{XY_1Y_2}(0, 1, 0) = 0.1625 & \bar{P}_{XY_1Y_2}(0, 1, 1) = 0.0875 \\ \bar{P}_{XY_1Y_2}(1, 0, 0) = 0.1375 & \bar{P}_{XY_1Y_2}(1, 0, 1) = 0.1125 \\ \bar{P}_{XY_1Y_2}(1, 1, 0) = 0.15 & \bar{P}_{XY_1Y_2}(1, 1, 1) = 0.1 \end{array} \right.$$



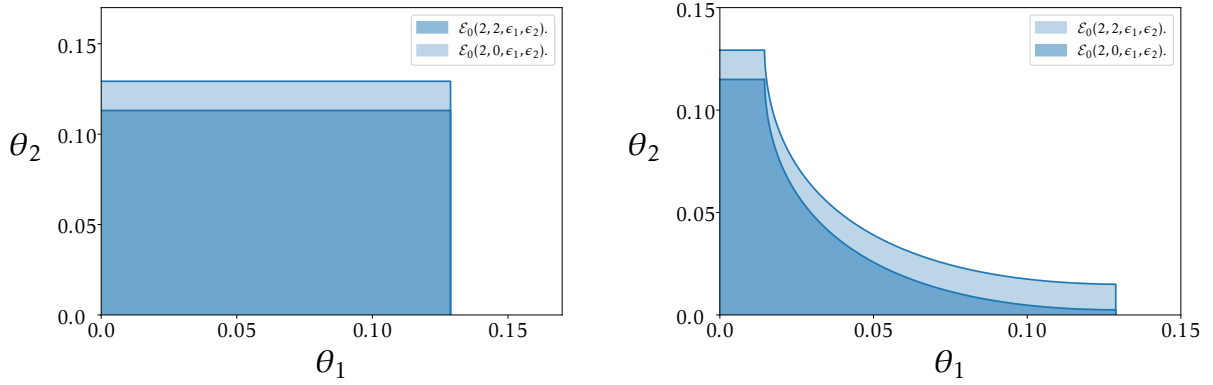


Figure 4.2: Exponents region of Example 4.1, see [12] for implementation details. On the left: exponent regions  $\mathcal{E}_0(2, 2, \epsilon_1, \epsilon_2)$  and  $\mathcal{E}_0(2, 0, \epsilon_1, \epsilon_2)$  for coherent detection. On the right: exponent regions  $\mathcal{E}_0(2, 2, \epsilon_1, \epsilon_2)$ , and  $\mathcal{E}_0(2, 0, \epsilon_1, \epsilon_2)$  for concurrent detection.

Figure 4.2 illustrates the exponent region for coherent and concurrent detection of Proposition 4.3 and of Theorem 4.4. Specifically, the figure on the left shows the exponent region with (when  $W_1 = 2$  and  $W_2 = 2$ ) or without (when  $W_1 = 2$  and  $W_2 = 0$ ) cooperation under coherent detection. In that cases the region is a square and each Detector behaves as if it was the only one in the network to take a decision under zero-rate compression. This occurs because the type signaling schemes which are optimal in the zero-rate regime when only one of the detectors has to take a decision, are sufficiently compact to be implemented at the same time when each message is constrained to be one bit long. The figure on the right shows these exponent regions under concurrent detection, it is non-convex. (Notice that time-sharing arguments cannot be applied to convexify the region.)

## 4.5 Benefits of Cooperation

To discuss the benefits of cooperation, we quickly state the optimal exponents region without cooperation, i.e., for

$$W_2 = 0. \quad (4.40)$$

They were determined in our work [9]. Under coherent detection or under concurrent detection with  $P_X = \bar{P}_X$  or  $W_1 \geq 3$ , the error-exponents region  $\mathcal{E}_0(W_1, W_2 = 0)$  are similar to Propositions 4.1–4.3 but with a modified constraint on  $\theta_2$ . More precisely, Propositions 4.1–4.3 remain valid for  $W_2 = 0$  if the constraints on  $\theta_2$ , (4.18), (4.20), (4.28) are replaced by

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_2}: \\ \tilde{P}_X = P_X, \\ \tilde{P}_{Y_2} = P_{Y_2}}} D\left(\tilde{P}_{XY_2} \parallel \bar{P}_{XY_2}\right).$$

So, in these scenarios, the exponents region is a rectangle both in the case with and without cooperation, and with cooperation the  $\theta_2$ -side of the rectangle is increased by the quantity

$$\min_{\substack{\tilde{P}_{XY_1Y_2}: \tilde{P}_X=P_X \\ \tilde{P}_{Y_1}=P_{Y_1}, \tilde{P}_{Y_2}=P_{Y_2}}} D\left(\tilde{P}_{XY_1Y_2} \parallel \bar{P}_{XY_1Y_2}\right) - \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X=P_X; \\ \tilde{P}_{Y_2}=P_{Y_2}}} D\left(\tilde{P}_{XY_2} \parallel \bar{P}_{XY_2}\right). \quad (4.41)$$

Under concurrent detection when  $P_X \neq \bar{P}_X$  and  $W_1 = 2$ , the exponents region is not a rectangle, but there is a tradeoff between the two exponents. In this case, it seems difficult to quantify the cooperation benefit in general.



---

Positive-rate distributed hypothesis testing

---

In this chapter, based on our work published in [9, 10, 11], we focus on the case of positive rate communication in the multiterminal binary hypothesis testing scenario shown in Figure 5.1 where communication is restricted to a maximum communication rate. Here  $R_1 > 0$ , and the cooperation rate  $R_2 \geq 0$ . Again we investigate the setup where both detectors aim at maximizing the error exponent under same hypothesis (*coherent detection*) or the setup where the two detectors aim at maximizing their exponent under different hypotheses (*concurrent detection*). Using the same notations as the ones used in previous chapter, we

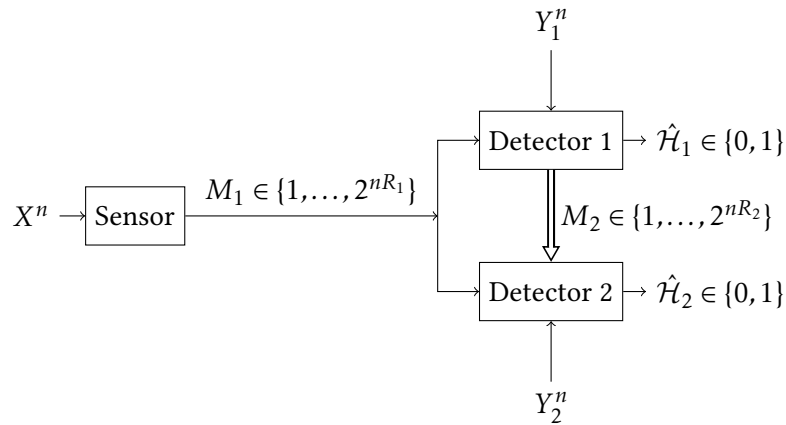


Figure 5.1: A Heegard-Berger type source coding model with unidirectional conferencing for multiterminal hypothesis testing.

introduce a distinct definition of achievable error-exponent pairs adapted to positive rate communication:

**Definition 5.1 (Achievability under Rate-Constraints)** Given rates  $R_1, R_2 \geq 0$ , an error-exponents pair

$(\theta_1, \theta_2)$  is said achievable if for all blocklengths  $n$  there exist functions  $\phi_{1,n}$ ,  $\phi_{2,n}$ ,  $\psi_{1,n}$  and  $\psi_{2,n}$  as in (3.43), (3.45), (4.5), and (4.8) so that the following limits hold:

$$\lim_{n \rightarrow \infty} \alpha_{1,n} = 0, \quad \lim_{n \rightarrow \infty} \alpha_{2,n} = 0, \quad (5.1)$$

$$\theta_1 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{1,n}, \quad \theta_2 \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_{2,n}, \quad (5.2)$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_{1,n} \leq R_1, \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log W_{2,n} \leq R_2. \quad (5.3)$$

We will use this new definition all along the chapter. We now define the error-exponent region under rate-constraints.

**Definition 5.2 (Error-Exponents Region under Rate-Constraints)** For rates  $R_1, R_2 \geq 0$  the closure of the set of all achievable exponent pairs  $(\theta_1, \theta_2)$  is called the error-exponents region  $\mathcal{E}(R_1, R_2)$ .

We will give some insights for special setup of coherent detection in section 5.1. In section 5.2, we describe a general testing scheme adapted from the one of Ahlswede, Csiszár and Han [14, 1] to the two setups of coherent and concurrent detection. We illustrate the special case of testing against independence with a numerical example in section 5.3.

## 5.1 Special cases

In this section, we study the setting in which the two detectors aim at maximizing the error exponents under the *same* hypothesis, i.e.,  $\bar{h}_1 = 0$ . For convenience, we assume that  $\bar{P}(x, y_1, y_2) > 0$  for all  $(x, y_1, y_2) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ .

### 5.1.1 High rate regime

We first consider the extreme case where both links are of high rates so that under hypothesis  $\mathcal{H} = h_1$ , Detector 1 can obtain the sequence  $X^n$  with high probability and under  $\mathcal{H} = 0$ , Detector 2 can obtain both sequences  $X^n$  and  $Y_1^n$  with high probability. We will see that in this case both Detector 1 attains the exponent of a centralized setup where it observes  $(X^n, Y_1^n)$  and Detector 2 attains the exponent of a centralized setup where it observes  $(X^n, Y_1^n, Y_2^n)$ .

We first consider coherent detection where  $\bar{h}_1 = 1$ . Pick a small  $\epsilon$ . The Sensor describes the sequence  $X^n$  to both detectors if  $X^n \in \mathcal{T}_\mu^n(P_X)$ , and otherwise it sends 0. Detector 1 describes the sequence  $Y_1^n$  to Detector 2 if  $(X^n, Y_1^n) \in \mathcal{T}_\mu^n(P_{XY_1})$ , and otherwise it sends 0. The described coding scheme requires rates

$$R_1 \geq H(X) + \epsilon \quad (5.4)$$

$$R_2 \geq H(Y_1|X) + \epsilon. \quad (5.5)$$

Detector 1 decides on  $\hat{\mathcal{H}}_1 = 1$ , if the Sensor sent 0 or itself it sent 0. Otherwise it decides on  $\hat{\mathcal{H}}_1 = 1$ . Detector 2 decides on  $\hat{\mathcal{H}}_2 = 1$ , if the Sensor or Detector 1 sent 0. Otherwise it decides on  $\hat{\mathcal{H}}_1 = 1$  if and only if its own observation  $Y_2^n$  and the received sequences  $X^n$  and  $Y_1^n$  are jointly typical,  $(X^n, Y_1^n, Y_2^n) \in \mathcal{T}_\mu^n(P_{XY_1Y_2})$ .

The described scheme achieves the set of all non-negative pairs  $(\theta_1, \theta_2)$  satisfying

$$\theta_1 \leq D(P_{XY_1} \| \bar{P}_{XY_1}) \quad (5.6a)$$

$$\theta_2 \leq D(P_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}). \quad (5.6b)$$

This set coincides with the optimal error-exponents region  $\mathcal{E}(R_1, R_2)$ , because it also coincides with the error-exponent region of a centralized setup where Detector 1 observes both  $X^n$  and  $Y_1^n$  and Detector 2 observes all  $X^n$ ,  $Y_1^n$ , and  $Y_2^n$ .

Consider now concurrent detection where  $\bar{h}_1 = 0$ . In this case, the Sensor describes the sequence  $X^n$  to both detectors if  $X^n \in \mathcal{T}_\mu^n(P_X)$  or if  $X^n \in \mathcal{T}_\mu^n(\bar{P}_X)$ . Otherwise it sends 0. Detector 1 describes the sequence  $Y_1^n$  to Detector 2 if  $(X^n, Y_1^n) \in \mathcal{T}_\mu^n(P_{XY_1})$ , and otherwise it sends 0. Detector 2 decides as above and Detector 1 decides on  $\hat{\mathcal{H}}_1 = 0$  if and only if its own observation  $Y_1^n$  and the described sequence  $X^n$  are jointly typical, i.e.,  $(X^n, Y_1^n) \in \mathcal{T}_\mu^n(\bar{P}_{XY_1})$ . The coding scheme requires rates

$$R_1 \geq \max\{H(\bar{X}), H(X)\} + \epsilon \quad (5.7)$$

$$R_2 \geq H(Y_1|X) + \epsilon. \quad (5.8)$$

and achieves the set of all nonnegative pairs  $(\theta_1, \theta_2)$  satisfying

$$\theta_1 \leq D(\bar{P}_{XY_1} \| P_{XY_1}) \quad (5.9a)$$

$$\theta_2 \leq D(P_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}). \quad (5.9b)$$

Again, this set coincides with the optimal error-exponents region  $\mathcal{E}(R_1, R_2)$  because it also coincides with the optimal exponents region when Detector 1 observes the pair  $X^n, Y_1^n$  and Detector 2 observes  $X^n, Y_1^n$ , and  $Y_2^n$ .

Both results remain valid without cooperation if the term  $D(P_{XY_1Y_2} \| \bar{P}_{XY_1Y_2})$  limiting the second exponent  $\theta_2$  is replaced by  $D(P_{XY_2} \| \bar{P}_{XY_2})$ . The benefit of cooperation is thus equal to

$$D(P_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}) - D(P_{XY_2} \| \bar{P}_{XY_2}) = \mathbb{E}_{P_{XY_2}} [D(P_{Y_1|XY_2} \| \bar{P}_{Y_1|XY_2})]$$

in both cases.

In some special cases, the described setup degenerates and the error-exponents region is the same as in a setup without cooperation or in a setup with a single centralized detector.

### 5.1.2 Degraded side information at first detector

We first consider a setup where cooperation is not beneficial.

**Proposition 5.1** *Assume the Markov chain  $X \dashv\!\!\!\dashv Y_2 \dashv\!\!\!\dashv Y_1$  under both hypotheses with identical law  $P_{Y_1|Y_2} = \bar{P}_{Y_1|Y_2}$ :*

$$P_{XY_1Y_2} = P_{XY_2} P_{Y_1|Y_2} \quad (5.10a)$$

$$\bar{P}_{XY_1Y_2} = \bar{P}_{XY_2} P_{Y_1|Y_2}. \quad (5.10b)$$

*In this case, irrespective of the cooperation rate  $R_2 \geq 0$  and of the value of  $\bar{h}_1 \in \{0, 1\}$ , the error-exponent region  $\mathcal{E}(R_1, R_2)$  coincides with the exponent region of the scenario without cooperation (see Figure 5.2).*

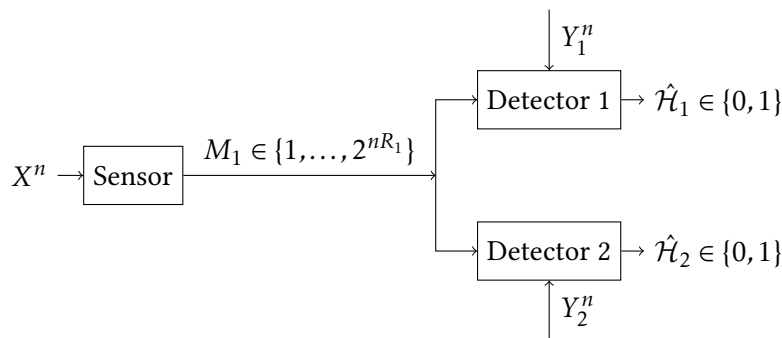


Figure 5.2: Equivalent system without cooperation when  $X \dashv\!\!\!\dashv Y_2 \dashv\!\!\!\dashv Y_1$  under both hypotheses.

*Proof:* The error-exponents region  $\mathcal{E}(R_1, R_2)$  of the original setup cannot be larger than the error-exponent regions of an enhanced setup (with cooperation) where Detector 2 not only observes  $Y_2^n$  but also  $Y_1^n$ . But in this new setup, the cooperation link is useless because Detector 2 can generate the cooperation message locally. Moreover, without cooperation, the observation  $Y_1^n$  is not beneficial because the conditional laws  $P_{Y_1|XY_2}$  and  $\bar{P}_{Y_1|XY_2}$  coincide and only depend on  $Y_2$ , and so Detector 2 can generate a statistically equivalent observation to  $Y_1^n$  itself based only on  $Y_2^n$ . By these arguments, the error-exponents region  $\mathcal{E}(R_1, R_2)$  of the original setup is not larger than the one of the setup without cooperation. But  $\mathcal{E}(R_1, R_2)$  can also not be smaller than the error-exponents regions of the same setup but without cooperation, because the latter setup can be mimicked in the former. This concludes the proof. ■

### 5.1.3 Degraded side information at second detector

We now consider a setup that is equivalent to a setup with a single centralized detector.

**Proposition 5.2** *Assume the Markov chain  $X \dashv\!\!\!\dashv Y_1 \dashv\!\!\!\dashv Y_2$  holds under both hypotheses with identical law  $P_{Y_2|Y_1} = \bar{P}_{Y_2|Y_1}$ . I.e.,*

$$P_{XY_1Y_2} = P_{XY_1} P_{Y_2|Y_1} \quad (5.11a)$$

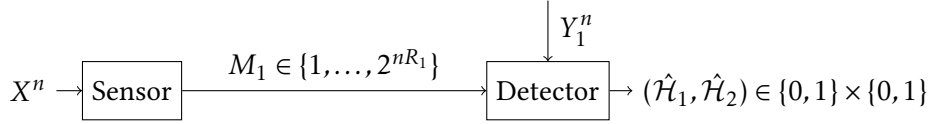


Figure 5.3: Equivalent point to point system when  $X \dashv\!\!\!\dashv Y_1 \dashv\!\!\!\dashv Y_2$  under both hypotheses.

$$\bar{P}_{XY_1Y_2} = \bar{P}_{XY_1}P_{Y_2|Y_1}. \quad (5.11b)$$

In this case, irrespective of the cooperation rate  $R_2 \geq 0$  and of the value of  $h_1 \in \{0, 1\}$ , the error-exponent regions  $\mathcal{E}(R_1, R_2)$  coincides with the error-exponents regions of the scenario in Figure 5.2 without cooperation and where both detectors observe  $Y_1^n$  but not  $Y_2^n$ . As a consequence, they also coincide with the error-exponents regions of the scenario in Figure 5.3 with a single detector observing  $Y_1^n$  that takes both decisions  $\hat{\mathcal{H}}_1$  and  $\hat{\mathcal{H}}_2$ .

*Proof:* The error-exponents region  $\mathcal{E}(R_1, R_2)$  of the original setup cannot be larger than the error-exponents regions of an enhanced setup (with cooperation) where Detector 2 not only observes  $Y_2^n$  but also  $Y_1^n$ . Since Detector 2 can generate an observation that is statistically equivalent to  $Y_2^n$  given  $Y_1^n$ , the error-exponents regions are no larger than in the setup where both detectors observe  $Y_1^n$  but not  $Y_2^n$ . Furthermore, since allowing the two detectors to fully cooperate in their decision can only increase the error-exponents regions, the region  $\mathcal{E}(R_1, R_2)$  must be included in the exponents regions of the setup in Figure 5.3 where a single detector takes both decisions.

On the other hand,  $\mathcal{E}(R_1, R_2)$  can also not be smaller than the error-exponents regions of the setup in Figure 5.3. In fact, in the original setup, detector 1 can mimick the single central detector and forward the decision  $\hat{\mathcal{H}}_2$  to detector 2, which follows this decision. This strategy requires only a single cooperation bit and can thus be implemented irrespective to the available cooperation rate  $R_2 \geq 0$ . This conclude the proof.  $\blacksquare$

#### 5.1.4 Testing against independence under coherent detection

We introduce a special case of “testing-against-independence” scenario under coherent detection,  $\bar{h}_1 = 1$ , where

$$P_{XY_1Y_2} = P_{X|Y_1Y_2}P_{Y_1}P_{Y_2} \quad (5.12)$$

$$\bar{P}_{XY_1Y_2} = P_XP_{Y_1}P_{Y_2}. \quad (5.13)$$

We assume a cooperation rate  $R_2 = 0$ , which means that Detector 1 can send a message  $M_2$  to Detector 2 that is described by a sublinear number of bits.

The simple scheme in the next subsection 5.2.1 achieves the following exponents region, which can be proved to be optimal.



**Theorem 5.3 (Testing Against Independence)** Assume  $\bar{h}_1 = 1$  and (5.12). Then,  $\mathcal{E}(R_1, 0)$  is the set of all nonnegative exponent pairs  $(\theta_1, \theta_2)$  for which

$$\theta_1 \leq I(U; Y_1) \quad (5.14a)$$

$$\theta_2 \leq I(U; Y_1) + I(U; Y_2), \quad (5.14b)$$

for some  $U$  satisfying the Markov chain  $U \text{---} X \text{---} (Y_1, Y_2)$  and the rate constraint  $R_1 \geq I(U; X)$ .

*Proof:* The achievability follows by specializing and evaluating Theorem 5.5 for this setup. The converse is proved in Appendix E. ■

**Remark 5.1** As seen in Theorem 5.3, in some special case exponents accumulate.

**Lemma 5.4 (Cardinality bound)** the right hand sides of (5.14) in Theorem 5.3 remain valid if we impose the cardinality bound  $|\mathcal{U}| = |\mathcal{X}| + 2$ .

Notice that for  $R_2 = 0$  the scheme in the following subsection sends only a single bit and that without cooperation, the term  $I(U; Y_2)$  needs to be removed on the right-hand side of (5.14b). This mutual information term thus represents the benefit of a single cooperation bit from Detector 1 to Detector 2.

## 5.2 General achievability results

### 5.2.1 A simple scheme with cooperation

In this subsection, we present coding schemes for both coherent and concurrent detection. Notice that this difficulty seems inherent to all multi-user hypothesis testing scenarios, see e.g. [29]. We first present a scheme for coherent detection,

$$\bar{h}_1 = 1, \quad (5.15)$$

and then explain how to change the scheme for concurrent detection.

#### Coherent Detection

Preliminaries: Fix a small  $\mu > 0$  and a pair of auxiliary random variables  $(U, V) \in \mathcal{U} \times \mathcal{V}$  satisfying the following Markov chains

$$U \text{---} X \text{---} (Y_1, Y_2) \quad (5.16)$$

$$V \text{---} (Y_1, U) \text{---} (Y_2, X) \quad (5.17)$$

and satisfying the rate constraints

$$R_1 > I(U; X) \quad (5.18)$$

$$R_2 > I(V; Y_1 | U). \quad (5.19)$$

Codebook Generation: we randomly generate the codebook

$$\mathcal{C}_U \triangleq \{u^n(m_1): m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}\} \quad (5.20)$$

by drawing each entry of each codeword  $u^n(m_1)$  i.i.d. according to  $P_U$ .

Furthermore, we superpose a codebook  $\mathcal{C}_V$  on codebook  $\mathcal{C}_U$ . So, for each index  $m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}$ , we randomly construct the codebook

$$\mathcal{C}_V(m_1) \triangleq \{v^n(m_2|m_1): m_2 \in \{1, \dots, \lfloor 2^{nR_2} \rfloor\}\} \quad (5.21)$$

by drawing the  $j$ -th entry of each codeword  $v^n(m_2|m_1)$  according to  $P_{V|U}(\cdot|u_j(m_1))$ , the conditional pmf where  $u_j(m_1)$  denotes the  $j$ -th component of codeword  $u^n(m_1)$ .

Reveal all codebooks to all terminals.

Sensor: Assume it observes the source sequence  $X^n = x^n$ .

Then, it first looks for a message  $m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}$  such that

$$(u^n(m_1), x^n) \in \mathcal{T}_{\mu/8}^n(P_{UX}). \quad (5.22)$$

If one or multiple such indices  $m$  are found, the Sensor selects  $m_1^*$  uniformly at random over these indices and sends

$$M_1 = (1, m_1^*).$$

Otherwise, it sends

$$M_1 = (0, 0).$$

Detector 1: If  $M_1 = (0, 0)$ , Detector 1 decides on the alternative hypothesis

$$\hat{\mathcal{H}}_1 = 1.$$

If

$$M_1 = (1, m_1) \quad \text{for some } m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\},$$

and given that  $Y_1^n = y_1^n$ , Detector 1 checks whether

$$(u^n(m_1), y_1^n) \in \mathcal{T}_{\mu/4}^n(P_{UY}). \quad (5.23)$$

If the test is successful, it decides on the null hypothesis

$$\hat{\mathcal{H}}_1 = 0.$$

Otherwise it decides on the alternative hypothesis. We now describe the communication to Detector 2. If

$$\hat{\mathcal{H}}_1 = 1$$

Detector 1 sends

$$M_2 = 0.$$

Otherwise, it looks for an index  $m_2 \in \{1, \dots, \lfloor 2^{nR_2} \rfloor\}$  such that

$$(u^n(m_1), v^n(m_2|m_1), y_1^n) \in \mathcal{T}_{\mu/2}^n(P_{UVY_1}). \quad (5.24)$$

If one or more such indices can be found, Detector 1 selects an index  $m_2^*$  among them uniformly at random and sends

$$M_2 = m_2^*.$$

Otherwise it sends  $M_2 = (0, 0)$ .

Detector 2: If

$$M_1 = (0, 0) \quad \text{or} \quad M_2 = 0,$$

Detector 2 decides on the alternative hypothesis

$$\hat{\mathcal{H}}_2 = 1.$$

If

$$M_1 = (1, m_1) \quad \text{for some } m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\} \text{ and } M_2 = m_2 \quad \text{for some } m_2 \in \{1, \dots, \lfloor 2^{nR_2} \rfloor\},$$

and given  $Y_2^n = y_2^n$ , Detector 2 checks whether

$$(u^n(m_1), v^n(m_2|m_1), y_2^n) \in \mathcal{T}_{\mu}^n(P_{UVY_2}). \quad (5.25)$$

If this check is successful, Decoder 2 decides on the null hypothesis

$$\hat{\mathcal{H}}_2 = 0.$$

Otherwise, it decides on the alternative hypothesis

$$\hat{\mathcal{H}}_2 = 1.$$

**Changes for concurrent detection when  $P_X = \bar{P}_X$**

We now consider the scenario of concurrent detection, so

$$\bar{h}_1 = 0.$$

We apply the same scheme as above, except for the decision at Detector 1, which is described next.

Detector 1: If

$$M_1 = (0, 0)$$

Detector 1 now decides

$$\hat{\mathcal{H}}_1 = 0.$$

If

$$M_1 = (1, m_1) \quad \text{for some } m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\},$$

and given that  $Y_1^n = y_1^n$ , Detector 1 checks whether

$$(u^n(m_1), y_1^n) \in \mathcal{T}_{\mu/4}^n(\bar{P}_{UY_1}). \quad (5.26)$$

If the test is successful, it decides

$$\hat{\mathcal{H}}_1 = 1.$$

Otherwise it decides

$$\hat{\mathcal{H}}_1 = 0.$$

Communication to Detector 2 is as described in the previous subsection.

### Changes for concurrent detection when $P_X \neq \bar{P}_X$

If  $\bar{h}_1 = 0$  and  $P_X \neq \bar{P}_X$ , the scheme should be changed as described in the previous paragraph. The following additional changes allow to obtain an improved scheme.

In this case, we choose  $\mu > 0$  so that the intersection

$$\mathcal{T}_{\mu/8}^n(P_X) \cap \mathcal{T}_{\mu/8}^n(\bar{P}_X) = \emptyset \quad (5.27)$$

is empty and we choose another auxiliary random variable  $\bar{U}_1$  satisfying

$$\bar{U}_1 \text{ --- } \bar{X} \text{ --- } (\bar{Y}_1, \bar{Y}_2) \quad (5.28)$$

$$R_1 \geq I(\bar{U}_1; \bar{X}) + \xi(\mu). \quad (5.29)$$

A third codebook

$$\mathcal{C}_{U,1} \triangleq \{u_1^n(m_1): m \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}\} \quad (5.30)$$

is drawn by picking the entries i.i.d. according to  $\bar{P}_{U_1}$ .

Encoding has to be changed as follows. If the test in (5.24) fails, then the Sensor looks for an index  $m_1 \in \{1, \dots, \lfloor 2^{nR_1} \rfloor\}$  such that

$$(u_1^n(m_1), x^n) \in \mathcal{T}_{\mu/8}^n(\bar{P}_{U_1X}). \quad (5.31)$$

If one or multiple such indices  $m_1$  are found, the Sensor selects  $m_1^*$  uniformly at random over these indices and sends

$$M_1 = (2, m_1^*)$$

Notice that by the condition (5.27), only one of the two tests (5.24) and (5.31) can be successful for any observed sequence  $x^n$ . It therefore does not matter which one is performed first.

Thus, now the Sensor sends three different types of messages:

$$M_1 = (0, 0) \quad \text{or} \quad M_1 = (1, m_1) \quad \text{or} \quad M_1 = (2, m_1).$$

The message  $M_1 = (1, m_1)$  indicates that the Sensor is tempted to guess  $\mathcal{H} = 0$ . After receiving such a message, Detector 1 therefore produces  $\hat{\mathcal{H}}_1 = 0$ . The same holds if  $M_1 = 0$ . In contrast, if  $M_1 = (2, m_1)$ , Detector 1 checks whether

$$(u_1^n(m_1), y_1^n) \in \mathcal{T}_{\mu/4}^n(\bar{P}_{U_1 Y_1}). \quad (5.32)$$

If successful it declares  $\hat{\mathcal{H}}_1 = 1$ , and otherwise  $\hat{\mathcal{H}}_1 = 0$ .

Communication from, Detector 1 to Detector 2 is as described before.

Similarly, the message  $M_1 = (2, m_1)$  now indicates that the Sensor is tempted to guess  $\mathcal{H} = 1$ . When receiving this message, Detector 2 therefore decides immediately  $\hat{\mathcal{H}}_2 = 1$ . Otherwise it acts as described in the original scheme.

### 5.2.2 Achievable regions

We now present the regions achieved by the coding scheme described in the previous subsection. Notice that the new achievable regions recover the extreme cases in the previous section 5.1.1, when the rates are set accordingly.

We first consider coherent detections  $\bar{h}_1 = 1$ .

For given rates  $R_1 \geq 0$  and  $R_2 \geq 0$ , define the following set of auxiliary random variables:

$$\mathcal{S}(R_1, R_2) \triangleq \left\{ (U, V) : \begin{array}{l} U \text{ --- } X \text{ --- } (Y_1, Y_2) \\ V \text{ --- } (Y_1, U) \text{ --- } (Y_2, X) \\ I(U; X) \leq R_1 \\ I(V; Y_1 | U) \leq R_2 \end{array} \right\}. \quad (5.33)$$

Further, define for each  $(U, V) \in \mathcal{S}(R_1, R_2)$ , the sets

$$\mathcal{L}_1(U) \triangleq \left\{ (\tilde{U}, \tilde{X}, \tilde{Y}_1) : \begin{array}{l} P_{\tilde{U}\tilde{X}} = P_{UX} \\ P_{\tilde{U}\tilde{Y}_1} = P_{UY_1} \end{array} \right\} \quad (5.34)$$

and

$$\mathcal{L}_2(UV) \triangleq \left\{ (\tilde{U}, \tilde{V}, \tilde{X}, \tilde{Y}_1, \tilde{Y}_2) : \begin{array}{l} P_{\tilde{U}\tilde{X}} = P_{UX} \\ P_{\tilde{U}\tilde{V}\tilde{Y}_1} = P_{UVY_1} \\ P_{\tilde{U}\tilde{V}\tilde{Y}_2} = P_{UVY_2} \end{array} \right\}, \quad (5.35)$$

and the random variables  $(\bar{U}, \bar{V})$  so as to satisfy

$$P_{\bar{U}|\bar{X}} = P_{U|X} \quad \text{and} \quad P_{\bar{V}|\bar{Y}_1, \bar{U}} = P_{V|Y_1, U} \quad (5.36)$$

and the Markov chains

$$\bar{U} \text{---} \bar{X} \text{---} (\bar{Y}_1, \bar{Y}_2) \quad (5.37)$$

$$\bar{V} \text{---} (\bar{Y}_1, \bar{U}) \text{---} (\bar{X}, \bar{Y}_2). \quad (5.38)$$

**Theorem 5.5 (Coherent Detection)** *If*

$$\bar{h}_1 = 1, \quad (5.39)$$

*the exponents region  $\mathcal{E}(R_1, R_2)$  contains all nonnegative pairs  $(\theta_1, \theta_2)$  that for some  $(U, V) \in \mathcal{S}(R_1, R_2)$  satisfy the two following conditions:*

$$\theta_1 \leq \min_{\bar{U}\bar{X}\bar{Y}_1 \in \mathcal{L}_1(U)} D(\bar{U}\bar{X}\bar{Y}_1 \| \bar{U}\bar{X}\bar{Y}_1) \quad (5.40a)$$

$$\theta_2 \leq \min_{\bar{U}\bar{V}\bar{X}\bar{Y}_1\bar{Y}_2 \in \mathcal{L}_2(UV)} D(\bar{V}\bar{U}\bar{X}\bar{Y}_1\bar{Y}_2 \| \bar{V}\bar{U}\bar{X}\bar{Y}_1\bar{Y}_2). \quad (5.40b)$$

*for some  $(U, V) \in \mathcal{S}(R_1, R_2)$*

*Proof:* The exponent region is achieved by the scheme described in Subsection 5.2.1. The proof is given in Appendix F. ■

For our second result, we also define for each auxiliary random variable  $U$  the set

$$\bar{\mathcal{L}}_1(U) \triangleq \left\{ (\bar{U}, \bar{X}, \bar{Y}_1) : \begin{array}{l} P_{\bar{U}\bar{X}} = \bar{P}_{UX} \\ P_{\bar{U}\bar{Y}_1} = \bar{P}_{UY_1} \end{array} \right\}. \quad (5.41)$$

**Theorem 5.6 (Concurrent Detection with  $P_X = \bar{P}_X$ )** *If*

$$\bar{h}_1 = 0, \quad \text{and} \quad P_X = \bar{P}_X, \quad (5.42)$$

*then the exponents region  $\mathcal{E}(R_1, R_2)$  contains all nonnegative pairs  $(\theta_1, \theta_2)$  that for some  $(U, V) \in \mathcal{S}(R_1, R_2)$  satisfy:*

$$\theta_1 \leq \min_{\bar{U}\bar{X}\bar{Y}_1 \in \bar{\mathcal{L}}_1(U)} D(\bar{U}\bar{X}\bar{Y}_1 \| UX Y_1) \quad (5.43a)$$

$$\theta_2 \leq \min_{\bar{U}\bar{V}\bar{X}\bar{Y}_1\bar{Y}_2 \in \mathcal{L}_2(UV)} D(\bar{V}\bar{U}\bar{X}\bar{Y}_1\bar{Y}_2 \| \bar{V}\bar{U}\bar{X}\bar{Y}_1\bar{Y}_2). \quad (5.43b)$$

*Proof:* Similar to the proof of Theorem 5.5 and omitted. ■

**Remark 5.2** *When  $R = 0$*

For a given rate  $R_1 \geq 0$ , define the following set of auxiliary random variables:

$$\mathcal{S}_{\text{nc}}(R_1) \triangleq \left\{ \bar{U}_1 : \begin{array}{l} \bar{U}_1 \text{ --- } \bar{X} \text{ --- } \bar{Y}_1 \\ I(\bar{U}_1; \bar{X}) \leq R_1 \end{array} \right\}. \quad (5.44)$$

and the random variable  $U_1$  so that  $P_{U_1|X} = \bar{P}_{U_1|X}$  and the Markov chain  $U_1 \text{ --- } X \text{ --- } Y_1$  holds.

**Theorem 5.7 (Concurrent Detection and  $P_X \neq \bar{P}_X$ )** *If*

$$\bar{h}_1 = 0 \quad \text{and} \quad P_X \neq \bar{P}_X, \quad (5.45)$$

*then the exponents region  $\mathcal{E}(R_1, R_2)$  contains all nonnegative pairs  $(\theta_1, \theta_2)$  that for some  $(U, V) \in \mathcal{S}(R_1, R_2)$  and  $\bar{U}_1 \in \mathcal{S}_{\text{nc}}(R_1)$  satisfy:*

$$\theta_1 \leq \min_{\tilde{U}_1 \tilde{X} \tilde{Y}_1 \in \mathcal{L}_1(\bar{U}_1)} D(\tilde{U}_1 \tilde{X} \tilde{Y}_1 \| U_1 X Y_1) \quad (5.46)$$

$$\theta_2 \leq \min_{\tilde{U} \tilde{V} \tilde{X} \tilde{Y}_1 \tilde{Y}_2 \in \mathcal{L}_2(UV)} D(\tilde{U} \tilde{V} \tilde{X} \tilde{Y}_1 \tilde{Y}_2 \| \bar{V} \bar{U} \bar{X} \bar{Y}_1 \bar{Y}_2). \quad (5.47)$$

*Proof:* The proof is given in Appendix G. It is based on the scheme of the previous Subsection 5.2.1.  $\blacksquare$

**Remark 5.3** *The exponents region in Theorem 5.7 is rectangular because  $\theta_1$  depends only on the auxiliary  $\bar{U}_1$  and  $\theta_2$  only on the pair of auxiliaries  $(U, V)$ . This implies that both exponents can be maximized at the same time without any tradeoff between them.*

*This is different in the first two Theorem 5.5 and 5.6 where both exponents depend on the same auxiliary, and therefore the regions exhibit a tension when maximizing the two exponents.*

### 5.3 Examples for coherent detection

We illustrate the benefit of cooperation with the following example.

**Example 5.1** *Consider a setup with coherent detection,  $\bar{h}_1 = 1$ , where  $X, Y_1, Y_2$  are ternary and under  $\mathcal{H} = 0$ :*

$$\begin{aligned} P_{XY_1Y_2}(0, 0, 0) &= 0.05 & P_{XY_1Y_2}(0, 0, 1) &= 0.05 & P_{XY_1Y_2}(0, 1, 0) &= 0.15 & P_{XY_1Y_2}(0, 1, 1) &= 0.083325 \\ P_{XY_1Y_2}(1, 0, 0) &= 0.05 & P_{XY_1Y_2}(1, 0, 1) &= 0.15 & P_{XY_1Y_2}(1, 1, 0) &= 0.05 & P_{XY_1Y_2}(1, 1, 1) &= 0.083335 \\ P_{XY_1Y_2}(2, 0, 0) &= 0.15 & P_{XY_1Y_2}(2, 0, 1) &= 0.05 & P_{XY_1Y_2}(2, 1, 0) &= 0.05 & P_{XY_1Y_2}(2, 1, 1) &= 0.083325 \end{aligned} \quad (5.48)$$

*whereas under  $\mathcal{H} = 1$  they are independent with same marginals as under  $\mathcal{H} = 0$ . Figure 5.4 illustrates an achievable error-exponent region obtained with Proposition 5.5 when the communication rate are  $R_1 = 0.1$  bits and  $R_2 = 1.0$  bits. It also shows the error-exponent region  $\mathcal{E}(0.1, 0)$  presented in Theorem 5.3 and the error-exponent region without cooperation when  $R_1 = 0.1$  bits derived in [41, Theorem 1]. It is difficult to directly infer the geometry of error-exponents regions from there single letter expressions. They must therefore be calculated numerically.*

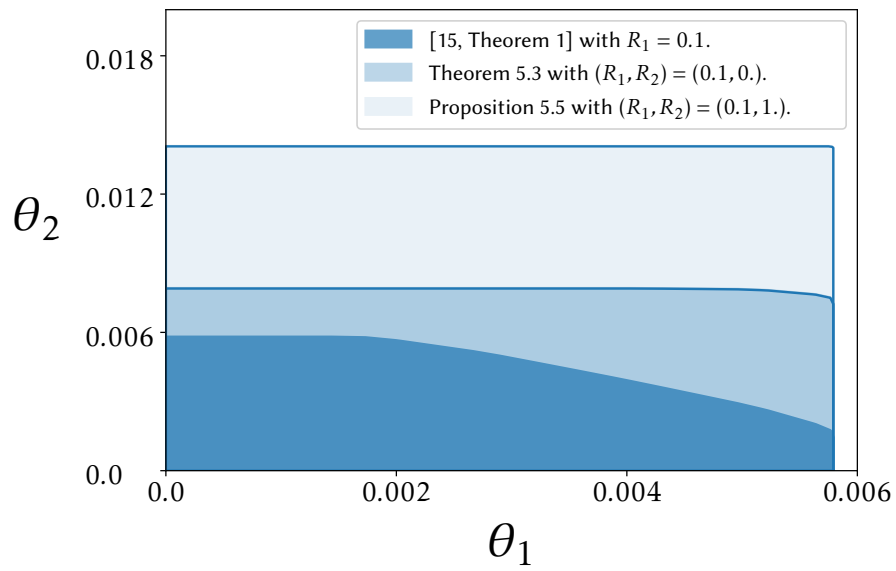


Figure 5.4: Error-exponent region of Example 5.1.





---

 On vector Gaussian hypothesis testing
 

---

Similar to its discrete memoryless counterpart, the Gaussian single-sensor single-detector hypothesis testing problem is yet to be solved for arbitrary covariance matrices. For instance, the solution of this problem is known only in few special cases such as the scalar Gaussian hypothesis testing against independence and the scalar Gaussian hypothesis testing against conditional independence. In this chapter, we study the problem of discriminating two multivariate Gaussian distributions. We find sufficient conditions under which the optimal exponent-rate function can be characterized explicitly. Perhaps interesting in its own right, in the proof of the result we also establish the optimal rate-exponent region of a slight generalization of the discrete memoryless hypothesis testing against conditional independence problem (a variation of it that accounts for distinct marginals under null and alternate hypotheses).

### 6.1 Formal problem statement

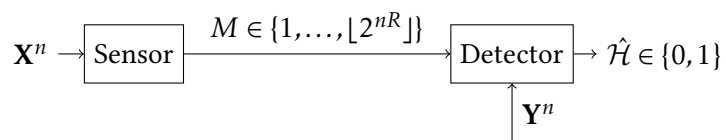


Figure 6.1: Vector Gaussian hypothesis testing problem

Consider the problem shown in Figure 6.1 in which a sensor observes  $n$  independent copies of a real-valued Gaussian vector  $\mathbf{X}$  of dimension  $m \geq 1$ , and communicates with a detector over a noise-free bit-pipe of rate  $R \geq 0$ . The detector observes  $n$  independent copies of a possibly correlated real-valued Gaussian vector

$\mathbf{Y}$  of dimension  $q \geq 1$ . The joint distribution of the pair of vectors  $(\mathbf{X}, \mathbf{Y})$  is multivariate Gaussian with zero-vector mean and covariance matrix that depends on the binary hypothesis  $\mathcal{H} \in \{0, 1\}$ . For instance,

$$\mathcal{H}=0: (\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{XY}} = \mathcal{N}(\mathbf{0}, \mathbf{K}), \quad (6.1)$$

$$\mathcal{H}=1: (\mathbf{X}, \mathbf{Y}) \sim \bar{P}_{\mathbf{XY}} = \mathcal{N}(\mathbf{0}, \bar{\mathbf{K}}), \quad (6.2)$$

with

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_X & \mathbf{K}_{XY} \\ \mathbf{K}_{XY}^\dagger & \mathbf{K}_Y \end{bmatrix} \quad (6.3)$$

and

$$\bar{\mathbf{K}} = \begin{bmatrix} \bar{\mathbf{K}}_X & \bar{\mathbf{K}}_{XY} \\ \bar{\mathbf{K}}_{XY}^\dagger & \bar{\mathbf{K}}_Y \end{bmatrix}. \quad (6.4)$$

The Sensor, which observes the sequence  $\mathbf{X}^n = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ , applies an encoding function

$$\phi_n: \mathbb{R}^{m \times n} \rightarrow \mathcal{M} = \{1, \dots, W_n\} \quad (6.5)$$

to it. It then sends the index

$$M = \phi_n(\mathbf{X}^n) \quad (6.6)$$

to the detector. In addition to  $M$ , the detector also observes the sequence  $\mathbf{Y}^n = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)$ . It applies a decision function

$$\psi_n: \mathcal{M} \times \mathbb{R}^{q \times n} \rightarrow \{0, 1\}. \quad (6.7)$$

to decides on the hypothesis

$$\hat{\mathcal{H}} = \psi_n(M, \mathbf{Y}^n). \quad (6.8)$$

The Type-I and type-II error probabilities at the detector are defined as:

$$\alpha_n = \Pr\{\hat{\mathcal{H}} = 1 \mid \mathcal{H} = 0\} \quad (6.9)$$

$$\beta_n = \Pr\{\hat{\mathcal{H}} = 0 \mid \mathcal{H} = 1\}. \quad (6.10)$$

**Definition 6.1** Given rate  $R \geq 0$ , an error-exponent  $\theta$  is said achievable if for all blocklengths  $n$  there exist functions  $\phi_n$  and  $\psi_n$  as in (6.5) and (6.7) so that the following limits hold:

$$\overline{\lim}_{n \rightarrow \infty} \alpha_n = 0, \quad (6.11a)$$

$$\theta \leq \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \beta_n \quad (6.11b)$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_n \leq R. \quad (6.11c)$$

**Definition 6.2 (exponent-rate function)** For any rate  $R \geq 0$ , the exponent-rate function is the supremum of the set of all achievable error-exponents and is denoted  $E(R)$ .

In essence, the problem of vector Gaussian hypothesis testing that we study here amounts to discriminating two covariance matrices. As we already mentioned the solution of this problem is known only in few special cases, namely the cases of testing against independence and testing against conditional independence, both in the scalar sources setting, i.e.,  $m = q = 1$  [25, Theorem 3]. For vector sources, the Gaussian hypothesis testing against conditional independence was solved recently in [39, 44].

## 6.2 Optimal exponent for a class of vector Gaussian hypothesis testing problems

Let  $\Xi$  denote any real-valued block matrix of the form

$$\Xi = \begin{bmatrix} \mathbf{K}_X & \mathbf{G} \\ \mathbf{G}^\dagger & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \mathbf{K}_Y \bar{\mathbf{K}}_Y^{-1} \bar{\mathbf{K}}_{XY}^\dagger \end{bmatrix}, \quad (6.12)$$

where the off-diagonal block  $\mathbf{G}$  is arbitrary. Also let  $\mathcal{C}$  denote the condition under which the matrices  $\mathbf{K}$  and  $\bar{\mathbf{K}}$  satisfy:

$$\mathcal{C}: \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \mathbf{K} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix}^\dagger = \arg \min_{\Xi} \log \left\{ \left| \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \bar{\mathbf{K}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix}^\dagger \right|_+ \right\} - \log \{ |\Xi|_+ \} \\ + \text{Tr} \left\{ \left( \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \bar{\mathbf{K}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \end{bmatrix}^\dagger \right)^\dagger \Xi \right\}. \quad (6.13)$$

The following theorem provides an explicit analytic expression of the exponent-rate function of the vector gaussian hypothesis testing problem of Figure 6.1 in the case in which the condition  $\mathcal{C}$  as given by (6.13) is fulfilled.

**Theorem 6.1** If  $\mathcal{C}$  is satisfied the exponent-rate function  $E(R)$  of the model of Figure 6.1 is given by

$$E(R) = \frac{m}{2} + \frac{q}{2} + \frac{1}{2} \log \frac{|\bar{\mathbf{K}}_Y|}{|\mathbf{K}_Y|} + \frac{1}{2} \text{Tr}(\bar{\mathbf{K}}_Y^{-1} \mathbf{K}_Y) \\ + \frac{1}{2} \log \frac{|\bar{\mathbf{K}}_{X|Y}|}{|\mathbf{K}_X - \mathbf{K}_{XY} \bar{\mathbf{K}}_Y^{-1} \bar{\mathbf{K}}_{XY}^\dagger (\bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \mathbf{K}_Y \bar{\mathbf{K}}_Y^{-1} \bar{\mathbf{K}}_{XY}^\dagger) + \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \mathbf{K}_{XY}^\dagger|} \\ + \frac{1}{2} \text{Tr} \left( \bar{\mathbf{K}}_{X|Y}^\dagger \left( \mathbf{K}_X - \mathbf{K}_{XY} \bar{\mathbf{K}}_Y^{-1} \bar{\mathbf{K}}_{XY}^\dagger (\bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \mathbf{K}_Y \bar{\mathbf{K}}_Y^{-1} \bar{\mathbf{K}}_{XY}^\dagger)^\dagger \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \mathbf{K}_{XY}^\dagger \right) \right) \\ + \max \min \left\{ R + \frac{1}{2} \log |\mathbf{I} - \Omega \mathbf{K}_{X|Y}|, \quad \frac{1}{2} \log \left| \mathbf{I} + \Omega \mathbf{K}_{XY} \left( \mathbf{K}_Y^{-1} - \mathbf{K}_Y^{-1} \bar{\mathbf{K}}_Y \bar{\mathbf{K}}_{XY}^\dagger \bar{\mathbf{K}}_{XY} \bar{\mathbf{K}}_Y^{-1} \right) \mathbf{K}_{XY}^\dagger \right| \right\},$$

where the maximization in the last term is over all matrices  $\mathbf{0} \leq \Omega \leq \mathbf{K}_{X|Y}^\dagger$  and  $\bar{\mathbf{K}}_{XY}^\dagger$  designates the Moore-Penrose pseudo inverse of  $\bar{\mathbf{K}}_{XY}$ .

*Proof:* see section 6.3 ■

**Remark 6.1** *The solution of the special case of scalar Gaussian sources, i.e.,  $m = q = 1$ , can be found by specializing that of the Gaussian many-help one hypothesis testing against independence problem of [25, Theorem 7] to the setting. More specifically, if  $Y$  is such that  $Y = X + N$  under the null hypothesis, where the noise  $N$  is independent of  $X$ ; and  $Y$  is independent of  $X$  under the alternate hypothesis, the optimal exponent-rate function is given by*

$$E(R) = \frac{1}{2} \log \left( \frac{\sigma_X^2 + \sigma_N^2}{\sigma_N^2 + e^{-2R} \sigma_X^2} \right) \quad (6.14)$$

where  $\sigma_X^2$  and  $\sigma_N^2$  designate the variance of  $X$  and the variance of the noise  $N$ , respectively. Through straightforward algebra it can be easily shown that this result can be recovered from our Theorem 6.1 (Note that the constraint  $C$  is always fulfilled in this case).

### 6.3 Proof of Theorem 6.1

For reasons that will become clearer, we first consider a slight generalization of the discrete memoryless single-sensor single-detector hypothesis testing against conditional independence problem (the generalization accounts for that the marginals are not restricted to be identical) for which we provide a single-letter characterization of its exponent-rate function.

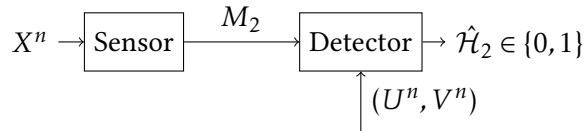


Figure 6.2: Problem  $P_2$ .

Consider the problem of Figure 6.2. Here,  $(X, U, V)$  denotes a triplet of discrete memoryless sources whose joint p.m.f. depends on the binary hypothesis  $\mathcal{H} \in \{0, 1\}$ . More precisely

$$\mathcal{H}=0: \quad (X, U, V) \sim P_{XUV} \quad (6.15)$$

$$\mathcal{H}=1: \quad (X, U, V) \sim \bar{P}_{XUV} = \bar{P}_U \bar{P}_{X|U} \bar{P}_{V|U}. \quad (6.16)$$

The definitions of acceptance and rejection regions, as well as the exponent-rate function, are similar to those of previous chapters; and, thus, we omit them here for reasons of brevity.

**Lemma 6.2** *If the joint p.m.f  $P_{XU}$  under the null hypothesis satisfies*

$$P_{XU} = \arg \min_{\substack{\bar{P}_{XU}: \bar{P}_X = P_X \\ \bar{P}_U = P_U}} D(\bar{P}_{XU} \| \bar{P}_{XU}), \quad (6.17)$$

the rate exponent function is given by

$$E(R) = D(P_{XU}||\bar{P}_{XU}) + D(P_V||\bar{P}_V|U) + \max I(S; V|U) \quad (6.18)$$

where in (6.18) the maximization is over all conditionals  $P_{S|X}$  for which  $I(S; X|U) \leq R$ .

*Proof of Lemma 6.2:*

We have

$$E(R) = \varliminf_{n \rightarrow \infty} E_n(R), \quad (6.19)$$

where

$$E_n(R) = \frac{1}{n} D(P_{\phi_n(X^n)U^n V^n} || \bar{P}_{\phi_n(\bar{X}^n)\bar{U}^n \bar{V}^n}). \quad (6.20)$$

It is easy to see that

$$\begin{aligned} \varliminf_{n \rightarrow \infty} E_n(R) &= \varliminf_{n \rightarrow \infty} \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n V^n} || \bar{P}_{\phi_n(\bar{X}^n)\bar{U}^n \bar{V}^n}) \\ &= \varliminf_{n \rightarrow \infty} \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} || \bar{P}_{\phi_n(X^n)U^n}) + D(P_V || \bar{P}_V|U) + \frac{1}{n} I(\phi_n(X^n); V^n|U^n) \\ &\stackrel{(a)}{\leq} D(P_V || \bar{P}_V|U) + \varliminf_{n \rightarrow \infty} \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} || \bar{P}_{\phi_n(X^n)U^n}) \\ &\quad + \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} I(\phi_n(X^n); V^n|U^n) \\ &\stackrel{(b)}{\leq} D(P_V || \bar{P}_V|U) + D(P_{XU} || \bar{P}_{XU}) + \varliminf_{n \rightarrow \infty} \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} I(\phi_n(X^n); V^n|U^n) \\ &\stackrel{(c)}{=} D(P_{XU} || \bar{P}_{XU}) + D(P_V || \bar{P}_V|U) + \max_{P_{S|X} : I(S; X|U) \leq R} I(S; V|U) \end{aligned} \quad (6.21)$$

where (a) holds since

$$\max_{x \in \mathcal{X}} f(x) + g(x) \leq \max_{x \in \mathcal{X}} f(x) + \max_{x \in \mathcal{X}} g(x)$$

(b) holds since the decentralized error-exponent function

$$\varliminf_{n \rightarrow \infty} \max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} || \bar{P}_{\phi_n(X^n)U^n}),$$

is always less than the centralized error-exponent  $D(P_{XU} || \bar{P}_{XU})$ ; (c) holds by noting that the term

$$(I(\phi_n(X^n); V^n|U^n))/n,$$

is the exponent of a hypothesis testing against conditional independence problem at rate  $R$  as given by [25, Theorem 3],

$$\max_{\phi_n : \log_2 |\phi_n| \leq nR} \frac{1}{n} I(\phi_n(X^n); V^n|U^n) = \max_{P_{S|X} : I(S; X|U) \leq R} I(S; V|U). \quad (6.22)$$

Combining (6.19) and (6.21) we get

$$E(R) \leq D(P_{XU} || \bar{P}_{XU}) + D(P_V || \bar{P}_V|U) + \max_{P_{S|X} : I(S; X|U) \leq R} I(S; V|U). \quad (6.23)$$

The proof ends by noticing that if

$$P_{XU} = \underset{\substack{\tilde{P}_{XU}: \tilde{P}_X = P_X \\ \tilde{P}_U = P_U}}{\operatorname{arg\,min}} D(\tilde{P}_{XU} \| \bar{P}_{XU}), \quad (6.24)$$

then both inequalities (a) and (b) in (6.21) hold with equality. In order to see that (b) holds with equality when (6.24) is satisfied observe that on the one hand we have

$$\begin{aligned} \max_{\phi_n: |\phi_n| \leq 2} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) &\leq \max_{\phi_n: \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) \\ &\leq D(P_{XU} \| \bar{P}_{XU}); \end{aligned} \quad (6.25)$$

and on the other hand we also have

$$\begin{aligned} \lim_{n \rightarrow \infty} \max_{\phi_n: |\phi_n| \leq 2} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) &\stackrel{(d)}{=} \min_{\substack{\tilde{P}_{XU}: \tilde{P}_X = P_X \\ \tilde{P}_U = P_U}} D(\tilde{P}_{XU} \| \bar{P}_{XU}) \\ &\stackrel{(e)}{=} D(P_{XU} \| \bar{P}_{XU}) \end{aligned} \quad (6.26)$$

where (d) holds by using [14, Theorem 5] and (e) holds by using (6.24). It remains to show that (a) in (6.21) holds with equality. To see this observe that on the one hand we have

$$\begin{aligned} \max_{\phi_n: \log |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) + I(\phi_n(X^n); V^n | U^n) \\ \leq \max_{\phi_n: \log |\phi_n| \leq nR} D(P_{XU} \| \bar{P}_{XU}) + I(\phi_n(X^n); V^n | U^n) \\ = D(P_{XU} \| \bar{P}_{XU}) + \max_{\phi_n: \log |\phi_n| \leq nR} I(\phi_n(X^n); V^n | U^n); \end{aligned} \quad (6.27)$$

and on the other hand, denoting by  $\phi_n^*$  the maximizer in the left hand side of (6.25), we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \max_{\phi_n: \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) + \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \\ = \lim_{n \rightarrow \infty} \max_{\phi_n: \log_2 |\phi_n| \leq nR+1} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) + \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \\ \geq \lim_{n \rightarrow \infty} \max_{(\tilde{\phi}_n, \phi_n^*): \log_2 |\tilde{\phi}_n| \leq nR} \frac{1}{n} D(P_{\tilde{\phi}_n(X^n)U^n} \| \bar{P}_{\tilde{\phi}_n(X^n)U^n}) + \frac{1}{n} I(\tilde{\phi}_n(X^n); V^n | U^n) \\ = \lim_{n \rightarrow \infty} \frac{1}{n} D(P_{\phi_n^*(X^n)U^n} \| \bar{P}_{\phi_n^*(X^n)U^n}) + \max_{\phi_n: \log_2 |\phi_n| \leq nR+1} \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \\ = D(P_{XU} \| \bar{P}_{XU}) + \max_{\phi_n: \log_2 |\phi_n| \leq nR+1} \frac{1}{n} I(\phi_n(X^n); V^n | U^n). \end{aligned}$$

Using the above we get

$$\begin{aligned} \lim_{n \rightarrow \infty} \max_{\phi_n: \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) + \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \\ = D(P_{XU} \| \bar{P}_{XU}) + \lim_{n \rightarrow \infty} \max_{\phi_n: \log_2 |\phi_n| \leq nR+1} \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \\ = \lim_{n \rightarrow \infty} \max_{\phi_n: \log_2 |\phi_n| \leq nR} \frac{1}{n} D(P_{\phi_n(X^n)U^n} \| \bar{P}_{\phi_n(X^n)U^n}) + \max_{\phi_n: \log_2 |\phi_n| \leq nR} \frac{1}{n} I(\phi_n(X^n); V^n | U^n) \end{aligned} \quad (6.28)$$

where the last equality follows since (b) in (6.21) holds with equality as we have already shown. ■

We now turn to the proof of Theorem 6.1. The result of Lemma 6.2 extends easily to the case of continuous alphabets using standard discretization (quantization) techniques. Let

$$\mathbf{U} = \mathbb{E}_{\bar{p}}[\mathbf{X}|\mathbf{Y}] \quad (6.29)$$

$$\mathbf{V} = \mathbf{Y}. \quad (6.30)$$

We can see that under  $\mathcal{H} = 1$ ,

$$\begin{aligned} I(\mathbf{X}; \mathbf{U}) &= H(\mathbf{X}) - H(\mathbf{X}|\mathbf{U}) \\ &= H(\mathbf{X}) - \log |\mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{U}}\mathbf{K}_{\mathbf{U}}^+\mathbf{K}_{\mathbf{X}\mathbf{U}}^+| \end{aligned} \quad (6.31)$$

$$= H(\mathbf{X}) - \log |\mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{Y}}\mathbf{K}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+ (\mathbf{K}_{\mathbf{X}\mathbf{Y}}\mathbf{K}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+)^+ \mathbf{K}_{\mathbf{X}\mathbf{Y}}\mathbf{K}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+| \quad (6.32)$$

$$= H(\mathbf{X}) - \log |\mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{Y}}\mathbf{K}_{\mathbf{Y}}^+\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+| \quad (6.33)$$

$$= I(\mathbf{X}; \mathbf{Y}), \quad (6.34)$$

and so:

$$\mathbf{X} \text{ --- } \mathbf{U} \text{ --- } \mathbf{V} \quad (6.35)$$

is a Markov chain.

Consider now the problem of testing against conditional independence in which the sensor observes  $\mathbf{X}$  and the detector observes the pair  $(\mathbf{U}, \mathbf{V})$  and aims to guess whether  $\mathbf{X}$  and  $\mathbf{V}$  are independent conditionally on  $\mathbf{U}$  or not. That is, the problem of Figure 6.2 in which  $\mathbb{E}_{\bar{p}}[\mathbf{X}|\mathbf{Y}]$  plays the role of  $\mathbf{U}$  and  $\mathbf{Y}$  plays the role of  $\mathbf{V}$ . In the remaining of this proof this problem will be referred to as Problem  $P_2$ . For given  $R$  the rate exponent function of the original vector Gaussian problem of Figure 6.1 is equal to that of the Problem  $P_2$ , i.e.,

$$E(R) = E_2(R). \quad (6.36)$$

(The interested reader may refer to Appendix H for a proof of (6.36)).

Now, invoking Lemma 6.2 on Problem  $P_2$  we get that if

$$P_{\mathbf{X}\mathbf{U}} = \arg \min_{\substack{\tilde{P}_{\mathbf{X}\mathbf{U}}: \tilde{P}_{\mathbf{X}}=P_{\mathbf{X}} \\ \tilde{P}_{\mathbf{U}}=P_{\mathbf{U}}}} D(\tilde{P}_{\mathbf{X}\mathbf{U}}||\tilde{P}_{\mathbf{X}\mathbf{U}}), \quad (6.37)$$

then we have

$$\begin{aligned} E(R) &= D(P_{\mathbf{X}\mathbf{U}}||\tilde{P}_{\mathbf{X}\mathbf{U}}) + D(P_{\mathbf{V}}||\tilde{P}_{\mathbf{V}}|\mathbf{U}) + \max_{P_{S|\mathbf{X}}: I(S;\mathbf{X}|\mathbf{U}) \leq R} I(S; \mathbf{V}|\mathbf{U}) \\ &= D(P_{\mathbf{X}\mathbf{U}}||\tilde{P}_{\mathbf{X}\mathbf{U}}) + D(P_{\mathbf{Y}}||\tilde{P}_{\mathbf{Y}}|\mathbf{U}) + \max_{P_{S|\mathbf{X}}: I(S;\mathbf{X}|\mathbf{U}) \leq R} I(S; \mathbf{Y}|\mathbf{U}) \\ &= D(P_{\mathbf{X}}||\tilde{P}_{\mathbf{X}}|\mathbf{U}) + D(P_{\mathbf{Y}}||\tilde{P}_{\mathbf{Y}}) + \max_{P_{S|\mathbf{X}}: I(S;\mathbf{X}|\mathbf{U}) \leq R} I(S; \mathbf{Y}|\mathbf{U}). \end{aligned} \quad (6.38)$$



In what follows we first show that the minimizer in

$$\arg \min_{\substack{\tilde{P}_{\mathbf{X}\mathbf{U}}: \tilde{P}_{\mathbf{X}}=P_{\mathbf{X}} \\ \tilde{P}_{\mathbf{U}}=P_{\mathbf{U}}}} D(\tilde{P}_{\mathbf{X}\mathbf{U}}\|\tilde{P}_{\mathbf{X}\mathbf{U}}) \quad (6.39)$$

is a multivariate Gaussian distribution on the pair  $(\mathbf{X}, \mathbf{U})$ . To see this let  $\tilde{P}_{\mathbf{X}\mathbf{U}}$  with  $\tilde{P}_{\mathbf{X}} = P_{\mathbf{X}}$  and  $\tilde{P}_{\mathbf{U}} = P_{\mathbf{U}}$  be given. Also let  $\tilde{P}_{\mathbf{X}\mathbf{U}}^G$  be a multivariate Gaussian with the same covariance matrix as that of  $\tilde{P}_{\mathbf{X}\mathbf{U}}$ . Then we have

$$\begin{aligned} D(\tilde{P}_{\mathbf{X}\mathbf{U}}\|\tilde{P}_{\mathbf{X}\mathbf{U}}) &= -h(\tilde{P}_{\mathbf{X}\mathbf{U}}) - \mathbb{E}_{\tilde{P}}\{\log \tilde{P}_{\mathbf{X}\mathbf{U}}\} \\ &\geq -h(\tilde{P}_{\mathbf{X}\mathbf{U}}^G) - \mathbb{E}_{\tilde{P}^G}\{\log \tilde{P}_{\mathbf{X}\mathbf{U}}\} \end{aligned} \quad (6.40)$$

where the inequality holds since among all distributions with the same covariance matrix the Gaussian maximizes entropy. Through straightforward algebra it can be shown that the condition

$$P_{\mathbf{X}\mathbf{U}} = \arg \min_{\substack{\tilde{P}_{\mathbf{X}\mathbf{U}}: \tilde{P}_{\mathbf{X}}=P_{\mathbf{X}} \\ \tilde{P}_{\mathbf{U}}=P_{\mathbf{U}}}} D(\tilde{P}_{\mathbf{X}\mathbf{U}}\|\tilde{P}_{\mathbf{X}\mathbf{U}}) \quad (6.41)$$

is equivalent to the constraint C as given by (6.13). (The interested reader may refer to Appendix I)

It remains to evaluate the right hand side (RHS) of (6.38). The relative entropy term in the RHS of (6.38) can be computed as

$$D(P_{\mathbf{Y}}\|\tilde{P}_{\mathbf{Y}}) = \frac{q}{2} + \frac{1}{2} \log \frac{|\tilde{\mathbf{K}}_{\mathbf{Y}}|}{|\mathbf{K}_{\mathbf{Y}}|} + \frac{1}{2} \text{Tr}(\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1} \mathbf{K}_{\mathbf{Y}}). \quad (6.42)$$

Similarly, the conditional entropy term of the RHS of (6.38) can be computed as

$$\begin{aligned} D(P_{\mathbf{X}}\|\tilde{P}_{\mathbf{X}}|\mathbf{U}) &= \frac{m}{2} + \frac{1}{2} \log \frac{|\tilde{\mathbf{K}}_{\mathbf{X}|\mathbf{Y}}|_+}{|\mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}^+ (\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}^+ + \tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+|_+} \\ &\quad + \frac{1}{2} \text{Tr}(\tilde{\mathbf{K}}_{\mathbf{X}|\mathbf{Y}}^+ (\mathbf{K}_{\mathbf{X}} - \mathbf{K}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}^+ (\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}^+ + \tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}\tilde{\mathbf{K}}_{\mathbf{Y}}^{-1}\mathbf{K}_{\mathbf{X}\mathbf{Y}}^+)). \end{aligned} \quad (6.43)$$

It remains to evaluate the maximum of the mutual information term  $I(S; \mathbf{Y}|\mathbf{U})$  where the maximization is over all test channels  $P_{S|\mathbf{X}}$  for which  $I(S; \mathbf{X}|\mathbf{U}) \leq R$ . This is done in [39, Theorem 5] where the maximizing  $S$  is shown to be multivariate Gaussian with

$$\begin{aligned} \max_{P_{S|\mathbf{X}}: I(S; \mathbf{X}|\mathbf{U}) \leq R} I(S; \mathbf{Y}|\mathbf{U}) &= \max \min \left\{ R + \frac{1}{2} \log |\mathbf{I} - \mathbf{O}\mathbf{K}_{\mathbf{X}|\mathbf{Y}}|, \right. \\ &\quad \left. \frac{1}{2} \log \left| \mathbf{I} + \mathbf{O}\mathbf{K}_{\mathbf{X}\mathbf{Y}} \left( \mathbf{K}_{\mathbf{Y}}^{-1} - \mathbf{K}_{\mathbf{Y}}^{-1} \tilde{\mathbf{K}}_{\mathbf{Y}} \tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}}^+ \tilde{\mathbf{K}}_{\mathbf{X}\mathbf{Y}} \tilde{\mathbf{K}}_{\mathbf{Y}}^{-1} \right) \mathbf{K}_{\mathbf{X}\mathbf{Y}}^+ \right| \right\}, \end{aligned} \quad (6.44)$$

Combining (6.42), (6.43) and (6.44) we get the RHS of (6.44); and this completes the proof of Theorem 6.1. ■

## 6.4 Discussion

The constraint C as given by (6.13) may seem too restrictive, but it is not. In what follows we show that this constraint may be satisfied of a large class of sources even in the case in which  $m = 1$  and  $q = 2$ . Let  $X$  be a scalar source that is observed at the sensor and  $\mathbf{Y} = (Y_1, Y_2)$  a 2-dimensional source that is observed at the detector. For convenience, let

$$\mathbf{K} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY_1} & \sigma_{XY_2} \\ \sigma_{XY_1} & \sigma_{Y_1}^2 & \sigma_{Y_1 Y_2} \\ \sigma_{XY_2} & \sigma_{Y_1 Y_2} & \sigma_{Y_2}^2 \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{K}} = \begin{bmatrix} \bar{\sigma}_X^2 & \bar{\sigma}_{XY_1} & \bar{\sigma}_{XY_2} \\ \bar{\sigma}_{XY_1} & \bar{\sigma}_{Y_1}^2 & \bar{\sigma}_{Y_1 Y_2} \\ \bar{\sigma}_{XY_2} & \bar{\sigma}_{Y_1 Y_2} & \bar{\sigma}_{Y_2}^2 \end{bmatrix}. \quad (6.45)$$

Also, let

$$a = (\bar{\sigma}_{XY_2} \bar{\sigma}_{Y_1 Y_2} - \bar{\sigma}_{XY_1} \bar{\sigma}_{Y_2}^2) \quad \text{and} \quad b = (\bar{\sigma}_{XY_1} \bar{\sigma}_{Y_1 Y_2} - \bar{\sigma}_{XY_2} \bar{\sigma}_{Y_1}^2). \quad (6.46)$$

For this example the constraint C as given by (6.13) reduces to

$$i) \quad \sigma_X^2 = \bar{\sigma}_X^2, \quad (6.47a)$$

$$ii) \quad a(\sigma_{XY} - \bar{\sigma}_{XY}) + b(\sigma_{XZ} - \bar{\sigma}_{XZ}) = 0 \quad (6.47b)$$

$$iii) \quad a^2(\sigma_Y^2 - \bar{\sigma}_Y^2) + 2ab(\sigma_{YZ} - \bar{\sigma}_{YZ}) + b^2(\sigma_Z^2 - \bar{\sigma}_Z^2) = 0 \quad (6.47c)$$

For example, if all components have unit variance under both  $P$  and  $\bar{P}$ , i.e.,  $\sigma_X^2 = \sigma_{Y_1}^2 = \sigma_{Y_2}^2 = 1$  and  $\bar{\sigma}_X^2 = \bar{\sigma}_{Y_1}^2 = \bar{\sigma}_{Y_2}^2 = 1$  then all definite positive matrices  $\mathbf{K}$  and  $\bar{\mathbf{K}}$  of the form

$$\mathbf{K} = \begin{bmatrix} 1 & a_{12} & h(\bar{a}_{12}, \bar{a}_{13}, \bar{a}_{23}, a_{12}) \\ a_{12} & 1 & \bar{a}_{23} \\ h(\bar{a}_{12}, \bar{a}_{13}, \bar{a}_{23}, a_{12}) & \bar{a}_{23} & 1 \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{K}} = \begin{bmatrix} 1 & a_{12} & \bar{a}_{13} \\ a_{12} & 1 & \bar{a}_{23} \\ \bar{a}_{13} & \bar{a}_{23} & 1 \end{bmatrix} \quad (6.48)$$

for some arbitrary parameters  $a_{12}$ ,  $\bar{a}_{12}$ ,  $\bar{a}_{13}$ ,  $\bar{a}_{23}$ , satisfy the constraint (6.47). Here

$$h(x, y_1, y_2, t) = y_1 - (t - y_2) \frac{y_1 y_2 - x}{x y_2 - y_1}. \quad (6.49)$$

**Example 6.1** Let

$$\mathbf{K} = \begin{bmatrix} 1. & 0.4 & \alpha \\ 0.4 & 1. & 0.1 \\ \alpha & 0.1 & 1. \end{bmatrix} \quad \text{and} \quad \bar{\mathbf{K}} = \begin{bmatrix} 1. & 0.1 & -0.8 \\ 0.1 & 1. & 0.1 \\ -0.8 & 0.1 & 1. \end{bmatrix}, \quad (6.50)$$

with  $\alpha \approx -0.73333$ . We can see that (6.47) is fulfilled. Figure 6.3 shows the evolution of the optimal exponent  $E$  as a function of the communication rate  $R$  as given by Theorem 6.1 for this example. It is interesting to observe that Han's scheme [14, Theorem 2] is strictly suboptimal for this example<sup>1</sup>, whereas Shimokawa-Han-Amari scheme [32, Theorem 1] is optimal.

<sup>1</sup>In the figure, Han's exponent as given by [14, Theorem 2] is computed using Gaussian test channels  $P_{U|X}$  and Gaussian  $\tilde{U}$

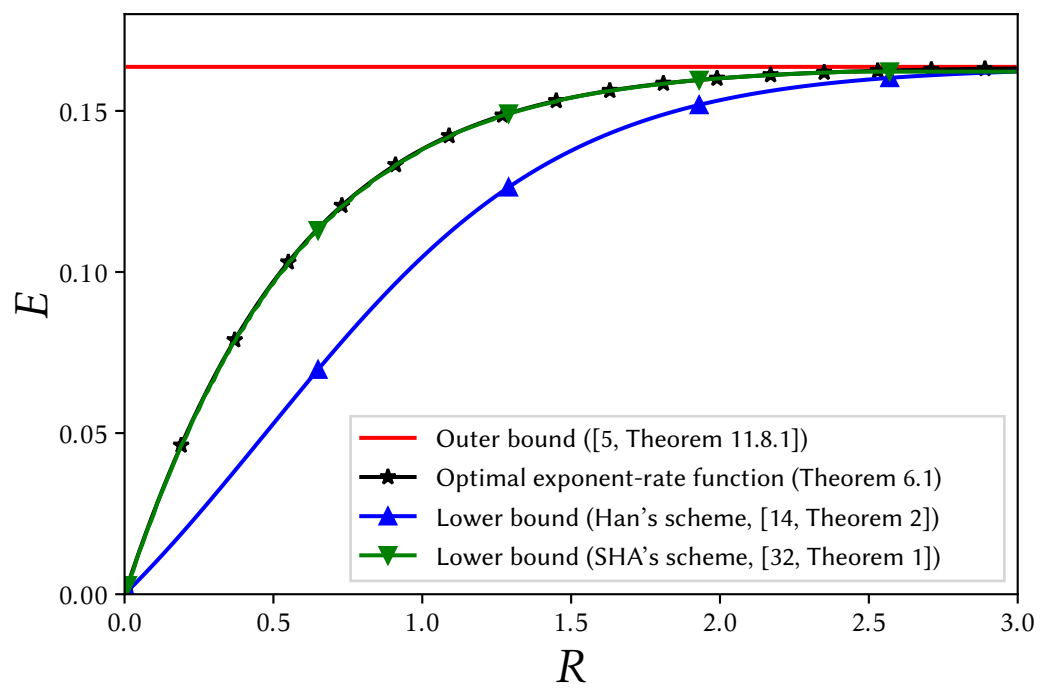


Figure 6.3: Rate-exponent region for Example 6.1.

## 7.1 Coherent detection and non-coherent detection with cooperation

In Chapters 4 and Chapters 5 we have studied hypothesis testing Type II error exponents in the context of coherent and concurrent detection with and without cooperation. In a first part, we studied the case of fixed-length communication and characterized the exponent regions when the Type I error probabilities are constrained to be less than a fixed threshold. These are rectangular in the case of a coherent detection or in the case of concurrent detection when the sensor can send more than one bit or when it cannot distinguish the two hypotheses. In this case each detector behaves as if it were the only one in the system. When the detection is concurrent and the sensor can send only one bit and can distinguish between the two hypotheses, there is a trade-off between the two exponents.

It may be interesting to generalize this work to the scenario in which  $K$  detectors share information from a single sensor in a network where communications are of fixed length. This could lead to a complete modeling of the sensor network as encountered into the IoT. We can also consider the problem where the sensor must compress its observation without knowing what test will be performed at the end but only that this test belongs to some family.

In a second part we studied for this same system the exponent-rate region for positive rate communication when the Type I error probabilities are constrained to decrease to zero with the number of observation going to infinity. It appears that in the concurrent case and when the sensor can distinguish the two hypotheses, the acceptance region is rectangular. A trade-off between the two exponents appears when the two hypotheses cannot be distinguished. In some particular cases, we have characterized this region and in the general case we have studied a communication scheme without binning. However, it is clear

that binning helps in the general case.

## 7.2 Distributed hypothesis testing in the Gaussian case

In chapter 6, we restricted attention to the vector Gaussian single-sensor single-detection hypothesis testing problem. In this case, the problem amounts to distinguishing two multivariate Gaussian distributions using only partial (compressed) information. Relying heavily on recent developments in this area [44, 39] we found sufficient conditions under which the optimal exponent-rate function can be characterized analytically.

---

Proof of Theorem 2.7

---

This proof follows along the same steps as [31, Theorem 2] proof, we reproduce this steps here for matter of completeness. For every  $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$  such that

$$P_X^{\otimes n}(C) \geq \frac{1 - \epsilon}{M_n}, \quad P_Y^{\otimes n}(F) \geq \frac{1 - \epsilon}{M_n}, \quad (\text{A.1})$$

where

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log M_n = 0, \quad (\text{A.2})$$

lemma 2.6 applies and there exist sequences  $k_n = o(n)$  and  $\gamma_n = o(1)$  such that

$$P_X^{\otimes n}(\Gamma^{k_n} C) \geq 1 - \gamma_n, \quad P_Y^{\otimes n}(\Gamma^{k_n} F) \geq 1 - \gamma_n \quad (\text{A.3})$$

The sequences  $k_n$  and  $\gamma_n$  depend only on  $|\mathcal{X}|$ ,  $|\mathcal{Y}|$ ,  $\epsilon$  and  $M_n$ . Using the fact that  $\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1$  and (A.3), we get

$$P_{XY}^{\otimes n}(\Gamma^{k_n} C \times \Gamma^{k_n} F) \geq 1 - 2\gamma_n. \quad (\text{A.4})$$

The same is true for the set  $\mathcal{T}_\eta^n(P_{XY})$ , where  $\eta = \eta_n = n^{-1/4}$ ,

$$P_{XY}^{\otimes n}(\mathcal{T}_\eta^n(P_{XY})) \geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4n\eta_n^2} = 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4n^{1/2}}. \quad (\text{A.5})$$

Using (A.4) and (A.5), it is clear that for large  $n$ ,

$$P_{XY}^{\otimes n}((\Gamma^{k_n} C \times \Gamma^{k_n} F) \cap \mathcal{T}_\eta^n(P_{XY})) \geq \frac{1}{2} \quad (\text{A.6})$$

Hereafter, for the sake of simplicity we abuse notation and dismiss the subscript  $n$  from  $k_n$  and  $\gamma_n$  and write simply  $k_n := k$  and  $\gamma_n := \gamma$ .

Now, using the fact that

$$\begin{aligned} \mathcal{T}_\eta^n(P_{XY}) = & \bigcup_{\substack{P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \\ |P_{\hat{X}\hat{Y}} - P_{XY}| \leq \eta P_{XY}}} \mathcal{T}^n(P_{\hat{X}\hat{Y}}) \end{aligned} \quad (\text{A.7})$$

and observing that all the elements of the type class  $\mathcal{T}^n(P_{\hat{X}\hat{Y}})$  are equiprobable under an i.i.d. measure, (A.6) can be rewritten as

$$\sum P_{XY}^{\otimes n}(\mathcal{T}^n(P_{\hat{X}\hat{Y}})) \frac{|(\Gamma^k C \times \Gamma^k F) \cap \mathcal{T}^n(P_{\hat{X}\hat{Y}})|}{|\mathcal{T}^n(P_{\hat{X}\hat{Y}})|} \geq \frac{1}{2} \quad (\text{A.8})$$

where, in the last inequality, the sum is over all types  $P_{\hat{X}\hat{Y}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$  such that,  $P_{\hat{X}\hat{Y}} \ll P_{XY}$  and

$$|P_{\hat{X}\hat{Y}} - P_{XY}| \leq \eta P_{XY}. \quad (\text{A.9})$$

As  $P_{XY}^{\otimes n}(\mathcal{T}^n(P_{\hat{X}\hat{Y}})) \leq 1$  and  $|\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}$ , there must exist at least one type  $P_{\hat{X}\hat{Y}}$  for which

$$\frac{|(\Gamma^k C \times \Gamma^k F) \cap \mathcal{T}^n(P_{\hat{X}\hat{Y}})|}{|\mathcal{T}_0^n(P_{\hat{X}\hat{Y}})|} \geq \frac{1}{2} \exp(-n\epsilon_n) \quad (\text{A.10})$$

with  $\mu_n \triangleq \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log(n+1) = \mathcal{O}(n^{-1} \log(n+1)) \rightarrow 0$  as  $n \rightarrow \infty$ . The equiprobability property is also true for the i.i.d measure  $Q_{XY}^{\otimes n}$ . Thus

$$\begin{aligned} Q_{XY}^{\otimes n}(\Gamma^k C \times \Gamma^k F) & \geq Q_{XY}^{\otimes n}((\Gamma^k C \times \Gamma^k F) \cap \mathcal{T}^n(P_{\hat{X}\hat{Y}})) \\ & = Q_{XY}^{\otimes n}(\mathcal{T}_0^n(P_{\hat{X}\hat{Y}})) \frac{|(\Gamma^k C \times \Gamma^k F) \cap \mathcal{T}^n(P_{\hat{X}\hat{Y}})|}{|\mathcal{T}_0^n(P_{\hat{X}\hat{Y}})|} \\ & \geq \frac{1}{2} \exp(-n\mu_n) Q_{XY}^{\otimes n}(\mathcal{T}_0^n(P_{\hat{X}\hat{Y}})). \end{aligned} \quad (\text{A.11})$$

where the last inequality follows using (A.10).

We thus have shown that the sets  $\Gamma^k C \times \Gamma^k F$  and  $\mathcal{T}^n(P_{\hat{X}\hat{Y}})$  have the same exponential order under the measure  $Q_{XY}^{\otimes n}$ . In what follows, we show that the same is true for the sets  $\Gamma^k C \times \Gamma^k F$  and  $C \times F$ .

Consider now an arbitrary triple of sequences  $(u^n, v^n) \in \Gamma^k C \times \Gamma^k F$ . By definition of the sets  $\Gamma^k C$ ,  $\Gamma^k F$  there exists at least one element  $(x^n, y^n) \in C \times F$  such that the vectors  $(u^n, v^n)$  and  $(x^n, y^n)$  differ in at most  $2k$  locations. Thus,

$$Q_{XY}^{\otimes n}(u^n, v^n) = \prod_{i=1}^n Q_{XY}^{\otimes n}(u_i, v_i) \leq \rho^{-2k} \prod_{i=1}^n Q_{XY}(x_i, y_i) = \rho^{-2k} Q_{XY}^{\otimes n}(x^n, y^n) \quad (\text{A.12})$$

where

$$\rho := \min_{x \in \mathcal{X}, y \in \mathcal{Y}} Q(x, y). \quad (\text{A.13})$$

As  $(u^n, v^n)$  ranges over  $\Gamma^k C \times \Gamma^k F$  each element of  $(x^n, y^n) \in C \times F$  will be selected as the closest neighbor at most  $|\Gamma^k(x^n)| \cdot |\Gamma^k(y_1^n)|$  times. Thus

$$Q_{XY}^{\otimes n}(\Gamma^k C \times \Gamma^k F) \leq \rho^{-2k} |\Gamma^k(x^n)| \cdot |\Gamma^k(y_1^n)| Q_{XY}^{\otimes n}(C \times F). \quad (\text{A.14})$$

From [7, Lemma 5.1], we have

$$|\Gamma^k(x^n)| \leq \exp \left[ -n \left( h_2\left(\frac{k}{n}\right) + \frac{k}{n} \log |\mathcal{X}| \right) \right] := \exp(n\xi_n) \quad (\text{A.15})$$

with  $h_2(\cdot)$  denotes the binary entropy function and  $\xi_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Hence:

$$Q_{XY}^{\otimes n}(\Gamma^k C \times \Gamma^k F) \leq \exp(n\xi_n) Q_{XY}^{\otimes n}(C \times F) \quad (\text{A.16})$$

Combining (A.11) and (A.16), we obtain

$$\begin{aligned} Q_{XY}^{\otimes n}(C \times F) &\geq \exp(-n\xi_n) Q_{XY}^{\otimes n}(\Gamma^k C \times \Gamma^k F) \\ &\geq \frac{1}{2} \exp \left[ -n(\xi_n + \mu_n) \right] Q_{XY}^{\otimes n}(\mathcal{T}^n(Q_{XY})) \\ &\geq \frac{(n+1)^{|\mathcal{X}||\mathcal{Y}|}}{2} \exp \left[ -n \left( D(P_{\hat{X}\hat{Y}} \| Q_{XY}) + \xi_n + \mu_n \right) \right] \\ &= \exp \left[ -n \left( D(P_{\hat{X}\hat{Y}} \| Q_{XY}) + \nu_n \right) \right]. \end{aligned} \quad (\text{A.17})$$

where  $\nu_n := \nu_n(\rho, \epsilon, M_n, |\mathcal{X}|, |\mathcal{Y}|) \rightarrow 0$  as  $n \rightarrow \infty$ . this completes the proof of the proof of Theorem 2.7.





---

Proof of Proposition 4.1

---

We present first a coding scheme for coherent detection that we will analyze to prove the achievability part. Fix a small  $\mu > 0$ .

Sensor: Assume it observes the source sequence  $X^n = x^n$ , then it checks whether

$$x^n \in \mathcal{T}_\mu(P_X)$$

if the test is successful it sends  $M_1 = 1$  else  $M_1 = 0$ .

Detector 1: If  $M_1 = 0$ , Detector 1 decides on the alternative hypothesis:

$$\hat{\mathcal{H}}_1 = 1.$$

If  $M_1 = 1$  and given that  $Y_1^n = y_1^n$ , Detector 1 checks whether

$$y_1^n \in \mathcal{T}_\mu(P_{Y_1}).$$

If the test is successful it decides on the null hypothesis

$$\hat{H}_1 = 0.$$

Otherwise it decides on the alternative hypothesis. We now describe the communication to Detector 2. If

$$\hat{\mathcal{H}}_1 = 0,$$

Detector 1 sends

$$M_2 = 1.$$

Else if

$$\hat{\mathcal{H}}_1 = 1,$$

Detector 1 sends

$$M_2 = 0.$$

Type I errors analysis : We have for  $n$  large enough:

$$\begin{aligned} \alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \\ &= \Pr[X^n \notin \mathcal{T}_\mu(P_X) \text{ or } Y_1^n \notin \mathcal{T}_\mu(P_{Y_1}) | \mathcal{H} = 0] \\ &\stackrel{(a)}{\leq} \Pr[(X^n, Y_1^n) \notin \mathcal{T}_\mu(P_{XY_1}) | \mathcal{H} = 0] \\ &\stackrel{(b)}{\leq} \epsilon_1. \end{aligned} \tag{B.1}$$

where (a) follows since  $\mathcal{T}_\mu(P_{XY_1}) \subseteq \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1})$ ; (b) follows by the weak law of large number. In the same way for  $n$  large enough:

$$\begin{aligned} \alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0] \\ &= \Pr[X^n \notin \mathcal{T}_\mu(P_X) \text{ or } Y_1^n \notin \mathcal{T}_\mu(P_{Y_1}) \text{ or } Y_2^n \notin \mathcal{T}_\mu(P_{Y_2}) | \mathcal{H} = 0] \\ &\stackrel{(a)}{\leq} \Pr[(X^n, Y_1^n, Y_2^n) \notin \mathcal{T}_\mu(P_{XY_1Y_2}) | \mathcal{H} = 0] \\ &\stackrel{(b)}{\leq} \epsilon_2. \end{aligned} \tag{B.2}$$

where (a) follows since  $\mathcal{T}_\mu(P_{XY_1Y_2}) \subseteq \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1}) \times \mathcal{T}_\mu(P_{Y_2})$ ; (b) follows by the weak law of large number.

Type II errors analysis : define the sets  $\mathcal{L}_{1,\mu}$  and  $\mathcal{L}_{2,\mu}$  so that:

$$\mathcal{L}_{1,\mu} \triangleq \left\{ \tilde{P}_{XY_1} : \begin{array}{l} |P_X - \tilde{P}_X| \leq \mu P_X, \\ |P_{Y_1} - \tilde{P}_{Y_1}| \leq \mu P_{Y_1} \end{array} \right\}, \quad \mathcal{L}_{2,\mu} \triangleq \left\{ \tilde{P}_{XY_1Y_2} : \begin{array}{l} |P_X - \tilde{P}_X| \leq \mu P_X, \\ |P_{Y_1} - \tilde{P}_{Y_1}| \leq \mu P_{Y_1}, \\ |P_{Y_2} - \tilde{P}_{Y_2}| \leq \mu P_{Y_2} \end{array} \right\}.$$

The acceptance region of the test performed at Detector 1 is:

$$\mathcal{A}_{1,n} = \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1}) = \bigcup_{\tilde{P}_{XY_1} \in \mathcal{L}_{1,\mu} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}_1)} \mathcal{T}(\tilde{P}_{XY_1}). \tag{B.3}$$

We have:

$$\begin{aligned} \beta_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \\ &= \Pr[(X^n, Y_1^n) \in \mathcal{A}_{1,n} | \mathcal{H} = 1] \\ &\stackrel{(a)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{Y}_1|} \exp\left[-n \min_{\tilde{P}_{XY_1} \in \mathcal{L}_{1,\mu}} D(\tilde{P}_{XY_1} \| P_{XY_1})\right], \end{aligned} \tag{B.4}$$

where (a) follows by Theorem 2.4. In the same way, the acceptance region of the test performed at Detector 2 is:

$$\mathcal{A}_{2,n} = \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1}) \times \mathcal{T}_\mu(P_{Y_2}) = \bigcup_{\tilde{P}_{XY_1Y_2} \in \mathcal{L}_{2,\mu} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2)} \mathcal{T}(\tilde{P}_{XY_1Y_2}). \quad (\text{B.5})$$

and so:

$$\begin{aligned} \beta_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \\ &= \Pr[(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_{2,n} | \mathcal{H} = 1] \\ &\stackrel{(a)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|} \exp\left[-n \min_{\tilde{P}_{XY_1Y_2} \in \mathcal{L}_{2,\mu}} D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2})\right], \end{aligned} \quad (\text{B.6})$$

where (a) follows by Theorem 2.4. Finally as  $\frac{|\mathcal{X}||\mathcal{Y}_1|}{n} \log(n+1)$  and  $\frac{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}{n} \log(n+1)$  are sequences decreasing to 0 when  $n \rightarrow \infty$ , the achievability part is conclude by letting  $n \rightarrow \infty$ .

We now continue with the converse part. Fix an achievable Type II error exponents pair  $(\theta_1, \theta_2)$ , then choose a small number  $\mu > 0$ , a sufficiently large blocklength  $n$ , and encoding and decision functions  $\phi_{1,n}, \phi_{2,n}, \psi_{1,n}, \psi_{2,n}$  satisfying

$$\begin{aligned} \alpha_{1,n} &\leq \epsilon_1, \\ \alpha_{2,n} &\leq \epsilon_2, \end{aligned} \quad (\text{B.7})$$

and

$$-\frac{1}{n} \log \beta_{1,n} \geq \theta_1 - \mu, \quad (\text{B.8a})$$

$$-\frac{1}{n} \log \beta_{2,n} \geq \theta_2 - \mu, \quad (\text{B.8b})$$

For the chosen encoding and decision functions, define for each pair  $(m_1, m_2) \in \{0, \dots, W_1 - 1\} \times \{0, \dots, W_2 - 1\}$  the subsets

$$\mathcal{C}_{m_1} \triangleq \{x^n \in \mathcal{X}^n : \phi_{1,n}(x^n) = m_1\}, \quad (\text{B.9})$$

$$\mathcal{G}_{m_1, m_2} \triangleq \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = m_2\}, \quad (\text{B.10})$$

$$\mathcal{F}_{1, m_1} \triangleq \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = 0\}, \quad (\text{B.11})$$

$$\mathcal{F}_{2, m_1, m_2} \triangleq \{y_2^n \in \mathcal{Y}_2^n : \psi_{2,n}(m_1, m_2, y_2^n) = 0\}. \quad (\text{B.12})$$

Moreover, the acceptance regions  $\mathcal{A}_{1,n}$  at Detector 1 and  $\mathcal{A}_{2,n}$  at Detector 2, defined through the relations

$$(X^n, Y_1^n) \in \mathcal{A}_{1,n} \iff \hat{\mathcal{H}}_1 = 0, \quad (\text{B.13})$$

and

$$(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_{2,n} \iff \hat{\mathcal{H}}_2 = 0, \quad (\text{B.14})$$

can be expressed as

$$\mathcal{A}_{1,n} \triangleq \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{F}_{1,m_1,m_2}. \quad (\text{B.15})$$

$$\mathcal{A}_{2,n} \triangleq \bigcup_{m_2=0}^{W_2-1} \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1,m_2} \times \mathcal{F}_{2,m_1,m_2}. \quad (\text{B.16})$$

By the constraint on the Type I error probability on Detector 1 and Detector 2, (C.8a),

$$\begin{aligned} P_{XY_1}^{\otimes n} \left[ (X^n, Y_1^n) \in \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{F}_{1,m_1,m_2} \right] &\geq 1 - \epsilon_1 \\ P_{XY_1Y_2}^{\otimes n} \left[ (X^n, Y_1^n, Y_2^n) \in \bigcup_{m_1=0}^{W_1-1} \bigcup_{m_2=0}^{W_2-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1,m_2} \times \mathcal{F}_{2,m_1,m_2} \right] &\geq 1 - \epsilon_2 \end{aligned} \quad (\text{B.17})$$

Now, by the union bound there exists an index pair  $(m_1^*, m_2^*) \in \{0, \dots, W_1 - 1\} \times \{0, \dots, W_2 - 1\}$  such that:

$$P_X^{\otimes n} [X^n \in \mathcal{C}_{m_1^*}] \geq \frac{1 - \epsilon}{W_1}, \quad (\text{B.18a})$$

$$P_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{F}_{1,m_1^*}] \geq \frac{1 - \epsilon}{W_1} \quad (\text{B.18b})$$

$$P_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{G}_{m_1^*, m_2^*}] \geq \frac{1 - \epsilon}{W_1 W_2}, \quad (\text{B.18c})$$

$$P_{Y_2}^{\otimes n} [Y_2^n \in \mathcal{F}_{2,m_1^*, m_2^*}] \geq \frac{1 - \epsilon}{W_1 W_2}, \quad (\text{B.18d})$$

Combining (C.18) with Theorem 2.7 and an extension of it to three pmfs (recall that  $P_{XY_1Y_2}(x, y_1, y_2) > 0$  by assumption, and thus  $P_{XY_1Y_2} \ll \tilde{P}_{XY_1Y_2}$ ), for sufficiently large  $n$ , we obtain:

$$\begin{aligned} \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] &\geq \max_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = P_X, \tilde{P}_{Y_1} = P_{Y_1}}} e^{-n(D(\tilde{P}_{XY_1} \| \tilde{P}_{XY_1}) + \mu)}, \\ \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] &\geq \max_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = P_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} e^{-n(D(\tilde{P}_{XY_1Y_2} \| \tilde{P}_{XY_1Y_2}) + \mu)}. \end{aligned} \quad (\text{B.19})$$

Taking  $n \rightarrow \infty$  and  $\mu \rightarrow 0$ , by the continuity of KL-divergence, we can conclude that for any achievable exponent  $\theta_2$ :

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = P_X, \\ \tilde{P}_{Y_1} = P_{Y_1}}} D(\tilde{P}_{XY_1} \| \tilde{P}_{XY_1}) \quad (\text{B.20})$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X = P_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1Y_2} \| \tilde{P}_{XY_1Y_2}). \quad (\text{B.21})$$

This concludes the proof.

---

Proof of Proposition 4.2

---

We present first a coding scheme for coherent detection that we will analyze to prove the achievability part. Fix a small  $\mu > 0$ .

Sensor: Assume it observes the source sequence  $X^n = x^n$ , then it checks whether

$$x^n \in \mathcal{T}_\mu(P_X) = \mathcal{T}_\mu(\bar{P}_X)$$

if the test is successful it sends  $M_1 = 1$  else  $M_1 = 0$ .

Detector 1: If  $M_1 = 0$ , Detector 1 decides on the alternative hypothesis:

$$\hat{\mathcal{H}}_1 = 1.$$

If  $M_1 = 1$  and given that  $Y_1^n = y_1^n$ , Detector 1 checks whether

$$y_1^n \in \mathcal{T}_\mu(\bar{P}_{Y_1}).$$

If the test is successful it decides on hypothesis

$$\hat{\mathcal{H}}_1 = 1.$$

Otherwise it decides on

$$\hat{\mathcal{H}}_1 = 0.$$

We now describe the communication to Detector 2. If

$$\hat{\mathcal{H}}_1 = 0,$$

Detector 1 sends

$$M_2 = 1.$$

Else if

$$\hat{\mathcal{H}}_1 = 1,$$

Detector 1 sends

$$M_2 = 0.$$

Type I errors analysis : We have for  $n$  large enough:

$$\begin{aligned} \alpha_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \\ &= \Pr[X^n \notin \mathcal{T}_\mu(\bar{P}_X) \text{ or } Y_1^n \notin \mathcal{T}_\mu(\bar{P}_{Y_1}) | \mathcal{H} = 0] \\ &\stackrel{(a)}{\leq} \Pr[(X^n, Y_1^n) \notin \mathcal{T}_\mu(\bar{P}_{XY_1}) | \mathcal{H} = 0] \\ &\stackrel{(b)}{\leq} \epsilon_1. \end{aligned} \tag{C.1}$$

where (a) follows since  $\mathcal{T}_\mu(\bar{P}_{XY_1}) \subseteq \mathcal{T}_\mu(\bar{P}_X) \times \mathcal{T}_\mu(\bar{P}_{Y_1})$ ; (b) follows by the weak law of large number. In the same way for  $n$  large enough:

$$\begin{aligned} \alpha_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 1 | \mathcal{H} = 0] \\ &= \Pr[X^n \notin \mathcal{T}_\mu(P_X) \text{ or } Y_1^n \notin \mathcal{T}_\mu(P_{Y_1}) \text{ or } Y_2^n \notin \mathcal{T}_\mu(P_{Y_2}) | \mathcal{H} = 0] \\ &\stackrel{(a)}{\leq} \Pr[(X^n, Y_1^n, Y_2^n) \notin \mathcal{T}_\mu(P_{XY_1Y_2}) | \mathcal{H} = 0] \\ &\stackrel{(b)}{\leq} \epsilon_2. \end{aligned} \tag{C.2}$$

where (a) follows since  $\mathcal{T}_\mu(P_{XY_1Y_2}) \subseteq \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1}) \times \mathcal{T}_\mu(P_{Y_2})$ ; (b) follows by the weak law of large number.

Type II errors analysis : define the sets  $\mathcal{L}_{1,\mu}$  so that:

$$\bar{\mathcal{L}}_{1,\mu} \triangleq \left\{ \bar{P}_{XY_1} : \begin{array}{l} |\bar{P}_X - \tilde{P}_X| \leq \mu \bar{P}_X, \\ |\bar{P}_{Y_1} - \tilde{P}_{Y_1}| \leq \mu \bar{P}_{Y_1} \end{array} \right\}.$$

The acceptance region of the test performed at Detector 1 is:

$$\mathcal{A}_{1,n} = \mathcal{T}_\mu(\bar{P}_X) \times \mathcal{T}_\mu(\bar{P}_{Y_1}) = \bigcup_{\tilde{P}_{XY_1} \in \bar{\mathcal{L}}_{1,\mu} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}_1)} \mathcal{T}(\tilde{P}_{XY_1}). \tag{C.3}$$

We have:

$$\begin{aligned} \beta_{1,n} &= \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \\ &= \Pr[(X^n, Y_1^n) \in \mathcal{A}_{1,n} | \mathcal{H} = 1] \end{aligned}$$

$$\stackrel{(a)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{Y}_1|} \exp\left[-n \min_{\tilde{P}_{XY_1} \in \tilde{\mathcal{L}}_{1,\mu}} D(\tilde{P}_{XY_1} \| P_{XY_1})\right], \quad (\text{C.4})$$

where (a) follows by Theorem 2.4. In the same way, the acceptance region of the test performed at Detector 2 is:

$$\mathcal{A}_{2,n} = \mathcal{T}_\mu(P_X) \times \mathcal{T}_\mu(P_{Y_1}) \times \mathcal{T}_\mu(P_{Y_2}) = \bigcup_{\tilde{P}_{XY_1Y_2} \in \mathcal{L}_{2,\mu} \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2)} \mathcal{T}(\tilde{P}_{XY_1Y_2}). \quad (\text{C.5})$$

and so:

$$\begin{aligned} \beta_{2,n} &= \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \\ &= \Pr[(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_{2,n} | \mathcal{H} = 1] \\ &\stackrel{(a)}{\leq} (n+1)^{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|} \exp\left[-n \min_{\tilde{P}_{XY_1Y_2} \in \mathcal{L}_{2,\mu}} D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2})\right], \end{aligned} \quad (\text{C.6})$$

where (a) follows by Theorem 2.4. Finally as  $\frac{|\mathcal{X}||\mathcal{Y}_1|}{n} \log(n+1)$  and  $\frac{|\mathcal{X}||\mathcal{Y}_1||\mathcal{Y}_2|}{n} \log(n+1)$  are sequences decreasing to 0 when  $n \rightarrow \infty$ , the achievability part is conclude by letting  $n \rightarrow \infty$ .

We now continue with the converse part. Fix an achievable Type II error exponents pair  $(\theta_1, \theta_2)$ , then choose a small number  $\mu > 0$ , a sufficiently large blocklength  $n$ , and encoding and decision functions  $\phi_{1,n}, \phi_{2,n}, \psi_{1,n}, \psi_{2,n}$  satisfying

$$\begin{aligned} \alpha_{1,n} &\leq \epsilon_1, \\ \alpha_{2,n} &\leq \epsilon_2, \end{aligned} \quad (\text{C.7})$$

and

$$-\frac{1}{n} \log \beta_{1,n} \geq \theta_1 - \mu \quad (\text{C.8a})$$

$$-\frac{1}{n} \log \beta_{2,n} \geq \theta_2 - \mu, \quad (\text{C.8b})$$

For the chosen encoding and decision functions, define for each pair  $(m_1, m_2) \in \{0, \dots, W_1 - 1\} \times \{0, \dots, W_2 - 1\}$  the subsets

$$\mathcal{C}_{m_1} \triangleq \{x^n \in \mathcal{X}^n : \phi_{1,n}(x^n) = m_1\}, \quad (\text{C.9})$$

$$\mathcal{G}_{m_1, m_2} \triangleq \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = m_2\}, \quad (\text{C.10})$$

$$\mathcal{F}_{1, m_1} \triangleq \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = 0\}, \quad (\text{C.11})$$

$$\mathcal{F}_{2, m_1, m_2} \triangleq \{y_2^n \in \mathcal{Y}_2^n : \psi_{2,n}(m_1, m_2, y_2^n) = 0\}. \quad (\text{C.12})$$

Moreover, the acceptance regions  $\mathcal{A}_{1,n}$  at Detector 1 and  $\mathcal{A}_{2,n}$  at Detector 2, defined through the relations

$$(X^n, Y_1^n) \in \mathcal{A}_{1,n} \iff \hat{\mathcal{H}}_1 = 0, \quad (\text{C.13})$$

and

$$(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_{2,n} \iff \hat{\mathcal{H}}_2 = 0, \quad (\text{C.14})$$



can be expressed as

$$\mathcal{A}_{1,n} \triangleq \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{F}_{1,m_1,m_2}. \quad (\text{C.15})$$

$$\mathcal{A}_{2,n} \triangleq \bigcup_{m_2=0}^{W_2-1} \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1,m_2} \times \mathcal{F}_{2,m_1,m_2}. \quad (\text{C.16})$$

By the constraint on the Type I error probability on Detector 1 and Detector 2, (C.8a),

$$\begin{aligned} \bar{P}_{XY_1}^{\otimes n} \left[ (X^n, Y_1^n) \in \bigcup_{m_1=0}^{W_1-1} \mathcal{C}_{m_1} \times \mathcal{F}_{2,m_1,m_2} \right] &\geq 1 - \epsilon_1 \\ P_{XY_1Y_2}^{\otimes n} \left[ (X^n, Y_1^n, Y_2^n) \in \bigcup_{m_1=0}^{W_1-1} \bigcup_{m_2=0}^{W_2-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1,m_2} \times \mathcal{F}_{m_1,m_2} \right] &\geq 1 - \epsilon_2 \end{aligned} \quad (\text{C.17})$$

Now, by the union bound there exists an index pair  $(m_1^*, m_2^*) \in \{0, \dots, W_1 - 1\} \times \{0, \dots, W_2 - 1\}$  such that:

$$\bar{P}_X^{\otimes n} [X^n \in \mathcal{C}_{m_1^*}] \geq \frac{1 - \epsilon}{W_1}, \quad \bar{P}_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{F}_{1,m_1^*}] \geq \frac{1 - \epsilon}{W_1} \quad (\text{C.18a})$$

$$P_X^{\otimes n} [X^n \in \mathcal{C}_{m_1^*}] \geq \frac{1 - \epsilon}{W_1}, \quad P_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{G}_{m_1^*, m_2^*}] \geq \frac{1 - \epsilon}{W_1 W_2}, \quad P_{Y_2}^{\otimes n} [Y_2^n \in \mathcal{F}_{2,m_1^*, m_2^*}] \geq \frac{1 - \epsilon}{W_1 W_2}. \quad (\text{C.18b})$$

Combining (C.18) with Theorem 2.7 and an extension of it to three pmfs (recall that we assumed  $P_{XY_1Y_2}(x, y_1, y_2) > 0$  and thus  $P_{XY_1Y_2} \ll \bar{P}_{XY_1Y_2}$ ), for sufficiently large  $n$ , we obtain:

$$\begin{aligned} \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] &\geq \max_{\substack{\bar{P}_{XY_1}: \\ \bar{P}_X = \bar{P}_X, \bar{P}_{Y_1} = \bar{P}_{Y_1}}} e^{-n(D(\bar{P}_{XY_1} \| P_{XY_1}) + \mu)}, \\ \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] &\geq \max_{\substack{\bar{P}_{XY_1Y_2}: \\ \bar{P}_X = P_X, \\ \bar{P}_{Y_1} = P_{Y_1}, \bar{P}_{Y_2} = P_{Y_2}}} e^{-n(D(\bar{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}) + \mu)}. \end{aligned} \quad (\text{C.19})$$

Taking  $n \rightarrow \infty$  and  $\mu \rightarrow 0$ , by the continuity of KL-divergence, we can conclude that for any achievable exponent  $\theta_2$ :

$$\theta_1 \leq \min_{\substack{\bar{P}_{XY_1}: \\ \bar{P}_X = \bar{P}_X, \\ \bar{P}_{Y_1} = \bar{P}_{Y_1}}} D(\bar{P}_{XY_1} \| P_{XY_1}) \quad (\text{C.20})$$

$$\theta_2 \leq \min_{\substack{\bar{P}_{XY_1Y_2}: \\ \bar{P}_X = P_X, \\ \bar{P}_{Y_1} = P_{Y_1}, \bar{P}_{Y_2} = P_{Y_2}}} D(\bar{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}). \quad (\text{C.21})$$

This conclude the proof.

---

Proof of Theorem 4.4

---

Error analysis of the scheme presented in subsection 4.3.2:

Notice first that for sufficiently small  $\mu$ , when  $X^n \in \mathcal{T}_\mu^n(\bar{P}_X)$  and  $Y_1^n \in \mathcal{T}_\mu^n(\bar{P}_{Y_1})$ , then Detector 1 declares  $\hat{\mathcal{H}}_1 = 1$ . Thus, by the weak law of large numbers, for sufficiently large  $n$ :

$$\alpha_{1,n} \leq 1 - \Pr \left[ (X^n, Y_1^n) \in \mathcal{T}_\mu^n(\bar{P}_X) \times \mathcal{T}_\mu^n(\bar{P}_{Y_1}) \right] \leq \epsilon_1.$$

In the same way, when  $X^n \in \mathcal{T}_\mu^n(P_X)$ ,  $Y_1^n \in \mathcal{T}_\mu^n(P_{Y_1})$  and  $Y_2^n \in \mathcal{T}_\mu^n(P_{Y_2})$ , then Detector 2 declares  $\hat{\mathcal{H}}_2 = 0$ . Thus, by the weak law of large numbers, for sufficiently large  $n$ :

$$\alpha_{2,n} \leq 1 - \Pr \left[ (X^n, Y_1^n, Y_2^n) \in \mathcal{T}_\mu^n(P_X) \times \mathcal{T}_\mu^n(P_{Y_1}) \times \mathcal{T}_\mu^n(P_{Y_2}) \right] \leq \epsilon_2.$$

Define now for  $r \in \mathbb{R}$ :

$$\mathcal{A}_{1,\mu}(\mathbf{r}) := \left\{ (x^n, y_1^n) : P_{x^n} \in \Gamma_{b(1)}(\mathbf{r}), \text{ and } y_1^n \in \mathcal{T}_\mu^n(\bar{P}_{Y_1}) \right\}. \quad (\text{D.1})$$

$$\mathcal{A}_{2,\mu}(\mathbf{r}) := \left\{ (x^n, y_1^n, y_2^n) : P_{x^n} \in \Gamma_{b(0)}(\mathbf{r}), y_1^n \in \mathcal{T}_\mu^n(P_{Y_1}), \text{ and } y_2^n \in \mathcal{T}_\mu^n(P_{Y_2}) \right\}. \quad (\text{D.2})$$

The type-II error probability at detector 1 satisfies:

$$\begin{aligned} \beta_{1,n} &\leq \Pr \left[ (X^n, Y_1^n) \in \mathcal{A}_{1,\mu} \mid \mathcal{H} = 0 \right] \\ &\leq \min_{\substack{\bar{P}_{XY_1} : \\ \bar{P}_X \in \Gamma_{b(1)}(\mathbf{r}) \\ |\bar{P}_{Y_1} - \bar{P}_{Y_1}| \leq \mu}} e^{-n \left( D(\bar{P}_{XY_1} \| P_{XY_1}) - \xi_1(\mu) \right)}, \end{aligned} \quad (\text{D.3})$$

where the last inequality holds for sufficiently large values of  $n$  and by Sanov's theorem. Here  $\xi_1(\mu) \rightarrow 0$  with  $\mu$ . In the same way, the type-II error probability at detector 2 satisfies:

$$\beta_{2,n} \leq \Pr \left[ (X^n, Y_1^n, Y_2^n) \in \mathcal{A}_{2,\mu} \mid \mathcal{H} = 1 \right]$$

$$\leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_{b(0)}(\mathbf{r}) \\ |\tilde{P}_{Y_1 - P_{Y_1}}| \leq \mu, |\tilde{P}_{Y_2 - P_{Y_2}}| \leq \mu}} e^{-n(D(\tilde{P}_{XY_1Y_2} \parallel \bar{P}_{XY_1Y_2}) - \xi_2(\mu))}, \quad (\text{D.4})$$

where the last inequality holds for sufficiently large values of  $n$  and by Sanov's theorem. Here  $\xi_2(\mu) \rightarrow 0$  with  $\mu$ . Taking  $\mu \rightarrow 0$  and  $n \rightarrow \infty$  establishes the desired achievability result.

Converse to Theorem 4.4:

Fix a real number  $r$  and an exponent pair  $(\theta_1, \theta_2) \in \mathcal{E}_0(2, 2)$  satisfying

$$\theta_2 = \theta_1 + r. \quad (\text{D.5})$$

Then fix a small number  $\epsilon > 0$ , a sufficiently large blocklength  $n$ , and encoding and decision functions  $\phi_{1,n}, \phi_{2,n}, \psi_{1,n}, \psi_{2,n}$  satisfying

$$\alpha_{1,n} \leq \epsilon, \quad (\text{D.6})$$

$$\alpha_{2,n} \leq \epsilon, \quad (\text{D.7})$$

and

$$-\frac{1}{n} \log \beta_{1,n} \geq \theta_1 - \epsilon, \quad (\text{D.8})$$

$$-\frac{1}{n} \log \beta_{2,n} \geq \theta_2 - \epsilon. \quad (\text{D.9})$$

For the chosen encoding and decision functions, define for each

$$m_1 \in \{0, 1, \dots, W_1 - 1\},$$

and

$$m_2 \in \{0, 1, \dots, W_2 - 1\},$$

the subsets

$$\mathcal{C}_{m_1} \triangleq \{x^n \in \mathcal{X}^n : \phi_{1,n}(x^n) = m_1\}, \quad (\text{D.10})$$

$$\mathcal{F}_{m_1}^1 \triangleq \{y_1^n \in \mathcal{Y}_1^n : \psi_{1,n}(m_1, y_1^n) = 1\}, \quad (\text{D.11})$$

$$\mathcal{G}_{m_1, m_2} \triangleq \{y_1^n \in \mathcal{Y}_1^n : \phi_{2,n}(m_1, y_1^n) = m_2\}, \quad (\text{D.12})$$

$$\mathcal{F}_{m_1, m_2}^2 \triangleq \{y_2^n \in \mathcal{Y}_2^n : \psi_{2,n}(m_1, m_2, y_2^n) = 0\} \quad (\text{D.13})$$

Notice that the sets  $\mathcal{C}_0, \dots, \mathcal{C}_{W_1-1}$  partition  $\mathcal{X}^n$  and for each  $m_1 \in \{0, \dots, W_1 - 1\}$  the sets

$$\mathcal{G}_{m_1, 0}, \dots, \mathcal{G}_{m_1, W_2-1},$$

partition  $\mathcal{Y}_1^n$ . Moreover, the acceptance regions  $\mathcal{A}_n^1$  and  $\mathcal{A}_n^2$  at detectors 1 and 2, defined through the relations

$$(X^n, Y_1^n) \in \mathcal{A}_n^1 \iff \hat{\mathcal{H}}_1 = 1, \quad (\text{D.14})$$

$$(X^n, Y_1^n, Y_2^n) \in \mathcal{A}_n^2 \iff \hat{\mathcal{H}}_2 = 0, \quad (\text{D.15})$$

can be expressed as

$$\mathcal{A}_{1,n} = \mathcal{C}_0 \times \mathcal{F}_0^1 \cup \mathcal{C}_1 \times \mathcal{F}_1^1 \quad (\text{D.16a})$$

and

$$\mathcal{A}_{2,n} = \bigcup_{m_2=0}^{W_2-1} \mathcal{C}_0 \times \mathcal{G}_{0,m_2} \times \mathcal{F}_{0,m_2}^2 \cup \bigcup_{m_2=0}^{W_2-1} \mathcal{C}_1 \times \mathcal{G}_{1,m_2} \times \mathcal{F}_{1,m_2}^2. \quad (\text{D.16b})$$

Define now for each  $m_1 \in \{0, 1\}$  the set

$$\Gamma_{m_1,n} := \left\{ \tilde{P}_X \in \mathcal{P}(\mathcal{X}) : \tilde{P}_X^{\otimes n} [X^n \in \mathcal{C}_{m_1}] \geq \frac{1-\epsilon}{2} \right\}, \quad (\text{D.17})$$

and for each pair  $(m_1, m_2) \in \{0, 1\} \times \{0, \dots, W_2 - 1\}$  the set

$$\Delta_{m_1, m_2, n} := \left\{ \tilde{P}_{Y_1} \in \mathcal{P}(\mathcal{Y}_1) : \tilde{P}_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{G}_{m_1, m_2}] \geq \frac{1-\epsilon}{2W_2} \right\}. \quad (\text{D.18})$$

Since the sets  $\mathcal{C}_0, \mathcal{C}_1$  cover  $\mathcal{X}^n$  and since for each  $\tilde{P}_X \in \mathcal{P}(\mathcal{X})$ , it holds that  $\tilde{P}_X^{\otimes n} [X^n \in \mathcal{X}^n] = 1$ , the subsets  $\Gamma_{0,n}, \Gamma_{1,n}$  cover the set  $\mathcal{P}(\mathcal{X})$ . Similarly, since for each  $m_1 \in \{0, 1\}$  the sets  $\mathcal{G}_{m_1, 0}, \dots, \mathcal{G}_{m_1, W_2-1}$  cover  $\mathcal{Y}_1^n$ , the subsets  $\Delta_{m_1, 0}^n, \dots, \Delta_{m_1, W_2-1, n}^n$  cover the set  $\mathcal{P}(\mathcal{Y}_1)$ . Moreover, by the constraint on the type-I error probability at detectors 1 and 2, (D.6) and (D.7):

$$\tilde{P}_{XY_1}^{\otimes n} \left[ (X^n, Y_1^n) \in \bigcup_{m_1=0}^1 \mathcal{C}_{m_1} \times \mathcal{F}_{m_1}^1 \right] \geq 1 - \epsilon \quad (\text{D.19})$$

$$P_{XY_1 Y_2}^{\otimes n} \left[ (X^n, Y_1^n, Y_2^n) \in \bigcup_{m_1=0}^1 \bigcup_{m_2=0}^{W_2-1} \mathcal{C}_{m_1} \times \mathcal{G}_{m_1, m_2} \times \mathcal{F}_{m_1, m_2}^2 \right] \geq 1 - \epsilon. \quad (\text{D.20})$$

By the union bound there exist thus an index  $\tilde{m}_1 \in \{0, 1\}$  and an index pair  $(m_1^*, m_2^*) \in \{0, 1\} \times \{0, \dots, W_2 - 1\}$  such that:

$$\tilde{P}_X^{\otimes n} [X^n \in \mathcal{C}_{\tilde{m}_1}] \geq \frac{1-\epsilon}{2}, \quad (\text{D.21a})$$

$$\tilde{P}_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{F}_{\tilde{m}_1}^1] \geq \frac{1-\epsilon}{2}, \quad (\text{D.21b})$$

and

$$P_X^{\otimes n} [X^n \in \mathcal{C}_{m_1^*}] \geq \frac{1-\epsilon}{2}, \quad (\text{D.22a})$$

$$P_{Y_1}^{\otimes n} [Y_1^n \in \mathcal{G}_{m_1^*, m_2^*}] \geq \frac{1-\epsilon}{2W_2}, \quad (\text{D.22b})$$

$$P_{Y_2}^{\otimes n} [Y_2^n \in \mathcal{F}_{m_1^*, m_2^*}^2] \geq \frac{1-\epsilon}{2W_2}, \quad (\text{D.22c})$$

Combining (D.21) with the definition of  $\Delta_{\tilde{m}_1, n}$  in (D.17) and with [31, Theorem 3] (recall that by assumption  $P_{XY_1}(x, y_1) > 0$ , for all  $(x, y_1) \in \mathcal{X} \times \mathcal{Y}_1$ ) yields that for any  $\mu > 0$  and sufficiently large  $n$ :

$$\Pr[\hat{\mathcal{H}}_1 = 1 | \mathcal{H} = 0] \geq \max_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_{\tilde{m}_1, n}, \\ \tilde{P}_{Y_1} = \tilde{P}_{Y_1}}} e^{-n(D(\tilde{P}_{XY_1} \| P_{XY_1}) + \mu)}.$$

In the same way, combining (D.22) (D.17) with (D.18) and extending [31, Theorem 3] to three pmfs (recall that by assumption  $P_{XY_1Y_2}(x, y_1, y_2) > 0$ , for all  $(x, y_1, y_2) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$ ), for sufficiently large  $n$ :

$$\Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1] \geq \max_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_{m_1^*, n}, \\ \tilde{P}_{Y_1} \in \Delta_{m_1^*, m_2^*, n}, \tilde{P}_{Y_2} = P_{Y_2}}} e^{-n(D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}) + \mu)}.$$

Taking now  $n \rightarrow \infty$  and  $\mu \rightarrow 0$ , by the continuity of the Kullback-Leiber divergence we can conclude that if the exponent pair  $(\theta_1, \theta_2)$  is achievable, then there exist subsets  $\Gamma_0, \Gamma_1$  that cover  $\mathcal{P}(\mathcal{X})$ , subsets  $\Delta_{0,0}, \dots, \Delta_{0,W_2-1}$  that cover  $\mathcal{P}(\mathcal{Y}_1)$ , and subsets  $\Delta_{1,0}, \dots, \Delta_{1,W_2-1}$  that cover  $\mathcal{P}(\mathcal{Y}_2)$  so that:

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_b, \\ \tilde{P}_{Y_1} = P_{Y_1}}} D(\tilde{P}_{XY_1} \| P_{XY_1}), \quad (\text{D.23a})$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} \in \Delta_{c,c_2}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}). \quad (\text{D.23b})$$

where the indices  $b, c \in \{0, 1\}$  and  $c_2 \in \{0, \dots, W_2 - 1\}$  are such that

$$\tilde{P}_X \in \Gamma_b, \quad (\text{D.24})$$

$$P_X \in \Gamma_c, \quad (\text{D.25})$$

$$P_{Y_1} \in \Delta_{c,c_2}. \quad (\text{D.26})$$

We continue to notice that the upper bounds in (D.23) become looser when elements are removed from the sets  $\Gamma_b, \Gamma_c$ , and  $\Delta_{c,c_2}$ . The converse statement thus remains valid by imposing

$$\Delta_{c,c_2} = \{P_{Y_1}\}. \quad (\text{D.27})$$

Moreover, if  $b = c$ , then we impose

$$\Gamma_b = \Gamma_c = \{P_X, \bar{P}_X\}, \quad (\text{D.28})$$

and if  $b \neq c$ , then we impose that  $\Gamma_b$  and  $\Gamma_c$  form a partition.

If  $b = c$ , this concludes the proof. Otherwise, if  $b \neq c$ , we obtain the intermediate result that

$$\theta_1 \leq \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_b, \\ \tilde{P}_{Y_1} = P_{Y_1}}} D(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}) \quad (\text{D.29a})$$

$$\theta_2 \leq \min_{\substack{\tilde{P}_{XY_1Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1Y_2} \| \bar{P}_{XY_1Y_2}) \quad (\text{D.29b})$$

for two sets  $\Gamma_b$  and  $\Gamma_c$  forming a partition of  $\mathcal{P}(\mathcal{X})$  and satisfying (D.24) and (D.25).

We now characterize the choice of the sets  $\{\Gamma_b, \Gamma_c\}$  that yields the loosest bound in (D.29). To this end, notice first that by assumption (D.5), constraints (D.29) are equivalent to:

$$\theta_1 \leq \min \left\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_b, \\ \tilde{P}_{Y_1} = \tilde{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}), \quad \min_{\substack{\tilde{P}_{XY_1 Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1 Y_2} \| \bar{P}_{XY_1 Y_2}) - r \right\}. \quad (\text{D.30})$$

We notice that the right-hand side of (D.30) is upper bounded as:

$$\min \left\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X \in \Gamma_b, \\ \tilde{P}_{Y_1} = \tilde{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}), \quad \min_{\substack{\tilde{P}_{XY_1 Y_2}: \\ \tilde{P}_X \in \Gamma_c, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1 Y_2} \| \bar{P}_{XY_1 Y_2}) - r \right\} \quad (\text{D.31})$$

$$\leq \min_{\pi_X \in \mathcal{P}(\mathcal{X}) \setminus \{P_X, \bar{P}_X\}} \max \left\{ \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = \tilde{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}); \quad \min_{\substack{\tilde{P}_{XY_1 Y_2}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1 Y_2} \| P_{XY_1 Y_2}) - r \right\}, \quad (\text{D.32})$$

and that the bound holds with equality when

$$(\pi_X \in \Gamma_b) \iff \left( \min_{\substack{\tilde{P}_{XY_1}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = \tilde{P}_{Y_1}}} D(\tilde{P}_{XY_1} \| \bar{P}_{XY_1}) \geq \min_{\substack{\tilde{P}_{XY_1 Y_2}: \\ \tilde{P}_X = \pi_X, \\ \tilde{P}_{Y_1} = P_{Y_1}, \tilde{P}_{Y_2} = P_{Y_2}}} D(\tilde{P}_{XY_1 Y_2} \| P_{XY_1 Y_2}) - r \right). \quad (\text{D.33})$$

This concludes the proof also for the case  $b \neq c$ .



---

Proof of converse to Theorem 5.3

---

Let  $R_2 = 0$ . Fix a rate  $R_1 \geq 0$  and a pair of exponents  $(\theta_1, \theta_2) \in \mathcal{E}_0(R_1, 0)$ . Then, choose an  $\epsilon \in (0, 1/2)$ , a sufficiently large blocklength  $n$ , encoding and decision functions  $\phi_{1,n}$ ,  $\phi_{2,n}$ ,  $\psi_{1,n}$ , and  $\psi_{2,n}$  that satisfy

$$\alpha_{1,n} \leq \epsilon, \tag{E.1}$$

$$\alpha_{2,n} \leq \epsilon, \tag{E.2}$$

and

$$-\frac{1}{n} \log \beta_{1,n} \geq \theta_1 - \epsilon, \tag{E.3}$$

$$-\frac{1}{n} \log \beta_{2,n} \geq \theta_2 - \epsilon. \tag{E.4}$$

Notice first that for each  $i \in \{1, 2\}$ :

$$\begin{aligned} D(P_{\hat{\mathcal{H}}_i|H} \| P_{\hat{\mathcal{H}}_i|\mathcal{H}=1}) &= -h_2(\alpha_{i,n}) - (1 - \alpha_{i,n}) \log(\beta_{i,n}) \\ &\quad - \alpha_{i,n} \log(1 - \beta_{i,n}) \end{aligned} \tag{E.5}$$

where  $h_2(p)$  denotes the entropy of a Bernoulli- $(p)$  memoryless source. Since  $\alpha_{i,n} \leq \epsilon < 1/2$ , for each  $i \in \{1, 2\}$ , Inequality (E.5) yields:

$$-\frac{1}{n} \log(\beta_{i,n}) \leq \frac{1}{n(1-\epsilon)} D(P_{\hat{\mathcal{H}}_i|\mathcal{H}=0} \| P_{\hat{\mathcal{H}}_i|\mathcal{H}=1}) + \mu_n$$

with  $\mu_n \triangleq \frac{1}{n(1-\epsilon)} h_2(\epsilon)$ . Notice that  $\mu_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Consider now:

$$\theta_1 - \epsilon \leq -\frac{1}{n} \log(\beta_{1,n})$$



$$\begin{aligned}
&\leq \frac{1}{n(1-\epsilon)} D(P_{\hat{\mathcal{H}}_1|\mathcal{H}=0} \| P_{\hat{\mathcal{H}}_1|\mathcal{H}=1}) + \mu_n \\
&\stackrel{(a)}{\leq} \frac{1}{n(1-\epsilon)} D(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1}) + \mu_n \\
&\stackrel{(b)}{=} \frac{1}{n(1-\epsilon)} I(Y_1^n; M_1) + \mu_n \\
&\stackrel{(c)}{=} \frac{1}{n(1-\epsilon)} \sum_{k=1}^n H(Y_{1k}) - H(Y_{1k}|M_1 Y_1^{k-1}) + \mu_n \\
&\stackrel{(d)}{\leq} \frac{1}{n(1-\epsilon)} \sum_{k=1}^n H(Y_{1k}) - H(Y_{1k}|M_1 Y_1^{k-1} X^{k-1}) + \mu_n \\
&\stackrel{(e)}{=} \frac{1}{n(1-\epsilon)} \sum_{k=1}^n H(Y_{1k}) - H(Y_{1k}|M_1 X^{k-1}) + \mu_n \\
&\stackrel{(f)}{=} \frac{1}{n(1-\epsilon)} \sum_{k=1}^n I(Y_{1k}; U_k) + \mu_n \\
&\stackrel{(g)}{=} \frac{1}{n(1-\epsilon)} I(Y_{1Q}; U_Q|Q) + \mu_n \\
&\stackrel{(h)}{=} \frac{1}{1-\epsilon} I(Y_1(n); U(n)) + \mu_n
\end{aligned}$$

where: (a) follows by the data processing inequality for relative entropy; (b) holds since  $M_1$  and  $Y_1^n$  are independent under the alternative hypothesis  $\mathcal{H} = 1$ . (c) is due to the chain rule for mutual information; (d) follows since conditioning reduces entropy; (e) is due to the Markov chain  $Y_1^{k-1} - \ominus (M_1, X^{k-1}) - \ominus Y_{1k}$ ; (f) holds by defining  $U_k \triangleq (M_1, X^{k-1})$ ; (g) is obtained by introducing a random variable  $Q$  that is uniform over the set  $\{1, \dots, n\}$  and independent of all previously defined random variables; and (h) holds by defining  $U(n) \triangleq (U_Q, Q)$  and  $Y_1(n) \triangleq Y_{1Q}$ .

In a similar way, one obtains:

$$\begin{aligned}
\theta_2 - \epsilon &\leq -\frac{1}{n} \log(\beta_{2,n}) \\
&\stackrel{(i)}{\leq} \frac{1}{n(1-\epsilon)} D(P_{Y_2^n M_1 M_2|\mathcal{H}=0} \| P_{Y_2^n M_1 M_2|\mathcal{H}=1}) + \mu_n \\
&\stackrel{(j)}{=} \frac{1}{n(1-\epsilon)} \left( I(Y_2^n; M_1, M_2) + D(P_{M_1 M_2|\mathcal{H}=0} \| P_{M_1 M_2|\mathcal{H}=1}) \right) + \mu_n \\
&\stackrel{(k)}{\leq} \frac{1}{n(1-\epsilon)} \left( I(Y_2^n; M_1) + I(Y_2^n; M_2|M_1) + D(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1}) \right) + \mu_n \\
&\stackrel{(l)}{\leq} \frac{1}{n(1-\epsilon)} \left( I(Y_2^n; M_1) + \log W_{2,n} + D(P_{Y_1^n M_1|\mathcal{H}=0} \| P_{Y_1^n M_1|\mathcal{H}=1}) \right) + \mu_n \\
&\stackrel{(m)}{=} \frac{1}{n(1-\epsilon)} \left( I(Y_2^n; M_1) + I(Y_1^n; M_1) \right) + \tilde{\mu}_n \\
&\stackrel{(o)}{\leq} \frac{1}{1-\epsilon} \left( I(Y_2(n); U(n)) + I(Y_1(n); U(n)) \right) + \tilde{\mu}_n,
\end{aligned}$$

where (i) follows by the data processing inequality for relative entropy; (j) holds by the independence of the pair  $(M, M_2)$  with  $Y_2^n$  under the alternative hypothesis  $\mathcal{H} = 1$ ; (k) by the data processing inequality

for relative entropy;  $(\ell)$  holds since conditioning reduces entropy;  $(o)$  follows by proceeding along the steps  $(b)$  to  $(h)$  above; and  $(m)$  holds by defining  $\tilde{\mu}_n \triangleq W_{2,n}/(n(1-\epsilon)) + \mu_n$ .

Notice that by the assumption  $R_2 = 0$ , the term  $1/n \log |W_{2,n}| \rightarrow 0$  as  $n \rightarrow \infty$ . Thus, also  $\tilde{\mu}_n \rightarrow 0$  as  $n \rightarrow \infty$ .

We next lower bound the rate  $R_1$ :

$$\begin{aligned}
nR_1 &\geq H(M_1) \\
&= H(M_1) - H(M_1|X^n) \\
&= I(M_1; X^n) \\
&= \sum_{k=1}^n I(M_1; X_k|X^{k-1}) \\
&= \sum_{k=1}^n I(X_k; U_k) \\
&= nI(X_Q; U_Q|Q) \\
&= nI(U(n); X(n))
\end{aligned}$$

For any blocklength  $n$ , the newly defined random variables  $X(n), Y_1(n), Y_2(n) \sim P_{X Y_1 Y_2}$  and  $U(n) \text{---} X(n) \text{---} (Y_1(n), Y_2(n))$ . Letting now the blocklength  $n \rightarrow \infty$ , and then  $\epsilon \rightarrow 0$ , by continuity of mutual information establishes the desired converse result.



---

Proof of Theorem 5.5

---

We analyze the scheme in Subsection 5.2.1. If  $M_1 \neq (0, 0)$  and  $M_2 \neq (0, 0)$ , let  $I, J$  be the random indices sent over the bit pipes and define the following events:

$$\begin{aligned}\mathcal{E}_{D_1} &= \{(U^n(I), Y_1^n) \notin \mathcal{T}_{\mu/2}(P_{UY_1})\}, \\ \mathcal{E}_{D_2} &= \{(U^n(I), V^n(J|I), Y_2^n) \notin \mathcal{T}_{\mu/2}(P_{UVY_2})\}.\end{aligned}\tag{F.1}$$

The Type I error probability at Detector 1 over the random code construction can be bounded for  $n$  large enough as follows:

$$\begin{aligned}\mathbb{E}_C[\alpha_{1,n}] &\leq \Pr[M_1 = (0, 0) \text{ or } \mathcal{E}_{D_1} | \mathcal{H} = 0] \\ &\leq \Pr[M_1 = (0, 0) | \mathcal{H} = 0] + \Pr[\mathcal{E}_{D_1} | \mathcal{H} = 0, M_1 \neq (0, 0)] \\ &\stackrel{(a)}{\leq} \frac{\epsilon_1}{2} + \Pr[\mathcal{E}_{D_1} | \mathcal{H} = 0, M_1 \neq (0, 0)] \\ &\stackrel{(b)}{\leq} \frac{\epsilon_1}{2} + \frac{\epsilon_1}{2} \\ &= \epsilon_1,\end{aligned}\tag{F.2}$$

where (a) holds by the covering lemma [8] and (b) holds by the Markov lemma [8].

In the same way the Type I error probability at Detector 2 over the random code construction can be bounded for  $n$  large enough as follows:

$$\begin{aligned}\mathbb{E}_C[\alpha_{2,n}] &\leq \Pr[M_1 = (0, 0) \text{ or } M_2 = (0, 0) \text{ or } \mathcal{E}_{D_1} \text{ or } \mathcal{E}_{D_2} | \mathcal{H} = 0] \\ &\leq \Pr[M_1 = (0, 0) | \mathcal{H} = 0] + \Pr[M_2 = (0, 0) \text{ or } \mathcal{E}_{D_1} | \mathcal{H} = 0, M_1 \neq (0, 0)] \\ &\quad + \Pr[\mathcal{E}_{D_2} | \mathcal{H} = 0, M_1 \neq (0, 0), M_2 \neq (0, 0)]\end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \frac{\epsilon_2}{4} + \Pr[M_2 = (0, 0) \text{ or } E_{D_1} | \mathcal{H} = 0, M_1 = (0, 0)] + \frac{\epsilon_2}{4} \\
&\stackrel{(b)}{\leq} \frac{\epsilon_2}{4} + \Pr[\mathcal{E}_{D_1} | \mathcal{H} = 0, M_1 = (0, 0)] + \Pr[M_2 = (0, 0) | \mathcal{H} = 0, M_1 = (0, 0)] + \frac{\epsilon_2}{4} \\
&\stackrel{(c)}{\leq} \frac{\epsilon_2}{4} + \frac{\epsilon_2}{4} + \frac{\epsilon_2}{4} + \frac{\epsilon_2}{4} \\
&= \epsilon_2,
\end{aligned} \tag{F.3}$$

where (a) holds by the covering lemma, (b) holds by the Markov lemma, (c) holds by the Markov lemma and the covering lemma.

We now analyze the probability of Type II error at Detector 1. We notice that  $\hat{\mathcal{H}}_1 = 0$  only if there exists an index  $m_1 \in \{1, \dots, 2^{nR_1}\}$  such that

$$(U^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(P_{UX}) \quad \text{and} \quad (U^n(m_1), Y_1^n) \in \mathcal{T}_{\mu/4}^n(P_{UY_1}) \tag{F.4}$$

Therefore, using the union bound:

$$\begin{aligned}
\mathbb{E}_C[\beta_{1,n}] &= \Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \\
&\leq \sum_{m_1=1}^{2^{nR_1}} \Pr\left[(U^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(P_{UX}), (U^n(m_1), Y_1^n) \in \mathcal{T}_{\mu/4}^n(P_{UY_1}) \mid \mathcal{H} = 1\right]
\end{aligned} \tag{F.5}$$

$$\stackrel{(a)}{\leq} 2^{nR_1} \cdot \max_{\substack{\tilde{P}: \\ |\tilde{P}_{UX} - P_{UX}| < \mu/8 \\ |\tilde{P}_{UY_1} - P_{UY_1}| < \mu/4}} 2^{-n(D(\tilde{P}_{UXY_1} \| P_U \tilde{P}_{XY_1}) - \xi_n)} \tag{F.6}$$

$$\stackrel{(b)}{<} \max_{\substack{\tilde{P}: \\ |\tilde{P}_{UX} - P_{UX}| < \mu/8 \\ |\tilde{P}_{UY_1} - P_{UY_1}| < \mu/4}} 2^{n(I(U; X) - D(\tilde{P}_{UXY_1} \| P_U \tilde{P}_{XY_1}) - \xi_n)} \tag{F.7}$$

$$= \max_{\substack{\tilde{P}: \\ |\tilde{P}_{UX} - P_{UX}| < \mu/8 \\ |\tilde{P}_{UY_1} - P_{UY_1}| < \mu/4}} 2^{-n(D(\tilde{P}_{UXY_1} \| P_{U|X} \tilde{P}_{XY_1}) - \xi_n)}, \tag{F.8}$$

where  $x_n$  is a sequence that tends to 0 as  $n \rightarrow \infty$ . Inequality (a) holds by Sanov's theorem and by the way the source sequences and the codewords are generated and Inequality (b) holds by the choice of  $R_1$  in (5.18).

To analyze the probability of Type II error at Detector 2, we notice that  $\hat{\mathcal{H}}_2 = 0$  only if there exists a pair of indices  $(m_1, m_2) \in \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}$  so that

$$\begin{aligned}
(U^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(P_{UX}) \quad \text{and} \quad (U^n(m_1), V^n(m_2|m_1), Y_1^n) \in \mathcal{T}_{\mu/2}^n(P_{UVY_1}) \\
\text{and} \quad (U^n(m_1), V^n(m_2|m_1), Y_2^n) \in \mathcal{T}_{\mu}^n(P_{UVY_2})
\end{aligned} \tag{F.9}$$

Therefore, applying similar steps as before:

$$\mathbb{E}_C[\beta_{2,n}] \Pr[\hat{\mathcal{H}}_2 = 0 | \mathcal{H} = 1]$$

---


$$\leq \sum_{m_1=1}^{2^{nR_1}} \sum_{m_2=1}^{2^{nR_2}} \Pr \left[ (U^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(P_{UX}), (U^n(m_1), V^n(m_2|m_1), Y_1^n) \in \mathcal{T}_{\mu/2}^n(P_{UVY_1}), \right. \\ \left. (U^n(m_1), V^n(m_2|m_1), Y_2^n) \in \mathcal{T}_{\mu/2}^n(P_{UVY_2}) \mid \mathcal{H} = 1 \right] \quad (\text{F.10})$$

$$< 2^{n(I(U;X)+I(V;Y_1|U))} \cdot \max_{\substack{\tilde{P}: \\ |\tilde{P}_{UX}-P_{UX}|<\mu/8 \\ |\tilde{P}_{UVY_1}-P_{UVY_1}|<\mu/2 \\ |\tilde{P}_{UVY_2}-P_{UVY_2}|<\mu}} 2^{-n(D(\tilde{P}_{UVXY_1Y_2}\|P_U P_{V|U} \tilde{P}_{XY_2})-\xi'_n)} \quad (\text{F.11})$$

$$= \max_{\substack{\tilde{P}: \\ |\tilde{P}_{UX}-P_{UX}|<\mu/8 \\ |\tilde{P}_{UVY_1}-P_{UVY_1}|<\mu/2 \\ |\tilde{P}_{UVY_2}-P_{UVY_2}|<\mu/4}} 2^{-n(D(\tilde{P}_{UVXY_1Y_2}\|P_{U|X} P_{V|UY_1} \tilde{P}_{XY_1Y_2})-\xi'_n)} \quad (\text{F.12})$$

where  $\xi'_n$  is a sequence that tends to 0 as  $n \rightarrow \infty$ . The proof is then concluded by letting  $n \rightarrow \infty$  and by noting that there must exist at least one pair of codebooks achieving the same exponents as the random ensemble.



---

Proof of Theorem 5.7

---

We analyze the scheme in Subsection 5.2.1. The type-II error probability at Detector 2 can be analyzed as in the preceding Appendix F.

To analyze the probability of type-II error at Detector 1, we notice that  $\hat{\mathcal{H}}_1 = 0$  only if there exists an index  $m_1 \in \{1, \dots, 2^{nR_1}\}$  such that

$$(U_1^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(\bar{P}_{U_1X}) \quad \text{and} \quad (U_1^n(m_1), Y_1^n) \in \mathcal{T}_{\mu/4}^n(\bar{P}_{U_1Y_1}) \quad (\text{G.1})$$

By now standard arguments:

$$\Pr[\hat{\mathcal{H}}_1 = 0 | \mathcal{H} = 1] \leq \sum_{m_1=1}^{2^{nR_1}} \Pr\left[(U_1^n(m_1), X^n) \in \mathcal{T}_{\mu/8}^n(\bar{P}_{U_1X}), (U_1^n(m_1), Y_1^n) \in \mathcal{T}_{\mu/4}^n(\bar{P}_{U_1Y_1}) \mid \mathcal{H} = 1\right] \quad (\text{G.2})$$

$$< \max_{\substack{\bar{P}: \\ |\bar{P}_{U_1X} - \bar{P}_{U_1X}| < \mu/8 \\ |\bar{P}_{U_1Y_1} - \bar{P}_{U_1Y_1}| < \mu/4}} 2^{-n(D(\bar{P}_{U_1X Y_1} \| P_{U_1|X} P_{X Y_1}) - \xi_n'')}, \quad (\text{G.3})$$

where the sequence  $\xi_n'' \rightarrow 0$  as  $n \rightarrow \infty$ .





---

Proof of Equation (6.36)

---

Given  $R \geq 0$ , if  $\theta$  is achievable in  $P_2$ , for every block-length  $n \geq 0$  there exist a pair of functions  $(\phi_{2,n}, \psi_{2,n})$  such that

$$\phi_{2,n}: \mathcal{X}^n \rightarrow \mathcal{M}_2 \triangleq \{1, \dots, W_{2,n}\}, \quad (\text{H.1})$$

$$\psi_{2,n}: \mathcal{M}_2 \times \mathcal{U}^n \times \mathcal{V}^n \rightarrow \{0, 1\}. \quad (\text{H.2})$$

and

$$\begin{aligned} & \overline{\lim}_{n \rightarrow \infty} \Pr\{\psi_{2,n}(\phi_{2,n}(\mathbf{X}^n), \mathbf{U}^n, \mathbf{V}^n) = 1 | \mathcal{H} = 0\} = 0, \\ \theta \leq & \underline{\lim}_{n \rightarrow \infty} -\frac{1}{n} \log \Pr\{\psi_{2,n}(\phi_{2,n}(\mathbf{X}^n), \mathbf{U}^n, \mathbf{V}^n) = 0 | \mathcal{H} = 1\}, \end{aligned} \quad (\text{H.3a})$$

and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log_2 W_{2,n} \leq R. \quad (\text{H.3b})$$

We can then define  $(\phi_{1,n}, \psi_{1,n})$ :

$$\phi_{1,n} = \phi_{2,n}, \quad (\text{H.4})$$

and for every  $(m, \mathbf{y}^n) \in \{1, \dots, \lfloor 2^{nR} \rfloor\} \times \mathbb{R}^{nq}$ :

$$\psi_{1,n}(m, \mathbf{y}^n) = \begin{cases} 0 & \text{if } \psi_{2,n}(m, \mathbb{E}_{\bar{p}}[\mathbf{x}^n | \mathbf{y}^n], \mathbf{y}^n) = 0 \\ 1 & \text{if } \psi_{2,n}(m, \mathbb{E}_{\bar{p}}[\mathbf{x}^n | \mathbf{y}^n], \mathbf{y}^n) = 1 \end{cases}, \quad (\text{H.5})$$

where  $\mathbb{E}_{\bar{p}}[\mathbf{y}^n | \mathbf{Y}^n] = \{\mathbb{E}_{\bar{p}}[\mathbf{x}_1 | \mathbf{y}_1], \dots, \mathbb{E}_{\bar{p}}[\mathbf{x}_n | \mathbf{y}_n]\}$ . The functions pair  $(\phi_{1,n}, \psi_{1,n})$  is a valid choice for  $P_1$ , and one can write that:

$$\alpha_{1,n} = \Pr\left(\psi_{1,n}(\phi_{1,n}(\mathbf{X}^n), \mathbf{Y}^n) = 1 | \mathcal{H}_2 = 0\right)$$

---


$$\begin{aligned}
&= \Pr\left(\psi_{2,n}\left(\phi_{2,n}(\mathbf{X}^n), \mathbb{E}_{\bar{P}}[\mathbf{X}^n|\mathbf{Y}^n], \mathbf{Y}^n\right) = 1 | \mathcal{H}_2 = 0\right) \\
&= \Pr\left(\psi_{2,n}\left(\phi_{2,n}(\mathbf{X}^n), \mathbf{U}^n, \mathbf{V}^n\right) = 1 | \mathcal{H}_1 = 0\right) \\
&= \alpha_{2,n},
\end{aligned}$$

and along similar lines one can prove that

$$\beta_{1,n} = \beta_{2,n}. \tag{H.6}$$

So conditions (6.11) are respected and exponent  $\theta$  is achievable in  $P_1$  so

$$E_1(R) \geq E_2(R). \tag{H.7}$$

Furthermore along similar lines one can prove that if  $\theta$  is achievable in  $P_1$  it is achievable in  $P_2$  and so

$$E_1(R) = E_2(R). \tag{H.8}$$

This concludes the proof of Equation (6.36). ■

---

Proof of Equivalence between (6.41) and constraint C

---

The Gaussian random vectors  $\mathbf{X}$  and  $\mathbf{U}$  always admit joint probability density functions,  $p_{\mathbf{XU}}$  and  $\bar{p}_{\mathbf{XU}}$  under  $\mathcal{H} = 0$  and  $\mathcal{H} = 1$  defined on a same vector subspace  $E \subseteq \mathbb{R}^{2m}$ , because their covariance matrices have same kernel. As the minimizer in (6.41) is a Gaussian distribution with same marginal in  $\mathbf{X}$  and  $\mathbf{U}$  than  $p_{\mathbf{XU}}$ , and that for any  $\tilde{p}_{\mathbf{XU}}$  if  $\tilde{p}_{\mathbf{XU}} \ll \bar{p}_{\mathbf{XU}}$  do not hold then  $D(\tilde{p}_{\mathbf{XU}}\|\bar{p}_{\mathbf{XU}}) = \infty$ , we can restrict the minimizer in (6.41) to be a normal distribution defined on  $E$  :

$$\tilde{p}_{\mathbf{XU}}(\mathbf{z}) = |2\pi\mathfrak{E}|_+^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{z}^\dagger\mathfrak{E}^+\mathbf{z}\right\}. \quad (\text{I.1})$$

Here  $\mathfrak{E}$  denote any semi-definite positive matrix of the form

$$\mathfrak{E} = \begin{bmatrix} \mathbf{K}_X & \mathbf{G} \\ \mathbf{G}^\dagger & \bar{\mathbf{K}}_{XY}\bar{\mathbf{K}}_Y^{-1}\mathbf{K}_Y\bar{\mathbf{K}}_Y^{-1}\bar{\mathbf{K}}_{XY}^\dagger \end{bmatrix}, \quad (\text{I.2})$$

where the off-diagonal block  $\mathbf{G}$  is arbitrary. We have

$$\begin{aligned} D(\tilde{p}_{\mathbf{XU}}\|\bar{p}_{\mathbf{XU}}) &= -h(\tilde{p}_{\mathbf{XU}}) - \mathbb{E}_{\tilde{p}_{\mathbf{XU}}} \{\log \bar{p}_{\mathbf{XU}}\} \\ &= -\frac{1}{2} \log |\mathfrak{E}|_+ + \frac{1}{2} \log \left| \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY}\bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \bar{\mathbf{K}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY}\bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \right|_+ + \frac{p}{2} \\ &\quad + \frac{1}{2} \text{Tr} \left\{ \left( \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY}\bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \bar{\mathbf{K}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{K}}_{XY}\bar{\mathbf{K}}_Y^{-1} \end{bmatrix} \right)^\dagger \mathfrak{E} \right\}. \end{aligned} \quad (\text{I.3})$$

and finally, since  $\frac{p}{2}$  is a constant term, (6.41) is equivalent to constraint C.



---

## Bibliography

---

- [1] R. Ahlswede and I. Csiszar. “Hypothesis testing with communication constraints”. *IEEE Transactions on Information Theory* 32.4 (July 1986), pp. 533–542. ISSN: 0018-9448. DOI: 10.1109/TIT.1986.1057194.
- [2] S.-I. Amari and T. S. Han. “Statistical inference under multiterminal rate restrictions: a differential geometric approach”. *IEEE Transactions on Information Theory* 35.2 (1989), pp. 217–227.
- [3] T. Berger. *Multiterminal source coding*. 1977.
- [4] R. Blahut. “Hypothesis testing and information theory”. *IEEE Transactions on Information Theory* 20.4 (July 1974), pp. 405–417. ISSN: 1557-9654. DOI: 10.1109/TIT.1974.1055254.
- [5] T. Cover and J. Thomas. *Elements of information theory 2nd edition*. Wiley-Interscience, July 2006. ISBN: 978-0-471-24195-9.
- [6] I. Csiszar. “The method of types [information theory]”. *IEEE Transactions on Information Theory* 44.6 (Oct. 1998), pp. 2505–2523. ISSN: 0018-9448. DOI: 10.1109/18.720546.
- [7] I. Csiszár and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [8] A. El Gamal and Y.-H. Kim. *Network information theory*. Cambridge University Press, 2011.
- [9] P. Escamilla, M. Wigger, and A. Zaidi. “Distributed hypothesis testing with concurrent detections”. *2018 IEEE International Symposium on Information Theory (ISIT’18)*. June 2018, pp. 166–170. DOI: 10.1109/ISIT.2018.8437906.
- [10] P. Escamilla, A. Zaidi, and M. Wigger. “Distributed hypothesis testing with collaborative detection”. *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Oct. 2018, pp. 512–518. DOI: 10.1109/ALLERTON.2018.8635828.

- 
- [11] P. Escamilla, M. Wigger, and A. Zaidi. “Distributed hypothesis testing: cooperation and concurrent detection”. *revision for publication in the IEEE Transactions of Information Theory* (2019).
- [12] P. Escamilla, A. Zaidi, and M. Wigger. *On distributed hypothesis testing with collaborative detection: implementation*. 2019. URL:  
<https://perso.telecom-paristech.fr/pescamilla/Implementation.ipynb>.
- [13] A. Gilani, S. Belhadj Amor, S. Salehkalaibar, and V. Y. F. Tan. “Distributed hypothesis testing with privacy constraints”. *Entropy* 21.5 (2019). ISSN: 1099-4300. DOI: 10.3390/e21050478. URL:  
<https://www.mdpi.com/1099-4300/21/5/478>.
- [14] T. S. Han. “Hypothesis testing with multiterminal data compression”. *IEEE Transactions on Information Theory* 33.6 (Nov. 1987), pp. 759–772. ISSN: 0018-9448. DOI:  
10.1109/TIT.1987.1057383.
- [15] T. S. Han and K. Kobayashi. “Exponential-type error probabilities for multiterminal hypothesis testing”. *IEEE Transactions on Information Theory* 35.1 (Jan. 1989), pp. 2–14. ISSN: 0018-9448. DOI:  
10.1109/18.42171.
- [16] C. Heegard and T. Berger. “Rate distortion when side information may be absent”. *IEEE Transactions on Information Theory* 31.6 (Nov. 1985), pp. 727–734. ISSN: 0018-9448. DOI:  
10.1109/TIT.1985.1057103.
- [17] W. Hoeffding. “Asymptotically optimal tests for multinomial distributions”. *The Annals of Mathematical Statistics* (1965), pp. 369–401.
- [18] A. H. Kaspi. “Rate-distortion function when side-information may be present at the decoder”. *IEEE Transactions on Information Theory* 40.6 (Nov. 1994), pp. 2031–2034. ISSN: 0018-9448. DOI:  
10.1109/18.340475.
- [19] G. Katz, P. Piantanida, and M. Debbah. “Collaborative distributed hypothesis testing with general hypotheses”. *2016 IEEE International Symposium on Information Theory (ISIT’16)*. July 2016, pp. 1705–1709. DOI: 10.1109/ISIT.2016.7541590.
- [20] B. G. Kelly and A. B. Wagner. “Reliability in source coding with side information”. *IEEE Transactions on Information Theory* 58.8 (Aug. 2012), pp. 5086–5111. ISSN: 0018-9448. DOI:  
10.1109/TIT.2012.2201346.
- [21] G. Kramer. “Topics in multi-user information theory”. *Foundations and Trends in Communications and Information Theory* 4.4–5 (2008), pp. 265–444. ISSN: 1567-2190. DOI: 10.1561/0100000028. URL:  
<http://dx.doi.org/10.1561/0100000028>.

- [22] J. Liao, L. Sankar, V. Y. F. Tan, and F. du Pin Calmon. "Hypothesis testing under mutual information privacy constraints in the high privacy regime". *IEEE Transactions on Information Forensics and Security* 13.4 (Apr. 2018), pp. 1058–1071. ISSN: 1556-6013. DOI: 10.1109/TIFS.2017.2779108.
- [23] J. Liao, L. Sankar, F. P. Calmon, and V. Y. Tan. "Hypothesis testing under maximal leakage privacy constraints". *2017 IEEE International Symposium on Information Theory (ISIT'17)*. IEEE. 2017, pp. 779–783.
- [24] M. L. Ngan. *Face Recognition Vendor Test (FRVT) Ongoing*. 2016.
- [25] M. S. Rahman and A. B. Wagner. "On the optimality of binning for distributed hypothesis testing". *IEEE Transactions on Information Theory* 58.10 (Oct. 2012), pp. 6282–6303. ISSN: 0018-9448. DOI: 10.1109/TIT.2012.2206793.
- [26] S. Salehkalaibar and M. Wigger. "Distributed hypothesis testing over multi-access channels". *2018 IEEE Global Communications Conference (GLOBECOM)*. Dec. 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647744.
- [27] S. Salehkalaibar and M. Wigger. "Distributed hypothesis testing over multi-access channels". *2018 Information Theory and Applications Workshop (ITA)*. Feb. 2018, pp. 1–5. DOI: 10.1109/ITA.2018.8503129.
- [28] S. Salehkalaibar, M. Wigger, and R. Timo. "On hypothesis testing against conditional independence with multiple decision centers". *IEEE Transactions on Communications* 66.6 (June 2018), pp. 2409–2420. ISSN: 0090-6778. DOI: 10.1109/TCOMM.2018.2798659.
- [29] S. Salehkalaibar, M. Wigger, and L. Wang. "Hypothesis testing in multi-hop networks". *arXiv:1708.05198* (2017).
- [30] S. Salehkalaibar, M. A. Wigger, and R. Timo. "On hypothesis testing against independence with multiple decision centers". *arXiv:1708.03941* (2017).
- [31] H. M. H. Shalaby and A. Papamarcou. "Multiterminal detection with zero-rate data compression". *IEEE Transactions on Information Theory* 38.2 (Mar. 1992), pp. 254–267. ISSN: 0018-9448. DOI: 10.1109/18.119685.
- [32] H. Shimokawa, T. S. Han, and S. Amari. "Error bound of hypothesis testing with data compression". *1994 IEEE International Symposium on Information Theory (ISIT'94)*. June 1994, p. 114. DOI: 10.1109/ISIT.1994.394874.
- [33] S. Sreekumar and D. Gündüz. "Distributed hypothesis testing over noisy channels". *2017 IEEE International Symposium on Information Theory (ISIT'17)*. June 2017, pp. 983–987. DOI: 10.1109/ISIT.2017.8006675.



- [34] S. Sreekumar, D. Gündüz, and A. Cohen. “Distributed hypothesis testing under privacy constraints”. *2018 IEEE Information Theory Workshop (ITW)*. Nov. 2018, pp. 1–5. DOI: 10.1109/ITW.2018.8613433.
- [35] V. Strassen. “Asymptotic estimates in Shannon’s information theory”. *Transactions of the Third Prague Conference on Information Theory*. 2009, pp. 689–723.
- [36] Te Sun Han and S. Amari. “Statistical inference under multiterminal data compression”. *IEEE Transactions on Information Theory* 44.6 (Oct. 1998), pp. 2300–2324. ISSN: 0018-9448. DOI: 10.1109/18.720540.
- [37] C. Tian and J. Chen. “Successive refinement for hypothesis testing and lossless one-helper problem”. *IEEE Transactions on Information Theory* 54.10 (Oct. 2008), pp. 4666–4681. ISSN: 0018-9448. DOI: 10.1109/TIT.2008.928951.
- [38] S.-Y. Tung. “Multiterminal source coding”. PhD thesis. Cornell University, Ithaca, NY, 1978.
- [39] Y. Ugur, I. E. Aguerri, and A. Zaidi. “Vector gaussian CEO problem under logarithmic loss and applications”. *CoRR abs/1811.03933* (2018). arXiv: 1811.03933. URL: <http://arxiv.org/abs/1811.03933>.
- [40] S. Watanabe. “Neyman-Pearson Test for zero-rate multiterminal hypothesis testing”. *IEEE Transactions on Information Theory* 64.7 (July 2018), pp. 4923–4939. ISSN: 0018-9448. DOI: 10.1109/TIT.2017.2778252.
- [41] M. Wigger and R. Timo. “Testing against independence with multiple decision centers”. *2016 International Conference on Signal Processing and Communications (SPCOM)*. Bangalore, India, June 2016, pp. 1–5. DOI: 10.1109/SPCOM.2016.7746613.
- [42] A. Wyner and J. Ziv. “The rate-distortion function for source coding with side information at the decoder”. *IEEE Transactions on Information Theory* 22.1 (1976), pp. 1–10.
- [43] Y. Xiang and Y. H. Kim. “Interactive hypothesis testing with communication constraints”. *Proc. of Allerton Conference on Comm., Control, and Comp.* Monticello (IL), USA, Oct. 2012, pp. 1065–1072. DOI: 10.1109/Allerton.2012.6483336.
- [44] A. Zaidi and I. E. Aguerri. “Optimal rate-exponent region for a class of hypothesis testing against conditional independence problems”. *2019 IEEE Information Theory Workshop (ITW)*. Aug. 2019, pp. 1–5.
- [45] W. Zhao and L. Lai. “Distributed testing against independence with multiple terminals”. *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Sept. 2014, pp. 1246–1251. DOI: 10.1109/ALLERTON.2014.7028598.

- [46] W. Zhao and L. Lai. “Distributed testing against independence with multiple terminals”. *Proc. of Allerton Conference on Comm., Control, and Comp.* Monticello (IL), USA, Oct. 2014, pp. 1246–1251.
- [47] W. Zhao and L. Lai. “Distributed testing With cascaded encoders”. *IEEE Transactions on Information Theory* 64.11 (Nov. 2018), pp. 7339–7348. ISSN: 0018-9448. DOI: 10.1109/TIT.2018.2833626.
- [48] W. Zhao and L. Lai. “Distributed testing with zero-rate compression”. *2015 IEEE International Symposium on Information Theory (ISIT’15)*. Hong Kong, June 2015, pp. 2792–2796.

**Titre :** Tests d'hypothèses distribués: coopération et détection concurrente

**Mots clés :** statistiques, théorie de l'information, test d'hypothèses

**Résumé :** L'inférence statistique prend une place prépondérante dans le développement des nouvelles technologies et inspire un grand nombre d'algorithmes dédiés à des tâches de détection, d'identification et d'estimation. Cependant il n'existe pas de garantie théorique pour les performances de ces algorithmes. Dans cette thèse, nous considérons un réseau simplifié de capteurs communicant sous contraintes pour tenter de comprendre comment des détecteurs peuvent se partager au mieux les informations à leur disposition pour détecter un même événement ou des événements distincts. Nous investiguons différents aspects de la coopération entre détecteurs et comment des besoins contradictoires peuvent être satisfaits au mieux dans le cas de tâches de détection. Plus spécifiquement nous étudions un problème de test d'hypothèse où chaque détecteur doit maximiser l'exposant de décroissance de l'erreur de Type II sous une contrainte d'erreur de Type I donnée. Comme il y a plusieurs détecteurs intéressés par des informations distinctes, un compromis entre les vitesses de décroissance atteignables va apparaître. Notre but est de caractériser la région des

compromis possibles entre exposants d'erreurs de Type II. Dans le cadre des réseaux de capteurs massifs, la quantité d'information est souvent soumise à des limitations pour des raisons de consommation d'énergie et de risques de saturation du réseau. Nous étudions donc, en particulier, le cas du régime de communication à taux de compression nul (i.e. le nombre de bits des messages croît de façon sous-linéaire avec le nombre d'observations). Dans ce cas, nous caractérisons complètement la région des exposants d'erreurs de Type II dans les configurations où les détecteurs peuvent avoir des buts différents. Nous étudions aussi le cas d'un réseau avec des taux de compressions positifs (i.e. le nombre de bits des messages augmente de façon linéaire avec le nombre d'observations). Dans ce cas, nous présentons des sous-parties de la région des exposants d'erreur de Type II. Enfin, nous proposons dans le cas d'un problème point à point avec un taux de compression positif une caractérisation complète de l'exposant de l'erreur de Type II optimal pour une famille de tests gaussiens.

**Title :** On cooperative and concurrent detection for distributed hypothesis testing

**Keywords :** statistics, information theory, hypothesis testing

**Abstract :** Statistical inference plays a major role in the development of new technologies and inspires a large number of algorithms dedicated to detection, identification and estimation tasks. However, there is no theoretical guarantee for the performance of these algorithms. In this thesis we try to understand how sensors can best share their information in a network with communication constraints to detect the same or distinct events. We investigate different aspects of detector cooperation and how conflicting needs can best be met in the case of detection tasks. More specifically we study a hypothesis testing problem where each detector must maximize the decay exponent of the Type II error under a given Type I error constraint. As the detectors are interested in different information, a compromise between the achievable decay exponents of the Type II error appears. Our goal is to characterize the region of possible trade-offs bet-

ween Type II error decay exponents. In massive sensor networks, the amount of information is often limited due to energy consumption and network saturation risks. We are therefore studying the case of the zero rate compression communication regime (i.e. the messages size increases sub-linearly with the number of observations). In this case we fully characterize the region of Type II error decay exponent. In configurations where the detectors have or do not have the same purposes. We also study the case of a network with positive compression rates (i.e. the messages size increases linearly with the number of observations). In this case we present subparts of the region of Type II error decay exponent. Finally, in the case of a single sensor single detector scenario with a positive compression rate, we propose a complete characterization of the optimal Type II error decay exponent for a family of Gaussian hypothesis testing problems.

